MA324 Project report:
Operational Plan for The Amazoff Company

Candidate number: 56019

April - May 2023

In Part I, we present results and steps for implementation from a practical perspective. This is followed by Part II, containing technical appendices, where we will discuss the mathematical formulations of each of the problems. In all cases where there is data generation or simulation, the seed is set to 1 for reproducibility purposes.

Contents

Part I
Report

1 Executive Summary

Objective:

This report aims to present an analysis conducted on behalf of The Amazoff Company. We also recommend some alterations in their production, order preparation, and delivery policies. These changes are aimed at minimising costs and enhancing efficiency.

Findings:

Within Section 2.1, we outline the allocation of facilities to customers in a way that would minimise costs, whilst catering to each customer's demand requirements. The evaluation suggests the most efficient cost of operation is £190.96, inclusive of the capacity restrictions inherent to each facility. Additionally, we offer feasible strategies to achieve a more equitable allocation of facility duties, ultimately aiming to enhance the overall efficiency of the organisation.

Next, Section 2.2 details the most efficient routes for delivery vehicles to take. The analysis demonstrates the suggested route planning framework results in an additional cost of £98.59, regardless of whether one, two or three delivery vehicles are available. In fact, in all three cases, the optimal solution is to just use one delivery vehicle, and the optimal route remains the same.

Section 2.3, explores the problem of uncertainty. In particular, in the context of the internal depot transportation, items to be delivered have uncertain attributes. These issues and possible solutions are analysed.

Lastly, Section 2.4 looks into Amazoff Insurance, an additional service provided by The Amazoff Company. We estimate the probability that your capital drops below a certain threshold by the end a 12 month cycle, which would lead to legal complications. The estimate comes out to be 26.1%. At the end of the 12 months we believe, Amazoff Insurance will have £37614.67 in capital, with a high level of confidence that we are within £500 of its true value. We additionally offer an enhanced method of estimation for this to improve the estimate.

Conclusion:

The document presents valuable findings that could greatly benefit The Amazoff Company. By following the recommended plans for facility opening, delivery routes, uncertainty considerations, and legal compliance, The Amazoff Company could make significant cost savings, which would increase company profits.

2 Management report

This section will focus on the relevant steps for implementation to achieve the optimal outcomes.

2.1 Warehousing

The problem faced is a decision of how many and which facilities to open. It turns out that it is unnecessary to open all facilities to satisfy client demands. In particular by omitting facilities 1, 2, and 11, you can save on opening and renting costs. Refer to Table 1 for implementation details, and Appendix A.2.1 for technical details on the modelling process.

Next, we expand upon the simple model above to now account for both the client demand and facility capacity. Perhaps surprisingly, the earlier solution remains optimal, at a cost of £190.96. The reason behind this is due to exponentially decreasing opening costs for each additional facility opened. Simultaneously, assignment costs between clients and facilities are much greater. Hence, it is logical to open multiple facilities in order to reduce the distance between the clients and the facilities. Appendix A.2.2 contains more technical insight.

Finally, it would be beneficial to have the power to regulate the variation of demand experienced by your opened facilities. To motivate this, Figure 1 shows that the distribution is noticeably uneven as facility 5 only experiences 12 units of demand, whilst facility 10 experiences 105 units. For details on how to counter this, see Appendix A.2.3.

For a comprehensive table of results, refer to Table 2 and Table 3. Technical details for the data generation for facility and client locations, and client demand can be found in Appendix A.1.

Facility	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Demand experienced	0	0	68	103	12	73	43	82	35	105	0	127	75	57	17
Assigned Customer IDs			20 37 59 60	3 6 40 21 46 53 55	51	12 13 30 31 43 57	5 15 36	7 10 35 48 54 56 58	25 32 33	1 8 11 19 21 23 26 45		2 4 18 28 29 38 44 47 49	14 17 22 27 34 42 50	9 39 52	16 24

Table 1: Warehousing - Results for Implementation

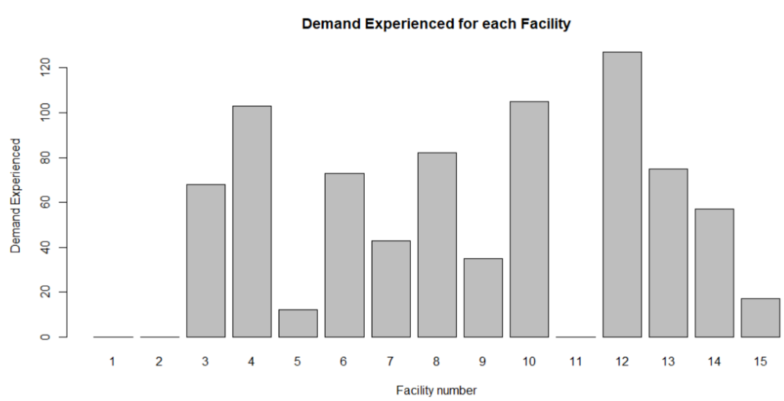


Figure 1: Warehousing - Plot of Demand Experienced by Each Facility

Limitations: Whilst the development of these models has improved upon the initial solution, there are a number of factors to consider for further enhancement.

