PRODUCT-LINE PLANNING UNDER UNCERTAINTY

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ŞAKİR KARAKAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
INDUSTRIAL ENGINEERING

JULY 2018

Approval of the thesis:

PRODUCT-LINE PLANNING UNDER UNCERTAINTY

submitted by **ŞAKİR KARAKAYA** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Director, Graduate School of **Natural and Applied Sciences** _______________

Prof. Dr. Yasemin Serin
Head of Department, **Industrial Engineering** _______________

Prof. Dr. Gülser Köksal
Supervisor, **Industrial Engineering Dept., METU** _______________

Examining Committee Members:

Assoc. Prof. Dr. Seçil Savaşaneril Tüfekci
Industrial Engineering Dept., METU _______________

Prof. Dr. Gülser Köksal
Industrial Engineering Dept., METU _______________

Assoc. Prof. Dr. Ayşe Kocabıyıkoğlu
Sabancı School of Management, Sabancı University _______________

Asst. Prof. Dr. Sakine Batun
Industrial Engineering Dept., METU _______________

Asst. Prof. Dr. Özlem Çavuş İyigün
Industrial Engineering Dept., Bilkent University _______________

Date: 24/07/2018

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Şakir Karakaya

Signature :

ABSTRACT

PRODUCT-LINE PLANNING UNDER UNCERTAINTY

Karakaya, Şakir
Ph.D., Department of Industrial Engineering
Supervisor: Prof. Dr. Gülser Köksal

July 2018, 355 pages

This study addresses the problem of multi-period mix of product-lines under a product- family, which incorporates launching decisions of new products, capacity expansion decisions and product interdependencies. The problem is modelled as a two-stage stochastic program with recourse in which price, demand, production cost and cannibalisation effect of new products are treated as uncertain parameters. The solution approach employs the Sample Average Approximation based on Monte Carlo bounding technique and multi-cut version of L-shaped method to solve approximate problems efficiently, which is tested on different cases considering VSS and EVPI performance measures. The data collected through two experimental studies is analysed using ANOVA and Random Forest methodology in order to understand which problem parameters are significant on the performance measures and to generate some rule-based inferences reflecting the relationship between significant parameters and the performance of the proposed stochastic model.

Keywords: Product-mix, Product-line, Capacity Planning, New Product Introduction, Product Interdependencies, Two-stage Stochastic Program, Sampling Average Approximation, L-shaped

ÖZ

BELİRSİZLİK ORTAMINDA ÜRÜN-GRUBU PLANLAMASI

Karakaya, Şakir
Doktora, Endüstri Mühendisliği Bölümü
Tez Yöneticisi: Prof. Dr. Gülser Köksal

Temmuz 2018, 355 sayfa

Bu çalışmada, yeni ürünlerin pazara sunulma zamanlarını, kapasite artırım kararlarını ve ürünler arası etkileşimleri içeren, bir ürün-ailesi altında yer alan ürün-grubu karması problemi ele alınmıştır. Ürün taleplerinin, satış fiyatlarının, birim üretim maliyetlerinin ve yeni ürünlerin mevcut ürünlerin satış hacmini azaltmaya yönelik etkisinin rassal parametreler olarak tanımlandığı bu problem, iki seviyeli stokastik programlama modeli olarak ele alınmıştır. Modelin çözümü için Monte Carlo sınırlama tekniğine dayalı ve stokastik modellerin etkin çözümü için L-shaped (Benders ayrışım) algoritmasını kullanan bir Örneklem Ortalaması Yakınsaması yaklaşımı geliştirilmiştir ve bu yaklaşımın performansı geliştirilen çeşitli problemler üzerinde test edilmiştir. Bu amaçla geliştirilen iki deneysel tasarımdan elde edilen veriler, problem parametrelerinin çözüm performansı üzerinde etkisini anlayabilmek ve önerilen stokastik yaklaşımın performansı ve anlamlı problem parametreleri arasındaki ilişkileri gösteren kurallar üretebilmek amacıyla ANOVA ve Random Forest metodolojileri kullanılarak analiz edilmiştir.

Anahtar Kelimeler: Ürün-karması, Ürün-grubu, Kapasite Planlama, Yeni Ürün Sunma, Ürün Etkileşimleri, İki-seviyeli Stokastik Programlama, Örneklem Ortalaması Yakınsaması, L-shaped.

To My Mom and Dad

ACKNOWLEDGEMENTS

I am grateful to my supervisor Prof. Dr. Gülser Köksal for her professional support, guidance and encouragement throughout the completion of this thesis work.

I would also like to express my sincere appreciation to Asst. Prof. Dr. Sakine Batun and Assoc. Prof. Dr. Ayşe Kocabıyıkoğlu who contributed to this thesis with valuable suggestions and comments.

Love and thanks to my family and my friends, particularly Yücel İnce, Erdoğan Cevher, Mustafa Özkubat, Alia Hindawi, Seiran Aliyev, Sedef Meral, Mahmut Özdemir, Yıldız Artar and Yücel Özkara for their never-ending patience, support and encouragement.

Ankara, July 2018

Şakir Karakaya

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

This study handles a product-mix problem that is described at strategic level and in the scope of product planning. This problem is formulated based on the observations obtained from industry and the related literature, and deals with the decisions used as an input to medium- and long-term production plans. Since it is identified under the product planning domain, in this chapter firstly, a general framework for product planning is drawn and then the motivation behind this study and the research questions of this study are presented.

Product planning is defined as a function including market analysis, competitive product analysis, pricing analysis, manufacturing analysis and decision-making process on whether a new product is introduced to identified markets (Ricci, 2012). *Market analysis* aims at getting information about potential customers, market size and trends in the market for existing products as well as new product requirements, which provide information for product portfolio of the firm (*Ibid*). *Competitive product analysis* focuses more on new products in defined markets and includes differentiation, market positioning and competition strategies for each product in the portfolio. Selling a leading product in new markets, establishing new marketing channels, improvement of existing/old products and developing new products for new and existing customers are some examples of those types of strategies (Thomson, 2014). Each of those strategies has a different impact on *product plans*, which are the main outputs of product planning activity. *Pricing analysis* aims at developing pricing strategies for new products, considering the information provided by market and competitive analysis; and *manufacturing analysis* focuses on determining product costs, expected profits and sales targets (Ricci, 2012). At the end of this process, the firm also develops its investment and resource allocation plans, product roadmaps considering the whole portfolio (existing and improved/new products) and execution plans. Prioritisation and timing of new product development projects, and technology selection are also some other decisions made within the context of product planning (Chen et al., 2006). Thomson (2014) proposes a well-structured and detailed product planning process integrated with the firm's corporate strategies, shown in Figure 1.

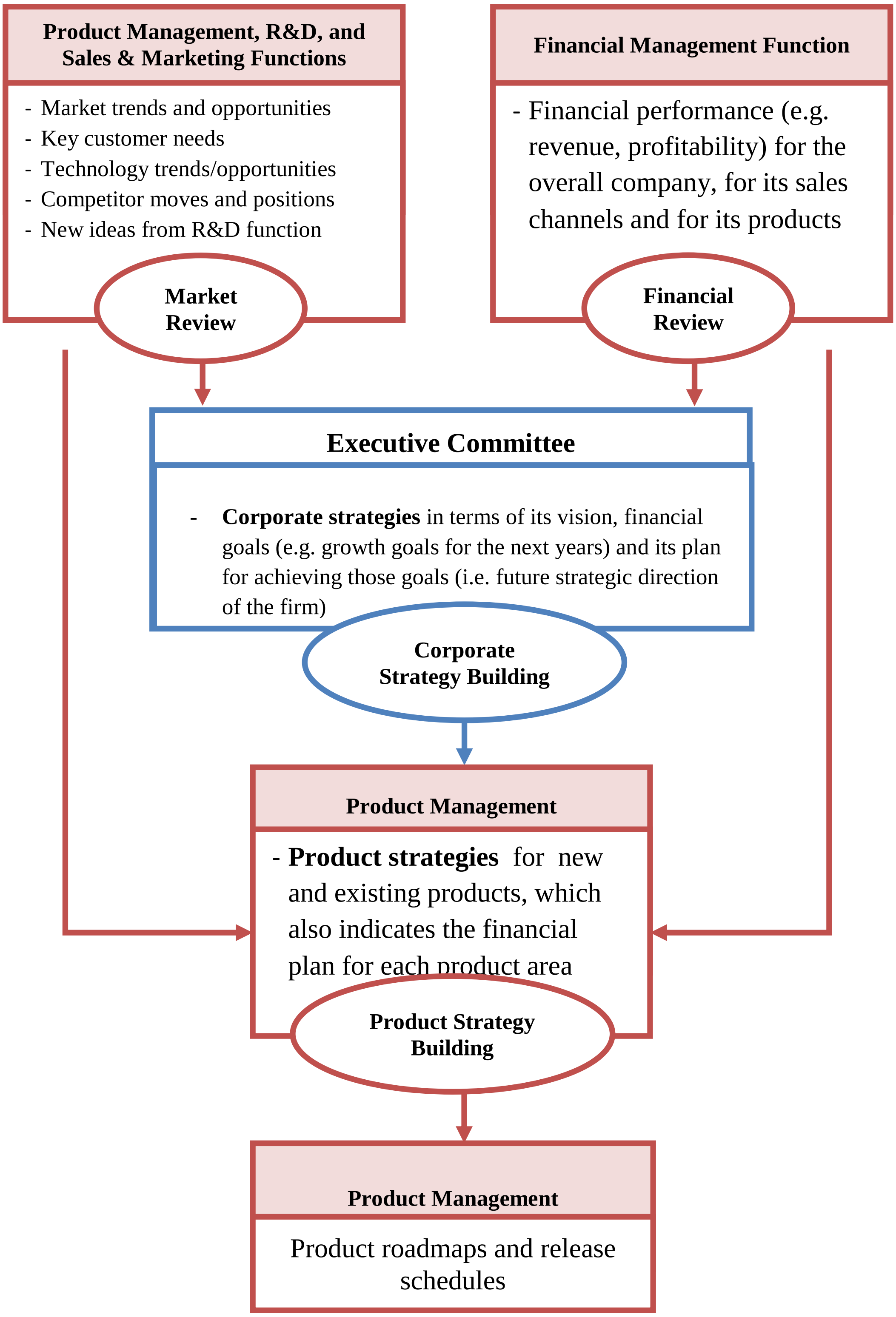


Figure 1.The flow of product planning process, adapted from Thomson (2014)

As shown in Figure 1, the product planning process consists of five steps: (1) Market review, (2) Financial review, (3) Corporate strategy building, (4) Product strategy building and (5) Product roadmaps and release schedules development.

In the first step, product management, and sales and marketing functions of the firm provide data about market trends and opportunities, key customer needs, technology trends and competitor moves and positions to the executive committee for supporting its corporate strategy building process. Meanwhile, research and development (R&D) function provides new ideas and new product/technology concepts. In the second step the financial department (function) presents the results of financial performance for overall company, for its sales channels and for its products, also for supporting the executive committee's strategy-building process. Then, the executive committee develops the corporate strategies specifying what changes to the products are required and indicating the financial plan for each product area (e.g. generating 50% of the next period revenue from new products, increasing the local market share by 10%, or increasing the overall profit by 30% etc.). In the fourth step, the product management department develops product strategies based on the corporate strategies as well as considering market dynamics, customer needs and the financial plan. In the last step, product roadmaps (or product plans) and also release schedules over a planning horizon are developed by the product management department as consistent with the corporate and product strategies built in the previous steps. It should be noted that although the product planning process displayed in Figure 1 may have some differences in the companies operating in different types of sectors, the main flow of this process remains the same. For instance, we have observed the same process in one of the leading Turkish companies operating in consumer durables goods sector, when we have interviewed the product manager of the company. Furthermore, the product planning process continues with activities such as product development/improvement, introducing products to target markets and post-launch product management activities, which include the implementation of product plans. At this point, it should be noted that the entire product plans of companies should be evaluated based on possible trade-offs between the decisions regarding those activities, before going to the implementation phases.

As mentioned above, the main outputs of a product planning process are product plans (a.k.a product roadmaps in companies having a well-structured and formal product management function) and release schedules specified in a timeline.

A product roadmap is a visualisation tool that shows the main decisions about a company's future product portfolio. Those decisions may roughly be on which existing products will be kept, which of them will be improved, when new products will be developed and introduced to target markets, which products will be manufactured in-house or outsourced (Albright and Nelson, 2004). As an example, Albright and Nelson (2004) present a product roadmap for a passenger car, developed based on product-platform in Figure 2.

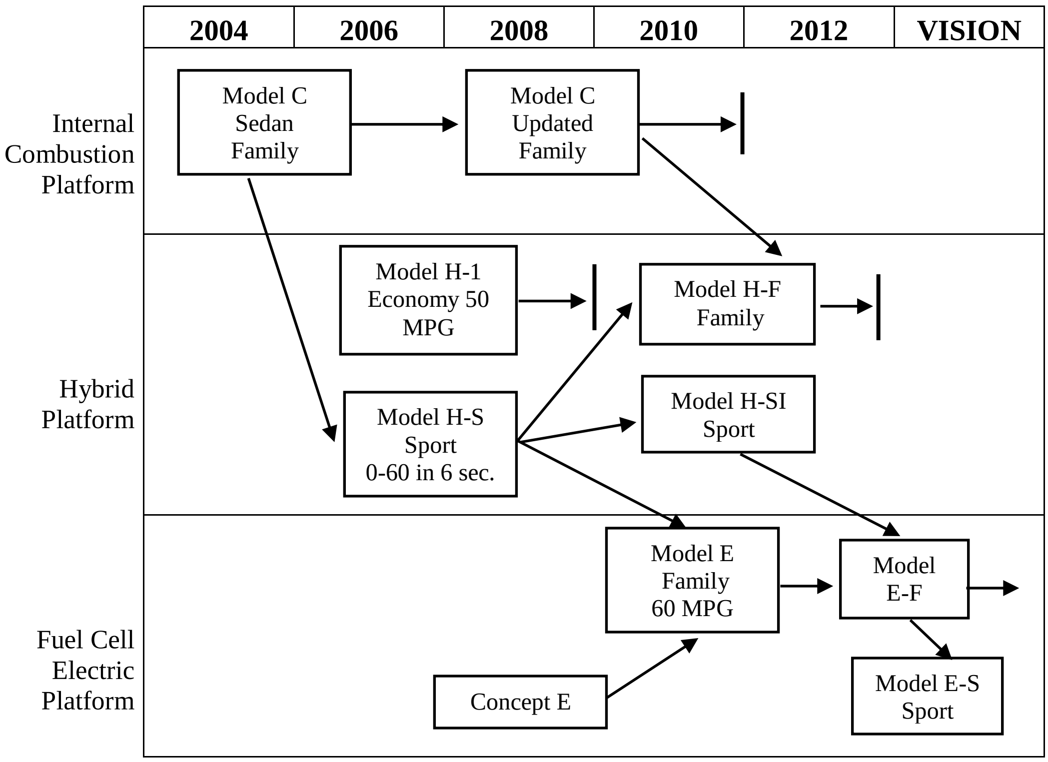


Figure 2. A product roadmap for a passenger car

The roadmap displayed in Figure 2 shows the plans of an automobile manufacturer over the time for its specific product. For instance, the firm plans to start up a concept development project for fuel cell electric platform in 2006 and to introduce a new product using that platform in 2010. Similarly, it plans to modify/renew the existing product, model C sedan family segment, using internal combustion platform and introduce that modified/renewed product to the market in 2006, and then that product will be deleted from the market in 2010. Furthermore, Model E is seen to be the basis of a continuing product line with two variations, and to consolidate the line, model C will be phased out in 2010 and the product using the hybrid platform, Model H, will be deleted from the market in 2012. This product roadmap also includes a vision at the end of planning horizon, which describes the ultimate goal for the product line.

Considering this roadmap, it should be noted that product roadmaps which currently display the plan of a firm's product lines need to be revised periodically according to changes in the dynamic market. For instance, market conditions might require the launching time of some products to be postponed or scheduled to an earlier date, or to extend the planned life of a product or delete this product in an earlier time. Another example might be related with adding a new market to the roadmap. Additionally, it should be noted that the product roadmap shown in Figure 2 is organized according to a key product driver which impacts the roadmap the most, e.g. powertrain platform in the given example. However, as stated by Albright and Nelson (2004), a product roadmap might also be organized according to any critical dimension of product performance or market segments. An example for a product roadmap developed based on a single platform and market segments is shown in Figure 3.

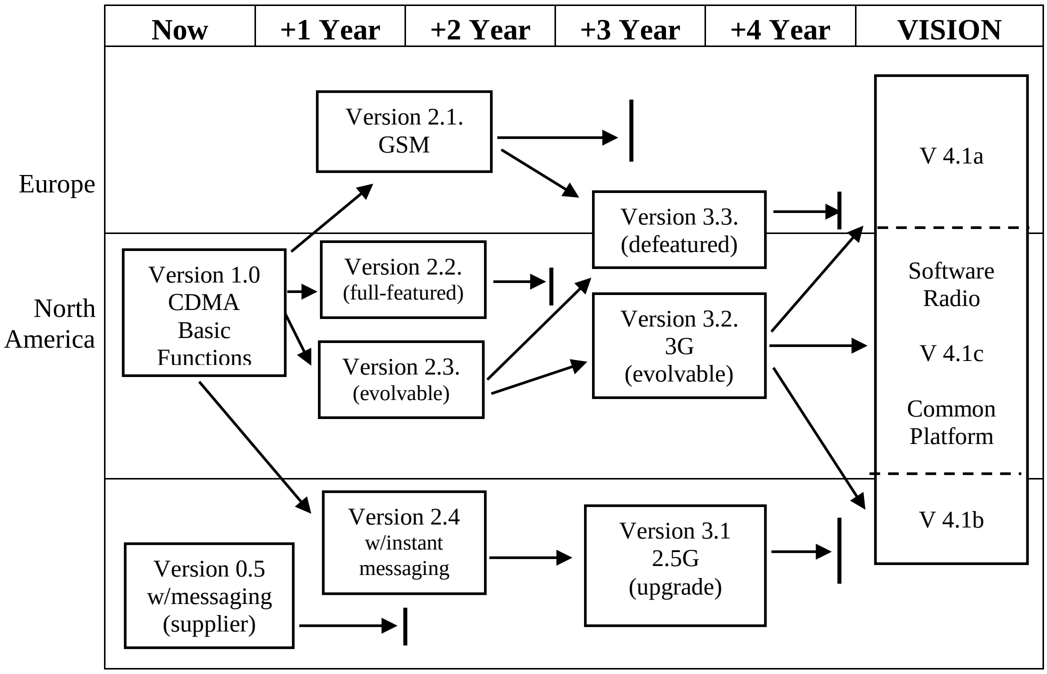


Figure 3. A product roadmap developed based on a single platform and market segments (Albright, 2002)

As seen from the example given in Figure 3, the platform roadmap begins with a single model designed for the North American market, while the needs of the Asian market are fulfilled with a model sourced from a supplier. Furthermore, the firm plans to introduce to the European market next year and to be in that market for nearly three years.

A product roadmap/plan is used as a nested source of the following information, each of which is considered for each period of a planning horizon and is linked to corporate strategies of the company:

- Target markets where the products would be sold,
- Which products to be developed and when to introduce to the target markets,
- Which products will be sold in which periods,
- Which product development projects are planned and when to be launched,
- Which products will be developed in-house or outsourced,

- The planned (or maximum) lives of products.

It should be noted that those kinds of information provided by roadmaps are used as inputs for the optimisation problem handled in this study.

Searching out the evolution of products, product selection, project evaluation and prioritization and visualizing all those decisions over a planning horizon are the main activities in a roadmapping process, which is the core of product planning framework, given in Figure 1. Market search analysis, quality function deployment, checklists, scoring models such as analytic hierarchy process and analytic network process, benefit measurement models based on subjective assessment, economic models particularly used in portfolio investment analysis, some specific models developed by consultancy firms, for instance, Boston Consulting Group Matrix, directional policy matrix and GE Nine Cell Matrix, strategic buckets and mathematical-based models such as linear, integer and dynamic programming are some examples of tools and methods used in the roadmapping process. Since the optimisation problem in this study deals with the outputs of the roadmapping process, i.e. roadmaps and release schedules, rather than how to generate roadmaps, the details of the roadmapping process as well as tools and techniques used for will not be discussed further in this study.

Product roadmaps are key sources for product portfolio management[1] and provide information that might be used for assessing trade-offs in a product portfolio (Albright and Nelson, 2004) and also for measuring the value created by a roadmap. Additionally, difficulties associated with allocating firm's scarce budget over multiple periods and multi-products, usually interdependent products competing for the financial (budget) and physical (e.g. production capacity) resources and

---

[1] "Product portfolio" and "product mix" are interchangeably used in the literature for describing set of all products of the company. However, there exists a main difference between them. While product portfolio describes all the products which a company offers to market, product-mix also describes the entire products but "in a specific period" (e.g. in the next three years) and also including the information about "product volumes" (i.e. how many of each product are to be sold).

uncertain nature of problem environment (e.g. uncertainties in future market demand and selling price) make the product portfolio management a challenging issue in practice. However, to the best of our knowledge, there is no study dealing with those difficulties in an integrated manner. Therefore, a decision support tool is needed to provide an opportunity for decision makers and also product managers to assess the trade-offs regarding product plans and to fine-tune/balance those plans by the way of determining best product portfolio/mix as well as balancing all outflows (e.g. investments for capacity and product development) and inflows (i.e. revenues) over a planning horizon. Furthermore, as Chen et al. (2014) stated, the decisions on the allocation of resources should be considered jointly with capacity expansion decisions in order to have the most profitable product-mix. A decision maker might also measure the created value (i.e. contribution to the firm's strategic goals regarding with market share, profit etc.) of prepared product roadmaps using this decision support tool which is developed in this study.

The goals of this tool are defined as follows:

1. Maximise the value (e.g., benefit, profit, strategic goals of firm) of product portfolio defined on roadmaps by assessing the trade-offs among capacity required, new product introducing plans, product deletion plans, new capital investment plans etc.,
2. Achieve a balance among the decisions regarding with existing and new/planned products, e.g. when to launch a new product, based on firm's future objectives,
3. Allocate financial and physical resources among the products in mix, i.e. adopting production capacity according to changing market conditions as well as firm future objectives,
4. Measure the created value of prepared product roadmaps.

Within the context of those goals, the research problem handled in this study deals with optimising the product mix of a firm at strategic-level, considering all existing and new/planned products, price or demand interdependencies among those products, target markets, all periods over a predetermined planning horizon (multi-period) and capacity expansion decisions, in order to have a balanced product roadmap, in other words strategic product plans of the firm. Thus, the decision support tool is designed to provide answers to the following questions:

1. Which products and how many of them will be made available in which (target) market in the next periods (e.g. 5 years)?
2. When will the planned (new) products be launched (i.e. release plans/time-to-market decisions)?
3. When and how much capacity will be added in each period within the planning horizon?
4. How will the capacity be allocated for each product in each period?

Those decisions are strongly interrelated, but the studies in literature have solved those problems separately. Except one study (Yilmaz et al., 2013), most of those handle capacity planning and product mix decisions together, but they ignore new product launching decisions. Regarding all those studies, an extensive literature review is presented in the next chapter.

The decision support tool is actually based on a mathematical model and a solution approach developed for this model in order to answer the questions above. In order to develop this tool, the following studies are performed in this thesis work:

- A detailed literature survey that is presented in Chapter 2 and a couple of interviews with product managers of four different manufacturers in Turkey (two of them operates in consumer durables sector, one operates in consumer goods including personal and household care products, one produces and sells batteries for vehicles) are done in order to understand the problem context well enough and formulating the problem properly. Then, considering product hierarchy and planning-level within the scope of management, main decisions within the problem context, i.e. product-mix and new product launching decisions and capacity expansion decisions, product interdependencies and parameter uncertainties in the problem environment, the problem is defined as presented in Chapter 3.

- Based on the problem definition, modelling environment and relevant assumptions, a mathematical model in which all parameters are deterministic is developed and then extended to a two-stage stochastic program with recourse in order to handle the uncertain parameters, i.e. demand, price, variable production cost and cannibalisation rate. In order to solve this stochastic problem with a manageable number of scenarios, multi-cut version of L-shaped method based on Benders' decomposition is employed. Both the deterministic and stochastic models, and L-shaped method are given in Chapter 3.
- Because of problem size and assumptions on random parameters, a tractable solution approach based on sample average approximation, bounding technique and random sampling method, as seen Chapter 4, is developed.
- Both the capability of the solution approach developed in Chapter 4 and the quality of solution obtained through the stochastic programming are tested via a computational study including nine illustrative cases. All problems generated for this study are coded in GAMS 22.2 and solved using CPLEX 10.0 solver. The results are presented in Chapter 5.
- The problem handled, the modelling and solution approach, and the results obtained through the computational study are discussed and accordingly future research directions are given in Chapter 6.

A summary including how the problem is handled from definition to solution is displayed in Figure 4.

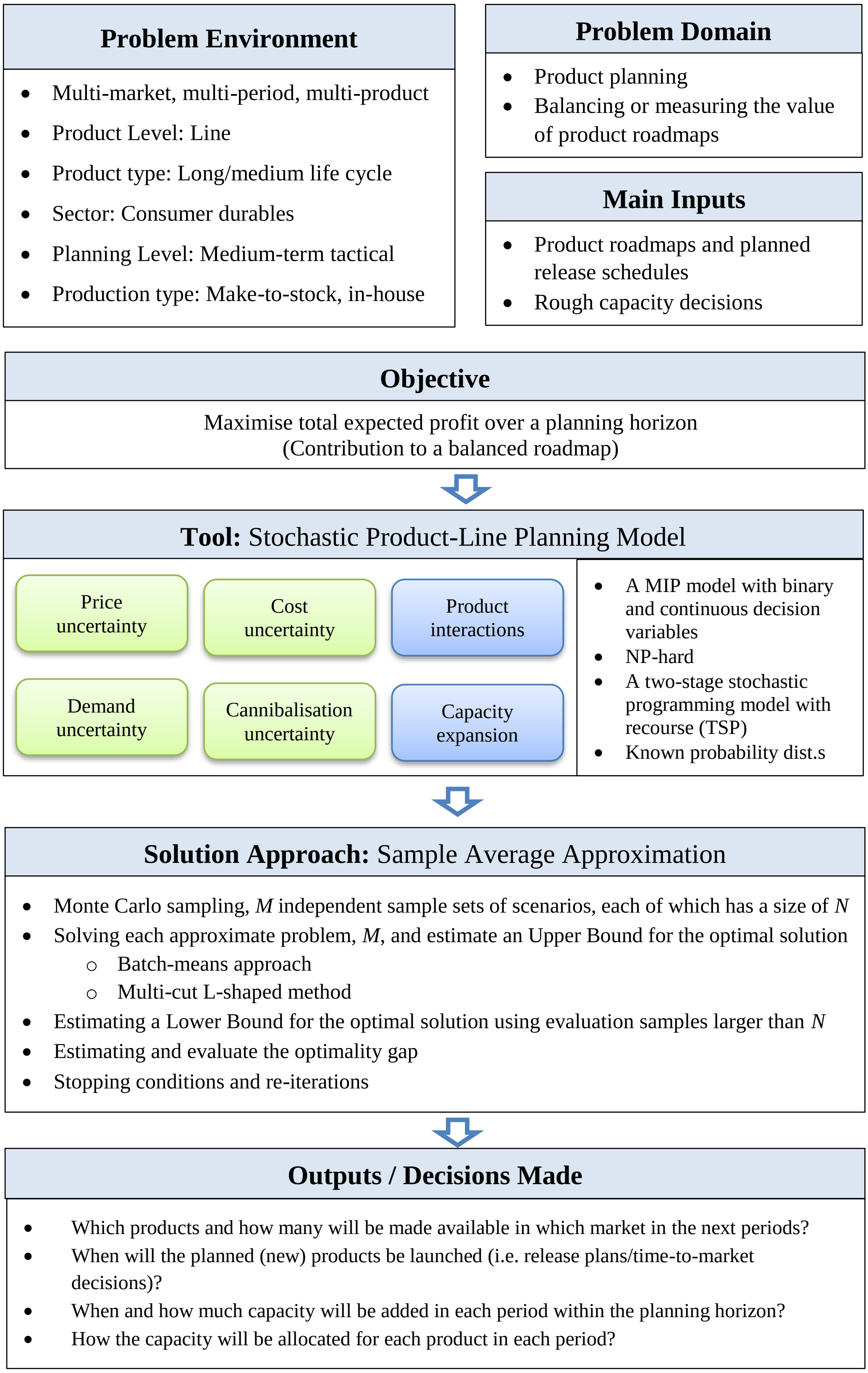


Figure 4. A summary of problem definition, modelling and solution approach, and main outputs

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

Considering the general problem context presented in the previous chapter, the problem handled can be described as an extension of multi-period "product-mix" problem, which is commonly studied in production management framework as a tactical or an operational problem, including new product launching and capacity expansion decisions (timing and sizing) issues.

"Product-mix" is defined as the set of all products offered by a company for sale in a certain period. Traditionally, a product-mix problem (a kind of resource allocation problem) includes the decisions on which product should be selected and how many of each should be produced in order to maximise the financial performance of a firm under demand and resource constraints (Sobreiro et al., 2014; Lea, 2007). In the Operational Research literature, the basic/traditional version of this problem, in which all parameters are deterministic and no capacity planning or product launching decisions are considered, is formulated using a mathematical model as follows:

Sets:
I : set of products that can potentially be produced, $i = 1,2, \dots, k$

Parameters:

p_i : unit selling price of product $i \in I$

c_i : unit production cost of product $i \in I$

cap_i : unit capacity usage of product $i \in I$

TC : total production capacity of product $i \in I$

d_i : total demand for product $i \in I$

Decision Variables:

y_i: number of units for product $i \in I$

Objective Function:

$maximize \ \sum_{i \in I}(p_i - c_i) \times y_i$

Subject to:

Capacity constraints: $\sum_{i \in I} cap_i \times y_i \leq TC$

Demand constraints: $y_i \leq d_i, \quad i \in I$ and $y_i \geq 0$ and integer

There are different types of this basic model, for instance, instead of maximising the profit a decision-maker may want to minimise the cost or to maximise the total revenue for the objective function of the model. The model can also be extended for multiple periods and multiple markets, including capacity acquisition decisions, uncertainties regarding some parameters such as demand and price, environmental considerations etc. (e.g. Tai et al., 2015; Hasuike and Ishii, 2009a; Hasuike and Ishii, 2009b; Wang et al., 2007; Alonso-Ayuso et al., 2005; Letmathe and Balakrishnan, 2005; Küttner, 2004).

Although the product-mix problem is studied since the end of 1950's (Küttner, 2004), to the best of our knowledge, there is no prior work that addresses multi-period and multi-market product-mix problem including the timing decisions of planned products of a firm and capacity expansion decisions, and taking product interdependencies and uncertainties associated to problem parameters into account. In accordance with this problem framework, there are a few related studies in the literature as summarised below.

Yilmaz et al. (2013) propose a multi-period model for a long-term capacity planning problem which considers short-life cycle products and their renewals. This study is a notable exception to the existing product-mix problems and the most related to the research problem handled in this study. Decisions for which products and how many of them will be produced by which technology in each planning period, whether to launch a new product in a planned period, and how much and when to expand the warehouses' capacities are considered in the model. However, this study does not explain the linkage between the problem considered, and product roadmaps. All parameters are considered as deterministic and the objective of the model is to maximise the cash balance at the end of the planning horizon, taking into account sales revenues, residuals of production technologies, warehouse assets in the final period, value of final stock and costs. In that study, they assume that the products are selected from a predetermined candidate list (including all existing and future products) and the production technologies are selected from a predetermined set (including all existing and new technologies). They also take cannibalization effects among the candidate products into account. Though this study is the most relevant one to the problem handled in this study, it assumes that all parameters are deterministic rather than considering a stochastic environment where market demand, price as well as costs are uncertain as in most of the real-life problems. Besides, there are significant differences between that study and ours in terms of objectives, constraints and assumptions in the model. Compared to traditional product-mix problems and their extensions including capacity planning decisions in the literature, this study is the unique one which incorporates new product plans over multiple periods into the problem framework. Furthermore, most of the product-mix determination and capacity planning studies given in this chapter assume that there exists a product set (can be considered as candidate set) which is determined at the beginning of planning horizon and does not change over periods. In our study, similar to that model, products that will be produced/sold for the next periods are selected from a set of candidate products. The product set displayed on the product roadmaps (also including alternative products) for each period can be considered as the candidate product set for our problem.

Mishra et al. (2017) study a problem optimising pricing and capacity planning decisions jointly. They consider a two-product and two-period model in which demand is considered as a stochastic parameter. A product is handled within a product-line (see Section 2.1 for product-line definition), and optimal pricing and capacity expansion decisions are made for a product-line setting through introducing a new product and deleting an existing product over time. This study actually contributes to the pricing and capacity management regarding product-line design literature.

Hasuike and Ishii (2009a) consider a multi-period and multi-criteria product-mix problem whose objective is to minimise total cost, to maximise total profit and to minimise inventory levels in an uncertain environment. Incorporating both randomness and fuzziness into the model is the focus of that study.

Hasuike and Ishii (2009b) present a single-period product-mix problem under uncertainty, including both fuzziness and randomness. They propose two models for this problem: (1) an optimisation model that maximises the total profit as much as possible while satisfying the probability of that the total profit is greater than or equal to a target value and the possibility that total cost is less than or equal to a target value, using chance constraints and (2) a model whose objective is to maximise the satisfaction level of decision maker who controls the total risk.

The last two studies given above (Hasuike and Ishii, 2009a and 2009b) are the unique studies dealing with product-mix models in which the uncertainty is modelled using both stochastic and fuzzy approaches in an integrated way.

Lin et al. (2007) develop a mixed integer programming (MIP) model for capacity and product mix planning problem, which can be solved iteratively under the objective of maximising the total contribution margin (=revenue - variable production cost - inventory costs - depreciation loss costs). In that model in which all parameters are deterministic, product mix and production quantity as well as the amount of additional capacity for each product group in a specific plant in each period are determined.

Letmathe and Balakrishnan (2005) take into account some environmental issues while making decisions on optimal product-mix. In this study, they take into account the threshold values for emissions specified by the regulations, the penalties and taxes that may occur because of exceeding those thresholds, emission allowances that can be exchanged between companies and product recycling policies while making product-mix decisions. The main objective of their study is to decide on the optimal product-mix which maximises the profit of a company, which equal to total revenue gained by selling products plus net revenue generated by exchanging emission allowances minus resource acquisition costs minus penalties occurred because of exceeding emission thresholds.It should be noted that this study is one the most important studies which incorporates the environmental issues (factors) into the product-mix decisions.

Alonso-Ayuso et al. (2003) present a model for supply chain planning problem at strategic level, which consider the decisions regarding the size/capacity of plants, mix of products in each plant, supplier selection for raw materials and production, stocking and transhipment volumes, under uncertainty in a multi-period environment. In this study, demand and selling price of the products, supplying cost of raw materials and production cost are handled as uncertain parameters.

Morgan and Daniels (2001) develop a mixed integer model that handles product-mix and new technology adoption decisions (i.e. deciding on the mix of incumbent and advanced tehnologies) jointly. In that model, they consider the interrelationship between the mix of products of a company and the decisions regarding the selection of technology that is required to manufacture products in the mix. The objective of the model is to maximise the profit over a relevant time horizon.

Stuart et al. (1999) develop a multi-period MIP model for selecting product and process alternatives in a deterministic environment, which take environmental considerations (including environmental costs, variables and constraints) into account. That study is a first attempt to consider environmental issues in product-mix decisions. The objective of the model is to maximise total net revenue gained from the sales of products in the mix and recyclable waste minus total net cost including overhead, variable, inventory (assembly and take-back inventory costs) and waste disposal costs.

Monroe et. al. (1976) develops a model for product-mix problem in marketing context to maximise the net cash flows. They take revenue interactions resulting from the coexistence of two products in a market into account and permit the addition and deletion of products over a multi-period planning horizon. All parameters are handled as deterministic and constraints regarding with product introduction, withdrawal and interaction effects are considered. However, budget, technology and capacity constraints are not considered.

Among those studies given above, the most related ones with our study can be summarised in Table 1.

Table 1. A summary of the most related product-mix studies in the literature

Paper	Capacity Planning	Uncertainties	New Product Launching	Product Interactions	Other
Yilmaz et al. (2013)	✓		✓	✓ (cannibalisation effect between successive generations)	• Short-life-cycle products • Warehouse capacity planning • Technology selection
Mishra et al. (2017)	✓	✓	✓		• Pricing decisions • 2-products, 2-periods • Uncertain: Demand
Hasuike and Ishii (2009a)		✓			• Randomness and fuzziness
Lin et al. (2007)	✓				• All parameters are deterministic
Alonso-Ayuso et al. (2003)	✓	✓			• Size/capacity of plants • Supplier selection • Uncertain: demand, price and cost
Monroe et. al. (1976)			✓	✓	• Marketing context • No capacity constraints
Our study	✓	✓	✓	✓	• Multi-market, Multi-period • Uncertain: demand, price, cost, cannb.rate

Based on the related studies given in Table 1, it is noteworthy to say that the model developed in this dissertation, which addresses a multi-period and multi-market product-mix problem including the timing decisions of planned products of a firm and capacity expansion decisions, and taking product interdependencies and uncertainties associated to problem parameters into account, is handled for the first time in this study. Additionally, this study is the first attempt to make a link between the product roadmaps (PRM) of a firm and described problem, which aims at balancing the PRM.

The characteristics of the problem handled in this study, which are given in Chapter 3, can be presented based on the following sub-titles underpinning the problem definition:

1- Product hierarchy and planning-level within the scope of management
2- Capacity expansion decisions
3- Product interdependencies
4- Parameter uncertainties in the problem environment

Therefore, in addition to the background, the related literature is also reviewed in accordance with those sub-titles and reported in Section 2.1-2.4.

2.1. Product Hierarchy and Planning-Level within the Scope of Management

A product-mix is generally defined as all the products offered by a company for sale within the context of this study. For better understanding the problem context and formulating the problem properly, we had a couple of interviews with product managers of four different manufacturers in Turkey; two of them operates in consumer durables sector, one operates in consumer goods including personal and household care products, one produces and sells batteries for vehicles. The first significant experience acquired from those interviews is to understand the need for clarifying the meaning of “product” while referring to any product-mix problem. A product might imply a family, a line, a model or a variant and the problem context differs according to this description; thereby the first issue at the beginning of this study was to decide on the product description. In the related literature as well as in real-life, those kinds of descriptions, such as product family, product line, model etc., are organised within a "product hierarchy" based on a grouping or aggregation strategy and the problem context and accordingly definition of product-mix is determined in accordance with the product level selected from this hierarchy. A product hierarchy lists out different levels to identify products based on consumer needs and items offered to consumers to fulfill their needs (Kotler and Keller, 2012). Grouping products in a business area generally starts with family level and ends up with item/variant/SKU from marketing and product management point of view. Those levels organised under a product hierarchy (Figure 5) are presented as follows (*Ibid.)*:

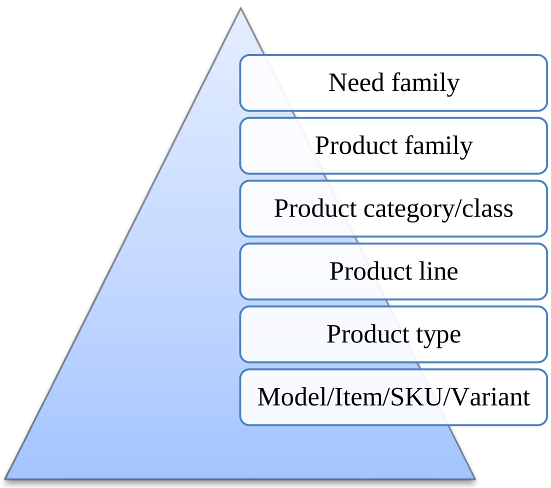


Figure 5. Product hierarchy developed by Kotler and Keller (2012)

Need Family is defined as the core need being fulfilled by any production or sales activity of a firm operating in any business area.

Product Family is defined as the set of product categories/classes satisfying a need family and the first product grouping level in the hierarchy.

Product Category (Class) is described as "a *group of products within the product family recognized as having a certain functional coherence*" (Kotler and Keller, 2012, p.336). A category includes all those objects that are close substitutes for the same needs despite differences in size, shape and technical characteristics (Govindarajan, 2007).

Product Line is defined as the set of products within a product category that are closely related, because of similar functionality, selling to same customer group or through the same channels or falling within certain price ranges (Kotler and Keller, 2012, p.336).

Product Type refers to "*a group of items within a product line that share one of several possible forms of the product*" (Kotler and Keller, 2012, p.336). If there is no such form, this level can be ignored and the models are defined under a product line directly.

Item/SKU/Variant/Model is a distinct unit under a product line that is seperable by some product attributes such as size, colour, price etc.

A real-life example of product hierarchy:

As an example, a real firm in Turkey, which operates in three different business areas of consumer durables sector, is considered: (1) white goods, (2) consumer electronics and (3) small household appliances, and different needs of customers are satisfied by the products grouped under each area. For instance, under the white goods area, food protecting by cooling, washing clothes, washing dishes and cooking needs of a house are met by the products of "cold family", "wet family" and "hot family", respectively. That relationship is shown in Figure 6.

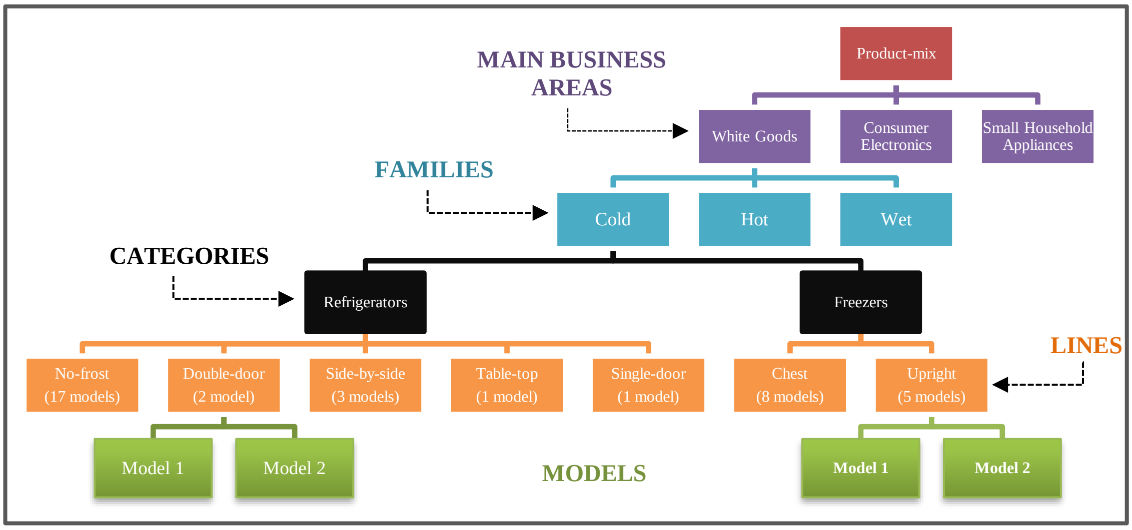


Figure 6. A real-life example for product hierarchy

Need Family: *Protecting foods from deteriorating*, defined as a need family by a firm operating in consumer durables sector, whose main business area is "producing white goods" for people.

Product Family: *Cooling and freezing devices (i.e. Cold Family)*

Product Category (Class):

Category 1: Refrigerators (designed for cooling and chilling)

Category 2: Freezers (designed for freezing)

Product Line:

Family: Cold

Category1: Refrigerators

Product Line 1: Single-door

Product Line 2: Double Door

Product Line 3: No-frost

Product Line 4: Side-by-Side

Product Line 5: Table Top

Here, the product lines are formed based on two criteria: (1) configuration (door type and size) and/or (2) cooling system (static vs. no-frost technology). For instance, the products having a static cooling system and two doors are grouped as "double-door" line; on the other hand the product having no-frost cooling system and two doors are grouped as "no-frost" line. The other lines are basically formed according to their configuration, e.g. the fridges having one door, regardless of its cooling system, are grouped as "single-door" line and the fridges like wardrobes and having a no-frost cooling system are grouped as "side-by-side" line.

Category 2: Freezers (Lines are formed based on customer groups)
Product Line 1: Chest (for workplaces such as butcher, bakery and fisher)
Product Line 2: Upright (for houses)

Product Type:

Cold Family
Category 1: Refrigerators
Product Line 2: No-frost
Product Type 1: with top-mounted freezer
Product Type 2: with bottom-mounted freezer

Item/SKU/Variant/Model:

Cold Family
Category 1: Refrigerators
Product Line 2: No-frosts
Product Type: with top-mounted freezer
Item 1: a white colour, 650 l cabinet volume, A++ energy class no- frost with top-mounted freezer
Item 2: a white colour, 475 l cabinet volume, A++ energy class no-frost with top-mounted freezer

Another example from tyre manufacturing industry is given in Appendix A.

The firm in this example has a large set of products, some of them are **long/medium life cycle (LMLC) products**, i.e. white goods and small household appliances, and others are short life cycle (SLC) products, i.e. such as mobile phones, TVs, computers etc. Technological changes are not so frequent and the rate of technological advance is not so rapid, e.g. airplanes, communication infrastructure, refrigerators, dishwashers, small households etc. for LMLC products. Those remain in use relatively longer periods compared to SLC products. Besides, since LMLC has relatively stable demand, it is possible to forecast demand based on products' sales history. However, technology for SLC products rapidly changes and because of high obsolescence, they are replaced with more technological new products within a few years after their introduction to markets (Prabhaker and Sandborn, 2013). Besides, forecasting demand is more difficult because of high uncertainty in the markets, and managing the performance of successive generations of SLC products and short-term marketing activities becomes vital for those products (Lee et al, 2006).

The product level considered in any product-mix problem might affect the main properties of the problem as well as mathematical model used for problem formulation. For instance, interactions among products in terms of demand and price are more frequent at line or item level than class or family level. Therefore, it needs to give more attention to include the product interactions into mathematical model at line or item level; on the other hand those effects might be ignored if the mix of classes or families are considered. In another example, modelling issues for capacity planning may also be changed. If one deals with determining the product-mix of models within a line (a.k.a *product-line design* on which there are too many studies in the literature, see e.g. Mohit et al., 2017; Liu et al., 2017; Bertsimas and Mišić, 2017; Müller and Haase, 2016; Anderson and Celik, 2015; Luo, 2011; Schön, 2010; Belloni et al., 2008; Tucker and Kim, 2008; Fruchter et al., 2006; Jiao and Zhang, 2005; Li and Azarm, 2002; Morgan et al., 2001; Chen and Haussman, 2000; Kohli and Sukumar, 1990), the investment decisions for new production technologies or large-scale capacity decisions such as opening a new plant would be out of concern. On the other hand, determining the mix of lines within a class, then medium-scale capacity decisions such as expanding warehouse capacity need to be represented in the model.

Another important issue that should be taken into account is if the products considered under a mix are financially evaluated together. For instance, on the condition that a firm does not monitor and evaluate the financial performance of its product classes or families under the same policy, it does not need to manage the mix of classes with together. However, since the product-lines under a class might use the same kind of capacity and their performance are monitored as a whole, a problem for optimizing the mix of product-lines becomes reasonable.

In addition to the factors given above, which should be taken into consideration while formulating any product-mix problem, another important issue is to define the relationships between the product level, planning context, decision levels also organised within in a hierarchical structure and planning horizon under consideration. Based on the related literature (e.g. Benton, 2014; Collier and Evans, 2009; Huh and Roundy, 2005; Miller, 2002; Chase et al., 1998; Buffa and Sarin, 1987; Ahrens, 1983) and the information obtained from interviews with industrial experts working in industrial organisations, three main hierarchical planning problems are described in order to define those kinds of relationships within the context of our problem:

1. Long-term Strategic Planning Problem
2. Medium-term Tactical Planning Problem
3. Short-term Tactical Planning Problem

Long-term Strategic Planning Problem (LSPP):

This problem handled at strategic decision level and in a long-term basis covers strategic decisions regarding with construction of a new plant, resource acquisition, supply chain design for a new product, determining which products will be manufactured and where etc. (Chou et. al., 2007; Huh and Roundy, 2005; Chen et al., 2002; Stuart et.al., 1999;). Those kinds of decisions are made at the highest level of the hierarchy of decision-making activities and are expected to affect long-term plans of the firm. LSPP concerns about long-term organisational objectives and decisions necessary to make those objectives real by providing proper resources (Miller, 2002).

Product aggregation/grouping level: Families and classes, e.g. refrigerators and freezers under cold family. For instance, assume that the firm in the example above does currently not sell freezers. The decision for introducing freezer category to any market is made at this level as a strategic decision.

Inputs used for decision-making process: profitability of families or classes, capacities of existing plants, main target markets, plant locations/relocations, growth and investment plans for families as well as categories, business dynamics in the sector, large-scale capacity expansion plans (acquisition, outsourcing, etc.), opportunity costs, financing alternatives, externally provided data such as competitors' capacity levels and total forecasted demand for whole industry (Miller, 2002) etc.

Decisions made (inputs for the medium-term tactical planning problem): when new capacity will be available (both expanded capacity of existing facilities and new capacity acquired from new plants), the planned capacity levels (existing and new plants) for each period, profit (or revenue) goals for a family and each class under this family, allocation of planned capacity to plants and to classes under the family, new markets to be entered and existing markets to be pulled out and their timing plans, decisions on launching a new class and on deleting an existing class under the family, technology migration from an old technology to the new technology decisions (Chien and Zheng, 2012), end-product outsourcing decisions at class level, opening new facilities (and their locations) for supporting new products, what plants should produce what and how (make-to-stock, make-to-order or assemble-to-order) etc.

<u>Medium-term Tactical Planning Problem (MTPP):</u>

This problem handled at tactical decision level and in a medium-term basis covers decisions regarding with allocations of production capacity, which is determined for a class in LSPP, to product-lines under this class, which product-lines to be introduced into target markets and their volumes in order to achieve the profit (revenue) goal determined at strategic level.

Product aggregation/grouping level: Product lines under a class.

Inputs used for decision-making process (provided by LSPP): total available (allocated) capacity for each class at each period, the goals defined in LSPP, expected profit determined for each class, expected number of products under each class to be sold for each period, new classes to be introduced, medium-scale capacity adjustment requirements etc.

Decisions made (inputs for the Short-term Tactical Planning Problem): allocation of production capacity to product lines, launching decisions for new lines, deleting decisions for existing lines, outsourcing decisions, medium-scale capacity adjustment decisions such as tool/machine purchasing, decommission (Chou et al., 2007) etc.

Compared to the capital investment decisions made at strategic level (i.e. LSPP), it should be noted that the capital-type decisions made at this tactical level are commonly smaller-scale (Miller, 2002). For instance, while setting a new plant up is related to the LSPP, decisions on buying a warehouse, adding a new production line to the factory, tool/machine purchasing etc. are considered at this tactical level as medium-scale capacity adjustments.

Short-term Tactical Planning Problem (STPP)

This problem also handled at tactical decision level as MTPP but in a short-term basis (typically one year) covers decisions regarding with determination of which models, aggregated under each product line, are sold in which markets, which new models are added to the mix, their volumes etc. in order to achieve the profit (revenue) goal determined at medium-term tactical level.

Product aggregation/grouping level: Models under the lines / Product Line Design problems in the related literature dealing with decisions such as optimal pricing, order quantities, differentiation, mix of product attributes, product-line extension, market positioning etc. and considering models such as conjoint, consumer preference etc. (see, e.g. Mohit et al., 2017; Liu et al., 2017; Bertsimas and Mišić, 2017; Müller and Haase, 2016; Anderson and Celik, 2015; Luo, 2011; Schön, 2010; Belloni et al., 2008; Tucker and Kim, 2008; Fruchter et al., 2006; Jiao and Zhang, 2005; Li and Azarm, 2002; Morgan et al., 2001; Chen and Haussman, 2000; Kohli and Sukumar, 1990).

Inputs used for decision-making process (provided by MTPP): total allocated capacity for each line at each period, goals for profit or revenue for each line determined in MTPP, expected number of products under each line to be sold for each period, new lines to be introduced into the mix etc.

Decisions made (inputs for the Aggregate Production Planning Problem): new model launching and existing model deleting decisions as a result of ensuring product variety in order to match customer needs, outsourcing decisions, smaller-scale capacity adjustment decisions typically at machine and process level while overall capacity is mostly fixed (Chou et al., 2007), differentiation, pricing, positioning and other marketing decisions etc.

It should be noted that decisions made at this level would be main input for aggregate production planning (including production volumes in each period, e.g. typically a month, item/model production schedules, work force levels, minor equipment planning, inventory levels, subcontracting etc.) and marketing.

2.2. Capacity Expansion Decisions

Capacity planning has recognised as a main stream in literature and becomes a challenging problem for organisational performance in practice (Benton, 2014; Collier and Evans, 2009; Zhang and Wang, 2009; Geng et al., 2009; Chou et. al., 2007; Miller, 2002; Chase et al., 1998). Decisions regarding timing particularly in the case of long lead-time required for new capacity installation as well as decisions for amounts of capacity increments make capacity planning tasks important but also difficult for the industries particularly including technologically complicated and capital intensive fabrication processes because of volatility of market demand due to uncertainties, rapid changes of technology, and long lead time and high cost for capacity increments (Vespucci et al., 2016; Alaniazar, 2013; Chien and Zheng, 2012 and Geng et al., 2009). Those factors may cause the capacity expansion decisions, which are costly and difficult to change later, to be under the risk of capacity shortage and surpluses along with time (Bish and Wang, 2004; Alaniazar, 2013). Furthermore, Ahmed et al. (2003) and Huang and Ahmed (2009) state that the economies of scale in capacity costs and the uncertainties in the costs make capacity expansion problems more complex. Therefore, in order to deal with those complexities quantitative approaches for capacity planning has been an intensive research area since the 1960s (Ahmed et al., 2003). Among those, mathematical programming-based approaches using linear, stochastic programming, simulation models etc. become more handy in the last decades (Chien and Zheng, 2012) in order to handle capacity planning problems mainly including capacity expansion and product-mix decisions together. Since capacity and product-mix planning decisions should be considered integratedly in the scope of our problem (Chou et al., 2007) in order to adjust/expand capacity to meet market demand of products in the mix, the studies in the related literature handling capacity and product mix planning together are presented elaborately in this section.

In the related literature, it can be seen that the studies dealing with both product mix and capacity planning decisions are highly focused on semiconductor industry and since those decisions are made over a long planning horizon, they are mostly handled in an uncertain environment. Because of the fact that uncertainties in markets regarding parameters such as demand and price are also taken into consideration in the scope of our problem, the studies that handle those kinds of uncertainties is the focal point of this section.

The first study that should be mentioned is conducted by Eppen et al. (1988) who consider a multi-product and multi-facility capacity planning problem with multiple periods under risk. They develop a stochastic mixed integer programming model with recourse, using scenarios for characterising demand-price combinations. The model, which maximises the expected value of the discounted cash flows, determines the production quantities for each product at each plant in each period, the configuration type of each site in each period (i.e. which site will produce which products in each period) as well as the time of the site reconfiguration, and the level of production capacity of each site.

Li and Tirupati (1994) develop a model and a solution algorithm for a multi-product dynamic investment model including technology (dedicated versus flexible) choices, and timing and sizing of capacity additions to minimise the total investment cost to satisfy the product demands over the planning horizon. One of the main points of that study is to consider a technology mix (i.e. deciding on the amount of both flexible and dedicated technologies) while determining a least-cost schedule of capacity additions for each type of technology.

Stafford (1997) proposes a capacity planning model for semiconductor industry, in order to find an optimal capacity expansion plan using a linear two-stage stochastic programming approach where demand is a stochastic parameter. The model minimises the weighted sum of unmet demand for each product and they make an analysis for comparing the unsatisfied demand for the "expected value" model with average unsatisfied demand for the stochastic demand model at varying budget levels.

Chen et al. (2002) develop a stochastic programming approach for capacity planning considering the product mix flexibility that is defined as the ability of a production plant to produce different types of products. In their model, they determine optimal capacity investments in each type of technology, i.e. the mix of dedicated and flexible technologies, over a planning horizon with multiple periods. Additionally, they present an extension of that model with a new product introduction, in which they deal with two questions: (1) Should the company launch the new product? (2) Which strategy for the capacity should be followed, if the new product is launched?

Karabuk and Wu (2003) present a scenario-based multi-stage stochastic programming model with a cost-minimisation objective under demand and capacity uncertainties. The main output of this model is to generate a capacity configuration in order to cope with the scenarios in which demand is extreme and capacity fluctuations is high. In that study, they also consider the outsourcing option when all demand is not satisfied from in-house resources (i.e. in the case of capacity shortfall).

Ahmet et al. (2003) propose a capacity expansion model in an uncertain environment with a single product, multiple resources and multiple periods. The model determines the time and level of capacity expansion or the type of technology in order to satisfy market demand by minimising total cost over the planning horizon.

Zhang et al. (2004) develop a capacity expansion model with multiple products and finite number of periods, which aims for determining optimal plans for tool purchase under stochastic demand, for semiconductor industry. In fact that study is not a product-mix problem; it aims at finding an optimal capacity expansion plan (i.e. deciding on timing and the number of purchasing tools) in order to satisfy the forecasted demand on the condition that the total lost sales and tool purchase costs are minimised.

Huang and Ahmed (2009) also address the multi-resource and multi-item capacity planning problem, which is defined as tool purchase planning problem for semiconductor industry. They propose a multi-stage stochastic programming model for the problem, in which demand, cost and resource utilization coefficients are handled as uncertain parameters.

Ren-qian (2007) describes the capacity planning decision as "*to determine an optimal schedule to replace older machines, equipments or activity centres by newer ones*". Based on that description, they develop a stochastic capacity expansion model, including the uncertainty involved in demand and unit capacity usage by a product. In that model, they decide on "*the number of activity centres purchased/constructed in each period*" as well as "*the sales amount of products in each period*" to satisfy demand for products.

Geng et al. (2009) presents a two-stage stochastic programming model for a capacity planning problem in a semiconductor firm, in which demand and capacity are considered as uncertain parameters. The model, which maximises the expected profit (total revenue minus tool procurement, production, underutilized capacity, inventory holding and stock-out costs), determines the number of new tools of each type that should be procured and production volumes of each product in each period.

Claro and Sousa (2012) present a multi-stage stochastic programming model for a capacity investment problem with multiple resources to determine the amount of capacity increments for each resource and optimal allocation of capacity to products. Their model, which handles demand and cost as uncertain parameters, consists of two objective functions: first is to minimise the total capacity investment cost and second is to minimise the Conditional-Value-at-Risk, which is one of the most commonly used risk measure in the literature.

Chien and Zheng (2012) deal with a capacity planning problem in semiconductor manufacturing wherein demand is characterised as an uncertain parameter. Based on a "*mini-max regret strategy*", they present a model that aims at generating a capacity expansion plan that is robust to all possible demand scenarios while minimising the maximum regret of capacity surplus and shortage.

Lin et al. (2014) focus on capacity planning problem with multiple sites in thin film transistor liquid crystal display (TFT-LCD) industry where demand is considered as a stochastic parameter and represented by a finite number of scenarios. They develop a stochastic dynamic programming model for that problem in order to find an optimal capacity expansion policy that provides a profitable product mix at a given site in a certain period.

In the capacity planning literature, one of the interesting studies that consider a capacity planning problem for short life-cycle products is conducted by Alaniazar (2013). He focuses on the capacity (in that study defined as capital equipment capacity) planning decisions in semiconductor industry which has high capital investments, high rate of obsolescence, irreversible capacity investments, high volatile demand and long procurement lead time of new capacity. The objective of the model developed is to maximise the expected profit by balancing over- and under-expansion of capacity over the product life cycle. The main decision in that model is when and how much capacity to expand in an uncertain environment.

One of the streams in capacity investment planning literature is the studies considering multi-resource investment planning (Eppen, 1988; Ahmed et al., 2003; Huang and Ahmed, 2005), which has received little attention. On the contrary to the main stream studies that consider investments for single resources (capital or labour) mostly considered as irreversible, Harrison and Mieghem (1999) handle a multi-resource investment problem wherein both reversible and irreversible investments are taken into consideration, and demand, price and cost are described as uncertain parameters. The main objective of the model (defined as a product mix model), is defined as follows: "*in each period, having chosen a resource vector and observed a demand vector, the firm chooses its production vector so as to maximise the profit (the expected value of operating profits minus resource adjustment costs*" over the planning horizon). One can refer to that study for detailed discussion on the multi-resource investment problems and their related models.

Based on the studies given above, some remarks about the modelling and solution approaches can be given as follows:

- In the related literature, mixed-integer programming and stochastic programming approaches using scenarios to model uncertainties are the most common models used for multi-period capacity planning problems.
- Two-stage stochastic programming model is extensively used for capacity planning problems, in which the first-stage decision aims at determining a capacity expansion schedule over a planning horizon, while second-stage decisions consists of the allocation of capacity that is determined in the first-stage. Ahmed et al. (2003), Huang and Ahmed (2005), Eppen et al. (1988), Thomas and Bollapragada (2010), Ren-qian (2007) and Geng et al. (2009) are some of those studies using two-stage stochastic programming approach.
- On the other hand, multi-stage stochastic models in which decisions made at a certain period depend on events and decisions up to that period and the decisions are revised when more information about the uncertainties is revealed at each period are also used in the related literature. Lin et al. (2014), Alaniazar (2013), Claro and Sousa (2012) Huang and Ahmed (2005), Karabuk and Wu (2003), Ahmed et al. (2003) and Chen et al. (2002) are some of the studies which use the multi-stage stochastic programming approach.
- Most of the given studies focus on determining a good capacity planning schedule including time phasing and size of capacity additions for satisfying the future demand (deterministic or stochastic) of products. However, they could not be considered as typical product mix problems, since they do not consider a product selection decision.
- It should be noted that the study of Ren-qian (2007) is different from the existing capacity planning studies in terms of dealing with unit capacity usage of a product as an uncertain parameter. The models dealing with capacity uncertainties can also be seen in the studies of Karabuk and Wu (2003) and Geng et al. (2009).

- All of the studies given above consider single resource capacity planning except Eppen (1989), Ahmed (2003), Huang and Ahmed (2005) and Harrison and Mieghem (1999) who consider multi-resource planning in their problems.

Above all, it should be noted that since the product-mix of a firm might be changed over the planning horizon because of unstable market demand and requirement for new product introductions to markets, capacity expansion as well as capacity allocation decisions must be adjusted/adopted dynamically in order to optimise firm objectives (mostly profit maximisation). Therefore, capacity expansion decisions are taken into consideration when formulating the problem handled in this study.

2.3. Product Interdependencies

One of the main characteristics of the product-mix problems in real life applications is to have interdependent products competing for financial (budget) and physical (e.g. production capacity) resources, and affecting the demand and production cost of each other. Therefore, in this section, the main properties of product interdependencies (interactions) are presented based on the studies in literature and the how this concept is handled in our study are presented.

In the related literature, one of the pioneering studies in which product interactions are taken into consideration in a *product-line decision model* is performed by Urban (1969), who states that the products in a line are not usually independent since the marketing mix established for one product may affect the sales of another product. He takes two basic kinds of interdependencies into account, i.e. complementarity and substitutability which are measured using cross-price elasticities.

Monroe et al. (1976) handle a product-mix problem that considers revenue interactions among products, which are assumed to be occurred due to coexistence of two products (all existing and new products) in the market. In this study, the interactions are subjectively estimated by the management. For instance, considering two specific products, the management may include the following interaction effect: the presence of product A on the market would decrease (or increase) the expected revenues of product B by 10 %. That interaction is also defined as *first order interaction, i.e. interaction between each pair of products* and they state that in product mix problems cost interactions and higher order interactions should be considered. Furthermore, it should be noted that the revenue interactions presented by Monroe et al. (1976) is similar to the concepts of *global substitutability*[2] and *global complementary*[3] defined by Devinney and Stewart (1988).

Two types of interdependencies among products in a *multi-product investment (portfolio) problem* can be considered: demand interdependency and supply interdependency. There are three types of demand interdependencies: demand substitutes, demand compliments and demand neuters.

Demand substitutes: The demands for two products are negatively related, i.e. if the demand for one of the products increases the demand for the other will decrease. This is also called as cannibalism (Kerin et al., 1988).

Demand complements: The demands for two products are positively related, i.e. if the demand for one of the products increases, the demand for the other will increase as well. In other words, coexistence of two products in a market affects the total demand of those products positively.

Demand neuters: The demands for two products are independent, i.e. changes in the demand for one product will not affect the demand for the other.

---

[2] There is a negative correlation between the expected return (profit) of one product and the investment in another, e.g. if the expected return of product A decreases with investment in product B, products A and B are defined as global substitutes.

[3] There is a positive correlation between the expected return (profit) of one product and the investment in another, e.g. if the expected return of product A increases with investment in product B, products A and B are defined as complements.

There are also three types of supply interdependency: supply substitutes, supply compliments and supply neuters (Devinney and Stewart, 1988):

Supply substitutes: The production cost of producing two (or more) products jointly is greater than the sum of the costs of producing those products individually.
Supply complements: This implies that the production cost of producing two (or more) products jointly is smaller than the sum of the costs of producing those products individually.

Supply neuters: production costs of different products are independent.

Regarding product interdependencies, Kerin et al. (1978) present the concept of cannibalism, i.e. product substitution, in which a new product obtains a proportion of its sales from an old product's sales. They state that new products acquire their sales revenue from three sources: (1) new customers not previously buyer of the product, (2) consumers of other companies and (3) consumers of an existing product who switch to the new product. The last of those sources refers to cannibalism and it becomes a problem when it provides no financial or competition benefit to the firm. Besides, as Lomax et al. (1997) highlighted, cannibalization effect of new products or brands on existing products or brands should be considered in decisions for product-line extensions. Roberts and McEvily (2005) describe this effect as "*new toy effect*" and state that the full effects of new product introductions, i.e. their effects on the products sold both in the same market and in other markets, should be taken into consideration while deciding on the product-mix of a firm. Shah and Avittathur (2007) also consider demand cannibalization effect in their multi-item (standard product and its customized extension) problem, which is described as "*customer-led demand substitution*" wherein the substitute product is purchased by the consumers when the primary product is out-of-stock.

Laruccia et al. (2012) presents different models of cannibalism occurred between products, which are adapted from Traylor (1986), as shown in Figure 7.

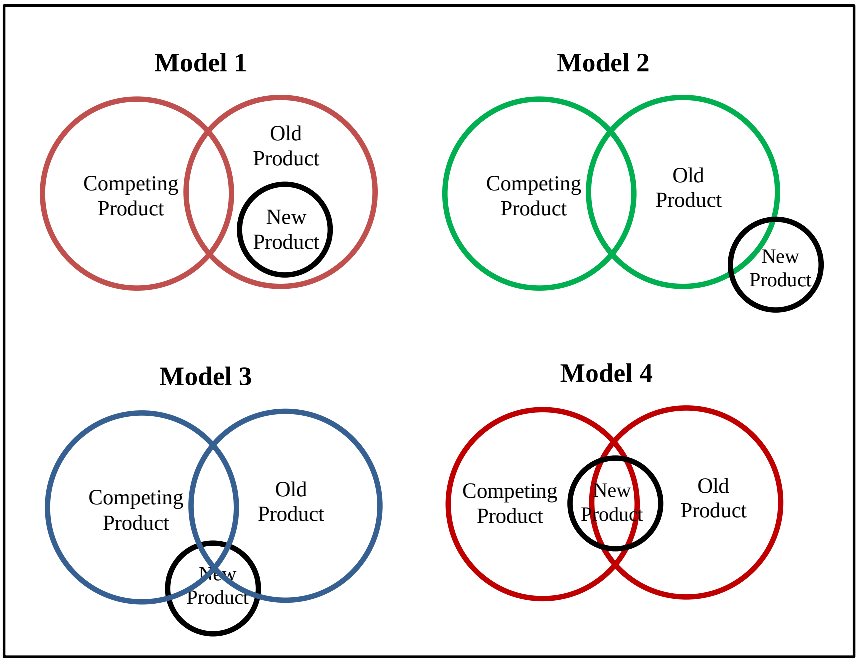


Figure 7. Models for cannibalism (Laruccia et al., 2012)

In Model 1, new product cannibalises the sales of old product currently sold in market and has no impact on firm's total sales provided by the old product. The firm may plan to introduce the new product on the condition that its contribution margin is greater than the old one. Model 2 shows another case in which new product cannibalises a proportion of the sales of old product currently sold in market, but provides firm's total sales to increase (market expansion). In Model 3, the new product cannibalises not only the sales of firm's old product but also the sales of competitor's product and increases the market size. Besides, the new product cannibalises the sales of firm's old product as well as competitor's product, however the market size does not change in Model 4. All these models may come to exist depending on market conditions, sectoral structure, and product features etc.

In addition to the studies given above, Yilmaz et al. (2013) also handle the cannibalisation effect of the new generation products on the sales of parent (old generations) products in their capacity planning model for short-life cycle products and their renewals are considered. That type of cannibalisation may occur between more than two products, i.e. higher-order interaction, and may be two-way in which new generation products and parent products may cannibalize each other.

Eppen et al. (1988) also consider the interaction effect when demand of a product exceeds capacity, i.e. in the case of having unsatisfied demand. In this situation, it is assumed that a proportion of unsatisfied demand for a product might be transmitted to another product. Morgan and Daniels (2001) study another case in parallel with this in their product mix and technology adoption model. They include *two-way substitution* between products, i.e. if a product is not chosed for the mix, a certain proportion of the demand for that product is met by similar products.

Blau et al. (2004) deal with different types of product dependencies in their portfolio management problem such as resource and production cost dependency that can be considered in the scope of substitutability, and technical success dependency that occurs between two candidate products under development when success or failure probability of one product is affected the success or failure of another product. Besides, Yayla-Küllü (2011) deal with a special case in which product with a high quality can substitute another product with a lower quality.

Based on the literature given above, some remarks about the product interdependencies that may be required to consider in a product-mix problem as follows:

- Demand-based substitution-type interactions are studied more than supply-based and complementary interactions.
- Demand-based substitution-type interaction, which is also described as cannibalism, is generally handled in product-mix problems, particularly in the case of new product introductions.
- First order interaction (i.e. interactions between two products) dominates the literature when compared to higher-order interactions occurring among more than two products.

2.4. Parameter Uncertainties in the Problem Environment

The conditions due to market dynamics, regulations, resource availability, fluctuations in stock markets, competitors' future strategies, new technologies etc. inherently affect future market demand, selling prices, costs, interest rates, machine capacities etc. and thereby force those parameters to be considered as uncertain in many problems such as capacity planning, portfolio selection, scheduling, location, routing, product design and resource allocation. Those uncertainties arise when the problem in question requires making decisions without having full or reliable information about the effects of problem parameters when the decision must be made or due to subjectivity of decision makers that naturally appear in a decision-making process (Rockfellar, 2001; Alonso-Ayuso et al., 2003; Hausike and Ishii, 2009a). Therefore, the uncertainties within a problem should be taken into consideration by the way of modelling their future effects on the decision-making process properly. In the case of ignoring uncertainties, a sub-optimal or mistaken solution might be obtained, which results in financial loss for the company.

In the related literature that can be entitled as "optimisation under uncertainty", a large number of approaches are developed in order to model the uncertainty in some parameters such as demand, selling price, cost etc. and to capture the dynamic characteristics of real-life accurately. There are typically four approaches developed for optimisation under uncertainty in the literature: (1) stochastic programming including two- and multi-stage models with recourse and probabilistic (chance-constrained) programming, (2) fuzzy programming, (3) simulation optimisation and (4) robust optimisation. In this section, first a brief description of those approaches and then which kind of uncertainties are considered in the scope of the product-mix problem handled in this study will be presented.

2.4.1. Stochastic Programming

Stochastic programming is one of the optimisation tools commonly used in decision-making problems under uncertainty, whose aim is to identify a feasible solution that minimises or maximises the expected value of an objective function over all scenarios, i.e. possible occurences of uncertain parameters (Solak, 2007). Those parameters are typically modelled using random variables and it is assumed that those have probability distributions estimated from historical data or subjective probabilities defined by decision makers for different scenarios that can be described as the possible future realizations of those parameters. For instance, assume that there are three different scenarios for a product's future demand, e.g. low demand (less than 1000 units), medium demand (units between 1000 and 2000) and high demand (more than 2000 units). The probability of each scenario might also be determined based on a subjective evaluation or statistical data, e.g. the demand may be low, medium or high with a probability of 2/9, 4/9 or 3/9, respectively. In this example, *scenarios*, which are particular representations of how the future might seem to be, are used in order to represent the uncertainty of the problem in a typical stochastic programming model. Besides, since the power of a stochastic programming model relies upon the ability of sufficiently modelling uncertainty through scenarios, firstly it is needed to generate descriptive and realistic scenarios for the problem in hand for the success of stochastic programming approach (Mitra, 2006). Therefore, a large number of scenario-generation methods are developed in the literature. The main objective of those methods is to generate all possible realisations of uncertain parameters in a problem, which is typically called as scenario or event tree and the probability of those realizations (Kaut and Wallace, 2007; Zhang and Wang, 2009; Tønnesen and Øveraas, 2012). A typical scenario tree can be seen in Figure 8.

