# Appendix

March 11, 2023

This document contains detailed description about the correlation and portfolio modelling for Carapace pools.

## 1 Correlation Analysis

For modelling the correlation of two lending pools, we must first define what correlation is. By correlation we intuitively mean the similarity of two pools. And to describe it more mathematically, it describes the jump in probability of default of one pool given the other has defaulted.

For getting some estimate of the correlation, we need to do the analysis of the constituents of the lending pools. The constituents of the lending pools refer to the companies they invest in, and since some company not returning the money would result in partial default of the lending pool, it is paramount to model the correlation of companies itself.

### 1.1 Company Correlation

For calculating the correlation between two companies A and B, we do the factor analysis of both companies.

#### 1.1.1 Factor Analysis

Factor analysis refers to looking at a company's fundamentals and assigning various factors about the company based on that information. Factors here could be the sector the company operates in, the geography the company operates in, and the business model of the company. So for every company, we analyze the fundamentals and assign factors to various discrete variables. Example - for Goldman Sachs, the factors could be finance (sector), USA (geography) and trading, investment banking and asset management (businesses). Note that here the business has multiple values.

#### 1.1.2 Company Correlation Calculation

For calculating the correlation between company A and company B, we compute the number of factors that are similar and divide that by the total number of factors. The following image contains the correlation matrix we got for a particular set of companies in Carapace pool 1.

