Does ESG Risk Exposure Affect Downside Risk Pricing of US Sector Specific Stocks? A Tail Dependency Analysis.

A Master Thesis submitted
for graduation of Econometrics and Management Science
Quantitative Finance track
January 14, 2023

for
KPMG Netherlands Financial Risk Management

by

Flip Jansen (455818)

University supervisor: **Prof. C. Zhou**
Company Supervisor: **B. Arendshorst**
Second assessor: **Prof. A. Tetereva**

Erasmus University Rotterdam
Erasmus School of Economics

Abstract

Sustainability - being able to perform a process indefinitely by taking into account the environment in which this process takes place - has, through increased societal awareness of our environment and the social pressures that followed, become an important characteristic of investments for constitutional and individual investors alike. To be able to transparently and reliably evaluate the sustainability of a business, the need for labelling and scoring arises. ESG ratings are a potential candidate to full fill this need. Understanding the influence that the ESG rating (see ESG risk exposure) has on the risk profile of investments is therefore important for sound investing. I evaluated the influence of ESG ratings on the systematic risk behaviour of US stocks by obtaining the tail dependency of high and low ESG-rated sector-specific portfolios with respective sector benchmarks. I used a bootstrap to obtain a confidence interval of the tail dependency difference between the ESG-sorted portfolios. I find a significant robust non-zero tail dependency difference for most sectors. The found values indicate that the markets price a lower ESG risk exposure of companies (high ESG rating) as being systemic downside risk averting. Moreover, most sectors are also found to have reduced systemic upside exposure stemming from higher ESG ratings. An interesting exception is the consumer staples sector, which only has a significant ESG risk influence on the loss part of the return distribution - upward shocks are unaffected. Uniquely, the materials sector companies are found to have *increased* exposure to systematic shocks related to higher ESG ratings.

Contents

1 Introduction

What is the core proposition of sustainability? One can argue that it is to complete a process in a matter that allows repetition of said process indefinitely (Brown, Hanson, Liverman, & Merideth, 1987). Consequently, this then involves considering a process's effect on its environment. Within business, this is akin to holding a more stakeholder-centric business model. That entails considering the influence of business practices on other factors besides shareholder value, e.g., the environment. In this instance, we take the environment as constituting anything or anyone that might affect the feasibility of business processes.

To address many of the significant problems facing current-day society, both institutional and individual investors feel the pressure of society to try and incorporate sustainability into their investments - through social and regulatory pressure (European Commission, 2018). This pressure is reflected in the drive of investors to integrate sustainability into their investment strategy (Henriksson, Livnat, Pfeifer, & Stumpp, 2019). For example, BlackRock, *the* largest institutional investor, reported that in 2021 Q4, a record one-third of long-term net capital flow was going into sustainable funds - for a sum of 32 billion USD (Brush, 2021). A way to incorporate these sustainable ambitions in a structured and transparent matter is to use a form of independent labelling of investments and companies. An example of a quickly expanding financial sustainability labelling instrument is the so-called Environment, Social, and Governance category, or ESG. In general, ESG scores are meant to indicate a level of ESG performance. With (low) high scores indicating (unsustainable) sustainable ESG behaviour and a (high) low exposure to ESG risks, such as physical and transition risks stemming from climate change. ESG asset market cap is projected at 53 trillion USD in 2025 and has had an average 30% year-over-year growth from 2016 to 2021 (Bloomberg, 2021). Therefore, from an asset management point of view, ESG can be considered a solidified asset class worthy of further investigation.

ESG labelling offers a way of potentially quantifying the extra-financial performance of companies or projects by evaluating their influence on the environment, social paradigm, and society at large. Moreover, ESG labels also seek to enforce this behaviour by evaluating company policies regarding the environment, social participation, and responsible governance structure.

Like De Jonghe (2010), I argue that a company's propensity to experience joint extreme adverse effects with the market reflects its systematic risk exposure. Sustainable practices can be argued to have a downward shock negating nature. An example of this behaviour is a more environmentally friendly company that might be less exposed to physical and transition risks affiliated with climate change (Setzer & Higham, 2021; Engle, Giglio, Kelly, Lee, & Stroebel, 2020). Or, as the OECD and American Bar Association state, a company with a responsible governance system can be argued to have less exposure to over-leveraging and litigation stemming from bad business practices (Organisation for Economic Co-operation and Development, 2014; American Bar Association, 2016). Therefore, I argue that companies' exposure to these environmental or social shocks to their business model can be quantified by their systematic risk. Given these proposed relations, I argue that ESG practices reduce the systematic risk exposure of assets by being sustainable and thus less susceptible to downward shocks in the respective three ESG components.

Consequently, selecting high ESG scores could be advantageous for risk management purposes. Additionally, selecting for high ESG scores offers a structured method of obtaining a more stakeholder-centric business portfolio. Therefore, if high ESG scoring assets (low ESG risk exposure) can be assumed to have reduced systematic risk, which most likely also involves a reduction in upside potential, that would make them a potential hedging product against the systematic risk of portfolios (E. Berg & Lange, 2020; Bax, Sahin, Czado, & Paterlini, 2021).

To evaluate the systematic risk, I use a widely accepted substitute measure called the tail dependency (Shirvani & Volchenkov, 2019; Embrechts, Lindskog, & McNeil, 2001). The tail dependency measures the codependency of two-time series during extreme price movements. Or, otherwise stated, as being the conditional probability of when one time series crosses a threshold that the other does as well. Therefore, we can take the tail dependency as a measure that can capture the systematic risk exposure of a company with the broader market. We can obtain the tail dependency by fitting a function that can measure the dependency between two time series.

Evaluating the tail dependency of both low- and high-rated ESG stocks with sector-specific indices can give insights into what extent or if high-scoring ESG stocks are priced as systematic risk averting. Doing this on a sector-by-sector basis can potentially uncover if this behaviour is sector specific. Following the proposed ESG risk dynamics, I expect a lower tail dependence with the broader market for highly-rated ESG instruments than lower-rated ESG instruments. A lower tail dependence can be argued to be associated with lower systematic risk. Also, It can be seen as a robust argumentation in favour of taking into account ESG ratings for risk management purposes. Therefore, evaluating ESG-related tail dependence could offer relevant information of how the market prices the societal and financial benefits of ESG ratings.

Several studies have focused on evaluating the risk structure of ESG instruments (Li, Wang, Sueyoshi, & Wang, 2021; Sassen, Hinze, & Hardeck, 2016). Moreover, some research has been focused explicitly on ESG-related tail behaviour (Górka & Kuziak, 2022; Bax et al., 2021). Nevertheless, the influence of ESG scores on the tail dependencies of sector-specific stocks with comparable sector benchmarks seems to be under-explored. The novel contributions of this thesis are; a sector-specific view of the influence of ESG scores on the tail risk behaviour of returns; and the formation of ESG portfolios from a wide scope of stocks (around eighty per cent of all public US companies are present in the analysed dataset), this is novel in this type of research. The widened scope is enabled by the expansive RepRisk ESG database. A more granulated search has the potential to uncover previously out-of-view tail risk characteristics that would have been hidden through aggregating the different sectors in singular ESG and market benchmarks. I assume that the tail dependency and ESG rating reflect the systematic risk and ESG risk exposure, respectively. Therefore I formulate the research question as the following

Do markets price systematic risk reduction stemming from reduced ESG risk exposure?

We can answer this question by obtaining the tail dependencies between highly rated ESG stocks and a benchmark index and observing if these differ from those of low-rated ESG stocks with the same benchmark.

To answer the research question, I use copula functions to capture the dependence structure between returns of aggregated ESG-sorted stocks and a benchmark index. From a copula, one can obtain an implied tail dependency. I form equal-weighted portfolios by aggregating returns of the highest- and lowest-ESG-scoring quartile of stocks to get a clearer view of ESG-score influence by smoothing out potential outliers and allowing more straightforward inference. The portfolio returns are filtered to obtain i.i.d residuals, which are then transformed to a $U \sim Uniform(0, 1)$ to allow fitting of the copula functions. An information criterion is used to decide which copula fit is preferred. Due to the availability of both ESG ratings and extensive stock returns in the US I will be reducing my scope to US markets.

I find that a significant robust non-zero ESG rating influence on the tail dependency is present. More specifically, a high ESG rating appears to cause a lower exposure to sector-wide shocks in stock return. An exception is the materials sector, where the opposite effect is present.

The remainder of this paper is structured as follows: Section 2 summarizes the literature and argues which method is most suited for answering the research question. Section 3 introduces the precise methodology. Section 4 describes the data used and the pre-processing. Then, Section 5 reports the empirical results with corresponding inference comments. Lastly, Section 6 concludes and provides an outlook for future research.

2 Literature

In this section, I provide an overview of relevant publications. I start with a broad scope - covering applications of different dependency measures within financial time series analysis. Then, we narrow our focus to using copula functions to this end and lay out the development of ESG (literature). Lastly, I close in on applying copula functions to study the risk characteristics of stocks with varying ESG risk exposure. More specifically, I will review papers analysing the co-dependency of these stocks with industry benchmarks during large shocks. This co-dependency is analysed by obtaining the corresponding tail dependency.

2.1 Dependency Measures

The study of the dependency between financial time series is relevant for risk management since it can offer a more comprehensive understanding of the movement of whole portfolios during different scenarios. An elementary introduction to these dependency measures (e.g. the relevance of Kendall's τ) can be found in Appendix A. To model the dependence between time series is to describe their co-movement, which involves capturing these dynamics using a model or measure. A first approach is to use multivariate GARCH models (Szetela, Mentel, & Gedek, 2016). However, this approach suffers from the curse of dimensionality and is challenging to apply for varying individual marginal distributions. Another method is to build more complex models that try to extract and reformulate information from linear correlations, e.g. looking at the Entropy information contained in correlations and extra-financial information (Sandoval Junior, Mullokandov, & Kenett, 2015).

Another example can be found in the research of He and Chen (2011). Who formulate that the study of the non-linear dependency between financial time series characteristics (volume and price) is of importance since it can offer an improved theoretical understanding of market dynamics and practical knowledge of market trading characteristics. They propose and apply a Detrended Cross-Correlation Analysis (DCCA), which uses features related to fractal geometry (generalized Hurst exponent) to estimate the non-linear dependency. Similarly, Morales, Di Matteo, and Aste (2014) also analyses the dependence of financial time series, those being daily stock returns, using cross-correlation analysis and the generalized Hurst exponent, offering insights into the dependency structures of these time series. However, DCCA still uses the linear and highly restricted Pearson's correlation. Therefore, reducing the flexibility and robustness of this method. Moreover, using only linear dependence is too restrictive for many fields, e.g. for insurance and finance (Embrechts, Lindskog, & McNeil, 2003; McNeil, Frey, & Embrechts, 2005).

The discussed papers in this Section share the same goal as my proposed method: to justify an economic theory by analyzing the dependency between relevant variables. The method I propose analyses the market-wide dependence of stocks stemming from the ESG risk exposure of a company. The underlying proposed economic theory is that a lower ESG risk exposure, as expressed by more sustainable business practices, protects companies from systematic shocks to their business model. The proposed effect is that a lower ESG risk exposure (high ESG ratings) would have a different dependency with the broader market than companies with higher ESG risk exposure (low ESG ratings).

2.2 Copulas in Financial Time Series Analysis

Another reason for exploring the use of alternative dependence models has also been suggested in light of evidence that asset returns are more highly connected during volatile markets and market downturns (Longin & Solnik, 2001; Ang & Chen, 2002). Notably, because this data feature can not be captured with a linear correlation measure. Consequently, conventional copula models were explored (i.e. Gaussian and Student-t) as possible alternative dependence models (Chen, Fan, & Patton, 2004). With the advancement of dependence research, more advanced copula models (e.g. Archimedean) were found to offer a better fit. This can be observed by the extent to which these copulas have been applied in many financial applications (Hofert & Scherer, 2011; Genest, Gendron, & Bourdeau-Brien, 2013).

Foundational copula research focused on the conditions needed for proper statistical inference (Hu, 1998; Beare, 2010). This foundation enabled extensive research, making inferences on a wide range of models using these more complex copulas. Complex copula research mostly has a focus on finance, more specifically risk management, e.g. in credit risk (Changqing, Yanlin, & Mengzhen, 2015), exchange rate risk (Boero, Silvapulle, & Tursunalieva, 2011), liquidity risk (Karimalis & Nomikos, 2018), or systematic risk analysis (Wen, Wei, & Huang, 2012; Aloui, Gupta, & Miller, 2016; BenSaïda & Litimi, 2021). However, complex copula application papers are also present in other fields, e.g. option pricing (Bajo, Barbi, & Romagnoli, 2015), weather simulations (Laux, Vogl, Qiu, Knoche, & Kunstmann, 2011), or hydraulics (Genest, Favre, et al., 2007).

I refer to Nelsen (2006) for a broad introduction to copulas. Moreover, a comprehensive overview of the copula in econometric models can be found in the work of Trivedi and Zimmer (2007). For further examples of copula application in practice, see Reboredo (2013); Tachibana (2018); Bressan and Romagnoli (2021). Specific examples use Gaussian, Archimedean or Dynamic copulas as done by Christoffersen, Errunza, Jacobs, and Langlois (2012), or in the case of multi-dimensional analysis, Vine-set copulas as done by Bax et al. (2021). The previously mentioned papers show how using copulas to analyze dependencies is a versatile and robust method for obtaining results. The proposed method of this thesis uses copulas to get dependencies that, through sorting of the data, can be attributed to my variable of interest - the ESG risk exposure of companies.

Evaluating the cross-dependencies in the tails of distributions is of importance in risk management, first, because these movements contribute to the risk and unpredictability of portfolios (Embrechts, McNeil, & Straumann, 2002). And second, it has been shown that stocks exhibit increased co-movement during extreme price shocks (Longin & Solnik, 2001). Additionally, copula-implied tail dependencies are found to be a good proxy for systematic risk (Chen, Xiao, & Wang, 2020). Moreover, De Luca and Zuccolotto (2011) find that this systematic risk estimation using implied tail dependencies can be valuable in portfolio construction. Therefore, the tail dependency can be argued to be an attractive measure to investigate dependencies. An example of its use can be found in the paper of Michelis and Ning (2010) who explores the relationship between the USD/CAD exchange rate and the Canadian stock market.

The papers mentioned in this Section specifically focus on the dependencies in the tails of distributions. They show that the tails of a distribution offer distinct characteristics from that of the whole. Distinct tail risk behaviour is especially relevant for this thesis's proposed hypothesis, which regards the extreme ends of the distributions. These are of interest when evaluating systematic shocks - that are at the core of the proposed ESG risk hypothesis. The extensive use of copula functions to analyze complex dependencies has also found its way into the study of ESG market dynamics.

