Application of Simulation and Optimization Approaches in Supply-Constrained Innovation Diffusion

by

Ashkan Negahban

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2016

Keywords: agent-based simulation, optimization, Monte Carlo simulation, new product diffusion, production-sales policies, viral marketing

Approved by

Jeffrey Smith, Chair, Professor of Industrial and Systems Engineering
Levent Yilmaz, Professor of Computer Science and Software Engineering
Daniel Butler, Professor of Marketing
Chase Murray, Assistant Professor of Industrial and Systems Engineering

Abstract

When introducing a new product, firms face a hierarchy of decisions at the strategic and operational levels including capacity sizing, time to market or starting sales, initial inventory required by the product's release time, and production management in response to changes in the demand. Firms also face the dilemma of how to support a fast and substantial take-off by targeting the right population of potential consumers for seeding. This dissertation explores the above inter-dependent decisions using a diverse set of analysis tools, namely agent-based modeling and simulation, Monte Carlo simulation, continuous-time mathematical models, and parametric and nonparametric statistical approaches. This work contributes to the marketing and operations management literature in five significant ways: (1) it shows that ignoring supply and demand uncertainties may lead to potentially incorrect decisions and that the optimal decision may change if risk is used as the primary performance measure instead of the commonly used expected (mean) profit; (2) it provides insights about the optimal introduction time of a new generation of a new product under market expansion and cannibalization; (3) it provides a joint analysis of marketing and production strategies and shows that a sequential decision-making process would lead to suboptimal decisions and reduced profit; (4) it explores the importance of the social network structure and individuals' interactions on the optimal combination of seeding and build-up policies; and, (5) it presents a more realistic analysis by relaxing many of the assumptions of previous studies and provides empirical evidence by a successful application to the case of the diffusion of Sony's PlayStation®3 game console in Europe. The findings of this work and its future extensions along the lines discussed in the dissertation have important implications for innovation diffusion research and can potentially help companies make better decisions regarding production and marketing of their new products.

Acknowledgments

This dissertation is dedicated to my mother, Mehri Shetabi, the most important person in my life and the most unconventional and unconditionally loving woman I know. I am grateful to my father, Azizollah Negahban, for his great support in my education, and most importantly, for teaching me how to think logically and act professionally. I am thankful to the best older brother in the world, Dr. Arash Negahban, for his constant encouragement, support, and kindness.

I would like to express my appreciation to my adviser, Dr. Jeffrey Smith, for training me to become a researcher and (hopefully) a successful academician. I would also like to thank the members of my committee, Dr. Jeffrey Smith (chair), Dr. Levent Yilmaz, Dr. Daniel Butler, and Dr. Chase Murray, for their support and enriching insights.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and research questions

The significance of new products is conspicuous in the fact that, on average, roughly one-third of firms' overall sales and profit comes from their new products (Hauser et al. 2006). With the increased rate of introduction of new products and shorter product life-cycles, firms are facing a higher level of competition and demand uncertainty. On the other hand, forecasting the demand for a new product during the course of its diffusion has been known as a very difficult problem due to the lack of historical sales data making capacity, production, and sales management for new products even more challenging. When introducing a new product/technology, selling as many units as possible without building an initial inventory (a *myopic* policy), can lead to supply shortages when demand exceeds production capacity (due to the high cost and long lead time for capacity expansions). This results in loss of potential profit. To avoid this problem, companies generally follow an alternative policy that involves building inventory prior to starting sales as a substitute for the need for capacity expansion (*build-up* policy). However, the build-up policy leads to higher inventory costs and also delays revenue collection from sales. Therefore, an appropriate build-up policy is critical to the overall sales and profit from the new product.

The initial phase of market penetration is critical to the success of new products. Companies generally rely on promotional activities to support a fast and substantial take-off to increase the chance of a successful launch and diffusion. A common approach to encourage adoption and positive word-of-mouth during this initial phase is to distribute reduced-price products to a set of target consumers in the hope that they will initiate a word-of-mouth grapevine. This approach is generally referred to as *viral marketing* or *seeding*. Given an arbitrary social network, individuals with high *influence* (for instance, consumers with a high number of ties) can potentially encourage more adoptions faster. Also, seeding a larger proportion of the population would increase the chance of faster diffusion and adoption. However, giving away more reduced-price products would decrease total revenue. A faster adoption rate would also increase the risk of supply shortages and lost sales (i.e., loss of potential profit). Therefore, targeting and level of seeding affect the overall profit and appropriate decisions regarding the two are crucial for a successful launch and diffusion. More importantly, these decisions are not independent of the decisions made on the supply side.

There are many cases where companies (even those with significant experience in successful new product launches) have faced considerable monetary losses due to inappropriate decisions regarding the above issues. For instance, due to incorrect over-anticipation of the demand for PlayStation®3 that resulted in excessive production and inventory costs, Sony Electronics Inc. lost $1.8B and laid off 3% of its workforce (*Los Angeles Times*, June 7, 2007). Another interesting example deals with the case of Tamagotchi™, the first virtual pet, that rapidly grew beyond expectations and led to excessive lost sales (Higuchi and Troutt 2004). The company (Bandai Co.) eventually expanded the production capacity to avoid further loss of potential profit. However, when the capacity expansion took place in 1998, the demand for the product had already begun to decline leading to a $123 million loss. There are many examples where companies had to delay the product's launch time due to production uncertainties. In 2001, Microsoft Co. postponed the launch of Xbox® in Japan for a year (*New York Times*, August 27, 2001) and in the US by a week (*New York Times*, September 22, 2001) as they failed to meet the targeted initial inventory of 1 million units by the originally announced released time. Other examples include the case of Apple's iPad® (*New York Times*, April 14, 2010) and iPod® (*New York Times*, March 26, 2004), Sony's PlayStation®3 (*New York Times*, January 25, 2007), and Nintendo's GameCube® (*CNN*, August 23, 2001). There are other examples where demand uncertainties caused significant problems for companies including the case of Apple's Power Mac G4 (*New York Post*, September 21, 1999) and PlayStation®4 (*New York Times*, December 19, 2014).

Motivated by the above, this dissertation aims at providing an integrated and joint analysis of the marketing and supply side of new product diffusion. More specifically, the stream of research performed in this dissertation explores the following research questions that the current literature leaves unanswered:

1. Does ignoring production and demand uncertainties lead to a potentially incorrect decision?

2. Does the decision on the optimal production-sales policy change if risk measures are considered instead of the expected profit?

3. What is the cost of making an incorrect decision if uncertainties are ignored?

4. When should the firm introduce the new generation of a new product to maximize its profit?

5. What is the optimal production capacity and build-up policy for the new generation?

6. What is the optimal sales plan for the new and previous generations after the new generation's launch?

7. Does making decisions sequentially on the seeding and build-up policies lead to sub-optimal decisions?

8. How does the optimal combination of production and seeding strategies vary for different product categories?

9. Is the seeding strategy that maximizes the adoption rate optimal in the presence of binding supply constraints?

10. What is the effect of the social network structure on the resulting diffusion dynamics and the optimal seeding and build-up policies?

Table 1.1: Positioning and contributions of this dissertation to the literature

	Single-generation				Multi-generation			
	UD	DSC	SSC	JPM	UD	DSC	SSC	JPM
Existing *Marketing* literature	✓				✓			
Existing *Operations Management* literature		✓						
Chapter 2 (Negahban and Smith 2016a)			✓					
Chapter 3 (Negahban and Smith 2016c)						✓		
Chapter 4 (Negahban and Smith 2016b)				✓				
Future research							☞	☞

UD: Unconstrained diffusion (unlimited supply)
DSC: Deterministic supply-constrained
SSC: Stochastic supply-constrained
JPM: Joint analysis of production and marketing strategies

1.2 Background and contributions

In this section, a critical analysis of the strengths and gaps of the two main related streams of research, namely the *marketing* and *operations management* literature is presented to establish the contributions of this dissertation (a detailed literature review of related studies is provided in each of the following chapters). The literature analysis and the contributions of the three papers that constitute this dissertation are summarized in Table 1.1 and explained below.

There is a long history of research on new product diffusion with early studies going back to the 1960's. The focus of the marketing research is primarily on demand forecasting. Perhaps the most fundamental diffusion model is the Bass model (Bass 1969), which is empirically tested and validated for hundreds of product categories (see Sultan et al. (1990) for a meta-analysis of 213 applications of the Bass model). Common extensions to the Bass model include incorporation of negative word-of-mouth, stochasticity in the demand, and modeling the demand of successive technology generations. For a comprehensive analysis of these extensions, see Hauser et al. (2006) and Peres et al. (2010). The strength of this stream of research is in the successful application of the Bass model and its extensions to forecast the demand for hundreds of products. These studies have also developed a better general understanding of the diffusion dynamics and the effect of heterogeneity, different marketing strategies, cross-market and cross-brand factors, differences in growth across countries, competition, and social interactions and its increasing complexity due to digital word-of-mouth. However, perhaps the most important deficiency in the marketing literature is an almost complete neglect of supply restrictions. The underlying assumption of these studies is that unlimited supply is available and thus the effect of capacity constraints on the future demand is not considered. As a result, some of the proposed policies seem unrealistic and are generally not consistent with industry practices. For instance, Wilson and Norton (1989) propose a two-generation demand model for durable products and develop analytical expressions for the total profit. Their analysis leads to the *Now or Never* market entry policy which essentially suggests that the second generation should either be introduced simultaneously with the first generation or not be introduced at all while many industries, namely fashion, high-technology, and pharmaceutical, report alternative timing strategies.

As a relatively new stream of research, the marketing-operations interface investigates the inter-dependency between the demand and supply for new products. The primary analysis tool used in these studies can be classified as either mathematical or simulation models. In two seminal works in this stream, Kumar and Swaminathan (2003) and Ho et al. (2002) independently propose a modified supply-restricted Bass model and develop mathematical models to determine the optimal production-sales policy (i.e., number of build-up periods, production capacity, and launch time) for a single generation of a new product. Similar analytical models are developed to study the optimal production-sales policy under different supply chain topologies, dynamic pricing, multi-stage ordering, learning phenomenon, etc. The studies under the marketing-operations management interface category provide valuable insights about supply-constrained innovation diffusion by showing how supply shortages and lost sales affect the future demand dynamics and the optimal production capacity, build-up policy, introduction time, and sales plan for a single generation of a new product. However, there are three major gaps in this stream of literature: (1) the majority of the proposed mathematical models assume deterministic demand and supply to make the problem analytically tractable and thus the effect of supply and demand uncertainties are ignored; (2) regardless of the type of the analysis tool, the studies under this general stream of research mainly consider the *expected* life-cycle profit as the primary performance measure while *risk* is ignored; and, (3) the effect of seeding and promotional activities on the new product's demand, and consequently on the optimal production-sales policy is generally ignored.

This dissertation contributes to the two streams of research by: (1) characterizing the simultaneous effect of demand and supply uncertainties on the optimal production and sales plan; (2) considering risk measures and the distribution of profit in the selection of the optimal production-sales policy; (3) extending previous multi-generation diffusion models to account for supply constraints, developing more realistic market entry policies, and providing empirical evidence for the efficacy of the proposed model through a case study on the launch of Sony's PlayStation®3 game console; and, (4) providing a joint analysis of production and viral marketing strategies that shows the optimal build-up policy varies under different seeding strategies, social network structures, and consumers' backlogging behavior.

1.3 Methodology

This dissertation explores the above research questions using a diverse set of analysis tools, namely agent-based modeling and simulation, Monte Carlo simulation, continuous-time mathematical models, and parametric and nonparametric statistical approaches. This section describes the methodology used in each of the following chapters.

The first three research questions are investigated in Chapter 2 (Negahban and Smith 2016a). Through extensive experimentation with a Monte Carlo simulation model of the supply-constrained new product diffusion, we demonstrate how the optimal production-sales policy changes when the stochasticity in supply and demand are explicitly considered and derive perspective results on the magnitude of these changes. As for the primary performance measures in our analysis, we use the expected net present value of profit over the product's life-cycle as well as the risk associated with it and show that the optimal policy may vary under different performance metrics. We use parametric and nonparametric statistical tests, namely the Welch's t-test and a double bootstrap method, to compare the mean and percentiles of profit, respectively, and establish statistically significant changes in the decision if uncertainties or risk are ignored. More than 300,000 market configurations are studied resulting in more than 7.1 billion total replications of the Monte Carlo model. The extensive experimentation generated more than 110 gigabytes of raw simulation output necessitating a timely and challenging data analysis. We developed and verified several automated programs to analyze the data to insure consistency; however, the computational aspects of the analysis process is out of the scope of this dissertation.

Chapter 3 (Negahban and Smith 2016c) explores research questions (4-6). We consider a single firm contemplating the introduction of a new generation of a product family where the technology for the new generation becomes available sometime after the introduction of the predecessor. We start by modifying an existing multi-generation demand model to account for supply constraints. We then formulate a continuous-time mathematical model in the context of *optimal control theory* to determine the optimal sales plan for successive generations. Closed-form solutions are then derived for the special case of *patient customers* (i.e., no lost sales). To gain insight about the optimal build-up policy, production capacity, and market entry policy for the general case of *impatient customers* (i.e., partially backlogged demand with lost sales), a comprehensive numerical study is performed using a discrete-time version of the model. We also present an application of the proposed model to the case of Sony's PlayStation®3 game console. The results are consistent with the empirical evidence collected on the sources of the product's poor performance. Our case study validates the proposed model and demonstrates its potential in helping companies choose an appropriate release time and build-up policy for successive generations of their products.

Chapter 4 (Negahban and Smith 2016b) investigates the last four research questions (7-10) using an agent-based simulation model consisting of a *firm* agent and socially networked *consumer* agents. We consider the following network structures: (1) regular lattice; (2) random; (3) small-world; and (4) scale-free. We consider both myopic and build-up policies as well as five seeding criteria, namely node degree, number of two-step ties, average path length, clustering coefficient, and random selection, to prioritize consumers and decide who should be targeted for seeding. Through extensive experimentation with the model, we jointly analyze the performance of different seeding and production polices in terms of the net present value of profit under various network structures, product categories, and demand backlogging behavior. We also systematically experiment with the parameters of the small-world and scale-free networks to study the effect of long-range connections and distribution of hubs (high-degree nodes) on the resulting diffusion process (including the adoption, waiting, and lost sales dynamics) as well as the product's life-cycle profit.

1.4 Organization of the dissertation

The following chapters of the dissertation are organized as follows. Chapters 2-4 start by discussing research motivations and the importance of the problem under consideration followed by a critical analysis of the related literature to identify research strengths and gaps and establish the contributions of the work presented in the respective chapter. The methodology, experimental design, and important results are then discussed. At the end of each chapter, we present a brief discussion on managerial implications, limitations, and potential extensions. Finally, Chapter 5 presents the concluding remarks by summarizing the important findings and their implications for innovation diffusion research and practice and outlining future research opportunities.

Bibliography

Bass, F. M. 1969. A new product growth for model consumer durables. *Management Science* **15**(5) 215–227.

Hauser, J., G. J. Tellis, A. Griffin. 2006. Research on innovation: A review and agenda for marketing science. *Marketing Science* **25**(6) 687–717.

Higuchi, T., M. D. Troutt. 2004. Dynamic simulation of the supply chain for a short life cycle product–lessons from the Tamagotchi case. *Computers & Operations Research* **31**(7) 1097–1114.

Ho, T.-H., S. Savin, C. Terwiesch. 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science* **48**(2) 187–206.

Kumar, S., J. M. Swaminathan. 2003. Diffusion of innovations under supply constraints. *Operations Research* **51**(6) 866–879.

Negahban, A., J. S. Smith. 2016a. The effect of supply and demand uncertainties on the optimal production and sales plans for new products. *International Journal of Production Research* **(Accepted, in press)**. URL `http://dx.doi.org/10.1080/00207543.2016.1157274`.

Negahban, A., J. S. Smith. 2016b. A joint analysis of production and seeding strategies for new products: An agent-based simulation approach. *Submitted to Annals of Operations Research* **(Under review)**.

Negahban, A., J. S. Smith. 2016c. Optimal entry time and production-sales policies for successive generations of new products. *Submitted to Production and Operations Management* **(Under review)**.

Peres, R., E. Muller, V. Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* **27**(2) 91–106.

Sultan, F., J. U. Farley, D. R. Lehmann. 1990. A meta-analysis of applications of diffusion models. *Journal of Marketing Research* **27**(1) 70–77.

Wilson, L. O., J. A. Norton. 1989. Optimal entry timing for a product line extension. *Marketing Science* **8**(1) 1–17.

Chapter 2

The effect of supply and demand uncertainties on the optimal production and sales plans for new products

2.1 Abstract

When introducing a new product, firms face a hierarchy of decisions at the strategic and operational levels including capacity sizing, time to market or starting sales, initial inventory required by the product's release time, and production management in response to changes in the demand (hereafter referred to as production-sales policies). The goal of this paper is to show the importance of considering both supply and demand uncertainties in the determination of the production-sales policy which has been overlooked in the existing literature. More specifically, we test two main hypotheses: (1) ignoring supply and demand uncertainties may lead to potentially incorrect decisions; and, (2) the decision could be different if risk is used as the primary performance measure instead of the commonly used expected (mean) profit. We perform extensive experimentation with a Monte Carlo simulation model of the stochastic supply-restricted new product diffusion and use different statistical procedures, namely, the Welch's t-test and a nonparametric double-bootstrap method to compare the average and percentiles of the profit for different policies, respectively. The results indicate that the correctness of the two hypotheses depends on the diffusion speed, consumers' backlogging behavior, production capacity, price, and variable production and inventory costs. The findings also have important implications for managers regarding market entry time, parameter estimation, production strategy, and the implementation of the proposed model.

keywords: Innovation diffusion; myopic and build-up policies; production uncertainties; stochastic Bass model; simulation

2.2 Introduction

There are many examples where companies had to delay the product's launch time due to a scarce supply caused by production uncertainties. In 2001, Microsoft Co. postponed the launch of Xbox® in Japan for a year (*New York Times*, August 27, 2001) and in the US by a week (*New York Times*, September 22, 2001) as they failed to meet the targeted initial inventory of 1 million units by the originally announced released time. Other examples include the case of Apple's iPad® (*New York Times*, April 14, 2010) and iPod® (*New York Times*, March 26, 2004), Sony's PlayStation®3 (*New York Times*, January 25, 2007), and Nintendo's GameCube® (*CNN*, August 23, 2001). Demand variability can also have a significant impact on the success of new products. For instance, due to incorrect over-anticipation of the demand for PlayStation®3 that resulted in excessive production and inventory costs, Sony Electronics Inc. lost $1.8B and laid off 3% of its workforce (*Los Angeles Times*, June 7, 2007). Another interesting example deals with the case of Tamagotchi™, the first virtual pet, that rapidly grew beyond expectations and led to excessive lost sales (Higuchi and Troutt 2004). The company (Bandai Co.) eventually expanded the production capacity to avoid further loss of potential profit. However, when the capacity expansion took place in 1998, the demand for the product had already begun to decline leading to a $123 million in after-tax loss. Other examples of the impact of demand uncertainties include the case of Apple's Power Mac G4 (*New York Post*, September 21, 1999) and PlayStation®4 (*New York Times*, December 19, 2014).

When introducing a new product/technology, selling as many units as possible without building an initial inventory (a *myopic* policy), can lead to supply shortages when demand exceeds production capacity (due to the high cost and long lead time for capacity expansions). This results in loss of potential profit. To avoid this problem, companies generally follow an alternative policy that involves building inventory prior to starting sales as a substitute for the need for capacity expansion (*build-up* policy). However, the build-up policy leads to higher inventory costs and also delays revenue collection from sales. It may also lead to lost sales during the build-up period as some of the potential consumers are not willing to wait until the product's launch. Developing an appropriate production-sales plan requires a deep understanding of the diffusion dynamics and necessitates a joint analysis of the impact of supply and demand uncertainties on future sales. Manufacturing systems exhibit significant uncertainties that affect the production throughput (Li et al. 2009) and thus supply levels for the new product. On the other hand, due to the lack of historical sales data, there is a high level of uncertainty associated with demand forecasts. The demand for new products is also subject to randomness due to disturbances in the market caused by economic and financial conditions, technological improvements, and competition.

In this paper, through extensive experimentation with a Monte Carlo simulation model of the supply-constrained new product diffusion, we demonstrate how the optimal production-sales policy changes when the stochasticity in supply and demand are explicitly considered and derive perspective results on the magnitude of these changes. As for the primary performance measures in our analysis, we use the expected net present value of profit over the product's life-cycle as well as the risk associated with it and show that the optimal policy may vary under different performance metrics. In order to address the general problem of considering the effect of demand and supply uncertainties in the determination of the production-sales plan, we explore the following important questions that the existing literature leaves unanswered:

1. Does ignoring uncertainties lead to a potentially incorrect decision? (Section 2.7.1)

2. Does the decision on the optimal production-sales policy change if risk measures are considered instead of the expected profit? (Section 2.7.1)

3. How does the effect of uncertainties vary for different production levels and product categories (with different diffusion characteristics)? (Section 2.7.2)

4. How different would the decision be from the optimal policy (in terms of the length of the build-up period) if uncertainties are ignored? (Section 2.7.3)

5. What is the cost of making an incorrect decision if uncertainties are ignored? (Section 2.7.4)

While the importance of production and demand uncertainties is understood from many real-world examples (e.g., the ones mentioned above), ignoring their impact on the diffusion dynamics and choice of the production-sales policy has been identified as one of the major gaps in the existing literature (see the results of survey papers by Bilginer and Erhun (2010), Hauser et al. (2006), Peres et al. (2010), and Negahban and Yilmaz (2014)). This paper contributes to both the marketing and operations literature by: (a) providing a formal analysis of the effect of supply and demand uncertainties and their interaction on demand and sales dynamics; (b) showing that ignoring these uncertainties could yield a potentially incorrect decision on the optimal production-sales policy; and, (c) evaluating risk and showing that the policy with the maximum expected profit is not necessarily optimal under risk measures.

The remainder of the paper is organized as follows. An overview of the literature is presented in Section 2.3. Section 2.4 provides a detailed description of the stochastic supply-constrained new product diffusion problem. The Monte Carlo simulation model is explained in Section 2.5. The experimental design, performance measures, and the statistical tests used for comparing policies are described in Section 2.6. Section 2.7 provides the analysis of simulation results and summarizes important findings. Finally, Section 2.8 presents the conclusions, a brief discussion on managerial implications for real-world applications, and potential future research opportunities.

2.3 Literature review

The related studies can be categorized into three main streams of research: the *marketing* literature, the *operations/production management* literature, and a more recent category on the interface of the two. Here, we provide a brief review of the literature with the goal to illustrate existing research gaps and characterize the main contributions of this work to the body of knowledge.

Early studies on new product and technology diffusion in the marketing literature go back to the 1960's (Fourt and Woodlock 1960). Perhaps the most fundamental diffusion model is the Bass model (Bass 1969). It is the most widely used forecasting tool in the industry and its predictive power has been empirically proven over decades. The majority of the models that were later proposed are essentially rooted in the Bass model (Bass 2004). These models include the following extensions: diffusion of successive technology generations (Norton and Bass 1987), effect of negative word-of-mouth (Mahajan et al. 1984), and stochastic Bass models (Niu 2002, Kanniainen et al. 2011, Skiadas and Giovanis 1997). The third group, which is of particular interest here, mainly focuses on modeling demand uncertainties to enhance forecasting. For critical analyses of this stream of research see the review papers by Peres et al. (2010), Bilginer and Erhun (2010), and Hauser et al. (2006). The strength of this stream of research is in the successful application of the Bass model and its extensions to forecast the demand for hundreds of products (see Sultan et al. (1990) for a meta-analysis of 213 applications of the Bass model). These studies have also developed a better general understanding of the diffusion dynamics and the effect of heterogeneity, different marketing strategies, cross-market and cross-brand factors, differences in growth across countries, competition, and social interactions and its increasing complexity due to digital word-of-mouth. However, perhaps the most important deficiency in the marketing literature is an almost complete neglect of supply restrictions. The underlying assumption of these studies, including the stochastic demand models, is that unlimited supply is available and thus the effect of capacity constraints on the future demand is not considered. This work contributes to this stream of research by considering the dependence of the demand on supply and investigating how demand and supply uncertainties interact.

The operations literature extensively considers different types of production uncertainty including but not limited to random processing times (Azadeh et al. 2012), machine failures and maintenance operations (Rezg et al. 2004), worker availability (Erel et al. 2001), lack of accurate information about the production process (Azadeh et al. 2011), and stochastic production throughput due to quality inspection (Li et al. 2009). See Stevenson et al. (2005) for a comprehensive review of classical production planning and control approaches and Mula et al. (2006) for an analysis of studies that consider different types of uncertainty. The strength of this stream of literature is in the development and application of numerous approaches (including conceptual models, analytical models, artificial intelligence, and simulation) for production planning, inventory management, and supply chain planning under uncertainty. These studies can be divided into two primary groups based on their demand model: (1) studies that assume either a constant demand or a changing demand with a deterministic and known pattern; and, (2) stochastic studies where the randomness in the demand follows a known probability distribution. Therefore, the main gap in this stream of research is that, regardless of the consideration of demand uncertainties, the demand process is assumed to be exogenous and thus independent of the decisions made on the supply side unlike most real-world applications. The current paper contributes to this literature by considering endogenous demand that is affected by the build-up policy and showing how supply uncertainties impact the diffusion dynamics.

As a relatively new stream of research, the studies in the marketing-operations interface investigate the inter-dependency between the demand and supply for new products. The primary analysis tool used in these studies can be classified as either analytical approaches or simulation models. In two independent studies, Kumar and Swaminathan (2003) and Ho et al. (2002) propose a modified supply-restricted Bass model and develop mathematical models to determine the optimal production-sales policy. Similar analytical studies provide valuable insights about the optimal production-sales policy by considering dynamic pricing (Shen et al. 2011, 2014), multi-period diffusion (Bilginer and Erhun 2015), learning phenomenon (Cantamessa and Valentini 2000), and different supply chain topologies (Amini and Li 2011). However, the majority of the analytical studies assume deterministic demand and supply to make the problem analytically tractable. On the other hand, the results of a comprehensive analysis on the application of simulation in marketing by Negahban and Yilmaz (2014) show that while the role of marketing strategies, positive and negative word-of-mouth, and theoretical and empirical social network structures have received considerable attention, only a few simulation studies address supply constraints while production uncertainties are ignored (see, Negahban et al. (2014), Amini et al. (2012), Negahban (2013)). Moreover, regardless of the type of the analysis tool, the studies under this general stream of research mainly consider the expected life-cycle profit as the primary performance measure while risk is ignored. Therefore, this paper contributes to this stream of research by characterizing the simultaneous effect of demand and supply uncertainties on the optimal production and sales plan. Also, to the best of our knowledge, this work is the first to consider risk measures in the selection of the optimal production-sales policy.

This paper extends a preliminary analysis by the authors (Negahban and Smith 2014) where the effect of production uncertainty is studied for a given production level under deterministic demand. Here, we generalize the results of the earlier study by incorporating the effect of demand uncertainty and its interaction with supply uncertainty under different production levels.

2.4 Supply-restricted new product diffusion

As discussed earlier, the Bass model (Bass 1969) lays the foundation for the supply-constrained diffusion models. Based on analogies from contagion models in epidemiology, product adoptions in the Bass model are driven by two sources: (1) mass media advertisement; and, (2) word-of-mouth. The Bass model formulates the first-time demand for a new *durable* product (i.e., no repeat purchases) at time period t, $d(t)$, by

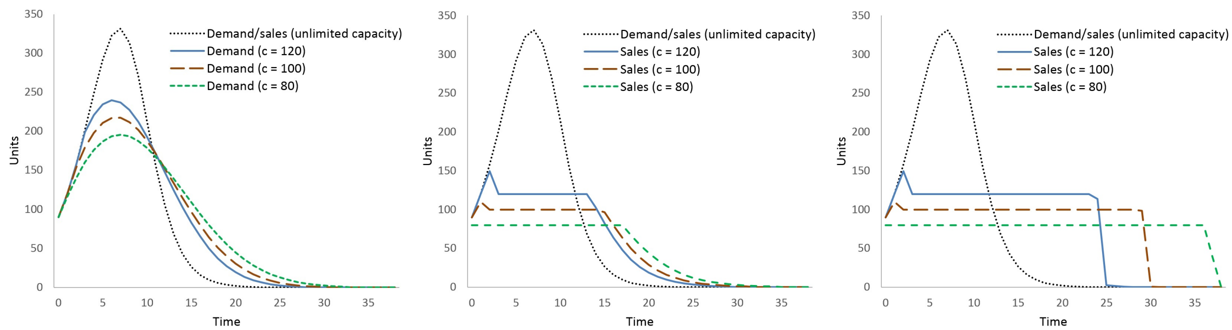
$$d(t) = p(m - D(t)) + (q/m)(m - D(t))D(t).$$

The model considers two main groups of consumers, namely *innovators* and *imitators* categorized based on their *innovativeness*, a personality trait that measures an individual's likelihood of adopting new ideas/technologies earlier than others in the market (Rogers 2003). For a product with a market size m, a proportion, p (coefficient of innovation) of the remaining potential adopters at time t will adopt the product independently (i.e., innovators) while the number of adopting imitators will be proportional to the cumulative demand up to time t, $D(t)$, representing the effect of word-of-mouth (where q denotes the coefficient of imitation). Finally, the unit of time represents the period over which the initial sales ($p \times m$) occur.

In essence, the Bass model assumes infinite supply by considering the diffusion of a product *class* that is marketed by many different companies which makes production constraints less relevant. However, when the new product is produced and marketed by a single firm, it is possible that the demand grows beyond the production capacity resulting in supply shortages and unsatisfied demand. Therefore, the cumulative sales up to time t, $S(t)$, may not necessarily be equal to the cumulative demand, $D(t)$. Assuming that only those adopters that have actually received the product will spread word-of-mouth, supply shortages can influence the number of imitator adopters. Moreover, in the presence of a "binding" capacity constraint, customers with unsatisfied demand will either impatiently abandon the new product (i.e., lost sales) or wait (i.e., demand backlogging) – common real-world phenomena that cannot be explored using the Bass model.

