# FLEXIBLE MODELS OF INTEGRATED MARKETING COMMUNICATIONS EFFECTS

Ceren Kolsarici Nalca

Desautels Faculty of Management

McGill University

Montreal, QC, Canada

September 8, 2009

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

# DEDICATION

To my lovely grandmother: Turkan Aytekin,

for always watching over me  –even when she's gone- ;

&

to Arcan Nalca,

for his unconditional love and support.

# ACKNOWLEDGEMENTS

Einstein once said, “Any fool can know; the point is to understand.” I did not even *know* enough when I first came to McGill five years ago; and I do not think I would have ever *understood* without the strong support and encouragement of the people I owe a great deal of appreciation to.

First and foremost, I would like to thank Demetrios Vakratsas, who has been my advisor, my mentor and my friend during my years at McGill University. Since the first time I knocked on his door long ago, in August 2004, and introduced myself – with a fair amount of diffidence, now that I confess-, Demetrios has been there for me at each step with his guidance, support, and valuable inputs. He has not only given me a strong foundation of quantitative methods in marketing but also encouraged me to improve my skills further, push my boundaries and learn new methodologies. I feel myself very lucky that, from the first day I stepped foot in the academic world, I had the chance to see closely, the ambition, hard work and diligence Demetrios possesses in his career. I thank him for never stopping challenging me, for always believing in me and for being the inspiration that he is. I cannot think of a better advisor and a role model than him.

I would also like to express my deepest gratitude to Prasad Naik, who acted as an external member of my thesis committee. Prasad generously provided me with precious ideas and feedback whenever I needed help. My conversations with him thought me a great deal about asking the right questions when doing research, positioning your work correctly into the related research stream as well as focusing on how to make an impact and a substantial contribution. I was astounded with his knowledge and dedication to marketing research every moment I spent with him. I thank Prasad for his support, his valuable insights as well as for accepting to make a conference call for my thesis proposal at 6am PST without hesitation.

I am deeply thankful to my other dissertation committee members, Emine Sarigollu, Georges Zaccour, Tamer Boyaci and Renaud Legoux for careful reading of the chapters and for their constructive feedback in addition to their very much appreciated presence at my defence.

I do not think I would have enjoyed the experience at McGill as much, if it wasn’t for the wonderful people of the Desautels Faculty of Management. I would like to take this opportunity to thank my friends in the Ph.D. community, Alice Labban, Caroline Roux, Robert Mackalski, Bo-Young Chang, Sarah Moran, Burcin Col and Chayawat Ornthanalai, for the fun times we had together - the fiestas, volleyball tournaments, spa trips, many night outs-, and more importantly for making me feel like a part of a big family with their heartfelt friendship. I will never forget the kindness of my office mates, Xiaoye Chen and Yanan Wang as well as the great assistance of Stella Scalia.

I have not once believed in distances; and for the last five years, the people I have felt closest to my heart were not less than an ocean away from where I was. I owe a lot to Ongun Saracbasi; I do not know how he did it from France, but he managed to make me feel his support everyday. When he wasn’t in the yellow box blinking on my taskbar, he was talking to me through my cell phone, or at some rare but precious times, he was right next to me sharing a unique Montreal moment. Special thanks goes to my dearest friends Berna Tari and Alev Kuruoglu for listening, sharing and making me laugh even at the most depressing moments, you made my days brighter from miles away; Deniz Tansel for our long inspiring conversations and her encouragements that help me go on no matter what; Nilay Donmez for being there for me for 18 years; and my childhood friends Ipek Yokus, Simay Batum and Sibel Acikel for keeping their promise: CISS forever!

I am more than grateful to my family as they have been with me through whatever challenges life might have brought. I thank my parents Nuray and Ozer Kolsarici for always encouraging me to have dreams and go after them, my aunt Tulay Aytekin for her cheerful presence, and my sister Sebnem Kolsarici for just being the way that she is; smart, loving, honest and fun – I wouldn’t be the same person without you.

Looking back in time, there was one magical person in my life, Arcan Nalca, who has managed to be everything to me, all at once… my colleague, my co-author, my best friend, my family and my love… At times, I felt like he believed in me more than I believed in myself – so glad that he did. Arcan Nalca, for your rock steady support, your patience with me, and your unconditional love, I dedicate this work to you. Thank you for being where you are - right next to me, always.

Ceren Kolsarici

# ABSTRACT

This thesis comprises three essays and investigates complex effects of integrated marketing communications, using advanced statistical and econometric models. The first essay focuses on the measurement of complex multi-media communications effects such as thresholds, saturation levels and cross-media synergies. We use, MARS, a non-parametric regression method based on multivariate adaptive splines, and show that it, successfully trading-off the bias reduction and variance increase, performs superior to parametric and non-parametric benchmarks in model fit and predictive validity. The results provide compelling evidence to one or more threshold points, saturation levels, early saturation for newspaper advertisements and support for possible supersaturation for certain media. Moreover, we quantify the observed threshold and saturation levels using non-parametric derivatives and find that majority of the media perform in inefficient spending ranges.

The second essay examines the dynamic effects of direct-to-consumer advertising (DTCA) in a market where regulations impose restrictions on the type and content of prescription pharmaceutical advertising. We identify three research questions that should be of great managerial interest: *Whether* DTCA is a reasonable option to choose under these regulations. If so, *which* type of DTCA is more effective, and *when*? We pursue these questions by examining data on new and refill prescriptions for a novel pharmaceutical through the

implementation of an Augmented Kalman Filter with continuous state and discrete observations (AKF(C-D)). Our findings suggest the presence of complex DTCA dynamics for the two types of regulation-induced advertising messages. We discuss implications and provide extensive validation tests that confirm the superiority of our modeling approach.

The final essay investigates the influence of market heterogeneity on the consumer and physician directed marketing communications effectiveness, and the diffusion pattern for a new prescription pharmaceutical. Based on clinical management and pharmacoepidemiological concepts that consider severity of health problems and the medical practice of watchful waiting, we propose a dual market model for the diffusion of new prescription pharmaceuticals. The model distinguishes between an “early” adoption market corresponding to prescriptions for patients with severe health problems for which demand is accumulated prior to the pharmaceutical’s launch and a “late” market, corresponding to prescriptions for patients with mild problems, which is developed after the product’s launch and potentially triggered by it. Empirical application to monthly data on new prescriptions and corresponding marketing activities for a new prescription category, representing the first oral disease treatment, suggested that the proposed model has good parameter face and forecasting validity, and that marketing communications affect the two distinct markets differently. Implications and avenues for future research are discussed.

# ABRÉGÉ

Cette thèse comprend trois essais et examine les effets complexes des communications marketing intégrées, utilisant des modèles statistiques et économétriques avancés. Le premier essai se concentre sur la mesure des effets complexes des communications multimédia comme les seuils minimums, les niveaux de saturation et les synergies des médias croisés. Nous utilisons, MARS, une méthode de régression non paramétrique basée sur des courbes adaptatives multivariables, et ce qui démontre qu'équilibrant avec succès la réduction de l'erreur moyenne et de l'écart de l'augmentation, MARS s'exécute mieux aux points de référence paramétriques et non paramétriques dans l'ajustement du modèle et la validité prédictive. Les résultats fournissent la preuve irréfutable d'un ou plusieurs points de seuil minimum, de niveaux de saturation, de la saturation précoce pour les publicités dans la presse écrite et d'un appui pour une possible sursaturation de certains médias. De plus, nous évaluons quantitativement le seuil observé et les niveaux de saturation en utilisant des dérivés non paramétriques et constatons que la majorité des médias s'exécute dans des gammes de dépenses inefficaces.

Le deuxième essai examine les effets dynamiques de la publicité directe au consommateur (DTCA) dans un marché où les règlements imposent des restrictions sur le type et le contenu de la publicité pour les prescriptions

pharmaceutiques. Nous identifions trois questions de recherche qui devraient être de grand intérêt en gestion, c'est-à-dire : Si la DTCA est une option raisonnable à choisir conformément à ces règlements ? S'il en est ainsi, quel type de DTCA est le plus efficace et quand ? Nous poursuivons ces questions en examinant des données sur les nouvelles prescriptions et les renouvellements de prescriptions pour un nouveau médicament par la mise en oeuvre d'un Filtre Kalman Augmenté avec état continu et observation discrète (AKF (C-D)). Nos découvertes suggèrent la présence de dynamiques de DTCA complexes pour les deux types de messages publicitaires avec réglementation induite. Nous discutons des implications et fournissons les tests de validation approfondis qui confirment la supériorité de notre approche de modélisation.

L'essai final examine l'influence de l'hétérogénéité du marché sur le consommateur et l'efficacité des communications marketing adressée au médecin, et le modèle de diffusion pour une nouvelle prescription pharmaceutique. Basé sur la gestion clinique et les concepts pharmaco-épidemiologiques qui considèrent la sévérité de problèmes de santé et la pratique médicale de surveillance étroite, nous proposons un modèle de marché dual pour la diffusion de nouveaux produits pharmaceutiques de prescription. Le modèle fait une distinction entre « un premier » marché d'adoption correspondant aux prescriptions pour des patients avec des problèmes de santé sévères pour

lequel la demande s'accumule avant même le lancement du produit pharmaceutique, et un marché « tardif », correspondant aux prescriptions pour des patients avec des problèmes légers, qui est développé après le lancement du produit et potentiellement déclenché par cela. L'application empirique aux données mensuelles sur de nouvelles prescriptions et les activités de marketing correspondantes à une nouvelle catégorie de prescription, représentant le premier traitement oral d'une maladie, ont suggéré que le modèle proposé ait de bons paramètres de la validité apparente et de la validité des prévisions, et que les communications marketing affectent les deux marchés distincts différemment. Les implications et les avenues pour la recherche future y sont traitées.

# CONTRIBUTION OF AUTHORS

The three essays in chapters two to four of this dissertation are written in collaboration with Demetrios Vakratsas, who is an associate professor of Marketing, in Desautels Faculty of Management at McGill University. Professor Vakratsas has acted as the supervisor of this thesis for five years.

For essays in chapters two and three, Ceren Kolsarici performed the data management, programming, estimation and analysis that make up the chapters. Ceren Kolsarici and Demetrios Vakratsas developed the main research ideas behind the chapters and discussed the revisions of the chapters.

For the last essay in chapter four, Demetrios Vakratsas developed the main research idea. Ceren Kolsarici performed the data management, and analysis that comprise the chapter. Ceren Kolsarici and Demetrios Vakratsas discussed the revisions of the chapter.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

The American Association of Advertising Agencies (AAAA) defines Integrated marketing communications (IMC) as the “concept of marketing communications planning that recognizes the added value of [a] comprehensive plan that evaluates the strategic roles of a variety of communication disciplines…and combines these disciplines to provide clarity, consistency, and maximum communications impact (Belch & Belch, 1999).” IMC ties together all of the company’s promotional mix and media communications messages, so that a consistent brand image and positioning are conveyed (Schultz, Tannenbaum, & Lauterborn, 1993). Short after its birth in early 1990s, the majority of the companies and advertising agencies embraced the concept of IMC and the idea of integrating all their marketing efforts.

The changes the marketplace has gone through since 1990s, such as consumer and media fragmentation, and increased competitive intensity, only added to the importance of IMC and measuring its effects. In this highly challenging business environment with a plethora of marketing tools available, correctly understanding the consequences of their actions is more valuable than ever for the managers. Indeed, the Marketing Science Institute (MSI) listed IMC effect measurement in its top research priorities for the period 2008-2010[1].

This collection of essays focuses on measuring the complex effects of IMC campaigns. Specifically, we try to accurately quantify the intricate effects of simultaneous marketing activities on firms' market performance measured by sales using advanced/sophisticated methodologies. In this thesis, an IMC campaign is operationalized as a set of concurrently employed multiple marketing tools, such as direct marketing, personal selling and/or advertising through multiple media channels, assuming the messages delivered through various channels are consistent.

In this chapter, we will go over the issues related to IMC effect measurement, present the research approach followed in the dissertation and briefly introduce the three essays.

---

[1] Every two years, research priorities list is prepared based on discussion groups held with MSI trustees and chief marketing officers and quantitative surveys sent to all MSI member company trustees.

### 1.1 IMC Effect Measurement

This section introduces the measurement of integrated marketing communications effects. We start with a brief overview of the promotional mix; and continue with the importance of measuring IMC effects. Next, we shortly discuss how marketing campaigns are evaluated by practitioners, followed by related studies in the marketing literature. We finish this section with the specific substantive and methodological research issues that need to be addressed for a comprehensive and accurate analysis of IMC effects.

#### 1.1.1 Overview of Promotional Mix Effects

A company's promotional mix, also known as the marketing communications mix, consists of five elements, namely, advertising, sales promotion, public relations, personal selling and direct marketing (Kotler, Armstrong, & Cunningham, 2007). These, essentially, are tools for the company to use in pursuit of its objectives, mainly to reach its consumers. Advertising includes print, broadcast, internet and outdoor; sales promotions includes point of purchase displays, premiums, discounts, coupons; public relations includes press releases and sponsorships; personal selling includes sales presentations, and incentive programs; and direct marketing includes catalogues, telephone marketing, internet and such.

The effects of the promotional mix are identified below. We distinguish between consumer-level effects and market-level effects. Although in this thesis the focus will be on the market-level effects, consumer responses deserve mentioning as well for the sake of completeness.

Consumer-level effects

- *Brand awareness*: Extent to which a brand or brand name is recognised by potential consumers, and correctly associated with the particular product category.
- *Category awareness*: Extent to which a new product category is recognized by consumers.
- *Adoption time acceleration*: Adopting the product sooner.
- *Brand loyalty*: Consistently preferring a particular brand over the others.
- *Brand switching*: A purchasing pattern characterized by a change from one brand to another.
- *Category switching*: A purchasing pattern characterized by a change from one product category to another.
- *Price sensitivity*: Price sensitivity measures the range of prices consumers are willing to pay for a product or service. Marketing activities can increase or decrease the price sensitivity of consumers depending on the firm's strategies and competitive environment.

- *Repeat purchasing*: A consumer purchasing the same brand as in the previous purchase occasion.
- *Purchase quantity acceleration*: Purchasing in larger quantity than usual in a specific purchase occasion.
- *Stockpiling*: A consumer having a higher stock at hand due to timing or quantity acceleration

Market-level effects

- *Performance stability*: Consistent increase in stock market value, sales and unwavering market share figures
- *Increased market share*
- *Increased sales*
- *Hysteresis*: a case when a temporary increase in advertising expenditures results in a permanent increase in the base sales level (Simon, 1997).

### 1.1.2 Importance of IMC effect measurement

As previously mentioned, the Marketing Science Institute (MSI) included measurement of IMC effects in its list of research priorities for the period 2008-2010. We discuss a number of reasons leading to this development.

Despite the negative effects of the 2001 economic recession in US, the amount of marketing spending has reached to significantly large numbers since early 2000s. Statistics show that the 100 Leading US Advertisers (LNA) spent a cumulative $150 billion in advertising in 2007, which marks a 6.37% increase from 2005 and a vast 4900% increase in the last 50 years[2]. With so much at stake today, possibly more than ever, companies need to correctly measure the consequences of their marketing communications.

The proliferation of new media is another reason driving attention to IMC measurement. MSI reports show that most high level managers emphasize their need to understand how to use alternative media, such as networking sites, blogs, mobiles and electronic billboards. Although these forms of communications only became available in 2002, lead by blog advertising, by the end of 2005 combined advertising spending on the top 3 online media -blog, podcast and RSS- had grown to $20.4 million, a 198.4% increase over the 2004 level[3]. Moreover, the same research shows that user-generated media, such as blogs, RSS, and podcasts, is forecasted to grow with a compound rate of 106.1% from 2005 to 2010, reaching $757.0 million by 2010. Two key areas of interest related to new media are how to allocate the promotional budget among these alternative media, and how to evaluate the effectiveness of the resulting spending.

---

[2] Source: Advertising Age Report on 100 Leading National Advertisers, January 23, 2008

[3] Source: PQ Media, Alternative Media Research Series I, April 2006.

Another related question shared by the MSI member companies is the role of the traditional media including TV, print and radio in this new communications environment. Although, driven by the continued audience fragmentation and the desire of companies to reach elusive younger markets, old media have started to be perceived as ineffective, they still accounted for around 96 % of the total ad spending in 2005.

### 1.1.3 IMC Effect Measurement in Practice

Accurate and comprehensive measurement of individual promotional mix effects, let alone analysis of simultaneous multiple marketing efforts, is quite limited in the corporate world, evident from the growing interest of managers in collaborating with external market research companies (e.g. AC Nielsen and IRI) and/or in organizations, such as MSI, dedicated to bridging the gap between marketing research and business practice.

Many reasons explain the lack of IMC effect measurement in practice. Most often, the rapid changes in the business environment hardly leave any time for managers to evaluate their actions.

In certain cases, management is ambitious about effect measurement but limited by data related problems. Just like in the case of insufficient data, abundance of data could lead to problems. Great detail and a high number of variables usually makes data management harder and driving any insights almost impossible for the practitioners. Increasing popularity of user-friendly software among companies to help make more credible decisions, also does not lead to much improvement in the quality of the marketing communications programs because of their inflexibility. Hence, rather than flexible methods and reliable quantitative measures which require a solid knowledge of econometrics, statistics, data management, and model building, the majority of the decisions regarding IMC programs have to be made based on managerial insights and competitors' actions.

### 1.1.4 IMC Effect Measurement in Marketing Literature

The relationship between advertising, pricing, promotions, and purchasing behaviour such as, brand choice, sales and market share has long been investigated by marketing researchers through market response models which could be classified into aggregate-level and individual-level models.

We will not provide a complete overview of literature on promotional mix effectiveness in this section. However, we would like to refer interested readers to helpful publications on generalizations regarding specific promotional mix elements. Vakratsas and Ambler (1999) offer a very comprehensive discussion of how advertising works, including a taxonomy of models and empirical generalizations. Blattberg, Briesch and Fox (1995) provide a summary of more than fifty articles on sales promotions. Bijmolt, Van Heerde and Pieters (2005) present an extensive discussion of new empirical generalizations on the determinants of price elasticity. Finally, quantitative generalizations on personal selling's effects are provided by Albers, Mantrala and Sridhar (2008) based on 46 empirical studies published in the last 40 years.

Research has shown one fact common across different marketing efforts; their effects separately could exhibit irregular and complex patterns. For example, Van Heerde, Leeflang and Wittink (2001) demonstrate high complexity for deal effects, due to thresholds and saturation levels. Similarly, advertising response has also been shown to exhibit thresholds (Bemmaor, 1984; Vakratsas, Feinberg, Bass, & Kalyanaram, 2004) and saturation effects (Hanssens, Parsons, & Schultz, 2001) requiring a flexible representation.

Despite the abundance of market response studies in the literature, frequently researchers focused on complex effects of a particular element of the promotional mix while assuming simple forms of effects for the rest of the marketing mix or ignoring them altogether. The main reason leading to this has been methodological limitations. We will discuss these in details in subsection 1.1.6; however, we will start with the substantive issues that are important for the comprehensive understanding of integrated marketing communications effects.

### 1.1.5 Important Issues to Consider in IMC Research

We list the following set of issues, the majority of which have not been sufficiently addressed in marketing literature. These are critical to IMC research and need to be acknowledged for a broad understanding of its effects.

- Irregular Market Response: The shape of marketing communications effects have been investigated by numerous researchers for various elements of the promotional mix, such as advertising (Bemmaor, 1984; Simon & Arndt, 1980; Vakratsas et al., 2004) and sales promotions (e.g. Van Heerde et al., 2001 for the shape of deal effects). There have been conflicting findings regarding the exact shape of the effects; for instance advertising response has been either shown to be concave(Simon & Arndt, 1980) or S-shaped (Vakratsas et al. 2004) by different researchers. However, academics agree that the response functions more often exhibit irregular and complex patterns. When investigating IMC, it is, therefore, important to build flexible enough models to capture the complexity of IMC effects.

- Competition: Over the years, competition has been claimed to be one of the most critical factors that affect the performance of many firm activities. Competitive intensity and the amount of competitive activity in particular, have significant influence on promotional mix effectiveness (Danaher, Bonfrer, & Dhar, 2008). Competitive interference increases clutter and, consequently, limits consumer attention to advertising (Webb, 1979; Webb & Ray, 1979), which may result in lower advertising effectiveness. Hence, it is essential to control for competitive efforts to better understand IMC effects.
- Dynamics: Marketing research over the years has suggested that promotional mix effects, specifically advertising, possibly follow a dynamic

pattern as a result of many influencing factors such as stage of product life cycle (PLC) (Parsons, 1975; Winer, 1979), message content, copy wear-out (Bass, Bruce, Murthi, & Majumdar, 2007; Naik, Mantrala, & Sawyer, 1998), quality restoration and copy replacement. Therefore, the dynamics of promotional mix effectiveness need to be recognized in the analysis of IMC effects.

- Media Synergies: Although the main theoretical argument behind IMC has been the concept of synergy, which suggests that the total effect of multiple simultaneous marketing activities would be greater than the sum of its parts, research focusing on synergistic effects of marketing efforts has been rather limited. Two of the few exceptions is Naik and Raman (2003) and Prasad and Sethi (2009), the former of which operationalized synergies through an interaction term directly included into the model. However, synergistic effects are quite likely to be more complex than what one can capture by a constant interaction term. Hence, cross-media synergies are possibly the most interesting and the least researched phenomenon regarding integrated marketing campaigns.

- Message Content: Content specific differences are also recognized to play a significant role in marketing communications effects. For instance, Chandy et al. (2001), relying primarily on behavioral theories, argue, and find, that informative messages are especially effective in new markets and for recently introduced products. However, as markets and products mature, emotional messages become the more effective alternative, suggesting their ability to produce persuasive effects (Becker & Murphy, 1993; Comanor & Wilson, 1974; Marshall, 1919) later on in the product life cycle. MacInnis, Rao and Weiss (2002) similarly concluded that frequently purchased brands in mature product categories are better off differentiating by using warmer and more likable messages based on affective executional cues, rather than relying on product-based information.
- Market Heterogeneity: The audience profile is an important determinant of the marketing effectiveness. Consumers with varying needs, wants and priorities are not affected the same way from the same message, which will have an effect of the shape of the market response. Hence, IMC

effects would certainly be understood better, once these market specific differences are acknowledged.

### 1.1.6 Methodological Challenges in IMC Effect Measurement

Some of the main methodological problems related to some of the aforementioned issues and the accurate measurement of integrated marketing communications effects are discussed below (mostly based on Leeflang, Wittink, Wedel, & Naert, 2000).

- Overparametrization: The more marketing tools a firm uses to promote its offerings, the greater the number of predictor variables a parametric model will need to capture the comprehensive effects of these activities. If one wishes to consider the possible interactions between all different marketing actions in a parametric specification, both the estimation and the interpretation of the models become problematic. In general, with K predictors, a full two-way interaction model consists of $2^K$ terms which, apparently, would diminish a model's ability to reflect reality. Thus, the majority of the time, the model builder has to specify the interaction variables to include in the model, and this action greatly undermines the model's flexibility and is likely to introduce selection bias.

- Curse of dimensionality: The most common methodological remedy to study complex marketing mix effects is to use semi/non parametric estimation techniques, such as spline and kernel methods(Abe, 1995; Kalyanam & Shively, 1998; Van Heerde et al., 2001). These techniques, due to their flexibility, enable capturing of the irregular effects of marketing efforts. However, with each additional predictor the dimension of the problem increases, the number of possible interaction effects explodes, and the required sample size increases drastically. Similar to the case of overparametrization in parametric models, this fact, referred to as the "curse of dimensionality" in the literature, affects the performance of most non-parametric models in addressing high dimensional problems.
- Bias Variance Trade-Off: A common issue modellers face in capturing irregular effects of marketing efforts is the bias-variance trade off. It is a known fact that models with less complexity and/or fewer parameters are subject to larger bias (due to limited flexibility) while models with high complexity and/or too many parameters are subject to larger variance

(due to high dependence on the sample). Although high complexity models achieve impressive in-sample predictions, their performance suffers considerably in out-of sample forecasts. Therefore, it is crucial to build models with optimal model complexity to minimize prediction errors in both training and test samples.

- Time interval bias: Another problem source in modeling of marketing effects is the discretization of continuous processes for estimation driven by the data being a collection of observations that are aggregated over certain time intervals. Most often, conventional estimation methodologies, such as OLS, require discrete approximation of continuous models which leads to a well-documented problem of time interval bias (Mahajan, Muller, & Wind, 2000; Putsis, 1996).
- Uncertainty: Although the data employed in any empirical study is subject to a certain amount of measurement error, this noise is not accounted for by most classical models. Explicit consideration of observation noise would undoubtedly result in more efficient estimates. In a similar fashion, the non-constancy of model parameters over time plays role as another factor of uncertainty. As a remedy to this problem, stochastic control

theory offers the use of dynamic stochastic measurement functions to determine whether model parameters change systematically or gradually (ie. Slow stochastic variation) over time.

These aforementioned points, although constituting a small part of the existing difficulties in understanding complex IMC effects, underline the need for advanced/sophisticated methodologies.

## 1.2 Research Approach

Given the proliferation of alternative media, and the changes in the communications environment, IMC management is crucial and could only be achieved by careful measurement of its effects. The purpose of this thesis is to develop models to accurately measure integrated marketing communications effects on market performance and provide insights for managers to improve their decisions and better allocate their recourses.

In this dissertation, we focus on aggregate-level market response models to study complex integrated marketing communications effects including multi-media thresholds, saturation points, synergies and dynamics. We use flexible and adaptive models, including non-parametric techniques and Kalman filtering, to adequately capture the complexity of IMC effects and provide better long-run forecasts of market response. This section provides a brief overview of the structure of the thesis, which consists of three essays (Figure 1.1). The categorization of the essays with respect to the issues discussed in the previous section can be found in Table 1.1.

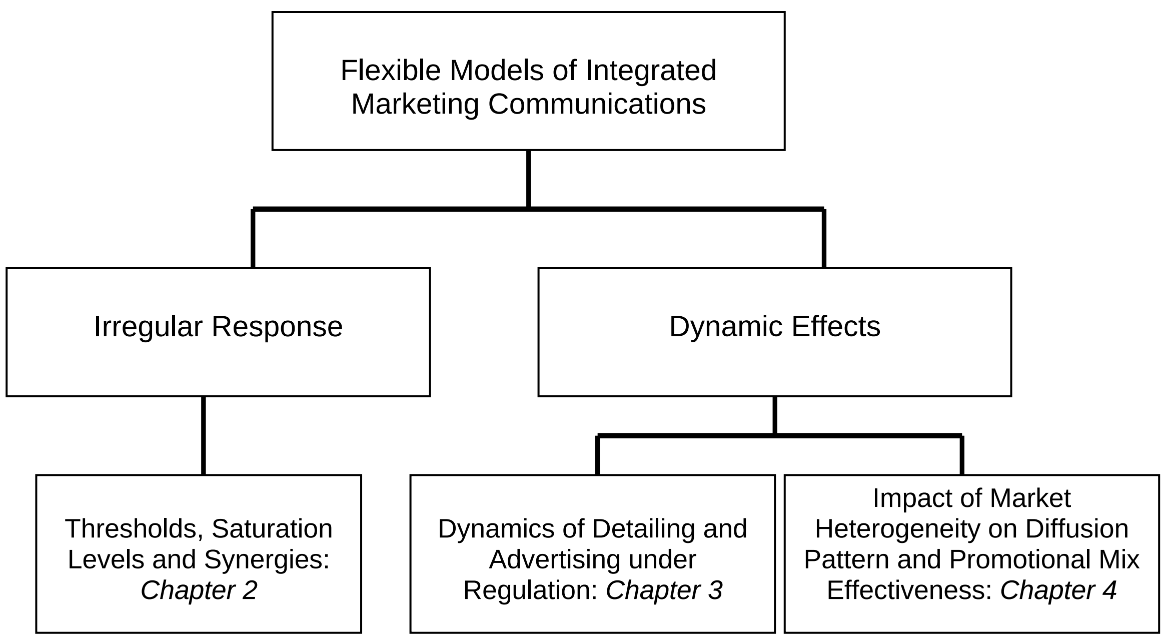



Figure 1-1: Thesis Framework

Chapter 2 focuses on measuring complex effects of multi-media communications including thresholds, saturation levels and cross-media synergies (see Kolsarici & Vakratsas, 2009). We address the issue of high dimensionality in the measurement of IMC effects by introducing a non-parametric regression method based on multivariate adaptive splines and comparing it to the existing alternatives in the marketing literature. We show that being highly adaptive and flexible, MARS, has several advantages over other parametric and non-parametric methods in dealing with multi-media problems, in terms of providing both accurate estimation of high dimensional response surfaces and reliable in sample and long-range forecasts of market performance. We also summarize the common response shapes to different media efforts across different data sets. We provide compelling evidence for threshold and saturation effects for multiple media, existence of multiple thresholds for a single medium, early saturation for newspaper advertisement, and some support for the possible supersaturation of magazine ads and cable TV ads. We also quantify the observed threshold and saturation levels using non-parametric derivatives; and show that the majority of the media do not operate in an efficient spending range, evident from the average saturation levels of 50% across media, confirmed for different product categories.

In this chapter, we aim to accurately measure the complex effects of multi-media communications and derive implications critical for the budgeting and allocation decisions. This study can lead the way for the optimization of the multi-media communications budgets, which is an interesting and relevant extension, and will be left for future research.

Table 1-1: Overview of Research Issues Studied in the Thesis

| | | Chapter 2 | Chapter 3 | Chapter 4 |
|---|---|---|---|---|
| SUBSTANTIVE ISSUES | Irregularities in Response | X | | |
| | Media Synergies | X | | |
| | Dynamics | | X | X |
| | Competition | | X | |
| | Message Content | | X | |
| | Market Heterogeneity | | | X |
| METHODOLOGICAL ISSUES | Overparametrization | X | | |
| | Curse of Dimensionality | X | | |
| | Bias-Variance Trade-Off | X | | |
| | Time Interval Bias | | X | X |
| | Uncertainty | | X | |

Chapters 3 and 4 investigate the promotional mix effects in the pharmaceutical context. Both essays utilize the same data set consisting of monthly market-level information on the number of prescriptions and matching marketing mix information for a recently developed therapeutic category, provided by one of the competing firms, in a non-U.S. market. Due to confidentiality reasons, the therapeutic category and market cannot be disclosed. However, the category concerns a new therapeutic class developed for the treatment of a lifestyle-related disease.

Chapter 3 examines the dynamics of direct-to-consumer advertising (DTCA) in a market where regulations impose considerable restrictions on the type and content of prescription drug advertising, allowing for brand only (reminder) or category only (generic) types of DTCA ads (see Kolsarici & Vakratsas, 2008). We investigate three managerially interesting issues: Is consumer directed advertising a reasonable option to choose under these regulations (whether)? If so, which type of DTCA is more effective? And when? These issues are addressed by analyzing data on new and refill prescriptions for a novel pharmaceutical product through the implementation of an Augmented Kalman Filter with continuous state and discrete observations (AKF(C-D)), a combination of Extended Kalman Filter (EKF) and adaptive filter. The findings suggest the presence of strong advertising dynamics as well as differences in the effects on new and refill prescriptions for two types of regulation-induced advertising messages.

Chapter 4 investigates the market-level effects of the promotional mix decisions for the diffusion of a new prescription pharmaceutical by taking into account the idiosyncratic elements of pharmaceutical markets (see Vakratsas & Kolsarici, 2008). A dual-market modeling framework is proposed, where an “early” adoption market, corresponding to patients with severe health problems, is followed by a “late” market, corresponding to patients with mild health problems. The early market may be formed prior to the launch of the pharmaceutical due to well-defined, diagnosed needs of the patients corresponding to this market, and the accumulated demand by this market is realized only when the prescription drug is eventually launched.