2.3 ESG

Significant growth of ESG investing during the last decade coincided with an increase in the number of studies on sustainable and ESG investing, e.g., looking at the impact of ESG ratings on the environment and the financial performance of the related financial instruments. Initially, ESG research focused on establishing the relevance of ESG ratings - to what extent ESG ratings are correlated with and to what extent they are a predictor of financial characteristics of financial instruments. An example is a study by Fatica and Panzica (2021), who evaluated the effect of green bonds on realized emissions and the extent to which the ratings of these bonds are reliable. Moreover, Lins, Servaes, and Tamayo (2017) find that high social capital firms - measured using corporate social responsibility (CSR) intensity - have been significantly less affected than low CSR firms during the 2008 financial crisis. Additionally, Jin (2018) shows that ESG-related risks are priced in for US equities by performing a Farma-French analysis adding an additional ESG-related factor to their 5-factor model (Fama & French, 2015). Here a non-zero significant beta and corresponding equity risk premium are found for the associated ESG-related factor.

Another example line of research tries to incorporate climate risk into the financial risk analysis (Battiston, Dafermos, & Monasterolo, 2021; Flori, Pammolli, & Spelta, 2021). The Network for Greening the Financial System (NGFS) published climate scenario models (Bertram et al., 2020), which can be used to evaluate financial stability (Allen et al., 2020). Moreover, these models could be incorporated into the larger ESG risk management frameworks. Pedersen, Fitzgibbons, and Pomorski (2021) developed a framework for integrating ESG scores into the portfolio selection process and found that high ESG scores can be incorporated into a selection strategy at low to no additional cost. Similarly, Coqueret et al. (2021) analyzed ESG ratings as factors in asset pricing and found that using ESG information as additional factors to a model using ordinary firm characteristics can improve portfolio performance. However, they find that performing the same factor analysis using only ESG information does not improve performance. Additionally, Martellini and Vallée (2021) researched the influence of specific ESG-constrained portfolio selections concerning the risk and performance of stocks and bonds. They found that although some positive effects are possibly present in the form of reduced borrowing costs for high environmental scores, generally, these strategies lead to a general increase in risk and reduction in return.

Berg, Kölbel, and Rigobon (2019) found that ESG rating providers have diverging ratings, stemming from different views on and measurements of ESG factors, or analyst-specific company bias among other factors. Moreover, van der Beck (2021) evaluated the dynamics driving ESG returns, finding that increased public interest in ESG and the resulting cash flow could cause the varying performance of ESG score sorting portfolio strategies. Both of these findings can be argued to obfuscate the true predictive nature of ESG ratings regarding financial performance. Therefore, it can be argued that the influences of ESG scores on risk dynamics can not yet be fully established, and a clear consensus has not yet been reached. One aspect of ESG ratings that I will be focusing on is the extent to which high ESG scores, which are taken to be a proxy of a low ESG risk exposure, could contribute to lower systematic risk. To this end, the existing literature mentioned in Section 2.2 seems to suggest that a copula implied tail dependency analysis appears to be the best-suited method. Moreover, I argue that the copula-implied tail dependency is an attractive research method for this problem due to the copula model's flexibility and the tail dependency's robustness and interpretability. I will elaborate on both of these arguments in Chapter 3.

2.4 Tail Dependency Through ESG Risk Exposure

The current literature on the risk characteristics of ESG-sorted stocks focuses primarily on the wider part of the distribution and to what extent ESG can be seen as a risk factor (Ashwin Kumar et al., 2016). Ashwin Kumar et al. (2016) find that ESG factors impact the return characteristics on a sector-specific basis. Moreover, a low ESG risk exposure seems to correlate with a lower relative risk.

Research on ESG risk exposure-related tail dependency has been done by comparing single ESG benchmarks with larger market benchmarks such as the S&P 500 (Górka & Kuziak, 2022). Bax et al. (2021) perform similar research with a more granulated view of the ESG classification. Zhang, De Spiegeleer, and Schoutens (2021) also study the relation between ESG ratings and implied tail risk. However, they do this by observing implied skewness and kurtosis. Bax et al. (2021), Zhang et al. (2021), and (Górka & Kuziak, 2022) all find that ESG-specific risk dynamics are not insignificant, especially in the tail regions of the return distribution.

In summary, I find that some research on the influence of ESG risk exposure on stocks has been done. Still, the research present has either not probed the influence of ESG risk directly – instead, using more restricted ESG ETFs and not sorting for ESG scores directly. Or, these studies do not probe possibly distinct risk characteristics of sectors. To address these underexplored aspects, I propose a sector-specific ESG sorted portfolio analysis to obtain the influence of ESG risk exposure on tail dependency with sector-specific benchmarks. A sector-specific view could expose ESG risk dynamics that would have been previously hidden in the aggregation of ESG performance in larger general benchmarks.

3 Methodology

In this section, I will describe the different components of the methodology that are required to answer the proposed hypothesis regarding the potential influence that the ESG risk exposure of US companies has on their tail dependency with sector benchmarks. I will be doing this in chronological order for clarity.

3.1 Filtering

To transform the financial data to (near) i.i.d., I will be making use of an autoregressive moving average (ARMA) - generalised autoregressive conditional heteroskedasticity (GARCH) model. I fit the ARMA-GARCH model to the return series of (ESG) portfolios by using maximum likelihood estimation (MLE). Consequently, I obtain the standardized residuals,$\{z_t\}$. I assume these residuals to be t distributed, with degrees of freedom, v, following $v = n - 1$, where n is the number of observations used in the estimation. A requirement for the i.i.d. assumption requires us to filter out any heteroskedasticity in the data. The GARCH component of our model should be able to address this. The GARCH model is a generalisation of the ARCH model proposed independently by both Taylor (1986) and Bollerslev (1986). It is characterised following the notation from Teräsvirta (2009) where we let ϵ_t be defined as the innovation sequence and y_t as the equal-weighted aggregated log returns of our respective (ESG) portfolios. To filter out any ARMA effects, an ARMA model is defined, following Box and Jenkins (1970), as

$$y_t - \phi_1 y_{t\text{-}1} - ... - \phi_k y_{t\text{-}k} = \epsilon_t - \theta_1 \epsilon_{t\text{-}1} - ... - \theta_m \epsilon_{t\text{-}m}, \tag{3.1}$$

where ϵ_t and y_t are defined as in Equation (3.2). I refer to the fitted model with its corresponding lag parameters (ϕ_n and θ_m) as ARMA(k, m).

It is assumed that ϵ_t can be decomposed as

$$\epsilon_t = s_t \sqrt{h_t}, \tag{3.2}$$

where $\{s_t\}$ is modeled as a sequence of i.i.d. t distributed random variables. The volatility, h_t, is defined as

$$h_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \epsilon_{t\text{ - }j}^2 + \sum_{j=1}^{p} \beta_j h_{t\text{-}j}, \tag{3.3}$$

where α_j and β_j represent the lag parameter coefficients for a GARCH(p, q) model. After fitting this model, the residuals, ϵ_t, are transformed to standardised residuals, $\{z_t\}$, using the assumed distributions of the fitted residuals (Student-t). To formally establish that the fitted GARCH model fully captures the ARCH dynamics, we can then use the augmented ARCH-LM test (as described in Section 3.1.1) to test the null of no ARCH effects.

I will be following a general-to-specific fitting regime; this means that my initial fitted ARMA-GARCH model will have order ((1,1),(1,1)). When the residuals from this fit do not pass the tests in Section 3.1.1, higher lag order models will be tried on a case-by-case basis. When the fitted model completely captures the ARMA dynamics of the data, then the Ljung-Box test (as described in Section 3.1.1) applied to the standardised residuals,$\{z_t\}$, will no longer be rejected. This allows us not to reject the null hypothesis of no serial correlation. Once the data can be assumed to be (near) i.i.d., we can start by obtaining the best fitting copula.

3.1.1 Testing

For robust statistical inference using the "inference function for margins" (IFM) method (Joe, 1997), it is required that the data can be assumed to be i.i.d. (Genest, Ghoudi, & Rivest, 1995). However, since we are working with financial data of a daily frequency, there are some often present stylized facts to take into account[1].

I apply multiple (portmanteau) tests to validate the possible presence of these stylized facts. The first is the Augmented Dickey-Fuller (ADF) test, see Cheung and Lai (1995). When we reject the null hypothesis of this test, we reject the stationarity of the data. If this were to be the case, I would have to change my data sampling frequency or sample size.

To verify the possible existence of serial correlation within the data, the Weighted Ljung-box test is applied, as proposed by T. J. Fisher and Gallagher (2012). This adapted form is preferred over the conventional Ljung-Box test, as done by McNeil et al. (2005, p 134) because it more accurately reflects how the statistics of the values from the estimated models

[1] Financial data often contain the following stylized facts: "(1) Return series are not iid although they show little serial correlation; (2) Series of absolute or squared returns show profound serial correlation ;(3) Conditional expected returns are close to zero; (4) Volatility appears to vary over time; (5) Return series are leptokurtic or heavy-tailed; (6) Extreme returns appear in clusters" (McNeil et al., 2005, p. 117).

are distributed.

Hetereoskadicity is the most often present stylized fact. It is usually tested using an ARCH-LM test, which allows us to reject the null of no hetereoskadicity, see Engle (1982). It does this by analyzing the serial correlation in the squared residual series. In the case of testing, data pre-model fit the squared observation value, y_t^2, is used. However, a more recent improvement on this test is the ARCH-LM test as proposed by T. J. Fisher and Gallagher (2012), which is argued to more accurately reflect how the statistics of the values from the estimated models are distributed. Therefore, I use a weighted Li-Mak test as proposed by T. J. Fisher and Gallagher (2012), which acts as an adapted ARCH-LM test.

The ADF test was rejected for every sector portfolio in our starting data set. I can therefore assume the taken input data to be stationary.

3.2 Copula

The idea of a "copula" or "copula function" originates from Sklar (1959), with Sklar being the first to describe the dependency of variables separately from their univariate marginals in this way. A copula describes the inner dependency between univariate distribution functions to form a multivariate function. I make use of the notation by Patton (2009). Consider a random vector $\mathbf{X} = [X_1, X_2, ..., X_n]'$ with joint distribution $\mathbf{F}$ and marginal distributions $F_1, F_2, ..., F_n$. The copula representation from Sklar (1959) theorem provides the mapping from the individual distribution functions to the joint distribution function as

$$\mathbf{F}(\mathbf{x}) = \mathcal{C}(F_1(x_1), F_2(x_2), ..., F_n(x_n)),\ \forall\, \mathbf{x} \in \mathbb{R}^n. \tag{3.4}$$

For any given multivariate distribution, $\mathbf{F}$, we can separate the copula, $\mathcal{C}$, and the marginal distributions, F_i. Since the marginals, F_i, contain all the univariate information on the variable X_i, and, the joint distribution $\mathbf{F}$ contains all univariate *and* multivariate information, it follows that the copula $\mathcal{C}$ must contain all the information on the dependence between the $X_i's$. The tail dependency is solely informed by dependence between the respective time series. Therefore, we can determine the implied tail dependency using only the information implied by the copula.

3.2.1 Tail Dependence

The stated hypothesis of this thesis formulates a possible difference in the tail ends of the dependency. Also, it is a given that the tails often require special attention due to non-conventional shapes in the far ends of the distributions, see (McNeil et al., 2005). Therefore, I will use a measure called the tail dependence to focus more precisely on the tail end of the return distribution. The tail dependence is defined following the notation from Nelsen (2006). Let X and Y be continuous random variables with distribution functions F and G, respectively. Then, the lower tail dependence parameter, λ_{Lower}, is the limit (conditioned on its existence) of the conditional probability that Y is less than the 100t-th percentile of G given that X is less than the 100t-th percentile of F as t approaches 0, written as

$$\lambda_{Lower}(X,Y) = \lim_{t \to 0^+} P(Y \leq G^{-1}(t) | X \leq F^{-1}(t)). \tag{3.5}$$

In the same matter, the upper tail dependence parameter, λ_{Upper}, is the limit (conditioned on its existence) of the conditional probability of Y being greater than the 100t-th of G given that X is greater than the 100t-th percentile of F as t approaches 1, that is

$$\lambda_{Upper}(X,Y) = \lim_{t \to 1^-} P(Y > G^{-1}(t) | X > F^{-1}(t)). \tag{3.6}$$

This result only depends on the copula of X and Y and can therefore be considered a non-parametric estimation, see Nelsen (2006). Consequently, we take the tail dependence, λ, as independent of the marginal distributions of the underlying series, (X, Y). Therefore, λ_{Lower} can fully capture and concisely present the propensity of X_t and Y_t to experience joint extreme adverse effects. The tail dependency is a multivariate dependency measure. Therefore, it is necessary to obtain a multivariate dependency fit. To this end, I will fit a copula function, $\mathcal{C}$, that allows us to calculate an implied tail dependency.

3.2.2 Copula Families

A copula can describe the dependency between marginals and join distributions in many ways. Different copula families can address different kinds of dependence structures. To test for the existence, direction (lower or upper dependence), and size of tail dependencies in a flexible but computationally feasible matter, I will be fitting copulas from a set of 21 different copulas types. Part of this set comprises ten distinct copula families that are also rotated by 180 degrees to obtain 20 different copula types. The set has 21 copula types since I will also fit the independence copula. Including the independence copula will allow testing for independence between the fitted series. The rotation of 180 degrees increases the flexibility of possible fits due to the asymmetry of the distribution of certain copula families. I will introduce these ten different copula families to better understand their respective characteristics and how they introduce flexibility in the tested for tail dependencies. Table 3.1 shows the numbers used to refer to the 21 different copula types chosen for the fitting set.

Table 3.1: Numbering system for chosen copula fit set

Label	Independence	Gaussian	Student t	Clayton	Gumbel	Frank	Joe	BB1	BB6	BB7	BB8
Indicator Number	0	1	2	3	4	5	6	7	8	9	10
Rotated Indicator Number	N/A	N/A	N/A	13	14	N/A	16	17	18	19	20

Note. Rotated refers to the 180 degrees of rotation of the respective copula families. N/A (not applicable) is used to indicating that adding these reference numbers is not of additional value since these rotated copulas are symmetrically distributed.

Elliptical Copula

In finance, two of the most commonly used elliptical copulas are the Gaussian and t copula (Bouyé, Durrleman, Nikeghbali, Riboulet, & Roncalli, 2000). Their generating formula uses normal and student-t distributions, respectively. These are elliptically shaped distributions. Hence, they are referred to as elliptical copula. Early financial applications of copulas made most use of the Gaussian copula. However, it can be argued that a t copula is preferred over the Gaussian since a Gaussian copula underestimates the probability of joint extreme downward movements relative to the t copula (Kole, Koedijk, & Verbeek, 2007). The bivariate function and corresponding rank correlation of Gaussian and t copula are shown in Table 3.2 (Trivedi & Zimmer, 2007).