To address the above issues, Ho et al. (2002) and Kumar and Swaminathan (2003) independently propose the following supply-restricted diffusion model, where at any given time t, the number of adopters due to the operation of word-of-mouth is proportional to the cumulative sales, $S(t)$, rather than the cumulative demand, $D(t)$:

$$d(t) = p(m - D(t)) + (q/m)(m - D(t))S(t).$$



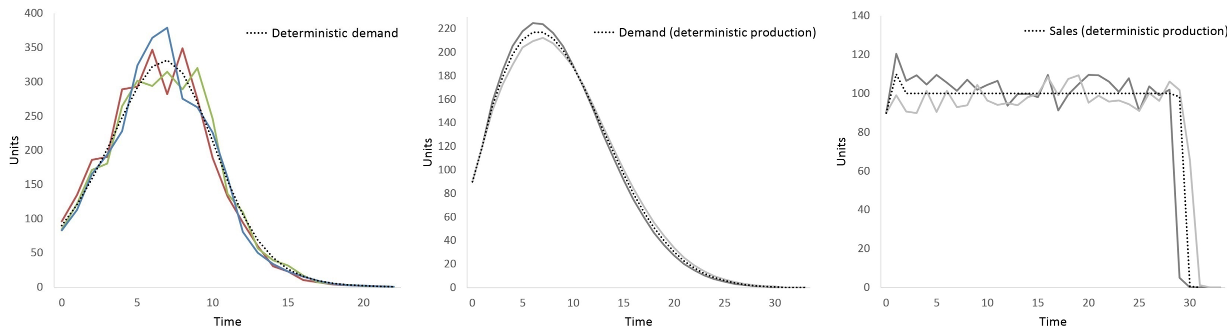
(a) Demand with no backlogging ($\beta = 0$). (b) Sales with no backlogging ($\beta = 0$). (c) Sales with no lost sales ($\beta = 1$).

Figure 2.1: Supply-restricted new product diffusion for the case of zero initial inventory build-up periods ($T_{Build-up} = 0$). The average values of $p = 0.03$ and $q = 0.38$ are used based on a meta-analysis of 213 applications of the Bass model (Sultan et al. 1990). We choose a market potential of $m = 3000$ which has been shown to be large enough to capture demand dynamics and provide statistically reliable results (Cowan and Jonard 2004, Goldenberg et al. 2007). Under these parameter choices, the market will be exhausted in approximately 30 time steps with an average demand of 100 per period. Note that production level, inventory, backlogged demand, and lost sales are not shown to make the figures easier to understand.

In this model, the market dynamics depend not only on past demand but also on past realized sales. It is worth noting that the model is valid for any probability of waiting/backlogging (β) and reduces to the Bass model under unlimited supply where we have $S(t) = D(t)$. Figure 2.1 illustrates the effect of capacity constraints and demand backlogging on demand growth rate, peak time, and magnitude of the peak demand.

2.4.1 Modeling production and demand uncertainties

We adopt a simple but general approach that is commonly used to incorporate randomness keeping in mind that our goal is not to develop complex stochastic analytical models for demand and production yield but rather to evaluate the impact of uncertainties on the optimal build-up policy (see the papers by Shen et al. (2014) and Xu (2010) that also use the same approach). We multiply the instantaneous demand rate at time t, $d(t)$, by a random variable $U_1(t)$ that is uniformly distributed between $1 - v_d$ and $1 + v_d$, where v_d is the maximum variation from the expected deterministic demand (Equation (2.1)). Similarly,



(a) Three sample paths of the stochastic demand ($v_d = 0.15$) under unlimited supply.

(b) Demand ($v_d = 0$) with random yield ($v_p = 0.10$), mean production level ($L_0 = 100$), full backlogging ($\beta = 1$).

(c) Sales with random yield ($v_p = 0.10$), deterministic demand ($v_d = 0$), mean production level ($L_0 = 100$), full backlogging ($\beta = 1$).

Figure 2.2: Sample paths generated under stochastic demand and supply ($p = 0.03$, $q = 0.38$, $m = 3000$, and $T_{Build-up} = 0$). The demand and sales dynamics presented in (b) and (c) correspond to the same two sample paths. The figure shows how the uncertainties affect sales and the timing and magnitude of the peak demand.

the production yield at time t, $y(t)$, is obtained by multiplying the average production level, $L(t)$, by a random variable $U_2(t)$ that is uniformly distributed between $1 - v_p$ and $1 + v_p$, where v_p is the maximum deviation from the mean production level (Equation (2.2)). With no variation (i.e., $v_d = 0$ or $v_p = 0$), the stochastic demand model reduces to the deterministic supply-constrained diffusion model and the throughput will always be equal to the production level. Figure 2.2 shows the impact of uncertainties on the demand and sales dynamics.

$$d(t) = [p(m - D(t)) + (q/m)(m - D(t))S(t)] \times U_1(t) \tag{2.1}$$

and

$$y(t) = L(t) \times U_2(t). \tag{2.2}$$

2.5 Monte Carlo simulation model

Monte Carlo simulation (aka *risk simulation*) is commonly used in many business applications and is widely known as an effective technique to evaluate risk (Loizou and French

2012, Zhang et al. 2015). The main advantage of Monte Carlo simulation is that it helps the decision maker examine the effect of uncertainties in the inputs on the distribution of the outcome(s) of interest. It also allows for sensitivity analysis to help better understand the dynamics of the system being modeled. As a result of its advantages, many business decision-making tools such as *Crystal Ball* and @*Risk* utilize Monte Carlo simulation to help managers assess the uncertainties pertaining to the outcomes. In the context of our problem, Monte Carlo simulation allows us to predict the distribution of the NPV of profit (outcome) and thus evaluate the risk associated with different production-sales policies under stochastic demand and supply (inputs).

Algorithm 1 summarizes the logic of the Monte Carlo simulation model. The simulation run starts by initializing model parameters at time $t = 0$. At the beginning of each time period, the *total* demand, $d_{total}(t)$, is determined by:

$$d_{total}(t) = d_{new}(t) + B(t-1), \tag{2.3}$$

where $d_{new}(t)$ is the demand from new adopters at t sampled using (2.1) and $B(t-1)$ denotes the backlogged demand (i.e., number of waiting customers). At each time period, a proportion (denoted by β) of the unmet demand will be backlogged for future fulfillment. The production yield, $y(t)$, is then sampled using (2.2). Given $I(t-1)$ as the remaining inventory at the end of the previous time step, $y(t)+I(t-1)$ will be the total supply at time t. Let $T_{Build-up}$ be the number of inventory build-up periods. If $t \leq T_{Build-up}$ (i.e., during the build-up period), sales will be zero since the product is not launched. Otherwise, if sales have begun, the company will sell as many units as possible (proven by Ho et al. (2002) and Kumar and Swaminathan (2003) to be the optimal sales plan) and sales ($s(t)$) will be the minimum of the current supply and total demand as given by

Algorithm 1 The logic of the Monte Carlo simulation model

Initialize model parameters/variables at time $t = 0$
while there is still demand for the product (either new or backlogged demand) **do**
  Determine total demand using Equation (2.3)
  Adjust mean production level using Equation (2.5)
  Sample production yield using Equation (2.2)
  Determine sales using Equation (2.4)
  Update cumulative demand and sales
  Determine remaining inventory, $I(t)$, and backlogged demand, $B(t)$, where $I(t) \times B(t) = 0$
  Calculate profit for the current time period
end while
Compute net present value of profit when diffusion is complete

$$s(t) = \begin{cases} 0 & t \leq T_{Build-up}, \\ s(t) = min(y(t) + I(t-1), d_{total}(t)) & t > T_{Build-up}. \end{cases} \tag{2.4}$$

At the end of every time period, the profit is obtained by subtracting the production cost, inventory cost, and cost of waiting customers from revenue. At the end of the simulation run when the market potential is entirely exhausted, the net present value of profit is calculated based on a given discount rate. To avoid unnecessary production and inventory costs near the end of the diffusion process and also to be consistent with the assumptions of the previous studies (Ho et al. 2002, Kumar and Swaminathan 2003), the company produces at the maximum mean production level (L_0) until the demand drops below L_0 for the first time (we represent this time by τ), after which $L(t)$ will be set to a level so that there will be enough supply to meet the demand, a common strategy known as the *lagging demand* (Olhager et al. 2001). We have:

$$L(t) = \begin{cases} L_0 & t \leq \tau, \\ min(L_0, max(d_{total}(t) - I(t-1), 0)) & t > \tau. \end{cases} \tag{2.5}$$

2.6 Experimentation

Extensive experimentation was conducted using the Monte Carlo simulation model. This section describes the experimental design, performance measures, scenario comparison and statistical tests.

2.6.1 Experimental design

The experimental design is summarized in Table 2.1. In order to provide a common ground for comparison, we adopt most of the parameter choices from (Kumar and Swaminathan 2003) and (Negahban and Smith 2014). Our experimental design results in $4 \times 3 \times 3 \times 4 \times 5 \times 4 \times 3 \times 3 \times 3 \times 4 = 311,040$ parameter configurations for which the optimal number of build-up periods is found by performing a one-dimensional search between 0 to 25, resulting in a total of 8,087,040 runs. We use the mean and percentiles of the net present value (NPV) of profit to compare different polices under each parameter configuration. While a preliminary analysis of the results indicates that the mean NPV can be estimated fairly accurately with a few hundred replications, since estimating the percentiles generally requires more data points, we set the number of replications to 1,000 which results in more than 7.1 billion total replications. It is worth noting that there are two main reasons behind our extensive experimentation. First, our objective is to demonstrate that, in general, demand and supply uncertainties can change the choice of the optimal policy and not necessarily to solve a specific case or family of cases. Secondly, the results (as will be discussed in Section 2.7) indeed show that the likelihood of making an incorrect decision varies based on different levels for most of these parameters - important findings that would not have been detected without a such comprehensive experimental design.

2.6.2 Performance measures and statistical tests

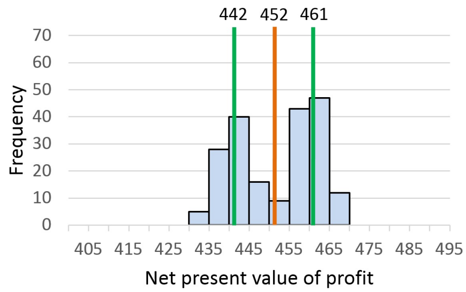
In classical decision theory, risk (commonly perceived as the variability of the distribution of outcomes) has been recognized as one of the most important factors for managers

Parameter	Levels	Value/Range
Demand-related parameters		
Demand variability, v_d	4	0%, 5%, 10%, 15%
Coefficient of innovation, p	3	0.01, 0.03, 0.05
Coefficient of imitation, q	3	0.2, 0.4, 0.6
Market size, m	1	3000
Backlogging percentage, β	4	0, 0.5, 0.8, 1
Production-related parameters		
Maximum mean production level, L_0	5	40, 80, 100, 120, 200
Variability in production yield, v_p	4	0%, 5%, 10%, 15%
Unit production cost, χ	1	1.0
Unit inventory/holding cost, h	3	0.001, 0.005, 0.01
Per customer waiting cost, w	3	0.001, 0.005, 0.01
Unit selling price, π	3	1.1, 1.2, 1.3
Discount rate, r	4	0, 0.003, 0.005, 0.01
Parameter related to the production-sales policy		
Number of inventory build-up periods, $T_{Build-up}$	26	0 – 25

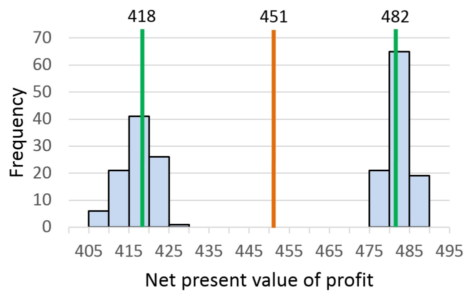
Table 2.1: Parameter choices for the Monte Carlo simulation experiments.

when making any financial investment (March and Shapira 1987). Uncertainty and the shape of the probability distribution continues to play a critical role in decision-making in modern operations management (Kaki et al. 2013) and inventory control (Liang et al. 2014). While previous innovation diffusion studies mainly use the expected NPV of profit as the primary performance measure, we will also assess the risk associated with different production-sales policies using the percentiles of the NPV distribution obtained from Monte Carlo simulation. As discussed by Nelson (2008), the focus on the mean statistic as the primary metric diminishes the power of the tool in characterizing risk. The mean is essentially an expected long-run performance indicator that can provide a good single-point estimate; however, unlike percentiles, it does not provide any information about the uncertainty in the future outcome.

In order to show the importance of percentiles in evaluating the level of uncertainty, Figure 2.3 provides the histogram, mean, 25^{th} and 75^{th} percentiles of the NPV of profit



(a) Myopic policy ($T_{Build-up} = 0$).



(b) Build-up policy ($T_{Build-up} = 8$).

Figure 2.3: Performance measures: mean vs. percentiles ($p = 0.01$, $q = 0.2$, $m = 3000$, $v_d = 0.15$, $L_0 = 80$, $v_p = 0.1$, $\beta = 0.8$, $\chi = 1$, $h = 0.001$, $w = 0.01$, $\pi = 1.2$, $r = 0.003$).

for 200 replications of the simulation model for a myopic ($T_{Build-up} = 0$) and a build-up policy with 8 build-up periods ($T_{Build-up} = 8$) under a given parameter configuration. The mean NPV value for the two policies is virtually the same making us indifferent towards the two policies based on this metric. However, there is a substantially higher uncertainty associated with the build-up policy. While there is a 25% chance of making more than 482 in the build-up policy (which seems not to be achievable with the myopic policy that has a maximum of 469), there is also a 25% possibility of making less than 418 (which is much lower than the minimum observation of 433 from the myopic policy). Therefore, ignoring the uncertainty could lead to a decision with limited potential in making high profits. We also note that the mean value of 451 was not observed in 200 replications of the build-up policy, which represents a case where the mean could be misleading by providing a point estimate of an *average outcome* that might never actually realize (the bi-modal shape of the NPV distribution for this scenario is justified in Section 2.7.2).

We evaluate different policies with respect to three performance measures: (1) average NPV of profit; (2) the 25$^{\text{th}}$ percentile of NPV of profit; and, (3) the 75$^{\text{th}}$ percentile of NPV of profit. It is worth noting that the choice of the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles is somewhat arbitrary, and in fact, any percentile of the NPV of profit could be used with the choice being dependent on how the decision maker defines and measures the uncertainty. From the experimentation point of view, an important implication of using extreme percentiles (such as the 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles) is that more data (i.e., replications) would be needed to collect *enough* observations of these less frequently occurring values to get an accurate estimate.

We use two statistical tests to compare the performance of different production-sales policies (all tests are performed at a 95% confidence level):

- *Welch's t-test for comparing average NPV:* Consider two policies (with different numbers of build-up periods, $T_{Build-up}$) under a specific parameter configuration. The Welch's t-test is performed on the replication results from the two policies to determine a statistical difference between their expected NPV of profit.

- *Nonparametric double bootstrap method for comparing percentiles:* Nonparametric tests are generally useful when standard statistical tests are not readily available, e.g., for comparing ordinal values of two populations (Conover 1980). We use a nonparametric double bootstrap method to determine a statistical difference between a particular percentile of two policies. The general steps of the method (which is also based on a Monte Carlo sampling process) are summarized in Algorithm (2) while the reader is referred to Spiegelman and Gates (2005) for more details.

2.7 Analysis of simulation results

The extensive experimentation performed in this study generated more than 110 gigabytes of raw simulation output necessitating a timely and challenging data analysis. We

Algorithm 2 The double bootstrap method for comparing percentiles (95% confidence level)

Let n denote the number of observations from each policy (i.e., number of replications)
Let p denote the percentile of interest
Step 1. Combine both samples into one column (under the hypothesis that the quantiles for the two samples are the same)
for the number of iterations in the first bootstrap procedure **do**
    Step 2. Draw two samples of size n from the pooled data
    Step 3. Compute the p^{th} percentile of the two samples
end for
Step 4. Compute estimates of the variances for the p^{th} percentile of the two samples from the collection of bootstrap samples
Step 5. Form a t-like statistic for the difference in the bootstrapped sample percentiles
for the number of iterations in the second bootstrap procedure **do**
    Step 6. Repeat steps 2 through 5 and compute the 95% percentile of the t-like statistic to determine the critical value ($t_{critical}$)
end for
Step 7. Compute the t-like statistic for the original data by using the results of the first bootstrap run, $\hat{t}$
Step 8. A statistical difference exists if $\hat{t} > t_{critical}$

developed and verified several automated programs to analyze the data to insure consistency (the computational aspects of the analysis process is out of the scope of this paper). This section presents selected results and outlines the most important findings. To test the main hypotheses of the paper, we consider two types of changes in the optimal policy as a result of the consideration of supply and demand uncertainties in the decision-making process:

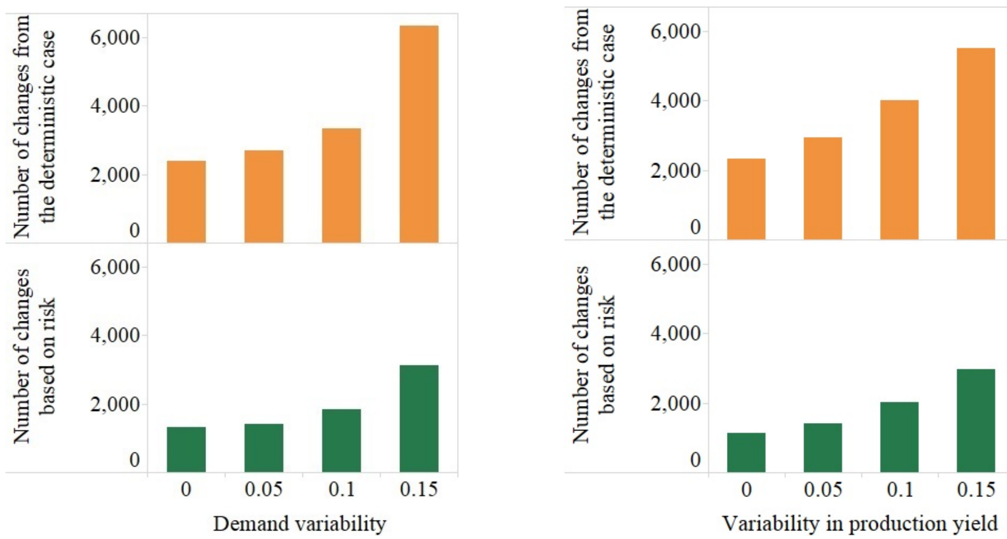
- **Type-1 (change in the decision from the deterministic case):** We are interested in finding the number of scenarios where ignoring uncertainties would yield to a potentially incorrect decision. Under each parameter configuration and level of demand and production uncertainty, we find the optimal policy and determine whether it is different from the optimal policy for the corresponding deterministic case. We then test the statistical difference between the performance of the two polices in the presence of uncertainties (we use the Welch's t-test and double bootstrap method to compare the mean NPV and its percentiles, respectively). A significant statistical difference suggests that the policy selected in the deterministic case is not optimal in the stochastic case under the given performance metric.

- **Type-2 (change in the decision based on risk):** We would like to identify the scenarios where the optimal build-up policy selected based on the expected NPV is different from the optimal policy selected based on risk (percentiles). We use the double bootstrap method to identify a significant statistical difference between the selected percentile of the two policies. A statistical difference then suggests that the policy with the maximum mean NPV is not optimal under the risk measure.

2.7.1 Production and demand uncertainties

Figure 2.4 illustrates the *main effects* of demand and production variability on the number of changes in the decision. As demand and production uncertainties increase, their effect on the optimal number of build-up periods increases so does the number of Type-1 and Type-2 changes and thus the likelihood of making a potentially incorrect decision if uncertainties are ignored.

Figure 2.5 shows how the two sources of uncertainty interact. By looking at the slope of the lines, an interesting finding is that production yield variability has the highest impact under deterministic demand. For higher demand variations, the slopes become gradually gentler suggesting a decrease in the effect of production yield variation. As the level of uncertainty in the demand increases, production uncertainties become less relevant meaning that most of the detected changes in the decision will be due to demand uncertainty. Another important finding is that under any demand variation level, the number of Type-1 and Type-2 changes increases with production uncertainty which is consistent with the general behavior observed in Figure 2.4(b).



(a) Effect of demand uncertainty. (b) Effect of production uncertainty.

Figure 2.4: Number of changes in the optimal build-up policy (out of 77,760 scenarios).

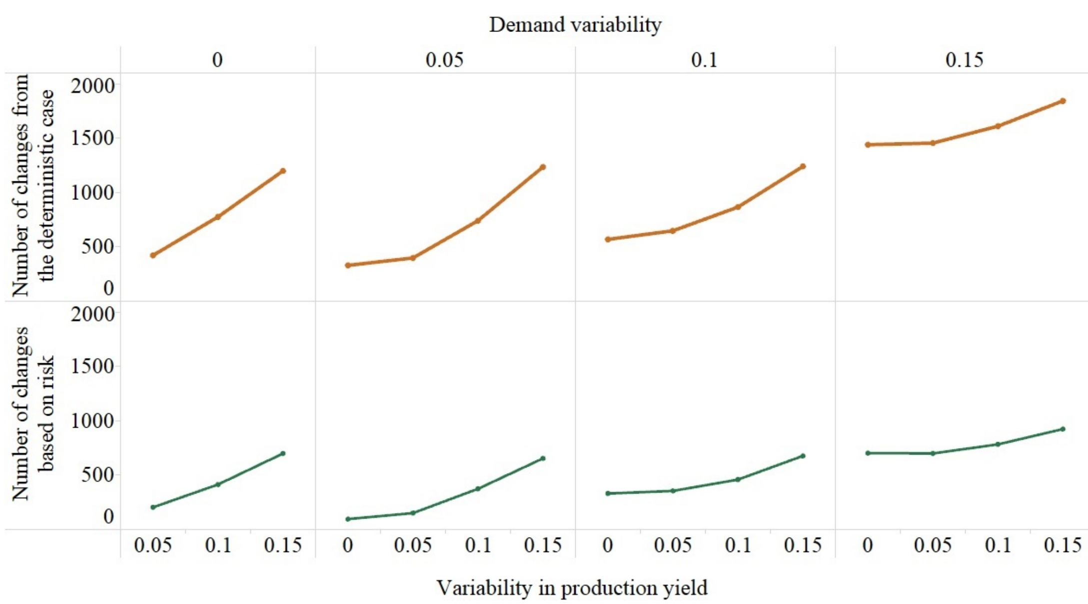


Figure 2.5: Number of changes in the optimal build-up policy (out of 19,440 scenarios).

2.7.2 Production level and diffusion parameters

Figure 2.6 shows that under different production levels (L_0) and coefficients of innovation (p) and imitation (q), the total number of Type-1 and Type-2 changes increases with higher production variability (consistent with the general findings previously discussed). However, perhaps the most interesting finding is that, in general, we see fewer changes in the decision under the following circumstances: (1) low production level and a relatively fast growing demand; and, (2) high production level and a relatively slow demand growth rate. Consider the case of $L_0 = 40$, where except for $p = 0.01$ and $q = 0.2$ (i.e., a very slow demand growth rate), in the other cases where the diffusion speed is relatively higher than production, we see smaller number of changes. In such situations, the demand will remain higher than supply for a long time. The excessive demand surplus reduces the effect of demand variations making sales strictly dependent on the supply level. As an extreme case, consider the case of an infinite demand for the product at each period which will result in sales being completely independent of demand fluctuations. On the other hand, with high production rates and a relatively slow demand process (take (L_0, p, q)=(120,0.01,0.2), (200,0.01,0.2), (200,0.01,0.4), (200,0.03,0.2), (200,0.05,0.2) for instance), due to the abundance of supply, small variations in the production throughput will have virtually no impact on sales (consider the pathological case of infinite production level).

The high number of Type-1 and Type-2 changes under intermediate levels of production level and diffusion speed, e.g., (L_0, p, q)=(120,0.03, 0.2), (80,0.01,0.2), (100,0.01,0.2), (200,0.03,0.4), where a mix of the above two effects exists, can be explained by Figure 2.7 that represents two replications of a parameter configuration. In Figure 2.7(a), the firm produces at maximum capacity until around time period $t = 30$ and then switches to the lagging demand strategy. However, Figure 2.7(b) shows a situation where misinterpreting a random decrease in the demand at time $t = 12$ as the beginning of the decline in demand, misled the firm into making the switch too early. While not plotted in the figure, due to the high

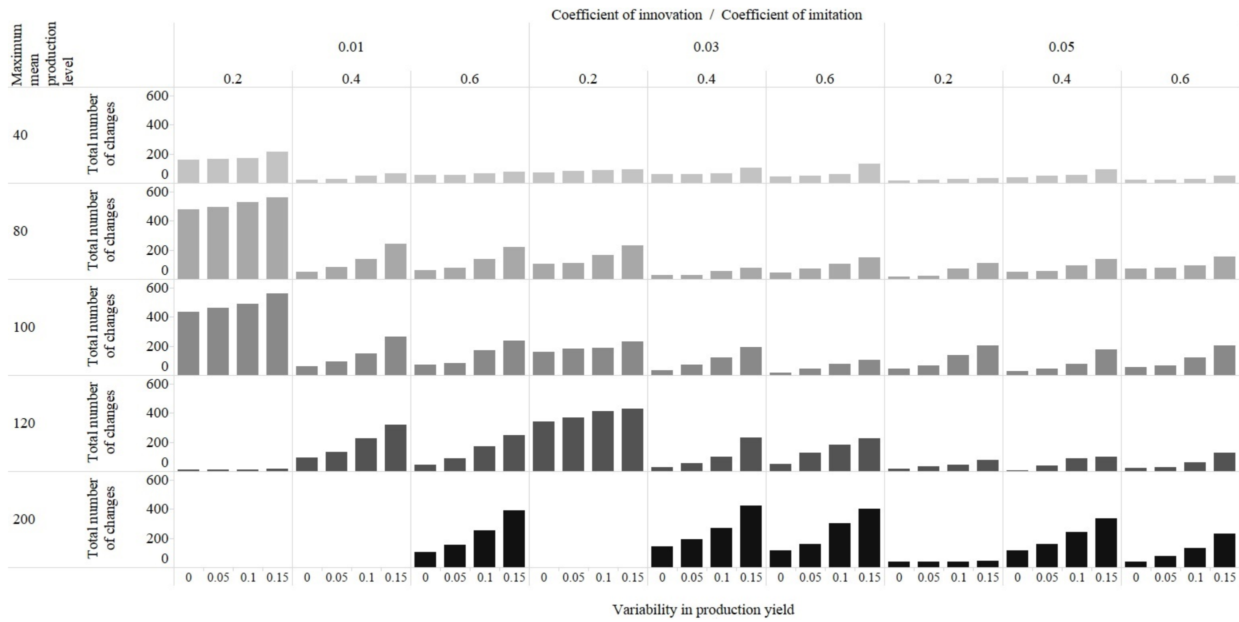
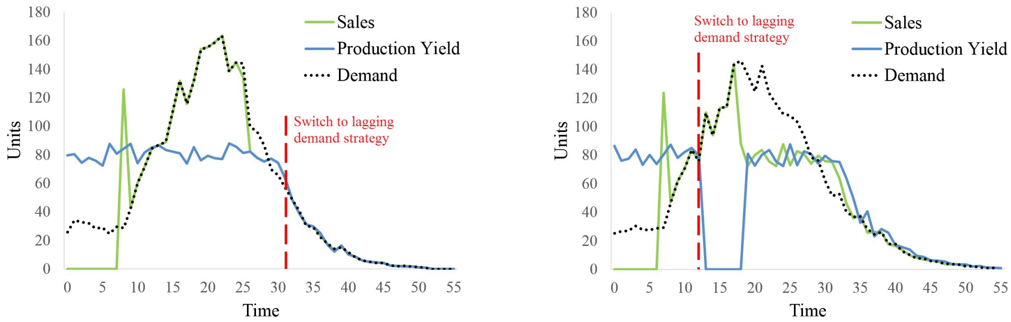


Figure 2.6: Number of changes in the decision broken down by mean production level vs. diffusion parameters (out of 1,728).

inventory level at the time of the switch (582 units), production is stopped and the inventory is used to satisfy demand. By the time that the firm detects the mistake and ramps up production to its maximum level ($t = 18$), the inventory is almost entirely exhausted leading to significant supply shortages and lost sales as demand remains higher than production for several periods (total sales for this scenario is 2,506 units compared to 2,822 for Scenario 1). It is worth noting that this example considers the same set of parameters used in Figure 2.3(b) in Section 2.6.2 which explains the bi-modal shape of the NPV distribution.

2.7.3 Change in the length of the build-up period

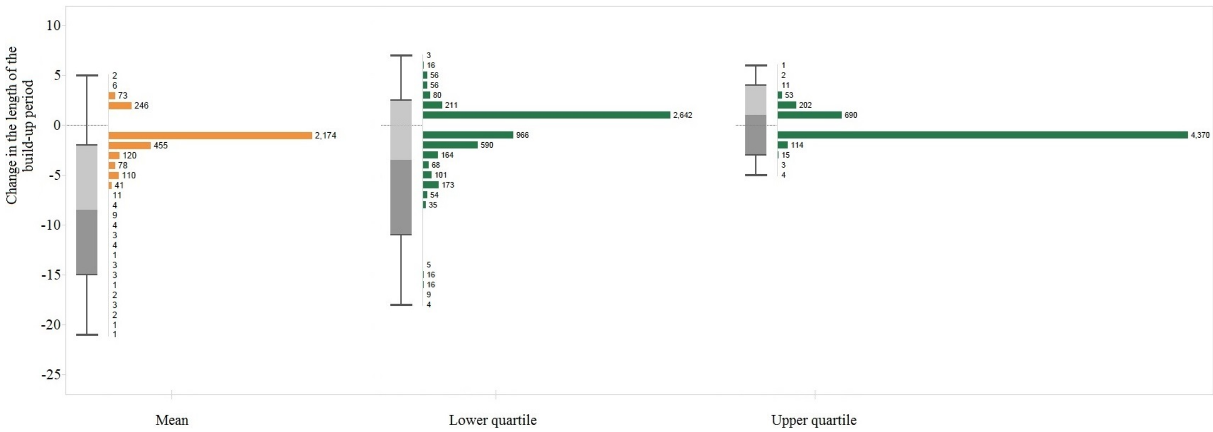
An important question is: how much does the optimal policy change? Figure 2.8 illustrates a significant perspective difference for Type-1 and Type-2 changes. The range of the box plots show that ignoring demand and production uncertainties could result in build-up periods that are from as much as 7 periods shorter to 21 periods longer than the optimal policy selected with the uncertainties taken into account with most of the differences (roughly



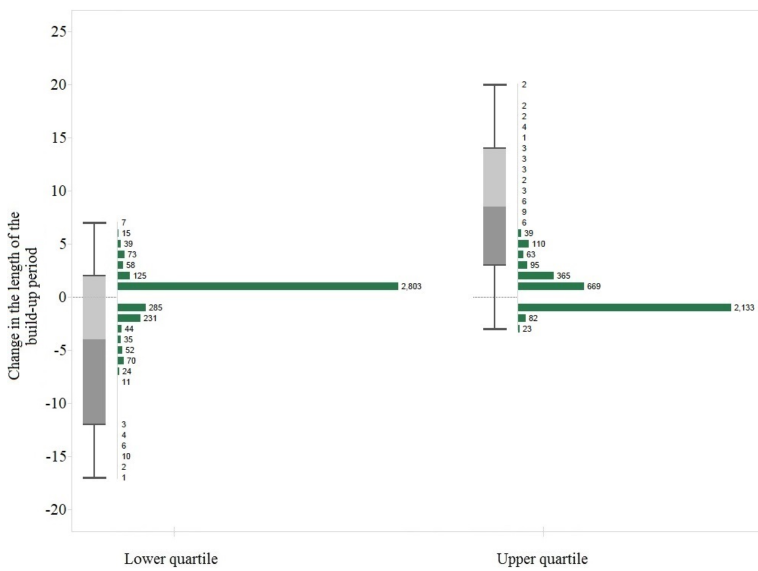
(a) Scenario 1 (NPV of profit = 481.2). (b) Scenario 2 (NPV of profit = 419.8).