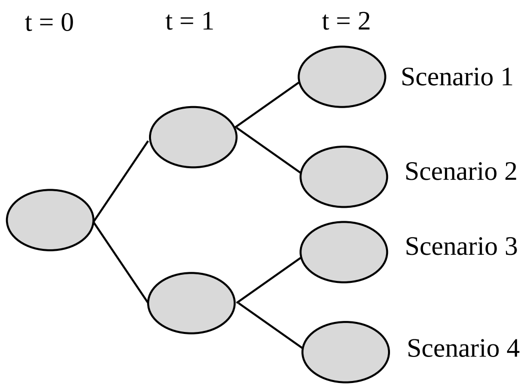


Figure 8. A scenario tree considered for a two-period problem

As seen from Figure 8, the initial state, two possible realisations at the time of t=1 and four possible realisations are represented by the root node at t=0, two nodes at t=1 and four nodes at t=2, respectively. Thus, a path from the initial state to any state at t=2 is called as a scenario each of which has a probability of occurrence. Though it is easy to deal with such a problem having less number of scenarios, the computational effort of stochastic programming for handling a large number scenario becomes extremely high, especially in multi-period problems wherein number of scenarios exponentially scales up and/or there are many uncertain parameters in the problem context; thereby, the usage of stochastic programming in real life is restricted. Yet, it's become a popular approach both in literature and practice (Chien and Zheng, 2012) in the last decades.

The scenario-generation methods commonly used in the related literature can be classified under four groups (Mitra, 2006):

1. *Statistical methods* that include statistical moment or property matching, principal component analysis and regression.

2. *Sampling methods* such as Monte Carlo/random sampling, importance sampling, bootstrap sampling, internal sampling, conditional sampling, stratified sampling, Monte Carlo based on Markov chain sampling.

3. *Simulation-based methods* such as stochastic process simulation, error correction model and vector auto-regressive model.
4. *Other scenario generation methods* such as artificial neural networks, clustering, scenario reduction and hybrid methods.

One can refer to Mitra (2006), and to Kaut and Wallace (2007), Linderoth et al. (2002) and Infanger (1992) for the details of those methods.

Stochastic programming approaches are mainly grouped under two categories: *(1) recourse-based stochastic programming* and *(2) probabilistic (chance-constraint) programming*.

Recourse-based Stochastic Programming

It can be said that the idea of incorporating uncertainty into a mathematical model starts with Dantzig (1955)'s work that was the first attempt to propose a recourse model that enables the solution to be adapted according to the consequence of a random event (Solak, 2007; Birge and Louveaux, 2011). Since then, recourse-based stochastic programming models' popularity has increased thanks to high-power solvers and sophisticated solution algorithms in the last decade (Karabuk and Wu, 2003). In stochastic programming approach, it is assumed that all possible scenarios and their attached probabilities for the subsequent stages are known (or at least estimated), but which of the scenarios will occur in reality is not known, and at this point recourse-based approach allows for making decisions "now" over all scenarios, which optimise the expected results (e.g. maximising the expected profit in a product-mix problem) and ensure adaptability of each scenario (Tønnesen and Øveraas, 2012). In another words, recourse-based stochastic models enables recourse actions or some decisions to be made after the uncertainty is revealed (Birge and Louveaux, 2011).

Due to the different characteristics of decision-making processes involved in real-life problems, the recourse-based stochastic problems are formulated as two-stage or multi-stage models with recourse (Chou et al., 2007).

Two-stage Recourse Models:

In a two-stage recourse model, a *stage* refers to a time point where new information regarding random variables is revealed and the decisions associated to stages are grouped under two sets:

1. Some of the decisions are taken before the actual realization of uncertain parameters, which are called as *first-stage decisions*.

2. Some of the decision are taken after the realization of uncertain (random) parameters is known, which are called as *second-stage decisions*. At this stage, further actions are employed by choosing, the values of second-stage or recourse variables interpreted as *corrective measures against any infeasibilities* arising due to a particular realization of uncertainty, i.e. in the second-stage, the first-stage actions are fine-tuned.

For instance, in a typical facility location problem, the facility locations as well as its capacity are determined in the first stage, and then the decision on how best to operate it is made in the second-stage after the realization of random events (e.g. different scenario generated for unit processing time). In another example of the capacity planning problem in semiconductor industry, the capacity expansion schedule (including the number of tools that will be purchased as the amount of capacity expanded and timing of procurement) over the planning horizon including multiple periods is decided in the first-stage and then the decisions associated to production planning, i.e. selected products and their quantities, as second-stage decisions, are made at the beginning of each period according to realised demand and available tool capacity determined in the first stage (Stafford, 1997; Ahmed, 2002). The main objective of two-stage stochastic programming with recourse is to make *the first-stage decisions considering the evolution of random parameters throughout the planning horizon by minimising the sum of first-stage costs and the expected value of second-stage costs* (Sahinidas, 2004). Besides, in two-stage recourse models decisions at all periods (e.g. *t =1* and *t =2* in Figure 8) are made at the beginning of time horizon (*t=0*, now) and it is assumed that the scenarios regarding each period are unrelated.

A typical formulation of a two-stage linear stochastic programme with recourse can be given as follows (Birge and Louveaux, 2011, p.10):

$$\min c^T x + E_\xi \; Q\;(x, \xi)$$
$$\text{subject to}$$
$$Ax = b \text{ and } x \;\geq 0$$

where $Q\;(x,\xi) \;=\; \min\{q^T y \mid Wy = h - Tx, y \geq 0\}$ which is the value of the second-stage for a certain realisation of random vector ξ which is framed by q^T, h^T, and *T,* and E_ξ denote expectation with respect to ξ. Here, $E_\xi\; Q\;(x,\xi)$ is called as recourse or value function and recourse matrix W can be considered as fixed (*fixed recourse)* or not (*relatively complete recourse, complete or simple recourse*). On can refer to Birge and Louveaux (2011) for a detailed analysis of this formulation as well the basic concepts.

Two-stage stochastic problems are tractable on the condition that there is a small number of scenarios or randomness is identified by discrete distributions. In those cases, the problem can be represented by equivalent deterministic linear program that can be solved using linear programming solution procedures. If the recourse function is convex, that type of problem, even large-scale ones (but finite number scenarios), can be efficiently solved using decomposition-based strategies such as Langrangian and Benders schemes (Ren-qian, 2007; Bertsimas et al., 2011; Birge and Louveaux, 2011).

For continuous distributions of the random parameters or the cases in which the number of discrete scenarios is too large, sampling-based, e.g. random sampling and importance sampling, approaches, decomposition and approximation methods, and gradient-based algorithms are developed to obtain a tractable and efficient solution (Sahinidas, 2004).

Furthermore, in some of the stochastic models it is assumed that scenarios and their associated probabilities are independent of the decisions that are taken. In the related literature, there are also some stochastic models that do not consider that assumption, i.e. decisions might influence the probability distributions (Linderoth, 2009).

Multi-stage Recourse Models:

A multi-stage recourse problem involves a series of decisions made over time based on the new information revealed at certain time points. This series typically has a sequential pattern such as “decide-observe-decide-observe-decide-…” (Tønnesen and Øveraas, 2012). The uncertainty in a multi-stage stochastic program is represented by a multi-tiered scenario tree (see Figure 8 as an example) and the optimisation strategy is to make decisions hedging against this scenario tree that shows the evolution of the future outcomes (Ahmed et al., 2003).

In two-stage recourse models decisions at all periods (e.g. t =1 and t =2 in Figure 7) are made at the beginning of planning horizon (t = 0) before the uncertainty is realised and subsequently a limited number of recourse actions are taken. However, in a multi-stage model the decisions made at a certain period depend on events and decisions up to that period and the decisions are revised when more information about the uncertainties is revealed at each period. The multi-stage models handle the problem in dynamic planning process better and provide more flexibility than the two-stage models (Huang, 2005; Huang and Ahmet, 2005). However, it is more difficult to solve these models than the two-stage models because of their complexity arising from a high number of stages and random parameters (Huang, 2005; Solak, 2007). Therefore, many solution strategies are proposed for multi-stage problems in the literature (see e.g. Ahmed et al., 2003; Karabuk and Wu, 2003; Sahinidas, 2004; Huang, 2005; Huang and Ahmed, 2005; Solak, 2007; Ahmed, 2002; Claro and Sousa, 2012; Alaniazar, 2013; Lin et al., 2014; Fattahi et al., 2017; Bertazzi and Maggioni, 2018).

Probabilistic (Chance-Constrained) Programming

While two-stage or multi-stage stochastic model with recourse given in the previous section requires the decision-maker to assign a certain penalty for the infeasibilities in the second stage, chance-constrained programming or programming with probabilistic constraints focuses on minimising the expected recourse costs. This type of program is suggested to use when the decision maker's main objective is to have an optimised solution with sufficiently higher probability of achieving a goal (Linderoth, 2009).

A typical deterministic linear programming model (including no uncertainties) given as (1) can be converted into a probabilistic programming model, (2), as follows:

$$(1) \quad : \{\max c^t x \quad \text{s.t.} \quad Ax \geq b, \quad x \geq 0\}$$

$$(2) \quad : \{\max c^t x \quad \text{s.t.} \quad P(Ax \geq b) \geq p, \quad x \geq 0\}$$

In model (2), it is assumed that there is uncertainty associated to the constraint matrix, A, and the vector b, and the corresponding constraints must be ensured with a probability p. It is suggested to refer to Prekopa (1995), Birge and Louveaux (2011) and Wang et al. (2017) for a detailed description of chance-constrained programming, its modeling and solution methodologies and applications in different management areas.

2.4.2. Fuzzy Programming

As explained in the previous section, *stochastic programming* is one of the approaches for dealing with uncertainties in a decision-making process through modelling uncertain parameters as random variables. Another approach that is commonly used for optimisation under uncertainty is *fuzzy programming* in which fuzzy numbers and fuzzy sets are preferred to model uncertainty.

The basic characteristic of a fuzzy programming model can be explained by the following example (Sahinidas, 2004):

Consider a constraint $ay \leq A$ with a decision vector y and random right-hand-side A which may take values within the range of $[m,\ m+\Delta]$, where $\Delta \geq 0$. Then the membership function, $f(y)$, which characterizes the fuzziness/uncertainty associated to this constraint (because of not knowing the exact value of A) can be defined as follows:

$$f(y) = \begin{cases} 1, & if\ ay \leq m \\ 1 - \frac{ay-m}{\Delta}, & if\ m\ < ay\ \leq m + \Delta, \\ 0, & if\ m + \Delta < ay. \end{cases}$$

This kind of membership function may be in different forms according to decision maker's preferences and problem framework. After modelling the uncertainties using membership functions and setting up the fuzzy optimisation model that may be in the form of linear, nonlinear, dynamic and multi-objective programming etc., this model is converted to a crisp optimisation programme (Tang et al., 2004).

In the related literature, there are two main types of fuzzy programming approaches: flexible programming involving right-hand-side uncertainties and possibilistic programming handling the uncertainties both in the coefficients of objective function and of constraints (Sahinidas, 2004). It is suggested to refer to Luhandjula (2015), Lodwick and Kacprzyk (2010) and Tang et al. (2004) for a detailed description of fuzzy programming, its modelling approaches and solution methodologies.

While two approaches, stochastic and fuzzy programming, can be used in order to handle uncertainties in an optimisation problem, one can ask the question of which approach should be used in which conditions. Bastin (2004) states that if historical data is available and well-defined probabilistic distributions associated with uncertain parameters are obtained, stochastic programming can be used since this approach is less sensitive to the modeller and provides a strong analytical point of view to the problem solution. On the other hand, if there is no historical data, modelling the uncertainties with probability distributions may not be possible. In this case, fuzzy programming is preferable on the condition that the modeller can identify fuzzy numbers and sets based on his/her knowledge and understanding about the problem (*Ibid.*).

2.4.3. Simulation Optimisation

Simulation optimisation is an approach including many techniques used for optimizing stochastic simulation model whose aim is to find optimal settings, which optimises the output, of the inputs, i.e. a target objective (Amaran et al., 2016). A generalised simulation optimisation model is given in Figure 9.

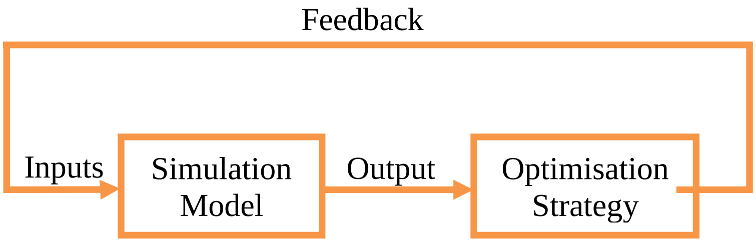


Figure 9. A simulation optimisation model (Carson and Maria, 1997)

As shown in Figure 9, the output of a simulation model is evaluated by an optimisation strategy and a feedback based on this evaluation is provided as a further input for the process of searching the optimal solution to the simulation model (Carson and Maria, 1997).

A general formulation of simulation optimisation problem is presented by Amaran et al. (2016) as follows:

$$\begin{aligned} \min \quad & E_{\xi}[f(x,\xi)] \\ s.t. \quad & E_{\xi}[g(x,\xi)] \leq 0 \\ & h(x) \leq 0 \\ & x_l \leq x \leq x_u \\ & x \in \mathbb{R}^n \end{aligned}$$

In this model, the output represented by function f is evaluated for an instance of inputs x and a certain realisation of the random variables ξ through simulation runs within the feasible region provided by the constraints described by function g

including the random variables, function h without the random variables (if there exists) and bounding constraints defined for the decision variables. Azadivar (1999) presents a classification for the approaches used in simulation optimisation, where each class can be handled as a special case of the general formulation above. Gradient based search methods (such as finite difference estimation, infinitesimal perturbation analysis, frequency domain analysis, Likelihood Ratio Estimators etc.), sample-path optimisation, response surface, heuristic search and statistical methods are some of the examples of simulation optimisation methods developed in the related literature (Chang, 2016; Olafsson and Kim, 2002; Azadivar, 1999; Carson and Maria, 1997).

It is suggested to refer to Amaran et al. (2016), Azadivar (1999), and Carson and Maria (1997) for the details of simulation optimisation concept, its modelling approaches, techniques and solution methodologies.

2.4.4. Robust Optimisation

Robust optimisation is another approach developed for solving problems under uncertainty, in which a decision-maker seeks for a robust solution that performs well considering any future realization of uncertainty (scenarios) defined in a set. In a general framework, the main objective of robust optimisation is to find a solution feasible for all scenarios in the set while minimising the deviation of the solution obtained for overall problem from the optimal solution obtained for each scenario (Better et al., 2008).

Robust optimisation is suggested to use when uncertainty regarding parameters can not be modelled as stochastic or distributional information is not available; and also the computationally tractable feature of this approach make this approach preferable compared to stochastic optimisation. However, in contrast to the robust optimisation, stochastic optimisation can provide an opportunity for decision makers to adjust the decision and take recourse actions in the next periods when the uncertain environment is changed.

It is suggested to refer to Gorissen et al. (2015), Gabrel et al. (2014), Bertsimas (2011), Linderoth (2009) and Better et al. (2008), for the details of robust optimisation, its modelling approaches, techniques and solution methodologies.

Considering the approaches given in this section and the studies dealing with capacity planning and product-mix determination summarised in this chapter it can be seen that recourse-based stochastic programming is one of the commonly used approaches for solving the multi-period problems under uncertainty. It also has some advantages such as the ability of handling a large number of scenarios, existing bounding techniques that provide to solve the stochastic problem with a higher efficiency, the opportunity for decision makers to adjust the decision and take recourse actions in the next periods when the uncertain environment is changed.

CHAPTER 3

PROBLEM DEFINITION AND A PROPOSED TWO-STAGE STOCHASTIC PROGRAMMING MODEL

In this chapter, firstly the problem definition and a summary for modelling environment with relevant assumptions are presented in Section 3.1; and then a deterministic mathematical model, a two-stage stochastic program which is developed based on the deterministic mathematical model and a solution approach for the two-stage stochastic program are given in Section 3.2, Section 3.3 and Section 3.4, respectively.

3.1. Problem Definition and Main Contributions to the Related Literature

The problem handled in this thesis work is described as a *medium-term tactical planning problem* (MTPP) considering *multiple periods* (set of equal periods) and the *mix of product-lines*, which are formed by grouping different models, under a class sold in *multi-markets*, and *uncertainty regarding some problem parameters*. The details of this problem as well the environment considered, which are described in accordance with the sub-titles given in Chapter 2, can be presented as follows:

<u>Long/medium life cycle products of a firm operating in consumer durables sector:</u>

According to problem context handled in this study, **long/medium life cycle products,** which remain in use significantly longer and have relatively stable demand are dealth with rather than short life cycle products. Since it is not feasible to develop a long-term plan for short life cycle products because of the fact that technological changes are very rapid and long-term demand is so volatile, short life cycle products managed based on short-term marketing strategies are out of concern in this study. Besides, a firm operating in **consumer durables** sector and having a make-to-stock (MTS) production environment in which demand is predictable over the course of the product life cycle from the market is considered.

Product- and Planning-level within the scope of management:

Considering the problem domain in this study, the problem is described as a **medium-term tactical planning problem (MTPP***)* considering the **mix of product-lines** defined under the concept of product hierarchy and medium-scale capacity expansion decisions such as the amount and timing of capacity increments (see Section 2.2.1).

Product-mix and new product launching decisions:

In the problem handled in this study, two main product-mix decisions, (1) for all of the existing and new products that will be in the mix and (2) merely for new products, are taken into consideration.

The decisions regarding first set can be described as follows:

1. Which products to be in the mix in each period of planning horizon, i.e. product list.
2. How many of them to be sold in which (target) market in each period of planning horizon, i.e. product quantities.

Considering Figure 6 (Section 2.2.1), for an example, there are two lines under freezer class and the firm may want to determine the product volumes (as sales targets) within the context of questions given above, based on its current product list (there is no new line that is planned to be launched in any markets or no new markets that the firm plans to introduce with its current line list). However, the firm plans to introduce new product lines in the next periods (e.g. a new-generation refrigerator which can generate the energy required for cooling using the heat inside foods rather than using electricity) according to its product roadmaps. The firm considers that this new refrigerator line can substitute the currently sold refrigerators sold in markets. Therefore, the firm wants to analyse if it is profitable to introduce this new line into markets. Furthermore, they need to allocate a budget for capacity investment for the new-generation refrigerator that is planned to introduce to the market two years later at the earliest but not after the third year. For that purpose, a feasible range for possible launching time of this new line is determined. The firm might also plan to introduce an existing line to a new market. At this point, another decision that will be made raises within the problem context:

3. When the planned (new) products to be launched, i.e. release plans/time-to-market decisions. In other words, if a new planned product is selected to the mix, when to launch that product within the range of their minimum and maximum launching time that are determined based on different strategies of the firm reflected in product roadmaps, and considering competition in markets.

Thus, a product list that will be sold in each market, sales targets for product quantities and launching times for new products and for existing products sold in new markets are the main decisions we are interested in. Some of the data required for the process of making those decisions, which are provided by product roadmaps and possible launching times, can be summarised as follows:

- Currently sold products, their maximum lifetimes and target markets in which those are planned to be sold,
- New planned products their maximum lifetimes and target markets in which those are planned to be sold,
- Release plans for the new products; i.e. expected maximum and minimum launching times.

It should be noted that a **minimum and maximum launching time**, which are identified based on the different strategies of a firm reflected in product roadmaps and considering competition in markets, are defined for the new products seen as candidate product-lines for a specific product-class.

Furthermore, since product quantities are determined as sales targets at strategic level in this study, and it is assumed that outsourcing decisions are made when product roadmaps are developed (i.e. only the products to be produced in-house are considered), inventory-holding and make/buy decisions are out of concern in the scope of our problem context.

Capacity expansion decisions and capacity-related assumptions:

Since the product-mix of a firm might be changed over the planning horizon because of unstable market demand and requirement for new product introductions to markets, capacity expansion as well as capacity allocation decisions must be adjusted/adopted dynamically in order to optimise firm objectives (mostly profit maximisation), as in most of the studies given in Section 2.2. Therefore, capacity expansion decisions are taken into consideration in accordance with our problem context.

It is assumed that main capacity decisions such as buildings, locations, new facility opening, closing existing plant decisions are pre-made; i.e. the design capacity of plants for each period is an input for the model (see Section 2.2.1), and medium-scale capacity adjustment decisions such as tool/machine purchasing and decommission (Chou et al., 2007), renting or buying a warehouse, adding a new production line to the factory etc. are considered in the scope of the problem handled in this thesis work.

The firm considered in the problem domain faces "soft capacity constraints", i.e. capacity constraints specified in a model can be relaxed at a penalty cost if capacity expansion generates additional profitable production. However, this relaxation will be limited by the design capacity of plants. In other words, the firm can install an additional capacity (or resource) on an as-needed basis at a cost of β_t per unit of capacity (i.e. unit penalty cost for augmenting initial/committed capacity). This cost including capital investments in the existing facilities such as supplying new equipment/production line, line modification or modernization (in this study those investments are considered as irreversible, i.e. capacity reduction decisions such as selling or renting the installed capacity in terms of machines, tools, etc. to other companies are out of concern) is determined considering different capacity addition options and their related costs. At this point, it is assumed that based on the data of previous years the probabilities related to the realization of different capital investment alternative and its average cost, which are used for calculating the "expected value" of β_t, can be estimated.

Furthermore, it is assumed that all products are in-house produced and interchangeable with respect to plant capacity (i.e. mix of lines does not cause significant changes in capacity requirements) and therefore all product-lines use the same plant capacity that is determined for the regarding class in the previous decision level which incorporates decisions related to the family as well as classes under this family.

Product interdependencies:

Considering the problem context and Secton 2.3, the following interactions between product-lines are taken into account in this study:

- Demand for two different product-lines are negatively interdependent; i.e. if the demand for one of the products decreases due to the price increment, the demand for the other will increase, and vice versa. This kind of interaction is modelled using cross-price elasticities and the modelling details are given in Section 3.2. For instance, consider the firm whose product hierarchy given in Figure 6 in Section 2.1. Although two different lines, e.g. single-door and double-door refrigerators, are designed based on configuration criteria (door type and cabinet volume), they have similarities in terms of their technical characteristics and may be marketed for the same consumer segment. On the condition that the price of single-door refrigerators increases (assume that the price of double-door remains constant), some of the potential consumers for single-door refrigerators may choose double-door refrigerators. Because, those consumers will make a comparison between two types of refrigerators based on price vs. performance (in terms of cabinet volume and functionality) and decide to buy the double-door instead of buying a single door with a higher price.

- Because of the definition of the selected product hierarchy, i.e. product line, the complementary-type interaction seems meaningless, therefore, only demand-based substitution-type interaction is considered.
- Two-way, i.e. interaction occurred due to coexistence of two products in a market, and first order interactions, i.e. interactions between each pair of products, are considered.
- Moreover, only interaction effects between two products sold in the same market are taken into account, but it is assumed that there is no interaction between two products sold in different markets.
- Because of interactions, the demand (sales) of an existing product will be affected by newly introduced products (cannibalisation effect) in addition to other existing products. However, it is assumed that the demand of a new product will be affected only by its own prices, i.e. it is independent of the existing products and other new products. In essence, this assumption seems meaningful, because the firm producing long/medium life cycle products will not prefer introducing more than one new product lines that will cannibalise each other within the same planning horizon and introducing a new product that will be cannibalised by an existing product.
- Regarding with cannibalisation effect of new products on the existing products, it is assumed that a new product cannibalises a percentage (or proportion) of old product's sales, but at the same time it expands the market share of the company. This assumption is Model 2 proposed by Laruccia et al. (2012) in Figure 7 (see Section 2.3). This percentage is usually defined as cannibalisation rate in the related literature. It can be expressed as, for

instance consider the cannibalisation rate of no-frost refrigerators (new product line) on double-door refrigerators (one of the existing product lines) is 20 %, the total potential demand of the double-door will be reduced by 20 % providing that no-frost refrigerator is introduced to the same market. Thus, the cannibalisation effect of new products on the existing ones is is represented by rates and demand-price dependencies (see Section 3.2).

Parameter uncertainties in the problem environment:

Considering the problem environment in this study, although some parameters are intrinsically handled as deterministic or the variability of those parameters do not have any significant impact on the solution, price, cost and demand should be dealt as uncertain parameters in such kind of problems including strategic decisions (Alonso-Ayuso et al., 2003). In addition to those parameters, the cannibalisation effect of new products on currently sold products in certain markets is also examined as an uncertain parameter within the context of this study.

Since those decisions highly depend on the future realizations of some main parameters, which are not known precisely with the present information, the uncertainty related to those parameters needs to be coped with. Actually, there are two sources of uncertainty in the problem. First, the conditions due to the dynamically changing market conditions may cause uncertainty regarding the value of some parameters, e.g. selling- price of a product. Second source of uncertainty stems from product aggregation. In this case, for instance, selling-price of a product line is obtained as the average of selling-price of each item under this product-line since a product-line is represented as a group of items. Because of this averaging process, the deviation from the average is also considered as a kind of uncertainty. In this study, selling-price, demand and variable production cost of products, and the cannibalisation rates are handled as *uncertain parameters*. Besides, it is assumed that probability distributions for those parameters are known, and the regarding uncertainty are structured in a set of scenarios.

The uncertainty regarding some parameters mentioned above is handled in a two-stage stochastic programming model whose details are presented in Section 3.3.

Above all, to the best of our knowledge, there is no prior work that addresses a multi-period product-mix problem including the timing decisions of new (planned) products of a firm and capacity expansion decisions, and taking product interdependencies and uncertainties associated to problem parameters into account. Additionally, this study is the first attempt to make a link between the product roadmaps (PRM) of a firm and described problem, which aims at balancing the PRM.

As a solution approach to the problem described above, firstly a mathematical model in which all parameters are deterministic is developed (Section 3.2), and then this model is extended to a two-stage stochastic programming model with recourse for incorporating uncertainties using a scenario-based approach (Section 3.3). Secondly, this model is solved for different cases using a bounding technique based on random sampling (Chapter 4).

3.2. Deterministic Mathematical Model

The problem whose characteristics are identified in Section 3.1 is represented by a model, called as "Product-Line Planning Model", whose input-output scheme is displayed in Figure 10.

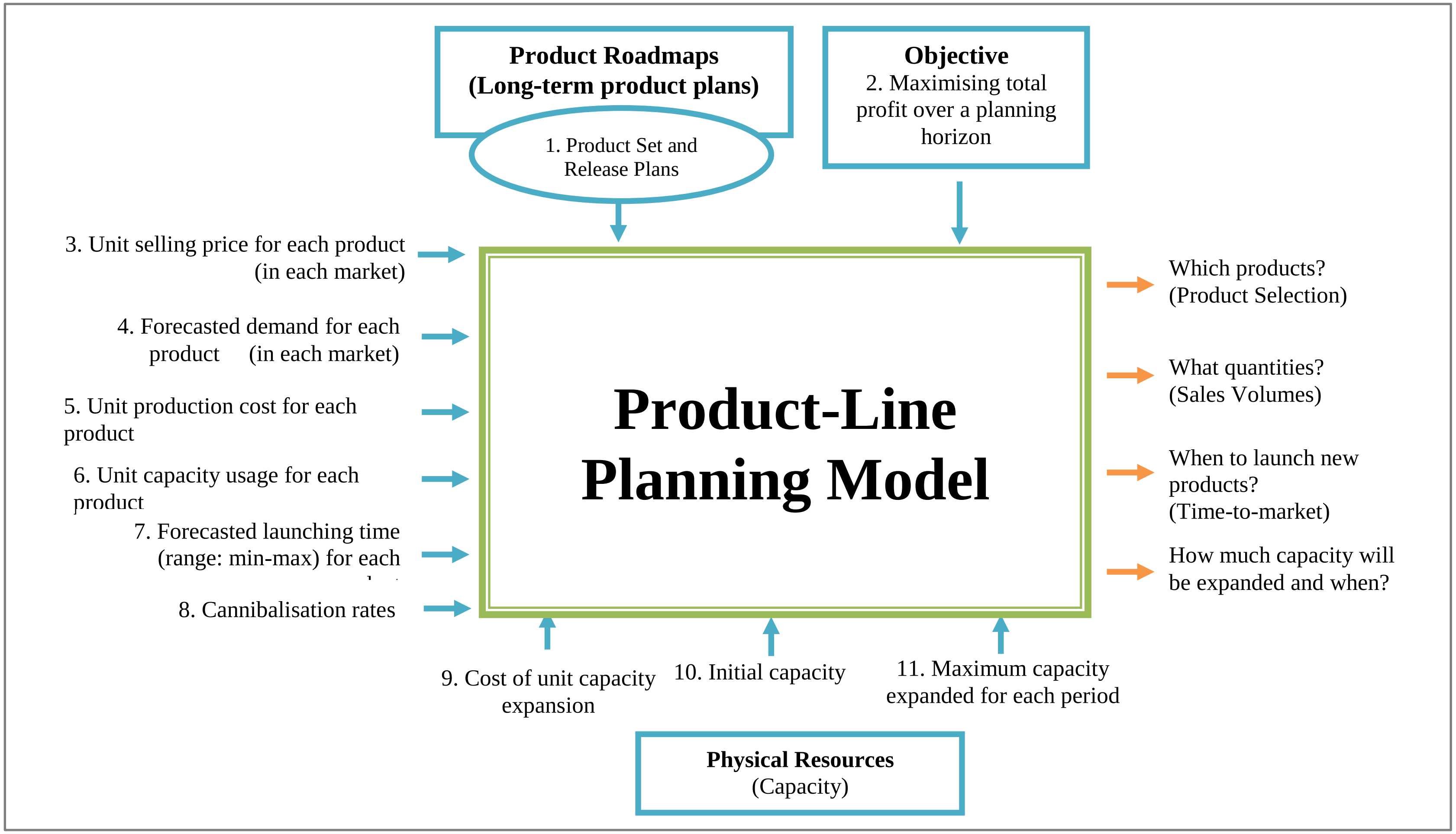


Figure 10. Input-output scheme of product-line planning model

Based on the input-output scheme presented in Figure 10, a deterministic mathematical formulation is developed for the problem and given in this section. Before presenting the mathematical formulation, in addition to the main characteristics of the problem and assumptions given in Section 3.1, some modelling issues and other assumptions required are summarised as follows:

- Decisions are made at the beginning of a period.
- Launching cost for new products such as cost of creating a new distribution channel, promotion and advertisement costs etc. introduced into the markets is negligible.
- The target markets in which a product is sold are determined for the planning horizon; i.e. set of target markets in which a product is sold cannot change on period-basis.
- There are two types of resources as regards to variable and fixed costs:
  - The first type, such as direct material, direct labour and variable overhead (e.g. indirect material that cannot be allocated to a certain product directly) costs, is required as an as-needed basis (Direct Costing System[4]).
  - For the second-type of resource, the firm operates at a level of capacity C_t in the beginning of the planning horizon, that will be available in *period t* at a cost b_t per unit of capacity. Thus, there will be "fixed overhead costs (FC_t)" at each period (such as insurance, rent, etc.); i.e. $FC_t = b_t * C_t$. Those costs are dependent on the committed capacity level and is not influenced by actual production level. Since those costs are fixed and considered as "constant values" in the objective function of

[4] Direct Costing System is similar to traditional full costing system, only the variable overhead costs are allocated to products, and fixed overhead costs are not considered in the cost of each unit of the product (Malik and Sullivan, 1995).

our problem, they are not taken into consideration in our optimisation problem.

Mathematical formulation of the product-line planning model

The sets, parameters and their related functions, decision variables, constraints and objective function of the (deterministic) mathematical model are given as follows:

Sets:

- $I = \{i | i = 1,2, \ldots, n_1 + n_2\}$: set of product lines (existing and new product lines considering all periods over the planning horizon)
  - $EP = \{i | i = 1,2, \ldots, n_1\}$: set of existing product lines
  - $NP = \{i | i = n_1 + 1, n_1 + 2, \ldots, n_1 + n_2\}$: set of new product lines planned to be introduced over the planning horizon.
- $T = \{t | t = 1,2, \ldots, |T|\}$: planning horizon (usually three to five years; divided into equal periods) and t represents periods/years
- $J = \{j | j = 1,2, \ldots, |J|\}$: set of target markets
- $M_i = \{j | j \in J$: set of markets in which product $i \in I$ is to be sold$\}$
- R_{ij}: set of products that interact with product $i \in I$ sold in market $j \in M_i$

Parameters:

c_i	:	unit capacity usage of product $i \in I$ (defined as the "average processing hours" required for producing one unit of a product)
$minL_{ij}$	:	earliest (planned) introducing time of product $i \in NP$ to market $j \in M_i$ (input from Product Road Maps)
EPD_{ijt}	:	potential demand for product $i \in EP$ in market $j \in M_i$ in period t, defined as a function of its own prices and the prices of other existing products (product interactions)
ND_{ijt}	:	demand for product $i \in NP$ in market $j \in M_i$ in period t, defined as a function of its own prices

P_{ijt} : selling price of product $i \in I$ in market $j \in M_i$ in period t

V_{it} : variable unit production cost of product $i \in I$ in period t, comprised of direct material, direct labour and variable overhead costs
It is assumed that variable cost of a product in a period does not depend on the launching time of the product.

β_t : cost of one unit of additional capacity
If additional capacity in period t is greater than zero ($\Delta_t > 0$), an additional cost of β_t for each additional unit of the capacity is incurred.

C_0 : Initial (existing) capacity at $t = 0$

$MAXC_t$: Maximum amount of capacity for each period t, i.e. the total capacity after a possible expansion cannot exceed $MAXC_t$ which represents the design capacity of the plants or the capacity allocated for the product class in period t

e_{ijlt} : Cannibalization rate of product $l \in NP$ on product $i \in EP$ (negative impact of product l on product i's demand, expressed as a percentage of potential demand of i) at time t; $0 < e_{ijlt} < 1$

Decision variables:

X_{ijt} : Number of units of product $i \in I$ to be available (sold) in market $j \in M_i$ in period t

Y_{ijd} : A binary variable taking 1 if product line $i \in NP$ is introduced to market $j \in M_i$ in period d and 0 otherwise

Δ_t : Additional capacity in period t (it is for adjusting the medium-term)

C_t : Total production capacity available in period t, allocated for the regarding product class, which includes the lines defined in set I (it is a real variable in the mathematical model)

Functions for demand-price relationship:

For the sake of simplicity, *a price dependent linear function* involving the product interdependencies is used for modelling the demand in this mathematical formulation of the problem. The details of other types of demand functions in the literature and a proposed procedure for selecting a function representing demand-price relationship in any problem is given in Appendix B.

The price dependent linear function is one of the most widely used functions in the literature and it is handled in an environment of single-firm without price competition. In real life, there are three types of this function:

- In single product case: Demand, *d(p)*, which is a linear function of price, *p*, can be formulised as follows: $d(p) = m - np,$ where $m, n > 0$ and $0 \leq \mathrm{p} \leq \mathrm{m/n}$
- In multi-product case: $d_i(p) = m_i - n_{ii}p_i + \sum_{j \neq i} n_{ij}p_j$ where n_{ii} (> 0): effect of product's own price on its demand; n_{ij}: the effect of other products' prices on its demand (if exists; > 0 if *j* is substitutable; < 0 if *j* is complement to product *i*)
- Piecewise linear demand function: In this function, price elasticity of demand takes different values in different price ranges.

In this study, the function in multi-product case and in single product case is used to represent the price dependent demand for existing products and for new (planned) products, respectively.

For existing products, $i \in EP$ ($j \in M_i$, $t \in T$), without any newly introduced products:

$$EPD_{ijt} = f(P_{ijt};\ P_{kjt}, k \in R_{ij} \cap EP) = m_{ijt} - n_{ij} \times P_{ijt} + \sum_{k \in R_{ij} \cap EP} n_{ikj} \times P_{kjt}$$

where $0 \leq P_{ijt} \leq m_{ijt}/n_{ij};\ m_{ijt} > 0;\ n_{ij} > 0$ denotes own-product elasticity for product *i* at market *j* in period *t* and $n_{ikj} > 0$ (since product *i* and *k* are substitutes) indicates the cross-price elasticity with respect to product *k*'s price at market *j* in period *t*. In this model, the price elasticities n_{ij} and n_{ikj} are price-independent constants.

For existing products, $i \in EP$, with new products to be launched within the planning horizon:

The demand function given above is revised by incorporating the cannibalisation effect of newly introduced products on existing products.

ED_{ijt} : demand for product $i \in EP$ at market $j \in M_i$ in period *t* (a parameter)

$$ED_{ijt} = EPD_{ijt} - \sum_{l \in SS_{ijt}} Y_{ljt} \times e_{ijlt} \times EPD_{ijt} = EPD_{ijt}\left(1 - \sum_{l \in SS_{ijt}} Y_{ljt} \times e_{ijlt}\right)$$

$$= \left(m_{ijt} - n_{ij} \times P_{ijt} + \sum_{k \in R_{ij} \cap EP} n_{ikj} P_{kjt}\right)\left(1 - \sum_{l \in SS_{ijt}} \sum_{d=minL_{lj}}^{t} Y_{ljd} \times e_{ijlt}\right)$$

$$SS_{ijt} = \left\{l \in R_{ij} \cap NP \;\middle|\; minL_{lj} \le t \le |T|\right\},\ \ i \in EP;\ \forall\, j \in M_i;\ \forall\, t \in T$$

Here, it is assumed that cannibalisation rate, e_{ijlt}, is dependent on the price of product $l \in R_{ij} \cap NP$, i.e. P_{ljt} and there exists a linear relationship between e_{ijlt} and P_{ljt}. Thus it can be expressed as follows:

$$e_{ijlt} = \alpha_{ijl} - \beta_{ijl} \times P_{ljt}$$

where α_{ijl} and β_{ijl} are constant and price dependent coefficient respectively.

For new products, $i \in NP$:

P_{ijt} : selling price of product i at period *t*

ND_{ijt} : (potential) demand of product *i* in market *j* at period *t*

$$ND_{ijt} = f(P_{ijt}) = m_{ijt} - n_{ij} \times P_{ijt}$$

$$\text{for } \forall\, i \in NP,\ j \in M_i,\ \forall\, t = minL_{ij}, minL_{ij} + 1, \dots, |T|$$

where $0 \leq P_{ijt} \leq m_{ijt}/n_{ij}$; $m_{ijt} > 0$; $n_{ij} > 0$ denotes own-product elasticity (which is a price-independent constant and can be estimated by analysing similar products or expert opinions) for product *i* at market *j* in period *t*.

Constraints:

1. New Product Launching Constraints: Those constraints ensure that a new product can be launched at a market in only one period that is selected from the range of $[minL_{ij}, |T|]$.

$$\sum_{d=minL_{ij}}^{|T|} \boldsymbol{Y_{ijd}} \leq \mathbf{1}, \qquad \forall i \in NP ; \forall j \in M_i$$

2. Demand Constraints: Those constraints ensure that total production volume for each product in each period cannot exceed the demand forecasted in that period.

(2.1) for existing products:

$$\boldsymbol{X_{ijt}} \leq \boldsymbol{EPD_{ijt}} * \left(\mathbf{1} - \sum_{(l,j)\in SS_{ijt}} \sum_{d=minL_{lj}}^{t} \boldsymbol{Y_{ljd}} \times \boldsymbol{e_{ijlt}}\right), \forall i \in EP; \forall j \in M_i; \forall t \in T$$

where

$$SS_{ijt} = \{(l,j) \in R_{ij} \cap NP \mid minL_{lj} \leq t \leq |T|\}, \ \forall i \in EP; \forall j \in M_i; \forall t \in T$$

$$EPD_{ijt} = m_{ijt} - n_{ij} \times P_{ijt} + \sum_{k\in R_{ij}\cap EP} n_{ikj} \times P_{kjt}, \ \forall i \in EP; \forall j \in M_i; \forall t \in T$$

$$e_{ijlt} = \beta_0 - \beta_1 \times P_{ljt}, \ \forall i \in EP; \ j \in M_i; \ l \in R_{ij} \cap NP$$

Regarding this constraint set, it should be noted that the potential demand for an existing product is not expected to be completely cannibalised by the new products. Since the problem considers the medium/long-term products and the product-line level, the cannibalisation at this level is not expected to be as severe as at model/variant level and the firm would not prefer to launch too many new product-lines in the same planning horizon (at most two new lines seem realistic). Therefore, the second term of ED_{ijt} (i.e. the term in the second parenthesis) is not exceeed 1 and an infeasibility problem because of the negativity of the right-hand-side of the constraint is not expected to occur.

(2.2) for new products:

$$\boldsymbol{X_{ijt}} \leq \boldsymbol{ND_{ijt}} \times \sum_{\boldsymbol{d=minL_{ij}}}^{\boldsymbol{t}} \boldsymbol{Y_{ijd}}, \quad \forall i \in NP, \forall\, j \in M_i, \forall\, t = \; minL_{ij}, minL_{ij}+1, \dots, |T|$$

where $ND_{ijt} = f(P_{ijt}) = m_{ijt} - n_{ij} \times P_{ijt}$

3. Capacity Constraints: There are three capacity constraints, defined as follows.

(3.1) Total resource consumption cannot exceed the capacity available in regarding period:

$$\sum_{(\boldsymbol{i},\ \boldsymbol{j}) \in \boldsymbol{S_t}} \boldsymbol{c_i} \times \boldsymbol{X_{ijt}} \leq \; \boldsymbol{C_t} \qquad \forall t \; = \; 1, 2, \dots, |T|$$

where

$$S_t = \left\{(i \in NP;\ j \in M_i) \;\middle|\; minL_{ij} \;\leq t \;\leq |T|\right\} \cup \{i \in EP; j \in M_i\}, \;\; \forall t \; = \; 1, 2, \dots, |T|$$

(3.2) Total production capacity available in each period after adding the amount of expansion:

$$\boldsymbol{C_t} = \; \boldsymbol{C_{t-1}} + \; \Delta_{\boldsymbol{t}} \qquad \forall t \; = \; 1, 2, \dots, |T|$$

(3.3) Total capacity in a period cannot exceed the maximum capacity, $MAXC_t$:

$$\boldsymbol{C_t} \; \leq \; \boldsymbol{MAXC_t} \qquad \forall t \; = \; 1, 2, \dots, |T|$$

4. Nonnegativity Constraints:

$$Y_{ijd} \;\; \in \{0,1\}, \;\; \forall\; i \in NP, \forall\; j \in M_i, \;\; d \; = \; minL_{ij} minL_{ij} + 1, \dots, |T|$$

$X_{ijt} \geq 0$ and integer[5] $\forall\, i \in I, \forall\, j \in M_i,\ \forall\, t \in T$ for $i \in EP$ and

$\forall\, t = minL_{ij}, minL_{ij} + 1, \ldots, |T|$ for $i \in NP$

$\Delta_t \geq 0$ $\forall t = 1, 2, \ldots, |T|$

Objective Function:

Maximise the profit = Quantity sold ×Contribution Margin (= Price - variable costs) – Total Capacity expansion costs

$$\boldsymbol{maximize\ OF} = \sum_{\boldsymbol{t=1}}^{|\boldsymbol{T}|} \sum_{(\boldsymbol{i},\ \boldsymbol{j}) \in \boldsymbol{S_t}} \boldsymbol{X_{ijt}} \times (\boldsymbol{P_{ijt}} - \boldsymbol{V_{it}}) - \sum_{\boldsymbol{t=1}}^{|\boldsymbol{T}|} \boldsymbol{\beta_t} \times \Delta_{\boldsymbol{t}}$$

The deterministic model given above is NP-hard, since it contains an integer bounded knapsack problem as a special case. To see this let

$$|\,\mathrm{J}\,| = 1\ (single\ market), |\,\mathrm{T}\,| = 1\ (single\ period) \text{ with } minL_i = 1 \text{ for all } i$$

(which leads the binary variable, Y_{ijd} , to become redundant) and make that sales of all products are independent of each other. If any capacity addition is not allowed (by defining the soft capacity constraints as hard constraints), the resulting problem reduces to an integer bounded knapsack problem that is known as NP-hard (Kellerer et al., 2004).

It should be noted that although it is stated an infeasibility problem in the case of that the second term of ED_{ijt} (i.e. the term in the second parenthesis) exceeds one and thereby the right-hand-side of the constraint set #2.1 becomes negative is not expected to occur in the context of this thesis work, in real life those kinds of situations may also arise. Therefore, in order to avoid from any infeasibilities in the

---

[5] The integrality restriction of this parameter may be relaxed in order to get easier solutions for the model. A noninteger solution can be obtained by relaxation, which may be rounded to its closest integer value. As a result, an infeasibility may occur; however, since the integer variables take very large values in this problem, the rounding up or rounding down operation would not significantly affect the optimal value of objective function. Besides, this relaxation also provides to solve the stochastic model more efficiently.

mathematical model it should be ensured that the right-hand-sides of those constraints take nonnegative values. For this purpose, the following revision should be made in the constraint set #2.1:

$$\boldsymbol{X_{ijt}} \leq \max\left\{\boldsymbol{0},\ \boldsymbol{EPD_{ijt}} \times \left(\boldsymbol{1} - \sum_{(\boldsymbol{l},\boldsymbol{j})\in \boldsymbol{SS_{ijt}}} \sum_{\boldsymbol{d}=\boldsymbol{minL_{lj}}}^{\boldsymbol{t}} \boldsymbol{Y_{ljd}} \times \boldsymbol{e_{ijlt}}\right)\right\},$$

$$\forall i \in EP;\ \forall\, j \in M_i;\ \forall\, t \in T$$

In order to preserve the linearity of the model, these constraints are linearised as given below and then added to the mathematical model instead of the constraint set #2.1. In addition to that, a new term is added to the objective function of the model.

- Define two nonnegative variables as τ_{ijt} and γ_{ijt}.
- For $\forall i \in EP, \forall\, j \in M_i, \forall\, t \in T$:
  - $X_{ijt} \leq \tau_{ijt} + \gamma_{ijt}$
  - $X_{ijt} \leq\ EPD_{ijt} \times \left(1 - \sum_{(l,j)\in SS_{ijt}} \sum_{d=minL_{lj}}^{t} Y_{ljd} \times e_{ijlt}\right) +\ \tau_{ijt} + \gamma_{ijt}$
  - $\tau_{ijt} \geq EPD_{ijt} \times \left(1 - \sum_{(l,j)\in SS_{ijt}} \sum_{d=minL_{lj}}^{t} Y_{ljd} \times e_{ijlt}\right)$
  - $\gamma_{ijt} \geq -\, EPD_{ijt} \times \left(1 - \sum_{(l,j)\in SS_{ijt}} \sum_{d=minL_{lj}}^{t} Y_{ljd} \times e_{ijlt}\right)$
  - $\tau_{ijt} \geq 0$ and $\gamma_{ijt} \geq 0$
- Revise the objective function of the model as follows:

  $maximize\ OF$

$$= \sum_{t=1}^{|T|} \sum_{(i\ ,\ \ j)\,\in\, S_t} \left(X_{ijt} \times (P_{ijt} - V_{it}) - (\tau_{ijt} + \gamma_{ijt})\right) - \sum_{t=1}^{|T|} \beta_t \times \Delta_t$$

Furthermore, as implied above, product-line planning problem involves both launching and capacity expansion decisions, and annual production amounts (in other words, sales targets). The decision process in this problem can be partitioned into two stages. Product launching and capacity expansion decisions are taken at the beginning of the planning horizon based on a limited information regarding selling price, demand, production cost and cannibalisation rate due to the long lead time for capacity installation and making new products ready for the target markets. On the other hand decisions for production amounts (sales targets) are postponed until the information about the uncertain parameters are revealed at the beginning of each period, i.e. a specific scenario is realised. Thus, product-line planning problem fits in with two-stage stochastic programming approach under the condition that the decision maker deals with uncertainities within the problem environment. Therefore, the deterministic model given above is extended to a two-stage stochastic optimisation model, in which the optimisation is done over all considered future realisations of uncertain parameters, i.e. scenarios, and it is assumed that uncertainty of selling price, demand, production cost and cannibalisation rate are represented by known probability distributions that are estimated using historical data. This stochastic model is given in the following section.

3.3. Two-Stage Stochastic Programming Model with Recourse

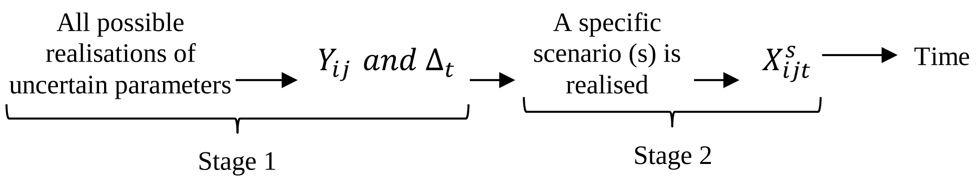
In this section, a two-stage stochastic programming (TSP) model with recourse in order to handle the uncertain parameters, i.e. demand, price, variable production cost and cannibalisation rate, is developed by revising the deterministic model presented in Section 3.2.

As mentioned in Section 2.4.1, the multi-period product-mix problem under uncertainty handled in this study can be formulated as a two-stage or multi-stage stochastic programming model. In the two-stage stochastic programming models, the decisions in the problem context are partitioned into two sets: (1) first-stage decisions and (2) second-stage decisions. The first-stage decisions, which include capacity expansion and new product launching decisions considered for the entire planning horizon, are made at the beginning of the planning horizon before the uncertainty is revealed, and are not changed over the periods. Then, the second-stage decisions, which include sales volumes, are made in each period based on the first-stage decisions and the certain realisations of the uncertain parameters. Though the two-stage stochastic programming models are commonly used in the literature, its main drawback is not to have any flexibility in the capacity expansion and new product launching plans in accordance with the realisations of parameters that are considered as uncertain at the beginning of the planning horizon (Huang and Ahmet, 2009). On the other hand, the multi-stage stochastic programming models allow the decision makers to revise the capacity expansion and new product launching plans when more information about the uncertainties is revealed at each time stage. Thus, as also mentioned in Section 2.4.1, the multi-stage stochastic programming models provide a dynamic decision making process over the entire planning horizon and more flexibility than the two-stage stochastic programming models (Huang, 2005; Huang and Ahmet, 2005; Huang and Ahmet, 2009). Therefore, the multi-period product-mix problem handled in this study should be formulated as a multi-stage stochastic programming model in order to represent the dynamism of the real life better. However, it is more difficult to solve multi-stage stochastic programming those models because of their computational complexity due to the number of stages and the number of variables than the two-stage models (Huang, 2005; Solak, 2007). In our problem context, many uncertain parameters involving demand, selling price, production cost and cannibalisation rate, product interactions as well as two main decisions, i.e. capacity expansion and new product launching decisions, are handled and it should be noted that there is already a complexity due to this problem framework. In addition to this complexity, if the problem is formulated as a multi-stage stochastic programming model, there will be another complexity, and thereby additional computational burden due to the structure of the multi-stage stochastic programming model. However, our main motivation in this study is to propose a comprehensive model and to be able to do analyses for detecting which problem parameters (both deterministic and uncertain parameters) are significant on the solutions. Therefore, in order to avoid from the additional computational burden that may cause difficulties while performing the analyses, we formulate the problem that is handled for the first time in this context as a two-stage stochastic model for all practical purposes.

The main assumptions of the proposed two-stochastic programming model are as follows:

1- The future realisations of uncertain parameters in a period (scenarios) does not affect the realisations of uncertain parameters in the previous stages, and scenarios and their associated probabilities are independent of the decisions taken.
2- The events, e.g. the evolution of selling prices or the demand-price relationships, for each demand market are independent.
3- Price, demand, variable production cost and cannibalisation rate are uncertain parameters whose probability distributions are known and the uncertainties are expressed by a set of scenarios.
4- All other parameters such as unit capacity usage, maximum capacity in each period and capacity addition cost are considered as deterministic, i.e. the expected value of those parameters are used.
5- In TSP model, two sets of decisions to be made are considered: the **first-stage decisions** are made before the actual realization of uncertain parameters and then the **second stage decisions** are made as recourse actions at a later stage, based on the information obtained from the first-stage decisions after the uncertainties are disclosed (Morales, 2007).

The sequence of events can be represented as follows:



In the beginning of the planning horizon, new products that will be produced (made available) over the planning horizon and their launching times, and the amount of additional capacity are determined as first-stage decisions, i.e. Y_{ij} and Δ_t . Those are implemented without waiting for the realization of random events (certain demand, price and cost values in a period). Since making any product-line available to be produced at the time it will be introduced to a market needs a preparation time (e.g. one or two years), waiting until the random variables becomes certain may cause not offering a new product to any market in a timely manner or not having enough capacity for satisfying the demand that will contribute to the profit. Therefore, the aim of TSP model is to make the first-stage decisions considering the evolution of random parameters throughout the planning horizon by means of minimising the sum of the first-stage and the expected second-stage costs (Morales, 2007; Sahinidas, 2004). Then, after a possible scenario, s, is realised, i.e. demand, price, cost and cannibalisation rate parameters becomes more certain, the volume of the products that will be sold in the second stage are determined by also considering the capacity and product launching decisions made in the first stage. Actually, the second-stage costs can be considered as precautions to be taken in order to hinder any possible infeasibilities resulting from a certain realisation of uncertain parameters (Morales, 2007).

In accordance with two sets of decisions to be taken, two sets of constraints are also defined: (1) first stage-constraints regarding first stage decisions (new product launching and capacity expansion decisions) and (2) second-stage constraints regarding the scenario-based decisions (volume of products at each market and in each period under a set of scenarios). Thus, the mathematical representation of TSP model can be defined.

In order to build the two-stage stochastic programming model, firstly a discrete set of scenarios that are represented by s is defined as follows:

$$\Omega = \{\mathrm{s} \,\big|\, \mathrm{s} = 1,2, \dots, |\mathrm{S}| \}$$

Here, the scenario set, Ω, denotes the set of all possible outcomes of random parameters and is assumed to be finite and discrete while constructing the stochastic model. For the continuous case, it is also possible to use this model by creating a finite Ω via discretisation or direct sampling from the probability density functions.

Secondly, the probability of occurrence for each scenario $s \in \Omega$ is defined as , p_s.

The representations for uncertain parameters considered in this study can be given as follows:

EPD_{ijt}^{s} : (potential) demand of product $i \in EP$ in market $j \in M_i$ at period t, under scenario $s \in \Omega$

ND_{ijt}^{s} : demand of product $i \in NP$ in market $j \in M_i$ at period t, under scenario $s \in \Omega$

P_{ijt}^{s} : selling price of product *i* in market $j \in M_i$ at period *t*, under scenario $s \in \Omega$

V_{it}^{s} : variable unit production cost of product *i* at period *t*, under scenario $s \in \Omega$

e_{ijlt}^{s} : cannibalization rate of product $l \in NP$ on product $i \in EP$ in market $j \in M_i$ at period t, under scenario $s \in \Omega$, $0 < e_{ijlt}^{s} < 1$

First-stage decision variables: fixed in the second stage of TSP model and their input data are free of uncertainty.

Y_{ijd} : 1, if product line $i \in NP$ is introduced to market $j \in M_i$ in period *d* and 0, otherwise.

Δ_t : Additional capacity in period *t* (for adjusting the medium-term capacity if required)

C_t : Total production capacity available at time *t*, allocated for the regarding product class, which includes the lines defined in set *I*.

Second-stage decision variables: depends upon both the first stage decisions and on the realization of the stochastic parameters.

X_{ijt}^{s} : Number of units of product *i* to be available (sold) in market *j* at period *t*, under scenario $s \in \Omega$

Objective Function:

maximise the expected profit over all scenarios, interpreted as the expected profit (i.e., long-run average profit) if the product plan given by the variables Y and production plan given by the variables Δ are implemented.

$$\max z = \sum_{s \in \Omega} \sum_{t=1}^{|T|} \sum_{(i,j) \in S_t} p_s \times X_{ijt}^s \times (P_{ijt}^s - V_{it}^s) - \sum_{t=1}^{|T|} \beta_t \times \Delta_t$$

First-stage Constraints:

1. New Product Launching Constraints:

$$\sum_{d=minL_{ij}}^{|T|} Y_{ijd} \leq 1, \qquad \forall i \in NP\ ;\ \forall\, j \in M_i$$

2. Capacity Constraints:

$$C_t = C_{t-1} + \Delta_t \qquad \forall t = 1, 2, \dots, |T|$$

$$C_t \leq MAXCAP_t \qquad \forall t = 1, 2, \dots, |T|$$

3. Nonnegativity Constraints:

$$Y_{ijd} \in \{0,1\} \qquad \forall\, i \in NP, \forall\, j \in M_i,\ d = minL_{ij} minL_{ij} + 1, \dots, |T|$$

$$\Delta_t \geq 0 \qquad \forall\, t = 1, 2, \dots, |T|$$

Second-stage Constraints:

1. Demand Constraints:

For existing products:

$$X_{ijt}^s \leq EPD_{ijt}^s \times \left(1 - \sum_{(l,j) \in SS_{ijt}} \sum_{d=minL_{lj}}^{t} Y_{ljd} \times e_{ijlt}^s\right)$$

$$\forall\, i \in EP, \forall\, j \in M_i, \forall\, t \in T, \forall\, s \in \Omega$$

where

$$SS_{ijt} \;=\; \left\{(l,j) \in R_{ij} \cap\; NP \;\middle|\; minL_{lj} \;\leq t \;\leq |T|\right\} \text{ for } i \in EP, \forall\, j \in M_i, \forall\, t \in T$$

$$EPD_{ijt}^{s} = m_{ijt} - n_{ij} \times P_{ijt}^{s} \;+ \sum_{k \in R_{ij} \cap\, EP} n_{ikj} \times P_{kjt}^{s}$$

$$\forall\, i \in EP, \forall\, j \in M_i, \forall\, t \in T, \forall\, s \in \Omega$$

$$e_{ijlt}^{s} = \;\alpha_{ijl} - \beta_{ijl} \times\; P_{ljt}^{s}\,, \quad \forall\, i \in EP,\; j \in M_i,\; l \in R_{ij} \cap\; NP$$

As implied for the deterministic mathematical model, in order to avoid from any infeasibilities in the stochastic model, which may be occurred due to the possibility of that the right-hand-sides of those demand constraints take negative values, the same operation as in the deterministic model should be made for those constraints by revising τ_{ijt} and γ_{ijt} variables as τ_{ijt}^{s} and γ_{ijt}^{s} .

For new products:

$$X_{ijt}^{s} \;\leq ND_{ijt}^{s} \times \sum_{d=minL_{ij}}^{t} Y_{ijd}$$

$$\forall\, i \in NP;\; \forall\, j \in M_i;\; \forall\, t = \; minL_{ij}, minL_{ij} + 1, \ldots, |T|;\; \forall\, s \in \Omega$$

where

$$ND_{ijt}^{s} = f(P_{ijt}^{s}) = m_{ijt} - n_{ij} \times P_{ijt}^{s}.$$

2. Capacity Constraints:

$$\sum_{(i\,,\;\, j) \in S_t} c_i \times X_{ijt}^{s} \leq\; C_t \qquad\qquad \forall\, t \;=\; 1, 2, \ldots, |T|, \forall\, s \in \Omega$$

$$S_t = \;\left\{(i \in NP;\; j \in M_i) \;\middle|\; minL_{ij} \;\leq t \;\leq |T|\right\} \cup \{i \in EP; j \in M_i\} \;\text{ for } \forall t \;=\; 1, 2, \ldots, |T|$$

3. Nonnegativity Constraints:

$$X_{ijt}^{s} \;\geq 0 \qquad \forall\, i \in I;\; \forall\, j \in M_i;\;\; \forall\, s \in \Omega$$

$$\forall\, t \in T \text{ for } i \in EP \;\text{ and} \;\forall\, t = \; minL_{ij}, minL_{ij} + 1, \ldots, |T| \text{ for } i \in NP$$

The model presented above, which considers a discrete set of scenarios, is equivalently formulated as a large-scale linear program (i.e. its extensive form) and can be solved using standard MIP solution procedures. In fact, a stochastic programming problem cannot be solved directly in its extensive form, because this will increase the complexity of the model and cause a long computation time if the scenario set is large-scale. Therefore, the stochastic programming model, under the condition that recourse function is convex, can be efficiently solved using decomposition-based strategies such as Langrangian and Benders schemes.

If there are a large but finite number of scenarios, TSP model is tractable by using large-scale linear programming techniques such as Benders' decomposition (Bertsimas, 2011). Because of this reason, for obtaining an efficient solution, Benders' decomposition method (also known as L-shaped method that converges to optimal solution**)** is used for solving the two-stage stochastic program that cannot be solved directly in its extensive form, given above**.**

In order to check whether the developed stochastic model above with a certain number of scenarios is solved in a reasonable time, three problem instances (one is small-sized, the other is medium-sized and the last one is large-sized) are tested with an increasing number of scenarios (those instances are also used for computational study given in Section 5.2 and one can refer to this section for the details of them). The results are given in Table 2[6].

[6] Those problem instances are solved by GAMS 22.2/CPLEX 10.0 on a computer with 2.2 GHz speed and 16 GB of RAM.

Table 2. Computational results for some problem instances in extensive form

Problem					# of scenarios	# of first stage decision variables	# of second stage decision variables	Solution time (seconds)
No.	# of existing lines	# of new lines	# of markets	# of periods				
1	4	1	10	3	500	36	75000	3
					1000	36	150000	9.2
					5000	36	750000	267
					10000	36	1500000	990
2	6	2	20	5	500	210	400000	193
					1000	210	800000	543
					1500	210	1200000	Out of memory
3	8	4	20	5	500	1210	600000	480
					1000	1210	1200000	Out of memory

Based on the results given in Table 2, the extensive form of Problem 1 can be solved in a very short time even if there are very large number of scenarios, i.e. 10000. However, when it is attempted to solve the extensive form of Problem 2 with 1500 scenarios and of Problem 3 with 1000 scenarios, CPLEX optimiser resulted in an out-of-memory error. Besides, the solution approach given in Section 4.1 may require more than 1000 number of scenarios for some cases in order to obtain an acceptable solution; therefore, L-shaped method converging to the optimal solution shoud be used at least for some medium and large-sized problem, e.g. the Problem 2 and Problem 3 given above. By applying L-shaped method, the extensive form of Problem 2 with 1500 scenarios and Problem 3 with 1000 scenarios could be solved optimally in 3900 seconds and 2188 seconds, respectively. Therefore, L-shaped method is adapted for TSP in order to get efficient solutions when it could not be possible to solve the entensive form of TSP.

In the next section, firstly, the generalised L-shaped method, secondly its special form developed for our TSP model will be presented. Then, an overall solution approach for solving TSP will be given.

3.4. L-shaped Method for Solving TSP with a Finite Number of Scenarios

L-shaped method is developed for solving the extensive form of stochastic programs (EFSP), which involves all possible outcomes (scenarios) $s \in \Omega$. Solving EFSP may not be easy because of having too many realizations of $s \in \Omega$, therefore, the most frequently used method, L-shaped method (stochastic version of Benders decomposition) is employed, which is based on "*building an outer linearization of the recourse cost function and a solution of the first stage problem plus this linearization*" (Birge and Louveaux, 2011, p.181). This method separates EFSP into a master problem and a set of independent sub-problems setting up for each scenario and which are used to generate cuts (Infanger, 1997). The master problem, sub-problems, cuts and the algorithm of this method is summarised as follows:

The Original Model (EFSP) with discrete set of scenarios:

$$\underset{y}{\text{minimise}} \quad c^T y + \sum_{s \in \Omega} p_s \ d_s^T x_s$$

s.t.

$$Ay = b$$

$$-T_s y + W_s x = h_s$$

$$y \geq 0, \ x_s \geq 0$$

where s denotes a scenario or possible outcome, $s \in \Omega$; p_s denotes the occurrence probability of the scenarios, $s \in \Omega$; y represents the first stage variables; x_s represents the second stage variables; and $Ay = b, y \geq 0$ represents the first stage constraints and $-T_s y + W_s x = h_s$, $x_s \geq 0$ second-stage constraints.

<u>L-shaped Method / Benders' Algorithm (A Generalised Form):</u>

The algorithmic steps of this method, for solving the deterministic equivalent linear model, given above, can be summarised as follows (Kalvelagen, 2003):

Step 1: Initialization

$k \quad := \quad 1$ {iteration number}

$UB := \quad \infty$ {Upper bound}

$LB := -\infty$ {Lower bound}

Solve the initial master problem:

$$\underset{y}{\text{minimise}} \quad c^T y$$

$$Ay = b$$

$$y \geq 0$$

$\bar{y}^k := y^*$ {optimal values}

Step 2: Sub problems (a series of problems dealing with the second-stage variables)

do

Solve the sub problem, $s \in \Omega$:

$$min \quad d_s^T x_s$$

$$W_s x = h_s + T_s \bar{y}^k$$

$$x_s \geq 0$$

$\bar{x}_s^k := x_s^*$ {optimal values}

$\bar{\pi}_s^k := \pi_s^*$ {optimal simplex multipliers}

end for

$UB := \min\{UB, \; c^T \bar{y}^k + \sum_{s \in \Omega} p_s \; d_s^T \bar{x}_s^k \}$

Step 3: Convergence test

if $|UB - LB| / |LB| \leq A\ specified\ tolerance\ value$

then Stop: convergence is achieved

Return $\bar{y}^k$

end if

else

Step 4: Master problem (focus on the first stage variables)

do $k = k + 1$.

Solve the master problem:

$$\begin{aligned} minimize \quad & c^T y + \theta \\ & Ay = b \\ & \theta \geq \textstyle\sum_{s\in\Omega} p_s \left[\bar{\pi}_s^l \left(h_s + T_s y \right) \right] \quad l = 1, 2, \dots, k-1 \\ & y \geq 0 \end{aligned}$$

$\bar{y}^k := y^*$ {optimal values of first stage variables}

$\bar{\theta}^k := \theta^*$

$LB := c^T \bar{y}^k + \bar{\theta}^k$

go to Step 2.

The constraint involving θ is the Benders' optimality cut generated by aggregating the simplex multipliers (single-cut). If any sub-problem becomes infeasible for some of the feasible values of first-stage variables, slightly different formulated cuts, called as feasibility cuts, need to be included to the model (for details see Birge and Louveaux, 2011, p.13). Besides, instead of adding feasibility cuts, it is also possible to derive some constraints (called as induced constraints) that must be satisfied in order to guarantee second-stage feasibility (Birge and Louveaux, 2011).

In other version of L-shaped method, the multi-cut version, is also proposed in the stochastic programming literature, in which one cut per scenario $s \in \Omega$ in the second stage is placed instead of one aggregated cut in single-cut version. By placing multiple cuts, it is possible to send more information to the master problem (Birge and Louveaux, 2011). Thus, the master problem at iteration $\boldsymbol{k}$ is revised as follows:

Step 4: Master problem at iteration $\boldsymbol{k}$ (focus on the first stage variables)

do $k = k + 1$.

Solve the master problem:

$$\begin{aligned} minimize \quad & c^T y + \textstyle\sum_{s\in\Omega} \theta_s \\ & Ay = b \\ \theta_s \ \geq \ & p_s[\bar{\pi}_s^r \,(h_s + T_s y)] \quad r = 1,2,\dots,k-1;\ s \in \Omega \\ & y \geq 0 \end{aligned}$$

$\bar{y}^k \ := \ y^*$ {optimal values of first stage variables}

$\bar{\theta}_s{}^k \ := \ \theta_s^*$

$LB \ := \ c^T \bar{y}^k + \sum_{s\in\Omega} \bar{\theta}_s{}^k$

go to Step 2.

Considering our problem, TSP, given in Section 3.3., the L-shaped method with multiple cuts can be employed to TSP with a set of finite number of scenarios, say Ω, (e.g. a scenario set obtained by random sampling from continuous probability distributions of uncertain parameters or subjectively determined scenarios) by following the steps below.

<u>L-Shaped Method Adapted to TSP:</u>

Assume that the problem involves n number of scenarios represented by a scenario set $\Omega = \{s|s = 1,2,3,\dots.n\}$ with p_s denoting the occurrence probability of the scenario $s \in \Omega$.

Step 1: Initialization

$k \quad := 1$ {iteration number}

$UB := \quad \infty$ {Upper bound}

$LB := -\infty$ {Lower bound}

do

Solve initial master problem without any cut:

$$\min \sum_{t=1}^{|T|} \beta_t \times \Delta_t$$

s.t.

$$\sum_{d=minL_{ij}}^{|T|} Y_{ijd} \leq 1, \quad \forall i \in NP\ ;\ \forall\, j \in M_i$$

$$C_t = C_{t-1} + \Delta_t \quad \forall t = 1, 2, \dots, |T|$$

$$C_t \leq MAXCAP_t \quad \forall t = 1, 2, \dots, |T|$$

$$Y_{ijd} \in \{0,1\}, \quad \forall\, i \in NP;\ \forall\, j \in M_i;\ \ d = minL_{ij} minL_{ij} + 1, \dots, |T|$$

$$\Delta_t \geq 0, \quad \forall t = 1, 2, \dots, |T|$$

$\bar{Y}^k := Y^*$ {optimal values}
$\bar{\Delta}^k := \Delta^*$ {optimal values}

Step 2: Sub-problems (a series of problems dealing with the second stage variables):

For each $s \in \Omega$;

Solve the sub problem, given $\bar{Y}^k$ and $\bar{\Delta}^k$ as the optimal solution to the master problem

$$z^s(\bar{Y}^k, \bar{\Delta}^k) = minimize\ - \sum_{t=1}^{|T|} \sum_{(i,j) \in S_t} X_{ijt}^s \times (P_{ijt}^s - V_{it}^s)$$

Subject to:

$$\pi_{ijt}^{k,s} : X_{ijt}^s \leq EPD_{ijt}^s \times \left(1 - \sum_{(l,j) \in SS_{ijt}} \sum_{d=minL_{lj}}^{t} \bar{Y}_{ljd}^k \times e_{ijlt}^s\right),\ i \in EP, \forall\, j \in M_i, \forall\, t \in T$$

$$\delta_{ijt}^{k,s} :\ X_{ijt}^s \leq ND_{ijt}^s \times \sum_{d=minL_{ij}}^{t} \bar{Y}_{ijd}^k,$$

$$\forall i \in NP, \forall\, j \in M_i, \forall\, t = minL_{ij}, minL_{ij} + 1, \dots, |T|$$

$$\mu_t^{k,s}: \quad \sum_{(i,\ j) \in S_t} c_i \times X_{ijt}^s \leq \bar{C}_t^k, \quad \forall t = 1, 2, \dots, |T|$$

$$X_{ijt}^s \geq 0, \ \ \forall\, i \in I, \forall\, j \in M_i,$$

$\forall\, t \in T$ for $i \in EP$ and $\forall\, t = minL_{ij}, minL_{ij} + 1, \dots, |T|$ for $i \in NP$

$\bar{X}_s^k := X_s^* \{optimal\ values\}$

$\pi_{ijt}^{k,s},\ \delta_{ijt}^{k,s},\ \mu_t^{k,s}$: optimal simplex multipliers of sub-problem s, given $\bar{Y}^k$, $\bar{C}^k$,the first stage decisions passed on iteration k to the subproblems

end for

$UB^{\,k} = \; \min\left\{UB^{k-1},\; \sum_{t=1}^{|T|} \beta_t * \bar{\Delta}^{\,k} \;+\; \sum_{s\in\Omega} p_s\, z^s(\bar{Y}^k, \bar{\Delta}^k)\;\right\}$

Step 3: Convergence test

if $|UB^{\,k} - LB^k| / \, |LB^k| \;\leq\; A\ specific\ tolerance\ value,$ **then**

Stop: convergence is achieved.