Table 3.2: Gaussian and t copula with their respective functions and dependence between Kendall's τ and the Pearson's correlation coefficient, ρ

Label	**Function**	**Kendall's $\boldsymbol{\tau}$**	**Tail Dependence ($\boldsymbol{\lambda_{Lower}}$,$\boldsymbol{\lambda_{Upper}}$)**
Gaussian	$\mathcal{C}_p = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$	$\tau = \frac{2}{\pi} arcsin(\rho)$	0
t	$\mathcal{C}_{p,v} = t_v(t_v^{-1}(u_1), t_v^{-1}(u_2))$	$\tau = \frac{2}{\pi} arcsin(\rho)$	$2t_{v+1}(-\sqrt{\frac{(v+1)(1-\rho)}{1+\rho}})$

Note. Φ represents the standard normal CDF and t_v represents the CDF of a student-t with v degrees of freedom; The implied lower and upper tail dependencies are represented by λ_{Lower} and λ_{Upper} respectively.

As indicated in Table 3.2, the Gaussian copula has zero tail dependency. The reason for including this family in the test set is to allow for testing the null of no tail dependency. If the Gaussian copula offers the best fit, it will allow me to conclude that no tail dependency is present.

Archimedean Copula

Using elliptical distributions to generate copulas greatly restricts the possible number of shapes a copula can take - thereby reducing the goodness-of-fit capacity. A less restrictive approach is to use a broader set of generator functions, an often applied class using such a broad set of generator functions is the Archimedean Copula. They are defined by a generator function within the copula formula that is strictly decreasing and convex, $\psi : [0,1]- > [0, \inf]$, such that $\psi(1) = 0$. It is a strict generator if $\psi(0) = 1$. A bivariate Archmidean copula is therefore defined as follows

$$\mathcal{C}(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2)). \tag{3.7}$$

Table 3.3 presents key properties of the most relevant bivariate Archimedean copulas (Trivedi & Zimmer, 2007). The copulas referred to as BB copulas are a set of bivariate as introduced by Joe (1997). These copulas are generalizations of the Gumbel and Clayton copulas, adapted to fit a wider range of possible probability density shapes.

Table 3.3: Copula class labels with their respective generator function ($\phi_\theta(t)$), parameter boundaries, Kendall's τ, and tail dependence

Label	**Generator $\boldsymbol{\phi_\theta(t)}$**	**Parameters**	**Kendall's $\boldsymbol{\tau}$**	**Tail Dependence ($\boldsymbol{\lambda_L}$,$\boldsymbol{\lambda_U}$)**
Frank	-ln($\frac{e^{-\theta t}-1}{e^{-\theta}-1}$)	$\theta \in \mathbb{R}$	1- $\frac{4}{\theta}$ + 4 $\frac{D_1(\theta)}{\theta}$	(0,0)
Clayton	$\frac{1}{\theta}(t^{-\theta}-1)$	$\theta \geq 0$	$\frac{\theta}{\theta+2}$	$(2^{-\frac{1}{\theta}}, 0)$
Gumbel	$(-ln(t))^\theta$	$\theta \geq 1$	1 - $\frac{1}{\theta}$	(0, 2 - $2^{\frac{1}{\theta}}$)
Joe	$-ln(1-(1-t)^\theta)$	$\theta \geq 1$	1 + $\frac{1}{\theta^2}\int_0^1 (t\ ln(t))(1-t)^{2(1-\theta)\theta}\ dt$	(0, 2 - $2^{\frac{1}{\theta}}$)
BB1	$(t^{-\theta}-1)^\delta$	$\theta \geq 0, \delta \geq 1$	1 - $\frac{2}{\delta(\theta+2}$)	$(2^{-\frac{1}{(\delta\theta)}}, 2^{-\frac{1}{\delta}})$
BB6	$(-log(1-(1-t)^\theta))^\delta$	$\theta \geq 1, \delta \geq 1$	$1+\frac{4}{\delta\theta}\int_0^1(-log(1-(1-t)^\theta)(1-t)(1-(1-t)^{-\theta}))dt$	$(0, 2-2^{-\frac{1}{(\delta\theta))}})$
BB7	$(1-(1-t)^\theta)^{-\delta}-1$	$\theta \geq 1, \delta \geq 0$	$1+\frac{4}{\delta\theta}\int_0^1(-(1-(1-t)^\theta)^{\delta+1}(\frac{(1-(1-t)^\theta)^{-\delta}-1}{(1-t)^{\theta-1}}))dt$	$(2^{-\frac{1}{\delta}}, 2-2^{\frac{1}{\theta}})$
BB8	$-log(\frac{1-(1-\delta t)^\theta}{1-(1-\delta)^\theta})$	$\theta \geq 1, \delta \in [0,1]$	$1+\frac{4}{\delta\theta}\int_0^1(-log(\frac{(1-t\delta)^\delta-1}{(1-\delta)^\theta-1})(1-t\delta)(1-1(1-t\delta)^{-\theta}))dt$	(0,0)

Note. The Debye function is represented as $D_1(\theta) = \theta^{-1}\int_0^\theta \frac{t}{(exp^t-1)}dt$; The copula parameters are represented by θ and δ. The implied lower and upper tail dependency is represented by $\lambda_{L(ower)}$ and $\lambda_{U(pper)}$, respectively.

3.2.3 Copula Estimation

High dimensional fitting of functions to data generally necessitates using some form of likelihood optimization. It often involves using the log-likelihood for a simplified analytical deduction. Therefore, it is helpful to obtain the log-likelihood of a copula. If we define the 'probability integral transform' of X_i as U_i. Or otherwise states as, $U_i \equiv F_i(X_i)$, it follows that $U_i \sim Uniform(0, 1)$ (R. Fisher, 1932; Casella & Berger, 1990; Diebold, Gunther, & Tay, 1998). In the two-dimensional case, it can be proven that $\mathbf{U} = [U_1, U_2]' \sim \mathcal{C}$, which is the copula of $\mathbf{X}$. If we are able to differentiate the joint distribution function *2*-times, followed by taking the 2^{th} cross-partial derivative of equation (3.4) we obtain

$$\mathbf{f}(\mathbf{x}) \equiv \frac{\partial^2}{\partial x_1 \partial x_2}\mathbf{F}(\mathbf{x}) = \prod_{i=1}^{2} f_i(x_i) * \frac{\partial^2}{\partial x_1 \partial x_2}\mathcal{C}(F_1(x_1), F_2(x_2)))$$
$$\equiv \prod_{i=1}^{2} f_i(x_i) * \mathbf{c}(F_1(x_1), F_2(x_2)). \quad (3.8)$$

Therefore, the joint density equals the product of the 'copula density', $\mathbf{c}$, and the marginal densities. Consequently, it is possible to estimate copula-based models in a more modular fashion since the joint log-likelihood is composed of the sum of univariate log-likelihoods and the 'copula log-likelihood. We can therefore denote the joint log-likelihood as

$$log\ \mathbf{f}(\mathbf{x}) = \sum_{i=1}^{2} log\ f_i(x_i) + log\ \mathbf{c}(F_1(x_1), F_2(x_2))). \quad (3.9)$$

The separation of the joint distribution into its copula and marginal distributions enables researchers to more flexibly specify models for the joint distribution. This is important when the goodness-of-fit and shape are the main criteria, as is the case when we want to obtain the implied tail dependency. Separation also allows the estimation of the univariate marginals and copula to be done separately when maximum likelihood (ML) fitting using the likelihood from Equation (3.9). This method is sometimes referred to as the "inference function for margins" (IFM) method (Joe, 1997). It allows for different methods of estimating the copula and marginals, e.g. we can estimate the marginals in a non- or semi-parametric fashion and the copula parametrically.

The standardized residuals obtained from the filtering as described in Section 3.1, $\{z_t\}$, need to be transformed to $U \sim Uniform(0, 1)$ before a copula can be fitted. I achieve this by applying a probability integral transform, which involves assuming that the empirical CDF of $\{z_t\}$ represents the true CDF. Then, plugging $\{z_t\}$ into the found empirical CDF will transform the data to $U \sim Uniform(0,1)$.

Since the question we are trying to answer is regarding the existence of a non-zero tail dependency difference it is of the most importance to correctly specify the copula in the case of no tail dependence. For this reason I chose to use a parametric estimation of the copula, assuming the copula to be part of a predefined set of copulas (class copula estimation). It is found that this method performs better then using non-parametric estimators of tail dependence when testing for the existence of non-zero tail dependency (Frahm, Junker, & Schmidt, 2015).

Therefore, I will fit the copula using the CML method following Chen et al. (2020), which is similar to the IFM method except that it estimates the marginal distribution through an empirical distribution of the filtered series. Similar to IFM, this is a semiparametric two-step copula estimator method, which is robust under misspecified copulas or nonstationarity. Moreover, it also facilitates statistical inferences, e.g. model selection test or hypothesis testing (Chen et al., 2020). Furthermore, it avoids many problems when performing likelihood maximization in a high-dimensionality setting by estimating one-dimensional marginals non-parametrically and then estimating the copula parametrically. For the complete theory of this i.i.d. case semi-parametric estimation, I refer to Genest et al. (1995).

3.2.4 Fitting Copula

A goodness-of-fit statistic is used to obtain the best-fitting copula family with their respective family parameters (θ and or δ). Brechmann (2010) and Manner et al. (2007) performed Monte Carlo simulation studies that validated the optimality of using AIC as the goodness-of-fit test statistics as compared to the other goodness-of-fit tests. Brechmann finds that the AIC has the best performance in selecting the correct copula, even in weak dependence scenarios. I will therefore be using AIC as my goodness-of-fit statistic and perform Maximum likelihood estimation (MLE) for selecting the most parsimonious copula fit from the defined scope of copula families (see Table 3.1, 3.2, and 3.3). The AIC is defined as

$$AIC = 2k - 2log(\hat{L}), \tag{3.10}$$

with k representing the number of parameters and $\hat{L}$ representing the maximum likelihood of the fitted model. Since the best suitable fit minimizes the AIC, it is clear that the $2k$ factor represents a penalty for less parsimonious models. Once we have obtained a copula fit, with a corresponding implied tail dependence, we can relate this result to our stated hypothesis.

3.3 Hypothesis Testing

The central hypothesis of this thesis requires us to verify whether the amount of ESG risk exposure of US public companies, as reflected by the ESG score, impacts the lower tail dependency - a substitute measure of systematic risk. To test this question, I propose the following null-hypothesis

$$H_0 : \delta_{Lower} = \lambda^{high}_{Lower} - \lambda^{low}_{Lower} \neq 0, \tag{3.11}$$

in which λ^{low}_{Lower} represents the lower tail dependency resulting from fitting a copula between the filtered aggregate daily stock log returns of low-scoring ESG stocks and an appropriate filtered market benchmark. Moreover, λ^{high}_{Lower} represents the same thing, except instead using aggregates from high-scoring ESG stocks. I will refer to the tail dependency difference between our ESG sorted portfolios as δ_x, in which x will refer to which side of the distribution it refers to (upper or lower). Therefore, I will refer to the test statistic of Null Hypothesis (3.11) as δ_{Lower}. I hypothesise that a higher exposure, and therefore, a lower ESG score, translates to a higher tail dependency - this I can test by verifying the result from testing Hypothesis (3.11) by a two-sided t-test and observing that the mean is lower than zero.

The problem that arises when one wants to test this null hypothesis is that using my selected methodology, only a point estimate of the tail dependency measure, and therefore, of δ_x will be found. A point estimate will not be sufficient to determine the statistical significance of δ_x in a two-sided t-test, which *is* needed for proper statistical inference regarding the null hypothesis. To obtain the statistical significance, I will be bootstrapping, which, unlike other methods, does not require knowledge of the distribution of the test statistic. Therefore, bootstrapping offers the most straightforward procedure since the distribution of the tail dependency is unknown.

3.3.1 Bootstrap

To test the hypothesis (3.11), we need to obtain a confidence interval of our estimate, δ_{Lower}. We can obtain this confidence interval through bootstrapping. Bootstrapping involves generating point estimates of δ_L from a number of bootstrapped sample data, $\mathbf{M}$, which are obtained by sampling with replacement from the standardized residuals after ARMA-GARCH fitting. The bootstrapped residuals can then be used to get a bootstrapped estimate of the tail dependencies through copula fitting. These tail dependencies can consequently be used to calculate our test statistic, δ_{Lower}.

I will be performing a tri-variate bootstrap. A tri-variate bootstrap involves drawing with replacement from the same location in three input series (residuals from ARMA-GARCH filtering) simultaneously, obtaining three bootstrapped output series. Consequently, performing this bootstrap with three identical input series will produce three identical output series. Performing a tri-variate bootstrap in this fashion allows the cross-dependency of the High ESG-, Low ESG-, and S&P- portfolio return to be preserved, which is a requirement if the bootstrapped test statistic estimates are to represent this cross-dependency.

Generating $\mathbf{M}$ test statistics estimates, $\hat{\delta_L}$, from $\mathbf{M}$ bootstraps will allow me to perform a two-sided t-test. When the t-test is rejected, I can assume the Null Hypothesis (3.11) not to be rejected. This is a valid method if the central limit theorem holds. I use the number of bootstraps, $\mathbf{M} = 1000$. This number of bootstraps should be sufficient given that $\mathbf{M} = 400$ is the minimally required amount to test using a p-value of 0.05 (Davidson & MacKinnon, 2000). I cannot increase the number of bootstraps due to computational constraints.

4 Data

4.1 Data

I obtained the ESG RepRisk Rating (RRR) (RepRisk, n.d.) of publicly traded US-based companies through the Wharton Research Data Services (WRDS) over the period 01-05-2012 until 31-12-2019. This period is chosen due to the benchmark index availability and overlap with the other data sets (CSRP US stock returns and RRR). The RRR is based on daily updated data from over 100 thousand public sources that are screened and classified through the use of machine learning, and this does not include any self-reported statistics (RepRisk, n.d.). This approach of RepRisk can be argued to be attractive since machine learning can offer benefits for deciding ESG scores through aggregate data analysis (Macpherson, Gasperini, & Bosco, 2021). Moreover, not including self-reported company statistics can be argued to increase the reliability of the ESG rating. This is because self-reported statistics are often a bad predictor of ESG risk exposure (Gyönyörová, Stachoň, & Stašek, 2021).