An empirical application to category-level monthly data on new prescriptions and corresponding marketing activities for a new drug, representing the first oral disease treatment, confirms the dual-market hypothesis, suggesting that the early market shows an exponential-like pattern and is not affected by marketing activities. The late market, on the other hand, follows a Bass-type adoption pattern and is significantly influenced by physician and patient-directed advertising. The results confirm the presence of differential effects of marketing mix decisions on two distinct markets with diverse characteristics and the advantages of flexible, custom-fit, approaches to better understand IMC effects.

We conclude the dissertation with Chapter 5 where we summarize the three essays, talk about the contributions of this research and managerial implications.

# CHAPTER 2

# INVESTIGATING COMPLEX EFFECTS OF MULTI-MEDIA COMMUNICATIONS

## 2.1 Introduction

American Association of Advertising Agencies (AAAA) define Integrated marketing communications (IMC) as the "concept of marketing communications planning that recognizes the added value of [a] comprehensive plan that evaluates the strategic roles of a variety of communication disciplines…and combines these disciplines to provide clarity, consistency, and maximum communications impact (Belch & Belch, 1999)." The effectiveness of IMC spending is a top priority for practitioners and academics[4]. A major contributing factor for the growing importance of this issue is the proliferation of new media and forms of communication such as online search, blogs, electronic billboards, community websites (e.g. facebook, myspace, twitter) and advergames. Thus, while new media present marketers with more options, they also prompt

[4] MSI 2008-2010 research priorities list

questions regarding the effectiveness and contribution of each medium employed (Havlena, 2008).

The evaluation of IMC spending, specifically multi-media communications, is a complex task due to the richness of the effects advertising can generate. For example, previous research has suggested that advertising effects possibly exhibit threshold and saturation levels (Bemmaor, 1984; Vakratsas et al., 2004). In addition, IMC campaigns frequently lead to synergies (Naik & Raman, 2003), resulting from the simultaneous use of different media. Thus, considering such intricate phenomena in evaluating the effects of multi-media IMC campaigns requires highly complex market response models, due to the plurality of media utilized, and the irregularity of the effects they can produce. This implies that the "blessing" of more media options to marketers as means of reaching their markets, can easily turn into a "curse" when it comes to assessing their productivity. Put in mathematical terms, the use of an increasing number of media increases the *dimensionality* of the problem of evaluating their effectiveness.

This paper addresses the issue of high dimensionality in the measurement of IMC spending effects by introducing a flexible modeling approach and comparing it to the existing alternatives in the marketing literature.  Specifically, we propose to use MARS, a non-parametric estimation methodology -based on multivariate adaptive splines - novel in the marketing literature, to simultaneously model multi-media communications and analyze cross-media synergies in the presence of possible threshold and saturation effects for media spending. As will be discussed in the Methodology section, MARS is highly adaptive and flexible, and has several advantages over existing parametric and non-parametric methods in dealing with multi-media problems, in terms of providing both accurate estimation of high dimensional response surfaces and reliable long-range forecasts of market performance. The implications of the findings on the budgeting and allocation decisions are also very critical however; our goal here is not to optimize IMC campaign allocations. Rather, we aim to accomplish accurate measurement of complex effects of multi-media communications, which is of crucial importance for the investigation of the optimal budget allocation across different media.

Our intended contribution is threefold. First, we address the problem of high-dimensionality in evaluating multi-media communications by flexibly estimating complex IMC effects including irregular main effects for multiple media and cross-media synergies. We provide empirical evidence to phenomena such as multiple thresholds, early and oversaturation with an aim to help generalize complex effect shapes that might vary across product categories and media. Second, we quantify the observed irregularities such as thresholds and saturation levels using non-parametric derivatives and show that much media spending lies in inefficient ranges for many brands. Third, we demonstrate that MARS, when compared to Kernel-based non-parametric methods, demonstrates superior performance in terms of prediction and long range forecasting of market response.

This chapter is structured as follows. First, we summarize the relevant literature with respect to complex market response to advertising and multi-media effects. We then introduce MARS, our proposed methodology, and discuss its advantages over other prevalent non-parametric methods. Next, we move to the empirical analysis where we first talk about the modeling framework, present the data sets, and then discuss the main results. We finish with the conclusion and discussions.

## 2.2 Related Literature

Many firms and advertising agencies have embraced the concept of IMC since its inception, and with the proliferation of new media, marketers now have more alternative means to promote their products. Despite the plethora of the market response studies, work on IMC effectiveness has been scarce in the literature. Although interactions between marketing efforts are emphasized as the basis of the concepts such as *marketing-mix* and *communications-mix* (for an extensive review see Gatignon, 1993; e.g. Gatignon & Hanssens, 1987; Naik & Raman, 2003), the overall effectiveness of an integrated marketing communications campaign, acknowledging the simultaneous effects of cross-media interactions and irregularities of market response, has not been addressed.

Marketing researchers have argued and shown that response to each marketing mix element such as advertising (Vakratsas et al., 2004), pricing (Kalyanam & Shively, 1998) and promotions (Van Heerde et al., 2001) separately can exhibit quite a complex pattern that cannot be captured by simple mathematical representations, but rather require flexible modeling and estimation techniques that are able to deal with the intricacies of the effects. Specifically, the effects of advertising on market response have been argued to exhibit threshold levels (Bemmaor, 1984; Simon & Arndt, 1980; Vakratsas et al., 2004) and saturation effects (Hanssens et al., 2001). Advertising threshold theory suggests that the market may be insensitive to advertising when carried out at low levels and responds to it only after a certain positive amount (i.e. defined either in terms of spending or awareness measured by GRPs) is invested. Vakratsas et al. (2004) provide empirical evidence for the fact that the market response to advertising is not necessarily globally concave and advertising thresholds indeed exist particularly for evolving product categories, putting to rest one of the longest debates in the advertising literature. Moreover, experimental studies on advertising response found evidence for the positive effects of decreased advertising levels, leading to patterns including V-shaped response and bi-modal M-shaped response (e.g. Ackoff & Emshoff, 1975; Hahn, Park, & Macinnis, 1992). Hence, it is very important to consider the complex effects for all the different media utilized in an IMC campaign.

Despite the proliferation of the media alternatives in marketing practice, to a large extent, marketing science literature has mainly focused on the effects of total or single-medium advertising spending rather than multi-media effects, frequently ignoring the dynamics and synergies between marketing actions. One of the few exceptions, is the recent study by Naik and Raman (2003) who explicitly investigate the synergies between TV and print advertising by employing Kalman Filtering methodology. The authors conceptualize media interactions through synergistic effects which lead the cohesive effectiveness of an IMC campaign to be greater than the added value of the individual elements making it up. They discuss that in a multi-media campaign, the effectiveness of a certain medium is significantly influenced by the spending on other media; a consumer may hear a radio commercial and recall a TV advertisement she has seen for the same brand or see a TV ad and remember some information, again for the same brand, she previously read in a magazine ad. Their results not only provide empirical evidence for the existence of cross-media synergies in multi-media communications, but also underline that these have important implications for the budget allocation decisions. Unlike what intuition suggests, under the presence of media synergies, managers should "…decrease (increase) the proportion of media budget allocated to the more (less) effective communications strategy" (Naik and Raman 2003, p. 382). This, being a very important finding by itself, also highlights the significance of considering smaller (less effective) media as well as the larger media (more effective) in the analysis, for the optimal allocation decisions.

So far, research on multi-media communications and cross media synergies has been brief and the few exceptional studies (e.g. Naik and Raman, 2003) looked at the linear interactions between media. In this paper, we argue that these interaction effects may be quite complex with non-linear or even non-monotonic shapes leading to “negative synergies” due to increased advertising levels in many media. The main barrier for the proper examination of the complex multi-media effects has been the lack of appropriate methodologies. In the next section, we first discuss two common methods used in the marketing literature to capture irregularities in response and then, introduce MARS as an alternative.

### 2.3 Methodology

In order to capture multi-media synergies and complex effects of media advertising spending, we propose MARS (Multivariate Adaptive Regression Splines) which is a flexible non-parametric statistical learning method introduced by Friedman (1991). MARS can offer substantial improvement over other commonly employed non-parametric methodologies, such as standard splines and kernel regression, in moderate sample sizes ($50 \le N \le 1000$) and moderate to high dimensions ($3 \le p \le 20$). In this section, we will briefly discuss these common flexible estimation techniques and their characteristics which will enable us to better highlight the advantages of MARS in comparison to these benchmarks in handling high dimensional problems.

### 2.3.1 Common Non-Parametric Methods

Standard Splines:

Spline regression is a piecewise polynomial estimation technique, where functions of usually low order polynomials are fit over sub-regions of the observation domain, D. The number of sub-regions (knots) in a standard spline estimation is of significant importance since it is the main factor controlling the trade off between flexibility and smoothness of the approximation barring the continuity constraints (Friedman, 1991) . Spline regression methods are fairly common and have applications in a wide range of fields including marketing (e.g. (Kalyanam & Shively, 1998; Sloot, Fok, & Verhoef, 2006; Stremersch & Lemmens, 2009). For example, Kalyanam and Shively (1998) discuss common retailer tactics such as odd pricing and price lining among the primary reasons for irregular market response to pricing and employ a stochastic spline approach to flexibly capture the irregularities. Stremersch and Lemmens (2009) use P-spline smoothing to model time varying parameters in investigating the effect of regulatory regimes (e.g. manufacturer price controls, physician prescription budgets, and prohibition of DTCA) on sales of new prescription drugs across globe.

Kernel Methods:

The other most common flexible method, Kernel, is a local non-parametric estimation technique. It works by approximating the criterion variable locally, using weighted least-squares fitting evaluated at all values of each predictor variable. The weights for the regression are defined by kernel functions (shown in Equation (1)) which are inversely related to the distances between the value of the predictor variable for which the criterion variable is estimated (i.e. $x^*$ in equation (1)) and the observed values of the predictor that lie in the proximity.

$$w(x, x^*) = K\left(\frac{|x - x^*|}{h}\right) \qquad (1)$$

The bandwidth (i.e. h in equation (1)) in Kernel regression plays a similar role as the number of knots in spline regression in determining the smoothness of the resulting approximation; hence, it is often referred to as the *smoothing constant*. The larger the bandwidth is, the smoother the estimated function will be, with the cost of being less flexible. Smaller bandwidths, on the other hand, will produce more complex and more flexible models which have more variance. Figure 2.1 depicts this phenomenon, also called as the "bias-variance trade off." (Silverman, 1986), discusses the optimal smoothing constant, $h_{opt}$, which minimizes the mean integrated squared error (MISE) and proposes several methods for its selection.

1. Subjective approaches: human judgement from the observation of the estimated density function plot, especially for lower-dimensional problems.
2. Cross validation: computing the *leave-one-out* Kernel estimator for each observation and comparing the prediction errors with respect to the actual value.
3. Approximating the true density, f, by a standard density: assuming f is d-variate standard normal and the kernel choice is

Gaussian, $h_{opt}$ which minimizes MISE can be approximated by equation (2), where *n* is the number of data points.

$$h_{opt} = \left\{ \frac{4}{n(2d+1)} \right\}^{1/(d+4)} \tag{2}$$

*Boundary effects* of Kernel method are also worthy of discussion. As the estimation point moves to the boundary of the domain, the interval of observations entering the determination of the kernel smoother becomes asymmetric which causes fewer observations to be averaged, hence flatter responses at those regions (Hardle, 1999). This fact, which is a concern for all smoothing methods, becomes more problematic for small to moderate sample sizes where a significant proportion of the observation interval can be affected by the boundary behaviour.

The extension of piecewise parametric modeling and kernel methods to higher dimensions, although straightforward in principle, is very difficult in practice, mainly due to the so-called “curse of dimensionality.” Framed in our context, this term, first coined by Bellman (1961), refers to the inability of most non-parametric techniques to efficiently and reliably handle multi-media response problems.

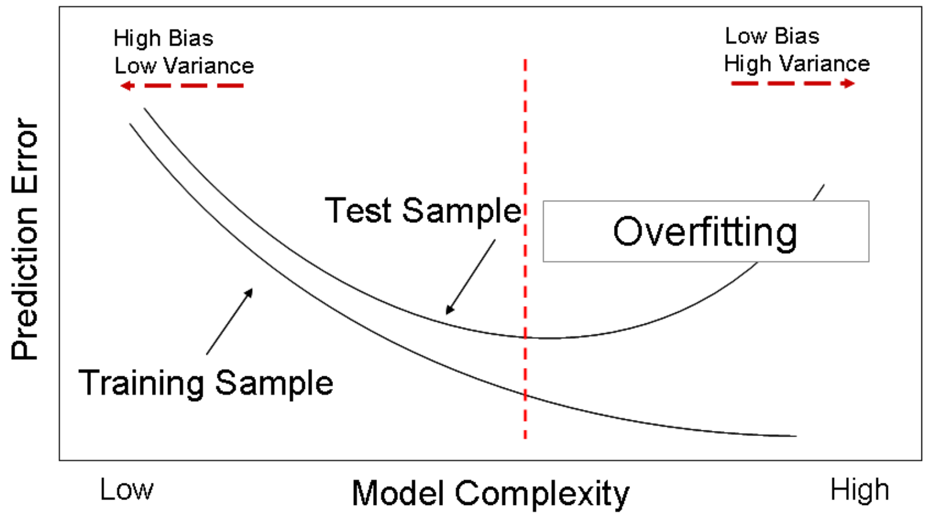



Figure 2-1: Bias-Variance Trade-off (Hastie, Tibshirani, & Friedman, 2001, p. 38)

Kernel based methods have been widely used to address marketing problems (e.g. Abe, 1995; Van Heerde et al., 2001). In one of the earlier applications, Abe (1995) employs Kernel density estimation to model consumer brand choice. The author also provides a theoretical discussion of the advantages and limitations of non-parametric methods as well as its role in developing, diagnosing and refining parametric models. More recently, Van Heerde et al. (2001) apply a Kernel based semi-parametric model to capture the deal effect curve and demonstrate the high complexity of market response to price promotions (i.e. deals) due to threshold, saturation levels, and interactions between discounts of different items.

### 2.3.2 MARS

Having briefly discussed the two prevalent non-parametric techniques in marketing literature, we now turn to Multivariate Adaptive Regression Splines as an alternative. MARS is very well suited for high-dimensional problems such as multi-media communications mostly because unlike standard splines and kernel-based methods which deal with all the variables simultaneously adopting a geometric approach of Euclidean space – hence, suffering from the "curse of dimensionality" - MARS handles individually each variable conditional on all the other variables taking on an analytic approach solely based on arithmetic concepts of adding and multiplying. MARS is adaptive, in the sense that it dynamically adjusts its strategy to take into account the behaviour of the function to be approximated. The density of the response variable is estimated by optimally dividing the domain for each predictor variable into sub-regions and fitting univariate splines (i.e. basis functions) at each region. The partitioning of each decision variable is conditional on the partitioning of all the other variables guaranteeing, thus, the optimality across all dimensions.

More specifically, the MARS algorithm works as follows. First, a set of piecewise linear basis functions, each of the form $(x-t)_+$ and $(t-x)_+$ where the subscript "+" refers to the positive part (i.e. $(x-t)_+ = \max((x-t),0)$, is created by forming reflected pairs for each variable $X_p$ ( $p \le P$) with knots at each observed value of that variable. Figure 2.2 depicts the basis functions $(x-0.5)_+$ and $(0.5-x)_+$, where the consecutive elements of the reflected pair are represented by the solid and the dashed lines respectively and the knot located at t=0.5.

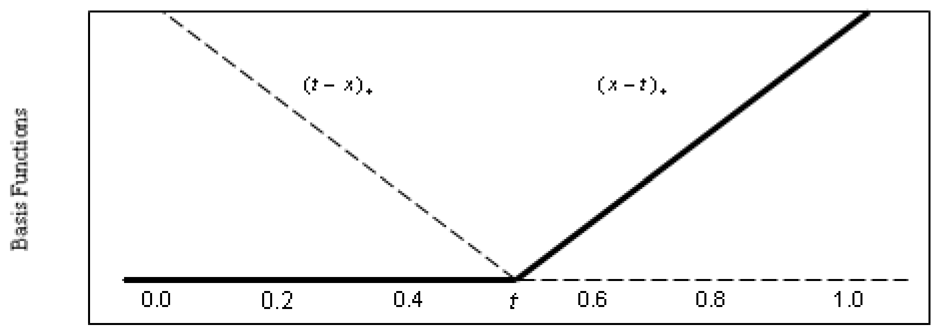



Figure 2-2: Reflected Pair of Piecewise Linear Basis Functions Used by MARS

The collection set, including all candidate functions, can be represented as in Equation (3) where p is the number of variables (i.e. dimensions) and N is the number of observations for each variable.

$$C = \left\{ (X_p - t)_+, (t - X_p)_+ \right\}_{\substack{t \in \{x_{1p}, x_{2p}, \ldots, x_{Np}\} \\ p=1,2,\ldots,P}} \qquad (3)$$

The rest of the MARS algorithm works very similarly to the forward stepwise linear regression, however, instead of the original variables, the elements of the collection set, C, and their interactions are allowed into the model. The model has the following form:

$$\hat{f}(\mathbf{X}) = \beta_0 + \sum_{m=1}^{M} \beta_m l_m(\mathbf{X}) \tag{4}$$

Starting with a constant function, $l_0(X) = 1$, a new basis function pair that produces the largest decrease in training error is included into the model at each stage. The products of all basis functions already in the model, $l_m(X)$, with each of the reflected pairs in the collection set, C, are considered as candidates for entry to the model at each selection step. The coefficients are estimated by least squares and the process continues until the model reaches a pre-set maximum number of terms.

Following the forward selection procedure, a backward elimination process is applied to control for overfitting of Equation (4) to the data. The term whose removal results in the smallest increase in the generalized cross-validation (GCV) is deleted from the model at each stage. Equation (5) represents the GCV which is used as the error measure to compute the knot locations and the optimal number of basis functions, λ, to have in the model at the end of the backward elimination procedure.

$$GCV = \frac{1}{N}\sum_{i=1}^{N}\frac{\left[y_i - \hat{f}_\lambda(X_i)\right]^2}{\left(1-\frac{C(\lambda)}{N}\right)^2} \tag{5}$$

The GCV criterion can be better understood by disentangling its elements; it is the averaged-squared error of the fit to the data (numerator) times a penalty (inverse denominator) to account for the increased variance associated with the increased model complexity.

Advantages of MARS:

MARS has a number of key advantages in tackling high dimensional problems. First, the basis functions, as shown in Figure 2.2, have the ability to operate locally, since they are only nonzero over a part of their range. When multiplied together before being added to the model, the resulting function is nonzero over a smaller part of the feature space where each of the component basis functions is nonzero. Hence, the regression surface is built up parsimoniously, using main effects and interactions locally, only when they are needed.

Second, MARS employs a hierarchical forward selection strategy. Higher order interactions are only built up with products involving terms already in the model. This is a logical assumption which increases the efficiency of the search and helps avoiding the exploration over an exponentially growing space of alternatives.

Moreover, local parametric estimation methods, such as Kernel, automatically incorporate all interaction effects of all orders into the model with the risk of including those that are unlikely to occur or are insignificant. MARS, on the other hand, not only assumes high-order interactions most likely to exist only if some of its lower order traces are already in the model, but also allows for setting a limit to the order of interactions. This unique characteristic of MARS to a large extent facilitates the model’s interpretability, which is a common flaw for the other nonparametric methods, in addition to controlling for overfitting to the data by eliminating spurious relations.

Unlike Kernel regression, boundary effects are not an issue for MARS. Most other non-parametric methods, work based on geometric concepts such as

Euclidean space, which is the main reason underlying the problem of "curse of dimensionality." When the number of explanatory variables increases, so does the dimensionality of the space, leading to sparsely populated "neighbourhoods". Consequently, a large number of observations are needed to fill the space resulting in non-local neighbourhoods and hence, increased bias of the estimation. Figure 2.3 illustrates the normalized length of the neighbourhood needed to capture a certain fraction of the data for different dimensions.

For instance, using a Kernel method which works on geometric spaces, we need to cover around 80% of the range of each coordinate to capture 10% of the data in a ten-dimensional problem. MARS, on the other hand, employs the arithmetic notions of adding and multiplying rather than the geometric concepts of Euclidean n-dimensional spaces, and thus avoids the curse of dimensionality. The high-dimensional model is built-up from individual dimensions and only variables or interactions that significantly improve model fit are included.

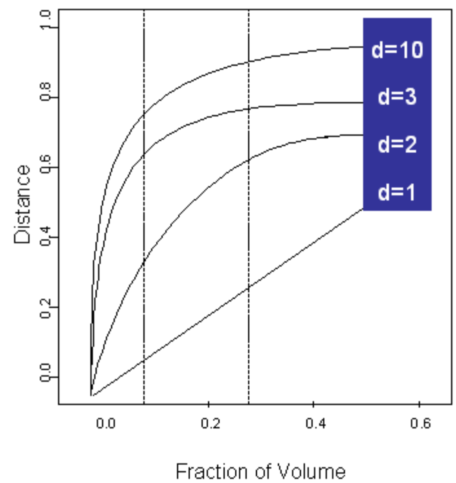



Figure 2-3: Curse of Dimensionality (Hastie et al., 2001, p. 23)

The characteristics brought up in this section highlight MARS as a front-runner method to address problems related to multi-media effectiveness. In the following, we provide the specifications of the MARS model and introduce a set of standard parametric and non-parametric benchmark models. We then describe the data, followed by the empirical analysis. In discussing the estimation results, we first show the fit and predictive validity statistics for MARS model and the three benchmark models, and conclude that the MARS model is superior for all data sets. Next, we provide illustrations of main and interaction effects between different media efforts and try to characterize the common patterns in shapes across data sets.

### 2.4 Model Specification

Having discussed primary methodologies to tackle irregular effects, in this section we summarize the models that are applied in this essay to capture complex effects of multi-media communications. For the rest of the essay, the bold letters in the equations are used to represent matrices, where non-bold letters symbolize scalar variables or parameters.

#### 2.4.1 MARS

The final MARS model after the backward elimination process has the following structure, where the right-most term in brackets corresponds to the basis functions, $l_m$, explained in the previous section.

$$S_i = \hat{f}(\mathbf{A}_i) = \beta_{0i} + \sum_{p=1}^{P}\sum_{m=1}^{M}\beta_{pmi}\prod_{k=1}^{K_m}\left[s_{p(k_m)i}(A_{p(k_m)i} - t_{p(k_m)i})\right]_+$$

(6)

where,

S, is sales in units for observation i, i=1,2,…N;

$A_p$, is the advertising variable through medium p, s.t. $p \in \{1,2,...,P\}$;

M, is the number of basis functions in the final model;

$K_m$, is the number of splits that gave rise to the $m^{th}$ basis function, in other words it is the level of interaction, s.t. $m \in \{1,2,...,M\}$,;

$t_{k_m}$ , is the knot point for the k^th interaction term of m^th basis function.

In equation (6), $\beta_0$ is the coefficient of the constant basis function (intercept) and the sum is over the basis functions $l_m$ which survive the backward elimination strategy. The term $s_{km}$ takes on values $\pm 1$ to imply right or left of the associated reflected pair.

The advertising variable in equation (6) is a goodwill variable for the time series data sets, which is calculated as below, for each medium.

$$G_{pi} = (1-\phi_p)G_{p(i-1)} + X_{pi} \tag{7}$$

where,

$A_{pi}$, is the spending on medium p, p=1,..,P, at time i, i=1,2,…N;

$\phi_p$, is the decay constant for medium p.

The reason for using stock variables, instead of the actual variables for advertising spending in analyzing time-series data is two fold. It enables us to account for the temporal effects of ad spending and to eliminate the problems due to zeros. Based on the previous literature and a set of preliminary analysis, we use ϕ=0.5 for magazines and ϕ=0.9 for all the other media. In the analysis of the cross-sectional data sets, the advertising variable in equation (6) is the actual spending amount, A. Also, following directly the original MARS paper by

Friedman (1991) we choose the degrees of freedom, which represent the cost of each basis function optimization, between 2 to 4 for each MARS model.

### 2.4.2 Benchmark Parametric Models

Parametric models are most commonly used in the marketing science literature to tackle problems regarding advertising effects (Leeflang et al., 2000). Although they are inflexible, parametric models have several optimality properties such as consistency and asymptotic efficiency (Leeflang et al., 2000, p. 397), in addition to their ease of interpretation and requirement of relatively few data points, when the true underlying function is close to the pre-specified parametric one. However, the model specification is more often subject to high uncertainty, and the parameter estimates end up being biased and inconsistent.

One parametric benchmark we use is the semi-log model of advertising goodwill, which allows for decreasing returns to scale for advertising. The exact specification is as follows:

$$S_i = \alpha_0 + \sum_{p=1}^{P} \alpha_{1p} \ln(A_{pi}) \qquad (8)$$

for i=1,…,N and p=1,…,P. The variables are defined as before.

The other parametric benchmark model is the multiplicative model represented in Equation (9). Multiplicative sales models have been quite popular in empirical marketing research, mainly due to their flexibility to allow for various response shapes based on the value of the estimated coefficient (i.e. increasing returns to scale if δ>1, and decreasing returns to scale if 0< δ<1 in equation (9)).

$$S_i = \prod_{p=1}^{P} (A_{pi})^{\delta_{1p}} \qquad (9)$$

We estimate equations (8) and (9) by ordinary least squares since it does not require any assumption regarding the distributions of the error terms unlike maximum likelihood estimation, and we aim to use the results mostly for prediction purposes.

### 2.4.3 Benchmark Non-Parametric Model

As a non-parametric benchmark we use the Kernel method with Gaussian kernels, following Abe (1995), and Van Heerde et al. (2001). For the multivariate extension, we employ the product Kernel as suggested by Hardle (1999); the regression model can then be represented by the following formulation using Nadaraya-Watson estimator (Nadaraya, 1970; Watson, 1964).

$$S^* = \frac{\sum_{i=1}^{N} S_i K\left(\frac{\mathbf{A}_i - \mathbf{A}^*}{\mathbf{h}}\right)}{\sum_{i=1}^{N} K\left(\frac{\mathbf{A}_i - \mathbf{A}^*}{\mathbf{h}}\right)} \tag{10}$$

In equation (10), the kernel function is,

$$K\left(\frac{\mathbf{A}_i - \mathbf{A}^*}{\mathbf{h}}\right) = K(A_{1i}, A_{2i}, ..., A_{Pi}) = (2\pi)^{-\frac{1}{2}P} \prod_{p=1}^{P} e^{-\frac{1}{2}\left(\frac{A_{pi} - A_p^*}{h}\right)^2} \tag{11}$$

Following the previous notation, the bold letters represent vectors of variables which are scalar values otherwise. where,

**A*,** is the vector of advertising variable (i.e. goodwill for the time series data, and nominal spending for the cross-sectional data) across media for which the response is estimated;

**$A_i$,** is the vector of advertising variable across media for which the response is observed;

p is the number of media in the model (i.e. dimensions).

For the bandwidth selection, we use the normal reference rule discussed in the previous section which provides a good approximation to the optimal bandwidth when the kernel choice is Gaussian.

## 2.5 Data

We investigate complex multi-media effects using various data sets from the US market. Time-series data consist of two subcategories of automobiles: SUVs and hybrids. We focus on the top two selling SUV brands, Ford Explorer and Jeep Grand Cherokee. The data set contains monthly information on unit sales for the brands since their launch in September 1990 and March 1992, respectively. Both brands started the promotional activities before the actual launch of the product; hence the data set covers a 129-month period between April, 1990 – December, 2000 for Explorer and 180-month period between January, 1992- December, 2006 for the Grand Cherokee.

Similarly, we focus on the top three brands with respect to their market share among the hybrid cars, namely Toyota Prius, Honda Civic and Toyota Camry. The hybrids subcategory data set also covers the sales of all three brands since their inception; more specifically, the data, consisting of monthly observations, starts from the beginning of the initial advertising campaign before the product launch (i.e. January 2001 for Prius, April 2002 for Civic and February 2006 for Camry) and ends December 2008. As a result, we have 96, 81 and 35 data points for Prius, Civic and Camry, respectively.

The first cross-sectional data set includes unit sales for the 44 beer brands in US for the year 2001. The data is part of a larger data set made available by Information Resources Inc. (IRI) which spans over 5 years of 2001-2005 and is comprised of weekly brand level store sales for a total of 30 product categories from 47 US markets (see Bronnenberg, Kruger, & Mela, 2008). The beer category is selected due to the high number of active brands with a substantial amount of variation in advertising levels for a wide range of media. The data for the beer category is pooled across stores for each brand in the market and aggregated over the year 2001 for the analysis. The second cross-sectional data is comprised of annual sales and media spending for the Leading National Advertisers in US for the year 2002. The data comes as a supplement to the Adage Special Report published in June 2003.

Table 2.1 includes a summary of these, presenting the number of observations, the range of media used to promote the brands in each data set, and the percentage spending across these media. Media information is obtained from TNS Media Intelligence (formerly Competitive Media Reporting, CMR) and includes 16 measures at the market/brand/month level covering 9 different media channels and different advertising weight metrics. Media channels comprise magazines, spot TV, network TV, cable TV, syndicated TV, newspapers, internet, radio and billboards. In this study, we use advertising variable is measured as expenditure (in $1,000s).

The shaded regions in Table 2.1 represent the media selected for the analysis for each data set. The selection process consisted of three criteria, the first one being the spending allocation percentage of total spending. Moreover, considering the fact that the cost for unit advertising through certain media could be less than others (for example a half-page newspaper ad costs less than that in a high-end magazine, which usually costs less than a 30-second spot in a TV network during prime time), we accounted for the frequency each medium is used. Last but not the least; to enhance the interpretability of the results and avoid collinearity as suggested by Friedman(1991), we eliminated the media with significantly high correlations with one or –at times- more other media. For example, newspaper ads for the Grand Cherokee is not included in the analysis because, although 6.6% of all the advertising expenditure is used to advertise through newspapers, it exhibits high correlations with magazine, network TV and cable TV ads, of values .62, .65 and .70, respectively. Similarly, cable TV is eliminated from the analysis for Civic, Camry and Prius, and internet is not included in the analysis for Camry and Prius due to high correlations. At this point, it is worth noting that despite the elimination of certain media from the analysis due to multicollinearity, we still have high dimensionality for the majority of the data sets (see Table 2.1), which provides additional support for the selection of MARS as the estimation method.

## 2.6 Estimation Results

In this section, we first report fit and predictive validity statistics for the MARS model in equation (6), parametric models in equations (8) and (9), and Kernel regression model in equation (10). We then show the nature of the multi-media effects.

### 2.6.1 Fit Statistics

In-Sample Prediction:

We report the fit of the four models in the top panel of Table 2.2 which displays the mean squared errors of the MARS estimation in the fist column and its percentage difference from the benchmark models in the columns on the right. These results indicate that non-parametric estimation improves fit to the sample data compared to both parametric models. This is an expected finding due to the greater flexibility of non-parametric methods in capturing complex and/or non-monotonic media effects.

More interesting observations, however, arise from the comparisons between the two non-parametric methods. MARS achieves an average improvement of around 55% in mean squared errors over Kernel. This difference is specifically higher for the Toyota Camry hybrid and the Leading National Advertisers, the former being the shortest, the latter being the highest dimensional data sets analyzed with 33 observations and 7 media, respectively. MARS seems to be better in handling “wider” and “shorter” data sets better, due to its arithmetic-based mechanism instead of the Euclidean geometry concept of Kernel.

Table 2-1: Summary of Data Sets

| | | | | MEDIA (% Spending) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | # of Observations | # of Dimensions | Magazine | Newspaper | Network TV | Spot TV | Syndicated TV | Cable TV | Radio | Internet |
| CROSS SECTIONAL | Lead National Adv.s. | 96 | 7 | 17.3 | 13.8 | 30.7 | 15.2 | 3.2 | 14.1 | 2.2 | 3.2 |
| | Beer | 44 | 3 | 5.2 | 0.2 | 64.1 | 6.8 | 2.2 | 19.9 | 0.6 | 0.8 |
| TIME SERIES | Ford Explorer | 129 | 5 | 28.5 | 4.5 | 41.3 | 18.6 | 1.8 | 3.8 | 0.2 | 1.2 |
| | Jeep Grand Cherokee | 180 | 4 | 22.7 | 6.6 | 25.2 | 37.4 | 0.4 | 7.0 | 0.4 | 0.2 |
| | Honda Civic Hybrid | 81 | 4 | 18.7 | 1.0 | 34.7 | 35.3 | 0 | 9.4 | 0.0 | 0.8 |
| | Toyota Camry Hybrid | 35 | 3 | 34.1 | 0.2 | 49.9 | 5.1 | 0 | 7.8 | 0.0 | 2.9 |
| | Toyota Prius Hybrid | 96 | 3 | 32.2 | 7.7 | 27.1 | 19.5 | 0.0 | 6.3 | 0.0 | 7.0 |

The dramatic differences could be attributed to a number of reasons which were discussed in details in the methodology section. Of these, probably the most widely conversed in the literature is the *curse of dimensionality*.