Figure 2.7: Two replications of the stochastic supply-restricted simulation model ($p = 0.01$, $q = 0.2$, $m = 3000$, $v_d = 0.15$, $L_0 = 80$, $v_p = 0.1$, $\beta = 0.8$, $\chi = 1$, $h = 0.001$, $w = 0.01$, $\pi = 1.2$, $r = 0.003$).

97%) ranging from 6 periods shorter to 3 periods longer than the respective deterministic cases (Type-1 changes). Moreover, the optimal build-up period selected based on risk can be from as much as 17 periods shorter to 21 periods longer than the policy selected based on the mean NPV of profit with most differences (roughly 98%) being between -6 to +5 (Type-2 changes). An interesting observation is that, when using the average NPV or the 75^{th} percentile of profit, in a considerable number of cases that the decision is affected, ignoring uncertainties would yield to a policy with a longer build-up period by one time unit which can be explained as follows. If the yield becomes more than the average production level in several consecutive periods early in the diffusion process while at the same time the demand is growing slower than expected (relative to the deterministic model), then the net effect would be two-fold. First, the targeted initial inventory is reached faster enabling the company to start sales earlier; and secondly, there will be fewer lost customers during the build-up period. This will move the revenues closer to the present time, reduce inventory costs, and increase cumulative sales and thus profit. On the other hand, when using the 25^{th} percentile, a considerable number of differences indicate that we would incorrectly select a build-up period that is one period shorter if uncertainties are ignored. This can be attributed to the fact that the company will generally need longer periods of build-up to guarantee a certain level of supply that reduces lost sales when demand variations make it grow faster

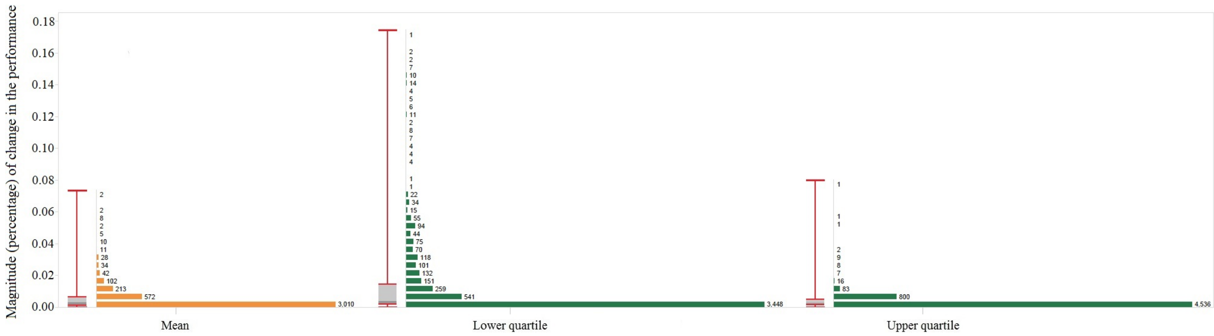


(a) Change in $T_{Build-up}$ from the deterministic case (Type-1).

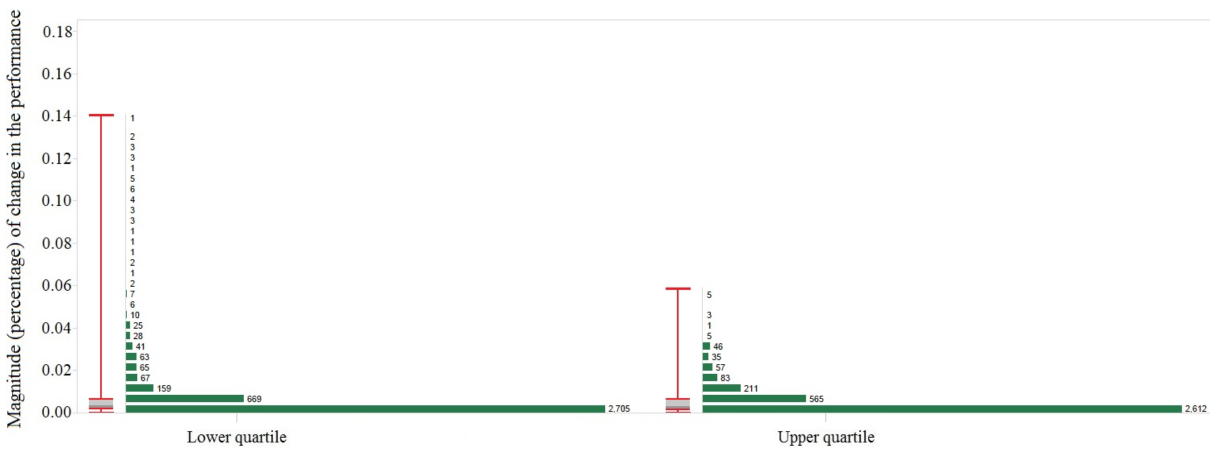


(b) Change in $T_{Build-up}$ based on risk as the primary objective (Type-2).

Figure 2.8: Box plots and histograms of the magnitude of change in the number of initial inventory build-up periods.



(a) Change in the performance from the deterministic case (Type-1).



(b) Change in the performance based on risk as the primary objective (Type-2).

Figure 2.9: Magnitude (percentage) of change in the performance.

than expectations. It is worth noting that these results are also consistent with industry practices as most of the changes in new products' launch time are not too long.

2.7.4 Magnitude of change in the performance measures

The magnitude of change in the performance metrics can be thought of as the *cost* of making a potentially incorrect decision if uncertainties are ignored and thus is important to managers. As shown in Figure 2.9(a), the policy selected based on a deterministic analysis could perform up to 18% worse than the policy that accounts for these uncertainties (Type-1). Figure 2.9(b) shows that the difference in the performance for Type-2 changes can be as high as 14%. In both cases, more than 90% of the cases report up to 4% difference in the performance. While the above differences are all statistically significant, the concept of *meaningful practical difference* (δ) becomes relevant which can be defined as the minimum difference that is important to detect. A difference of less than δ is considered *practically insignificant* meaning that the statistical difference between the two policies is negligible and the decision maker will be indifferent towards selecting either of the two policies. It is then up to the firm to decide whether the identified differences are practically significant. Also, note that a sufficiently large sample size enables us to detect even minute statistical differences (at any chosen confidence level). As a result, the number of replications of the simulation model can be determined by sequentially increasing the number of replications and then stop if a statistical difference of less than δ is detected.

2.7.5 Selling price, backlogging percentage, waiting cost, and inventory cost

Under any discount rate, a higher selling price increases the impact of uncertainties on the profit and thus we expect a higher number of Type-1 and Type-2 changes. Also, high discount rates reduce the effect of sales variations on the NPV for the periods later in the diffusion process and the number of changes in the decision decreases (Figure 2.10). As shown in Figure 2.11, for any waiting cost, the number of Type-1 and Type-2 changes decreases with higher levels of demand backlogging due to the decreased impact of demand and supply uncertainties on lost sales. Finally, as inventory costs increase, long periods of build-up become less attractive even though they are more likely to reduce lost sales. In other words, this reduction in lost sales does not outweigh the higher inventory cost. Therefore, we expect a decrease in the number of Type-1 and Type-2 changes, particularly, for cases where a longer build-up period would have otherwise been preferable (Figure 2.12).

2.8 Conclusions

Motivated by real-world problems, we investigate the effect of supply and demand uncertainties on the optimal production-sales policy for new products. We develop a Monte

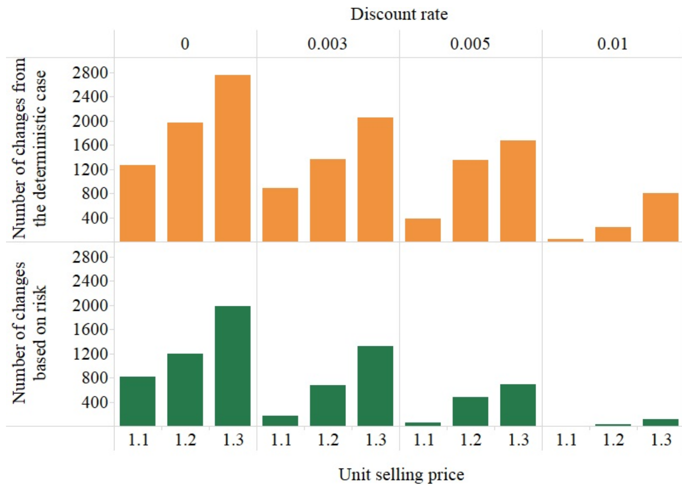
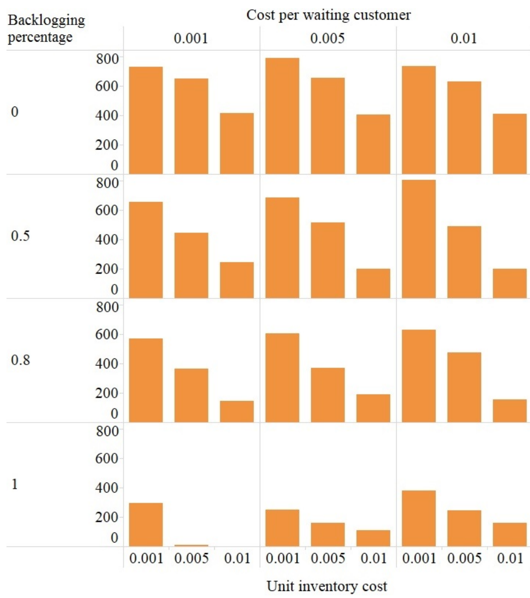
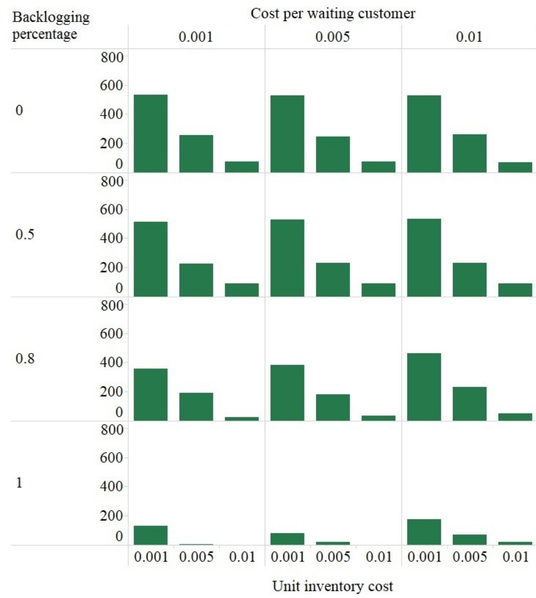


Figure 2.10: Number of changes (out of 25,920) in the decision broken down by discount rate and unit selling price.

Figure 2.11: Number of changes (out of 25,920) in the decision broken down by waiting cost and backlogging percentage.



(a) Number of changes from the deterministic case.



(b) Number of changes based on risk.

Figure 2.12: The effect of inventory cost on the number of changes in the decision (out of 8,640 cases).

Carlo simulation model of the stochastic supply-restricted new product diffusion and through extensive experimentation, we investigate the five important questions posed in Section 2.2:

1. We show that, regardless of the primary performance measure, the optimal policy from a deterministic analysis may not necessarily be optimal under stochastic supply and demand.

2. Using percentiles as measures of risk, we show that the policy selected based on the expected profit is not necessarily optimal under risk measures.

3. The results show that as the level of uncertainty in the demand increases, production uncertainties become less important. We also show that under any demand variation level, the likelihood of making an incorrect decision increases with production uncertainty. The effect of the diffusion speed on this likelihood is shown to vary depending

on the production level. Moreover, this likelihood is found to have an inverse relationship with the cost per waiting customer, unit inventory cost, and discount factor while a direct relationship with selling price is observed.

4. We provide perspective results on the effect of uncertainties on the optimal number of build-up periods. Our analysis suggests that, in most cases, ignoring uncertainties or focusing solely on the expected NPV of profit can lead to a sub-optimal policy with a build-up period that is from as much as 6 periods shorter to 6 periods longer than that of the optimal policy.

5. The results indicate that the cost of overproduction or underproduction as a result of ignoring supply and demand uncertainties can lead to as much as 18% less profit.

Managerial implications: The above findings have important implications for managers in real-world applications:

- Ignoring demand and supply constraints may result in either underestimation or overestimation of the required pre-launch inventory, both leading to decreased profit. Insufficient build-up inventory increases the likelihood of stock-outs and lost sales while overproduction incurs excessive production and inventory costs. However, the negative impact of longer build-up periods and thus later market entry is aggravated in the presence of competition. The marketing literature provides strong and consistent support suggesting that market-share rewards to pioneers. Pioneers also have other advantages, namely shaping consumers preferences and establishing consumer loyalty, avoiding consumer switching cost compensations, gaining performance advantages from early sales, establishing and maintaining standards, and preempting preferred patents and suppliers (Hauser et al. 2006). Therefore, consideration of supply and demand uncertainties becomes even more important in an oligopolistic market structure and thus our assumption of a monopolistic market will not undermine our major findings.

- Our findings can also help managers allocate resources to parameter estimation more effectively. In general, the likelihood of making an incorrect decision is more sensitive to the uncertainty in the demand rather than production yield which implies it is more important to obtain a precise estimate of the diffusion parameters and market volatility. However, for the case of a relatively slow demand, it is critical to accurately predict the production yield variation. A precise estimate of the price is crucial when the discount rate is small or for products with very short life-cycles. The company should also give more importance to estimating the backlogging percentage rather than the cost associated with waiting customers. In situations where a high percentage of customers are willing to wait for new supply, such as the monopolistic market of landline phone in Israel (Jain et al. 1991), supply and demand uncertainties become less relevant.

- Managers can easily misinterpret random fluctuations as suggesting general patterns in the demand process. We showed the high cost of stopping production due to misjudging a random decrease in the demand as the beginning of the decline in the diffusion process and its dramatic impact on the distribution of the NPV of profit. Therefore, a wise strategy would involve smooth changes in production level until there is sufficient evidence of decline. This balanced strategy will depend on the trade-off between the cost of possible overproduction and lost sales due to an early and dramatic production ramp-down.

- The non-linear first- and higher-order interaction effects of different factors prohibit any sort of extrapolation/interpolation on the results. Therefore, experimentation would still be necessary to determine the optimal policy for the estimated parameters. Moreover, while our focus is on statistically significant changes in the decision, further analysis of the simulation results reveals that for many parameter configurations, two or more policies perform virtually the same in terms of the net present value of profit. In such cases, multi-criteria decision-making processes can be used to compare inventory, backlogged demand, production cost, and sales trajectories to select an appropriate policy.

Our model involves several assumptions and certainly does not capture every dimension of this complex problem. We consider random fluctuations in the demand process that cause deviations from a base pattern predicted by a deterministic demand model. Our assumption is that accurate estimates of the actual diffusion parameters are available to the decision maker. In the real world, this may not always be the case and a robust production-sales policy needs to be selected through sensitivity analysis on a range of possible parameter values to account for forecasting errors. The model can be modified to implement any complex production planning approach. Moreover, an agent-based version of the model can also be developed to account for more realistic behavioral and adoption rules. Alternative approaches to incorporate stochasticity into the demand and production processes could also be investigated. For the case of the production yield variability, if the probability distribution turns out to have a significant effect, then it would be necessary to explicitly analyze the production/supply process to estimate an appropriate distribution (using discrete event simulation or any other analysis tool). Other interesting extensions include consideration of competition (using a multi-product supply-restricted demand model), outsourcing options, and other types of uncertainties that exist in a supply chain for a new product.

Bibliography

Amini, M., H. Li. 2011. Supply chain configuration for diffusion of new products: An integrated optimization approach. *Omega* **39**(3) 313–322.

Amini, M., T. Wakolbinger, M. Racer, M. G. Nejad. 2012. Alternative supply chain production-sales policies for new product diffusion: An agent-based modeling and simulation approach. *European Journal of Operational Research* **216**(2) 301–311.

Azadeh, A., M. Moghaddam, S. Asadzadeh, A. Negahban. 2011. An integrated fuzzy simulation-fuzzy data envelopment analysis algorithm for job-shop layout optimization: The case of injection process with ambiguous data. *European Journal of Operational Research* **214**(3) 768–779.

Azadeh, A., A. Negahban, M. Moghaddam. 2012. A hybrid computer simulation-artificial neural network algorithm for optimisation of dispatching rule selection in stochastic job shop scheduling problems. *International Journal of Production Research* **50**(2) 551–566.

Bass, F. M. 1969. A new product growth for model consumer durables. *Management Science* **15**(5) 215–227.

Bass, F. M. 2004. Comments on "A new product growth for model consumer durables". *Management Science* **50**(12) 1833–1840.

Bilginer, O., F. Erhun. 2010. *Managing Product Introductions and Transitions*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Bilginer, O., F. Erhun. 2015. Production and sales planning in capacitated new product introductions. *Production and Operations Management* **24**(1) 42–53.

Cantamessa, M., C. Valentini. 2000. Planning and managing manufacturing capacity when demand is subject to diffusion effects. *International Journal of Production Economics* **66**(3) 227–240.

Conover, W. J. 1980. *Practical Nonparametric Statistics*. 3rd ed. John Wiley Inc., New York, NY, USA.

Cowan, R., N. Jonard. 2004. Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control* **28**(8) 1557–1575.

Erel, E., I. Sabuncuoglu, B. A. Aksu. 2001. Balancing of U-type assembly systems using simulated annealing. *International Journal of Production Research* **39**(13) 3003–3015.

Fourt, L., J. Woodlock. 1960. Early prediction of market success for new grocery products. *Journal of Marketing* (2) 31–38.

Goldenberg, J., B. Libai, S. Moldovan, E. Muller. 2007. The NPV of bad news. *International Journal of Research in Marketing* **24**(3) 186–200.

Hauser, J., G. J. Tellis, A. Griffin. 2006. Research on innovation: A review and agenda for marketing science. *Marketing Science* **25**(6) 687–717.

Higuchi, T., M. D. Troutt. 2004. Dynamic simulation of the supply chain for a short life cycle product–lessons from the Tamagotchi case. *Computers & Operations Research* **31**(7) 1097–1114.

Ho, T.-H., S. Savin, C. Terwiesch. 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science* **48**(2) 187–206.

Jain, D., V. Mahajan, E. Muller. 1991. Innovation diffusion in the presence of supply restrictions. *Marketing Science* **10**(1) 83–90.

Kaki, A., A. Salo, S. Talluri. 2013. Impact of the shape of demand distribution in decision models for operations management. *Computers in Industry* **64**(7) 765–775.

Kanniainen, J., S. Makinen, R. Piche, A. Chakrabarti. 2011. Forecasting the diffusion of innovation: A stochastic bass model with log-normal and mean-reverting error process. *IEEE Transactions on Engineering Management* **58**(2) 228–249.

Kumar, S., J. M. Swaminathan. 2003. Diffusion of innovations under supply constraints. *Operations Research* **51**(6) 866–879.

Li, J., D. E. Blumenfeld, N. Huang, J. M. Alden. 2009. Throughput analysis of production systems: Recent advances and future topics. *International Journal of Production Research* **47**(14) 3823–3851.

Liang, C., S. P. Sethi, R. Shi, J. Zhang. 2014. Inventory sharing with transshipment: Impacts of demand distribution shapes and setup costs. *Production and Operations Management* **23**(10) 1779–1794.

Loizou, P., N. French. 2012. Risk and uncertainty in development: A critical evaluation of using the monte carlo simulation method as a decision tool in real estate development projects. *Journal of Property Investment & Finance* **30**(2) 198–210.

Mahajan, V., E. Muller, R. A. Kerin. 1984. Introduction strategy for new products with positive and negative word-of-mouth. *Management Science* **30**(12) 1389–1404.

March, J. G., Z. Shapira. 1987. Managerial perspectives on risk and risk taking. *Management Science* **33**(11) 1404–1418.

Mula, J., R. Poler, J. Garcia-Sabater, F. Lario. 2006. Models for production planning under uncertainty: A review. *International Journal of Production Economics* **103**(1) 271–285.

Negahban, A. 2013. A hybrid simulation framework for the newsvendor problem with advertising and viral marketing. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, M. E. Kuhl, eds., *Proceedings of the 2013 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., 1613–1624.

Negahban, A., J. S. Smith. 2014. Production-sales policies for new product diffusion under stochastic supply. *Proceedings of the Sixth International Conference on Advances in System Simulation*. International Academy, Research, and Industry Association (IARIA), Inc., Red Hook, New York, 197–202.

Negahban, A., L. Yilmaz. 2014. Agent-based simulation applications in marketing research: An integrated review. *Journal of Simulation* **8**(2) 129–142.

Negahban, A., L. Yilmaz, T. Nall. 2014. Managing production level in new product diffusion: An agent-based simulation approach. *International Journal of Production Research* **52**(17) 4950–4966.

Nelson, B. 2008. The MORE plot: Displaying measures of risk & error from simulation output. S. J. Mason, R. R. Hill, L. Monch, O. Rose, T. Jefferson, J. W. Fowler, eds., *Proceedings of the 2008 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 413–416.

Niu, S.-C. 2002. A stochastic formulation of the bass model of new-product diffusion. *Mathematical Problems in Engineering* **8**(3) 249–263.

Norton, J. A., F. M. Bass. 1987. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science* **33**(9) 1069–1086.

Olhager, J., M. Rudberg, J. Wikner. 2001. Long-term capacity management: Linking the perspectives from manufacturing strategy and sales and operations planning. *International Journal of Production Economics* **69**(2) 215–225.

Peres, R., E. Muller, V. Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* **27**(2)

91–106.

Rezg, N., X. Xie, Y. Mati. 2004. Joint optimization of preventive maintenance and inventory control in a production line using simulation. *International journal of production research* **42**(10) 2029–2046.

Rogers, E. M. 2003. *Diffusion of innovations*. 5th ed. Free Press, New York, NY, USA.

Shen, W., I. Duenyas, R. Kapuscinski. 2011. New product diffusion decisions under supply constraints. *Management Science* **57**(10) 1802–1810.

Shen, W., I. Duenyas, R. Kapuscinski. 2014. Optimal pricing, production, and inventory for new product diffusion under supply constraints. *Manufacturing & Service Operations Management* **16**(1) 28–45.

Skiadas, C. H., A. N. Giovanis. 1997. A stochastic bass innovation diffusion model for studying the growth of electricity consumption in greece. *Applied Stochastic Models and Data Analysis* **13**(2) 85–101.

Spiegelman, C., T. J. Gates. 2005. Post hoc quantile test for one-way analysis of variance using a double bootstrap method. *Transportation Research Record: Journal of the Transportation Research Board* **1908** 19–25.

Stevenson, M., L. C. Hendry, B. G. Kingsman. 2005. A review of production planning and control: The applicability of key concepts to the make-to-order industry. *International Journal of Production Research* **43**(5) 869–898.

Sultan, F., J. U. Farley, D. R. Lehmann. 1990. A meta-analysis of applications of diffusion models. *Journal of Marketing Research* **27**(1) 70–77.

Xu, H. 2010. Managing production and procurement through option contracts in supply chains with random yield. *International Journal of Production Economics* **126**(2) 306–313.

Zhang, W., P. Padmanabhan, C.-H. Huang. 2015. Sequential capital investment decision making under extreme cash flow situations: Evidence using monte carlo simulation. *Journal of Business Economics and Management* **16**(5) 877–900.

Chapter 3

Optimal entry time and production-sales policies for successive generations of new products

3.1 Abstract

This paper explores the optimal entry timing, capacity, initial inventory at launch, and sales plan for a new generation of a product. We modify an existing multi-generation demand model to account for supply constraints. A mathematical model of the supply-restricted multi-generation diffusion problem is then developed and solved both analytically and numerically to find the optimal solution to the above set of inter-dependent decision-making problems. The paper provides insights about the effect of consumers' backlogging, cost of production capacity, unit profit margin for the products, market expansion by the new generation, and cannibalization of older generations. We show that the consideration of supply restrictions and lost sales adds realism to the analysis and generalizes previous findings on market entry policy by showing that there is a continuous range of optimal introduction times depending on the level of cannibalization and market extension. Empirical implications of the proposed model are explored for the diffusion of Sony's PlayStation®3 game console in Europe. The model describes the diffusion of the successive generations of the game console fairly accurately. The results of our application suggests that the company introduced the product too late and overproduced inventory which, as supported by empirical evidence, had a negative impact on the product's performance. Limitations and considerations for the application of the model are also discussed.

keywords: multi-generation innovation diffusion, market entry timing, myopic and build-up policies, PlayStation®3 game console

3.2 Introduction

When introducing a new generation of a product, firms need to consider several key factors including high investment costs, cannibalization of the older generations, the risk of supply shortages and lost sales due to insufficient production capacity or inventory at launch, and the risk of not reaching the full market potential and losing potential profit due to a non-optimal launch time (pre-matured or delayed). Firms need to determine an appropriate combination of initial inventory and production capacity to satisfy future demand. Due to the high cost of production capacity and long lead time for capacity expansions (or reductions), companies generally try to build sufficient inventory prior to starting sales as a cushion in case the demand grows rapidly and exceeds the production capacity (*build-up* policy). While this policy delays revenue collection from sales and leads to additional inventory costs, the risk of substantial lost sales associated with the alternative *myopic* policy (i.e., starting sales without building an initial inventory) is often too high for managers to take. These supply-related decisions also depend on the choice of the product's launch time which in turn is dependent on the expansion in the potential market as well as the cannibalization effect. Cannibalization is when the new generation of the product gains additional sales that would have otherwise gone to the existing product generations. Cannibalization becomes critically important when the older generation has a higher unit profit margin which necessitates evaluating the trade-off between the market extension versus losing a percentage of high-margin sales from the existing product. Finally, the company needs an appropriate sales plan for its product generations.

Due to the complexity of the problem, even companies with significant experience in successful new product launches have faced considerable monetary losses due to inappropriate decisions. For instance, Sony Electronics Inc. lost \$1.8B in its game division and laid off 3% of its workforce due to delayed launch and excessive production and inventory costs for PlayStation®3 (PS3) when, opposite to expectations, the product's growth was slower than its predecessor (*Los Angeles Times*, June 7, 2007). While demand uncertainties and misjudging the diffusion parameters could be part of the problem, understanding the interdependence of these decisions is the first step towards better decisions. This paper aims at taking this first step by exploring three main research questions about the supply-constrained multi-generation diffusion that the current literature leaves unanswered:

1. When should the firm introduce the new generation to maximize its profit?

2. What is the optimal production capacity and build-up policy for the new generation?

3. What is the optimal sales plan for the new and existing generations?

Figure 3.1 summarizes the research methodology. We consider a single firm contemplating the introduction of a new generation of a durable product (i.e., our focus is on the first-time demand with no repeat purchases) where the technology for the new generation becomes available sometime after the introduction of the predecessor. We start by modifying an existing multi-generation demand model to account for supply constraints. We then formulate the problem in the context of *Optimal Control Theory* (Sethi and Thompson 2000) and determine the optimal sales plans for successive generations (Section 3.4). Closed-form solutions are then derived and analyzed for the special case of *patient customers* with no lost sales (Section 3.5). To gain insight about the optimal build-up policy, production capacity, and market entry policy for the general case of *impatient customers* (i.e., partially backlogged demand with lost sales), a comprehensive numerical study is performed using a discrete-time version of the model (Section 3.6). While previous studies suggest that the new generation must be introduced either simultaneously with the previous generation, at the maturity of the previous generation, or should never be introduced at all, our model suggests a continuous range from *Now* to *Never* for the optimal entry time. We present an application of the proposed model to the case of Sony's PlayStation®3 game console in Europe (Section 3.7). The results indicate a late introduction and overproduction of the product which are inline with the empirical evidence on the product's poor performance.

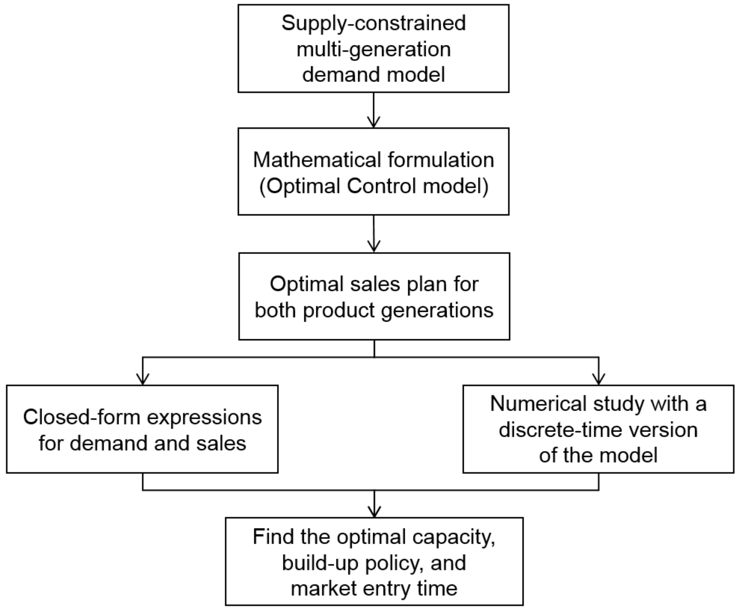


Figure 3.1: The research methodology.

