Week 5 Report

Kristina Schiffhauer, Sagar Dubey, Hewitt Trinh, Jinwoo Oh, Stefanie Montgomery

This week we attempted to improve investment strategies from last week by framing them as optimization problems. With the variables-objective-constraints framework, we explored several ideas to construct the portfolios for Jasmine and compared the expected returns of the optimal solutions. While details of each construction are discussed in details in the following sections, we observed that:

* Optimization was not only a tool that resulted in better returns but also **allowed further customization in constructing loan portfolios to fit an investor profile**.
* We optimized the predicted return of the portfolio over various combinations of constraints such as maximum budget, loan number range, and portfolio diversification (grades), and found that the return averaged in the range of 5.4% - 6.73%. Interestingly, **a model with budget constraint performed *worse* than when the portfolio-diversification constraints were added**. It goes to show the value of portfolio diversification, although such a portfolio without any budget constraint may not be feasible for the investor.
* We also predicted the **standard deviation of a loan to maximize the ‘risk-adjusted’ return of the overall portfolio**. The risk-adjusted return of the portfolio performed similarly to the model with budget and diversification constraints.
* We discovered that overall, changing the number of loans picked, the number of loans required from each grade, and the budget did not have a large impact on the rate of return. However, we did see some small tendencies. The first is that the return rate increased until around a budget of $700,000, and then flattened off. We also see that imposing a minimum threshold of the number of loans from each grade has increasing pay-off until around ten loans, followed by a decrease in the payoff as the threshold increases.
* Varying optimization constraints significantly altered the mix of loans picked for our portfolio. **We applied Principal Component Analysis to explain the results of our optimization models**.

# Optimization Models

We used binary integer programming to build our optimization model. The decision variable was whether or not to select a loan as part of our portfolio. As discussed earlier, the objective was to maximize predicted return of the portfolio. On top of the constraint of a max and min loans of 100 and 50 respectively, we added the following combinations of constraints:

* No constraint (NC) - We ran models without constraints to compare different models’ performance. We did not completely remove all constraints but we kept constraints for the min/max number of loans which was applied to all other optimization models.
* Budget Constraint (BU) - The budget constraint we initially set was 1 million. In the process of setting budget constraints, we utilized the multi-knapsack formulation by linking budget amount to our decision binary variable to select loans within the budget we defined. We found out that budget constraints significantly dropped the maximized profit and returns compared to no constraints (from 6.89% to 5.41%) and selected loans decreased to 61 from 100. This makes sense since loans with high returns tend to have high loan amounts.
* Portfolio Diversification or Grade Constraint (GR) - To diversify our loan portfolio, we applied grade constraints. The constraint we applied was to have at least 10 loans for each grade. We first anticipated that grade constraint would lower the number of loans and maximized profit further but this was not the case. Interestingly, all 100 loans were selected even though the amount of maximized profit slightly dropped.
* Budget and Grade constraint (BG) - After exploring each grade and budget constraint, we applied both constraints together to see how it performed. As expected, the maximized profit was the lowest, but interestingly the average return and the number of selected loans were higher than the model with only budget constraint.
* Risk Adjusted Return (RA) - In the absence of a particular loan’s history, it is difficult to calculate its standard deviation and hence its risk. As a workaround, we ran k-means on the test set of loans and computed the standard deviation based on the predicted returns. We used the k-means trained model to predict the cluster of a new loan, and used the cluster standard deviation as a measure of risk of that loan. We used the ‘elbow-method’ to find out the optimal number of clusters (which turned out to be 100 in this case). We then used the standard deviation of the loan to maximize the ‘risk-adjusted’ return of the overall portfolio. We used a control parameter ‘beta’ to adjust the risk-impact to suit the investor’s risk tolerance.

The objective function below maximizes the return of each loan adjusted by the inherent risk of that loan.