To further investigate the curse of dimensionality issue, we pick two data sets and perform in-sample prediction by adding one variable at a time. We then compare the marginal improvement of MARS over Kernel within each data set across different dimensions, rather than comparing performances across data sets. Hence, we aim to eliminate the effects of differences in data specific characteristics on model performances. The results are presented in Tables 2.3 and 2.4 for Ford Explorer and LNA, respectively. For each data set, we added variables based on their allocated percentage spending, as shown in Table 2.1, starting with the biggest-spending media first. The error measures used for the comparison are mean absolute deviation (MAD), mean squared error (MSE) and mean absolute percentage deviation (MAPD)[5].

---

[5] $MAD = \frac{1}{N}\sum_{n=1}^{N}\left|y_n - \hat{y}_n\right|, MSE = \frac{1}{N}\sum_{n=1}^{N}\left[y_n - \hat{y}_n\right]^2, MAPD = \frac{1}{N}\sum_{n=1}^{N}\frac{100\left|y_n - \hat{y}_n\right|}{y_n}$

Table 2-2: Fit and Predictive Validity Statistics for MARS and Benchmark Models

| | | MSE | Difference From (in %s)[6] | | |
|---|---|---|---|---|---|
| | | MARS | Kernel | Semi-Log | Multiplicative |
| | **Estimation** | | | | |
| TIME SERIES | H.Civic | 3.05E+05 | -58.1 | -79.7 | -78.9 |
| | T. Camry | 3.90E+05 | -74.7 | -79.5 | -85.0 |
| | T. Prius | 4.25E+06 | -44.1 | -46.1 | -79.4 |
| | Explorer | 3.42E+06 | -55.6 | -78.6 | -82.5 |
| | Jeep | 8.60E+06 | -30.1 | -53.8 | -59.0 |
| CROSS SECTIONAL | L. Nat. Adv.s | 1.16E+08 | -84.3 | -93.1 | -92.8 |
| | Beer | 1.05E+11 | -42.9 | -95.3 | -98.9 |
| | **Validation** | | | | |
| TIME SERIES | H.Civic | 1.39E+06 | -35.6 | -12.6 | -40.3 |
| | T. Camry | 2.10E+06 | -8.7 | -63.8 | -69.1 |
| | T. Prius | 2.60E+07 | -30.1 | -84.5 | -100.0 |
| | Explorer | 5.64E+07 | -42.7 | -11.9 | -99.6 |
| | Jeep | 3.05E+07 | -15.3 | -12.9 | -93.9 |
| CROSS SECTIONAL | L. Nat. Adv.s | 2.50E+09 | -23.1 | -8.8 | -13.8 |
| | Beer | 2.80E+11 | -67.1 | -78.6 | -99.9 |

Table 2-3: Curse of Dimensionality for the Ford Explorer

| | | Difference From (in %s) | | Difference From (in %s) | | Difference From (in %s) |
|---|---|---|---|---|---|---|
| | MARS | Kernel | MARS | Kernel | MARS | Kernel |
| *Ford Explorer* | 3-Dim | | 4-Dim | | 5-Dim | |
| MAD | 2278 | -5.35 | 1674.9 | -21.44 | 1397 | -33.44 |
| MAPD | 8.13 | -8.65 | 6.5 | -17.72 | 5 | -36.38 |
| MSE | 8.30E+06 | -11.70 | 5.14E+06 | -35.26 | 3.42E+06 | -55.52 |

[6] Negative numbers denote improvement by MARS.

Table 2-4: Curse of Dimensionality for the Leading National Advertisers

| | MARS | Difference From (in %s) Kernel | MARS | Difference From (in %s) Kernel | MARS | Difference From (in %s) Kernel | MARS | Difference From (in %s) Kernel |
|---|---|---|---|---|---|---|---|---|
| *Leading National Advertisers* | 3-Dim | | 4-Dim | | 5-Dim | | 6-Dim | |
| MAD | 17732 | 5.4 | 13712 | -12.15 | 11881 | -11.80 | 9812 | -22.47 |
| MAPD | 120.1 | 1.1 | 89.7 | -12.17 | 70.4 | -18.13 | 64 | -19.73 |
| MSE | 6.11E+08 | -28.12 | 3.61E+08 | -54.19 | 2.62E+08 | -63.91 | 1.56E+08 | -78.57 |

It is clear for both data sets that as the number of variables considered in increases, the marginal improvement of MARS over Kernel increases as well. In fact, both methods perform notably close to each other in three dimensions, with the Kernel method achieving smaller MAD and MAPD values for the LNA data set. Hence, MARS represents an improved alternative for prediction in high-dimensional problems.

Hold-out Forecasting:

Improvement in fit is virtually guaranteed with more flexible methods; therefore, a more stringent test involves the comparison of performances in the validation samples. The lower panel in Table 2.2 presents prediction errors for all models, using a 50%-50% ratio for estimation/validation sample sizes following Van Heerde et al. (2001). For all data sets, MARS achieves better predictive validity than the parametric benchmarks. MSE values for Kernel regression, on the other hand, are larger than the best fitting parametric model for three different data sets. Semi-log model attains smaller out of sample MSE than Kernel, for

Honda Civic Hybrid, and Ford Explorer. For the LNA data, Kernel's predictive validity is inferior to both semi-log and multiplicative models. Thus, the reduction in bias achieved by Kernel does not necessarily compensate for the increase in variance at every occasion.

Comparing the performances of Kernel and MARS, we see that MARS achieves better out of sample fits, confirmed by all data sets. The average improvement in fit with MARS is around 30%. For instance, for the Toyota Camry data which has a validation sample size of 16 and three variables, Kernel and MARS perform quite similarly; whereas, for the beer data, with a validation sample size of 22 and the same number of variables as in Camry, performance difference is the largest of all by 67%. Naturally, intrinsic characteristics of the data sets, such as the discrepancy of each variable in the validation sample, affect the marginal performance improvement by changing the approximate optimal bandwidth in Kernel (i.e. equation (2)).

The validation statistics reveal the benefits of the MARS algorithm where the regression surface is built up parsimoniously using main and interaction effects locally only when they are needed. This characteristic prevents overfitting the data and leads to better hold-out forecasts.

### 2.6.2 Multi-Media Effect Curves

Main Effects:

In this section, we discuss the substantive findings regarding multi-media communications effects. For the main effects, we quantify the threshold and saturation levels for each medium. We apply the MARS algorithm by constraining the interaction effects to be zero, and compute the threshold and saturation levels from the resulting MARS model after the backward elimination (as shown in equation (6)). Therefore, the final MARS model consists, only, of the significant single variable basis functions.

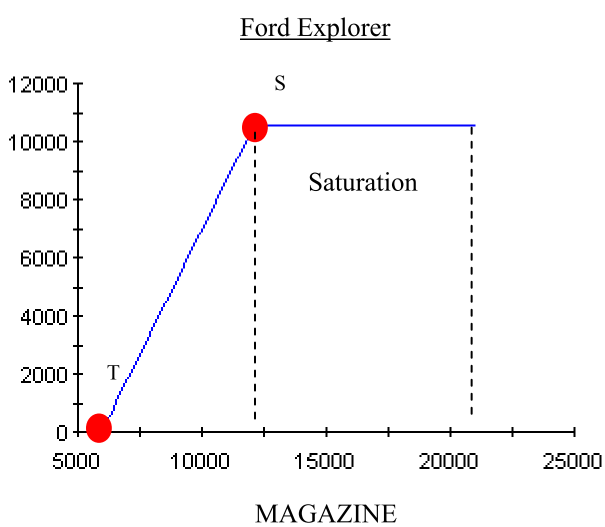
Since MARS employs simple first degree polynomials to form the basis functions, we use the first derivatives of the corresponding main effect curves in our calculations. More specifically, we represent the threshold and saturation levels of each medium by the following mathematical formulations, where ε is a very small, positive real number and Y is the response measure (i.e. sales in units, in this chapter). Equation (12) and equation (13) represent the threshold - $x_i^*$ - and saturation - $x_i^{**}$ - levels for medium *i,* respectively:

$$x_i^* = \left\{ x_i \mid \frac{\partial Y}{\partial x_i}\bigg|_{x_i = x_i^* - \varepsilon} \leq 0, \frac{\partial Y}{\partial x_i}\bigg|_{x_i = x_i^* + \varepsilon} > 0, \varepsilon > 0 \right\} \qquad (12)$$

$$x_i^{**} = \left\{ x_i \mid \frac{\partial Y}{\partial x_i}\bigg|_{x_i = x_i^{**} - \varepsilon} > 0, \frac{\partial Y}{\partial x_i}\bigg|_{x_i = x_i^{**} + \varepsilon} \leq 0, \varepsilon > 0 \right\} \qquad (13)$$

In this operationalization, a threshold point is the minimum local level after which the incremental effect of a marginal increase in spending on sales is positive. Following a similar logic, the saturation point is the maximum local level before which the incremental effect of a marginal increase in spending on sales is positive.

We illustrate some main effect curves in Figures 2.4-2.7, with advertising variable per medium on the x-axis and contribution to sales on the y-axis. Red circles mark the corresponding threshold (T) and saturation (S) points. The set of curves presented here are selected to exemplify different effect shapes that the analysis reveals most of them common across media and brands.



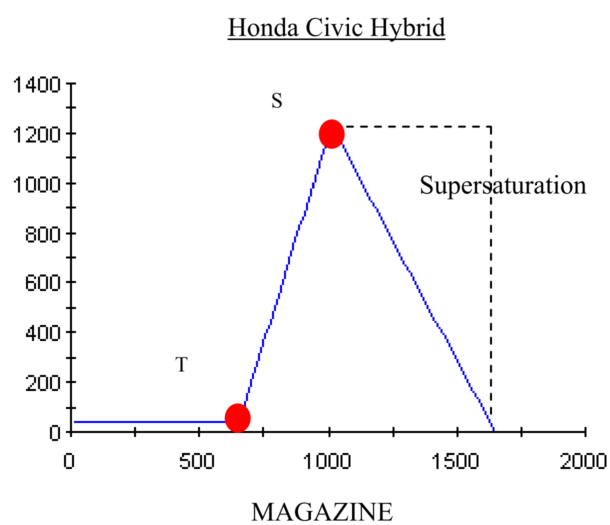




Figure 2-4: S-shaped pattern

Figure 2.4 shows examples of advertising response curves typically found in the literature. The curve on the left displays the response to magazine advertising for the Ford Explorer brand, and we see an S-shaped curve with a threshold and saturation levels. On the right, magazine ad response for the Civic hybrid portrays a similar S-shape, however, the contribution to sales decreases towards the upper end of the spending range. This oversaturated advertising response, despite not being widespread across the seven data sets we analysed, is also encountered for cable TV in the cross sectional data sets of beer and Leading National Advertisers. Hanssens et al. (2001), who describe this phenomena as supersaturation, explain that too much marketing effort is likely to cause negative response, and mostly give examples of marketing decision variables other than advertising. In this analysis, we show that supersaturation of advertising efforts is possible, especially under the presence of multiple media where consumers may view the excessive efforts as repetitive or intolerable, completely block advertising or even react negatively.

Another emerging pattern in advertising response is the possible existence of multiple thresholds. The bimodal patterns in Figure 2.5 each reveal two different threshold levels for the responses to cable TV and magazine ads for Ford Explorer and Jeep Grand Cherokee, respectively. A certain level of spending seems to be necessary for marginal advertising to have a positive impact on sales; although there seems to be an early saturation for communications efforts in both cases, after a second threshold the marginal impact of advertising turns positive again. This kind of response pattern has been documented in the literature and could very well be the result of the existence of multiple segments of consumers with distinct thresholds and saturation points (see Ackoff & Emshoff, 1975; Kalyanam & Shively, 1998).

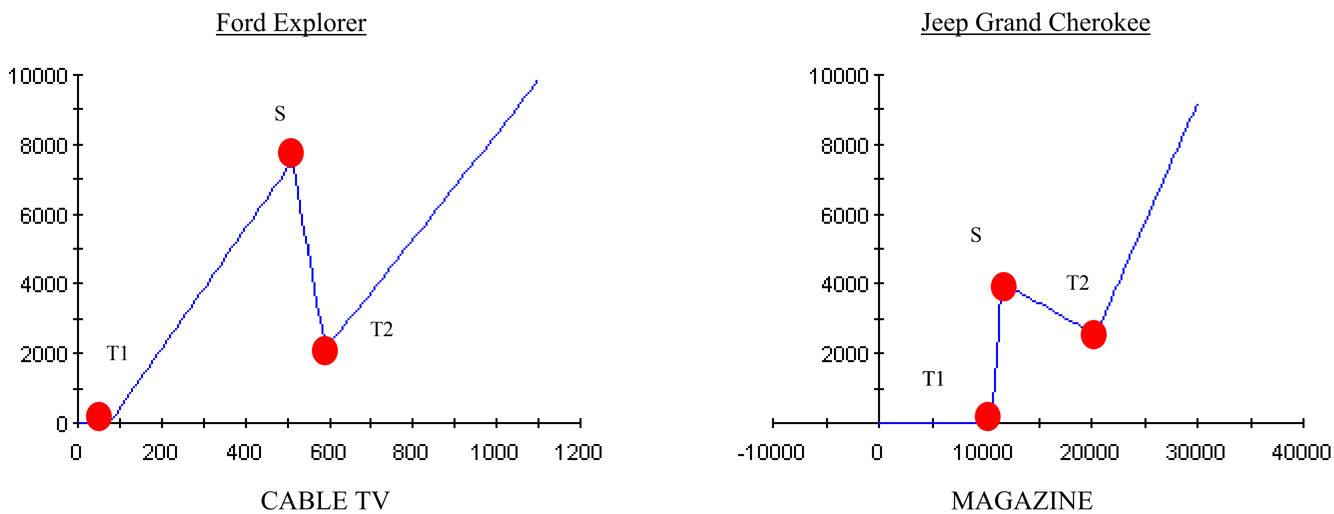
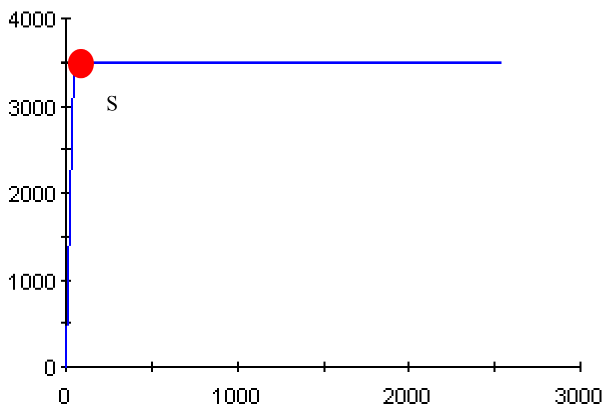



Figure 2-5: Multiple Thresholds

A third pattern revealed in the analysis is early saturation of some media efforts. Specifically, newspaper advertising is observed to have the smallest saturation levels. Figure 2.6 exhibits two response curves to newspaper advertising. For each of these curves, the saturation level is around 5% of the maximum spending – a significantly low value compared with those for the rest of the media alternatives.

Ford Explorer




Lead National Advertisers

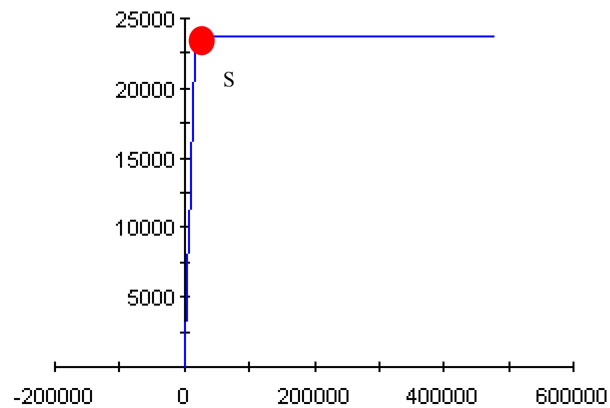



Figure 2-6: Early Saturation

As a final pattern, although in a limited number of instances, the results show decreasing effects of advertising. Two of these reverse S-shaped effects are presented in Figure 2.7, depicting the response to internet and cable TV ads for Honda Civic and Jeep Grand Cherokee, respectively. This decreasing pattern is interesting for a variety of reasons. Most importantly, this type of pattern for advertising response has not been commonly observed in the previous literature, which mainly focused on the effects of single medium advertising or the cumulative advertising effort. In this study, we are able to show that the response shapes for multi-media advertising campaigns could be significantly different, and even follow a decreasing pattern for some of the less effective media, due to the excessive repetition of the messages.

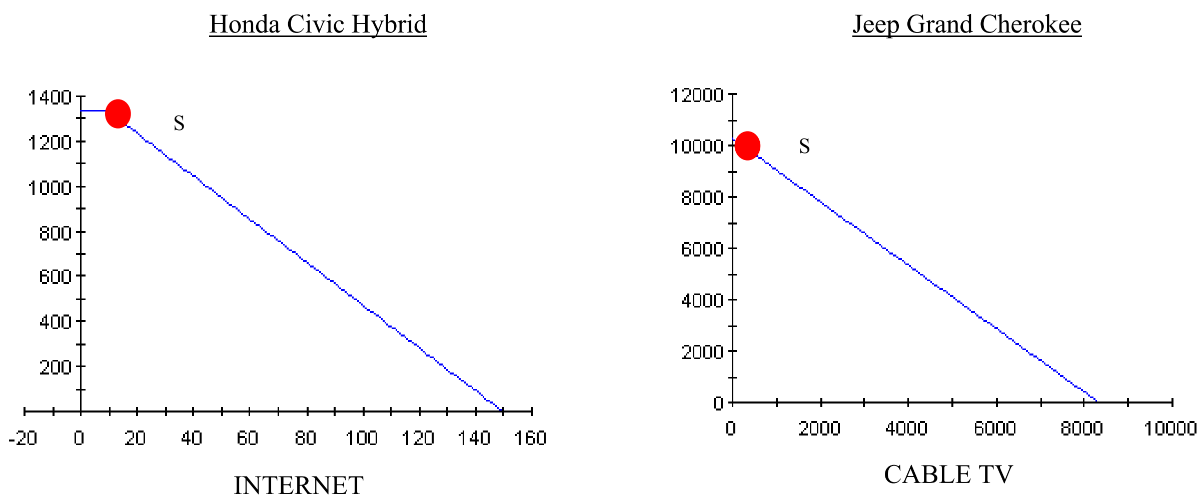



Figure 2-7: Decreasing Response

Having discussed these emerging shapes and findings, we summarize the key characteristics of the response curves of each medium based on the final MARS model for each data set in Table 2.5, which provides a more general overview of the results. The threshold and saturation levels presented are calculated values rounded to the nearest multiple of 5%. The threshold levels for different media and data sets range from 5% to 30%, saturation levels have an average of approximately 50% excluding newspaper ads. As mentioned before, newspaper messages are exposed to very early saturation common across data sets of this paper. It is clear from Table 2.5 that the majority of the media do not operate in an efficient spending range and overspending is a common finding across data sets and media.

Table 2-5: Description of Main Effects

| | Effect Shape | MARS Model: Threshold Level | MARS Model: Saturation Level |
|---|---|---|---|
| *Toyota Prius Hybrid* | | | |
| Magazine | Convex | 5% | - |
| Network TV | Concave | - | 15% |
| *Toyota Camry Hybrid* | | | |
| Magazine | Concave | - | 45% |
| Network TV | Concave (no saturation) | 5% | - |
| *Honda Civic Hybrid* | | | |
| Magazine | S (supersaturation) | 35% | 65% |
| Network TV | Convex | 30% | - |
| Internet | Reverse S | - | 10% |
| *Ford Explorer* | | | |
| Magazine | S | 15% | 40% |
| Network TV | Reverse S | - | 55% |
| Spot TV | Concave (no saturation) | 5% | - |
| Cable TV | Double S | T1: 5% T2: 25% | S1:20% S2: - |
| Newspaper | Concave | - | 5% |
| *Jeep Grand Cherokee* | | | |
| Magazine | Double S | T1:30% T2:65% | S1:40% S2: - |
| Spot TV | Double S | T1: 5% T2:55% | S1:20% S2: - |
| Cable TV | Reverse S | - | 5% |
| *Leading National Adv.s (CS)* | | | |
| Magazine | S (no saturation) | 30% | - |
| Spot TV | S (no saturation) | 10% | - |
| Cable TV | Reverse S | - | 25% |
| Newspaper | Concave | - | 5% |
| *Beer (CS)* | | | |
| Network TV | S(no saturation) | 15% | - |
| Cable TV | Concave (supersaturation) | - | 35% |

Interaction Effects:

One of our main goals is to estimate the cross-media interactions in multi-media communications. For this purpose, we illustrate the interaction effects based on the resulting MARS models for each data set. To highlight the differences between MARS and Kernel, we compare the estimated synergy effects by the two methods. Among the significant two-way interactions based on the MARS model with maximum level of interactions set to two, a selected set is presented in Figures 2.8-2.10. In the MARS figures, the vertical axis represents the predicted contribution to the sales volume, and the other two axes represent the spending for the corresponding media. For the Kernel figures, the vertical axis is the predicted sales volume instead of the contribution, and the other two axes are the same as those for MARS.

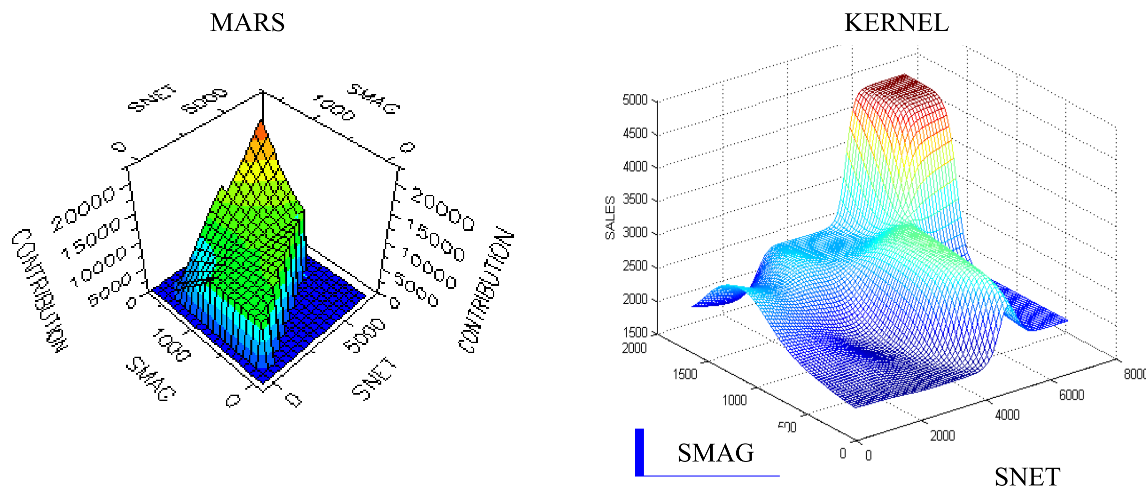



Figure 2-8: Three-Dimensional Multi-Media Synergy Surfaces - Honda Civic

The right panel of Figure 2.8 portrays the Kernel output for the magazine-network TV interaction for Honda Civic. Comparing this with the corresponding MARS figure on the left clearly reveals that, despite the two surface plots being quite similar, Kernel clearly suffers from boundary conditions and gives flatter responses closer to the boundaries of the observation domain. Figure 2.8 is chosen solely to serve as an example; similar boundary effects are common across data sets and media pairs. Hence we can confidently say that the Kernel method often understates the effects of very small and very large media spending by showing threshold and saturation effects that may not actually be in play.

Figure 2.9 demonstrates two sample interaction effects for Jeep Grand Cherokee. The top panel compares estimated surfaces by MARS (left) and Kernel (right) models for the spot TV- magazine interaction. Similarly, the bottom panel presents the estimated network TV-magazine interaction surface. Kernel and MARS surfaces are quite similar in general for both interactions with the exception of a specific region. More specifically, for the top panel Kernel method represents the interaction effect where spot TV spending is moderate to high (e.g. greater than 10,000) and magazine spending is very low (e.g. smaller than 10,000) as a flat surface, reminiscent of a saturation level; whereas, in the corresponding MARS plot the same region is left empty, without any significant interactions.

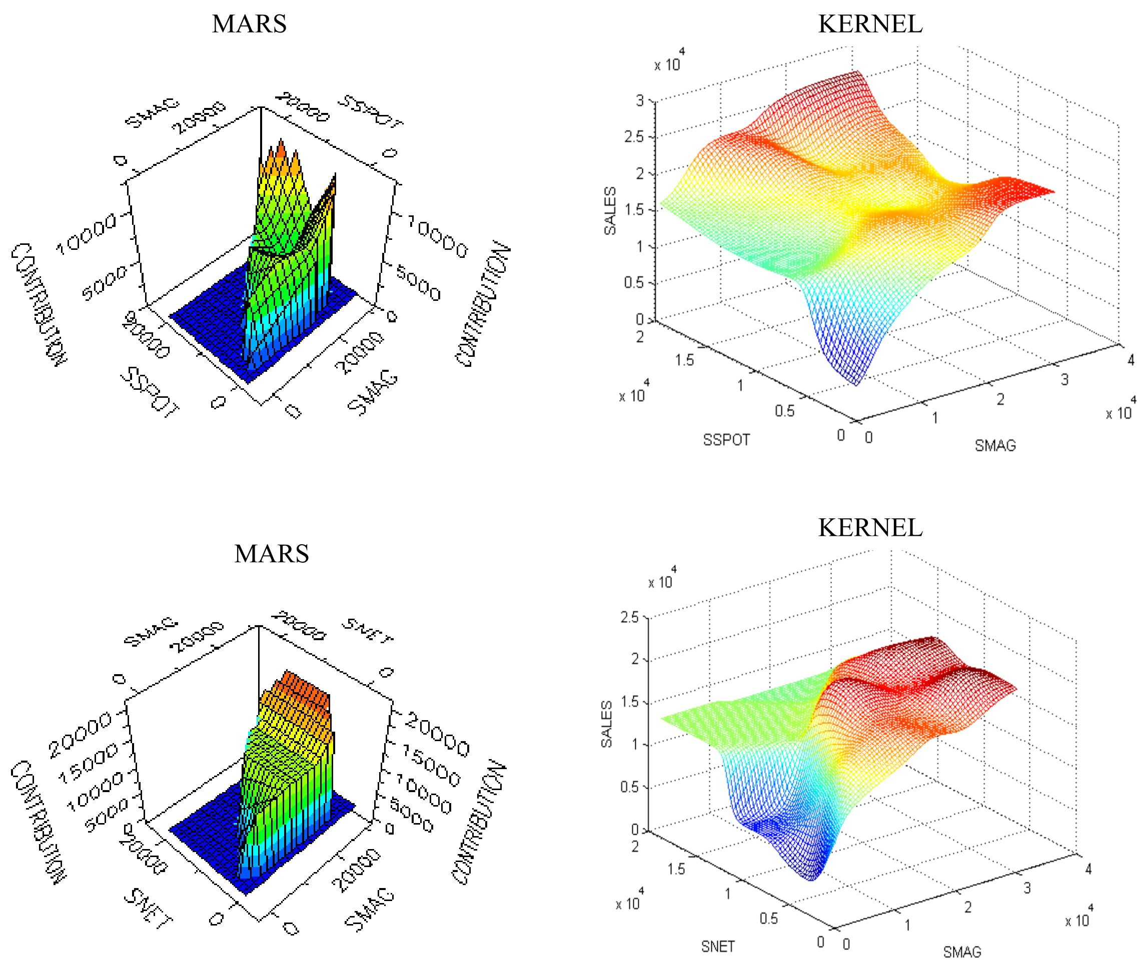



Figure 2-9: Three-Dimensional Multi-Media Synergy Surfaces - Jeep G.C.

A similar situation is apparent for the network TV-magazine interaction effect for moderate to high network TV spending and very low magazine spending. More careful examination of the data reveals that a small number of observations actually fall into those flat regions of the Kernel plots. Specifically the number of observations with high spot TV (network TV) and very low magazine values are 2

(4) out of 180 available points in the data set. Hence, the Kernel-estimated interactions can possibly be misleading and present spurious effects in regions with limited observations. MARS overcomes this problem by building up the regression surface parsimoniously and using interactions locally only when they exist. Figure 2.10 represents two sample interaction surfaces for Ford Explorer where the Kernel estimates are still subject to problems but the extent of these are much less than the previous examples.

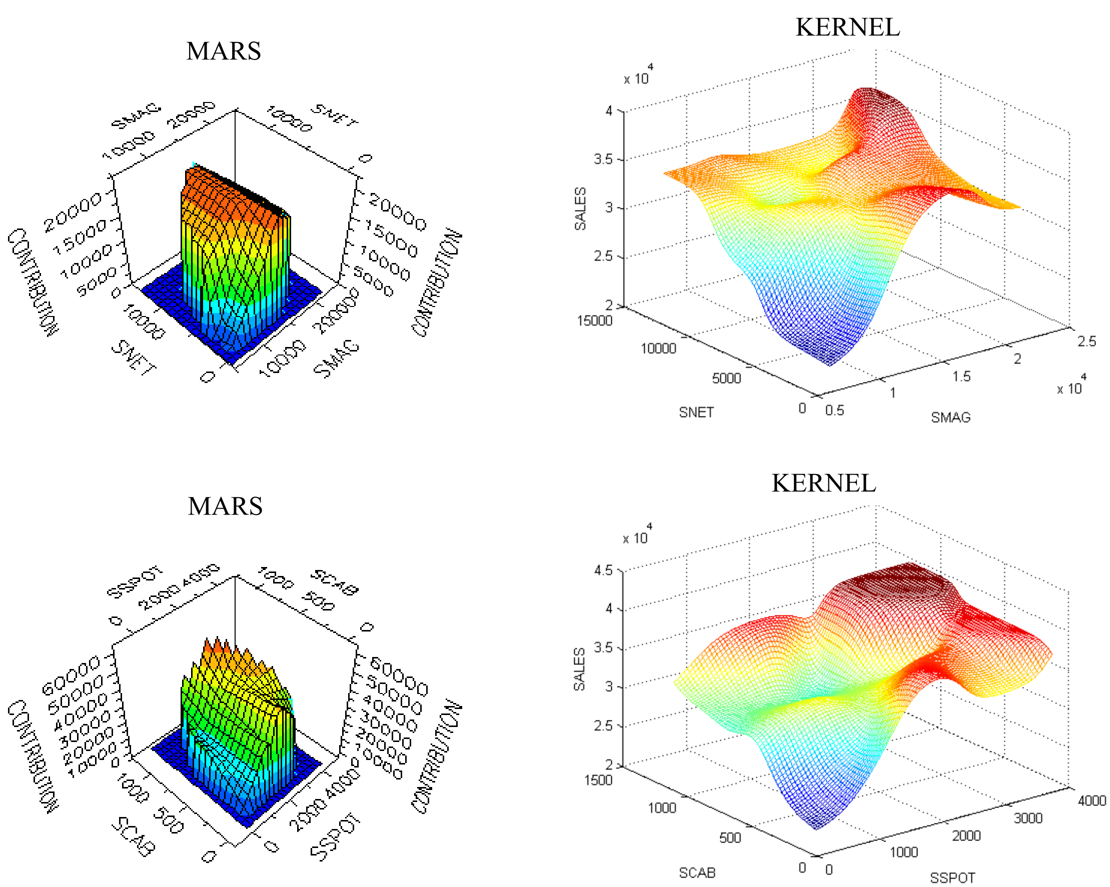



Figure 2-10: Three-Dimensional Multi-Media Synergy Surfaces - Ford Explorer

Overall, we see that the interaction effects are complex and irregular, which would be difficult to model parametrically. The advantage of MARS is that, it simplifies these effects by (1) fitting combinations of simple linear functions to the observation domain and (2) eliminating the non-significant effects all together, hence, increasing the interpretability of the results. As a direct consequence of the former argument, the synergy surfaces estimated by MARS turn out to be more edgy and less smooth when compared to Kernel.

