# Data Collection & Analysis

This section will outline a comprehensive methodology for analysing the competing risks of a financial portfolio which we have assumed, consisting of the S&P500 index and US Treasury Bonds using copulas. Our objective is to provide a robust and flexible approach for modelling the dependence structure between these two assets, which is crucial for evaluating the risk profile of the portfolio.

To achieve this, we began by collecting historical data on the S&P500 index and US Treasury Bonds, using a Bloomberg terminal. We used this data to estimate their individual marginal distributions, which provided us with an understanding of the behaviour of each asset in isolation. In particular, we looked for certain statistical properties of these distributions, including their moments, skewness, and kurtosis, to gain insights into the underlying risk characteristics of each asset.

To establish the marginal distributions of the data, we conducted a thorough analysis of the historical data for the S&P500 index and US Treasury Bonds. We began by collecting daily price data for both assets over a period of several years. We then calculated the daily returns for each asset, which were defined as the percentage change in price from the previous day's closing price. Next, we conducted a series of statistical tests to evaluate the distributional properties of the returns for each asset. We first plotted histograms of the returns, as shown in Figure 5.1 below, and visually inspected them for symmetry and skewness. We also calculated summary statistics for the returns, including the mean, standard deviation, skewness, and kurtosis, to gain further insights into the underlying distributional properties.

Chart, histogram

Description automatically generated

Based on our analysis, we found that the returns for the S&P500 index exhibited slight positive skewness and leptokurtosis, indicating that they were not normally distributed. The returns for US Treasury Bonds, on the other hand, exhibited negative skewness and platykurtosis, indicating that they were negatively skewed and had thinner tails than a normal distribution. In order to model the marginal distributions of the two assets, we considered several different distributions, including normal, log-normal, and Student's t-distributions. We fit each distribution to the empirical data using maximum likelihood estimation and evaluated the goodness-of-fit using the Kolmogorov-Smirnov test.

From this, we found that the log-normal distribution provided the best fit to the S&P500 index returns, while the Student's t-distribution provided the best fit to the returns for US Treasury Bonds. We therefore used these distributions to model the marginal distributions of the two assets in our subsequent analyses.

# Fitting a Joint Distribution to the Data

To fit a joint distribution to the S&P500 index and US Treasury Bond returns data, we first attempted to fit a multivariate Gaussian distribution. However, we found that this did not do a good job of capturing the dependence structure between the two assets, which is known to be non-linear and non-normal. Specifically, the Gaussian distribution failed to capture the tail dependence between the two assets, which is particularly important in financial crises when the prices of both assets can be strongly affected by a common shock.

To address this limitation, we then fit a Gaussian copula to the data. A copula allows us to link the marginal distributions of multiple variables to their joint distribution. The Gaussian copula is a popular choice because it allows for flexible dependence structures, and can be easily parameterized based on the correlation matrix of the data. Our analysis showed that the Gaussian copula did a better job of capturing the dependence structure between the S&P500 index and US Treasury Bond returns compared to the multivariate Gaussian distribution. However, we also found that the Gaussian copula understated the risk of joint extreme events in the tails of the distribution. This is because the Gaussian copula assumes a symmetric dependence structure, which may not be realistic in financial crises when the returns of both assets can move together in a non-linear and asymmetric way. Therefore, while the Gaussian copula provides a better approximation to the joint distribution of the data than the multivariate Gaussian distribution, it may underestimate the risk of extreme events, particularly in the tails of the distribution.

We then proceeded to model the dependence structure between the two assets using copulas. As explained previously, copulas are a powerful mathematical tool for modelling the joint distribution of two or more random variables, and this allowed us to capture any complex dependencies that may exist between the S&P500 index and US Treasury Bonds. In particular, copulas enabled us to model the tail dependencies between the two assets, which are often crucial in assessing the risk of financial portfolios.

We used various copula models to evaluate the dependence structure between the two assets, including Archimedean and elliptical copulas. We analysed the goodness-of-fit of each model using statistical tests, such as the Kolmogorov-Smirnov test and the Cramer-von Mises test, to determine the most appropriate copula for our analysis.

In summary, our methodology provides a comprehensive and flexible approach for modelling the competing risks of a financial portfolio composed of the S&P500 index and US Treasury Bonds using copulas. By utilizing this approach, we aim to provide valuable insights into the risk profile of the portfolio, which can guide investment decision-making.