Return $\bar{Y}^k$ and $\bar{\Delta}^k$ as the optimal solution to the original problem.

else $k = k + 1$

Step 4: Master problem (focus on the first stage variables)

Solve the master problem:

$$v^k = minimize \quad \sum_{t=1}^{|T|} \beta_t \times \Delta_t \;+\; \sum_{s\in\Omega} \theta_s$$

Subject to.

$$\sum_{d=minL_{ij}}^{|T|} Y_{ijd} \;\leq 1, \qquad \forall i \in NP\,, \forall\, j \in M_i$$

$$C_t = \; C_{t-1} + \;\Delta_t \qquad \forall t \;=\; 1, 2, \dots, |T|$$

$$C_t \;\leq\; MAXCAP_t \qquad \forall t \;=\; 1, 2, \dots, |T|$$

Optimality Cut:

$$\theta_s \geq \sum_{s\in S} p_s \Bigg[\sum_{i\,\in\,EP} \sum_{j\in M_i} \sum_{t\in\,T} \pi_{ijt}^{r,s} \times EPD_{ijt}^{s} + \sum_{i\,\in\,NP} \sum_{j\in M_i} \sum_{t=minL_{ij}}^{|T|} \delta_{ijt}^{r,s} \times ND_{ijt}^{s} \times \sum_{d=minL_{ij}}^{t} Y_{ijd} - \sum_{i\,\in\,EP} \sum_{j\in M_i} \sum_{t=min\{minL_{fj}, f\,\in\,R_{ij}\}}^{|T|} \pi_{ijt}^{r,s} \times EPD_{ijt}^{s} \times \sum_{l\in\,SS_{it}} e_{ijlt}^{s} \sum_{d=minL_{lj}}^{t} Y_{ljd} \;+\; \sum_{t\in\,T} \mu_t^{r,s} \times\; C_t \Bigg]$$

$$r = 1,2,\ldots,k-1^{\,7};\ \forall\ s \in \Omega$$

$Y_{ijd} \in \{0,1\}$ $\forall\ i \in NP;\ \forall\ j \in M_i;\ \ d = minL_{ij}minL_{ij} + 1, \ldots, |T|$

$\Delta_t \geq 0$ $\forall t = 1,2,\ldots,|T|$

$\theta \geq 0$

$\bar{Y}^k := Y^*$ {optimal values of first stage variables}

$\bar{\Delta}^k := \Delta^*$ {optimal values of first stage variables}

$\bar{\theta}_s^k := \theta_s^*$

$$LB^k := \sum_{t=1}^{|T|} \beta_t * \bar{\Delta}^k + \sum_{s\in\Omega} \bar{\theta}_s^k$$

go to Step 2.

As seen above, the algorithm starts with finding a solution set, $\bar{Y}^1$ and $\bar{\Delta}^1$, obtained from the relaxed master problem (step 1)[8]. The master problem at iteration k is solved by adding optimality cuts to obtain a trial solution, $\bar{Y}^k, \bar{\Delta}^k$ and $\bar{\theta}^k$. Given this solution, each sub-problem defined for each scenario $s \in$ S is optimally solved in order to calculate the expected second-stage costs, z $(\bar{Y}^1, \bar{\Delta}^1)$, and the coefficients and the right hand side of the optimality cut that will be add to the master problem in the next iteration. At each iteration of L-shaped method, a first-stage solution ($\bar{Y}^k, \bar{\Delta}^k$ and $\bar{\theta}^k$), an upper bound (UB^k) and a lower bound (LB^k) for the stochastic problem is solved. The algorithm continues until the stopping criterion,

---

[7] In this constraint, the first part, $\sum_{s\in\Omega} p_s\left[\sum_{i\,\in\,EP} \sum_{j\in M_i} \sum_{t\in T} \ \pi_{ijt}^{r,s} * EPD_{ijt}^s\ \right]$ denotes the constant of the optimality cut, and the remaining parts are the coefficients of the cut, which are calculated in each iteration.

[8] Since the potential demand for an existing product will not be completely cannibalised by the new products (such a scenario will not be realistic), an infeasibility problem because of the negativity of the right-hand-side of the demand constraints defined for existing products in the second stage will not occur. Furthermore, this problem will also not be possible for the capacity constraints. Therefore, any feasibility cuts are added to the master problem.

$|UB^{\,k} - LB^{k}| / |LB^{k}| \leq A\ specific\ tolerance\ value$, is met. Then, the optimal solution for the original problem is reported as the best solution obtained:

$$\{\bar{Y}^{*}, \bar{\Delta}^{*}\} \in \arg\min \left\{\sum_{t=1}^{|T|} \beta_t \times \bar{\Delta}^{k} + \sum_{s \in \Omega} p_s\, z^{s}(\bar{Y}^{k}, \bar{\Delta}^{k}),\ k = 1,2,\ldots\right\}.$$

L-shaped method whose main steps are presented above provides an exact solution for the original problem (TSP) on the condition that there are finite and discrete set of scenarios. However, for the problems wherein there are large number of scenarios, which make L-shaped method be not manageable/tractable for even discrete cases or there are infinite number of scenarios that are obtained through continuous probability distributions, it requires to reduce the scenario set to a tractable/manageable size. One of the most commonly used strategy for solving those kind of problems is to use Monte Carlo (random) sampling procedure to generate a certain number of scenarios instead of considering all realisations of random parameters and to find a solution for the original problem through solving a series of approximated problems (Mak et al.,1999; Shapiro and Homem-de-Mello, 1998, Kleywegt et al., 2001; Bayraksan and Morton, 2011; Kim et al., 2015; Bidhandi and Patrick, 2017; Linderoth et al., 2002; Infanger, 1997; Infanger, 1992). This solution approach is known as Sample-Average Approximation (SAA) in the related literature. The next chapter presents a solution methodology based on a bounding technique consisting of SAA approach based on Monte Carlo random sampling.

CHAPTER 4

THE SAMPLE AVERAGE APPROXIMATION APPROACH FOR THE TWO-STAGE STOCHASTIC PROGRAMMING MODEL

The problem handled in this study deals with a two-stage stochastic programming model developed for the product-mix problem under uncertainty (TSP), in which there are too many random parameters characterised by continuous probability distributions. Since the scenarios representing the possible realisations of random parameters are generated from those distributions and thereby there are infinite number of scenarios, it is not possible to solve the deterministic equivalent model using the L-shaped method. Therefore, a solution approach based on approximation, bounding technique and random sampling method is developed.

Firstly, the theoretical background of this approach and a flow diagram for the solution approach of TSP is presented in Section 4.1 and then performance measures for computational study that will be given in Chapter 5 is presented in Section 4.2.

4.1. Solution Approach for TSP

Instead of solving a large-scale stochastic program with a nonmanageable set of scenarios exactly, it requires *approximately* solving the *original problem (OP)* using a certain number of scenarios generated by a sampling procedure.

The original problem (Mak et al., 1999) :

$$OP: \quad z^* = \max_{x \in X} Ef(x,\omega) \qquad\qquad x^* = \operatorname{argmax}_{x \in X} Ef(x,\omega)$$

$$\text{where } f(x,\omega)^9 = cx + \min_{y \geq 0} by$$

$$s.t. \qquad Dy = Ax + e$$

f: real-valued function
x: a vector of decision variables with feasible set X
ω: a vector of random variables

In order to solve OP in a tractable and efficient way, at the beginning of the solution procedure, an independent sample S with a size of $|S| = N$ is drawn from the underlying distribution of possible outcomes (external sampling) and then the *approximate problem (AP)* is solved based on this sample.

Approximate Problem:

$$OP_N: \quad z_N^* = \max_{x \in X} \frac{1}{N} \sum_{i=1}^{N} Ef\left(x,\omega^i\right), \qquad x_N^* = \operatorname{argmax}_{x \in X} \frac{1}{N} \sum_{i=1}^{N} Ef\left(x,\omega^i\right)$$

$\omega^i: i = 1, \ldots, N,$ are independent and identically distributed from the distribution of ω

By virtue of sampling, OP_N with N number of scenarios can be solved using L-shaped method (all iterations proceed with the same sample drawn at the beginning of the solution approach) in order to estimate the expected value of second-stage costs, and optimality and feasibility cuts at each iteration (Linderoth et al., 2006; Kim et al., 2011; Bayraksan and Norton, 2009; Kaut and Wallace, 2003; Kleywegt et al., 2001; Infanger, 1997; Infanger, 1992; Mak et al.,1999; Shapiro and Homem-de-Mello, 1998). The procedure follows the same steps as the original L-shaped method, except from that the true expected second-stage costs, and optimality and feasibility cuts are approximated over a random sample S of size N instead of Ω (Dantzig and Infanger, 2005). Actually, the scenario set Ω and occurrence

[9] It is assumed that the first and second moments of $f(x,\omega)$ exist for all $x \in X$ (Mak et al., 1999).

probability of scenario s (i.e. p_s) interchange with the sample S of *size* N and $\frac{1}{N}$, respectively in the second and fourth steps of L-shaped method given in Section 3.4.

Approximate problems of the form OP_N with different samples S of size N generated randomly (a.k.a. Monte Carlo sampling) are used to determine a candidate solution $\hat{x}$ for OP and to evaluate the quality of this solution by the way of *bounding the optimality gap* expressed as $Ef(\hat{x}, \omega) - z^*$ (the difference between the objective value of candidate solution and optimal solution of the original problem). Now, we give the details of the process of finding a candidate solution and of bounding the optimality gap that is defined as the difference between an upper bound and a lower bound, known as Monte Carlo bounds (Mak et al., 1999).

Step 1: Monte Carlo sampling

M independent sample sets of scenarios, each of which has a size of N, i.e. $\omega_j^1, \omega_j^2, \ldots, \omega_j^N$ (i.i.d. from the distribution of ω in a general case or the scenario set Ω in discrete case) for $j = 1, 2, \ldots, M$ (M number of batches) are randomly generated.

Step 2: Solving each approximate problem in the form of OP_N and estimating an upper bound (UB) for the optimal solution z^*

i) For each j, an optimal solution is found *using multi-cut version of L-shaped method*, given in Section 3.4:

$$z_N^j = \max_{x \in X} \frac{1}{N} \sum_{i=1}^{N} Ef\left(x, \omega_j^i\right)$$

Optimal first-stage solutions for $j = 1, 2, \ldots, M$: $\hat{x}_N^1, \hat{x}_N^2, \ldots, \hat{x}_N^M$.

ii) The expected value of z_N, i.e. Ez_N^*, greater than or equal to the optimal value z^*,is calculated as follows[10]:

$$\bar{z}_{N,M} = \frac{1}{M}\sum_{j=1}^{M} z_N^j$$

From the central limit theorem:

$$\sqrt{M}\left[\bar{z}_{N,M} - Ez_N^*\right] \overset{11}{\Rightarrow} Normal(0, \sigma_u^2)\ as\ M \to \infty \text{ where } \sigma_u^2 = varz_N^*.$$

This theorem, with $s_u^2(M)$, the standard sample variance estimator of σ_u^2, pave the way for building a confidence interval for Ez_N^*.

$$s_u^2(M) = \frac{1}{M(M-1)}\sum_{j=1}^{M}\left(z_N^j - \bar{z}_{N,M}\right)^2$$

$\bar{z}_{N,M}$ is an unbiased estimator of the expected value of z_N, therefore, $\bar{z}_{N,M}$ provides an upper bound (UB) for the optimal solution, since it is an unbiased estimator of z_N.

Let $s_{\bar{z}_{N,M}}^2(M)$ denote the standard sample variance estimator of σ_u^2 and define:

$$\tilde{\varepsilon}_U = \frac{t_{M-1,\alpha}\ s_u(M)}{\sqrt{M}}$$

This step, also known as batch-means approach (since the mean value of M number of batches are used to calculate the UB), provides an upper bound on the optimal

---

[10] *Proof (Mak et al, 1999, p.49):*

$\omega^i: i = 1, \dots, N,$ are i. i. d from the distribution of ω:

$$z^* = \max_{x\in X} Ef(x,\omega) = \max_{x\in X} E\,\tfrac{1}{N}\textstyle\sum_{i=1}^{N} Ef(x,\omega^i) \le E\max_{x\in X} \tfrac{1}{N}\textstyle\sum_{i=1}^{N} Ef(x,\omega^i) = Ez_N^*; \text{ thus, } z^* \le Ez_N^*.$$

This leads to confidence intervals on a upper bound for z^*.

[11] It stands for "convergence in distribution".

objective function value for the original stochastic (TSP) model, i.e. $\boldsymbol{z}^*$ and to build a confidence interval used for the solution quality.

Step 3: Estimating a Lower Bound for the optimal solution $\boldsymbol{z}^*$

The following procedure is applied to find an estimator for the lower bound on the optimal solution z^*:

i) Find a candidate solution, $\hat{x} \in X$, using the solutions obtained from *M* number of approximate problems in Step 2:

- A larger sample size N' (*evaluation sample*, much larger than and independent of N) is generated to obtain an accurate estimate $z_{N'}^{j}$, of the objective value z_N^j of an optimal solution $\hat{x}_N^j$ of the approximate problem (Kleywegt et al., 2001).

  This strategy of using a larger sample size for finding the best solution among M number of approximate solutions (batches) to determine the candidate solution, $\hat{x}$, is the result of the *monotonicity property* and the fact that given the specific values of decision variables found by solving the approximate problem using N number of scenarios (Step 2), calculating the objective value of the approximate problem with a larger sample size N' requires *less computational effort* .

  The monotonicity property can be explained by the following theorem (Mak et al., 1999):

  $\omega^i : i = 1, \dots, N, N+1$, are i.i.d from the distribution of ω, and z_N^* and z_{N+1}^* is the objective function value of the approximate problem with N and $N+1$, respectively. Then,

$$Ez_{N+1}^* = E\max_{x \in X} \left[\frac{1}{N+1}\sum_{i=1}^{N+1} Ef\left(x, \omega^i\right)\right] = E\max_{x \in X} \left[\frac{1}{N+1}\sum_{i=1}^{N+1} \frac{1}{N} \sum_{j=1, j\neq i}^{N+1} f\left(x, \omega^i\right)\right]$$

$$\leq \frac{1}{N+1} \sum_{i=1}^{N+1} E \max_{x \in X} \frac{1}{N} \sum_{j=1, j \neq i}^{N+1} f(x, \omega^i)$$

$$= E z_N^*$$

This result shows that a better UB and thereby a narrower confidence interval can be obtained by the way of increasing the sample size.

- For each j, calculate

$$z_{N'}^j(\hat{x}_N^j) = \frac{1}{N'} \sum_{i=1}^{N'} Ef(\hat{x}_N^j, \omega_j^i)$$

- Choose $\hat{x}_N^j$, which gives the best solution, i.e. the highest $z_{N'}^j$ value, as the candidate solution, $\hat{x}$.

ii) Since choosing any feasible solution of the first-stage problem will provide a lower bound for the optimal value, z^*,

- a much larger sample size N'' (*final evaluation sample*, larger than and independent of N') is generated and a lower bound (LB) is calculated for the optimal solution z^*:

$$\bar{z}_{N''}(\hat{x}) = \frac{1}{N''} \sum_{i=1}^{N''} Ef(\hat{x}, \omega^i)$$

$Ef(\hat{x}, \omega)$ is the expected profit at the solution $x = \hat{x}$ and it can be estimated using the standard sample mean estimator that is an unbiased estimator of the objective function value of the original problem at the solution $x = \hat{x}$, i.e., $E\bar{z}_N = Ef(\hat{x}, \omega) \leq z^*$, and satisfies the central limit theorem (Mak et al., 1999).

From the central limit theorem:

$\sqrt{N''}\,[\bar{z}_{N''} - Ef(\hat{x}, \omega)] \Rightarrow Normal(0, \sigma_l^2)\ as\ n \to \infty$ where $\sigma_l^2 = var f(\hat{x}, \omega)$.

This theorem, with $s_l^2(N'')$, the standard sample variance estimator of σ_l^2, pave the way for building a confidence interval for the objective function value at $x = \hat{x}$.

$$s_l^2(N'') = \frac{1}{N''(N''-1)} \sum_{i=1}^{N''} \left(Ef\left(\hat{x}, \omega^i\right) - \bar{z}_{N''} \right)^2$$

Let $s_l^2(N'')$ denote the standard sample variance estimator of σ_l^2 and define:

$$\tilde{\varepsilon}_l = \frac{t_{N''-1,\alpha}\, s_l(N'')}{\sqrt{N''}}$$

<u>Step 4: Estimating the Optimality Gap</u>

The optimality gap is constructed based on the following inequality (Mak et al., 1999):

$$\begin{aligned} P\left\{\bar{z}_{N''} - \tilde{\varepsilon}_l \le Ef(\hat{x},\omega) \le z^* \le Ez_N^* \le \bar{z}_{N,M} + \tilde{\varepsilon}_u\right\} \\ \ge 1 - P\{\bar{z}_{N''} - \tilde{\varepsilon}_l \le Ez_N^*\} \\ - P\left\{\bar{z}_{N,M} + \tilde{\varepsilon}_u \le Ef(\hat{x},\omega)\right\} \\ \approx 1 - 2\alpha. \end{aligned}$$

Thus, $\left[0,\ \bar{z}_{N,M} - \bar{z}_{N''}(\hat{x}) + \tilde{\varepsilon}_l + \tilde{\varepsilon}_u\right]$ is an approximate 100(1-2α)% confidence interval for the optimality gap at $\hat{x}$. Because of sampling error, it is possible to come across $\bar{z}_{N,M} < \bar{z}_{N''}$. Therefore, the following more conservative confidence interval is recommended by Mak et al. (1999):

$$\left[0, \max\left\{\bar{z}_{N,M} - \bar{z}_{N''}(\hat{x}),\ 0\right\} + \tilde{\varepsilon}_l + \tilde{\varepsilon}_u\right].$$

This optimality gap estimator can be separated into its components as follows (Kleywegt, 2001):

$$\bar{z}_{N,M} - \bar{z}_{N''}(\hat{x}) + \tilde{\varepsilon}_l + \tilde{\varepsilon}_u = \underbrace{\left(\bar{z}_{N,M} - z^*\right)}_{1} + \underbrace{\left(z^* - z(\hat{x})\right)}_{2} + \underbrace{\left(z(\hat{x}) - \bar{z}_{N''}(\hat{x})\right)}_{3} + \underbrace{\tilde{\varepsilon}_l + \tilde{\varepsilon}_u}_{4}$$

Expression # 1: **bias term** that has positive expected value decreasing in the sample N. It is the difference between the unbiased estimator of the expected value of z_N ($\bar{z}_{N,M}$) and the optimal objective function value of the original problem (z^*).

Expression # 2: **true optimality gap** that is the difference between the optimal objective function value of the original problem (z^*) and the optimal objective function value given by the candidate solution, $\hat{x}$.

Expression # 3: the term that is the difference between the optimal objective function value given by the candidate solution, $\hat{x}$ and the unbiased estimator of it, and its expected value is zero.

Expression # 4: **accuracy term** that decreases in *M* number of replications and the sample size N''.

The gap estimator may be large on the condition that the number of batches (M), the size of the sample used for solving the approximate problem (N) or the size of final evaluation sample (N'') is small, even if $\hat{x}$ is an optimal solution, i.e. $z^* - z(\hat{x}) = 0$ (*Ibid.*). Because of that reason, if the confidence interval is not satisfied sufficiently for any setting of M, N and N'', the recommended strategy is to increase M, N and N''.

<u>Step 5: Evaluating the optimality gap estimator</u>

i) **If** the width of the confidence interval (WCI) as a percentage of lower bound estimate, $\bar{z}_{N''}(\hat{x})$, the solution found by Step 1-2, is less than or equal to 2 %, **STOP**.

$ii)$ **Else** increase the number of batches M from the initial value to a new value (e.g. from 30 to 40)[12]. **Go to Step 1.**

$iii)$ **If** $WCI \leq 2\ \%\ of\ lower\ bound\ estimate$, **STOP.**

$iv)$ **Else** increase N and N' (e.g. increase *N* from 500 to 1000 and $N'$from 1000 to 2000). **Go to Step 1.**

$v)$ **If** $WCI \leq 2\ \%\ of\ lower\ bound\ estimate$, **STOP.**

$vi)$ **Else** apply Step 4 (ii)- (v) till total solution time or M reaches to a maximum level (e.g. 12 hours or $M = 60$).

In the next chapter, different cases are developed to test the effect of some deterministic and random parameters on certain performance measures given in Section 4.2. The solution procedure presented in this section is implemented to solve all those cases based on the following settings:

- The initial value of $M = 30$ is increased firstly to 40, then to 50 and to 60 finally.
- The initial value of N=500 and N'=1000 is increased to 1000 and 2000, respectively. After this increment is performed for the first time within the solution process, if an acceptable solution can not be obtained and it is required to increase M in the following stage, N and N' are increased to 1500 and 3000, respectively, and then for all following iterations wherein M is increased to any value, those new values are used.
- The size of sample generated for final evaluation is determined as $N'' = 5000$.

[12] *M* approximate problems are solved on *5 parallel processors*, if the WCI is not satisfied and requires more than 30 replications. Thus, $\frac{M}{5}$ number of problems are solved on each processor simultaneously and thereby the solution time decreases dramatically.

- Stopping condition for the width of the confidence interval and time limit is determined as $\frac{WCI}{\bar{z}_{N''}} \times 100 \leq 2$ and 12 hours, respectively.
- An entire new set of scenarios is generated for each Monte Carlo replication as well as for evaluation sampling process to avoid "the risk of persisting with a bad set of samples" (Bayraksan and Morton, 2011, p.898).

A summarised flow diagram for the solution approach is displayed in Figure 11.

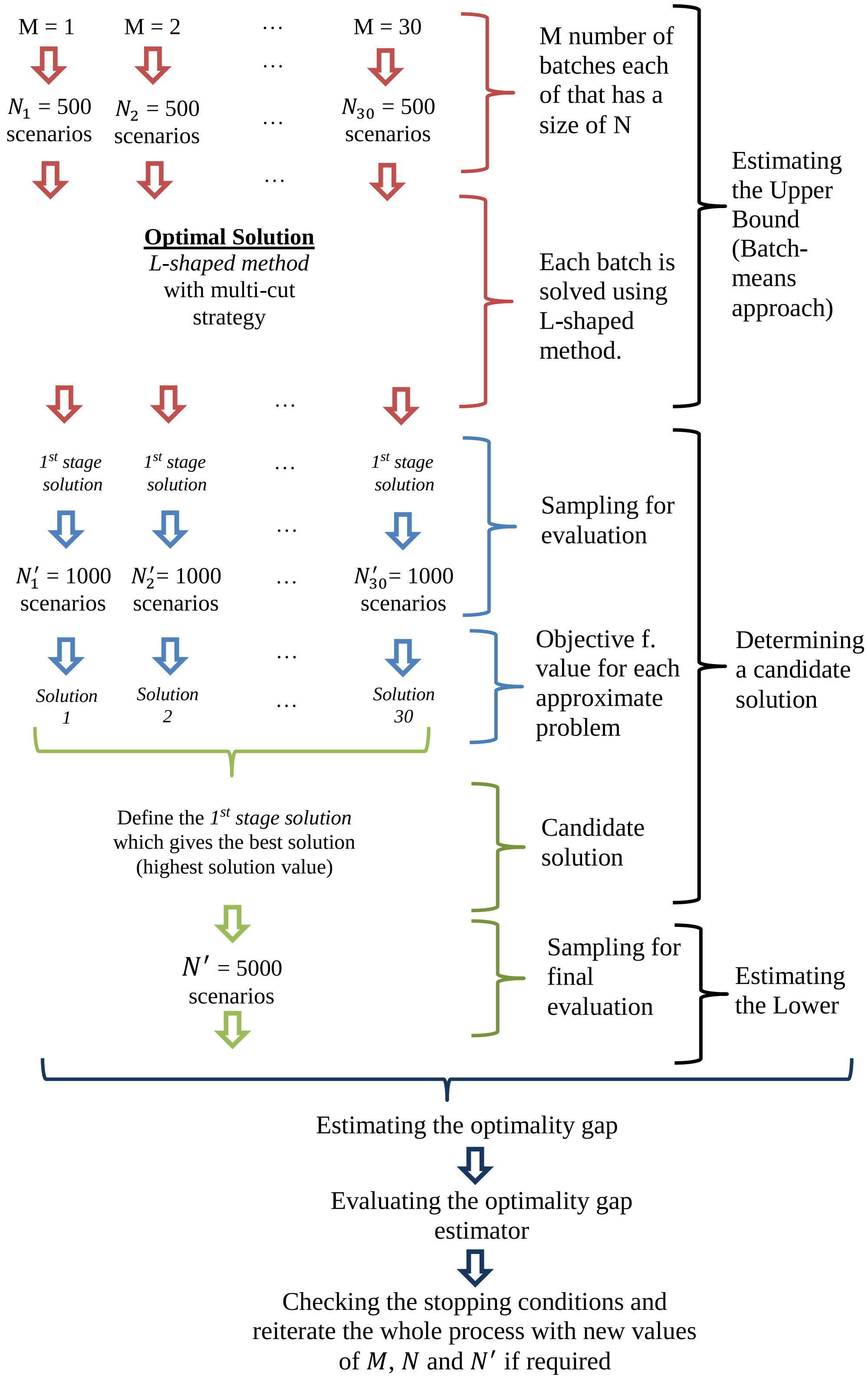


Figure 11. A summary of the sample-average approximation approach

4.2. Performance Measures for Stochastic Solution

In this section, two main measures are defined in order to test both the capability of the solution approach given in the previous section and the quality of solution obtained through the stochastic programming: (1) The value of stochastic solution and (2) The expected Value of perfect information.

The value of stochastic solution (VSS):

It is a measure used to compare the solution obtained by the stochastic model and the expected objective function value calculated based on a deterministic model in which the values of all random parameters are fixed to their mean values. VSS indicates the gain obtained if stochastic programming is used, in other words the cost of ignoring uncertainty of parameters while choosing a decision (Birge and Louveaux, 2011). The larger VSS means that the better results are obtained through stochastic solution, and thereby motivates the decision maker to use the stochastic programming model.

VSS is calculated as follows (*Ibid.*):

- Consider the stochastic problem in the form of OP: $\max_{x \in X} Ef(x, \omega)$ (see Section 4.1.) and let the objective function value of the solution obtained through this model or sample average approximation approach be SP.

- Replace all random variables by their mean values and solve the expected value problem, EV:

$$EV = \max_{x \in X} f(x, \overline{\omega}), \text{ where } \overline{\omega} = E(\omega).$$

Let the expected value solution be $\bar{x}(\overline{\omega})$, as the optimal solution to EV.

- Calculate the expected result of using the optimal solution to EV, i.e. EEV, by considering the uncertainties, i.e.

$$EEV = E_{\omega}\big(f(\bar{x}(\overline{\omega}), \omega)\big)$$

- The value of stochastic solution, VSS,

$$VSS = SP - EEV.$$

While assessing VSS in the computational study, $VSS\%$ is used, which is expressed as a percentage of SP, i.e. $VSS\% = (VSS/SP) \times 100$.

The expected value of perfect information (EVPI):

It is a measure that denotes the difference between the expected value of the stochastic solution obtained by solving the stochastic model for each scenario and the solution obtained by optimizing the stochastic model over all scenarios.

EVPI indicates how much the decision maker would be ready to pay to obtain perfect information about uncertain parameters (Birge and Louveaux, 2011), in other words how much it is worth to invest in better forecasting technology. If EVPI is smaller, it can be said that digging more (perfect) information about the future does not contribute to the expected net profit in our problem context. Moreover, if EVPI is zero, it can be stated that first stage decisions are independent of the realization random parameters.

EVPI is calculated as follows (Birge and Louveaux, 2011):

- Consider the stochastic problem in the form of OP: $\max_{x \in X} Ef(x, \omega)$ (see Section 4.1) and let the objective function value of the solution obtained through this model or sample average approximation approach be SP.

- Solve the stochastic problem for each realisation of random parameters (in our solution approach, $N'' = 5000$ scenarios) and let $\bar{x}(\omega)$ be optimal solutions to those problems.

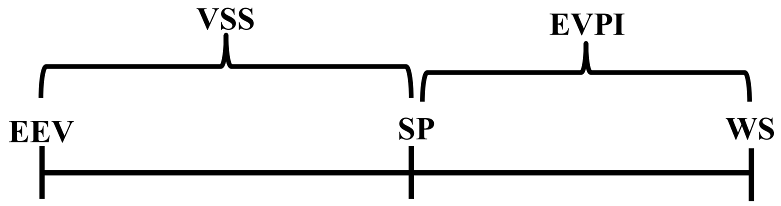
- Calculate the mean of all objective function values obtained for each stochastic problem, a.k.a. wait-and-see-solution, i.e.

$$WS = E_{\omega}\left[\max_{x \in X} f(x, \omega)\right] = E_{\omega}\big(f(\bar{x}(\omega), \omega)\big).$$

- The expected value of perfect information, EVPI,

$$EVPI = WS - SP.$$

While assessing $EVPI$ in the computational study, $EVPI\%$ is used, which is expressed as a percentage of SP, i.e. $EVPI\% = (EVPI/SP) \times 100$. The relationship among VSS, EVPI, EEV, SP and WS can be displayed as follows:



Furthermore, the solution capacity of the model and the solution approach developed is also tested in terms of solution time and the results are given in the next chapter.

CHAPTER 5

COMPUTATIONAL ANALYSIS

In this chapter, the results of a computational study based on nine illustrative cases are presented. The main objectives of this study can be summarised as follows:

- to analyse the effect of deterministic parameters (*capacity expansion cost, variable unit production cost as a percentage of price, unit capacity usage and maximum amount of capacity available in each period*) on stochastic solution performance in terms of the Value of Stochastic Solution (VSS) and the Expected Value of Perfect Information (EVPI),
- to analyse how and to what extent the random parameters (*demand, cost, price and cannibalisation rate*) affect the solution performance, and
- to see the solution capacity of the model and the solution approach developed in terms of solution time.

Six cases (Set 1) in which some of the data are collected through firms' web sites and asking to the firms to which we have interviewed, and the rest is estimated based on the collected data and three cases (Set 2) in which the value of parameters are assigned mostly randomly but in accordance with real life in order to test the solution time are developed for the computational analysis.

This chapter proceeds in two main sections each of which is organised based on the experiments regarding Set 1 and Set 2. Each section including a separate subsection for each case starts with the definition of cases, continues with the related computational analysis and general remarks based on experimental studies, and ends up with a summary of the results obtained for all cases *(see Table 25-28 and Table 31*) and some general inferences based on this summary. The analysis of the cases in Set 1 and Set 2 is presented in Section 5.1 and 5.2, respectively.

5.1. Illustrative Cases – Set 1

In this section six cases are presented in sequence; first the data related to each case and then the results based on two experimental design studies are given: one is based on deterministic parameters (factors in Experiment 1) and the other on uncertain parameters (factors in Experiment 2).

The cases and factors handled in this section are developed considering a crossed array design (Montgomery, 2009, p.488). In each experiment, seven factors are studied to determine their effect on VSS and EVPI. Number of new lines, number of markets and number of periods are the common factors studied in both experiments. In addition to those, deterministic parameters, i.e. capacity expansion cost, variable unit production cost, unit capacity usage and maximum amount of capacity available in each period, are handled in the first experiment, and variability of price, cost, demand and cannibalisation rate are considered in the second experiment. The settings for each factor are defined as follows:

<u>Factors and Their Levels Used in the Experiments</u>

Factor 1 (A): Number of new lines

Two levels of this factor are considered: Level-1: one new line and Level-2: two new lines.

Factor 2 (B): Number of markets

Two aggregated markets, (1) national market in which the firm is located and (2) global market, and two levels for this factor, Level-1: single market and Level-2: two markets, are considered in the experiment.

Factor 3 (C): Number of periods (the length of planning horizon)

This factor is handled in two levels: (1) three-periods and (2) five-periods.

It should be noted that only two existing lines are taken into account, which is considered as a fixed factor for the sake of simplicity while generating the cases. All other factors such as entry price of a new line which is planned to introduce to a market, expected cannibalisation effect of a new line on the demand of existing lines, price elasticity of demand for all products, the characteristics of new lines, market share of existing products, the effect of a new line on market expansion, the relative price and demand of products compared to other products etc. are determined in line with real-like environment considered for each case.

The design for those common factors is a 2^{3-1} fractional factorial design with generator $C = -AB$ (this can be called as the inner array design), which is shown in Table 3.

Table 3. The inner array design, $\mathbf{2^{3-1}}$

Cases	**Factors**		
	A (Number of new lines) - : 1 line, + : 2 lines	B (Number of markets) - : 1 market, + : 2 markets	C (Number of periods) - : 3, + : 5 periods
Case 1	-	-	-
Case 2	-	+	+
Case 3	+	-	+
Case 4	+	+	-

Four deterministic parameters for Experiment 1 are considered, as presented below.

Factor 4 (D): Capacity expansion cost

Three levels for this factor are determined. For each case in Table 1, in order to get a general idea about the reasonable value of capacity expansion cost and to ensure a possible trade-off between capacity expansion cost and unit contribution margin in accordance with the proposed optimisation model, firstly the expected contribution margin including all products and all periods is calculated based on 50 000 scenarios generated randomly. As a result, a value is obtained, say ECM, and this value is assumed as Level-1 (Mid) in the experimental design. A high value and a low value of the capacity expansion cost, given below, are studied in order to see the effect of this parameter on performance measures.

Level-1 (Mid)	:	ECM
Level-2 (High)	:	ECM×2
Level-3 (Low)	:	ECM/2

Factor 5 (E): Variable unit production cost as a percentage of price (profitability)

This factor can also be evaluated as the *profitability* of product, which is specified using the unit production cost as a percentage of price. The levels of this factor are determined as follows:

Level-1	:	New product line(s) is(are) LESS PROFITABLE than the existing lines
Level-2	:	New product line(s) is(are) MORE PROFITABLE than the existing lines
Level-3	:	All existing and new lines have the SAME PROFITABILITY.

Factor 6 (F): Unit capacity usage

Five levels for the experimental study are used:

Level-1	:	Same for all products
Level-2	:	New line consumes less resource
Level-3	:	New line consumes much less resource
Level-4	:	New line consumes more resource
Level-5	:	New line consumes much more resource

Factor 7 (G): Maximum amount of capacity available in each period

Four levels are considered for this factor. For each case shown in Table 1, in order to get a general idea about the reasonable value of maximum amount of capacity available for each period, firstly the total expected demand of all products over all periods under 50 000 scenarios is calculated. Then four different levels are determined for this parameter, as a percentage of total expected demand calculated:

Level-1	:	5 % of total expected demand calculated (very limited capacity)
Level-2	:	25 % of total expected demand calculated (limited capacity)
Level-3	:	50 % of total expected demand calculated (abundant capacity)
Level-4	:	100 % of total expected demand calculated (overabundant capacity)

The design for the factors D-G can be called as the outer array design in which total number of runs performed is 180[13]. Each run in the outer design is performed for all cases in the inner array, by this way the crossed array structure for Experiment 1 is generated (see Table 4).

Table 4. The crossed array design for Experiment 1 with seven factors

Cases	**Factors**						
	A	B	C	D	E	F	G
Case 1	-	-	-	180 runs obtained by combining all levels of those factors			
Case 2	-	+	+	180 runs obtained by combining all levels of those factors			
Case 3	+	-	+	180 runs obtained by combining all levels of those factors			
Case 4	+	+	-	180 runs obtained by combining all levels of those factors			

In order to explore which and to what extent the uncertain parameters are significant on the solution performance in terms of VSS and EVPI, the levels given below are defined for Experiment 2. Here, the variability regarding with random parameters is expressed by the coefficient of variation (CV), the ratio of the standard deviation to the mean value. For this experimental design study, the levels of CV for each parameter are defined as 0 (this means there is no variability regarding all parameters considered as uncertain), 0.15 (this means the standard deviation is 15 % of the mean) and 0.30. Since the variability of demand and price for new products

[13] $Total\ \#\ of\ runs = (\#\ of\ levels\ of\ D) \times (\#\ of\ levels\ of\ E) \times (\#\ of\ levels\ of\ F) \times (\#\ of\ levels\ of\ factor\ G) = 3 \times 3 \times 5 \times 4 = 180$

and new markets is higher than for the old products and the old markets, it is assumed that these levels are higher for the new products and new markets. The random factors and base levels are given as follows:

Factor 8 (H): Coefficient of variation of price (CV_Price)

Level-1 (Low)	:	0
Level-1 (Mid)	:	0.15
Level-2 (High)	:	0.30

Factor 9 (I): Coefficient of variation of demand (CV_Demand)

Level-1 (Low)	:	0
Level-1 (Mid)	:	0.15
Level-2 (High)	:	0.30

Factor 10 (J): Unit Production Cost (CV_Cost)

Level-1 (Low)	:	0
Level-1 (Mid)	:	0.15
Level-2 (High)	:	0.30

Factor 11 (K): Cannibalisation Rate (CV_CanR)

Level-1 (Low)	:	0
Level-1 (Mid)	:	0.15
Level-2 (High)	:	0.30

It should be noted that for this experiment the levels of deterministic parameters (i.e. unit capacity usage, capacity expansion cost, profitability and capacity) are fixed to one of the values where VSS takes higher values in Experiment 1.

The experiment for the factors H-K can be called as the outer array design in which total number of runs performed is 81[14]. Each run in the outer design is performed

---

[14]

$$Total\ of\ runs = (\#\ of\ levels\ of\ K) \times (\#\ of\ levels\ of\ I) \times (\#\ of\ levels\ of\ J) \times (\#\ of\ levels\ of\ factor\ K) = 3^4 = 81$$

for all cases in the inner array, by this way the crossed array structure for Experiment 2, given in Table 5, is generated:

Table 5. The crossed array design for Experiment 2 with seven factors

Cases	**Factors**						
	A	B	C	H	I	J	K
Case 1	-	-	-	81 runs obtained by combining all levels of those factors			
Case 2	-	+	+	81 runs obtained by combining all levels of those factors			
Case 3	+	-	+	81 runs obtained by combining all levels of those factors			
Case 4	+	+	-	81 runs obtained by combining all levels of those factors			

Since total number of runs is high which increases the workload, it is decided to use a D-optimal design with less number of runs for each case requiring $3^4 = 81$ runs in total. The main factors and some interactions that are expected as significant on performance measures, i.e. the interaction between the variability of demand and price, the variability of price and cost, and the variability of demand and cannibalisation rate, are considered for that optimal design. As a result, the D-optimal design with 30 runs is generated using MINITAB (see Table 6).

Table 6. The D-optimal design for Experiment 2 with four factors

Run	Factors			
	CV_Demand	**CV_Price**	**CV_Cost**	**CV_CanR**
1	0.15	0.15	0.15	0.15
2	0.15	0.15	0.15	0.30
3	0.15	0.15	0.30	0.15
4	0.15	0.15	0.30	0.30
5	0.15	0.30	0.15	0.15
6	0.15	0.30	0.15	0.30
7	0.15	0.30	0.30	0.15
8	0.15	0.30	0.30	0.30
9	0.30	0.15	0.15	0.15
10	0.30	0.15	0.15	0.30
11	0.30	0.15	0.30	0.15
12	0.30	0.15	0.30	0.30
13	0.30	0.30	0.15	0.15
14	0.30	0.30	0.15	0.30
15	0.30	0.30	0.30	0.15
16	0.30	0.30	0.30	0.30
17	0	0.30	0	0.30
18	0	0	0.30	0.30
19	0	0	0	0
20	0	0.15	0	0.15
21	0.15	0	0.30	0
22	0.30	0.30	0	0
23	0	0	0.15	0.15
24	0.15	0	0	0.15
25	0.30	0.15	0	0
26	0	0.15	0.15	0.30
27	0	0.30	0.15	0.15
28	0	0	0	0.15
29	0.15	0.15	0	0
30	0.30	0	0.15	0.15

As seen from Table 4-6, four cases are developed based on fractional factorial design and in total 720 problems (runs) for Experiment 1 and 120 problems (runs) for Experiment 2 are solved. In addition to Case 1-4, two more cases as an extension of Case 1 and Case 2 with different settings are also studied (overall 1270 runs are studied). Since we have found that a problem which is larger than the problems in

Set 1 can be solved in a reasonable time with a very large number of scenarios, i.e. 10000, those problems are solved by their extensive forms using GAMS 22.2/CPLEX 10.0 on a computer with 16 GB of RAM and 2.0 GHz speed. On the other hand, the larger-sized problems in Set 2 are solved using L-shaped method in order to guarantee the optimal solution for the approximated problems by GAMS 22.2/CPLEX 10.0 on a computer with 4 processors each of which has 14 processing units (those problems are solved using parallel processors), 2.2. GHz speed and 16 GB of RAM based on the solution methodology mentioned in Section 4.1.

The Analysis of Data Collected Through the Experiments

The analysis of data collected through those experimental studies are performed using two methodologies: (1) Analysis of Variance (ANOVA) to reveal which factors and their interactions are significant (Montgomery, 2009) and (2) Association Analysis to extract rules (frequent variable interactions) based on the predictions made from Random Forest (RF) method (Tan et al., 2006).

The RF is an ensemble learning method developed for decision tree classifiers (*Ibid.*), in which each successive tree is generated independently using bootstrap sampling from the data set and then a majority voting procedure is applied in order to make predictions (Mashayekhi and Gras, 2015). For the problem handled in this study, firstly the package *randomForest*[15] developed in R, which implements Breiman's random forest algorithm for classification, is used to generate a tree ensemble (i.e. forest). Then, the package *inTrees*[16] developed in R is used to extract and summarise classification rules from the generated tree ensemble. Here, a rule refers to an expression of the form $A \rightarrow B, A \cap B = \emptyset$ and indicates the co-occurrence of A and B, not a causal relationship between them.

---

[15] https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

[16] https://cran.r-project.org/web/packages/inTrees/inTrees.pdf

The strength of a rule is generally specified by its *support*[17] which determines the frequency of a rule in a given data set and *confidence*[18] which determines how frequently item in B appears to transactions containing A (Tan et al., 2006, p.329 - 330). Since a rule having high confidence may occur by chance, this kind of misleading rule is detected by determining whether the left-hand-side and right-hand-side of the rule are statistically independent, i.e. by calculating *lift*[19] value (Azevedo and Jorge, 2007). If the lift takes a value around 1, it can be said that the rule is not interesting. Actually, those kinds of measures are called as interestingness measures in data mining literature and support, confidence and lift are commonly used ones in order to extract interesting rules in a data. Within the context of this study, an exemplar of a rule can be given as follows:

$$A\text{: }\{e\text{xpansion cost is high and capacity available is abundant}\} \rightarrow B\text{: }VSS \;\geq 20\ \%$$

That means that if the capacity expansion cost is high and the capacity available at each period is abundant compared to the total expected demand, the VSS takes a value greater than 20 %. At this point, the strength of this rule in terms of support, confidence and lift should be considered and the resultant rule should be accepted if the support and confidence exceed predetermined thresholds, i.e. *minsup* and *minconf*, and the *lift is far from 1*.

The procedure for extracting rules for each problem instance can be given as follows:

Step 1: The performance measures, VSS and EVPI, are transformed into three different categories as shown below:

---

[17] Support is defined as the fraction of transactions involving both A and B, i.e. $support(A \rightarrow B) = number\ of\ transactions\ (A + B)/total\ number\ of\ transactions$

[18] $confidence\ (A \rightarrow B) = \ support(A \rightarrow B)/support(A)$

[19] $lift\ (A \rightarrow B) = \ confidence(A \rightarrow B)/support(B)$

VSS / EVPI (%)	Category
$[0\,;10)$	Low (L)
$[10\,;20)$	Medium (M)
≥ 20	High (H)

It should be noted that those categories, their limits and meanings may be changed according to the problem context as well as the sector in which the firm operates and market conditions. For instance, while a VSS smaller than 10 % may be considered as low for a firm, a VSS greater than 10 % may be considered as high instead of medium. Furthermore, it is also possible to define more than or less than three categories. However, since the number and the limits of the categories do not affect the procedure applied, in this thesis work, those categories are arbitrarily defined in order to be able to apply the rule-extraction procedure.

Step 2: Decide on the settings for the algorithm. For this study $minsup = 0.05$[20] and $confidence = 0.75$[21], $mtry = \log_2[number\ of\ variables] + 1$[22] and $ntree = 1000$[23] are applied (see Appendix C). Though higher values for the measures cared, i.e. support, confidence and lift, indicate better rules, the $minsup$ is considered low in order to get the rules of low support but high confidence, which may be considered as *exception rules*. A rule that has a high predictive power (i.e. high confidence) and strong relationship between the condition and consequent (i.e. high lift) but occurs rarely (i.e. low support) is called as exception rule deriving noteworthy information about patterns in the data (Wang, 2018). Since the problems handled include small number of data, it is expected to have the rules that may not occur very frequently, however, the strength of the relationship, which refers to have a high confidence and lift, is more cared in this study. Besides, after

[20] A user-specified threshold value for support. It means that a rule is interesting if its support is greater than or equal to 5 %.
[21] A user-specified threshold value for confidence. It means that a rule is interesting if its confidence is greater than or equal to 75 %.
[22] Number of variables randomly sampled used in each tree.
[23] Number of trees used in the forest.

an exception rule is generated through the Random Forest algorithm, it is evaluated based on the common sense about the relationship between the condition and the result, which is gained from ANOVA and then the rule is accepted.

Furthermore, for the association analysis in this study, a rule is extracted from the data if its support and confidence are greater than the threshold values, i.e. *minsup* and *minconf*. However, how to specify the *minsup* or *minconf* values to obtain interesting rules becomes a problem at this point. If the threshold values chosen are increased, the number of rules extracted will be lower; but the more distinctive patterns will be obtained. In this study, we have some generalised information about the patterns in the data through ANOVA results before applying the association analysis. Considering this information, and the number of the extracted rules and their meaningfulness at different threshold values, i.e. via a kind of trial-and-error process, we have defined the *minsup* and *minconf* as defined above.

Step 3: Apply the Random Forest algorithm (package *randomForest* in R) to generate the forest, check out-of-bag (OOB) error which shows the accuracy of the model and the variable importance across the whole forest, considering Mean Decrease Accuracy and Mean Decrease Gini (see Appendix C).

Step 4: Obtain the rules extracted from the forest generated using the package *inTrees* in R.

The source code for this procedure in R is given in Appendix C.

<u>Scenario Generation</u>

For all cases, it is possible to determine a discrete set of unstructured scenarios that can be defined based on the historical data regarding the uncertain parameters or subjective evaluations of decision-makers or a data preparation team involving people from marketing, finance and production departments. If there is enough data, it is also possible to estimate the probability distributions (discrete or continuous) of uncertain parameters. In the computational study, it is assumed that all uncertain parameters are normally distributed. It should be noted that different types of probability distributions can also be used, but the normal distribution is chosen in this study because of its commonality. Since there are infinite number of scenarios in the case of having a continuous probability distribution such as normal distribution, a sample set of scenarios is generated for each step of Sample Average Approximation. The following procedure is defined to generate those sample sets:

Step 1: The mean value of the selling price for each product in each market at each period is defined and the standard deviation is calculated as a percentage of the mean value. Then, a specific value for the selling price for each product in each market at each period is randomly generated from the normal distributions with those parameters and regarding probability of occurrences are calculated.

$$P_{ijt} \; : \; \text{selling price of product } i \in I \text{ at period } t \; ; \; P_{ijt} \sim \mathrm{N}\,(\mu_{P_{ijt}}, \, \sigma^2_{P_{ijt}})$$

Assume that a value for the selling price of each product in each market at each period, say $\tilde{P}_{ijt}$, is generated.

Step 2: Using the values for the selling prices, which are randomly generated from the regarding normal distributions, and the functions representing the demand-price relationships, the mean value of the demand for each product in each market at each period is defined and the standard deviation is calculated as a percentage of the mean value.

EPD_{ijt} : potential demand of product $i \in EP$ in market $j \in M_i$ at period t;

$$EPD_{ijt} \sim \mathrm{N}\,(\mu_{EPD_{ijt}}, \, \sigma^2_{\varepsilon_{ijt}})$$

$$\text{where } \; \mu_{EPD_{ijt}} = m_{ijt} - n_{ij} \times \tilde{P}_{ijt} + \textstyle\sum_{k \in R_{ij} \cap EP} n_{ikj} \times \tilde{P}_{kjt}$$

$$EPD_{ijt} = \mu_{EPD_{ijt}} + \varepsilon_{ijt} \, , \quad \varepsilon_{ijt} \sim \mathrm{N}\,(0, \, \sigma^2_{\varepsilon_{ijt}})$$

Then, a specific value for the demand for each product in each market at each period is randomly generated from the normal distributions, given above, with those parameters and regarding probability of occurrences are calculated.

Step 3: Using the values for the selling prices, which are randomly generated from the regarding normal distributions, and the functions representing the cannibalisation rate-price relationships, the mean value of the cannibalisation rates is defined and the standard deviation is calculated as a percentage of the mean value.

e_{ijlt} : Cannibalization rate (as a percentage) of product $l \in NP$ on product $i \in EP$ in market $j \in M_i$ at period t

$$e_{ijlt} \sim \text{N}\,(\mu_{e_{ijlt}},\, \sigma^2_{\delta_{ijlt}}) \text{ where } \mu_{e_{ijlt}} = \alpha_{ijl} - \beta_{ijl} \times \tilde{P}_{ljt}$$

$$e_{ijlt} = \mu_{e_{ijlt}} + \delta_{ijlt}\,, \quad \delta_{ijlt} \sim \text{N}\,(0,\, \sigma^2_{\delta_{ijlt}})$$

Then, a specific value for each cannibalisation rate is randomly generated from the normal distribution, given above, with those parameters and regarding probability of occurrences are calculated.

Step 4: The mean value of the variable production cost for each product at each period is defined and the standard deviation is calculated as a percentage of the mean value.

V_{it}: variable unit production cost of product $i \in I$ at period t; $V_{it} \sim$ N $(\mu_{V_{it}},\, \sigma^2_{V_{it}})$

Then, a specific value for the variable production cost for each product at each period is randomly generated from the normal distributions with those parameters and regarding probability of occurrences are calculated.

Step 5: The values for each parameters determined in Step 1-4 are specified as a scenario and the regarding probability of occurrence for this scenario is calculated by multiplying the probability of occurences found in Step 1-4. All those steps are repeated N times, and thus, a discrete set of scenarios with a size of N is obtained and used in the Sample-Average-Approximation procedure.

It should be noted that one random number is generated for each demand parameter mentioned in Step 2 and for each cannibalisation rate parameter mentioned in Step

3 at a specific selling price level that is generated randomly from the regarding distributions. As another scenario generation strategy, in order to have well-representative samples it is also possible to generate more than one random number for each demand and cannibalisation rate at the same selling price level, since both the demand and the cannibalisation rate parameters have some randomness at any selling price level. However, this increases the number of scenarios (the sample size) dramatically. For instance, assume that 500 scenarios are obtained by generating one parameter level for each scenario. If, for instance, at least two levels for the demand and cannibalisation rate parameters are determined by randomly sampling at a specific price level, the number of scenarios is increased to 2000[24] from 500. Therefore, in order to have tractable solutions for the computational study by making the number of scenarios as less as possible, in this study it is not prefeerred generating more than one random number for each demand and cannibalisation rate parameter at a specific selling price level.

ANOVA and Association Analysis (AA) are applied for each case as well as for the crossed array design (considering Case 1-4 in a single-design) for both Experiment 1 and Experiment 2 performed based on the scenario generation procedure given above. In Section 5.1.1, 5.1.2, 5.1.3 and 5.1.4, the results of ANOVA and AA analysis for Case 1, 2, 3 and 4, respectively, are presented. Then the results regarding the crossed array design and general conclusions considering all four cases are given in Section 5.1.5 and Section 5.1.6. Furthermore, Case 5 and Case 6 are also presented in Section 5.1.7 and Section 5.1.8.

---

[24] $$\text{Total number of scenarios} = 500 \text{ replications} \times 1 \text{ value for price} \times 2 \text{ values for demand} \times 2 \text{ values for cannibalization rate} \times 1 \text{ value for cost} = 2000$$

5.1.1. Case 1: Mix of Two Existing Lines with a New Higher-Priced Line and a Single Market

In this case a firm producing major household appliances (white goods) is considered. This firm has two product lines under "freezer" category (defined under "cold family") according to the product hierarchy, given in Figure 6.

- 2 existing (old) lines
- 1 new line
- 3 periods (years)
- 1 market (national)

5.1.1.1. Basic Data

a) Products:

Existing Line 1 : Chest Freezer

Existing Line 2 : Upright Freezer

New Line : Drawer Freezer

Chest Freezer

Upright Freezer

Drawer Freezer

b) Markets:

The products are sold in a single market which is aggregated as the "national market" where the manufacturing facility of the firm is also located.

c) Product Prices:

P_{ijt} : selling price of product $i \in I$ in market $j \in M_i$ at period t

$$P_{ijt} \sim \mathrm{N}\,(\mu_{P_{ijt}},\ \sigma^2_{P_{ijt}})$$

	Mean Price(TL[25]) $j = 1,\ t = 1$	Mean Price (TL) $j = 1,\ t = 2$	Mean Price (TL) $j = 1,\ t = 3$
Existing line-1 ($i = 1$)	$\mu_{P_{111}} = 1765$[26]	$\mu_{P_{112}} = 1765$	$\mu_{P_{113}} = 1765$
Existing line-2 ($i = 2$)	$\mu_{P_{211}} = 1530$	$\mu_{P_{212}} = 1530$	$\mu_{P_{213}} = 1530$
New Line ($i = 3$)	$\mu_{P_{311}} = 2200$	$\mu_{P_{312}} = 2200 \times 0.9 = 1980$	$\mu_{P_{313}} = 1980 \times 0.9 = 1782$

The existing lines are considered as mature products, and thereby it is assumed that the mean price is stable and doesn't change over the planning horizon (i.e. next three years). However, the new product is launched with a higher price which diminishes in the next three periods by a rate of 10 %. At the end of the planning horizon the price of this new product is expected to be nearly the same as of the existing product with higher price, i.e. chest freezer.

The standard deviation of price for existing lines: $\sigma_{P_{ijt}} = 0.30 \times \mu_{P_{ijt}}$

It is also assumed that the uncertainty regarding the price of new line is higher than of the existing lines, and thereby, the standard deviation of the price of the new line is considered as 20 % higher than of the existing lines, i.e.

The standard deviation of price for new line: $\sigma_{P_{ijt}} = 0.36 \times \mu_{P_{ijt}}$

d) Demand:

EPD_{ijt} : potential demand of existing product $i \in EP$ in market $j \in M_i$ at period t

$$EPD_{ijt} \sim \text{N}\ (\mu_{EPD_{ijt}}, \sigma^2_{EPD_{ijt}})$$

ND_{ijt} : demand of new product $i \in NP$ in market $j \in M_i$ at period t

$$ND_{ijt} \sim \text{N}\ (\mu_{ND_{ijt}}, \sigma^2_{ND_{ijt}})$$

---

[25] TL : Turkish Lira

[26] This value is calculated as the average price of different models under the chest freezer line of a firm operating in Turkey, Arçelik A.Ş. Retrieved on 5 July, 2017, from https://www.arcelik.com.tr/derin-dondurucu/.

	Demand (unit)
Existing line-1 ($i = 1,\ j = 1,\ t$)	$\mu_{EPD_{11t}} = 150000 - 36.5 \times P_{11t} + 10 \times P_{21t}$
Existing line-2 ($i = 2,\ j = 1,\ t$)	$\mu_{EPD_{21t}} = 90000 - 37.5 \times P_{21t} + 10 \times P_{11t}$
New Line ($i = 3,\ j = 1,\ t$)	$\mu_{ND_{31t}} = 160000 - 22.5 \times P_{31t}$

The standard deviation of demand for existing lines: $\sigma_{EPD_{ijt}} = 0.30 \times \mu_{EPD_{ijt}}$

It is also assumed that the uncertainty regarding the demand of new line is higher than of the existing lines, and thereby, the standard deviation of the demand of the new line is considered as 20 % higher than of the existing lines, i.e.

The standard deviation of demand for new line: $\sigma_{NP_{ijt}} = 0.36 \times \mu_{ND_{ijt}}$

In the current situation in which the new product is not introduced the sales volume of the chest and upright freezer is 2/3 and 1/3 of total sales, respectively. By introducing the new line, it is expected that the market size (total expected demand) would increase by a rate of 1/3.

Own-price elasticities of demand and cross elasticities don't change over the planning horizon. Also, own-price elasticities of demand are higher for existing (old) products, i.e. more sensitive to price changes.

e) Cannibalisation Rate:

e_{ijlt} : Cannibalisation rate (as a percentage) of product $l \in NP$ on product i in market j at period t

$$e_{ijlt} \sim \mathrm{N}\,(\mu_{e_{ijlt}},\ \sigma^2_{e_{ijlt}})$$

	Cannibalisation rate of new line on existing lines
Existing line-1 ($i = 1, j = 1, l = 3$, t)	$\mu_{e_{113t}} = 0.5 - 0.000045 \times P_{31t}$
Existing line-2 ($i = 2, j = 1, l = 3$, t)	$\mu_{e_{213t}} = 0.5 - 0.000045 \times P_{31t}$

The variance of cannibalisation rate: $\sigma^2_{e_{ijlt}} = 0.30 \times \mu_{e_{ijlt}}$

If the new line is introduced to the market, it is expected that potential demand for both of the existing lines would be cannibalised by the same rate. Considering demand, price and cannibalisation rate are at their mean values, the following figures can be obtained:

	Chest	Upright	Drawer (New Line)	Total
Expected demand before the introduction of the new line	100 000	50 000	-	150 000
Expected demand cannibalised by 40 %	40 000	20 000	-	60 000
Expected demand if the new line is introduced	60 000	30 000	110 000	200 000

f) Variable unit production cost:

The mean value of cost is expressed as a percentage of the mean price. Different percentages are considered for the experimental study given in the next section.

V_{it} : variable unit production cost of product $i \in I$ at period t;

$V_{it} \sim \mathrm{N}\left(\mu_{V_{it}}, \sigma^2_{V_{it}}\right)$

The variance of cost: $\sigma^2_{V_{it}} = 0.30 \times \mu_{V_{it}}$

g) Minimum launching time of the new line:

$$minL_{31} = 1.$$

5.1.1.2. Other Case Data and Experiment 1 for Deterministic Parameters

5.1.1.2.1. Experiment 1: Settings

Four deterministic parameters are studied as factors for the outer array of the experiment given in Table 4. Those are

1. Capacity expansion cost with three-levels
2. Variable unit production cost as a percentage of price with three-levels
3. Unit capacity usage with five-levels

4. Maximum amount of capacity available for each period with four-levels

Factor D: Capacity expansion cost

In order to get an idea about the reasonable value of capacity expansion cost and to provide a possible trade-off between capacity expansion cost and unit contribution margin in accordance with the proposed optimisation model, firstly the total expected demand of all products over all periods under 50 000 scenarios is calculated. As a result, the value of TL 683 is obtained and set as Level-1 (Mid) in the study. A high value and a low value of the capacity expansion cost, given below, are also experimented in order to see the effect of this parameter on performance measures.

Level-1 (Mid)	:	683
Level-2 (High)	:	683×2 = 1366
Level-3 (Low)	:	683/2 = 342

Factor E: Variable unit production cost as a percentage of price (profitability)

This factor can also be evaluated as the *profitability* of products which is specified using the unit production cost as a percentage of price. The levels used in this experiment are determined as follows:

Level-1	:	New product line is LESS PROFITABLE than the existing lines
Level-2	:	New product line is MORE PROFITABLE than the existing lines
Level-3	:	All existing and new lines have the SAME PROFITABILITY.

V_{it} : variable unit production cost of product $i \in I$ at period t

$$V_{it} \sim \mathrm{N}\,(\mu_{V_{it}},\ \sigma^2_{V_{it}})$$

	Chest	Upright	Drawer (New Line)
Level-1	$\mu_{V_{1t}} = 0.6 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$
Level-2	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.6 \times P_{31t}$
Level-3	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.8 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$

Factor F: Unit capacity usage

Five different levels for the experiment are studied:

Level-1 : Same for all products
Level-2 : New line consumes less resource
Level-3 : New line consumes much less resource
Level-4 : New line consumes more resource
Level-5 : New line consumes much more resource

Levels	Unit Capacity Usage		
	Chest	Upright	Drawer (New Line)
Level-1	1 unit	1 unit	1 unit
Level-2	1.5 unit	1.5 unit	1 unit
Level-3	2 unit	2 unit	1 unit
Level-4	1 unit	1 unit	1.5 unit
Level-5	1 unit	1 unit	2 unit

Factor G: Maximum amount of capacity in each period

In order to get an idea about the reasonable value of maximum amount of capacity available for each period, firstly the total expected demand of all products over all periods under 50 000 scenarios are calculated. Then, four different levels for this parameter are determined as a percentage of total expected demand calculated:

Level-1 : 5 % of total expected demand calculated (very tight capacity)
Level-2 : 25 % of total expected demand calculated (tight capacity)
Level-3 : 50 % of total expected demand calculated (loose/abundant capacity)
Level-4 : 100 % of total expected demand calculated (very loose/overabundant capacity)

For each level, the related percent of total expected demand is used for the maximum capacity at t = 2, and for ICAP, t =1 and t =3, 70 %, 85 % and 115 % of that value calculated for t=2 are used, respectively. All the capacity data is shown in the following table:

	ICAP	*t = 1*	*t = 2*	*t = 3*
Level 1	9 300	11 250	13 250	15 250
Level 2	46 500	56 250	66 250	76 250
Level 3	93 000	112 500	132 500	152 500
Level 4	186 000	225 000	265 000	305 000

Total number of runs performed for the experimental study = 3×3×5×4 = 180.

5.1.1.2.2. Experiment 1: Results

<u>Response 1: Value of Stochastic Solution (VSS)</u>

After solving 180 runs, the related data is firstly analysed using MINITAB 17 software to see which factors including interactions among them are significant. This analysis provides to understand the general behaviour of the factors and contributes to evaluation of the interestingness of rules generated by Association Analysis. The results are shown in Appendix C.1.1.1.

Remarks:

- **Main effects**
    - Considering ANOVA Table and related figures given in Appendix C.1.1.1, it can be seen that the factors profitability, unit capacity usage and maximum capacity available at each period are significant whilst capacity expansion cost is insignificant.
- **The most important interactions (based on ANOVA table and related figures in Appendix C.1.1.1)**
    - <u>Profitability & Unit Capacity Usage</u>
        - When the new line is less profitable than the existing lines and unit capacity usage is the same for all products, VSS has

its maximum value. However, when the profitability of the new line increases relatively compared to the existing lines, VSS decreases. This relationship between the profitability and unit capacity usage is also similar to the case of that the new consumes fewer resources compared to the old lines.

    - When the new is more profitable, VSS has its maximum (minimum) value in the case of that the new uses more (less than or equal) capacity compared to the existing lines.
    - When the unit capacity usage and profitability of the new line is less than or equal to of the old lines, VSS takes its higher values compared to the case of that the new is more profitable.

- Profitability & Capacity
    - When the new is more profitable, at all levels of capacity VSS tends to take its lowest value.
    - When the profitability of all products is the same, VSS tends to take its highest values at all levels of capacity.
    - It should be noted that VSS is not considerably different depending on the maximum capacity when the new is less profitable or has the same profitability with the old lines.
- Unit Capacity Usage & Capacity
    - In general, VSS is higher at 5 % capacity level than at other capacity levels.
    - At all levels of capacity (i.e. 5 %, 25 %, 50 % and 100 %), VSS tends to take its highest values when the unit capacity usage of the new product is the same as or more than of the old products. However, if the new product consumes much more resources compared to the old products, VSS decreases.
    - VSS tends to take lower values when the new consumes less production resources than the old products.

- <u>Capacity expansion cost & Capacity</u>
    - When capacity is limited, i.e. 5 % and 25 %, VSS takes its highest value if capacity expansion cost is very high; and decreases as capacity expansion cost decreases.
    - When capacity is abundant, i.e. 50 % and 100 %, VSS takes its lowest value if capacity expansion cost is very high; and increases as capacity expansion cost decreases.

- **VSS over 180 runs**
    - Average VSS (%) = 3.7
    - Min VSS (%) = 0
    - Max VSS (%) = 36.5

- **Association Analysis: Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.1.1.1 indicates that *unit capacity usage*, *maximum capacity* and *profitability* are the significant parameters on VSS, as obtained by ANOVA. Considering the rules extracted given in Appendix C.1.1.1, many of them are eliminated because of having a lift value close to 1. If the rules with minimum-support = 0.10 and a confidence level greater than 0.75 are considered, only one rule is accepted. Other three rules have a high confidence and lift even it has a low support. Those types of rules are called as *exception rules* (Wang, 2008) which have a high predictive power (i.e. high confidence) but occur rarely (i.e. low support), and give noteworthy information about patterns in the data (Ibid.). Therefore, the rules no. 2-4 are also taken into account.

Table 7. Rules extracted for Case 1 (Deterministic parameters: VSS)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (VSS)
					Profitability	**Unit Cap. U.**	**Capacity**	
1	2	0,103	0,819	4,791	1	1	…	M
2	2	0,047	0,865	6,606	1	2	…	H
3	3	0,045	1	7,637	1	2	3 or 4	H
4	3	0,043	1	7,637	3	1	3 or 4	H

According to Table 7:

- When the new product competes for the same amount of production resources with the existing products, but less profitable than the existing ones, the VSS is predicted to take a value between 10-20 % (i.e. “Medium” value, as ANOVA also indicates) and thus it can be said that in this case the decision makers can use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.
- Considering the exception rules, when the new product is less profitable and also consumes less amount of resources, VSS is predicted to take a value greater than 20 %. In addition to this condition, if the capacity is loose/abundant, the confidence as well as lift increases and thus rule no.3 becomes stronger than the rule no.2 are obtained. Besides, when the new product competes for the same amount of production resources, has the same profitability as existing ones and the capacity is loose/abundant, VSS is predicted to take a value greater than 20 %. Those conditions do not contradict with ANOVA results in which VSS becomes maximum. Therefore, under these conditions, the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.
- Besides, lift value for each rule is greater than 1 and this indicates a strong relationship between the left-hand-side and right-hand-side of

the rule and thus it can be said that the chance that the rule occurs by coincidence is lower.

Response 2: Expected Value of Perfect Information (EVPI)

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.1.1.2.

Remarks

- **Main effects**
    - Considering ANOVA Table and related figures in Appendix C.1.1.2, it can be seen that profitability, unit capacity usage, maximum capacity available at each period and capacity expansion cost are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.1.1.2)**
    - Capacity expansion cost & Unit Capacity Usage
        - If the new product consumes the same as or more production resources than the existing products, EVPI is insensitive to capacity expansion cost.
        - When the unit capacity usage of the new product is less than of the existing products, EVPI increases as the capacity expansion cost increases (i.e. less capacity usage and expensive capacity expansion results in higher EVPI).
    - Profitability & Unit Capacity Usage
        - When the new product is more profitable and consumes the same as or less production resources than the existing products, EVPI has its minimum value. However, when the profitability of the new relatively decreases compared to the existing lines, EVPI increases.

        - When the new consumes more production resources, if it is equally or more (less) profitable than the existing products, EVPI takes its maximum (minimum) value.
    - Capacity expansion cost & Capacity
        - When capacity is limited, i.e. 5 % and 25 %, EVPI takes its highest value if capacity expansion cost is very high; and decreases as capacity expansion cost decreases.
        - When capacity is abundant, i.e. 50 % and 100 %, the interaction becomes less significant.
    - Profitability & Capacity
        - At all levels of capacity EVPI tends to take its lowest value when the new is more profitable; but it increases when the new is less profitable than the old products.
        - When the profitability of all products is the same, EVPI tends to take its highest values at all levels of capacity.
    - Unit Capacity Usage & Capacity
        - At 5 % capacity level, EVPI increases as unit capacity usage of the new increases relatively to the unit capacity usage of the old products.
        - At all other levels, EVPI takes its highest values when the unit capacity usages of old and new products are the same or the new product's capacity usage is more than (not too much) the olds'. However, if the difference between the unit capacity usage of old and new products increases, EVPI tends to decrease.

- **EVPI over 180 runs**

    - Average EVPI (%) = 6.5
    - Min EVPI (%) = 0.25
    - Max EVPI (%) = 21

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.1.1.1 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on EVPI. Considering the rules extracted given in Appendix C.1.1.1, many of them are eliminated because of having a lift value close to 1 (the rules having a lift value around 1.6 are not considered as strong rules). Thus, only one rule with a high support and a lift but low support is obtained as an exception rule (Table 8).

Table 8. Rules extracted for Case 1 (Deterministic parameters: EVPI)

Rule	Length	Support	Confidence	Lift	Condition			Prediction (EVPI)
					Profitability	Unit Capacity Usage	Capacity	
1	2	0,054	0,974	2,681	1	1	…	M

From Table 8:

- When the new product competes for the same amount of production resources with the existing products, but less profitable than the existing ones, the EVPI is predicted to take a value between 10-20 % (i.e. "Medium" value) and thus it can be said that based on the context and firm's goals the need to invest on better forecasting technologies may be taken into account.

5.1.1.3. Experiment 2 for Uncertain Parameters

5.1.1.3.1. Experiment 2: Settings

The levels regarding the uncertain parameters considered as factors for the related experiment are specified as follows:

1. Factor H: Coefficient of variation of price (CV_Price)

   Level 1: $L_E = 0$, $L_N = 0$
   L_E: $level\ 1\ for\ the\ existing\ products\ and\ L_N$: $level\ 1\ for\ the\ new\ products$

   Level 2: $M_E = 0.15$, $M_N = 0.18$
   M_E: $level\ 2\ for\ the\ existing\ products\ and\ M_N$: $level\ 2\ for\ the\ new\ products$

   Level 3: $H_E = 0.30$, $H_N = 0.36$
   H_E: $level\ 3\ for\ the\ existing\ products\ and\ H_N$: $level\ 3\ for\ the\ new\ products$

2. Factor I: Coefficient of variation of demand (CV_Demand)

   Level 1: $L_E = 0$, $L_N = 0$
   L_E: $level\ 1\ for\ the\ existing\ products\ and\ L_N$: $level\ 1\ for\ the\ new\ products$

   Level 2: $M_E = 0.15$, $M_N = 0.18$
   M_E: $level\ 2\ for\ the\ existing\ products\ and\ M_N$: $level\ 2\ for\ the\ new\ products$

   Level 3: $H_E = 0.30$, $H_N = 0.36$
   H_E: $level\ 3\ for\ the\ existing\ products\ and\ H_N$: $level\ 3\ for\ the\ new\ products$

3. Factor J: Unit Production Cost (CV_Cost)

   Level 1: $L = 0$, ($L\ level\ 1\ for\ all\ products$)
   Level 2: $M = 0.15$, ($M\ level\ 2\ for\ all\ products$)
   Level 3: $H = 0.30$ (H: $level\ 3\ for\ all\ products$)

4. Factor K: Cannibalisation Rate (CV_CanR)

   Level 1: $L = 0$, ($L\ level\ 1\ for\ all\ products$)
   Level 2: $M = 0.15$, ($M\ level\ 2\ for\ all\ products$)
   Level 3: $H = 0.30$ (H: $level\ 3\ for\ all\ products$)

It should be noted that for this experiment the levels of deterministic parameters (such as unit capacity usage, capacity expansion cost etc. considered as factors in Section 5.1.1.2.1) are fixed to the one of the values where VSS takes the maximum values mentioned in Section 5.1.1.2.2.

The D-optimal design with 30 runs/problems (Table 6) is generated based on the levels given above and the results are presented in the following section.

5.1.1.3.2. Experiment 2: Results

The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model which ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

<u>Response 1: Value of Stochastic Solution (VSS)</u>

Remarks

- **VSS over 30 runs**
    - Average VSS (%) = 21.3
    - Min VSS (%) = 0
    - Max VSS (%) = 36.7 (when the variability of both of price and cost is higher)
- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.1.2.1 indicates that *coefficient of variation of price* and *cost (CV_Price, CV_Cost)*, are the significant parameters on VSS. The rules with high support, high confidence and high lift, which are shown in Table 9, are obtained.

Table 9. Rules extracted for Case 1 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition		Prediction (VSS)
					CV_Price	CV_Cost	
1	1	0.204	1	2.446	0.30	…	H
2	2	0.180	1	2.600	0.15	0 or 0.15	M
3	2	0.170	1	2.446	0.15	0.30	H
4	2	0.170	1	2.600	0	0.30	M
5	2	0.162	1	4.843	0	0 or 0.15	L

According to Table 9:

- When the variability of price is 0.30 (high variability), with a high support and confidence VSS is predicted to take a value greater than 20 % (i.e. "High" value), therefore in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach. Considering rule 3, with relatively lower support but high confidence, VSS is also expected to become greater than 20 % providing that the variability of price is 0.15 and variability of cost is 0.30 (High variability).
- When the coefficient of variation of price and cost is 0.15 (medium variability) and 0.0 (no variability) or 0.15, respectively, OR the coefficient of variation of price is 0.0 and of cost is 0.30, VSS is predicted to take a value between 10 % and 20 % (i.e. "Medium" value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- When there is no variability regarding price and the variability of cost is 0.00 (no variability) or 0.15 (low variability), VSS is predicted to take a value less than 10 % (i.e. "Low" category) and it can be said that stochastic programming approach can be used. In this case, the stochastic programming approach would not gain a noteworthy contribution and therefore it would be better to use a

deterministic approach considering the mean values of uncertain parameters.