## 2.7 Conclusions and Implications

In this chapter, we focus on capturing complex/irregular effects of multi-media communications on market response. We develop a non-parametric regression model based on multivariate adaptive splines, MARS, for this purpose. We show that MARS is highly suitable for addressing problems, such as multi-media effects, which require flexible modeling of large number of variables; whereas, most other non-parametric methods suffer from the well known issue of *curse of dimensionality* when working in higher dimensions.

Comparisons between our MARS-based non-parametric model with benchmark parametric models and a non-parametric Kernel-based model show that (1) MARS provides a better fit and shows better average forecast performance than both the parametric benchmarks and Kernel regression confirmed by all data sets, and (2) the marginal improvement of MARS over Kernel significantly increases as the number of variables in the non-parametric model increases.

Following the validation of the MARS approach, we turn into the investigation of the shapes of multi-media communications effects. We present a set of main effect shapes common across various data sets. We find compelling evidence to (1) S-shaped response to multiple media efforts with typical threshold and saturation levels, (2) existence of multiple thresholds separated by a single saturation point, (3) possible supersaturation and decreasing response for certain media such as internet and cable TV – although not prevalent-, and (4) early saturation of response to newspaper ads. We also quantify the threshold and saturation levels using non-parametric derivatives. The results also show that most product categories and media do not operate in the most efficient spending range for advertising, evident from the average saturation levels of around 50% of the maximum spending across media and data sets.

We demonstrate cross-media synergies by three dimensional interaction surface plots. Synergy surfaces are noticeably complex and irregular supporting the use of flexible non-parametric estimation methods. Comparison of respective plots for MARS and Kernel-based estimations reveal that significant differences in regions which i) are close to the boundaries of the observation domain and ii) have limited number of data points. The results clearly show that Kernel suffers from *boundary effects* and tends to falsely display threshold and saturation levels due to the sparsely populated data regions.

Given the reliability of MARS in capturing multi-media effects, future research can take these findings further by working on optimal scheduling and budget allocation decisions for multi-media communications problems.

# CHAPTER 3

# DYNAMIC EFFECTS OF HIGHLY REGULATED ADVERTISING MESSAGES

## 3.1 Introduction

Regulations can impose significant restrictions on the communication of advertising messages to consumers, especially in closely monitored markets such as pharmaceuticals (e.g. Gellad & Lyles, 2007; Lyles, 2002; Rosenthal, Berndt, Donohue, & Frank, 2002; Stremersch & Lemmens, 2009). Such restrictions may increase the already high uncertainty regarding the effectiveness of advertising and naturally pose a dilemma for brand managers who, faced with high advertising costs, should: (a) contemplate whether or not to advertise under such conditions and (b) assess the potential effectiveness of regulated advertisements. In such a case, findings on advertising effectiveness based on studies of mass-marketed consumer packaged goods or durables (e.g. Assmus, Farley, & Lehmann, 1984; Leone & Schultz, 1980; Lodish et al., 1995; Vakratsas & Ambler, 1999) may not serve as appropriate benchmarks due to the varying degrees, or even lack, of regulation in the markets covered. Furthermore, strict regulatory environments frequently require that advertisers should choose from a particular menu of advertising claims. For example, in the case of nutritional supplements regulations demand that advertising messages can either be “nutritional support” statements or “health” claims, imposing restrictions on the format of advertising messages as well (e.g. Nestle, 1999). Then, additional considerations for potential advertisers include: which type of message would be more effective and when? Thus, potential advertisers should consider whether or not to use regulated messages, choose the most appropriate message and decide when it would be preferable to do it.

The previous discussion suggests that advertising effects in regulated markets may be highly idiosyncratic and thus warrant a customized and comprehensive approach. In this paper, we investigate the dynamic effects of Direct-to-Consumer Advertising (DTCA) messages on sales of a new prescription pharmaceutical in a non-U.S. market, where regulations impose strict requirements on the type and content of prescription pharmaceutical advertising. Specifically, only two types of DTCA messages are allowed: One so-called disease-related or help-seeking advertising message, akin to a generic advertising message, where only information about the disease, and not a specific brand or device, can be communicated to the consumer. The other, is a brand-specific, or reminder, message containing information only about the brand, void of any therapeutic-related information. These regulation conditions give rise to some interesting research questions: Can advertising be effective under these strict regulatory conditions? Which type of advertising message is more effective: disease-related (generic) or branded? Disease-related messages can have high informational value but may not be able to create differentiation due to the lack of any brand-related communication. Branded messages, on the other hand, may be more differentiating provided consumers can connect it with a specific disease. A final question regarding the effectiveness of regulated advertising messages is time-related: When is one type of advertising message more effective than the other? Can informational disease-related messages still be effective as the market becomes more educated and matures or would branded messages be then more preferable? In sum, our study seeks to provide comprehensive answers to the following questions: *whether* highly regulated DTCA messages can be effective; *which* type of DTCA is more effective and *when*? (e.g. Tellis, 2004; Tellis, Chandy, MacInnis, & Thaivanich, 2005; Tellis, Chandy, & Thaivanich, 2000).

We investigate these issues by examining a comprehensive database consisting of monthly information on new and refill prescriptions of a novel prescription pharmaceutical and all major marketing mix instruments (i.e. DTCA, detailing and physician journal advertising). Data are available since the launch of the pharmaceutical, which marks the inception of the therapeutic category, and cover a period of seven years. We employ the Augmented Kalman Filter with continuous state and discrete observations (AKF(C-D)) (Xie, Song, Sirbu, & Wang, 1997) to estimate dynamic effects of DTCA and the rest of the marketing mix. The use of Kalman filtering is highly appropriate since it allows us to accurately estimate flexible dynamic advertising effects using non-linear models, continuous by nature (such as the Bass model of diffusion of innovations), and limit any discretization biases. From a methodological point of view, we extend the Xie et al. (1997) approach by incorporating parameter estimation in the AKF (C-D) algorithm and obtaining corresponding confidence bounds using Monte Carlo simulation.

Our study intends to make two contributions towards the advancement of current knowledge on advertising effectiveness. First, it investigates dynamic effects of DTCA. Although many previous studies using U.S. data have examined the effects of DTCA, none has focused on its dynamics. The only study concerned with dynamic effects in pharmaceutical markets (Narayanan, Manchanda, & Chintagunta, 2005) focuses on detailing and other physician-directed activities, using a learning model of physician prescription behavior. Hence, our work intends to complement that of Narayanan, Manchanda and Chintagunta (2005) by extending the investigation of dynamics to consumer-directed, regulated activities which could exhibit different patterns due to their distinct nature and objectives. Exploring the dynamics of regulated advertising messages should be useful to managers, as it would provide them with a better sense of timing and duration of their advertising. Surprisingly, there is a general dearth of research in advertising dynamics, and only recently some notable studies have attempted to enhance our knowledge (Bass et al., 2007; Bruce, 2007; Chandy, Tellis, MacInnis, & Thaivanich, 2001; Naik et al., 1998). Second, our study is the first, to our knowledge, to empirically assess the effectiveness of generic advertising in a novel, non-commodity market and compare it with that of branded messages. Marketing studies on the subject of generic advertising are mainly theoretical and concerned with the optimal allocation of generic advertising funds (Bass, Krishnamoorthy, Prasad, & Sethi, 2005; Krishnamurthy, 2000, 2001). Our study complements the previously cited theoretical work by furnishing relevant empirical evidence.

The rest of the paper is organized as follows. In Section 2, we provide a discussion of the regulation environment. Section 3 discusses the related literature on DTCA, generic advertising and dynamic advertising effects. Section 4 provides a detailed description of the data, followed by the modeling approach in section 5 and findings in section 6. Section 7 discusses the implications of our findings, and Section 8 addresses validation issues. Section 9 contains a summary and the conclusions from our study.

### 3.2 Drug Advertising in a Regulated Environment

We consider a market with strict regulations regarding pharmaceutical advertising targeted directly to consumers (DTCA). In the market of interest, direct-to-consumer advertising of prescription drugs was originally prohibited according to the Food and Drugs Act (FDA), which regulates the advertising of pharmaceutical products in the focal market. An amendment, introduced in 1978 to allow price advertising of prescription-only drugs to the public, states that: “Where a person advertises to the general public (…), the person shall not make any representation other than with respect to the brand name, proper name, common name, price and quantity of the drug”. Moreover, another section of the Act provides a list of diseases including impotence, diabetes, baldness and asthma for which treatments and cures may not be presented to the public at all. Finally deceptive and misleading advertising is also prohibited by the Act. Two additional major policy shifts and relaxations in the interpretation of law since 1996, contributed to an increase in DTCA volume. These shifts mainly concern the two different types of DTCA eventually allowed as described below:

- *Disease-oriented or help-seeking ads*: These discuss a specific disease or health condition and prompt viewers or readers to ask their doctor about an unspecified treatment, but do not mention a specific brand or device or make any representation or suggestion concerning a particular drug or device. No risk information is required. In a sense, generic messages may not be considered as advertising but rather as informational announcements since they do not contain brand-related information.

- *Reminder ads*: These may contain only the brand name, the established name of each active ingredient and, optionally, information relating to quantitative ingredient statements, dosage form, quantity, price, and other limited information. They should not include health claims or by any means use representation about the product's use, such as listing of medical specialties. No risk information is required.

Thus, in the focal market, only these two types of ads are allowed which draw a clear distinction between disease (therapeutic category) and reminder (branded) advertising. In other words, prescription drug firms cannot combine promotional information for a specific prescription drug brand and the particular disease or condition it addresses in a single advertisement. Further, if a typical consumer could easily link two announcements and the messages contained in those announcements, this similarly contravenes this prohibition. Therefore, the airing of such announcements sufficiently close in time with the same tone and manner or with the same actors is also prohibited by the rule. From here on, we will refer to the disease ads as “generic” and reminder ads as “branded.”

By contrast, the FDA in the U.S., in addition to the previous two types of ads, also allows the following:

- *Full product ads*: These include the brand name and health claims and must by law include risk information[7].

Table 3-1: Types of DTCA Defined by Regulations

| | **Full Product Ads** | **Generic Ads** | **Branded Ads** |
|---|---|---|---|
| Therapeutic Category Information | √ | √ | X |
| Symptoms | √ | √ | X |
| Health Claims | √ | √ | X |
| Risk Information | √ | NR | NR |
| Price Information | √ | X | √ |
| Brand Name | √ | X | √ |
| Dosage Information | √ | X | √ |
| Direct to a physician | √ | √ | NR |

NR : Not required
√ : Required/allowed
X : Not allowed
: Focal Market
: US Only

[7] Full product ads are not allowed in the focal market. The only other market where full-product, US-style, ads are allowed is New Zealand. Regulations regarding physician-directed activities are very similar between the U.S. and the focal market.

Table 3.1 presents a summary of the characteristics of the different types of DTCA discussed. It should be noted that the type of advertising message (generic versus branded) essentially dictates the ad content. Generic advertising messages are typically more informational since they focus more on health claims, symptoms, risk, etc. Branded advertising messages, on the other hand, are more likely to be emotional since they are restricted to convey very little information, beyond the mention of the brand name, and the communication of any other benefits can only be implied through the use of visuals, music, and actors, in other words of different creative executional elements. Thus, the choice of the DTCA type becomes a critical strategic issue.

## 3.3 Related Literature

### 3.3.1 Research on DTCA Effectiveness

Research on the effects of DTCA has exhibited considerable growth during the last decade, mainly due to the deregulation of such advertising in the U.S. and the subsequent availability of relevant data. Table 3.2 summarizes the most relevant studies on DTCA with their objectives, data used, methodology employed and findings. Some researchers specifically focus on the market expansion and switching effects of DTCA (Berndt, Bui, Reiley, & Urban, 1995; Iizuka & Jin, 2005; Wosinska, 2002) while others more generally attempt to compare DTCA effectiveness to other marketing activities typically employed for prescription drug advertising, such as physician journal advertising, detailing and free samples (Calfee, Winston, & Stempski, 2002; Rosenthal et al., 2002). In general, studies on DTCA found its effects to be modest in comparison to detailing. Kremer et al.(2008), upon investigating 58 studies on pharmaceutical promotions, report average detailing and DTC advertising elasticities of 0.33 and 0.07, respectively.

A careful examination of Table 3.2 reveals two points that deserve attention. First, there is no empirical work which employs data that differentiate between branded and generic DTCA. All of the studies mentioned in Table 3.2 are based on U.S. data sets where firms typically prefer the product claim DTCA format (Table 3.1). We believe that the distinction we draw in this study between branded and generic advertising, forced by regulation, will shed more light on the role of regulations in the effectiveness of advertising messages, and thus provide implications for advertising message strategy and the impact of advertising messages in strict regulatory environments. Second, no study has examined dynamic DTCA effects. Narayanan, Manchanda and Chintagunta (2005) have investigated dynamic detailing effects but the different nature, objectives, and

Table 3-2: Summary of Previous Studies on DTCA

| Author Reference | Objective | Marketing Variables | Generic vs Branded DTCA | Dynamic Effects | Methodology | Main Findings |
|---|---|---|---|---|---|---|
| Berndt et al. (1995) | Examine the role of marketing communication on the expansion if a therapeutic category and the market shares of the top brands | Price and, goodwill stocks for DTCA, detailing, and physician journal ads | No | No | NL-2SLS regressions for a category and a brand level model | •Sales elasticity is largest for detailing, followed by journal ads and smallest for DTCA |
| Wosinska (2002) | Investigate the market expanding and business stealing effects of DTCA | DTCA, detailing | No | No | Utility maximization of the physician through mixed logit framework | •DTCA increases choice probability for only the drugs in the formulary •Marginal effect of detailing is greater than the marginal effect of DTCA |
| Calfee,Winston&Stempski (2002) | Assess the effects of promotional activity on category demand | DTCA, detailing, physician journal ads | No | No | Time series regression | •No significant demand effects of DTCA •Patient doctor interaction and information dissemination re: drug efficacy are main drivers of demand |
| Bowman, Heilman, & Seetharaman (2004) | Explain the factors affecting product-use compliance | DTCA, patient costs and benefits, distribution | No | No | OLS regression for compliance & MLE of a latent class model for segment level analysis (based on demographics) | •Consumer heterogeneity in response to DTCA •DTCA has positive effect on some consumer segments and negative effect on others |
| Wosinska (2005) | Quantify the economic magnitude of DTCA effects on brand sales through decreased interpurchase time | Own DTCA, competitor DTCA | No | No | OLS estimation of a negative binomial model with missed therapy days as dependent variable | •Own brand DTCA increases compliance among patients taking competing brands •DTCA has a positive but economically insignificant effect on the advertising brand's compliance |
| Iizuka&Jin (2005) | Examine the market expanding effects of DTCA through doctor visits | goodwill stock for DTCA | No | No | NLS regression with doctor visits as a dependent variable | •Significant positive effects of DTCA on doctor visits •The effect becomes stronger after 1997 FDA clarification |
| Narayanan, Machanda&Chintagunta (2005) | Examine the effect of marketing communication over the life cycle of a new product category | Price, goodwill stocks for DTCA, detailing and other marketing expenditures | No | Detailing only | Discrete choice model of physician learning through a Bayesian learning process | •Significant direct and indirect effects of detailing •Higher elasticities for detailing when compared to DTCA and OME •Firms should allocate more resources to detailing in the introductory phase when it leads to faster learning |

regulatory restrictions on DTCA could produce different effectiveness paths. Thus our study intends to complement that of Narayanan, Manchanda and Chintagunta (2005) in providing insights on the dynamics of consumer-directed activities.

### 3.3.2 Generic Advertising Effects

Empirical studies on generic advertising effectiveness (Alston, Chalfant, & Piggott, 2000; Brester & Schroeder, 1995) mainly focused on commodity-type food markets, such as animal products and by-products and farm commodities (e.g. beef, pork, eggs, corn, soybeans). Despite their merit, these studies have two major limitations: First, they do not examine non-food differentiated markets such as the one considered in this paper, and second, they do not clearly distinguish between branded and generic advertising effects, since advertisers in these markets are allowed to provide brand-specific information in addition to category-specific information in the same message. In other words, branded and generic advertising content are not mutually exclusive in these settings. To our knowledge, no study has examined the effectiveness of purely generic advertising messages in differentiated markets.

Most marketing studies on generic advertising are theoretical and concerned with normative implications on the allocation of generic advertising (Bass et al., 2005; Krishnamurthy, 2000, 2001). These studies assume that generic advertising, by providing information about the general qualities and attributes of a product category, should be expected to increase primary demand. Krishnamurthy (2000, 2001) discusses free riding issues related to generic advertising and shows that dominant firms should be indifferent to free riding by lesser firms because they benefit the most from generic advertising even when they incur the entire industry advertising expense. This finding is also confirmed by Bass et al. (2005) who further suggest that optimal allocation of generic advertising, as a proportion of the total budget, should be initially high but then decrease progressively in favor of higher branded advertising allocation. The implication is that market conditions should initially favor generic advertising. This applies well to our case, since we observe a period of monopoly for the pioneering brand from which it can benefit through generic advertising. However, as the market becomes more educated and competition enters, branded advertising should be a more appropriate vehicle. Thus, we should expect branded advertising to be more effective later in the product life cycle.

### 3.3.3 Advertising Dynamics

Previous research suggested that advertising effects are likely to follow a dynamic pattern as a result of many influencing factors such as stage of product life cycle (PLC), message content (type), and competition. In this section, we will review these factors and their potential impact on advertising effectiveness over time.

- *Product Life Cycle*: Parsons(1975), in a seminal study, analyzed the effects of the time a product spent in the market and the stage of its product life cycle on advertising elasticities. He concluded that the absolute magnitude of advertising elasticities declines in a nonlinear fashion over time. Since then, the idea of declining advertising elasticities over the product lifecycle has become a commonly held view among practitioners and academics. Winer(1979) also explored the possibility of time varying advertising effects using the Lydia Pinkham data set. His study suggests that in the presence of carryover effects, advertising elasticities may exhibit an increasing pattern over time, while carryover effects may decline. Meta-analytic studies have also shown advertising elasticities to be lower for mature markets (e.g. Assmus et al., 1984; Vakratsas & Ambler, 1999). Hence, we should expect both generic and branded advertising effectiveness to exhibit a declining trend over time.

- *Message Content*: Chandy et al. (2001), relying primarily on behavioral theories, argue, and find, that informative advertising is especially effective in new markets and for recently introduced products. However, as markets and products mature, emotional advertising becomes the more effective alternative, suggesting its ability to produce persuasive effects (Becker & Murphy, 1993; Comanor & Wilson, 1974). MacInnis, Rao and Weiss (2002) similarly concluded that frequently purchased brands in mature product categories are better off differentiating by using warmer and more likable advertisements based on affective executional cues, rather than relying on product-based information.

To the extent that generic advertising messages in our case tend to be more informational by regulation requirements (focus on the disease, its symptoms etc.), whereas branded advertising messages are likely to be more emotional, due to the restriction on the information they can communicate, one should expect that generic advertising is more effective early in the product's lifecycle. Similarly, branded advertising, having to rely more on executional elements as discussed earlier, may achieve persuasive or reinforcing effects and hence should be more effective later in the lifecycle.

- *Competition*: Research has shown that competitive intensity and the amount of competitive advertising activity are important factors influencing advertising effectiveness (Danaher et al., 2008; Vakratsas et al., 2004). Competitive interference typically leads to higher levels of advertising clutter and, consequently, limits consumer attention to advertising (Webb, 1979; Webb & Ray, 1979), which may result in lower advertising effectiveness. Danaher, Bonfrer and Dhar (2008) investigated the effect of competitive advertising on a focal brand's advertising elasticity and sales. Their results indicate strong negative influence of competitive interference on the advertising elasticity of the focal brand and, in turn, on its sales when at least one competing brand advertises in the same week as the brand of interest.

Our discussion of contributing factors to advertising effectiveness unequivocally suggests that advertising effects should exhibit considerable dynamics. It further indicates that the combined effects of product life cycle, message types, and competition could produce complex advertising effectiveness patterns, requiring a flexible modeling approach. The Augmented Kalman Filter methodology employed in this study can accommodate such complex advertising effectiveness paths, allowing us to pursue our main research questions.

## 3.4 Data

We use monthly data on number of prescriptions and marketing mix information for the pioneering drug in a recently developed therapeutic category in a non U.S. market. The data were obtained directly from the brand and cover both new (NRx) and refill (RRx) prescriptions over an 86-month period (March 1999 to May 2006). Each new prescription can be used for a certain number of automatic refills. After using up all the refills, a patient has to visit a physician again to renew the prescription, first use of which counts as a new prescription (NRx). Thus, new prescriptions correspond to both first-time adopters (*triers*) and ones who renew their prescription (*renewers*). Refills, on the other hand, correspond to prescribers refilling their existing prescriptions. Figure 3.1 depicts total, new and refill prescriptions for the focal brand during the observation period, with the entry times of the two competitors also marked. The drug was approved as the first effective oral treatment for the category in early 1999 and has received considerable media attention since, which may explain the high introductory sales in Figure 3.1. It treats a chronic disease suffered by a considerable percentage of the population (higher among certain demographic groups) and is covered by insurance plans. Although alternative (invasive, non-pharmaceutical) treatments were available prior to the launch of their pioneering brand, their incidence was negligible. Two competitors entered the market in the observation window, September 2003 and March 2004, with products of comparable efficacy and side effects.

Related marketing mix information includes television spending on generic and branded DTCA, discussed in Section 2, physician journal advertising expenditures ("PJ"), detailing ("DET") and sample pack volume ("SPV"). Price is not critical, possibly due to the lack of significant variation during the observation window. Management decided to air DTCA ads for the first time in early 2001, and the campaigns conformed to the regulations discussed in Section 2.

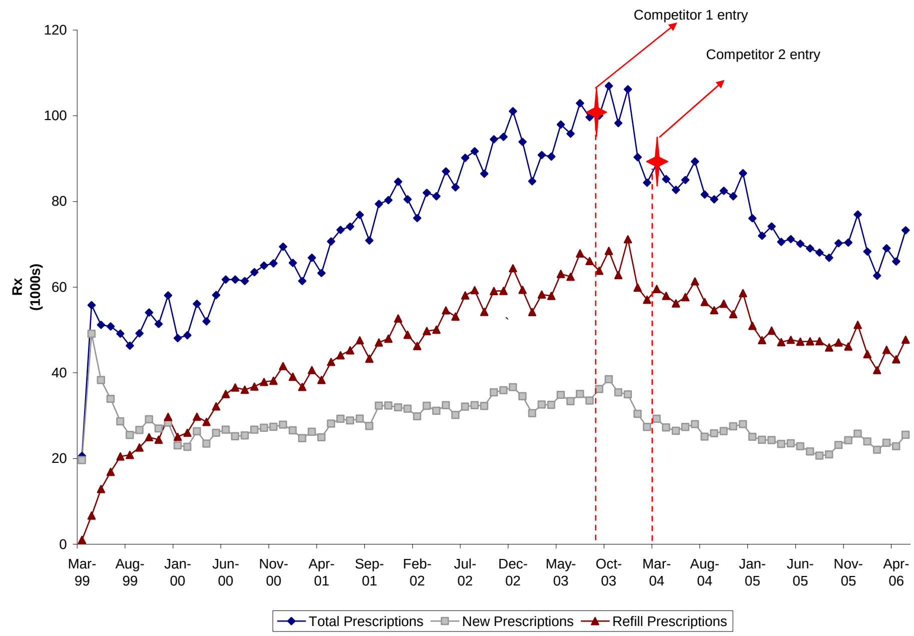



Figure 3-1: New, Refill and Total Prescriptions

The timing of the DTCA campaign was exclusively a management decision and was not dictated by a change or re-interpretation of regulations. Brand-commissioned research suggested that doctor visits and consultations increased by 50% as a result of the DTCA campaign. Table 3.3 contains summary statistics for the marketing mix variables[8]. PJ expresses the total

[8] The data is masked to preserve confidentiality.

expenditure for the medical journal advertisements mainly directed to the physicians. SPV, on the other hand, is the number of sample packages left with the physician during a product discussion with the sales representative. Detailing (DET) is a measure of effort spent by salesmen visiting doctors' offices explaining the benefits of their brand, informing the physician about the quality, specifications, indications and effectiveness of the drug.

Table 3-3: Descriptive Statistics

| | DET (units) | SPV (units) | PJ ($1000) | Branded ($1000) | Generic ($1000) |
|---|---|---|---|---|---|
| **Mean** | 4621 | 15846 | 64.2 | 232.9 | 65.8 |
| **Median** | 4467 | 15913 | 43 | 47.8 | 0 |
| **Standard Deviation** | 1829 | 10158 | 73.9 | 305.8 | 144.9 |
| **Minimum** | 998 | 872 | 0 | 0 | 0 |
| **Maximum** | 9059 | 38474 | 317.7 | 1141.7 | 730.6 |
| **Sum** | 397423 | 1362733 | 5523.9 | 20265.7 | 5722.3 |
| **Count** | 86 | 86 | 86 | 86 | 86 |

Because the majority of the detailing discussions end up with a drop of free samples by the sales person sampling and detailing efforts are usually highly correlated (e.g. Mizik & Jacobson, 2004; Wosinska, 2002). This is precisely the case in our data set as well, where correlation between SPV and DET was close to 80%. As a result, we decided to use only detailing in our analysis since it is typically used in marketing studies and captures a more frequent direct-to-physician activity than sampling since the latter requires detailing but not vice-versa.

## 3.5 Empirical Analysis

### 3.5.1 Modeling Framework

Our model focuses on capturing prescription growth and marketing mix effects for the pioneering brand. Due to the availability of both new and refill prescriptions in our database we can express total prescriptions (sales) of the focal brand as:

$$S_t = NRx_t + RRx_t \qquad (1)$$

Where:

$S_t$ =Total prescriptions at time t

$NRx_t$ =New prescriptions at time t

$RRx_t$ =Refill prescriptions at time t

As already discussed, NRx data include both first-time adoptions (trials) and renewals. Thus, NRx can be expressed as:

$$NRx_t = T_t + R_t + v_{1t} \qquad (2)$$

Where:

$T_t$ =Trial prescriptions at time t

$R_t$ =Renewal prescriptions at time t

and $v_1$ is an error term.

We model trials via a brand-level diffusion model with separable marketing mix effects (e.g. Dockner & Jorgensen, 1988; Kalish, 1983; Parker & Gatignon, 1994):

$$T_t = f(T_t^c)g(X_t) \tag{3}$$

The first function on the right-hand side of equation (3), f, represents the effect of cumulative trial prescriptions, $T^c$, whereas g captures the effects of other variables, X, such as marketing mix effort and competition. We assume that cumulative adoptions affect current adoptions according to the well-known Bass Model (Bass, 1969):

$$f(T_t^c) = \left[p + \frac{q}{M}T_t^c\right]\left[M - T_t^c\right] \tag{4}$$

where p and q are the coefficients of innovation and imitation respectively, and $M$ is the market potential. Prior research has suggested diminishing returns to scale for marketing mix effects on diffusion (e.g Bass, Krishnan, & Jain, 1994; Robertson & Gatignon, 1986). Hence, we model marketing mix effects using an element-wise square-root transformation of the vector of lagged marketing mix variables, X1. We account for competitive effects by including competitor sales, X2. Finally, we use an exponential specification to ensure non-negativity. Thus, we arrive at the following specification:

$$g(X_t) = \exp(\alpha_{1t}\sqrt{X_{1t}} + \alpha_{2t}X_{2t}) \tag{5}$$

Our use of lagged marketing mix variables, to allow for delayed effects, is consistent with previous pharmaceutical marketing studies (e.g. Hahn, Park, Krishnamurthi, & Zoltners, 1994). Incorporating (4) and (5) into (3) leads to the following expression for trial prescriptions:

$$T_t = \left[p_t + \frac{q_t}{M}T_t^c\right]\left[M - T_t^c\right]\exp(\alpha_{1t}\sqrt{X_{1t}} + \alpha_{2t}X_{2t}) \tag{6}$$

It should be noted that the use of a diminishing returns function for marketing mix effects does not preclude the optimality of pulsing advertising strategies, since our specification accounts for dynamic advertising effects (Naik et al., 1998; Vakratsas & Naik, 2007).

We model renewal prescriptions as a dynamic percentage of cumulative trials of the drug up to that point.

$$R_t = \rho_{1t}T_t^c \tag{7}$$

Combining (6), (7) and (2), new prescriptions can be written as:

$$NRx_t = \left[ p_t + \frac{q_t}{M} T_t^c \right]\left[M - T_t^c\right]\exp\left(\alpha_{1t}\sqrt{X_{1t}} + \alpha_{2t} X_{2t}\right) + \rho_{1t} T_t^c + \nu_{1t} \tag{8}$$

Finally, we model refill prescriptions as:

$$RRx_t = L_t + \nu_{2t} \tag{9}$$

where,

$$L_t = \rho_{2t} T_t^c \tag{10}$$

and $\nu_2$ is an error term. In other words, refill prescriptions are also assumed to be a dynamic percentage of cumulative trials. Our formulation of renewal and refill prescriptions as a percentage of cumulative adopters follows closely the treatment of repeat prescriptions by Shankar, Carpenter and Krishnamurthi (1998).

Thus, our model decomposes total prescriptions into three components: Trials, Renewals and Refills. Our formulation builds upon and extends the two-segment model of Shankar, Carpenter and Krishnamurthi (1998) who partition total prescriptions into trials and repeats. We are able to further decompose repeats into renewal prescriptions and refills due to the availability of both new (NRx) and refill (RRx) prescriptions data. Hence, our approach is well-grounded in previous pharmaceutical marketing research. It should also be noted that new and refill prescriptions are linked through the trial prescription process (6).

Equations (8) and (9) constitute our estimation model. However, the number of cumulative first-time adopters, $T_t^c$ , is not directly observable, since our new prescriptions data (NRx) do not differentiate between trials and renewals. This problem can be appropriately addressed using the Kalman filter (KF) methodology, which enables estimation of an unobserved process through the use of noisy observations. In our case, we are able to estimate the three processes given in equations (6), (7) and (10) by using the observation equations (8) and (9) for new and refill prescriptions. Further details will be provided in the estimation subsection that follows.

### 3.5.2 Estimation

We employ the Augmented Kalman Filter with continuous state and discrete observations, AKF(C-D) (Xie et al. 1997), to properly capture dynamic advertising effects. It is a recursive, adaptive technique that allows for efficient estimation of unobserved state variables. One distinctive element of our model is that we incorporate model parameters in the augmented state vector. Thus, we dynamically estimate model parameters simultaneously with market performance (i.e. number of prescriptions). The AKF(C-D) is a combination of extended Kalman filter (EKF) (Harvey, 1991) with continuous state and discrete observations, which estimates the state of a continuous system with known parameters, and an augmented filter (Stengel, 1986), which estimates unknown parameters of a continuous Kalman filter model.

Xie et al. (1997) use the AKF(C-D) for the univariate, continuous estimation of the Bass model. Among the few other studies which employed Kalman filtering in the estimation of dynamic models, methodologically the foremost ones are Naik, Mantrala and Sawyer (1998) and Naik, Prasad and Sethi (2008). The former estimated multivariate but linear dynamics of the modified Nerlove-Arrow model, while the latter estimated a discretized multivariate, nonlinear model of brand awareness. The dynamic system described in the previous section is multivariate and nonlinear, but unlike Naik, Prasad and Sethi (2008) it will be estimated in its continuous form without discretization in this paper, with the aim to eliminate biases resulting from estimating discrete models with analogous structure (e.g. Rao, 1986; Schmittlein & Mahajan, 1982).