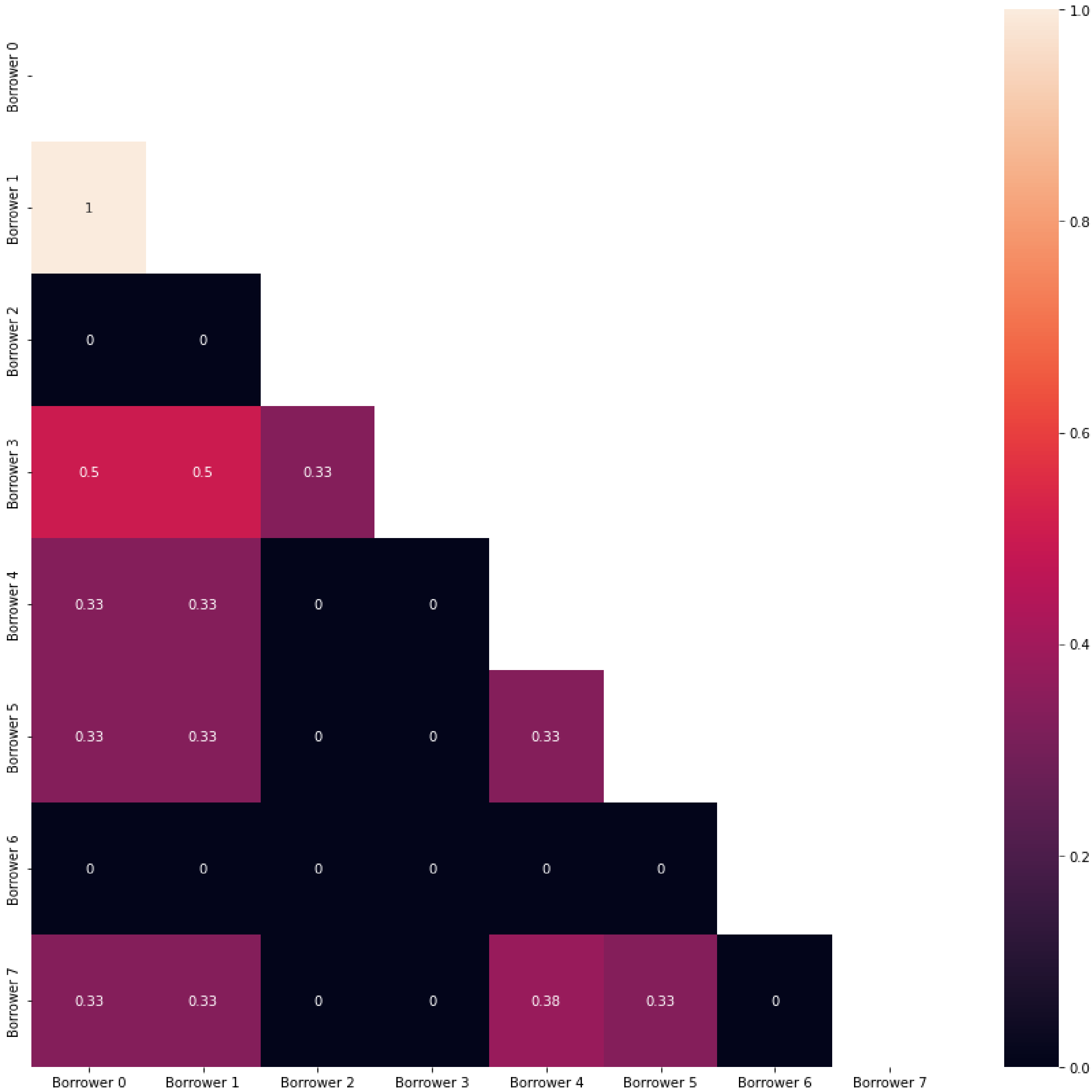


Figure 1: Company Correlation Matrix

## 1.2 Lending Pool Correlation

Given that we have correlation matrix for any company pair with us, the problem of finding the correlation between lending pool A and lending pool B reduces to a problem of finding correlation between a portfolio of companies A and portfolio of companies B.

Let the companies lending pool A invests is $A_1, A_2, .......A_m$ and the companies lending pool B invests is $B_1, B_2, B_3........B_n$, then the correlation of lending pools is given by

$$\frac{\sum_{m=1}^{i}\sum_{n=1}^{j}A_iB_j}{mn}$$

For each pool, you can find the resultant lending pool correlation matrix in the investor doc.

# 2 Portfolio Modelling

For getting various risk measures, we quantitatively model the portfolio of lending pools. We simulate the portfolio to come up with various risk measures.

## 2.1 Lending Pool Modelling

We take publicly available data regarding the financing rate, the payment dates and the amortization type for a lending pool to its investors, and model a stream of payments which we can use. So for a given lending pool with a financing rate 10%, payment dates every 6 months from now, and bullet amortization, we get a stream of payments equal to 0.1 dollars for every 1 dollar principal on each day, and a final payment of 1 dollar on the last day.

## 2.2 Credit Riskiness Modelling

Modelling the riskiness of a lending pool is a tricky thing, and hence we only imply a risk factor based on the financing rate. We use the Carapace premium formula to imply a risk factor. For example if the financing rate is 15% and the tenor of the lending pool is 360 days, then the risk factor is implied by the following formula.

$$1 - e^{(-\lambda*360)} = 0.15$$

where $\lambda$ is the risk factor.

The choice to imply the risk factor provides us a way to model various riskiness regimes. If we for example take the interest rates for each lending pool in Goldfinch, then the risk factors are implied according to how the Goldfinch protocol interprets it to be. One can also assign the same interest rate to each lending pool, and have all loans equally risky.

## 2.3 Default Probability

We model the default of a lending pool using a Poisson process, as done in most financial institutions. Hence the probability of survival between time t1 and t2 is given by $e^{-\lambda*(t2-t1)}$. One can see that a higher risk factor means the survival probability is less (i.e pool is more riskier)

This also has a property that the probability loan defaults in our models over its tenor is exactly equal to the interest rate. That means a lending pool having 15% rate will have a 15% percent default probability in our simulations.

## 2.4 Correlation Incorporation

For incorporating the correlation matrix for each lending pool, we do the following.

Assuming pool j defaults, we increase the risk factors of other pools based on the correalation matrix. The increase is defined as -

$$\lambda_i = \lambda_j * (1 + c(i, j))$$

where $\lambda$ is the risk factors, j is the defaulted pool and c(i,j) is the correlation between i and j derived earlier.

## 2.5 Simulations

We simulate the entire portfolio multiple times. On each simulation date, we use the survival probability distribution to sample defaults, and then subsequently use the correlation matrix to increase the risk factors of other loans, increasing their subsequent default probabilities.. As we do this multiple times, for each time we get the dates on which pools defaulted. And using those dates, we can get the missed payments for each pool. The sum of the missed payments across all pool will be the capital loss of the investor. Note here by the investor we refer to a person who puts money into a portfolio of all the lending pools, equally divided among all. We also intend to provide analytics for putting money into the portfolio unequally, based on how uneven the demand for protection is.