The problem assumes demand for goods is known and fixed. Uncertainty is completely unaccounted for. The model is solely deterministic, with no consideration of any kind of sensitivity analysis. Despite the likely presence of minor variations in demand, the substantial capacity of the facilities allows us to maintain a high level of assurance in the dependability of our outcome.

Another concern, specific to this problem, is that it seems unlike that a factory experiencing 1 unit of demand to a given client costs the same as delivering (for example) 100 units of demand, due to additional fuel and administrative costs. Perhaps this is another area for future refinement.

2.2 Vehicle Routing

Now the focus is on an optimal delivery route from a set depot to each of the high priority clients, referred to as 'composite clients'. Note that here, demand is satisfied from products at the depot (now ignoring facilities). Delivery vehicles are assumed to have unlimited capacity, and we are permitted to use either one or three delivery vehicles for the chosen route.

Perhaps unexpectedly, the analysis conducted suggests the minimum total cost (and optimal route) is the same, regardless of if one or three delivery vehicles are used. This cost came to £98.59. See Figure 2 for a visual representation [1] of the optimal route.

Upon further thought, this is plausible intuitively speaking. As all utilised vehicles must start and end at the depot, the use of additional vehicles would incur unnecessary costs due to the potential for route overlap. For a technical discussion on this, refer to Appendices B.2.1 and B.2.2.

For a comprehensive table of results, refer to Table 2 and Table 3. Note that technical details on the data generation and simulations for selecting composite clients can be found in Appendix B.1.

The optimal route is:
$S \rightarrow c25 \rightarrow c32 \rightarrow c23 \rightarrow c11 \rightarrow c19 \rightarrow c50 \rightarrow c60 \rightarrow c52 \rightarrow c9 \rightarrow c43 \rightarrow c57 \rightarrow c22 \rightarrow c12 \rightarrow c49 \rightarrow c34 \rightarrow c54 \rightarrow c5 \rightarrow c36 \rightarrow c38 \rightarrow c44 \rightarrow c4 \rightarrow c16 \rightarrow c53 \rightarrow c10 \rightarrow c33 \rightarrow S$

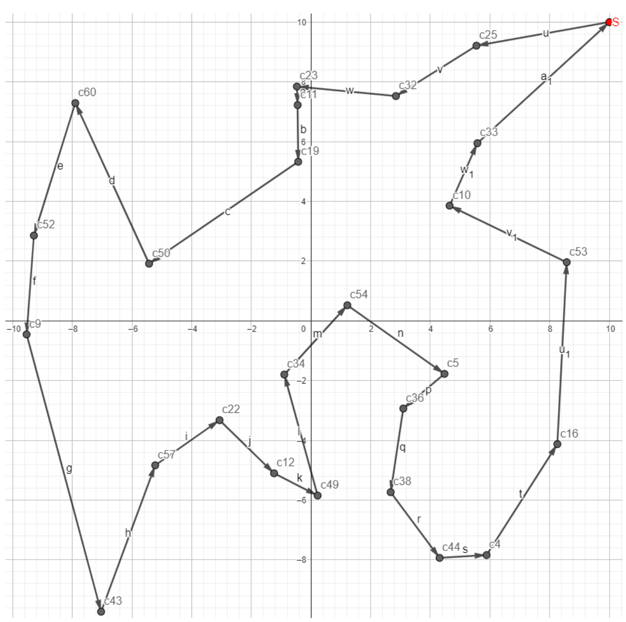


Figure 2: Routing - Recommended Route for Delivery

[1]The plot was produced from Geogebra

Limitations: The 'Travelling Salesman Problem' is NP-hard and computationally intractable for large problem sizes, making it difficult to handle increasing client numbers.

The problem also assumes the distance between any two cities is well-defined, symmetric, and fixed. However, in reality, this is unlikely, due to traffic conditions, road closures, or other factors.

Lastly, it seems unlikely that delivery vehicles actually have unlimited capacity. Therefore, it would be desirable to allow for more vehicles, and take that adjusted route.

Facility	x	y	**Facility**	x	y	**Facility**	x	y
f1	-4.69	-2.558	f6	-5.881	-6.469	f11	8.694	-5.757
f2	1.457	8.164	f7	3.74	-2.318	f12	3.033	-7.489
f3	-5.966	7.968	f8	5.397	-0.046	f13	-4.656	-2.278
f4	8.894	3.216	f9	4.352	9.838	f14	-9.732	-2.352
f5	2.582	-8.764	f10	-2.399	5.549	f15	7.394	-3.193