I obtain the company-specific stock prices over the same period from the CSRP stock database (Center for Research in Security Prices, n.d.). Lastly, I obtained the respective S&P Composite 1500 sector-specific indices over the same period from the S&P indices website (Standard and Poor's Dow Jones Indices). S&P Composite 1500 sector-specific indices track the value-weighted aggregated return of the constituents from the S&P Composite 1500 belonging to that specific sector. The S&P Composite 1500 comprises US-based large, mid, and small-cap enterprise indices (SP -500, -400, and -600). It covers 90% of the US public company market cap, which makes it a comprehensive, and therefore attractive, US market benchmark.

Table 4.1: Summary statistics log returns S&P 1500 sector indices

Sector	**Mean**	**Std.**	**Min**	**Max**	**Skewness**	**Kurtosis**
Info Tech	0.0007	0.011	-0.051	0.058	-0.408	5.843
Health Care	0.0006	0.009	-0.046	0.045	-0.388	5.122
Financials	0.0005	0.010	-0.053	0.045	-0.401	5.437
Cons Disc	0.0006	0.009	-0.040	0.060	-0.411	5.800
Communication Services	0.0003	0.010	-0.050	0.052	-0.234	5.600
Industrials	0.0005	0.009	-0.046	0.047	-0.429	5.149
Cons Staples	0.0004	0.007	-0.038	0.030	-0.401	5.045
Energy	0.0000	0.013	-0.067	0.063	-0.193	4.857
Utilites	0.0005	0.008	-0.045	0.029	-0.553	4.814
Real Estate	0.0004	0.009	-0.049	0.034	-0.516	5.002
Materials	0.0004	0.010	-0.046	0.044	-0.302	4.379

4.2 Data Pre-Processing

The raw data acquired comprises the prices of stocks and benchmark indices over time. However, I will use log returns for my analysis, as this normalizes the data and allows time additivity. I will obtain log returns from these prices by converting them as

$$r_t = log(\frac{P_{t+1}}{P_t}), \tag{4.1}$$

where r_t represents the log return of price time series P_t.

Table 4.2: Sector specific number of companies (N) in data set and portfolio formation

Sector	**N Companies Data Set**	**N Companies Portfolios**
Info Tech	191	48
Health Care	353	89
Financials	466	117
Cons Disc	339	85
Communication Services	109	28
Industrials	567	142
Cons Staples	252	63
Energy	232	58
Utilites	108	27
Real Estate	53	14
Materials	70	18

To observe the effect an ESG score has on the codependency in extreme price movements, I sort based on the ESG score. To this end, I perform a yearly sort on ESG scores based on the earliest available ESG rating for that year. I filter out companies with a year without a single ESG rating. Moreover, to prevent missing value problems, I also filtered out any companies with missing price data. The resulting input data set contains daily stock prices and yearly ESG scores on 3550 public US companies, representing coverage of 86.5 % of the total (4012) in the year 2012 (World Bank, n.d.).

Given that the broadest ESG index covers around 1500 US public companies, the filtered ESG RepRisk ratings provide considerably more extensive coverage than conventional ESG benchmarks (S&P Dow Jones Sector Indices, n.d.). I argue that more comprehensive coverage data improves the strength and robustness of possible inference on ESG risk since the data is more representative of US public companies. A downside of this more expansive database is that the component-specific ESG score is not always available (e.g. no indication of how much of the ESG score is derived from the Environment category), which restricts my analysis to only studying the composite ESG score.

I will be using daily stock price data (1930 days of observations after pre-processing) to increase the availability of extreme events, which can help fit a correctly specified copula. A problem with this approach is that the high frequency of data can mean that return outliers of individual companies can obfuscate results. To address this and to allow more straightforward inference on the influence of the ESG score, I will be aggregating the stock returns of the highest and lowest ESG score quartiles. This method is sufficient to average out possible individual outlier companies. I chose to use quartiles due to the widespread use of this measure and its found robustness for US equities (Fama & French, 2015). Table 4.2 shows the number of companies in the corresponding data sets and subsequent portfolios.

In short, to evaluate ESG score-specific characteristics, I will form portfolios by taking quartiles of the sorted scores of companies and aggregating their performance through equal weighting.

I categorized the RepRisk ESG Scores on their respective sectors and applied the methodology to these distinct categories. The different sectors are classified as one of the eleven general S&P sectors (Communication Services, Cons Disc, Cons Staples, Energy, Financials, Health Care, Industrials, Info Tech, Materials, Real Estate and Utilities) (Standard and Poor's Dow Jones Indices, n.d.).

Table 4.3: Summary statistics of log returns from the low ESG score aggregate

Sector	**Mean**	**Std.**	**Min**	**Max**	**Skewness**	**Kurtosis**
Info Tech	0.0004	0.012	-0.056	0.099	-0.107	7.103
Health Care	0.0004	0.010	-0.043	0.083	-0.056	6.117
Financials	0.0006	0.011	-0.051	0.046	-0.367	4.528
Cons Disc	0.0006	0.009	-0.043	0.051	-0.323	4.609
Communication Services	0.0005	0.011	-0.048	0.048	-0.286	4.302
Industrials	0.0004	0.010	-0.048	0.045	-0.474	4.826
Cons Staples	0.0004	0.009	-0.048	0.035	-0.340	4.811
Energy	0.0002	0.016	-0.082	0.073	-0.160	5.360
Utilites	0.0005	0.008	-0.043	0.032	-0.258	4.319
Real Estate	0.0005	0.013	-0.064	0.050	-0.248	4.171
Materials	-0.0002	0.014	-0.059	0.062	-0.024	4.320

Table 4.4: Summary statistics of log returns from the high ESG score aggregate

Sector	**Mean**	**Std.**	**Min**	**Max**	**Skewness**	**Kurtosis**
Info Tech	0.0009	0.014	-0.080	0.174	0.502	18.015
Health Care	0.0008	0.015	-0.056	0.104	0.324	7.161
Financials	0.0005	0.009	-0.039	0.041	-0.183	4.466
Cons Disc	0.0005	0.011	-0.044	0.047	-0.153	3.783
Communication Services	0.0004	0.011	-0.047	0.045	-0.257	4.105
Industrials	0.0006	0.012	-0.046	0.045	-0.298	3.940
Cons Staples	0.0006	0.011	-0.042	0.084	0.218	5.936
Energy	0.0000	0.019	-0.097	0.128	0.247	7.297
Utilites	0.0005	0.008	-0.044	0.050	-0.193	4.962
Real Estate	0.0006	0.013	-0.060	0.048	-0.259	4.049
Materials	0.0002	0.013	-0.068	0.049	-0.195	3.980

The log returns summary statistics of the S&P 1500 sector indices, low ESG score aggregate, and high ESG score aggregate can be found in Table 4.1, 4.3, and 4.4 respectively. These tables show that very few sector log returns follow a normal return pattern since we have non-zero skewness and kurtosis that is different from four. However, this is as expected for financial time series.

In Figure 4.1 the log-returns of the S&P 1500 sector indices are plotted versus the same-day log returns of respective sector-specific high/low ESG score aggregate log returns. I observe distinct relations within different sectors, e.g. highly different behaviour when comparing the log returns of both high and low ESG score aggregate versus the S&P 1500 index or the more correlated behaviour visible in the Real Estate sector plot. From this, we can conclude that sector-specific analysis seems sensible. Additional analysis on the return characteristics of the portfolio returns after data pre-processing can be found in Appendix B.

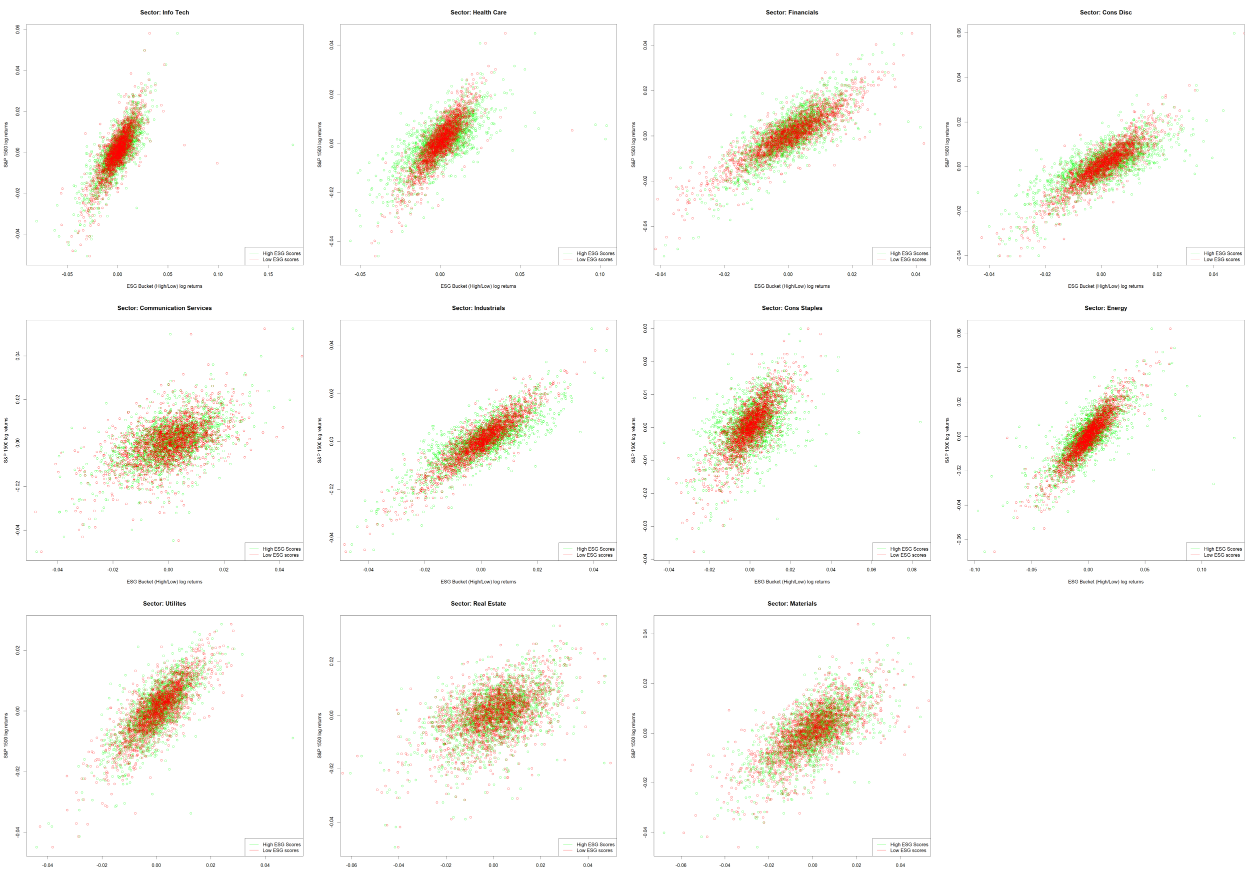

Figure 4.1: Log returns of S&P 1500 indices versus log returns of high/low ESG score aggregate.

5 Results

5.1 Filtering

Following a general-to-specific fitting methodology, the initial fit is an ARMA-GARCH (1,1)(1,1) model; this involves fitting a model where each component is fitted a lag parameter for a lag of 1-time step. I test the standardised residuals to verify the successful filtering of unwanted financial data features (serial correlation and hetereoskadicity).

The Ljung-Box test, as introduced in Section 3.1.1, is used to test for the presence of serial correlation after filtering. Table 5.1 shows that the Ljung-Box test is not rejected for a single portfolio; from this, we can conclude that serial correlation is not present in the standardised residuals.

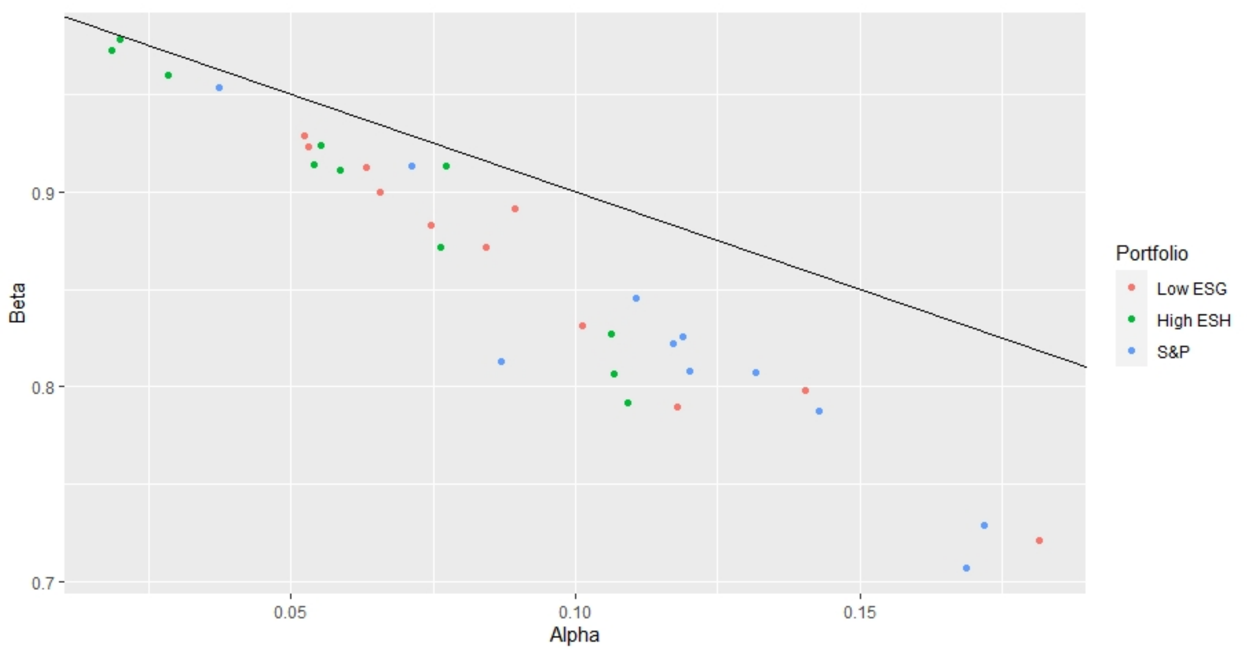

Figure 5.1: Alpha versus Beta, which are the fitted GARCH parameters of the ARMA-GARCH models; the black line represent Alpha plus Beta equals one.

I apply the ARCH-LM, as introduced in Section 3.1.1, and the Ljung-Box on squared residuals to test for hetereoskadicity. The former testing for hetereoskadicity directly, while the latter tests for serial correlation in the squared residuals. Both can indicate the presence of hetereoskadicity in the standardised residuals. From both Table 5.2 and 5.3, we observe that the standardised residuals from the reference S&P Utilities portfolio still possess hetereoskadicity. Moreover, from Table 5.3, we find the same for the low ESG Industrials portfolio. This calls for additional fitting attempts to obtain the desired residuals suited for copula fitting.