We discuss important considerations for decision-making in real-world applications and the limitations of the proposed model (Section 3.8).

To the best of our knowledge, this work is the first to study the supply-constrained multi-generation diffusion problem. A recent survey paper by Bilginer and Erhun (2010) reports that there is no paper that analyzes capacity constraints for successive generations of innovations. Based on our critical literature analysis (discussed in Section 3.3), the contributions of this paper can be summarized as follows: (1) it presents a modified multi-generation demand model that accounts for supply constraints and develops an integrated model that considers production and inventory costs, demand backlogging, and lost sales; (2) it establishes the optimal sales plan that maximizes the total profit; (3) it yields more realistic market entry policies that generalize previously proposed *Now or Never* and *Now or Maturity* policies; and, (4) it provides insights on the optimal combination of production capacity and build-up inventory for the new product generation. Our case study also has important practical implications for managers by showing that the model can potentially help companies avoid a significant financial loss due to an inappropriate introduction time or production-sales policy for successive generations of their products.

3.3 Related literature

We present a critical analysis of the strengths and gaps of the two main streams of research, namely the *marketing* and *operations management* literature, to establish the contributions of this paper.

The focus of the marketing research is primarily on demand forecasting. The majority of the proposed models assume infinite supply and are generally rooted in the Bass model (Bass 1969) which is perhaps the most commonly used tool for forecasting the demand of new products in the industry (Bass 2004) and its efficacy has been verified empirically for hundreds of products (Sultan et al. 1990, Mahajan et al. 1995). Common extensions to the Bass model include incorporation of negative word-of-mouth, stochasticity in the demand, and modeling the demand of successive technology generations. For a comprehensive analysis of these extensions, see Peres et al. (2010) and Mahajan et al. (1991). There are also simulation studies, especially agent-based simulation models, that investigate the diffusion of new products under various social network structures, population heterogeneity, and other micro-level factors that lead to the emergence of macro-level market dynamics (see Negahban and Yilmaz (2014) for comprehensive analysis of these studies). Among these extensions, multi-generation demand models are of particular interest for our purpose. In a seminal work, Norton and Bass (1987) propose a multi-generation model of adoption and substitution that captures both market expansion and cannibalization. The model is validated empirically for several products, however no analysis is provided on the optimal entry time for successive generations. In another seminal work, Wilson and Norton (1989) propose a two-generation demand model for durable products and develop analytical expressions for the total profit from both generations. Their analysis leads to the *Now or Never* market entry policy. This policy suggests that the second generation should either be introduced simultaneously with the first generation or not be introduced at all. Under a slightly different set of assumptions, Mahajan and Muller (1996) suggest that the optimal entry policy for the new generation follows a *Now or Maturity* rule as a generalization of the *Now or Never* policy suggesting that the firm should either introduce the new generation as soon as it is available or at the maturity of the predecessor. For other related studies, see Norton and Bass (1992) and Kim et al. (2000).

The strength of the marketing stream is primarily in the development of multi-generation diffusion models and their successful implementation in forecasting the demand for different product families. These studies also provide valuable insights about the diffusion dynamics, market entry timing, cannibalization and substitution of successive product generations. However, perhaps the most important deficiency in this stream of research is that the proposed "*Now or Never*" and "*Now or Maturity*" policies seem to be not applicable to many product categories. Different industries, namely fashion, high-technology, and pharmaceutical, report alternative timing strategies. We believe the main reason that the proposed policies are not consistent with many industry practices lies in the fact that these models ignore capacity constraints, inventory cost and build-up policies, and lost sales. Moreover, a clear definition of *maturity* is missing. This paper contributes to this stream by extending the Wilson-Norton demand model to account for supply constraints and developing an integrated mathematical model that considers the above factors. Our analysis yields more realistic market entry policies that generalize the previously proposed policies and enables the decision-maker to define different diffusion phases (i.e., *early*, *middle*, or *maturity*).

The second stream of research involves the marketing-operations interface. To account for supply restrictions for a single generation, a modified Bass model is independently proposed by Kumar and Swaminathan (2003) and Ho et al. (2002). The papers develop virtually similar models and determine the optimal sales plan and build-up policy in the presence of lost sales, capacity and inventory costs, and discounting. For similar studies, see Shen et al. (2011), Bilginer and Erhun (2015), Shen et al. (2014), Cantamessa and Valentini (2000), Amini and Li (2011). Several agent-based simulation studies also consider supply-related aspects of the single-product diffusion (Negahban et al. 2014, Amini et al. 2012, Negahban 2013). While these studies provide a good understanding of the supply-restricted diffusion and timing decisions, the effect of market extension, substitution, and cannibalization in the presence of multiple generations are ignored. Very few studies consider the supply side of successive product generations. Ke et al. (2013) consider the diffusion of two generations under only one supply replenishment during the entire planning horizon. The first order is made at the launch of the first generation and satisfies its future demand. The second order occurs at the launch of the second generation and will satisfy its future demand. While they consider inventory costs, the diffusion of both generations is unconstrained and thus lost sales are ignored. Moreover, they ignore time value of money as well as fixed and variable (per unit) costs associated with replenishments. Although their model yields a continuous range for the optimal introduction time for the second generation in the single-replenishment case, for the more realistic and general case of frequent replenishments (i.e., production process), their model yields *Now or Never*. In two related studies (Wilhelm and Xu 2002, Wilhelm et al. 2003), suppliers, production, backorders, inventory, and pricing are considered while in neither of the studies the demand follows diffusion dynamics – a strong assumption that was mainly used to make the problem analytically tractable.

Despite their theoretical contributions, the above limitations inhibit any real-world applications. We relax many of these assumptions in a more realistic model and provide empirical evidence by successfully applying the model to the case of Sony's PlayStation®3 game console in Europe. The results validate the model's potential in helping companies choose an appropriate release time and build-up policy for successive generations of their products.

Table 3.1: Model parameters and notation

Notation	Description
i	Index of generation number
c_i	Production capacity for the i$^{\text{th}}$ generation
H_i	Cost per unit of capacity for generation i
τ	Launch time of the 2$^{\text{nd}}$ generation
A_1	Cumulative awareness at the launch of the 2$^{\text{nd}}$ generation ($A(\tau)$)
T_B	Build-up period for the 2$^{\text{nd}}$ generation
h_i	Unit holding/inventory cost for generation i
l_i	Rate of loss of waiting customers for generation i
$\alpha_i(t)$	Profit margin of the i$^{\text{th}}$ generation at time t
p	Coefficient of innovation
q	Coefficient of imitation
N	Number of consumers in the market (market size)
m_1	Generation 1 adoption rate before generation 2
m_2	Generation 1 adoption rate after generation 2
m_3	Generation 2 adoption rate
θ	Discount factor
$a(t)$, $A(t)$	Instantaneous and cumulative awareness at time t
$d_i(t)$, $D_i(t)$	Demand rate and cumulative demand for generation i at time t
$s_i(t)$, $S_i(t)$	Sales rate and cumulative sales for generation i at time t
$I_i(t)$	Inventory level of the i^{th} generation at time t
$L_i(t)$	Cumulative number of lost customers for the i^{th} generation at time t
$W_i(t)$	Waiting customer population for the i^{th} generation at time t
$r_i(t)$	Production rate of the i^{th} generation at time t

3.4 Model formulation

In this section, we propose a demand model that accounts for supply restrictions and establish the optimal sales plan that maximizes the total profit. Table 3.1 summarizes the notations.

3.4.1 Demand model

Wilson and Norton (1989) model the flow of information about the product among consumers as a Bass process. Adopters stop searching for and processing information. At any given time t, from the remaining uninformed population, a proportion of p become aware of the product's generations independently while the number of individuals informed through word-of-mouth is proportional to the number of previous adopters. The instantaneous awareness rate at time t, $a(t)$, is given by

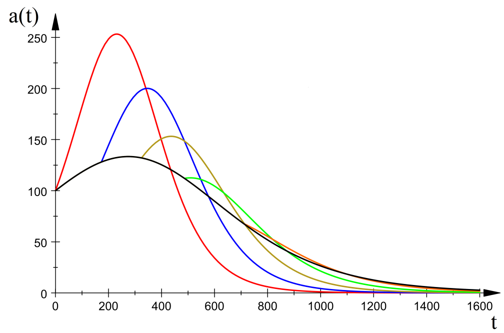
$$a(t) = p(N - A(t)) + (q/N)(N - A(t))[D_1(t) + D_2(t)], \tag{3.1}$$

Let $t = 0$ be the release time of the first generation (Product 1) and τ the launch time of the second generation (Product 2). For $0 \leq t < \tau$, a fraction of m_1 of the informed population adopt Product 1. After the introduction of the second generation ($t \geq \tau$), besides its own new market potential, some of the consumers who would otherwise have adopted Product 1 will adopt Product 2. Given a choice of both generations, let m_2 and m_3 be the fraction of informed population that adopt Product 1 and 2, respectively. The cumulative demands can be expressed as

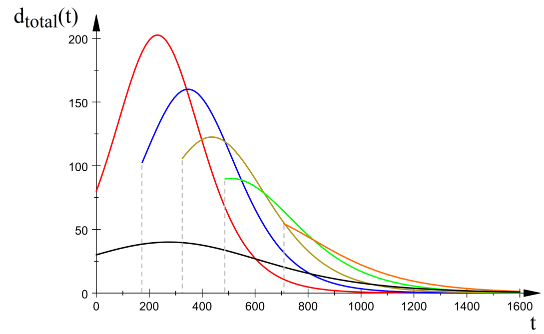
$$D_1(t) = \begin{cases} m_1 A(t) & 0 \leq t < \tau, \\ m_1 A(\tau) + m_2(A(t) - A(\tau)) & t \geq \tau. \end{cases} \tag{3.2}$$

$$D_2(t) = \begin{cases} 0 & 0 \leq t < \tau, \\ m_3(A(t) - A(\tau)) & t \geq \tau. \end{cases} \tag{3.3}$$

The underlying assumption of the Wilson-Norton model is an infinite supply (i.e., no limits on the firm's ability to meet demand). Capacity constraints can lead to supply shortages and unsatisfied demand. As a result, the cumulative sales for generation i up to time t, $S_i(t)$, is not necessarily equal to its cumulative demand, $D_i(t)$. Assuming that word-of-mouth is spread by adopters that actually received the product, supply shortages influence future demand and sales dynamics. Lost sales and demand backlogging become relevant in the presence of a "binding" capacity. We propose a modified supply-constrained Wilson-Norton model by representing the effect of word-of-mouth as being proportional to the cumulative



(a) Information diffusion rate.



(b) Combined demand rate for both generations.

Figure 3.2: Information acquisition rate and combined demand rate for the two generations in the original Wilson-Norton model plotted for different values of introduction time of the second generation (τ). The lower curve (an initial segment of which is common to any upper curve) represents the rate for Product 1 before the introduction of Product 2. The parameters are $N = 100,000$, $p = 0.001$, $q = 0.1$, $m_1 = 0.3$, $m_2 = 0.1$, $m_3 = 0.7$, and τ is selected such that $A(\tau) = 0, 20,000, 40,000, 60,000, 80,000$.

sales of the generations rather than the cumulative *demand*:

$$a(t) = p(N - A(t)) + (q/N)(N - A(t))[S_1(t) + S_2(t)]. \tag{3.4}$$

Our modification to (3.1) uses the same analogy as in Kumar and Swaminathan (2003) and Ho et al. (2002) that independently propose a supply-constrained Bass model for a single generation. It is also worth noting that our modified model reduces to the original model under unlimited supply, where $S_i(t) = D_i(t)$, and is valid regardless of the rate of loss of waiting customers (l_i).

3.4.2 Optimal sales plan

The supply-constrained multi-generation diffusion problem can be formulated as a continuous-time model within the *optimal control* framework (Sethi and Thompson 2000). The model is used to find the optimal sales plan that maximizes the net present value (NPV) of profit from both generations after the introduction of Product 2 by considering the sales rates ($s_i(t)$) as the *control variables*. To ensure continuously differentiable functions, the time origin is shifted to the launch of the second generation, τ. We consider fixed and separate production capacities c_1 and c_2 for the two generations. A fixed capacity is reasonable as capacity expansions/reductions are generally practically infeasible in short- or medium-term during the diffusion of a new product due to their excessive cost and long lead time. It is also reasonable to assume independent sources of capacity when: (1) the two generations are manufactured by different suppliers; or, (2) the new generation is substantially different and requires different production processes and equipment. The firm manages the production rate by producing at maximum capacity until the demand starts to decline and falls below capacity after which it is set to satisfy the demand (the production management strategy is explained in more detail in the following section). The resulting optimization model is as follows

$$\text{Maximize} \left\{ NPV = \int_0^{+\infty} \left(\sum_{i=1}^{2} \alpha_i(t) s_i(t) - \sum_{i=1}^{2} h_i I_i(t) \right) e^{-\theta t} dt \right\} \tag{3.5}$$

subject to

$$\frac{dA}{dt} = a(t) \tag{3.6}$$

$$\frac{dS_i}{dt} = s_i(t) \qquad \forall i = 1, 2 \tag{3.7}$$

$$\frac{d^2A}{dt^2} = \frac{q}{N} \left(N - A(t)\right) \left(s_1(t) + s_2(t)\right) - a(t) \left(p + \frac{q}{N} \left(S_1(t) + S_2(t)\right)\right) \tag{3.8}$$

$$\frac{dL_i}{dt} = l_i W_i(t) \qquad \forall i = 1, 2 \tag{3.9}$$

$$\frac{dW_i}{dt} = d_i(t) - s_i(t) - l_i W_i(t) \qquad \forall i = 1, 2 \tag{3.10}$$

$$\frac{dI_i}{dt} = r_i(t) - s_i(t) \qquad \forall i = 1, 2 \tag{3.11}$$

$$W_i(t), L_i(t), I_i(t) \geq 0 \qquad \forall i = 1, 2 \tag{3.12}$$

$$W_1(0) = W_1^0, I_1(0) = I_1^0, L_1(0) = L_1^0, W_2(0) = I_2(0) = L_2(0) = 0 \tag{3.13}$$

$$A(0) = A_1, S_1(0) = m_1 A_1, S_2(0) = 0, a(0) = p(N - A_1) + \frac{q}{N}(N - A_1) S_1(0) \tag{3.14}$$

The objective function (3.5) seeks to maximize the discounted profit from both generations after the launch of the second generation by subtracting inventory costs from sales. Constraints (3.6) and (3.7) are self-explanatory. Constraint (3.8) is the time derivative of (3.4). Equation (3.9) describes lost sales where a proportion of l_i of the waiting customers are lost (i.e., waiting customers for product i abandon the product after $\frac{1}{l_i}$ time periods on average). The number of waiting customers at any given time is expressed by (3.10). The change in the inventory level for product i will be a function of its production rate ($r_i(t)$) and sales ($s_i(t)$) as described in (3.11). The non-negativity constraints (3.12) and initial values (3.13 and 3.14) are natural. The following theorem states the optimal control (sales) policy at any given time t (the proof can be found in Appendix A):

Theorem 3.1 *For any profit margin $\alpha_i(t) > 0$, inventory holding cost $h_i > 0$, and launch time of the new generation $\tau \geq 0$, the optimal sales rate at any time t for generation i is given by*

$$s_i^*(t) = \begin{cases} r_i(t) & W_i^*(t) > 0, \\ \min(r_i(t), d_i^*(t)) & W_i^*(t) = I_i^*(t) = 0, \\ d_i^*(t) & I_i^*(t) > 0, \end{cases} \tag{3.15}$$

where W_i^*, I_i^*, and $d_i^*(t)$ are the number of waiting customers, inventory level, and demand rate under optimality. In other words, for both product generations, the optimal sales plan is to sell as many units as possible. Furthermore, $W_i^* I_i^* = 0$ for all $t \geq 0$.

The finding is consistent with the optimal sales plan for a single product as identified by Ho et al. (2002) and Kumar and Swaminathan (2003) and suggests that although immediate demand fulfillment may accelerate the diffusion and lead to increased lost sales due to an earlier and higher peak demand, the benefit from the immediate revenue outweighs the potential lost sales.

3.5 Analytical solutions for the case of patient customers

We use the optimal sales plan to derive closed-form expressions for the demand, sales, and backlogging dynamics as well as the total profit. To make the problem analytically tractable, we assume unconstrained diffusion for the first generation which results in $I_1(t) = W_1(t) = L_1(t) = 0$ for all t. This also allows us to focus on decision-making regarding the optimal entry-time, capacity, and build-up period for the second generation. Even with this assumption, the problem is analytically solvable only for the special case of *patient customers* with no lost sales ($l_2 = 0$). This section provides the closed-form solutions to this special case while the general case of partial demand backlogging is investigated numerically in Section 3.6. For any introduction time (τ), capacity (c_2), and inventory build-up period (T_B) for Product 2, the objective function of the model becomes

$$\text{Maximize}\left\{NPV(\tau, c_2, T_B) = \int_0^{+\infty}\left(\sum_{i=1}^{2}\alpha_i(t)s_i(t) - h_2 I_2(t)\right)e^{-\theta t}dt\right\} \tag{3.16}$$

Depending on the choice of τ, c_2, and T_B, one of the following three cases can occur for the diffusion of the new generation: (1) unconstrained diffusion with sufficient supply (i.e., capacity and/or initial inventory); (2) initially unconstrained diffusion followed by constrained diffusion and then a second unconstrained phase towards the end of the life-cycle; or, (3) initially constrained diffusion followed by an unconstrained phase once production becomes sufficient to satisfy demand.

3.5.1 Case 1: Unconstrained diffusion

Under unconstrained diffusion for both generations, we have $S_1(t) = D_1(t)$ and $S_2(t) = D_2(t)$. By replacing (3.2) and (3.3) in (3.4), the differential equation for the awareness rate is expressed by

$$A'(t) = p\ (N - A(t)) + \frac{q\ (N - A(t))\ (A_1\, u + m\, A(t))}{N}, A(0) = A_1, \tag{3.17}$$

where $m = m_2 + m_3$, $u = m_1 - m_2 - m_3$ and A_1 is the cumulative awareness at $t = 0$ when Product 2 is released. The cumulative awareness during the unconstrained diffusion (UD) phase is given by

$$A^{UD}(t) = \frac{N(\sigma_1 - p) - A_1\, q\, u}{\sigma_1 + m\, q}, \tag{3.18}$$

where $\sigma_1 = \mathrm{e}^{\left(\omega_1 + \frac{t}{N}\right)\chi}, \omega_1 = \dfrac{\ln\!\left(\frac{N\,p + m_1 A_1 q}{N - A_1}\right)}{\chi}$, and $\chi = N\,(p + m\,q) + A_1\,q\,u$.

$a^{UD}(t)$, $D_1^{UD}(t)$, $D_2^{UD}(t)$, $d_1^{UD}(t)$, and $d_2^{UD}(t)$ can be obtained from $A^{UD}(t)$. Let t^+ be the last time that the demand for Product 2 becomes equal to its production capacity, i.e., $t^+ = \max(t|c_2 = d_2(t))$:

$$t^+ = N\left(\frac{\ln\!\left(\frac{\chi(\sigma_2 + \chi)}{2Nc_2} - m\,q\right)}{\chi} - \omega_1\right), \tag{3.19}$$

where $\sigma_2 = \sqrt{m_3\ (\chi^2 - 4Nc_2 mq)}$.

Given zero build-up periods ($T_B = 0$), the minimum capacity required to maintain unconstrained diffusion for Product 2, c_2^{UD}, satisfies $c_2 t^+ = D_2(t^+)$. For $c_2 < c_2^{UD}$, the shortest build-up period required is the solution to $c_2(T_B + t^+) = D_2(t^+)$. Therefore, for any c_2 and A_1, the new generation follows an unconstrained diffusion if and only if T_B exceeds the critical value T_B^{UD} given by

$$T_B^{UD} = \begin{cases} 0, & c_2 \geq c_2^{UD}, \\ \dfrac{(m_3\,\chi + \sigma_2)(N - A_1) - 2\,N\,c_2}{c_2\,(\sigma_2 + \chi)} - t^+, & c_2 < c_2^{UD}. \end{cases} \tag{3.20}$$

As shown in Figure 3.3, this expression provides the critical surface $T_B = T_B^{UD}$ in the (c_2, A_1, T_B) space that separates the unconstrained space ($c_2 \geq c_2^{UD} \vee T_B \geq T_B^{UD}$)

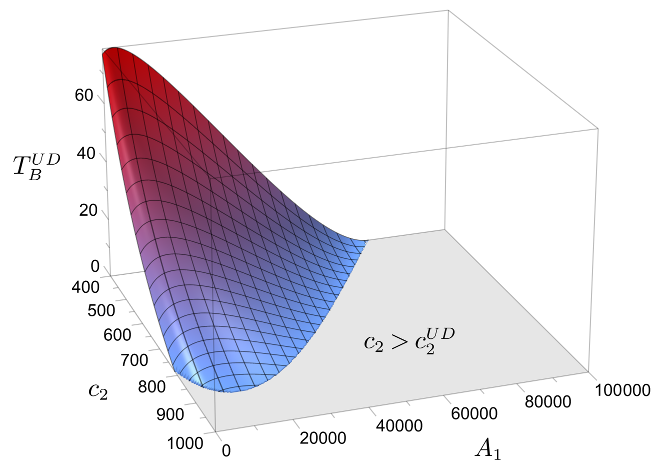


Figure 3.3: Minimum build-up period necessary to maintain unconstrained diffusion for the new generation (T_B^{UD}) as a function of its production capacity (c_2) and awareness level at its launch (A_1). To better illustrate the critical curve of $c_2 = c_2^{UD}$ in the (c_2, A_1) plane, the region $c_2 > c_2^{UD}$ where $T_B^{UD} = 0$ is represented in gray. The parameters are $p = 0.001$, $q = 0.1$, $m_1 = 0.3$, $m_2 = 0.1$, $m_3 = 0.7$, and $N = 100,000$.

from the constrained space ($c_2 < c_2^{UD} \wedge T_B < T_B^{UD}$), spaces above and below the surface, respectively. The figure also illustrates the critical curve of $c_2 = c_2^{UD}$ in the (c_2, A_1) plane that separates the region where a myopic policy will not lead to supply shortages ($c_2 > c_2^{UD}$) from the region where a build-up policy is required to avoid supply shortages ($c_2 < c_2^{UD}$). For any introduction time, T_B^{UD} decreases with higher capacity (similar to the findings of Ho et al. (2002) for a single generation). As an important observation, T_B^{UD} peaks for early introductions of the second generation for any c_2. For simultaneous introduction, in early periods when demand is low, part of production can be used to build inventory requiring less initial inventory to maintain unconstrained diffusion. For an introduction long after Product 1, fewer build-up periods will suffice as most of the market potential is already exhausted by the first generation. We also observe that the initial awareness (A_1) at which T_B^{UD} peaks decreases with c_2. Note that here, for illustrative purposes, we allow the preproduction of Product 2 to begin before the launch of Product 1. In our analysis of the general case of impatient customers in Section 3.6, the technology for generation 2 will not become available before the launch of the first generation.

3.5.2 Case 2: Initially unconstrained diffusion

When the combination of initial inventory and capacity is insufficient for unrestricted diffusion of Product 2 (in the constrained space ($c_2 < c_2^{UD} \wedge T_B < T_B^{UD}$) below the $T_B = T_B^{UD}$ surface in Figure 3.3), there will be an initial phase of unconstrained diffusion (UD1) followed by a constrained phase (CP) that will switch back to unconstrained diffusion (UD2) once the demand falls below capacity. Awareness and demand dynamics during the first phase (UD1) is similar to the Case 1. The ending time of the first unconstrained phase, T_{UD1}, is determined by $c_2(T_B + T_{UD1}) = D_2(T_{UD1})$. When the constrained phase begins at $t = T_{UD1}$, there will be unsatisfied demand ($W_2(t) > 0$) and the instantaneous sales for Product 2 will be bounded by its production capacity ($s_2(t) = c_2$). During CP, $S_2(t) = D_2(T_{UD1}) + c_2(t - T_{UD1})$. By moving the time origin to T_{UD1} and given $A_{CP} = A(T_{UD1})$ (cumulative awareness at the beginning of CP), the information flow during CP is described by

$$A'(t) = (N - A(t))\left(p + \frac{q\,(m_1 A_1 + m_2\,(A(t) - A_1) + m_3\,(A_{CP} - A_1) + c_2\,t)}{N}\right), \quad (3.21)$$

where $A(0) = A_{CP}$ and the cumulative awareness during CP is obtained as

$$A^{CP}(t) = N + \frac{1}{\mathrm{e}^{t\,(p+m_2\,q)+\frac{t\,(2\,A_1\,q\,u+2\,A_{CP}\,m_3\,q+c_2\,q\,t)}{2\,N}}\left(\sigma_3 - \frac{1}{N-A_{CP}}\right)}, \quad (3.22)$$

$$\sigma_3 = \int_0^t \frac{m_2\,q}{N\,\mathrm{e}^{x\,(p+m_2\,q)+\frac{x\,(2\,A_1\,q\,u+2\,A_{CP}\,m_3\,q+c_2\,q\,x)}{2\,N}}}\,\mathrm{d}x. \quad (3.23)$$

The closed-form expression for (3.23) is provided in Appendix B. Given $A^{CP}(t)$, expressions for $a^{CP}(t)$, $D_1^{CP}(t)$, $D_2^{CP}(t)$, $d_1^{CP}(t)$, and $d_2^{CP}(t)$ can be derived. During CP, given $s_2(t) = c_2$ and $l_2 = 0$ (patient customers), the demand backlogging dynamics for Product 2 is described by

$$W_2'(t) = d_2(t) - c_2, W_2(0) = 0, \quad (3.24)$$

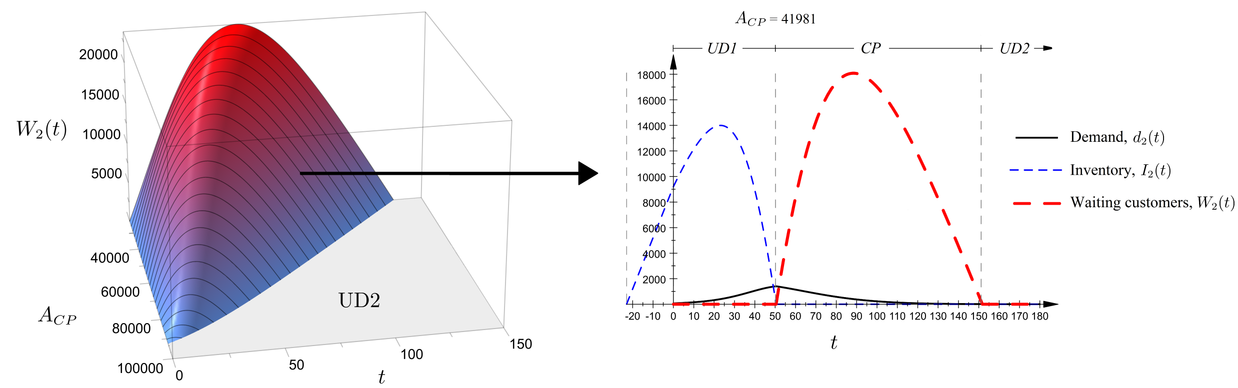


Figure 3.4: Backlogged demand for the new generation ($W_2(t)$) during the constrained phase (CP) for $c_2 = 400$ and simultaneous release of both generations ($A_1 = 0$). $t = 0$ marks the beginning of CP. To better illustrate the length of CP, the second unconstrained diffusion phase (UD2) is represented in gray. With zero initial inventory (i.e., $T_B = 0$), CP begins at $T_{UD1} = 39.6$ with $A(T_{UD1}) = A_{CP} = 22614.6$ (the origin of the A_{CP} axis). A_{CP} increases with T_B which delays CP. The figure on the right pertains to $T_B = 23$. Other parameters are $p = 0.001$, $q = 0.1$, $m_1 = 0.3$, $m_2 = 0.1$, $m_3 = 0.7$, and $N = 100,000$.

and the number of waiting customers is given by $W_2^{CP}(t) = m_3\left(A^{CP}(t) - A_{CP}\right) - c_2 t$. The constrained phase ends at T_{CP} when $W_2^{CP}(t)$ becomes zero for the first time after T_{UD1}, i.e., $T_{CP} = min(t|t > T_{UD1}, W_2^{CP}(t) = 0)$ which can be solved numerically. As shown in Figure 3.4, a longer build-up period for the new generation not only delays the constrained phase but also reduces its length and the peak number of waiting customers.

Let A_{UD2} represent the cumulative awareness at the beginning of the second unconstrained diffusion phase (UD2), i.e., $A_{UD2} = A(T_{CP})$. During this phase, $s_2(t) = d_2(t)$ once again. By moving the time origin to T_{CP}, the information dynamics can be described by

$$A'(t) = p(N - A(t)) + \frac{q(N - A(t))(A_1 u + mA(t))}{N}, A(0) = A_{UD2}. \tag{3.25}$$

The cumulative awareness during UD2 is given by

$$A^{UD2}(t) = -\frac{(N\,p - N\,\sigma_4 + A_1\,q\,u)}{(\sigma_4 + m\,q)}, \tag{3.26}$$

$$\sigma_4 = \mathrm{e}^{\left(\frac{\ln\left(-\frac{(N\,p+A_{UD2}\,m\,q+A_1\,q\,u)}{(A_{UD2}-N)}\right)}{\chi}+\frac{t}{N}\right)\chi}. \tag{3.27}$$

$a^{UD2}(t)$, $D_1^{UD2}(t)$, $D_2^{UD2}(t)$, $d_1^{UD2}(t)$, and $d_2^{UD2}(t)$ can be obtained. This concludes Case 2.

3.5.3 Case 3: Initially constrained diffusion

When the initial inventory and capacity for Product 2 is insufficient from the beginning of its diffusion, i.e., $c_2T_B + c_2 < m_3a(\tau)$, there will be an initial constrained phase during which $W_2(t) > 0$ followed by an unconstrained diffusion phase once the number of waiting customers becomes zero. The awareness, demand, and sales dynamics in this case is similar to the second and third phases of the initially unconstrained case discussed above (i.e., CP and UD2) with $A_{CP} = A_1$ and $T_{UD1} = 0$.