In order to estimate our system of equations using Kalman filtering, we first write it in a state space formulation:

$$\begin{aligned}
&\text{Process Equation} \quad : \frac{dy(t)}{dt} = f_y(y(t), X(t)) + w \\
&\qquad \text{where,} \quad w \sim (0, W) \\
&\text{Measurement Equation} \quad : z_t = Hy_t + v_t \\
&\qquad \text{where,} \quad v \sim (0, V)
\end{aligned} \tag{11}$$

In (11), the process equation represents the evolution of the augmented system state, y(t), of sales and model parameters. Specifically, $y = \begin{bmatrix} T^c & R^c & L^c & \theta \end{bmatrix}'$ is a 12x1 vector, where $T^c, R^c \text{ and } L^c$ are the cumulative prescriptions from trials, renewals and refills $\left( i.e. \frac{dT^c(t)}{dt} = T, \frac{dR^c(t)}{dt} = R, \frac{dL^c(t)}{dt} = L \right)$, respectively, and θ[9] is the vector of parameters including $p, q, \rho_1, \rho_2, \alpha_1$ (for the four marketing mix variables) and $\alpha_2$. Moreover, $f_y = \begin{bmatrix} f_{T^c} & f_{R^c} & f_{L^c} & f_\theta \end{bmatrix}'$ is the function of how the state evolves over time, where the first three elements depict the prescription

[9] To obtain more stable parameter estimates, market potential M was excluded from the state vector due to lack of dynamics. Instead, we used a value based on managerial estimates of the population suffering from the disease the drug treats. These estimates are very accurate due to the availability of frequently updated statistics. As discussed in Mahajan, Muller and Bass (1990), using managerial estimates of market potential can lead to better forecasting results.

evolution defined by equations (6), (7) and (10), and parameter evolution $f_\theta$, or $\frac{d\theta(t)}{dt}$, is assumed to follow a random walk pattern (i.e. $\frac{d\theta(t)}{dt} = \sigma dw$).

Although the state vector is of interest at each time point, one cannot observe it directly but only through noisy observations represented by the measurement equations in (11). Specifically, $z$ represents the 2x1 observation vector of cumulative sales from new and refill prescriptions separately and $v_{2x1}$ is the measurement noise. $H_{2x12}$ is of the form [1 1 0 0..0;0 0 1 0..0] and links the system state with the observation vector. W and V are process noise covariance and measurement noise covariance matrices, respectively, which are assumed to be i.i.d.

The AKF(C-D) works as follows. At each time point, the state prior, $y_t^-$, is estimated conditional on all observations up to that point. The density of the state prior is assumed to follow a distribution with mean $\hat{y}_{t|t-1}$, and covariance $P_{t|t-1}$, which are both estimated within the AKF(C-D) algorithm. Once a new observation becomes available, the state prior is updated, and the mean, $\hat{y}_{t|t}$, and the variance, $P_{t|t}$, of the conditional posterior density are computed.

Briefly, the following relationship is assumed to hold (Ristic, Arulampalam, & Gordon, 2004), where $Z_t = \{z_1, z_2, ..., z_t\}$.

$$\text{prior density}: p(y_t^- \mid Z_{t-1}) \sim (y_t ; \hat{y}_{t|t-1}, P_{t|t-1})$$
$$\text{posterior density}: p(y_t \mid Z_t) \sim (y_t ; \hat{y}_{t\}t}, P_{t|t}) \qquad (12)$$

The algorithm starts immediately after the initial values for the state vector and estimation error covariance matrix, P, are set. Although there are major technical differences between the two as shown in Appendix 1, AKF(C-D) evolves by using a similar logic to the standard Kalman filter, and the state vector is continually updated as new observations become available. That is, once the filter starts at time $t_0$, the time update equations project the current state and error covariance forward in time to obtain a priori estimates for $t_1$, $\hat{y}_{t_1|t_0}$. Then the measurement update equations update the a priori estimates by weighing them proportionately to their distance from the noisy measurement to obtain an improved a posterior estimate of the state vector for $t_1$, $\hat{y}_{t_1|t_1}$. At this stage, AKF(C-D) enables the updating of the model parameters via equation (A4) in Appendix 1. Incorporating simultaneous estimation of uncertain parameters and statistics in the state estimation through the use of an augmented state vector, improves the quality (efficiency) of the estimates.

Additional information from the measurements is used to adapt the filter gains and, in turn, the parameters by a random effects type of approach.

In our model, we assume stochastic constant coefficients and do not impose a certain pattern for the dynamics of the model parameters over time (i.e. we assume $\frac{d\theta(t)}{dt} = \sigma dw$). However, their estimates may still vary in time and we can capture these changes in parameters through the prediction error ($v_t$) in a random effects fashion. In other words, the information about the change in model parameters which is reflected in the prediction error is used to generate new parameter estimates (Xie et al., 1997).

<u>Parameter Inference in AKF(C-D):</u>

The choice of starting values is critical for the Kalman filter algorithm. Deciding on the initial values of model parameters solely based on intuition could lead to biases, thus we employ a systematic grounded approach to determine them. We follow the recommendation of Sultan, Farley and Lehmann(1990) and use meta-analytic results to initialize starting values. Specifically, we initialize the coefficients of innovation and imitation based on the meta-analytic findings of Sultan, Farley and Lehmann (1990) who found that the grand means for the coefficient of innovation, p, and the coefficient of imitation, q, are 0.04 and 0.3, respectively. They also showed that including marketing mix variables into the diffusion model decreases the coefficient of imitation by 0.06, which leads to a value of 0.24 for q. We need to adjust for two more factors: first, the use of monthly observations, unlike the typical annual interval in diffusion studies and, second, the expectation that innovative effects should be higher than imitative effects at the inception of the product (e.g. Rogers, 1983). We adjust for monthly observations by dividing 0.24 by 12 resulting in a value of 0.02. To take into account possibly small imitation effects in the beginning of the product life cycle, we scale down the calculated monthly average q, 0.02, by one order of magnitude (i.e. a factor of 10) and end up with 0.002 as the starting value for the coefficient of imitation. We finally assume that the innovation effect would be initially 10 times that of imitation's, due to higher innovation effects at the introduction of the new product, and set the starting value for p to 0.02.

Similarly, we rely on previous studies reporting market-level elasticities of pharmaceutical promotions and analytically calculate the initial priors of corresponding parameters. Based on the meta-analytic results of Kremer et al.

(2008), we set the initial priors by assuming a value of 0.07 for starting DTCA elasticities[10]. For physician journal advertising (PJ) and detailing (DET) initial elasticities were assumed to be higher than those of DTCA as argued by the literature (see Manchanda et al., 2005), and set at 0.1[11] and 0.32[12], respectively. To further control the sensitivity to the initial conditions we calculate confidence intervals.

Given the high dimension of our state vector and the extent of nonlinearity in our model, the distributions of the various random variables are unlikely to be normal. Hence, confidence intervals will be difficult to obtain using the posterior distribution. However, making inferences about the parameter estimates is desirable. Hence, we randomize the initial prior vector, $y_0^-$, and employ Monte Carlo simulations to obtain confidence intervals for the means of the parameter estimates. Specifically, we start $m$ independent filtering processes in parallel which enables us to obtain actual empirical densities of the mean, $y_{t|t}$, and covariance, $P_{t|t}$, of the target posterior pdf, $p(y_t \mid Z_t)$ at each time point, t. We sample independently from the state prior at time $t_0$ [13], m=10,000 times, and update those independent filters within the AKF algorithm

[10] Rosenthal et al. (2003) and Kremer et al. (2008) report DTCA elasticities of 0.1and 0.06 respectively.
[11] Berndt et al. (1995), and Montgomery and Silk (1972) report journal elasticities of 0.18, and 0.15 respectively.
[12] Kremer et al. (2008) report an average elasticity of 0.326 for detailing.
[13] We use the multivariate normal distribution to randomize $y_0$ .

iteratively for the whole observation period, hence obtaining simulated distributions of mean states. Then, we derive 95% confidence intervals for the state variables using $2.5^{th}$ and $97.5^{th}$ percentiles of the corresponding simulated distributions, which also provide a measure of parameter sensitivity to the initial priors.

Thus, our proposed estimation extends the Xie et al (1997) approach by incorporating and estimating marketing mix effects and providing corresponding confidence intervals for all model parameters, which allow us to examine the sensitivity of posterior estimates to the initial priors at each time point. The full details of the Monte-Carlo AKF(C-D) algorithm are available in Appendix 1.

### 3.6 Findings

Our findings are best presented by plotting the estimated parameter paths over time[14]. Figures 3.2 and 3.3 depict the dynamic effects of DTCA and the rest of the marketing mix on total prescriptions. The dotted lines denote confidence intervals. Consistent with terminology used in Kalman Filter studies, we refer to the dynamic effect of a marketing mix variable, $\theta$, as “quality.” A cursory examination reveals that the model parameters exhibit considerable

[14] Our discussion focuses on the dynamics of marketing mix effects. The complete set of results is available to the interested reader by the authors upon request.

dynamics, providing an implicit validation of our approach. Generic and branded DTCA effects follow different paths[15]. Generic DTCA effectiveness is almost monotonically decreasing, whereas branded DTCA is characterized by buildup with an eventual decline.

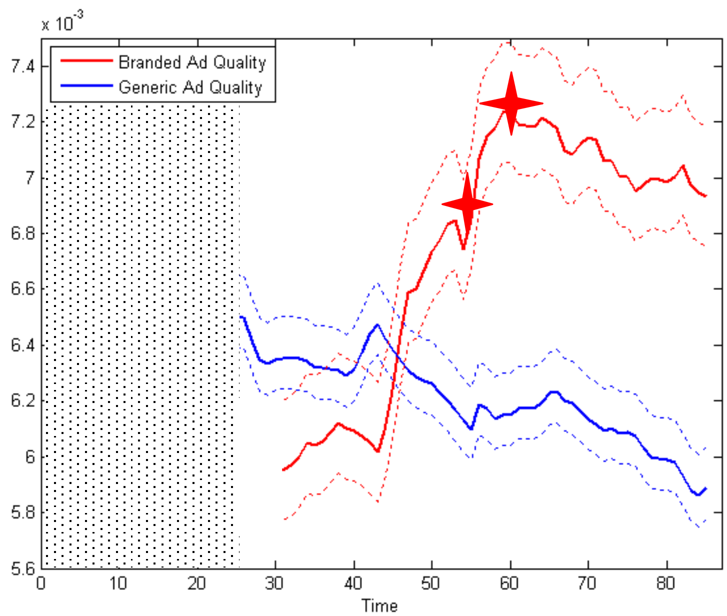



Figure 3-2: Generic vs. Branded Advertising Quality Dynamics

Although such paths can be partially attributed to spending allocation decisions, the brand started with generic advertising hence initially allocating more money to it, branded advertising spending had surpassed that of generic long before it

[15] Due to the lack of DTCA in the first two years following the launch of the brand, the parameter paths start at the 25th month.

reached its peak (see also Figure 3.4). Thus, the buildup in branded DTCA effectiveness should also be attributed to a delayed market response to this type of advertising, possibly due to limited public knowledge about the name of the available treatment and its benefits. Hence, branded advertising only starts to become effective as the market becomes more educated.

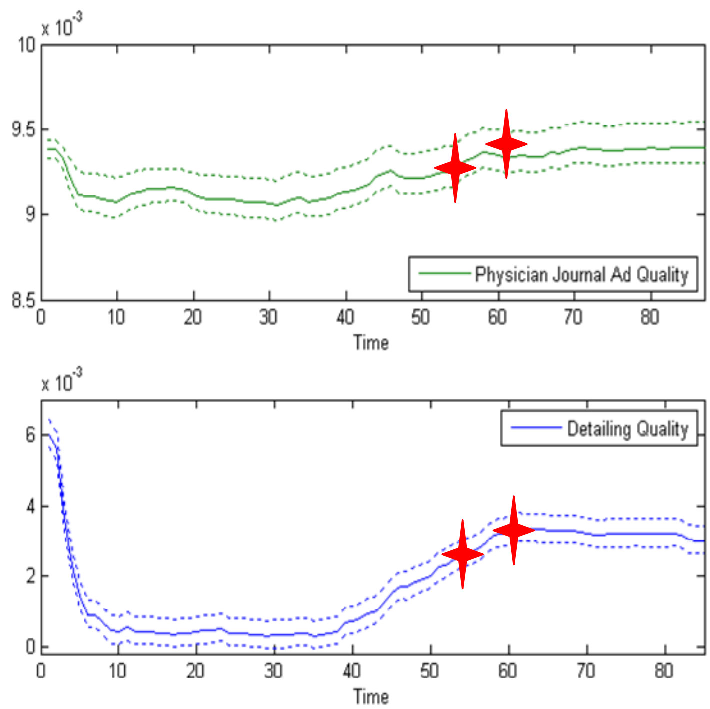



Figure 3-3: Physician Journal Advertising and Detailing Quality Dynamics

Similarly, the high initial effectiveness of generic advertising should be attributed to its high informational value for a young market largely unfamiliar with the product. As the market becomes more educated, the influence of generic advertising diminishes, following a classic product-life cycle pattern (e.g. Parsons, 1975). It should be noted, however, that even in the beginning, with the exception of the first few months of advertising, branded and generic advertising qualities are within each other's confidence bounds suggesting lack of significant differences. The peak of branded advertising quality is higher than that of generic and clearly outside the latter's confidence bounds, suggesting a significantly higher maximum value. Another interesting observation is that the end point of generic advertising quality is outside the confidence bounds of its peak although the same cannot be said of branded. This suggests that branded advertising is less susceptible to competitive effects than generic[16].

The above observations collectively indicate that branded advertising appears to be more effective than generic and more resistant to competition. The latter feature could be attributed to the reinforcing nature of branded advertising (mention of the brand name) in later stages of the brand's life cycle, as hypothesized in our background discussion. However, our findings also suggest that in the early stages of the product life cycle, generic DTCA is at

[16] Higher allocation could be an alternative explanation for higher branded advertising effectiveness. But it would also make branded advertising more susceptible to saturation, thus offsetting any potential gains.

least as effective as its branded counterpart. This should be good news for regulators who are concerned with adverse effects of DTCA. It suggests that the influence of “combative” branded advertising on young markets is rather limited, whereas there is an opportunity for informative generic messages to educate. A definitive answer regarding the size of the discussed effects will be provided in the next section with the examination of the corresponding elasticities.

Of the physician-directed marketing mix instruments, journal advertising remains relatively stable throughout the observation period as its movement is within the maximum and minimum observed confidence bounds. Detailing, on the other hand, exhibits considerable dynamics starting off very high, then dropping but recovering towards the end of the observation period after competitive entry. Thus, it appears that detailing can be effectively used to defend against competition although post-competitive levels never reach the high initial quality levels. Again, further examination through the calculation of elasticities will shed more light on this issue. The dynamic pattern of detailing quality is similar to that estimated by Narayanan, Manchanda and Chintagunta (2005) for “direct” detailing elasticities, lending thus validity to our findings.

Furthermore, the detailing quality path is different from those of generic and branded DTCA, highlighting the merit of investigating the dynamics of direct-to-consumer activities in addition to those directed to physicians.

One concern with adaptive estimation methods, such as the one employed here, is that estimates may be largely driven by the data, in our case observed prescriptions. To alleviate such concerns we compare quality and prescription paths. A careful examination of Figures 3.4 to 3.6 suggests that the qualities of the marketing mix variables are not driven by prescriptions (sales).

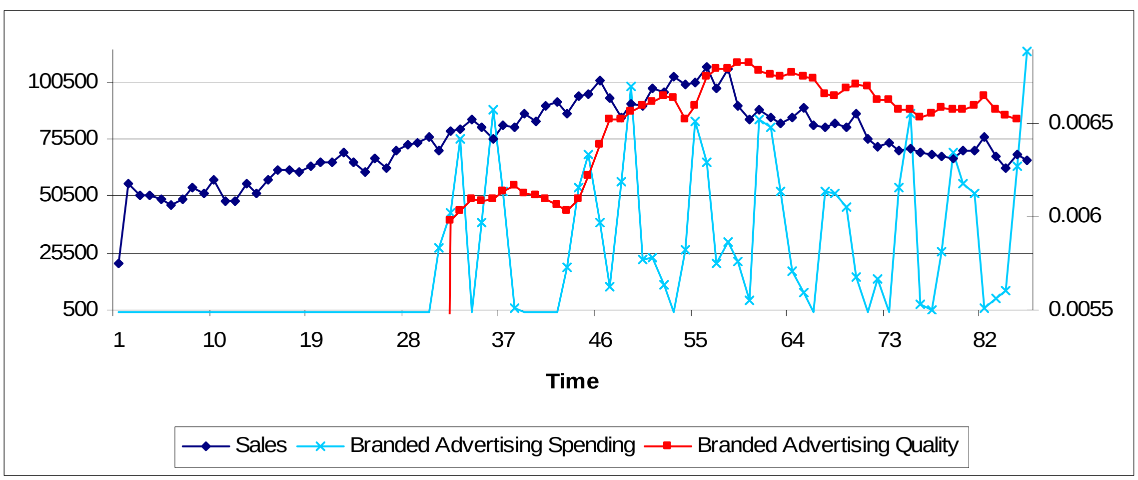



Figure 3-4: Sales vs. Branded DTCA

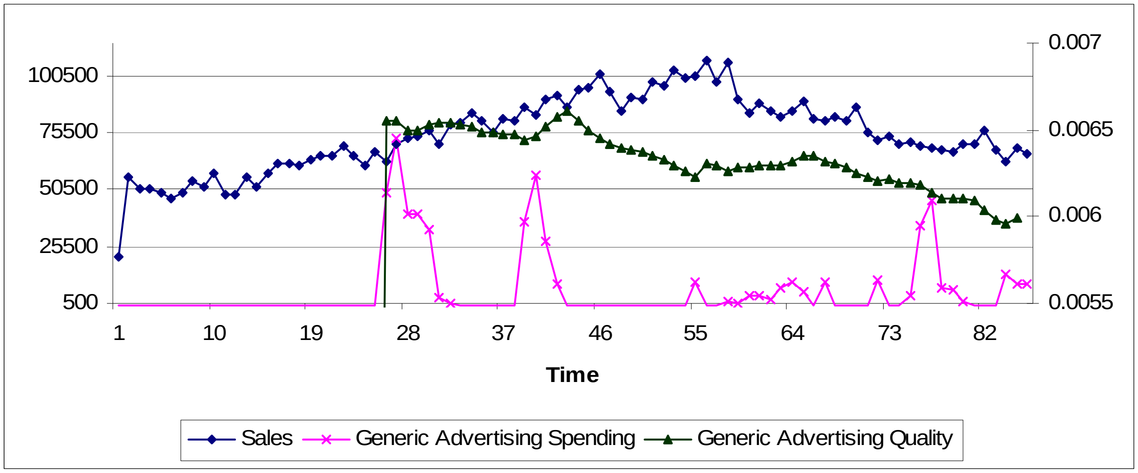



Figure 3-5: Sales vs. Generic DTCA

Specifically, quality is crossing paths with prescriptions for all three marketing mix instruments that exhibit dynamics (generic and branded DTCA and detailing). While the peak of branded advertising quality roughly coincides with that of prescriptions, it also corresponds to the period of highest advertising spending. More specifically, in the one year period which covers both branded advertising quality and sales peaks (i.e. months 49-61), the firm spent approximately one-third of its total branded advertising allocation. These observations were further formalized through correlation analyses which rejected any notion of systematic relationship between quality and prescriptions[17].

[17] Correlations between marketing mix qualities and prescriptions were low and insignificant ranging from 0.05 to 0.36.

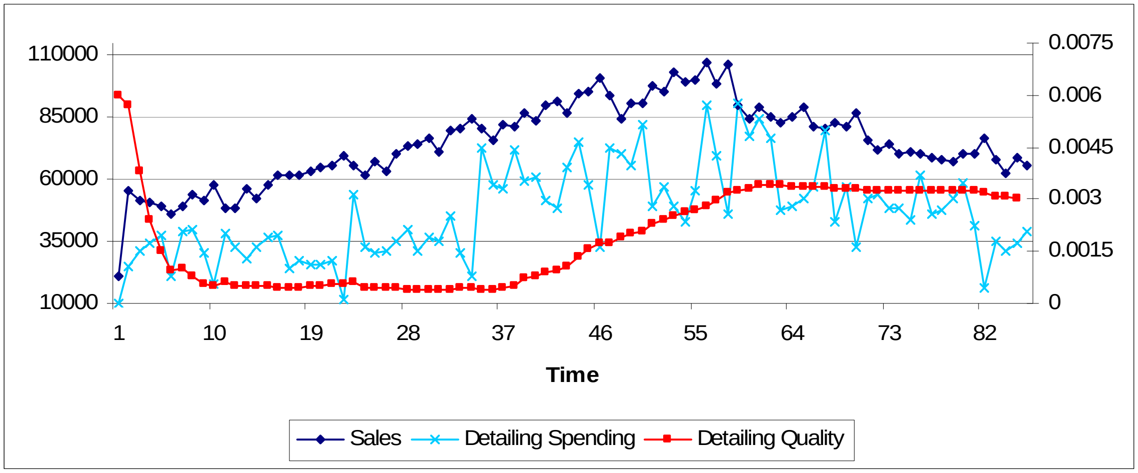



Figure 3-6: Sales vs. Detailing

### 3.7 Implications

Our findings suggest that DTCA and direct-to-physician effectiveness varies considerably over time and there are gains to be made from a timely placement of DTCA messages. To further substantiate these conclusions, we calculate the corresponding total prescriptions (sales) elasticities. This would allow us to answer the questions of whether DTCA is effective, which type of advertising is more effective and when.

Since our model specification allows for carryover effects through past prescription feedback, we calculate long-term elasticities. Specifically, we calculate the average 12- month effect of a 1% increase in spending using a moving window approach. We divide our observation period into “pre-” and “post-” competition to measure competitive effects. The 1 % increase in spending was applied only to marketing mix instruments for which actual spending was non-zero, to respect the firm's decision not to invest in a marketing mix instrument a certain period, preserving thus the original marketing mix scheduling. To obtain average pre-competition long-term elasticities, we covered the observation period until 11 months before competitive entry. We applied the same logic for the calculation of post-competition elasticities where we stopped 11 months before the end of the observation period. It should be noted that this method should lead to conservative elasticities, since only the months in which the brand advertised were included in the analysis, and spending in “off” advertising periods was kept at zero to preserve the original pulsing schedule. If increases from a zero base were to be implemented, they would have yielded higher returns due to our diminishing returns specification.

Table 3-4: Average Elasticities

| | Total Prescriptions (Sales) | | |
|---|---|---|---|
| | **Pre-Competition** | **Post-Competition** | **Overall** |
| **Generic Advertising** | 0.049 | 0.017 | 0.038 |
| **Branded Advertising** | 0.056 | 0.048 | 0.053 |
| **Physician Journal Adv.** | 0.037 | 0.033 | 0.035 |
| **Detailing** | 0.134 | 0.092 | 0.118 |

According to the results in Table 3.4, DTCA elasticities are generally modest and within the range of the values reported in many review and meta-analytic studies (e.g. Lodish et al. 1995, Vakratsas and Ambler 1999, Leone and Schultz 1980). Considering that the calculated DTCA elasticities are within the “operating range” of many other consumer goods, DTCA should be viewed as an effective instrument even under strict regulations. From this perspective, the answer to the whether question should be “yes.” The elasticities also formalize casual observations based on the examination of quality paths in Figure 3.2. Specifically, overall elasticities are higher for branded advertising. Thus, the answer to “which advertising format” question of is “branded.” This is largely due to the continued post-competition effectiveness of branded advertising, as suggested by the corresponding entries of Table 3.4. Thus, an answer to the when question should be, “branded advertising after competition." The answer for pre-competition is less straightforward as generic, despite lower allocation, is almost as effective as branded[18]. Thus, on the basis of effectiveness and the merit of informing the public, generic advertising should be considered by managers at least at the early stages of product life cycle. This suggestion is also consistent with the normative prescriptions of Bass et al. (2005) who recommend that relative allocation of generic advertising spending should be higher in the beginning. Hence, our findings provide an empirical validation for the Bass et al. (2005) recommendations.

Finally, detailing is the most effective marketing mix instrument as suggested by the calculated elasticities. Physician journal advertising is unaffected by competition but is also characterized by low effectiveness, lower than that of DTCA. This is a somewhat surprising finding, compared to the results of previous studies, which could be due to the particular product category and the presence of considerable DTCA activity.

[18] The elasticities for months 25-40, when cumulative spending was the same for each type of advertising, were as follows: generic: 0.033, branded: 0.0316.

## 3.8 Validation

To establish the appropriateness of our model and estimation approach we test its forecasting ability and compare it against alternative benchmarks. We compare the one-step ahead forecasting performance of our three-segment AKF (C-D) to different specifications and different methods of estimation. More specifically, we test the following benchmarks[19]:

- The two-segment model of Shankar, Carpenter and Krishnamurthi (1998) estimated using the AKF(C-D)
- A three-segment model estimated with a multivariate extended Kalman filter (MV-EKF)
- A Koyck model estimated with a Standard (linear) Kalman Filter (SKF)
- A three-segment model estimated with nonlinear least squares (NLS)
- A dynamic three-segment model estimated with NLS[20]

The NLS and MV-EKF techniques require a discretization of equation (11) prior to estimation. The process equation for the MV-EKF estimation is:

[19] We also tested the diminishing returns assumption by using a logistic functional form for g(X) (equation (3)). The prediction error statistics were as follows: MAD=486, MAPD=1.1 and MSE=1.02e+06.

[20] In order to reproduce dynamics similar to those suggested by our model, we used quadratic functions of time to introduce dynamic marketing mix effects in the NLS-estimated model.

$$\begin{bmatrix} T_t^c \\ R_t^c \\ L_t^c \end{bmatrix} = \begin{bmatrix} T_{t-1}^c + \left( p + \frac{q}{M} T_{t-1}^c \right) \left( M - T_{t-1}^c \right) \exp\left( \alpha_1 \sqrt{X_{1t}} + \alpha_2 X_{2t} \right) \\ R_{t-1}^c + \rho_1 T_{t-1}^c \\ L_{t-1}^c + \rho_2 T_{t-1}^c \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{bmatrix} \tag{13}$$

The Koyck model (Naik & Raman, 2003; Palda, 1964), which is relatively more parsimonious and commonly used in practice (Bucklin & Gupta, 1999, p. 262), has the following form:

$$S_t = \alpha_0 + \lambda S_{t-1} + \alpha_1 \sqrt{X_{1t}} + \alpha_2 X_{2t} \tag{14}$$

The MV-EKF and Koyck models are non-dynamic, providing an opportunity to test the relative importance of dynamic effects when using an adaptive estimation method.

Table 3-5: Prediction Error Comparison

| | **Adaptive Estimation** | | | | **Conventional Estimation (NLS)** | |
|---|---|---|---|---|---|---|
| | **2-segment model AKF(C-D) (Dynamic)** | **3-segment model AKF(C-D) (Dynamic)** | **3-segment model EKF** | **Koyck model SKF** | **3-segment model** | **3-segment model (Dynamic)** |
| **MAD** | 929.11 | 469.62 | 971.57 | 3096.9 | 4475.8 | 3348 |
| **MAPD** | 1.68 | 0.88 | 1.42 | 5.39 | 6.51 | 5.57 |
| **MSE** | 2.62E+06 | 9.72E+05 | 1.44E+06 | 1.49E+07 | 3.48E+07 | 2.12E+07 |
| **RANK** | 3 | 1 | 2 | 4 | 6 | 5 |
| **AIC** | | | | | 1509.40 | 1474.78 |
| **BIC** | | | | | 1529.04 | 1504.23 |

Table 3.5 contains prediction statistics for all models. A number of points deserve mention. For each estimation method (adaptive and conventional), dynamic models perform better than non-dynamic[21]. Prediction errors for the two-segment model are approximately twice as large as those from the three-segment AKF(C-D) model. Models estimated with Kalman Filtering outperform NLS, advocating an adaptive estimation method. Among the adaptive estimation methods, MV-EKF errors are about twice those of AKF(C-D), although considerably lower compared to the rest of the benchmarks. This difference should be attributed to (1) the relative importance of dynamic effects, and (2) gains due to the continuous-discrete estimation. Thus, our validation results suggest that dynamics are important regardless of the estimation method, but adaptive is superior to conventional estimation. Figure 3.7 depicts the tracking performance of AKF(C-D).

[21] The AIC for the dynamic NLS model was 1474.8 vs. 1509.4 for the non-dynamic one.

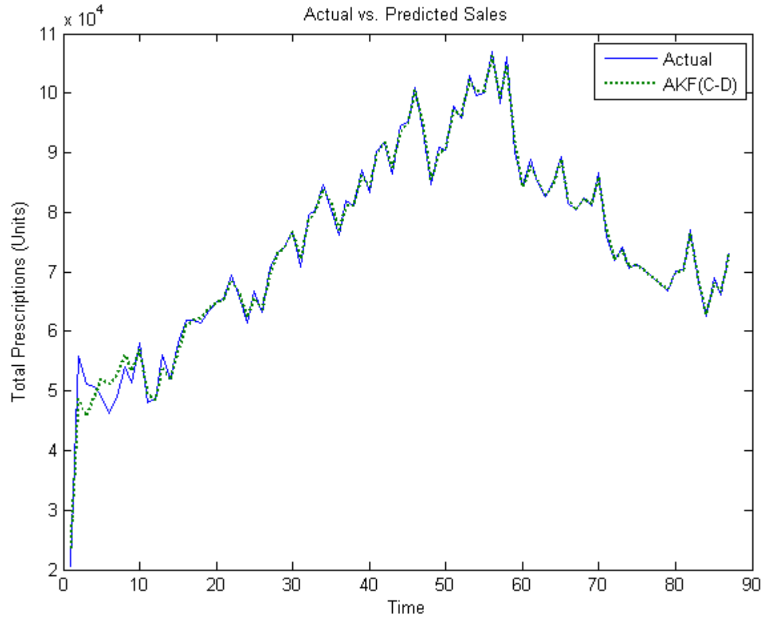



Figure 3-7: Actual vs. Predicted Sales

## 3.9 Conclusion

Regulations impose considerable restrictions on the format and content of advertising messages, thus introducing high uncertainty regarding their effectiveness. Consequently, managers of brands competing in regulated markets should consider whether or not to advertise under these conditions and assess potential gains from such advertising. In this paper, we explore dynamic DTCA effects in a highly regulated market, where firms should choose between branded and generic messages each time they wish to advertise. We pose three research questions which should be of great interest to managers:

Whether advertising is effective under such strict regulation conditions; Which type of advertising message is more effective? When? We pursue these questions by examining data on a pioneering brand of a new therapeutic category. We apply the Augmented Kalman Filter with continuous state and discrete observations (AKF(C-D)), to estimate dynamic marketing mix effects without discretization biases. To our knowledge, the implementation of this multivariate continuous nonlinear estimation is new to the marketing literature.

Our findings suggest that DTCA, as well as direct-to-physician activities, exhibit complex dynamics. DTCA elasticities are modest, consistent with previous findings in the pharmaceutical literature, but within the operating range of mass-marketed consumer goods, suggesting that brands can benefit from advertising even under strict regulatory conditions. Thus, our answer to the whether question is “yes.” When comparing the effectiveness of generic and branded advertising messages, our results clearly indicate that branded advertising is more effective than generic. Thus, the answer to the which question is “branded.” We attribute the higher effectiveness of branded messages to their reinforcing nature: the mention of the brand name, absent from generic messages, can defend the brand against competition and the lower informational content makes this type of message more appealing to a mature, educated market (Chandy et al., 2001). Generic advertising effectiveness, on the other hand, is limited to early pre-competition stages potentially due to its higher informational value. The early effectiveness of generic advertising, coupled with a corresponding low initial branded advertising effectiveness should be good news for regulators, who are concerned with adverse effects of DTCA. It suggests that young markets rely less on “combative” branded advertising messages whereas there is an opportunity for informative generic messages to educate through their initial effectiveness. Hence, the final when question can be answered with “branded after competition,” but generic should be useful at the early stages of the product life cycle.