I manually fit ARMA-GARCH models with increasingly higher lag orders for the two portfolios that failed the respective test. The corresponding ARMA-GARCH orders that have been found to produce residuals with the desired characteristics are (1,1)(2,2) and (1,1)(3,3) for the low ESG Industrials and S&P Utilities portfolios, respectively. Table 5.4 shows that these portfolio-specific ARMA-GARCH fits pass the tests, allowing the assumption of the standardised residuals being i.i.d. The accompanying final fit statistics (AIC and likelihood values) can be found in Appendix C. Figure 5.1 shows the values of both the predicted alpha and beta in the ARMA-GARCH models for the GARCH(1,1) equation. A black line is drawn to indicate where alpha plus beta equals one. The parameters are not above or on this line, which shows that the fit model is stationary. Therefore we can conclude that the resulting standardised residuals are suited for copula fitting.

Table 5.1: Test statistics for the Ljung-Box test applied to the standardized residuals

Test: Ljung-Box Res.		**Low ESG**		**High ESG**		**S&P**	
Sector	**Lag**	**Statistic**	**p-value**	**Statistic**	**p-value**	**Statistic**	**p-value**
Info Tech	1	0.108	0.742	0.050	0.822	0.057	0.811
	5	0.787	1.000	0.537	1.000	0.528	1.000
	9	2.062	0.982	1.397	0.998	2.463	0.955
Health Care	1	0.112	0.737	0.004	0.950	0.220	0.639
	5	1.396	0.999	1.411	0.999	1.682	0.993
	9	2.158	0.977	3.138	0.869	2.885	0.907
Financials	1	0.010	0.922	0.013	0.909	0.103	0.748
	5	1.221	1.000	1.665	0.993	2.517	0.770
	9	3.971	0.698	3.962	0.700	5.494	0.348
Cons Disc	1	0.059	0.809	0.019	0.891	0.760	0.383
	5	0.792	1.000	1.728	0.990	1.239	1.000
	9	2.216	0.973	3.475	0.806	2.624	0.939
Communication Services	1	0.078	0.781	0.204	0.652	0.001	0.975
	5	0.953	1.000	1.344	1.000	1.049	1.000
	9	2.222	0.973	2.695	0.931	1.532	0.997
Industrials	1	0.139	0.710	0.022	0.883	0.042	0.838
	5	0.743	1.000	0.490	1.000	2.608	0.718
	9	1.475	0.997	1.478	0.997	5.670	0.314
Cons Staples	1	0.068	0.795	0.018	0.893	0.049	0.825
	5	2.281	0.879	1.033	1.000	0.820	1.000
	9	4.032	0.683	2.242	0.972	1.825	0.991
Energy	1	0.405	0.524	0.152	0.696	0.091	0.763
	5	0.884	1.000	2.050	0.948	0.714	1.000
	9	1.650	0.995	3.303	0.840	2.006	0.984
Utilites	1	0.421	0.516	0.001	0.971	0.209	0.647
	5	3.866	0.093	0.607	1.000	2.993	0.476
	9	5.750	0.300	2.346	0.965	5.228	0.403
Real Estate	1	0.004	0.951	0.032	0.859	0.014	0.907
	5	1.518	0.998	2.109	0.934	1.403	0.999
	9	3.427	0.816	4.639	0.538	4.674	0.530
Materials	1	0.017	0.895	0.001	0.973	0.455	0.500
	5	2.835	0.577	3.338	0.279	3.067	0.431
	9	4.352	0.607	5.246	0.399	5.382	0.371

Note. Test statistics for the Ljung-Box test applied to the standardized residuals of the ARMA-GARCH (1,1)(1,1) model fit for the respective sector data; Not rejecting the test at a 5% confidence value is indicated with a grey cell colouring. High ESG, Low ESG and S&P refer to the respective ESG sorted and benchmark portfolios.

Table 5.2: Test statistics for the Ljung-Box test applied to the squared standardized residuals

Test: Ljung-Box Sqrd. Res.		**Low ESG**		**High ESG**		**S&P**	
Sector	**Lag**	**Statistic**	**p-value**	**Statistic**	**p-value**	**Statistic**	**p-value**
Info Tech	1	3.251	0.071	0.028	0.867	0.519	0.471
	5	3.704	0.293	0.128	0.997	1.760	0.676
	9	4.392	0.524	0.221	1.000	3.154	0.733
Health Care	1	0.249	0.618	0.489	0.484	0.724	0.395
	5	0.742	0.915	0.566	0.947	2.404	0.527
	9	1.794	0.928	1.043	0.984	6.257	0.270
Financials	1	0.029	0.864	0.086	0.770	0.250	0.617
	5	0.227	0.991	0.755	0.912	1.212	0.810
	9	0.770	0.994	1.656	0.942	2.693	0.808
Cons Disc	1	0.005	0.943	0.083	0.773	0.057	0.812
	5	2.608	0.483	1.688	0.694	1.361	0.774
	9	6.581	0.237	5.020	0.427	3.613	0.654
Communication Services	1	0.001	0.974	0.146	0.702	0.506	0.477
	5	1.190	0.815	0.551	0.950	1.693	0.693
	9	2.481	0.841	3.204	0.725	3.491	0.676
Industrials	1	1.258	0.262	0.435	0.510	0.306	0.580
	5	5.142	0.142	0.841	0.895	1.127	0.830
	9	8.306	0.112	1.702	0.937	3.109	0.741
Cons Staples	1	0.056	0.813	1.186	0.276	0.972	0.324
	5	0.566	0.947	1.643	0.705	1.843	0.656
	9	1.391	0.964	4.256	0.546	3.112	0.740
Energy	1	0.010	0.921	2.934	0.087	1.228	0.268
	5	0.875	0.887	3.316	0.352	2.095	0.596
	9	1.818	0.926	4.180	0.559	3.274	0.713
Utilites	1	2.204	0.138	0.113	0.737	3.003	0.083
	5	3.311	0.353	2.544	0.497	8.110	0.028
	9	5.439	0.368	4.311	0.537	15.437	0.003
Real Estate	1	0.316	0.574	0.012	0.914	0.731	0.393
	5	1.699	0.691	2.017	0.614	0.896	0.883
	9	3.773	0.628	4.093	0.573	2.350	0.859
Materials	1	0.084	0.772	0.001	0.979	1.746	0.186
	5	0.273	0.986	1.067	0.844	2.477	0.511
	9	0.446	0.999	2.099	0.893	3.580	0.661

Note. Test statistics for the Ljung-Box test applied to the squared standardized residuals of the ARMA-GARCH (1,1)(1,1) model fit for the respective sector data; Not rejecting the test at a 5% confidence value is indicated with a grey cell colouring. High ESG, Low ESG and S&P refer to the respective ESG sorted and benchmark portfolios.

Table 5.3: Test statistics for the ARCH LM test applied to the standardized residuals

Test: ARCH LM		**Low ESG**		**High ESG**		**S&P**	
Sector	**Lag**	**Statistic**	**p-value**	**Statistic**	**p-value**	**Statistic**	**p-value**
Info Tech	1	0.307	0.580	0.113	0.737	0.030	0.864
	5	0.718	0.818	0.170	0.972	1.698	0.542
	9	1.287	0.863	0.179	0.998	2.395	0.634
Health Care	1	0.197	0.657	0.013	0.909	0.267	0.606
	5	1.234	0.665	0.158	0.975	3.321	0.247
	9	1.595	0.802	0.414	0.986	6.104	0.134
Financials	1	0.054	0.816	0.154	0.695	0.029	0.865
	5	0.400	0.913	1.611	0.564	1.434	0.610
	9	0.878	0.933	1.873	0.744	2.320	0.650
Cons Disc	1	0.110	0.740	0.151	0.698	0.013	0.910
	5	6.064	0.058	3.645	0.209	3.303	0.249
	9	8.065	0.051	5.512	0.178	4.525	0.277
Communication Services	1	0.027	0.871	0.247	0.620	1.124	0.289
	5	2.249	0.419	0.599	0.854	2.206	0.428
	9	2.857	0.541	3.494	0.425	3.761	0.382
Industrials	1	1.785	0.182	0.040	0.841	0.004	0.949
	5	6.367	0.049	0.668	0.833	2.248	0.419
	9	7.830	0.057	1.235	0.873	3.527	0.420
Cons Staples	1	0.000	0.995	0.280	0.597	0.114	0.736
	5	1.369	0.627	0.844	0.780	1.141	0.692
	9	1.638	0.793	1.522	0.817	1.737	0.773
Energy	1	0.750	0.387	0.168	0.682	0.179	0.672
	5	1.742	0.531	0.722	0.817	1.079	0.710
	9	1.954	0.727	1.172	0.884	2.075	0.702
Utilites	1	0.241	0.624	0.587	0.444	1.894	0.169
	5	3.060	0.281	2.244	0.420	13.575	0.001
	9	4.037	0.341	3.197	0.477	15.845	0.001
Real Estate	1	1.041	0.308	0.227	0.634	0.053	0.819
	5	1.912	0.491	0.495	0.885	0.205	0.964
	9	3.463	0.431	2.008	0.716	1.828	0.754
Materials	1	0.025	0.873	0.108	0.742	0.218	0.641
	5	0.362	0.923	0.511	0.880	1.014	0.729
	9	0.445	0.983	0.891	0.931	1.800	0.760

Note. Test statistics for the ARCH LM test applied to the standardized residuals of the ARMA-GARCH (1,1)(1,1) model fit for the respective sector data; Not rejecting the test at a 5% confidence value is indicated with a grey cell colouring. High ESG, Low ESG and S&P refer to the respective ESG sorted and benchmark portfolios.

Table 5.4: Refitted data test outcomes

Fit Information	**Test**	**Lag**	**statistic**	**p-value**
Sector (Data Source)	LB Stan. Res.	1	0.130	0.719
Industrials (Low ESG)		5	0.758	1.000
Refitted ARMA-GARCH		9	1.518	0.997
(1,1)(2,2)	LB Stan. Squared Res.	1	0.245	0.620
		5	8.714	0.163
		9	13.004	0.194
	ARCH LM	5	2.408	0.121
		7	3.960	0.203
		9	4.429	0.337
Sector (Data Source)	LB Stan. Res.	1	0.159	0.690
Utilities (S&P)		5	2.466	0.797
Refitted ARMA-GARCH		9	4.568	0.555
(1,1)(3,3)	LB Stan. Squared Res.	1	0.245	0.620
		5	8.714	0.163
		9	13.004	0.194
	ARCH LM	7	0.515	0.473
		9	1.088	0.745
		11	2.797	0.632

Note. The ARMA-GARCH fit is the case-specific fit offering the best fit when increasing the lag order from (1,1)(1,1); LB Stan. Res. refers to the Ljung-Box test on the standardized residuals resulting from the ARMA-GARCH fit. LB Stan. Squared Res. refers to the Ljung-Box test on the squared standardized residuals resulting from the ARMA-GARCH fit. ARCH LM refers to the ARCH Lagrange Multiplier test applied to standardized residuals resulting from the ARMA-GARCH fit.

5.2 Copula Fit

I applied the MLE AIC fit to the standardized residuals as described in Section 3.2 and 3.2.4. Table 5.5 presents the resulting copula fit output, which was the result of fitting on the original standardized (non-bootstrapped) residuals. The BB1 copula or its rotated version offers the best fit for most sectors. As opposed to the Student-t copula, the BB1 copula has an asymmetrical tail dependency, indicating tail-side specific risk behaviour - significant losses appear to be co-moving distinctly from significant gains in the stock. The copula fit results support the stated hypothesis; there appears to be a clear difference between the fitted copula parameters of the Low and High ESG portfolios. The most pronounced difference is present for the Cons Staples sector, with a difference of factor ten in the parameter value. However, since this is a single estimate without a confidence range, I cannot yet describe the significance of these findings.

Table 5.5: Copula fit output original data set

Parameter:	**Copula Type**		**Par. 1**		**Par. 2**	
Sector/ESG Score	**High**	**Low**	**High**	**Low**	**High**	**Low**
Info Tech	17	17	0.169	0.156	2.036	2.177
Health Care	17	17	0.189	0.207	1.777	2.307
Financials	17	7	0.244	1.001	2.062	1.894
Cons Disc	17	17	0.167	0.183	1.959	2.653
Communication Services	17	17	0.243	0.178	1.359	1.379
Industrials	17	17	0.224	0.343	2.142	2.759
Cons Staples	17	2	0.159	0.749	1.362	12.687
Energy	17	17	0.210	0.329	2.026	2.988
Utilites	2	2	0.737	0.793	7.240	5.743
Real Estate	2	2	0.474	0.465	10.402	9.010
Materials	7	2	0.675	0.690	1.398	12.429

Note. Number 2 refers to the Student t copula family, number 7 (17) refers to the (rotated) BB1 copula; The whole set of copula fit numbering can be found in Table

3.1.

5.3 Tail Dependence

Table 5.6 shows the implied tail dependencies (see Table 3.2 and 3.3) from the copula fit on the standardized residuals (non-bootstrapped). The tail dependency point estimates from the original data indicate a distinct difference between the High and Low ESG portfolios. Moreover, there is an even more pronounced difference between the estimates of the upper or lower side of the distribution. These findings indicate a noticeable sector-specific and tail-side-specific effect of ESG risks on the tail dependency with sector benchmarks. However, the same precaution regarding the significance due to a lack of confidence interval should be taken.

Table 5.6: Implied tail dependencies copula fit original data set

Sector/ λ (ESG Score)	**Upper (High)**	**Lower (High)**	**Upper (Low)**	**Lower (Low)**
Info Tech	0.594	0.133	0.625	0.131
Health Care	0.523	0.127	0.650	0.235
Financials	0.600	0.253	0.694	0.558
Cons Disc	0.576	0.121	0.701	0.239
Communication Services	0.334	0.123	0.347	0.059
Industrials	0.618	0.235	0.714	0.480
Cons Staples	0.337	0.040	0.184	0.184
Energy	0.592	0.196	0.739	0.494
Utilites	0.296	0.296	0.408	0.408
Real Estate	0.068	0.068	0.085	0.085
Materials	0.480	0.358	0.140	0.140

Note. The tail dependency, λ, is either "Upper" or "Lower", representing the upper- or lower- tail dependency, respectively. The ESG sorted portfolio of the upper quartile (High ESG scores) or the lower quartile (Low ESG scores) are represented by "High" or "Low" respectively.