3.5.4 Total profit from both product generations

We now turn to the characterization of the net present value of total profit from both generations at time zero (launch of the first generation). Given a production capacity c_2, length of the build-up period T_B, and release time of the second generation τ, the life-cycle profit after the introduction of the second product described in (3.16) is given by the sum of the discounted profit during UD1, CP, and UD2 phases (note that not all three phases may exist depending on which of the above three cases occurs). Given $t_1 = t-\tau$, $t_2 = t-\tau-T_{UD1}$, and $t_3 = t - \tau - T_{UD1} - T_{CP}$, we have

$$\begin{aligned} NPV^{\tau^+}(\tau, c_2, T_B) = &\int_{\tau}^{\tau+T_{UD1}} \left(\sum_{i=1}^{2} \alpha_i(t_1)d_i^{UD1}(t_1) - \sum_{i=1}^{2} h_2(c_2(T_B+t_1) - D_2^{UD1}(t_1))\right) e^{-\theta t}dt \\ &+ \int_{\tau+T_{UD1}}^{\tau+T_{UD1}+T_{CP}} \left(\sum_{i=1}^{2}(\alpha_1(t_2)d_1^{CP}(t_2) + \alpha_2(t_2)c_2)\right) e^{-\theta t}dt \\ &+ \int_{\tau+T_{UD1}+T_{CP}}^{+\infty} \left(\sum_{i=1}^{2} \alpha_i(t_3)d_i^{UD2}(t_3)\right) e^{-\theta t}dt. \end{aligned} \tag{3.28}$$

The NPV of prelaunch inventory cost during the build-up period can be expressed as

$$NPV_B(\tau, c_2, T_B) = \int_{\tau - T_B}^{\tau} h_2\, c_2\, t_4\, e^{-\theta t} dt = \frac{h_2\, c_2\, \left(\mathrm{e}^{\theta T_B} - \theta T_B - 1\right)}{\theta^2\, \mathrm{e}^{\theta \tau}}, \tag{3.29}$$

where $t_4 = t - \tau + T_B$. Next, we calculate the NPV of profit from generation 1 before the launch of Product 2, $NPV^{\tau^-}(\tau)$. The following differential equation describes the information diffusion dynamics before the introduction of the second generation

$$A'(t) = p\, (N - A(t)) + \frac{m_1\, q\, A(t)\, (N - A(t))}{N}, A(0) = 0. \tag{3.30}$$

The cumulative awareness and instantaneous awareness rate before Product 2 are given by

$$A^{\tau^-}(t) = \frac{N\, (\sigma_5 - p)}{\sigma_5 + m_1\, q}, \tag{3.31}$$

$$a^{\tau^-}(t) = \frac{N\, \sigma_5\, (p + m_1\, q)}{\sigma_5 + m_1\, q} + \frac{N\, \sigma_5\, (p + m_1\, q)\, (p - \sigma_5)}{(\sigma_5 + m_1\, q)^2}, \tag{3.32}$$

where $\sigma_5 = \mathrm{e}^{N\,(p + m_1\, q)\, \left(\frac{t}{N} + \frac{\ln(p)}{N\,(p + m_1\, q)}\right)}$.

Given $d_1^{\tau^-}(t) = m_1 a^{\tau^-}(t)$, the NPV of profit generated by Product 1 before Product 2 will be

$$NPV_1^{\tau^-}(\tau) = \int_0^{\tau} \alpha_1(t) d_1^{\tau^-}(t) e^{-\theta t} dt. \tag{3.33}$$

Let H_i be the cost of unit of capacity for product i, the NPV of total capacity investment is

$$NPV^c(\tau, c_2, T_B) = H_1 c_1 + H_2 c_2 (1 + \beta)^{-(\tau - T_B)}, \tag{3.34}$$

where $c_1 = \max_{t \geq 0}(d_1(t))$ and $\beta = e^{\theta} - 1$. The total profit from both generations will be

$$NPV^{Total}(\tau, c_2, T_B) = NPV^{\tau^+}(\tau, c_2, T_B) - NPV_B(\tau, c_2, T_B) + NPV_1^{\tau^-}(\tau) - NPV^c(\tau, c_2, T_B). \tag{3.35}$$

Given the above, we can now formalize the hierarchy of decisions that the firm needs to make. For a given introduction time τ and capacity c_2 for the second generation, the optimal build-up period T_B^* can be determined through a one-dimensional search as follows

$$Profit^*_{\tau^+}(\tau, c_2) = \max_{0 \leq T_B \leq \tau} \left(NPV^{\tau^+}(\tau, c_2, T_B) - NPV_B(\tau, c_2, T_B) \right), \tag{3.36}$$

where $0 \leq T_B \leq \tau$ will ensure the availability of the technology for generation 2 after the launch of generation 1. Next, the optimal capacity is determined by a one-dimensional search as follows

$$c_2^* = \max_{c_2} = \left(Profit^*_{\tau^+}(\tau, c_2) - NPV^c(\tau, c_2, T_B^*) \right). \tag{3.37}$$

Finally, the optimal entry time for the new generation of the product may be found by

$$\tau^* = \max_{\tau} = NPV^{Total}(\tau, c_2^*, T_B^*). \tag{3.38}$$

Steps (3.28)-(3.38) also apply to the case of impatient customers. However, the closed-form solution to the differential equation in (3.10) is not available when $l_2 \neq 0$. Therefore, we leave the analysis of the optimal decisions to the next section where the general case of the problem is solved numerically.

3.6 The general case of impatient customers: A numerical study

In this section, we numerically solve a discrete-time version of the model and outline the most important findings for the general case of impatient consumers.

3.6.1 Experimental design

The experimental design is summarized in Table 3.2. The values/ranges for the parameters are selected based on numerical studies by Wilson and Norton (1989) and Kumar and Swaminathan (2003) to provide a common ground for comparison. We consider equal costs of production capacity ($H_1 = H_2 = H$) and fixed profit margins (i.e., $\alpha_i(t) = \alpha_i$). We use the proportion of informed consumers at the launch of the second generation ($\psi(\tau)$) instead of the introduction time (τ) in our analysis to reduce the sensitivity of the results to the absolute market size N making them generalizable to different market sizes. Similar to Wilson and Norton (1989), we consider a market space (M) frequently found in high-technology, publishing, and pharmaceutical industries as follows

$$M = \{(\alpha_1, \alpha_2, m_1, m_3, m_2) | \alpha_1 \geq \alpha_2, 0 < m_1 \leq 1, 0 \leq m_2 \leq m_1, m_1 \leq m_2 + m_3 \leq 1\}. \quad (3.39)$$

Similar to Wilson and Norton (1989), we use the ratio of the adoption rate of Product 1 after and before Product 2 ($\frac{m_2}{m_1}$) as a measure of *cannibalization* of generation 1 by the new generation with $\frac{m_2}{m_1} = 1$ suggesting no cannibalization and $\frac{m_2}{m_1} = 0$ full cannibalization. Also, $\frac{\alpha_2 m_3}{\alpha_1 m_1}$ is used to measure relative *profitability* of the two product generations. In this marketplace, the first generation makes a higher unit contribution to profit but has a smaller market penetration while the second generation has a higher market potential and a lower profit margin. The resulting trade-off is between making high-margin sales slowly (through the first generation only) or cannibalize part of the sales from Product 1 to make low-margin sales more rapidly by introducing the new generation. The experimental design results in more than 15.2 million different parameter configurations.

3.6.2 Optimal entry time

Figure 3.5 provides the optimal introduction time under various market settings. In the plot for each combination of production capacity for the second generation (c_2) and cost per

Table 3.2: Parameter choices for numerical experiments

Parameter	Description	Value/Range
p	Coefficient of innovation	0.001, 0.002, 0.003, 0.004, 0.005
q	Coefficient of imitation	0.1, 0.02, 0.3, 0.4, 0.5
N	Market size	100,000
c_2	Production capacity for Generation 2	50, 100, 150, 200, 300, 400, 500, 600
H	Cost per unit of capacity	0-40 in increments of 10
$\psi(\tau)$	proportion of informed consumers at τ	0-1 in increments of 0.05
T_B	Build-up period for Generation 2	0-τ
l	Rate of loss of waiting customers	0, 0.01, 0.05, 0.1, 0.5, 1
θ	Discount factor	0.003
$\frac{m_2}{m_1}$	Cannibalization level	0-1 in increments of 0.1
$\frac{\alpha_2 m_3}{\alpha_1 m_1}$	Relative profitability	0-1 in increments of 0.1
h_2	Unit holding cost for Generation 2	$0.01\alpha_2$

unit of capacity (H), the x-axis represents the relative profitability of the two generations ($\frac{\alpha_2 m_3}{\alpha_1 m_1}$) and the y-axis is the cannibalization level ($\frac{m_2}{m_1}$). For each plot, the plane is divided into an 11×11 lattice and the optimal entry time $\psi(\tau)$ is represented by the darkness of each lattice point with darker shades indicating a later release. Therefore, the white color suggests the optimality of simultaneous release of both generations (*Now* policy) and black suggests it is optimal not to introduce the second generation at all (*Never* policy).

Perhaps the most important finding is that the optimal introduction time can vary continuously from *now* to *never* which generalizes the previously proposed *Now or Never* and *Now or Maturity* rules. We never introduce the second generation when the cannibalization effect is high and it has low profitability compared to the first generation. Under high cannibalization, we introduce the second generation if it is *sufficiently* profitable to outweigh the reduction in the high-margin sales of the first generation. When the relative profitability is small, we introduce the new generation only if it does not cannibalize first generation too much. Therefore, for most of the grid points below the diagonal, the *never* rule is optimal.

When capacity is cheap, consider $H = 0$ or 10, the size of the *now* region increases with capacity c_2. For $H = 30$ or 40, we observe a general reduction in the size of the *now* region as capacity increases (we see a mixed effect for intermediate capacity levels, e.g., $H = 20$). For

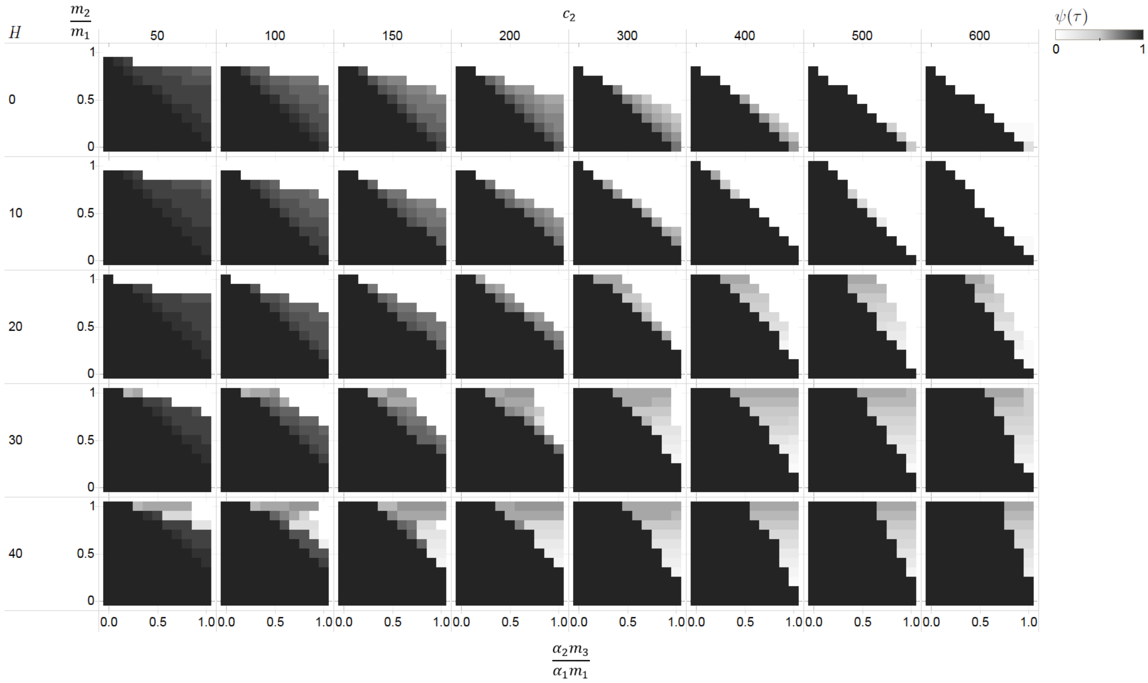


Figure 3.5: Optimal entry time for the case of impatient consumers ($l = 0.1$) broken down by production capacity for Product 2 c_2 and the cost per unit of capacity H.

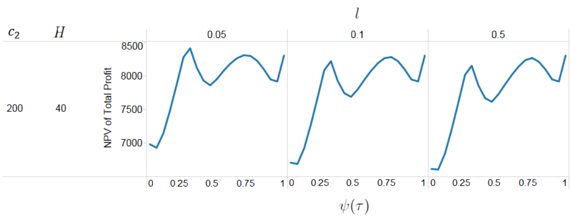
any c_2, there is a general increase in the size of the *never* region and decrease in size of the *now* region with H. Similar patterns are observed for other levels of lost sales (i.e., l). These findings can be explained by the trade-off between capacity investment, delayed sales of the new generation, and lost sales. When capacity is expensive, it is beneficial to delay capacity investment even though it would also delay revenue collection from the sales of Product 2. When capacity is cheap, earlier sales of the second generation justifies an earlier investment in capacity. On the other hand, an early launch accelerates the information and demand dynamics and potentially leads to higher lost sales. A high capacity level is expected to reduce lost sales making an early introduction more preferable only if capacity is cheap. As capacity becomes more expensive, a later introduction is beneficial.

Figure 3.6 presents a set of representative market entry policies. In each plot, we systematically change the rate of loss (l) and observe the resulting NPV of total profit as a function of the proportion of informed consumers at the release of the second generation

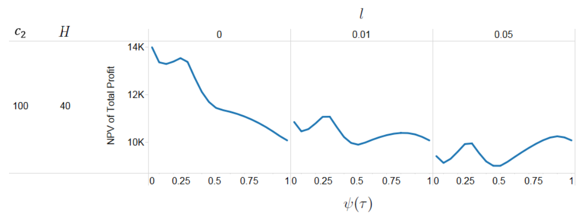
$(\psi(\tau))$. We use the following conditions to define introduction policies. The optimal policy is *Now* if $\psi(\tau^*) = 0$, *Early* if $0 < \psi(\tau^*) \leq 0.3$, *Middle* if $0.3 < \psi(\tau^*) \leq 0.6$, *Maturity* if $0.6 < \psi(\tau^*) < 1$, and *Never* if $\psi(\tau^*) = 1$ (i.e., the market potential is fully exhausted by the first generation). The choice of the above boundaries is somewhat arbitrary and different ranges can be used depending on the specific product under consideration. However, it is important to note that using $\psi(\tau)$ allows the decision-maker to define these regions regardless of the duration of the diffusion for the two product generations (unlike Mahajan and Muller (1996) that fail to clearly characterize the *Now or Maturity* policy as their analysis focuses on τ). An important observation is the formation of a local maximum in the *middle*, *maturity*, or *never* regions as l increases which can be explained by the trade-off between lost sales and the delay in capacity investment and sales of the new generation as discussed above.

3.6.3 Optimal build-up policy and capacity

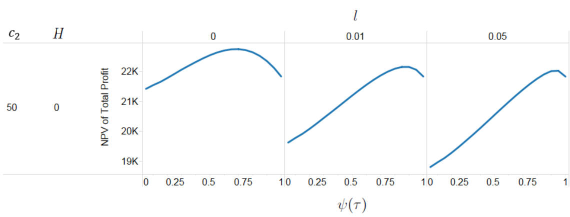
As depicted in Figure 3.7, for any unit capacity cost (H), the optimal build-up period (T_B^*) becomes shorter with more capacity c_2. Since the technology for Product 2 is not available before the launch of Product 1, $t = 0$ is the earliest time that inventory build-up may begin (i.e., $0 \leq T_B^* \leq \tau$). For small capacity levels, T_B^* increases with $\psi(\tau)$ as longer build-up periods become permissible. When capacity is free, T_B^* peaks for introduction in the *early* region which results in a fast growth and high peak demand. Moreover, the company would not have as much time to build inventory after the product's launch compared to the case of simultaneous introduction. T_B^* starts to decline for later introductions as a larger proportion of the market would be already exhausted by Product 1 reducing the peak demand for Product 2. As H increases, an interesting pattern is observed in the *early* region where T_B^* decreases mainly because additional sales through higher preproduction does not justify earlier capacity investment and delaying preproduction becomes beneficial. This effect is augmented for higher c_2 and H. The general decrease in T_B^* as capacity becomes



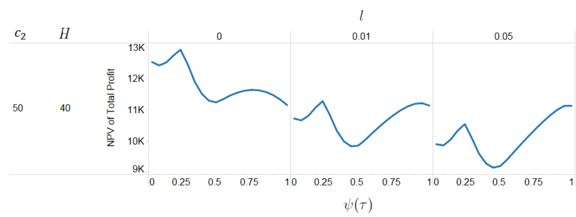
(a) *Early-Maturity-Never* policy.



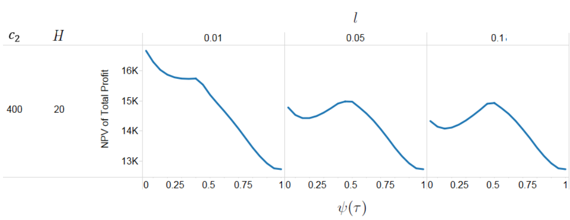
(b) *Now-Early-Maturity* policy.



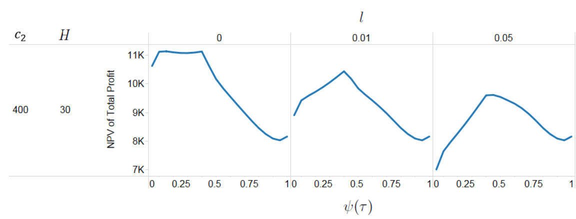
(c) *Maturity* policy.



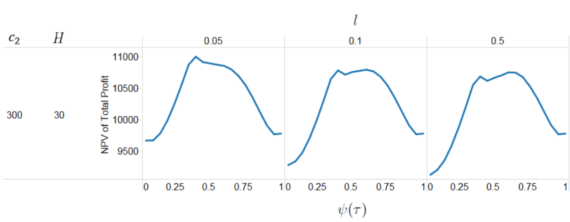
(d) *Early-Maturity* policy.



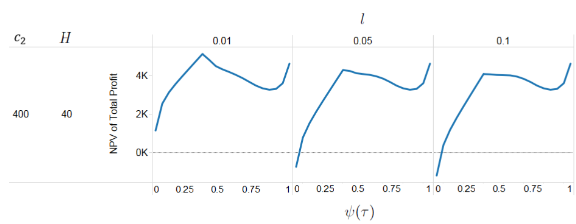
(e) *Now-Middle* policy.



(f) *Early-Middle* policy.



(g) *Middle-Maturity* policy.



(h) *Middle-Never* policy.

Figure 3.6: Selected market entry policies ($\frac{\alpha_2 m_3}{\alpha_1 m_1} = 0.7$, $\frac{m_2}{m_1} = 0.7$)

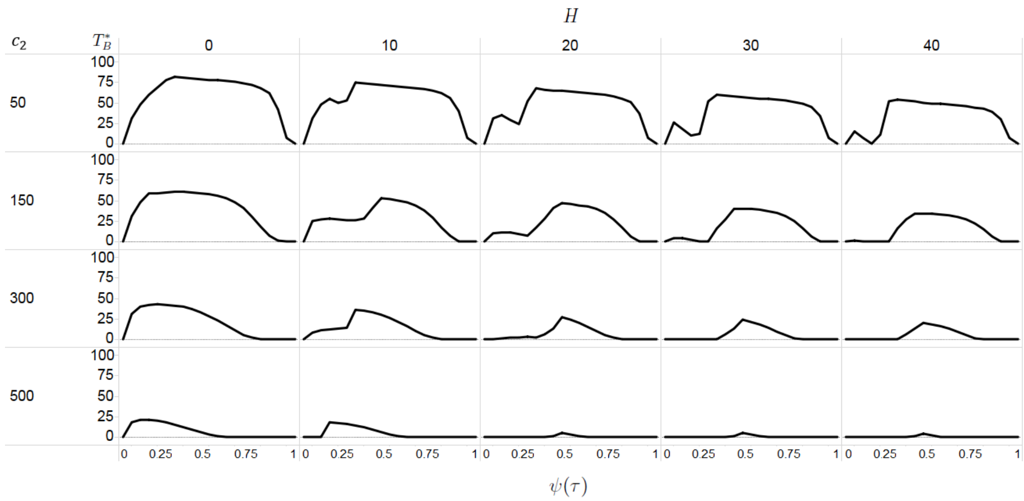


Figure 3.7: Optimal build-up policy for the second generation for different proportions of informed consumers at its launch ($\psi(\tau)$) broken down by its capacity (c_2) and unit capacity cost (H). The other parameters are $l = 0.05$, $\frac{\alpha_2 m_3}{\alpha_1 m_1} = 0.7$, and $\frac{m_2}{m_1} = 0.7$.

more expensive is due to the same reason. Similar patterns are observed for other market configurations.

Given T_B^*, the optimal production capacity is determined through a one-dimensional search as in (3.37). According to Figure 3.8, when capacity is free ($H = 0$), the total profit is an increasing function of c_2 as more capacity would reduce the cost of carrying build-up inventory. As H increases, large capacity levels become less preferable while the optimal capacity will depend on the remaining market potential. When capacity is too expensive, a small capacity will become preferable regardless of the introduction time although it leads to longer build-up periods and higher inventory costs.

3.7 Empirical evidence: PlayStation®3 game console

Until this date, Sony has launched four generations of its popular game console. The original PlayStation (PS), launched in 1994, was a huge success with sales exceeding 40 million units worldwide in just two years. PS2, launched in 2000, became the fastest game console to reach 100 million shipments worldwide in 5 years and 9 months. The most recent

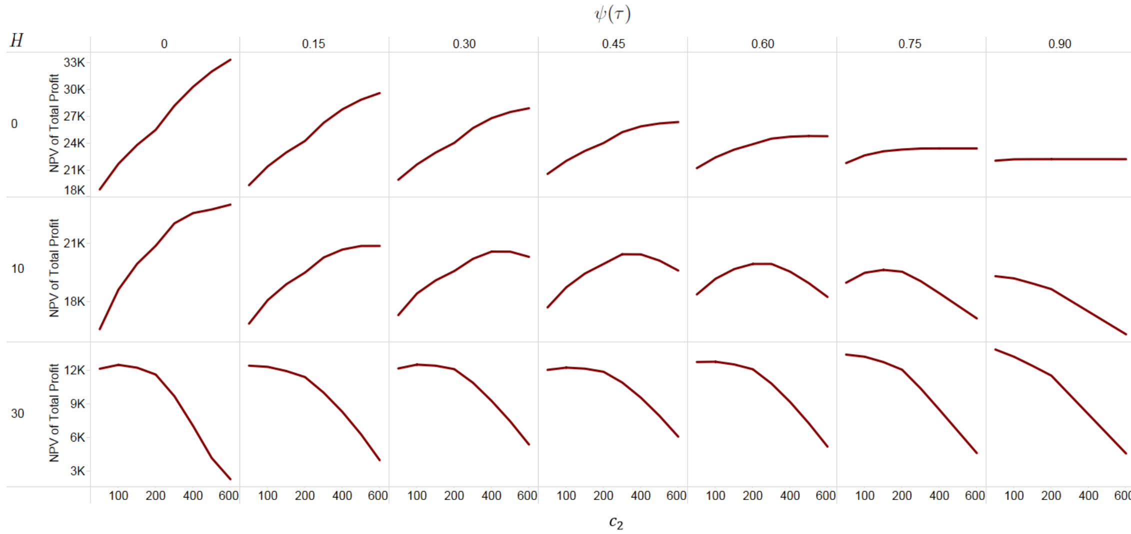


Figure 3.8: NPV of total profit for different capacity levels (c_2) broken down by unit capacity cost (H) and awareness at the launch of the second generation ($\psi(\tau)$). Other parameters are $l = 0.05$, $\frac{\alpha_2 m_3}{\alpha_1 m_1} = 0.7$, and $\frac{m_2}{m_1} = 0.7$.

generation, PS4, became the fastest selling PlayStation with sales exceeding 4.2 million worldwide in 3 months. However, the case of PS3 is of particular interest as it was not as successful as the other three generations.

In March 2006, Sony announced the first delay in the launch of PS3, originally scheduled for Spring 2006, until November saying that the console would be released simultaneously in Japan, America, and Europe for the first time (both PS1 and PS2 were first released in the Japan and America with Europe coming a belated third). The head of Sony Computer Entertainment Europe (SCEE) said at the time: "This is an exciting first for Europe, and is a huge endorsement and vote of confidence in the strength of the European market and its importance globally." However, the company postponed the launch in Europe again from November 2006 to March 2007. Both delays were due to problems with the production of the console's Blu-ray disc drive (*BBC News*, September 6, 2006). Besides the long delays, Sony was selling the console at a strategic price much lower than its production cost during the introductory period of the platform (a negative margin of about \$200 per console) hoping to recover the loss from the revenue from games/software. However, the console's high price

($599 compared to $399 for Microsoft's Xbox 360 and $249 for Nintendo Wii) was a barrier to its widespread penetration. There was a significant drop in the sales of the console (82%) and popular games (almost 60%) in Europe on the second week after the launch leading to additional loss due to unsold inventory. We did not find any evidence about significant supply challenges for PS3 after its launch in Europe which can be another sign of overproduction. Sony reported a loss of $1.8B within the Game segment for the fiscal year ended March 31, 2007. In April 2007, Ken Kutaragi, President of Sony Computer Entertainment and the lead architect for PlayStation, announced his retirement saying that he had been planning his retirement for six months while many news agengies attributed this to poor sales (*The Wall Street Journal*, April 26, 2007 and *Times Online*, April 27, 2007). Despite the significant reduction in PS3 production cost from $800 to $450 in 2008, the company still reported a $1.2B and $0.6B loss in its Game division for the fiscal years ended March 31, 2008 and 2009, respectively, mainly due to slow sales for PS3.

Motivated by the above, we explore the optimal entry time and build-up policy for PS3 in Europe. Our discussion highlights two issues that had a negative impact on PS3's performance: (1) late market entry; and, (2) demand overestimation leading to excessive inventory. Our goal is not to provide an exact replication of the real case, but rather show that the proposed model is capable of providing insights about the optimal introduction time and build-up policy for PS3 that could potentially help Sony make more informed decisions and avoid/reduce the incurred financial loss.

The underlying assumption of our model is that product generations are introduced in a monopolistic market while in many real-world markets (including the case of PS3), the optimal timing may also be function of competition (e.g., the launch of Xbox 360 or Nintendo Wii). We justify this assumption by the fact that PlayStation is historically strong in Europe. In fact, Sony dominated the market (in most European countries) during the years included in our analysis. In the UK, for instance, one in four households owned a PS at the launch of PS2 (*Reuters*, November 27, 2000) and PS2 gained 80% of the market share.

PS2, the best selling game console of all history, gave Sony 70% of the market share (*The Economist*, November 16, 2006). Even PS3, despite entering the market more than a year after Microsoft's Xbox 360, had a record-breaking launch and outsold its closest competitor in Europe starting from October 2007 (*EuroGamer.net*, May 6, 2008) and was the best-selling console in Europe in 2012. It is worth noting that PS4 (although not considered in our analysis) has also maintained the market leadership with a 70% to 90% market share in every country in Europe (*EuroGamer.net*, July 1, 2015). As in our numerical study, we also assume unconstrained diffusion for PS and PS2 to be able to focus on supply-related decisions for PS3.

3.7.1 Model parameters

Parameter choices for the case study are provided in Table 3.3. The market potential was estimated to be the 216 million households in Europe (Eurostat 2015). Sales (units) are obtained from Sony's financial statements. The initial inventory and the length of preproduction were obtained from news releases from which the production capacity was estimated. We ran the model with PS3's actual release time and build-up period and used the nonlinear regression procedure of the SAS statistical package to estimate the diffusion and cannibalization parameters. We performed this analysis multiple times with different number of historical data points. The model's predictions are better when based on 12 years of data yielding an $R^2 = 0.85$ and, more importantly, accurate predictions of the demand peaks and general pattern (Figure 3.9). The model did not perform well when fitted to the demand for all four generations. We believe this is mainly due to the underlying assumption of the Wilson-Norton model that the market potential is constant. As the number of generations and/or the planning horizon increases, the accuracy of the model tends to decrease as the market size may change.

The discount factor is estimated from Sony's annual reports (the average value between 1999 and 2015 was used). The profit margin includes the profit from games. The average

Table 3.3: Parameter estimates for the empirical study

Parameter	Description	Estimated value/range
N	Market size	216 million
p	Coefficient of innovation	0.019
q	Coefficient of imitation	0.88
m_{11}	Fraction that purchase PS before PS2	0.59
m_{12}	Fraction that purchase PS after PS2	0.02
m_{21}	Fraction that purchase PS2 before PS3	0.35
m_{22}	Fraction that purchase PS2 after PS3	0.24
m_{31}	Fraction that purchase PS3 before PS4	0.26
c_3	Production capacity for PS3	318,000 units per month
β	Discount factor	5.6%
α_2	Profit margin per PS2 console	$300
α_3	Profit margin per PS3 console	$163
C	R&D and capacity costs	$0.4-$1.0 billion in 0.1 increments
h	Holding cost per console per year	$10, $20, $30, $40, $50
l	Rate of loss of waiting customers	0, 0.05, 0.1, 0.15, 0.2, 0.5, 1
τ_3	Introduction time of PS3 (fiscal year)	2003, 2004, 2005, 2006
T_B	Build-up period for PS3	0-24 months

number of games sold per console is 9.23, 10.54, and 10.71 for PS, PS2, and PS3, respectively. We consider a console owner fee of 11.5%, that is $6.9 profit on a $60 game (*Forbes*, December 19, 2006). We used the average unit production cost as it changed over the years. For instance, the production cost of a PS3 console was around $800 at launch and was reduced to $450 after two years and dropped to $240 five years later. The average production cost per PS3 console was estimated by the weighted average. Similarly, we calculated the average

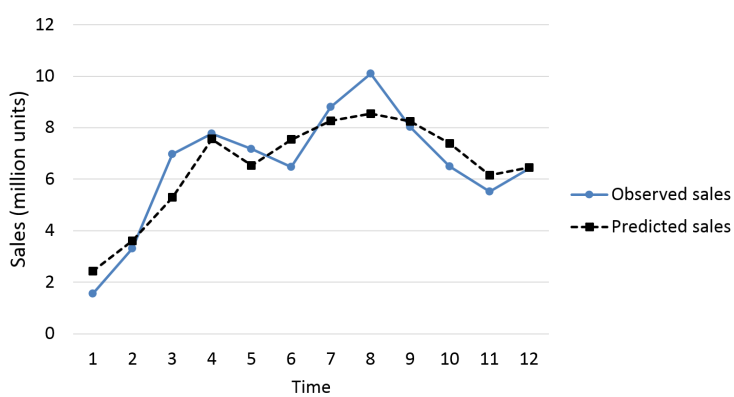
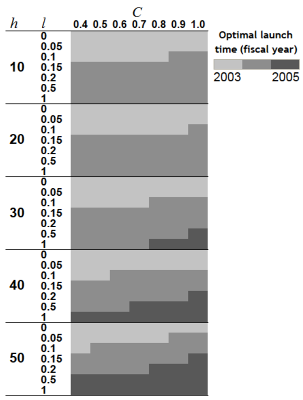


Figure 3.9: Model predictions vs the actual sales of PlayStation console generations

price of each console over their life-cycle. The average profit margin per console was then estimated from the above. We estimated the possible range for the investment in Research and Development (R&D) and production capacity from Sony's annual reports. However, we were not able to find any information about the inventory cost per console and the rate of loss of waiting customers and thus analyzed different levels of these parameters. We then varied the introduction time of PS3 from 2003 to 2006 (the actual launch time) and the build-up period from 0 to 24 months to find their optimal value.