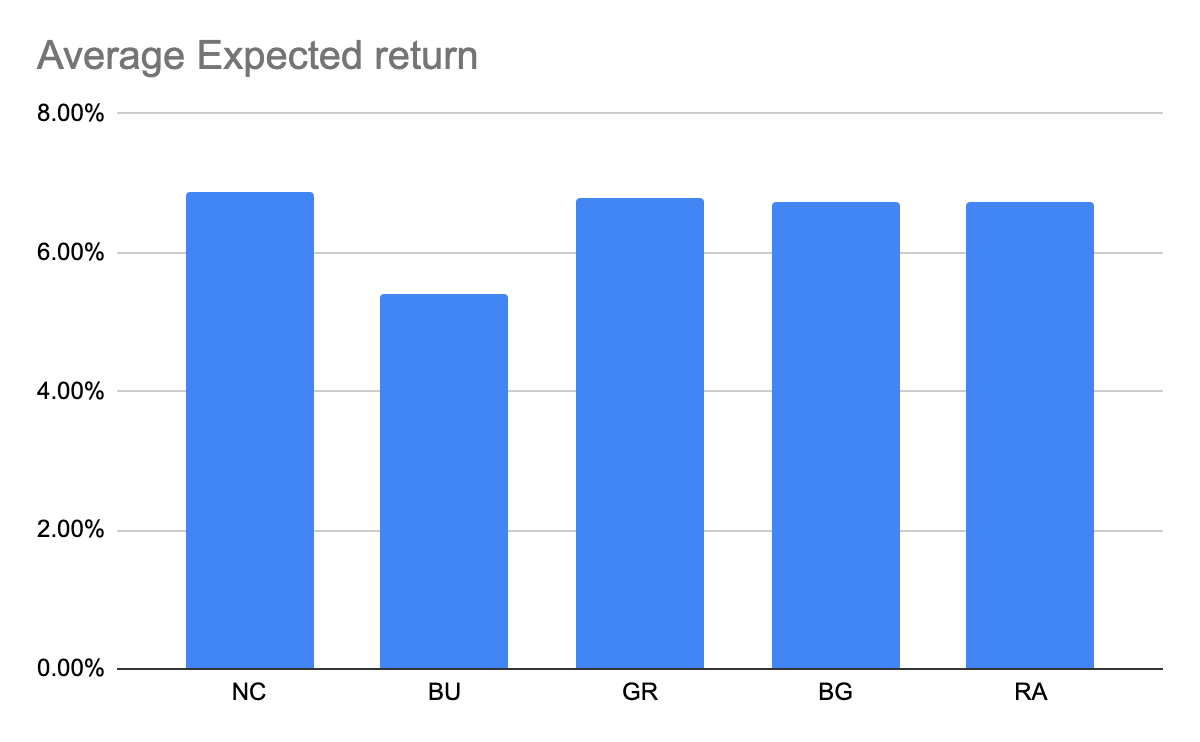
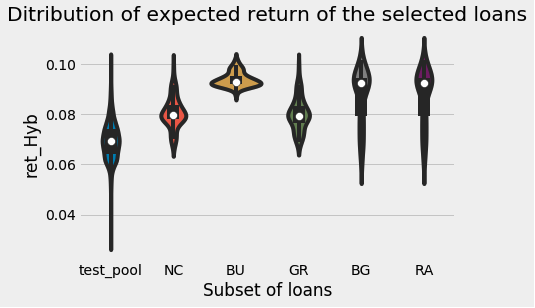
c[i] = return of loan i

v[i] = volatility or standard deviation of loan i’s predicted return

SelectLoans[i] = 1 if loan i is selected, 0 otherwise

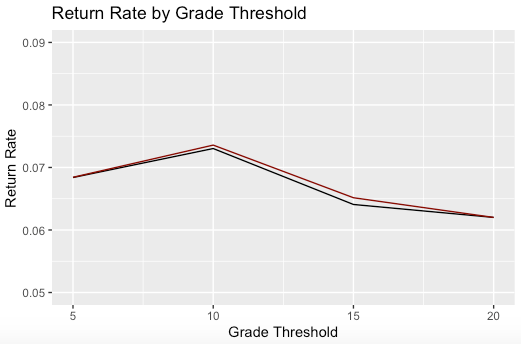
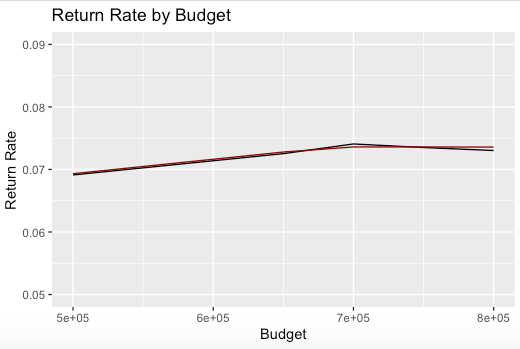
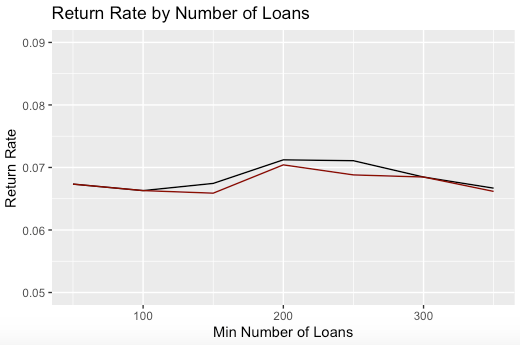
beta = control parameter / risk tolerance indicator

