2018 INFORMS O.R. & Analytics Student Team Competition – ENTRY FORM

Entry Number: [2018ORASTC252]

Executive Summary (not to exceed 2 pages)

Team Makeup & Process

Framing the Problem

Data

1. The Structure of the Data

   The data from Principal consists of 3 parts which are timeseries data, riskmodels data, and result template data.

   - Time Series Data :

   - Riskmodels Data :

   - Result Template Data :

2. Data Pre-processing and Rescailing

   There are some difficulties in applying the data given from Principal directly to the model. Therefore, we present the following data preprocessing process. To make the data composed of three parts more flexible, a process of data preprocessing formed dataframe using the Pandas Library(http://pandas.pydata.org) in Python.

   In particular, as rebalancing is carried out, it is configured to extract not all time series data, but only the data needed for the iteration. The parameters used in model were extracted from data frames formed for each period, when each parameter (Alpha Score(α), Beta(β),4 Weekly Returns (r), etc) was set with the 'SEDOL' index as the key value. This dictionary data type is proper to consider a list of assets that may change every period.

   The parameter Omega (Ω) is specified as the covariance matrix for a given data. In this case, the index in the columns and rows are 'SEDOL', the key value of the time series dictionary derived from the above, and are constructed in the form of full matrix.

3. The Analysis of the Data

Methodology Approach & Model Building

1. Global Approach

   Formulation Approach : Before buiding the model, we tried to solve the Quadratic Constraint Quadratic Programming (QCQP) problem by using the IBM CPLEX, which is well known Optmization tool, to determine the difficulty of the problem. Since CPLEX can not deal with non-linear constraints,the given formulation is required to be reformulated. First, the constraints (9) are reformulated as follows:

$$
\begin{aligned}
(9^*) \quad & y_i \geq w_i, \quad \forall i \in N \\
& y_i \leq w_i + 0.999 \quad \forall i \in N \\
& 50 \leq \sum_{i \in N} y_i \leq 70 \\
& y_i \in \{0, 1\}, \quad \forall i \in N
\end{aligned}
$$

   The decision variables y_i is 1 if asset $i \in N$ is selected where w_i is bigger than 0. Because $w_i \leq 0.001$ is considered to $w_i = 0$, $y_i = 0$ if $w_i = 0$. The opposite is the same as well. Constraints (10) and (11) are also non-linear. Furthermore, if non-convex constraints are linearized, they significantly impact speed by increasing the number of variables to determine and the number of constraints to consider. In other words, it is very inefficient to reformulate Constraints (10) and (11), which represent the limits of Active Share and Tracking Error, respectively. Therefore, we propose a bisearch algorithm to adjust Active share and Tracking Error.

   Evolutionary Approach :

   Neural Network Approach :

2. Hybrid Approach

   When solving the problem with the formulation appraoch, we could not find a reasonable solution because to consider about 500 assets per iteration the number of decision variables to decide and the number of constraints to be satisfied are too high . Especially, constraints (9) which select cardinalities considering the target range on number of stocks make the problem difficult. In the case of Neural Network Approach, there is a great advantage in that a solution that does not get significantly out from the whole constraints and that has a good objective value can be considered at the same time. However, it is difficult to find a soluton satisfying the feasibility for all the constraints and the fact that the training time takes too much time.

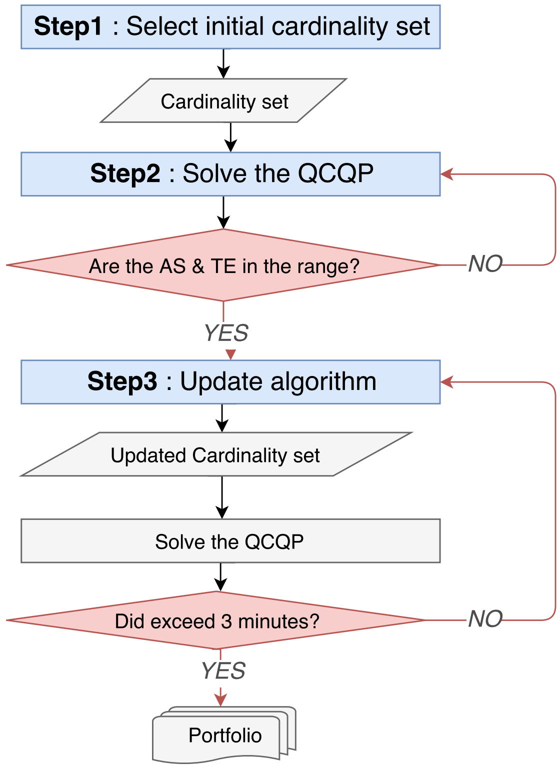

Figure 1: A flowchart of ...

Analytics Solution and Results

- This Entry Form: Present your solution and results in this section, including completion of the Portfolio Performance Statistics chart below. You may supplement your analysis with additional charts, diagrams and/or other visualization; these supplements must be incorporated into this section of the Entry From.

Portfolio Performance Statistics

- Results Template: Populate and submit your numerical results using the Results Template. This template is provided as a separate file on the Competition download site. You can use either the Excel or CSV version.

References

Please follow guidelines in the *Chicago Manual of Style*, 16$^{\text{th}}$ Edition. Here are examples:

- Journal article: Flynn J, Gartska SK (1990) A dynamic inventory model with periodic auditing. *Oper. Res.* 38(6):1089–1103.

2007-01-01 to 2016-12-31	**Portfolio**	**Benchmark**
Cumulative Return	%	%
Annualized Return	%	%
Annualized Excess Return	%	–
Annualized Tracking Error	%	–
Sharpe Ratio		
Information Ratio		–

– Book: Makridakis S, Wheelwright SC, McGee VE (1983) *Forecasting: Methods and Applications*, 2nd ed. (John Wiley & Sons, New York).

– Edited Book: Martello S, Toth P (1979) The 0-1 knapsack problem. Christofides N, Mingozzi A, Sandi C, eds. *Combinatorial Optimization* (John Wiley & Sons, New York), 237–279.

– Online reference, fictional example: American Mathematical Institute (2005) Better predictors of geospatial variability. Retrieved June 14, 2005, www.mathematicsinstitute.