- Besides, lift value for each rule is greater than 1 and this indicates a strong relationship between the left-hand-side and right-hand-side of the rule and thus it can be said that the chance that the rule occurs by coincidence is lower.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

Remarks

- **EVPI over 30 runs**
    - Average EVPI (%) = 10.8
    - Min EVPI (%) = 0.0
    - Max EVPI (%) = 13.3

- When there is no variability regarding price and cost, EVPI takes a value close to zero. In all of the other cases, EVPI falls into "Medium" category.

5.1.2. Case 2: Mix of Two Existing Lines Sold in Two Markets with a New Higher-Priced Line Sold in both of the Markets

In this case, the same firm in Case 1, which produces major household appliances (white goods), is considered. This firm has two product lines under “refrigerator” category (defined under “cold family”) according to the product hierarchy, given in Figure 6.

- 2 existing (old) lines
- 1 new line
- 2 markets
- 5 periods (years)

5.1.2.1. Basic Data

a) Products:

Existing Line 1 : Single-door
Existing Line 2 : Double-door
New Line : No-frost (higher-technology)

Single-door Double-door No-frost

b) Market:

There are two markets, national and global, in which the products are sold. The national market is considered more stable (low variability), and so in this market demand and price parameters have low coefficient of variation. However, the global market is more volatile (higher variability) and so in this market demand and price parameters have high coefficient of variation. Besides, it is assumed that (1) the global market is more profitable than the national market, (2) both the existing lines and new line are sold in both of the markets and (3) they can be introduced to those markets at the same time.

c) Product Prices:

P_{ijt} : selling price of product $i \in EP$ in market $j \in M_i$ at period t

$$P_{ijt} \sim \text{N}\,(\mu_{P_{ijt}}, \sigma^2_{P_{ijt}})$$

	National Market ($j = 1,\ t$)	Global Market ($j = 2,\ t$)
Existing line-1 ($i = 1$)	$\mu_{P_{ijt}} = 1450$[27]	$\mu_{P_{ijt}} = 1750$
Existing line-2 ($i = 2$)	$\mu_{P_{ijt}} = 2550$[28]	$\mu_{P_{ijt}} = 3000$
New Line ($i = 3$)	$\mu_{P_{311}} = \mu_{P_{312}} = \mu_{P_{313}} = 3000$ $= \mu_{P_{314}} = \mu_{P_{315}}$ $= 2500$[29]	$\mu_{P_{321}} = \mu_{P_{322}} = \mu_{P_{323}} = 3600$ $\mu_{P_{324}} = 3600 \times 0.83 = 3000$ $\mu_{P_{325}} = 3000$

The existing products are considered as mature, thereby it is assumed that the mean price is stable and doesn't change over the planning horizon (next five years). However, the new product is launched with a higher price which diminishes after three periods with a rate of 20 %. At the end of the planning horizon the price of this new product is expected to be nearly the same as of the existing product with higher price, i.e. double-door fridge.

The standard deviation of price for existing lines in the national market:

$$\sigma_{P_{i1t}} = 0.30 \times \mu_{P_{i1t}}$$

[27] This value is calculated as the average price of different models under the single-door fridge line of a firm operating in Turkey, Arçelik A.Ş.. Retrieved 5 July, 2017, from https://www.arcelik.com.tr/tek-kapili-buzdolabi/.

[28] This value is calculated as the average price of different models under the double-door fridge line of a firm operating in Turkey, Arçelik A.Ş.. Retrieved 5 July, 2017, from https://www.arcelik.com.tr/cift-kapili-buzdolabi/.

[29] This value is calculated as the average price of different models under the no-frost line of a firm operating in Turkey, Arçelik A.Ş.. Retrieved 5 July, 2017, from https://www.arcelik.com.tr/no-frost-buzdolabi/.

Since the global market is more volatile (higher variability), in this market the price parameter has high coefficient of variation, as follows:

The standard deviation of price for existing lines in the global market:

$$\sigma_{P_{i2t}} = 0.33 \times \mu_{P_{i2t}}$$

It is also assumed that the uncertainty regarding the price of new line is higher than to the existing lines, and thereby, the standard deviation of the price of the new line is considered as 10 % higher than of the existing lines, i.e.,

the standard deviation of price for new line sold in the national market:

$$\sigma_{P_{31t}} = 0.33 \times \mu_{P_{31t}} ,$$

the standard deviation of price for new line sold in the global market:

$$\sigma_{P_{32t}} = 0.36 \times \mu_{P_{32t}} .$$

d) Demand:

EPD_{ijt} : potential demand of existing product $i \in EP$ in market $j \in M_i$ at period t

$$EPD_{ijt} \sim \mathrm{N}\,(\mu_{EPD_{ijt}}, \sigma^2_{EPD_{ijt}})$$

ND_{ijt} : demand of new product $i \in NP$ in market $j \in M_i$ at period t

$$ND_{ijt} \sim \mathrm{N}\,(\mu_{ND_{ijt}}, \sigma^2_{ND_{ijt}})$$

	National Market ($j = 1,\ t$)	Global Market ($j = 2,\ t$)
Existing line-1 ($i = 1$)	$\mu_{EPD_{11t}} = 110000 - 24 \times P_{11t} + 10 \times P_{21t}$	$\mu_{EPD_{12t}} = 135000 - 35 \times P_{12t} + 10 \times P_{22t}$
Existing line-2 ($i = 2$)	$\mu_{EPD_{21t}} = 240000 - 27.5 \times P_{21t} + 23 \times P_{11t}$	$\mu_{EPD_{22t}} = 290000 - 46 \times P_{22t} + 27 \times P_{12t}$
New Line ($i = 3$)	$\mu_{ND_{311}} = 320000 - 40 \times P_{311}$ $\mu_{ND_{312}} = 340000 - 40 \times P_{312}$ $\mu_{ND_{313}} = 380000 - 45 \times P_{313}$ $\mu_{ND_{324}} = 375000 - 45 \times P_{324}$ $\mu_{ND_{325}} = 375000 - 45 \times P_{325}$	$\mu_{ND_{321}} = 353000 - 30 \times P_{321}$ $\mu_{ND_{322}} = 413000 - 30 \times P_{322}$ $\mu_{ND_{323}} = 480000 - 33 \times P_{323}$ $\mu_{ND_{324}} = 460000 - 33 \times P_{324}$ $\mu_{ND_{325}} = 470000 - 33 \times P_{325}$

The existing products are considered as mature, thereby it is assumed that the mean demand is stable and doesn't change over the planning horizon (next five years). For the new product (no-frost fridge), the average demand is expected to be relatively low in the introduction phase which is assumed for the first two years after the launching; then increases by the third year and becomes stable over the rest of the planning horizon. On the other hand, the price elasticity of demand of this new product is less than of the existing products that are more sensitive to price changes because of their maturity, but increases over the planning horizon after the introduction phase, i.e. becomes recognised in the markets[30].

In the current situation in which the new product is not introduced, the sales volume of single and double-door fridge in each market is 1/3 and 2/3, respectively. By introducing the new line to the markets, it is expected that the market size (total expected demand) would increase by 20 % after the period *t=3*.

Own-price elasticities of demand and cross elasticities don't change over the planning horizon for the existing products. Since the global market is more volatile, the own-price elasticities in this market is relatively higher than the national market, i.e. the demand in the global market is more sensitive to price changes. Besides, the

[30] This case is created on the condition that the new product will be successful. However, it is also possible to consider another case in which the new product will fail in the introduction phase.

demand of the existing line “double-door fridge” is more sensitive to price changes than of the “single-door fridge”.

The standard deviation of demand for existing lines in the national market:

$$\sigma_{P_{i1t}} = 0.30 \times \mu_{P_{i1t}}$$

Since the global market is more volatile (higher variability) and so in this market demand parameter has high coefficient of variation, as follows:

The standard deviation of demand for existing lines in the global market:

$$\sigma_{P_{i2t}} = 0.33 \times \mu_{P_{i2t}}$$

It is also assumed that the uncertainty regarding the demand of new line is higher than the existing lines, and thereby, the standard deviation of the demand of the new line is considered as 10 % higher than of the existing lines, i.e.

The standard deviation of demand for new line sold in national market:

$$\sigma_{P_{31t}} = 0.33 \times \mu_{P_{31t}}$$

The standard deviation of demand for new line sold in the global market:

$$\sigma_{P_{32t}} = 0.36 \times \mu_{P_{32t}}$$

e) Cannibalisation Rate:

e_{ijlt} : Cannibalisation rate (as a percentage) of product $l \in NP$ on product i in market j at period t

$$e_{ijlt} \sim \mathrm{N}\,(\mu_{e_{ijlt}}, \sigma^2_{e_{ijlt}})$$

	Cannibalisation rate of new line on existing lines in the national market	Cannibalisation rate of new line on existing lines in the global market
Existing line-1 ($i = 1, l = 3$, t)	$\mu_{e_{113t}} = 0.1 - 0.00017 \times P_{31t}$	$\mu_{e_{123t}} = 0.1 - 0.00020 \times P_{32t}$
Existing line-2 ($i = 2, l = 3$, t)	$\mu_{e_{213t}} = 0.1 - 0.00017 \times P_{31t}$	$\mu_{e_{223t}} = 0.1 - 0.00008 \times P_{32t}$

The variance of cannibalisation rate: $\sigma^2_{e_{ijt}} = 0.30 \times \mu_{e_{ij3t}}$

If the new line is introduced to the global market, it is expected that potential demand for both of the existing lines would be cannibalised by different rates, i.e. the “double-door fridge” will be cannibalised much more than the “single-door fridge” since the double-door is more similar to the no-frost fridge in term of functionality. Considering demand, price and cannibalisation rate are at their mean values, the following figures can be obtained (when the related parameters are at their mean values):

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$
Expected price of the new line that can be introduced to the global market	3500	3500	3500	3000	3000
Expected cannibalised demand from the existing line “single-door fridge” (by 30 %) in the global market	30000	36000	45000	54000	60000
Expected cannibalised demand from the existing line “double-door fridge” (by 70 %) in the global market	70000	84000	105000	126000	140000
Market expansion by introduction of the new line in the global market	150000	180000	200000	180000	160000
Expected demand after the introduction of the new line in the global market	250000	300000	360000	360000	360000
Expected price of the new line that can be introduced to the national market	3000	3000	3000	2500	2500
Expected cannibalised demand from the existing line “single-door fridge” (by 50 %)	25000	50000	50000	75000	75000
Expected cannibalised demand from the existing line “double-door fridge” (by 50 %)	25000	50000	50000	75000	75000
Market expansion by introduction of the new line in the national market	150000	120000	160000	110000	110000
Expected demand after the introduction of the new line in the national market	200000	220000	260000	260000	260000

As it can come out from the table above, it is assumed that a new product cannibalises a percentage (or proportion) of old product's sales, but it expands the market share of the company as well (Laruccia et al., 2012).

f) Variable unit production cost:

The mean value of cost is determined as a percentage of the mean price and different percentages are considered for the experimental study.

V_{it} : variable unit production cost of product $i \in I$ at period t

$V_{it} \sim \text{N}\,(\mu_{V_{it}},\ \sigma^2_{V_{it}})$

The variance of cost: $\sigma^2_{V_{it}} = 0.30 \times \mu_{V_{it}}$

g) Minimum launching time of the new line in the markets:

$$minL_{31} = minL_{32} = 1.$$

5.1.2.2. Other Case Data and Experiment 1 for Deterministic Parameters

5.1.2.2.1. Experiment 1: Settings

As in Case 1, four deterministic parameters considered as factors for the outer array of the experiment given in Table 4, which are

1. Capacity expansion cost with three-levels
2. Variable unit production cost as a percentage of price with three-levels
3. Unit capacity usage with five-levels
4. Maximum amount of capacity available for each period with four-levels

Factor D: Capacity expansion cost

In order to get an idea about the reasonable value of capacity expansion cost and to provide a possible trade-off between capacity expansion cost and unit contribution margin in accordance with our optimisation model, firstly we calculate the expected contribution margin including all products and all periods, calculated using 50 000 scenarios. As a result, the value of TL 1200 is obtained and set this as Level-1 (Mid) in the study. Besides, a high value and a low value of the capacity expansion cost are experimented in order to see the effect of this parameter on performance measures:

Level-1 (Mid)	:	1200
Level-2 (High)	:	1200×2 = 2400
Level-3 (Low)	:	1200/2 = 600

Factor E: Variable unit production cost as a percentage of price (profitability)

This factor can also be evaluated as the *profitability* of products which is specified using the unit production cost as a percentage of price. The levels used in this experiment are determined as follows:

Level-1	:	New product line is LESS PROFITABLE than the existing lines
Level-2	:	New product line is MORE PROFITABLE than the existing lines
Level-3	:	All existing and new lines have the SAME PROFITABILITY.

V_{it} : variable unit production cost of product $i \in I$ at period t

$$V_{it} \sim \mathrm{N}\,(\mu_{V_{it}},\ \sigma^2_{V_{it}})$$

	Single-door fridge	Double-door fridge	No-frost fridge (New Line)
Level 1	$\mu_{V_{1t}} = 0.6 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$
Level 2	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.6 \times P_{31t}$
Level 3	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.8 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$

Factor F: Unit capacity usage

Five different levels for the experiment are considered as in Case 1, as follows:

Level-1	:	Same for all products
Level-2	:	New line consumes less resource
Level-3	:	New line consumes much less resource
Level-4	:	New line consumes more resource
Level-5	:	New line consumes much more resource

Levels	Unit Capacity Usage		
	Chest	Upright	Drawer (New Line)
Level-1	1 unit	1 unit	1 unit
Level-2	1.5 unit	1.5 unit	1 unit
Level-3	2 unit	2 unit	1 unit
Level-4	1 unit	1 unit	1.5 unit
Level-5	1 unit	1 unit	2 unit

Factor G: Maximum amount of capacity in each period

In order to get an idea about the reasonable value of maximum amount of capacity available for each period, firstly we calculate the total expected demand of all products over all periods under 50 000 scenarios. Then, four different levels for this parameter are determined as a percentage of total expected demand calculated as follows:

Level-1	:	5 % of total expected demand calculated (very limited capacity)
Level-2	:	25 % of total expected demand calculated (limited capacity)
Level-3	:	50 % of total expected demand calculated (abundant capacity)
Level-4	:	100 % of total expected demand calculated (overabundant capacity)

For each level, the related percent of total expected demand is used for the maximum capacity available at t = 3, and 70 %, 80 %, 90 %, 110 % and 120 % of that value calculated for t=3 are used for ICAP, t =1, t =3, t=4 and t=5, respectively. All the capacity data is shown in the following table:

	ICAP	*t = 1*	*t = 2*	*t = 3*	*t = 4*	*t = 5*
Level 1	42 000	48 000	54 000	60 000	66 000	72 000
Level 2	210 000	240 000	270 000	300 000	330 000	360 000
Level 3	420 000	480 000	540 000	600 000	660 000	720 000
Level 4	840 000	960 000	1 080 000	1 200 000	1 320 000	1 440 000

Total number of runs performed for the experimental study = 3×3×5×4 = 180.

5.1.2.2.2. Experiment 1: Results

<u>Response 1: Value of Stochastic Solution (VSS)</u>

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.2.1.1.

Remarks

- **Main effects**
    - Considering ANOVA Table and related figures in Appendix C.2.1.1, it can be seen that the factors profitability, unit capacity usage, capacity expansion cost and maximum capacity are significant.
- **The most important interactions (based on ANOVA table and related figures in Appendix C.2.1.1)**
    - <u>Capacity expansion cost & Unit Capacity Usage</u>
        - When the unit capacity usage of the old and new products is the same or the unit capacity usage of the new product is less than of the old products, VSS increases as the capacity expansion cost reduces.
        - At all other levels of the unit capacity usage, VSS increases as the capacity expansion cost increases.

- Profitability & Unit Capacity Usage
    - When the new line is less profitable than the existing lines and unit capacity usage is the same for all products, VSS has its maximum value. However, when the profitability of the new relatively increases compared to the existing lines, VSS decreases. This relationship between the profitability and unit capacity usage is also similar to the case of that the new consumes fewer resources compared to the old lines.
    - When the new is more profitable, VSS has its maximum (minimum) value in the case of that the new uses more (less than or equal) capacity compared to the existing lines.
    - When the profitability of the new and old products is the same, the unit capacity usage becomes insignificant and VSS doesn't change.
- Profitability & Capacity
    - When the capacity is at the levels of 25, 50 and 100 %, VSS tends to take its lowest value when the new is more profitable. But VSS takes higher values when the profitability of all products is the same or the new is less profitable than the old products.
    - When the capacity is at the levels of 5 %, the profitability factor becomes insignificant and VSS doesn't change.
    - Furthermore, VSS takes its lowest value when the capacity is at 25 % level and the new is more profitable than the old products.
- Unit Capacity Usage & Capacity
    - When the maximum allowable capacity is at 5 % and 25 % levels (the capacity is limited), VSS takes its highest value when the unit capacity usage is the same for all products and tends to decrease when the unit capacity usage of the new is less or more than of the old products.

  - When the maximum allowable capacity is at 50 % level, VSS takes higher values when the unit capacity usage is the same for all products or when the unit capacity usage of the new product is more than of the old products; in all other levels of unit capacity usage, VSS tends to decrease.
  - When the maximum allowable capacity is at 100 % level (the capacity is abundant), VSS takes its lowest value when the unit capacity usage is the same for all products and tends to increase when the unit capacity usage of the new is less or more than of the old products.

- **VSS over 180 runs**
  - Average VSS (%) = 2.8
  - Min VSS (%) = 0
  - Max VSS (%) = 22.8

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.2.1.1 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on VSS. Considering the rules extracted given in Appendix C.2.1.1, many of them are eliminated because of having a lift value close to 1 (the rules having a lift value around 1.3 are considered as weak rules). Thus, eight rules with high confidence and lift but low support are considered as exception rules (Table 10). Amongst them, rule no.1 outpeforms rule no.2, rule no.5 outperforms rule no.3, 4 and 6, and rule no.7 outperforms rule no.8.

Table 10. Rules extracted for Case 2 (Deterministic parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition			Prediction (VSS)
					Profitability	**Unit Cap. U.**	**Capacity**	
1	3	0,087	1	6,23	1	2	3	M
2	2	0,087	0,978	6,10		2	3	M
3	2	0,079	0,959	8,39	1		1 or 2	H
4	1	0,079	0,908	7,94			1 or 2	H
5	3	0,078	1	8,74	1	2	1 or 2	H
6	2	0,078	0,972	8,50		2	1 or 2	H
7	3	0,073	1	6,23	3	1	1	M
8	2	0,073	0,789	4,92	3	1		M

According to Table 10:

- When the new product is less profitable, uses less production resources than the old products and capacity is loose (i.e. the level of 50 %) OR the profitability and unit capacity usage of both the new line and existing lines are the same and the capacity is tight/scarce, VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- When the new product is less profitable, consumes less amount of production resources than the old ones and capacity is tight (the level of 5% and 25 %), VSS is predicted to take a value over 20 % (this fact can be seen as compatible with the results obtained based on the interaction of factors given by ANOVA results) and thus it can be that in this case the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.2.1.2.

Remarks

- **Main effects**
    - Considering ANOVA table and related figures in Appendix C.2.1.2, it can be seen that profitability, unit capacity usage, maximum capacity available at each period and capacity expansion cost are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.2.1.1)**
    - <u>Profitability & Unit Capacity Usage</u>
        - When the new product is more profitable and consumes less production resources than the existing products, EVPI has its minimum value. However, when the profitability of the new relatively decreases compared to the existing lines, EVPI increases.
        - When the new consumes more production resources, if it is equally or more (less) profitable than the existing products, EVPI takes its maximum (minimum) value.
    - <u>Capacity expansion cost & Capacity</u>
        - When capacity is limited, 5 % and 25 %, EVPI takes its highest value if capacity expansion cost is very high; and decreases as capacity expansion cost decreases.
        - When capacity is abundant, 50 % and 100 %, the interaction becomes less significant.

- Profitability & Capacity
  - At all levels of capacity, EVPI tends to take its lowest value when the new is more profitable; but it increases when the new is less profitable than the old products or the profitability of all products is the same.
  - At all levels of profitability, EVPI increases as maximum level to which the total capacity could be expanded and takes its highest value when the capacity is abundant.
- Unit Capacity Usage & Capacity
  - At all capacity levels, EVPI takes its highest values when the unit capacity usage of old and new products is the same or the new product's capacity usage is more than (not too much) the olds'. However, if the difference between the unit capacity usage of old and new products increases, EVPI tends to decrease.

- **EVPI over 180 runs**

  - Average EVPI (%) = 6.20
  - Min EVPI (%) = 0.20
  - Max EVPI (%) = 18.4

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.2.1.2 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on EVPI. Considering the rules extracted given in Appendix C.2.1.2, six rules with high confidence and lift but low support are considered as exception rules (Table 11).

Table 11. Rules extracted for Case 2 (Deterministic parameters: EVPI)

Rule	Length	Support	Confidence	Lift	Condition			Prediction (EVPI)
					Profitability	Unit Capacity U.	Capacity	
1	1	0,081	0,986	1,906	…	…	3 or 4	M
2	1	0,077	1	2,071	…	…	1 or 2	L
3	1	0,071	1	2,071	…	2 or 3	…	L
4	2	0,070	1	1,933	…	5	3 or 4	M
5	2	0,063	1	2,071	2	1	…	L
6	2	0,055	1	2,071	…	4 or 5	1	L

According to Table 11:

- Considering rule no.1 which also outperforms rule no.4, when the capacity is loose (i.e. the level of 50 and 100 %), EVPI tends to become between 10% and 20 % and for these cases it can be said that based on the context and firm's goals the need to invest on better forecasting technologies may be taken into account. This rule is also compatible with ANOVA output.
- When the capacity is tight (i.e. the level of 5 and 25 %) OR the new product consumes less resources than the old ones OR the new product is more profitable but uses the same amount of unit production resources OR unit capacity usage of the new is more than of the old ones and capacity is very tight, EVPI is expected to be less than 10 %, thereby it can be said that having a better forecast about the uncertain parameters would not gain a noteworthy contribution for those cases.

5.1.2.3. Experiment 2 for Uncertain Parameters

5.1.2.3.1. Experiment 2: Settings

In this case, the levels of coefficient of variation for each parameter are also considered as 0, 0.15 and 0.30. Since it is assumed that the uncertainty regarding the price and the demand of new line is higher than the existing lines in both markets, the standard deviation of the price of the new line is considered as 10 % higher than of the existing lines. Plus, since the global market is more volatile than the national market, the variability in the global market is 10 % higher than the variability in the national market. The levels for the uncertain parameters are specified as follows:

1. Factor H: Coefficient of variation of price (CV_Price)

Level 1: $L_{Existing\ lines,market\ 1}$ = 0, $L_{Existing\ lines,market\ 2}$ = 0,
$L_{New\ line\ ,\ \ market\ 1} = 0, \quad L_{New\ line\ ,market\ 2} = 0$

Level 2: $M_{Existing\ lines,market\ 1}$ = 0.15, $M_{Existing\ lines,market\ 2}$ = 0.17,
$M_{New\ line\ ,\ \ market\ 1} = 0.18$, $M_{New\ line\ ,market\ 2} = 0.20$

Level 3: $H_{Existing\ lines,market\ 1}$ = 0.30, $H_{Existing\ lines,market\ 2}$ = 0.33,
$H_{New\ line\ ,market\ 1} = 0.33$, $H_{New\ line\ ,market\ 2} = 0.36$

2. Factor I: Coefficient of variation of demand (CV_Demand)

Level 1: $L_{Existing\ lines,market\ 1}$ = 0, $L_{Existing\ lines,market\ 2}$ = 0,
$L_{New\ line\ ,\ \ market\ 1} = 0, \quad L_{New\ line\ ,market\ 2} = 0$

Level 2: $M_{Existing\ lines,market\ 1}$ = 0.15, $M_{Existing\ lines,market\ 2}$ = 0.17,
$M_{New\ line\ ,\ \ market\ 1} = 0.18$, $M_{New\ line\ ,market\ 2} = 0.20$

Level 3: $H_{Existing\ lines,market\ 1}$ = 0.30, $H_{Existing\ lines,market\ 2}$ = 0.33,
$H_{New\ line\ ,market\ 1} = 0.33$, $H_{New\ line\ ,market\ 2} = 0.36$

3. Factor J: Unit Production Cost (CV_Cost)

Level 1: $L = 0, \quad (L\ level\ 1\ for\ all\ products)$
Level 2: M = 0.15, ($M\ level\ 2\ for\ all\ products$)
Level 3: H = 0.30 (H: $level\ 3\ for\ all\ products$)

4. Factor K: Cannibalisation Rate (CV_CanR)

Level 1: $L = 0, \quad (L\ level\ 1\ for\ all\ products)$
Level 2: M = 0.15, ($M\ level\ 2\ for\ all\ products$)
Level 3: H = 0.30 (H: $level\ 3\ for\ all\ products$)

It should be noted that for this experiment the levels of deterministic parameters (such as unit capacity usage, capacity expansion cost etc. considered as factors in Section 5.1.2.2.1) are fixed to the one of the values where VSS takes high values mentioned in Section 5.1.2.2.2.

The D-optimal design with 30 runs/problems (Table 6) is generated based on the levels given above and the results are presented in the following section.

5.1.2.3.2. Experiment 2: Results

The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model that ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

Response 1: Value of Stochastic Solution (VSS)

Remarks

- **VSS over 30 runs**
  - Average VSS (%) = 19.6
  - Min VSS (%) = 0
  - Max VSS (%) = 35.2 (when the variability of both of price and cost is higher)

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.2.2.1 indicates that *coefficient of variation of price* and *cost (CV_Price, CV_Cost)*, are the significant parameters on VSS. The rules with high support, high confidence and high lift, which are shown in Table 12, are obtained (those rules are the same as the rules regarding Case 1).

Table 12. Rules extracted for Case 2 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition		Prediction
					CV_Price	CV_Cost	(VSS)
1	1	0,203	1	2,487	0.30	…	H
2	2	0,179	1	2,623	0.15	0 or 0.15	M
3	2	0,170	1	2,623	0	0.30	M
4	2	0,168	1	2,487	0.15	0.30	H
5	2	0,164	1	4,614	0	0 or 0.15	L

According to Table 12:

- When the variability of price is 0.30 (high variability), with a high support, confidence and a lift greater than 1 VSS is predicted to take a value greater than 20 % (i.e. "High" value), therefore in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach. Considering rule 4, with relatively lower support but high confidence and a lift greater than 1, VSS is also expected to become greater than 20 % providing that the variability of price is 0.15 and variability of cost is 0.30 (High variability).
- When the coefficient of variation of price and cost is 0.15 (medium variability) and 0 (no variability) or 0.15, respectively, OR the coefficient of variation of price is 0 and of cost is 0.30, VSS is predicted to take a value between 10 % and 20 % (i.e. "Medium" value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- When there is no variability regarding price and the variability of cost is 0 (no variability) or 0.15 (low variability), VSS is predicted to take a value less than 10 % (i.e. "Low" category) and it can be said that stochastic programming approach can be used. In this case, the stochastic programming approach would not gain a noteworthy

contribution and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.

Response 2: Expected Value of Perfect Information (EVPI)

The minimum, average and maximum values of EVPI, which are 0 %, 1.1 % and 1.8 %, respectively, are very small, and all responses fall into "Low" category. Since it needs to have at least two classes (categories) to do classification in Random Forest, any rules regarding EVPI for this Case could not be obtained.

5.1.3. Case 3: Mix of Two Existing Lines Sold in a Single Market with Two New Lines and Five-Periods

In this case a tyre manufacturing company is considered. This firm has four product lines under "passenger cars" category/class according to the product hierarchy given in Appendix A. The details of the case are as follows:

- 2 existing (old) lines
- 2 new lines
- 5 periods (years)
- 1 market (national)

5.1.3.1. Basic Data

a) Products:

Existing Line 1 : Winter tyres
Existing Line 2 : Summer tyres
New Line 1 : 4-seasons tyres
New Line 2 : Ultra high performance

b) Market:

The products are sold only in the national market.

c) Product Prices:

P_{ijt} : selling price of product $i \in EP$ in market $j \in M_i$ at period t

$$P_{ijt} \sim \mathrm{N}\,(\mu_{P_{ijt}},\ \sigma^2_{P_{ijt}})$$

	Mean Price $j = 1,\ t$ (TL)
Existing line-1 $(i = 1)$:	$\mu_{P_{11t}} = 170$; not changed over all periods
Existing line-2 $(i = 2)$:	$\mu_{P_{21t}} = 160$; not changed over all periods
New line-1 $(i = 3)$:	$\mu_{P_{311}} = 210$, $\mu_{P_{312}} = 210$, $\mu_{P_{313}} = 210 \times 0.95 = 200$, $\mu_{P_{314}} = 200 \times 0.95 = 190$, $\mu_{P_{315}} = 190$; it is introduced with a higher price which reduces after the introduction phase
New line-2 $(i = 4)$:	$\mu_{P_{411}} = 280$, $\mu_{P_{412}} = 280$, $\mu_{P_{413}} = 280 \times 0.95 = 266$, $\mu_{P_{414}} = 266 \times 0.95 = 253$, $\mu_{P_{415}} = 253$; it is introduced with a higher price which reduces after the introduction phase

Since the existing products are considered as mature, it is assumed that the mean price is stable and does not change over the planning horizon (next five years). However, the new products are launched with a higher price diminishing in the next three periods with a rate of 5 %.

The standard deviation of price for existing lines: $\sigma_{P_{ijt}} = 0.30 \times \mu_{P_{ijt}}$

It is also assumed that the uncertainty regarding the price of new lines is higher than to the existing line, and thereby, the standard deviation of the price of new line-1 (4-season) and new line-2 (ultra high performance) is considered as 15 % and 5 %, respectively, higher than of the existing lines.

The standard deviation of price for new line -1: $\sigma_{P_{31t}} = 0.35 \times \mu_{P_{31t}}$

The standard deviation of price for new line -2: $\sigma_{P_{41t}} = 0.32 \times \mu_{P_{41t}}$

d) Demand:

EPD_{ijt} : potential demand of existing product $i \in EP$ in market $j \in M_i$ at period t

$$EPD_{ijt} \sim \mathrm{N}\,(\mu_{EPD_{ijt}},\ \sigma^2_{EPD_{ijt}})$$

ND_{ijt} : demand of new product $i \in NP$ in market $j \in M_i$ at period t

$$ND_{ijt} \sim \mathrm{N}\,(\mu_{ND_{ijt}},\ \sigma^2_{ND_{ijt}})$$

	Demand (unit)
Existing line-1 $(i = 1,\ j = 1,\ t)$	$\mu_{EPD_{11t}} = 250000 - 150 \times P_{11t}$

Existing line-2	$(i = 2,\ j = 1,\ t)$	$\mu_{EPD_{21t}} = 250000 - 170 \times P_{21t}$
New line-1	$(i = 3,\ j = 1,\ t)$	$\mu_{ND_{31t}} = 200000 - 100 \times P_{31t}$
New line-2	$(i = 4,\ j = 1,\ t)$	$\mu_{ND_{41t}} = 125000 - 50 \times P_{41t}$

The standard deviation of demand for existing lines: $\sigma_{EPD_{ijt}} = 0.30 \times \mu_{EPD_{ijt}}$

The first assumption for the demand functions is that there are no interactions between the existing line-1 and -2. Secondly, the uncertainty regarding the demand of new lines is assumed to be higher than to the existing lines, and thereby, the standard deviation of the demand of new line-1 (4-season tyre) and new line-2 (ultra high performance tyre) is considered as 15 % and 5 %, respectively, higher than of the existing lines.

The standard deviation of demand for new line -1: $\sigma_{ND_{31t}} = 0.35 \times \mu_{ND_{31t}}$

The standard deviation of demand for new line -2: $\sigma_{ND_{41t}} = 0.32 \times \mu_{ND_{41t}}$

Price elasticities of demand and cross elasticities do not change over the planning horizon. Also, price elasticities of demand are higher for existing (old) products, i.e. more sensitive to price changes.

e) Cannibalisation Rate:

e_{ijlt} : Cannibalisation rate (as a percentage) of product $l \in NP$ on product i in market j at period t

$$e_{ijlt} \sim \mathrm{N}\,(\mu_{e_{ijlt}},\ \sigma^2_{e_{ijlt}})$$

	Cannibalisation rate of new lines on existing lines
Existing line-1 $(i = 1, j = 1, l = 3,$ t):	$\mu_{e_{113t}} = 0.80 - 0.0020 \times P_{31t}$
Existing line-2 $(i = 2, j = 1, l = 3,$ t)	$\mu_{e_{213t}} = 0.08 - 0.0020 \times P_{31t}$
Existing line-1 $(i = 1, j = 1, l = 4,$ t)	$\mu_{e_{114t}} = 0.35 - 0.0008 \times P_{41t}$
Existing line-2 $(i = 2, j = 1, l = 4,$ t)	$\mu_{e_{214t}} = 0.35 - 0.0008 \times P_{41t}$

The variance of cannibalisation rate: $\sigma^2_{e_{ijlt}} = 0.30 \times \mu_{e_{ijlt}}$

In this case, both of the old lines are assumed to be cannibalised by both of the new lines; however, the new line-1 (4 season) may cannibalise the existing lines more than the new line-2 (ultra high performance/UHP tyre).

If the new line is introduced to the market, it is expected that potential demand for both of the existing lines would be cannibalised by the same rate. Considering demand, price and cannibalisation rate are at their mean values, the following figures can be obtained:

	Winter	Summer	4-season	UHP	Total
Expected demand before the introduction of the new line-1 & new line-2	240 000	240 000	0	0	480 000
Expected demand cannibalised by the new line-1 with a rate 38 %	90 000	90 000	0	0	180 000
Expected demand cannibalised by the new line-2 with a rate 13 %	30 000	30 000	0	0	60 000
Expected demand after the introduction of the new line	120 000	120 000	180 000	110 000	530 000

f) Variable unit production cost:

The mean value of cost is determined as a percentage of the mean price and different percentages are considered for the experimental study, given in the next section.

V_{it} : variable unit production cost of product $i \in I$ at period t

$$V_{it} \sim \mathrm{N}\,(\mu_{V_{it}},\ \sigma^2_{V_{it}})$$

The variance of cost: $\sigma^2_{V_{it}} = 0.30 * \mu_{V_{it}}$

g) Minimum launching time of the new line:

$$minL_{31} = minL_{41} = 1.$$

5.1.3.2. Other Case Data and Design of Experiments for Deterministic Parameters

5.1.3.2.1. Experiment 1: Settings

Four deterministic parameters are considered as factors for the outer array of the experiment given in Table 4:

1. Capacity expansion cost with three-levels
2. Variable unit production cost as a percentage of price with three-levels
3. Unit capacity usage with five-levels
4. Maximum amount of capacity available for each period with four-levels

Factor D: Capacity expansion cost

In order to get an idea about the reasonable value of capacity expansion cost and to provide a possible trade-off between capacity expansion cost and unit contribution margin in accordance with the optimisation model, firstly the expected contribution margin including all products and all periods under 50 000 scenarios is calculated, then the value of 65 is obtained and set as Level-1 (Mid) in the experimental study. Besides, a high value and a low value of the capacity expansion cost, given below, are experimented in order to see the effect of this parameter on performance measures.

Level-1 (Mid)	:	65
Level-2 (High)	:	$65\times2 = 130$
Level-3 (Low)	:	$65/2 = 33$

Factor E: Variable unit production cost as a percentage of price (Profitability)

This factor can also be evaluated as the *profitability* of products that is specified using the unit production cost as a percentage of price. The levels used in this experiment are determined as follows:

Level-1	:	The new product line is LESS PROFITABLE than the existing lines
Level-2	:	The new product line is MORE PROFITABLE than the existing lines
Level-3	:	All existing and new lines have the SAME PROFITABILITY.

	Winter	Summer	4-season (New Line-1)	U.High Perf. (New Line-2)
Level 1	$\mu_{V_{1t}} = 0.6 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$	$\mu_{V_{4t}} = 0.8 \times P_{31t}$
Level 2	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.6 \times P_{31t}$	$\mu_{V_{4t}} = 0.6 \times P_{31t}$
Level 3	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.8 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$	$\mu_{V_{4t}} = 0.8 \times P_{31t}$

Factor F: Unit capacity usage

We use five different levels for the experimental study, as follows:

Level-1	:	Same for all products
Level-2	:	New lines consume less resource
Level-3	:	New lines consume much less resource
Level-4	:	New lines consume more resource
Level-5	:	New lines consume much more resource

Unit Capacity Usage

	Winter	Summer	4-season (New Line-1)	U.High Perf. (New Line-2)
Level 1	1 unit	1 unit	1 unit	1 unit
Level 2	1.5 unit	1.5 unit	1 unit	1 unit
Level 3	2 unit	2 unit	1 unit	1 unit
Level 4	1 unit	1 unit	1.5 unit	1.5 unit
Level 5	1 unit	1 unit	2 unit	2 unit

Factor G: Maximum amount of capacity available for each period

In order to get an idea about the reasonable value of maximum amount of capacity available for each period, firstly the total expected demand of all products over all periods under 50 000 scenarios is calculated. Then, four different levels for this parameter are determined as a percentage of total expected demand calculated as follows:

Level-1	:	5 % of total expected demand calculated (very limited capacity)
Level-2	:	25 % of total expected demand calculated (limited capacity)
Level-3	:	50 % of total expected demand calculated (abundant capacity)
Level-4	:	100 % of total expected demand calculated (overabundant capacity)

For each level, the related percent of total expected demand is used for the maximum capacity available at t = 3, and 70 %, 80 %, 90 %, 110 % and 120 % of that value calculated for t=3 are used for ICAP, t =1, t =3, t=4 and t=5, respectively. All the capacity data is shown as follows:

	ICAP	*t = 1*	*t = 2*	*t = 3*	*t = 4*	*t = 5*
Level 1	28 000	32 000	36 000	40 000	44 000	48 000
Level 2	140 000	160 000	180 000	200 000	220 000	240 000
Level 3	280 000	320 000	360 000	400 000	440 000	480 000
Level 4	560 000	640 000	720 000	800 000	880 000	960 000

Total number of runs performed for DOE study = 3×3×5×4 = 180.

5.1.3.2.2. Experiment 1: Results

<u>Response 1: Value of Stochastic Solution (VSS)</u>

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.3.1.1.

Remarks

- **Main effects**
  - Considering ANOVA Table and related figures in Appendix C.3.1.1, it can be seen that the factors profitability, unit capacity usage, capacity expansion cost and maximum capacity available at each period are significant.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.3.1.1)**

  - <u>Capacity expansion cost & Unit Capacity Usage</u>
    - At all levels of unit capacity usage, VSS tends to increase as the capacity expansion cost increases.
    - VSS is higher when the unit capacity usage of new products is more than of the old products regardless of the levels of capacity expansion cost.
  - <u>Profitability & Unit Capacity Usage</u>
    - When the new products are more profitable than the existing products and the unit capacity usage is the same for all products or it is low for the new products, VSS has its minimum value. However, when the profitability of the new relatively increases compared to the existing products, VSS increases.
    - When the new products consume more resources compared to the old products but they are less profitable, VSS takes its maximum value.
  - <u>Capacity expansion cost & Capacity</u>
    - When the capacity at the level of 5 %, the capacity expansion cost is insignificant, but at this capacity level, VSS takes its highest value.

    - When the capacity is at the levels of 25 % and 50 %, VSS increases as the capacity expansion cost increases.
    - When the capacity is at the level 100 %, VSS decreases as the capacity expansion cost increases.
- Profitability & Capacity
    - When the capacity at the level of 5 %, the profitability is insignificant, but at this capacity level, VSS takes its highest value.
    - At all other capacity levels, VSS takes lower values, when the profitability of the new products is more than of the old products. If the profitability of the new products decreases compared to the old products, VSS takes higher values.
- Unit Capacity Usage & Capacity
    - When the maximum allowable capacity is at 5 % and 25 % levels (the capacity is limited), VSS takes higher values when the unit capacity usage of new products is less than of old products and tends to decrease when the unit capacity usage of the new is more than of the old products. Besides, the level of 5 % gives higher VSS values than the level of 25 %.
    - When the maximum allowable capacity is at 50 % level, VSS takes higher values when the unit capacity usage is the same for all products or when the unit capacity usage of the new product is more than of the old products; in all other levels of unit capacity usage, VSS tends to decrease.
    - When the maximum allowable capacity is at 100 % level (the capacity is abundant), VSS takes its lowest value when the unit capacity usage is the same for all products and tends to increase when the unit capacity usage of the new is less or more than of the old products.

- **VSS over 180 runs**
  - Average VSS (%) = 5.7
  - Min VSS (%) = 0
  - Max VSS (%) = 26.4

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.3.1.1 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on VSS. Considering the rules extracted given in Appendix C.3.1.1, many of them are eliminated because of having a lift value not far enough from 1 (the rules having a lift value around 1.7 are considered as weak rules). Thus, two rules with high confidence and lift but low support are considered as exception rules (Table 13).

Table 13. Rules extracted for Case 3 (Deterministic parameters: VSS)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (VSS)
					Profitability	Unit Capacity U.	Capacity	
1	2	0,084	0,943	3,250	2	…	1	M
2	2	0,059	0,891	3,071	1	…	3	M

According to Table 13:

- When the new products are less profitable and capacity is loose (the level of 50 %) OR the new one are more profitable but the capacity is tight (the level of 5 %), VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- No rules whose predicted category is “High” or “Low” with a strong support and high confidence are obtained.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.3.1.2.

Remarks

- **Main effects**
    - Considering ANOVA table and related figures in Appendix C.3.1.2, it can be seen that profitability, unit capacity usage, maximum capacity and capacity expansion cost are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.3.1.1)**
    - <u>Profitability & Unit Capacity Usage</u>
        - At all levels of the unit capacity usage, when the new product is more profitable, EVPI has its minimum value. However, when the profitability of the new relatively decreases compared to the existing lines, EVPI increases.
        - When the new consumes more production resources, almost at all levels of profitability, EVPI takes higher value.
    - <u>Capacity expansion cost & Capacity</u>
        - When capacity is limited, 5 % and 25 %, EVPI takes its highest value if capacity expansion cost is very high; and decreases as capacity expansion cost decreases.
        - When capacity is abundant, 50 % and 100 %, the interaction becomes less significant.

- Profitability & Capacity
    - At all levels of capacity, EVPI tends to take its lowest value when the new is more profitable; but it increases when the new is less profitable than the old products or the profitability of all products is the same.
    - At levels of profitability, EVPI increases as maximum level to which the total capacity could be expanded and takes its highest value when the capacity is abundant.
- Unit Capacity Usage & Capacity
    - At 5 % capacity level, EVPI increases as unit capacity usage of the new increases relatively to the unit capacity usage of the old products.
    - At all other levels, EVPI takes its highest values when the unit capacity usage of old and new products are the same or the new product's capacity usage is more than the old products. However, if the difference between the unit capacity usage of old and new products increases, EVPI tends to decrease.

- **EVPI over 180 runs**
    - Average EVPI (%) = 3.4
    - Min EVPI (%) = 0
    - Max EVPI (%) = 13

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.3.1.2 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on EVPI. Considering the rules extracted given in Appendix C.3.1.2, many of them are eliminated because of having a lift value not far enough from 1 (the rules having a lift value around 1.4 are considered as weak rules). Thus, only two rules with a high support and a lift as well as a high confidence are obtained (Table 14).

Table 14. Rules extracted for Case 3 (Deterministic parameters: EVPI)

Rule	**Length**	**Support**	**Conf.**	**Lift**	**Condition**			**Prediction (EVPI)**
					Profitability	**Unit Capacity U.**	**Capacity**	
1	2	0,116	0,949	3,338	3	4	…	M
2	2	0,100	0,935	3,289	1	1	…	M

From Table 14:

- When all products have the same profitability and the new products use more production resources than the old products OR the new products are less profitable and the unit capacity usage are the same for all products, EVPI is predicted to take a value between 10 % and 20 % (i.e. "Medium" value) and it can be said that based on the context and firm's goals the need to invest on better forecasting technologies may be taken into account.

5.1.3.3. Experiment 2 for Uncertain Parameters

5.1.3.3.1. Experiment 2: Settings

In this case, the levels of coefficient of variation for each parameter are also considered as 0, 0.15 and 0.30. Since the variability of demand and price for new products is higher than for the old products, these levels are taken 15 % and 5 % higher for the new line-1 and new line-2. The levels for the uncertain parameters are specified as follows:

Factor H: Coefficient of variation of price (CV_Price)

Level 1 $L_E = 0$, $L_{New\_Line\_1} = 0$, $L_{New\_Line\_2} = 0$

L_E: $level\ 1\ for\ the\ existing\ products\ and\ L_N$: $level\ 1\ for\ the\ new\ products$

Level 2: M_E = 0.15, $M_{New\_Line\_1} = 0.18$, $M_{New\_Line\_2} = 0.16$

$M_E: level\ 2\ for\ the\ existing\ products\ and\ M_N: level\ 2\ for\ the\ new\ products$

Level 3: H_E = 0.30, $H_{New\_Line\_1} = 0.35$, $H_{New\_Line\_2} = 0.32$

$H_E: level\ 3\ for\ the\ existing\ products\ and\ H_N: level\ 3\ for\ the\ new\ products$

Factor I: Coefficient of variation of demand (CV_Demand)

Level 1 L_E = 0, $L_{New\_Line\_1} = 0$, $L_{New\_Line\_2} = 0$

$L_E: level\ 1\ for\ the\ existing\ products\ and\ L_N: level\ 1\ for\ the\ new\ products$

Level 2: M_E = 0.15, $M_{New\_Line\_1} = 0.18$, $M_{New\_Line\_2} = 0.16$

$M_E: level\ 2\ for\ the\ existing\ products\ and\ M_N: level\ 2\ for\ the\ new\ products$

Level 3: H_E = 0.30, $H_{New\_Line\_1} = 0.35$, $H_{New\_Line\_2} = 0.32$

$H_E: level\ 3\ for\ the\ existing\ products\ and\ H_N: level\ 3\ for\ the\ new\ products$

Factor J: Unit Production Cost (CV_Cost)

Level 1: L = 0, ($L: level\ 1\ for\ all\ products$)

Level 2: M = 0.15, ($L: level\ 2\ for\ all\ products$)

Level 3: H = 0.30 ($H: level\ 3\ for\ all\ products$)

Factor K: Cannibalisation Rate (CV_CanR)

Level 1: L = 0, ($L: level\ 1\ for\ all\ products$)

Level 2: M = 0.15, ($L: level\ 2\ for\ all\ products$)

Level 3: H = 0.30 ($H: level\ 3\ for\ all\ products$)

It should be noted that for this experiment the levels of deterministic parameters (such as unit capacity usage, capacity expansion cost etc. considered as factors in

Section 5.1.3.2.1) are fixed to the one of the values where VSS takes high values mentioned in Section 5.1.3.2.2.

The D-optimal design with 30 runs/problems (Table 6) is generated based on the levels given above and the results are presented in the following section.

5.1.3.3.2. Design of Experiments (DOE): Results

The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model that ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

<u>Response 1: Value of Stochastic Solution (VSS)</u>

Remarks

- **VSS over 30 runs**
    - Average VSS (%) = 21.3
    - Min VSS (%) = 0
    - Max VSS (%) = 34.7 (when the variability of both of price and cost is higher)

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.3.2.1 indicates that *coefficient of variation of price* and *cost (CV_Price, CV_Cost)*, are the significant parameters on VSS. The rules with high support, high confidence and high lift, which are shown in Table 15, are obtained.

Table 15. Rules extracted for Case 3 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition		Prediction
					CV_Price	CV_Cost	(VSS)
1	1	0,306	0,851	2,730	0	…	L
2	1	0,236	1	2,130	0.15	0.30	H
3	1	0,207	1	2,130	0.30	…	H
4	1	0,206	0,767	3,505	0.15	…	M
5	2	0,174	1	4,569	0.15	0 or 0.15	M
6	1	0,142	0,808	2,592	…	0	L
7	2	0,137	1	3,208	0	0	L
8	2	0,113	1	3,208	0	0.15	L
9	1	0,113	0,772	2,477	…	0.15	L

According to Table 15:

- When the coefficient of variation of price or cost is 0.30 (high variability), with a high support and confidence and a lift greater than 1, VSS is predicted to take a value greater than 20 % (i.e. “High” value), therefore in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.
- When the coefficient of variation of price is 0.15 (medium variability) OR the coefficient of variation of price is 0.15 (medium variability) and the coefficient of variation of cost is 0 or 0.15 (low and medium variability) VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- When there is no variability regarding the price and/or cost OR the coefficient of variation of cost is 0.15 (medium variability) while the variability regarding the price is at its low levels, VSS is predicted to take a value less than 10 % (i.e. “Low” category) and it can be said that stochastic programming approach can be used. In this case,

the stochastic programming approach would not gain a noteworthy contribution and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.

Response 2: Expected Value of Perfect Information (EVPI)

All responses are in “Low” category; and since it needs to have at least two classes (categories) to do classification in Random Forest, any rules regarding EVPI for this case could not be found.

5.1.4. Case 4: Mix of Two Existing Lines with Two New Lines, Two Markets and Three Periods

In this case the same problem described in Section 5.1.3 is considered, but there is one additional market, i.e. global market and the planning horizon is comprised of three time-periods instead of five periods.

- 2 existing (old) lines
- 2 new lines
- 3 periods (years)
- 2 markets (national and global)

5.1.4.1. Basic Data

In this section, the data only different from Case 3 is presented.

a) Product Prices:

The mean prices of products in the global market ($j = 2$) are considered as 30 % higher than the prices in the national market. The firm plans to introduce the new line-1 (4-season tyres) to the global market with a lower price which increases by a rate of10 % in the first two years because of firm's market positioning strategy. It follows the same strategy for the new line-2, which is described in Case 5.

P_{ijt} : selling price of product $i \in EP$ in market $j \in M_i$ at period t

$$P_{ijt} \sim \mathrm{N}\,(\mu_{P_{ijt}},\ \sigma^2_{P_{ijt}})$$

		Mean Price; $j = 2,$ t (TL)
Existing line-1	$(i = 1)$	$\mu_{P_{12t}} = 220$; not changed over all periods
Existing line-2	$(i = 2)$	$\mu_{P_{22t}} = 208$; not changed over all periods
New line-1	$(i = 3)$	$\mu_{P_{321}} = 270$, $\mu_{P_{322}} = 270 \times 1.10 = 300$, $\mu_{P_{323}} = 300$
New line-2	$(i = 4)$	$\mu_{P_{421}} = 365$, $\mu_{P_{422}} = 365$, $\mu_{P_{423}} = 365 \times 0.95 = 345$

The standard deviation of price for existing lines in the market-1: $\sigma_{P_{i1t}} = 0.30 \times \mu_{P_{i1t}}$

Since the global market is more volatile than the national market, the variability in the global market is 10 % higher than the variability in the national market.

The standard deviation of price for existing lines in the market-2: $\sigma_{P_{i2t}} = 0.33 \times \mu_{P_{i2t}}$

The uncertainty regarding the price of new lines is assumed to be higher than to the existing lines in the national market, and thereby, the standard deviation of the price of new line-1 (4-season) and new line-2 (ultra high performance) is considered as 15 % and 5 %, respectively, higher than of the existing lines.

The standard deviation of price for new line -1 in the market-1: $\sigma_{P_{31t}} = 0.35 \times \mu_{P_{31t}}$
The standard deviation of price for new line -2 in the market-1: $\sigma_{P_{41t}} = 0.32 \times \mu_{P_{41t}}$
The standard deviation of price for new line -1 in the market-2: $\sigma_{P_{32t}} = 0.38 \times \mu_{P_{32t}}$
The standard deviation of price for new line -2 in the market-2: $\sigma_{P_{42t}} = 0.35 \times \mu_{P_{42t}}$

b) Demand:

EPD_{ijt} : potential demand of existing product $i \in EP$ in market $j \in M_i$ at period t

$$EPD_{ijt} \sim \mathrm{N}\,(\mu_{EPD_{ijt}}, \sigma^2_{EPD_{ijt}})$$

ND_{ijt} : demand of new product $i \in NP$ in market $j \in M_i$ at period t

$$ND_{ijt} \sim \mathrm{N}\,(\mu_{ND_{ijt}}, \sigma^2_{ND_{ijt}})$$

It is also assumed that the market size of the firm in the global market is 50 % of the national market's size of the firm. Based on this assumption, the following demand functions are obtained.

		Demand (unit) in market-2
Existing line-1	$(i = 1,\ j = 2,\ t)$	$\mu_{EPD_{12t}} = 170000 - 180 \times P_{12t}$
Existing line-2	$(i = 2,\ j = 2,\ t)$	$\mu_{EPD_{22t}} = 170000 - 210 \times P_{22t}$
New line-1	$(i = 3,\ j = 2,\ t)$	$\mu_{ND_{32t}} = 120000 - 100 \times P_{32t}$
New line-2	$(i = 4,\ j = 2,\ t)$	$\mu_{ND_{42t}} = 100000 - 120 \times P_{21t}$

The variability of the demand is assumed to be the same as the variability of the price for all products. Thus, the following figures are obtained:

The standard deviation of demand for existing lines in the market-1:

$$\sigma_{EPD_{i1t}} = 0.30 \times \mu_{EPD_{i1t}}$$

The standard deviation of demand for existing lines in the market-2:

$$\sigma_{EPD_{i2t}} = 0.33 \times \mu_{EPD_{i2t}}$$

The standard deviation of demand for new line -1 in the market-1:

$$\sigma_{ND_{31t}} = 0.35 \times \mu_{ND_{31t}}$$

The standard deviation of demand for new line -2 in the market-1:

$$\sigma_{ND_{41t}} = 0.32 \times \mu_{ND_{41t}}$$

The standard deviation of demand for new line -1 in the market-2:

$$\sigma_{ND_{32t}} = 0.38 \times \mu_{ND_{32t}}$$

The standard deviation of demand for new line -2 in the market-2:

$$\sigma_{ND_{42t}} = 0.35 \times \mu_{ND_{42t}}$$

c) Minimum launching time of the new line:

$$minL_{32} = minL_{42} = 1.$$

5.1.4.2. Other Case Data and Design of Experiments for Deterministic Parameters

5.1.4.2.1. Experiments 1: Settings

As all previous cases, four factors are considered for the experimental study:

1. Capacity expansion cost with three-levels
2. Variable unit production cost as a percentage of price with three-levels
3. Unit capacity usage with five-levels
4. Maximum amount of capacity available for each period with four-levels

Factor D: Capacity expansion cost

In order to get an idea about the reasonable value of capacity expansion cost, firstly the expected contribution margin including all products and all periods, and under 50000 scenarios is calculated and as a result of this calculation, the value of 160 is obtained and set as Level-1 (Mid) in the experimental study. Besides, a high value and a low value of the capacity expansion cost, given below, are considered in order to see the effect of this parameter on performance measures.

Level 1: 160 (Mid)
Level 2: 160×2 = 320 (High)
Level 3: 160/2 = 80 (Low)

Factor E: Variable unit production cost as a percentage of price (Profitability)

This factor can also be seen as the *profitability* of products that is specified using the unit production cost as a percentage of price. The levels used in this DOE study are as follows:

Level 1: the new lines are LESS PROFITABLE than the existing lines
Level 2: the new lines are MORE PROFITABLE than the existing lines
Level 3: SAME PROFITABILITY for all existing and new lines

	Winter	Summer	4-season (New Line-1)	U.High Perf. (New Line-2)
Level 1	$\mu_{V_{1t}} = 0.6 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$	$\mu_{V_{4t}} = 0.8 \times P_{31t}$
Level 2	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.7 \times P_{21t}$	$\mu_{V_{3t}} = 0.6 \times P_{31t}$	$\mu_{V_{4t}} = 0.6 \times P_{31t}$
Level 3	$\mu_{V_{1t}} = 0.8 \times P_{11t}$	$\mu_{V_{2t}} = 0.8 \times P_{21t}$	$\mu_{V_{3t}} = 0.8 \times P_{31t}$	$\mu_{V_{4t}} = 0.8 \times P_{31t}$

Factor F: Unit capacity usage

We use five different levels for the DOE study, as follows:

Level 1: Same for all products

Level 2: New lines consumes less resource

Level 3: New lines consumes much less resource

Level 4: New lines consumes more resource

Level 5: New lines consumes much more resource

	Unit Capacity Usage			
	Winter	Summer	4-season (New Line-1)	U.High Perf. (New Line-2)
Level 1	1 unit	1 unit	1 unit	1 unit
Level 2	1.5 unit	1.5 unit	1 unit	1 unit
Level 3	2 unit	2 unit	1 unit	1 unit
Level 4	1 unit	1 unit	1.5 unit	1.5 unit
Level 5	1 unit	1 unit	2 unit	2 unit

Factor G: Maximum amount of capacity available for each period

In order to get an idea about the reasonable value of maximum amount of capacity available for each period, firstly, the total expected demand of all products over all periods under 50 000 scenarios is calculated. Then, based on this value four different levels for this parameter as a percentage of total expected demand calculated are obtained:

Level 1: 5 % of total expected demand calculated (very limited capacity)

Level 2: 25 % of total expected demand calculated (limited capacity)

Level 3: 50 % of total expected demand calculated (abundant capacity)

Level 4: 100 % of total expected demand calculated (overabundant capacity)

For each level, the related percent of total expected demand is used for the maximum capacity at t = 2, and 70 %, 85 % and 115 % of that value calculated for t=2 are used for ICAP, t =1 and t =3, respectively. All the capacity data is shown in the following table:

	ICAP	*t = 1*	*t = 2*	*t = 3*
Level 1	38 500	46 750	55 000	63 250
Level 2	192 500	233 750	275 000	316 250
Level 3	385 000	467 500	550 000	632 500
Level 4	770 000	935 000	1 100 000	1 265 000

Total number of runs performed for DOE study = 3×3×5×4 = 180.

5.1.4.2.2 Experiments 1: Results

<u>Response 1: Value of Stochastic Solution (VSS)</u>

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.4.1.1.

Remarks

- **Main effects**
  - Considering ANOVA Table and related figures in Appendix C.4.1.1, it is seen that the factors profitability, unit capacity usage and maximum capacity available at each period are significant whilst capacity expansion cost is insignificant.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.4.1.1)**
  - <u>Profitability & Unit Capacity Usage</u>
    - When the unit capacity usage of new products is less than or the same as the unit capacity usage of old products, VSS decreases as the relative profitability of new products compared to the existing products increases.
    - When the new lines are less profitable than the existing lines and consume more resources, VSS has its minimum values. However, as the profitability of the new relatively increases

compared to the existing lines and the new ones consume more resources, VSS takes its highest values.

- Profitability & Capacity
  - At all levels of profitability, VSS takes its highest value when the capacity is limited (i.e. at levels of 5 % and 25 %) and is not affected significantly by the levels of profitability.
  - When the capacity is abundant (i.e. at level of 50 %), VSS takes its lowest value when the new lines have the same profitability as the old ones and highest value when the new lines are less profitable.
  - When the capacity is overabundant (i.e. at level of 100 %), VSS takes its highest value when the new lines have the same profitability as the old ones and lowest value when the new lines are less profitable.
- Unit Capacity Usage & Capacity
  - When the capacity is limited (i.e. at levels of 5 % and 25 %) VSS tends decrease as the unit capacity usage of the new products increases relatively compared to the old products.
  - When the capacity is abundant (i.e. at levels of 50 % and 100 %) VSS tends decrease as the unit capacity usage of the new products increases relatively compared to the old products.
- Capacity expansion cost & Capacity
  - At all levels of capacity expansion cost, VSS takes its highest value when the capacity is limited (i.e. at levels of 5 % and 25 %).
  - If capacity expansion cost is very high and the capacity is limited (i.e. at level of 5 %), VSS takes higher values.
  - When capacity is abundant (i.e. at level of 50 %), VSS takes its lowest value if capacity expansion cost is very high; and increases as capacity expansion cost decreases.

- **VSS over 180 runs**

- Average VSS (%)   = 4.7
- Minimum VSS(%) = 0
- Maximum VSS (%) = 59

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.4.1.1 indicates that *unit capacity usage* and maximum *capacity* available at each period are the significant parameters on VSS. Considering the rules extracted given in Appendix C.4.1.1, many of them are eliminated because of having a lift value close 1. Thus, only one rule with a high support, confidence and lift (Table 16).

Table 16. Rules extracted for Case 4 (Deterministic parameters: VSS)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (VSS)
					Profitability	Unit Capacity U.	Capacity	
1	2	0,105	1	10,94	…	1	1 or 2	M

According to Table 16:

- When both the new products and the old products consume the same amount of production resources and maximum capacity available at each period is at 5 or 25 % level (tight capacity), VSS is predicted to take a value between 10-20 % (i.e. “Medium” value) and thus at in those cases the decision makers can use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.4.1.2.

Remarks

- **Main effects**
  - Considering ANOVA table and related figures in Appendix C.4.1.2, we can see that profitability, unit capacity usage, maximum capacity available at each period and capacity expansion cost are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.4.1.2)**

  - <u>Profitability & Unit Capacity Usage</u>
    - When the new products are more profitable and consume the same as or less production resources than the existing products, EVPI has its minimum value. However, when the profitability of the new relatively decreases compared to the existing lines, EVPI increases.
    - When the new products consume more production resources, if they are equally or more (less) profitable than the existing products, EVPI takes its maximum (minimum) value.
  - <u>Profitability & Capacity</u>
    - At all levels of capacity EVPI tends to take its lowest value when the new is more profitable; but it increases when the new is less profitable than the old products.
    - When the profitability of all products is the same or the new ones are less profitable than the old ones, EVPI tends to take its highest values at all levels of capacity.
  - <u>Unit Capacity Usage & Capacity</u>
    - At 5 % capacity level, EVPI increases as unit capacity usage of the new increases relatively to the unit capacity usage of the old products.

- At all other levels, EVPI takes its highest values when the unit capacity usages of old and new products are the same or the new product's capacity usage is more than (not too much) the olds'. However, if the difference between the unit capacity usage of old and new products increases, EVPI tends to decrease.

- **EVPI over 180 runs**
  - Average EVPI (%) = 7.40
  - Min EVPI (%) = 0.03
  - Max EVPI (%) = 21.4

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.4.1.2 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on EVPI. Appendix C.4.1.2 shows that only two rules with a high support, confidence and a lift greater than 1 are obtained (Table 17).

Table 17. Rules extracted for Case 4 (Deterministic parameters: EVPI)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (VSS)
					Profitability	Unit Capacity U.	Capacity	
1	1	0,109	1	2,077	…	…	1 or 2	L
2	2	0,100	1	2,077	2	1	…	L

According to Table 17:

- When the capacity is tight (i.e. the level of 5 and 25 %) OR the new products are more profitable and consume the same amount of unit production resources, EVPI is expected to be less than 10 %, thereby it can be said that having a better forecast about the uncertain parameters would not gain a noteworthy contribution for those cases.

5.1.4.3. Experiment 2 for Uncertain Parameters

5.1.4.3.1. Experiment 2: Settings

In this case, the levels of CV for each parameter are also considered as 0, 0.15 and 0.30. Since the uncertainty regarding the price and demand of new lines is assumed to be higher than to the existing lines in the national market, the standard deviation of the price of new line-1 (4-season) and new line-2 (ultra high performance) is considered as 15 % and 5 %, respectively, higher than of the existing lines. In addition, since the global market is more volatile than the national market, the variability in the global market is 10 % higher than the variability in the national market.

Factor H: Coefficient of variation of price (CV_Price)

Level 1: $L_{Existing\ lines,market\ 1} = 0$, $L_{Existing\ lines,market\ 2} = 0$,

$L_{New\ line\ 1\ ,market\ 1} = 0$, $L_{New\ line\ 2\ ,market\ 1} = 0$

$L_{New\ line\ 1\ ,market\ 2} = 0$, $L_{New\ line\ 2\ ,market\ 2} = 0$

Level 2: $M_{Existing\ lines,market\ 1} = 0.15$, $M_{Existing\ lines,market\ 2} = 0.17$,

$M_{New\ line\ 1\ ,market\ 1} = 0.18$, $M_{New\ line\ 2\ ,market\ 1} = 0.16$

$M_{New\ line\ 1\ ,market\ 2} = 0.20$, $M_{New\ line\ 2\ ,market\ 2} = 0.18$

Level 3: $H_{Existing\ lines,market\ 1} = 0.30$, $H_{Existing\ lines,market\ 2} = 0.33$,

$H_{New\ line\ 1\ ,market\ 1} = 0.35$, $H_{New\ line\ 2\ ,market\ 1} = 0.32$

$H_{New\ line\ 1\ ,market\ 2} = 0.38$, $H_{New\ line\ 2\ ,market\ 2} = 0.35$

Factor I: Coefficient of variation of demand (CV_Demand)

Level 1: $L_{Existing\ lines,market\ 1} = 0$, $L_{Existing\ lines,market\ 2} = 0$,

$L_{New\ line\ 1\ ,market\ 1} = 0$, $L_{New\ line\ 2\ ,market\ 1} = 0$

$L_{New\ line\ 1\ ,market\ 2} = 0$, $L_{New\ line\ 2\ ,market\ 2} = 0$

Level 2: $M_{Existing\ lines,market\ 1}$ = 0.15, $M_{Existing\ lines,market\ 2}$ = 0.17,

$M_{New\ line\ 1\ ,market\ 1} = 0.18$, $M_{New\ line\ 2\ ,market\ 1} = 0.16$

$M_{New\ line\ 1\ ,market\ 2} = 0.20$, $M_{New\ line\ 2\ ,market\ 2} = 0.18$

Level 3: $H_{Existing\ lines,market\ 1}$ = 0.30, $H_{Existing\ lines,market\ 2}$ = 0.33,

$H_{New\ line\ 1\ ,market\ 1} = 0.35$, $H_{New\ line\ 2\ ,market\ 1} = 0.32$

$H_{New\ line\ 1\ ,market\ 2} = 0.38$, $H_{New\ line\ 2\ ,market\ 2} = 0.35$

Factor J: Unit Production Cost (CV_Cost)

Level 1: L = 0 , (L: $level\ 1\ for\ all\ products$)

Level 2: M = 0.15, (M: $level\ 1\ for\ all\ products$)

Level 2: H = 0.30 (H: $level\ 2\ for\ all\ products$)

Factor K: Cannibalisation Rate (CV_CanR)

Level 1: L = 0 , (L: $level\ 1\ for\ all\ products$)

Level 2: M = 0.15, (M: $level\ 1\ for\ all\ products$)

Level 2: H = 0.30 (H: $level\ 2\ for\ all\ products$)

It should be noted that for this experiment the levels of deterministic parameters (such as unit capacity usage, capacity expansion cost etc. considered as factors in Section 5.1.4.2.1) are fixed to the one of the values where VSS takes high values mentioned in Section 5.1.4.2.2.

The D-optimal design with 30 runs/problems (Table 6) is generated based on the levels given above and the results are presented in the following section.

5.1.4.3.2. Experiment 2: Results

The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model which ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

<u>Response 1: Value of Stochastic Solution (VSS)</u>

Remarks

- **VSS over 30 runs**

    - Average VSS (%) = 11.9
    - Min VSS (%) = 0
    - Max VSS (%) = 23.9 (when the variability of both of price and cost is higher)

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.4.2.1 indicates that *coefficient of variation of price* and *cost (CV_Price, CV_Cost)*, are the significant parameters on VSS. The rules with high support, high confidence and high lift, which are shown in Table 18, are obtained.

Table 18. Rules extracted for Case 4 (Uncertain parameters: VSS)

Rule	**Length**	**Support**	**Confidence**	**Lift**	**Condition**		**Prediction**
					CV_Price	**CV_Cost**	**(VSS)**
1	1	0,205	1	2,478	0.30	…	H
2	1	0,178	1	2,623	0.15	0 or 0.15	M
3	1	0,167	1	2,623	0	0.30	M
4	1	0,164	1	4,648	0	0 or 0.15	L
5	2	0,164	1	2,478	0.15	0.30	H

According to Table 18:

- When the coefficient of variation of price is 0.30 (high variability), and price and cost factors are at their medium and high levels respectively, VSS is predicted to take a value greater than 20 % (i.e.

"High" value), therefore in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.

- When the coefficient of variation of price is 0.15 (medium variability) and the coefficient of variation of cost is 0 or 0.15 OR the coefficient of variation of cost is 0.30 (high variability), VSS is predicted to take a value between 10 % and 20 % (i.e. "Medium" value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- When there is no variability regarding the price OR the coefficient of variation of cost is 0 or 0.15 (low and medium variability), VSS is predicted to take a value less than 10 % (i.e. "Low" category). In this case, the stochastic programming approach would not gain a noteworthy contribution and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.
- Based on all of the rules conditions and their consequences, it can be seen that the variability of price is more significant on the VSS than the variability of cost.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

Remarks

- **EVPI over 30 runs**
    - Average EVPI (%) = 9
    - Min EVPI (%) = 0
    - Max EVPI (%) = 13.3 (when the variability of both of price and cost is higher)

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.4.2.2 indicates that *coefficient of variation of price* and *cost (CV_Price, CV_Cost)*, are the significant parameters on EVPI. The rules with high support, high confidence and high lift, which shown in Table 19, are obtained.

Table 19. Rules extracted for Case 4 (Uncertain parameters: EVPI)

Rule	Length	Support	Confidence	Lift	Condition		Prediction (EVPI)
					CV_Price	CV_Cost	
1	1	0,259	1	1,944	0.30	…	M
2	1	0,248	0,926	1,907	...	0 or 0.15	L
3	1	0,239	0,909	1,872	0 or 0.15	…	L
4	2	0,235	1	2,059	0 or 0.15	0 or 0.15	L
5	1	0,225	1	2,059	0	…	L
6	2	0,205	1	2,059	0	0.30	L

According to Table 19:

- When the coefficient of variation of price is 0.30 (high variability), EVPI is predicted to take a value between 10 % and 20 % (i.e. "Medium" value) and it can be said that based on the context and firm's goals the need to invest on better forecasting technologies may be taken into account.
- When there is no variability or medium variability regarding the price and/or cost OR the coefficient of variation of cost is 0.30 (high variability) while there is no variability regarding the price, EVPI is predicted to take a value less than 10 % (i.e. "Low" category) and it should be noted that having a better forecast about the uncertain parameters would not gain a noteworthy contribution for those cases.

5.1.5. The Crossed Array Design for Experiment 1 with All Deterministic Parameters (Full Design)

In this section, the full design including all deterministic parameters (capacity expansion cost, profitability, unit capacity usage and capacity) plus common design factors (number of new lines, number of markets and number of periods) is considered. Thus a crossed arrayed design, shown in Table 4, is generated with 720 runs. The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model which ensures a detailed analysis for revealing the significance of the parameters.

Rules extracted from the Random Forest application

Response: VSS

The *Random Forest* output obtained from *RStudio* given in Appendix C.5.1 indicates that *number of new lines, number of markets, number of periods, profitability, unit capacity usage* and maximum *capacity* available at each period are the significant parameters on VSS. All the rules with minimum-support = 0.05 (though very low support is selected) and a confidence level greater than 0.75 are generated, however no rules satisfying all three criteria, particularly lift regarding each rule is around 1, are obtained from this crossed array design (see Appendix C.5.1).

Response: EVPI

The *Random Forest* output obtained from *RStudio* given in Appendix C.5.2 indicates that *number of new lines, number of markets, number of periods, profitability*, *unit capacity usage* and *capacity a*re the significant parameters on EVPI. All the rules with minimum-support = 0.05 (though very low support is selected) and a confidence level greater than 0.75 are generated, however no rules satisfying all three criteria, particularly lift regarding each rule is around 1, are obtained from this crossed array design (see Appendix C.5.2).

5.1.6. The Crossed Array Design for Experiment 2 with Uncertain Parameters (Full Design)

In this section, the full design including all uncertain parameters (coefficient of variation of demand, price, cost and cannibalisation rate) plus common design factors (number of new lines, number of markets and number of periods) is considered. Thus, another crossed arrayed design of which the outer array is constructed based on the design given in Table 5 is generated with 120 runs in total. The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model that ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

Rules extracted from the Random Forest application

Response: VSS

The *Random Forest* output obtained from *RStudio* given in Appendix C.6.1 indicates that *number of new lines, number of markets, number of periods, CV_Price and CV_Cost a*re the significant parameters on VSS. Considering the rules extracted given in Appendix C.6.1, six rules with high confidence and lift but relatively low support are obtained as exception rules in addition to one rule (rule no.1) satisfying all three measures (Table 20).