The distribution and average return of each of the above models are depicted below. As mentioned earlier, the portfolio with budget-only constraint (BU) delivered low average return, although its distribution is less. Tellingly, all the portfolios showed less return-distribution than that of the universe of loans (aka test pool).

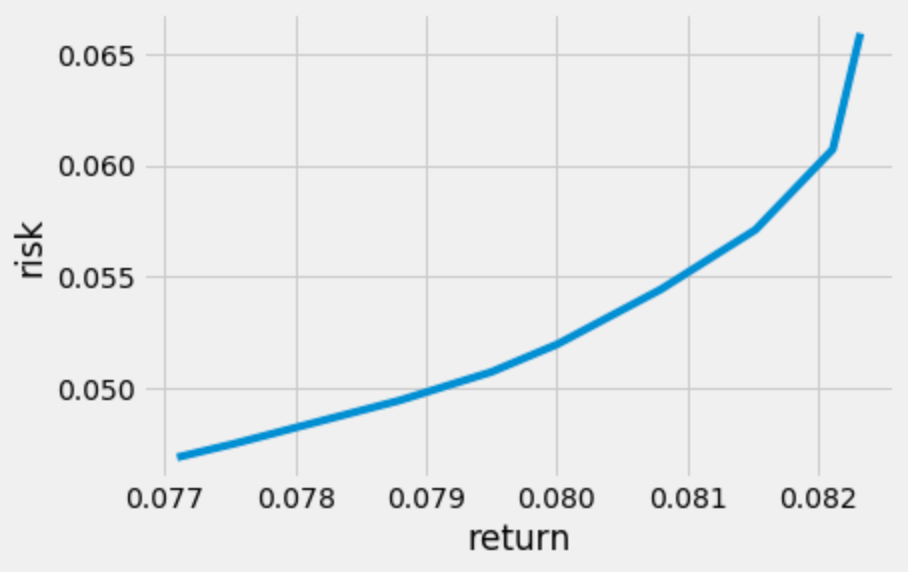


# Sensitivity Analysis

Below we see the results of our sensitivity analysis. Each chart shows the rate of return for our budget and grade constraint optimization model (black) and risk adjusted return model (red) against the minimum chosen number of loans, budget, and grade threshold, respectively. We can see that the variable with the largest impact is the grade threshold, which we imposed as the minimum number of loans picked from each grade. Each of these charts has increasing pay-off until a point, where it either decreases or levels off. Using these trials, we chose our optimal parameters to lie around choosing 200 -250 loans, with a $700,000 budget, and a minimum of 10 loans from each grade.



# Risk v Return

We ran the risk-adjusted optimization model with varying degrees of investor’s risk tolerance, using the ‘beta’ value to reflect the risk tolerance limit. As can be seen from the chart below, the average return increases as the investor is willing to take on more risks. However, an increase in risk from 5% to 6.5% allows the average return to increase from 7.7% to only 8.2%. So, the cost of undertaking higher risk may not be worth the benefit. 

# Model Interpretation

While optimal solutions from integer programming models can produce high-return solutions, they are also black boxes. We applied descriptive analytics to the results of our prescriptive models in an effort to uncover basic rules of our models.

The primary method of descriptive analysis we used was Principal Components Analysis. You will recall that we performed PCA several weeks ago in an effort to better understand the data and to uncover the loan grade structure. Here, we will apply PCA on the entire set of 10,000 test loans as the basis for a visual comparison of optimization strategies. We performed PCA using borrower characteristics only. The (rounded) loadings of the first two principal components are shown below.