Our study contributes to the advancement of knowledge on advertising effects in two ways: First, using an appropriate model representation, it explores dynamic effects of DTCA. No other study has examined dynamic DTCA effects, and relatively few studies have examined the issue of advertising dynamics in general. Thus, our work complements the seminal study of Narayanan, Manchanda and Chintagunta (2005), which investigates the dynamics of physician-directed activities. The different dynamic paths estimated for DTCA and detailing effects, emphasize the merit of uncovering the dynamics of direct-to-consumer activities in addition to those directed to physicians. Second, it is the first field study in marketing to empirically assess the effectiveness of purely generic advertising messages for a differentiated product. Thus, it complements theoretical studies on the subject (Bass et al. 2005), which prescribe optimal allocation policies for generic and branded advertising. Our findings on higher branded advertising effectiveness in the latter stages of the product life-cycle provide support for the Bass et al. (2005) recommendation on progressively higher branded advertising allocation. Also, from a methodological point of view, we extend the AKF(C-D) algorithm by randomizing initial state priors through the use of Monte Carlo simulation to provide confidence intervals for the posterior estimates of the parameters at every time point. Validation results confirm the superiority of our proposed approach compared to existing alternatives.

A limitation of our study is that our findings are confined to a single category in a single market (country). Applying the proposed methodology to multiple markets/countries representing different regulatory environments (e.g. Stremersch & Lemmens, 2009) would provide further valuable insights on the role of regulations in advertising effectiveness. Also, we did not consider interaction effects of the marketing mix (Naik & Raman, 2003) largely due to the prohibitive requirements they impose on the dimension of the state vector. Exploring dimension reduction techniques which efficiently select the most important interaction and main effects should be a topic worthy of future research (see also Vakratsas, 2005).

# CHAPTER 4

# DUAL-MARKET DIFFUSION FOR A NEW PRESCRIPTION PHARMACEUTICAL

## 4.1 Introduction

Prescription pharmaceuticals have become an increasingly important component of patient care and a critical factor in the economics of modern health care systems. Trends such as rapid growth in prescription pharmaceutical spending, a shift in health care spending towards prescription drugs and away from inpatient care, and finally an upwards shift in age demographics are all contributing factors to the importance of prescription drug-related matters (Zuvekas & Cohen, 2007). As a result, disciplines such as pharmacoeconomics and pharmacoepidemiology, which are concerned with the use and the effects of drugs in large numbers of people, have experienced tremendous growth lately (Strom, 2000). The study of prescription pharmaceuticals has also become increasingly important from a marketing perspective as recent years have seen a soaring in marketing expenditures, especially in North America largely due to deregulation in patient-directed marketing activities such as direct-to-consumer advertising (Findlay, 2001). Since many new drugs represent breakthrough innovations in the treatment of diseases, a fundamental research question for the marketer is how new prescription pharmaceuticals are initially adopted. This can be further refined into two issues of particular importance: What kind of pattern does the diffusion of a novel prescription pharmaceutical exhibit? What is the effect of marketing spending on such diffusion? Interestingly, despite the proliferation of diffusion models and a plethora of applications in many different markets and domains, relatively little research has been conducted in the area of pharmaceutical drug diffusion (for recent examples see Desiraju, Nair, & Chintagunta, 2004; Hahn et al., 1994; Van den Bulte & Lilien, 2001). Yet, pharmaceutical markets are characterized by distinctive elements that may lead to unique diffusion patterns, calling for a “custom-made” approach in the modeling of novel prescription drugs.

The uniqueness of prescription pharmaceutical markets is worthy of elaboration. First, pharmaceutical markets can be quite different from “typical” product-markets in the way they are created. While for a typical consumer market the product may “create” the market (e.g. demand for CDs or Music Players), in the case of pharmaceuticals the market is typically created from the patient’s physical need for treatment as diagnosed by his physician. This need will be unfulfilled if no prescription treatment is available at the time of the diagnosis. Thus, in the case of pharmaceuticals, markets may precede the launch of a prescription treatment, suggesting an accumulation of demand prior to product launch. When a prescription drug is eventually launched, this accumulated demand could be immediately realized, due to the high need of already diagnosed patients, leading to atypical diffusion patterns. Second, in addition to the prescription market for patients with severe, “classic” and thus easily diagnosed health problems, a second market may exist, corresponding to patients with mild health problems or lack of visible, persistent symptoms. Physicians may not be able to diagnose the problem for such patients, due to the mildness of the symptoms, or opt to delay any treatment and exercise “watchful waiting” instead (e.g. Hyde et al., 2005; Strom, 2000). Thus, in this case, demand is unlikely to be accumulated prior to product launch. However, the launch of a new pharmaceutical would raise awareness about the problems it treats, and may trigger demand for the market corresponding to patients with mild problems by increasing physician and patient vigilance about symptoms. Furthermore, the availability of the prescription drug may present physicians with an easy-to implement or least harmful treatment for patients with mild symptoms. This is consistent with findings in the medical literature regarding the effects of drug launches on prescriptions. For example, studies have found that the launch of SSRIs (Selective Serotonin Re-uptake Inhibitors) in the late eighties, one of which is the well known Prozac, prompted a phenomenal growth in the prescriptions of antidepressants both in the US and the UK (Martin, Hilton, Kerry, & Richards, 1997; Middleton, Gunnell, Whitley, Dorling, & Frankel, 2001; Olfson & Klerman, 1993). Thus, prescriptions for this market will be adopted only after the launch of the pharmaceutical, but not necessarily immediately, suggesting a more “mainstream” adoption pattern that exhibits similarities to those for typical product markets, captured by established diffusion models (Bass, 1969).

Due to the difference in the severity of health problems and symptoms for the corresponding patients, and the time lag in prescription adoption, these two markets may be at best weakly connected. This essentially suggests the presence of two disconnected markets: an “early” market corresponding to prescriptions for patients with severe problems for which demand is potentially accumulated before the launch of the pharmaceutical and a “late” market corresponding to prescriptions for patients with mild problems. Physicians could promptly write prescriptions for patients with severe problems as soon as the pharmaceutical becomes available, especially if these have been diagnosed before the launch of the pharmaceutical, but may delay prescription treatment (watchful waiting) for those with mild problems. Thus, the timing of physician prescription adoption will vary depending on the severity of health problems of patients leading to two temporally separated markets: One corresponding to patients with severe health problems for which prescription treatment will be adopted immediately, and another corresponding to patients with mild problems for which prescription treatment may be delayed. Consistent with our previous discussion, we will refer to the first market as the “early” market and the second as the “late” market.

The importance of these two characteristics of pharmaceutical markets, accumulated demand for the product and the presence of early and late markets, for the diffusion pattern can be demonstrated through the following stylized example. Patients with severe health problems and classic symptoms are diagnosed early and typically before an appropriate prescription treatment is available. This is not an unreasonable assumption since most disease epidemics precede pharmaceutical launches, in fact they trigger them. The demand of these patients cannot be satisfied until the first such drug is launched in the market at time, say, T. In other words, physicians ideally would like to provide a prescription treatment for patients with severe problems as soon as they come up with a diagnosis, but are unable to do so until the time T of such drug launch. Thus, the ideal adoption times for these patients cannot be immediately realized, since they precede the drug launch time T. This gives rise to an ideal adoption curve for patients with severe, clearly diagnosed health problems, the early market. We assume that the hazard rate h of the ideal adoption times $\underline{t}$ for the early market follows the Bass Model:

$$h(\underline{t}) = p + qF(\underline{t})$$

with F denoting the corresponding cumulative distribution function. If the pharmaceutical is launched late enough, i.e. $T > \underline{t}$ for most $\underline{t}$, then at the time of launch, T, the hazard rate of actual adoption for the early market will be approaching p+q, the maximum value of the ideal times hazard rate, since no one actually adopts the product before T due to its unavailability. This corresponds to an exponential distribution with parameter p+q for the realized adoption times suggesting that the early market's actual adoption times, t, will exhibit a diffusion pattern similar to that of a market consisting exclusively of innovators[22]. The difference, however, is that in our case the "imitation" effect q of the ideal adoption process is built in the exponential parameter (p+q) of the realized adoption curve, leading to a much larger scale of adoption than that typically realized for a "purely innovative" market[23]. In other words, although the adoption curve is *seemingly* driven purely by innovators, its generating mechanism and scale are quite different. For the late market we assume that the ideal adoption times perfectly coincide with the realized adoption times t. In other words, demand is not accumulated prior to launch due to mildness of problems or unidentifiable symptoms, but could be "triggered" by it as

[22] This can also be seen using the closed form expression of the Bass hazard rate which increases sharply initially, but reaches a plateau for medium to high values of time. Thus, even if there is some overlap between ideal and actual adoption times past the ideal times peak for the early market, in other words $\underline{t} > T$ for some $\underline{t}$ past the peak of the ideal adoption times, the hazard rate will be roughly constant in the actual adoption time domain corresponding to an exponential-like adoption distribution. More formal mathematical arguments for the exponential-like shape of the distribution for the early market's realized adoption times are provided in Appendix 2.

[23] Our illustration is somewhat similar to that of Muller, Peres and Mahajan (2007) in their elaboration of the "shadow diffusion" concept. However, in their illustration (Figure 7.1), pre-launch adoption decisions do not translate into accumulated demand realized immediately after the launch of the product. The concept of shadow diffusion can be traced back to the work of Givon, Mahajan and Muller (1995) on software piracy, which suggests that there are two parallel markets, a legal one and a "shadow" piracy market. However, our conceptualization of shadow diffusion is closer to that of Muller, Peres and Mahajan (2007) in that the shadow diffusion precedes the realized one.

previously discussed. Ideal and realized adoption times for the early market and actual adoption times for the late market are depicted in Figure 4.1 for the case where ideal adoption times for the early market precede product launch time T as discussed above. In other words, the ideal and actual adoption time domains for the early market are entirely separated by the launch time T of the pharmaceutical.

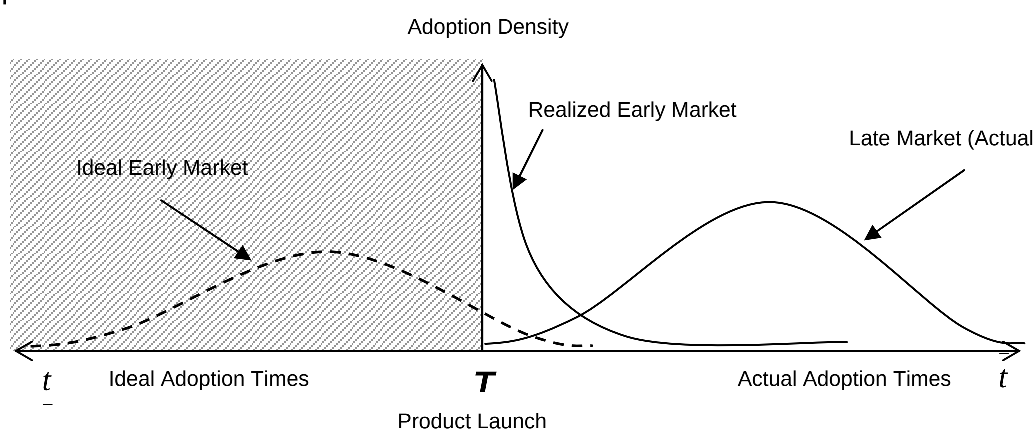



Figure 4-1: Stylized Diffusion Pattern for Early and Late Markets

The figure captures the following sequence of events (from left to right). The early market corresponds to adoptions for patients with severe, diagnosed problems and thus a clear need for treatment that ideally should be satisfied before the launch of the product (shaded part of Figure 4.1). Product launch follows and "sets the clock" for actual adoption times. At the time of the launch, the hazard rate of adoption for the early market is constant and equal to its maximum value, since ideal adoption times precede launch but no adoption has been made yet. Consequently, accumulated demand for the drug is realized for the early market in the form of an exponential, seemingly pure innovation-style adoption curve. In other words, the realized adoption curve for the early market is the result of the accumulation of (unsatisfied) pre-launch demand represented by the ideal adoption times. Finally, adoption begins for the late market whose need may also be potentially triggered by the launch of the pharmaceutical. Hence for the late market, actual and ideal adoption times are identical.

In sum, Figure 4.1 suggests that the realized adoption times of the early market are shifted to the right due to the prior lack of the prescription pharmaceutical to satisfy the needs of that market, truncating thus its adoption process. This also results in a closer temporal proximity of the two markets as they appear to be more like two segments of the same market rather than two separate markets! Furthermore, the realized adoption curve for the early market appears to be purely driven by innovation although it can be the product of a demand accumulation. Thus, the launch of the pharmaceutical essentially “left-truncates” the adoption process, leading to unusually high initial adoption rates and shortening the time distance between the two markets.

It should be noted that this dual-market notion we propose is conceptually very different from that suggested for technological markets (e.g. Goldenberg, Libai, & Muller, 2002; Moore, 1992). In the latter case, the early market consists of high-risk product enthusiasts with a high appreciation of novel technological features regardless of their practical usefulness (or “effectiveness”) and the late market consists of more conservative, utilitarian adopters concerned with a product’s functionality.

Thus, proneness for innovative behavior appears to be the driving force for early adoption of technological products. In our case, the two markets differ in terms of their demand intensity, with the early market being characterized by a high demand for the drug, driven by a well-defined, diagnosed, physical need rather than innovativeness. This also implies that realized adoption by the early market will depend on the product’s efficacy and not much else. We further suggest that early market adoption is the result of a pre-launch, shadow-type diffusion capturing the market’s unsatisfied demand. The idea of pre-launch diffusion has been previously discussed or studied in other contexts. What uniquely characterizes our case is that pre-launch demand is independent of the product launch, or its anticipation, and is driven by a pre-existent physical need for the product rather than hype, as in the case of movies (e.g. Muller, Peres, & Mahajan, 2007), launch pre-announcement, restriction of supply (Dye, 2000), or the option of pre-ordering (Hui, Eliashberg, & George, 2008; Moe & Fader, 2002). We therefore expect that the early market will most likely not be influenced by marketing activities, a hypothesis that is directly testable within the framework of the model we propose. Prescription adoption for the late market, on the other hand, may exhibit more typical product-market characteristics such as marketing mix influence as it is characterized by lower demand intensity due to the mildness of their health problems. For example, it has been suggested that marketing activities such as Direct-to-Consumer Advertising (DTCA) may prompt physicians to "overprescribe" for patients with mild problems and encourage patients to seek unnecessary treatments (Kravitz et al., 2005). This further implies that the late market's potential may be considerably larger than that of the early market, since it corresponds to a much broader population with less well defined needs and diagnosed symptoms.

Our dual-market explanation for new prescription pharmaceutical adoption is particularly appropriate for drugs that treat a disease of which the severity is located on a bipolar spectrum (ranging from mild to severe), rather than a disease manifested in a dichotomous manner (presence/absence of disease). Many lifestyle-induced diseases such as dyslipidemia, diabetes, ED, depression and dementia fit this profile. In such cases, a distinction can be made between patient categories at the extreme ends of the spectrum (mild/severe), however for patients located at intermediate points of the spectrum no substantial discontinuities are observed (e.g. Peralta & Cuelta, 2007; Shah & Reichman, 2006). This discernible difference between patients at opposite ends of the severity spectrum is capable of leading to a dual market adoption pattern. In particular the case of drugs with rare and mild side effects is highly applicable, as they allow patients with less severe problems to adopt a treatment at relatively low risk. Adoption of drugs for infectious diseases, on the other hand, may not fit the proposed pattern very well, since such diseases manifest in a dichotomous manner and exhibit heavy epidemic patterns (e.g. tuberculosis).

This line of argument is further corroborated by medical literature advocating a physician classification based on prescription volume, distinguishing between “high” and “low” prescribers (Prosser, Almond, & Walley, 2003; Prosser & Walley, 2003). These studies suggest that low prescribers adopt a “wait and see” approach, similar to the concept of watchful waiting, lending additional support for the idea of a late market. Also, low prescribers tend to follow a gradual, cumulative adoption process, indicating the strong influence of accumulated information, consistent with the Bass model of adoption of innovations. High prescribers, on the other hand, tend to adopt a new drug early only when the benefit-to-risk ratio would be greatest, suggesting the presence of an early market for severe health cases, which is also argued in this study (see also Stremersch & Lemmens, 2009).

In this paper, we propose a switching regime dual-market diffusion model that is designed to capture the idiosyncratic elements of pharmaceutical markets. Our model extends the well-known Generalized Bass Model (GBM) due to Bass, Krishnan and Jain (1994) by explicitly accommodating the possibility of the presence of an early market which is disconnected from the late main market. We examine new prescriptions only, thus focusing on first time adoptions, consistent with the diffusion literature. From a diffusion perspective, our study contributes to the growing body of work published in this journal (e.g. Cestre & Darmon, 1998; Jiang, Bass, & Bass, 2006), aiming to enhance our knowledge of the phenomena underlying the innovation adoption process. From a pharmaceutical marketing and health perspective, our study is similar in spirit to that of Desiraju, Nair, and Chintagunta (2004), published in this journal, in linking pharmaceutical marketing activities to new drug adoption behavior.

The rest of the paper is organized as follows. In Section 2 we briefly discuss related literature. Section 3 provides a detailed exposition to our model. In Section 4 we discuss the data and our estimation procedure and in Section 5 we elaborate on our findings. We conclude in Section 6 by summarizing our contribution, discussing the implications of our findings and identifying opportunities for future research.

### 4.2 Related Literature

Two literature streams are related to our study. One is concerned with pharmaceutical diffusion research and the other is concerned with research on dual markets in the context of diffusion of innovations. In terms of the first literature stream, recent papers closely related to our study are those by Hahn et al. (1994) and Desiraju, Nair, and Chintagunta (2004)[24]. Hahn et al. (1994) present a comprehensive four-segment trial and repeat model and test it with data on 21 pharmaceutical products from seven different therapeutic classes. They find it has good forecasting ability and yields appropriate estimated values. However, their model is at the brand level and assumes that the number of category adopters is constant over time. In other words, there is no evolution in adoption but rather the market is saturated and brands compete for a static pool of customers who have previously adopted one of the available brand alternatives[25]. A version of this model was also used by Shankar, Carpenter, and Krishnamurthi (1998) in a study of the effects of late entry in prescription pharmaceutical markets. The study by Desiraju et al. (2004) is more closely related to our approach in that it considers category-level diffusion. They use a logistic growth model proposed by Van den Bulte (2000) to compare diffusion speeds and pricing effects in developed and developing countries. However, they do not examine the possibility of a dual market mechanism and do not investigate the effects of marketing efforts such as

---

[24] More, but earlier, related references are discussed in Hahn et al (1994) and thus will not be further discussed here. We refer the interested reader to the original article for such discussion.

[25] In the Hahn et al (1994) model, prescription evolution is captured through growth in usage rate of adopters.

detailing or advertising, as our study does. Furthermore, they use total prescriptions (or "sales in kilograms") whereas we focus on first adoptions only, consistent with diffusion studies, by examining new prescriptions.

In terms of the dual-market research stream, the study by Goldenberg, Libai, and Muller (2002) is the most relevant to our approach. The authors argue for the presence of two distinct markets, early and late, for technological products. The early market consists of product enthusiasts with an appreciation of advanced technological features whereas the late market consists of utilitarian adopters seeking functionality. Applying cellular automata at the individual level, they demonstrate that the saddle phenomenon,[26] frequently observed in technology adoption data, can be attributed to the limited communication between the early and late markets. This effectively is confirmed by Van den Bulte and Joshi(2007) who show, through an examination of different cases, that a saddle is more likely to be observed for low values of "w," which represents the cross-influence parameter in their Asymmetric Influence Model (AIM). Goldenberg, Libai, and Muller (2002) also estimate an aggregate-level dual market model, the exact details of which are

[26] Goldenberg, Libai, and Muller (2002) define the "saddle" as a pattern in which an initial peak predates a trough of sufficient depth and duration to exclude random fluctuations, which is followed by sales eventually exceeding the initial peak.

not provided to the reader, by using mean values of the stochastic cellular automata model. This does not exactly correspond to a Bass diffusion model or an extension of it, which has typically formed the basis of most aggregate diffusion empirical analyses. Furthermore, the effects of the marketing mix, including potential differences in the influence of early and late markets, are not considered, a dimension we introduce in this study. Finally, they assume an imitation parameter for the early market, whereas we claim that due to early market pre-launch demand, the realized adoption rate will be exponential in shape. In the next section, when we discuss the details of our model specification, we will revisit differences and similarities with models previously proposed in the literature.

### 4.3 Dual-Market Model

Based on our opening discussion, the fundamental premise of the dual market hypothesis in a new prescription pharmaceutical category lies in the assumption that the two markets are distinct in the following two ways:

1) There is no link between the two markets, i.e. their adoption processes are disconnected

2) There is a temporal order in the adoption of the two markets, i.e. there is an early market adopting first, followed by a late market.

The above assumptions, which are also in agreement with the Goldenberg, Libai, and Muller (2002) view on the dual-market phenomenon, can then translate into three defining characteristics of a potential dual-market diffusion model:

a) the two markets have different potentials
b) each market's adoption is not influenced by the other, i.e. there is no cross-market influence
c) prescriptions due to the early market precede prescriptions due to the late market.

We express our model in terms of prescriptions due to the availability of aggregate prescription data. Guided by the three defining characteristics above, we propose the following model:

$$N_t = I_{[d_t=1]} N_{1t} + (1 - I_{[d_t=1]}) N_{2t} + \varepsilon_t \qquad (1)$$

Where N denotes new prescriptions ("NRx"), t denotes time, 1 corresponds to the early market and 2 corresponds to the late market. I is an indicator function and $d_t$ is a dummy variable denoting the market from which prescriptions originate. Specifically:

$$d_t = \begin{cases} 1 \text{ if prescriptions at time t are due to the early market} \\ 0 \text{ if prescriptions at time t are due to the late market} \end{cases} \tag{2}$$

We subsequently express prescriptions due to each market using the adoption time domain, which has shown to capture well the underlying adoption pattern in diffusion studies (Srinivasan & Mason, 1986). Thus, we employ the following expressions for the prescriptions corresponding to the two markets.

$$N_{1t} = M_1[F_1(t) - F_1(t-1)] \tag{3}$$

$$N_{2t} = M_2[F_2(t) - F_2(t-1)] \tag{4}$$

Where F denotes the cumulative distribution function (cdf) for adoption times. Thus, equations (3) and (4) link the time and prescription domains. We use differences in cdf's since we employ discrete data.

Following up on our introduction discussion and the stylized example, we model the adoption timing of the early market using the exponential distribution to account for the phenomenon of pre-launch accumulated demand. Further, we assume that the influence of marketing activities will be negligible since the need of this market for the prescription pharmaceutical is clearly defined and driven by severe health problems:

$$F_1(t) = 1 - \exp(-p_1 t) \quad (5)$$

where $p_1$ can be considered as the innovation parameter in a Bass Model with a zero imitation coefficient[27]. As we discussed in the introduction, the exponential form of adoption for the early market may have seemingly the characteristics of a pure innovator adoption, however this may be the result of pre-launch accumulated demand.

For the late market, keeping with the long-standing tradition of using the Bass Model, or one of its extensions, as the standard for modeling innovation

[27] The assumptions of exponential adoption times and zero marketing mix influence are both testable for the early market. To test the exponential hypothesis, the standard Bass model can be estimated to allow for inferences about the imitation coefficient. Similarly, to test the presence of marketing mix effects a model similar to the one proposed for the late market can be assumed for the early market.

diffusion, we use the Generalized Bass Model (GBM) (Bass et al., 1994), to allow for cumulative marketing-mix effects:

$$F_2(t) = \frac{1 - \exp[-(p_2 + q_2)X(t)]}{1 + \frac{q_2}{p_2}\exp[-(p_2 + q_2)X(t)]} \tag{6}$$

where $X(t) = \sum_{j=1}^{t} \exp(\beta' x_j)$ captures the cumulative marketing effort with x being the current marketing effort and β the corresponding parameter vector[28].

If prescriptions due to each market were perfectly sequenced in time, and the time of transition from one market to the other, c, was known, then:

$$d_t = \begin{cases} 1 & \text{if t} \leq \text{c} \\ 0 & \text{if t} > \text{c} \end{cases} \tag{A1}$$

and prescriptions could be expressed as:

$$N_t = I_{[t \leq c]} N_{1t} + (1 - I_{[t \leq c]}) N_{2(t-c)} + \varepsilon_t \tag{A2}$$

[28] This particular version of the GBM was proposed by Danaher et al. (2001) and ensures non-negativity of the vector of cumulative marketing effects, which is critical for the properties of the corresponding cdf.

c could then be determined through a grid search by using either maximum likelihood based on the distributional assumptions for the error term or minimizing least-squares (e.g. Vakratsas et al., 2004).

A more general specification would introduce uncertainty regarding the transition time expressed as a probability with which $d_t$ takes on a certain value:

$$P(d_t = 1) = \pi_t = \frac{\exp(-\rho t)}{1 + \exp(-\rho t)} \text{ with } \rho > 0 \tag{7}$$

The negative time dependence of the early market probability satisfies the temporal ordering requirement for the two markets[29]. A "steeper" decrease in this probability indicates a stronger separation between the two markets. We will refer to the probability in (7) as "switching probability" consistent with the terminology of switching regime models (Vakratsas et al., 2004). Finally, rounding up our model exposition, we assume that $\varepsilon_t \sim N(0, \sigma^2)$.

A few observations are in order here, to highlight the features of the proposed model, in particular its ability to capture the dual-market phenomenon, and compare it with alternative model specifications. Our model

---

[29] However, the negative dependence is tested rather than imposed.

is capable of satisfying the “separability” requirement of the dual-market hypothesis. First, it assumes that the two markets have different potentials. Second, it assumes that prescriptions due to the two markets are temporally distinct via the time-dependent switching probability in (7) or the time-dependent market dummy in (A1). Finally, the additive form of the prescription model (1) and the separate modeling of the two adoption processes do not impose any interaction, and thus link, between the two markets allowing them to be disconnected. Thus, our model satisfies the three basic principles of a dual-market diffusion model outlined in the beginning of this section.

Although our model bears similarities with that of Moe and Fader (2002) and the Asymmetric Influence Model (AIM) of Van den Bulte and Joshi (2007) for w=0, it is fundamentally different as it allows for complete separation of the two markets via different potentials and temporal ordering of adoptions due to each market. In both the aforementioned models, the two types of adopters are “drawn” from one pooled market, implying that they belong to two segments of a single market rather than belonging to two separate markets. Further, the proportion of each type of adoption or the “mixing” probability (the equivalent of our switching probability), is constant rather than time-dependent. Thus, the mixing probability, assumed to be time-independent, determines both the size of each segment and the proportion of adoptions due to each segment at every time point. In our model, by controlling for market size via the different market potentials, our switching probability determines exclusively the proportion of adoptions due to each market at a given time. Moreover, the time dependence of the switching probability allows for further temporal separation of the two markets. In addition, our model also accounts for the effects of marketing effort. Although the AIM of Van den Bulte and Joshi (2007) provides an elegant and comprehensive closed form solution, capable of nesting many different models as special cases, it does not control for marketing effort[30]. These model features are all testable within our model's framework and will be directly compared to alternative specifications. Estimation specifics and competing model specifications are discussed in the estimation section following the data set description.

[30] Accounting for cumulative marketing effort in the GBM sense may not yield a closed-form solution for AIM.

### 4.4 Empirical Application

#### 4.4.1 Data

We use monthly category-level data on the number of new prescriptions (NRx) and matching marketing mix information for a recently developed therapeutic category, provided by one of the competing firms , in a non-U.S. market. Due to confidentiality reasons, the therapeutic category and market cannot be disclosed. However, the category concerns a new therapeutic class developed for the treatment of a lifestyle-related disease. The three drugs competing in the therapeutic class exhibit high efficacy and mild side effects, thus, based on our introduction discussion, the category is appropriate for the study of the dual market phenomenon. The pioneering drug was approved as the first effective oral treatment for the category just before April 1999 and the two competitors followed, with products of comparable efficacy, in September 2003 and March 2004. The new prescription (NRx) data were collected by IMS Health and are based on new prescriptions dispensed in a representative national sample of pharmacies in the market of concern. Marketing mix information includes IMS Health data on physician journal advertising dollar expenditures (PJ), detailing effort (SDV), expressed in numbers of details for the specific product, sample pack volume (SPV), and A.C. Nielsen-audited data on direct-to consumer advertising (DTC) dollars. The data set covers an 85-month period between April 1999 and May 2006. DTC advertising was initially employed in early 2001 by the pioneering drug, primarily through the use of television advertising. Regulations governing the advertising of pharmaceuticals directly to consumers in this market are much stricter than in the U.S. Specifically, only two types of DTC are allowed, one informing consumers about the disease (disease-related), which should not mention any brand-specific information, and another (brand-related) which can mention the drug's brand name, price and quantity but cannot indicate the disease which it treats. To ensure that consumers cannot combine information or infer a link between them, the two types of messages have to use different executional elements as well, and they cannot be aired or placed simultaneously. The competing firms in our sample opt mostly for brand-related advertising and we combine the expenditures on both types of messages to construct our advertising measure. Due to the high correlation between detailing and sampling, which is typical for pharmaceutical data, we included only detailing in our empirical analysis. Prices are not critical for this class possibly due to lack of variation in the observation window. New category prescriptions and the marketing variables included in our empirical analysis are shown in Figure 4.2 for the entire observation period (figures are masked to preserve confidentiality). The entry times of the two later competitors are also indicated.

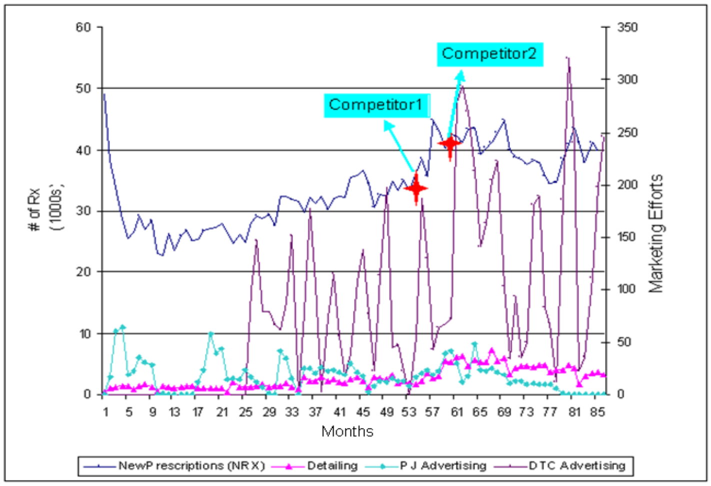



Figure 4-2: New Prescriptions and Marketing Effort

An interesting observation from Figure 2, consistent with our discussion in the introduction, is that prescriptions do not follow the typical diffusion pattern (e.g. Bass, 1969). Rather, the prescription curve over time is characterized by a very early spike (the biggest throughout the observation period), which is followed by an exponential-like drop only then to resume a more “Bass-like” diffusion pattern.

Thus, the existence of two markets, one characterized by high need and demand accumulated even prior to the pharmaceutical's launch and the other, closely resembling a typical product market is possible and will be tested in our empirical application.

### 4.4.2 Estimation

We write our model in an "estimation-ready" form. Specifically, equations (1)-(6) suggest that new prescriptions can be expressed in the following fashion:

$$N_t = \begin{cases} M_1[F_1(t) - F_1(t-1)] + \varepsilon_t & \text{with prob } \pi_t \\ M_2[F_2(t) - F_2(t-1)] + \varepsilon_t & \text{with prob } 1\text{-}\pi_t \end{cases} \tag{8}$$

with $F_1$, $F_2$ and $\pi_t$ defined as in (5), (6) and (7) respectively. Utilizing the normality assumption for the error term, the likelihood for the model can be written out as follows:

$$L = \prod_{t=1}^{T} \left\{ \pi_t \frac{1}{\sigma} \phi\left(\frac{N_t - M_1[F_1(t) - F_1(t-1)]}{\sigma}\right) + (1-\pi_t) \frac{1}{\sigma} \phi\left(\frac{N_t - M_2[F_2(t) - F_2(t-1)]}{\sigma}\right) \right\} \tag{9}$$

where $\phi$ denotes the standard normal density[31]. We lag marketing mix variables by one month to explain current month prescriptions, to capture the typically delayed effects of marketing effort on prescriptions (e.g. Hahn et al., 1994), and use their square root to account for diminishing returns. Parameter estimates are obtained by using the maximum likelihood method implemented in GAUSS.

Following up on our model section discussion, a number of alternative specifications can be tested against the proposed one (which we will refer to as M0). We list those that can directly test the main assumptions of our model:

a) M1: Early market with an imitation coefficient (but no marketing mix influence). This assumes a Bass Model specification for the early market specified as: $F_1(t) = \dfrac{1-\exp[-(p_1+q_1)t]}{1+\dfrac{q_1}{p_1}\exp[-(p_1+q_1)t]}$

b) M2: Early market with marketing mix influence (but no imitation). This assumes a Generalized Bass Model for the early market specified as: $F_1(t) = 1-\exp[-p_1 X_1(t)]$, i.e. $q_1$=0[32].

[31] The hypothesis of equal variances for the two regimes could not be rejected and thus was adopted for reasons of model parsimony.
[32] A competing model with both an imitation coefficient and marketing mix influence for the early market is possible, but, as our empirical results will show, redundant.

c) M3: Static switching probability, i.e. $\pi_t = \pi = \dfrac{\exp(-\rho)}{1+\exp(-\rho)}$.

d) M4: Deterministic switch. This corresponds to the model specified through equations (A1) and (A2) and can be estimated using grid search.

In addition, we test two other specifications:

e) M5: A single market GBM included for completeness and the purpose of forming a "benchmark" for our empirical results.

f) M6: A "pooled market" model with a constant switching probability, specified as follows: $N_t = M\{\pi[F_1(t)-F_1(t-1)]+(1-\pi)[F_2(t)-F_2(t-1)]\}+\varepsilon_t$ with $F_1$, $F_2$ defined as in our proposed dual-market model and $\pi$ defined as in M3 (static). This formulation is similar in spirit to the models proposed by Moe and Fader (2002) and Van den Bulte and Joshi (2007), the latter for w=0. Despite the apparent similarities with M0, this model is conceptually very different as it assumes that adoptions originate from one "pooled" market of potential M (hence the label). Thus, it essentially assumes that adoptions come from two different segments of the same market rather than two different markets and hence does not satisfy the requirements of the dual-market hypothesis. This is also reflected in the prescription equation characterizing this model above, since prescriptions due to each adoption process are not completely separable but rather factored by the common market potential parameter M. It does allow, however, for non-uniform marketing mix influence on the two types of adoptions.

g) M7: A pooled market model with a dynamic switching probability $\pi_t$, as defined in (7), to investigate whether assuming time-dependent segment sizes in the pooled market specification can lead to considerable gains in model fit.

## 4.5 Findings

Estimation results for all competing model specifications are presented in Table 4.1. The models are compared on the basis of the Bayesian Information Criterion (BIC). The proposed model M0 has the lowest (best) BIC and its parameter estimates have the expected signs most of which are significant. Better performance of the dual-market model suggests that its basic premises cannot be rejected. Specifically:

1) The dual market hypothesis cannot be rejected as M0 does better than both the single-regime (M5) and the pooled market models M6 and M7.

In fact, all dual market models perform better than both the pooled market and the single regime models, with the exception of the deterministic switch model M4. The latter is only marginally inferior to M6 and M7 in terms of the BIC. The fact that M7 is only marginally better than M6 suggests that assuming a dynamic switching probability without accounting for separate market potentials cannot considerably improve model fit, further validating the dual-market hypothesis.

2) Adoption by the two markets is temporally sequenced. The time parameter in the switching probability is significant in all dual market models and M0 does better than the static switching probability model (M3). Figure 4.3 shows the time path of the switching probability superimposed on the new prescriptions curve. It can be easily detected that the switching probability tracks very well the initial exponential prescriptions curve, which we attributed to the early market, and then drops dramatically soon after the prescriptions curve “picks up” again. This sharp drop in the switching probability is also in broad agreement with the time switch of the 10th month suggested by the deterministic switch model M4 since the membership probability for the early market is

approaching 0.05 at that time. However, the relatively poor performance of M4 suggests that the transition is smooth and somewhat gradual (although quickly resolving) rather than abrupt.

3) The early market is not influenced by marketing efforts as suggested by the better performance of M0 versus M2 and the lack of statistically significant marketing mix effects for the early market.

4) The early market realized adoption rate follows an exponential pattern. The proposed model M0 does better than M1 which includes an "imitation" coefficient q. Thus, the assumption of an exponential adoption rate for the early market due to accumulated demand cannot be rejected.

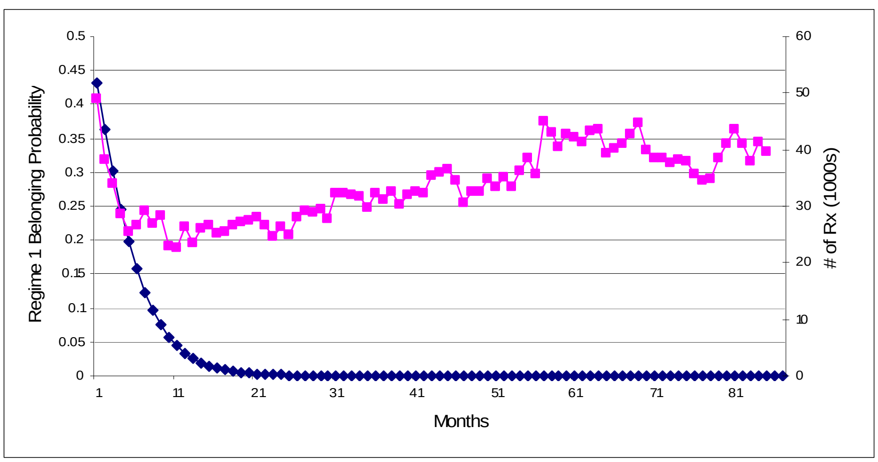



Figure 4-3: Regime Switching Probability and Prescriptions

Table 4-1: Estimation Results for Competing Model Specification

| | **Dual Market Models** | | | | | | | | | | **Alternative Models** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **M0 Proposed** | | **M1 Early Market Imitation Effects** | | **M2 Early Market Marketing Mix*** | | **M3 Static Switching Probability** | | **M4 Deterministic Switch**** | | **M5 Single Regime** | **M6 Pooled Market Static Switching Probability** | | **M7 Pooled Market Dynamic Switching Probability** | |
| | Early | Late | Early | Late | Early | Late | Early | Late | Early | Late | | Early | Late | Early | Late |
| p | 0.1814 | 0.0018 | 0.1818 | 0.00184 | 0.0609 | 0.0020 | 0.1913 | 0.0017 | 0.0283 | 0.0033 | 0.0003 | 0.1065 | 0.0030 | 0.0092 | 0.0029 |
| | (5.07) | (2.61) | (5.04) | (2.62) | (2.55) | (2.87) | (4.52) | (2.21) | (2.43) | (4.44) | (0.35) | (2.15) | (10.20) | (9.33) | (8.62) |
| q | | 0.0090 | 0.001693 | 0.00898 | | 0.0095 | | 0.0085 | | 0.0136 | 0.0038 | | 0.0197 | | 0.0187 |
| | | (4.15) | (0.01) | (4.14) | | (4.12) | | (3.78) | | (4.17) | (2.57) | | (2.11) | | (3.81) |
| Market potential (M) | 347700 | 11446100 | 346500 | 11430600 | 858200 | 10533500 | 336500 | 12346573 | 1242100 | 6742125 | 92911800 | 5662700 | | 5921000 | |
| | (8.52) | (2.46) | (7.89) | (2.47) | (2.12) | (2.73) | (7.61) | (2.10) | (2.77) | (4.02) | (0.35) | (5.70) | | (5.10) | |
| Advertising (DTC) | | 0.0024 | | 0.002368 | | 0.0024 | | 0.0024 | | 0.0023 | 0.0025 | | 0.0026 | | 0.0026 |
| | | (2.55) | | (2.54) | | (2.55) | | (2.56) | | (1.67) | (1.72) | | (1.70) | | (2.02) |
| Physician Jour. Adv. (PJ) | | 0.0077 | | 0.007722 | -0.0177 | 0.0075 | | 0.0081 | | 0.0043 | 0.0055 | | 0.0048 | | 0.0055 |
| | | (3.11) | | (3.12) | -(1.56) | (2.89) | | (3.24) | | (1.09) | (1.65) | | (1.22) | | (1.55) |
| Detailing (SDV) | | 0.0060 | | 0.006083 | 0.0005 | 0.0061 | | 0.0059 | | 0.0050 | 0.0109 | | 0.0020 | | 0.0009 |
| | | (1.05) | | (1.05) | (0.07) | (1.06) | | (1.03) | | (0.58) | (1.18) | | (0.16) | | (0.16) |
| Variance ($\sigma^2$) | 5.29E+06 | | 5.29E+06 | | 5.29E+06 | | 5.29E+06 | | 1.02E+07 | | 1.52E+07 | 1.02E+07 | | 1.02E+07 | |
| | (6.53) | | (6.53) | | (6.47) | | (6.34) | | (6.56) | | (6.55) | (6.56) | | (6.56) | |
| $\varrho$ | -0.27 | | -0.26 | | -0.2 | | -3.32 | | | | | -3.41 | | -0.13 | |
| | -(2.31) | | -(2.31) | | -(2.07) | | -(5.50) | | | | | -(4.64) | | -(2.50) | |
| LL | -792.09 | | -792.10 | | -791.51 | | -800.31 | | -818.25 | | -833.57 | -817.94 | | -817.69 | |
| BIC | 1628.75 | | 1633.20 | | 1636.47 | | 1645.17 | | 1676.59 | | 1698.31 | 1675.97 | | 1675.62 | |
| **RANK** | **1** | | **2** | | **3** | | **4** | | **7** | | **8** | **6** | | **5** | |

*DTC was not employed until the 25th month, thus we did not include a DTC effect for the early market.
** The switch time c in M4 is the 10th month after launch, identified by a grid search.

In sum, our findings on the dual market model suggest that there is not sufficient evidence to counter the hypothesis of two disconnected markets, an early and a late one, as suggested by the better performance of the dual market model compared to the pooled market[33]. The early market is characterized by an exponential distribution attributed to accumulated pre-launch demand that is not influenced by marketing effort (rejection of imitation effects and marketing mix influence). Also as predicted, the exponential parameter for the early market is considerably high and larger than typical p values of diffusion models reported in meta-analyses (Sultan et al., 1990)[34]. The pooled market models M6 and M7 underestimate p for the early market and thus the corresponding adoption rate, suggesting that they cannot adequately capture the accumulated demand phenomenon due to the early market. Furthermore, in the single-regime model (M5) the estimates of both p and the market potential M are insignificant with the latter's value being rather

[33] We further tested the separability or lack of cross-market communication assumption by estimating an additional model "MX," where the switching probability $\pi_t$ is a function of the early markets lagged cdf $F_1$, specifically $\pi_t = \frac{\exp(-\rho F_1(t-1))}{1+\exp(-\rho F_1(t-1))}$. This model suggests that the late market's adoption decision is dependent on the past adoption behavior of the early market, thus introducing cross-market communication. Estimation results revealed that the model's fit is not superior to that of M7, lending further validity to the proposed model and its fundamental assumptions.

[34] It should also be noted that the reported p and q values in meta-analytic studies are based on findings of annual data analyses and hence are expected to be higher than those produced by monthly data analyses such as ours.

unrealistic. While troublesome, this lack of stability in the parameters of the single-regime model can be attributed to its inability to capture the early-market phenomenon, highlighting the need for a different approach[35]. As shown by the dual market model (M0) estimates, the late market possesses more typical product market characteristics. Both p and q are significant as well as marketing effort variables such as physician journal and DTC advertising[36].

It is worth noting that despite the larger number of parameters employed by the dual market model, it produces more efficient estimates for the marketing mix effects than both the pooled and single regime models. Specifically, only DTC advertising is significant in the pooled market models M6 and M7, at the 10% and 5% levels respectively, whereas in the single regime model DTC advertising and physician journal advertising are marginally significant at the 10% level. By contrast, both these variables are confidently significant at the 1% level in the dual-market model M0. This should be attributed to the ability of M0 to separate adoptions due to the two markets and thus better identify the influence of marketing activities on each of them, a property that does not appear to be shared by the other two specifications. In addition, the effects of

[35] Standard statistical packages like SAS frequently reported convergence problems on the estimation of the single-regime model.

[36] The corresponding empirical short-term elasticities, calculated by using the estimated coefficients and increasing marketing effort by 1%, are 0.06 for DTC and 0.07 for physician journal advertising.

physician journal advertising are underestimated by both the pooled market and the single regime models, which, combined with the underestimation of their significance, could lead to erroneous policy implications both from a management and a public perspective. These implications will be further discussed in the next section.

Although the lack of significance of detailing may be somewhat surprising, it is consistent with the results of Narayanan, Desiraju and Chintagunta (2004), who found that only DTC advertising affects category prescriptions, and those of Rosenthal et al. (2003) who report greater category demand elasticities for DTC advertising than for detailing. The latter study attributes the greater influence of DTCA advertising to its higher growth due to the change in US FDA regulations regarding this form of patient-directed marketing activity. This is not dissimilar to our case, since DTC advertising started only two years after the category inception. Compared to the other physician-directed marketing activity, physician journal advertising, the relative ineffectiveness of detailing could be due to higher saturation. In other words, because of management decisions to periodically stop advertising spending, quality of physician advertising may have been restored during these “off” advertising periods, eventually increasing the overall effectiveness of physician journal advertising (Naik et al., 1998). Detailing, on the other hand, is practiced continuously, thus being more prone to saturation or over-saturation (Hanssens et al., 2001). This finding on physician journal advertising effectiveness hopefully furthers our understanding of the effectiveness of this marketing activity since frequently in pharmaceutical studies, detailing and physician journal advertising are merged (Hahn et al., 1994; Shankar et al., 1998). Two alternative explanations for the lack of significant detailing effects are the absence of serious side effects for the drugs in the category in question (Venkataraman & Stremersch, 2007), which renders detailing less critical, and potential brand aggregation effects due to the application of a category sales model (Fischer & Albers, 2007).

### 4.6 Predictive Validity

A frequent use of diffusion models is for forecasting purposes, so testing the predictive ability of the major competing specifications both in the short- and the long-run is not only appropriate but also of practical importance. We pick four models: the best of the dual-market models, M0, the single-regime

Table 4-2: One Step-Ahead Prediction Errors

Panel A: Full sample

| | **M0 Dual Market** | **M5 Single Regime** | **M6 Pooled Market Static Switching Probability** | **M7 Pooled Market Dynamic Switching Probability** |
|---|---|---|---|---|
| MAPD | **7.09** | 7.40 | 7.19 | 7.31 |
| MAD | **2119.25** | 2520.47 | 2224.61 | 2360.36 |
| MSE | **10260297** | 15459521 | 11345028 | 12063341 |

*Panel B: First seven observations*

| | **M0 Dual Market** | **M5 Single Regime** | **M6 Pooled Market Static Switching Probability** | **M7 Pooled Market Dynamic Switching Probability** |
|---|---|---|---|---|
| MAPD | **12.49** | 18.67 | 22.98 | 22.71 |
| MAD | **4582.11** | 6918.11 | 6535.57 | 6550.74 |
| MSE | **33645088** | 101934232 | 73411023 | 73786937 |

model, M5, and the pooled market models, M6 and M7; and compare them on the basis of in-sample, one-step-ahead forecasts, using the full sample estimation results reported in Table 4.1, as well as out-of sample forecasts. The latter is performed by leaving the last ten observations out of the estimation sample and forecasting them using the parameters estimated with the remainder of the sample[37]. Tables 4.2 and 4.3 contain a summary of the short-

[37] Because we estimate a relatively large number of parameters, we are somewhat limited in the number of observations we can leave out of the sample without losing sufficient statistical power.

term and long-term forecasting results respectively. In both cases, the dual market model (M0) performs better than its counterparts on all measures[38].

In the in-sample one step-ahead predictions, all models do quite well underlying the importance of updating cumulative adoptions for forecasting purposes (Panel A of Table 4.2). Once the cumulative adoptions are updated, every time a new observation becomes available, even the single regime model can do reasonably well. However, the major advantage of the dual market model lies in its ability to capture the early market. Hence, a more stringent test for all models would be to check their ability to predict early market adoptions. In Panel B of Table 4.2, we compare all prediction metrics for the three models on the first seven observations, corresponding to the steep decline observed in the new prescriptions data of Figure 4.2 and attributed to the early market. The differences among the three models now are quite dramatic with M0 errors being roughly 50% less than those of the other two. Remarkably, the single-regime model does better than the pooled market models M6 and M7 in terms of MAPD, but worse in terms of MAD because of its inability to capture the high number of adoptions in the first month, where the error is significantly larger in

[38] We use Mean Absolute Percent Deviation (MAPD), Mean Absolute Deviation (MAD), and Mean Squared Errors (MSE).

absolute terms. Somewhat surprisingly, M6 does better than M7 in terms of the one step-ahead prediction, both for the full sample and the first seven observations, further confirming that assuming a dynamic switching probability cannot alone improve the performance of the pooled market model.

Table 4-3: Holdout Forecasting Performance

| | M0<br>Dual Market | M5<br>Single Regime | M6<br>Pooled Market<br>Static Switching<br>Probability | M7<br>Pooled Market<br>Dynamic Switching<br>Probability |
|---|---|---|---|---|
| MAPD | **7.86** | 22.94 | 9.79 | 9.10 |
| MAD | **2992.76** | 9084.99 | 3881.41 | 3581.58 |
| MSE | **13264964** | 93544302 | 19666050 | 16951226 |
| LL | **-100.1** | -121.7 | -107.6 | -107.2 |

In terms of the long-term, holdout, performance the dual market model does three times better than the single-regime model and 20-30% better than the pooled market models, as indicated by the error statistics. This also highlights the importance of correctly specifying marketing mix effects. Since in the holdout long-range forecasting case forecasts are not updated with actual cumulative adoption information, the only actual updates the manager-forecaster can make are only those concerning the marketing mix allocation which can be planned and thus controlled. Thus, better estimates of the marketing mix effects lead to superior forecasts especially when no other information is available. The rather poor performance of the single-regime model further exposes the quality of the estimates produced, which, in the absence of adoption information updating, cannot effectively forecast adoption. The forecasting performance of M0 is shown in Figures 4.4 and 4.5.

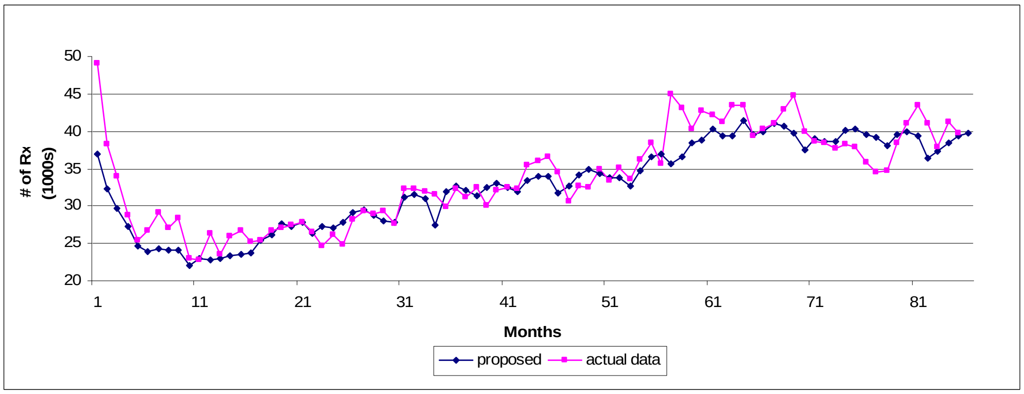



Figure 4-4: Tracking Performance of the Proposed Model

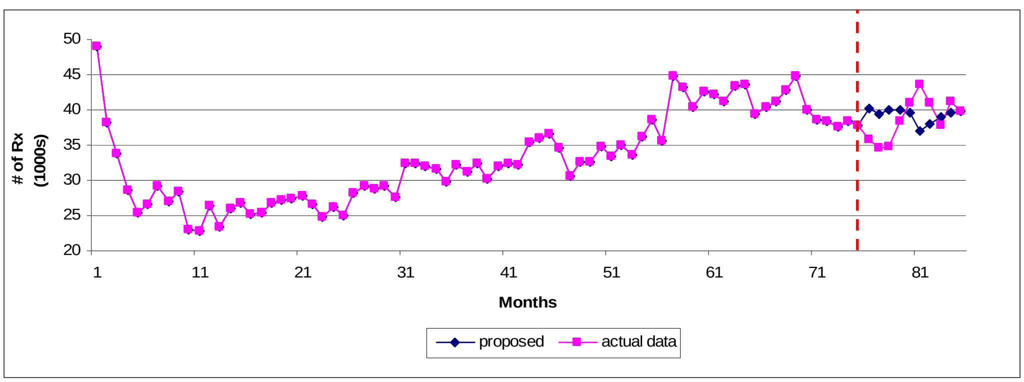



Figure 4-5: Holdout Forecast Performance of Proposed Model

Thus, the dual-market model, M0, shows both good forecasting and parameter face validity (in terms of size and significance of the parameter estimates). The main implications of our approach and findings will be discussed and summarized in the ensuing final section.

### 4.7 Summary, Discussion and Further Research

We argued for a custom-fit approach to modeling the diffusion of new prescription pharmaceuticals due to the idiosyncratic elements of prescription markets. The premise of our argument is based on clinical management and pharmacoepidemiological concepts that consider the severity of patients' health problems and symptoms, and the medical practice of "watchful waiting." Specifically, we suggest that due to persistent and severe symptoms suffered by a class of patients, an early market for prescriptions is created. This market may be formed even before the launch of the pharmaceutical due to well-defined, diagnosed needs of the patients corresponding to this market. When the prescription drug is eventually launched, the accumulated demand by the early market is realized, typically in the form of an exponential adoption curve. The launch may also trigger adoption by a second, "late," market corresponding to patients with milder conditions or lack of "classic" or persistent symptoms for whom the drug may be chosen as the most convenient solution to treating the condition. This is particularly likely in the case of drugs treating diseases the severity of which is located in a continuum (mild to severe) rather than those manifested in a dichotomous manner (presence/absence of disease). Lifestyle-related diseases treated by drugs with mild side effects fit well this profile.

To address the idiosyncratic elements of pharmaceutical adoption we propose a switching regime dual-market diffusion model that extends the well-known Generalized Bass Model (GBM) by explicitly accommodating the possibility of an early market disconnected from the late main market. We use a dynamic switching probability to capture the transition from one market to the other. Our model shows good parameter face and forecasting validity (including a holdout sample exercise) compared to the standard (single-regime) GBM and one that assumes that both early and late adoptions originate from a common, “pooled,” market with two segments. The latter model bears similarities with Moe and Fader’s (2002) model and a special case of Van den Bulte and Joshi’s (2007) Asymmetric Influence Model (AIM).

The model’s good forecasting ability could be of great value to drug manufacturers in terms of inventory management, production, and logistics, particularly since it is capable of predicting well turning points in the data such as the early drop due to the untapping of the accumulated early demand. Specifically, the early market peak should alert planners regarding production and logistics requirements, particularly since early adoption concerns patients with the highest need for the treatment. The quick realization of early market demand, expressed by the sharply declining exponential curve, and the ensuing saddle should also ease concerns about inventory buildup and may lead to higher efficiencies. Such predictions can also be used by top management in order to communicate to the various stakeholders (investors, analysts etc.) and better manage expectations regarding revenues and earnings forecasts. Similarly, health organizations and governments may use these forecasts to estimate and project coverage demanded for a drug therapy and balance costs and benefits of a potential treatment adoption. The evidence furnished here for the dual-market phenomenon in the pharmaceutical market is consistent with similar findings for the case of technological innovations, although the driving forces behind the phenomenon are very different in the two contexts. This convergence of evidence could eventually lead to two distinct benefits. First, it should alert managers about the significance of such phenomena and thus make the communication of relevant expectations to stakeholders more credible. Second, it contributes toward the generalization of these phenomena and thus their legitimization both from an academic and a practitioner perspective.

Impressively, although our dual-market model employs more parameters, it yields more efficient estimates for the marketing mix effects, underlying the importance of accounting for the presence of two separable markets. The pooled market model, on the other hand, underestimates the significance and, in the case of physician journal advertising, the size of marketing mix effects. This could lead to erroneous policy implications that could be detrimental both from a management and public perspective. For example, underestimation of physician journal advertising may prompt a decision to under-fund this marketing activity, resulting in fewer prescriptions. From a public benefit perspective, less spending on physician journal advertising may lead to less information communicated to the physicians and thus lower awareness of potential treatments. This is particularly important for the case of general practitioners or family physicians treating a lot of patients with varying medical problems, prescribing potentially a large gamut of pharmaceuticals and facing severe time constraints.

Our study carries the obvious limitation that its findings are specific to the novel pharmaceutical class analyzed. Although the availability of data on a single therapeutic category prevented us from comprehensively testing the separability assumption for the two markets, we believe we provided strong enough arguments for the presence of two disconnected markets by thoroughly examining multiple alternative formulations. We hope that our study will stimulate further research on the dual market issue, which could significantly contribute to our knowledge of complex, "non-standard" diffusion phenomena. Future research should examine additional therapeutic categories and further investigate the underlying reasons for the dual market phenomenon, the conditions under which such a pattern emerges, as well as the differences in the timing of the switch from the early to the late market across different categories. However, we should note that not many prior studies have examined the diffusion of prescription pharmaceuticals, especially using monthly observation data, new prescriptions, and a full set of marketing mix variables. For example, we considered the effects of physician journal advertising, a variable frequently merged with other physician-directed activities such as detailing, and found that its effects are more significant than those of detailing possibly due to smaller saturation effects since advertising is subject to “on” and “off” cycles. Although it is possible that this finding is specific to the category studied here, it also suggests that the effects of physician journal advertising, frequently overlooked by pharmaceutical studies, warrant more systematic investigation. We also do not consider supply-side effects, which may not be applicable in our context but could be critical for other markets.

The lack of marketing mix effectiveness on early market adoption suggests that this market has a pre-defined need for the product and adoption considerations are most likely based on efficacy. However, it is possible that early markets may be following pre-announcements or early press releases before the launch of a drug, which will eventually influence their adoption decisions. Unfortunately, this type of data were not available for our study, however this could be a topic worthy of further research. The effectiveness of marketing activities, at least DTC and physician journal advertising, on the late market may indicate that the product “creates” this market just like in typical product markets. This is consistent with views in the medical field that marketing may lead to overprescription by promoting overuse. Although our study cannot provide a definite answer to this issue, we hope that it will stimulate further research on the topic. While our study suggested that the influence of marketing mix on the late market is significant, it also found, consistent with most marketing studies, that such effects are economically modest. Yet, as noted above, the medical literature frequently attributes prescription inflation to heavy “marketization.” One possible explanation, that may reconcile medical and marketing views, is that most category expansion effects, and potential overprescribing phenomena, are caused by the mere entry of new products and the accompanying launch campaigns and buzz as was apparently the case with SSRIs (Selective Serotonin Re-uptake Inhibitors) in the antidepressant market. Databases combining brand entry, publicity and marketing mix information can allow for the testing of lead and lag effects of all these factors and shed light on whether marketing activities may create markets and potentially lead to pharmaceutical overprescribing.

A final potentially promising avenue of future research is one that attempts to connect the (unobserved to the marketer) ideal adoption times of Figure 4.1 to the actual adoption times. This would require combining pre- and post-marketing data with the former being collected from medical sources. Specifically, ideal adoption times may be estimated from disease epidemic data before the launch of an appropriate prescription treatment. In other words, the times when individuals are diagnosed with identifiable severe symptoms of a disease can be considered as ideal adoption times. Thus, disease and product adoption epidemics can be combined, in the fashion presented in the stylized example of Figure 4.1, and used for predicting the need and accumulated demand for potential pharmaceutical treatments. This would not only be useful for the supply side, pharmaceutical firms and managers planning for launch and adequate supply of a new prescription drug, but also on the demand side, institutional clients such as hospitals who would like to satisfy the needs of patients with severe health problems in the fastest possible manner.

# CHAPTER 5

# CONCLUSION

## 5.1 Summary of Findings

This dissertation aims to develop flexible models to accurately capture complex effects of integrated marketing communications campaigns, where companies and advertising agencies use a multitude of marketing tools to promote their products and grab the attention of their target consumers. In this thesis we operationalize an IMC campaign as a set of concurrently employed multiple marketing tools, such as direct marketing, personal selling and/or advertising through multiple media channels, assuming the messages delivered through various channels are consistent.

Over the three essays that make up chapters 2-4 of the thesis, we focus on a number critical substantive issues regarding marketing communications effects - some with limited and/or conflicting findings in the literature - , including temporal variations, media synergies, content specific effects, market heterogeneity and irregular market response. To address the methodological limitations preventing the proper examination of these topics, we adopt advanced statistical and econometrical methods such as Kalman filtering, Monte Carlo simulations, multivariate adaptive regression splines and kernel estimation, and broadly discuss the advantages of these methods over more traditional benchmarks. Table 5.1 summarizes the research conducted in chapters 2-4.

Table 5-1: A Brief Summary of Research in this Thesis

| | Chapter 2 | Chapter 3 | Chapter 4 |
|---|---|---|---|
| *Research Focus* | Complex Multi-Media Communications Effects on Market Response | Dynamics of Direct-to-Consumer Advertising Under Regulation | Dynamics of the Diffusion Pattern: The Effects of Promotional Mix and Market Heterogeneity |
| *Model* | Multivariate Adaptive Regression Splines Model | Augmented Kalman Filter Model with Continuous State and Discrete Observations | Dual-Market Generalized Bass Model with a Dynamic Probabilistic Switch |
| *Data*[39] | TS: 2 SUV Brands (US)<br>3 Hybrid Car Brands (US)<br>CS: Beer Product Category (US)<br>Leading National Advertisers (US) | TS: A Focal Prescription Drug Brand (non-US)<br>Two Competitive Drugs | TS: Category-Level Data on a Novel Therapeutic Class (non-US) |

In chapter 2, we focus on measuring the complex effects of multi-media communications including thresholds, saturation levels and cross-media interactions. In order to handle flexible modeling in higher dimensions and sacrifice as little as possible from the efficiency of the estimation, we employ MARS, a non-parametric regression model based on multivariate adaptive

[39] Abbreviations TS and CS stand for time-series and cross-sectional, respectively.

splines. We show that MARS is highly suitable for addressing problems such as multi-media effects, which require flexible modeling of a large number of variables whereas, most other non-parametric methods suffer from the well known issue of *curse of dimensionality* when working in higher dimensions.

We analyze five time series data sets including sales and media spending information of the top three hybrid cars and the top two SUV brands, as well as two cross-sectional data sets referring to beer category and Leading National Advertisers (LNA). All data belong to the US market. Comparisons between our MARS-based non-parametric model with benchmark models show that (1) MARS provides better fit and better average forecast performance than both the parametric benchmarks and Kernel regression confirmed by all data sets, and (2) the marginal improvement of MARS over Kernel significantly increases as the number of variables in the non-parametric model increases.