Table 20. Rules extracted for the Crossed Array Design with uncertain parameters: VSS

					Condition				
Rule	**Length**	**Support**	**Conf.**	**Lift**	**Number of markets**	**Number of periods**	**CV_Price**	**CV_Cost**	**Prediction**
1	1	0,181	0,754	1,952	...	...	0.30	…	H
2	2	0,095	0,754	2,155	...	...	0.15	0.15	M
3	2	0,073	0,786	2,176	1	...	...	0.30	H
4	2	0,071	0,779	2,157	...	2	…	0.30	H
5	2	0,066	1	3,461	...	…	0	0 or 0.15	L
6	2	0,055	0,754	2,578	...	1	0	…	L

According to Table 20:

- The significant parameters, which generate a rule amongst the common factors on VSS, are the number of periods and markets and furthermore the variability of price and cost, as it's also detected for each individual case, is significant on VSS considering the rules given in this table. When the variability of price is 0.30 (high variability), with a high support and confidence and a lift value greater than 1, VSS is predicted to take a value greater than 20 % (i.e. "High" value). Therefore, for the companies operating in the markets where the price is highly uncertain, the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach. Considering rules # 3 and # 4, with relatively lower support but higher confidence, VSS is also expected to become greater than 20 % in the case of high variability regarding cost, and longer planning horizon or lower number of markets.
- When both of the coefficient of variation of price and cost are 0.15 (medium variability), VSS is predicted to take a value between 10 % and 20 % (i.e. "Medium" value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used in those kinds of scenarios.
- When price is not handled as an uncertain parameter plus cost variability is lower OR price is not handled as an uncertain parameter plus the problem is considered in an environment with shorter planning horizon, VSS is predicted to take a value less than 10 % (i.e. "Low" value). In this case, the stochastic programming approach would not gain a noteworthy contribution and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.

Response: EVPI

The *Random Forest* output obtained from *RStudio* given in Appendix C.6.2 indicates that *number of new lines, number of markets, number of periods,* *coefficient of variation of price and cost* are the significant parameters on EVPI. The rules with minimum-support = 0.05 (i.e. a condition is frequent if it comes forward more than 72 times over all observations) and a confidence level greater than 0.75 are considered. Based on those conditions, the following rules are obtained for this case (see Table 21).

Table 21. Rules extracted for the Crossed Array Design with uncertain parameters: EVPI

Rule	Length	Support	Confidence	Lift	Condition			Prediction
					Number of periods	CV_Price	CV_Cost	
1	1	0,229	1,0	1,775	...	0	…	L
2	1	0,164	1,0	1,775	2			L
3	2	0,140	0,762	1,745	1	...	0.30	M
4	1	0,130	0,812	1,860	…	0.30	0	M
5	2	0,117	1,0	2,290	1	0.15 or 0.30	…	M
6	2	0,103	0,751	1,333	1		0 or 0.15	L
7	2	0,061	1,0	1,775		0	0 or 0.15	L

According to Table 21:

- When the variability of price is high (CV = 0.30) and the planning horizon shorter OR the variability of price is high (CV = 0.30) and there is no variability regarding cost OR the planning horizon is shorter plus if the price is considered as an uncertain parameter, EVPI is predicted to take a value between 10 % and 20 % (i.e. “Medium” value). Thus it can be said that based on the context and firm’s goals the need to invest on better forecasting technologies may be taken into account in those kinds of scenarios
- When there is no variability regarding price OR the planning horizon is longer OR the coefficient of variation of price is 0 and the variability of cost is 0 or 0.15 (no/low variability) EVPI is predicted to take a value less than 10 % (i.e. “Low” value). Thus, it can be said that having a better forecast about the uncertain parameters would not gain a noteworthy contribution for those scenarios.

- In addition to those rules, the rule number 6 is not considered as an interesting rule since the related lift value is near zero, i.e. the condition has almost no effect on the occurrence of related prediction.

5.1.7. Case 5: Mix of Two Existing Lines with a New Lower-Priced Line, a Single Market and Three-Periods (A Revised Version of Case 1)

5.1.7.1. Basic Data

In Case 1 the new line is launched with a higher price compared to the existing products. This case is the same as Case 1 in Section 5.1.1, however t**he new line is launched with a lower price** in this case. The objective of studying this case is to see if the relative price of the new line is influential on stochastic solution performance. The main differences of this case can be summarised as follows:

- The new product line is for lower-income families with lower price, i.e. its price is less than the price of old products. The mean price is:

- P_{ijt} : selling price of product $i \in EP$ in market $j \in M_i$ at period t

  $P_{ijt} \sim \text{N}\ (\mu_{P_{ijt}},\ \sigma^2_{P_{ijt}})$

  $\mu_{P_{31t}} = 1200$ TL,

  The standard deviation of price for new line: $\sigma_{P_{ijt}} = 0.36 \times \mu_{P_{31t}}$

- The expected demand of the new line:

  ND_{ijt} : demand of new product $i \in NP$ in market $j \in M_i$ at period t

  $ND_{ijt} \sim \text{N}\ (\mu_{ND_{ijt}},\ \sigma^2_{ND_{ijt}})$

  $\mu_{ND_{31t}} = 180000 - 58 \times P_{31t}$

  The standard deviation of demand for new line: $\sigma_{NP_{ijt}} = 0.36 \times \mu_{ND_{ijt}}$

- The price elasticity of demand for the new line is higher than in Case 1, since it is more price-sensitive product, which targets lower-income families.

5.1.7.2. Other Case Data and Experiment 1 for Deterministic Parameters

5.1.7.2.1. Experiment 1: Settings

All four deterministic parameters are considered as design factors for the other individual cases given in Section 5.1.1- 5.1.4.

1. Capacity expansion cost with three-levels
2. Variable unit production cost as a percentage of price with three-levels
3. Unit capacity usage with five-levels
4. Maximum amount of capacity available for each period with four-levels

Since all the levels and their values are the same as Case 1, the details would not be repeated.

5.1.7.2.2. Experiment 1: Results

Response 1: Value of Stochastic Solution (VSS)

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.7.1.1.

Remarks

- **Main effects**
    - Considering ANOVA Table and related figures in Appendix C.7.1.1, it can be seen that the factors profitability, unit capacity usage and maximum capacity available at each period are significant whilst capacity expansion cost is insignificant.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.7.1.1)**
    - Profitability & Unit Capacity Usage
        - When the new line is less (more) profitable than the existing lines, VSS has its maximum value if the unit capacity usage of new product is less (more) than of the existing products.
        - When the profitability of the new and existing lines is the same, VSS takes its maximum value if the unit capacity usage of the new and existing products is the same. Besides, when the profitability of the new relatively increases compared to the existing lines, VSS tends to decrease.
    - Profitability & Capacity
        - When the new is more profitable and capacity is limited (5 % and 25 %), VSS tends to take its lowest values.
        - When the profitability of all products is the same or the new is less profitable and capacity is limited (5 % and 25 %), VSS tends to takes its highest values.
        - When the new is more profitable and capacity is abundant (100 %), VSS tends to take its highest value.
    - Unit Capacity Usage & Capacity
        - In general, VSS is higher at 5 % capacity level than at other capacity levels.
        - At all levels of capacity (i.e. 5 %, 25 %, 50 % and 100 %), VSS tends to take its highest values when the unit capacity usage of the new product is the same as or less than of the old products. However, if the new product consumes much more or much less resources compared to the old products, VSS decreases.

- **VSS over 180 runs**
  - Average VSS (%) = 3.0
  - Min VSS (%) = 0
  - Max VSS (%) = 35

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.7.1.1 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on VSS. Considering the rules extracted given in Appendix C.7.1.1, many of them are eliminated because of having a lift value close to 1 (the rules having a lift value around 1.3 are considered as weak rules). Thus, seven rules with high confidence and lift but low support are considered as exception rules (Table 22). Amongst them, rule no.1 outpeforms rule no.2 and rule no.6 outperforms rule no.7.

Table 22. Rules extracted for Case 5 (Deterministic parameters: VSS)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (VSS)
					Profitability	**Unit Capacity U.**	**Capacity**	
1	3	0,073	1	7,598	1	2	4	M
2	2	0,073	0,971	7,378	…	2	4	M
3	2	0,066	0,963	9,717	1	…	1 or 2	H
5	2	0,066	0,975	9,838	…	2	1 or 2	H
6	3	0,057	1	7,598	3	1	1	M
7	2	0,057	0,787	5,980	3	1	…	M

According to Table 22:

- When the new product is less profitable and the capacity is tight (5% or 25%) OR the new product consumes less unit production resources and the capacity is tight, VSS is predicted to take a value over 20 % and thus in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.

- When the new product is less profitable and consumes less unit capacity than the old ones, and the capacity is loose (level of 100%) OR both the new and old products have the same profitability, consume the same amount of unit capacity and the capacity is tight, VSS is expected to be between 10 % and 20 %.
- Comparing this case with Case 1 (the only difference between Case 1 in which the new product introduces to the market with a relatively higher price compared to the old products and Case 5 in which the new product introduces to the market with a relatively lower price compared to the old products is the entry price of the new product), it can be observed that the main effect of the difference between two cases is based on the capacity available in each period. The highest VSS is obtained when capacity is loose in Case 1 and when capacity is tight in Case 5. Therefore it can be said in the case of that the price of new product is planned to be higher than of old ones and there is no capacity-scarcity problem, the stochastic approach provides a higher expected profit.

Response 2: Expected Value of Perfect Information (EVPI)

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.7.1.2.

Remarks

- **Main effects**
  - Considering ANOVA Table and related figures in Appendix C.7.1.2, it can be seen that profitability, unit capacity usage, maximum capacity available at each period are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.7.1.2)**

    - Profitability & Unit Capacity Usage
        - When the new product is more profitable and consumes the same as or less production resources than the existing products, EVPI has its minimum value. However, when the profitability of the new relatively decreases compared to the existing lines, EVPI increases.
        - When the new consumes more production resources, if it is equally or more (less) profitable than the existing products, EVPI takes its maximum (minimum) value.
    - Capacity expansion cost & Capacity
        - When capacity is limited, i.e. 5 % and 25 %, EVPI takes its highest value if capacity expansion cost is very high; and decreases as capacity expansion cost decreases.
        - When capacity is abundant, i.e. 50 % and 100 %, the interaction becomes less significant.
    - Unit Capacity Usage & Capacity
        - At 5 % capacity level, EVPI increases as unit capacity usage of the new increases relatively to the unit capacity usage of the old products.
        - At all other levels, EVPI takes its highest values when the unit capacity usages of old and new products are the same or the new product's capacity usage is more than (not too much) the olds'. However, if the difference between the unit capacity usage of old and new products increases, EVPI tends to decrease.

- **EVPI over 180 runs**
    - Average VSS (%) = 6.5

- Min VSS (%) = 0
- Max VSS (%) = 21.0

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.7.1.2 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on EVPI. Considering the rules extracted given in Appendix C.7.1.2, many of them are eliminated because of having a lift value close to 1 (the rules having a lift value around 1.6 are considered as weak rules). Thus, one rule with high confidence and lift but low support is obtained as exception rules (Table 23).

Table 23. Rules extracted for Case 5 (Deterministic parameters: EVPI)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (EVPI)
					Profitability	Unit Cap. U.	Capacity	
1	2	0,062	0,968	2,65	2	4 or 5	…	M

From Table 23:

- When the new product is more profitable and consumes more production resources than the existing products, EVPI is predicted to take a value between 10-20 % and thus it can be said that based on the context and firm’s goals the need to invest on better forecasting technologies may be taken into account.

5.1.7.3. Experiment 2 for Uncertain Parameters

5.1.7.3.1. Experiment 2: Settings

For this case, the settings given in Section 5.1.1.3 for Case 1 are also used. But, the only difference is that for this experimental study the levels of deterministic parameters (such as unit capacity usage, capacity expansion cost etc. considered as factors in Section 5.1.1.2.1) to the one of the values where VSS takes maximum values mentioned in Section 5.1.7.2.2.

5.1.7.3.2. Experiment 2: Results

The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model which ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

Response 1: Value of Stochastic Solution (VSS)

Remarks

- **VSS over 30 runs**
    - Average VSS (%) = 14.0
    - Min VSS (%) = 0
    - Max VSS (%) = 23.4 (when the variability of both of price and cost is higher)

- **Rules extracted from the Random Forest application**

  The *Random Forest* output obtained from *RStudio* given in Appendix C.7.2.1 indicates that *coefficient of variation of price, cost and cannibalisation rate (CV_Price, CV_Cost, CV_Cannb.Rate)*, are the significant parameters on VSS. The rules with high support, high confidence and high lift, which are shown in Table 24, are obtained.

Table 24. Rules extracted for Case 5 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition		Prediction (VSS)
					CV_Price	CV_Cost	
1	1	0,218	1	4,580	0.30	…	H
2	1	0,190	0,991	2,507		0	L
3	1	0,171	0,832	2,104	0.15		M
4	2	0,159	1	2,589	0 or 0.15	0	L
5	2	0,130	1	2,529	0.15	0.15 or 0.30	M

According to Table 24:

- When the coefficient of variation of price is 0.30 (high variability), with a high support, confidence and lift VSS is predicted to take a value greater than 20 % (i.e. “High” value), therefore in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.
- When the coefficient of variation of price is 0.15 (low variability) and, VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.
- When there is no variability regarding cost, VSS is predicted to take a value less than 10 % (i.e. “Low” category). In this case, the stochastic programming approach would not gain a noteworthy contribution and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

Since the minimum, average and maximum values of EVPI, which are 0.00 %, 0.19 % and 0.035 %, respectively, are close to zero and all responses fall into “Low” category, and it needs to have at least two classes (categories) to do classification in Random Forest, any rules regarding EVPI for this case could not be obtained.

Compared to Case 1 (the only difference is the new line is introduced to market with a lower price in Case 5), it can be said that VSS values decreases by 34 % in Case 5 compared to Case 1. Though price and cost are significant uncertain parameters in both cases, the variability of cannibalisation rate becomes significant when the new line is introduced to market with a lower price (i.e. Case 5). Therefore, under the condition that it is intended to introduce the new product with a higher price compared to the existing products, i.e. the new product is very competitive within the mix, it is highly recommended to use stochastic programming approach in order to get higher expected profit compared to the deterministic approach. Besides, if the new product is introduced to the market with a lower price compared to the existing products, it is beneficial to take the variability of cannibalisation rate into account.

However, though EVPI is between 9-13 % in Case 1 (except from the cases in which there is no variability regarding price and cost) , it is very close to zero in Case 5. Therefore, under the condition that it is intended to introduce the new product with a higher price compared to the existing products, digging more (perfect) information about the future will moderately contribute to the expected net profit.

5.1.8. Case 6: Mix of Two Existing Lines Sold in Two Markets with a New Higher-Priced Line Sold in one of the Markets (An Extension of Case 2)

5.1.8.1. Basic Data

In Case 2, the new line with a higher price is sold in both of the markets. This case is almost the same as the Case 2 in Section 5.1.2, however the new line can be sold in one of the markets (i.e. global market) in this Case.

5.1.8.2. Other Case Data and Experiment 1 for Deterministic Parameters

5.1.8.2.1. Experiment 1: Settings

All four deterministic parameters are considered as design factors for the other individual cases given in Section 5.1.1- 5.1.4.

1. Capacity expansion cost with three-levels
2. Variable unit production cost as a percentage of price with three-levels
3. Unit capacity usage with five-levels
4. Maximum amount of capacity available for each period with four-levels

Factor D: Capacity expansion cost

Level-1 (Mid)	:	900
Level-2 (High)	:	900×2 = 1800
Level-3 (Low)	:	900/2 = 450

Factor E: Variable unit production cost as a percentage of price (Profitability)

Level-1	:	New product line is LESS PROFITABLE than the existing lines
Level-2	:	New product line is MORE PROFITABLE than the existing lines
Level-3	:	All existing and new lines have the SAME PROFITABILITY.

Factor F: Unit capacity usage

We use five different levels for the DOE study, as follows:

Level-1 : Same for all products

Level-2 : New line consumes less resource

Level-3 : New line consumes much less resource

Level-4 : New line consumes more resource

Level-5 : New line consumes much more resource

Factor G: Maximum amount of capacity available for each period

Level-1	:	5 % of total expected demand calculated (very limited capacity)
Level-2	:	25 % of total expected demand calculated (limited capacity)
Level-3	:	50 % of total expected demand calculated (abundant capacity)
Level-4	:	100 % of total expected demand calculated (overabundant capacity)

	ICAP	*t = 1*	*t = 2*	*t = 3*	*t = 4*	*t = 5*
Level 1	31 500	36 000	40 500	45 000	49 500	54 000
Level 2	157 500	180 000	202 500	225 000	247 500	270 000
Level 3	315 000	360 000	405 000	450 000	495 000	540 000
Level 4	630 000	720 000	810 000	900 000	990 000	1 080 000

Total number of runs performed for DOE study = 3×3×5×4 = 180.

5.1.8.2.2. Experiment 1: Results

Response 1: Value of Stochastic Solution (VSS)

After solving 180 runs, the related data is analysed using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.8.1.1.

Remarks

- **Main effects**
    - Considering ANOVA Table and related figures in Appendix C.8.1.1, it can be seen that profitability, unit capacity usage,

maximum capacity available at each period and capacity expansion cost are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.8.1.1)**
    - Profitability & Unit Capacity Usage
        - When the new is more profitable than the olds, VSS tends to take its lowest values regardless of the levels of the unit capacity usage.
        - When the new and the old products have the same profitability level and the same capacity usage (it can be said that the competition between the new and old products for the capacity sharing is higher in this case), VSS is extremely high compared to other combinations of those factors. Besides, in the case of that the new is less profitable, if the unit capacity usage of the new is less than or equal to the unit capacity usage of the old products, VSS increases.
    - Profitability & Capacity
        - In general, VSS is higher at 50 % and 100 % capacity levels (i.e. capacity is abundant) than the other levels at which maximum allowable capacity is limited.
        - At all levels of capacity (i.e. 5 %, 25 %, 50 % and 100 %), VSS tends to take its lowest values when the new is more profitable than the old products. But it increases as the profitability of the new decreases compared to the profitability of the old products.
    - Unit Capacity Usage & Capacity
        - In general, at all levels of unit capacity usage, VSS tends to take higher values when maximum allowable capacity is abundant (i.e. levels of 50% and 100 %).

- At all levels of capacity (i.e. 5 %, 25 %, 50 % and 100 %), VSS tends to take its highest value when the unit capacity usage of the new product is the same as of the old products. However, if the new product consumes much more or much less resources compared to the old products, VSS takes its lowest values.

- **VSS over 180 runs**
    - Average VSS (%) = 2.8
    - Min VSS (%) = 0
    - Max VSS (%) = 26.1

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.8.1.1 indicates that *unit capacity usage, capacity* and *profitability* are the significant parameters on VSS. Considering the rules extracted given in Appendix C.8.1.1, many of them are eliminated because of having a lift value close to 1 (the rules having a lift value around 1.5 are considered as weak rules). Thus, in addition one rule having high support, confidence and lift, two rules seen in Table 25 with high confidence and lift but low support are also considered as exception rules. Amongst the exception rules, it can also be seen that rule no.2 outperforms rule no.3.

Table 25. Rules extracted for Case 6 (Deterministic parameters: VSS)

Rule	Length	Support	Conf.	Lift	Condition			Prediction (VSS)
					Profitability	**Unit Cap. U.**	**Capacity**	
1	2	0,142	0,793	2,786	1	2		H
2	3	0,076	1	11,341	3	1	1 or 2	M
3	2	0,076	0,892	10,116	…	1	1 or 2	M

According to Table 25:

- When the new line is less profitable and the unit capacity usage of new line is less than the unit capacity usage of the existing lines, VSS is predicted to take a value greater than 20 % (i.e. “High” value) and therefore decision makers should use the stochastic programming approach in order to get higher expected profit compared to the deterministic approach.
- When both the new products and the old products have the same profitability, consume the same amount of production resources and maximum capacity available at each period is at 5 or 25 % level (tight capacity), VSS is predicted to take a value between 10-20 % (i.e. “Medium” value) and thus at in those cases the decision makers can use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.

Response 2: Expected Value of Perfect Information (EVPI)

After solving 180 runs, analyse the related data using MINITAB 17 software to see which factors including interactions among them are significant. The results are shown in Appendix C.8.1.2.

Remarks

- **Main effects**
    - Considering ANOVA table and related figures in Appendix C.8.1.2, it can be seen that profitability, unit capacity usage, maximum capacity and capacity expansion cost are significant factors.

- **The most important interactions (based on ANOVA table and related figures in Appendix C.8.1.2)**

    - Capacity expansion cost & Unit Capacity Usage

    - If the unit capacity usage of the new product is the same as or more than of the existing products, EVPI is insensitive to capacity expansion cost.
    - However, when capacity expansion cost is at its minimum value, EVPI decreases and its highest value is given by the case of that the unit capacity usage of all products is the same.
- Profitability & Unit Capacity Usage
    - When the new product is more profitable and consumes less production resources than the existing products, EVPI has its minimum value. However, when the profitability of the new relatively decreases compared to the existing lines, EVPI increases.
    - When the new consumes more production resources, if it is equally or more (less) profitable than the existing products, EVPI takes its maximum (minimum) value.
- Capacity expansion cost & Capacity
    - When capacity is limited, i.e. 5 % and 25 %, EVPI takes its highest value if capacity expansion cost is very high; and decreases as capacity expansion cost decreases.
    - When capacity is abundant, i.e. 50 % and 100 %, the interaction becomes less significant.
- Profitability & Capacity
    - At all levels of capacity, EVPI tends to take its lowest value when the new is more profitable; but it increases when the new is less profitable than the old products or the profitability of all products is the same.
    - At levels of profitability, EVPI increases as maximum level to which the total capacity could be expanded and takes its highest value when the capacity is abundant.

- Unit Capacity Usage & Capacity
    - At all capacity levels, EVPI takes its highest values when the unit capacity usages of old and new products are the same or the new product's capacity usage is more than (not too much) the olds'. However, if the difference between the unit capacity usage of old and new products increases, EVPI tends to decrease.

- **EVPI over 180 runs**
    - Average EVPI (%) = 4.9
    - Min EVPI (%) = 0.3
    - Max EVPI (%) = 14

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.8.1.2 indicates that *unit capacity usage*, *capacity* and *profitability* are the significant parameters on EVPI. Considering the rules extracted given in Appendix C.8.1.2, many of them are eliminated because of having a lift value close to 1 (the rules having a lift value around 1.6 are considered as weak rules). Thus, in addition to one rule having high support, confidence and lift, two rules seen in Table 26 with high confidence and lift but low support are also considered as exception rules. Amongst the exception rules, it can also be seen that rule no.2 outperforms rule no.3.

Table 26. Rules extracted for Case 6 (Deterministic parameters: EVPI)

Rule	Length	Support	Confidence	Lift	Condition			Prediction (EVPI)
					Profitability	Unit Cap. U.	Capacity	
1	2	0,105	0,902	2,34	1	1	…	M
2	2	0,074	0,846	2,20	2 or 3	4 or 5	…	M
3	2	0,067	0,767	1,99		4 or 5	3 or 4	M

According to Table 26:

- When the new line is less profitable and the unit capacity usage of both the new and products are the same OR the profitability of new product is greater than or equal to the profitability of old producs and the new one consumes more unit production resources than the old ones OR the new one consumes more unit production resources than the old ones and capacity is loose (50 or 100 % levels), EVPI is predicted to take a value between 10-20 % (i.e. “Medium” value). Thus, in those cases it can be said that based on the context and firm’s goals the need to invest on better forecasting technologies may be taken into account.

5.1.8.3. Experiment 2 for Uncertain Parameters

5.1.8.3.1. Experiment 2: Settings

For this case, the settings given in Section 5.1.2.3 for Case 2 are also used. But, the only difference is that for this experimental study the levels of deterministic parameters (such as unit capacity usage, capacity expansion cost etc.) to the one of the values where VSS takes maximum values mentioned in Section 5.1.8.2.2.

5.1.8.3.2. Experiment 2: Results

The analysis is merely based on Random Forest method, which enables to extract some rules including different levels of the parameters, since it is not possible to find a good parametric regression model which ensures a detailed analysis for revealing the significance of the parameters. The results are shown below.

<u>**Response 1: Value of Stochastic Solution (VSS)**</u>

Remarks

- **VSS over 30 runs**
    - Average VSS (%) = 28.3

- Min VSS (%) = 0.0
- Max VSS (%) = 50.7

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.8.2.1 indicates that *the coefficient of variation of demand and price (CV_Demand, CV_Price)* are significant parameters on VSS. The rules with high support, high confidence and high lift, which are shown in Table 27, are obtained.

Table 27. Rules extracted for Case 6 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition		Prediction
					CV_Demand	CV_Price	
1	1	0,227	1	2,536	…	0.30	H
2	2	0,163	1	2,917	0	0.15	M
3	1	0,162	1	2,536	0.15	0.15	H
4	2	0,151	1	3,803	0	0	L

According to Table 27:

- When the coefficient of variation of price is 0.30 (high variability) OR both the coefficient of variation of demand and price is 0.15, with a high support and confidence and a lift greater than 1, VSS is predicted to take a value greater than 20 % (i.e. “High” value), therefore in those cases the decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.
- When the coefficient of variation of price is 0.15 (low variability) but there is no variability regarding demand, VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.

- When there is no variability regarding both demand and price, VSS is predicted to take a value less than 10 % (i.e. “Low” category). In this case, the stochastic programming approach would not gain a noteworthy contribution and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.

Response 2: Expected Value of Perfect Information (EVPI)

Since the minimum, average and maximum values of EVPI, which are 0.00 %, 0.2 % and 0.03 %, respectively, are close to zero and all responses fall into “Low” category, and it needs to have at least two classes (categories) to do classification in Random Forest, any rules regarding EVPI for this Case could not be obtained.

Compared to Case 2 (the only difference is the new line is introduced to only one of the markets in Case 6), it can be said that VSS increases by 43 % on the average in Case 6 compared to Case 2. Though price is a significant uncertain parameter in both cases, the variability of demand and cost is significant in Case 2 and Case 6, respectively. Based on this result, it can be said that introducing a product to only one or to more than one markets can affect VSS as well as significant factors on VSS.

The summary of all cases and the most important results are given in Table 28-31.

Table 28. Summary of Case 1- 6: VSS performance and rules extracted (Deterministic parameters) (*)

No	**Case**	**# of existing lines**	**# of new lines**	**# of markets**	**# of periods**	**Characteristics of the new product(s)**	**Average VSS**	**Highest VSS**	**Rules**			
									Significant Factors	**VSS: Less than 10% (Low)**	**VSS: Between 10% and 20% (Medium)**	**VSS: Greater than 20 % (High)**
1	Mix of two existing lines with a new higher-priced line, a single-market and three-periods	2	1	1	3	• launched with a higher price which reduces over the planning horizon • an improved version of the old products • cannibalises both of the old products	3.7 %	36.5 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Less UCU: Same	• PROF: Less UCU: Less CAP: Loose • PROF:Same UCU: Same CAP: Loose
2	Mix of two existing lines sold in two markets with a new higher-priced line sold in both of the markets and five-periods	2	1	2	5	• more technological and innovative product • higher price than old products • sold in both of the markets • higher price in one of the markets	2.8 %	22.8 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Less UCU: Less CAP: Loose • PROF: Same CAP: Tight	• PROF: Less UCU: Less CAP: Tight
3	Mix of two existing lines with two new lines with a single-market and five-periods	2	2	1	5	• more functional and preferable by a certain customer segment • one is more technological • demand of more functional product is higher than the demand of more technological product • price is higher for the more technological products • more functional one cannibalises the old products much more than the other	5.7 %	26.4 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Less CAP: Loose • PROF: More CAP: Tight	…

(*) PROF: Variable unit production cost as a percentage of price expressed as the profitability of new line(s) compared to the existing lines
UCU : Unit capacity usage of the new line(s), expressed as the relative resources used compared to the existing products
CAP : Maximum amount of capacity available in each period

Table 28. Summary of Case 1- 6: VSS performance and rules extracted (Deterministic parameters) (cont'd)

No	**Case**	**# of existing lines**	**# of new lines**	**# of markets**	**# of periods**	**Characteristics of the new product(s)**	**Average VSS**	**Highest VSS**	**Rules**			
									Significant Factors	**VSS: Less than 10% (Low)**	**VSS: Between 10% and 20% (Medium)**	**VSS: Greater than 20 % (High)**
4	Mix of two existing lines with two new lines, two markets and three-periods	2	2	2	3	• the same as Case 3	4.7 %	59.0 %	• unit capacity usage • max. capacity available in each period	…	• UCU: Same CAP: Tight	….
5	Mix of two existing lines with a new lower-priced line, a single market and three-periods (revised version of Case 1)	2	1	1	3	• sold for lower-income families with lower price, i.e. its price is less than the price of old products. • in other respects, the same as Case 1	3.0 %	35.0 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Less UCU: Less CAP: Loose • PROF: Same UCU: Same CAP: Tight	• PROF: Less UCU: Tight • UCU: Less CAP: Tight
6	Mix of two existing lines sold in two markets with a new higher-priced line sold only in one of the markets and five-periods (revised version of Case 2)	2	1	2	5	• more technological and innovative product • higher price than old products • sold in one of the markets	2.8 %	26.1 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Same UCU: Same CAP: Tight	• PROF: Less UCU: Less

Table 29. Summary of Case 1- 6: EVPI performance and rules extracted (Deterministic parameters) (*)

No	**Case**	**# of existing lines**	**# of new lines**	**# of markets**	**# of periods**	**Characteristics of the new product(s)**	**Average EVPI**	**Highest EVPI**	**Significant Factors**	**Rules**		
										EVPI : Less than 10% (Low)	**EVPI : Between 10% and 20% (Medium)**	**EVPI : Greater than 20 % (High)**
1	Mix of two existing lines with a new higher-priced line, a single-market and three-periods	2	1	1	3	• launched with a higher price which reduces over the planning horizon • an improved version of the old products • cannibalises both of the old products	6.5 %	21.0 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Less UCU: Same	…
2	Mix of two existing lines sold in two markets with a new higher-priced line sold in both of the markets and five-periods	2	1	2	5	• more technological and innovative product • higher price than old products • sold in both of the markets • higher price in one of the markets	6.2 %	18.4 %	• profitability • unit capacity usage • max. capacity available in each period	• CAP: tight • PROF: More UCU: Same • UCU: Less • UCU: More CAP: Tight	• CAP: Loose	…
3	Mix of two existing lines with two new lines with a single-market and five-periods	2	2	1	5	• more functional and preferable by a certain customer segment • one is more technological • demand of more functional product is higher than the demand of more technological product • price is higher for the more technological products • more functional one cannibalises the old products much more than the other	3.4 %	13.0 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Same UCU: More • PROF: Less UCU: Same	…

(*) PROF: Variable unit production cost as a percentage of price expressed as the profitability of new line(s) compared to the existing lines
UCU : Unit capacity usage of the new line(s), expressed as the relative resources used compared to the existing products
CAP : Maximum amount of capacity available in each period

Table 29. Summary of Case 1- 6: EVPI performance and rules extracted (Deterministic parameters) (cont'd)

No	**Case**	**# of existing lines**	**# of new lines**	**# of markets**	**# of periods**	**Characteristics of the new product(s)**	**Average EVPI**	**Highest EVPI**	**Significant Factors**	**Rules**		
										EVPI : Less than 10% (Low)	**EVPI : Between 10% and 20% (Medium)**	**EVPI : Greater than 20 % (High)**
4	Mix of two existing lines with two new lines, two markets and three-periods	2	2	2	3	• the same as Case 3	7.3 %	21.4 %	• profitability • unit capacity usage • max. capacity available in each period	• CAP: tight • PROF: More UCU: Same	…	…
5	Mix of two existing lines with a new lower-priced line, a single market and three-periods (revised version of Case 1)	2	1	1	3	• sold for lower-income families with lower price, i.e. its price is less than the price of old products. • in other respects, the same as Case 1	6.5 %	21.0 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: More UCU: More	…
6	Mix of two existing lines sold in two markets with a new higher-priced line sold only in one of the markets and five-periods (revised version of Case 2)	2	1	2	5	• more technological and innovative product • higher price than old products • sold in one of the markets	4.9 %	14 %	• profitability • unit capacity usage • max. capacity available in each period	…	• PROF: Less UCU: Same • PROF: Same and More UCU: More CAP: Loose	…

Table 30. Summary of Case 1- 6: The significant uncertain parameters (factors) on VSS and EVPI (*)

No	Case	VSS			EVPI		
		Average and Highest values (%)	**Significant main effects**	**Significant interactions**	**Average and Highest values (%)**	**Significant main effects**	**Significant interactions**
1	Mix of two existing lines with a new higher-priced line, a single-market and three-periods	21.3 36.7	• Price • Cost	• Price & Cost	10.8 13.3	• Price • Cost	• Price & Cost
2	Mix of two existing lines sold in two markets with a new higher-priced line sold in both of the markets and five-periods	19.6 35.2	• Price • Cost	• Price & Cost	Close to zero	---- (**)	---- (**)
3	Mix of two existing lines with two new lines with a single-market and five-periods	21.3 34.7	• Price • Cost	• Price & Cost	Close to zero	---- (**)	---- (**)
4	Mix of two existing lines with two new lines, two markets and three-periods	11.9 23.9	• Price • Cost	• Price & Cost	9.0 13.3	• Price • Cost	• Price & Cost
5	Mix of two existing lines with a new lower-priced line, a single market and three-periods (revised version of Case 1)	14.0 23.4	• Price Cost	• Price & Cost	Close to zero	---- (**)	---- (**)
6	Mix of two existing lines sold in two markets with a new higher-priced line sold only in one of the markets and five-periods (revised version of Case 2)	28.3 50.7	• Price • Demand	• Price & Demand	Close to zero	---- (**)	---- (**)

(*): This table is designed based on the results of parametric regression analysis. An additional analysis based on random forest method is also given for those individual cases (see Table 31).

(**): Since EVPI values are very small and near zero, the analysis is not done for these cases.

Table 31. Summary Case 1-6: Rules extracted (Uncertain parameters: VSS)

Rule	Condition			Prediction (*)	Case
	Variability of Demand	**Variability of Price**	**Variability of Cost**		
1	…	High	…	VSS ≥ 20 %	1, 2, 3, 4, 5, 6
2	…	Medium	High		1, 2, 3, 4
3	Medium	Medium	…		6
4	…	Medium	Low or Medium	10 % ≤ VSS < 20%	1, 2, 3, 4, 5
5	Low	Medium	…		6
6	...	Low	High		1, 2, 4
7	…	Medium	High		5
8	…	Low	Low	VSS < 10 %	1, 2, 3, 4, 5
9	…	Low	Medium		1, 2, 3, 4
10	…	Medium	Low		5
11	Low	Low	…		6

5.1.9. General Conclusions for Case 1-6

Based on Table 28-31 and Section 5.1.1-5.1.6, the following generalised inferences can be detected:

- **Considering the significant main deterministic factors (parameters) and their interactions on VSS:**
  - As it can be seen from Table 28, VSS performance of all different types of problems depends on relative *unit capacity usage* of the new and old products and *maximum allowable capacity in each period* for all six cases; besides relative *profitability* of the new and old products is a significant factor for all cases except from one case.
  - The interactions, *profitability&unit capacity usage*, *profitability&capacity* and *unit capacity usage&capacity* are significant for all cases. However, their levels which give the maximum or minimum VSS values and their effect are different for cases.
  - Considering the rules extracted for each case, it can be seen that it is hard to find any common conditions and to make some generalisations covering all cases. However, It should be noted that (on the average) VSS tends to be high when the new products are less profitable or have the same profitability as the old products and the unit capacity usage of the new products are the same as or less than the old products.
  - VSS is expected to be 23 % - 59 % at maximum considering all six cases. Based on this result, it can be said that the competition between the new and old products for sharing the limited capacity and for making contribution to the expected net profit increases, VSS also tends to increase.

- Since the only difference between Case 1 and Case 5 is the entry price of the new product, it is noteworthy to see if there would be any interesting result about the effects of factors on VSS because of this difference. It can be observed that the main effect of this difference reveals based on the capacity available in each period. The highest VSS is obtained when capacity is loose in Case 1 and when capacity is tight in Case 5. Therefore it can be said in the case of that the price of new product is planned to be higher than of old ones and there is no capacity-scarcity problem, the stochastic approach provides a higher expected profit.
- For Case 2 and Case 6 (which is the revised version of Case 2), VSS tends to become higher (at least 10 %, which can be seen from the rules) when new and old (existing) products consume the same or less amount of resources. Also VSS tends to become lower when the new product is more profitable. Thus, in general it can be said that when the new products are less profitable or as same profitable as the old products, unit capacity usage of new products are less than or equal to the unit capacity usage of new products, VSS becomes higher. Therefore, it should be noted that there is no remarkable difference between those cases in terms of VSS values (average, minimum and maximum values) and the factors affecting the VSS.
- For Case 3, VSS tends to increase when the unit capacity usage of the new products increases compared to the unit capacity usage of old products and the new products are less profitable and decrease as the capacity becomes tight..
- For Case 4, similar to Case 2 and Case 6, VSS decreases as the difference between the unit capacity usage of new and old products increases and it takes higher values in all other levels of unit capacity usage. In contrast to all other cases, VSS becomes highest when the

new products are more profitable and decreases almost linearly as maximum allowable capacity becomes more loose in Case 4.

- **Considering main deterministic factors (parameters) and interactions significant for EVPI**

    - As it can be seen from Table 29, EVPI performance of all different types of problems depends on relative *profitability* of new products compared to existing products, *unit capacity usage* of them and maximum allowable *capacity* in each period. Besides, *capacity expansion cost* is not a significant factor on EVPI performance in all of the cases.
    - The interactions, *profitability&unit capacity usage* and *unit capacity usage&capacity*, are significant for all cases, and their effects on EVPI are almost the same for all cases. Besides, the interactions, *profitability&capacity* and *capacity expansion cost&capacity*, are also significant for most of the cases.
    - The highest EVPI is expected to be between 13 % and 21 % and considering the rules extracted for each case, it can be seen that it is hard to find any common conditions and to make some generalisations covering all cases.

- **Considering the rules extracted from the crossed array design with deterministic parameters;**

    - Based on the results obtained from the Random Forest application, number of new lines, number of markets, number of periods, profitability, unit capacity usage and maximum capacity available at each period are the significant parameters on both VSS and EVPI; however, any rules satisfying all three criteria are obtained from this crossed array design.

- **Considering main uncertain parameters, their interactions significant for VSS and EVPI, and rules extracted;**
    - As seen from Table 30:
        - The uncertain parameters, price and cost, and their interaction are the most significant factors on VSS for all of the cases except from Case 6 in which demand is significant rather than cost in addition to price. As the variability of price and cost parameters for Case 1-5 and the variability of price and demand for Case 6 increase, VSS also increases and it takes its highest values (23.4 % at minimum and 36.7 % at maximum) when the variability of both of these parameters is at their highest values (i.e. coefficient of variation = 0.30). This result motivates decision makers to use stochastic programming for the cases in which unit contribution margin (= price - cost) of products is highly uncertain.
        - The variability of price and cost are also significant on EVPI just for Case 1 and 4, and EVPI is expected to become around 10 % on the condition that price has higher variability. However if the variability of price and/or cost is not high (i.e. 0 or 0.5), EVPI is expected to be close to zero.
    - According to Table 31 which shows the rules extracted from association analysis based on Random Forest model:
        - The coefficient of variation of price for all the cases and cost for most of the cases, and demand for only one case are the significant factors which determine the VSS.
        - VSS becomes more than 20 % on the condition that the variability of price is high, i.e. coefficient of variation is 0.30, for all of the cases. Besides, when the variability of price is

medium and of cost is high, VSS is expected to be more than 20 % for cases 1-4 and at least 10 % for all of the cases where price and cost are significant factors on VSS. Thereby, the stochastic programming approach is recommended for the price-volatile markets as well as cost-volatile markets in order to have a high expected profit.

- When the variability of price is medium but the variability of cost is low or medium, VSS is expected to be at least 10 % for cases 1-5 in which price and cost are significant factors on VSS.
- When the variability of price is low, there is no variability regarding price, and the variability of cost is zero or less than 0.15, VSS is expected to be less than 10 %. Thereby, it can be said that for the cases in which price is handled as a deterministic parameter and the variability regarding cost is not too high the stochastic programming approach would not gain a noteworthy contribution, and therefore it would be better to use a deterministic approach considering the mean values of uncertain parameters.

- According to Table 20 and Table 21:
    - Considering the crossed array design based on Case 1-4, the significant deterministic and uncertain parameters on VSS are "*the number of markets*", "*the number of periods*", "*variability of price*" and "vari*ability of cost*".
    - When price has a high variability, VSS is expected to be more than 20 %. Besides, if the variability of cost is high and there are fewer markets or longer planning horizon, VSS is also expected to be more than 20 %.

- VSS is expected to be at least 10 % providing that both the variability regarding price and cost is not too high.
- If there is no variability regarding price and cost or price is handled as a deterministic parameter but the planning horizon is shorter, i.e. 3 periods, VSS is expected to become less than 10 %.
- As a generalised inference, it can be said that stochastic programming approach is highly recommended in order to handle high uncertainty regarding price and cost.
- Considering EVPI, it can be said that the need to invest on better forecasting technologies may be taken into account in cases with a shorter planning horizon with a high cost variability OR in cases in which the variability of price is high. Plus it should be noted that in cases where both the variability of price and cost is low, having a better forecast about the uncertain parameters would not gain a noteworthy contribution to the expected profit.

5.2. Illustrative Cases – Set 2

In this section, three larger-sized cases are developed to make inferences about the solution capacity of the model and the solution approach developed in terms of solution time.

5.2.1. Data for Cases

The problem sizes that are compatible with the context, i.e. product-mix problem at product-line level, are given as follows:

Case	# of existing lines	# of new lines	# of markets	# of periods	Problem size
7	4	1	10	3	Small
8	6	2	20	5	Medium
9	8	4	20	5	Large

While generating the problem data for each case, the following settings are taken into account:

- In order to get a general idea about the reasonable value of maximum amount of capacity available for each period, firstly the total expected demand of all products over all periods under 50 000 scenarios are calculated. Then the value of this parameter is determined as 25 % of total expected demand calculated.
- In order to get an idea about the reasonable value of capacity expansion cost and to provide a possible trade-off between capacity expansion cost and unit contribution margin in accordance with the optimisation model, firstly the expected contribution margin including all products and all periods under 50 000 scenarios is calculated. Then, this roughly calculated value is used for the capacity expansion cost for each problem.

- A product can be sold in every market but with different prices (variability is not too much) and different amounts of demand.
- An old product to be sold in a new market is also handled as a new product.
- Half of the new products are introduced with high (low) prices which goes down (up) over the periods with a rate of 10 % and the prices over the periods of old (mature) products are relatively stable (i.e. does not change over the periods).
- The constant term of demand function is the same for each period for all products and selected randomly from a discrete distribution.
- The demand of some of the new products will grow at a certain rate (e.g. 3 %) per year after introducing to a market. In addition, for other new products, the demand is high when they are launched, but decreases over the planning horizon at a certain rate per year thereafter.
- Price elasticities of demand are higher for old products (more sensitive to price changes) than for new products, increase over the planning horizon with a rate of 10 % for new products and, on the other hand, stable for the old products.
- The mean prices, constant term of demand functions and price elasticities (including cross-elasticities) are randomly chosen from a continuous distribution (normal distribution) whose parameters are exogenously assigned.
- For unit production costs, firstly 50 000 scenarios are generated for each market price parameter of each product, and then the expected values are calculated over 50 000 scenarios. Finally, the minimum of these expected market prices for each product is used for unit production costs. Thus, some markets would be more profitable than the others, as in real business environment would.

- Cannibalisation rate depends on the price of the new product (s) and for it a linear function is determined considering minimum and maximum values of cannibalisation rate. Some new product(s) cannibalise the potential demand of the old products more than the other(s). If a new line(s) is (are) introduced to any market, some of the old products, not all them, will be cannibalised by this (these) new line (s).
- Minimum launching time for each product sold in each market is assigned randomly from a discrete distribution.

  $If\ t = 3, minL_{ij} = \{1, 2\}\ and\ if\ t = 5, minL_{ij} = \{1, 2,3\}$.
- The uncertainty regarding the price and demand of new lines is higher than of existing lines, and thereby, the standard deviation of the price of new line(s) is considered as 20 % higher than of the existing lines.
- All random parameters are uniformly distributed in those cases.
- By introducing new line(s) to the market, it is expected that the market size (total expected demand) would increase.

5.2.2. Experimental Design: Settings

The main objective of this experiment[31] is to explore if the stochastic solution approach is sensitive to problem size and the variability regarding the uncertain parameters. Since the stochastic environment of the problem is considered, one of the levels, i.e. the coefficient of variation is zero, handled in the experiments given in Section 5.1 is not taken into account in this experiment. Furthermore, the medium level determined for the experiments in Section 5.1., i.e. the coefficient of variation (CV) for each parameter = 0.15, is changed to CV = 0.05 in order to have at least some stochasticity for the problem handled. Thereby, in order to test the solution

[31] The experiment designed for exploring the effect of the deterministic parameters on stochastic solution quality could not be performed because of requiring very long solution time which makes it impossible to perform all required runs (i.e. 180 runs).

time as a performance measure two levels for the variability of uncertain parameters are determined: (1) the coefficient of variation (CV) for each parameter = 0.05 and (2) CV = 0.30. Since the variability of demand and price for new products is higher than for the old products, these levels are taken as 20 % higher for the new products, i.e. 0.06 and 0.36. Thus, the levels can be stated as follows:

Factor H: Coefficient of variation of price (CV_Price)

Level 1: L_E = 0.05, $L_N = 0.06$
(
L_E: $level\ 1\ for\ the\ existing\ products\ and\ L_N$: $level\ 1\ for\ the\ new\ products$)

Level 2: H_E = 0.30, $H_N = 0.36$
(
H_E: $level\ 2\ for\ the\ existing\ products\ and\ H_N$: $level\ 2\ for\ the\ new\ products$)

Factor I: Coefficient of variation of demand (CV_Demand)

Level 1: L_E = 0.05, $L_N = 0.06$
(
L_E: $level\ 1\ for\ the\ existing\ products\ and\ L_N$: $level\ 1\ for\ the\ new\ products$)

Level 2: H_E = 0.30, $H_N = 0.36$
(
H_E: $level\ 2\ for\ the\ existing\ products\ and\ H_N$: $level\ 2\ for\ the\ new\ products$)

Factor J: Unit Production Cost (CV_Cost)

Level 1: L = 0.05, (L: $level\ 1\ for\ all\ products$)

Level 2: H = 0.30 (H: $level\ 2\ for\ all\ products$)

Factor K: Cannibalisation Rate (CV_CanR)

Level 1: L = 0.05, (L: $level\ 1\ for\ all\ products$)

Level 2: H = 0.30 (H: $level\ 2\ for\ all\ products$)

The experiment is well-known full-factorial design with $2^4 = 16$ runs. For each case, 16 problems are generated by different combinations of the uncertain parameters having two levels and the results are shown in the following section. It should be noted that in addition to data collected regarding the solution time as a performance measure, VSS and EVPI are also recorded for each run in order to perform an additional analysis for testing the effect of the uncertain parameters on those measures.

5.2.3. Experimental Design: Results

5.2.3.1. Case 7: 4 existing lines, 1 new line, 10 markets, 3 periods

After solving 16 runs, the related data is analysed using MINITAB software to see which uncertain parameters and their interactions are significant on the solution time and solution performance in terms of VSS and EVPI, and using Random Forest method to extract rules including different levels of the uncertain parameters. The results are shown below.

Response 1: Value of Stochastic Solution (VSS)

Remarks

- **Main effects and the most important interactions**

  Considering ANOVA table and related figures in Appendix C.9, it can be seen that:

  - the variability of price, cost and interaction between them are the most significant factors on VSS,
  - VSS increases when the variability of the price and the cost increases, and it takes its maximum value (around 15 %) when the variability of both factors, i.e. price and cost, is higher,
  - the variability of demand and the interaction of demand and cannibalisation rate are very slightly significant on VSS, and
  - the variability of the cannibalisation rate is insignificant.

- **VSS over 16 runs**
  - Average VSS (%) = 10.6
  - Min VSS (%) = 5.0
  - Max VSS (%) = 15.4

- **Solution time**
  - Nearly 3600 seconds for each run

- **Stochastic solution quality: optimality gap of solution approach**
  - Min=0.8 %, Average =1.2 %, Max = 1.85 %.
  - All runs are solved with 30 replications (scenario batches).

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.9 indicates that *CV_Price (variability of price)* is the significant parameter on VSS. The rules with high support, high confidence and high lift, which are shown in Table 32, are obtained.

Table 32. Rules extracted for Case 7 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition	Prediction
					CV_Price	**(VSS)**
1	1	0.50	1	2.00	0.05	L
2	1	0.50	1	2.00	0.30	M

According to Table 32:

- When the coefficient of variation of price is 0.30 (high variability), with a high support, confidence and a lift, VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value).
- When the coefficient of variation of price is 0.05 (low variability), with a high support, confidence and lift, VSS is predicted to take a value less than 10 % (i.e. “Low” value).

- At first sight, there may seem a contradiction between the results of ANOVA and association analysis done through Random Forest about the significance of the CV_Cost parameter; i.e. the coefficient of variation of cost is significant factor in ANOVA whilst it is insignificant in Random Forest. Since the VSS for each experiment is categorised and thus in both cases the response (VSS) falls into the same category and becomes indifferent, the Random Forest concludes that only CV_Price is significant and CV_Cost is insignificant. However, in ANOVA the small difference between two cases becomes indicative, and thereby both the CV_Price and CV_Cost are labelled as significant.

<u>Response 2: Expected Value of Perfect Information (EVPI)</u>

Since the minimum, average and maximum values of EVPI, which are 1.4 %, 1.5 % and 1.6 %, respectively, are very small, an ANOVA is not performed. However, it can be said that digging more (perfect) information about the future does not contribute to the expected net profit in this case.

Besides, all of sixteen experiments have an EVPI very close to one percent, and so all responses fall into “Low” category. Since it needs to have at least two classes (categories) to do classification in Random Forest, any rules regarding EVPI could not be obtained for this Case.

5.2.3.2. Case 8: 6 existing lines, 2 new lines, 20 markets, 5 periods

<u>Response 1: Value of Stochastic Solution (VSS)</u>

Remarks

- **Main effects and the most important interactions**

  Considering ANOVA table and related figures in Appendix C.10, it can be seen that,

- the variability of price is the most significant factor on VSS; following the price, cost, and the interaction between price and cost are the other significant factors. VSS increases when the variability of the price and the cost increases, and it takes its maximum value (around 9.5 %) when the variability of both factors, i.e. price and cost, is higher,
- the variability of demand and cannibalisation rate, and the interaction of demand and price, and demand and cost are very slightly significant on VSS.

- **VSS over 16 runs**
  - Average VSS (%) = 6.0
  - Min VSS (%) = 2.2
  - Max VSS (%) = 9.5 (when the variability of both of price and cost is higher)

- **Solution time**
  - Nearly 12600 seconds (around 3.5 hours) for each run

- **Stochastic solution quality: optimality gap of solution approach**
  - Min=0.18 %, Average =0.32 %, Max = 0.53 %.
  - All runs are solved with 30 replications.

- **Rules extracted from the Random Forest application**

All of sixteen experiments have a VSS less than 9 %, and so all responses fall into “Low” category. Since it needs to have at least two classes (categories) to do classification in Random Forest, any rules regarding VSS could not be obtained for this Case.

Response 2: Expected Value of Perfect Information (EVPI)

Since the minimum, average and maximum values of EVPI, which are 0.07 %, 0.4 % and 0.8 %, respectively, are very small and near zero, the ANOVA is not performed. However, it can be said that digging more (perfect) information about the future does not contribute to the expected net profit in this case.

Besides, all of sixteen experiments have an EVPI very close to zero, and so all responses are in "Low" category. Since it needs to have at least two classes (categories) to do classification in Random Forest, any rules could not be found regarding EVPI for this Case.

5.2.3.3. Case 9: 8 existing lines, 4 new lines, 20 markets, 5 periods

Response 1: Value of Stochastic Solution (VSS)

Remarks

- **Main effects and the most important interactions**

  Considering ANOVA table and related figures in Appendix C.11, it can be seen that:

  - the variability of price is the most significant factor on VSS, following this, cost, and the interaction between price and cost are the other more significant factors. VSS increases when the variability of the price and the cost increases, and it takes its maximum value (around 11 %) when the variability of both factors, i.e. price and cost, is higher,
  - the interaction between price and the cannibalisation rate is also very slightly significant on VSS.

- **VSS over 16 runs**
  - Average VSS (%) = 8.3
  - Min VSS (%) = 5.3

- Max VSS (%) = 11 (when the variability of both of price and cost is higher)

- **Solution time**
  - Nearly 33000 seconds (nearly 9.17 hours) for each run.
- **Stochastic solution quality: optimality gap of solution approach**
  - Min=0.19 %, Average =0.39 %, Max = 0.96 %.
  - The solution is obtained with 40 replications and $N = 1000$ scenarios.

- **Rules extracted from the Random Forest application**

The *Random Forest* output obtained from *RStudio* given in Appendix C.11 indicates that *the coefficient of variation of price (CV_Price)* and *cost (CV_Cost)* are significant parameters on VSS. Considering the rules extracted given in Appendix C.11, many of them are eliminated because of having a lift value not far enough from 1 (the rules having a lift value around 1.5 are considered as weak rules). Thus, only one rule with high confidence, support and lift are obtained (Table 33).

Table 33. Rules extracted for Case 9 (Uncertain parameters: VSS)

Rule	Length	Support	Confidence	Lift	Condition		Prediction
					CV_Price	CV_Cost	
1	2	0,329	1	3,041	0.30	…	M

According to Table 33:

- When the coefficient of variation of price is 0.30 (high variability), VSS is predicted to take a value between 10 % and 20 % (i.e. “Medium” value) and it can be said that based on the satisfaction level of the decision maker, stochastic programming approach can be used.

Response 2: Expected Value of Perfect Information (EVPI)

Since the minimum, average and maximum values of EVPI, which are 0.02 %, 0.3 % and 0.97 %, respectively, are very small and near zero, the ANOVA is not performed. However, it can be said that digging more (perfect) information about the future does not contribute to the expected net profit in this case.

Besides, all of sixteen experiments have an EVPI very close to zero, and so all responses are in “Low” category. Since it needs to have at least two classes (categories) to do classification in Random Forest, any rules could not be found regarding EVPI for this Case.

The summary of the cases and the most important results are given in Table 34.

Table 34. The significant random factors on VSS for Case 7-9

Problem	**Solution Time (sec's)**	**VSS (%)**			**The most significant factors on VSS**	**Rules**
		Min	Avg	Max		
4-1-10-3	≅ 3600	5	10.6	15.4	• Price • Cost • Interaction of price & cost	$CV_{Price} = 0.05 \Rightarrow VSS: [0\,;10\%)$ $CV_{Price} = 0.30 \Rightarrow VSS: [10; 20\%)$
6-4-20-5	≅ 12600	2.2	6.2	9.5	• Price • Cost • Interaction of price & cost	…
8-4-20-5	≅ 33000	5.3	8.3	11.1	• Price • Cost • Interaction of price & cost	$CV_{Price} = 0.30 \Rightarrow VSS: [10; 20\%)$

5.2.4. General Conclusions

Based on Table 34, the following general inferences can be made for these large-sized problems:

- For all of the cases, the variability of price and cost, and their interaction are the most significant factors on VSS, as in most of the cases in Section 5.1.

- VSS increases when the variability of the price and the cost increases, and it takes its maximum value when the variability of both factors, i.e. price and cost, is higher, as in most of the cases (5 of 6) in Section 5.1. Therefore, it can be said that the possible gain obtained from stochastic programming approach increases when the uncertainty of the price and cost parameters increases in general. Besides, in most of the problems, the variability of demand and cannibalisation rate is very slightly significant on VSS.

- On the average, while VSS takes values between 6.2 % and 10.6 %, EVPI is very near to zero. This result tells if a decision maker benefits from a stochastic approach, the expected profit may increase by a rate of between 6.2 % – 10.6 % depending upon the problem. However, perfect knowledge of the future will not improve the expected profit, thereby more information should not be dug in these cases, as in most of the cases in Section 5.1.

- The maximum solution time which is for the largest problem (8-4-20-5) considered in the problem context is nearly 33000 seconds (9.17 hours). This can be seen as reasonable for these kinds of strategic problems. Furthermore, the magnitude of the variability of uncertain parameters doesn't make any difference on the solution time, i.e. the problems are insensitive to the magnitude of the variability in terms of solution time.

- Based on the rules extracted by association analysis through Random Forest application, it can be seen that the variability of price has the huge impact on the stochastic solution performance. If it has high variability, VSS takes a value between 10 % - 20% and if it has low variability, VSS takes a value less than 10 %.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

In this study, a multi-period and multi-market product-mix problem, which involves timing decisions of new planned products of a firm and capacity expansion decisions, and takes product interdependencies and uncertainties associated to problem parameters into account, is handled. To the best of our knowledge, there is no prior work that deals with such a problem in the related literature. Besides, this study is the first attempt to optimise (balance) product roadmaps (strategic product plans) of a firm using operational research tools in a holistic way.

This problem is formulated as a two-stage stochastic programming model with recourse, in which the first-stage decisions, i.e. new product launching and capacity expansion decisions, are taken before the actual realization of uncertain parameters, i.e. demand, price, cost and cannibalisation rate. Then the second stage decisions, i.e. production volumes (sales targets) and capacity allocated for each product-line, are taken in later stages (periods) based on the information obtained from the first-stage decisions after the uncertainties are revealed. In order to solve this two-stage stochastic programming model efficiently, first, L-shaped method with multi-cuts are employed. Then, a sample average approximation approach based on Monte Carlo bounding technique is developed to solve the model under the assumption of having a large unmanageable set of scenarios or that the random parameters have continuous probability distributions (infinite number of scenarios).

The developed model as well as the solution approach is tested on different cases through two experimental studies in order to understand which problem parameters are significant on the solution obtained, make inferences about the solution capacity of the model and the solution approach developed in terms of solution time, explore if using stochastic programming approach provides an advantage based on the value of stochastic solution (VSS) and deduce the potential worth of more accurate forecasts based on the expected value of perfect information (EVPI).

Some of the following most important results are obtained through two experimental studies:

- All problems are solved using 30 number of batches for the sample average approximation approach with higher number of sample size (scenarios), except from Case 9 wherein the solution is obtained with 40 batches and a sample set of 1000 scenarios. Actually, the solution time highly depends on this number of batches and the size of sample taken for approximate problems. The solution time is very small, between 1800 and 3600 seconds, for the small-sized problems in Set 1 without using any parallel processor. However, when the problem increases as the cases in Set 2, the solution time dramatically increases. However, since the batch-means approach allows solving the problems in each batch separately, parallel processors are used to solve a certain number of batches simultaneously. Thus, the largest problem consisting of 8 existing lines, 4 new (candidate) lines, 20 markets and 5 periods, could be solved in 9.17 hours. This solution time is acceptable for such a strategic and complex problem. Even though this problem may be rarely seen in real-life because of the product aggregation level, i.e. product-line, considered in this study, the solution time would be around 45.83 hours without using parallel processors to find a satisfactory approximate solution, which is also acceptable for a large problem. However, one can concern about strategies for accelerating the solution in order to reduce total solution time. For that purpose, L-shaped method could be accelerated (e.g. Bidhandi and Patrick, 2017) or an optimal sample size, which may be less than the size given in this study (i.e. 500, 1000, 5000 scenarios), could be calculated at the beginning of the Monte Carlo sampling step.

- Considering deterministic parameters in Experiment 1, unit capacity usage, maximum allowable capacity (MAC) in each period, relative profitability of the new and old products, the interaction of profitability and unit capacity usage, of profitability and MAC, and of unit capacity usage and MAC are significant on VSS.

- VSS tends to be high providing that new products are profitable the same as or less than existing products and unit capacity usage of new products are the same as or less than of existing products. Therefore, it can be stated that competition between the new and old products for sharing the limited capacity and for contributing to the expected net profit increases, VSS also tends to increase. In those cases, decision makers should use stochastic programming approach in order to get higher expected profit compared to the deterministic approach.

- Relative profitability of new products compared to existing products, unit capacity usage of them and maximum allowable capacity in each period, and the interaction of profitability and unit capacity usage, and of unit capacity usage and MAC are significant factors on EVPI.

- Considering uncertain parameters in Experiment 2, the variability of price, cost and the interaction of those factors are the most significant factors on VSS. If price has a high variability or cost has a high variability but price variability is at medium levels, VSS is expected to be more than 20 %. Therefore, stochastic programming is recommended for the price-volatile markets as well as cost-sensitive production environment in order to have a high expected profit.

- During this thesis work, both the research on literature and interviews show that the product level as well as planning level considered in any product-mix problem affect the main properties of the problem as well as the mathematical model used for problem formulation. Therefore, clarifying the meaning of “product” in the scope of “product hierarchy” while referring to any product-mix problem and the fact that planning level at which the problem is formulated should be specified at the beginning of the problem definition step becomes crucial in such kind of studies.

- The solution generated by our model, i.e. expected profit obtained by the optimised product-line mix, can be used as a constraint on expected profit gained from the mix of model/SKU/variant under a product-line. In other words, the profit obtained through the optimal mix of two lines, e.g. one is refrigerators and the other is freezers (see Figure 7), can be separated for each line. Then for the problem of finding optimal mix of models/variants under each of the product lines, the expected profit obtained for each product-line can be added to the model as a constraint, as follows:

$$maximize \quad z = \text{Total profit of the optimal mix of variants/models}$$

$$s.t:$$
$$z \geq \text{Total profit of the mix of product} - \text{line obtained}$$

Based on those results given above, stochastic programming provides an advantage to handle parameter uncertainties, particularly uncertainty for selling price of products and for variable unit production cost, and to have more expected profit than mean-value solution approach. Thus, our modelling and solution approach can be useful tool for decisions makers to handle their strategic product-mix problems under uncertainty. However, there are some possible enhancements for our model based on experiences gained during this thesis and some extensions of the model and solution approach developed, as explained below.

Multi-stage and/or decision-dependent stochastic programming approach:

The product-mix problem in this study is handled as a two-stage stochastic programming model wherein it is assumed that the scenarios regarding each period are unrelated, i.e. a future realisation of uncertain parameters in a period (scenarios) does not affect the realisations of uncertain parameters in the previous stages, and scenarios and their associated probabilities are independent of the decisions taken. The first assumption might be relaxed in order to handle the problem in a dynamic planning environment and provide more flexibility than the two-stage stochastic models (Huang and Ahmet, 2005). Thus, the problem might be modelled as a multi-stage stochastic program in which decisions made at a certain period depend on events and decisions up to that period and the decisions are revised when more information about the uncertainties is revealed at each period. However, it is more difficult to solve these models because of their complexity due to the number of stages and the number of random parameters than the two-stage models (Huang, 2005; Solak, 2007). Therefore, many solution strategies are proposed for multi-stage problems in the literature (see e.g. Ahmed et al., 2003; Karabuk and Wu, 2003; Sahinidas, 2004; Huang, 2005; Huang and Ahmed, 2005; Solak, 2007; Ahmed, 2002; Claro and Sousa, 2012; Alaniazar, 2013; Lin et.al., 2014; Fattahi et al., 2017; Bertazzi and Maggioni, 2018).

The second assumption might also be relaxed for the cases in which optimisation decisions taken at any stages can influence the underlying stochastic process. Thus, the stochastic programming model can be extended to a decision-dependent stochastic programming approach (see, for instance, Zhan and Zheng, 2018 and Tarhan et al., 2009).

Incorporating fuzziness for some of the uncertain parameters

In the problem handled in this study, it is assumed that uncertain parameters such as demand, price and cost, associated with existing products are adequately represented by random variables due to having historical data about the performance of those products. Therefore, they can be modelled as random variables. However, when the uncertain parameters cannot be modelled by random variables because of lack of data, particularly for the new products, fuzzy numbers can be incorporated in order to model uncertain parameters. Thus, an alternative modelling and solution approach that considers fuzziness for new products and randomness for existing (old) products can be developed, as proposed by Hasuike and Ishii (2009a and 2009b). However, this approach may complicate the stochastic model further.

Different probability distributions to model randomness

All experimental results in this study are obtained under the assumption of that all random parameters are normally distributed, but different probability distributions might also be considered and the effect of them on the solution performance might be studied although Linderoth (2009) highlights that " … *solutions obtained from stochastic programs are often quite stable with respect to changes in the input probability distribution*".

Different demand-price models

In this study, for the sake simplicity and to avoid from any nonlinearities in demand constraints, which causes the problem difficult to solve, a linear function to model demand-price relationship is used. However, if this model does not comply with the problem, different kinds of models might be used. For instance, if demand is a function of price in an iso-elastic or exponential form, the demand constraints becomes nonlinear and this makes the problem harder to solve than the linear stochastic models.

Besides, it is also assumed that the demand for a new product is independent of existing products and other new products, since a firm under consideration producing long/medium life cycle products will not prefer introducing more than one new product-lines that will cannibalise each other within the same planning horizon and introducing a new product that will be cannibalised by an existing product. However, in real life there may be some cases in which the demand of a new product might be cannibalised by other new products if both of them are sold in market at the same time. If such a case occurs, the *demand constraints 2.2* in the deterministic mathematical model should be revised by multiplying at least two binary variables (e.g. two new lines which influence the demand of each other may be planned to launch) representing new product launching decisions in order to incorporate the interaction between those products. Thereby, the related demand constraints becomes nonlinear and this also makes the problem harder to solve than the linear stochastic models.

In order to handle this problem, nonlinear stochastic programming approaches and solution strategies are proposed in the related literature (see, for instance, Kulkarni and Shanbhag, 2012; Beraldi et al., 2009; Shastri and Diwekar, 2009).

Uncertainties regarding capacity-related parameters

In this study demand, price, cost and cannibalisation rate are handled as uncertain parameters; on the other hand, unit capacity usage and cost of one unit of additional capacity are considered as deterministic. There may be uncertainty regarding unit capacity usage because of production disruptions and unit capacity expansion cost whose mean value is used considering different capacity addition options, and their related costs and probability of occurrence. Therefore, these parameters can also be modelled as random parameters in the model, as studied in a few studies in the literature (Ren-qian, 2007; Karabuk and Wu, 2003; Hasuike and Ishii, 2009a; Hasuike and Ishii, 2009a; Hasuike and Ishii, 2009a; and Geng et al., 2009).

Incorporating risk measures into the stochastic model

If there exists uncertainty regarding any problem, there is also a decision risk that negatively influence the goal that is desired to achieve. In that context, one of the objectives in any problem under uncertainty is to determine some actions to reduce the risk to an acceptable level (Better et al., 2008). On the other hand, since our problem's objective is to maximise the expected profit, it can be stated that the decision-maker is neutral against risk. However, as Zhang and Wang (2009), and Eppen et al.(1998) highlighted, the decision-maker might also be concerned about the risk involved in a decision problem and therefore our model can also be revised in order to incorporate risk measures. In the related literature, for instance, Alonso-Ayuso et al. (2003) consider *value-at-risk* and *reaching probability* as commonly used risk measures and the objective function of their model is to maximise the expected benefit and probability of achieving a benefit target or maximise Value-at-Risk (VaR) for a given probability. Eppen et al. (1988) also take one of the risk measures, i.e. down-side risk, into account, which is calculated as the failure to meet the target profit specified by the decision-maker. Claro and Sousa (2012) consider another commonly used risk measure, i.e. the conditional-value-at-risk (CVaR), which is calculated as the expected value of the losses of improbably scenarios beyond the specified confidence level over a time horizon in their capacity expansion model under uncertainty. It is recommended to refer to Rockafellar (2007) for the details of the risk measures.

Considering our problem, some targets on the objective function value may be set, e.g. increasing the overall profit by 30 % for the next five years. Those goals can be incorporated into the model, which can be organized in order to maximise the probability (or minimise the risk) of reaching those targets.

Additional constraints:

In our model, only the physical constraints for maximum capacity available at each period is considered. Besides, financial constraints such as available budget for total capacity expansion over the planning horizon or a budget in each period, managerial constraints such as target values for profit and decision risk, and production constraints such as minimum production amounts (because of pre-taken orders or economies of scale) can be added to the model.

Make/Buy Decisions:

In this study, it is assumed that outsourcing decisions are made when the candidate set (existing and new product-lines) is developed, and thereby this set includes only the products to be produced in-house. However, the outsourcing option for some products can be considered and a decision on outsourcing when the expected demand is not satisfied using in-house resources or this option is more advantageous for some products compared to the cost of expanding the amount of in-house resources (i.e. capacity expansion) can also be incorporated into the model.

Capacity-related revisions:

The model developed in this study assumes that the capacity might be expanded to a maximum level continuously. However, in many real-world problems, it may not be feasible to decide on any amount of increments on capacity; instead it is preferred to increase capacity from a certain level to another (higher) feasible level. For instance, a production plant may be operated in three feasible capacity levels, e.g. Level-0 (currently, 100 000 unit/year), Level 1 (150 000 unit/year) and Level-2 (maximum: 220 000 unit/year). Therefore, the capacity of this plant could be expanded to Level 1 or Level 2 in a period. At this point, it is also possible to open a new plant if the existing one reaches its maximum level (Level-2) in a period. Since building a new plant is out of concern at the product and planning level considered in this study, the maximum capacity that will be available at a period might still be constrained by Level-2.

On the other hand, capacity expansion cost may also be in a piecewise form, i.e. the cost of expanding capacity might be constant until the next capacity level and it jumps to a new level after the plants is operated at a new capacity level.

Another assumption of our model is that capital investments are irreversible, i.e. the decisions for reducing capacity by renting or selling are not considered. However, it is also possible to incorporate those kind of decisions into the model when the current capacity is abundant for satisfying the demand in a period.

REFERENCES

Ahmed, S. (2002), “Semiconductor tool planning via multi-stage stochastic programming”, Proceedings of the international conference on modeling and analysis of semiconductor manufacturing (MASM 2002), Editors G. T. Mackulak, J. W. Fowler, and A. Schoemig, Tempe, AZ, April 10-12, 2002. 152-157.

Ahmed, S., King, A.J. and Parija, G. (2003), “A multi-stage stochastic integer programming approach for capacity expansion under uncertainty”, Journal of Global Optimisation, 26, 3-24.

Ahrens, R. (1983), “Basics of capacity planning and control”, BPICS Control, 9 (1), 27-30.

Alaniazar, S. (2013), “Demand modeling and capacity planning for innovative short life-cycle products”, Wayne State University, Dissertations, Paper 826.

Albright, R.E. and Nelson, B. (2004), “Product and technology mapping tools for planning and portfolio decision making”, In: The PDMA Toolbook 2 for New Product Development. Chapter 15. Eds.: Belliveau, P., Griffin, A., Somermeyer, S. Jonh Wiley&Sons, Inc.

Albright, R.E. (2002), “How to use roadmapping for global platform products”, PDMA Visions, 24 (4), 19-22.

Alonso-Ayuso, A., L.F. Escudero, A. Garin, M.T. Ortuno, G. Perez, (2005), “On the Product Selection and Plant Dimensioning Problem under Uncertainty”, Omega 33, 307-318.

Alonso-Ayuso A., Escudero L.F., Garin A., Ortuno M.T. and Perez G. (2003), “An approach for strategic supply chain planning under uncertainty based on stochastic 0-1 programming”, Journal of Global Optimisation, 26, 97-124.

Amaran, S., Sahinidis, N.V., Sharda, B. and Bury, S.J. (2016), "Simulation optimisation: A review of algorithms and applications", Annals of Operations Research, 240:351–380.

Anderson, S.P. and Celik, L. (2015), "Product line design", Journal of Economic Theory, 157, 517-526.

Azadivar, F. (1999), "Simulation optimisation methodologies", Proceedings of the 1999 Winter Simulation Conference.

Azevedo, P.J. and Jorge, A.M. (2007), "Comparing rule measures for predictive association rules", In: Kok J.N., Koronacki J., Mantaras R.L., Matwin S., Mladenič D., Skowron A. (eds) Machine Learning: ECML 2007, ECML 2007, Lecture Notes in Computer Science, vol 4701, Springer, Berlin, Heidelberg.

Bass, F. M. (1969), "New product growth for model consumer durables", Management Science, 15(5), 215-227.