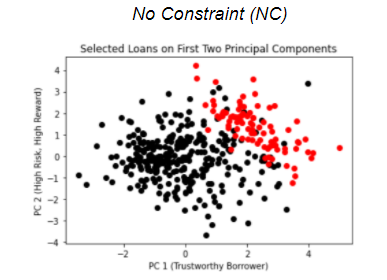
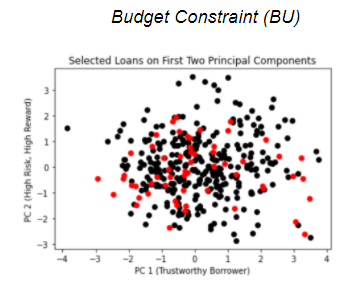
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | fico\_range\_high | int\_rate | annual\_inc | loan\_amnt | emp\_length |
| PC 1 | **0.39** | -0.26 | **0.64** | **0.56** | 0.24 |
| PC 2 | **-0.54** | **0.69** | 0.20 | 0.42 | 0.10 |

Based on our understanding of borrower characteristics and how they correspond to loan performance, we can refer to the above principal components as:

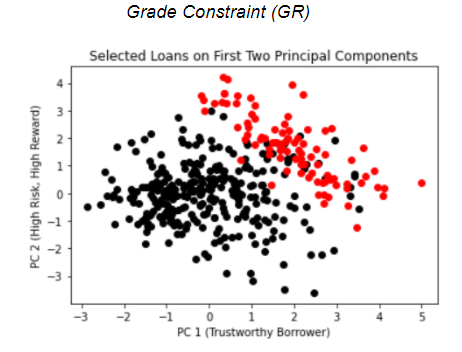
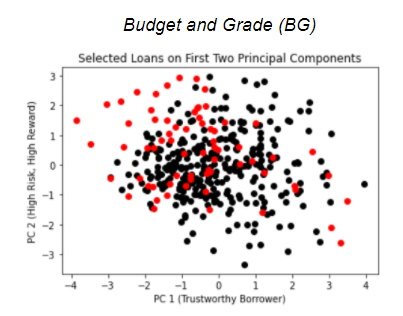
* PC 1 - “Trustworthy borrowers”: This PC is characterized by high annual income, high loan amount, and relatively high fico score. We have seen these borrowers associated with low-default, safe loans. However, we may see that a budget constraint prevents us from picking too many of these expensive but safe loans.
* PC 2 - “High risk, high reward”: This PC is characterized by very low fico score and high interest rate. As we have seen previously, these loans are the most likely to default but can be very lucrative if the borrower does pay back the loan in full.

The below scatter plots show a sampling of loans we have selected (red) and not selected (black) for each optimization model plotted against the first two principal components. In other words, the red points can tell us characteristics of the kinds of loans each model is choosing and the black points can tell us which loans the model avoids.

The **No Constraint** model allows us to pick loans across the riskiness spectrum, and purely optimize for expected return. It turns out that investing in a mix of safe but expensive (no budget constraint) and risky loans satisfies the objective. The **Budget Constraint** model shows loans that don’t appear to be particularly safe, expensive, nor risky and rewarding. One might interpret this as the budget constraint forcing us to invest in “middle of the road” loans across both PCs shown.

**

The **Grade Constraint** model, without any budget constraint, has us again investing in a strong mix of loans. The grade constraint encourages diversification of the portfolio, which as we saw earlier, leads to higher returns. The **Budget and Grade** model, again due to the budget constraint, produces a cloud of loans skewed to the negative of the first PC. This makes sense, as we expect the first PC to consist of expensive loans and the budget constraint prohibits us from investing in many of these. It is worth noting that this model has a higher expected return than the Budget Only model shown above - this again proves the value of portfolio diversification.

**

Finally, the **Risk-Adjusted** model shows loans skewed toward the center of both principal components. This is expected along PC2, which we can think of as measuring riskiness. If we removed the budget constraint from this model, we hypothesize that we would see the red loans clustered more positively along the PC1 axis. As is, our budget is not high enough to see a cluster form on the positive PC1 axis.