Table 2: Facilities - Locations

Client	x	y	Demand	**Client**	x	y	Demand
c1	-0.358	1.991	7	c31	-5.207	-8.821	13
c2	-0.129	-6.276	10	**c32	2.846	7.525	15
c3	6.547	3.369	15	**c33	5.578	5.946	13
**c4	5.885	-7.841	23	**c34	-0.895	-1.798	5
**c5	4.474	-1.775	6	c35	6.217	2.099	18
c6	6.419	2.941	23	**c36	3.094	-2.936	10
c7	5.659	1.061	24	c37	-4.595	9.854	20
c8	0.594	5.787	17	**c38	2.67	-5.736	13
**c9	-9.533	-0.455	16	c39	-7.413	-0.438	18
**c10	4.646	3.855	2	c40	8.481	1.975	25
**c11	-0.448	7.224	6	c41	9.523	4.636	10
**c12	-1.238	-5.104	5	c42	-2.865	-1.371	20
c13	-8.586	-8.011	18	**c43	-7.036	-9.738	24
c14	-3.675	0.373	10	**c44	4.311	-7.936	6
c15	3.24	-1.863	20	c45	-1.074	2.802	17
**c16	8.258	-4.128	13	c46	9.837	-0.088	20
c17	-0.819	-3.352	18	c47	-0.313	-6.531	3
c18	3.017	-4.84	25	c48	5.096	-0.922	19
**c19	-0.429	5.326	10	**c49	0.223	-5.849	11
c20	-8.315	7.506	20	**c50	-5.427	1.914	21
c21	-3.219	6.789	24	c51	1.497	-8.459	12
**c22	-3.066	-3.325	6	**c52	-9.289	2.856	22
**c23	-0.473	7.844	17	**c53	8.572	1.962	11
c24	7.287	-2.2	4	**c54	1.218	0.521	7
**c25	5.546	9.212	7	c55	9.702	0.153	2
c26	-1.307	4.25	10	c56	3.656	2.031	3
c27	-2	-3.493	1	**c57	-5.223	-4.837	8
c28	5.142	-5.946	10	c58	4.586	-0.949	13
c29	4.222	-7.566	22	c59	-6.497	4.934	17
c30	-5.09	-7.134	9	**c60	-7.9	7.291	11

Table 3: Clients - Composite (denoted by **), Locations, Demand

2.3 Internal Depot Transportation with Uncertainty

This section examines uncertainty in internal depot transportation, where delivery vehicles' attributes (e.g. volume and weight capacity) are known, but not all attributes of the items to be delivered are fully known. Two models are presented: one is always feasible and conservative, while the other is flexible and allows tailored uncertainty levels. Both models balance item profits with capacity constraints and account for variation as a risk representation.

The recommended and more flexible model (like the 'Balanced Loads' solution in the warehousing section) allows the firm to have a degree of control over the amount of variation (and hence the risk and uncertainty) allowed. On the other hand, the conservative model does not give this kind of adaptability, which could be seen as a disadvantage if control is desired. Ultimately, the model chosen heavily depends on company priorities. For example if the Amazoff Company wants to increase potential profits, it may be more desirable to use the flexible model. Although of course this involves a pre-specified degree of risk. On the other hand, to be on the safe side (for example, you do not want to disappoint any customers by not delivering all the items), then the conservative model is better.

Limitations: A limitation of both models is assuming items can be 'squeezed' into available volume capacity, which is unrealistic for some types unless they are soft and non-fragile.

A different method for more accurate results is to incorporate uncertainty using a scenario based approach. This would involve generating a set of possible scenarios to represent the possible realisations for the uncertain parameters (volume and weights of each item). However, this would be computationally expensive. Alternatively, a computationally cheaper solution is to run simulations to estimate the uncertain parameters. Once these estimates have been obtained, the problem effectively has been converted from a stochastic to deterministic nature.

2.4 Product Specific Insurance

In this final section, we exclusively explore The Amazoff Insurance branch. To aid budgeting plans and other related management decisions we consider three components, as detailed below: likelihood of legal complications, forecasting of capital stock, and improved estimation techniques.

Legal complications

We first deliver the likelihood of legal complications, where legal complications arise when capital drops below £30000 at the end of the 12 months. Using $K = 1000$ simulations, we estimate it to be at 26.1%, which should be a cause for concern.