Bastin F. (2004), "Trust-region algorithms for nonlinear stochastic programming and mixed logit models", PhD Dissertation, Facultés Universitaires Notre-Dame de la Paix, Belgium.

Bayraksan, G. and Morton, D.P. (2011), "A sequential sampling procedure for stochastic programming", Operations Research, 59(4), 898-913.

Belloni, A., Freund, R., Selove, M. and Simester, D. (2008), "Optimizing product line designs: efficient methods and comparisons", Management Science, 54(9), 1544-1552.

Benton, W.C. (2013), "Supply chain focused manufacturing planning and control", 1st edition, Cengage Learning, the USA.

Beraldi, P., Bruni, M.E. and Conforti, D. (2009), "A solution approach for two-stage stochastic nonlinear mixed integer programs", Algorithmic Operations Research, 4, 76-85.

Bertazzi, L. and Maggioni, F. (2018), “A stochastic multi-stage fixed charge transportation problem: Worst-case analysis of the rolling horizon approach”, European Journal of Operational Research, 267(2), 555-569.

Bertsimas, D., Brown, D.B. and Caramanis, C. (2011), “Theory and applications of robust optimization”, SIAM, 53(3), 464-501.

Bertsimas, D. and Mišić, V.V. (2017), “Robust product line design”, Operations Research, 65(1), 19-37.

Better, M., Glover, F., Kochenberger, G. and Wang, H. (2008), “Simulation optimisation: Applications in risk management”, International Journal of Information Technology and Decision Making, 7(4), 571-587.

Bidhandi, H.M. and Patrick, J. (2017), “Accelerated sample average approximation method for two-stage stochastic programming with binary first-stage variables”, Applied Mathematical Modelling, 41, 582-595.

Birge, J.R and Louveaux, F. (2011), “Introduction to Stochastic Programming”, Second edition, Springer Series in Operations Research and Financial Engineering, Springer-Sience+Business Media.

Bish, E.K. and Wang Q. (2004), “Optimal investment strategies for flexible resources, considering pricing and correlated demand”, Operations Research, 52 (6), 954-964.

Blau G.E., Pekny J.F., Varma V.A. and Bunch P.R. (2004), “Managing a portfolio of interdependent new product candidates in the pharmaceutical industry”, Journal of Innovation Management, 21, 227-245.

Breiman, L. (2001), “Random forests”, Machine Learning, 45 (1), 5–32.

Buffa, E.S. and Sarin, R.K. (1987), “Modern production/operations management”, 8th edition, Wiley, New York.

Carson Y. and Maria A. (1997), “Simulation optimisation: methods and applications”, Proceedings of the 1997 Winter Simulation Conference.

Chang, K.-H. (2016), “Risk-controlled Product mix planning in semiconductor manufacturing using simulation optimization”, IEEE Transactions on Semiconductor Manufacturing, 29(4), 411-418.

Chase, R.B., Aquilano, N.J. and Jacobs, F.R. (1998), “Production and operations management: manufacturing and services”, Boston, Mass.: Irwin/McGraw-Hill.

Chen, K.D., W.H. Haussman (2000), “Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis”, Management Science, 46 (2), 327-332.

Chen Z-L., Li S. and Tirupati D. (2002), “A scenario-based stochastic programming approach for technology and capacity planning”, Computers and Operations Research, 29, 781-806.

Chen H.H., Leeb, A.H.I. and Tong Y. (2006), “Analysis of new product mix selection at TFT-LCD technological conglomerate network under uncertainty”, Technovation (26), 1210-1221.

Chen, C.-K., Liao, C. and Weng, T.-C. (2007), “Optimal replenishment policies for the case of a demand function with product-life-cycle shape in a finite planning horizon”, Expert Systems with Applications, 32, 65-76.

Chen T-L., Lin, J.T. and Wu C-H. (2014), “Coordinated capacity planning in two-stage thin-film-transistor liquid-crystal-display (TFT-LCD) production networks”, Omega (42), 141-156.

Chien C-F. and Zheng J.-N. (2012), “Mini-max regret strategy for robust capacity expansion decisions in semiconductor industry”, Journal of Intell Manufacturing, 23, 2151-2159

Chou Y.-C., Cheng C.-T., Yang F.-C. and Liang Y.-Y. (2007), “Evaluating alternative capacity strategies in semiconductor manufacturing under uncertain demand and price scenarios”, International Journal of Production Economics, 105 , 591-606.

Claro J. and Sousa P. (2012), "A multi-objective metaheuristic for a mean-risk multistage capacity investment problem with process flexibility", Computers and Operations Research, 39, 838-849.

Collier, D.A. and Evans, J.R. (2009), "Operations management: Concepts, techniques, and applications", South Western/Cengage Learning India.

Dale, L. and Fujita, K.S. (2008), "An analysis of the price elasticity of demand for household appliances", Retrieved on 14 December 2014, from Energy Citations Database on EBSCOhost.

Dantzig, G.B. and Infanger, G. (2005), "A Probabilistic Lower Bound for Two-Stage Stochastic Programs", Chapter 2 in Stochastic Programming, Volume 150 of the series International Series in Operations Research & Management Science, 13-35.

Devinney T. M. and Stewart D. W. (1988), "Rethinking the Product Portfolio: A Generalized Investment Model", Management Science, 34(9), 1080-1095.

Eppen, G., Martin, K. and Schrage L. (1988), "A scenario approach to capacity planning, Operations Research", 37 (4), 517-527.

Fattahi, M., Govindan, K. and Keyvanshokooh, E. (2017), "A multi-stage stochastic program for supply chain network redesign problem with price-dependent uncertain demands", Computers & Operations Research, In Press.

Fruchter, G. E., A. Fligler and R.S. Winer (2006), "Optimal product line design: Genetic algorithm approach to mitigate cannibalization", Journal of Optimisation Theory and Applications, 131(2), 227-244.

Gabrel, V., Murat, C. and Thiele, A. (2014), "Recent advances in robust optimisation: An overview", European Journal of Operational Research, 235 (3), 471-483.

Geng N., Jiang Z. and Chen F. (2009), "Stochastic programming based capacity planning for semiconductor wafer fab with uncertain demand and capacity", European Journal of Operational Research, 198, 899-908.

Gorissen, B.L., Yanıkoğlua, I. and den Hertog, D. (2015), "A practical guide to robust optimization", Omega, 53, 124-137.

Govindarajan, M. (2007), "Marketing management: Concepts, cases, challenges and trends", 2nd Edition, Prentice Hall of India Private Limited.

Goyal S.K. and Giri B.C. (2001), "Recent trends in modeling of deteriorating inventory", European Journal of Operational Research, 134, 1-16.

Harrison J.M., Mieghem J.A.V. (1999), "Multi-resource investment strategies: Operational hedging under demand uncertainty", European Journal of Operational Research, 113, 17-29.

Hasuike, T. and H. Ishii (2009a), "A multi-criteria product mix problem considering multi-period and several uncertainty conditions", International Journal of Management Science and Engineering Management, 4 (1), 60-71.

Hasuike, T. and H. Ishii (2009b), "Product mix problems considering several probabilistic conditions and flexibility of constraints", Computers & Industrial Engineering 56, 918-936.

Huang, K. (2005), "Multi-stage stochastic programming models in production planning", PhD Thesis, submitted to School of Industrial and Systems Engineering, Georgia Institute of Technology.

Huang K. and Ahmed S. (2009), "The value of multi-stage stochastic programming in capacity planning under uncertainty", Operations Research, 57(4)), 893–904.

Huang J., Leng M. and Parlar, M. (2013), "Demand functions in decision modeling: A comprehensive survey and research directions", Decision Sciences, 44(3), 557-609.

Huh, W.T. and Roundy, R.O. (2005), "A continuous-time strategic capacity planning model", Naval Research Logistics (52), 329-343.

Infanger, G. (1992), "Monte Carlo (importance) sampling within a benders decomposition algorithm for stochastic linear programs", Annals of Operations Research, 39, 69-95.

Infanger, G. (1997), “DECIS user guide: A system for solving large-scale stochastic programs”, Retrieved on 16 Jan 2016, from https://www.gams.com/24.8/docs/solvers/decis/index.html.

Jiao, J. and Zhang, Y. (2005), “Product portfolio planning with customer-engineering interaction”, IIE Transactions. 37, 801-814.

Kalvelagen, E. (2003), “Benders decomposition for stochastic programming with GAMS”, Retrieved on 14 Jan 2016, from http://www.amsterdamoptimisation.com/pdf/stochbenders.pdf.

Karabuk S. and Wu S.D. (2003), “Coordinating strategic capacity planning in the semiconductor industry”, Operations Research, 51 (6), 839-849.

Kaut, M. and Wallace, S. (2007), “Evaluation of Scenario-Generation Methods for Stochastic Programming”, Pacific Journal of Optimisation, 3(2), 257–271.

Kellerer, H., Pferschy, U. and Pisinger, D. (2004), “Knapsack problems”, Springer, Berlin Heidelberg New York.

Kerin R.A., Harvey, M.G. and Rothe, J.T. (1978), “Cannibalism and new product development”, Business Horizons, 21(5), 25-31

Kim, S., Pasupathy, R. and Henderson, S.G. (2015), “A guide to sample-average approximation”, International Series in Operations Research and Management Science, 216, 207-243.

Kleywegt, A.J., Shapiro, A. and Homem-De-Mello, T. (2001), “The sample average approximation method for stochastic discrete optimization”, SIAM Journal on Optimisation, 12(2), 479-502.

Kohli, R. and Sukumar, R. (1990), “Heuristics for product-line design using conjoint analysis”, Management Science, 36(12), 1464-1478.

Kotler, P. and Keller, K.L., (2012), “Marketing Management”, 14th Edition, Pearson Prentice Hall Upper Saddle River, New Jersey, USA.

Kulkarni, A.A. and Shanbhag, U.V. (2012), “Recourse-based stochastic nonlinear programming: Properties and Benders-SQP algorithms”, Computational Optimization and Applications, 51(1), 77-123.

Küttner, R. (2004), “Optimal planning of product mix for subcontracting companies”, 4th International DAAAM Conference, Industrial Engineering–Innovation as Competitive Edge for SME, 29-30 April 2004, Tallinn, Estonia.

Laruccia, M., Tristão, J. & Loureiro, S.M.C. (2012), “A conceptual approach for cannibalism between goods”, Chinese Business Review, 11 (11), 989-995.

Letmathe P. and Balakrishnan N. (2005), “Environmental considerations on the optimal product mix”, European Journal of Operational Research, 167, 398-412.

Lea, B.-R. (2007), “Management accounting in ERP integrated MRP and TOC environments”, Industrial Management & Data Systems, 107 (8), 1188–1211.

Lee J., Cho Y., Lee J.-D. and Lee C.-Y. (2006), “Forecasting future demand for large-screen television sets using conjoint analysis with diffusion model”, Technological Forecasting & Social Change, 73, 362 – 376.

Li, H. and Azarm, S. (2002), “An approach for product line design selection under uncertainty and competition”, Journal of Mechanical Design, 124, 385-392.

Li S. and Tirupati D. (1994), “Dynamic capacity expansion problem with multiple products: technology selection and timing of capacity additions”, Operations Research, 42 (5), 958-976.

Lilien G.L., Kotler, P. and Moorthy, K.S. (1992), “Marketing models”, Prentice Hall International, Inc. New Jersey, USA.

Lin J.T., Chen T. and Chen W. (2007), “Capacity and product mix planning problem for TFT array multi-plant”, Journal of Chinese Institute of Industrial Engineers, 24 (6), 489-504.

Lin, J.T., Chen, T.-L. and Chu, H.-C. (2014), “A stochastic dynamic programming approach for multi-site capacity planning in TFT-LCD manufacturing under demand uncertainty”, Int. Journal of Production Economics, 148, 1–36.

Linderoth, J. (2009), “Models and algorithms for stochastic programming”, presented in Enterprise-Wide Optimisation Meeting held in Carnegie-Mellon University.

Linderoth, J., Shapiro, A. and Wright, S. (2002), “The empirical behavior of sampling methods for stochastic programming”, Optimisation Technical Report 02-01, Retrieved on 5 June 2018, from http://pages.cs.wisc.edu/~swright/stochastic/sampling/lsw.pdf.

Liu, X., Du, G., Jiao, R.J. and Xia, Y. (2017), “Product line design considering competition by bilevel optimisation of a stackelberg-Nash game”, IISE Transactions, 49(8), 768-780.

Lodwick, W.A. and Kacprzyk, J. (2010), “Fuzzy optimisation: Recent advances and applications”, Springer-Verlag, Berlin–Heidelberg.

Luhandjula, M.K. (2015), “Fuzzy optimisation: Milestones and perspectives, Fuzzy Sets and Systems”, 274, 4-11.

Luo, L. (2011), “Product line design for consumer durables: An integrated marketing and engineering approach”, Journal of Marketing Research, 48(1), 128-139.

Mak, W-K. Morton, D.P. and Wood, R.K. (1999), “Monte Carlo bounding techniques for determining solution quality in stochastic programs”, Operations Research Letters, 24, 47–56.

Mashayekhi, M., and Gras, R. (2015), “Rule extraction from random forest: the RF+HC Methods”, 28th Canadian Conference on Artificial Intelligence, Canadian AI 2015, Halifax, Nova Scotia, Canada, June 2-5, 2015, Proceedings. Cham: Springer International Publishing.

Miller, T. (2002), “Hierarchical operations and supply chain planning”, Springer-Verlag London.

Mishra, B.K., Prasad, A., Srinivasan, D. and ElHafsi, M. (2017), “Pricing and capacity planning for product-line expansion and reduction”, International Journal of Production Research, 55:18, 5502-5519.

Mitra, S. (2006), “A white paper on scenario generation for stochastic programming”, Retrieved on 5 June, 2018, from http://optirisk-systems.com/papers/opt004.pdf.

Mohit, G., Yash, D. and Tiwari, M.K. (2017), “An integrated framework for product line design for modular products: product attribute and functionality-driven perspective”, International Journal of Production Research, 55(13), 3862-3885.

Monroe, K. B.,S. Sundar, W. A. Wells and A. A. Zoltners, (1976), “A multi-product integer programming approach to the product mix problem”, Proceedings of the 1976 AMA Meeting, Bernhardt, K.(ed.), American Marketing Association, Chicago, IL, 493-497.

Montgomery, D. C. (2009), “Design and analysis of experiments”, Seventh Ed., John Wiley and Sons Inc.

Morgan, L. O., Daniels, R.L. and Kouvelis, P. (2001), “Marketing/manufacturing trade-offs in product line management”, IIE Transactions 33, 949-962.

Morales, J.A.A. (2007), “Strategic development in the petrochemical industry”, PhD Thesis, The University of London.

Morgan L. O. and Daniels R.L. (2001), “Integrating product mix and technology adoption decisions: a portfolio approach for evaluating advanced technologies in the automobile industry”, Journal of Operations Management, 19, 219-238.

Müller, S. and Haase, K. (2016), “On the product portfolio planning problem with customer-engineering interaction”, Operations Research Letters, 44(3), 390-393.

Ólafsson, S. and Kim, J. (2002), “Simulation optimization”, Proceedings of the 2002 Winter Simulation Conference.

Oum H. T. (1989), "Alternative demand models and their elasticity estimates", Journal of Transport Economics and Policy, 23 (2), 163-187.

Parker P.M. and Neelamegham R. (1997), "Price elasticity dynamics over the product life cycle: A study of consumer durables", Marketing Letters, 8(2), 205-216.

Prabhakar, V.J. and Sandborn, P. (2013), "A model for comparing sourcing strategies for parts in long life cycle products subject to long-term supply chain disruptions", International Journal of Product Lifecycle Management, 6 (3), 228-249.

Prekopa, A. (1995). "Stochastic programming", Dordrecht: Kluwer Academic.

Ren-qian Z. (2007), "Research on capacity planning under stochastic production and uncertain demand", Systems Engineering-Theory and Practice, 27(1), 51-59.

Ricci, T. (2012), "Product planning: A key to entrepreneurial success", Retrieved on 20 May 2018, from https://www.asme.org/engineering-topics/articles/product-planning/product-planning-a-key-to-entrepreneurial-success.

Roberts P. W. and McEvily S., (2005), "Product line extension and resource cannibalism", Journal of Economic Behavior and Organization, 57, 49-70.

Rockafellar, R.T. (2007), "Coherent approaches to risk in optimisation under uncertainty", Tutorials in Operations Research, Retrieved from https://pubsonline.informs.org/doi/abs/10.1287/educ.1073.0032, on June 12, 2018.

Rockafellar, R.T. (2001), "Optimisation under uncertainty", Lecture Notes, University of Washington.

Sahinidas, N.V. (2004), "Optimisation under uncertainty: The state-of-the-art and opportunities", Computers and Chemical Engineering, 28, 971-983.

Schön, C. (2010), “On the product line selection problem under attraction choice models of consumer behavior”, European Journal of Operational Research, 206, 260-264.

Shah J. and Avittathur (2007), “The retailer multi-item inventory problem with demand cannibalization and substitution”, International Journal of Production Economics, 106, 104-114.

Shapiro, A. and Homem-de-Mello, Tito (1998), “A simulation-based approach to two-stage stochastic programming with recourse”, Mathematical Programming, 81, 301-325.

Shastri, Y. and Diwekar, U. (2009), “An efficient algorithm for large-scale stochastic nonlinear programming problems”, Computers and Chemical Engineering, 30, 864–877.

Sobreiro, V.A., Mariano, ED.B. and Nagano, M.S. (2014), “Product mix: The approach of throughput per day”, Production Planning & Control, 25(12), 1015-1027.

Solak, S. (2007), “Efficient solution procedures for multistage stochastic formulations of two problem classes”, PhD Thesis, submitted to School of Industrial and Systems Engineering, Georgia Institute of Technology.

Stafford, M. (1997), “A product-mix capacity planning model”, Technical Report, ORP97-03, Graduate Program in Operations Research Technical Report Series, The University of Texas at Austin, available at http://www.me.utexas.edu/~orie/techrep.html.

Stuart, J. A., J. C. Ammons, L. J. Turbini (1999), “A product and process selection model with multidisciplinary environmental considerations”, Operations Research 47 (2), 221-234.

Tai, Y.T., Pearn, W.L. and Chen, C.C. (2015), P”roduct mix determination and capacity allocation for heterogeneous products in thin film transistor liquid crystal display manufacturing”, Computers & Industrial Engineering, 80, 201–211.

Tan, P.-N., Steinbach, M. and Kumar, V. (2006), “Introduction to data mining”, Pearson Education Inc.

Tang, J.F., Wang, D.W, Fung, Y.K. and Yung, K.L. (2004), “Understanding of fuzzy optimisation: theories and methods”, Journal of Systems Science and Complexity, 17(1), 117–136.

Tarhan, B., Grossman, I.E. and Goel, V (2009), “Stochastic programming approach for the planning of offshore oil or gas field infrastructure under decision-dependent uncertainty”, Industrial and Engineering Chemistry Research, 48(6), 3078-3097.

Thomas B.G. and Bollapragada, S. (2010), “General Electric uses an integrated framework for product costing, demand forecasting, and capacity planning of new PV technology products”, Interfaces, 40(5), 353-367.

Thomson, B. (2014), “Creating a strategic product plan”, http://www.pragmaticmarketing.com//resources/creating-a-strategic-product-plan?p=0 accessed at 10.05.2014.

Tønnesen C.R. and Øveraas S.T, (2012), “Production optimisation in the salmon farming industry”, MS Thesis in Norwegian University of Science and Technology.

Tucker, C.S. and Kim, H.M. (2008), “Optimal product portfolio formulation by merging predictive data mining with multilevel optimization”, Journal of Mechanical Design, 130(4), 041103-1-15.

Urban, G. L. (1969), “A mathematical modeling approach to product line decisions”, Journal of Marketing Research 6 (1), 40-47.

Vespucci, M.T., Bertocchi, M., Pisciella, P. and Zigrino, S. (2016), “Two-stage stochastic mixed integer optimisation models for power generation capacity expansion with risk measures”, Optimisation Methods & Software, 31(2), 305-327.

Wang, J. (2008), “Data warehousing and mining: Concepts, methodologies, tools and application”, Information Science Reference, Newyork.

Wang, F.K. Du, T. and Wen, F.C. (2007), “Product mix in the TFT-LCD industry”, Production Planning & Control, 18(7), 584–591.

Wang, T., Wang, S. and Meng, Q. (2017), “Chance-constrained programming”, Chapter four in Liner Ship Fleet Planning Models and Algorithms, 2017, 49–59.

Yayla-Küllü, H.M. (2011), “Capacity investments and product mix choice under economic uncertainty”, 2011 MSOM Annual Conference, Michigan.

Yılmaz G., Lusa A. and Benedito E. (2013), “Long-term capacity planning with products’ renewal”, 7th International Conference on Industrial Engineering and Industrial Management.

Zhan, Y., and Zheng, Q.P. (2018), “A multistage decision-dependent stochastic bilevel programming approach for power generation investment expansion planning”, IISE Transactions, 50(8), 720-734.

Zhang F., Roundy R., Çakanyıldırım M. and Huh W.T., (2004), “Optimal capacity expansion for multi-product, multi-machine manufacturing systems with stochastic demand”, IIE Transactions, 36, 23-36.

Zhang R. and Wang R. (2009), “Scenario-based stochastic capacity planning model and decision risk analysis”, Systems Engineering-Theory and Practice, 29(1), 55-63.

APPENDIX A

THE PRODUCT HIERARCHY FOR A RUBBER TYRE MANUFACTURER

A firm operating in rubber tyre manufacturing sector, which produces very high volumes of fully standardised products, is considered. It produces and sales rubber tyres for different type of vehicles. According to the product hierarchy developed by Kotler and Keller (2012), the firm has one product family and nine product classes, which are formed according to area of use. Under each type of product class, there are different lines and under each line there are specific models (variants).

The product hierarchy of this firm is displayed as follows:

Table A1. The product hierarchy for a rubber tyre manufacturer

Family	Class	Line	Model
Rubber tyre	1. Passenger Car	1.1. Winter 1.2. Summer 1.3. Four Seasons	Models with different sizes and patterns
	2. SUV / 4X4	2.1. Four Seasons 2.2. Winter 2.3. Summer	Models with different sizes and patterns
	3. Agricultural	3.1. Floatation 3.2. Forestry 3.3. Tractor-front 3.4. Tractor Rear 3.5. Tractor Radial	Models with different sizes and patterns
	4. Industrial Tires	4.1. Grader 4.2. Loader&Earthmover 4.3. Compactor 4.4. Implement 4.5. Excavator	Models with different sizes and patterns
	5. Light Truck	5.1. Four Seasons 5.2. Winter 5.3. Summer	Models with different sizes and patterns
	6. Truck/Bus	5.1. Radial 5.2. Bias	Models with different sizes and patterns
	7. Military Aircraft	7.1. Main 7.2. Nose	Models with different sizes and patterns

APPENDIX B

DEMAND MODELS

In marketing and operations management literature, a variety of mathematical functions is developed in order to model demand. Based on the problem environment and the variables (independent) significant on demand (dependent variable), different kinds of models can be defined. For instance, it can be considered that the sales of a firm may depend on its and competitors' advertising expenditures, its and competitors' prices, its product quality, consumer utilities etc., and thereby those factors should be incorporated into any model which is used to forecast the future demand for this firm's products. As another example, a firm's sales may be independent of competitors' decisions on advertising, pricing or other marketing efforts, but it can be modeled in terms of product's life cycle that shapes in the market, and this shape should be followed in order to forecast the future sales. Considering those examples, two main groups regarding demand models can be specified:

- **Group A:** Models used to describe demand which is defined in a market where customers are sensitive to firms' operational and marketing activities, thereby decisions on price, rebate, lead-time, shelf-space, product quality and advertising, or the integration of more than one of those activities (e.g., both price and advertising expenditures)
- **Group B:** Models that are independent of a firm's decision, e.g. a model in which demand is formulated in accordance with product's life cycle.

Models in Group A:

In this group, there are different models which characterize the relation between the independent variables and market response variable. Those models may be in

deterministic form in which there is no probabilistic elements because of having the full knowledge about demand process or *stochastic form* in which the parameters in the model are handled as uncertain (Lilien et al., 1992; Alonso-Ayuso et al., 2003; Huang et al., 2013).

They may be in *static form* or *dynamic form* in which intertemporal effects, e.g. demand (expected sales) in a period may be influenced by the sales of previous period as in diffusion models (Bass, 1969), or dynamic behavior of elasticities over the product life-cycle, e.g. the price elasticity of demand declines over time in automobile industry and increases over time in white goods industry (Dale and Fujita, 2008) and price is not so sensitive during early phases of life cycle and becomes very sensitive during maturity or decline phase (Parker, 1997), are incorporated.

Considering the *mathematical form of the models* in this group, there may be defined three main subgroups:

- Single-firm single-variable models without competition (the most common: price-dependent models)
- Multi-firm single-variable models under competition (the most common: price-dependent models)
- Multi-variable demand models

Single-firm single-variable demand models (the most common: price-dependent models)

This group covers different types of mathematical functions considering one independent variable, such as price, advertising expenditure, rebate, lead-time, space, quality and so on. Those functions has also different characteristics according to whether there is a competition in the market, e.g. price-competition among the rivals. For both of the case the following mathematical forms are widely used to model demand and the factors on which demand is dependent, e.g. price:

- *Simple Linear Model* in which the shape of the function is linear and the nonlinear effects arising in the market are not taken into consideration.
- *Power Series Model* in which the shape of the function is various such as cubic, quadratic etc., e.g. a price-dependent model may include the curvature effect of price on demand through adding a term of the square of price.
- *Iso-elastic Models*, as a special case of a fractional root model, in which price elasticity of demand is constant everywhere in a price-dependent model (Oum, 1989).
- *Semilog Model* in which the shape is concave and handle situations, for instance, that consider a threshold value for the marketing effort; i.e. total marketing expenditure (as a factor on which the sales is dependent) need to be exceeded a threshold value before it affects the sales.
- *Exponential Model* in which the shape is convex and handle situations where there an increasing return to scale, e.g. in a price-dependent model there may be increasing returns to price decrements (Lilien et al., 1992).
- *Logistic Model* which has an S-shape, e.g. the sales of a firm may have an inflection point where the S-curve transforms from convex to concave when the independent variable price changes, and is widely used to represent the relationship between time phases and sales over a product-life cycle.

In addition to those functional forms, there are some other forms in the literature, e.g. *Gompertz model, Adbudg model, log-reciprocal, modified exponential* etc. For instance, one can refer to Huang et al. (2013) for other types of mathematical forms of price-dependent demand models.

Multi-firm single-variable demand models (the most common: price-dependent models):

In those models, since there is a competition among the firms in market, if the marketing effort of a firm changes, the demand of other firm(s) is affected by this change. For instance, in the price-dependent models that are the most widely used, there is a price competition in the market where a price increase by one firm reduces the demand for its products, but increases the demand for competitors' products.

There are two different cases in those models:

(1) Demand models for homogenous products (perfectly competitive market) in which all products sold by different companies are identical and thereby substitutable, and each firm is price-taker (the mathematical functions developed for single-firm single-variable case can be used for this kind of models).

(2) Demand models for differentiated products in which the products sold by different firms are not identical and thereby not totally substitutable. Those models are used in the cases where there is price competition among the firms in the market, e.g. when a firm increases its price, this decreases its demand; but it would cause the demand of the other firms in the market to increase. In the related literature, *linear models* and *attraction models (market share model)* are commonly used in order to represent the demand-price relationship under competition. Among the attraction models, *multinomial logit (MNL)* and *multiplicative competitive interaction (MCI)* models are most-commonly used. Besides, *linear model, Hotelling model*, *Cobb-Douglas* (constant elasticity log-linear) model and *constant expenditure model* are also used to characterize the price-dependent in the case of two or more competing firms (for details; see Huang et al., 2013; Lilien et al., 1992).

Multi-variable demand models:

In addition to price which is an independent variable in a demand model, customer purchasing decisions may be influenced by the *promotions* offered by the firms, *lead-time* for delivery, *advertisements*, *quality of the products* and so on. Therefore, in addition to price-dependent models in which the price is the only independent variable in a demand model, some other models considering the promotion level, advertisement expenditures, lead-time, quality perception, which can be characterized by consumers utility functions etc. are developed in the related literature (for a detailed survey on those types of models, refer to Huang et al., 2013).

Models in Group B:

This group covers the models which are independent of a firm's decision, e.g. pricing decision, instead demand is formulated in accordance with product's life cycle (Chen et al., 2007; Goyal and Giri, 2001; Yilmaz et al, 2013). Some assumptions regarding those models can be given as follows:

- After launching a product to a market, total amount of product expected to be adopted by consumers throughout its life cycle can be estimated.
- The expected sales (demand) is formulated as a function of time-stages over the life cycle and parameters by which adoption rate is influenced.
- The pattern (e.g. S-shape: product sales increases exponentially or linearly in the introduction and growth phase on a product life cycle and then after an inflection point the sales starts to decrease in the maturity phase) which a firm follows over its own product's life cycle is the same as or similar to the pattern occurred in the market.

Considering those assumptions, it is noted that those models can be considered for the markets where there is only one seller (monopolistic case) or there are many sellers each of which knows its market share and the products for all sellers follow the same life cycle pattern. In literature, for instance, Chen et al. (2007) and Yilmaz et al. (2013) use a demand function which follows the product life cycle S-shape, which is derived from a Beta distribution represented by parameters such as total life of product, cumulative quantity of demand in the planning horizon and constant parameters.

A summary of the demand models in the related literature, given above, can be displayed as flow diagram in the following figure:

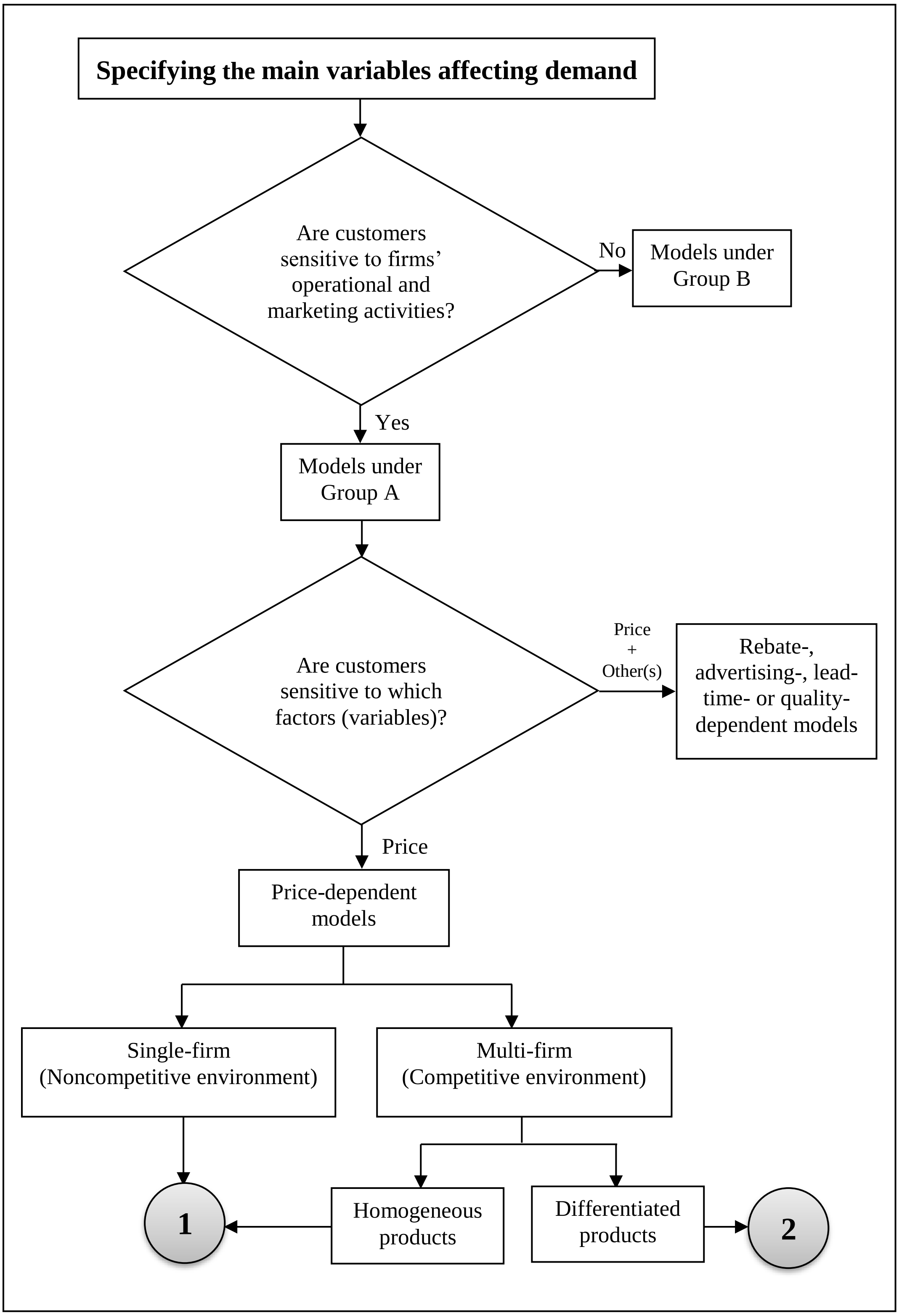


Figure B1. A summary of the demand models in the related literature (cont'd)

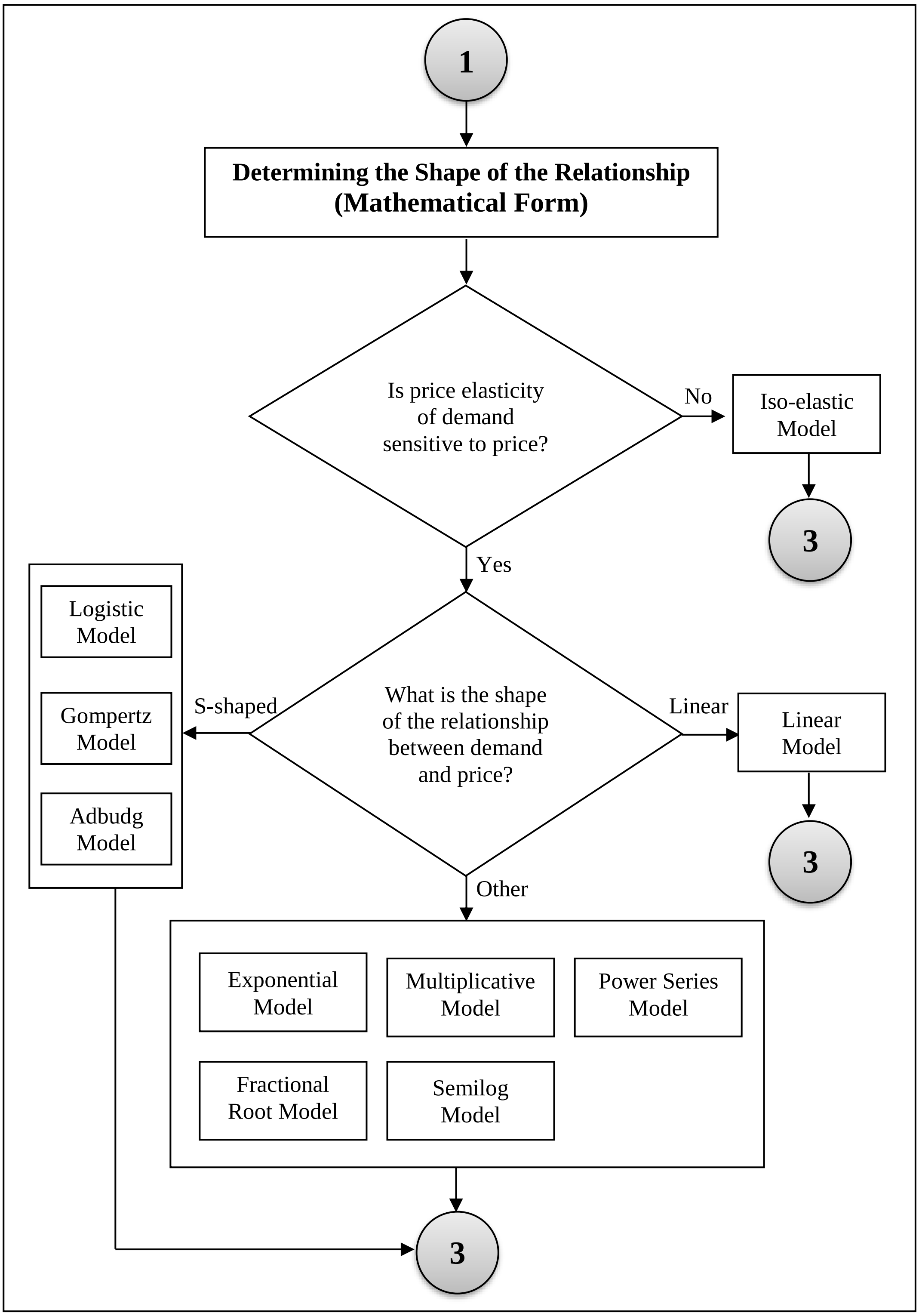


Figure B1. A summary of the demand models in the related literature (cont'd)

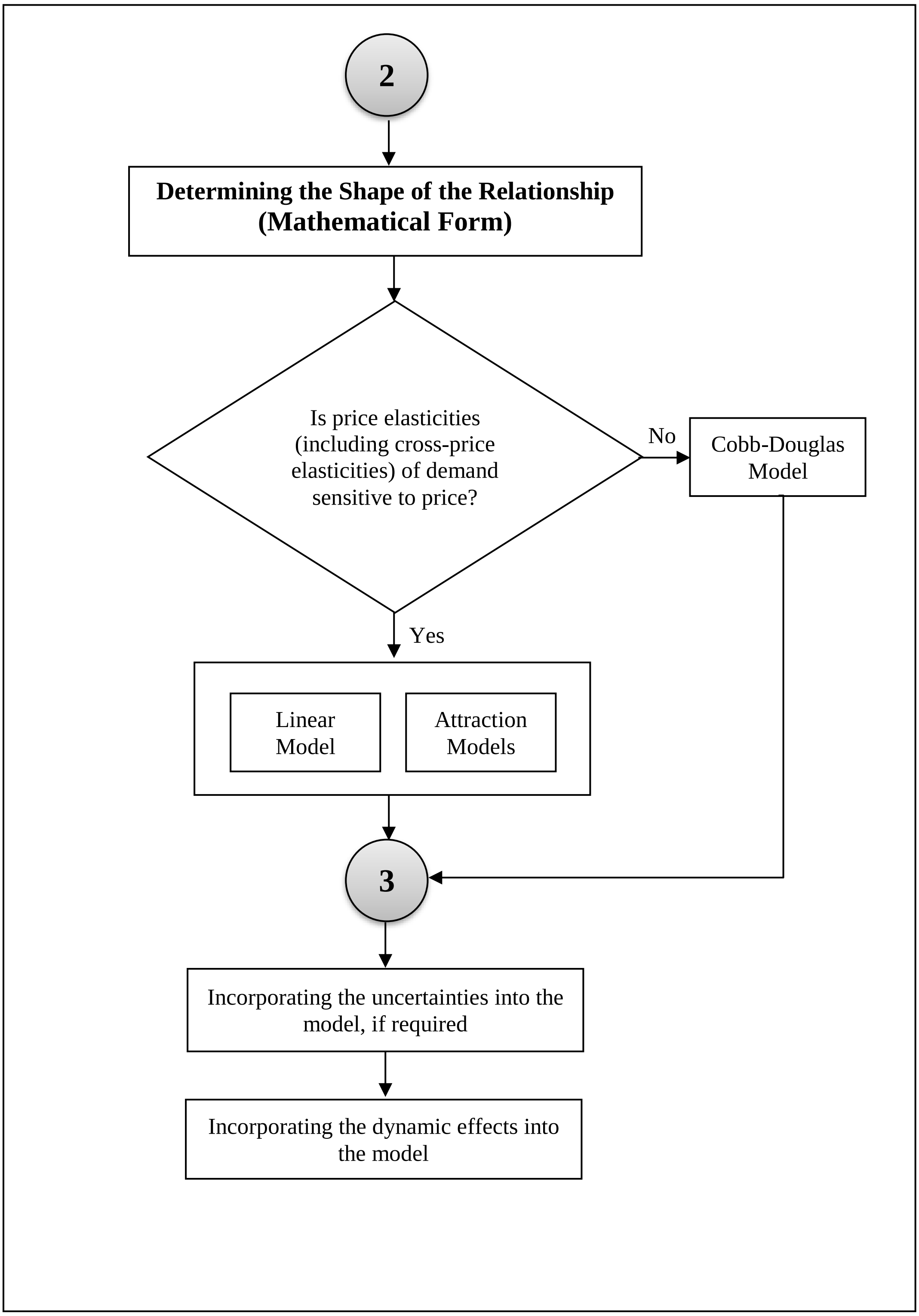


Figure B1. A summary of the demand models in the related literature (cont'd)

APPENDIX C

COMPUTATIONAL RESULTS

CASE 1

C.1.1. Experiment 1

C.1.1.1. Response: VSS

ANOVA Table for General Factorial Regression (Asin_VSS[32] versus factors)

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	107	3,82416	0,035740	32,87	0,000
Linear	11	0,67651	0,061501	56,56	0,000
Cap_Exp_Cost	2	0,00268	0,001340	1,23	0,298
Profitability	2	0,11521	0,057607	52,98	0,000
Unit_Cap_Usage	4	0,44222	0,110556	101,67	0,000
Capacity	3	0,11639	0,038797	35,68	0,000
2-Way Interactions	44	1,48846	0,033829	31,11	0,000
Cap_Exp_Cost*Profitability	4	0,01621	0,004053	3,73	0,008
Cap_Exp_Cost*Unit_Cap_Usage	8	0,00876	0,001095	1,01	0,439
Cap_Exp_Cost* Capacity	6	0,11216	0,018693	17,19	0,000
Profitability*Unit_Cap_Usage	8	0,74988	0,093735	86,20	0,000
Profitability* Capacity	6	0,15874	0,026456	24,33	0,000
Unit_Cap_Usage* Capacity	12	0,44273	0,036894	33,93	0,000
3-Way Interactions	52	1,65919	0,031907	29,34	0,000
Cap_Exp_Cost*Profitability*Unit_Cap_Usage	16	0,03575	0,002234	2,05	0,020
Cap_Exp_Cost*Profitability* Capacity	12	0,04896	0,004080	3,75	0,000
Profitability*Unit_Cap_Usage* Capacity	24	1,57448	0,065603	60,33	0,000
Error	72	0,07829	0,001087		
Total	179	3,90245			

[32] In data analysis we benefit from special transformations in order to satisfy the assumptions of the regression model. Since the response $r = VSS/EEV$ is a proportion, we replace "r" by "the angular transformation of r, i.e. $arcsin\sqrt{r}$. The transformed response is seen as Asin_VSS in ANOVA table and related figures.

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0329751	97,99 %	95,01%	87,46%

Residual Plots:

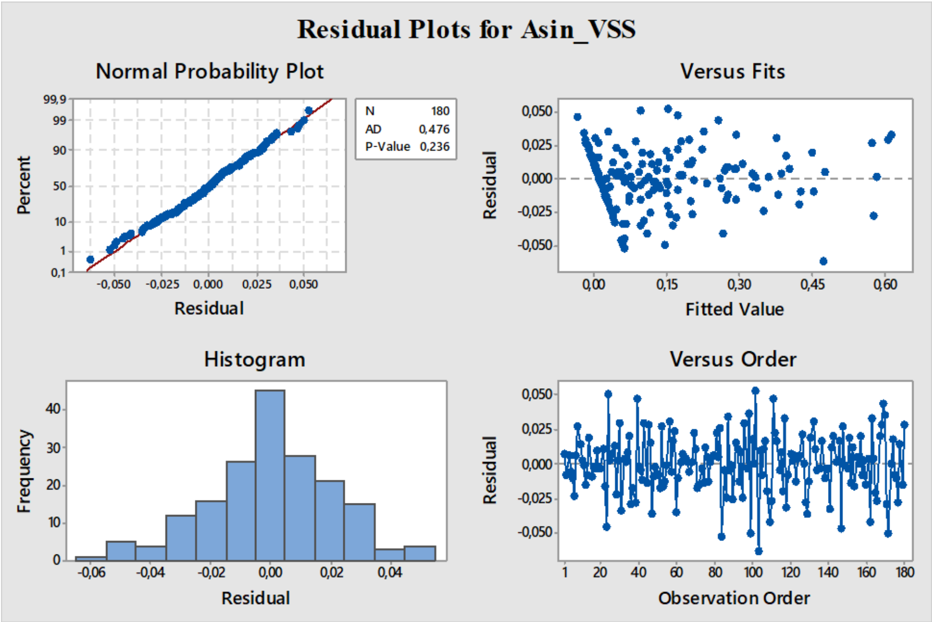


Figure C1. Residual plots for Case 1 (Deterministic Parameters: VSS)

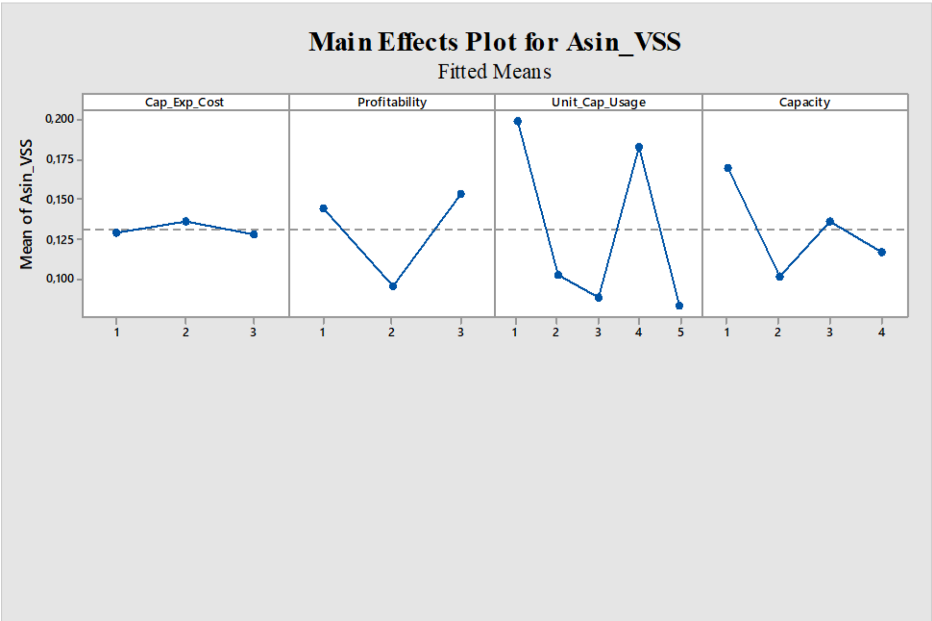


Figure C2. Main effects plot for Case 1 (Deterministic Parameters: VSS)

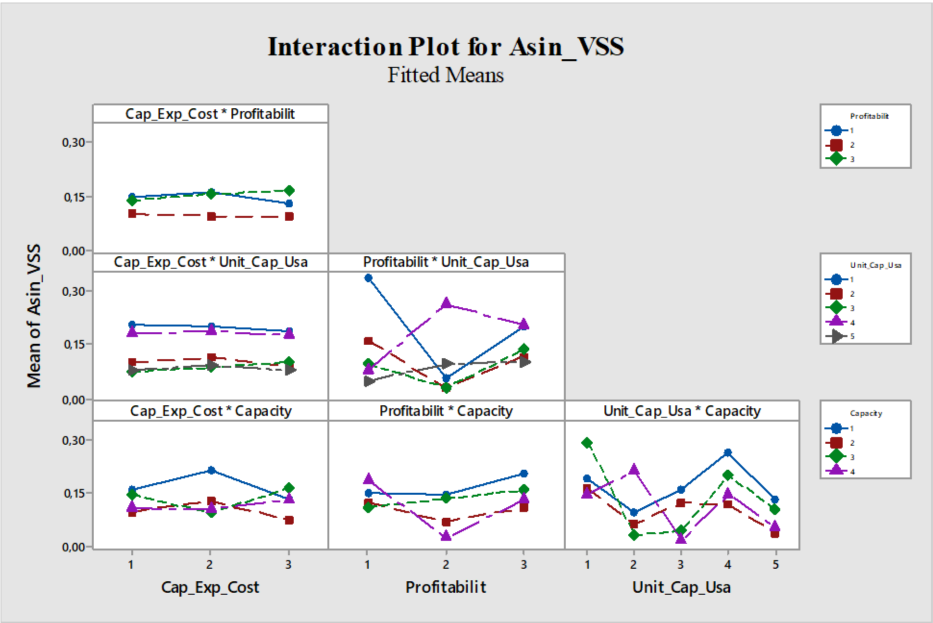


Figure C3. Interaction plot for Case 1 (Deterministic Parameters: VSS)

According to Figure C1, the assumptions of the regression model whose R-sqr (adjusted) is 95 % are satisfied.

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four deterministic parameters:

Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacity .f, data = veri temiz, mtry = 3,importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB[33] estimate of error rate: 12.22%

Confusion matrix[34]:

	H	L	M	class.error
H	2	5	0	0.71428571
L	0	154	3	0.01910828
M	0	14	2	0.87550000

[33] OOB stands for Out-of-Bag eroor and accuracy of the model = 1-OOB.
[34] Confusion matrix shows how many cases were guessed properly by the model fitted by Random Forest.

Importance of variables:

	H	L	M	MeanDecreaseAccuracy[35]	MeanDecreaseGini[36]
Cap_Exp_Cost.f	-10.17	-26.23	-14.69	-31.16	4.81
Profitability.f	21.94	16.37	21.68	27.03	11.18
Unit_Cap_Usage.f	26.50	18.81	30.81	33.47	13.59
Capacity.f	27.37	7.76	2.36	14.27	9.62

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

***R output for the revised model including three deterministic parameters*:**

Call:
randomForest(formula = VSS.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f, data = cleandata , mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 7.78%

Confusion matrix:

	H	L	M	class.error
H	7	0	0	0.00000000
L	0	153	4	0.02547771
M	2	8	6	0.62500000

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Profitability.f	37.09	50.71	43.04	73.62	12.19
Unit_Cap_Usage.f	41.42	50.84	55.62	77.26	14.23
Capacity.f	41.51	38.25	22.60	54.78	8.70

[35] Mean Decrease Accuracy shows how much the model fitted by Random Forest decreases when a variable is dropped.
[36] Mean Decrease Gini shows the explanatory power of the related variable in the model fitted.

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,112	1	1,432	X[,1] %in% c('2','3')	L
2	2	0,103	0,819	4,791	X[,1] %in% c('1') & X[,2] %in% c('1')	M
3	1	0,072	1	1,432	X[,2] %in% c('1','4')	L
4	2	0,072	1	1,432	X[,1] %in% c('3') & X[,3] %in% c('1')	L
5	2	0,067	1	1,432	X[,1] %in% c('1') & X[,2] %in% c('4')	L
6	1	0,063	1	1,432	X[,2] %in% c('2','3','5')	L
7	1	0,058	1	1,432	X[,2] %in% c('5')	L
8	2	0,054	1	1,432	X[,1] %in% c('2') & X[,2] %in% c('1')	L
9	1	0,051	1	1,432	X[,2] %in% c('3','5')	L
10	2	0,047	0,865	6,606	X[,1] %in% c('1') & X[,2] %in% c('2')	H
11	3	0,045	1	7,637	X[,1] %in% c('1') & X[,2] %in% c('2') & X[,3] %in% c('3', '4')	H
12	2	0,044	1	1,432	X[,1] %in% c('1') & X[,3] %in% c('3')	L
13	1	0,043	0,989	1,417	X[,1] %in% c('1','2')	L
14	3	0,043	1	7,637	X[,1] %in% c('3') & X[,2] %in% c('1') & X[,3] %in% c('3', '4')	H
15	1	0,042	0,999	1,431	X[,3] %in% c('1','2','3')	L
16	2	0,04	1	1,432	X[,1] %in% c('2','3') & X[,3] %in% c('4')	L

C.1.1.2. Response: EVPI

ANOVA Table for General Factorial Regression (Logit_EVPI[37] versus factors)

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Max_Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	51	151,618	2,9729	49,65	0,000
Linear	11	116,889	10,6263	177,46	0,000
Cap_Exp_Cost	2	3,242	1,6212	27,07	0,000
Profitability	2	7,017	3,5083	58,59	0,000

[37] In data analysis we benefit from special transformations in order to satisfy the assumptions of the regression model. Since the response r =*EVPI/RP* is a proportion, we replace "r" by "the logit transformation of r", i.e. $logit\ r = log\left(\frac{r}{1-r}\right)$. The transformed response is seen as Logit_EVPI in ANOVA table and related figures.

Unit_Cap_Usage	4	42,821	10,7052	178,78	0,000
Capacity	3	63,810	21,2699	355,21	0,000
2-Way Interactions	40	34,729	0,8682	14,50	0,000
Cap_Exp_Cost*Unit_Cap_Usage	8	1,820	0,2276	3,80	0,000
Cap_Exp_Cost*Capacity	6	9,221	1,5369	25,67	0,000
Profitability*Unit_Cap_Usage	8	13,198	1,6497	27,55	0,000
Profitability*Capacity	6	3,211	0,5352	8,94	0,000
Unit_Cap_Usage*Capacity	12	7,278	0,6065	10,13	0,000
Error	128	7,665	0,0599		
Total	179	159,283			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,244704	95,19%	93,27%	90,48%

Residual Plots

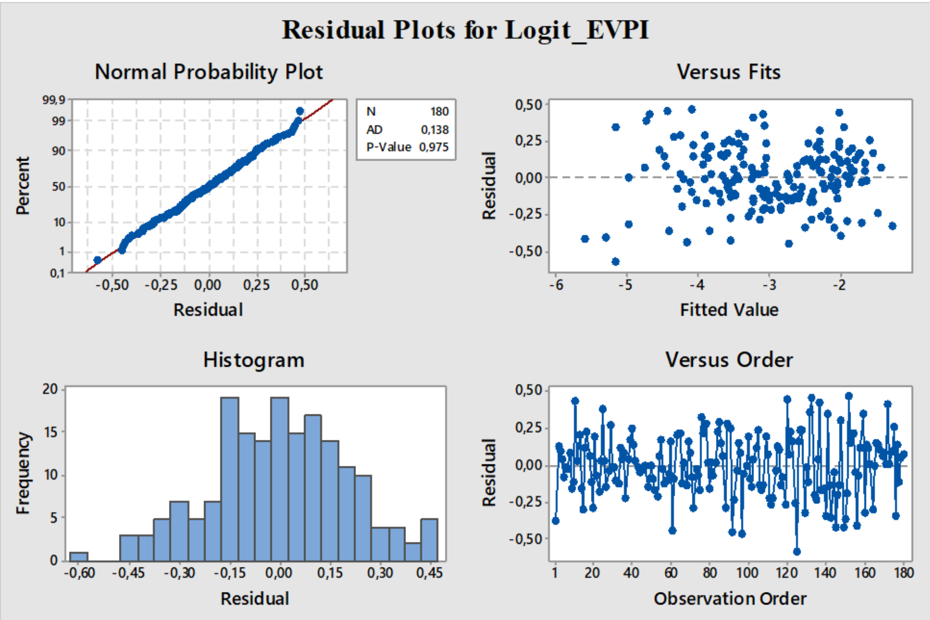


Figure C4. Residual plots for Case 1 (Deterministic Parameters: EVPI)

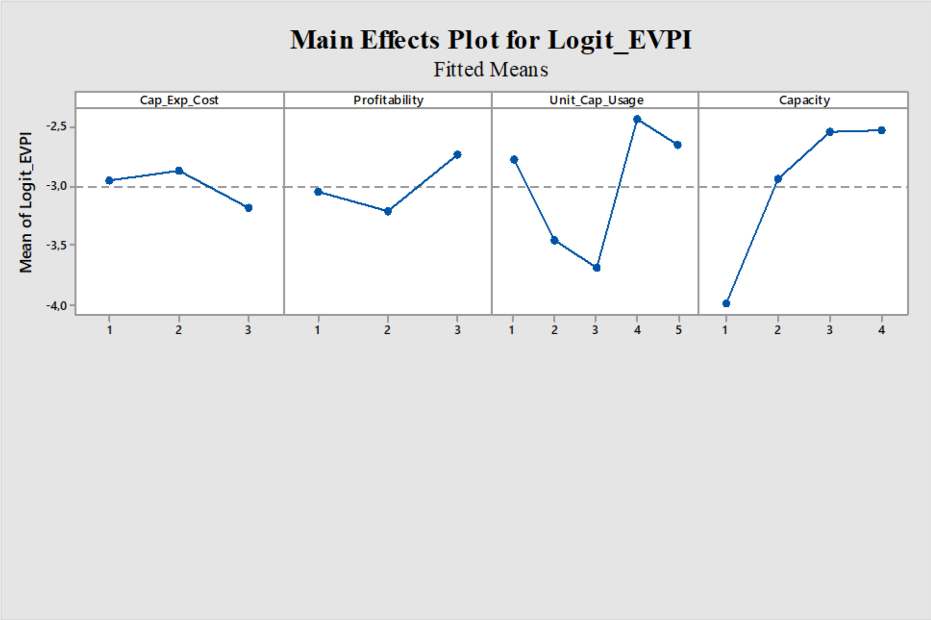


Figure C5. Main effects plot for Case 1 (Deterministic Parameters: EVPI)

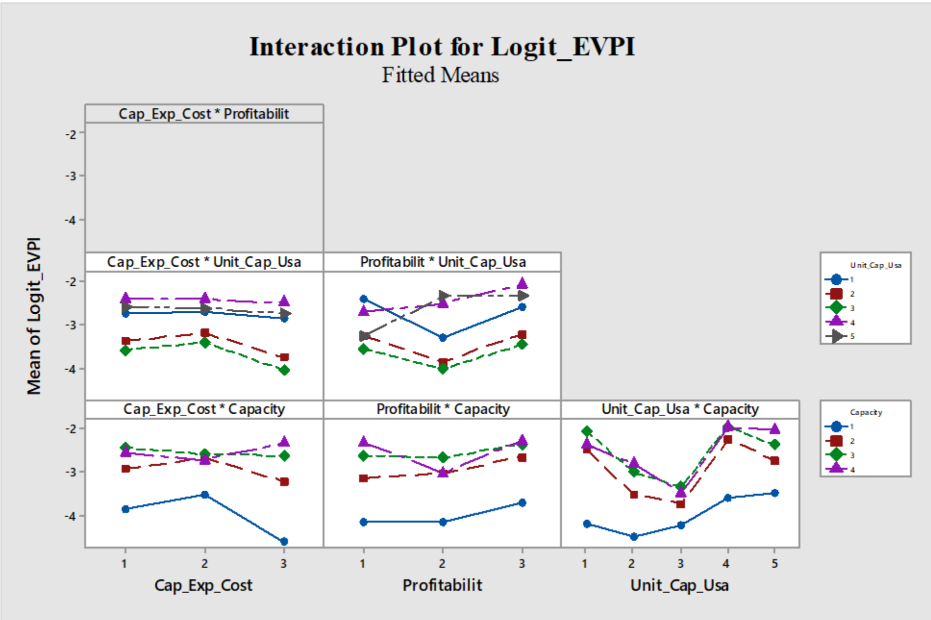


Figure C6. Interaction plot for Case 1 (Deterministic Parameters: EVPI)

According to Figure C4, the assumptions of the regression model whose R-sqr (adjusted) is 93 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:
 randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Profitability.f +  Unit_Cap_Usage.f +
                          Capacity.f, data =cleandata, mtry = 3, importance = TRUE, ntree = 1000,
                          na.action = na.omit)
            Type of random forest: classification
                Number of trees: 1000
No. of variables tried at each split: 3

            OOB estimate of  error rate: 10.56 %
```

Confusion matrix:

	H	L	M	class.error
H	0	0	3	1.00000000
L	0	128	8	0.05882353
M	1	7	33	0.19512195

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-6.69	-24.19	-25.54	-33.45	5.86
Profitability.f	9.06	28.12	22.85	32.92	12.56
Unit_Cap_Usage.f	8.68	53.93	54.69	67.40	26.99
Capacity.f	8.13	35.45	33.80	44.24	19.93

Since the mean decrease accuracy of capacity expansion cost is negative, we exclu de this parameter and re-run the model with the remaining parameters.

R output for the revised model including three deterministic parameters:

```
Call:
 randomForest(formula = EVPI.f ~ Profitability.f + Unit_Cap_Usage.f  +  Capacity.f, data = clean
                                    data, mtry = 3, importance = TRUE,      ntree = 1000, na.action
                                    = na.omit)
            Type of random forest: classification
                Number of trees: 1000
No. of variables tried at each split: 3

      OOB estimate of  error rate: 9.44%
```

Confusion matrix:

	H	L	M	class.error
H	0	0	3	1.00000000
L	0	132	4	0. 02941176
M	3	7	31	0.2390244

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Profitability.f	16.35	60.72	56.19	82.72	12.75
Unit_Cap_Usage.f	19.03	81.31	92.09	113.38	26.54
Capacity.f	17.57	60.30	61.98	82.23	20.62

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,084	1	1,644	X[,2] %in% c('2','3')	L
2	1	0,063	1	1,644	X[,3] %in% c('1')	L
3	2	0,054	0,974	2,681	X[,1] %in% c('1') & X[,2] %in% c('1')	M
4	2	0,045	1	1,644	X[,1] %in% c('2') & X[,2] %in% c('1')	L
5	1	0,044	0,945	1,554	X[,2] %in% c('1','4','5')	L
6	2	0,043	1	1,644	X[,2] %in% c('5') & X[,3] %in% c('2')	L
7	2	0,041	1	1,644	X[,2] %in% c('1','4','5') & X[,3] %in% c('1')	L
8	2	0,041	1	1,644	X[,1] %in% c('1') & X[,2] %in% c('5')	L

C.1.2. Experiment 2

C.1.2.1. Response: VSS

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four uncertain parameters:

Call:
randomForest(formula = VSS.f ~ CV_Demand.f + CV_Price.f + CV_Cost.f + CV_Canb.Rate.f, da
ta = cleandata, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 6.67%

Confusion matrix:

	H	L	M	class.error
H	15	0	0	0.0
L	0	4	1	0.2
M	0	1	9	0.1

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-2.56	-4.92	-9.82	-9.20	0.94
CV_Price.f	47.05	35.09	38.54	59.01	10.97
CV_Cost.f	22.80	14.11	21.28	31.23	4.28
CV_Canb.Rate.f	-4.43	4.36	-4.82	-1.98	1.47

Since the mean decrease accuracy of CV_Demand and CV_Cannibalisation Rate are negative, we exclude these parameters and re-run the model with the remaining parameters.

***R output for the revised model including two uncertain parameters*:**

Call:
randomForest(formula = VSS.f ~ CV_Price.f + CV_Cost.f, data = cleandata, mtry = 3, importance
= TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 0%

Confusion matrix:

	H	L	M	class.error
H	15	0	0	0.0
L	0	5	0	0.0
M	0	0	106	0.0

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_ Price.f	62.15	50.65	58.80	91.37	11.87
CV_Cost.f	33.94	19.69	36.43	49.98	5.82

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,204	1	2,446	X[,1] %in% c('0,30')	H
2	2	0,180	1	2,600	X[,1] %in% c('0,15') & X[,2] %in% c('0','0,15')	M
3	2	0,170	1	2,446	X[,1] %in% c('0,15') & X[,2] %in% c('0,30')	H
4	2	0,170	1	2,600	X[,1] %in% c('0') & X[,2] %in% c('0,30')	M
5	2	0,162	1	4,843	X[,1] %in% c('0') & X[,2] %in% c('0','0,15')	L

CASE 2

C.2.1. Experiment 1

C.2.1.1. Response: VSS

After solving 180 runs, we analyse the related data using MINITAB software to see which factors including interactions among them are significant. The results are shown below.

ANOVA Table for General Factorial Regression: Logit_VSS versus factors

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	91	951,07	10,451	8,49	0,000
Linear	11	183,64	16,694	13,55	0,000
Cap_Exp_Cost	2	8,50	4,248	3,45	0,036
Profitability	2	45,46	22,730	18,45	0,000
Unit_Cap_Usage	4	45,77	11,443	9,29	0,000
Capacity	3	83,91	27,970	22,71	0,000
2-Way Interactions	44	579,63	13,173	10,70	0,000
Cap_Exp_Cost*Profitability	4	7,28	1,821	1,48	0,216
Cap_Exp_Cost*Unit_Cap_Usage	8	22,49	2,812	2,28	0,029
Cap_Exp_Cost*Capacity	6	15,83	2,638	2,14	0,056
Profitability*Unit_Cap_Usage	8	130,25	16,281	13,22	0,000
Profitability*Capacity	6	61,17	10,194	8,28	0,000
Unit_Cap_Usage*Capacity	12	342,61	28,551	23,18	0,000
3-Way Interactions	36	187,80	5,217	4,24	0,000
Cap_Exp_Cost*Profitability*Capacity	12	31,43	2,619	2,13	0,023
Profitability*Unit_Cap_Usage*Capacity	24	156,37	6,516	5,29	0,000
Error	88	108,39	1,232		
Total	179	1059,46			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,10983	89,77%	79,19%	57,20%

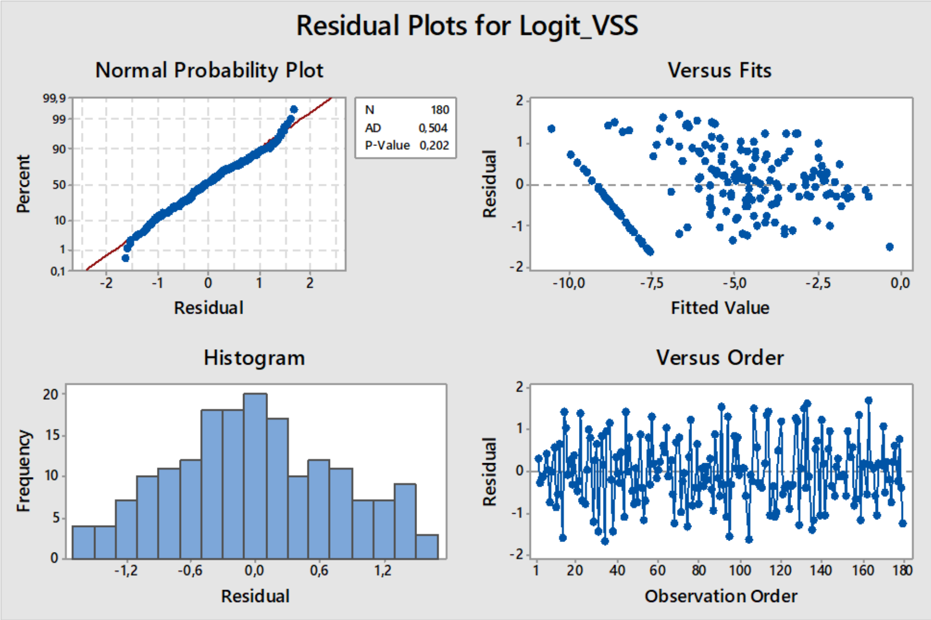


Figure C7. Residual plots for Case 2 (Deterministic Parameters: VSS)

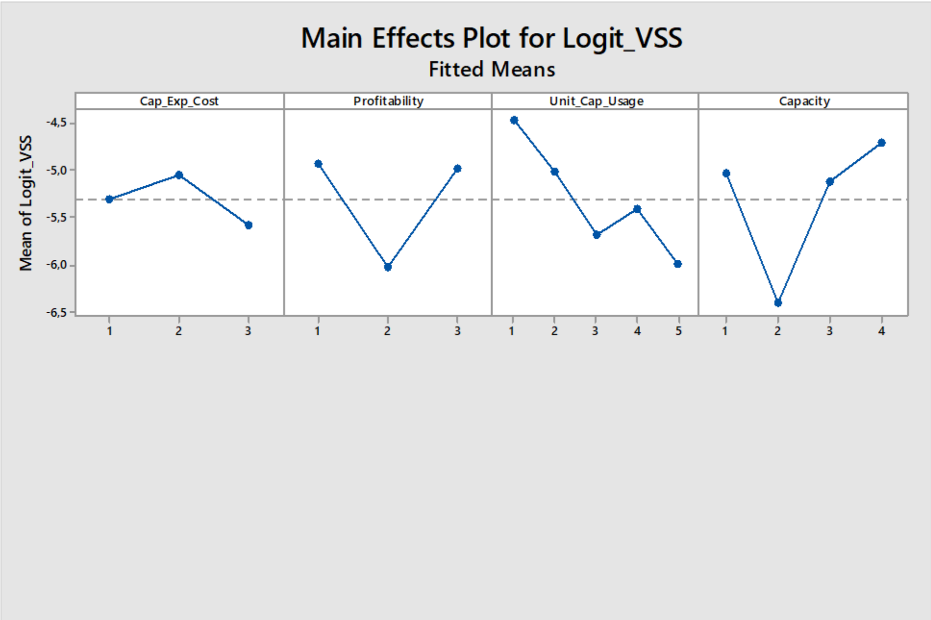


Figure C8. Main effects plot for Case 2 (Deterministic Parameters: VSS)

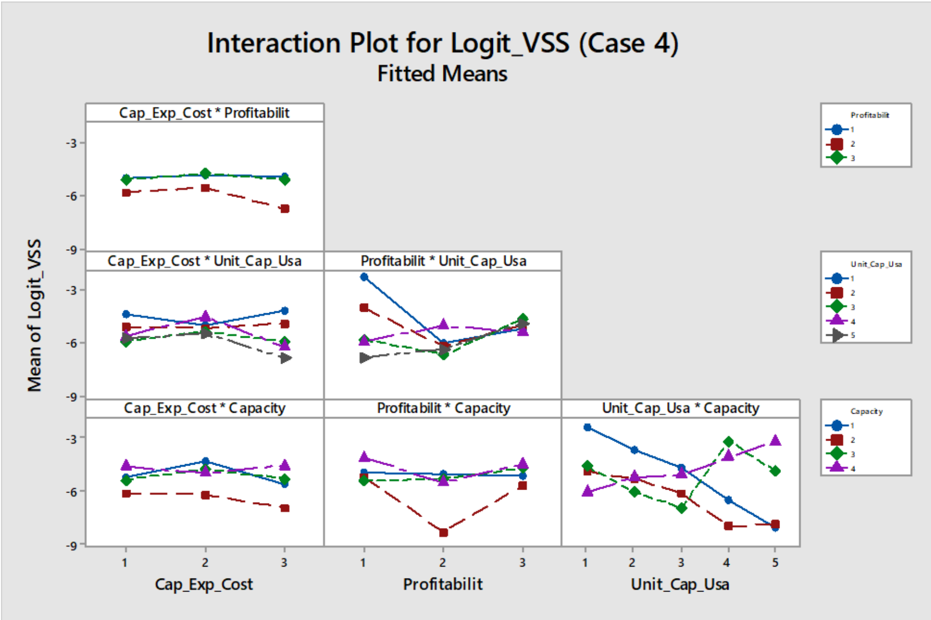


Figure C9. Interaction plot for Case 2 (Deterministic Parameters: VSS)

According to Figure C7, the assumptions of the regression model are satisfied and R-sqr (adjusted), 79.2 % can be accepted.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacit
y.f, data = cleandata, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.
omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 2.22%

Confusion matrix:

	H	L	M	class.error
H	6	0	1	0.1428571
L	0	167	0	0.0000000
M	0	3	3	0.5000000

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-11.18	-12.49	-8.26	-18.43	1.75
Profitability.f	34.23	34.40	23.75	49.53	8.23
Unit_Cap_Usage.f	39.29	33.79	18.60	45.52	6.71
Capacity.f	21.69	16.34	24.59	30.57	7.06

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for the revised model including three deterministic parameters:

```
Call:randomForest (formula = VSS.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f,
               data = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.action = na.omit)
          Type of random forest: classification
               Number of trees: 1000
No. of variables tried at each split: 3

       OOB estimate of error rate: 2.22%

Confusion matrix:
     H    L    M    class.error
  H  6    0    1    0.1428571
  L  0  167    0    0.0000000
  M  0    3    3    0.5000000

Importance of variables:

                      H      L      M   MeanDecreaseAccuracy   eanDecreaseGini
  Profitability.f   35.39  39.02  24.26        51.96                6.70
  Unit_Cap_Usage.f  41.43  40.73  22.19        56.09                6.96
  Capacity.f        22.33  17.46  25.44        31.93                5.51
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,148	1	1,379	X[,1] %in% c('2','3')	L
2	1	0,12	1	1,379	X[,2] %in% c('5')	L
3	3	0,087	1	6,237	X[,1] %in% c('1') & X[,2] %in% c('2') & X[,3] %in% c('3')	M
4	2	0,087	0,978	6,100	X[,2] %in% c('2') & X[,3] %in% c('3')	M
5	1	0,082	1	1,379	X[,3] %in% c('4')	L
6	2	0,081	1	1,379	X[,1] %in% c('1') & X[,3] %in% c('4')	L
7	2	0,08	1	1,379	X[,1] %in% c('2','3') & X[,2] %in% c('2')	L
8	2	0,079	0,959	8,390	X[,1] %in% c('1') & X[,3] %in% c('1','2')	H
9	1	0,079	0,908	7,944	X[,3] %in% c('1','2')	H
10	3	0,078	1	8,749	X[,1] %in% c('1') & X[,2] %in% c('2') & X[,3] %in% c('1','2')	H
11	2	0,078	0,972	8,504	X[,2] %in% c('2') & X[,3] %in% c('1','2')	H
12	2	0,077	1	1,379	X[,2] %in% c('2') & X[,3] %in% c('4')	L
13	3	0,076	1	1,379	X[,1] %in% c('1') & X[,2] %in% c('2') & X[,3] %in% c('4')	L
14	1	0,075	0,997	1,374	X[,3] %in% c('2','3','4')	L
15	3	0,073	1	6,237	X[,1] %in% c('3') & X[,2] %in% c('1') & X[,3] %in% c('1')	M

C.2.1.2. Response: EVPI

After solving 180 runs, we analyse the related data using MINITAB software to see which factors including interactions among them are significant. The results are shown below.