## 2.6 Analytics

We provide the simulation results in four different risk and correlation regimes. The result of a simulation is a plot with x-axis containing the capital loss percentage and the y - axis representing the probability of that much capital loss percentage. Capital loss percentage is defined as the amount of missed payments divided by the value of the investment. Again to reiterate, the capital loss here refers to the missed payments for an investor putting money equally into all the lending pools. In real life, when a company defaults, one is able to recover some amount, and so the capital losses will be less. Nevertheless, we do our analysis independent of what the recovery amount will be.
Apart from this we provide order statistics for each distribution the 10 percentile, 25 percentile, 50 percentile, 75 percentile and 99 percentile values of the capital loss distribution, apart from the mean capital distribution. This allows an investor judge the credit risk of the portfolio better.

### 2.6.1 Low Risk Regime

This regime corresponds to the case when the interest rate for each loan is set to 5%. This means each loan has 5% probability of defaulting over its tenor. This might be still riskier than real world scenarios as till now there has been no default in Goldfinch. Yet, we believe that this case is closest to the real world scenario given historical records.

### 2.6.2 Normal Regime

Here we imply the risk factors from the financing rate for each lending pool in Goldfinch. This regime takes into account the relative riskiness of each pool, and simulates a more riskier regime for the investor.

### 2.6.3 High Risk Regime

This regime is a high stress regime. Given our correlation effects come into picture only when a default happens, assigning a 20% equal rate to all pools makes all of them equally risky - probability of default 20% over the tenor (which is pretty high). This regime has the effects of the correlation

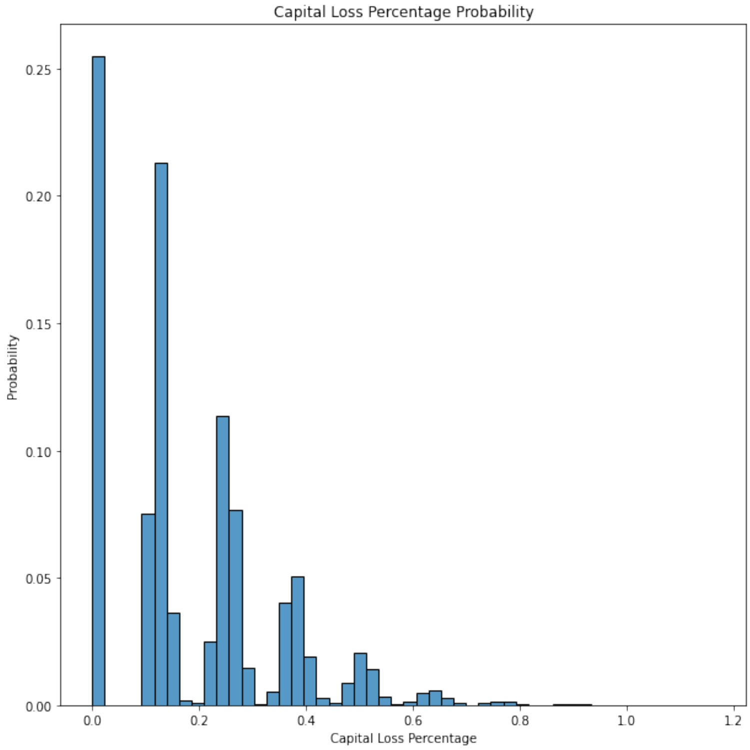


Figure 2: Sample Result - This picture represents the plot used to assess riskiness. The height of the bar at any point refers to the probability of an investor that amount of capital. For example, in this image the probability of 0 capital loss is close to 25%. From a high level POV, the more curve is right tail oriented, the higher the chance of huge tail losses (this should usually happen in extremely correlated portfolio)

matrix the highest, given risk factors are pretty high and the probability of further increases is even higher.

## 2.7 Uncorrelated High Risk Regime

This is an again high regime event, but we assume all lending pools are uncorrelated with each other. We assign a 20% equal rate to all pools, and this regime can be compared with the correlated high risk regime to understand the effect of the correlation matrix. This allows investors to see how correlation increases the tail riskiness of the portfolio.