Figure 5.2: Sector specific bootstrapped means and 95% standard confidence interval error bars of tail dependency differences between ESG sorted portfolios ($\hat{\delta} = \lambda$ High ESG - λ Low ESG) for both the upper and lower tail dependencies.

The Histograms of the sector-specific values for our bootstrapped tail dependency difference, $\hat{\delta}$, can be found in Appendix D. From these, we observe that a t-test appears viable, and the found means appear to approach the true mean.

Table 5.7 and 5.8 show the sector-specific values for $\hat{\delta}$. Figure 5.2 visualizes the found means of $\hat{\delta}$ with the corresponding 95% error bars. For both the upper and lower side of the distribution, almost all the found means of $\hat{\delta}$ are significantly different from zero, finding this through the rejection of two-sided t-tests. Only the found mean of the upper tail dependency difference ($\hat{\delta}_{Upper}$) for the Cons Staples sector is not significantly different from zero. This means no influence of ESG risk is present for the upper tail dependence of public US Consumer Staples companies. The values in Table 5.7 do not lead to rejection of the Null Hypothesis 3.11 for most sectors - since most means of $\hat{\delta}_{Lower}$ are lower than and significantly different from zero. However, I find the opposite effect for the Materials and Communication Services sectors. I find that $\hat{\delta}_{Lower}$ is significantly higher than zero. From this, we can conclude that ESG risk exposure significantly influences the lower tail dependency of all US sectors - downside risk pricing is affected. Therefore, I can conclude that ESG Risk exposure generally lowers the downside risk co-movement of US stocks.

The upper and lower sides of the distribution of $\hat{\delta}$ indicate that for US-based public Materials companies, a low ESG risk exposure (high ESG rating) *increases* the co-movement of companies with the S&P benchmark during both significant gains and losses. I find the same behaviour for US public Communication Services companies, but only regarding the co-movement during significant gains - not losses.

Table 5.7: Bootstrapped lower tail dependency difference ($\hat{\delta}_{Lower}$) statistical output

Sector	**Mean** $\hat{\delta}_{Lower}$	**t statistic**	**p value**
Info Tech	-0.022	-6.826	0.000
Health Care	-0.129	-43.377	0.000
Financials	-0.162	-42.847	0.000
Cons Disc	-0.065	-12.664	0.000
Communication Services	0.043	16.932	0.000
Industrials	-0.258	-90.741	0.000
Cons Staples	-0.069	-19.122	0.000
Energy	-0.244	-72.117	0.000
Utilites	-0.120	-65.408	0.000
Real Estate	-0.051	-21.150	0.000
Materials	0.050	10.245	0.000

Table 5.8: Bootstrapped upper tail dependency difference ($\hat{\delta}_{Upper}$) statistical output

Sector	**Mean** $\hat{\delta}_{Upper}$	**t statistic**	**p value**
Info Tech	-0.065	-18.405	0.000
Health Care	-0.288	-49.027	0.000
Financials	-0.091	-54.270	0.000
Cons Disc	-0.184	-44.525	0.000
Communication Services	-0.061	-13.879	0.000
Industrials	-0.118	-62.404	0.000
Cons Staples	0.007	1.479	0.139
Energy	-0.211	-55.199	0.000
Utilites	-0.134	-49.993	0.000
Real Estate	-0.020	-11.423	0.000
Materials	0.201	35.842	0.000

5.3.1 Interpretation

For the economic explanation of the found distinct Material and Communication Services sector downside risk behaviours, I propose the following dynamics to explain the observed ESG risk influence anomalies. High sector consolidation causes downside ESG risk to affect larger companies less than smaller ones due to regulatory capture. Moreover, one can argue that smaller companies generally tend to have a higher ESG rating than larger ones because large companies are shielded from litigation and responsibilities due to the aforementioned regulatory capture. This shielding allows larger companies to function with a lower ESG score than their smaller counterparts - who do need to take into account ESG risks and adhere to ESG policies.

Additionally, due to the commoditized nature of the products in these sectors, economies of scale will be the primary competitive driver. Also, the commoditized nature of products will cause smaller companies to seek differentiation through a higher ESG rating. Therefore, the difference between the Low and High ESG portfolios is not only one of ESG rating but also one of company size - between large and small companies, respectively. This size-through-ESG-sort causes the market to price exposure to systematic risks of the Low ESG portfolio *less* than for the High ESG portfolio.

Moreover, I propose that this effect is not present in other consolidated sectors that also obtained regulatory capture (e.g. Info Tech) because these companies have a higher end-product exposure to the public domain, as compared to companies in the Material sector. Companies with consumers as their end customers will be more exposed to the downsides of not conforming to higher ESG standards. Therefore, the markets price them as such. We do not observe this in the relatively consumer-facing Communication Services sector because this sector is consolidated to such an extent that, for many customers, no viable alternatives are present.

The economic explanation for the distinct Material sector systematic upside return behaviour could be that the markets do expect the smaller high ESG-rated companies to outperform their more sluggish large consolidated counterparts during market rallies.

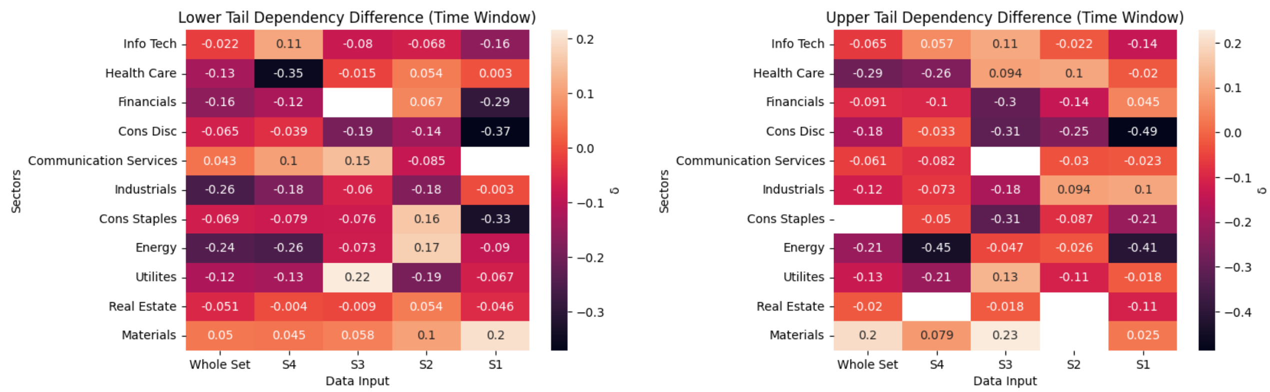

Figure 5.3: Sector-specific heat-maps of the lower and upper tail dependency differences ($\hat{\delta}_{Lower}$ & $\hat{\delta}_{Upper}$), which are calculated by applying the methodology (Bootstrap number M = 1000) to equally sized continuous subsets of the original data set (indicated by the column indicator of S1 to S4 respectively), e.g. S1 represents the first quarter of data points; The values from the main experiment on the original data set are included as a reference and referred to as 'Whole Set'; All insignificant results at a 5% confidence value are replaced with white space.

5.3.2 Robustness

To probe the time robustness of the found tail dependency differences, I performed a sub-window analysis - applying the methodology to subsections of the residuals from ARMA-GARCH filtering (using M = 1000 bootstraps). With this approach, we can probe if the significance, sign, and size of the observed tail dependencies differences are a time-dependent feature or remain constant.

The results are displayed for the lower and upper tail dependency differences in Figure 5.3. From these tables, we can consider the existence of a non-zero tail dependency to be time robust since all values are significant and non-zero. The sign of the lower tail dependency difference also appears robust when considering the outcomes for subsections S4, S4, and S1, since almost each sector value in each subsection has the same sign as that of the 'Whole Set' outcome. The time robustness of the upper tail dependency difference sign appears less solid. The observed time subset difference could result from outliers. Still, I argue that the most likely explanation is that the upper tail dependency differences are dynamic over time, reflecting shifting market attitudes towards the pricing of ESG risks.

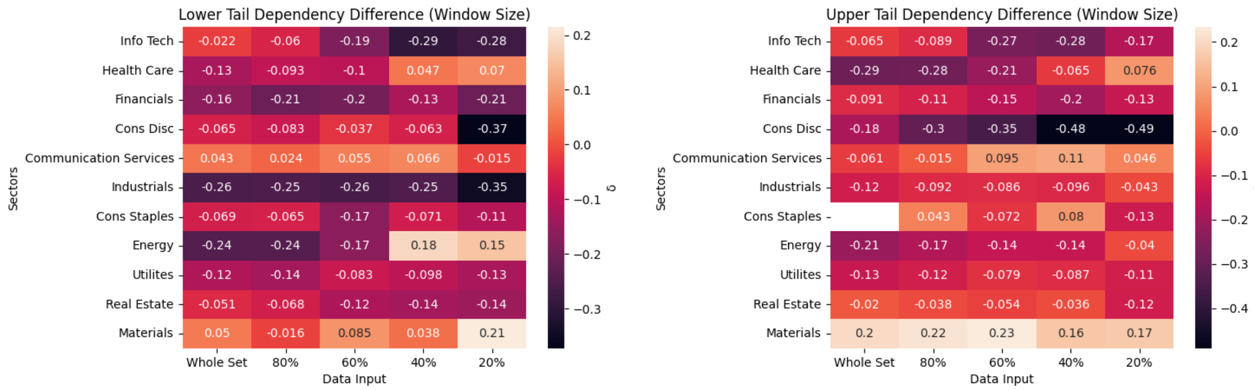

Figure 5.4: Sector-specific heat-maps of the lower and upper tail dependency differences ($\hat{\delta}_{Lower}$ & $\hat{\delta}_{Upper}$), which are calculated by applying the methodology (Bootstrap number M = 1000) to an increasing size of input data set; starting with the first 20% of the data points and expanding with 20% for each subsequent measurement; The values from the main experiment on the whole data set are included as a reference and referred to as 'Whole Set'. All insignificant results at a 5% confidence value are replaced with white space.

Additionally, I test the robustness of our results to differing sample sizes by applying the methodology to an expanding sample size. Applying the methodology to a sample of the first twenty % of the data, then applying it to the first forty %, until we reach the output of the original experiment on the whole data set.

From figure 5.4 we observe that from a sample size of 60 % the whole data set (1930 observations) $\hat{\delta}_{Lower}$ & $\hat{\delta}_{Upper}$ are robust in significance, sign and size when changing the input sample size. More dynamic findings are found for applying the methodology to the first 20 and 40 % of the data. Still, the significance and sign are largely maintained. From these findings, we can conclude that our estimates of the ESG-driven tail dependency differences are robust in both time- and window size - they reflect a significant, maintained but dynamic effect over time and sample size.

6 Conclusion

I have performed a sector-specific ESG sort on US-based public companies to obtain a High and Low ESG rating portfolio return. These ESG portfolio returns and their respective industry benchmark indices (S&P) were filtered using ARMA-GARCH modelling. The resulting standardized residuals were transformed to allow copula fitting. Copula implied tail dependency estimates were obtained by fitting a copula between the High or Low ESG portfolios transformed standardized residuals and those of their benchmarks. The difference between the tail dependencies of High and Low ESG portfolios was the test statistic of choice. I bootstrapped the standardized residuals to obtain a confidence interval and answered the stated research question.

I find that US markets robustly price systematic risk reduction stemming from reduced ESG risk exposure. More specifically, a low ESG risk exposure (high ESG rating) reduces the co-movement of US stocks for both significant gains and losses. This finding indicates that markets price the influence of ESG ratings (assume ESG risk exposure) as both upside and downside systematic shock co-movement averting. Except for the Material, Communication Services, and Cons Staples sectors, these appear to have different risk dynamics. For the latter, no significant influence of ESG risk on the upside shock exposure appears to be present.

Interestingly, the Communication Services sector is found to have increased co-movement stemming from ESG risk exposure for the lower end of the distribution - if markets have high losses, high ESG-rated companies will have a higher chance of also having high losses. However, for the upper end of the distribution, the opposite is true - high ESG-rated Communication Services companies also have a lower chance of having significant gains in tandem with the whole sector. High ESG-rated Communication Services companies are priced as having increased expected sector-wide losses while lowering expected sector-wide gains.

6.1 Shortcomings

A possible problem with the methodology of this thesis could be with the used period of the data. It could be that a slightly shifted period with years outside my current set exposes different dynamics - or show nonexistent ESG risk influence. I could not analyze a shifted set due to data availability constraints.

I find another possible issue in the assumption of the distribution of the test statistic, δ. It is unclear whether the test statistic's bootstrapped distribution is always suitable for a t-test. Moreover, a possible issue related to the RepRisk ESG rating is the underlying assumption that the ESG rating is a good approximation of the amount of ESG risk exposure of a company. The extent to which these correlate is hard to verify due to the opaque nature of the RepRisk ESG rating algorithm.

Another point of attention is that the comparisons between portfolios only measure the difference between the extreme ends of the ESG ratings. Further studies could address this by performing portfolio formation and comparisons that also include the middle parts of the ESG rating distributions or using a more coarse portfolio selection.

6.2 Policy Recommendations

The found significant influence of ESG risk could be a reason for including ESG ratings as risk factors in asset pricing. Moreover, these findings support using ESG ratings as a basis for diversifying portfolios against systematic risk.

Given the finding that a high ESG exposure reduces both upward and downward co-movement during systematic shocks, it can be argued that ESG ratings can be used as a diversification factor in systematic risk management. Moreover, the existence of ESG rating-driven risk dynamics supports the possibility of ESG ratings being a relevant factor in asset pricing. The finding that high ESG exposure reduces downward co-movement during systematic shocks supports the stated hypothesis of this thesis - namely, that sustainable practices protect against ESG shocks. The finding that high ESG exposure reduces upward co-movement during systematic shocks can be explained by arguing that the additional costs of ESG compliance are priced as reducing the potential of growth and or profitability of companies. This proposed effect could explain the observed upper tail side behaviour since these are important factors for pricing companies during stock rallies.

The found unique properties of US Communication Services companies indicate the viability of an ESG portfolio shorting strategy. This would involve selling the high- and buying the low- ESG Communication Services portfolios. Resulting in a decrease in the systematic risk and an increase in the upward market movement exposure. Also, high ESG ratings of the Consumer Staples sector companies appear to only lower downside systematic risk exposure, but not their upside exposure. Consequently, performing high ESG portfolio formation of Consumer Staples companies appears attractive. Moreover, the finding that US Material companies are priced as having increased systematic shock exposure stemming from reduced ESG risk exposure could be explained by the market stating the materials sector has unique challenges related to ESG policies. This could be used in policy recommendations or portfolio formation.