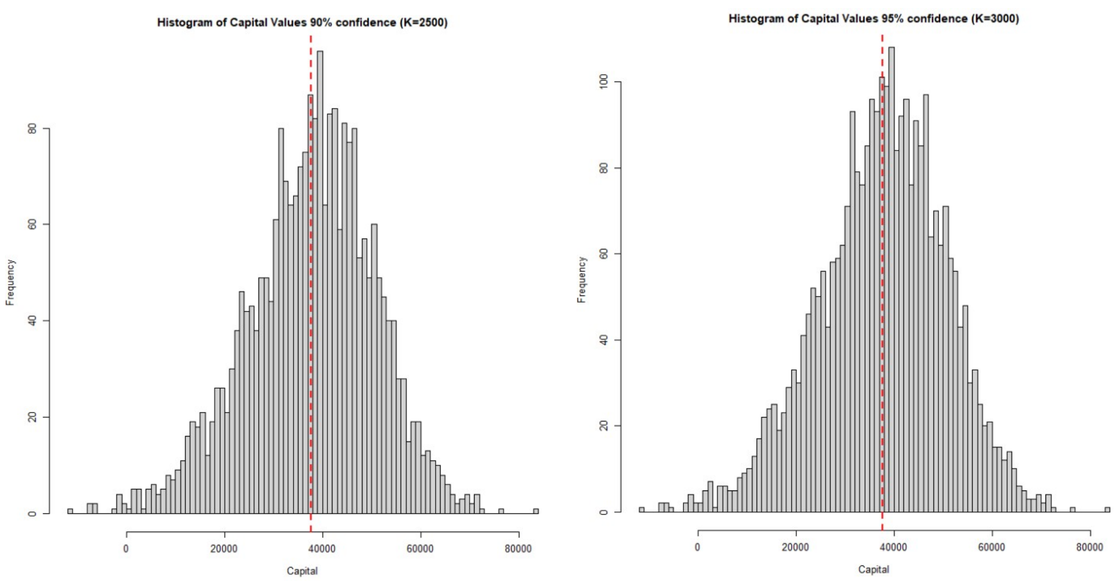
A potential solution is raise the monthly membership fee M from the current value of £300 to £350. By re-running the simulation with this modified parameter, likelihood of legal complications drops drastically to 10.5%. This would be a relatively easy fix, and would not be too unreasonable for the customers to accept. It should be noted though that raising the monthly rate too much would likely cause a drop the in arrival rate, and a potential rise in the departure rate, which in turn will increase the legal complications likelihood again. See Appendix D.1 for technical details.

Capital forecasting

To detail more context and useful considerations, we present our forecast of capital stock remaining for the end of the 12 months, which is at £37740.80. More specifically, this represents our estimate with 90% confidence we are within £500 if the true value. This required $K = 2500$ simulations. The 95% version of confidence within £500 is at £37614.67 and required $K = 3000$ simulations. These (and further results) are summarised in Table 4. The distributions of average simulation results can be found in Figure 3. See Appendix D.2 for technical details on how these estimates came about.

Measure (£)	90% confidence	95% confidence
Estimated $E[C12]$ within £500 of true	37740.80	37614.67
Lower bound of confidence interval	37313.57	37151.92
Upper bound of confidence interval	38168.03	38077.41
Difference in upper and lower bounds	854.47	925.49
Variance of estimate for $E[C12]$	12986.96	12931.60
Number of simulations K	2500	3000

Table 4: Summary of the simulation results for 90% confidence



(a) 90% confidence, K=2500

(b) 95% confidence, K=3000

Figure 3: Insurance: Distribution of sample means with different confidence levels

Enhanced estimation
We are aware that it would be more useful to have more certainty of correctness for the capital forecasting, particularly as the Insurance branch are small and hence likely more susceptible to shocks. Therefore we have developed an improved methodology for estimating the capital stock that reduces the variation in the estimate. See Appendix D.3 for technical details.

Limitations: To improve estimations further, we can run simulations with even more iterations. However, since the number of iterations is already high, the marginal improvement in estimates would not be significant, although confidence interval ranges should reduce further.

A key assumption made in this modelling and simulation is that customers will be charged for the full month from when they join (and until the month they leave). However, in reality, this sounds unfair and it would be more likely customers will be charged on a pro rata rate. This could therefore be a factor to consider in a development.

Additionally, currently Amazoff Insurance charges all customers the same fixed amount M, independent of age or type of technology product. However, as Amazoff Insurance grows, it would seem more reasonable to have different rates for different groups of people. For example younger clients may use their devices more, making the rate of accidents and rate of claims greater. Then these clients should be charged more.

Part II
Appendix

A Warehousing

A.1 Simulations and Data Generation

A.1.1 Client and Facility Locations

`Related files: 1_warehouse_gen_locations.R`

Before any modelling, we randomly (uniformly) generate the locations of the 15 warehouses, and the 60 clients. Each location and their respective distance from one other will be used to calculate costs in later parts of the problem. The methodology used for location generation is outlined below.

Algorithm 1 Generating Client and Facility Locations

Summary: Randomly generate coordinate locations of clients and facilities

1. Initialise empty vectors for the x and y location coordinates for all 75 (50 clients + 15 facilities)
2. Random generate numbers using uniform distributions between 0 and 1
3. Obtain final coordinates by re-scaling to make them each between -10 and 10 (inclusive)
4. Combine the x and y values to form coordinates and extract the first 15 coordinates as facility locations, and the remaining 60 coordinates as client locations

---

Note that we complete step 3 by multiplying by 20 and subtracting 10. In the code, we also round the resulting values to 3 decimal places. Additionally, since coordinates are completely random and independent, we maintain randomness by simply letting the first 15 (x, y) pairs represent the facility coordinates, and the remaining 60 (x, y) coordinates represent the client coordinates

A.1.2 Client Demands

`Related files: 1_warehouse_gen_demand.R`

Client demand data is needed for the 'Capacitated Model', with demand uniformly distributed from 1 to 25 units (inclusive). We assume these to be discrete quantities. See method below.

Algorithm 2 Generating Client Demands

Summary: Randomly generate the demands of the clients

1. Initialise an empty vector to hold the demand values
2. For each demand value, generate a random number between 0 and 1 using a uniform distribution
3. Rescale the random number to be between 1 and 25 (inclusive) by multiplying it by 25, rounding down to the nearest integer, and adding 1

Note that if we do not round down in the recaling step, we could be generating a demand that is greater than 25, which is not permitted.

A.2 Mathematical Modelling

A.2.1 Simple Model

`Related files: 1_warehouse_1_simple.mod, 1_warehouse_1_simple.dat, 1_warehouse_1_simple.run`

Here, we need to decide on the optimal way of opening and assigning facilities to clients [2]. For this most simplistic model, we do not take into account the client demands and facility capacity.

Sets

- we have 15 facilities where each are denoted by $f \in F$
- we have 60 clients where each are denoted by $c \in C$

Parameters[3]

- b_f = associated opening costs for facility f
- a_{fc} = cost of assigning facility f to client c
- $p_c = (x, y)$ coordinate of client c
- $p_f = (x, y)$ coordinate of facility f

Decision variables

- δ_f = indicator (binary) for if facility f is opened
- α_{fc} = indicator (binary) for if facility f is assigned to client c

Objective

$$\text{minimise} \quad \sum_{f\in F} b_f \delta_f + \sum_{(c,f)\in C\times F} a_{fc}\alpha_{fc} \tag{1}$$

Constraints

$$\sum_{f\in F} \alpha_{fc} = 1 \qquad \forall c \in C \tag{2}$$

$$\alpha_{fc} \leq \delta_f \qquad \forall f \in F,\ \forall c \in C \tag{3}$$

The objective (1) will minimise the total costs. Note that there are two types of costs: facility opening costs, and assignment costs. Constraint (2) shows we must assign exactly 1 facility per client. Constraint (3) reflects that a client can only be served by a facility if that facility is open.

A.2.2 Capacitated Model

`Related files: 1_warehouse_2_capacitated.mod, 1_warehouse_2_capacitated.dat, 1_warehouse_2_capacitated.run`

Now we wish to account for both the client demand and facility capacity. We call this the Capacitated Model. For this, we will extend the previous model (Simple Model) to now introduce 2 additional parameter types and add 1 additional constraint. All the rest remains.

Additional parameters[4]

[2] Reference: MA324 Course Pack, page 53 on Facility Location

[3] Note that we are given $b_f = 100 \times 3^{-f}$, and a_{fc} is the l_1-distance between the coordinates of a client c and facility f. Additionally, we have the coordinate points p_c and p_f, as generated by Algorithm 1

[4] Note that we have values d_c from the generating the client demands earlier in Algorithm 2, and $c_f = 100 \times 2^f$

- d_c = demand of client c
- c_f = capacity of facility f

Additional constraint

$$\sum_{c \in C} d_c \alpha_{fc} \leq c_f \delta_f \qquad \forall f \in F \tag{4}$$

This additional constraint (4) enforces that for each opened facility (denoted by when $\delta_f = 1$), it can only supply up to and including its capacity for all the clients it is responsible for.

A.2.3 Balanced Loads Model

As a further model development, we now wish to have more control over the outcome, where the difference between the facilities facing the most and least demand (among the operational facilities) can be constrained. For such a model, we will adjust the Capacitated Model to include 2 more decision variables, 1 new parameter, and 3 additional constraints. See below for model progression details.

Additional decision variables

- e_{max} = maximum demand experienced by a facility
- e_{min} = minimum demand experienced by a facility

Additional parameter

- T = tolerance allowed

Additional constraints

$$\sum_{c \in C} d_c \alpha_{fc} \leq c_f e_{max} \qquad \forall f \in F \tag{5}$$

$$\sum_{c \in C} d_c \alpha_{fc} \geq c_f e_{min} \qquad \forall f \in F \tag{6}$$

$$e_{max} - e_{min} \leq T \tag{7}$$

The first constraint (5) guarantees that the demand experienced by the facility providing the most demand does not exceed e_{max}. Similarly, the second constraint (6) guarantees that the demand experienced by the facility providing the least demand does not fall below e_{min}. Finally, the constraint (7) enforces that the difference between these maximum and minimum demands is no larger than a pre-specified tolerance, denoted by T.

B Vehicle Routing

B.1 Simulations and Data Generation

B.1.1 Selecting Composite Clients

`Related files: 2_route_part1_composite.R`

Like before, the initial step before modelling was to generate some random data to be used in the model. Here, we take a random subset of 25 'composite clients' of the total 60 clients to act these composite clients who are prioritised for delivery. The methodology used is outlined below.

Algorithm 3 Determining which clients are composite clients

Summary: Generates a random subset of clients and exports their locations as a CSV file

1. Define a function that takes two inputs: N and R
   (N = total number of clients = 60
   R = size of the random subset that will be generated = 25)

2. The function creates a list of all clients from 1 to N inclusive, and randomly shuffles them.

3. Once the clients are thoroughly shuffled, select the last R clients from the shuffled list.

4. Returns the subset of composite clients as a list and sorts in ascending order (for readability).

Note that this does indeed take a random subset of size 25 from the original set of clients. Since the list of all clients is shuffled randomly, each client in the original set of 60 has an equal chance of being selected for the subset. [5]

B.2 Mathematical Modelling

Note that in the following routing problems, we will only focus on the 25 composite clients [6].

B.2.1 Routing with 1 Vehicle

`Related files: 2_route_1_simple.mod, 2_route_1_simple.dat, 2_route_1_simple.run`

The initial model concerns 1 vehicle which must start at the depot, deliver to each and every composite client, and end back at the depot. As such, a version of the travelling salesman problem (TSP) formulation should be used. However, there are many different approaches for the TSP, such as edge-based formulation, node-based formulation, sub-tour elimination, which give different formulation variations. In particular, there is the Dantzig, Fulkerson, and Johnson (DFJ) formulation, the Gavish and Graves (GG) commodity flow formulation, and the Miller, Tucker, and Zemlin (MTZ) formulation [7]. We must select the most appropriate formulation for our needs.

We chose to incorporate the last of these methods, MTZ. DFJ was eliminated as it involves a very large number of constraints and variables, making it the most computationally expensive method. GG was also eliminated as it is not the easiest to extend to a multiple TSP, which we will need to do later. Hence, MTZ was selected.

Note that MTZ involves considering subsets of cities that form subtours. The decision variables in this formulation are binary variables that indicate whether a particular subtour is used or not. Additionally, with regards to the objective function, the only cost is travelling costs (calculated byt the l_2 norm). Therefore, the objective function (8) is to minimise this.

Sets

- we have the set of depot(s) D[8]
- we have the set of 25 composite clients C_{com}
- we have connecting nodes $(i, j) \in V = D \cup C_{com}$ [9]

[5] Reference: MA324 Week 6 Lecture, 'generate_subset.R'
[6] Note that composite clients were selected in Algorithm 3
[7] Reference: MA324 Week 4 Notes, 'Remarks on Modelling the TSP'.
[8] Note: in the specification, there is only 1
[9] Note $i \in \{1, ...n\}$ and $j \in \{1, ..., n\}$, where n denotes the cardinality of set V

Parameters

- c_{ij} = cost of travelling from node i to node j
- n = cardinality of set V

Decision variables

- x_{ij} = indicator (binary) for if there is a vehicle travelling from i to j
- u_i = index (natural number), based on the order composite client i comes in the route (e.g. for a tour 1 - 4 - 2 - 3 - 1, then $u_1 = 1, u_2 = 3, u_3 = 4, u_4 = 2$)

Objective

$$\text{minimise} \sum_i \sum_{j \neq i} c_{ij} x_{ij} \tag{8}$$

Constraints

$$\sum_{j \neq i} x_{ij} = 1 \qquad \forall i \in \{1, ..., n\} \tag{9}$$

$$\sum_{i \neq j} x_{ij} = 1 \qquad \forall j \in \{1, ..., n\} \tag{10}$$

$$u_i - u_j + n x_{ij} \leq n - 1 \qquad \forall (i, j) \in V \ s.t. \ j \neq 1, i \tag{11}$$

Here, constraint (9) specifies that each node i in the set of nodes must be visited exactly once. Then constraint (10) specifies that from every node j in the set of nodes, there must be exactly one outgoing edge. Lastly, the constraint (11) is known as the subtour elimination constraint. It ensures that the solution does not contain any subtours, which are loops that do not include all the nodes in the set of nodes.

B.2.2 Routing with q Vehicles

`Related files: 2_route_2_extended.mod, 2_route_2_extended.dat, 2_route_2_extended.run`

Now we have exactly 3 vehicles which must all start and end at the given depot. Between the vehicles, they must deliver to each and every composite client. Like previously, there are no vehicle capacity constraints. Since we used the MTZ formulation for the routing with 1 vehicle model, it is now easy to extend the problem earlier to transform it into the multiple TSP model by defining an additional parameter and adding some additional constraints, as outlined below.

Additional Parameter

- s = the starting node (the depot)
- q = number of delivery vehicles

Additional Constraints

$$\sum_i x_{si} = q \tag{12}$$

$$\sum_i x_{is} = q \tag{13}$$

Here, constraint (12) enforces that the delivery vehicles must start from the depot node. Then constraint (13) reflects ensures that the total number of delivery vehicles must end at the starting depot node. Hence these additional constraints make sure that all delivery vehicles start and end at the inital depot.

C Internal Depot Transportation with Uncertainty

In the previous section, the models assumed unlimited vehicle capacity. We will now address these concerns in the context of a different scenario in which the objective is to maximise the total profit. Here, we have multiple vehicles where the vehicle attributes (volume and weight capacity) are known. Items to be delivered from these vehicles have uncertain attributes, although the mean and variance are known. Vehicles are identical, and can contain multiple different types of items (possible several of the same type).

We propose two models, the first is the recommended model, and the second is a more robust approach. Whilst the recommended model is more flexible and adjustable (depending on your company's tolerance for variation in the item attributes), the robust approach is more conservative and will always be feasible.

C.1 Assumptions

- Items can be 'squeezed' into the available capacity for volume. This only seems likely if the items are soft and non-fragile, such as clothes, or bedding.
- Number of item types is finite. This should be true in any case.
- Mean and variance of the weight and volume of each item type are known. (Given)
- Maximum capacity (in volume) and total weight each vehicle can carry are known. (Given)
- Each vehicle can carry at most a given finite number of certain items (different limits for different item types) due to insurance concerns. (Given)

C.2 Recommended Model

Here is the modelling for the recommended model. It is the more flexible model that can be tailored this depending on your company's tolerance for variation in the item attributes. Note that it does not guarantee feasibility in every scenario.

Sets

- we have item types $t \in T$

Parameters

- V = vehicle capacity for volume
- W = vehicle capacity for weight
- p_t = profit of item type t
- N_t = limit on number of items type t that can be loaded onto a vehicle
- w_t = mean weight of item type t
- v_t = mean volume of item type t
- $\sigma^2_{w,t}$ = variance of the weight of item type t
- $\sigma^2_{v,t}$ = variance of the volume of item type t
- ϵ = allowance of variation

Decision variables

- x_t = number (a positive integer) of item type t assigned to vehicle

Objective

$$\text{maximise} \sum_t p_t x_t \tag{14}$$

Constraints

$$\sum_t v_t x_t \leq V \tag{15}$$

$$\sum_t w_t x_t \leq W \tag{16}$$

$$\sum_t x_t \leq N_t \tag{17}$$

$$\sum_t x_t(\sigma_{v,t}^2 + \sigma_{w,t}^2) \leq \epsilon \tag{18}$$

The objective function, given by (14), maximises profit through summing the profit of each item type multiplied by the number of items of that type loaded onto the vehicle.

Here, constraints (15) and (16) reflect the volume capacity and weight capacity of each vehicle. Then constraint (17) reflects limit on the number of items of type t that can be loaded onto a vehicle. These constraints arise as a result of the insurance policy, as outlined in the specification. Finally, to account for the uncertainty considerations, constraint (18) limits the variation by ensuring the weighted total variation does not surpass a pre-specified allowance, given by ϵ.

C.3 Conservative Model

This is the model for the more robust approach, where the solution should always be feasible. The model is fairly similar but uses more conservative constraints for vehicle capacity limits.
Sets

- we have item types $t \in T$

Parameters

- V = vehicle capacity for volume
- W = vehicle capacity for weight
- p_t = profit of item type t, for each $t \in T$
- N_t = limit on number of items type t that can be loaded onto a vehicle, for each $t \in T$
- w_t^{max} = maximum weight of item for each type $t \in T$
- v_t^{max} = maximum volume of item for each type $t \in T$

Decision variables

- x_t = number of item type t assigned to vehicle (a positive integer)

Objective

$$\text{maximise} \sum_t p_t x_t \tag{19}$$

Constraints

$$\sum_t v_t^{max} x_t \leq V \tag{20}$$

$$\sum_t w_t^{max} x_t \leq W \tag{21}$$

$$\sum_t x_t \leq N_t \tag{22}$$

Note that the objectve function (19) is of course the same. Here, constraint (20) is the more conservative version of the recommended model constraint (15). It is more conservative in the sense that it takes the maximum weight for that item type. The same approach is made for the new constraint (21) over old constraint constraint (16). These constraints serve as the uncertainty considerations. Lastly, nothing has or needs to be changed for the limit on the number of items of type t, as reflected by the constraint here (22).

D Product specific insurance

Before we commence on the specific simulations, we provide a brief overview of two discrete-event simulation progressions: "next-event time progression" [10] and "incremental time progression". The main difference is in how the simulation clock advances. The former identifies the next event and sets the simulation clock accordingly, while the latter advances the clock by a fixed increment.

Since we are told your insurance company uses a simple insurance model where clients are charged a 'fixed' and 'monthly' premium, we decided to make use of the incremental time progression, which gives more flexibility in time granularity. In particular, the simulation moves forward on a monthly basis. This ensures the accurate calculation of the capital, which is affected by the monthly membership premium paid. Note, the key assumption made is that customers are charged for the full month of when they join (and subsequent months they remain as customers), regardless of if they join on the 1st of the month, or the last day of the month.

D.1 Likelihood of Legal Complications Estimation

`Related files: 4_insurance_1_likelihood.R`

Here we estimate the probability of capital falling below C after 12 months using (as explained above) an incremental time progression discrete event simulation model, resetting monthly, reflecting the fee payment structure.

[10]Reference: MA324 Week 7, '11_4_Poisson_process.R', '11_5_Poisson_process_nonhomogeneous.R'
Reference: MA324 Week 8, 'time_next_arrival.R'

Algorithm 4 Estimating likelihood of legal complications

Summary: estimate likelihood company's capital will drop below C by the end of the year

1. Initialise parameters: n_0, c_0, M, C, λ, $\mu(t)$, α
2. Create helper functions: `num_hom`, `num_nonhom_leave`, `rv_bernoulli`, `rv_specific_discrete`
3. **for** $i = 1$ **to** K
   (a) Initialise: $n = n_0$, $c = c_0$, arrival_events, departure_events, accident_events
   (b) **for** month = 1 **to** 12
      i. Simulate: `num_join_month`, `num_leave_month` (use `num_hom`, `num_nonhom_leave`)
      ii. Update: n, c, arrival_events, departure_events
      iii. Initialise: `num_acc_month` = 0
      iv. **for** $j = 1$ **to** n
         A. Simulate: `num_acc_customer` (use `num_hom`)
         B. Update: `num_acc_month`
      v. Update: accident_events
      vi. **for** $k = 1$ **to** `num_acc_month`
         A. Simulate: `claim_made_customer`
         B. Update: c **if** `claim_made_customer` (use `rv_bernoulli`, `rv_specific_discrete`)
   (c) Record result: success or failure in the simulated year, based on c
4. Calculate: likelihood of legal complications

Customer arrivals and accidents are modeled as homogeneous Poisson processes with constant rates, while the departure process is a non-homogeneous Poisson process with a time-dependent rate function, mu(t). Helper functions, num_hom and num_nonhom_leave, simulate the number of interested events within a given time horizon. Claims follow a Bernoulli distribution with a 60% success probability. Finally, claim amounts are determined using a specific discrete random variable, rv_specific_discrete, yielding amounts in the form 300X + 500, where X is a uniformly distributed integer from 0 to 10.

The simulation iterates over K trials that each cover a span of 12 months. In each iteration, it calculates customer arrivals, departures, and accidents, updating capital stock accordingly. Claims are processed, and the capital is adjusted based on claim amounts. The algorithm computes the likelihood of legal complications as the ratio of failed trials, where capital falls below the threshold (C), to the total trials (K). A summary table of events for each month is generated to display the results.

D.2 Confidence of Estimations

`Related files: 4_insurance_2_confidence_90.R, 4_insurance_2_confidence_95.R`

In addition to having an estimate for the likelihood of legal complications, it would also be helpful to have an estimate for the mean capital the company has at the end of the 12 months. In this section, we focus on this. More specifically, we will find such an estimate for which we are 90% confident it is within £500 of its true value. See the process below in Algorithm 5.

Algorithm 5 Estimating mean capital within a tolerance margin

Summary: estimate E[C12] with a specified confidence that it is within £500 of the true value

1. Set `desired_margin_of_error` to 500
   Set `confidence_level` to 0.9
   Set `K_increment` to 500
2. Calculate the sample mean and sample standard deviation of capital
3. Calculate the actual margin of error for 90% confidence using the critical value, sample standard deviation, and sample size (K)
4. Check if the `desired_margin_of_error` is met:
   (a) If the `actual_margin_of_error` is $\leq$ `desired_margin_of_error`:
      i. Calculate the confidence interval bounds
      ii. Output the estimated $E(C_{12})$ with 90% confidence, confidence interval bounds, difference in bounds, variance of the estimate, and the number of iterations
   (b) Else:
      i. Increase the number of iterations by `K_increment`
      ii. Display a message that the calculated margin is too big and that the number of iterations has been increased. Ask user to rerun with this increased value of K

The key idea here is to produce an e.g. 90% confidence interval for the estimated population mean. Then we calculate the margin of error. If the margin of error is larger than the tolerance of being within £500, number of iterations (this can be interpreted as the sample size used) will be increased. We will then reconstruct the 90% confidence interval. This process will be repeated, until the tolerance requirement is satisfied. At te end, results will be shown.

To estimate the population mean ($E[C12]$), we use the sample mean. The use of this estimator is motivated by the Law of Large Numbers: as the sample size increases, the sample mean converges to the population mean. The variance of this estimator is: *sample variance* $\div$ *sample size*, as a result of the Central Limit Theorem.

D.3 Variance Reduction of Estimation

In this final section of analysis, we aim to reduce variance in the estimator for the sample mean of C12. We suggest using the variance reduction technique known as Control Variates. This method is based on using a known random variable (called the control variate) that is correlated with the random variable of interest. Note that the population mean of one of these random variables must be known, and correlations can be positive or negative. We can use the control variate to improve the estimate of the population mean by reducing the variance of the estimator.

In our case, using the amount claimed at the end of 12 months could serve as a control variate. This would be valid, since insurance claims are measurable and can easily be tracked. Hence we can can easily obtain an estimate for the population mean of amount claimed. Additionally, we would expect a strong negative correlation between this and our estimate of E[C12], since a larger total amount claimed (by the end of the year), which is a direct cause of capital reduction.