ANOVA Table for General Factorial Regression: Logit_EVPI versus factors

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Max_Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	43	164,981	3,8368	91,36	0,000
Linear	11	129,593	11,7812	280,54	0,000
Cap_Exp_Cost	2	0,322	0,1610	3,83	0,024
Profitability	2	10,579	5,2896	125,96	0,000
Unit_Cap_Usage	4	62,832	15,7081	374,05	0,000
Capacity	3	55,859	18,6197	443,38	0,000
2-Way Interactions	32	35,388	1,1059	26,33	0,000
Cap_Exp_Cost*Capacity	6	1,360	0,2267	5,40	0,000
Profitability*Unit_Cap_Usage	8	14,485	1,8106	43,11	0,000
Profitability*Capacity	6	2,899	0,4832	11,51	0,000
Unit_Cap_Usage*Capacity	12	16,644	1,3870	33,03	0,000
Error	136	5,711	0,0420		
Total	179	170,692			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,204927	96,65%	95,60%	94,14%

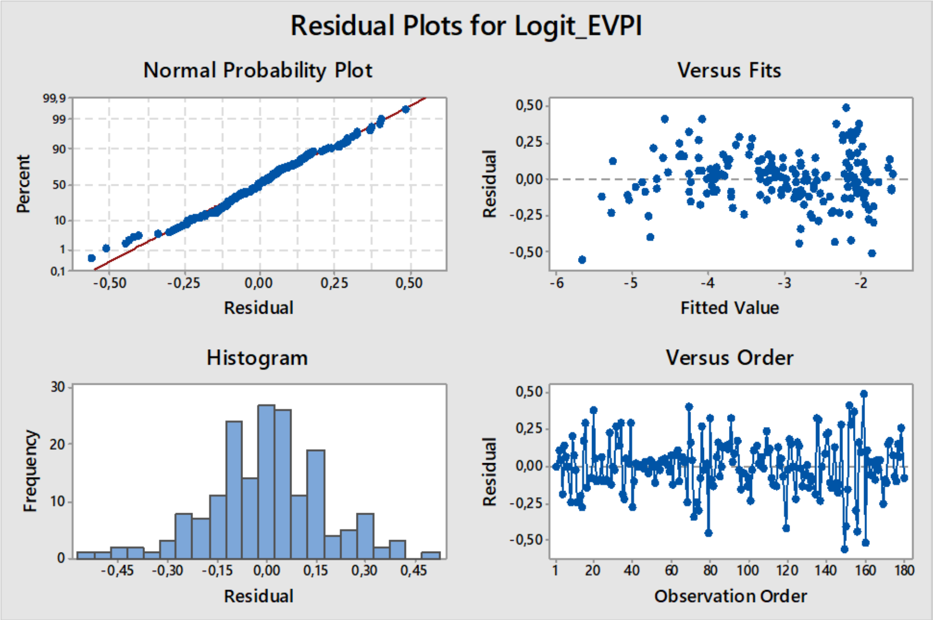


Figure C10. Residual plots for Case 2 (Deterministic Parameters: EVPI)

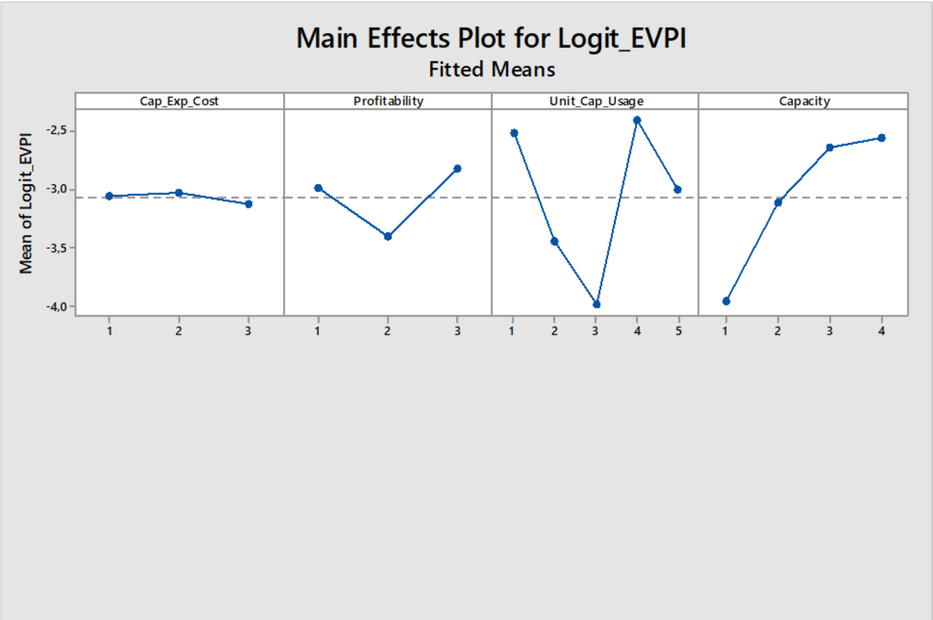


Figure C11. Main effects plot for Case 2 (Deterministic Parameters: EVPI)

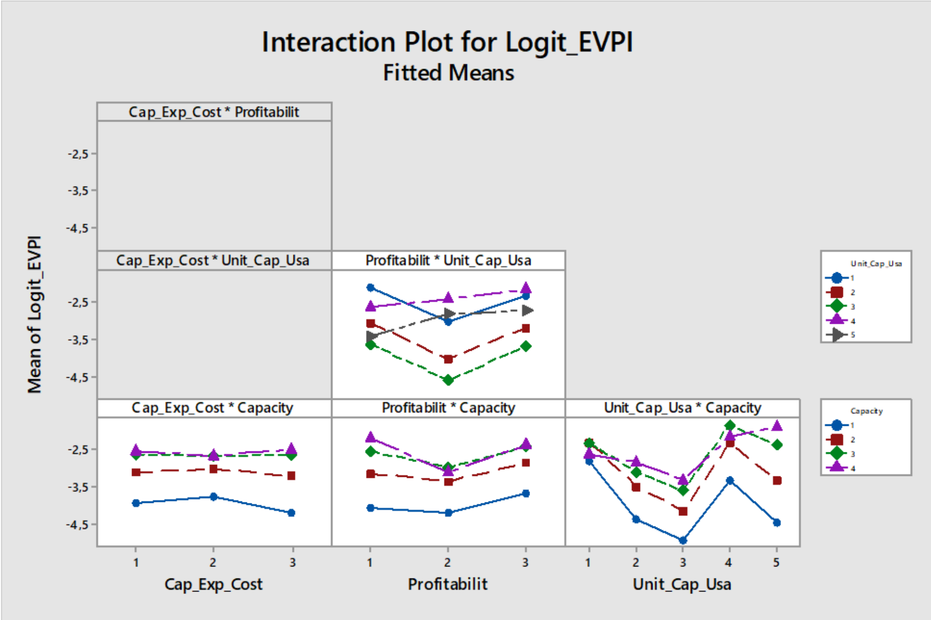


Figure C12. Interaction plot for Case 2 (Deterministic Parameters: EVPI)

According to Figure C10, the assumptions of the regression model whose R-sqr (adjusted) is 95.6 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

Call:
randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacity.f, data = cleandata, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 9.44%

Confusion matrix:

	L	M	class.error
L	122	11	0.08270677
M	6	41	0.12765957

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-26.18	-24.00	-32.27	3.94
Profitability.f	25.18	21.11	30.65	10.31
Unit_Cap_Usage.f	74.25	79.77	96.16	30.13
Capacity.f	49.65	54.19	66.08	23.61

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for a model including three deterministic parameters:

```
Call:
randomForest(formula = EVPI.f ~ Profitability.f + Unit_Cap_Usage.f +  Capacity.f, data = cleanda
              ta, mtry = 3, importance = TRUE,  ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate: 5.56%

Confusion matrix:
    L   M  class.error
L 129   4  0.03007519
M   6  41  0.12765957

Importance of variables:
                     L     M  MeanDecreaseAccuracy  MeanDecreaseGini
Profitability.f  40.41 40.00                 54.67              9.03
Unit_Cap_Usage.f 82.49 96.99                112.45             30.21
Capacity.f       66.97 72.80                 91.40             21.53
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,081	0,986	1,906	X[,3] %in% c('3','4')	M
2	1	0,077	1	2,071	X[,3] %in% c('1','2')	L
3	1	0,071	1	2,071	X[,2] %in% c('2','3')	L
4	2	0,070	1	1,933	X[,2] %in% c('5') & X[,3] %in% c('3','4')	M
5	2	0,063	1	2,071	X[,1] %in% c('2') & X[,2] %in% c('1')	L
7	2	0,055	1	2,071	X[,2] %in% c("4','5') & X[,3] %in% c('1')	L

C.2.2. Experiment 2

C.2.2.1. Response: VSS

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four uncertain parameters:

```
Call:
RandomForest (formula = VSS.f ~ CV_Demand.f + CV_Price.f + CV_Cost.f +CV_Canb.Rate.f,
                              data = cleandata, mtry = 3, importance = TRUE,  ntree = 1000,
                                     na.action = na.omit)
           Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 3

     OOB estimate of error rate: 10%
```

Confusion matrix:

	H	L	M	class.error
H	15	0	0	0.00
L	0	4	1	0.20
M	0	2	8	0.20

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-3.90	-4.51	-8.57	-9.43	0.87
CV_Price.f	48.11	36.32	37.72	61.45	11.03
CV_Cost.f	24.03	13.27	20.96	30.29	4.39
CV_Canb.Rate.f	-5.62	1.90	-4.96	-4.75	1.43

Since the mean decrease accuracy of CV_Demand and CV_Cannibalisation Rate are negative, we exclude these parameters and re-run the model with the remaining parameters.

***R output for the revised model including three uncertain parameters*:**

```
Call:
randomForest(formula = VSS.f ~ CV_Price.f + CV_Cost.f,  data = cleandata, mtry = 2,
                            importance = TRUE,  ntree = 1000,  na.action = na.omit)
            Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 2
     OOB estimate of  error rate: 0%
```

Confusion matrix:

	H	L	M	class.error
H	15	0	0	0
L	0	5	0	0
M	0	0	10	0

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Price.f	59.24	52.46	57.68	90.66	11.90
CV_Cost.f	32.78	21.49	33.86	50.58	5.83

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,203	1	2,487	X[,1] %in% c('0,30')	H
2	2	0,179	1	2,623	X[,1] %in% c('0,15') & X[,2] %in% c('0','0,15')	M
3	2	0,170	1	2,623	X[,1] %in% c('0') & X[,2] %in% c('0,30')	M
4	2	0,168	1	2,487	X[,1] %in% c('0,15') & X[,2] %in% c('0,30')	H
5	2	0,164	1	4,614	X[,1] %in% c('0') & X[,2] %in% c('0','0,15')	L

CASE 3

C.3.1. Experiment 1

C.3.1.1. Response: VSS

After solving 180 runs, we analyse the related data using MINITAB software to see which factors including interactions among them are significant. The results are shown below.

ANOVA Table for General Factorial Regression: Logit_VSS versus factors

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,2

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	131	1358,370	10,369	23,050	0,000
Linear	11	666,630	60,603	134,690	0,000
Cap_Exp_Cost	2	13,410	6,703	14,900	0,000
Profitability	2	24,090	12,045	26,770	0,000
Unit_Cap_Usage	4	228,580	57,145	127,000	0,000
Capacity	3	400,550	133,518	296,740	0,000
2-Way Interactions	44	458,030	10,410	23,140	0,000
Cap_Exp_Cost*Profitability	4	2,700	0,675	1,500	0,217
Cap_Exp_Cost*Unit_Cap_Usage	8	9,030	1,129	2,510	0,023
Cap_Exp_Cost*Capacity	6	26,280	4,380	9,730	0,000
Profitability*Unit_Cap_Usage	8	28,900	3,612	8,030	0,000
Profitability*Capacity	6	20,600	3,433	7,630	0,000
Unit_Cap_Usage*Capacity	12	370,530	30,877	68,620	0,000
3-Way Interactions	76	233,710	3,075	6,830	0,000
Cap_Exp_Cost*Profitability*Unit_Cap_Usage	16	9,930	0,621	1,380	0,192
Cap_Exp_Cost*Profitability*Capacity	12	17,010	1,417	3,150	0,002
Cap_Exp_Cost*Unit_Cap_Usage*Capacity	24	58,600	2,442	5,430	0,000
Profitability*Unit_Cap_Usage*Capacity	24	148,170	6,174	13,720	0,000
Error	48	21,600	0,450		
Total	179	1379,970			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,670782	98,43%	94,16%	77,99%

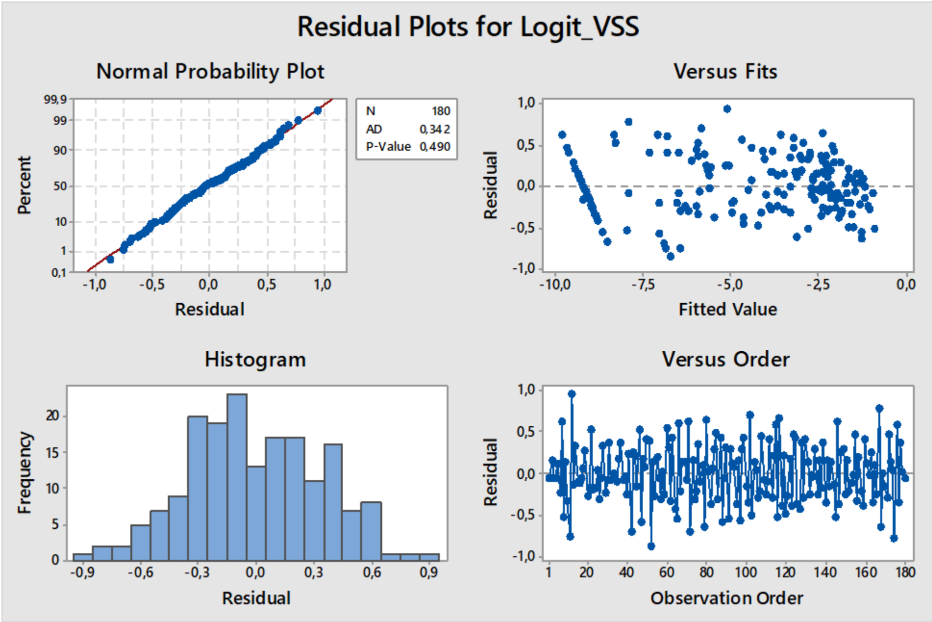


Figure C13. Residual plots for Case 3 (Deterministic Parameters: VSS)

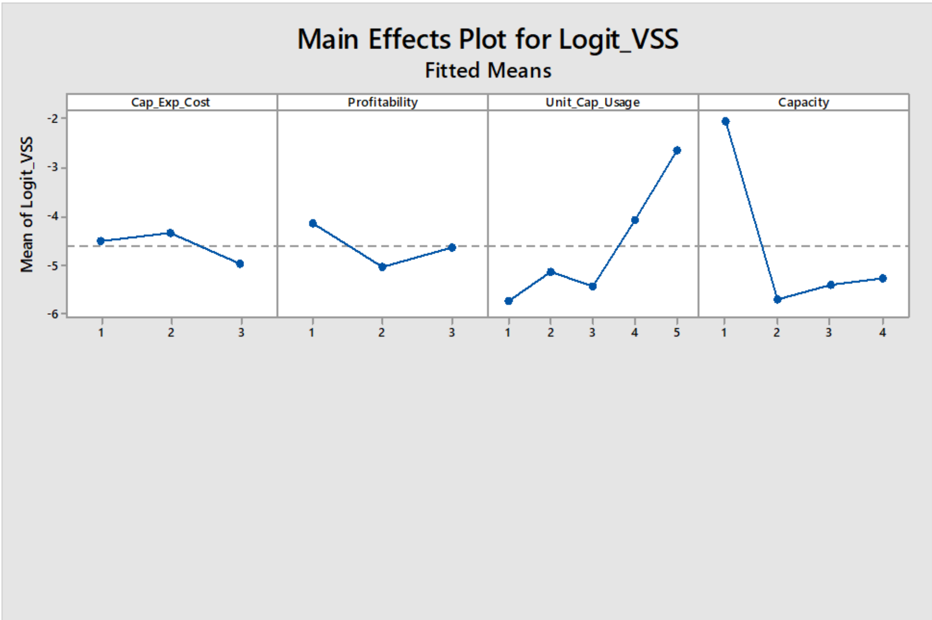


Figure C14. Main effects plot for Case 3 (Deterministic Parameters: VSS)

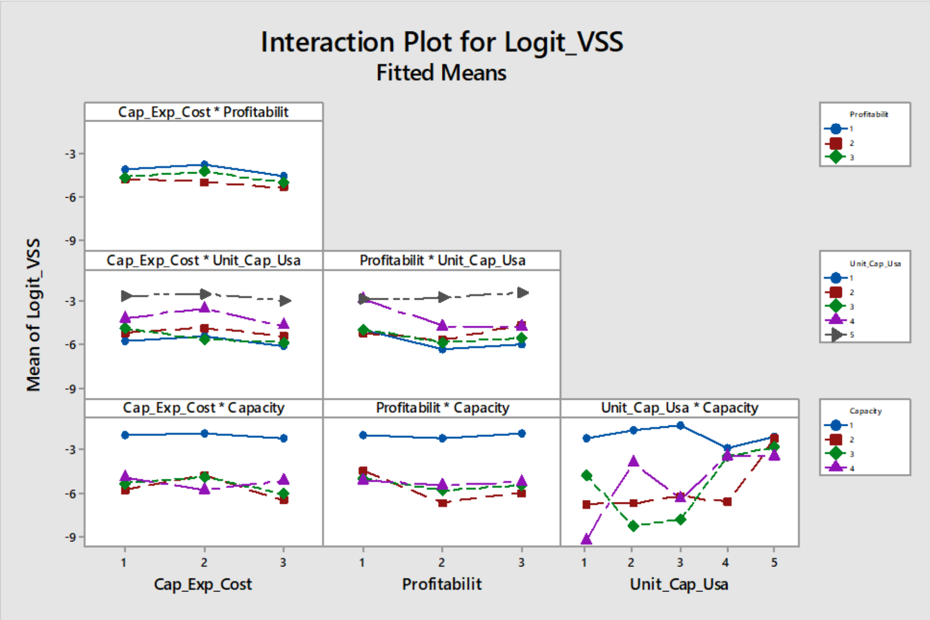


Figure C15. Interaction plot for Case 3 (Deterministic Parameters: VSS)

According to Figure C13, the assumptions of the regression model are satisfied and R-sqr (adjusted), 94.6 % is accepted.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacit
                y.f,
                data = cleandata, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
            Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 3

     OOB estimate of  error rate: 10.56%
```

Confusion matrix:

	H	L	M	class.error
H	6	2	2	0.40000000
L	0	129	5	0.03731343
M	2	8	26	0.27777778

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-12.57	-21.08	-22.73	-31.80	6.16
Profitability.f	20.37	16.40	30.70	34.93	15.16
Unit_Cap_Usage.f	36.34	33.69	33.53	52.18	22.11
Capacity.f	43.52	70.60	57.67	87.92	26.25

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for a model including all three deterministic parameters:

```
Call:
 randomForest(formula = VSS.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f, data = cleandat
                  a,   mtry = 3,   importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate: 7.78%

Confusion matrix:
       H    L    M    class.error
  H    6    0    4    0.40000000
  L    0  131    3    0. 02238806
  M    2    5   29    0. 19444444

Importance of variables:
                       H        L        M    MeanDecreaseAccuracy   MeanDecreaseGini
  Profitability.f    21.97    28.68    41.63        49.61                  12.31
  Unit_Cap_Usage.f   39.72    47.79    50.71        71.87                  19.48
  Capacity.f         47.51    84.60    71.64       107.67                  25.65
```

Rules Extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,193	0,916	1,582	X[,2] %in% c('4')	L
2	1	0,123	1	1,727	X[,3] %in% c('4')	L
3	2	0,084	0,943	3,250	X[,1] %in% c('2') & X[,3] %in% c('1')	M
4	2	0,078	1	1,727	X[,2] %in% c('4') & X[,3] %in% c('1')	L
5	1	0,077	0,912	1,575	X[,3] %in% c('2','4')	L
6	2	0,059	0,891	3,071	X[,1] %in% c('1') & X[,3] %in% c('3')	M
7	1	0,057	1	1,727	X[,3] %in% c('2','3','4')	L
8	2	0,05	1	1,727	X[,1] %in% c('2','3') & X[,2] %in% c('4')	L
9	2	0,043	1	1,727	X[,2] %in% c('5') & X[,3] %in% c('4')	L
10	2	0,041	1	1,727	X[,1] %in% c('2') & X[,3] %in% c('3')	L

C.3.1.2. Response: EVPI

After solving 180 runs, we analyse the related data using MINITAB software to see which factors including interactions among them are significant. The results are shown below.

ANOVA Table for General Factorial Regression: ASIN_EVPI versus factors

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	43	1,494	0,035	143,810	0,000
Linear	11	1,323	0,120	497,890	0,000
Cap_Exp_Cost	2	0,001	0,001	2,350	0,099
Profitability	2	0,125	0,062	258,560	0,000
Unit_Cap_Usage	4	0,550	0,138	569,230	0,000
Capacity	3	0,647	0,216	892,680	0,000
2-Way Interactions	32	0,171	0,005	22,090	0,000
Cap_Exp_Cost*Capacity	6	0,013	0,002	8,930	0,000
Profitability*Unit_Cap_Usage	8	0,056	0,007	29,030	0,000
Profitability*Capacity	6	0,042	0,007	28,680	0,000
Unit_Cap_Usage*Capacity	12	0,060	0,005	20,750	0,000
Error	136	0,033	0,000		
Total	179	1,527			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0155439	97,85%	97,17%	96,23%

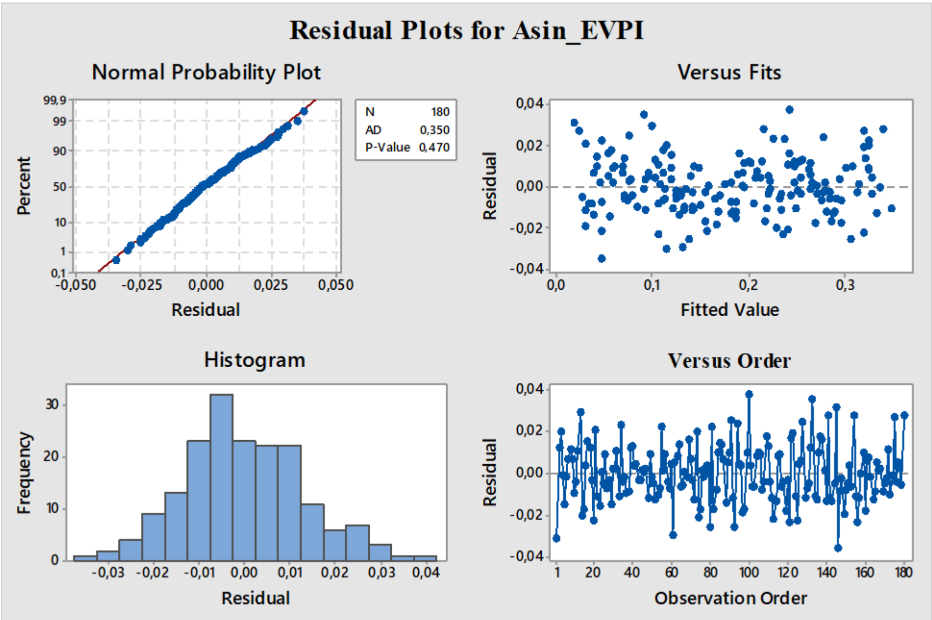


Figure C16. Residual plots for Case 3 (Deterministic Parameters: EVPI)

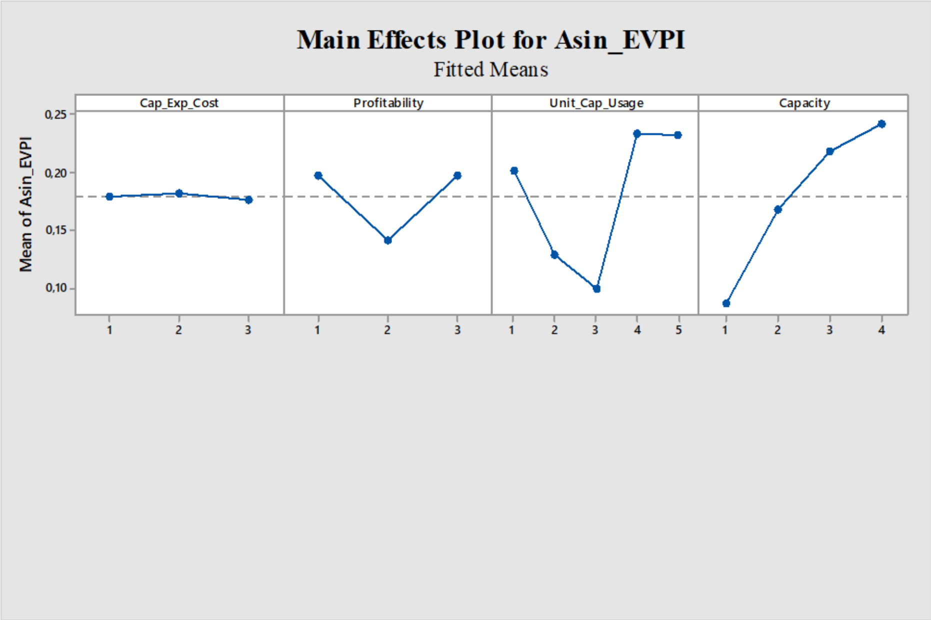


Figure C17. Main effects plot for Case 3 (Deterministic Parameters: EVPI)

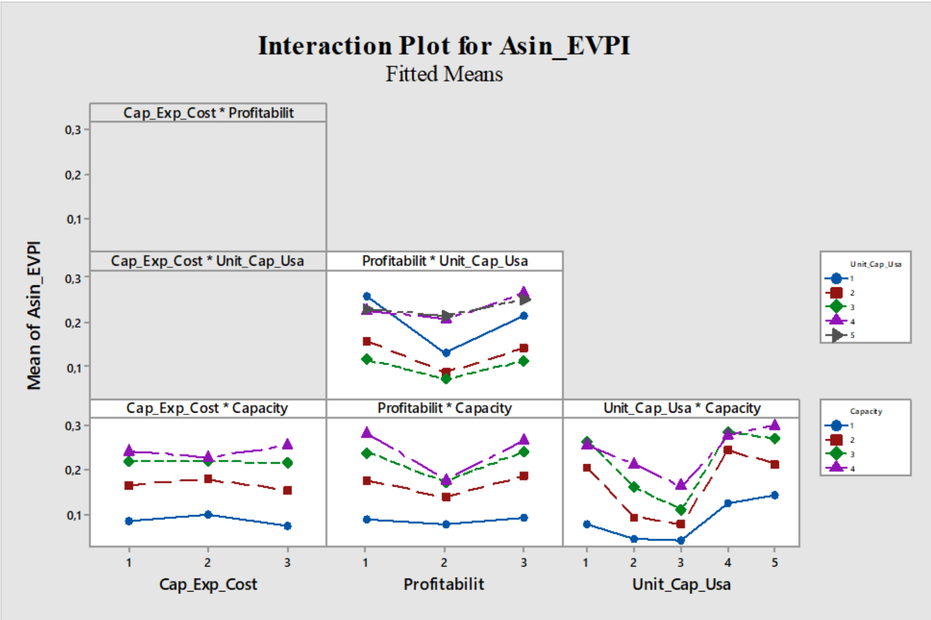


Figure C18. Interaction plot for Case 3 (Deterministic Parameters: EVPI)

According to Figure C16, the assumptions of the regression model whose R-sqr (adjusted) is 97.2 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:
randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capaci
                 ty.f,
                           data = cleandata, mtry = 3, importance = TRUE, ntree = 1000,
                                  na.action = na.omit)
               Type of random forest: classification
                      Number of trees: 1000
No. of variables tried at each split: 3

      OOB estimate of  error rate: 7.78%

Confusion matrix:
        L    M    class.error
  L   161    4    0. 02424242
  M    10    5    0. 66666667

Importance of variables:
                          L       M     MeanDecreaseAccuracy   MeanDecreaseGini
  Cap_Exp_Cost.f      -11.72   -8.72         -14.41                 3.34
  Profitability.f      21.26   30.44          33.67                 8.33
  Unit_Cap_Usage.f     21.27   30.74          33.16                 8.91
  Capacity.f           20.48   27.70          30.10                 5.83
```

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for the revised model including three deterministic parameters:

```
Call:
 randomForest(formula = EVPI.f ~ Profitability.f + Unit_Cap_Usage.f  +  Capacity.f, data = clean
                   data, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                      Number of trees: 1000
No. of variables tried at each split: 3

      OOB estimate of  error rate: 4.44%

Confusion matrix:
        L    M    class.error
  L   163    2    0. 01212121
  M     6    9    0. 40000000

Importance of variables:
                          L       M     MeanDecreaseAccuracy   MeanDecreaseGini
  Profitability.f      29.76   39.40          45.00                 7.55
  Unit_Cap_Usage.f     30.99   44.07          48.22                 8.14
  Capacity.f           25.63   35.35          39.53                 5.27
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	2	0,116	0,949	3,338	X[,1] %in% c('3') & X[,2] %in% c('4')	M
2	1	0,108	1	1,397	X[,1] %in% c('2')	L
3	2	0,100	0,935	3,289	X[,1] %in% c('1') & X[,2] %in% c('1')	M
4	2	0,096	1	1,397	X[,1] %in% c('1') & X[,2] %in% c('4')	L
5	1	0,091	1	1,397	X[,3] %in% c('1','2')	L
6	1	0,078	1	1,397	X[,2] %in% c('2','3')	L
7	2	0,072	1	1,397	X[,1] %in% c('3') & X[,2] %in% c('1')	L
8	1	0,064	0,923	1,290	X[,2] %in% c('1','4')	L
9	2	0,059	1	1,397	X[,1] %in% c('2') & X[,3] %in% c('3','4')	L
10	2	0,051	1	1,397	X[,1] %in% c('1') & X[,2] %in% c('5')	L

C.3.2. Experiment 2

C.3.2.1. Response: VSS

Rules extracted from the Random Forest Application

R output for a model including all four uncertain parameters:

```
Call:
randomForest(formula = VSS.f ~ CV_Demand.f + CV_Price.f + CV_Cost.f + CV_Canb.Rate.f,dat
                 a = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate: 10%

Confusion matrix:
    H   L   M   class.error
H  17   0   0   0.0000000
L   0   2   2   0.5000000
M   0   1   8   0.1111111

Importance of variables:
                   H      L      M     MeanDecreaseAccuracy   MeanDecreaseGini
CV_Demand.f      -6.37   5.26  -5.57        -2.65                  1.41
CV_Price.f       35.45  24.52  32.48        50.49                  8.67
CV_Cost.f        33.59   9.88  26.39        39.18                  5.79
CV_Canb.Rate.f   -4.47  -3.43  -2.72        -7.27                  0.69
```

Since the mean decrease accuracy of CV_Demand.f and CV_Canb.Rate.f are negative, we exclude those parameters and re-run the model with the remaining parameters.

R output for a model including two uncertain parameters:

Call: randomForest(formula = VSS.f ~ CV_Price.f + CV_Cost.f, data = cleandata, mtry = 2,
importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 2

OOB estimate of error rate: 6.67%

Confusion matrix:

	H	L	M	class.error
H	17	0	0	0.0000000
L	0	3	1	0.2500000
M	0	1	8	0.1111111

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Price.f	41.35	35.05	42.12	62.71	9.40
CV_Cost.f	43.49	20.83	32.11	53.18	6.66

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,306	0,851	2,730	X[,1] %in% c('0')	L
2	1	0,236	1	2,130	X[,1] %in% c('0,15') & X[,2] %in% c('0,30')	H
3	1	0,207	1	2,130	X[,1] %in% c('0,30')	H
4	1	0,206	0,767	3,505	X[,1] %in% c('0,15')	M
5	2	0,174	1	4,569	X[,1] %in% c('0,15') & X[,2] %in% c('0','0,15')	M
6	1	0,142	0,808	2,592	X[,2] %in% c('0')	L
7	2	0,137	1	3,208	X[,1] %in% c('0') & X[,2] %in% c('0')	L
8	2	0,113	1	3,208	X[,1] %in% c('0') & X[,2] %in% c('0,15')	L
9	1	0,113	0,772	2,477	X[,2] %in% c('0,15')	L

CASE 4

C.4.1. Experiment 1

C.4.1.1. Response: VSS

After solving 180 runs, we analyse the related data using MINITAB software to see which factors including interactions among them are significant. The results are shown below.

ANOVA Table for General Factorial Regression: Logit_VSS versus factors

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	43	1278,26	29,727	23,10	0,000
Linear	11	743,57	67,598	52,52	0,000
Cap_Exp_Cost	2	0,16	0,081	0,06	0,939
Profitability	2	11,81	5,903	4,59	0,012
Unit_Cap_Usage	4	42,88	10,720	8,33	0,000
Capacity	3	688,73	229,575	178,38	0,000
2-Way Interactions	32	534,69	16,709	12,98	0,000
Cap_Exp_Cost*Capacity	6	15,13	2,521	1,96	0,076
Profitability*Unit_Cap_Usage	8	168,85	21,106	16,40	0,000
Profitability*Capacity	6	63,71	10,619	8,25	0,000
Unit_Cap_Usage*Capacity	12	286,99	23,916	18,58	0,000
Error	136	175,03	1,287		
Total	179	1453,29			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
1,13447	87,96%	84,15%	78,90%

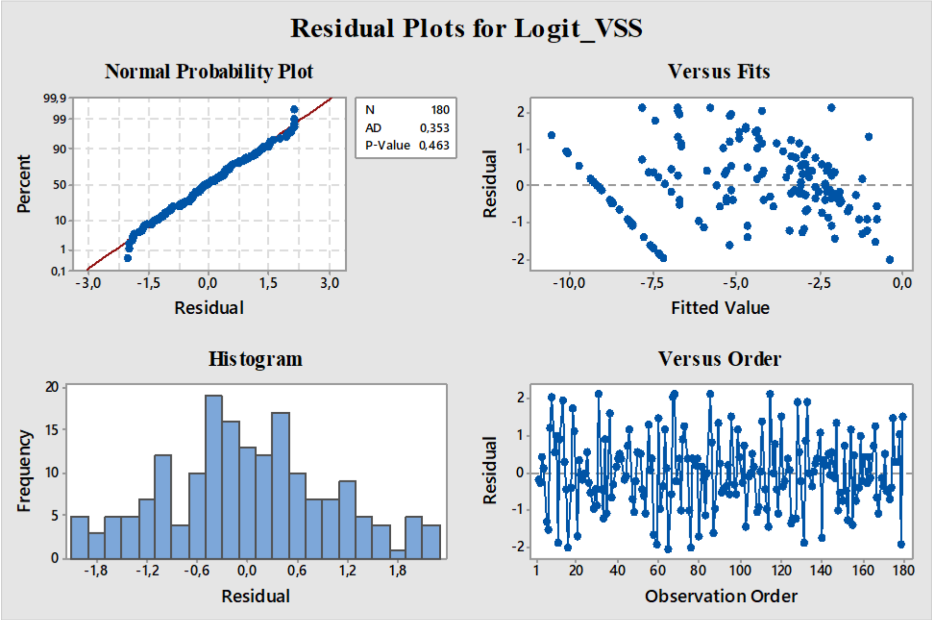


Figure C19. Residual plots for Case 4 (Deterministic Parameters: VSS)

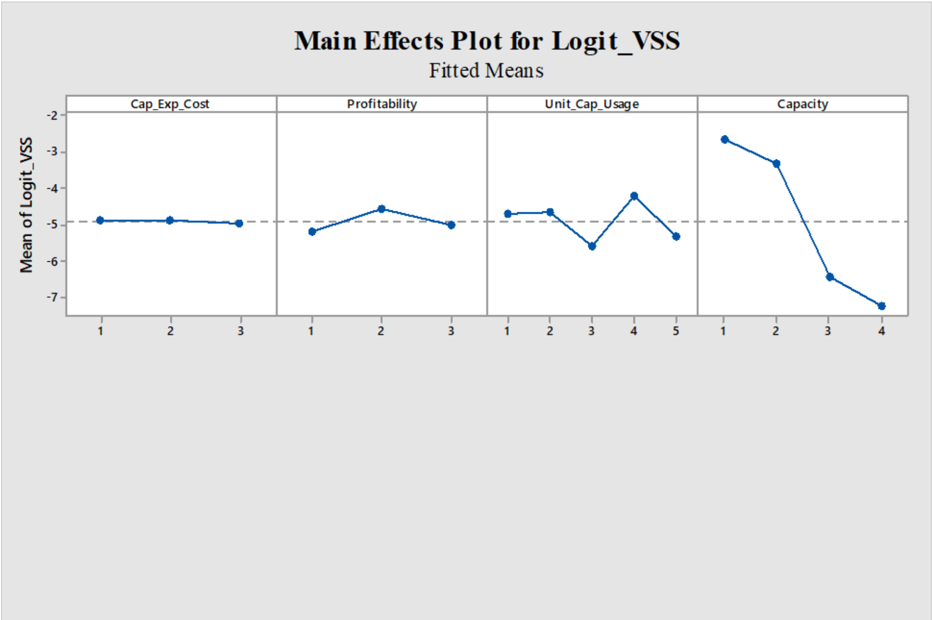


Figure C20. Main effects plot for Case 4 (Deterministic Parameters: VSS)

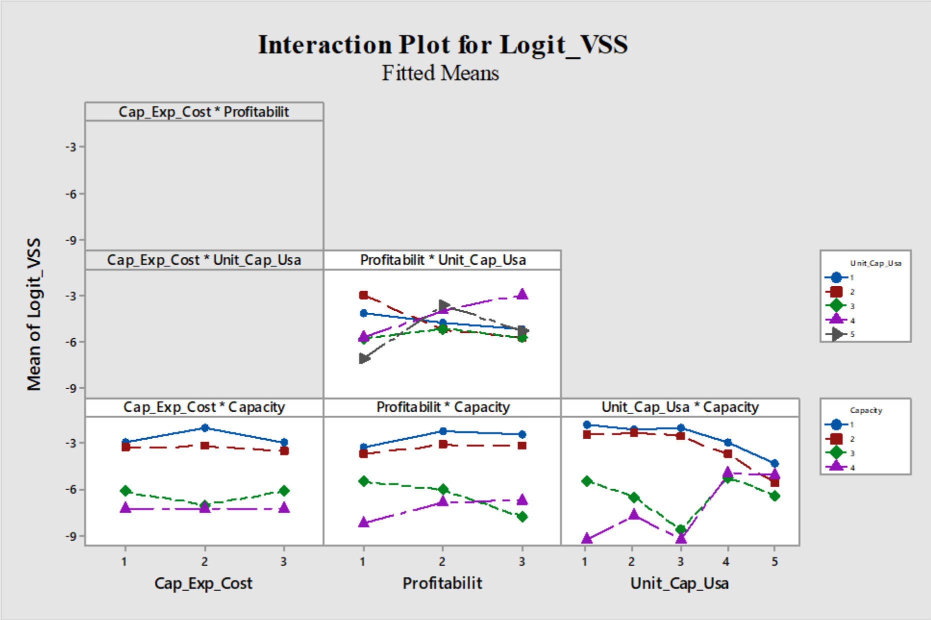


Figure C21. Interaction plot for Case 4 (Deterministic Parameters: VSS)

According to Figure C19, the assumptions of the regression model are satisfied and R-sqr (adjusted), 84.1 % is accepted.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacit
              y.f,
                data = cleandata, mtry = 3,  importance = TRUE, ntree = 1000, na.action = na.omit)
            Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 3

      OOB estimate of  error rate: 12.78%

Confusion matrix:
   H   L   M   class.error
 H  0   3   1   1.0000000
 L  0  154  4   0.02531646
 M  0   15  3   0.83333333
```

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	11.55	-1.80	-10.10	-4.37	6.74
Profitability.f	9.55	-6.16	-2.40	-4.13	7.66
Unit_Cap_Usage.f	-6.55	14.99	19.21	19.90	12.13
Capacity.f	10.75	19.94	30.53	32.53	11.12

Since the mean decrease accuracy of capacity expansion cost and profitability are negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for the revised model including two deterministic parameters:

```
Call:
 randomForest(formula = VSS.f ~ Unit_Cap_Usage.f  +  Capacity.f, data = cleandata, mtry = 2,
                  importance = TRUE,      ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 2

        OOB estimate of error rate: 8.89 %

Confusion matrix:
        H     L     M       class.error
   H    0     3     1     1.000000000
   L    0   157     1     0. 006329114
   M    0    11     7     0. 611111111

Importance of variables:
                          H         L         M     MeanDecreaseAccuracy   MeanDecreaseGini
  Unit_Cap_Usage.f    -2.65     33.05     45.89              50.29                   9.14
  Capacity.f          -1.00     31.73     46.09              51.86                   8.14
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,319	1	1,101	X[,1] %in% c('2')	L
2	1	0,3	1	1,101	X[,2] %in% c('2')	L
3	1	0,193	1	1,101	X[,1] %in% c('3')	L
4	1	0,105	1	1,101	X[,2] %in% c('3')	L
5	2	0,105	1	10,948	X[,1] %in% c('1') & X[,2] %in% c('1', '2')	M
6	2	0,09	1	1,101	X[,1] %in% c('2') & X[,2] %in% c('2')	L
7	1	0,089	1	1,101	X[,2] %in% c('3','4')	L
8	2	0,088	1	1,101	X[,1] %in% c('3') & X[,2] %in% c('2')	L
9	2	0,085	1	1,101	X[,1] %in% c('3') & X[,2] %in% c('1')	L
10	2	0,083	1	1,101	X[,1] %in% c('2') & X[,2] %in% c('1')	L

C.4.1.2. Response: EVPI

After solving 180 runs, we analyse the related data using MINITAB software to see which factors including interactions among them are significant. The results are shown below.

ANOVA Table for General Factorial Regression: ASIN_EVPI versus factors

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Max_Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	115	2,79507	0,024305	627,40	0,000
Linear	11	2,38973	0,217248	5607,92	0,000
Cap_Exp_Cost	2	0,00695	0,003473	89,64	0,000
Profitability	2	0,10677	0,053387	1378,10	0,000
Unit_Cap_Usage	4	1,02823	0,257058	6635,55	0,000
Capacity	3	1,24777	0,415925	10736,46	0,000
2-Way Interactions	44	0,36451	0,008284	213,85	0,000
Cap_Exp_Cost*Profitability	4	0,00024	0,000061	1,57	0,194
Cap_Exp_Cost*Unit_Cap_Usage	8	0,00377	0,000471	12,16	0,000
Cap_Exp_Cost*Capacity	6	0,03565	0,005942	153,38	0,000
Profitability*Unit_Cap_Usage	8	0,18340	0,022925	591,76	0,000
Profitability*Capacity	6	0,04775	0,007959	205,44	0,000
Unit_Cap_Usage*Capacity	12	0,09370	0,007808	201,55	0,000
3-Way Interactions	60	0,04084	0,000681	17,57	0,000
Cap_Exp_Cost*Profitability*Capacity	12	0,00485	0,000404	10,43	0,000
Cap_Exp_Cost*Unit_Cap_Usage*Capacity	24	0,00322	0,000134	3,46	0,000
Profitability*Unit_Cap_Usage*Capacity	24	0,03277	0,001365	35,24	0,000
Error	64	0,00248	0,000039		
Total	179	2,79755			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0062241	99,91%	99,75%	99,30%

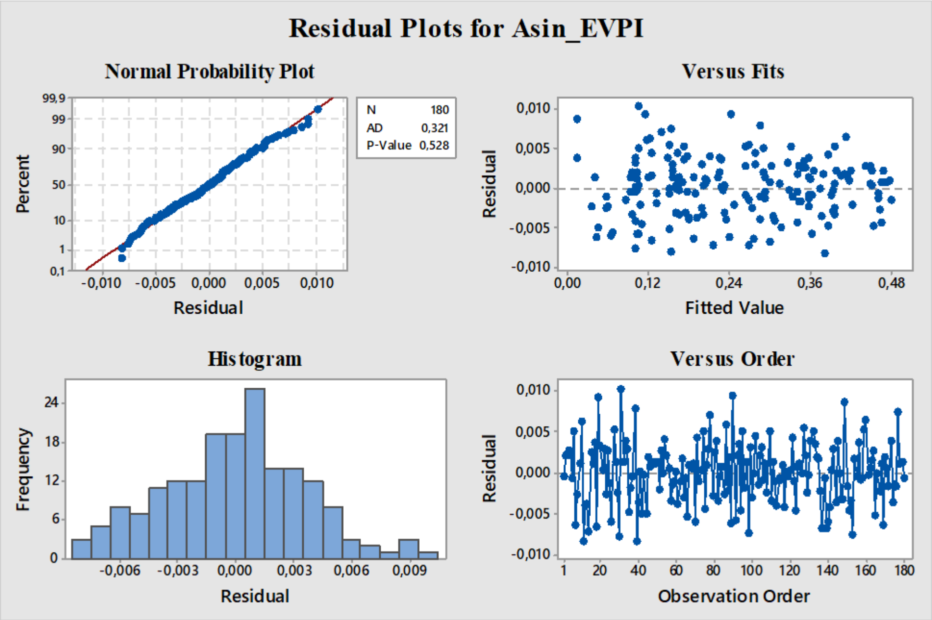


Figure C22. Residual plots for Case 4 (Deterministic Parameters: EVPI)

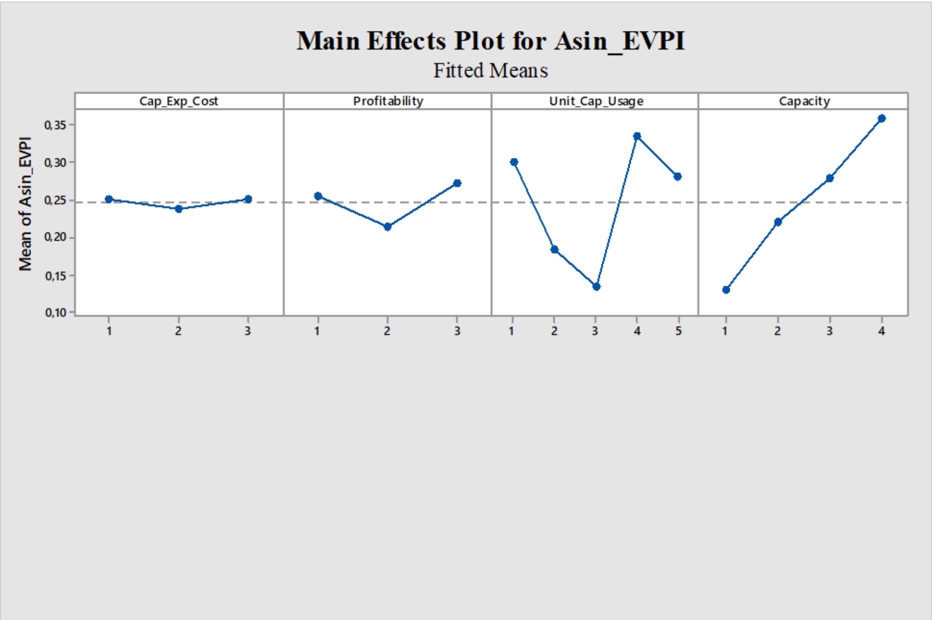


Figure C23. Main effects plot for Case 4 (Deterministic Parameters: EVPI)

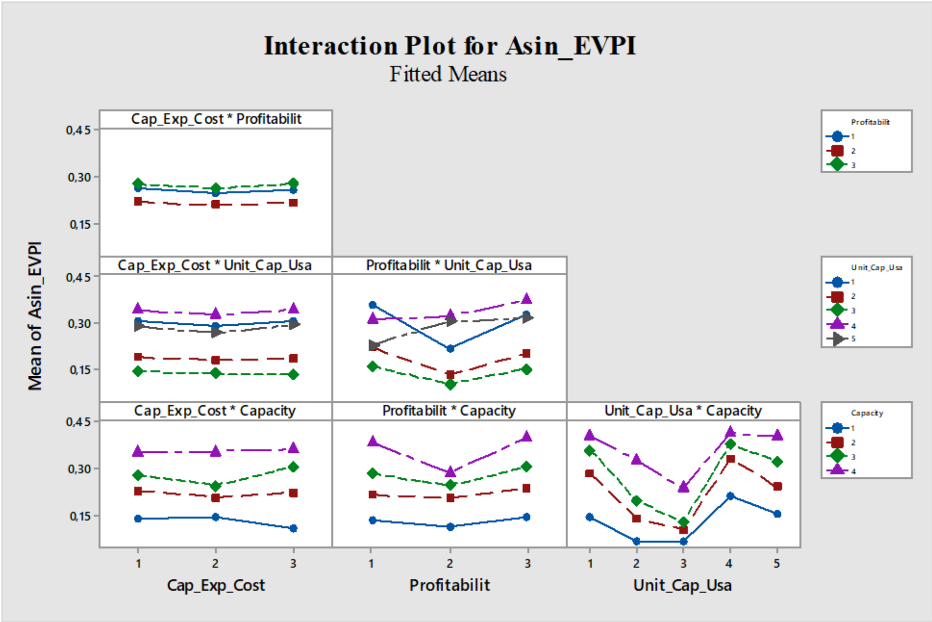


Figure C24. Interaction plot for Case 4 (Deterministic Parameters: EVPI)

According to Figure C22, the assumptions of the regression model whose R-sqr (adjusted) is 99.8 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:
randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Profitability.f  + Unit_Cap_Usage.f + Capac
             ity,data = cleandata, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omi
             t)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate: 7.22 %

Confusion matrix:
     H    L    M    class.error
  H  4    0    4    0.50000000
  L  0  118    3    0. 02479339
  M  2    4   45    0. 11764706

Importance of variables:
                         H        L        M     MeanDecreaseAccuracy   MeanDecreaseGini
  Cap_Exp_Cost.f     -14.45   -25.45   -25.40         -36.08                 4.08
  Profitability.f     28.05    37.22    33.36          51.45                16.30
  Unit_Cap_Usage.f    27.63    77.64    75.72          98.82                32.63
  Capacity.f          24.61    78.92    69.84          93.38                29.18
```

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for the revised model including three deterministic parameters:

```
Call:
randomForest(formula = EVPI.f ~ Profitability.f  + Unit_Cap_Usage.f + Capacity.f, data = cleanda
               ta, mtry = 2,   importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 2

        OOB estimate of  error rate: 2.78 %

Confusion matrix:
     H    L   M   class.error
  H  6    0   2   0.25000000
  L  0  120   1   0.00826446
  M  0    2  49   0.03921569

Importance of variables:
                      H      L      M    MeanDecreaseAccuracy   MeanDecreaseGini
  Profitability.f    33.62  45.98  42.37         65.03                14.04
  Unit_Cap_Usage.f   34.14  85.44  86.14        110.61                31.37
  Capacity.f         33.28  94.15  87.73        119.06                28.38
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,109	1	2,077	X[,3] %in% c('1','2')	L
2	2	0,100	1	2,077	X[,1] %in% c('2') & X[,2] %in% c('1')	L

C.4.2. Experiment 2

C.4.2.1. Response: VSS

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four uncertain parameters:

```
Call:
randomForest(formula = VSS.f ~ CV_Demand.f + CV_Price.f + CV_Cost.f + CV_Canb.Rate.f,dat
               a = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

          OOB estimate of  error rate: 10%
```

Confusion matrix:

	H	L	M	class.error
H	15	0	0	0.0
L	0	4	1	0.2
M	0	2	8	0.2

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-1.76	-5.37	-8.89	-8.31	0.96
CV_Price.f	47.10	36.81	37.91	60.77	10.79
CV_Cost.f	24.65	12.25	21.42	31.97	4.45
CV_Canb.Rate.f	-4.36	3.57	-4.75	-3.05	1.45

Since the mean decrease accuracy of CV_Demand.f and CV_Canb.Rate.f are negative, those parameters are excluded and the model is re-runned with the remaining parameters.

R output for a model including two uncertain parameters:

Call: randomForest(formula = VSS.f ~ CV_Price.f + CV_Cost.f, data = cleandata, mtry = 2,
importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 2

OOB estimate of error rate: 0%

Confusion matrix:

	H	M	class.error
H	8	0	0.000000
M	0	8	0.000000

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Price.f	60.89	51.78	60.78	91.11	11.96
CV_Cost.f	33.29	21.35	37.37	53.54	5.79

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,205	1	2,478	X[,1] %in% c('0,30')	H
2	2	0,178	1	2,623	X[,1] %in% c('0,15') & X[,2] %in% c('0','0,15')	M
3	2	0,167	1	2,623	X[,1] %in% c('0') & X[,2] %in% c('0,30')	M
4	2	0,164	1	4,648	X[,1] %in% c('0') & X[,2] %in% c('0','0,15')	L
5	2	0,164	1	2,478	X[,1] %in% c('0,15') & X[,2] %in% c('0,30')	H

C.4.2.2. Response: EVPI

Rules extracted from the Random Forest Application

R output for a model including all four uncertain parameters:

```
Call:
randomForest(formula = EVPI.f ~ CV_Demand.f + CV_Price.f + CV_Cost.f + CV_Canb.Rate.f,
                  data = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.action = na.omi
               t)
            Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 3

   OOB estimate of  error rate: 6.67%
```

Confusion matrix:

	L	M	class.error
L	13	2	0.1333333
M	0	15	0.0000000

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-5.70	-0.78	-4.28	0.80
CV_Price.f	43.88	46.26	54.61	9.42
CV_Cost.f	19.11	24.98	29.51	3.56
CV_Canb.Rate.f	-2.98	-2.71	-4.03	0.68

Since the mean decrease accuracy of CV_Demand.f and CV_Canb.Rate.f are negative, those parameters are excluded and the model is re-runned with the remaining parameters.

R output for a model including two uncertain parameters:

```
Call: randomForest(formula = EVPI.f ~ CV_Price.f + CV_Cost.f, data = cleandata, mtry = 2,
                          importance = TRUE,  ntree = 1000, na.action = na.omit)
            Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 2

     OOB estimate of  error rate: 0%
```

Confusion matrix:

	L	M	class.error
L	15	0	0.000
M	0	15	0.000

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Price.f	56.05	60.10	72.25	10.86
CV_Cost.f	30.35	35.06	42.40	3.65

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,259	1	1,944	X[,1] %in% c('0,30')	M
2	1	0,248	0,926	1,907	X[,2] %in% c('0','0,15')	L
3	1	0,239	0,909	1,872	X[,1] %in% c('0','0,15')	L
4	2	0,235	1	2,059	X[,1] %in% c('0','0,15') & X[,2] %in% c('0','0,15')	L
5	1	0,225	1	2,059	X[,1] %in% c('0')	L
6	1	0,206	0,750	1,458	X[,1] %in% c('0,15')	M
7	2	0,205	1	2,059	X[,1] %in% c('0') & X[,2] %in% c('0,30')	L

THE CROSSED ARRAY DESIGN FOR EXPERIMENT 1 WITH ALL FACTORS (FULL DESIGN)

C.5.1. Response: VSS

Rules extracted from the Random Forest Application

R output for a model including all seven deterministic parameters:

```
Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Newline.f + Market.f + Period.f + Profitabilit
                         y.f +Unit_Cap_Usage.f + Capacity.f,      data = cleandata, mtry = 4, impor
                         tance = TRUE,      ntree = 1000, na.action = na.omit)
             Type of random forest: classification
                   Number of trees: 1000
No. of variables tried at each split: 4

     OOB estimate of error rate: 11.53%

Confusion matrix:
       H     L     M     class.error
  H    6    14     3    0. 73913043
  L    0   605     8    0. 01305057
  M    0    58    26    0. 69047619

Importance of variables:
                        H        L        M     MeanDecreaseAccuracy   MeanDecreaseGini
  Newline.f           28.12    34.61    38.25          47.00                 14.65
  Market.f            21.96    19.22    26.59          29.84                 11.56
  Period.f            23.20    20.01    26.70          30.24                 12.17
  Cap_Exp_Cost.f     -21.19   -33.22   -34.63         -48.93                 16.59
  Profitability.f     24.64    13.02    32.23          29.53                 29.78
  Unit_Cap_Usage.f    30.70    28.28    45.60          48.72                 46.65
  Capacity.f          41.34    40.58    50.44          63.28                 31.74
```

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for a model including six deterministic parameters:

```
Call:
randomForest(formula = VSS.f ~ Newline.f + Market.f + Period.f + Profitability.f +
                          Unit_Cap_Usage.f  +    Capacity.f,  data = cleandata, mtry = 4,
                          importance = TRUE, ntree = 1000, na.action = na.omit)
             Type of random forest: classification
                   Number of trees: 1000
No. of variables tried at each split: 4

     OOB estimate of error rate: 7.92 %
```

Confusion matrix:

	H	L	M	class.error
H	16	3	4	0. 30434783
L	0	602	11	0. 01794454
M	3	36	45	0. 46428571

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Newline.f	36.11	53.84	59.75	71.35	14.59
Market.f	32.13	40.80	42.55	53.86	11.51
Period.f	30.79	39.25	43.49	52.40	11.24
Profitability.f	44.84	45.43	63.13	77.19	26.18
Unit_Cap_Usage.f	53.27	64.81	83.77	103.03	43.29
Capacity.f	56.75	74.66	82.99	113.86	28.74

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,245	0,844	1,028	X[,1] %in% c('1')	L
2	1	0,217	0,854	1,040	X[,2] %in% c('2')	L
3	1	0,206	0,757	0,922	X[,4] %in% c('1')	L
4	1	0,205	0,865	1,054	X[,3] %in% c('1')	L
5	1	0,154	0,887	1,081	X[,4] %in% c('2','3')	L
6	2	0,105	0,833	1,015	X[,1] %in% c('1') & X[,6] %in% c('1')	L
7	1	0,100	0,923	1,124	X[,5] %in% c('4')	L
8	1	0,097	0,846	1,031	X[,4] %in% c('3')	L
9	1	0,094	0,923	1,124	X[,6] %in% c('2')	L
10	1	0,093	0,99	1,206	X[,6] %in% c('2','3','4')	L
11	1	0,09	0,856	1,043	X[,6] %in% c('3')	L
12	2	0,081	0,751	0,915	X[,2] %in% c('2') & X[,6] %in% c('1')	L
13	1	0,077	0,827	1,007	X[,4] %in% c('1','3')	L
14	1	0,068	0,977	1,190	X[,5] %in% c('4','5')	L
15	1	0,065	0,977	1,190	X[,6] %in% c('3','4')	L
16	2	0,065	0,82	0,999	X[,1] %in% c('2') & X[,4] %in% c('1')	L
17	1	0,064	0,96	1,169	X[,6] %in% c('2','4')	L
18	1	0,063	0,912	1,111	X[,6] %in% c('4')	L
19	1	0,062	0,863	1,051	X[,5] %in% c('5')	L
20	1	0,061	0,764	0,931	X[,5] %in% c('3')	L
21	2	0,057	0,797	0,971	X[,1] %in% c('2') & X[,2] %in% c('2')	L
22	2	0,056	1	1,218	X[,1] %in% c('1') & X[,4] %in% c('2','3')	L
23	2	0,055	0,849	1,034	X[,1] %in% c('1') & X[,2] %in% c('1')	L
24	2	0,050	0,755	0,920	X[,2] %in% c('1') & X[,4] %in% c('1')	L
25	2	0,050	0,799	0,973	X[,1] %in% c('2') & X[,3] %in% c('1')	L

C.5.2. Response: EVPI

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all seven deterministic parameters:

```
Call:
randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Newline.f + Market.f + Period.f + Profitabili
                             ty.f + Unit_Cap_Usage.f  +  Capacity.f,     data = cleandata, mtry = 4,
                               importance = TRUE,  ntree = 1000, na.action = na.omit)
             Type of random forest: classification
                    Number of trees: 1000
No. of variables tried at each split: 4

      OOB estimate of error rate: 7.22%

Confusion matrix:
        H     L     M      class.error
   H    1     0     10    0. 90909091
   L    0    536    21    0. 03770197
   M    0     21   131    0. 13815789

Importance of variables:
                            H           L           M        MeanDecreaseAccuracy    MeanDecreaseGini
  Newline.f             12.32     19.32     19.07              26.30                        10.52
  Market.f              12.90     25.46     24.84              34.47                        12.80
  Period.f              17.11     21.89     24.65              31.04                        11.49
  Cap_Exp_Cost.f      -13.18    -41.34    -43.50             -58.95                        14.39
  Profitability.f       24.80     53.24     48.35              67.22                        38.45
  Unit_Cap_Usage.f    23.12    109.15   127.87            147.47                        85.57
  Capacity.f            23.19     92.12     93.60            118.79                        64.81
```

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

R output for a model including six deterministic parameters:

```
Call: randomForest(formula = EVPI.f ~ Newline.f + Market.f + Period.f + Profitability.f
                                       + Unit_Cap_Usage.f + Capacity.f,  data = cleandata, mtry = 4,
                                       importance = TRUE, ntree = 1000,   na.action = na.omit)
             Type of random forest: classification
                    Number of trees: 1000
No. of variables tried at each split: 4

      OOB estimate of error rate: 4.31 %

Confusion matrix:
        H     L     M      class.error
   H    6     0     5     0. 45454545
   L    0    545    12    0. 02154399
   M    1     13   138    0. 09210526
```

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Newline.f	18.23	33.30	35.38	44.22	10.96
Market.f	22.35	44.68	43.95	56.98	13.46
Period.f	22.43	38.14	37.89	48.48	11.74
Profitability.f	34.89	87.88	82.57	116.92	40.05
Unit_Cap_Usage.f	35.85	146.54	165.21	198.92	82.81
Capacity.f	34.84	125.79	129.86	168.81	63.10

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,219	0,77	1,187	X[,2] %in% c('1')	L
2	1	0,14	0,762	1,175	X[,6] %in% c('2')	L
3	1	0,126	0,752	1,160	X[,4] %in% c('2')	L
4	1	0,084	1	1,542	X[,5] %in% c('2','3')	L
5	2	0,067	0,835	1,288	X[,2] %in% c('1') & X[,5] %in% c('1','4','5')	L
6	2	0,061	0,959	1,479	X[,2] %in% c('1') & X[,3] %in% c('2')	L
7	1	0,056	1	1,542	X[,6] %in% c('1')	L
8	1	0,055	0,755	1,164	X[,4] %in% c('1','2')	L
9	2	0,049	0,968	1,493	X[,1] %in% c('2') & X[,2] %in% c('1')	L
10	2	0,042	0,796	1,227	X[,3] %in% c('2') & X[,5] %in% c('1','4','5')	L
11	2	0,042	0,82	1,264	X[,2] %in% c('1') & X[,6] %in% c('4')	L

THE CROSSED ARRAY DESIGN FOR EXPERIMENT 2 WITH ALL FACTORS (FULL DESIGN)

C.6.1. Response: VSS

Rules extracted from the Random Forest Application

R output for a model including all seven parameters and factors:

```
Call:
 randomForest(formula = VSS.f ~ Newline.f + Market.f + Period.f + CV_Price.f + CV_Cost.f +
                                CV_Demand.f + CV_Canb.Rate.f, data = cleandata, mtry = 4,
                                importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 4

        OOB estimate of  error rate: 10 %

Confusion matrix:
      H    L    M    class.error
  H  49    0    1    0. 05769231
  L   0   27    2    0. 06896552
  M   6    1   32    0. 17948718

Importance of variables:
```

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Newline.f	21.29	20.82	31.90	36.92	7.54
Market.f	20.16	21.97	29.99	35.52	7.08
Period.f	20.32	20.72	28.99	35.08	7.21
CV_Price.f	78.35	80.70	72.33	109.87	32.36
CV_Cost.f	54.89	34.47	52.94	73.92	13.93
CV_Canb.Rate.f	-2.10	22494	-8.36	-1.28	3.75
CV_Demand.f	1.20	1.81	-8.29	-3.37	4.12

Since the mean decrease accuracy of the coefficient of variation of demand and cannibalisation rate are negative, these parameters are excluded and the model is re-runned with the remaining parameters.

R output for a model including five parameters and factors:

```
Call:
randomForest(formula = VSS.f ~ Newline.f + Market.f + Period.f + CV_Price.f + CV_Cost.f, data
                  = cleandata,  mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate: 2.5 %
```

Confusion matrix:

	H	L	M	class.error
H	51	0	1	0.01923077
L	0	28	1	0.03448276
M	0	1	38	0.02564103

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Newline.f	25.96	24.15	38.50	43.81	6.94
Market.f	28.23	23.92	35.54	44.61	7.07
Period.f	26.80	25.93	33.86	42.19	7.03
CV_Price.f	98.27	90.75	99.07	148.76	34.40
CV_Cost.f	66.25	45.51	66.07	94.16	15.08

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,181	0,754	1,952	X[,4] %in% c('0,30')	H
2	2	0,095	0,754	2,155	X[,4] %in% c('0,15') & X[,5] %in% c("0,15')	M
3	2	0,073	0,786	2,176	X[,2] %in% c('1') & X[,5] %in% c('0,30')	H
4	2	0,071	0,779	2,157	X[,3] %in% c('2') & X[,5] %in% c('0,30')	H
5	2	0,066	1	3,461	X[,4] %in% c('0') & X[,5] %in% c('0','0,15')	L
6	2	0,055	0,754	2,578	X[,3] %in% c('1') & X[,4] %in% c('0')	L

C.6.2. Response: EVPI

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all seven parameters and factors:

Call:
randomForest(formula = EVPI.f ~ Newline.f + Market.f + Period.f + CV_Price.f + CV_Cost.f + CV_Demand.f+ CV_Canb.Rate.f, data = cleandata, mtry = 4, importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 4

OOB estimate of error rate: 2.5%

Confusion matrix:

	L	M	class.error
L	79	1	0.0125
M	2	38	0.0500

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Newline.f	18.58	18.77	20.30	3.67
Market.f	19.19	17.82	20.21	3.46
Period.f	77.19	77.19	86.78	25.88
CV_Price.f	51.11	40.51	58.00	12.16
CV_Cost.f	16.56	17.85	22.71	4.49
CV_Canb.Rate.f	3.39	-5.88	-1.35	1.55
CV_Demand.f	-9.63	-5.85	-11.15	1.31

Since the mean decrease accuracy of the coefficient of variation of demand and cannibalisation rate are negative, we exclude these parameters and re-run the model with the remaining parameters.

R output for a model including five parameters and factors:

```
Call:
randomForest(formula = EVPI.f ~ Newline.f + Market.f + Period.f + CV_Price.f + CV_Cost.f, dat
                 a = cleandata,  mtry = 4, importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate:  0 %

Confusion matrix:
         L    M    class.error
   L    80    0         0
   M     0   40         0
```

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Newline.f	18.58	18.77	20.30	3.67
Market.f	19.19	17.82	20.21	3.46
Period.f	77.19	77.19	86.78	25.88
CV_Price.f	51.11	40.51	58.00	12.16
CV_Cost.f	16.56	17.85	22.71	4.49
CV_Canb.Rate.f	3.39	-5.88	-1.35	1.55
CV_Demand.f	-9.63	-5.85	-11.15	1.31

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,229	1,0	1,775	X[,4] %in% c('0')	L
2	1	0,164	1,0	1,775	X[,3] %in% c('2')	L
3	2	0,140	0,762	1,745	X[,3] %in% c('1') & X[,5] %in% c('0,30')	M
4	1	0,130	0,812	1,860	X[,4] %in% c('0,30')	M
5	2	0,117	1,0	2,290	X[,3] %in% c('1') & X[,4] %in% c('0,15', '0,30')	M
6	2	0,103	0,751	1,333	X[,3] %in% c('1') & X[,5] %in% c('0','0,15')	L
7	2	0,061	1,0	1,775	X[,4] %in% c('0') & X[,5] %in% c('0','0,15')	L

CASE 5

C.7.1. Experiment 1

C.7.1.1. Response: VSS

ANOVA Table for General Factorial Regression (Asin_VSS versus factors)

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of VarianceSource	DF	Adj SS	Adj MS	F-Value	P-Value
Model	119	2,93651	0,024677	15,42	0,000
Linear	11	0,72718	0,066107	41,31	0,000
Cap_Exp_Cost	2	0,00609	0,003043	1,90	0,158
Profitability	2	0,16608	0,083040	51,90	0,000
Unit_Cap_Usage	4	0,41116	0,102789	64,24	0,000
Capacity	3	0,14385	0,047951	29,97	0,000
2-Way Interactions	44	1,69088	0,038429	24,02	0,000
Cap_Exp_Cost*Profitability	4	0,01708	0,004269	2,67	0,041
Cap_Exp_Cost*Unit_Cap_Usage	8	0,03247	0,004059	2,54	0,019
Cap_Exp_Cost*Capacity	6	0,02747	0,004578	2,86	0,016
Profitability*Unit_Cap_Usage	8	1,16832	0,146040	91,27	0,000
Profitability*Capacity	6	0,23009	0,038349	23,97	0,000
Unit_Cap_Usage*Capacity	12	0,21545	0,017954	11,22	0,000
3-Way Interactions	64	0,51846	0,008101	5,06	0,000
Cap_Exp_Cost*Profitability*Unit_Cap_Usage	16	0,04897	0,003061	1,91	0,037
Cap_Exp_Cost*Unit_Cap_Usage*Capacity	24	0,10405	0,004335	2,71	0,001
Profitability*Unit_Cap_Usage*Capacity	24	0,36544	0,015227	9,52	0,000
Error	60	0,09600	0,001600		
Total	179	3,03252			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0400010	96,83%	90,56%	71,51%

Residual Plots

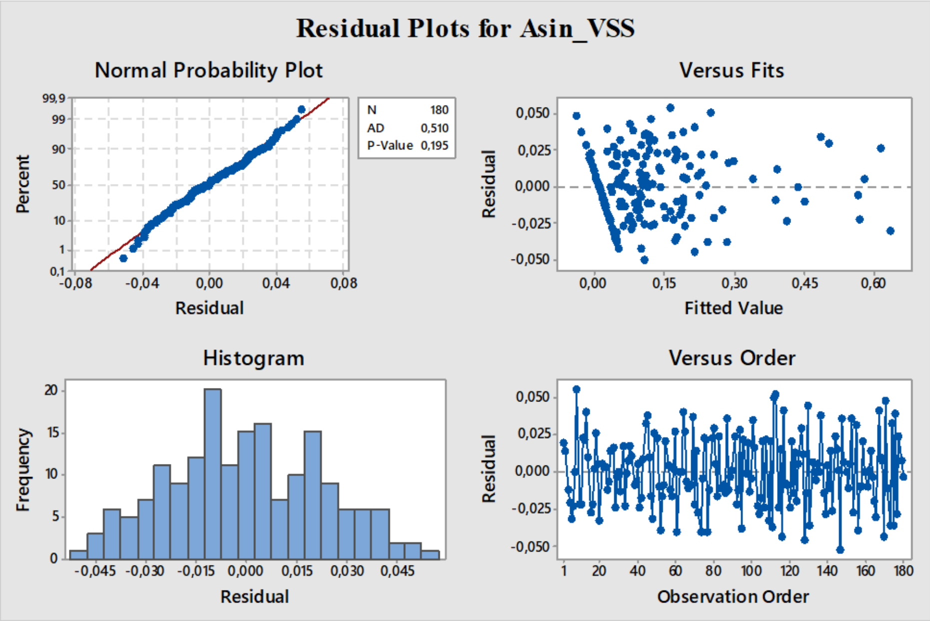


Figure C25. Residual plots for Case 5 (Deterministic Parameters: VSS)

Figure C26. Main effects plot for Case 5 (Deterministic Parameters: VSS)

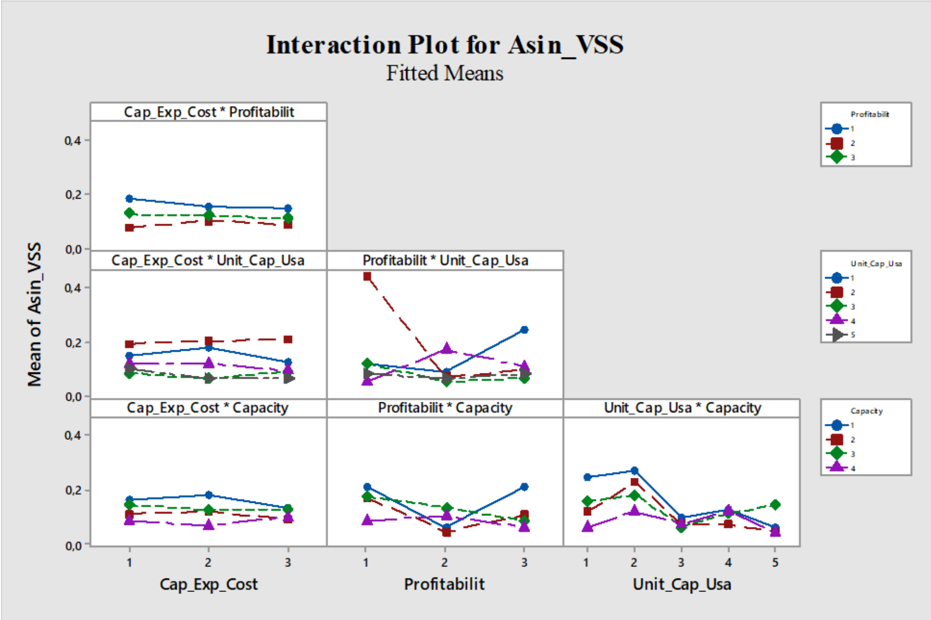


Figure C27. Interaction plot for Case 5 (Deterministic Parameters: VSS)

According to Figure C25, the assumptions of the regression model whose R-sqr (adjusted) is 90.6 % are satisfied.