Lastly, the finding that the height of the ESG rating correlates with systematic risk reduction can be used to include ESG ratings as a factor in sustainable portfolio formation.

6.3 Future Research

Future research could focus on applying a similar sector-specific methodology to different periods to observe possibly time-specific dynamics. Moreover, future research could look at the effect of larger size time windows on the resulting copula fits.

Additionally, an ESG-size double sort could provide insights into the proposed economic reasoning for the observed divergent downside risk behaviour of the Material and Consumer Services sector. Given that sector-specific ESG risk influence on tail risk dynamics is found to be present, future research could also focus on verifying this presence using other measures or data from regions other than the US. One could apply the methodology to EU data. As an example finding, I expect that the found Consumer Staples upper tail dependency difference, δ_{Upper}, will be even higher due to European consumers valuing ESG policies more for their staple goods [1]. Therefore applying a similar methodology to the EU market could be of interest. More broadly, differing results could reflect how consumers value ESG differently than in the US. That, in turn, would be reflected in how the market perceives the ESG risk effects.

[1] More information on the Nielsen Global Survey can be found on: $https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/Global20Sustainability20Report\_October202015.pdf$

References

Allen, T., Dees, S., Caicedo Graciano, C. M., Chouard, V., Clerc, L., de Gaye, A., ... others (2020). *Climate-related scenarios for financial stability assessment: an application to France.* Working paper, Banque de France.

Aloui, R., Gupta, R., & Miller, S. M. (2016). Uncertainty and crude oil returns. *Energy Economics*, *55*, 92–100.

American Bar Association. (2016). *The interplay between corporate governance issues and litigation: What is corporate governance and how does it affect litigation?* Available at `https://www.americanbar.org/groups/business_law/publications/blt/2016/12/05\_grace/` (25-02-2022).

Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, *63*(3), 443–494.

Ashwin Kumar, N., Smith, C., Badis, L., Wang, N., Ambrosy, P., & Tavares, R. (2016). ESG factors and risk-adjusted performance: A new quantitative model. *Journal of Sustainable Finance & Investment*, *6*(4), 292–300.

Bajo, E., Barbi, M., & Romagnoli, S. (2015). A generalized approach to optimal hedging with option contracts. *The European Journal of Finance*, *21*(9), 714–733.

Battiston, S., Dafermos, Y., & Monasterolo, I. (2021). Climate risks and financial stability. *Journal of Financial Stability*, *54*, 100867.

Bax, K., Sahin, Ö., Czado, C., & Paterlini, S. (2021). ESG, risk, and (tail) dependence. *Available at SSRN 3846739*.

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica*, *78*(1), 395–410.

BenSaïda, A., & Litimi, H. (2021). Financial contagion across G10 stock markets: A study during major crises. *International Journal of Finance & Economics*, *26*(3), 4798–4821.

Berg, Kölbel, J., & Rigobon, R. (2019). Aggregate confusion: The divergence of ESG ratings.

doi: http://dx.doi.org/10.2139/ssrn.3438533

Berg, E., & Lange, K. W. (2020). *Enhancing ESG-risk modelling-a study of the dependence structure of sustainable investing* (Master's thesis, Royal Institute of Technology). Retrieved from `https://www.diva-portal.org/smash/get/diva2:1385199/FULLTEXT01.pdf`

Bertram, C., Hilaire, J., Kriegler, E., Beck, T., Bresch, D., Clarke, L., ... others (2020). *NGFS climate scenarios database: Technical documentation.* Potsdam Institute for Climate Impact Research.

Bloomberg. (2021). *ESG assets may hit $53 trillion by 2025, a third of global aum.* Available at `https://www.bloomberg.com/professional/blog/esg-assets-may-hit-53-trillion-by\-2025-a-third-of-global-aum/` (31-03-2022).

Boero, G., Silvapulle, P., & Tursunalieva, A. (2011). Modelling the bivariate dependence structure of exchange rates before and after the introduction of the euro: a semi-parametric approach. *International Journal of Finance & Economics*, *16*(4), 357–374.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bouyé, E., Durrleman, V., Nikeghbali, A., Riboulet, G., & Roncalli, T. (2000). Copulas for finance-a reading guide and some applications. *Available at SSRN 1032533*.

Box, G. E., & Jenkins, G. (1970). *Time series analysis: forecasting and control.* Holden Day, San Francisco.

Brechmann, E. C. (2010). *Truncated and simplified regular vines and their applications* (Doctoral dissertation, Technical University of Munich). Retrieved from `https://mediatum.ub.tum.de/doc/1079285/file.pdf`

Bressan, G. M., & Romagnoli, S. (2021). Climate risks and weather derivatives: A copula-based pricing model. *Journal of Financial Stability*, *54*, 100877.

Brown, B. J., Hanson, M. E., Liverman, D. M., & Merideth, R. W. (1987). Global sustainability: Toward definition. *Environmental Management*, *11*(6), 713–719.

Brush, S. (2021). *Blackrock's jump into sustainable investing leads to record haul.* Available at `https://www.bloomberg.com/news/articles/2021-10-13/blackrock-s-jump-into\-sustainable-investing-leads-to-record-haul` (31-03-2022).

Casella, G., & Berger, R. (1990). *Statistical inference.* Duxbury Press, U.S.A.

Center for Research in Security Prices. (n.d.). *CSRP documentation.* Available at `https://www.crsp.org/documentation` (01-03-2022).

Changqing, L., Yanlin, L., & Mengzhen, L. (2015). Credit portfolio risk evaluation based on the pair copula var models. *Journal of Finance and Economics*, *3*(1), 15–30.

Chen, X., Fan, Y., & Patton, A. J. (2004). Simple tests for models of dependence between multiple financial time series, with applications to us equity returns and exchange rates. *London Economics Financial Markets Group Working Paper*(483).

Chen, X., Xiao, Z., & Wang, B. (2020). Copula-based time series with filtered nonstationarity. *Journal of Econometrics*, *228*(1), 127–155.

Cheung, Y.-W., & Lai, K. S. (1995). Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, *13*(3), 277–280.

Christoffersen, P., Errunza, V., Jacobs, K., & Langlois, H. (2012). Is the potential for international diversification disappearing? A dynamic copula approach. *Review of Financial Studies*, *25*(12), 3711–3751.

Coqueret, G., Stiernegrip, S., Morgenstern, C., Kelly, J., Frey-Skött, J., & Österberg, B. (2021). Boosting ESG-based optimization with asset pricing characteristics. *Available at SSRN 3877242*.

Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, *19*(4), 497–515.

Davidson, R., & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews*, *19*(1), 55–68.

De Jonghe, O. (2010). Back to the basics in banking? A micro-analysis of banking system stability. *Journal of Financial Intermediation*, *19*(3), 387–417.

De Luca, G., & Zuccolotto, P. (2011). A tail dependence-based dissimilarity measure for financial time series clustering. *Advances in Data Analysis and Classification*, *5*(4), 323–340.

Diebold, F., Gunther, T., & Tay, A. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review,*, *39*(4), 863-883.

Embrechts, P., Lindskog, F., & McNeil, A. (2001). Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich*, *14*.

Embrechts, P., Lindskog, F., & McNeil, A. (2003). *Modelling dependence with copulas and applications to risk management. handbook of heavy tailed distributions in finance, edited by ST Rachev.* Elsevier, North-Holland, Amsterdam.

Embrechts, P., McNeil, A., & Straumann, D. (2002). *Correlation and dependence in risk management: properties and pitfalls* (Vol. 1). Camebridge University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987–1007.

Engle, R. F., Giglio, S., Kelly, B., Lee, H., & Stroebel, J. (2020). Hedging climate change news. *Review of Financial Studies*, *33*(3), 1184–1216.

European Commission. (2018). *Renewed sustainable finance strategy and implementation of the action plan on financing sustainable growth.* Available at `https://ec.europa.eu/info/publications/sustainable-finance-renewed-strategy_en` (31-03-2022).

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

Fatica, S., & Panzica, R. (2021). Green bonds as a tool against climate change? *Business Strategy and the Environment*, *30*(5), 2688–2701.

Fisher, R. (1932). *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

Fisher, T. J., & Gallagher, C. M. (2012). New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association*, *107*(498), 777–787.

Flori, A., Pammolli, F., & Spelta, A. (2021). Commodity prices co-movements and financial stability: A multidimensional visibility nexus with climate conditions. *Journal of Financial Stability*, *54*, 100876.

Frahm, Junker, & Schmidt. (2015). Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, *37*(1), 80-100.

Fredricks, G. A., & Nelsen, R. B. (2007). On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, *137*(7), 2143–2150.

Genest, C., Favre, A.-C., et al. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, *12*(4), 347–368.

Genest, C., Gendron, M., & Bourdeau-Brien, M. (2013). The advent of copulas in finance. In *Copulae and multivariate probability distributions in finance* (pp. 13–22). Routledge, UK.

Genest, C., Ghoudi, K., & Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*(3), 543–552.

Górka, J., & Kuziak, K. (2022). Volatility modeling and dependence structure of ESG and conventional investments. *Risks*, *10*(1), 20.

Gyönyörová, L., Stachoň, M., & Stašek, D. (2021). ESG ratings: Relevant information or misleading clue? Evidence from the S&P global 1200. *Journal of Sustainable Finance & Investment*, 1–35.

He, L.-Y., & Chen, S.-P. (2011). Nonlinear bivariate dependency of price–volume relationships in agricultural commodity futures markets: A perspective from multifractal detrended cross-correlation analysis. *Physica A: Statistical Mechanics and its Applications*, *390*(2), 297–308.

Henriksson, R., Livnat, J., Pfeifer, P., & Stumpp, M. (2019). Integrating ESG in portfolio construction. *Journal of Portfolio Management*, *45*(4), 67–81.

Hofert, M., & Scherer, M. (2011). CDO pricing with nested archimedean copulas. *Quantitative Finance*, *11*(5), 775–787.

Hu, H.-L. (1998). *Large sample theory for pseudo-maximum likelihood estimates in semiparametric models*. University of Washington.

Jin, I. (2018). Is ESG a systematic risk factor for US equity mutual funds? *Journal of Sustainable Finance Investment*, *8*(1), 72-93.

Joe, H. (1997). *Multivariate models and dependence concepts* (Vol. 73). Chapman and Hall, London.

Karimalis, E. N., & Nomikos, N. K. (2018). Measuring systemic risk in the european banking sector: A copula covar approach. *The European Journal of Finance*, *24*(11), 944–975.

Kole, E., Koedijk, K., & Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance*, *31*(8), 2405–2423.

Laux, P., Vogl, S., Qiu, W., Knoche, H. R., & Kunstmann, H. (2011). Copula-based statistical refinement of precipitation in rcm simulations over complex terrain. *Hydrology and Earth System Sciences*, *15*(7), 2401–2419.

Li, T.-T., Wang, K., Sueyoshi, T., & Wang, D. D. (2021). ESG: Research progress and future prospects. *Sustainability*, *13*(21), 11663.

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *Journal of Finance*, *72*(4), 1785–1824.

Longin, F., & Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, *56*(2), 649–676.

Macpherson, M., Gasperini, A., & Bosco, M. (2021). Artificial intelligence and fintech technologies for ESG data and analysis. *Available at SSRN 3790774*.

Manner, H., et al. (2007). *Estimation and model selection of copulas with an application to exchange rates*. Citeseer.

Martellini, L., & Vallée, L.-S. (2021). Measuring and managing ESG risks in sovereign bond portfolios and implications for sovereign debt investing. *The Journal of Portfolio Management*, *47*(9), 198–223.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton University Press, Princeton and Oxford.

Michelis, L., & Ning, C. (2010). The dependence structure between the canadian stock market and the usd/cad exchange rate: a copula approach. *Canadian Journal of Economics/Revue canadienne d'économique*, *43*(3), 1016–1039.

Morales, R., Di Matteo, T., & Aste, T. (2014). Dependency structure and scaling properties of financial time series are related. *Scientific Reports*, *4*(1), 1–9.

Nelsen, R. B. (2006). *An introduction to copulas*. Springer Science & Business Media, New York.

Organisation for Economic Co-operation and Development. (2014). *Risk management and corporate governance* (Vol. 6; Tech. Rep.). Paris, France.

Patton, A. J. (2009). Copula–based models for financial time series. In *Handbook of financial time series* (pp. 767–785). Springer.

Pedersen, L. H., Fitzgibbons, S., & Pomorski, L. (2021). Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics*, *142*(2), 572–597.

Reboredo, J. C. (2013). Is gold a safe haven or a hedge for the us dollar? implications for risk management. *Journal of Banking & Finance*, *37*(8), 2665–2676.

RepRisk. (n.d.). *Reprisk methodology overview.* Available at `https://www.reprisk.com/news-research/resources/methodology#:~:text=\The%20RepRisk%20Rating%20(RRR)%2C,of%20the%20ten%20UNGC%20Principles./` (27-02-2022).

Sandoval Junior, L., Mullokandov, A., & Kenett, D. Y. (2015). Dependency relations among international stock market indices. *Journal of Risk and Financial Management*, *8*(2), 227–265.

Sassen, R., Hinze, A.-K., & Hardeck, I. (2016). Impact of ESG factors on firm risk in europe. *Journal of Business Economics*, *86*(8), 867–904.

Setzer, J., & Higham, C. (2021). Global trends in climate change litigation: 2021 snapshot. *Grantham Institute on Climate Change and the Environment*.

Shirvani, A., & Volchenkov, D. (2019). A regulated market under sanctions: On tail dependence between oil, gold, and tehran stock exchange index. *arXiv preprint arXiv:1911.01826*.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, *8*, 229–231.

S&P Dow Jones Sector Indices. (n.d.). *Core ESG.* Available at `https://www.spglobal.com/spdji/en/index-family/esg/core-esg/?utm_medium=next_gen&utm_source=google&utm_campaign=paid_campaign&utm_term=esg&indices&utm_content=Intl_ESG&gclid=Cj0KCQiAg_KbBhDLARIsANx7wAzw2f7l3ld9FWmudMHZIZaJpJq8f3Vg2hrih-vBKMo3sv_Lrtx_-R4aAiiFEALw_wcB#overview` (22-11-2022).

Standard and Poor's Dow Jones Indices. (n.d.). *S&P Dow Jones Sector Indices.* Available at `https://www.spglobal.com/spdji/en/landing\/investment-themes/sectors` (27-02-2022).

Szetela, B., Mentel, G., & Gedek, S. (2016). Dependency analysis between bitcoin and selected global currencies. *Dynamic Econometric Models*, *16*, 133–144.

Tachibana, M. (2018). Safe-haven and hedge currencies for the US, UK, And Euro area stock markets: A copula-based approach. *Global Finance Journal*, *35*, 82–96.