## Methods

In order to fit a joint probability distribution to the data, we used a variety of different methodologies. Firstly, we assumed normality of the underlying marginal distributions, i.e. of the S&P 500 Returns and then Bond Returns, and hence we fit the data using a Gaussian Multivariate Distribution. This was done by calculating the means and covariance of the data, and from this we could then extract the probabilities of the portfolio returning less than our inputted values, using the CDF.

In a similar way, we next fit the data with a Gaussian Copula. The key difference between using a Gaussian Copula and a Gaussian Multivariate Distribution is that the multivariate distribution assumes that the underlying marginals are normally distributed. In contrast, the Gaussian Copula allows for the underlying marginals to be non-normally distributed, and in this case Student’s t-distributed. As a result of this, we expect that the extreme values in the tails of our distribution will be incorporated more effectively into our probabilities.

In order to compare these different methods of modelling probabilities of returns, we used financial data going back to the early 2000s, in particular before the 2008 financial crash. Copulas were (and in some cases still are) used to analyse risk associated with multi-asset portfolios, and we aimed to simulate their shortcomings using the data that would have been available at the time. Hence, we trained our models using financial data from 2000-2007, and then tested using empirical probabilities of returns realised in 2008-2009.

## Results

The analysis of the benefits of using a Gaussian Multivariate Distribution and Copulas to model the probability distribution of a financial portfolio containing a stock and a bond has produced some interesting results. As can be seen in the output table below, it is clear that the Gaussian Multivariate Distribution severely understates the probabilities of extreme returns from the portfolio, when compared with the Empirical Probabilities. For example when we examine the tails of the distribution, i.e. the top and bottom 10 percentiles, the probabilities given by our models are significantly lower than the empirical values.

|  |  |  |  |
| --- | --- | --- | --- |
| Empirical Probabilities | Gaussian Copula Computed Probabilities | Multivariate Gaussian Computed Probabilities | Percentile |
| 0.001912 | 0.000158 | 0.000004 | 0.05 |
| 0.007648 | 0.001002 | 0.000466 | 0.10 |
| 0.019120 | 0.002903 | 0.002860 | 0.15 |
| 0.028681 | 0.009086 | 0.012417 | 0.20 |
| 0.047801 | 0.020151 | 0.028293 | 0.25 |
| 0.072658 | 0.043013 | 0.055975 | 0.30 |
| 0.103250 | 0.087566 | 0.099104 | 0.35 |
| 0.130019 | 0.145100 | 0.147149 | 0.40 |
| 0.175908 | 0.208382 | 0.194912 | 0.45 |
| 0.233270 | 0.250169 | 0.225856 | 0.50 |
| 0.290631 | 0.334699 | 0.286860 | 0.55 |
| 0.346080 | 0.414427 | 0.347244 | 0.60 |
| 0.409178 | 0.511061 | 0.425685 | 0.65 |
| 0.483748 | 0.586786 | 0.494185 | 0.70 |
| 0.550669 | 0.665998 | 0.574012 | 0.75 |
| 0.636711 | 0.760119 | 0.686548 | 0.80 |
| 0.718929 | 0.833340 | 0.793381 | 0.85 |
| 0.801147 | 0.898404 | 0.901016 | 0.90 |
| 0.896750 | 0.943967 | 0.968367 | 0.95 |

On the other hand, the Gaussian Copula also understates the probabilities in the tails, but it is noticeably more accurate than the Multivariate Distribution. This is because the Copula allows for the underlying marginal distributions to be t-distributed, whereas the Multivariate Distribution assumes normally distributed underlying variables. This is an important factor, as the assumption of normality is not always appropriate for financial data, which often exhibits heavy tails and skewness. In particular, throughout the financial crash extreme values became more common and hence higher probabilities should be attributed to them. The Copula can capture these characteristics of the data and provide a better fit than the Multivariate Distribution.

It is worth noting that even though the Copula does a better job than the Multivariate Distribution, it still understates the probabilities in the tails. This is an important consideration for financial risk management, as tail events can have a significant impact on portfolio performance. Therefore, it is crucial to use models that can capture the tails of the distribution accurately.

In conclusion, this analysis highlights the importance of using appropriate models to capture the characteristics of financial data, especially when it comes to estimating the probability distribution of portfolio returns. The results suggest that the Gaussian Copula is a better model than the Multivariate Distribution, but it is not perfect and should be used in conjunction with other methods to provide a more accurate estimate of the distribution.

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated