Markov Decision Processes for Services Opportunity Pipeline Optimization

Aurélie Glerum
Master project

Project supervised by:
Dr. Gianluca Antonini,
Information Analytics, IBM Research – Zurich

Professor: Michel Bierlaire,
Transport and Mobility Laboratory, EPFL

January 2010

Disclaimer

The data reported in this document are simulated based on various scenarios, and are used for illustration purposes only.

Abstract

The dynamics of sales opportunities can be modelled by a Markov Decision Process. The latter can be solved using dynamic programming and assigns to each state an optimal action. In this project, states are modelled by the number of opportunities at five different maturity levels called ranks, actions are represented by investments and rewards by profits from signed contracts. Transitions are simulated using the probabilities that an opportunity moves from one rank to another. Two different types of policy appear recurrently in the model outcome, i.e. a low-investment policy when the opportunities are rather uniformly distributed across ranks and a high-investment policy, when a larger number of opportunities have reached mature status or have just entered the pipe.

Acknowledgements

I would like to thank Dr. Gianluca Antonini for his help with this project and for making corrections to the present report, and Stefan Woerner for his help on the modelling part.

Contents

1 Introduction **5**

2 Data description **6**

3 Data splitting **7**

3.1 Tree structure 7

3.2 Average transitions 9

3.2.1 Existing opportunities 9

3.2.2 New opportunities 10

4 The model **11**

4.1 Theory of Markov Decision Processes 11

4.2 Application of the model to pipeline data 13

4.2.1 State space 13

4.2.2 Action space 18

4.2.3 Transition function 22

4.2.4 Reward function 25

5 Results **29**

5.1 Application of the model to the leaves of Tree 1 29

5.2 Variation of the dimensions of the discretised action space 41

5.2.1 Action space with 3 variables for I_0 and 2 variables for c 41

5.2.2 Action space with 2 variables for I_0 and 3 variables for c 45

5.2.3 Action space with 3 variables for I_0 and 3 variables for c 48

5.2.4 All historical actions considered individually 49

5.3 Variation of the number of possible occurences in each rank 52

5.4 Variation of the state space discretisation 54

5.5 Variation of the discount factor γ 55

6 Conclusion **56**

Chapter 1

Introduction

Pipeline data collect the information relative to the evolution of sales opportunities. The latter appear and disappear over time and between these two points in time, their statuses evolve through different maturity states. When a sales opportunity appears in the pipe, it is given a low-level maturity rank, which indicates that its likelihood to be won is still low. Over time, it gets mature and reaches higher ranks that make it more likely to be concretized. In the best cases, it is eventually won.

The opportunities are generated and maintained in the pipe due to investments injected at the beginning of each quarter, which translates, amongst other things, into sales workforce. When opportunities are won, contracts are signed and the signings produce a certain amount of money.

The aim of this project is to model the dynamics of sales opportunities from IBM data by a *Markov Decision Process (MDP)* in order to find an optimal policy of investments at the beginning of each quarter. To do so, we use the data from the opportunity flows, i.e. the number of opportunities at each point in time, together with the historical investments and signings. In an MDP framework, finding an optimal policy implies associating a best action to each state. In the sales opportunity context, this means performing an optimal investment depending on the number of opportunities and their status. Optimality is reached when some criterion is satisfied. From a business point of view, it would be relevant to maximize the amount of money obtained from the signings. Hence, given that a certain number of opportunities are in the pipe at a particular point in time, we would like to perform the best investments in order to maximize the rewards, i.e. the amount of money given by the signings.

The available data come from different world geographies. Moreover, they are split across three different business lines and eight different industrial sectors. Using this information, we will attempt to apply a model to each category. In this report, we will only cover the analysis for one particular region.

This work is split into four parts. First, we will give a brief description of the data. After this, we will explain how the data are split across several categories and perform some preliminary computations on the dynamics of the opportunities. The third part of this report provides some theory about the sort of model we use and then presents its specification in the context of sales opportunities. Finally, the last part of this report shows results of the application of the model to different categories of opportunities identified in the pipeline data.

Chapter 2

Data description

The data consist of *pipeline data* describing the evolution over time of sales opportunities. Precisely, a sales opportunity is rated according to the likelihood it might be won. There are five ranks describing the latter: **1**, **2**, **3**, **4** and **5**.

The first ranking, i.e. 1, indicates the very early stage of an opportunity and the last one, i.e. 5, means that the opportunity is won and hence, the contract signed. The opportunity evolution is observed over three years, denoted as Y1, Y2 and Y3, and each of them is divided into four quarters of 13 weeks. The data relative to the rank of an opportunity are collected every week.

Let us note that in this project, we only consider the data from the first quarter of year Y1 to the first quarter of year Y3 included. This is due to the fact that data from the three last quarters of year Y3 are incomplete.

The pipeline data are obtained for five different world regions. In this report, we will only present the results for the data from one of them. Additional characteristics of the opportunities are known, such as the *Business Line* or the *Sector* they belong to. There are three possible business lines, denoted as BL1, BL2 and BL3, and eight possible sectors, denoted as Sec1, Sec2, Sec3, Sec4, Sec5, Sec6, Sec7 and Sec8. For each opportunity, we also have an identifier for the salesman that is responsible for the opportunity.

Quantitative information about each opportunity are also provided. They are of three types:

Expected gain: For each opportunity at each point in time, the salesman has an expectation about the amount of money he can obtain in case the opportunity is won.

Signings: If an opportunity is won, a contract is signed and the *signings* indicate the amount of money which is obtained from the total won opportunities in a quarter. The signings are given by business line, by sector and by quarter.

Investments: We have available investment data by quarter/by business line. They combine salaries, commissions and travel expenses for executives and salespeople plus advertising costs and other items.

Chapter 3

Data splitting

This chapter first presents the repartition of the opportunities according to a tree structure. Then it explains the method of computation of transition rates from one rank to another. Statistics were performed on these transition rates but are not shown here.

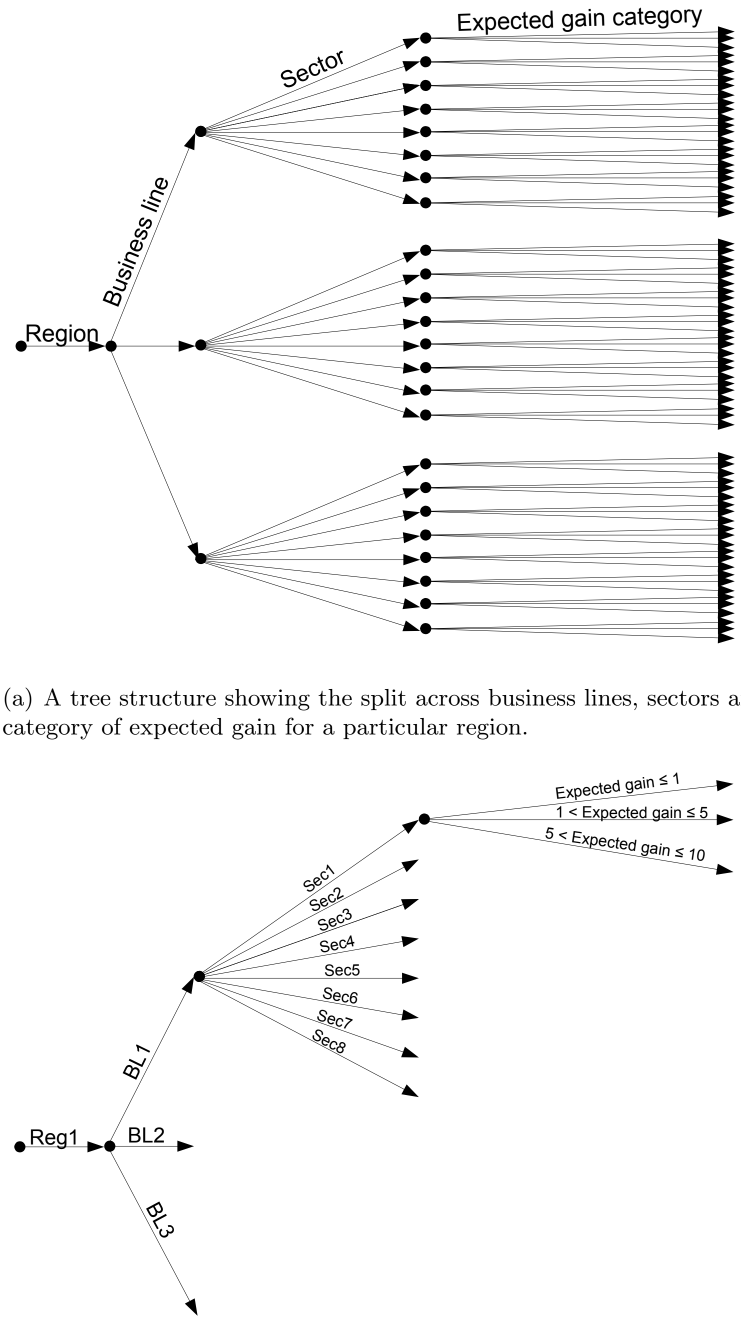
3.1 Tree structure

After performing some data cleaning, the opportunities are organised in a tree structure based on the geography and the industrial sector characterising the different opportunities as well as their business size. More precisely, we use three categorical variables in the data set 'region', 'business line' and 'sector', to naturally split the opportunities. A fourth 3-value categorical variable is built from the average expected gain for each opportunity as follows:

- Opportunities with an average expected gain below 1 million USD.
- Opportunities with an average expected gain between 1 and 5 million USD.
- Opportunities with an average expected gain between 5 and 10 million USD.

Opportunities with average expected gain values above 10 million USD were discarded.

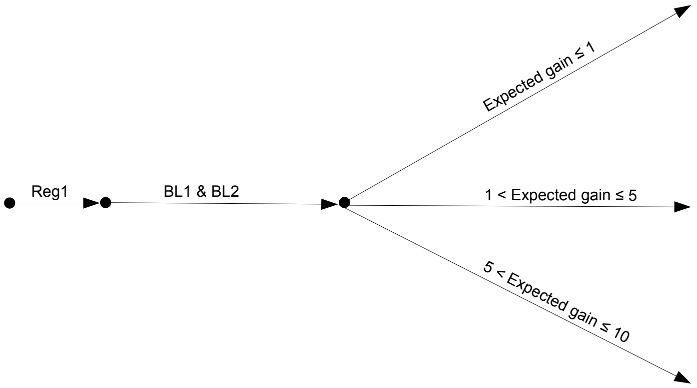
A complete tree for a given region is shown as an example in Figure 3.1. Several other trees can be easily generated aggregating the opportunities across one or more dimensions (e.g. the business line and/or the sector) as shown in Figure 3.2.



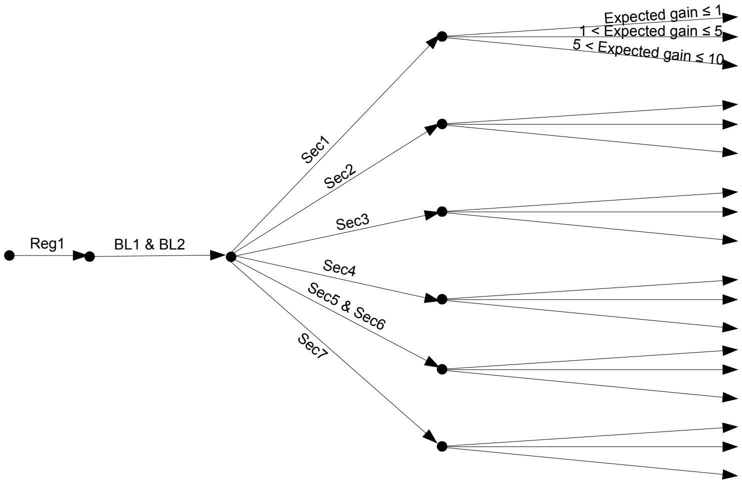
(a) A tree structure showing the split across business lines, sectors and category of expected gain for a particular region.

(b) Detailed version of the split.

Figure 3.1: A tree structure showing a complete split of the opportunities according to their business line, sector and category of expected gain in the example of the data from a region Reg1.



(a) Tree 1



(b) Tree 2

Figure 3.2: Tree structures showing the two splittings of the opportunities considered in this project for a region Reg1.

3.2 Average transitions

To get an insight into the dynamics of the pipeline data, we compute the opportunity transition rates from a rank to another using different time lags: 1-week, 2-week, 1-month and 1-quarter.

3.2.1 Existing opportunities

Let us denote one of the possible lags by l. The transition rates are always computed using the data from a current week w and the week at the previous time lag $w - l$, for $w = 1, \ldots, W$, where $w = 1$ refers to the first week of the second quarter of year Y1 and $w = W$ refers to the last (13th) week of the first quarter of year Y3. A transition rate from a rank i to a rank j, for $i, j = 1, \ldots, 5$ is obtained by counting the number of opportunities $N_{\text{prev}}[i]$ that were in rank i at a particular week $w - l$ and counting the number of opportunities $N_{\text{curr}}[i,j]$ out of the $N_{\text{prev}}[i]$ that moved to rank j at week w. The rate is then computed as:

$$r_{ij} := \frac{N_{\text{curr}}[i,j]}{N_{\text{prev}}[i]}.$$

Such rates are computed for all previous ranks i at a week $w-l$, to all current ranks j at a week w. Moreover, we also consider the transitions from any previous rank i to a sort of sixth rank called 'out', which represents the opportunities that were in rank i at week $w-l$ and that were not in the pipe anymore at week w. For each couple of weeks $(w-l,w)$, the transition rates r_{ij} can be represented as a 5×6 matrix $M(w-l,w)$:

$$M(w-l,w) = \begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} & r_{1\text{out}} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} & r_{2\text{out}} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} & r_{3\text{out}} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} & r_{4\text{out}} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} & r_{5\text{out}} \end{pmatrix}$$

Hence, for each leaf of a particular opportunity tree, we can create a list of transition matrices $M(w-l,w)$ of length W.

3.2.2 New opportunities

In an analogous way as for the lists of transition matrices for existing opportunities, we create lists of new opportunities, that is, opportunities that appeared at a particular week w in each rank and that were not in the pipe at the previous time lag $w-l$. For each transition matrix $M(w-l,w)$, there is a corresponding five-component vector $v_{\text{new}}(w-l,w)$ of the new opportunities in each rank:

$$v_{\text{new}}(w-l,w) = (n_1, n_2, n_3, n_4, n_5).$$

Chapter 4

The model

The aim of this project is to find the best investment policy for each state, that is, the investment that maximizes some objective function for each given state. To do so, we apply a *Markov Decision Process (MDP)* to the pipeline data. The first section of this chapter describes the basic concepts of an MDP.

4.1 Theory of Markov Decision Processes

The purpose of a Markov Decision Process is to find an optimal sequence of decisions, i.e. a *policy*, that maximizes a particular criterion given by an *objective function*. An MDP can have a finite or infinite number of stages, i.e. time steps, and this concept defines whether one deals with a *finite-horizon* or an *infinite-horizon* model. In this project, the infinite-horizon case will be considered. A justification will be given in section 4.2.

Definition 1. *A* Markov Decision Process (MDP) *is defined by the following four components.*

State space: *The state space S describes all the possible situations that can occur in the system. There can be finite or infinite state spaces, but in the model we present in this project, we will only consider finite state spaces.*

Action space: *The action space A describes all possible decisions that can be taken from all states in the state space. In our case, we will consider that every action a in the action space can be taken from any state $s \in S$.*

Transition function: *A transition function $P : S \times S \times A \longrightarrow [0, 1]$ is a rule that gives the probability that a state $s \in S$ is mapped to a next state $s' \in S$, given that an action $a \in A$ was chosen. A basic assumption of an MDP is the* Markov Property, *which states that the choice of the next state s' only depends on the current state s and the chosen action a, and is independent of the previously visited states.*

Reward function: *A reward function $R : S \times S \times A \longrightarrow \mathbb{R}$ determines the obtained reward when a transition is performed from state s to state s', given that an action a was chosen.*

The definition of the basic components of an MDP enables us to give more formal definition of a policy:

Definition 2. *A* policy *is a function $\pi : S \longrightarrow A$, that associates to any state $s \in S$ an action $a \in A$.*

The aim of an MDP is to find an optimal policy, that is, a policy that maps every state $s \in S$ to the best possible action. The search for an optimal policy implies some optimality criterion to be verified. One of the four components of an MDP is the reward function R which is defined for each triple $(s, s', a) \in S \times S \times A$ and the idea is to maximize the rewards in a particular way. If there was just one time step t, the aim would be to optimize the direct expected reward $E[r_t]$. But we are interested in finding a best sequence of actions, and hence we must optimize the rewards over the whole horizon. A possible optimality criterion is to maximize the expected sum of the discounted rewards, given by the following formula:

$$E\left[\sum_{t=0}^{\infty} \gamma^t r_t\right], \tag{4.1}$$

with $\gamma \in (0, 1]$ called discount factor. Its purpose is to give less importance to the rewards that are obtained far in the future.

We would like to define an expression such as equation (4.1) for each possible state $s \in S$. For a particular policy $\pi : S \longrightarrow A$, the latter is given by a *value function*:

$$V^\pi(s) = E_\pi\{\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t = s\}.$$

The value function equation can be defined recursively by the so-called *Bellman equation*:

$$V^\pi(s) = E_\pi\left\{r_t + \gamma V^\pi(s_{t+1}) | s_t = s\right\} = \sum_{s' \in S} P(s, s', a)\left(R(s, s', a) + \gamma V^\pi(s')\right) \tag{4.2}$$

For each state $s \in S$, the aim is to find the action $a \in A$ that maximizes the value function. Hence the optimal value function $V^*(s)$ for each state $s \in S$ is given as follows.

$$V^*(s) = \max_{a \in A} \sum_{s' \in S} P(s, s', a)\left(R(s, s', a) + \gamma V^*(s')\right). \tag{4.3}$$

The optimal policy π^* can be deduced from equation (4.3) taking the argument of the maximum:

$$\pi^*(s) = \arg\max_{a \in A} \sum_{s' \in S} P(s, s', a)\left(R(s, s', a) + \gamma V^*(s')\right) \tag{4.4}$$

In practice, in order to solve equations (4.3) and (4.4), we use an algorithm called *value iteration*, which solves the Bellman equation (4.2) by updating it using an intermediate *state-action value function* $Q : S \times A \longrightarrow \mathbb{R}$:

$$Q(s, a) = \sum_{s' \in S} P(s, s', a)\left(R(s, s', a) + \gamma V(s')\right) \tag{4.5}$$

The algorithm starts with an arbitrary value function $V(s)$ for each $s \in S$ and then computes the state-action value function $Q(s, a)$ for each $s \in S$ and each $a \in A$. For each $s \in S$, it updates the value function $V(s)$ taking the maximum of expression (4.5) over all possible actions $a \in A$ that are reachable from s. This update algorithm is performed until some convergence criterion is reached and is summarized in Algorithm 1.

Algorithm 1 Value Iteration

```
$V(s) \leftarrow 0, \forall s \in S$
repeat
  for $s \in S$ do
    $V_{\text{prev}}(s) \leftarrow V(s)$
    for $a \in A$ do
      $Q(s,a) \leftarrow \sum_{s' \in S} P(s,s',a)\,(R(s,s',a) + \gamma V(s'))$
    end for
    $V(s) \leftarrow \max_{a \in A} Q(s,a)$
  end for
until Distance between $V$ and $V_{\text{prev}}$ is small enough.
```

4.2 Application of the model to pipeline data

We use an infinite-horizon model, as the aim is to obtain an optimal action associated to each state independently of the time step. Even if the historical data collected from quarter 1 of year Y1 to quarter 1 of year Y3 are split into 9 quarters, we would like to obtain a decision model for a larger time scale.

The dynamics are modelled by the number of opportunities that transit from one rank to another. Using the historical data, we constructed the four necessary components to an MDP: the state space, the action space, the transition function and the reward function.

4.2.1 State space

In order to have a global view of the opportunity flow from one rank to another, we represented one state as a five-dimensional vector $s \in \mathbb{N}^5$, corresponding to the number of opportunities in each rank. For example, a state vector could be given by

$$s = (100, 230, 261, 83, 21), \tag{4.6}$$

indicating 100 opportunities in rank 1, 230 in rank 2 and so on.

This representation implies that the state space S_{tot} can be defined as a bounded subset of $\mathbb{N}^5$. Using this representation for the state space, the key issue is represented by the size of the state space itself, $|S_{\text{tot}}|$. The latter can be computed multiplying together the number of possible occurences for $s[i]$, with $i = 1, \dots, 5$. To do so, we need to define the maximal number of opportunities in each rank, $\max_i$, with $i = 1, \dots, 5$. We estimate the latter using historical data obtaining for the dimension of the state space the expression:

$$|S_{\text{tot}}| = \prod_{i=1}^{5} \max_i.$$

As an example, if the maximal number of opportunities in each rank was equal to 100, the dimension of the state space would equal $|S_{\text{tot}}| = 10'000'000'000$, which is far too high to be handled. Hence, it is necessary to reduce the number of states using an appropriate approximation.

In order to have a smaller state space, we use a projection that maps each 'real' state into an approximated version, where quantiles of the empirical distribution of the number of opportunities per rank, obtained from historical data, are used to represent the state vectors instead of the actual number of opportunities per rank. For example, if we wish to have five possible occurences in each rank, we have to build the 0% (minimal value), 20%, 40%, 60%, 80% and 100% (maximal value) quantiles.

Let $q_i \in \mathbb{N}^n$ denote the quantile vector for rank i, $i = 1, \ldots, 5$ and $q_i[j]$ its j^{th} element, $j = 1, \ldots, n$. The componentwise projection function $\pi_{\text{state}} : \mathbb{N} \longrightarrow \mathbb{N}$ is given by formula (4.8). For values of $s[i]$ inbetween the 0%, 20%, 40%, 60%, 80% to 100% quantiles, the projection is given by the average of the the two quantiles between which $s[i]$ is situated. An exception occurs for the values of $s[i]$ which are higher than the maximal value $q_i[n]$[1]. These values of $s[i]$ are then mapped to the average of the 80% and 100% quantiles of q_i.

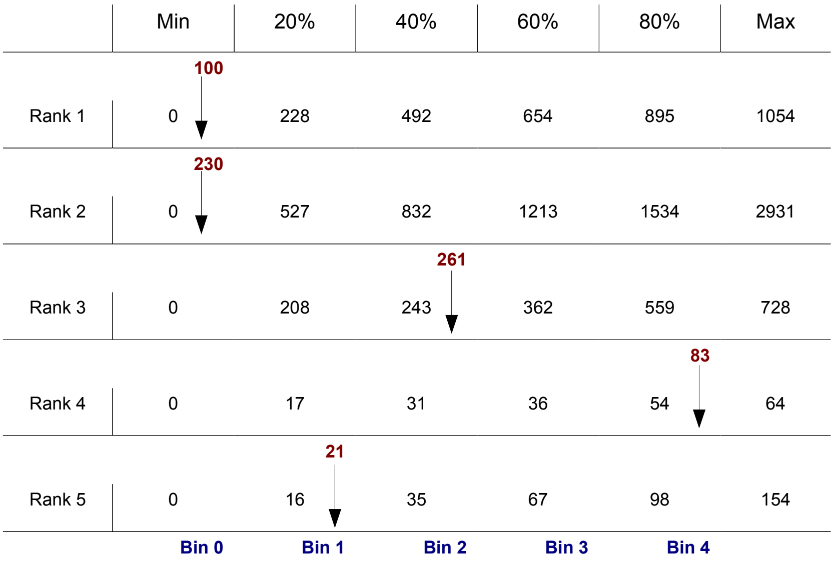
$$\pi_{\text{state}} \quad : \quad \mathbb{N} \longrightarrow \mathbb{N} \tag{4.7}$$

$$\pi_{\text{state}}(s[i]) \quad = \quad \begin{cases} \left\lceil \frac{q_i[1]+q_i[2]}{2} \right\rceil & \text{if } q_i[1] \leq s[i] \leq q_i[2] \\ \left\lceil \frac{q_i[j]+q_i[j+1]}{2} \right\rceil & \text{if } q_i[j] < s[i] \leq q_i[j+1], \text{ for } j = 2, \ldots, n-1 \\ \left\lceil \frac{q_i[n-1]+q_i[n]}{2} \right\rceil & \text{if } s[i] > q_i[n] \end{cases} \tag{4.8}$$

Let us illustrate this by an example. Consider the second leaf in Tree 1 (see Figure 3.2(a)), where the expected gain is between 1 and 5 million USD, the quantiles corresponding to the distribution of the number of opportunities per rank are shown in Figure 4.1. The number of possible occurences in each rank is equal to 5, i.e. to the number of percentiles chosen for the discretization minus 1. The components of the example state $s = (100, 230, 261, 83, 21)$ are mapped into categories formed by the quantiles, e.g. $s[1] = 100$ falls between the minimal value and the 20%-quantile for rank 1. According to formula (4.8), the projection of $s[1]$ is equal to

$$\pi_{\text{state}}(s[1]) = \frac{q_1[1] + q_1[2]}{2} = \left\lceil \frac{0 + 228}{2} \right\rceil = 114.$$

[1] Indeed, this situation can occur when we apply the model, as we will not reproduce exactly the states of the historical data.



	Min	20%	40%	60%	80%	Max
Rank 1	0	228	492	654	895	1054
Rank 2	0	527	832	1213	1534	2931
Rank 3	0	208	243	362	559	728
Rank 4	0	17	31	36	54	64
Rank 5	0	16	35	67	98	154

Figure 4.1: Example of the projection of a state $s = (100, 230, 261, 83, 21)$ into bins determined by quantiles computed on the historical distribution of the number of opportunities in each rank. For example, component $s[i]$ is projected onto a value given by the average of the quantiles between which it is situated, i.e. the 0% and the 20% quantiles. This value is $\lceil \frac{0+228}{2} \rceil = 114$.

We can extend the projection π_{state} defined in equation (4.8) at a state level, i.e. let us define ϕ_{state} as follows:

$$\phi_{\text{state}} \quad : \quad S_{\text{tot}} \longrightarrow S_{\text{new}} \tag{4.9}$$

$$\phi_{\text{state}}(s) \quad = \quad (\pi_{\text{state}}(s[1]), \ldots, \pi_{\text{state}}(s[5])) \tag{4.10}$$

The state space is finally defined as the set of all possible projections by ϕ_{state}, that is, the image of ϕ_{state}. The projection ϕ_{state} leads to a reduction of the original state space S_{tot}. Indeed, the new state space S_{new}, which consists of the projected states only, is reduced to size:

$$|S_{\text{new}}| = (n-1)^5$$

In the example, n is equal to 6, hence the dimension of the state space is $|S_{\text{new}}| = 5^5 = 3'125$, which is a much lower number than the dimension of the original state space $|S_{\text{tot}}|$.

Later in this chapter, we will need to define the transition function, which can be expressed by a matrix of size $|S_{\text{new}}| \times |S_{\text{new}}| \times |A|$, where $|A|$ is the dimension of the action space. To do so, we need to enumerate the state space S_{new}, i.e. find a bijection that maps every state of S_{new} to a number x that is an element of a subset of $\mathbb{N}$. This is done in two steps.

1. First, we find a binning for every possible component of a projected state of S_{new}.

Following the same pattern as formula (4.8), we define function π_{bin}:

$$\begin{aligned} \pi_{\text{bin}} &: \mathbb{N} \longrightarrow \{0, \ldots, n-2\} & (4.11) \\ \pi_{\text{bin}}(s[i]) &= \begin{cases} 0 \text{ if } q_i[1] \leq s[i] \leq q_i[2] \\ j-1 \text{ if } q_i[j] < s[i] \leq q_i[j+1], \text{ for } j = 2, \ldots, n-1 \\ n-1 \text{ if } s[i] > q_i[n] \end{cases} & (4.12) \end{aligned}$$

The same way as for π_{state}, we can define a binning function ϕ_{bin} for each state $s \in \mathbb{N}^5$ as follows

$$\begin{aligned} \phi_{\text{bin}} &: S_{\text{new}} \longrightarrow \{1, \ldots, n-2\}^5 & (4.13) \\ \phi_{\text{bin}}(s) &= (\pi_{\text{bin}}(s[1]), \ldots, \pi_{\text{bin}}(s[5])) & (4.14) \end{aligned}$$

2. By formula (4.14), we have $\phi_{\text{bin}}(s) \in \{0, \ldots, n-2\}^5$. Using this property, we define a second bijection that maps every binned state to an integer x in a subset of $\mathbb{N}$, that is, an *enumeration.* This definition depends on the maximum value each component of a binned state can take. As any binned state s_b belongs to set $\{0, \ldots, n-2\}^5$, the maximum value of $s_b[i]$ is $n-2$, for $i = 1, \ldots, 5$. Let us denote this maximum value by $s_b^{\max}$. Using this, we can define the enumeration as follows:

$$\begin{aligned} \psi &: \{0, \ldots, n-2\}^5 \longrightarrow \{1, \ldots, (n-1)^5\} & (4.15) \\ \psi(s_b) &= s_b[1](s_b^{\max}+1)^4 + s_b[2](s_b^{\max}+1)^3 + s_b[3](s_b^{\max}+1)^2 & (4.16) \\ &+ s_b[4](s_b^{\max}+1) + s_b[5] + 1 & (4.17) \end{aligned}$$

For example, such enumeration gives the following mappings:

$$\begin{aligned} (0,0,0,0,0) &\longmapsto 1 \\ (0,0,0,0,1) &\longmapsto 2 \\ &\vdots \\ (0,0,0,0,s_b^{\max}) &\longmapsto s_b^{\max}+1 \\ (0,0,0,1,0) &\longmapsto s_b^{\max}+2 \\ &\vdots \\ (s_b^{\max}, s_b^{\max}, s_b^{\max}, s_b^{\max}, s_b^{\max}) &\longmapsto (s_b^{\max}+1)^5 \end{aligned}$$

The composition of the two bijections $\psi \circ \phi_{\text{bin}}$ enables us to map every state $s \in S_{\text{new}}$ onto a unique integer $x \in \{1, \ldots, (n-1)^5\}$. Later on, in the construction of the transition matrix, we will also need the inverse functions of ψ and ϕ_{bin}. Let us define these two functions.

1. Similarly as for ϕ_{bin}, we first define the componentwise function π_{bin}^{-1} that maps every component of a binned state $s_b \in \{0, \ldots, n-2\}^5$ to a component of a projected state $s \in S_{\text{new}}$.

$$\begin{aligned} \pi_{\text{bin}}^{-1} &: \{0, \ldots, n-2\} \longrightarrow \mathbb{N} & (4.18) \\ \pi_{\text{bin}}^{-1}(s_b[i]) &= \left\lceil \frac{q_i[j]+q_i[j+1]}{2} \right\rceil \text{ if } s_b[i] = j-1, \text{ for } j = 1, \ldots, n-1 & (4.19) \end{aligned}$$

Let us define the corresponding function ϕ_{bin}^{-1} at a state level:

$$\begin{aligned} \phi_{\text{bin}}^{-1} &: \{0, \ldots, n-2\}^5 \longrightarrow S_{\text{new}} && (4.20) \\ \phi_{\text{bin}}^{-1}(s_b) &= (\pi_{\text{bin}}^{-1}(s_b[1]), \ldots, \pi_{\text{bin}}^{-1}(s_b[5])) && (4.21) \end{aligned}$$

2. We now define the inverse function of the enumeration, i.e. the *inverse enumeration*. Before giving its expression, we define the following expressions, for $x \in \{1, \ldots, (n-1)^5\}$:

$$\begin{aligned} m_1 &:= (s_b^{\max} + 1)^4 \\ m_2 &:= (s_b^{\max} + 1)^3 \\ m_3 &:= (s_b^{\max} + 1)^2 \\ m_4 &:= (s_b^{\max} + 1) \\ m_5 &:= 1 \\ p_1(x) &:= (x-1) \mod m_1 \\ p_2(x) &:= p_1(x) \mod m_2 \\ p_3(x) &:= p_2(x) \mod m_3 \\ p_4(x) &:= p_3(x) \mod m_4 \\ p_5(x) &:= p_4(x) \mod m_5 \end{aligned}$$

Let us now define the inverse enumeration, using the expressions defined above:

$$\begin{aligned} \psi^{-1} &: \{1, \ldots, (n-1)^5\} \longrightarrow \{0, \ldots, n-2\}^5 \\ \psi^{-1}(x) &= \left(\frac{(x-1) - p_1(x)}{m_1}, \frac{p_1(x) - p_2(x)}{m_2}, \frac{p_2(x) - p_3(x)}{m_3}, \frac{p_3(x) - p_4(x)}{m_4}, \frac{p_4(x) - p_5(x)}{m_5} \right) \end{aligned}$$

Though the expression of the inverse enumeration might look quite complicated, it simply performs the inverse mapping of enumeration ψ, e.g. it maps the binned state $(0, 0, 0, 0, 0) \in \{0, \ldots, n-2\}^5$ to 1.

All functions presented in this section can be summarized by Figure 4.2. Under the bijection scheme, we presented a numerical application using example state (4.6) in order to illustrate the different transformations we apply to an original state $s \in S_{\text{tot}}$.

$$\begin{tikzcd}
S_{tot} \arrow[r, "\phi_{state}"] & S_{new} \arrow[r, shift left, "\phi_{bin}"] & \{0, \ldots, n-2\}^5 \arrow[l, shift left, "\phi_{bin}^{-1}"] \arrow[r, shift left, "\psi"] & \{1, \ldots, (n-1)^5\} \arrow[l, shift left, "\psi^{-1}"] \\
(100, 230, 261, 83, 21) \arrow[r, mapsto] & (114, 264, 303, 59, 26) \arrow[r, mapsto, shift left] & (0, 0, 2, 4, 1) \arrow[l, mapsto, shift left] \arrow[r, mapsto, shift left] & 72 \arrow[l, mapsto, shift left]
\end{tikzcd}$$

Figure 4.2: Summary of the functions described in this section, i.e. projection ϕ_{state}, bijections ϕ_{bin}, ψ and their inverse functions. The numerical application below the scheme shows the projection of example state $s = (100, 230, 261, 83, 21)$ onto S_{new} by ϕ_{state}, that is, $\phi_{\text{state}}(100, 230, 261, 83, 21) = (114, 264, 303, 59, 26)$, its binning $\phi_{\text{bin}}(168, 218, 272, 50, 18) = (0, 0, 2, 4, 1)$ and its enumeration $\psi(0, 0, 2, 4, 1) = 72$.

4.2.2 Action space

In chapter 2, we introduced the data of investments at the beginning of each quarter, for each business line. In our MDP framework, investments decisions represent the actions a decision maker will perform and an optimal sequence of investments will represent the optimal policy, which is the output of the MDP.

As the aim is to apply a model to a leaf, i.e. a category of opportunities defined by particular characteristics such as the business line, the sector and the category of expected gain, and the investments are only available by business line, as a preliminary step, we need to split the investments given at a business line level across the different leaves of a particular tree, according to some weighting scheme.

We decided to compute the weight for each leaf as the number of opportunities per salesman and per leaf, normalized by the sum of all number of opportunities per salesman in the business line. Again, this is performed quarterly. Let us denote the quarterly investment for a particular business line a by $I_{\mathrm{BL}}[a, t]$, where t denotes a quarter, with $t = 1, \ldots, 9$, index 1 corresponding to the first quarter of year Y1 and index 9 corresponding to the first quarter of year Y3. For example, if the opportunities of a are split across N leaves and leaf $l \in \{1, \ldots, N\}$ has a total number of opportunities $N_{\mathrm{opp}}[l, t]$ and a total number of salesmen $N_{\mathrm{s}}[l, t]$ for a quarter t, the quarterly ratio for leaf l is:

$$r[l, t] := \frac{N_{\mathrm{opp}}[l, t]}{N_{\mathrm{s}}[l, t]}$$

The weight $w[l, t] \in [0, 1]$ for leaf l is hence given by:

$$w[l, t] := \frac{r[l, t]}{\sum_{i=1}^{N} r[i, t]}$$

and its corresponding quarterly investment is:

$$I[l, t] := w[l, t] \cdot I_{\mathrm{BL}}[a, t].$$

This computation enabled us to have a quarterly investment value for each leaf of a particular tree. As an example, figure 4.3 shows a scheme of the split of the investments across the different leaves of business line BL1.

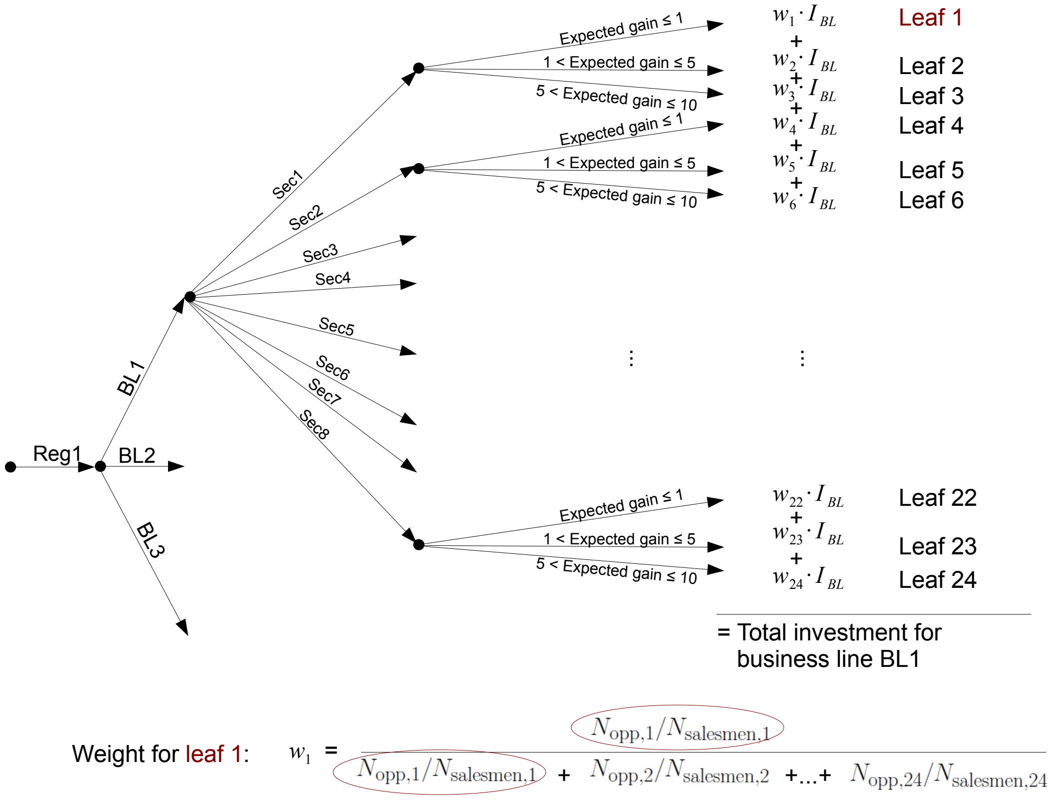

Reg1
BL1
BL2
BL3
Sec1
Sec2
Sec3
Sec4
Sec5
Sec6
Sec7
Sec8
Expected gain ≤ 1
1 < Expected gain ≤ 5
5 < Expected gain ≤ 10
$w_1 \cdot I_{BL}$ + $w_2 \cdot I_{BL}$ + $w_3 \cdot I_{BL}$ + $w_4 \cdot I_{BL}$ + $w_5 \cdot I_{BL}$ + $w_6 \cdot I_{BL}$
Leaf 1
Leaf 2
Leaf 3
Leaf 4
Leaf 5
Leaf 6
⋮
$w_{22} \cdot I_{BL}$ + $w_{23} \cdot I_{BL}$ + $w_{24} \cdot I_{BL}$
Leaf 22
Leaf 23
Leaf 24
= Total investment for business line BL1
Weight for leaf 1: $w_1 = \frac{N_{\text{opp},1}/N_{\text{salesmen},1}}{N_{\text{opp},1}/N_{\text{salesmen},1} + N_{\text{opp},2}/N_{\text{salesmen},2} + \ldots + N_{\text{opp},24}/N_{\text{salesmen},24}}$

Figure 4.3: Split of the investments across the different leaves of business line BL1. The investment of each leaf l is given by a proportion w_l of the total investments for BL1. At the bottom of the figure, an example of computation of weight is given for leaf 1. It is given by the number of opportunities per salesman divided by the sum of the number of opportunities per salesman across all leaves of the business line. Let us recall this is performed for the investments of each quarter.

Throughout all quarters of years Y1, Y2 and Y3 (for which data are available), opportunities appear and disappear at any weekly time step. Therefore, having in mind the goal of an optimal investment strategy, we would like to know the amount of money to invest at the beginning of each quarter for two sorts of opportunities:

Existing opportunities: Opportunities that appear before the beginning of the quarter and are still in the pipe at the end of the quarter.

New opportunities: Opportunities that appear after the beginning of the quarter.

All the following formulas only refer to one leaf, hence let us define the simplified notation $I[t] := I[l, t]$ for the quarterly investment at a leaf level. To take into account the investment for both kinds of opportunities, let us define $I_0[t]$ as the quarterly investment for the new opportunities, $c[t]$ as the quarterly investment for one existing opportunity and $N_{\text{ex}}[t]$ the number of existing opportunities at the right beginning of quarter t, that is, in its first week. We split the quarterly investement per leaf $I[t]$ as follows:

$$\begin{aligned} I[t] &= I_0[t] + c[t] \cdot N_{\text{ex}}[t] \qquad (4.22) \\ &= \alpha[t]I[t] + (1 - \alpha[t])I[t] \qquad (4.23) \end{aligned}$$

Precisely, a percentage $\alpha[t]$ of the quarterly investment $I[t]$ is given for the new opportunities and the remaining percentage $1 - \alpha[t]$ is left out for the existing ones. Extracting $\alpha[t]$ from the data enables us to compute both the quarterly investment for the new opportunities $I_0[t]$ and the quarterly investment for each existing opportunity $c[t]$.

In order to compute $\alpha[t]$, we first build a distribution D_α from the whole timeline of the historical data:

$$D_\alpha = \{\alpha_1, \ldots, \alpha_W\},$$

where each α_w is computed as follows:

$$\alpha_w = \frac{N_w^{\text{new}}}{N_{w-13}^{\text{ex}} + N_w^{\text{new}}} \qquad (4.24)$$

Here, index w refers to the current week. It ranges from 1 to W, where 1 is the first week of the second quarter of year Y1 and W is the last (thirteenth) week of the first quarter of year Y3. The quantities N_w^{new} and N_{w-13}^{ex} are the number of new opportunities at week w and the number of existing opportunities at week $w - 13$, respectively. Let us note that, if we have $1 \leq w \leq 13$, N_{w-13}^{ex} will refer to the number of existing opportunities of weeks 1 to 13 in the first quarter of year Y1.

Each α_w computed using formula (4.24) represents the proportion of new opportunities relatively to the number of opportunities in the pipe one month ago at the same week. For example, in order to compute α_2, we need to know the number of new opportunities at week 2 of the second quarter of year Y1 and the total number of opportunities at week 2 of the first quarter of year Y1. As $I[t]$ is computed quarterly, we need to find a quarterly proportion of new opportunities $\alpha[t]$ as well. To do so, we compute the quarterly average of the α_w's, that is, for quarter t, with $t = 1, \ldots, 8$:

$$\overline{\alpha_t} := \frac{1}{13} \sum_{w=13(t-1)+1}^{13t} \alpha_w.$$

Let us note that we compute $\overline{\alpha_t}$ for $t = 1, \ldots, 8$ and not for the first quarter of year Y3, that is, the quarter corresponding to index 9, because each element of distribution D_α is computed using data from the current quarter $t + 1$ and the previous quarter t. Hence, we do not have the value of $\overline{\alpha_t}$ for the quarter corresponding to index 9. Nevertheless, we need a value for the latter and use the median of all $\overline{\alpha_t}$, for $t = 1, \ldots, 8$ as a replacement, i.e.,

$$\overline{\alpha_9} := \text{median}(\overline{\alpha_1}, \ldots, \overline{\alpha_8}).$$

We now have a quarterly percentage of the investment $I[t]$ and can define $\alpha[t] := \overline{\alpha_t}$, for $t = 1, \ldots, 9$. Hence we can deduce the quarterly investment for each individual opportunity $c[t]$ as follows:

$$c[t] := \frac{(1 - \alpha[t]) \cdot I[t]}{N_{\text{ex}}[t]},$$

where $N_{\text{ex}}[t]$ is the number of existing opportunities in the first week of quarter t.

To define the action space within a leaf, we group together all quarterly investments that have 'close' values for $I_0[t]$ and $c[t]$. The latter form an *action*. In order to define 'closeness', we create a two-dimensional grid. The x-axis displays the $I_0[t]$-values and the y-axis the $c[t]$-values. The endpoints of the mesh are given by the maximum an minimum values of $I_0[t]$ on the x-axis and $c[t]$ on the y-axis, respectively. The number of cells in the grid is to be chosen by setting a number of possible bins for $I_0[t]$ on the x-axis and a number of possible bins for $c[t]$ on the y-axis. The resulting mesh will define the number of actions depending on the number of cells in the grid which are occupied by the $(I_0[t], c[t])$-couples.

As an example, Figure 4.4 shows the actions created by a 3×2 discretised action space. The points are the $(I_0[t], c[t])$-couples computed on the data of the historical investments for the leaf with average expected gain between 1 and 5 mio USD. Let us note that it may happen that some of the cells are not filled. Hence the number of actions is given by the number of filled cells. Here all cells are filled, therefore the number of actions is 6.

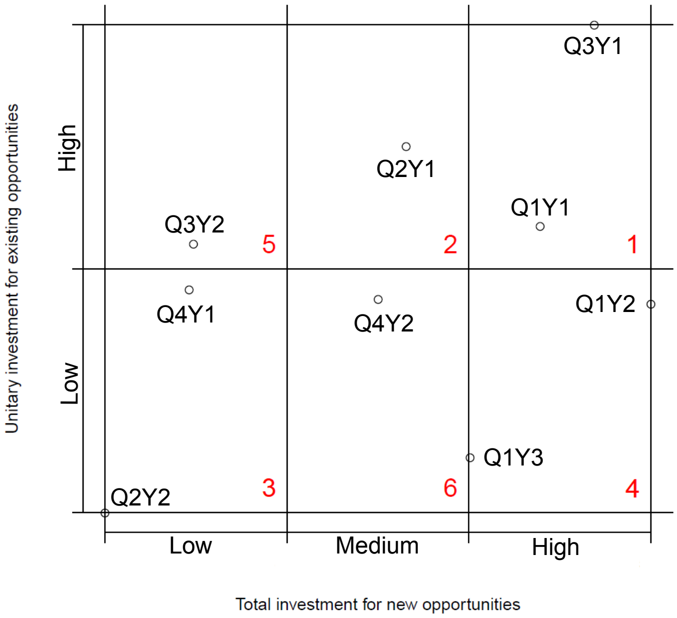


Figure 4.4: Example of a 3×2 discretised action space to group together the $(I_0[t], c[t])$-couples of investments with close values, for the leaf with average expected gain between 1 and 5 mio USD. The dimensions of the mesh is to be chosen by the modeller. The filled cells form the actions.

The boundaries of the filled cells of the mesh define the possible actions in the action space. Each dimension (i.e. the x or y-axis) is labelled. For example, an action could be to invest a **low** amount of money for the new opportunities and a **low** amount of money for each existing opportunity.

As for the state space, we need to enumerate the action space, i.e. to put labels to all actions. This is done by assigning a number to each action the first time it is visited. For example, the action represented by the upper right cell on Figure 4.4 is given label 1 as the investments $I_0[1]$ and $c[1]$ of the very first quarter of year Y1 belong to it. Similarly, the investments of quarter 2 of year Y1 belong to the upper middle cell. Hence the latter is assigned label 2. This labelling goes on till all visited cells are numbered.

4.2.3 Transition function

Let us introduce the third component of an MDP defined in section 4.1, i.e. the transition function $P : S \times S \times A \longrightarrow [0, 1]$. According to the previous definitions of the state space and the action space, we define sets S and A as follows:

$$S := \{1, \ldots, (n-1)^5\}$$

$$A := \{1, \ldots, N_a\}$$

Let us recall that n is the number of computed quantiles we want in order to map the components of a state vector of S_{tot}. Moreover, we define N_a as the number of actions.

The transition function can be represented by a three-dimensional matrix denoted by the same name P, where each cell $P(s, s', a)$ represents the probability that state $s \in S$ moves to state $s' \in S$ when action $a \in A$ is chosen.

In section 3.2, we created lists of matrices of the transition probabilities from one rank to another and lists of five-component vectors of the new opportunities in each rank. This was performed for each leaf of the two trees described in section 3.1 and for four different time lags, i.e. lag 1 for the weekly transitions, lag 2 for fortnightly transitions, lag 4 for monthly transitions and lag 13 for quarterly transitions.

As part of the available historical data, i.e. the investments and the signings, is given on a quarterly scale, we want the MDP to be on a quarterly scale as well. This is one of the main issues we met for the modelling part, as the rest of the data, such as the number of opportunities or the values of expected gain for each opportunity, are given on a weekly scale. We first considered applying a *Partially Observable Markov Decision Process (POMDP)*, but discarded this possibility due to the difficulty of defining the observation process and its relationship with the core process. We finally opted for a simulation of a quarterly transition, sampling from the lists of weekly transition matrices.

Due to the fact that we want to model a quarterly transition, every cell $P(s, s', a)$ of the transition matrix P represents the probability that, given that action $a \in A$ is chosen, state $s \in S$ moves to state $s' \in S$ in one quarter. To build these quarterly transitions, we will perform a simulation using the matrices of transition probabilities and the vectors of new opportunities computed for a lag of one week (lag 1).

Before presenting this simulation, we assign an action to each couple of 5×6 matrix of transition probabilities and each 5-dimensional vector of new opportunities. In the previous section describing the action space, each cell of the mesh was given a number representing a label from 1 to N_a, the number of actions, which is the number of occupied cells. Each quarterly couple of investments $(I_0[t], c[t])$ was then mapped to a particular action $a \in A$. Knowing the correspondance between a quarter $t \in \{1, \ldots, 9\}$ and an action a, we can construct lists of matrices of transitions by action. For example, the investments of quarters Q1 and Q3 belong to action 1 of Figure 4.4. This implies that the list of matrices of transitions for action 1 will consist of all matrices of transition probabilities that occur *within* quarter Q1 or quarter Q3. Figure 4.5 summarizes the binning of the transition matrices in the example of the 3×2 mesh of Figure 4.4. For example, as the couples of investments $(I_0[t], c[t])$ corresponding to the first and third quarters of year Y1 were represented in the same cell in the 3×2 mesh, the weekly matrices of transition within these two quarters are mapped into the same action.

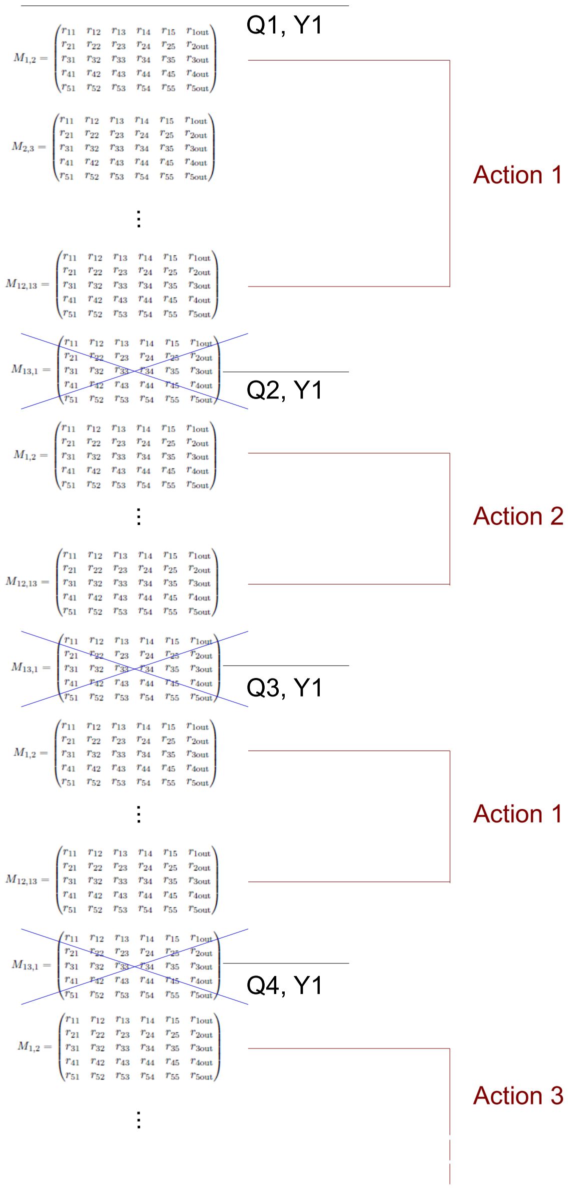


Figure 4.5: Mapping of the weekly transition matrices into the actions determined by the 3×2 mesh of Figure 4.4 for the four quarters of year Y1. Inter-quarter transitions are not considered here.

In order to build P, we perform a simulation. Let us first assume a fixed initial state $s \in S$ and a fixed action $a \in A$. The aim is to obtain the probability for state s to reach any state $s' \in S$ given action a. The idea is first to transform s into an element s_{proj} of S_{new} and then to multiply it by a 5×6 matrix of transition M *randomly selected* amongst the list of weekly transition matrices corresponding to action a. The result will be a 1×6 vector v that consists of a new repartition of the opportunities that were in s_{proj}. The first five components of v consist of the number of opportunities in the five ranks and the sixth component corresponds to the number of opportunities that went out of the pipe after a one-week transition. We are only interested in the vector consisting of the five first components as it represents the opportunities that are still in the pipe. Let us denote the latter by w. The matrix multiplication only deals with the existing opportunities. We now need to add the vector of new opportunities v_{new} corresponding to the same transition as M in the historical data in order to model the dynamics completely. The resulting state s'_{proj} after the simulation of this weekly transition is given by:

$$s'_{\text{proj}} := w + v_{\text{new}}.$$

To obtain a quarterly transition, we perform this one-week step twelve times and map the resulting state back into space S. The enumeration s' of this state will give us the second index in matrix P. In terms of implementation, all elements of P are initialized to zero. After the twelve simulation steps, the value of the corresponding $P(s, s', a)$ is incremented by 1. We repeat this procedure a large number of times N_{sim} for the initial state s and do the same for all other initial states of S and all other actions of A.

Finally each cell of matrix P is normalized by N_{sim} in order to obtain the probabilities of transition. The whole procedure is summarized by the next three algorithms. Algorithm 2 calls Algorithm 3 and Algorithm 3 calls Algorithm 4. Function `makeP` of Algorithm 2 presents the construction of the transition matrix P and function `simulationQuarter` of Algorithm 3 calls twelve times the weekly simulation function `simulationLag` implemented in Algorithm 4.

Algorithm 2 makeP

```
for $a \in A$ do
  for $s \in S$ do
    $s_{\text{bin}} \leftarrow \psi^{-1}(s)$
    $s_{\text{proj}} \leftarrow \phi_{\text{bin}}^{-1}(s_{\text{bin}})$
    for $i = 1, \ldots, N_{\text{sim}}$ do
      $s'_{\text{real}} \leftarrow$ simulationQuarter$(s_{\text{proj}})$
      $s'_{\text{bin}} \leftarrow \phi_{\text{bin}} \circ \phi_{\text{state}}(s'_{\text{real}})$
      $s' \leftarrow \psi(s'_{\text{bin}})$
      $P(s, s', a) \leftarrow P(s, s', a) + 1$
    end for
    $P(s, :, a) \leftarrow \frac{P(s,:,a)}{N_{\text{sim}}}$
  end for
end for
return $P$
```

Remark 1. *The notation $P(s, :, a)$ of Algorithm 2 means that we consider the row of matrix P that has a fixed $s \in S$ and a fixed $a \in A$. Performing the operation*

$$P(s, :, a) \leftarrow \frac{P(s, :, a)}{N_{sim}}$$

means that all elements of row $P(s, :, a)$ are divided by N_{sim}.

Algorithm 3 simulationQuarter

```
$s_{\text{proj}}[5] \leftarrow 0$
for $i = 1, \dots, 12$ do
    $s_{\text{proj}} \leftarrow \text{simulationLag}(s_{\text{proj}})$
end for
return $s_{\text{proj}}$
```

Remark 2. *Due to the fact that we set $s_{proj}[5]$ to 0 in Algorithm 3, Algorithm 2 needs only to be performed every $n-1$ step. Hence, it is not necessary to loop over all $s \in S$, but over all $\tilde{s} \in \{1, 6, \dots, (n-1)^5\}$, and in the computation of P, all $n-2$ arrays below array $P(\tilde{s}, :, :)$ are just copies of the latter.*

Algorithm 4 simulationLag

```
$r \leftarrow \text{dice}_a()$
$M \leftarrow l^a_M[r]$
$v_{\text{new}} \leftarrow l^a_{v_{\text{new}}}[r]$
$s'_{\text{proj}} \leftarrow (s_{\text{proj}} \cdot M)[1:5] + v_{\text{new}}$
return $s'_{\text{proj}}$
```

Remark 3. *Function dice_a of Algorithm 4 randomly draws an integer r between 1 and the size of the list with the matrices of transitions for action a, l^a_M. The matrix M and the vector of new opportunities v_{new} are selected using the same index r in list l^a_M of matrices of transitions for action a and list $l^a_{v_{new}}$ of vectors of new opportunities for a, respectively.*

4.2.4 Reward function

In this section, we will define the reward function $R : S \times S \times A \longrightarrow \mathbb{R}$. In chapter 2, we introduced the signings data, given for each quarter and at a sector level. We will use these data in order to find an expression of the reward function R.

As for the investments, we need to split each quarterly value of signings given by business line and by sector across the different leaves of our tree representation. The rationale we use to split the values of signings is that the amount of money corresponding to the expected gain for opportunities that are actually won (i.e. that reached rank 5) should be highly correlated with the money values reported in the signings data. We define a weight for each leaf l of a particular sector b at a quarter t. Let us assume that leaf l has $N^{r_5}_{\text{opp}}[l,t]$ opportunities that reach rank 5 for quarter t and that sector b has N leaves. Let us denote the quarterly value of signings for sector b by $\text{S}[b,t]$ The weight $w[l,t]$ corresponding to quarter t of leaf l is defined as the sum of the average values of expected gain of all opportunities of leaf l for quarter t, divided by the sum of all average values of expected gains in sector b for quarter t:

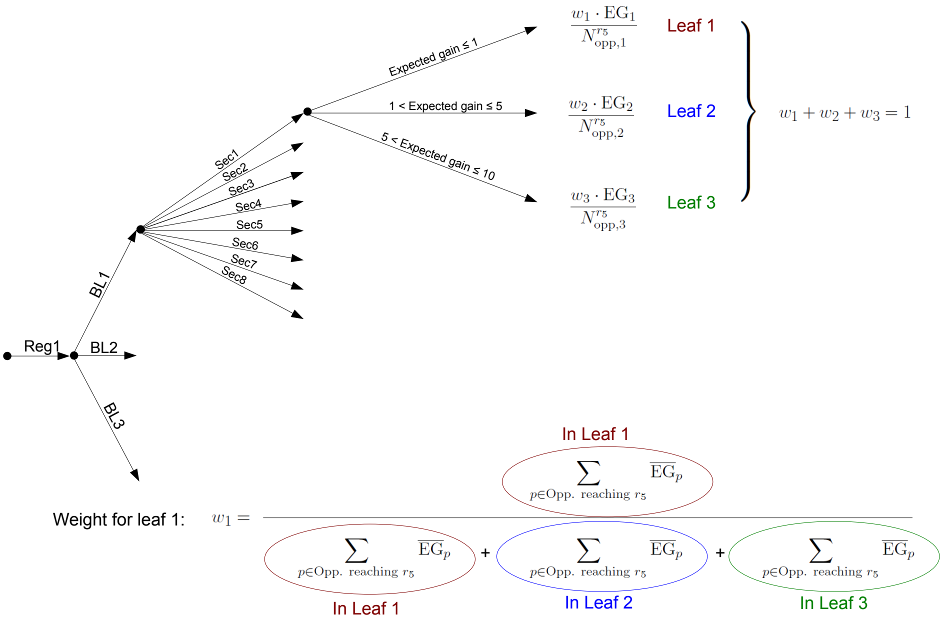
$$w[l,t] := \frac{\sum_{p=1}^{N^{r_5}_{\text{opp}}[l,t]} \overline{\text{EG}}[p]}{\sum_{i=1}^{N} \sum_{j=1}^{N^{r_5}_{\text{opp}}[i,t]} \overline{\text{EG}}[j]}.$$

Here, $\overline{\text{EG}}[p]$ denotes the average value of expected gain for an opportunity p of leaf l. In the computation of the weights for the investments, all weights within the same business line and for the same quarter summed up to 1. In the case of the signings, all weights $w[l,t]$ within the same sector and for the same quarter sum up to 1.

We deduce the *unitary* reward $r[l,t]$ for an opportunity of leaf l at quarter t as

$$r[l,t] := \frac{w[l,t] \cdot \mathrm{S}[b,t]}{N_{\text{opp}}^{r_5}[l,t]}.$$

Dividing by $N_{\text{opp}}^{r_5}[l,t]$ implies that we only give rewards for the opportunities that reach rank 5. An example of splitting of the values of signings for sector Sec1 of business line BL1 is shown in Figure 4.6.



Reg1
BL1
BL2
BL3
Sec1
Sec2
Sec3
Sec4
Sec5
Sec6
Sec7
Sec8
Expected gain ≤ 1
1 < Expected gain ≤ 5
5 < Expected gain ≤ 10
$\frac{w_1 \cdot \mathrm{EG}_1}{N_{\text{opp},1}^{r_5}}$
Leaf 1
$\frac{w_2 \cdot \mathrm{EG}_2}{N_{\text{opp},2}^{r_5}}$
Leaf 2
$\frac{w_3 \cdot \mathrm{EG}_3}{N_{\text{opp},3}^{r_5}}$
Leaf 3
$w_1 + w_2 + w_3 = 1$
In Leaf 1
Weight for leaf 1:
$w_1 = \frac{\sum_{p \in \text{Opp. reaching } r_5} \overline{\mathrm{EG}}_p}{\sum_{p \in \text{Opp. reaching } r_5} \overline{\mathrm{EG}}_p + \sum_{p \in \text{Opp. reaching } r_5} \overline{\mathrm{EG}}_p + \sum_{p \in \text{Opp. reaching } r_5} \overline{\mathrm{EG}}_p}$
In Leaf 1
In Leaf 2
In Leaf 3

Figure 4.6: Split of the values of signings corresponding to sector Sec1 of business line BL1 across the different leaves determined by the categories of average expected gain. The computation of the weight for the leaf corresponding to values of average expected gain below 1 million USD is shown at the bottom of the figure. Let us note that we divide each value of the weighted average expected gain by the number of opportunities that reached rank 5 in order to obtain a unitary reward. This split is performed for all quarters.

Let us note that as a model will be applied individually to each leaf, we can simplify the notation for the unitary reward as follows:

$$r[t] := r[l,t].$$

Let us recall that the aim of this project is to apply an MDP to the data and to solve it using the value iteration algorithm described in section 4.1. The reward function $R : S \times S \times A \longrightarrow \mathbb{R}$ given in the same section can be represented as a $|S| \times |S| \times |A|$ matrix with the same name R, in a similar way as for the transition function P. Before giving the expression of matrix R, let us define two other matrices:

- $\mathrm{Rew}(s',a)$, which is the average unitary reward in action cell a multiplied by the number of opportunities in rank 5 of s'.
- $\mathrm{Inv}(s,a)$, which is the average unitary investment in action cell a multiplied by the number of opportunities in ranks 1 to 4 of s.

We define each element $R(s, s', a)$ as the difference between $\text{Rew}(s', a)$ and $\text{Inv}(s, a)$. Let us note that we need to discount the first term of the difference by a factor $\gamma \in (0, 1]$ as it is situated one step forward in the future. This difference is expressed as follows:

$$R(s, s', a) := \gamma \cdot \text{Rew}(s', a) - \text{Inv}(s, a). \tag{4.25}$$

Equation (4.25) enables us to build each element of matrix R. Nevertheless, it is possible to build a $|S| \times |A|$ matrix PR that represents the reward component of the model in a slightly different way. The use of this matrix of smaller dimension is a result from the value iteration algorithm and can reduce significantly the computation time of the reward component of the model.

Let us give a justification for the use of PR. Let us consider again equation (4.5) and rewrite it as follows:

$$\begin{aligned} Q(s,a) &= \sum_{s' \in S} P(s, s', a) \left(R(s, s', a) + \gamma V(s') \right) \\ &= \sum_{s' \in S} P(s, s', a) R(s, s', a) + \gamma \sum_{s' \in S} P(s, s', a) V(s') \end{aligned}$$

The first term of the last expression is the expectation of $R(s, s', a)$ as there is a probability component $P(s, s', a)$. Let us denote it by $PR(s, a)$ and rewrite it as follows:

$$\begin{aligned} PR(s,a) &:= \sum_{s' \in S} P(s, s', a) R(s, s', a) \\ &= E\left(R(s, s', a) \right) \\ &= E\left(\gamma \cdot \text{Rew}(s', a) - \text{Inv}(s, a) \right) \\ &= \gamma \cdot E\left(\text{Rew}(s', a) \right) - \text{Inv}(s, a) \\ &= \gamma \cdot \sum_{s' \in S} P(s, s', a) \text{Rew}(s', a) - \text{Inv}(s, a) \end{aligned}$$

All values of $\text{Rew}(s', a)$ and $\text{Inv}(s, a)$ can be represented in two $|S| \times |A|$ matrices Rew and Inv, respectively. The computation of PR can be summarized by the following three algorithms. Algorithm 5 presents function `makeRew` which builds matrix Rew, Algorithm 6 explains function `makeInv`, which builds matrix Inv and finally, Algorithm 7 presents function `makePR`, which builds matrix PR using the computations of matrices Rew and Inv and also transition matrix P.

Algorithm 5 makeRew

```
for $a \in A$ do
  $r_{\text{mean}} \leftarrow \text{average}(r[t], t \in \text{action cell } a)$
  for $s' \in S$ do
    $s'_{\text{bin}} \leftarrow \psi^{-1}(s')$
    $s'_{\text{state}} \leftarrow \phi_{\text{bin}}^{-1}(s'_{\text{bin}})$
    for $k = 1, 6, 11, \ldots, (n-1)^5$ do
      $\text{Rew}(s' + k, a) \leftarrow r_{\text{mean}} \cdot s'_{\text{state}}[5]$
    end for
  end for
end for
return Rew
```

Remark 4. *Let us note that as only the fifth component of a state vector $s'_{state} \in S_{new}$ is used in the computation of a reward, every fifth row of matrix Rew is identical.*

Algorithm 6 makeInv

```
for $a \in A$ do
    $I_0^{\text{mean}} \leftarrow \text{average}(I_0[t], t \in \text{action cell } a)$
    $c^{\text{mean}} \leftarrow \text{average}(c[t], t \in \text{action cell } a)$
    for $s = 1, 6, 11, \ldots, (n-1)^5$ do
        $s_{\text{bin}} \leftarrow \psi^{-1}(s)$
        $s_{\text{state}} \leftarrow \phi_{\text{bin}}^{-1}(s_{\text{bin}})$
        $\text{Inv}(s : (s+4), a) \leftarrow I_0^{\text{mean}} + c^{\text{mean}} \cdot \sum_{i=1}^{4} s_{\text{state}}[i]$
    end for
end for
return Inv
```

Remark 5. *The notation Inv$(s : (s+4), a)$ means that cells Inv(s, a) to Inv$(s+4, a)$ are filled at the same time. Indeed, only the four first components of a state vector $s'_{state} \in S_{new}$ are considered in the computation of the investments as we do not invest for opportunities that are already won, i.e. opportunities in rank 5. Therefore, each five consecutive rows of matrix Inv are identical.*

Algorithm 7 makePR

```
for $a \in A$ do
    $PR(:, a) \leftarrow \gamma \cdot (P(:, :, a) \cdot \text{Rew}(:, a)) - \text{Inv}(:, a)$
end for
return $PR$
```

Chapter 5

Results

This chapter presents the results of the application of the MDP presented in chapter 4 to the different leaves of Tree 1.

We report the results of different experiments, obtained varying the values of the following parameters:

- The discretisation of the action space;
- The number of possible occurences in each rank for the states;
- The discretisation of the state space;
- The discount factor γ.

Without any explicit indication, the size of the discretised action space is 2×2, allowing a binary high/low investment for the new opportunities as well as for the existing ones. This choice is due to the fact that for the first results, we just would like to see the tendency between investing more for the existing opportunities and/or investing more for the new ones. We set the number of possible occurences for each rank of a state to 5, which implies that the total number of possible states is $5^5 = 3'125$. The choice of 5 occurences per rank comes from the fact that a state space size of $3'125$ is already very large to handle in the simulation of the transition matrix P. The number of opportunities are binned according to quantiles determined on the historical data from the first quarter of year Y1 to the second quarter of year Y3[1] and the discount factor γ is set to 0.9, which is high enough to give importance to future rewards but does not give them too much weight either.

5.1 Application of the model to the leaves of Tree 1

In this section, we present the results of the application of the MDP to the three leaves of Tree 1 with the default parameters.

Let us first consider the leaf of Tree 1 with average expected gain per opportunity lower than 1 million USD. The couples of investments $(I_0[t], c[t])$ for $t = 1, \dots, 9$ extracted from the historical data for this leaf were discretised as shown in Figure 5.1.

[1] The reason for including the second quarter of year Y3 is to have a wider range of possible states of S_{tot}. Indeed, as the data from the second quarter of year Y3 is incomplete, it can occur that some states do not have many opportunities per rank or do not have any at all. Therefore, the distributions of the number of opportunities in each rank will be based on a wider range of values.

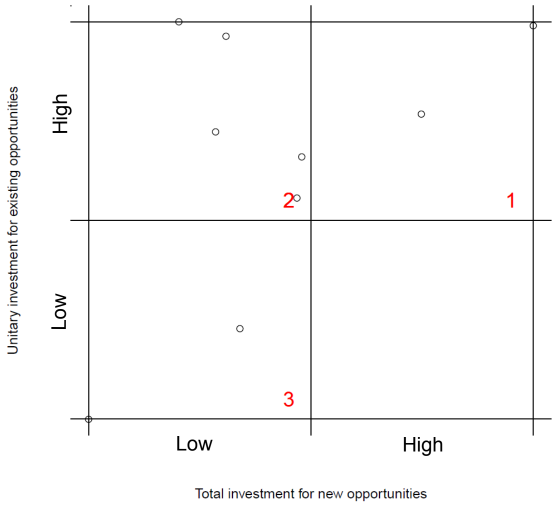
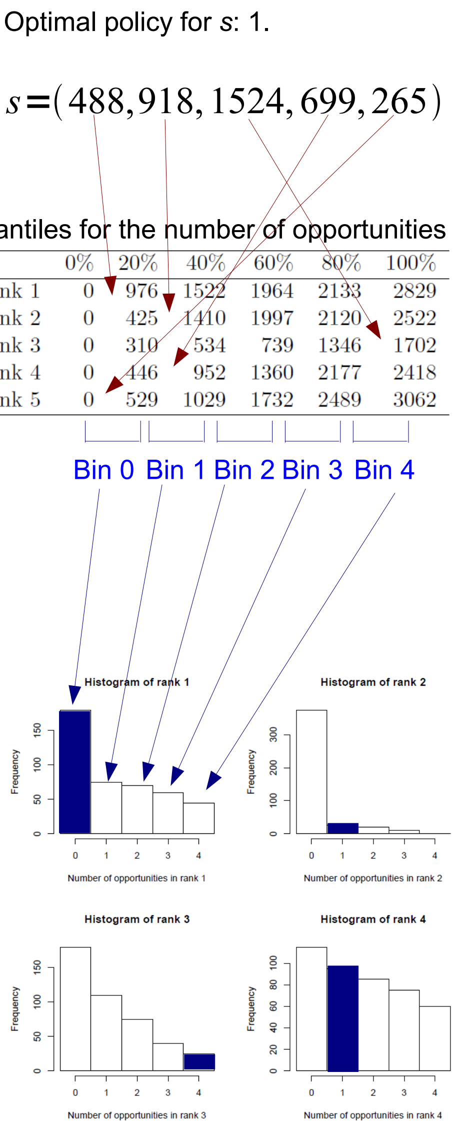


Figure 5.1: 2 × 2 discretisation of the action space. The number of actions is determined by the number of filled cells, for which action labels are given at their bottom right. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain below 1 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

The application of the value iteration algorithm presented in section 4.1 lead to the selection of two kinds of optimal actions amongst the $3'125$ states, that is, the actions corresponding to actions 1 and 3 of the discretised action space of Figure 5.1. For most of the states, the optimal actions is 3, which implies a low investment both for the existing opportunities and the new ones. For the rest of the states, the optimal actions is 1, that is, a high investment for both the existing opportunities and the new ones.

In order to identify the states for which action 1 or action 3 were selected as optimal, we plotted histograms of the number of opportunities in each of the four first ranks. Figure 5.2 summarizes their construction. Let us assume that action 1 was selected as optimal for a particular state vector $s = (488, 918, 1524, 699, 265)$. The latter is discretised to a category between 0 and 4 using the quantiles of the distribution of the number of opportunities per rank. We create the histograms per rank counting the number of occurences of each category for all states with optimal action 1. For example, as the number of opportunities of s in rank 1 is in category 0, it will appear in the first column of the histogram for rank 1.


Optimal policy for *s*: 1.
$s = (488, 918, 1524, 699, 265)$
Quantiles for the number of opportunities
Bin 0
Bin 1
Bin 2
Bin 3
Bin 4
Histogram of rank 1
Histogram of rank 2
Histogram of rank 3
Histogram of rank 4
Frequency
0
50
100
150
200
300
20
40
60
80
1
2
3
4
Number of opportunities in rank 1
Number of opportunities in rank 2
Number of opportunities in rank 3
Number of opportunities in rank 4

	0%	20%	40%	60%	80%	100%
Rank 1	0	976	1522	1964	2133	2829
Rank 2	0	425	1410	1997	2120	2522
Rank 3	0	310	534	739	1346	1702
Rank 4	0	446	952	1360	2177	2418
Rank 5	0	529	1029	1732	2489	3062

Figure 5.2: Construction of the histograms of the number of opportunities per rank for the states with a particular optimal action, e.g. action 1. The quantiles of the distribution of number of opportunities between which the components of s are situated determine its projection onto the discretised state space represented as five bins numbered from 0 to 4 for each rank. Category 0 represents a low number of opportunities and category 4 a high number of opportunities. The categories can be quantified reading in the table of quantiles. To create the histograms per rank, we count the number of occurences in each of the five categories amongst all states with optimal action 1.

The histograms for the states with optimal policy 1 are plotted in Figure 5.3 and the histograms for the states with optimal policy 3 in Figure 5.4. The histogram corresponding to rank 5 is not plotted as states with the four first ranks equal have the same resulting policy, independently of the number of opportunities in rank 5.

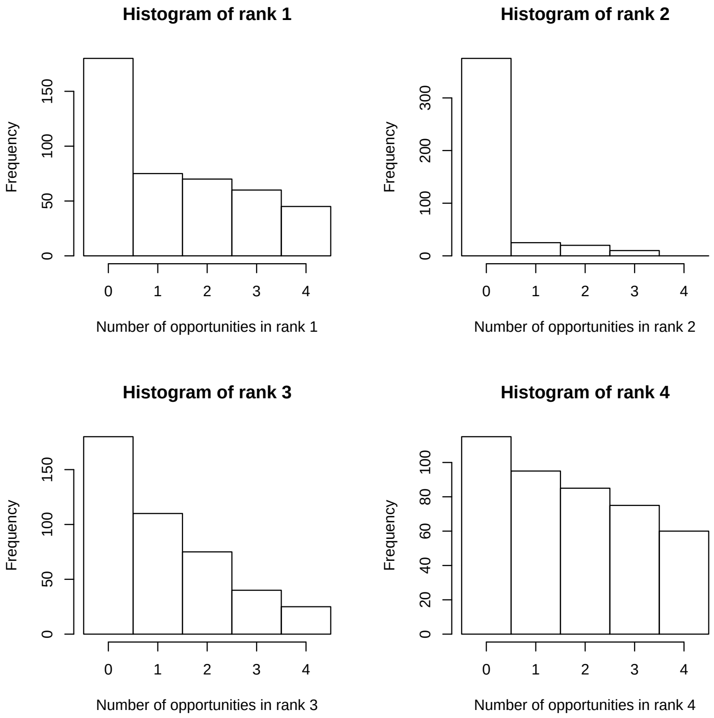


Figure 5.3: Histograms of the number of opportunities per rank for the states for which the best action is 1. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2 × 2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

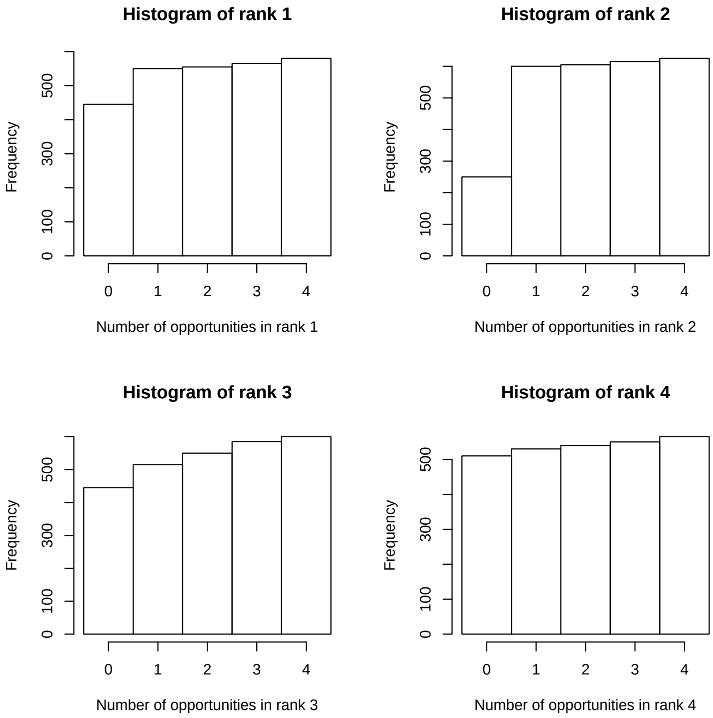


Figure 5.4: Histograms of the number of opportunities per rank for the states for which the best action is 3. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2 × 2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

Figure 5.3 shows that when there are only a few number of opportunities in each rank (label 0), particularly for ranks 1, 2 and 3, the optimal action is to invest the highest amount of money for the new opportunities. This means that when the opportunities have not reached a mature rank yet, it is not worth investing money for them and a better strategy is to invest money for the new opportunities. For rank 4, there seems to be more opportunities, as the histogram shows more occurences in categories 1, 2, 3 and 4. This justifies the fact that the optimal action is to invest also the highest amount of money for the existing ones, as there are proportionally more opportunities in rank 4 that in the other ranks.

Figure 5.4 shows the histograms with the opposite configuration as the ones of Figure 5.3. There is a slight tendency for these states to have a higher number of opportunities in each rank, compared to the states for which the optimal action was 1. This justifies the fact that the investment for the new opportunities is low.

In order to see what the labels 0 (low number of opportunities) to 4 (high number of opportunities) represent in the case of this particular leaf, Table 5.1 gives a fictive example of list of the quantiles computed on the distribution of the opportunities in each rank, similar to the one presented as an example in Figure 4.1. For example, a low number of opportunities (label 0) in rank 1 corresponds to a number of opportunities between 0 and 976, and a high number of opportunities (label 4) in rank 1 corresponds to a number of opportunities between 2133 and 2829.

	0%	20%	40%	60%	80%	100%
Rank 1	0	976	1522	1964	2133	2829
Rank 2	0	425	1410	1997	2120	2522
Rank 3	0	310	534	739	1346	1702
Rank 4	0	446	952	1360	2177	2418
Rank 5	0	529	1029	1732	2489	3062

Table 5.1: Example of list of quantiles based on the distribution of the number of opportunities in each rank. The number of possible occurences in each rank is 5.

Let us now compare the policy given by the model with the historical policy, i.e. the sequence of actions that were actually taken at the beginning of each quarter, from the first quarter of year Y1 to the first quarter of year Y3. The states for which this comparison is performed are the states corresponding to the first week of each quarter in the historical data. The sequence of the optimal actions and the historical actions for these particular states is given in Table 5.2. The model seems to give a different outcome than the historical data. It occurs only twice that the optimal actions match the historical ones, i.e. for the first and fourth quarters of year Y1. Let us notice that in the historical data, the most frequent action is 2, i.e. a low investment for the new opportunities and a high investment for the existing ones, and that it is never selected by the model. Let us note that apart from the first quarters of years Y1, Y2 and Y3, the optimal actions always correspond to a lower investment than the historical ones.

	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2	Q3Y2	Q4Y2	Q1Y3
Historical actions	1	1	2	3	3	2	2	2	2
Optimal actions	1	3	3	3	1	3	3	3	1

Table 5.2: Historical actions and optimal actions given by the application of an MDP to the leaf of Tree 1 with average expected gain per opportunity below 1 million USD. The discretised action space is of dimensions 2×2, the number of possible occurences in each rank is 5, the distributions of the ranks are based on the data from the first quarter of year Y1 to the second quarter of year Y3 and the discount factor γ is equal to 0.9.

Let us analyse the outcome of the MPD for the second leaf of Tree 1, i.e. with average expected gain per opportunity between 1 and 5 million USD. Figure 5.5 shows the discretised action space determined by the historical investments. The number of actions is 4 as all cells of the grid are filled.

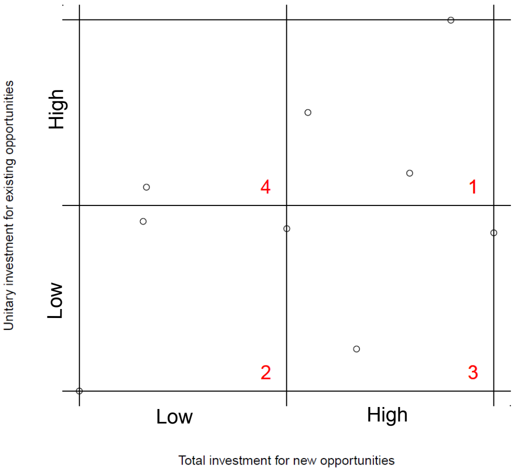


Figure 5.5: 2 × 2 discretisation of the action space. The number of actions is determined by the number of filled cells, for which action labels are given at their bottom right. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain between 1 and 5 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

Three different optimal actions were selected after applying the MDP to the second leaf of Tree 1. For most of the states, the optimal action was to invest the lowest amount of money for both existing and new opportunities. For some states, though, two other types of action were selected as optimal: action 1 and action 3. Action 1 corresponds to a high investment for both existing and new opportunities and action 3 corresponds to a high investment for new opportunities and a low investment for the existing ones.

In order to identify the type of states for which each of the three policies was selected as optimal, we plot similar histograms as for the first leaf of Tree 1.

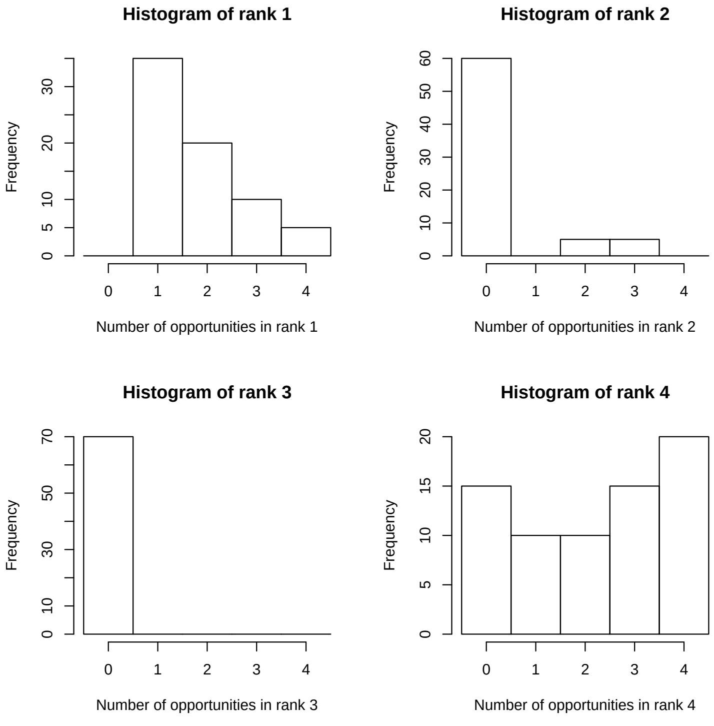


Figure 5.6: Histograms of the number of opportunities per rank for the states for which the best action is 1. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain between 1 and 5 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2×2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

Figure 5.6 shows histograms of the number of opportunities in each of the four first ranks of the states for which the optimal action is to invest a high amount of money for the existing opportunities and for the new ones. As there are only a few opportunities in ranks 1, 2 and 3, it is relevant to invest for new opportunities rather than for the latter which are not in an advanced state yet. But on the other hand, the number of opportunities in rank 4 is rather high in proportion to the number of opportunities in the other ranks. This also justifies a high investment for the existing opportunities.

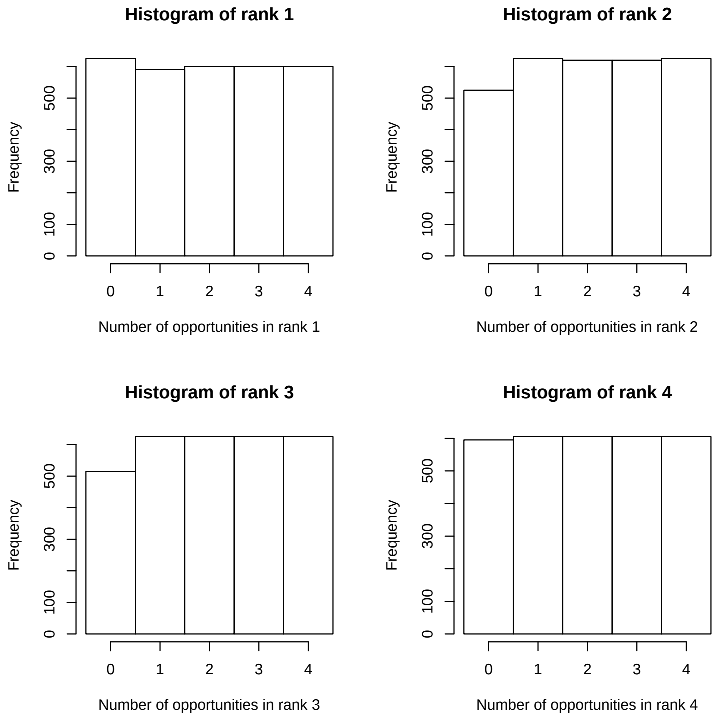


Figure 5.7: Histograms of the number of opportunities per rank for the states for which the best action is 2. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain between 1 and 5 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2×2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

The histograms for the states for which the optimal action is 2, i.e. to invest a low amount of money for both new and existing opportunities are shown in Figure 5.7. The number of opportunities seem equally distributed across the ranks and hence no particular tendency can be identified for that policy.

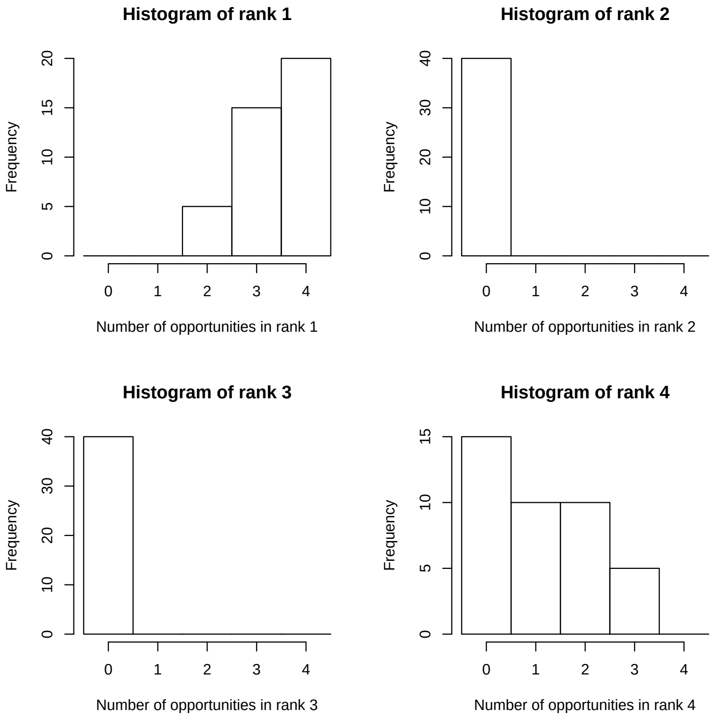


Figure 5.8: Histograms of the number of opportunities per rank for the states for which the best action is 3. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain between 1 and 5 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2×2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

Figure 5.8 shows the histograms for states for which the optimal action is 3, that is, to invest the lowest amount of money for the existing opportunities and the highest amount of money for the new ones. As there are very few opportunities in ranks 2 and 3 and also a rather low number of opportunities in rank 4 for most of the states, it is relevant not to invest a lot of money for them. The highest number of opportunities seem to be in rank 1. As these opportunities have not reached a mature rank yet, it is more relevant to invest for new opportunities than for the latter.

In order to compare the results of the MDP applied to this leaf with the historical data, we present a table of the historical action and the optimal actions (see Table 5.3) as for the first leaf of Tree 1. The optimal action for the states that appeared in the historical data is always 2, i.e. investing a low amount of money for the new and existing opportunities, according to the model. We notice again that the results given by the model are very different from the ones of the historical data.

	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2	Q3Y2	Q4Y2	Q1Y3
Historical actions	1	1	2	2	3	4	3	3	1
Optimal actions	2	2	2	2	2	2	2	2	2

Table 5.3: Historical actions and optimal actions given by the application of an MDP to the leaf of Tree 1 with average expected gain per opportunity between 1 and 5 million USD. The discretised action space is of dimensions 2 × 2, the number of possible occurences in each rank is 5, the distributions of the ranks are based on the data from the first quarter of year Y1 to the second quarter of year Y3 and the discount factor γ is equal to 0.9.

The last leaf of Tree 1 is the leaf for which the average expected gain per opportunity is between 5 and 10 million USD. The 2 × 2 discretisation of the action space determined three actions, which are shown in Figure 5.9.

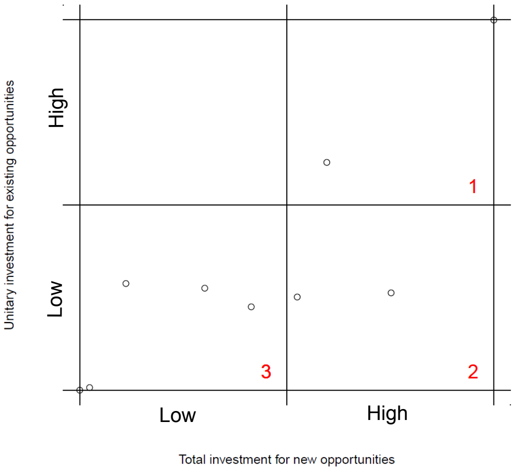


Figure 5.9: 2 × 2 discretisation of the action space. The number of actions is determined by the number of filled cells, for which action labels are given at their bottom right. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain between 5 and 10 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

The model selected action 1, i.e. investing the most for the existing and the new opportunities, for all $3'125$ states, except for 5 states, for which the optimal action was to invest the lowest amount of money for both existing and new opportunities. Figure 5.10 shows the histograms of the number of opportunitites in each rank for these particular 5 states. We do not show the histograms for all other states as they are nearly uniform.

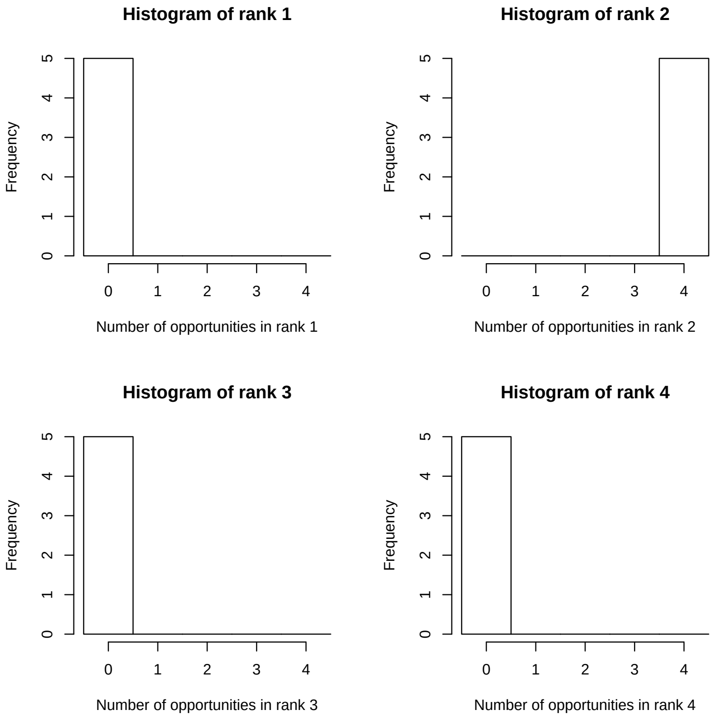


Figure 5.10: Histograms of the number of opportunities per rank for the states for which the best action is 3. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain between 5 and 10 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2×2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

The histograms of Figure 5.10 show that the states for which this holds are the states with a lot of opportunities in rank 2 and very few in the other ranks. It is hard to give a justification for investing more money for these states rather than for all $3'120$ others, as these histograms are based only on five values each.

Finally, let us have a look at the historical policy compared to the optimal policy (see Table 5.4). As for the two other leaves, we notice that the optimal actions given by the model only match twice with the historical actions, that is, for the first and second quarters of year Y1.

	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2	Q3Y2	Q4Y2	Q1Y3
Historical actions	1	1	2	3	3	3	3	2	3
Optimal actions	1	1	1	1	1	1	1	1	1

Table 5.4: Historical actions and optimal actions given by the application of an MDP to the leaf of Tree 1 with average expected gain per opportunity between 5 and 10 million USD. The discretised action space is of dimensions 2×2, the number of possible occurences in each rank is 5, the distributions of the ranks are based on the data from the first quarter of year Y1 to the second quarter of year Y3 and the discount factor γ is equal to 0.9.

We noticed that, for the two first leaves presented in this section, the optimal policy lead to a lower investment for a majority of the states than the historical policy.

5.2 Variation of the dimensions of the discretised action space

After presenting results for actions determined by a 2×2 discretisation of the action space, we considered finer grids. First, we considered three variables for the investments for the new opportunities and two for the existing ones, then we created the opposite grid, that is, with two variables for the new opportunities and three for the existing ones. Finally, we considered a 3×3 discretisation of the action space.

We applied the model for the leaf of Tree 1 with average expected gain per opportunity below 1 million USD. The other parameters are the default ones presented at the beginning of this chapter, i.e. 5 possible occurences in each rank of a state, the rank distributions based on the data from the first quarter of year Y1 to the second quarter of year Y3 and a discount factor of $\gamma = 0.9$.

5.2.1 Action space with 3 variables for I_0 and 2 variables for c

Choosing a 3×2 discretisation of the action space allows for an intermediate investment for the new opportunities and keeps the binary high/low investments for the existing ones. Figure 5.11 shows the repartition of the $(I_0[t], c[t])$-couples, for $t = 1, \ldots, 9$, which determine 5 actions cells.

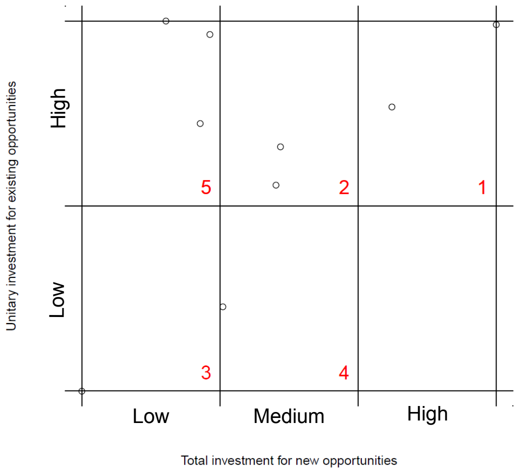


Figure 5.11: 3×2 discretisation of the action space. The number of actions is determined by the number of filled cells, for which action labels are given at their bottom right. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain below 1 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

Three actions were selected as optimal by the model, that is, actions 1, 3 and 4. The first two actions corresponds to the same actions that were selected for the model applied to the same leaf for the 2×2 discretisation of the action space. Moreover, histograms of the number of opportunities in each rank for the states for which either action 1 or action 3 were selected as optimal (see Figures 5.12 and 5.13) showed very similar trends as the ones for the 2 × 2 discretisation of the action space (see Figures 5.3 and 5.4). With a 3 × 2 discretisation of the action space, a third action (action 4) was selected as optimal for 27% of the states. This action corresponds to a middle investment for the new opportunities and a low investment for the existing ones. Comparing the histograms corresponding to optimal action 3 (see Figure 5.13) or optimal action 4 (see Figure 5.14), we notice that states with optimal action 4 have proportionally less opportunities in ranks 1 and 2 than states with optimal action 3. Hence, this justifies a slightly higher investment for the new opportunities, as few money should be invested for the existing ones, as they have not reached mature ranks yet.

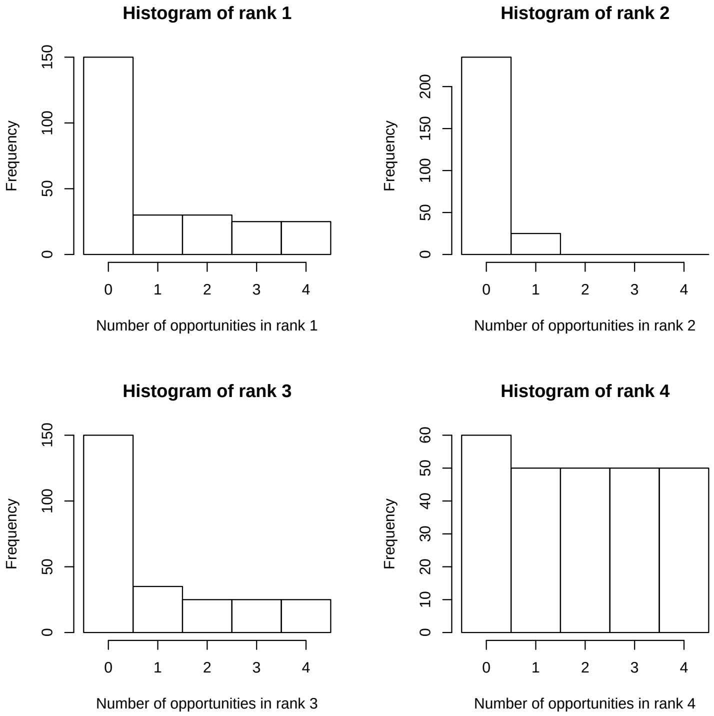


Figure 5.12: Histograms of the number of opportunities per rank for the states for which the best action is 1. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 3 × 2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

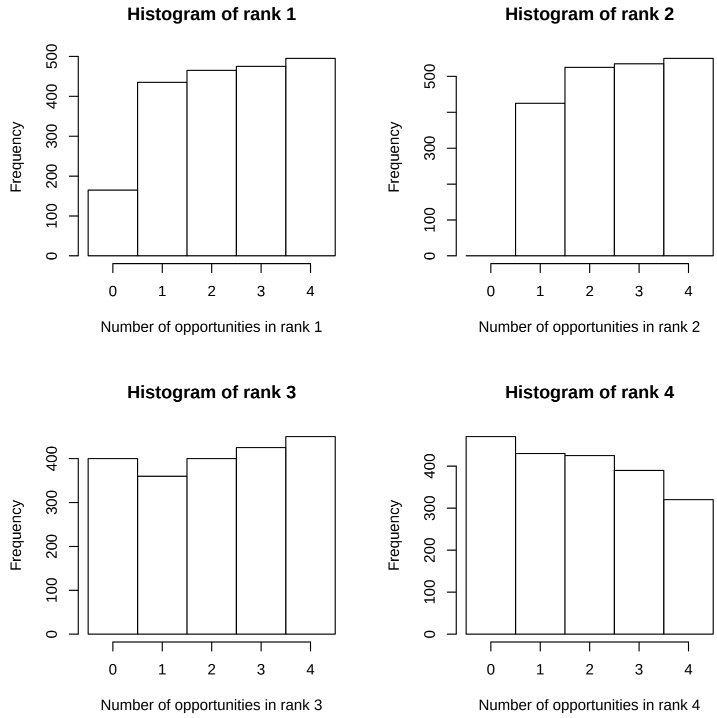


Figure 5.13: Histograms of the number of opportunities per rank for the states for which the best action is 3. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 3 × 2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

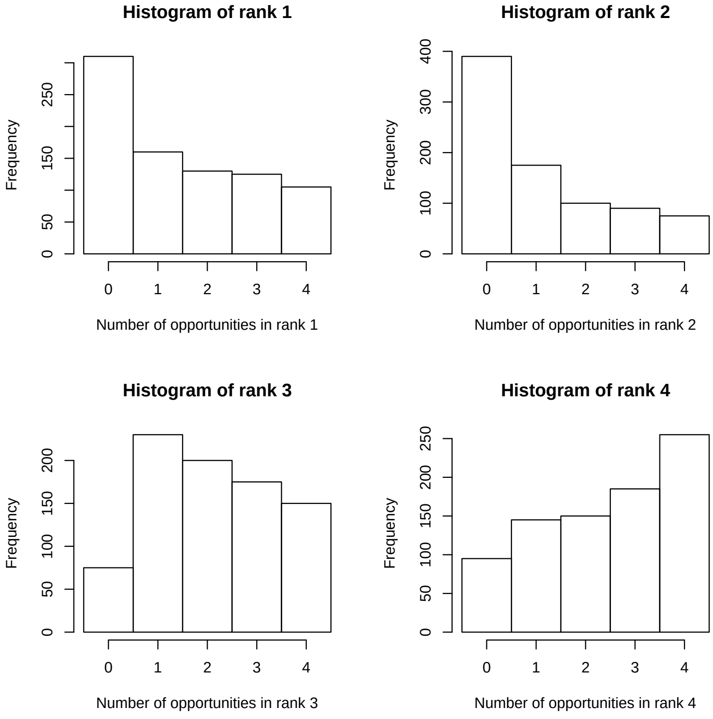


Figure 5.14: Histograms of the number of opportunities per rank for the states for which the best action is 4. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 3 × 2. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

In order to compare the historical policy with the one given by the model, Table 5.5 shows the historical actions and the optimal actions. For the states that appeared in the historical data, the action which is most chosen by the model is 3, i.e. a low investment for both existing and new opportunities. Two exceptions occurs, that is, for the first quarter of year Y1, for which the action chosen by the model is 4 and for the first quarter of year Y3, for which the action chosen by the model is 1. Apart from the first quarter of year Y3, let us note that the actions selected by the model always show lower investments than the ones of the historical data.

	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2	Q3Y2	Q4Y2	Q1Y3
Historical actions	1	1	2	3	4	5	5	2	5
Optimal actions	4	3	3	3	3	3	3	3	1

Table 5.5: Historical actions and optimal actions given by the application of an MDP to the leaf of Tree 1 with average expected gain per opportunity below 1 million USD. The discretised action space is of dimensions 3 × 2, the number of possible occurences in each rank is 5, the distributions of the ranks are based on the data from the first quarter of year Y1 to the second quarter of year Y3 and the discount factor γ is equal to 0.9.

5.2.2 Action space with 2 variables for I_0 and 3 variables for c

Let us now analyse the results of the application of the model to a 2 × 3 discretisation of the action space given in Figure 5.15. In the opposite way as for 3 × 2 discretisation of the action space presented in section 5.2.1, we include an intermediate investment for the existing opportunities and keep the binary high/low investment for the new ones. Four actions were identified.

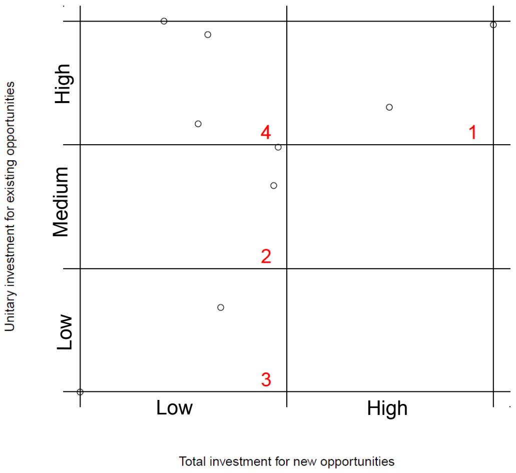


Figure 5.15: 2 × 3 discretisation of the action space. The number of actions is determined by the number of filled cells, for which action labels are given at their bottom right. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain below 1 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

As for the discretised action space of dimension 2 × 2, two of the optimal actions selected were a high investment for both new and existing opportunities (action 1) and a low investment for both new and existing opportunities (action 3). The histograms (see Figures 5.16 and 5.17) show the same trends as for the discretised action space of dimension 2 × 2.

In addition to optimal actions 1 and 3, a third action, i.e. action 4, was selected as optimal for 45 states. It corresponds to a high investment for the existing opportunities and a low investment for the new ones. Histograms for these 45 states are shown in Figure 5.18. Most of these states have a high number of opportunities in each ranks, especially ranks 2, 3 and 4. This justifies a high investment for the existing opportunities in this particular case.

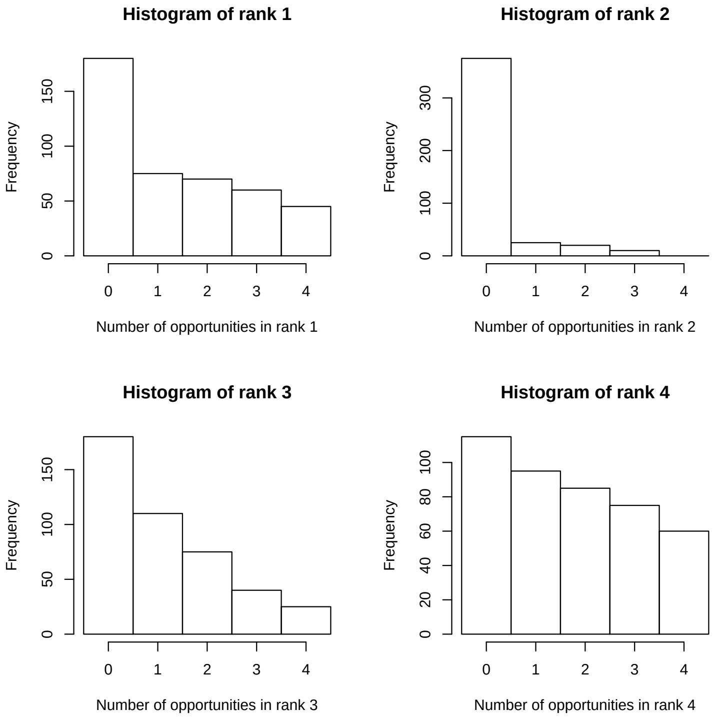


Figure 5.16: Histograms of the number of opportunities per rank for the states for which the best action is 1. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2×3. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

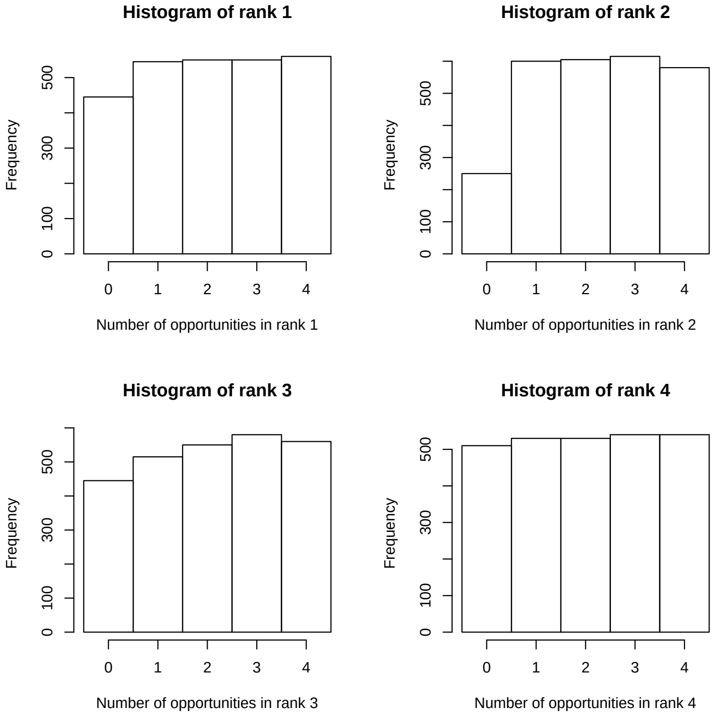


Figure 5.17: Histograms of the number of opportunities per rank for the states for which the best action is 3. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2 × 3. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

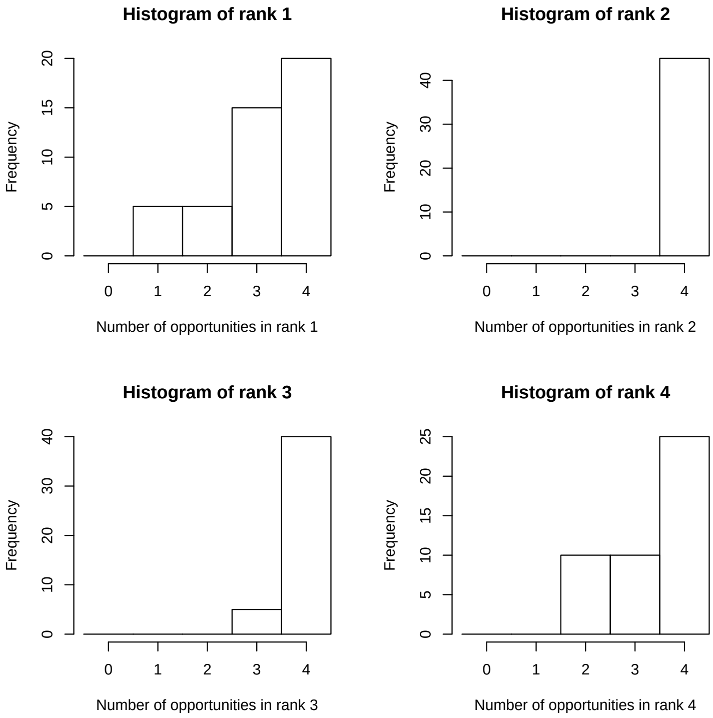


Figure 5.18: Histograms of the number of opportunities per rank for the states for which the best action is 4. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The action space discretisation is 2 × 3. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

Let us finally compare the investments given by the model with the historical ones (see Table 5.6). The optimal actions are the same as the ones given for a discretised action space of dimension 2 × 2 (see Table 5.2) and are always lower or equal to the historical ones, except for the investment on the first quarter of year Y3.

	Q1Y1	Q2Y1	Q3Y1	Q4Y1	Q1Y2	Q2Y2	Q3Y2	Q4Y2	Q1Y3
Historical actions	1	1	2	3	3	4	4	2	4
Optimal actions	1	3	3	3	1	3	3	3	1

Table 5.6: Historical actions and optimal actions given by the application of an MDP to the leaf of Tree 1 with average expected gain per opportunity below 1 million USD. The discretised action space is of dimensions 2 × 3, the number of possible occurences in each rank is 5, the distributions of the ranks are based on the data from the first quarter of year Y1 to the second quarter of year Y3 and the discount factor γ is equal to 0.9.

5.2.3 Action space with 3 variables for I_0 and 3 variables for c

The last discretised action space we present is of dimensions 3 × 3 and is plotted in Figure 5.19. It allows for three discrete values for both the investments for the existing opportunities and the new ones. Six actions were identified.

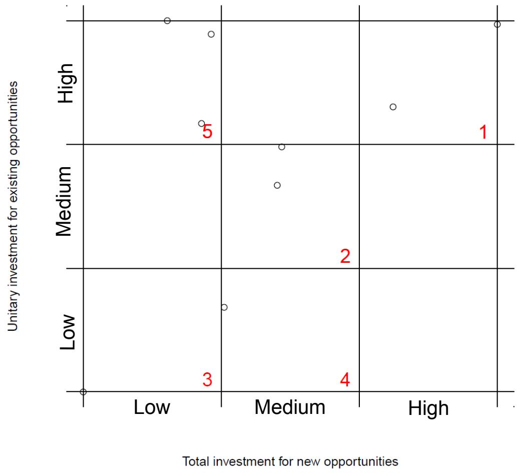


Figure 5.19: 3 × 3 discretisation of the action space. The number of actions is determined by the number of filled cells, for which action labels are given at their bottom right. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain below 1 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

Three types of optimal actions were selected by the model, actions 1, 3 and 4. As for all other meshes we considered, two of the selected actions were the one with the highest investment for the existing and new opportunities and the one with a low investment for the existing and new opportunities. The third action that was selected is action 4 and corresponds to a middle investment for the new opportunities and a low investment for the existing ones. The optimal actions that were identified are exactly the same as the ones identified for the discretised action space of dimension 3 × 2. Moreover, the states for which each optimal action was selected matched exactly.

This enables us to conclude that a discretised action space of dimension 3 × 3 has not brought any further precision to the outcome of the model than a discretised action space of size 3 × 2 for the leaf of Tree 1 with average expected gain per opportunity below 1 million USD.

5.2.4 All historical actions considered individually

Previously, we have always grouped actions with close values using a mesh. Let us consider the case where all historical actions for the nine quarters are considered individually. A plot of the $(I_0[t], c[t])$ investment couples is shown in Figure 5.20 in the example of the leaf of Tree 1 with average expected gain below 1 million USD.

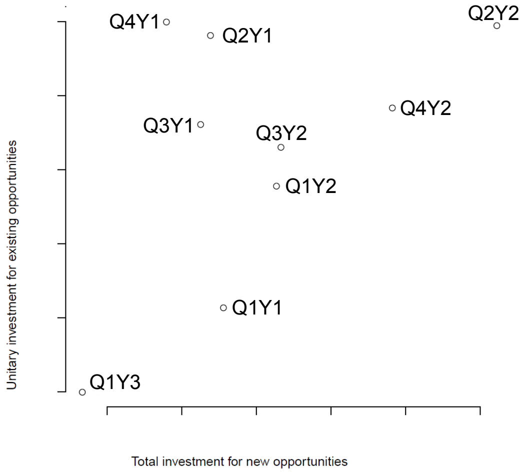


Figure 5.20: Repartition of the investments of the 9 quarters. The leaf to which the grid corresponds is the leaf of Tree 1 with average expected gain below 1 million USD. The investment for the new opportunities is represented by the x-axis and the investment for the existing ones is represented on the y-axis.

Two optimal policies were selected by the model, i.e. the one corresponding to the investment of the second quarter of year Y2, for a few states, and the one corresponding to the investment of the first quarter of year Y2, for a larger number of states. Histograms of the number of opportunities in each state are plotted in Figure 5.21 for the states with optimal action corresponding to the investment of the second quarter of year Y2 and Figure 5.22 for the states with optimal action corresponding to the investment of the first quarter of year Y2.

Let us note that the action that was selected as optimal for most of the states is the one corresponding to a middle investment for both existing and new opportunities (see Figure 5.20). This was not identifiable when we were binning all actions in a 2×2 discretised action space. Therefore, the precision brought by considering all quarterly investments as individual actions lead to a slighly different optimal policy. Let us recall that the optimal action for most of the states with the 2×2 discretised action space was to invest the lowest amount of money for both new and existing opportunities.

For a smaller number of states, the optimal action was to invest the highest amount of money for both new and existing opportunities. The histograms of Figure 5.21 show the same trend as the ones for optimal action 1 of the 2×2 discretised action space (see Figure 5.1). Thus, in the case of that particular action, considering all quarterly actions individually has not brought much precision.

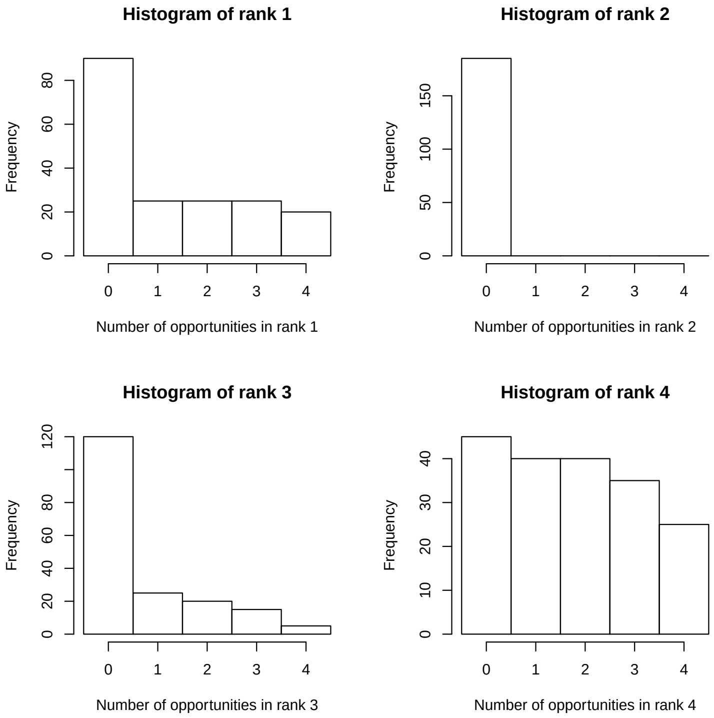


Figure 5.21: Histograms of the number of opportunities per rank for the states for which the best action is the investment of the second quarter of year Y2. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

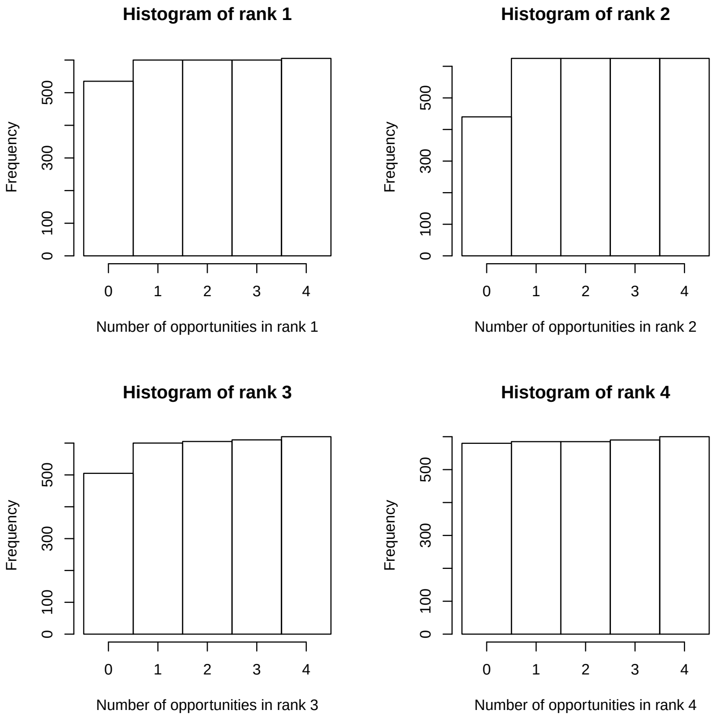


Figure 5.22: Histograms of the number of opportunities per rank for the states for which the best action is the investment of the first quarter of year Y2. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 4 corresponds to a high number of opportunities. The number of possible occurences in each rank is 5 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

5.3 Variation of the number of possible occurences in each rank

In this section, we vary another parameter, which is the number of possible occurences in each rank. So far, we have considered a number of occurences equal to 5. Hence, in the previous histograms, the bins were numbered from 0 to 4. Here we applied the MDP to the data from the leaf of Tree 1 with average expected gain below 1 million USD for a number of occurences in each rank equal to 4 or 6.

In the case of 4 occurences per rank, the histograms of the number of opportunities in each rank showed the same trend as the ones for 5 occurences in each rank (see Figures 5.3 and 5.4), but with less precision.

When we allow 6 occurences per rank, the histograms of the number of opportunities per rank show a greater precision (see Figures 5.23 and 5.24). For example, a clearer tendency to the lowest opportunity bin (label 0) can be noticed in the histogram of the number of opportunities in rank 3 for the states that have optimal action 1 (see Figure 5.23).

It would be interesting to apply the model for higher numbers of occurences per rank. Indeed, the precision of the histograms would be greater as the number of states increases. For example, with a number of occurences per rank of 6, the number of states becomes $6^5 = 7'776$, which is about 2.5 times the number of states with a number of occurences per rank equal to 5. Unfortunately, if we increase it even more, it becomes computationally unfeasible.

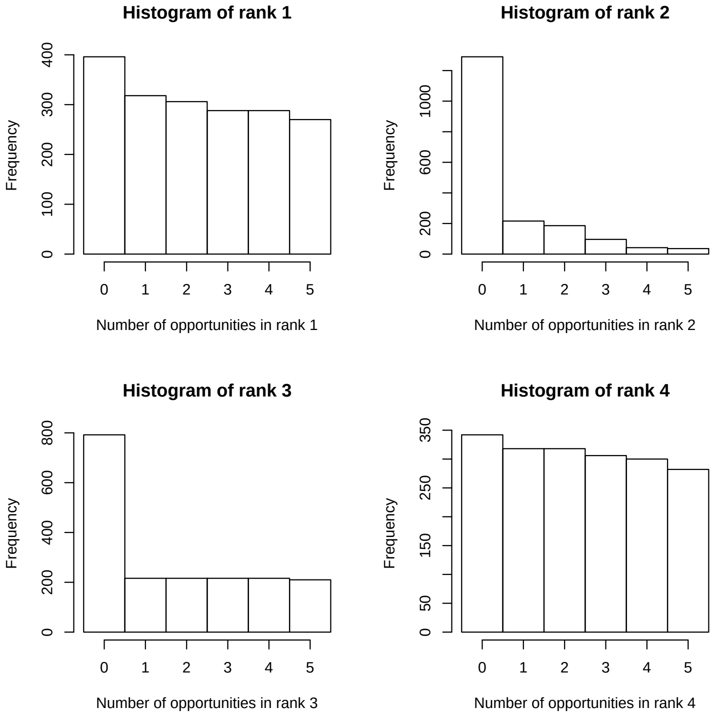


Figure 5.23: Histograms of the number of opportunities per rank for the states for which the best action is 1. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 5 corresponds to a high number of opportunities. The action space discretisation is 2 × 2. The number of possible occurences in each rank is 6 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

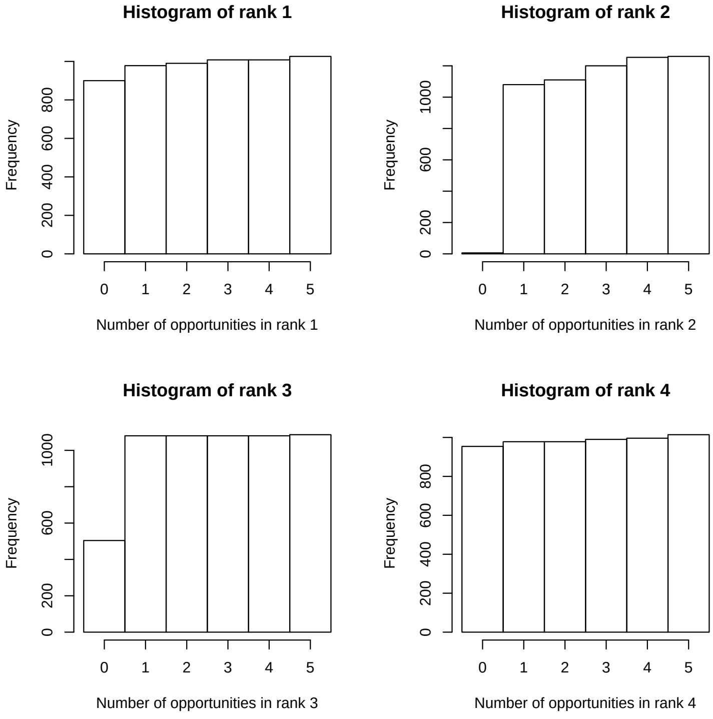


Figure 5.24: Histograms of the number of opportunities per rank for the states for which the best action is 3. The leaf to which the histograms correspond is the leaf of Tree 1 with average expected gain below 1 million USD. The label 0 corresponds to a low number of opportunities and the label 5 corresponds to a high number of opportunities. The action space discretisation is 2 × 2. The number of possible occurences in each rank is 6 and the number of opportunities in each rank are binned according to the data from the first quarter of year Y1 to the second quarter of year Y3. The chosen discount factor is $\gamma = 0.9$.

5.4 Variation of the state space discretisation

We mentionned in the definition of our default parameters that the distribution of the number of opportunities per rank on which the quantile computation is based comes from the data from the first quarter of year Y1 to the second quarter of year Y3. Let us now consider the distributions of the number of opportunities excluding the second quarter of year Y3, as all other elements of the model are computed using the data from the nine first quarters only.

In this section, we present the results of the application of the model using these new distributions to the three leaves of Tree 1.

For the leaf with average expected gain per opportunity below 1 million USD, each state of the historical data is mapped via ϕ_{state} to a projected version using a similar table of quantiles such as Table 5.1. But this time, the range of the number of opportunities per rank is much tighter than the one for the quantiles of the distributions including the data of the second quarter of year Y3 as all the data refering to the latter is discarded. A unique optimal action was selected by the model for all states, i.e. action 3, which corresponds to a low investment for both existing and new opportunities. This action is the one that already occurs for most of the states with the previous action binning. Hence restricting the range of the number of opportunities per rank leads to less variation is the number of possible optimal actions.

For both other leaves of Tree 1, a unique optimal action is selected for all states as well. For the leaf with average expected gain per opportunity between 1 and 5 million USD, the best action is to invest the lowest amount of money for both new and existing opportunities (action 2). This action is the most frequently selected by the model for a distribution of the number of opportunities per rank including the second quarter of year Y3.

For the leaf with average expected gain per opportunity between 5 and 10 million USD, the optimal action is to invest the highest amount of money for both new and existing opportunities and it similarly to the two other leaves, it is the most frequently selected by the model when we consider the previous distribution of the number of opportunities per rank.

5.5 Variation of the discount factor γ

So far, we had set the discount factor γ to 0.9. Let us now vary this parameter in order to see which impact it has on the model. A discount factor close to 1 implies that a rather high importance is given to rewards far in the future and a discount factor close to 0 means that only rewards close in time have an impact on the model. In this section, we presents results of the application of the MDP to the leaf of Tree 1 with average expected gain below 1 million USD and for a discount factor of 0.8, 0.95 and 0.99 and compare them with the results for a discout factor of 0.9. All other parameters are the default ones.

For all four discount factors, the most frequent optimal action is 3, i.e. investing the least amount of money for existing and new opportunities, and the other possible optimal action is 1, i.e. investing the highest amount of money for existing and new opportunities. The difference is that the number of states for which either action 3 or action 1 is selected varies, if we change parameter γ. The number of states with optimal action 3 decreases as γ increases. This is summarized in Table 5.7.

γ	Action 1	Action 3
0.8	300	2825
0.9	430	2695
0.95	505	2620
0.99	555	2570

Table 5.7: Number of states with optimal actions 1 or 3 for different values of the discount factor γ after applying the MDP to the leaf of Tree 1 with average expected gain below 1 million USD.

This results implies that if we give importance to rewards far in the future, for a higher number of states, the best action will be to invest the highest amount of money. Let us finally note that varying the discount factor γ does not change a lot the trend of the histograms. Hence, we do not plot them here as they are very similar to the ones for $\gamma = 0.9$ (see Figures 5.3 and 5.4).

Chapter 6

Conclusion

This work showed an application of an MDP to the dynamics of sales opportunities. In each category of data to which the model was applied, from one to three different kinds of policies were selected as optimal for various types of states. Moreover, two types of policies seemed to show some recurrence across categories: a cautious policy which consists of not investing a lot of money when the number of opportunities is quite uniform across ranks and a bolder strategy that consists of investing more money when part of the opportunities have reached a more mature rank and the other part has just entered the pipe.

Increasing the size of the discretised action space lead to the apparition of an additional policy. This was the case for the action space allowing three types of investments for the new opportunities, i.e. low/medium/high, and two types of investments for the existing ones, i.e. low/high, compared to the action space with two types of investments for the new and existing opportunities, i.e. low/high. But when we included three types of investments for the existing opportunities as well, the policy did not change.

Considering the quarterly investments as individual actions did not increase the number of different optimal actions in the best policy, but it contributed to a higher precision of the investments than with a mesh discretisation.

Different types of state discretisation gave different policies. Restricting the range of the distributions of the number of opportunities per rank lead to the selection of only one type of action as optimal. Hence, it seems more relevant to keep a wider range for these distributions.

Finally, we noticed in the example of the leaf of Tree 1 with average expected gain below 1 million USD that increasing the discount factor lead to the selection of the action for which the investment is the highest for a larger number of states. This implies that if we want to give more importance to signings further in the future, we have to invest more.

Generally, the historical policy does not match much the policy selected by the model. Very often, the policy selected by the model involves lower investments than the historical one. This shows that the investment strategy could be improved and lower amount money could be spent for an optimal strategy. But this should be confirmed by historical data over a longer period than nine quarters.

The state space we considered is very large and this generated two major issues. First, the analysis of the optimal actions selected for all states can only be performed using histograms summarizing common properties of the states for which this particular action was selected. Second, the simulation of the transition matrix was computationally intensive. To remedy these two issues, we plan to use another algorithm to solve MDPs, which is the SARSA algorithm of reinforcement learning. Moreover, we will consider a different state space of a much smaller dimension, for which each state is a rank and not a combination of number of opportunities across the different ranks.

Bibliography

[1] Bertsekas, D. P., *Dynamic programming and optimal control*, Athena Scientific, Belmont, Massachusetts, 2000.

[2] Garcia, F. et al., *Markov Decision Process (MDP) Toolbox v2.0*, Biometry and Artificial Intelligence Unit of INRA Toulouse, 2005.

[3] Gerstner, W., *Neural networks and biological modelling*, Lecture Notes, 2009.

[4] Hillier, F. S., Lieberman, G. J., *Introduction to operations research*, McGraw-Hill Inc., New York, 1995.

[5] Kaelbling, L. P., Littman, M. L., Cassandra, A. R., *Planning and acting in partially observable stochastic domains*, Artificial Intelligence, 101, pp. 99-134, 1998.

[6] Monahan, G. E., *A survey of Partially Observable Markov Decision Processes: Theory, Models and Algorithms*, Management Science, 28, pp. 1-16, 1982.

[7] Nielsen, L. R., Kristensen, A., R., *Finding the K best policies in finite-horizon Markov decision processes*, Dina Research Report, 110, 2004.

[8] van Otterlo, M., *Markov Decision Processes: Concepts and Algorithms*, Course on 'Learning and Reasoning', 2009.

[9] Puterman, M. L., *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & sons Inc., New York, 1994.

[10] Singh, S. P., Sutton, R. S., *Reinforcement learning with replacing eligibility traces*, Maching Learning, 22, pp. 123-158, Kluwer Academic Publishers, Boston, 1996.

[11] Sutton, R. S., Barto, A. G., *Reinforcement learning: An introduction*, MIT Press, Cambridge MA, 1998.

[12] Sutton, R. S., Precup, D., Singh, S., *Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning*, Artificial Intelligence, 112, pp. 181-211, 1999.

[13] Tirenni, G. R., *Allocation of Marketing Resources to Optimize Customer Equity*, PhD Dissertation of the University of St. Gallen, 2005.