3.7.2 PS3: Analysis of build-up period and introduction time decisions

Figure 3.10(a) suggests that the optimal introduction time varies between 2003 and 2005 fiscal years indicating the late introduction of PS3 in Europe as supported by the empirical evidence (i.e, long delays). While a later launch would be preferable for higher initial investment and/or rate of loss of waiting customers, a 2006 launch is found to be late even for the highest levels of these parameters in our analysis. Figure 3.10(b) shows that for the 2006 fiscal year (actual release time), the optimal build-up period is either 0 or 1 month. Sony built one million consoles for PS3's launch in Europe suggesting 3 months of preproduction (*Financial Times*, March 28, 2007). Therefore, our model was able to capture both identified problems. The results suggest that Sony would need longer build-up periods with earlier launch which results in a faster growth and higher peak demand requiring more supply to avoid lost sales. In general, as the cost of R&D and capacity increases, additional sales through higher preproduction does not justify an earlier investment and the company would benefit from shorter build-up periods. Moreover, as consumers become more impatient, supply-shortages become more costly and thus a longer preproduction would be preferable.



(a) Optimal launch time.

Unit Inventory Cost ($) / Rate of loss of waiting customers

Launch Time	Investment (Billion $)	10							20							30							40							50						
		0	0.05	0.1	0.15	0.2	0.5	1	0	0.05	0.1	0.15	0.2	0.5	1	0	0.05	0.1	0.15	0.2	0.5	1	0	0.05	0.1	0.15	0.2	0.5	1	0	0.05	0.1	0.15	0.2	0.5	1
2003	0.4	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	9	10	10	12	12	12	12	8	10	10	10	10	12	12
	0.5	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	9	10	10	12	12	12	12	0	10	10	10	10	12	12
	0.6	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	9	10	10	11	12	12	12	0	10	10	10	10	12	12
	0.7	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	9	10	10	11	12	12	12	0	10	10	10	10	12	12
	0.8	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	8	10	10	11	12	12	12	0	10	10	10	10	12	12
	0.9	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	8	10	10	11	12	12	12	0	10	10	10	10	12	12
	1.0	12	12	12	12	12	12	12	10	12	12	12	12	12	12	10	10	12	12	12	12	12	8	10	10	11	12	12	12	0	10	10	10	10	12	12
2004	0.4	8	12	12	12	12	12	12	7	8	11	12	12	12	12	6	8	8	11	11	12	12	0	7	8	8	8	12	12	0	7	7	8	8	12	12
	0.5	8	12	12	12	12	12	12	7	8	11	12	12	12	12	6	8	8	10	11	12	12	0	7	8	8	8	12	12	0	5	7	8	8	12	12
	0.6	8	12	12	12	12	12	12	7	8	11	12	12	12	12	6	8	8	10	11	12	12	0	7	8	8	8	12	12	0	5	7	8	8	12	12
	0.7	8	12	12	12	12	12	12	7	8	11	12	12	12	12	5	7	8	10	11	12	12	0	7	8	8	8	12	12	0	4	7	8	8	12	12
	0.8	8	12	12	12	12	12	12	7	8	11	12	12	12	12	1	7	8	10	10	12	12	0	7	8	8	8	12	12	0	4	7	8	8	12	12
	0.9	8	11	12	12	12	12	12	7	8	11	12	12	12	12	0	7	8	8	10	12	12	0	7	8	8	8	12	12	0	4	7	8	8	12	12
	1.0	8	11	12	12	12	12	12	6	8	11	11	12	12	12	0	7	8	8	10	12	12	0	7	7	8	8	12	12	0	4	7	7	8	12	12
2005	0.4	4	4	5	5	5	5	6	2	4	4	4	5	5	6	0	3	4	4	4	5	5	0	0	3	4	4	5	5	0	0	2	4	4	5	5
	0.5	4	4	5	5	5	5	6	0	4	4	4	5	5	5	0	3	4	4	4	5	5	0	0	3	4	4	5	5	0	0	2	3	4	5	5
	0.6	3	4	5	5	5	5	6	0	4	4	4	5	5	5	0	3	4	4	4	5	5	0	0	3	4	4	5	5	0	0	2	2	4	5	5
	0.7	3	4	5	5	5	5	6	0	4	4	4	5	5	5	0	2	4	4	4	5	5	0	0	3	4	4	5	5	0	0	2	2	4	5	5
	0.8	3	4	5	5	5	5	6	0	3	4	4	5	5	5	0	2	4	4	4	5	5	0	0	2	4	4	5	5	0	0	0	2	4	4	5
	0.9	3	4	4	5	5	5	6	0	3	4	4	5	5	5	0	2	4	4	4	5	5	0	0	2	4	4	5	5	0	0	0	2	4	4	5
	1.0	3	4	4	5	5	5	6	0	3	4	4	4	5	5	0	0	3	4	4	5	5	0	0	2	4	4	5	5	0	0	0	2	4	4	5
2006	0.4	0	0	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	0.5	0	0	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	0.6	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	0.7	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	0.8	0	0	0	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	0.9	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
	1.0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1

(b) Optimal build-up period.

Figure 3.10: Optimal build-up period (months) and introduction time (fiscal year) for PS3 broken down by initial investment (C), unit inventory cost (h), and rate of loss of waiting customers (l).

3.8 Discussion and conclusions

We investigate the optimal market entry and production-sales policy for successive generations of innovations. We modify a classical diffusion model and propose a supply-restricted model while retaining a parsimonious analytical representation. A mathematical model is then developed with the objective to maximize the total discounted profit that accounts for the simultaneous effects of inventory and production costs and lost sales. We establish the optimal sales plan and derive the closed-form expressions for demand, sales, and backlogging dynamics for the special case of patient customers. A numerical study is carried out for the general case of impatient customers. The model supports the hierarchy of inter-dependent decisions that a firm needs to make when introducing a new generation by answering the three research questions posed in Section 3.2:

1. *Optimal market entry policy:* We generalize the previously proposed *now or never* and *now or maturity* policies by showing that the optimal launch time for the second generation varies continuously from *now* to *never* depending on the cannibalization level and the generations' relative profitability. The consideration of supply constraints,

capacity and inventory costs, and lost sales adds realism to the analysis and yields policies that are more consistent with industry practices.

2. *Optimal production capacity and build-up policy:* The build-up inventory can be used as a substitute for production capacity and vice versa. We show that the optimal combination of the two is determined by the trade-off between the capacity investment and inventory cost, and is also affected by cannibalization, the generations' relative profitability, lost sales, and market entry time.

3. *Optimal sales plan:* We show that the total life-cycle profit after the launch of the second generation is maximized if the firm sells as many units as possible even though this might accelerate the demand growth rate and lead to higher lost sales.

Our empirical study suggests that Sony introduced PS3 too late and overestimated the demand which resulted in unsold inventory as supported by empirical evidence. However, our model ignores competition by assuming a monopolistic market. We believe this limitation does not undermine our analysis as the pressure from competitors makes an earlier introduction even more favorable strengthening our finding about the late introduction of PS3. The marketing literature strongly supports that market-share rewards to pioneers. Shaping consumers preferences and establishing consumer loyalty, avoiding consumer switching cost compensations, performance advantages from early sales, establishing and maintaining standards, and preempting preferred patents and suppliers are among other advantages for pioneers (Hauser et al. 2006). Therefore, we can conclude that a late market entry as a monopolist will also be late in an oligopolistic market. Another important assumption of our application is that the demand is distributed uniformly throughout the year (i.e., constant daily/monthly demand rate). For most game consoles, sales are much higher over the first few days after launch and before the holiday season (e.g., PS3 sold 600,000 units in Europe in just 2 days after its launch). While the optimal build-up period suggested by our model ignores this *seasonality*, it still provides helpful insights for managers under high seasonal demand variations.

In real-world applications, the profit margin can be obtained internally while historical data (either from the first generation or similar products in the past) can help estimate the coefficients of innovation and imitation, market size, and rate of loss. An analysis of the technological enhancements in the new generation and their impact on consumer buying motives would be necessary to estimate the market expansion and cannibalization. Finally, due to demand and supply uncertainties, we suggest sensitivity analysis on different model parameters and developing expectations for possible outcomes including the best and worst cases to select a robust policy. We believe these considerations in the application of the model would help companies make more informed decisions.

Bibliography

Amini, M., H. Li. 2011. Supply chain configuration for diffusion of new products: An integrated optimization approach. *Omega* **39**(3) 313–322.

Amini, M., T. Wakolbinger, M. Racer, M. G. Nejad. 2012. Alternative supply chain production-sales policies for new product diffusion: An agent-based modeling and simulation approach. *European Journal of Operational Research* **216**(2) 301–311.

Bass, F. M. 1969. A new product growth for model consumer durables. *Management Science* **15**(5) 215–227.

Bass, F. M. 2004. Comments on "A new product growth for model consumer durables". *Management Science* **50**(12) 1833–1840.

Bilginer, O., F. Erhun. 2010. *Managing Product Introductions and Transitions*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Bilginer, O., F. Erhun. 2015. Production and sales planning in capacitated new product introductions. *Production and Operations Management* **24**(1) 42–53.

Cantamessa, M., C. Valentini. 2000. Planning and managing manufacturing capacity when demand is subject to diffusion effects. *International Journal of Production Economics* **66**(3) 227–240.

Eurostat, E. C. 2015. Private households by household composition. `http://ec.europa.eu/eurostat/web/products-datasets/-/lfst_hhnhwhtc`. Accessed: November 5, 2015.

Hauser, J., G. J. Tellis, A. Griffin. 2006. Research on innovation: A review and agenda for marketing science. *Marketing Science* **25**(6) 687–717.

Ho, T.-H., S. Savin, C. Terwiesch. 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science* **48**(2) 187–206.

Ke, T., Z.-J. Shen, S. Li. 2013. How inventory cost influences introduction timing of product line extensions. *Production and Operations Management* **22**(5) 1214–1231.

Kim, N., D. R. Chang, A. D. Shocker. 2000. Modeling intercategory and generational dynamics for a growing information technology industry. *Management Science* **46**(4) 496–512.

Kumar, S., J. M. Swaminathan. 2003. Diffusion of innovations under supply constraints. *Operations Research* **51**(6) 866–879.

Mahajan, V., E. Muller. 1996. Timing, diffusion, and substitution of successive generations of technological innovations: The {IBM} mainframe case. *Technological Forecasting and Social Change* **51**(2) 109–132.

Mahajan, V., E. Muller, F. M. Bass. 1991. New product diffusion models in marketing: A review and directions for research. N. Nakicenovic, A. Grubler, eds., *Diffusion of Technologies and Social Behavior*. Springer Berlin Heidelberg, 125–177.

Mahajan, V., E. Muller, F. M. Bass. 1995. Diffusion of new products: Empirical generalizations and managerial uses. *Marketing Science* **14**(3 supplement) G79–G88.

Negahban, A. 2013. A hybrid simulation framework for the newsvendor problem with advertising and viral marketing. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, M. E. Kuhl, eds., *Proceedings of the 2013 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., 1613–1624.

Negahban, A., L. Yilmaz. 2014. Agent-based simulation applications in marketing research: An integrated review. *Journal of Simulation* **8**(2) 129–142.

Negahban, A., L. Yilmaz, T. Nall. 2014. Managing production level in new product diffusion: An agent-based simulation approach. *International Journal of Production Research* **52**(17) 4950–4966.

Norton, J. A., F. M. Bass. 1987. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science* **33**(9) 1069–1086.

Norton, J. A., F. M. Bass. 1992. Evolution of technological generations: the law of capture. *Sloan Management Review* **33**(2) 66–77.

Peres, R., E. Muller, V. Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* **27**(2) 91–106.

Sethi, S. P., G. L. Thompson. 2000. *Optimal Control Theory: Applications to Management Science and Economics*. 2nd ed. Kluwer Academic Publishers, Boston, MA, USA.

Shen, W., I. Duenyas, R. Kapuscinski. 2011. New product diffusion decisions under supply constraints. *Management Science* **57**(10) 1802–1810.

Shen, W., I. Duenyas, R. Kapuscinski. 2014. Optimal pricing, production, and inventory for new product diffusion under supply constraints. *Manufacturing & Service Operations Management* **16**(1) 28–45.

Sultan, F., J. U. Farley, D. R. Lehmann. 1990. A meta-analysis of applications of diffusion models. *Journal of Marketing Research* **27**(1) 70–77.

Wilhelm, W. E., P. Damodaran, J. Li. 2003. Prescribing the content and timing of product upgrades. *IIE Transactions* **35**(7) 647–663.

Wilhelm, W. E., K. Xu. 2002. Prescribing product upgrades, prices and production levels over time in a stochastic environment. *European Journal of Operational Research* **138**(3) 601–621.

Wilson, L. O., J. A. Norton. 1989. Optimal entry timing for a product line extension. *Marketing Science* **8**(1) 1–17.

Chapter 4

A joint analysis of production and seeding strategies for new products:

An agent-based simulation approach

4.1 Abstract

When introducing a new product, firms often face the dilemma of how to support a fast and substantial take-off by targeting the right population of potential consumers for seeding and how to avoid supply shortages and lost sales by building sufficient inventory before the product's launch. The goal of this paper is to provide a joint analysis of seeding and inventory build-up policies for new products. We propose and experiment with an agent-based simulation model to evaluate different seeding criteria (used for prioritizing individuals for seeding), fraction of the market to seed, and inventory build-up policies under various social network structures, demand backlogging levels, and product categories. In contrast to previous findings (that are mainly based on the assumption of unlimited supply), we show that the seeding strategy that maximizes the adoption rate is not necessarily optimal in the presence of supply constraints. More importantly, we show that determining seeding and build-up policies sequentially may lead to suboptimal decisions and that the optimal combination of seeding and build-up policy varies for different product categories. We perform systematic experimentation with different small-world and scale-free networks and illustrate how the distribution of long-range connections and influential nodes would affect the adoption, demand backlogging, and lost sales dynamics as well as the overall profit. The important implications of the findings for diffusion research as well as marketing and operations management practice are also discussed.

keywords: innovation diffusion, viral marketing, myopic and build-up policies, social network, agent-based modeling and simulation

4.2 Introduction

The initial phase of market penetration is critical to the success of new products. Companies generally rely on promotional activities to support a fast and substantial take-off to increase the chance of a successful launch and diffusion. A common approach to encourage adoption and positive word-of-mouth during this initial phase is by distributing reduced-price products to a set of target consumers in the hope that they will initiate a word-of-mouth grapevine. This approach is generally referred to as *viral marketing* or *seeding*. Given an arbitrary social network, individuals with high *influence* (for instance, consumers with a high number of ties) can potentially encourage more adoptions faster. Also, seeding a larger proportion of the population would increase the chance of faster diffusion and adoption. However, giving away more reduced-price products would decrease total revenue. A faster adoption rate would also increase the risk of supply shortages and lost sales (i.e., loss of potential profit). Therefore, targeting and level of seeding affect the overall profit and appropriate decisions regarding the two are crucial for a successful launch and diffusion. More importantly, these decisions are not independent of the supply level.

When introducing a new product/technology, firms also need to determine an appropriate supply level to satisfy future demand. Starting sales without building an initial inventory (the *myopic* policy) can lead to supply shortages and lost sales when demand exceeds production capacity. With the increased rate of introduction of new products and shorter product life-cycles, common capacity management strategies would fail to respond to changes in the demand levels due to the high cost and long lead time of capacity expansions. A common strategy to avoid this problem is to build inventory prior to starting sales as a supply cushion replacing the need for capacity expansions (the *build-up* policy). While the build-up policy would reduce the chance of excessive lost sales, it incurs higher inventory costs and delays revenue collection from sales. Therefore, an appropriate build-up policy is critical to the overall sales and profit from the new product. More importantly, the required initial inventory at launch depends on the diffusion speed (the faster the demand dynamics, the more build-up inventory is needed) and thus ignoring the effect of seeding on the future demand could lead to incorrect decisions on the appropriate build-up policy.

The trade-offs between seeding and build-up policies can be summarized as follows. Seeding requires giving away reduced-price products to accelerate demand dynamics and revenue collection from sales necessitating higher initial inventory levels to reduce the risk of supply shortages and lost sales while longer inventory build-up periods delay sales and incur higher inventory costs. Moreover, whom to seed depends on the social network structure and the distribution of influential nodes. Therefore, a joint analysis of seeding and build-up policies that accounts for the effect of the social network structure is crucial for effective decision-making. As will be discussed in Section 4.3, while the two problems are addressed separately, ignoring the above trade-offs is identified as one of the major gaps in the existing literature. This paper aims at taking the first step towards a joint analysis of marketing and production strategies for new products by exploring the following research questions that the current literature leaves unanswered:

- Does making decisions sequentially on the seeding and build-up policies lead to sub-optimal decisions?
- How does the optimal combination of production and seeding strategies vary for different product categories?
- Is the seeding strategy that maximizes the adoption rate optimal in the presence of binding supply constraints?
- What is the effect of the social network structure on the optimal seeding fraction, seeding criterion, build-up policy, and the resulting diffusion dynamics?

We develop an agent-based diffusion model to address the above questions. Agent-based modeling and simulation (ABMS) is capable of capturing the effect of individuals' decision-making behavior and their interactions in the context of a social network (i.e., micro-level factors) on the emergence of diffusion dynamics (i.e., macro-level market behavior). ABMS is considered as a powerful analysis tool for studying innovation diffusion and is extensively used to investigate different marketing or production strategies, positive and negative word-of-mouth, and theoretical and empirical social network structures (for a comprehensive list of these studies and critical analysis of the literature, see the survey papers by Negahban and Yilmaz (2014), Garcia (2005), Kiesling et al. (2011), Zenobia et al. (2009), and Jager (2007)). Through extensive experimentation with the model, we evaluate the performance of seeding and build-up polices in terms of the net present value (NPV) of profit under various network structures, product categories, and demand backlogging behavior. Moreover, we analyze the resulting diffusion process (low-level results) for different small-world and scale-free networks by systematically changing the parameters of the two network types and comparing the adoption, demand backlogging, and lost sales dynamics.

The remainder of the paper is organized as follows. A critical analysis of the literature is presented in Section 4.3. Section 4.4 provides a detailed description of the simulation logic and different components of the agent-based model including the consumer decision-making behavior, social network structure, seeding strategy, and build-up policy. Experimental design and important findings are discussed in Section 4.5. Finally, Section 4.6 presents the conclusions, managerial implications, limitations, and future research opportunities.

4.3 Literature review

There is a long history of research on new product diffusion with early studies going back to the 1960's with the introduction of several demand models including the Bass model (Bass 1969), which is empirically tested and validated for hundreds of product categories (see Sultan et al. (1990) for a meta-analysis of 213 applications of the Bass model). Many extensions of these models have been proposed to enhance demand forecasting for new products, namely diffusion of successive technology generations (Norton and Bass 1987), effect of negative word-of-mouth (Mahajan et al. 1984), stochastic Bass models (Kanniainen et al. 2011), and diffusion of short-life cycle technologies (Aytac and Wu 2010). For critical analyses of the innovation diffusion literature, see the review papers by Hauser et al. (2006) and Peres et al. (2010). The Bass model and its extensions are also used extensively to analyze the performance of different marketing and production strategies. However, perhaps the most important deficiency in the literature is the lack of a joint analysis of these decisions. Here, we provide a critical review of two streams of research on seeding and production policies to illustrate research strengths and gaps and characterize the main contributions of this work to each of the two streams.

As a relatively recent stream of research, the literature on the interface of marketing and operations management investigates the inter-dependency between the demand and supply for new products using both analytical and simulation approaches. In the context of analytical models, Kumar and Swaminathan (2003) and Ho et al. (2002) independently propose a modified supply-restricted Bass model and develop mathematical models to determine the optimal production-sales policy (i.e., number of build-up periods, production capacity, and launch time) for a single generation of a new product. Similar analytical models are developed to study the optimal production-sales policy under different supply chain topologies (Amini and Li 2011), dynamic pricing (Shen et al. 2014, Swami and Khairnar 2006), multi-stage ordering and diffusion (Bilginer and Erhun 2015), learning phenomenon (Cantamessa and Valentini 2000), and multi-generation innovation diffusion (Negahban and Smith 2016b). In the context of simulation approaches, Amini et al. (2012) develop an agent-based model to evaluate the performance of myopic and build-up policies under both positive and negative word-of-mouth and various network structures. Negahban et al. (2014) evaluate different production management strategies and lengths of the planning horizon using an agent-based model where the manufacturing firm periodically updates its forecast of the future demand by using the Bass model and adjusts its production level accordingly. In two related studies (Negahban and Smith 2016a, 2014), Monte Carlo simulation and statistical approaches are used to illustrate that ignoring demand and supply uncertainties could lead to potentially incorrect decisions on the optimal build-up policy. The papers also show that the optimal decision could be different if risk measures are used as the primary performance metric instead of the expected value (mean) of profit.

The studies under the marketing-operations management interface category provide valuable insights about supply-constrained innovation diffusion by showing how supply shortages and lost sales affect the future demand dynamics, as well as the optimal production capacity, build-up policy, introduction time, and sales plan for new products. However, the effect of marketing strategies on the new product's demand, and consequently on the optimal production-sales policy is generally ignored. This paper contributes to this stream of research by providing a joint analysis of production and viral marketing strategies and shows that the optimal build-up policy varies under different seeding criteria and levels, social network structures, and consumers' backlogging behavior.

There is a rich stream of research on the effect of viral marketing and social network structure in both marketing and social sciences literature. Leskovec et al. (2007) analyze the propagation of person-to-person recommendations in a real network consisting of four million people to identify populations, product, and pricing categories for which viral marketing would be effective. Through agent-based simulation, Delre et al. (2007) study the efficacy of various promotional strategies and find that the absence, inappropriate timing and/or targeting of promotional efforts may lead to diffusion failure. Goldenberg et al. (2009) identify two types of hubs (individuals with exceptionally large number of ties), namely innovative and follower hubs, and find that innovative hubs have a greater impact on the diffusion speed while follower hubs have a greater impact on market share. In another study, Delre et al. (2010) use ABMS to test the effect of highly connected agents (VIPs) and network topology on the success and failure of innovation diffusion and conclude that the persuasive power of VIPs is not as important as their capacity to inform many consumers. Through ABMS, Libai et al. (2005) compare the *uniform*, *support-the-strong*, and *support-the-weak* strategies to find conditions under which each strategy would be more effective.

In another ABMS study, Haenlein and Libai (2013) compare targeting of *revenue leaders* (customers with high lifetime value) versus *opinion leaders* (individuals with disproportional effect on others) and show that the seeding fraction and distribution of revenue leaders in the population determine which seeding strategy is preferable. For a list of related studies, see Goldenberg et al. (2010), van Eck et al. (2011), Iyengar et al. (2011), Negahban (2013), and Yenipazarli (2014).

The strength of this stream of research is in the development and assessment of viral marketing strategies and enhancing our understanding of the role of *influentials* in the success and failure of new product diffusion under different social network structures. Exploring promotional activities using empirical data and network structures has provided valuable insights not only for researchers but also marketing practitioners on the effectiveness of viral marketing. However, perhaps the most important gap in these studies is an almost complete neglect of production constraints by assuming unlimited supply and thus ignoring the effect of supply shortages on the future demand dynamics. While these studies generally define the optimal marketing strategy so as to maximize the diffusion speed, in the presence of limited supply, this would lead to high lost sales and decreased profit. Moreover, some of the findings seem somewhat unrealistic and counterintuitive as a result of ignoring the supply side of new product diffusion. For instance, Stonedahl et al. (2010) find that there is virtually no difference between the performance trends for different seeding strategies under high and low virality (internal influence). The current paper contributes to this stream of research by considering seeding in conjunction with the build-up policy and shows that the optimal seeding fraction and criterion depends on the decisions made on the supply level. We also show that under supply constraints, viral marketing strategies may indeed perform differently in terms of their impact on diffusion speed and overall profit. We also show how the preferences change under different social network structures. To the best of our knowledge, this work is the first to address the above issues.

4.4 The agent-based model

The proposed agent-based model consists of a *Firm* agent that produces and markets a new product and *Consumer* agents that reside in a social network. Consumers interact with each other through positive and negative word-of-mouth (WOM) and make decisions regarding the adoption of the new product. Initially, all consumers are *undecided* meaning that they have not made a decision on whether to adopt or reject the new product. At every time tick, some of the remaining undecided consumers may decide to adopt the product in which case they become *adopters*. Depending on supply availability, some of the adopters may receive the product (i.e., adopters with met demand) in which case they will either become *satisfied* or *dissatisfied* with the product. Those adopters that do not receive the product due to supply shortages will either decide to wait for another period (*waiting* adopters) or completely withdraw (*lost* customers). Rejecters, dissatisfied, and lost customers communicate negative WOM, satisfied consumers communicate positive WOM, and undecided and waiting consumers engage in neither positive or negative WOM. The above categorization of the adoption status of consumers and operation of word-of-mouth is based on existing agent-based diffusion models (Goldenberg et al. 2007, Negahban et al. 2014, Amini et al. 2012) and is summarized in Figure 4.1.

The general logic of the simulation process is summarized in Figure 4.2 while different components of the model are described in detail in the following subsections. The simulation starts by generating a population of potential consumers and connecting them based on the choice of the social network structure (Section 4.4.2). The product is not launched until the end of the build-up period. Therefore, during the build-up period, the firm produces at the maximum production level and sales will be zero. Seeding is performed one period before the product is released into the market while the seeded individuals are selected based on the choice of the seeding strategy (Section 4.4.3). The seeded consumers will be either satisfied or dissatisfied with the product. At every time period after the product's launch, undecided consumers will evaluate positive and negative feedback from their peers and make their

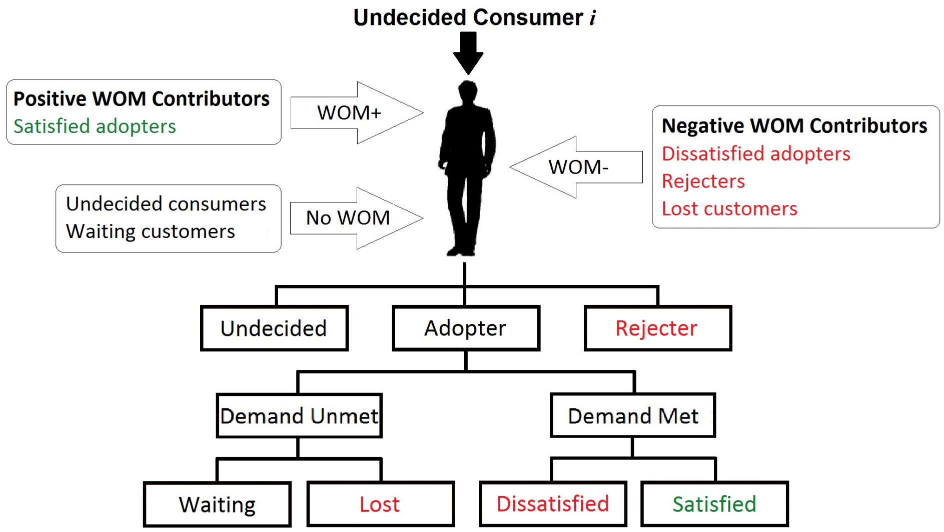


Figure 4.1: Adoption status and the word-of-mouth operation in the proposed agent-based model

adoption decision according to their behavioral rules (Section 4.4.1). Once an undecided consumer adopts the product, she will receive the product if there is supply and become either satisfied or dissatisfied with the product; otherwise, she will either decide to wait or withdraw.

The supply at each time period is used to satisfy the demand of waiting customers as well as new adopters and is determined by the current production level and inventory carried over from the previous period. While the firm continues to produce at maximum capacity after the product's launch, to avoid excessive production and inventory towards the end of the diffusion, once the demand starts to decline and drops below the maximum capacity, the production level will be set to the demand in the previous period – a strategy known as *lagging demand* (Olhager et al. 2001) that is commonly used in related studies (Ho et al. 2002, Kumar and Swaminathan 2003, Negahban and Smith 2016a). The simulation ends when the market is exhausted (i.e., all individuals have made their adoption/rejection decision). At the end of the simulation run, the net present value of profit is calculated based on the choice of the discount factor where the profit for each time period is calculated by subtracting the cost of production, inventory, and backlogged demand from sales.

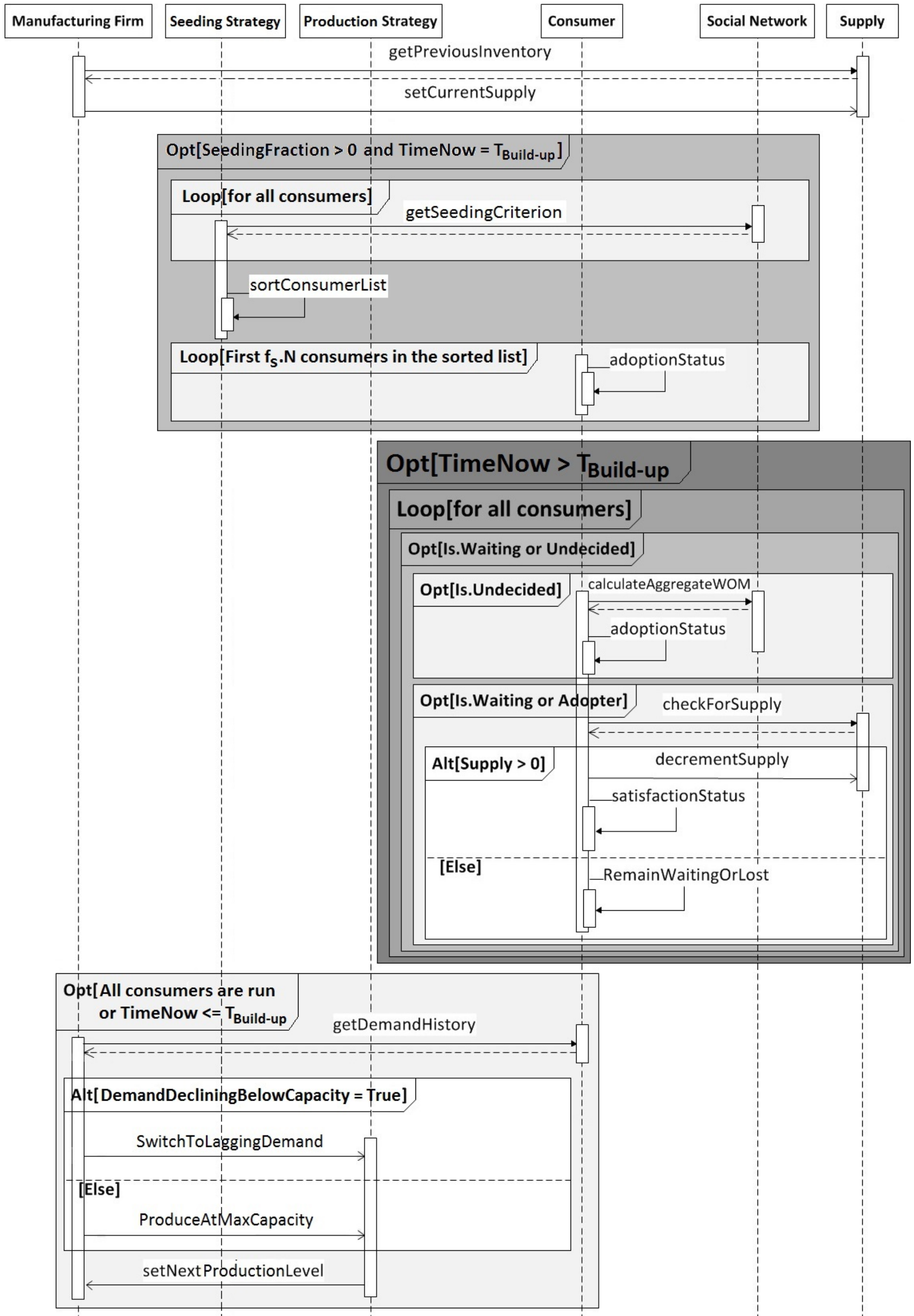


Figure 4.2: The Unified Modeling Language (UML) sequence diagram of the simulation logic. For more information about UML diagrams, see Rumbaugh et al. (2004).

4.4.1 Consumer decision-making behavior

Several approaches have been proposed to model the consumer decision making behavior in the context of agent-based simulation of virtual markets (see the review paper by Negahban and Yilmaz (2014) for a comprehensive list of these approaches). Here, we adopt a commonly used model from the literature (Goldenberg et al. 2007, Negahban et al. 2014, Amini et al. 2012) that considers the effect of both external (e.g., promotions and mass media advertising) and internal (i.e., positive and negative word-of-mouth) influences. The model is also consistent with the assumptions of the Bass model (Bass 1969) where the magnitude of the effect of external and internal influences is captured by the coefficient of innovation, p, and the coefficient of imitation, q, respectively. As suggested by Goldenberg et al. (2007), p is the same for both the individual and aggregate levels while the population-level q is adjusted for each individual i by dividing it by the number of its direct social ties to obtain q_i. Therefore, at any period t, the probability that an individual i is influenced by positive feedback ($P^{+}_{i,t}$) or negative WOM ($P^{-}_{i,t}$) is given by

$$p^{+}_{i,t} = 1 - (1-p)\prod_{j}(1-q_j), \text{ where } j \in L(i)^{+}, \tag{4.1}$$

$$p^{-}_{i,t} = 1 - \prod_{j}(1-M \times q_j), \text{ where } j \in L(i)^{-}, \tag{4.2}$$

where $(1-p)$ is the probability of not receiving external influences and $L(i)^{+}$ is the set of direct social ties that communicate positive WOM. Therefore, the probability of not receiving positive feedback from each $j \in L(i)^{+}$ is $(1-q_j)$. Subtracting the product of the two terms from 1 gives the probability of positive influence about the new product. Similarly, let $L(i)^{-}$ be the set of social ties that communicate negative WOM. Then, $(1 - M * q_j)$ denotes the probability of not receiving any negative feedback from neighbor $j \in L(i)^{-}$ where the impact of negative WOM is M times greater than positive WOM.

At time t, the probability that consumer i receives *only* positive WOM is $(1 - p_{i,t}^-)p_{i,t}^+$. Similarly, $(1-p_{i,t}^+)p_{i,t}^-$ is the probability of receiving only negative WOM, and $p_{i,t}^+ p_{i,t}^-$ the probability of receiving both positive and negative feedback. For the last group, the individual adopts the product with a probability of $\alpha_{i,t} = p_{i,t}^+/(p_{i,t}^+ + p_{i,t}^-)$ and rejects with a probability of $1 - \alpha_{i,t}$ which leads to the final state transition probabilities:

$$p_{i,t}^{adopt} = (1 - p_{i,t}^-)p_{i,t}^+ + \alpha_{i,t}\ p_{i,t}^+\ p_{i,t}^-, \tag{4.3}$$

$$p_{i,t}^{reject} = (1 - p_{i,t}^+)p_{i,t}^- + (1 - \alpha_{i,t})p_{i,t}^+\ p_{i,t}^-, \tag{4.4}$$

$$p_{i,t}^{undecided} = (1 - p_{i,t}^+)(1 - p_{i,t}^-), \tag{4.5}$$

where Equations (4.3)-(4.5) are the adoption, rejection, and remaining undecided probabilities, respectively, and will sum to 1. If there is supply, the adopter will receive the product and become dissatisfied with a probability d and satisfied with a probability $1 - d$. If there is no supply, the adopter will either wait for another period with a probability of backlogging β or withdraw with a probability $1 - \beta$.

4.4.2 Social network structure

We model a market consisting of $N = 3000$ socially networked consumers. Previous studies have shown that this is sufficiently large to effectively capture the diffusion dynamics and provide statistically valid results (Cowan and Jonard 2004). Other studies also confirm that there is no significant statistical difference between the simulation results for networks of size 3000 and above (Goldenberg et al. 2007, Amini et al. 2012). We consider the following four network structures:

- *Regular lattice:* In a lattice, every node has the exact same number of ties. In this paper, we consider a lattice where each node has exactly 26 neighbors in order to provide a common ground for comparison with previous studies (Stonedahl et al. 2010, Negahban et al. 2014, Amini et al. 2012). An average of 26 ties is also consistent with the findings of an empirical study on the number of ties by Goldenberg et al. (2007).

- *Random:* In a random network of size N, any two nodes will be connected with a probability of ϕ resulting in a network with $\phi \times N$ as the expected number of edges for each node. Based on the above, in our implementation of the random network, ϕ is set to 26/3000 which leads to an average of 26 edges per node.

- *Small-world:* Introduced by Watts and Strogatz (1998), a small-world network lies between a lattice and a random network (two extremes). Many real-world systems are found to be highly clustered (like regular lattices) while having short path lengths (like random networks) and thus can be modeled with an appropriate small-world network (for example, see the paper by Vieira et al. (2009) that studies the dynamics of HIV infection). A small-world network is generated by starting with a lattice where nodes are connected to their nearest neighbors and then some edges are rewired randomly with a probability of $P_{rewiring}$. This would result in long-range connections between nodes leading to a small-world phenomenon often referred to as *six degrees of separation* (Milgram 1967, Guare 1990). In our implementation of the small-world network, we start with a lattice with a degree of 26 and systematically experiment with different rewiring probabilities ($P_{rewiring}$) from 0 (lattice) to 1.0 (random network).

- *Scale-free:* The scale-free network, introduced by Barabási and Albert (1999), strives to incorporate two generic aspects of many real networks, namely growth and preferential attachment. In the above models, we start with a fixed number of nodes (N) while many existing real-world networks in fact grew over their lifetime. Secondly, highly connected nodes are particularly absent in the above networks while they can be found in many real networks where the distribution of the nodes' connectivity decays as a power law. To incorporate the growth mechanism, a scale-free network is generated by starting with a small number of nodes (m_0) and then new nodes are sequentially added to the network. To incorporate preferential attachment, the probability π that a new node will be connected to an existing node i is related to the connectivity of node i, k_i, so that $\pi(k_i) = (k_i / \sum_j k_j)^{\gamma_{power}}$. In this paper, we start with $m_0 = 10$ initial nodes and experiment with different values for the exponent γ_{power}.

4.4.3 Production and seeding strategies

We consider both myopic and build-up policies for the firm agent. In the myopic policy with zero build-up periods ($T_{Build-up} = 0$), the firm starts the production and releases the product at the same time while for the build-up policy ($T_{Build-up} > 0$), during the periods of inventory build-up, consumer agents will remain inactive and thus the diffusion does not start until after the build-up period is over. Just before the product is launched, a fraction f_{seed} of the population will be targeted for seeding. In this model, seeded consumers receive reduced-price products at the production cost (i.e., the company does not make any profit from the seeded products). We consider the following five seeding criteria to prioritize consumers and decide who should be targeted for seeding (in all cases, ties are broken randomly):

- *Degree:* The nodes with more neighbors (i.e., higher degree) can directly encourage more adoption. Consumer agents are prioritized based on their degree so that high-degree nodes will be seeded first.

- *Two-step:* In the two-step strategy, a node that has a higher number of nodes reachable within two steps (i.e., two edges) will have a higher priority for seeding – an extension to the degree strategy.

- *Average path length:* This measure corresponds to the average number of links between the target node and any other node in the network. A small average path length is preferable as it suggests faster connection to the entire network leading to potentially more adoptions faster. It is worth noting that this method can be computationally intensive, especially for large networks, as it requires finding the *shortest path* between every pair of nodes in the network.

- *Clustering coefficient:* This measure determines how close a node and its neighbors are to being a clique (complete graph). Let k_i denote the number of neighbors for node i. The clustering coefficient for node i, cc_i, is calculated by the number edges between the neighbors of node i, $e_i^{neighbors}$, divided by the total number of edges possible between its neighbors, i.e., $cc_i = e_i^{neighbors}/C_{2,k_i}$, where C_{2,k_i} is the combination of k_i choose 2. A node with a lower clustering coefficient is given a higher priority for seeding as there is less overlap among its neighbors encouraging a wider adoption more quickly.

- *Random:* In this method, nodes are randomly selected for seeding by assigning a random priority $u \sim Uniform[0, 1]$ to each node.

4.5 Model implementation and results

The simulation model is implemented in Repast Symphony (North et al. 2006), a Java-based ABMS platform. Before performing the simulation experiments, verification and validation (V&V) are performed according to the guidelines by Rand and Rust (2011), Balci (1994), and Sargent (2005) to gain confidence about the correct implementation of the model. For the sake of conciseness, a brief description of the V&V steps is provided here.

Through structured walk-through and tracing, different processes including agents activation and scheduling of methods at each time step, random number generation and probabilistic decisions, synchronous updating of consumers' adoption status, identification of agent's personal social network, and calculation of seeding criteria are extensively debugged and verified to ensure that the model is error-free and properly implemented. Operational graphics of the firm's production level and number of undecided consumers, rejectors, and adopters (including satisfied, dissatisfied, waiting, and lost) are developed to observe the emergence of diffusion dynamics over time and verify the model's behavior. Degenerate

Table 4.1: Parameter choices for the agent-based simulation model

Parameter	Value/Range
The Firm agent	
Build-up period ($T_{Build-up}$)	0-20
Seeding fraction (f_{seed})	0-0.07 in 0.005 increments
Seeding Strategy	Degree, Two-step, Average path length, Clustering coefficient, Random
Discount factor (θ)	0.005, 0.01, 0.02
Initial production level	40
Unit selling price	1.2
Unit production cost	1
Unit inventory holding cost	0.001
Cost per waiting customer	0.001
The Consumer agent	
Population size (N)	3000
Coefficient of innovation (p)	0.001, 0.005, 0.01
Coefficient of imitation (q)	0.3, 0.5, 0.7
Influence of negative to positive WOM (M)	2
Dissatisfaction probability (d)	5%
Probability of waiting (β)	0, 0.8
The social network	
Network type	Random, Lattice, Small-world, Scale-free
Random network connection probability (ϕ)	26/3000
Regular lattice degree	26
Small-world degree	26
Small-world rewiring probability ($P_{rewiring}$)	0, 0.001, 0.005, 0.01, 0.05, 0.1, 1.0
Scale-free initial number of nodes (m_0)	10
Scale-free exponent (γ_{power})	1.05, 1.1, 1.15, 1.2, 1.5, 2.0

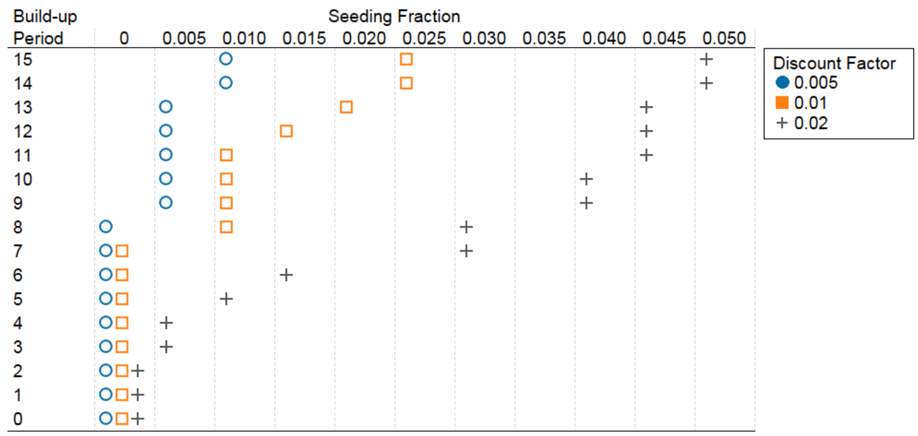
tests are performed to verify the behavior of the model in extreme conditions, including 100% dissatisfaction probability, coefficient of innovation of 0 and 1 as well as fully connected and disconneted networks. We also perform sensitivity analysis on a number of predictable scenarios to ensure that results match the expected behavior. For instance, we systematically increased the dissatisfaction probability from 0 to 1 and observed an increase in the number rejecters. We also observed increased rejections as we systematically decreased the probability of waiting/backlogging. This is the expected behavior as both cases would lead to a higher number of individuals that communicate negative WOM. A similar behavior is observed as we increase M (the relative power of negative WOM to positive WOM).

4.5.1 Experimentation

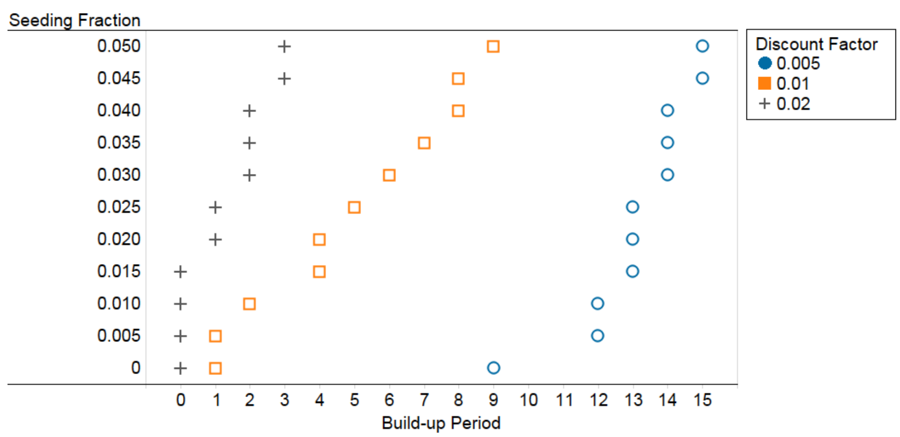
Extensive experimentation was conducted using the agent-based simulation model with more than 10.2 million parameter configurations (the experimental design is summarized in Table 4.1). In order to provide a common ground for comparison, most of the parameter choices are adopted from related studies (Negahban et al. 2014, Amini et al. 2012, Stonedahl et al. 2010, Goldenberg et al. 2007). For each scenario, we collect simulation outputs on the number of undecided consumers, rejectors, and adopters (including satisfied, dissatisfied, waiting, and lost) at each time tick as well as the realized net present value of profit. Based on a preliminary analysis of the confidence interval half-widths from a set of pilot runs, the number of replications is set to 50 to provide statistically accurate results on the above metrics. Therefore, more than 510 million replications are run in total which required approximately 25,000 hours (1040 days) of computational time (less than a month in real time as the experiments were distributed across a computing cluster).

4.5.2 Optimal production and seeding strategies

In general, as shown in Figrue 4.3(a), a higher seeding fraction becomes preferable with longer build-up periods as more supply increases the ability of the firm to handle a fast growth in the demand for the new product. Seeding essentially accelerates the diffusion process resulting in a higher demand more quickly and, given supply availability, the firm would enjoy faster revenue collection from sales. With high discounting, accelerated sales would yield a higher NPV of profit that outweighs the loss in potential profit that would otherwise have been gained from the seeded individuals if they had adopted the product. As a result, given any build-up period, the optimal seeding fraction is an increasing function of the discount factor. On the other hand, as shown in Figure 4.3(b), more periods of inventory build-up are necessary to handle the increase in the diffusion speed as a result of a higher seeding level. However, longer periods of build-up would delay launch and revenue collection



(a) Optimal seeding fraction for different build-up periods ($p = 0.01$, $q = 0.7$, and impatient customers with $\beta = 0$)



(b) Optimal build-up period for different seeding fractions ($p = 0.005$, $q = 0.5$, and impatient customers with $\beta = 0$)

Figure 4.3: Optimal seeding fraction and build-up period (aggregate results over all network structures and seeding strategies based on 50 replications)

from sales. Therefore, for any seeding fraction, the optimal number of build-up periods decreases with higher discounting.

In Figure 4.4, we analyze the combination of production and seeding policies for different levels of coefficients of innovation (p) and imitation (q). In this plot, the blocks correspond to different combinations of p and q (i.e., different product categories). In each block, the x-axis represents the build-up period, the y-axis is the seeding fraction, and the NPV of profit is color-coded with the darkest shade of green corresponding to a *high-profit* region and the darkest shade of red a *high-loss* region. These results have important managerial implications as the size of the high-profit region can be thought of as a measure of difficulty in making a near-optimal decision regarding the combination of production and seeding strategies (a small region indicates a difficult decision as only a few policies would yield relatively high profit levels). For products with smaller innovation factors, it becomes critically important which and how many individuals to seed as there are fewer innovators in the population to kick off the diffusion process. With higher innovativeness levels, absolutely *right* targeting and level of seeding promotions become less critical as there are more innovators in the population that would adopt the product independently and help accelerate the demand for the new product. As a result, we observe a general increase in the size of the high-profit region with higher levels of innovation. On the other hand, a high coefficient of imitation essentially leads to high adoption pressure from peer-based word-of-mouth and thus consumers are likely to make their adoption/rejection decision more quickly. Therefore, it becomes critical to target the *right* individuals as the diffusion process will be increasingly sensitive to this initial population of adopters. As a result, the size of the high-profit region shrinks as the imitation factor increases.

We also observe that for high levels of innovation and imitation, the likelihood of a negative profit (i.e., *loss* region) diminishes as the demand is relatively higher than the production rate making even large initial inventory levels profitable. However, for products with a low p and q, long build-up periods would lead to excessive production and inventory costs due to a slow demand growth rate. Further analysis of the results indicates that the global optimal build-up period and seeding fraction depend on the seeding criterion (Table 4.2). The table also confirms our previous finding on the effect of discounting that in general, short build-up periods are preferable for high discounting levels. With a low discount factor, longer build-up periods and higher seeding fractions will be more profitable. However, similar patterns are observed in terms of the size of the high-profit and loss regions for different seeding criteria. The above results indicate the importance of a joint analysis

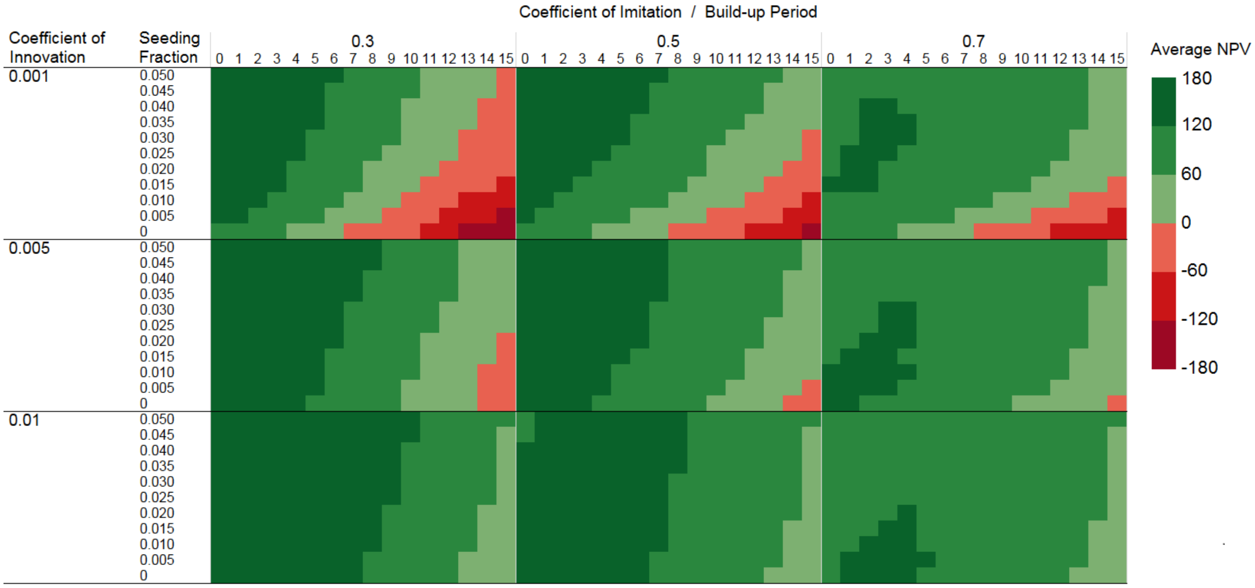


Figure 4.4: Optimal combination of seeding fraction and build-up period under different levels of innovation and imitation factors (Degree seeding strategy, small-world network with $P_{rewiring} = 0.05$, $\theta = 0.02$, and partial backlogging with $\beta = 0.8$)

on the build-up and seeding policies by showing that the firm would potentially make a suboptimal decision if these decisions are made sequentially.

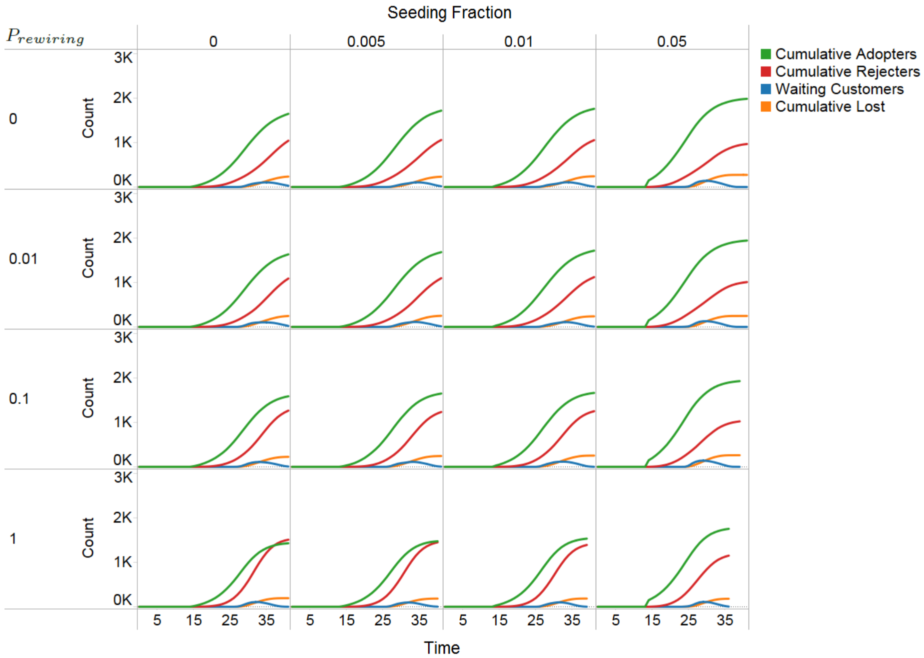
Table 4.2: Optimal combination of build-up period and seeding fraction ($T^*_{Build-up}, f^*_{seed}$) per seeding criteria for different discounting levels θ (other parameters are $p = 0.005$, $q = 0.5$, $\beta = 0.8$, and small-world network with $P_{rewiring} = 0.05$)

Seeding criteria	Discount factor		
	$\theta = 0.005$	$\theta = 0.01$	$\theta = 0.02$
Degree	(15, .035)	(7, .030)	(0, .010)
Two-step	(15, .040)	(8, .045)	(0, .010)
Average path length	(15, .040)	(7, .030)	(0, .010)
Clustering coefficient	(15, .035)	(7, .030)	(0, .005)
Random	(15, .045)	(9, .045)	(0, .010)

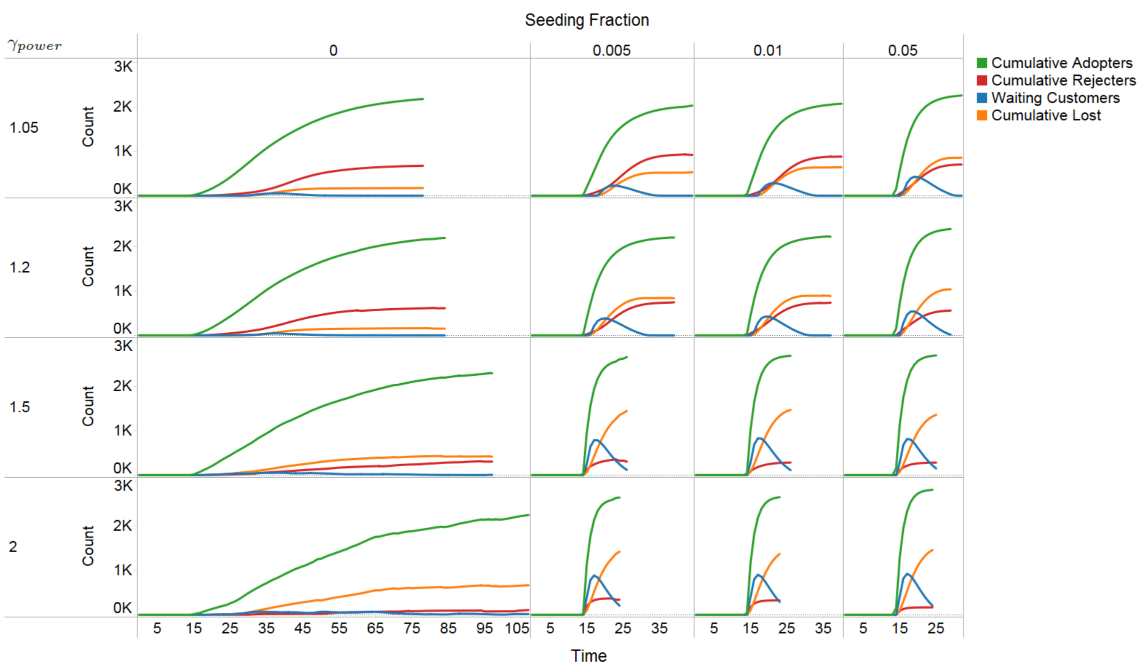
4.5.3 The effect of social network structure

In order to provide insights about the diffusion of innovations in the context of social networks, we systematically experiment with different rewiring probability ($P_{rewiring}$) and offset exponent (γ_{power}) values for the small-world and scale-free network structures and analyze the diffusion dynamics at a lower level. By looking at the rows in Figure 4.5(a), for any $P_{rewiring}$, a higher seeding fraction accelerates the diffusion process and encourages more adoption (the difference between the cumulative adoptions (green) and rejections (red) facilitates identification of the patterns discussed in this section). In the case of a scale-free network, we observe a similar behavior for any γ_{power} except that the effect of seeding fraction on the adoption dynamics is significantly stronger than in the small-world network. This is mainly due to the presence of high-degree nodes (i.e., influentials) who, if seeded, would immediately influence a large number of potential consumers resulting in a much faster take-off for the product, more backlogged demand (waiting customers), and shorter diffusion time. By looking at columns under $f_{seed} = (0.005,\ 0.01,\ 0.05)$ in Figure 4.5(b), we observe that this effect is augmented with a higher γ_{power} as it essentially leads to increased preferential attachment by decreasing the probability of connection to low-degree nodes as new nodes are added during the network generation. Therefore, there will be a large number of nodes with very few connections. As a result, with no seeding ($f_{seed} = 0$) the diffusion process slows down with higher γ_{power} as innovators (early adopters) may not necessarily be among the high-degree individuals. Finally, we observe that as $P_{rewiring}$ increases (columns in Figrue 4.5(a)), the cumulative adoptions decreases. A high $P_{rewiring}$ would increase the number of long-range connections in the initial lattice encouraging a *wider* adoption, while at the same time, it essentially decreases the magnitude of word-of-mouth and adoption pressure from peers on the nearest neighbors delaying their adoption decision which would result in a slower adoption curve.

Figure 4.6 illustrates the adoption and waiting dynamics under the two network structures for different seeding criteria. As expected, regardless of prioritization or the network structure, seeding accelerates the adoption process leading to earlier and higher demand backlogging (i.e., supply shortage). While different seeding criteria perform similarly in a small-world network (Figure 4.6(a)), we observe a significant difference between their performance under a scale-free network (Figure 4.6(b)). Interestingly, we observe that the adoption



(a) Small-world network and the effect of $P_{rewiring}$



(b) Scale-free network and the effect of γ_{power}

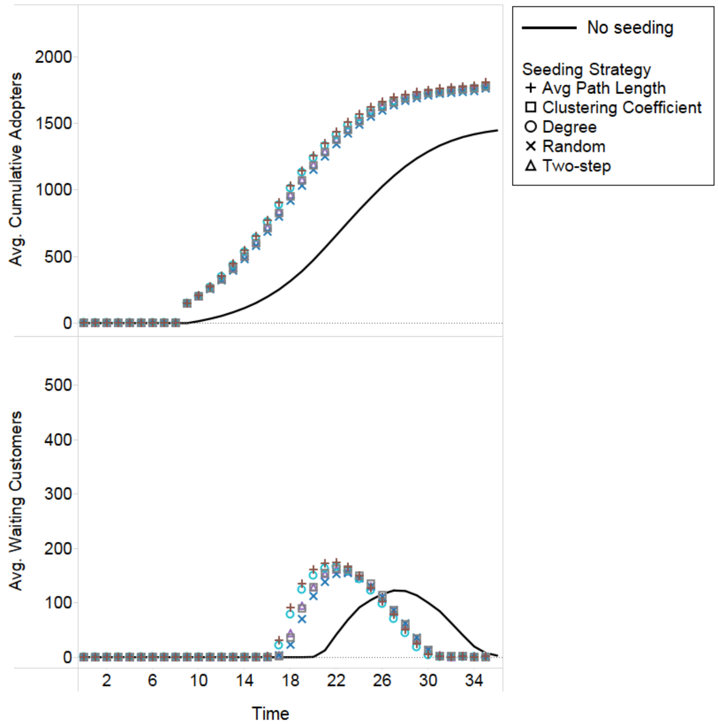
Figure 4.5: Backlogged demand and cumulative adoption, rejection, and lost customers. The difference between the cumulative adoptions (green) and rejections (red) facilitates identification of the patterns (average values are based on 50 replications, $T_{Build-up} = 15$, *Avg. path length* seeding strategy, $p = 0.005$, $q = 0.5$, and partial backlogging with $\beta = 0.8$).

curve for the clustering coefficient criterion is only marginally faster than random seeding in a scale free network, which can be explained by the presence of a large number of very low-degree nodes (with only a few neighbors), which simply have a clustering coefficient of 0 (i.e., high seeding priority) but are ineffective seeding choices. On the other hand, the adoption curve for the two-step method is found to be slower than the degree or average path length strategies. This is because of the low-degree nodes that are direct neighbors of high-degree nodes and thus have a high number of two-step neighbors giving them a high seeding priority. However, if one of these nodes is seeded, it takes two time periods for the information to reach the large number of two-hop neighbors. This short lag makes such nodes not as effective as the actual high-degree nodes in creating a fast take-off.

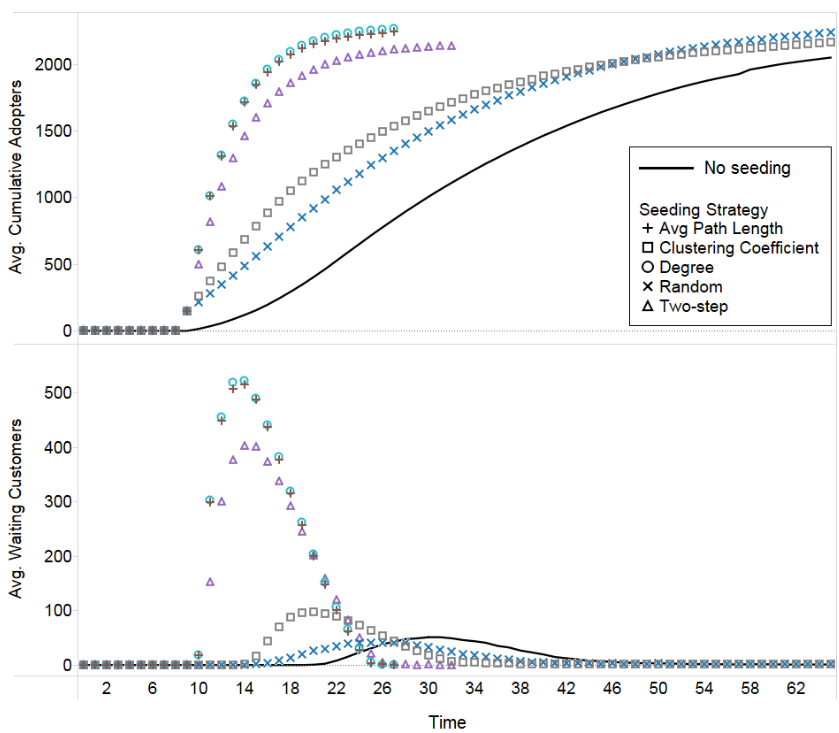
Figure 4.7 compares the NPV of profit for different seeding strategies for small-world and scale-free networks. Of course, fast-growing demand is preferable under the assumption of unlimited capacity. However, in the presence of supply constraints, it would lead to lost sales and decreased NPV of profit which is why in a scale-free network, the three seeding strategies with fastest adoption rates (i.e., average path length, degree, and two-step) yield lower profits than random and clustering coefficient criteria that, in turn, perform better than no seeding. Therefore, our results challenge the findings of previous studies that ignore supply restrictions (for example, see Stonedahl et al. (2010)) by showing that clustering coefficient and random seeding strategies may actually be optimal under a scale-free network despite the fact that they do not accelerate the demand dynamics as much as the other three seeding strategies. Finally, as expected based on Figure 4.6(a), no significant difference is observed between different seeding criteria in a small-world network.

4.6 Discussion and conclusions

We conducted simulation experiments using an agent-based diffusion model to provide insights on the optimal seeding and inventory build-up policies for new products. Our findings have important managerial implications. The results indicate the importance of



(a) Small-world network with $P_{rewiring} = 0.1$



(b) Scale-free network with $\gamma_{power} = 1.1$

Figure 4.6: Cumulative adoption and backlogged demand (50 replications, $T_{Build-up} = 10$, $f_{seed} = 0.05$, $p = 0.005$, $q = 0.5$, and partial backlogging with $\beta = 0.8$)

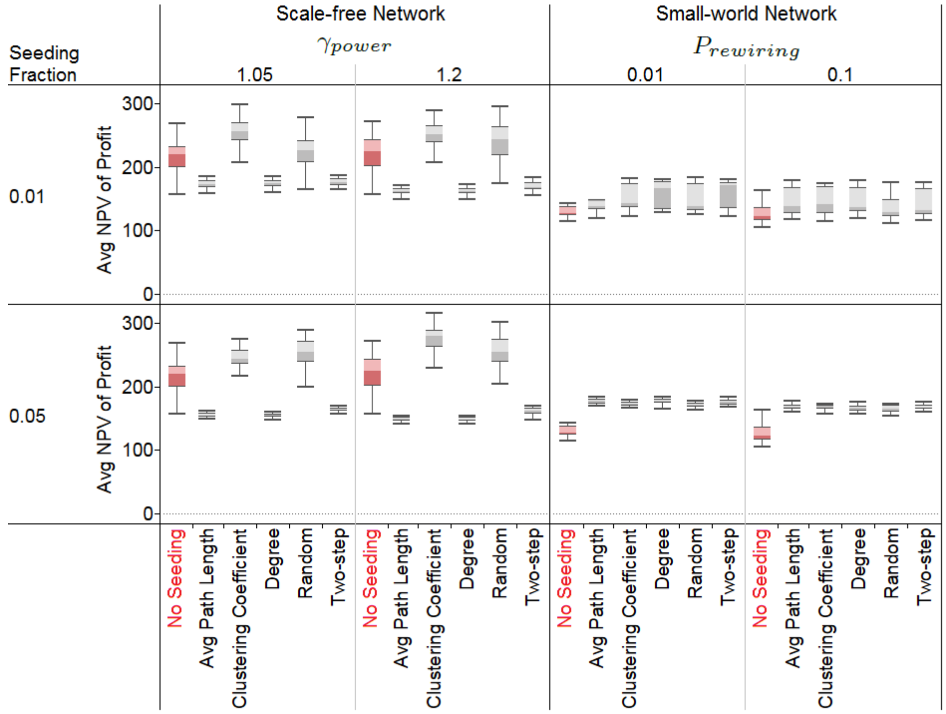


Figure 4.7: Net present value of profit (50 replications, $T_{Build-up} = 10$, $\theta = 0.01$, $p = 0.005$, $q = 0.5$, and partial backlogging with $\beta = 0.8$)

a joint analysis of viral marketing and production policies by showing that these decisions are inter-dependent and that a sequential decision-making process would lead to suboptimal decisions and reduced profit. While previous studies assume unlimited supply and focus on finding the seeding strategy that maximizes the adoption rate and diffusion speed, our results indicate that under supply constraints, such strategies may actually perform poorly causing excessive supply shortages and lost sales. These findings are especially important for short life-cycle products where capacity expansions are not feasible due to long lead times and high costs. The results provide valuable insights about decision-making for different product categories and show that for products with high innovation and imitation factors, there is a smaller set of policies that would yield high profit values requiring a much more careful assessment of seeding and build-up policies before launching the new product.

The results also indicate the importance of the social network structure on the optimal combination of seeding and build-up policies. We study the effect of long-range connections in small-world networks and distribution of hubs (high-degree nodes) in scale-free networks on the resulting diffusion process (including the adoption, waiting, and lost sales dynamics) as well as the product's life-cycle profit. We show that random seeding can in fact be a good strategy for a scale-free network as it would result in a slower diffusion process and thus reduce supply shortages and lost sales as opposed to other strategies that suggest seeding the high-degree nodes which would result in a significantly faster demand growth. Finally, we show that the company's discount rate can have a significant impact on the choice of the seeding and build-up policies and should be explicitly considered in the decision-making process.

There are several important things to note. The above results are generalizable and do not depend on a particular network structure or market configuration. Since many real networks show characteristics of a small-world or scale-free network, we believe our experimental design covers a wide range of real-world networks. Our approach is also applicable to any empirical network developed based on real data. In many cases, data regarding the global network structure may not be available. Even in the case of social networking websites (e.g., Facebook, Twitter, or LinkedIn), only some of the data may be available due to privacy considerations. In such cases, we suggest a heuristic approach where an appropriate small-world or scale-free network is fitted to the data and then the proposed ABMS approach is used to determine the optimal seeding and build-up policies. Sensitivity analysis on the estimated parameters of the fitted theoretical network could also be performed to assess risk. With regard to the other parameters of the model, production and inventory costs can be obtained internally while analysis of purchasing motives as well as historical data from similar products in the past can help estimate coefficients of innovation and imitation, market size, and rate of loss of waiting customers. Surveys and feedback from the individuals that test the product prior to its release would provide some insight about the satisfaction probability. Finally, we also suggest sensitivity analysis on these parameters to develop expectations for different possible outcomes including the best and worst cases to help select a robust policy and reduce the risk of overproduction and/or ineffective seeding.

To the best of our knowledge, this is the first study to consider simultaneous optimization of the seeding and build-up policies. Our model involves several assumptions since our goal in this first step was to understand the basic underlying dynamics of the problem. We only use network-related characteristics as the primary seeding criteria and do not consider the *persuasive power* of individuals. We believe this assumption would not undermine the main findings of the current paper for two reasons. First, existing studies show that the importance of influential nodes is in their capacity to inform many consumers and not in a stronger persuasive power (Delre et al. 2010). Secondly, we argue that, in most real-world situations, individuals with many connections generally have stronger persuasive power due to their political or social position that caused them to have many connections in the first place – as of March 2016, the 100 most followed Twitter accounts belong to celebrities, politicians, major news or government organizations, and innovative corporations (*Twitter, Inc.* 2016). However, it is easy to modify the proposed model to incorporate individuals' persuasiveness level and investigate the performance of alternative seeding criteria. Moreover, we focus on a single seeding criterion at a time while different combinations of these strategies are possible through weighting functions (i.e., linear or non-linear combinations of criteria that assign weights to different criteria to prioritize nodes for seeding). The use of multi-product consumer decision-making models to investigate the problem under an oligopolistic market would be another interesting area for future research. Finally, a number of parameters were fixed for the purpose of our study. Future research could investigate the effect of these parameters, namely the selling price, unit inventory cost, cost per waiting customer, relative influence of negative to positive word-of-mouth, and probability of dissatisfaction with the product. We believe the findings of this work and its future extensions along the above lines would help companies make more informed decisions regarding production and marketing strategies for their new products.

Bibliography

Amini, M., H. Li. 2011. Supply chain configuration for diffusion of new products: An integrated optimization approach. *Omega* **39**(3) 313–322.

Amini, M., T. Wakolbinger, M. Racer, M. G. Nejad. 2012. Alternative supply chain production-sales policies for new product diffusion: An agent-based modeling and simulation approach. *European Journal of Operational Research* **216**(2) 301–311.

Aytac, B., S. D. Wu. 2010. Characterization of demand for short life-cycle technology products. *Annals of Operations Research* **203**(1) 255–277.

Balci, O. 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research* **53**(1) 121–173.

Barabási, A.-L., R. Albert. 1999. Emergence of scaling in random networks. *Science* **286** 509–512.

Bass, F. M. 1969. A new product growth for model consumer durables. *Management Science* **15**(5) 215–227.

Bilginer, O., F. Erhun. 2015. Production and sales planning in capacitated new product introductions. *Production and Operations Management* **24**(1) 42–53.

Cantamessa, M., C. Valentini. 2000. Planning and managing manufacturing capacity when demand is subject to diffusion effects. *International Journal of Production Economics* **66**(3) 227–240.

Cowan, R., N. Jonard. 2004. Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control* **28**(8) 1557–1575.

Delre, S., W. Jager, T. Bijmolt, M. Janssen. 2007. Targeting and timing promotional activities: An agent-based model for the takeoff of new products. *Journal of Business Research* **60**(8) 826–835.

Delre, S. A., W. Jager, T. H. A. Bijmolt, M. A. Janssen. 2010. Will it spread or not? the effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management* **27**(2) 267–282.

Garcia, R. 2005. Uses of agent-based modeling in innovation/new product development research. *Journal of Product Innovation Management* **22**(5) 380–398.

Goldenberg, J., S. Han, D. R. Lehmann, J. W. Hong. 2009. The role of hubs in the adoption process. *Journal of Marketing* **73**(2) 1–13.

Goldenberg, J., B. Libai, S. Moldovan, E. Muller. 2007. The NPV of bad news. *International Journal of Research in Marketing* **24**(3) 186–200.

Goldenberg, J., B. Libai, E. Muller. 2010. The chilling effects of network externalities. *International Journal of Research in Marketing* **27**(1) 4–15.

Guare, J. 1990. *Six degrees of separation: A play*. Vintage Books, New York.

Haenlein, M., B. Libai. 2013. Targeting revenue leaders for a new product. *Journal of Marketing* **77**(3) 65–80.

Hauser, J., G. J. Tellis, A. Griffin. 2006. Research on innovation: A review and agenda for marketing science. *Marketing Science* **25**(6) 687–717.

Ho, T.-H., S. Savin, C. Terwiesch. 2002. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science* **48**(2) 187–206.

Iyengar, R., C. V. den Bulte, T. W. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 195–212.

Jager, W. 2007. The four P's in social simulation, a perspective on how marketing could benefit from the use of social simulation. *Journal of Business Research* **60**(8) 868–875.

Kanniainen, J., S. Makinen, R. Piche, A. Chakrabarti. 2011. Forecasting the diffusion of innovation: A stochastic Bass model with log-normal and mean-reverting error process. *IEEE Transactions on Engineering Management* **58**(2) 228–249.

Kiesling, E., M. Gunther, C. Stummer, L. Wakolbinger. 2011. Agent-based simulation of innovation diffusion: a review. *Central European Journal of Operations Research* **20**(2) 183–230.

Kumar, S., J. M. Swaminathan. 2003. Diffusion of innovations under supply constraints. *Operations Research* **51**(6) 866–879.

Leskovec, J., L. A. Adamic, B. A. Huberman. 2007. The dynamics of viral marketing. *ACM Transactions on the Web* **1**(1) 228–237.

Libai, B., E. Muller, R. Peres. 2005. The role of seeding in multi-market entry. *International Journal of Research in Marketing* **22**(4) 375–393.

Mahajan, V., E. Muller, R. A. Kerin. 1984. Introduction strategy for new products with positive and negative word-of-mouth. *Management Science* **30**(12) 1389–1404.

Milgram, S. 1967. The small-world problem. *Psychology Today* **1**(1) 60–67.

Negahban, A. 2013. A hybrid simulation framework for the newsvendor problem with advertising and viral marketing. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, M. E. Kuhl, eds., *Proceedings of the 2013 Winter Simulation Conference*. Institute of Electrical and Electronics Engineers (IEEE), Inc., 1613–1624.

Negahban, A., J. S. Smith. 2014. Production-sales policies for new product diffusion under stochastic supply. *Proceedings of the Sixth International Conference on Advances in System Simulation (SIMUL'14)*. International Academy, Research, and Industry Association (IARIA), Inc., Red Hook, New York, 197–202.

Negahban, A., J. S. Smith. 2016a. The effect of supply and demand uncertainties on the optimal production and sales plans for new products. *International Journal of Production Research* **(Accepted, in press)**. URL `http://dx.doi.org/10.1080/00207543.2016.1157274`.

Negahban, A., J. S. Smith. 2016b. Optimal entry time and production-sales policies for successive generations of new products. *Submitted to Production and Operations Management* **(Under review)**.

Negahban, A., L. Yilmaz. 2014. Agent-based simulation applications in marketing research: An integrated review. *Journal of Simulation* **8**(2) 129–142.

Negahban, A., L. Yilmaz, T. Nall. 2014. Managing production level in new product diffusion: An agent-based simulation approach. *International Journal of Production Research* **52**(17) 4950–4966.

North, M. J., N. T. Collier, J. R. Vos. 2006. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation* **16**(1) 1–25.

Norton, J. A., F. M. Bass. 1987. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science* **33**(9) 1069–1086.

Olhager, J., M. Rudberg, J. Wikner. 2001. Long-term capacity management: Linking the perspectives from manufacturing strategy and sales and operations planning. *International Journal of Production Economics* **69**(2) 215–225.

Peres, R., E. Muller, V. Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* **27**(2) 91–106.

Rand, W., R. T. Rust. 2011. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing* **28**(3) 181–193.

Rumbaugh, J., I. Jacobson, G. Booch. 2004. *The Unified modeling language reference manual*. 2nd ed. Pearson Higher Education, Boston, MA.

Sargent, R. G. 2005. Verification and validation of simulation models. *Proceedings of the 2005 Winter simulation Conference*. Institute of Electrical and Electronics Engineers, Inc., Piscataway, New Jersey, 130–143.

Shen, W., I. Duenyas, R. Kapuscinski. 2014. Optimal pricing, production, and inventory for new product diffusion under supply constraints. *Manufacturing & Service Operations Management* **16**(1) 28–45.

Stonedahl, F., W. Rand, U. Wilensky. 2010. Evolving viral marketing strategies. *Proceedings of the 12th annual conference on Genetic and evolutionary computation*. 1195–1202.

Sultan, F., J. U. Farley, D. R. Lehmann. 1990. A meta-analysis of applications of diffusion models. *Journal of Marketing Research* **27**(1) 70–77.

Swami, S., P. J. Khairnar. 2006. Optimal normative policies for marketing of products with limited availability. *Annals of Operations Research* **143**(1) 107–121.

Twitter, Inc. 2016. Twitter top 100 most followed. URL `http://www.twittercounter.com/pages/100/`. Accessed: 2016-03-07.

van Eck, P. S., W. Jager, P. S. H. Leeflang. 2011. Opinion leaders' role in innovation diffusion: A simulation study. *Journal of Product Innovation Management* **28**(2) 187–203.

Vieira, I. T., R. C. H. Cheng, P. R. Harper, V. Senna. 2009. Small world network models of the dynamics of HIV infection. *Annals of Operations Research* **178**(1) 173–200.

Watts, D. J., S. H. Strogatz. 1998. Collective dynamics of 'small-world'networks. *Nature* **393**(6684) 440–442.

Yenipazarli, A. 2014. A road map to new product success: warranty, advertisement and price. *Annals of Operations Research* **226**(1) 669–694.

Zenobia, B., C. Weber, T. Daim. 2009. Artificial markets: A review and assessment of a new venue for innovation research. *Technovation* **29**(5) 338–350.

Chapter 5

Concluding remarks

Motivated by real-world problems, this dissertation explores the set of inter-dependent decisions that firms face when introducing a new product into the market, namely capacity sizing, time to market or starting sales, initial inventory at launch, production management in response to changes in demand, and targeting and level of seeding and promotional activities. A diverse set of analysis tools, namely agent-based modeling and simulation, Monte Carlo simulation, continuous-time mathematical models, and parametric and nonparametric statistical approaches, are used to address the research questions posed in Chapter 1.

We show that, regardless of the primary performance measure, the optimal build-up policy from a deterministic analysis may not necessarily be optimal under stochastic supply and demand. We demonstrate the importance of considering risk when making such decisions by showing that the mean (expected) profit (the most commonly used metric by previous studies) can be flawed and misleading. We generalize previous findings on the optimal launch time for successive generations of innovations and show that the optimal market entry policy for the product's second generation varies continuously from simultaneous introduction with the predecessor (the *Now* policy) to not introducing at all (the *Never* policy). We also provide insight on the optimal combination of production capacity and build-up inventory for the new generation. The efficacy of the supply-constrained multi-generation model is empirically validated through its application to the case of Sony's PlayStation®3 game console. To the best of our knowledge, this is the first analysis of the supply-constrained multi-generation diffusion problem. Finally, we demonstrate the importance of a joint analysis of marketing and production strategies for new products by showing that determining seeding and build-up policies sequentially may lead to suboptimal decisions. The results also indicate the importance of the social network structure on the optimal combination of seeding and build-up policies by investigating the effect of long-range connections in small-world networks and distribution of hubs (high-degree nodes) in scale-free networks on the emergence of word-of-mouth and demand dynamics.

The stream of research performed in this dissertation aims at laying the foundation for more realistic decision-making tools that provide holistic and practical solutions for real-world problems faced by innovative companies. While we relax many of limiting and unrealistic assumptions of the previous studies (e.g., unlimited supply, exogenous demand, deterministic supply and demand, etc.), our models certainly do not capture every dimension of this complex problem. Important extensions to the proposed models include:

- *Forecasting error:* Our assumption is that accurate estimates of the actual diffusion parameters (i.e., innovation and imitation coefficients and market potential) are available to the decision maker. In the real world, this may not always be the case and a robust production-sales policy needs to be selected through sensitivity analysis on a range of possible parameter values to account for forecasting errors.

- *Competition:* Our models ignore competition by assuming a monopolistic market. While we discussed that this assumption does not undermine the main findings of this dissertation, the pressure from competitors may affect these decisions. Extending the proposed models to support an oligopolistic market would further enhance decision-making in real-world situations.

- *Social network structure:* While our experimental design covers a wide range of real-world networks, the question of how to estimate the social network structure still remains unanswered. While social networking websites provide some information about the individuals' connections, they do not represent all connections in the real-world. Therefore, there is a need for alternative approaches to calibrate this component of diffusion models.

We believe the findings of the work presented in this dissertation and its future extensions along the above lines would help companies increase the chance of a successful launch for their new products through appropriate and informed decisions about their supply and marketing strategies.

Appendices

Appendix A

Proof of Theorem 3.1

We use the Pontryagin's maximum principle to prove the optimality of selling at maximum possible rate for both product generations at any given time t. In order to make the proof easier to read, we omit t when representing functions of time (e.g., $f(t)$ is represented by f). The Hamiltonian of the optimal control model is given by

$$\begin{aligned} H(A,&S_1, S_2, a, L_1, L_2, W_1, W_2, I_1, I_2, s_1, s_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7, \lambda_8, \lambda_9, \lambda_{10}, t) = \\ &(\alpha_1 s_1 + \alpha_2 s_2 - h_1 I_1 - h_2 I_2)e^{-\theta t} + \lambda_1 a + \lambda_2 s_1 + \lambda_3 s_2 + \lambda_4(\frac{q}{N}(s_1 + s_2)(N - A) \\ &- a(p + \frac{q}{N})(S_1 + S_2)) + \lambda_5 l W_1 + \lambda_6 l W_2 + \lambda_7(d_1 - s_1 - lW_1) + \lambda_8(d_2 - s_2 - lW_2) \\ &+ \lambda_9(r_1 - s_1) + \lambda_{10}(r_2 - s_2), \end{aligned} \tag{A.1}$$

and the following are the system of equations for the adjoint variables λ_1, ..., λ_{10}:

$$\frac{d\lambda_1}{dt} = \frac{q}{N}(s_1 + s_2)\lambda_4 \tag{A.2}$$

$$\frac{d\lambda_2}{dt} = \frac{d\lambda_3}{dt} = \frac{q}{N}a\lambda_4 \tag{A.3}$$

$$\frac{d\lambda_4}{dt} = -\lambda_1 + \lambda_4(p + \frac{q}{N}(S_1 + S_2)) \tag{A.4}$$

$$\frac{d\lambda_5}{dt} = \frac{d\lambda_6}{dt} = 0 \tag{A.5}$$

$$\frac{d\lambda_7}{dt} = l_1(\lambda_7 - \lambda_5) \tag{A.6}$$

$$\frac{d\lambda_8}{dt} = l_2(\lambda_8 - \lambda_6) \tag{A.7}$$

$$\frac{d\lambda_9}{dt} = h_1 e^{-\theta t} \tag{A.8}$$

$$\frac{d\lambda_{10}}{dt} = h_2 e^{-\theta t} \tag{A.9}$$

$$\lambda_1(\infty) = \lambda_2(\infty) = ... = \lambda_{10}(\infty) \tag{A.10}$$

From (A.6) and (A.7) we immediately get $\lambda_7(t) = \lambda_8(t) = 0$. From (A.8) and (A.9) we have $\lambda_9 = -\frac{h_1}{\theta}e^{-\theta t}$ and $\lambda_{10} = -\frac{h_2}{\theta}e^{-\theta t}$. By differentiating (A.4) and replacing (A.2) into the resulting equation and using the final condition $\frac{\lambda_4}{dt}(\infty) = 0$, we have $\lambda_4(t) = 0$. By replacing $\lambda_4(t) = 0$ into (A.2) and (A.3) and using the final conditions $\lambda_1(\infty) = \lambda_2(\infty) = \lambda_3(\infty) = 0$, we get $\lambda_1(t) = \lambda_2(t) = \lambda_3(t) = 0$. Hence, the Hamiltonian becomes

$$H = (\alpha_1 s_1 + \alpha_2 s_2 - h_1 I_1 - h_2 I_2 - \frac{h_1}{\theta}(r_1 - s_1) - \frac{h_2}{\theta}(r_2 - s_2))e^{-\theta t}. \tag{A.11}$$

We know the optimal controls $s_1^*(t)$ and $s_2^*(t)$ maximize the Hamiltonian. Since $s_1^*(t)$ and $s_2^*(t)$ enter linearly into the Hamiltonian, the optimal policy follows a *generalized bang-bang* control. Let x and λ denote the state vector $(\alpha_1, \alpha_2, h_1, h_2, I_1, I_2, r_1, r_2)$ and the vector of adjoint variables $(\lambda_1, \lambda_2, ..., \lambda_{10})$, respectively. Then

$$H(x^*, s_1^*, s_2^*, \lambda^*, t) \geq H(x^*, s_1, s_2^*, \lambda^*, t), \qquad \forall t \in [0, \infty],$$
$$H(x^*, s_1^*, s_2^*, \lambda^*, t) \geq H(x^*, s_1^*, s_2, \lambda^*, t), \qquad \forall t \in [0, \infty].$$

Given positive coefficients for the two control variables, at any given time t, an extremal choice will be optimal for both controls, i.e., $s_1^*(t) \geq s_1(t)$ and $s_2^*(t) \geq s_2(t)$. By selling as much as possible, it is guaranteed that $I_i^*(t)W_i^*(t) = 0$ for all t. For product i, consider the following cases to determine the bound for $s_i^*(t)$. When $I_i^*(t) > 0$ (i.e., supply surplus during the first unconstrained diffusion phase), then the maximum sales is is equal to the demand rate $d_i^*(t)$. When $W_i^*(t) > 0$ (i.e, supply shortage during the constrained diffusion phase), $s_i^*(t)$ is bounded by the maximum production rate. Finally, in the case where $I_i^*(t) =$

$W_i^*(t) = 0$ (i.e., during the second unconstrained diffusion phase), $s_i^*(t)$ is bounded by $\min(r_i^*(t), d_i^*(t))$.

Appendix B

Closed-form expression for σ_3

$$\sigma_3 = \frac{\sqrt{\pi}\, m_2\, q}{\omega_2}\, (\omega_4 - \omega_5)\, \mathrm{e}^{\frac{(p + m_2\, q)(A_1\, m_1 + A_{CP}\, m_3) + N\, m_2\, p}{c_2}} \mathrm{e}^{\frac{A_1^2\, q\, (m_1^2 + m_2^2 + m_3^2) + A_{CP}^2\, m_3^2\, q + 2\, A_1\, m_3\, q\, (A_1\, m_2 + A_{CP}\, m_1)}{2\, N\, c_2}} \mathrm{e}^{\frac{N\, (p^2 + m_2^2\, q^2)}{2\, c_2\, q}}, \quad \text{(B.1)}$$

where $\omega_2 = \omega_3\, \mathrm{e}^{\frac{A_1\, m\, (p + m_2\, q)}{c_2}} \mathrm{e}^{\frac{A_1\, m\, q\, (A_1\, m_1 + A_{CP}\, m_3)}{N\, c_2}}$, $\omega_3 = \sqrt{2\, N\, c_2\, q}$, $\omega_4 = \mathrm{erf}\left(\frac{\omega_6}{\omega_3}\right)$, $\omega_5 = \mathrm{erf}\left(\frac{\omega_6 + c_2\, q\, t}{\omega_3}\right)$, and $\omega_6 = N\, (p + m_2\, q) + A_1\, q\, u + A_{CP}\, m_3\, q$.

Also, $\mathrm{erf}(x)$ is twice the integral of the Gaussian distribution with mean 0 and variance of 0.5 as given by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} dt$$