Taylor, S. (1986). *Modelling financial time series.* Wiley, New York, NY.

Teräsvirta, T. (2009). An introduction to univariate garch models. In *Handbook of financial time series* (pp. 17–42). Springer.

Trivedi, P. K., & Zimmer, D. M. (2007). *Copula modeling: An introduction for practitioners.* Now Publishers Inc.

van der Beck, P. (2021). Flow-driven ESG returns. *Available at SSRN 3929359*.

Wen, X., Wei, Y., & Huang, D. (2012). Measuring contagion between energy market and stock market during financial crisis: A copula approach. *Energy Economics*, *34*(5), 1435–1446.

World Bank. (n.d.). *Listed domestic companies, total - United States.* Available at `https://data.worldbank.org/indicator/CM.MKT.LDOM.NO?name_desc=false&locations=US` (22-11-2022).

Zhang, J., De Spiegeleer, J., & Schoutens, W. (2021). Implied tail risk and ESG ratings. *Mathematics*, *9*(14), 1611.

A Dependency Measures

To quantify the dependence, one can make use of dependency measures. The Pearson correlation coefficient is one of the most commonly used bivariate dependency measures. Its widespread use is most likely due to its straightforwardness in its interpretability and computation. The problem with using Pearson's correlation is that only linear dependence is being measured, which is an issue for financial data due to the often present non-linear features (e.g. fat tails and volatility clustering). Moreover, due to the marginal transformations required in obtaining our dependency estimate Pearson's ρ correlation is not suited since it is not invariable to these transformations.

Two main candidates for evaluating the dependency in a matter invariant to marginal transformations are Kendall's τ and Spearman's ρ. Unsurprisingly, these dependency measures are closely related to the copula function. This is because they also represent cross-dependencies without taking marginal information into account. Kendall's τ has a smaller gross error sensitivity (GES) and asymptotic variance (AV) in most scenarios (Croux & Dehon, 2010). Moreover, Kendall's τ is much more prevalent within the copula and financial dependency literature. Therefore, I will focus on Kendall's τ over Spearman's ρ. This correlation is defined as

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y]}{\sigma_X \sigma_Y}, \tag{A.1}$$

with $corr$(X,Y) and $\rho_{(X,Y)}$ representing the Pearson correlation coefficient of the random variables X and Y. Moreover, σ represents their respective standard deviation, and μ their means.

A.0.1 Kendall's τ

Kendall's τ is a measure of concordance, which indicates the correlation between the ranks of two random variables, e.g. X and Y. In this way, it can capture the dependence between any two monotonic functions. We have concordance when $(X_1 - X_2)(Y_1 - Y2) > 0$ and discordance when $(X_1 - X_2)(Y_1 - Y2) < 0$. Kendall's τ is defined as the probability of concordance minus the probability of discordance. Therefore, it can be represented as

$$\tau(X,Y) = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0), \tag{A.2}$$

where the sets (X_1, Y_1) and (X_2, Y_2) are considered two independent draws from the same join distribution (X,Y), and $\tau \in [-1, 1]$. This can be estimated empirically as

$$\hat{\tau}(X,Y) = \binom{n}{2}^{-1} \sum_{1 \leq t < s \leq n} sign((X_t - X_s)(Y_s - Y_t)), \tag{A.3}$$

where n represents the amount of observations in the dataset and *sign* is a function that returns one when the input is positive and minus one when it is negative. I use the formulation of the relation between copulas and Kendall's τ by Fredricks and Nelsen (2007), which in the bivariate scenario is represented as

$$\tau = 4 \int_{u_1} \int_{u_2} \mathcal{C}(u_1, u_2) d\mathcal{C}(u_1, u_2) - 1. \tag{A.4}$$

For the scope of copula used in this thesis, the copula is fully described by Kendall's τ and vice versa. The Kendall's τ can therefore be used to obtain a goodness-of-fit indication by comparing the empirical value with the implied value from the copula fit. The Kendall's τ is a measure describing the dependence of the whole distribution.

B Input Data Processing Analysis

Summary statistics comparing the *merged* dataset resulting from the data pre-processing described in Section 4.2 can be found in Table B.1. From Table B.1, we can observe that for highly consolidated sectors, such as Information Technology or Materials, our *merged* dataset roughly covers the same number of companies as the benchmark index. For less consolidated industries the size bias of the S&P leads to a larger number of companies being present in the merged dataset.

Table B.1: Relevant statistics of input data thesis

Sector	$\boldsymbol{N_{merged}}$	$\boldsymbol{N_{benchmark}}$	***merged* avg. %** $\frac{r}{y}$	***benchmark* avg. %** $\frac{r}{y}$
Info Tech	191	199	20.453	17.570
Health Care	351	187	20.824	15.468
Financials	466	235	15.548	13.734
Cons Disc	339	209	17.171	14.242
Communication Services	109	52	15.924	8.051
Industrials	567	227	17.198	12.465
Cons Staples	252	80	16.955	10.884
Energy	232	63	6.776	0.054
Utilites	108	51	14.146	12.114
Real Estate	53	113	18.794	9.804
Materials	70	89	5.757	8.820

Note. *merged* refers to the data set resulting from merging the CRSP stock data with the Reprisk ESG rating data. Moreover, *benchmark* refers to the respective S&P1500 sector composite indices - these track the aggregated performance of constituents of the S&P1500 composite index belonging to specific sectors. $\mathbf{N}_{dataset}$ refers to the number of companies present in the dataset classified to be within their respective sector. Lastly, $\frac{r}{y}$ refers to the yearly return over the period 01-05-2012 until 31-12-2019.

C ARMA-GARCH Fit Statistics

Table C.1: Likelihood values ARMA-GARCH fit

Sector/ Data	**High ESG**	**Low ESG**	**S&P**
Info Tech	5604.330	5877.438	6223.112
Health Care	5466.291	6183.475	6503.159
Financials	6383.911	6066.428	6248.848
Cons Disc	5960.829	6386.683	6491.617
Communication Services	6064.356	6115.709	6284.063
Industrials	5909.621	5592.930	6374.318
Cons Staples	6086.854	6543.176	6906.096
Energy	5179.926	5429.250	5824.638
Utilites	6585.798	6541.574	5592.930
Real Estate	5673.522	5688.107	6463.443
Materials	5652.226	5592.930	6224.060

Note. High ESG, Low ESG and S&P refer to the respective ESG sorted and benchmark portfolios.

Table C.2: AIC values ARMA-GARCH fit

Sector/ Data	**High ESG**	**Low ESG**	**S&P**
Info Tech	-5.801	-6.084	-6.443
Health Care	-5.658	-6.402	-6.733
Financials	-6.609	-6.280	-6.469
Cons Disc	-6.171	-6.612	-6.721
Communication Services	-6.278	-6.331	-6.506
Industrials	-6.118	-5.790	-6.599
Cons Staples	-6.301	-6.774	-7.150
Energy	-5.362	-5.620	-6.030
Utilites	-6.818	-6.773	-6.769
Real Estate	-5.873	-5.888	-6.692
Materials	-5.851	-5.790	-6.444

Note. High ESG, Low ESG and S&P refer to the respective ESG sorted and benchmark portfolios.

D Histograms Bootstrapped $\hat{\delta}_{Lower}$ & $\hat{\delta}_{Upper}$

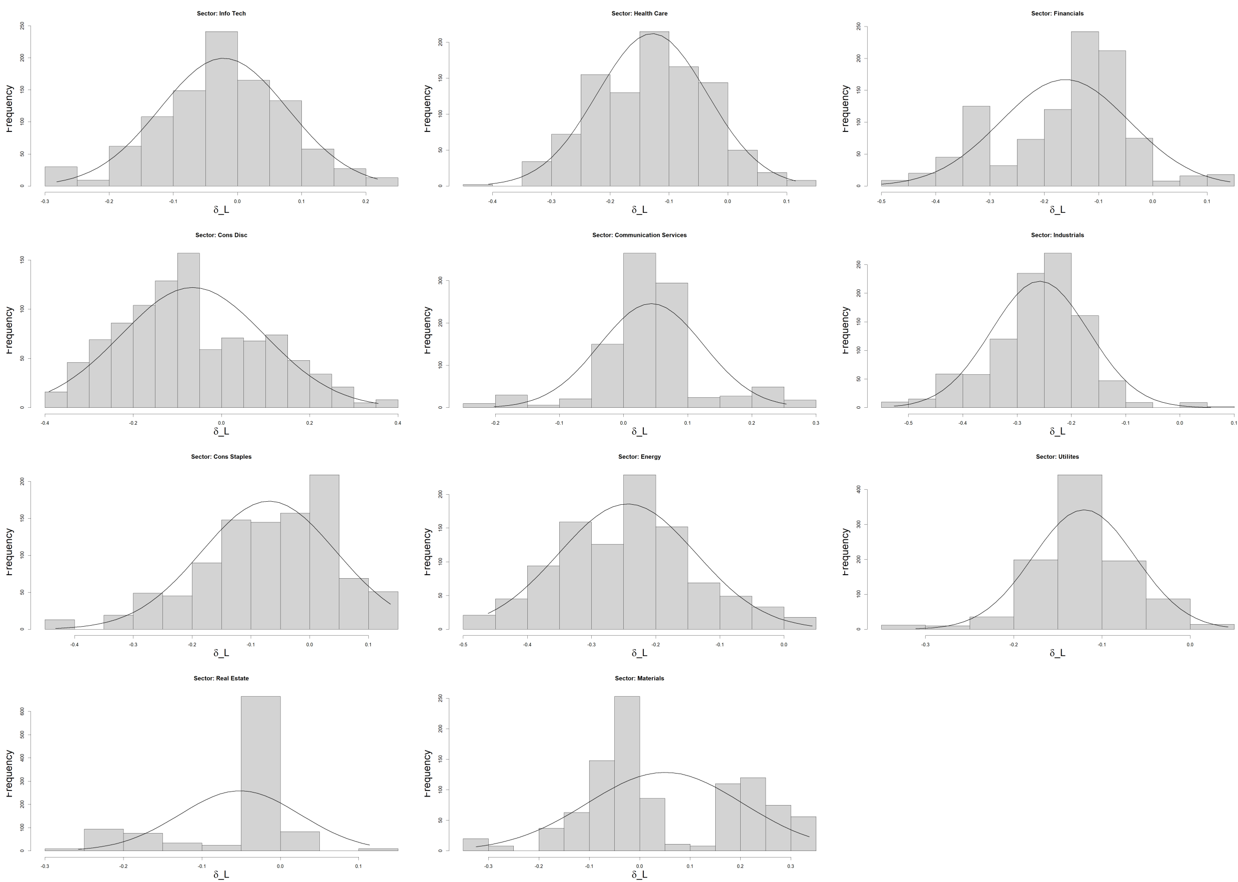

Figure D.1: Histograms of the sector specific values for $\hat{\delta}_{Lower}$
;The black line represents the normal distribution of the mean and variance from the bootstrapped $\hat{\delta}_{Lower}$.

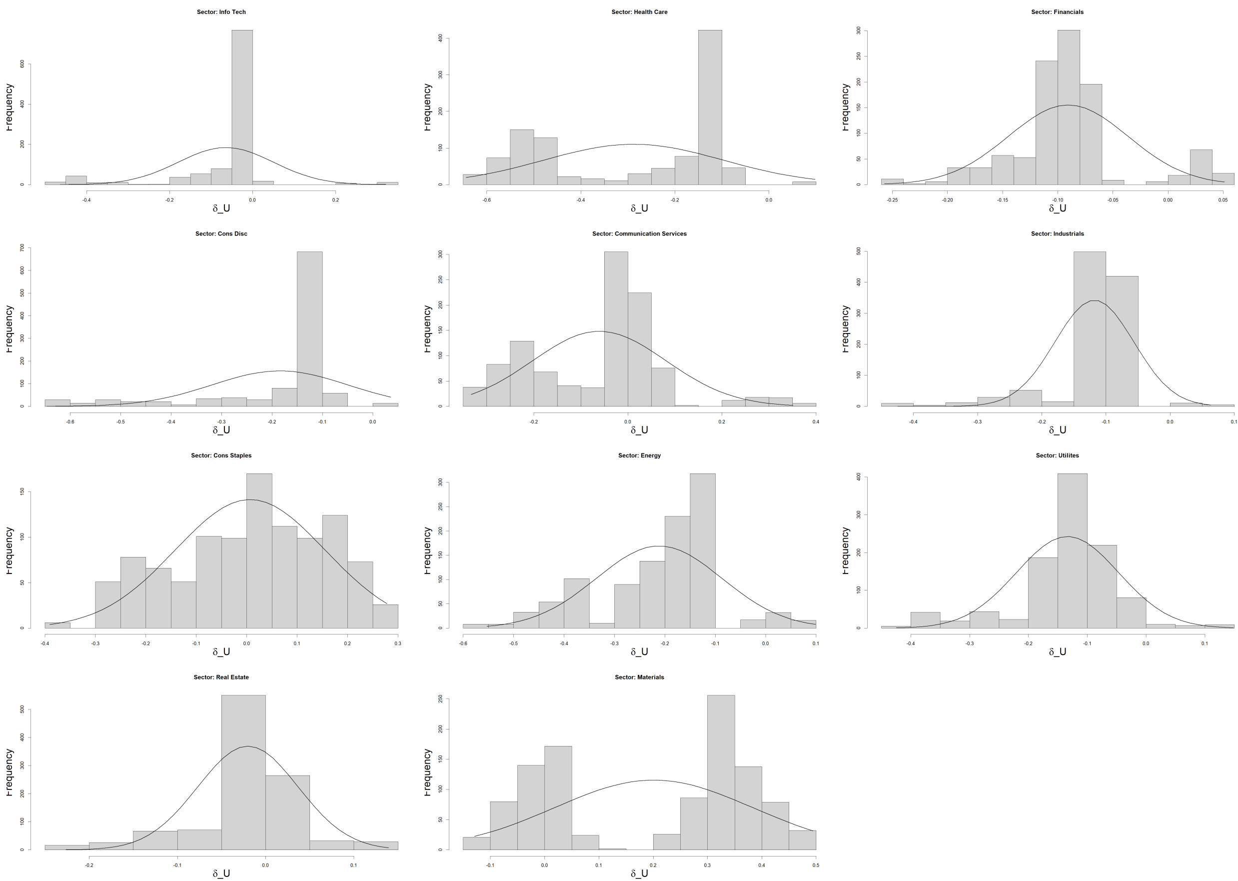

Figure D.2: Histograms of the sector specific values for $\hat{\hat{\delta}}_{Upper}$;The black line represents the normal distribution of the mean and variance from the bootstrapped $\hat{\hat{\delta}}_{Upper}$.