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four deterministic parameters:

Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacity .f, data = cleandata, mtry = 3,importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 2.78%

Confusion matrix:

	H	L	M	class.error
H	6	1	0	0. 1428571
L	0	166	0	0. 0000000
M	0	4	3	0. 5714286

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-10.07	-14.44	-11.42	-19.74	1.98
Profitability.f	34.15	30.61	21.82	44.15	8.55
Unit_Cap_Usage.f	39.85	29.14	17.84	41.42	7.20
Capacity.f	22.48	11.79	23.7	25.95	7.81

Since the mean decrease accuracy of capacity expansion cost is negative, we exclude this parameter and re-run the model with the remaining parameters.

***R output for the revised model including three deterministic parameters*:**

```
Call: randomForest(formula = VSS.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f,
                               data = cleandata, mtry = 3, importance = TRUE, ntree = 1000,
                               na.action = na.omit)
          Type of random forest: classification
                Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of  error rate: 2.78%

Confusion matrix:
      H    L    M    class.error
  H   6    0    1    0.1428571
  L   0  166    0    0.0000000
  M   2    2    3    0.5714286

Importance of variables:
                      H      L      M    MeanDecreaseAccuracy  MeanDecreaseGini
  Profitability.f   42.44  40.77  31.14         60.22               9.43
  Unit_Cap_Usage.f  52.09  44.77  30.20         65.99               5.77
  Capacity.f        30.77  26.15  34.10         48.92               8.48
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,158	1	1,300	X[,3] %in% c('4')	L
2	1	0,133	1	1,300	X[,1] %in% c('2','3')	L
3	1	0,096	1	1,300	X[,1] %in% c('2')	L
4	1	0,085	1	1,300	X[,2] %in% c('4')	L
5	1	0,077	0,996	1,295	X[,1] %in% c('1','2')	L
6	1	0,074	1	1,300	X[,2] %in% c('5')	L
7	3	0,073	1	7,598	X[,1] %in% c('1') & X[,2] %in% c('2') & X[,3] %in% c('4')	M
8	2	0,073	0,971	7,378	X[,2] %in% c('2') & X[,3] %in% c('4')	M
9	2	0,072	1	1,300	X[,1] %in% c('1') & X[,3] %in% c('4')	L
10	2	0,068	1	1,300	X[,1] %in% c('2','3') & X[,2] %in% c('2')	L
11	2	0,066	0,963	9,717	X[,1] %in% c('1') & X[,3] %in% c('1','2')	H
14	2	0,065	1	1,300	X[,2] %in% c('2') & X[,3] %in% c('4')	L
15	1	0,06	1	1,300	X[,2] %in% c('3','4','5')	L
16	3	0,057	1	7,598	X[,1] %in% c('3') & X[,2] %in% c('1') & X[,3] %in% c('1')	M
17	2	0,057	0,787	5,980	X[,1] %in% c('3') & X[,2] %in% c('1')	M
18	1	0,056	0,997	1,296	X[,3] %in% c('2','3','4')	L

C.7.1.2. Response: EVPI

ANOVA Table for General Factorial Regression (Asin_EVPI versus factors)

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Max_Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	37	1,48167	0,040045	57,08	0,000
Linear	11	1,06154	0,096504	137,57	0,000
Cap_Exp_Cost	2	0,00188	0,000942	1,34	0,265
Profitability	2	0,10714	0,053572	76,37	0,000
Unit_Cap_Usage	4	0,34602	0,086506	123,31	0,000
Capacity	3	0,60649	0,202164	288,18	0,000
2-Way Interactions	26	0,42013	0,016159	23,03	0,000
Cap_Exp_Cost*Capacity	6	0,06876	0,011460	16,34	0,000
Profitability*Unit_Cap_Usage	8	0,22592	0,028239	40,26	0,000
Unit_Cap_Usage*Capacity	12	0,12545	0,010454	14,90	0,000
Error	142	0,09961	0,000702		
Total	179	1,58129			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0264861	93,70%	92,06%	89,88%

Residual Plots

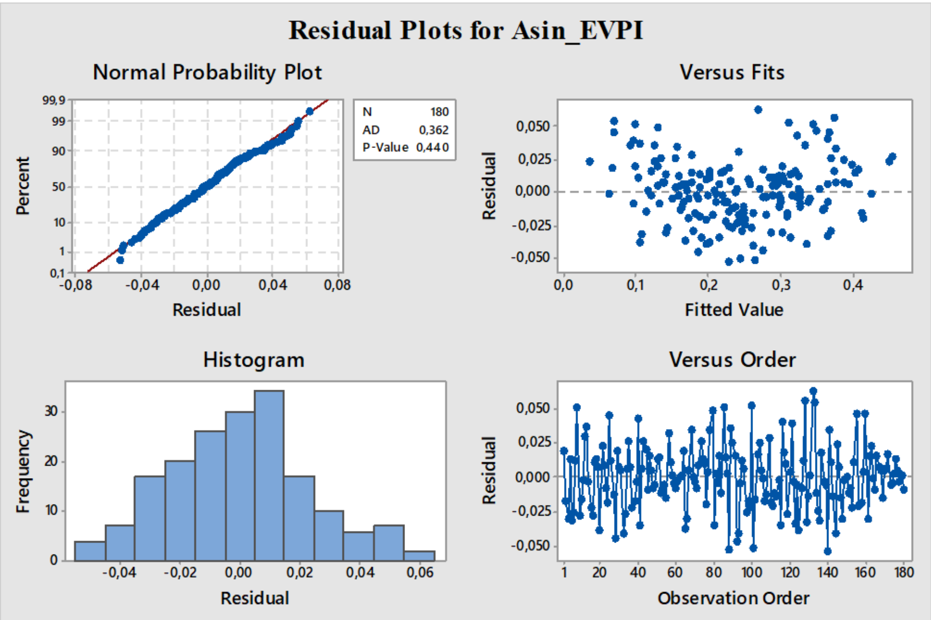


Figure C28. Residual plots for Case 5 (Deterministic Parameters: EVPI)

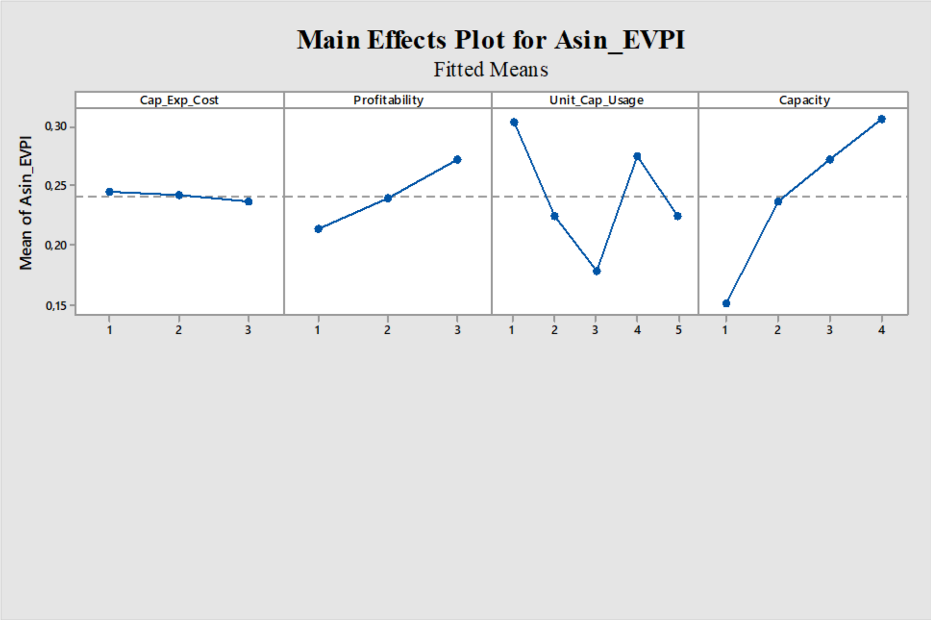


Figure C29. Main effects plot for Case 5 (Deterministic Parameters: EVPI)

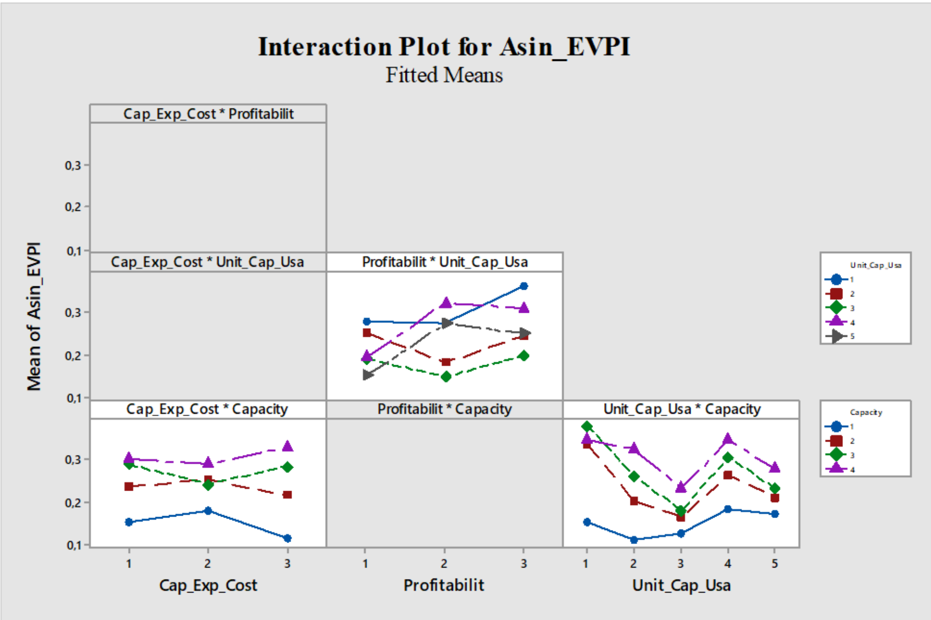


Figure C30. Interaction plot for Case 5 (Deterministic Parameters: EVPI)

According to Figure C28, the assumptions of the regression model whose R-sqr (adjusted) is 92 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:  randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f +
                                       Capacity.f,   data =cleandata, mtry = 3,  importance = TR
UE,
                                       ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

                OOB estimate of  error rate: 13.9 %
```

Confusion matrix:

	H	L	M	class.error
H	0	0	2	1.00000000
L	0	123	12	0. 08888889
M	0	11	32	0. 25581395

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-5.68	-22.32	-21.70	-30.25	6.54
Profitability.f	9.58	18.44	21.59	26.65	12.83
Unit_Cap_Usage.f	10.77	49.95	50.36	63.83	24.80
Capacity.f	10.95	44.12	45.57	57.29	22.06

Since the mean decrease accuracy of capacity expansion cost is negative, this para meter is excluded and the model is re-runned with the remaining parameters.

R output for the revised model including three deterministic parameters:

```
Call:
 randomForest(formula = EVPI.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f, data = cleanda
              ta, mtry = 3,   importance = TRUE,  ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

        OOB estimate of  error rate: 8.33%
```

Confusion matrix:

	H	L	M	class.error
H	0	0	2	1.000000000
L	0	132	3	0. 02222222
M	1	9	33	0. 23255814

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Profitability.f	13.60	33.17	37.06	47.98	11.05
Unit_Cap_Usage.f	14.61	64.30	70.45	83.78	23.06
Capacity.f	14.26	56.80	62.78	78.15	20.78

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,196	0,763	1,267	X[,1] %in% c('1')	L
2	1	0,075	1	1,660	X[,2] %in% c('3')	L
3	2	0,071	1	1,660	X[,2] %in% c('3') & X[,3] %in% c('4')	L
4	2	0,062	0,968	2,648	X[,1] %in% c('2') & X[,2] %in% c('4', '5')	M
5	1	0,061	1	1,660	X[,3] %in% c('1')	L
6	2	0,06	0,776	1,288	X[,1] %in% c('2') & X[,2] %in% c('1')	L
7	2	0,051	1	1,660	X[,1] %in% c('1') & X[,2] %in% c('4')	L
8	2	0,049	0,82	1,361	X[,1] %in% c('1') & X[,3] %in% c('3')	L
9	1	0,047	1	1,660	X[,3] %in% c('1','2')	L
10	1	0,047	1	1,660	X[,2] %in% c('2','3','5')	L

C.7.2. Experiment 2

C.7.2.1. Response: VSS

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four uncertain parameters:

```
Call:
randomForest(formula = VSS.f ~  CV_Demand.f  +  CV_Price.f  +  CV_Cost.f  +  CV_Canb.Rate.
                 f,                data = cleandata, mtry = 3, importance = TRUE,      ntree = 1000, na.ac
                 tion = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

      OOB estimate of  error rate: 6.67 %
```

Confusion matrix:

	H	L	M	class.error
H	11	0	0	0.00000000
L	0	7	1	0.12500000
M	0	1	10	0.09090909

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-3.15	-5.48	-2.46	-6.06	0.71
CV_Price.f	76.39	29.51	39.89	74.91	11.88
CV_Cost.f	-3.68	30.50	25.96	34.15	4.95
CV_Canb.Rate.f	-3.44	21520	43438	11.15	1.57

Since the mean decrease accuracy of CV_Demand is negative, we exclude this parameter and re-run the model with the remaining parameters.

***R output for the revised model including three uncertain parameters*:**

```
Call:
randomForest(formula = VSS.f ~ + CV_Price.f + CV_Cost.f + CV_Cannb.Rate.f, data = cleandata
              , mtry = 3, importance = TRUE,   ntree = 1000,  na.action = na.omit)
           Type of random forest: classification
                 Number of trees: 1000
No. of variables tried at each split: 3

     OOB estimate of  error rate: 0%

Confusion matrix:
    H  L  M   class.error
H  11  0  0   0.00000000
L   0  7  1   0.12500000

Importance of variables:
                    H      L      M    MeanDecreaseAccuracy   MeanDecreaseGini
CV_Price.f       80.21  36.24  46.11          83.12                 11.95
CV_Cost.f        -0.43  35.47  29.55          40.74                  5.99
CV_Canb.Rate.f   -1.00  12.63   1.13          10.67                  1.20
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,218	1	4,580	X[,1] %in% c('0,30')	H
2	1	0,19	0,991	2,507	X[,2] %in% c('0')	L
3	1	0,171	0,832	2,104	X[,1] %in% c('0,15')	M
4	2	0,159	1	2,589	X[,1] %in% c('0','0,15') & X[,2] %in% c('0')	L
5	2	0,13	1	2,529	X[,1] %in% c('0,15') & X[,2] %in% c('0,15','0,30')	M

CASE 6

C.8.1. Experiment 1

C.8.1.1. Response: VSS

ANOVA Table for General Factorial Regression (Asin_VSS versus factors)

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	61	2,83436	0,046465	27,16	0,000
Linear	11	1,26589	0,115081	67,28	0,000
Cap_Exp_Cost	2	0,01724	0,008618	5,04	0,008
Profitability	2	0,33485	0,167423	97,88	0,000
Unit_Cap_Usage	4	0,72802	0,182005	106,40	0,000
Capacity	3	0,18579	0,061931	36,21	0,000
2-Way Interactions	26	1,11222	0,042778	25,01	0,000
Profitability*Unit_Cap_Usage	8	0,87174	0,108968	63,70	0,000
Profitability*Capacity	6	0,05155	0,008592	5,02	0,000
Unit_Cap_Usage*Capacity	12	0,18892	0,015744	9,20	0,000
3-Way Interactions	24	0,45625	0,019010	11,11	0,000
Profitability*Unit_Cap_Usage*Capacity	24	0,45625	0,019010	11,11	0,000
Error	118	0,20185	0,001711		
Total	179	3,03621			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0413588	93,35%	89,92%	84,53%

Residual Plots

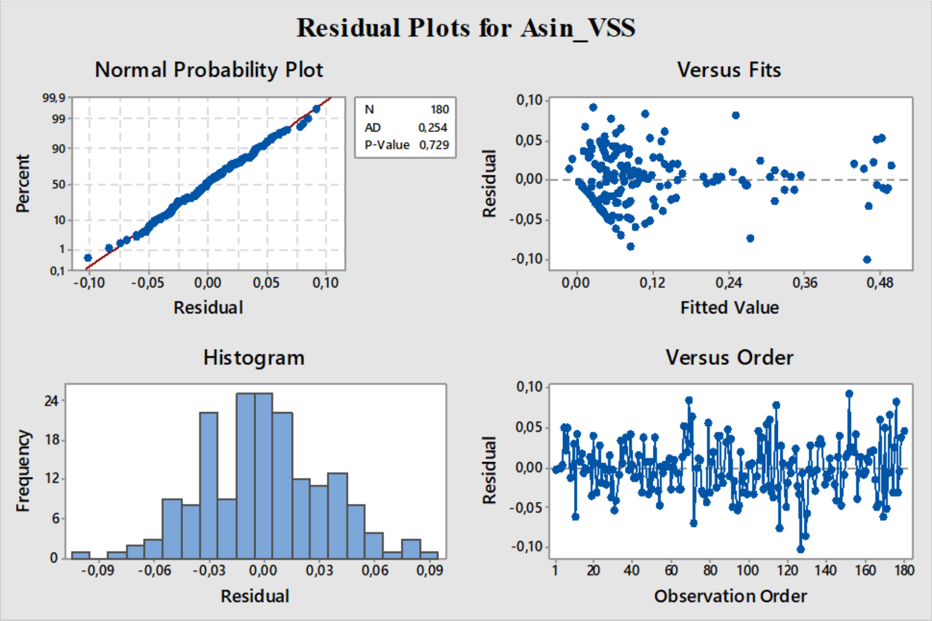


Figure C31. Residual plots for Case 6 (Deterministic Parameters: VSS)

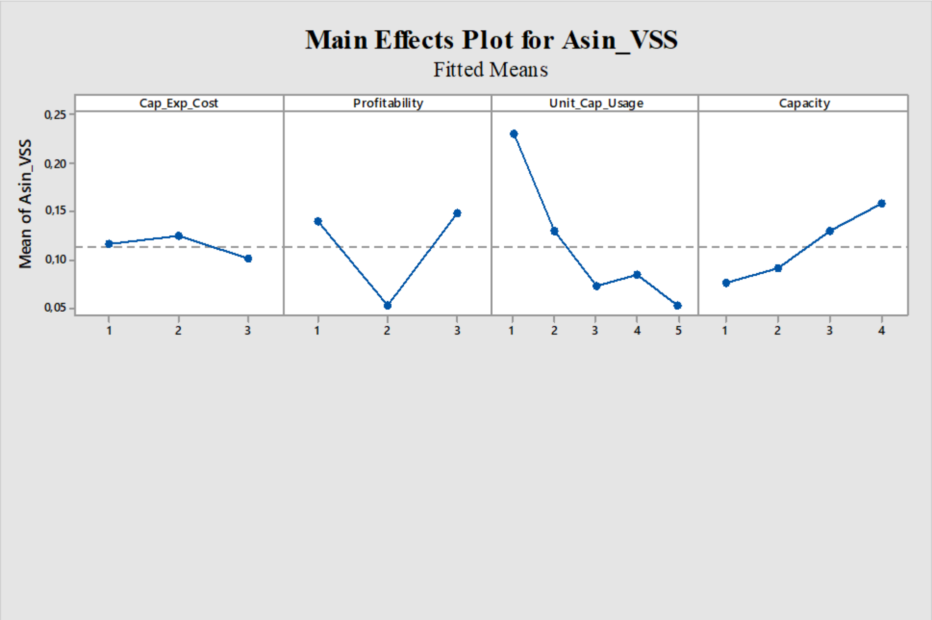


Figure C32. Main effects plot for Case 6 (Deterministic Parameters: VSS)

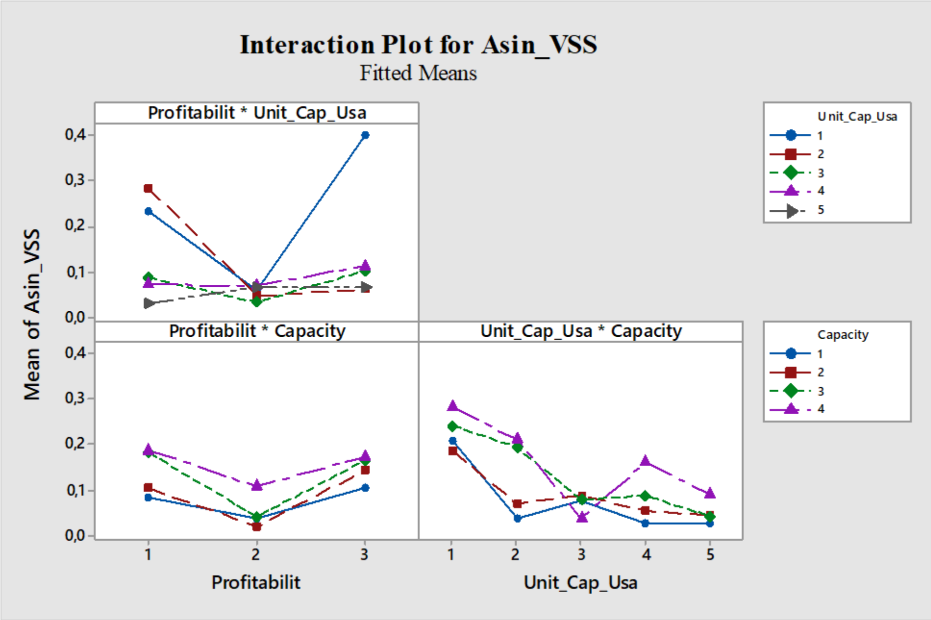


Figure C33. Interaction plot for Case 6 (Deterministic Parameters: VSS)

According to Figure C31, the assumptions of the regression model whose R-sqr (adjusted) is 89.9 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four deterministic parameters:

```
Call:
randomForest(formula = VSS.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacity
               .f,          data = cleandata, mtry = 3,importance = TRUE, ntree = 1000, na.action
               = na.omit)
            Type of random forest: classification
                  Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of  error rate: 2.22 %

Confusion matrix:
      H    L    M    class.error
  H   9    0    1    0. 1000000
  L   0  161    0    0. 0000000
  M   2    1    6    0. 3333333

Importance of variables:
                      H        L       M     MeanDecreaseAccuracy   MeanDecreaseGini
 Cap_Exp_Cost.f    -13.39  -15.55   -9.32         -21.56                 2.55
 Profitability.f    35.74   47.36   36.45          63.12                12.24
 Unit_Cap_Usage.f   42.71   50.01   42.03          69.92                12.33
 Capacity.f         22.20    4.32   14.07          20.68                 7.12
```

Since the mean decrease accuracy of capacity expansion cost is negative, this parameter is excluded and the model is re-runned with the remaining parameters.

R output for the revised model including three deterministic parameters:

```
Call:
randomForest(formula = VSS.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f, data = cleandata
,
                          mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of  error rate:  2.22 %

Confusion matrix:
      H    L    M    class.error
  H   9    0    1    0. 1000000
  L   0  161    0    0. 0000000
  M   2    1    6    0. 3333333

Importance of variables:
                      H      L      M    MeanDecreaseAccuracy  MeanDecreaseGini
  Profitability.f   38.39  56.26  40.19         74.20              11.30
  Unit_Cap_Usage.f  46.14  59.64  49.20         79.35              11.94
  Capacity.f        26.51  14.46  17.68         32.65               5.46
```

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	2	0,142	0,793	2,786	X[,1] %in% c('1') & X[,2] %in% c('2')	H
2	1	0,123	0,997	1,590	X[,1] %in% c('1','2')	L
3	1	0,101	1	1,594	X[,2] %in% c('3','4','5')	L
4	1	0,093	1	1,594	X[,2] %in% c('4')	L
5	2	0,079	1	1,594	X[,2] %in% c('4') & X[,3] %in% c('4')	L
6	3	0,076	1	11,341	X[,1] %in% c('3') & X[,2] %in% c('1') & X[,3] %in% c('1','2')	M
7	2	0,076	0,892	10,116	X[,2] %in% c('1') & X[,3] %in% c('1','2')	M
8	2	0,068	1	1,594	X[,1] %in% c('3') & X[,2] %in% c('2')	L
9	2	0,065	1	1,594	X[,2] %in% c('2') & X[,3] %in% c('1','2')	L
10	1	0,063	0,908	1,448	X[,3] %in% c('1','2','3')	L

C.8.1.2. Response: EVPI

ANOVA Table for General Factorial Regression (Logit_EVPI versus factors)

Factor Information

Factor	Levels	Values
Cap_Exp_Cost	3	1; 2; 3
Profitability	3	1; 2; 3
Unit_Cap_Usage	5	1; 2; 3; 4; 5
Max_Capacity	4	1; 2; 3; 4

Backward Elimination of Terms, α to remove = 0,1

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	51	143,128	2,8064	230,46	0,000
Linear	11	105,858	9,6234	790,27	0,000
Cap_Exp_Cost	2	0,933	0,4665	38,31	0,000
Profitability	2	5,794	2,8968	237,89	0,000
Unit_Cap_Usage	4	45,600	11,4001	936,17	0,000
Capacity	3	53,531	17,8435	1465,30	0,000
2-Way Interactions	40	37,271	0,9318	76,52	0,000
Cap_Exp_Cost*Unit_Cap_Usage	8	0,637	0,0796	6,54	0,000
Cap_Exp_Cost*Capacity	6	1,975	0,3291	27,03	0,000
Profitability*Unit_Cap_Usage	8	19,089	2,3862	195,95	0,000
Profitability*Capacity	6	0,615	0,1026	8,42	0,000
Unit_Cap_Usage*Capacity	12	14,954	1,2462	102,34	0,000
Error	128	1,559	0,0122		
Total	179	144,687			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,110351	98,92%	98,49%	97,87%

Residual Plots

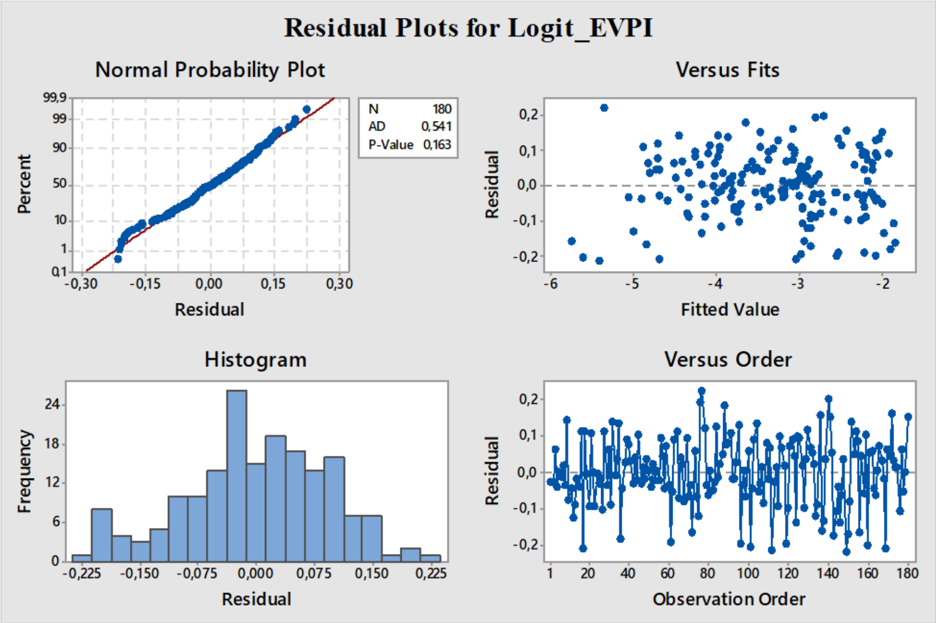


Figure C34. Residual plots for Case 6 (Deterministic Parameters: EVPI)

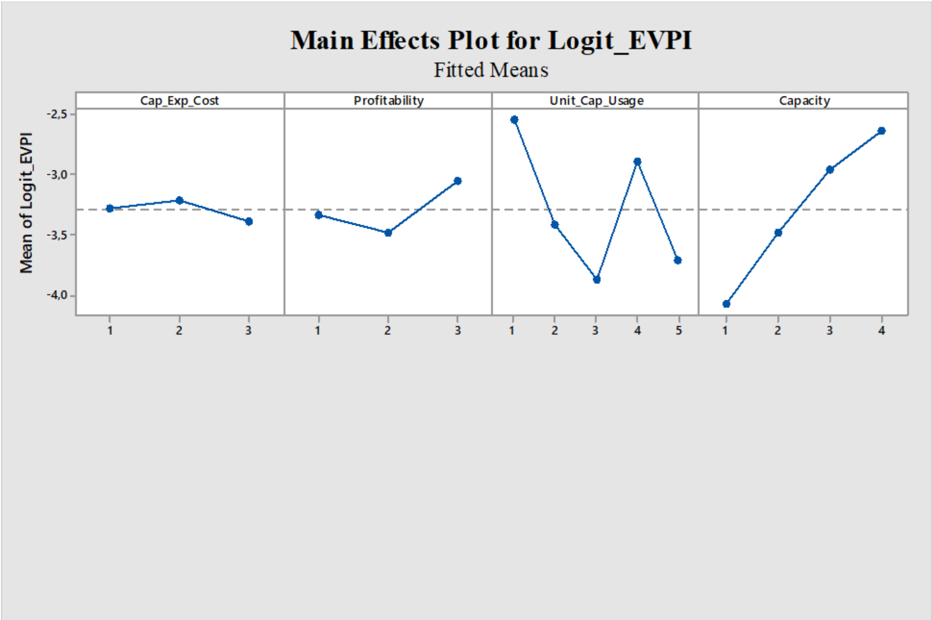


Figure C35. Main effects plot for Case 6 (Deterministic Parameters: EVPI)

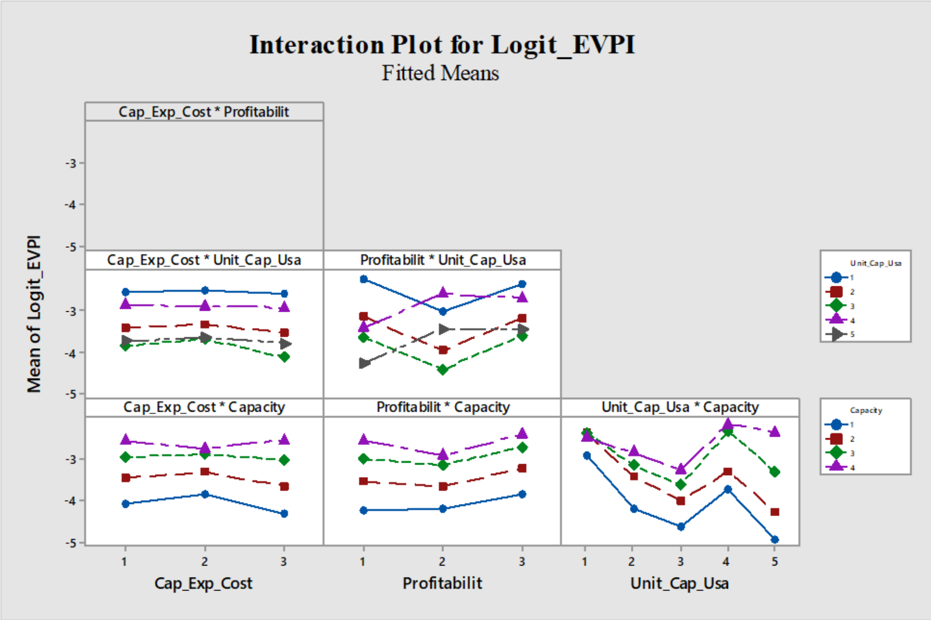


Figure C36. Interaction plot for Case 6 (Deterministic Parameters: EVPI)

According to Figure C34, the assumptions of the regression model whose R-sqr (adjusted) is 98.5 % are satisfied.

<u>Rules extracted from the Random Forest Application</u>

R output for a model including all four deterministic parameters:

Call:
randomForest(formula = EVPI.f ~ Cap_Exp_Cost.f + Profitability.f + Unit_Cap_Usage.f + Capacity.f, data =cleandata, mtry = 3, importance = TRUE, ntree = 1000, na.action = na.omit)

Type of random forest: classification

Number of trees: 1000

No. of variables tried at each split: 3

OOB estimate of error rate: 8.9 %

Confusion matrix:

	L	M	class.error
L	148	6	0. 03896104
M	10	16	0. 38461538

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
Cap_Exp_Cost.f	-17.12	-16.82	-22.28	4.69
Profitability.f	20.11	22.51	27.21	10.83
Unit_Cap_Usage.f	41.75	57.12	60.04	16.00
Capacity.f	18.24	25.51	27.37	10.86

Since the mean decrease accuracy of capacity expansion cost is negative, this parameter is excluded and the model is re-runned with the remaining parameters.

***R output for the revised model including three deterministic parameters*:**

```
Call:
 randomForest(formula = EVPI.f ~ Profitability.f + Unit_Cap_Usage.f + Capacity.f, data = cleanda
                    ta, mtry = 2,   importance = TRUE,  ntree = 1000, na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 2

       OOB estimate of  error rate: 8.89 %

Confusion matrix:
        L     M     class.error
   L    148    6    0. 03896104
   M     10    16   0. 38461538

Importance of variables:
                            L         M      MeanDecreaseAccuracy    MeanDecreaseGini
  Profitability.f        33.20    35.46             45.57                    8.86
  Unit_Cap_Usage.f   52.35    73.60             76.69                   16.11
  Capacity.f             31.25    37.48             43.72                    9.77
```

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	2	0,105	0,902	2,348	X[,1] %in% c('1') & X[,2] %in% c('1')	M
2	2	0,08	1	1,624	X[,1] %in% c('2') & X[,2] %in% c('1')	L
3	2	0,077	1	1,624	X[,1] %in% c('1') & X[,2] %in% c('4')	L
4	2	0,074	0,846	2,202	X[,1] %in% c('2','3') & X[,2] %in% c('4','5')	M
5	2	0,067	0,767	1,997	X[,2] %in% c('4','5') & X[,3] %in% c('3','4')	M
6	1	0,063	1	1,624	X[,3] %in% c('1')	L
7	1	0,062	1	1,624	X[,2] %in% c('2','3','5')	L
8	2	0,059	0,97	2,525	X[,1] %in% c('3') & X[,3] %in% c('3')	M

C.8.2. Experiment 2

C.8.2.1. Response: VSS

<u>Rules extracted from the Random Forest Application</u>

***R output for a model including all four uncertain parameters*:**

```
Call:
randomForest(formula = VSS.f ~  CV_Demand.f  + CV_Price.f  +  CV_Cost.f  +  CV_Canb.Rate.,
                 data = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.action = na.omit)
```

Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 3

OOB estimate of error rate: 20%

Confusion matrix:

	H	L	M	class.error
H	20	0	1	0.04761905
L	0	4	1	0.20000000
M	2	2	0	100000000

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	20.95	12.93	8.69	24.25	3.60
CV_Price.f	46.39	33.16	15.05	51.99	8.06
CV_Cost.f	1.68	-2.70	-6.28	-3.04	1.12
CV_Canb.Rate.f	1.00	-7.78	-7.34	-6.90	0.73

Since the mean decrease accuracy of CV_Cost and CV_Cannibalisation Rate are negative, we exclude these parametesr and re-run the model with the remaining parameters.

***R output for the revised model including two uncertain parameters*:**

Call:
randomForest(formula = VSS.f ~ + CV_Demand.f + CV_Price.f, data = cleandata, mtry = 2,
importance = TRUE, ntree = 1000, na.action = na.omit)
Type of random forest: classification
Number of trees: 1000
No. of variables tried at each split: 2

OOB estimate of error rate: 10%

Confusion matrix:

	H	L	M	class.error
H	21	0	0	0.00
L	0	4	1	0.20
M	0	2	2	0.50

Importance of variables:

	H	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	26.11	22.05	16.55	34.77	3.82
CV_Price.f	51.84	44.44	21.72	62.13	9.07

Rules extracted:

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,227	1	2,536	X[,2] %in% c('0.30')	H
2	2	0,163	1	2,917	X[,1] %in% c('0') & X[,2] %in% c('0.15')	M
3	1	0,162	1	2,536	X[,1] %in% c('0.15') & X[,2] %in% c('0.15')	H
4	2	0,151	1	3,803	X[,1] %in% c('0') & X[,2] %in% c('0')	L

CASE 7

Response: VSS

ANOVA Table for General Factorial Regression (Logit_VSS versus factors)

Backward Elimination of Terms
α to remove = 0,15

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	12	3,85021	0,32085	8328,21	0,000
Linear	4	3,71549	0,92887	24110,42	0,000
CV_Demand	1	0,00106	0,00106	27,50	0,013
CV_Price	1	3,41041	3,41041	88522,85	0,000
CV_Cost	1	0,30401	0,30401	7891,02	0,000
CV_Canb.Rate	1	0,00001	0,00001	0,29	0,630
2-Way Interactions	6	0,13420	0,02237	580,57	0,000
CV_Demand*CV_Price	1	0,00006	0,00006	1,48	0,311
CV_Demand*CV_Cost	1	0,00010	0,00010	2,67	0,201
CV_Demand*CV_Canb.Rate	1	0,00033	0,00033	8,55	0,061
CV_Price*CV_Cost	1	0,13262	0,13262	3442,43	0,000
CV_Price*CV_Canb.Rate	1	0,00091	0,00091	23,61	0,017
CV_Cost*CV_Canb.Rate	1	0,00018	0,00018	4,67	0,119
3-Way Interactions	2	0,00052	0,00026	6,70	0,078
CV_Demand*CV_Price*CV_Canb.Rate	1	0,00016	0,00016	4,13	0,135
CV_Demand*CV_Cost*CV_Canb.Rate	1	0,00036	0,00036	9,28	0,056
Error	3	0,00012	0,00004		
Total	15	3,85033			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0062069	100,00%	99,98%	99,91%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		-2,22781	0,00155	-1435,70	0,000	
CV_Demand	0,01628	0,00814	0,00155	5,24	0,013	1,00
CV_Price	0,92337	0,46168	0,00155	297,53	0,000	1,00
CV_Cost	0,27568	0,13784	0,00155	88,83	0,000	1,00
CV_Canb.Rate	-0,00166	-0,00083	0,00155	-0,53	0,630	1,00
CV_Demand*CV_Price	0,00377	0,00189	0,00155	1,22	0,311	1,00
CV_Demand*CV_Cost	0,00507	0,00254	0,00155	1,63	0,201	1,00
CV_Demand*CV_Canb.Rate	0,00907	0,00454	0,00155	2,92	0,061	1,00
CV_Price*CV_Cost	-0,18209	-0,09104	0,00155	-58,67	0,000	1,00
CV_Price*CV_Canb.Rate	0,01508	0,00754	0,00155	4,86	0,017	1,00
CV_Cost*CV_Canb.Rate	-0,00671	-0,00335	0,00155	-2,16	0,119	1,00
CV_Demand*CV_Price*CV_Canb.Rate	- 0,00631	-0,00315	0,00155	-2,03	0,135	1,00
CV_Demand*CV_Cost*CV_Canb.Rate	-0,00945	-0,00473	0,00155	-3,05	0,056	1,00

Regression Equation in Uncoded Units

Logit_VSS = -2,22781 + 0,00814 CV_Demand + 0,46168 CV_Price + 0,13784 CV_Cost
- 0,00083 CV_Canb.Rate + 0,00189 CV_Demand*CV_Price + 0,00254 CV_Demand*CV_Cost
+ 0,00454 CV_Demand*CV_Canb.Rate - 0,09104 CV_Price*CV_Cost
+ 0,00754 CV_Price*CV_Canb.Rate - 0,00335 CV_Cost*CV_Canb.Rate
- 0,00315 CV_Demand*CV_Price*CV_Canb.Rate
- 0,00473 CV_Demand*CV_Cost*CV_Canb.Rate

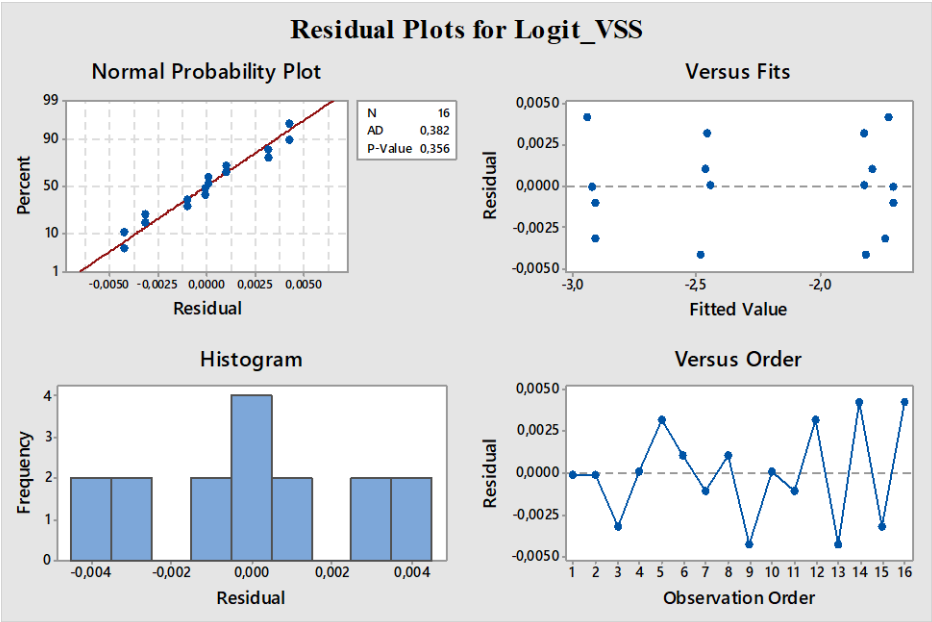


Figure C37. Residual plots for Case 7 (Uncertain parameters: VSS)

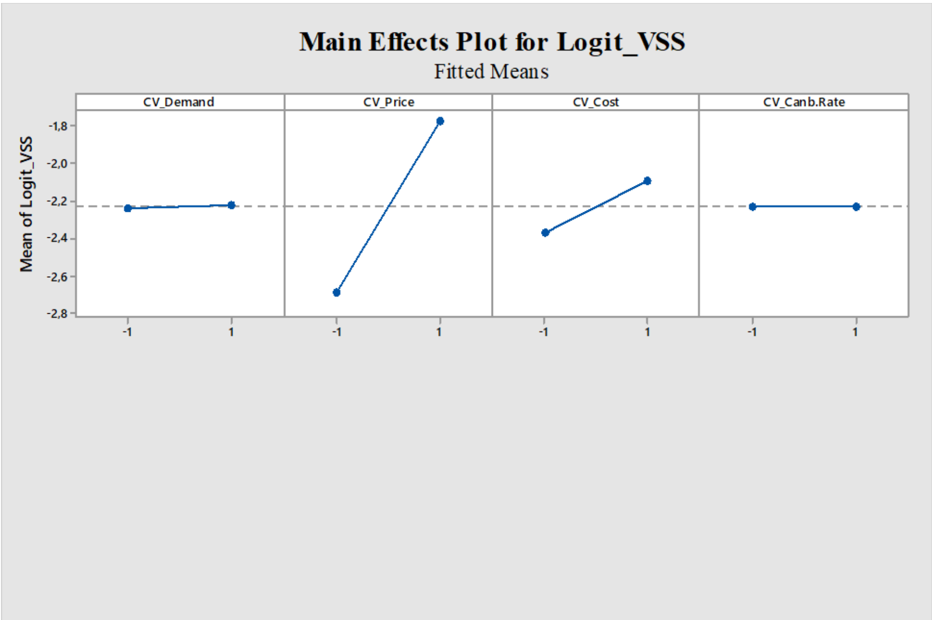


Figure C38. Main effects plot for Case 7 (Uncertain parameters: VSS)

Figure C39. Interaction plot for Case 7 (Uncertain parameters: VSS)

Figure C40. Pareto chart of the standardised effects for Case 7 (Uncertain parameters: VSS)

According to Figure C37, the assumptions of the regression model whose R-sqr (adjusted) is 99.9 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four uncertain parameters:

```
Call:
randomForest(formula = VSS.f ~  CV_Demand.f  +  CV_Price.f  +  CV_Cost.f  +  CV_Canb.Rate.
              f,              data = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.actio
              n = na.omit)
           Type of random forest: classification
                 Number of trees: 1000
No. of variables tried at each split: 3

     OOB estimate of  error rate: 0%
```

Confusion matrix:

	L	M	class.error
L	8	0	0.0000000
M	0	8	0. 0000000

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-3.76	-2.52	-4.24	0.10
CV_Price.f	46.77	47.79	55.01	7.19
CV_Cost.f	-3.58	-1.86	-3.99	0.10
CV_Canb.Rate.f	-4.19	-3.48	-5.32	0.10

Since the mean decrease accuracy of CV_Demand, CV_Cost and CV_Cannibalisation Rate are negative, we exclude these parameters and re-run the model with the remaining parameters.

R output for the revised model including one uncertain parameter:

```
Call:
randomForest(formula = VSS.f ~ + CV_Price.f, data = cleandata, mtry = 2, importance = TRUE,
                          ntree = 1000,  na.action = na.omit)
           Type of random forest: classification
                 Number of trees: 1000
No. of variables tried at each split: 1

     OOB estimate of  error rate: 0%
```

Confusion matrix:

	L	M	class.error
L	8	0	0.0000000
M	0	8	0. 0000000

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_ Price.f	51.63	51.27	60.28	7.52

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,5	1	2,00	X[,1] %in% c('0.30')	M
2	1	0,5	1	2,00	X[,1] %in% c('0.05')	L

CASE 8

Response: VSS

ANOVA Table for General Factorial Regression (Logit_VSS versus factors)

Backward Elimination of Terms
α to remove = 0,15

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	7	7,65716	1,09388	64515,38	0,000
Linear	4	7,61292	1,90323	112249,61	0,000
CV_Demand	1	0,00008	0,00008	4,59	0,065
CV_Price	1	7,51895	7,51895	443456,41	0,000
CV_Cost	1	0,09363	0,09363	5521,93	0,000
CV_Canb.Rate	1	0,00026	0,00026	15,53	0,004
2-Way Interactions	3	0,04424	0,01475	869,73	0,000
CV_Demand*CV_Price	1	0,00021	0,00021	12,34	0,008
CV_Demand*CV_Cost	1	0,00007	0,00007	4,14	0,076
CV_Price*CV_Cost	1	0,04396	0,04396	2592,71	0,000
Error	8	0,00014	0,00002		
Total	15	7,65730			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0041177	100,00%	100,00%	99,99%

Coded Coefficients

Term	Effect	Coef SE	Coef	T-Value	P-Value	VIF
Constant		-2,96783	0,00103	-2883,01	0,000	
CV_Demand	0,00441	0,00220	0,00103	2,14	0,065	1,00
CV_Price	1,37104	0,68552	0,00103	665,93	0,000	1,00
CV_Cost	0,15299	0,07650	0,00103	74,31	0,000	1,00
CV_Canb.Rate	0,00811	0,00406	0,00103	3,94	0,004	1,00
CV_Demand*CV_Price	-0,00723	-0,00362	0,00103	-3,51	0,008	1,00
CV_Demand*CV_Cost	0,00419	0,00210	0,00103	2,04	0,076	1,00
CV_Price*CV_Cost	-0,10483	-0,05242	0,00103	-50,92	0,000	1,00

Regression Equation in Uncoded Units
Logit_VSS = -2,96783 + 0,00220 CV_Demand + 0,68552 CV_Price + 0,07650 CV_Cost
+ 0,00406 CV_Canb.Rate - 0,00362 CV_Demand*CV_Price + 0,00210 CV_Demand*CV_Cost
- 0,05242 CV_Price*CV_Cost

Residual Plots:

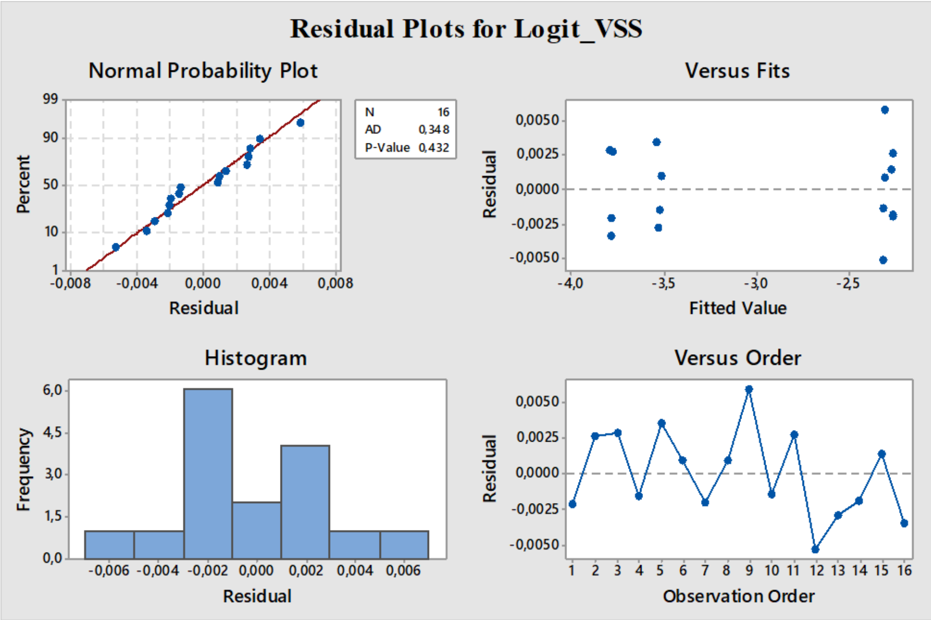


Figure C41. Residual plots for Case 8 (Uncertain parameters: VSS)

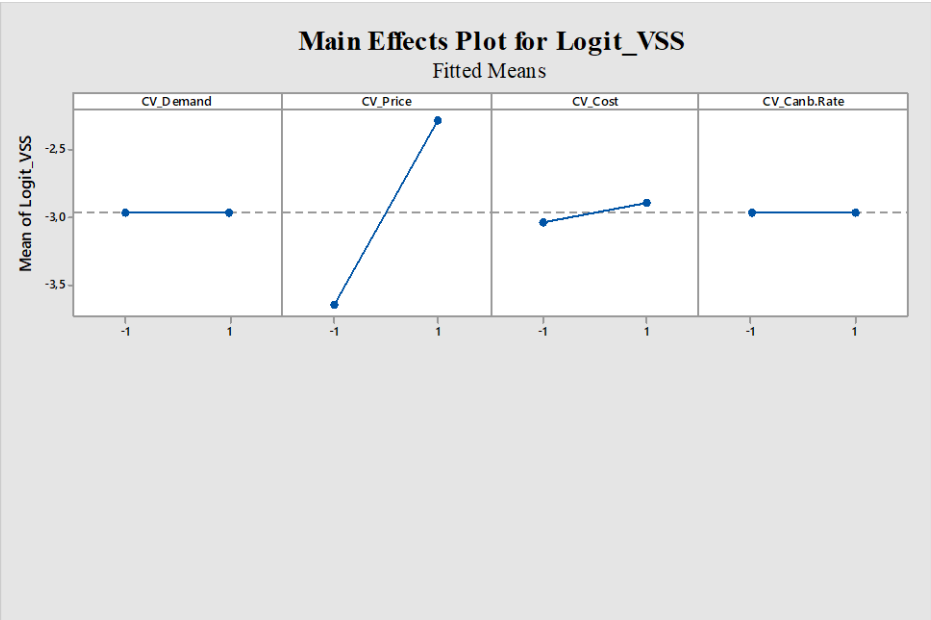


Figure C42. Main effects plot for Case 8 (Uncertain parameters: VSS)

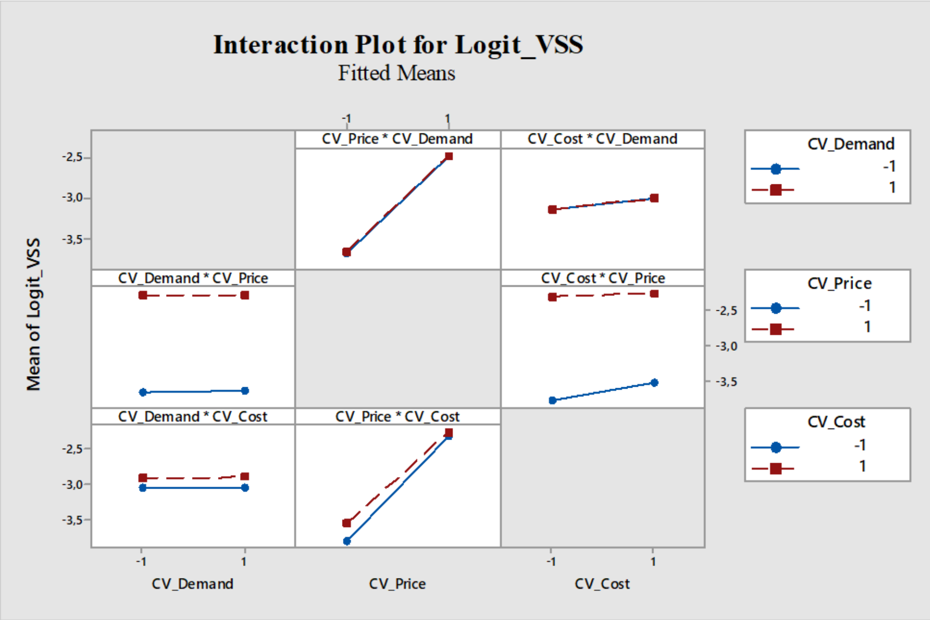


Figure C43. Interaction plot for Case 8 (Uncertain parameters: VSS)

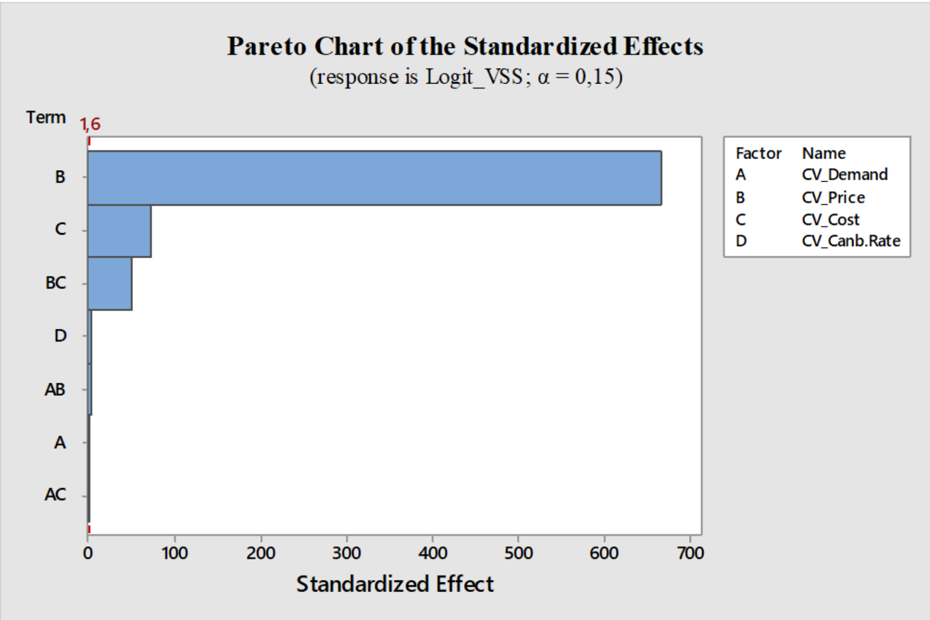


Figure C44. Pareto chart of the standardised effects for Case 8 (Uncertain parameters: VSS)

According to Figure C41, the assumptions of the regression model whose R-sqr (adjusted) is 100 % are satisfied.

CASE 9

Response: VSS

ANOVA Table for General Factorial Regression (Logit_VSS versus factors)

Backward Elimination of Terms
α to remove = 0,15

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	5	1,36580	0,27316	8419,31	0,000
Linear	3	1,31749	0,43916	13535,90	0,000
CV_Price	1	1,09479	1,09479	33743,57	0,000
CV_Cost	1	0,22217	0,22217	6847,63	0,000
CV_Canb.Rate	1	0,00054	0,00054	16,50	0,002
2-Way Interactions	2	0,04831	0,02415	744,43	0,000
CV_Price*CV_Cost	1	0,04730	0,04730	1457,79	0,000
CV_Price*CV_Canb.Rate	1	0,00101	0,00101	31,08	0,000
Error	10	0,00032	0,00003		
Total	15	1,36612			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0,0056960	99,98%	99,96%	99,94%

Coded Coefficients

Term	Effect	Coef	SE Coef	T-Value	P-Value	VIF
Constant		-2,43644	0,00142	-1710,98	0,000	
CV_Price	0,52316	0,26158	0,00142	183,69	0,000	1,00
CV_Cost	0,23567	0,11784	0,00142	82,75	0,000	1,00
CV_Canb.Rate	-0,01157	-0,00579	0,00142	-4,06	0,002	1,00
CV_Price*CV_Cost	-0,10874	-0,05437	0,00142	-38,18	0,000	1,00
CV_Price*CV_Canb.Rate	-0,01588	-0,00794	0,00142	-5,58	0,000	1,00

Regression Equation in Uncoded Units
Logit_VSS = -2,43644 + 0,26158 CV_Price + 0,11784 CV_Cost - 0,00579 CV_Canb.Rate
- 0,05437 CV_Price*CV_Cost - 0,00794 CV_Price*CV_Canb.Rate

Fits and Diagnostics for Unusual Observations

Obs	Logit_VSS	Fit	Resid	Std Resid	
11	-2,87739	-2,86807	-0,00932	-2,07	R

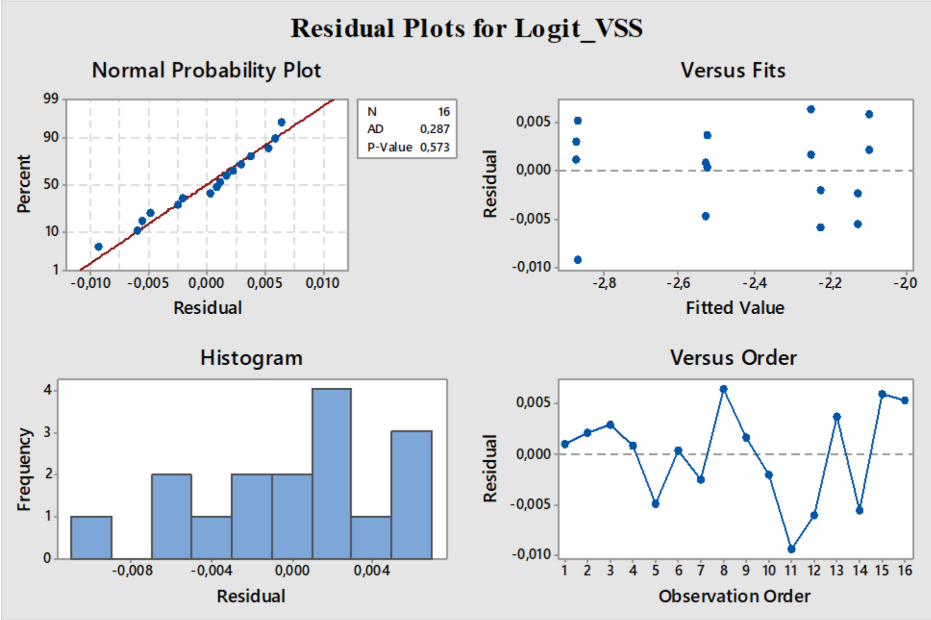


Figure C45. Residual plots for Case 9 (Uncertain parameters: VSS)

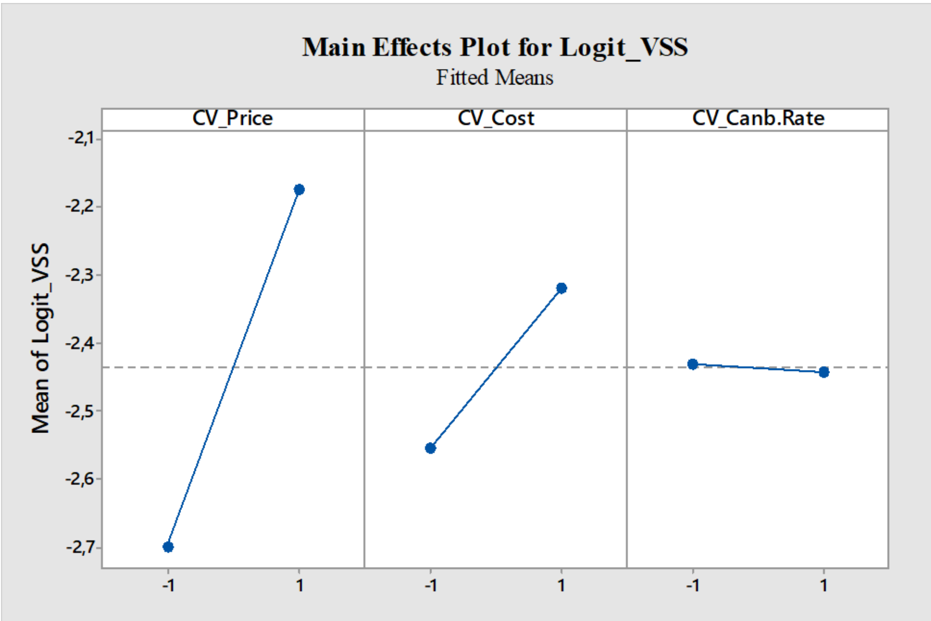


Figure C46. Main effects plot for Case 9 (Uncertain parameters: VSS)

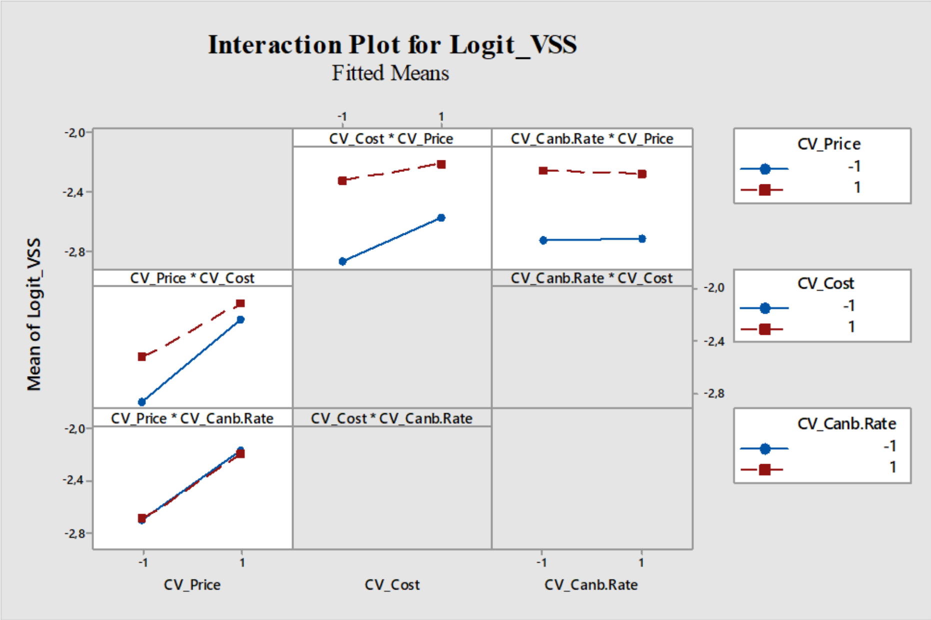


Figure C47. Interaction plot for Case 9 (Uncertain parameters: VSS)

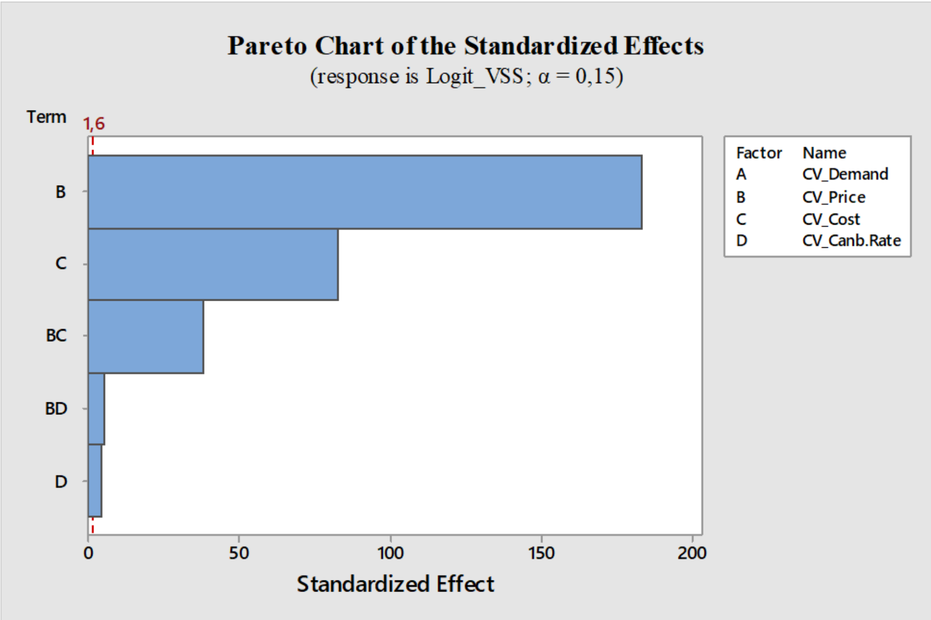


Figure C48. Pareto chart of the standardised effects for Case 9 (Uncertain parameters: VSS)

According to Figure C45, the assumptions of the regression model whose R-sqr (adjusted) is 99.9 % are satisfied.

Rules extracted from the Random Forest Application

R output for a model including all four uncertain parameters:

```
Call:
randomForest(formula = VSS.f ~  CV_Demand.f  +  CV_Price.f  +  CV_Cost.f  +  CV_Canb.Rate.
                   f,                  data = cleandata, mtry = 3, importance = TRUE,  ntree = 1000, na.actio
                   n = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 3

       OOB estimate of  error rate: 0%
```

Confusion matrix:

	L	M	class.error
L	12	0	0.0000000
M	0	4	0. 0000000

Importance of variables:

	H	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_Demand.f	-5.47	-5.80	-7.28	0.13
CV_Price.f	24.56	25.74	31.51	2.67
CV_Cost.f	24.22	26.08	31.69	2.68
CV_Canb.Rate.f	-6.13	-6.73	-8.11	0.14

Since the mean decrease accuracy of CV_Demand and CV_Cannibalisation Rate are negative, we exclude these parametesr and re-run the model with the remaining parameters.

R output for the revised model including two uncertain parameters:

```
Call:
randomForest(formula = VSS.f ~ + CV_Price.f + CV_Cost.f, data = cleandata, mtry = 2, importan
                   ce = TRUE,   ntree = 1000,  na.action = na.omit)
               Type of random forest: classification
                     Number of trees: 1000
No. of variables tried at each split: 2

       OOB estimate of  error rate: 0%
```

Confusion matrix:

	L	M	class.error
L	12	0	0.0000000
M	0	4	0. 0000000

Importance of variables:

	L	M	MeanDecreaseAccuracy	MeanDecreaseGini
CV_ Price.f	29.80	28.83	37.55	2.90
CV_Cost.f	28.48	29.66	37.40	2.85

***Rules extracted*:**

Rule No.	Length	Support	Confidence	Lift	Condition	Prediction
1	1	0,332	1	1,490	X[,1] %in% c('0,15')	L
2	1	0,332	1	1,490	X[,2] %in% c('0,15')	L
3	2	0,329	1	3,041	X[,1] %in% c('0,3') & X[,2] %in% c('0,3')	M
4	2	0,177	1	1,490	X[,1] %in% c('0,3') & X[,2] %in% c('0,15')	L
5	2	0,152	1	1,490	X[,1] %in% c('0,15') & X[,2] %in% c('0,3')	L

R CODE FOR RULE EXTRACTION THROUGH RANDOM FOREST

```
library (readxl)
library (randomForest)
library (inTrees)
data                    <- read_excel("D://DetCase1VSS.xlsx")
data                    <- as.data.frame(data)
data$Cap_Exp_Cost.f     <- factor(data $Cap_Exp_Cost)
data$Profitability.f    <- factor(data$Profitability)
data$Unit_Cap_Usage.f   <- factor(data$Unit_Cap_Usage)
data$Capacity.f         <- factor(data$Capacity)
data$VSS.f              <- factor(data$VSS)
cleandata               <- data[,6:10]
pmix.rf1                <- randomForest (VSS.f ~ Cap_Exp_Cost.f + Profitability.f +
                             Unit_Cap_Usage.f + Capacity.f, data=cleandata, mtry=3,
                             importance=TRUE, na.action=na.omit, ntree=1000)
round(importance(pmix.rf1), 2)
target                  <- cleandata[,"VSS.f"]
pmix.rf3                <- randomForest(target ~ Cap_Exp_Cost.f + Profitability.f +
                            Unit_Cap_Usage.f + Capacity.f, data=cleandata, mtry=3,
                            importance=TRUE, na.action=na.omit, ntree=1000)
treeList                <- RF2List(pmix.rf3)
exec                    <- extractRules(treeList, pmix.rf1, ntree=1000)
ruleMetric              <- getRuleMetric(exec,pmix.rf1,target)
readableRules           <- presentRules(ruleMetric, colnames(pmix.rf1))
readableRules
view(readableRules)
freqPattern             <- getFreqPattern(ruleMetric,minsup = 0.05, minconf = 0.75,
                             minlen = 1, maxlen = 5)
```

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Karakaya, Şakir
Nationality: Turkish (TC)
Date and Place of Birth: 10 August 1981, Sivas
Marital Status: Single
Phone: +90 312 201 53 82
email: shakir.karakaya@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MSc	London School of Economics and Political Science (LSE), Public Management and Governance	2016
MSc	Middle East Technical University (ODTÜ) Industrial Engineering	2008
BSc	Gazi University, Industrial Engineering	2004
High School	Malatya Turgut Özal Anatolian High School	1999

WORK EXPERIENCE

Year	Place	Enrollment
2017 Nov-Present	Ministry of Industry and Technology	Head of Product Safety Department
2011 Agt- 2017 Nov	Ministry of Science, Industry and Technology	Industry and Technology Expert
2004 Dec – 2011 Agt	National Productivity Centre	Productivity Expert

FOREIGN LANGUAGES

English, Spanish