In terms of the shapes of the multi-media communications effects the results reveal a set of response shapes common across various data sets. We find compelling evidence to (1) S-shaped response to multiple media efforts with typical threshold and saturation levels, (2) existence of multiple thresholds separated by a single saturation point, (3) possible supersaturation for certain media such as magazines and cable TV – although not prevalent-, and (4) early saturation of response to newspaper ads. We also quantify the threshold and saturation levels using non-parametric derivatives. The results show that most product categories and media do not operate in the most efficient spending range for advertising, evident from the average saturation levels of around 50% of the maximum spending across media and data sets. These findings are interesting since they provide empirical evidence for long discussed theories about advertising response, such as the S-shaped response curve. Moreover, they present significant proof for inefficient budgeting and scheduling of media communications efforts (e.g. early and supersaturation).

We demonstrate cross-media synergies by three dimensional interaction surface plots. Synergy surfaces are noticeably complex and irregular, supporting the use of flexible non-parametric estimation methods. Comparison of respective plots for MARS and Kernel-based estimations reveal significant differences in regions which i) are close to the boundaries of the observation domain and ii) have limited number of data points. The results clearly show that Kernel suffers from *boundary effects* and tends to falsely display threshold and saturation levels due for the sparsely populated data regions.

Chapter 2 contributes to the extant literature in three ways. First, we address the problem of high-dimensionality in evaluating multi-media communications by flexibly estimating complex IMC effects including irregular main effects for multiple media and cross-media interactions. We provide empirical evidence to phenomena such as multiple thresholds, early and supersaturation, with an aim to help generalize complex effect shapes that might vary across product categories and media. Second, we quantify the observed irregularities such as thresholds and saturation levels using non-parametric derivatives and show that most media operate in inefficient spending ranges for various product categories. Third, we demonstrate that MARS, when compared to Kernel-based non-parametric methods, exhibits superior performance in terms of accurate prediction and long range forecasting of market response. Given the reliability of MARS in capturing multi-media effects, future research can take these findings further by working on optimal scheduling and budget allocation decisions for multi-media communications problems.

Unlike chapter 2 which investigates the complex multi-media communications effects, the following two chapters focus on the dynamics of the promotional mix activities in the pharmaceutical industry.

In chapter 3, we explore dynamic direct-to-consumer advertising (DTCA) effects in a highly regulated market, where regulations impose considerable restrictions on the format and content of advertising messages, thus introducing high uncertainty regarding their effectiveness. Regulations force firms to choose between two mutually exclusive types of messages (i.e. branded or generic) each time they wish to advertise. We investigate three research questions which should be of great interest to managers: Whether advertising is effective under such strict regulation conditions; which type of advertising message is more effective? When? We pursue these questions by examining data on a pioneering brand of a new therapeutic category. We apply the Augmented Kalman Filter with continuous state and discrete observations (AKF(C-D)), to estimate dynamic marketing mix effects without discretization biases. To our knowledge, the implementation of this multivariate continuous nonlinear estimation is new to the marketing literature.

Our findings suggest that DTCA, as well as direct-to-physician activities, exhibit complex dynamics which would have been hard to capture using standard time-varying models. DTCA elasticities, although modest, consistent with the previous findings in the pharmaceutical literature, are within the operating range of mass-marketed consumer goods. This implies that the answer to the *whether* question is “yes;” and that, brands can benefit from advertising even under strict regulatory conditions. The elasticities also reveal that branded advertising is clearly more effective than generic advertising. Thus, the answer to the *which* question is “branded.” We attribute the higher effectiveness of branded messages to their reinforcing nature: the mention of the brand name, absent from generic messages, can defend the brand against competition and the lower informational content makes this type of message more appealing to a mature, educated market (Chandy et al., 2001). Generic advertising effectiveness, on the other hand, is limited to early pre-competition stages potentially due to its higher informational value. Hence, the final *when* question can be answered with “branded advertising after competition,” but generic should be useful at the early stages of the product life cycle.

Chapter 3 contributes to the advancement of knowledge on advertising effects in two ways: First, using an appropriate model representation, it explores dynamic effects of DTCA. No other study has examined dynamic DTCA effects, and relatively few studies have examined the issue of advertising dynamics in general. The different dynamic paths estimated for DTCA and detailing effects, emphasize the merit of uncovering the dynamics of consumer and physician directed marketing activities separately. Second, it is the first field study in marketing to empirically assess the effectiveness of purely generic advertising messages for a differentiated product. Thus, it complements theoretical studies on the subject (Bass et al. 2005), which prescribe optimal allocation policies for generic and branded advertising. Our findings on higher branded advertising effectiveness in the latter stages of the product life-cycle provide support for the Bass et al. (2005) recommendation on progressively higher branded advertising allocation. Finally, from a methodological point of view, we extend the AKF(C-D) algorithm by randomizing initial state priors through the use of Monte Carlo simulation to provide confidence intervals for the posterior estimates of the parameters at every time point. Validation results confirm the superiority of our proposed approach compared to existing alternatives.

In chapter 4, we argue for a custom-fit approach to measuring the effect of promotional mix on the diffusion of new prescription pharmaceuticals due to the idiosyncratic elements of prescription markets. Specifically, we suggest that an early market for prescriptions, which may be formed even before the launch of the pharmaceutical, is created due to persistent and severe symptoms suffered by a class of patients. When the prescription drug is eventually launched, the accumulated demand by the early market is realized, typically in the form of an exponential adoption curve. The launch of the drug also triggers adoption by a second, “late,” market corresponding to patients with milder conditions or lack of “classic” or persistent symptoms for whom the drug may be chosen as the most convenient solution to treating the condition. This is particularly likely in the case of drugs treating diseases the severity of which is located in a continuum (mild to severe) rather than those manifested in a dichotomous manner (presence/absence of disease). Lifestyle-related diseases treated by drugs with mild side effects fit well this profile.

To address the idiosyncratic elements of pharmaceutical adoption, we propose a switching regime dual-market diffusion model that extends the well-known Generalized Bass Model (GBM) by explicitly accommodating the possibility of an early market disconnected from the late main market. We use a dynamic switching probability to capture the transition from one market to the other. Our model shows good parameter face and forecasting validity compared to several benchmarks including the standard GBM and one that assumes that both early and late adoptions originate from a common, “pooled,” market with two segments. The latter model bears similarities with Moe and Fader’s (2002) model and a special case of Van den Bulte and Joshi’s (2007) Asymmetric Influence Model (AIM).

The results also suggest that marketing activities affect the two distinct markets in quite different ways. While the early market adoption seems to be resistant to all promotional efforts, the late market adoption is significantly affected by consumer directed advertising and physician journal advertising. The lack of marketing mix effectiveness on early market adoption suggests that this market has a pre-defined need for the product and adoption considerations are most likely based on efficacy. However, it is possible that early markets may be following pre-announcements or early press releases before the launch of a drug, which will eventually influence their adoption decisions. Unfortunately, this type of data were not available for our study, however this could be a topic worthy of further research. The effectiveness of marketing activities, at least DTC and physician journal advertising, on the late market may indicate that the product "creates" this market just like in typical product markets. This is consistent with views in the medical field that marketing may lead to overprescription by promoting overuse. Although our study cannot provide a definitive answer to this issue, we hope that it will stimulate further research on the topic. While we suggest that the influence of marketing mix on the late market is significant, we also found, consistent with most marketing studies, that such effects are economically modest. Yet, as noted above, the medical literature frequently attributes prescription inflation to heavy "marketization." One possible explanation, that may reconcile medical and marketing views, is that most category expansion effects, and potential overprescribing phenomena, are caused by the mere entry of new products and the accompanying launch campaigns and buzz as was apparently the case with SSRIs (Selective Serotonin Re-uptake Inhibitors) in the antidepressant market. Databases combining brand entry, publicity and marketing mix information can allow for the testing of lead and lag effects of all these factors and shed light on whether marketing activities may create markets and potentially lead to pharmaceutical overprescribing.

## 5.2. Managerial Implications

The results in this thesis have several implications for managers which can be used to improve the efficiency of marketing communications campaigns, and hence, facilitate more resource allocation for consumer satisfaction. The conceptual implications include:

- There may be significant threshold levels for various media in multi-media communications campaigns, below which advertising efforts are ineffective (Chapter 2);
- There may be saturation levels in multi-media communications campaigns, over which advertising efforts are ineffective (Chapter 2);
- Simultaneous use of multiple media might lead to early saturation or supersaturation for advertising. The former indicates that for

the majority of the spending range the effectiveness of that certain medium is constant; whereas, the latter connotes negative market response over the saturation point, due to consumers getting overwhelmed with too many messages (Chapter 2);

- There are important interaction effects present across media, which essentially suggest that the overall effect of multi-media campaign is in fact different and, most often, greater than the sum of individual effects of each medium (Chapter 2);
- There are usually multiple segments of consumers with different thresholds, which notably affect the response pattern for advertising (Chapter 2);
- Advertising effects possess significant dynamics (Chapter 3);
- Advertising can still be effective under regulatory environments; and the important issue is to recognize when different types of messages are more effective (Chapter 3);
- Generic advertising has an informative role and is more effective during the initial periods following the launch of a pioneer brand in a product category. Although, the application in this thesis is

within the pharmaceutical context and concerns the pioneer drug in a new therapeutic category, this result could be helpful in other contexts where managers are limited to restrict their advertising content due to regulations (Chapter 3);

- Branded advertising mainly acts as a reminder and is more effective at the later stages of the product life cycle, only after the market becomes informed about the product category (Chapter 3);
- Although direct-to-consumer advertising has attracted the majority of attention regarding marketing of prescription drugs, it's effects are modest; and detailing is still a more effective tool for prescription drug promotion (Chapter 3);
- Marketing efforts may affect the adoption process of different markets in quite distinct ways, and the idiosyncratic elements of the corresponding product category are important factors in understanding the diffusion pattern and its determinants (Chapter 4);

- The evidence of dual-market phenomenon in the pharmaceutical industry, consistent with similar findings in technological innovations, is critical for top management in terms of communication with the various stakeholders and better management of expectations regarding revenues and earnings forecasts. (Chapter 4)

In this thesis, we try to highlight the importance of sound methodologies in handling the complex problems in marketing communications. In the various chapters of the thesis, we demonstrate that adaptive and flexible models (1) provide a more accurate understanding of complex integrated marketing communications effects, (2) achieve noticeably better fit and hold-out forecasts, and more importantly (3) enable tackling of interesting managerial questions, which may not necessarily be addressed with simpler, more traditional, approaches. The tools we employ and the models we develop in this thesis can be used by managers to better forecast the short and long-term consequences of their actions on market performance, perform "what-if" analysis and gain insights to help improve their decisions.

# APPENDICES

## APPENDIX 1: Monte Carlo AKF (C-D) algorithm

The Monte Carlo AKF(C-D) algorithm is presented below.

1. Set t=0 and fix a large integer m. Provide the initial value of the state covariance matrix, $P_t$ ,[40] and set the process noise covariance, W[41], and measurement noise covariance, V[42], to reasonable values. Specify the initial value of the state variables for sales, $T^c, R^c$ and $L^c$.

2. Draw $m$ random samples for the parameter vector, $\theta_t$, from the assumed initial prior, $\mathrm{MVN} \sim (\mu_t, \sigma_t)$. Call each of these random variables $\theta_{tj}^i$ for i=1,..., $m$ and j=1,..,J representing marketing variables of interest[43].

3. Start the time update process.

Using the initial value of the augmented state vector, $y_{tj}^i = [T_t^{ci} \quad R_t^{ci} \quad L_t^{ci} \quad \theta_t^i]^T$, generate N values of $\hat{y}_{t+1}^{-i}$ by integrating the deterministic part of the transition equation over the interval $(t, t+1)$ for each i=1,..., $m$.

---

[40] Because of the fact that state covariance is updated at every step of the algorithm, the resulting values are found to be not dependent on the initial guess of $P_{t_0}$ which converges to more or less the same values after a certain number of observations. This is also supported by Stengel (1986).

[41] Usually one assumes small covariance for the process noise. Although letting W=0 is possible, a non-zero value gives more flexibility in "tuning" the filter (Welch & Bishop, 2001)

[42] Although observation noise of prescription drugs would not be high in general, it is not necessarily negligible. To be on the safe side, we set the standard error of observation noise to 1% of the observed number of prescriptions.

[43] In our case, J=9 including of effectiveness parameters for branded advertising, generic advertising, physician journal adv., detailing and competition among others as explained in the modeling section.

(Welch & Bishop, 2001)

$$\int_{t}^{t+1} \frac{dy^i}{dt} dt = \int_{t_k}^{t_k+1} f_{y^i}(y(t), \mathrm{X(t)}) dt \qquad \text{(A1)}$$

$$where,$$

$$f_{y^i} = \begin{bmatrix} f_{T^{ci}} & f_{R^{ci}} & f_{L^{ci}} & f_{\theta^i} \end{bmatrix}'$$

4. Similarly generate $m$ values of a priori state covariance matrix, $P_{t+1}^{-i}$ , by computing the integral below over the same interval.

$$\int_{t}^{t+1} \frac{dP^i}{dt} dt = F(y,t)P^T + PF^T(y,t) + W$$

$$where, \qquad \text{(A2)}$$

$$F(y,t) = \frac{\partial f_y[y(t), \mathrm{X(t)}]}{\partial y} \mid y(t) = y$$

When the new observation (t+1) becomes available start the measurement update (correction) process.

5. Compute the Kalman Gain for each $m$.

$$\varphi_{t+1}^{i} = \hat{P}_{t+1}^{-i} H^T (H\hat{P}_{t+1}^{-i} H^T + V)^{-1} \qquad \text{(A3)}$$

6. Compute the posterior state estimates, $\hat{y}_{t+1}^{i}$, by updating each of the $m$ state vectors proportional to their distance from their respective measurement error which can be computed from the observation equation.

$$\hat{y}_{t+1}^{i} = \hat{y}_{t+1}^{-i} + \hat{\varphi}_{t+1}^{i}(z_{t+1} - H\hat{y}_{t+1}^{-i}) \tag{A4}$$

7. Update the state covariance matrix for each $m$ to get the posterior covariance, $\hat{P}_{t+1}^{i}$.

$$\hat{P}_{t+1}^{i} = \left[I - \hat{\varphi}_{t+1}^{i}H\right]\hat{P}_{t+1}^{-i} \tag{A5}$$

8. Set t=t+1 and go back to step 3.

## APPENDIX 2: Simulations for the realized adoption curve

It can be shown that the realized adoption curve for the early market exhibits an exponential-like pattern as long as the time of launch T is past the ideal times peak, in other words $\underline{t} > T$ for some $\underline{t}$ past the peak of the ideal adoption times. At the time of launch T, the actual adoption hazard rate of those early market adopters with ideal adoption times less than T will be constant and equal to p+q*F(T) (equation 1), since they have not yet adopted the product due to its unavailability. The proportion of those early market adopters with a constant hazard rate is F(T) and their corresponding adoption density is exponential with rate p+q*F(T), representing the accumulated demand density, which can be expressed as:

$$f_a(t) = (p + qF(T))e^{-(p+qF(T))t}$$

For the remaining of the early market (1-F(T)), ideal and realized adoption times are identical since the launch time precedes their ideal adoption times. Thus, based on equation (1), their adoption density follows a truncated Bass distribution:

$$f_c(t)=\frac{1}{1-F(T)}\frac{(p+q)^2}{p}\frac{e^{-(p+q)t}}{(1+\frac{q}{p}e^{-(p+q)t})^2}I_{[t\geq T]}$$

where I is an indicator function. The mode of this density function is at T since the launch takes place past the peak of ideal adoption times. Thus, the corresponding hazard rate is non-increasing.

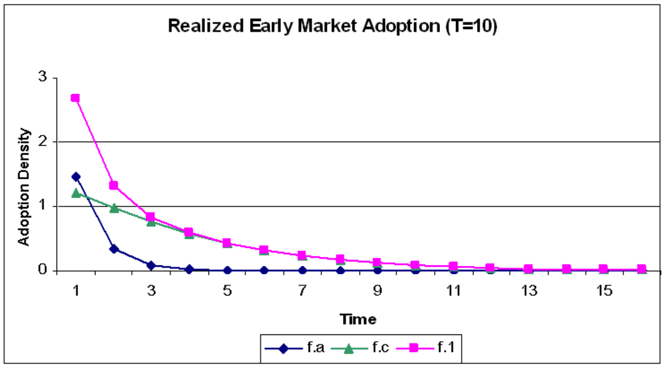
It follows then that the resulting density function of the realized adoption times for the early market has the following mixture form:

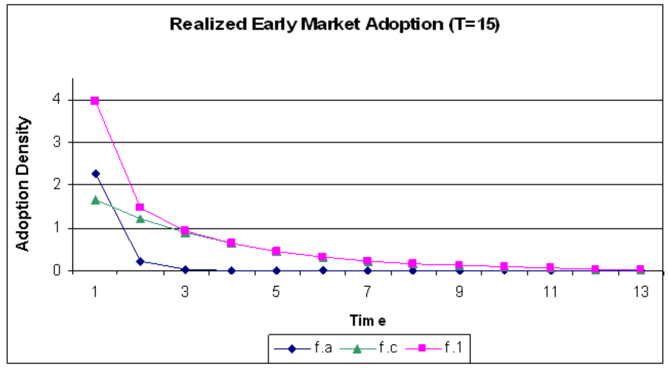
$$f_1(t)=F(T)*f_a(t)+(1-F(T))*f_c(t)$$

This is the mixture of two density functions with non-increasing failure rates and according to Proschan's theorem (Proschan, 1963), its failure rate is also non-increasing suggesting that the early market realized adoption curve has an exponential-like shape with a mode at T.

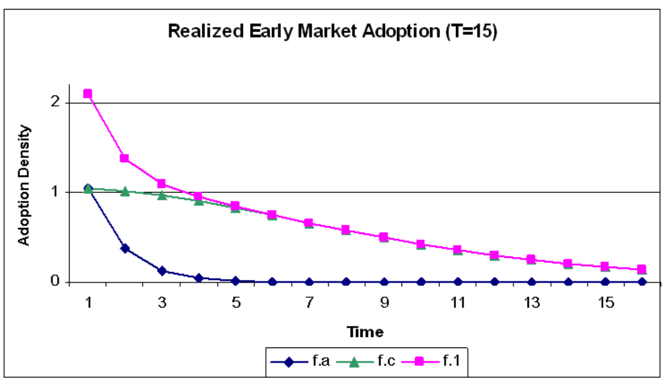
To demonstrate numerically that such mixtures actually have monotonically declining densities, we simulated the realized adoption density for the early market using different sets of plausible diffusion parameters for different launch times, T, past the peak of the ideal adoption density. Below, two representative simulations are presented. Example 1, uses diffusion parameter values from the meta-analytic study of Sultan, Farley and Lehmann (1990) with launch times T=10 and 15 respectively. A similar exercise is carried out for a slower diffusion process, shown in Example 2 with launch times T=15 and 20. Both examples show that the realized early market densities follow an exponential-like, monotonically decreasing pattern independent of the magnitude of the diffusion parameters. It is also noteworthy that the realized adoption curve quickly converges to the truncated one, which represents the "laggards" of the early market, lending further validity to our arguments. Hence, an exponential function should adequately capture early market adoption.

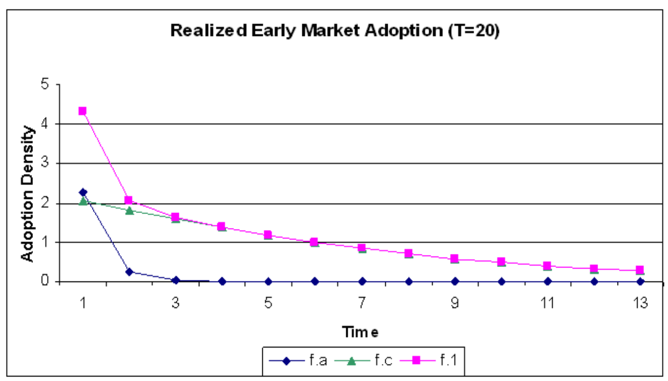
**Example 1**: p=0.04, q=0.3, Ideal Adoption Peak: t=6




**Example 2**: p=0.01, q=0.2, Ideal Adoption Peak: t=13



# BIBLIOGRAPHY


Abe, M. (1995). A Nonparametric Density-Estimation Method for Brand Choice Using Scanner Data. *Marketing Science, 14*(3), 300-325.

Ackoff, R. L., & Emshoff, J. R. (1975). Advertising Research at Anheuser-Busch, Inc. (1963-68). *Sloan Management Review* (Winter), 1-15.

Albers, S., Mantrala, M., & Sridhar, S. (2008). Meta-Analysis of Personal Selling Elasticities [Electronic Version]. *MSI Reports*, No:08-001, 3-24. Retrieved April 2008.

Alston, J. M., Chalfant, J. A., & Piggott, N. E. (2000). The Incidence of The Costs and Benefits of Generic Advertising. *American Journal of Agricultural Economics, 82*(3), 665-671.

Assmus, G., Farley, J. U., & Lehmann, D. R. (1984). How Advertising Affects Sales - Meta-Analysis of Econometric Results. *Journal of Marketing Research, 21*(1), 65-74.

Bass, F. M. (1969). New Product Growth for Model Consumer Durables. *Management Science Series a-Theory, 15*(5), 215-227.

Bass, F. M., Bruce, N., Murthi, B. P. S., & Majumdar, S. (2007). Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Ad-Sales Relationship. *Marketing Science, 26*(2), 179-195.

Bass, F. M., Krishnamoorthy, A., Prasad, A., & Sethi, S. P. (2005). Generic and Brand Advertising Strategies In A Dynamic Duopoly. *Marketing Science, 24*(4), 556-568.

Bass, F. M., Krishnan, T. V., & Jain, D. C. (1994). Why the Bass Model Fits without Decision Variables. *Marketing Science, 13*(3), 203-223.

Becker, G. S., & Murphy, K. M. (1993). A Simple Theory of Advertising as a Good or Bad. *Quarterly Journal of Economics, 108*(4), 941-964.

Belch, G. E., & Belch, M. A. (1999). *Advertising and Promotion – An Integrated Marketing Communications Perspective*. Boston, MA: McGraw Hill.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Bemmaor, A. C. (1984). Testing Alternative Econometric-Models on the Existence of Advertising Threshold Effect. *Journal of Marketing Research, 21*(3), 298-308.

Berndt, E. R., Bui, L., Reiley, D. R., & Urban, G. L. (1995). Information, Marketing, and Pricing in the Us Antiulcer Drug Market. *American Economic Review, 85*(2), 100-105.

Bijmolt, T. H. A., Van Heerde, H. J., & Pieters, R. G. M. (2005). New empirical generalizations on the determinants of price elasticity. *Journal of Marketing Research, 42*(2), 141-156.

Blattberg, R. C., Briesch, R., & Fox, E. J. (1995). How Promotions Work. *Marketing Science, 14*(3), G122-G132.

Bowman, D., Heilman, C. M., & Seetharaman, P. B. (2004). Determinants of Product-use Compliance Behavior. *Journal of Marketing Research, 41*(3), 324-338.

Brester, G. W., & Schroeder, T. C. (1995). The Impacts Of Brand And Generic Advertising On Meat Demand. *American Journal Of Agricultural Economics, 77*(4), 969-979.

Bronnenberg, B. J., Kruger, M. W., & Mela, C. (2008). The IRI Marketing Data Set. *Marketing Science, 27*(4), 745-748.

Bruce, N. (2007). Pooling and Dynamic Forgetting Effects In Multi-Theme Advertising: Tracking The Ad Sales Relationship With Particle Filters. *Marketing Science, Forthcoming*.

Bucklin, R. F., & Gupta, S. (1999). Commercial Use of UPC Scanner Data: Industry and Academic Perspectives. *Marketing Science, 18*(3), 247.

Calfee, J., Winston, C., & Stempski, R. (2002). Direct-to-Consumer Advertising and the Demand for Cholesterol-Reducing Drugs. *Journal of Law and Economics, 45*(2), 673-690.

Cestre, G., & Darmon, R. Y. (1998). Assessing consumer preferences in the context of new product diffusion. *International Journal of Research in Marketing, 15*(2), 123-135.

Chandy, R. K., Tellis, G. J., MacInnis, D. J., & Thaivanich, P. (2001). What to Say When: Advertising Appeals in Evolving Markets. *Journal of Marketing Research, 38*(4), 399-414.

Comanor, W. S., & Wilson, T. A. (1974). *Advertsing and Market Power*. Cambridge MA.: Harvard University Press.

Danaher, P. J., Bonfrer, A., & Dhar, S. (2008). The Effect of Cometitive Advertising Interference on Sales for Packaged Goods. *Journal Of Marketing Research, Forthcoming*.

Danaher, P. J., Hardie, B. G. S., & Putsis, W. P. (2001). Marketing-Mix Variables and the Diffusion of Successive Generations of a Technological Innovation. *Journal of Marketing Research, 38*(4), 501-514.

Desiraju, R., Nair, H., & Chintagunta, P. (2004). Diffusion of New Pharmaceutical Drugs in Developing and Developed Nations. *International Journal of Research in Marketing, 21*(4), 341-357.

Dockner, E., & Jorgensen, S. (1988). Optimal Advertising Policies for Diffusion-Models of New Product Innovation in Monopolistic Situations. *Management Science, 34*(1), 119-130.

Dye, R. (2000). The Buzz on Buzz. *Harvard Business Review, 78*(6), 139-146.

Findlay, S. D. (2001). Direct-to-consumer promotion of prescription drugs - Economic implications for patients, payers and providers. *Pharmacoeconomics, 19*(2), 109-119.

Fischer, M., & Albers, S. (2007). Patient-or Physician-Oriented Marketing: What Drives Primary Demand for Prescription Drugs? *MSI Working Paper*, [07-117].

Friedman, J. (1991). Multivariate Adaptive Regression Splines. *Annals of Statistics, 19*(1), 1-67.

Gatignon, H. (1993). Marketing Mix Models. In J. Eliashberg & G. L. Lilien (Eds.), *Handbook of OR and MS* (Vol. 5, pp. 697–732). New York: North-Holland.

Gatignon, H., & Hanssens, D. M. (1987). Modeling Marketing Interactions with Application to Salesforce Effectiveness. *Journal of Marketing Research, 24*(3), 247-257.

Gellad, Z. F., & Lyles, K. W. (2007). Direct-to-consumer advertising of pharmaceuticals. *American Journal of Medicine, 120*(6), 475-480.

Givon, M., Vijay, M., & Eitan, M. (1995). Software Piracy: Estimation of Lost Sales and The Impact of Software Diffusion. *Journal of Marketing, 59*(1), 29-37.

Goldenberg, J., Libai, B., & Muller, E. (2002). Riding the saddle: How cross-market communications can create a major slump in sales. *Journal of Marketing, 66*(2), 1-16.

Hahn, M., Park, S., Krishnamurthi, L., & Zoltners, A. A. (1994). Analysis of New Product Diffusion Using a 4-Segment Trial-Repeat Model. *Marketing Science, 13*(3), 224-247.

Hahn, M., Park, W. C., & Macinnis, D. J. (1992). The Adaptive Information Processing Hypothesis: Accounting for the V-Shaped Advertisign Response Function. *Journal of Advertising, 21*(2), 37-46.

Hanssens, D. M., Parsons, L. J., & Schultz, R. L. (2001). *Market Response Models, Econometric and Time Series Analysis* (2 ed.). Boston, MA: Kluwer Academic Press.

Hardle, W. (1999). *Applied Non-Parametric Regression* (Vol. 19). Cambridge, UK: Cambridge University Press.

Harvey, A. C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Springer Science and Business Media.

Havlena, W. J. (2008). *Measuring Marketing Effectiveness in a New Media World.* Paper presented at the MSI Conference, IESE Business School, Barcelona, Spain.

Hui, S. K., Eliashberg, J., & George, E. I. (2008). Modeling DVD Pre-Order and Sales: An Optimal Stopping Approach. *Marketing Science, Forthcoming*.

Hyde, J., Calnan, M., Prior, L., Lewis, G., Kessler, D., & Sharp, D. (2005). A Qualitative Study Exploring How GPs Decide to Prescribe Antidepressants. *British Journal of General Practice, 55*(519), 755-762.

Iizuka, T., & Jin, G. Z. (2005). The Effect of Prescription Drug Advertising on Doctor Visits. *Journal Of Economics & Management Strategy, 14*(3), 701-727.

Jiang, Z., Bass, F. M., & Bass, P. I. (2006). Virtual Bass Model and the Left-Hand Data-Truncation Bias in Diffusion of Innovation Studies. *International Journal of Research in Marketing, 23*(1), 93-106.

Kalish, S. (1983). Monopolist Pricing with Dynamic Demand and Production Cost. *Marketing Science, 2*, 135-159.

Kalyanam, K., & Shively, T. S. (1998). Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach. *Journal of Marketing Research, 35*(1), 16-29.

Kolsarici, C., & Vakratsas, D. (2008). *Dynamic Effects of Highly Regulated Advertising Messages*.Unpublished manuscript, Submitted.

Kolsarici, C., & Vakratsas, D. (2009). *Investigating Complex Effects of Multi-Media Communications: Cross-Media Synergies, Thresholds and Saturation Levels*.Unpublished manuscript, Working Paper.

Kotler, P., Armstrong, G., & Cunningham, P. H. (2007). *Principles of Marketing* (7 ed.): Pearson Education Canada.

Kravitz, R. L., Epstein, R. M., Feldman, M. D., Franz, C. E., Azari, R., Wilkes, M. S., Hinton, L., & Franks, P. (2005). Influence of Patients' Requests for Direct-to-Consumer Advertised Antidepressants - A Randomized Controlled Trial. *Journal of the American Medical Association, 293*(16), 1995-2002.

Kremer, S. T. M., Tammo H.A. Bijmolt , Jaap E. Wieringa, & Leeflang, P. S. H. (2008). Generalizations on the Effectiveness of Pharmaceutical Promotional Expenditures. *International Journal of Research in Marketing, 25*(4), 234-246.

Krishnamurthy, S. (2000). Enlarging the Pie vs. Increasing One's Slice: An Analysis of the Relationship Between Generic and Brand Advertising. *Marketing Letters, 11*(1), 37.

Krishnamurthy, S. (2001). The Effect of Provision Points On Generic Advertising Funding. *Marketing Letters, 12*(4), 315-325.

Leeflang, P. S. H., Wittink, D. R., Wedel, M., & Naert, P. A. (2000). *Building Models for Marketing Decisions*. Boston: Kluwer Academic Publishers.

Leone, R. P., & Schultz, R. L. (1980). Study of Marketing Generalizations. *Journal of Marketing, 44*(1), 10-18.

Lodish, L. M., Abraham, M., Kalmenson, S., Livelsberger, J., Lubetkin, B., Richardson, B., & Stevens, M. E. (1995). How Tv-Advertising Works - a Metaanalysis of 389 Real-World Split Cable Tv-Advertising Experiments. *Journal of Marketing Research, 32*(2), 125-139.

Lyles, A. (2002). Direct Marketing of Pharmaceuticals to Consumers. *Annual Review of Public Health, 23*, 73-91.

MacInnis, D. J., Rao, A. G., & Weiss, A. M. (2002). Assessing when increased media weight of real-world advertisements helps sales. *Journal of Marketing Research, 39*(4), 391-407.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New Product Diffusion-Models In Marketing - A Review And Directions For Research. *Journal Of Marketing, 54*(1), 1-26.

Mahajan, V., Muller, E., & Wind, Y. (2000). *New-Product Diffusion Models*. Boston: Kluwer Academic Publishers.

Manchanda, P., D.R. Wittink, A. Ching, P. Cleanthous, M. Ding, X.J. Dong, P.S.H. Leeflang, S. Misra, N. Mizik, S. Narayanan, T. Steenburgh, J.E. Wieringa, M. Wosinska, & Xie, Y. (2005). Understanding Firm, Physician and Consumer Choice Behavior In The Pharmaceutical Industry. *Marketing Letters, 16*(4), 293-308.

Marshall, A. (1919). *Industry and Trade*. London: MacMillan and Co.

Martin, R. M., Hilton, S. R., Kerry, S. M., & Richards, N. M. (1997). General Practitioners' Perceptions of the Tolerability of Antidepressant Drugs: A Comparison of Selective Serotonin Reuptake Inhibitors and Tricyclic Antidepressants. *British Medical Journal, 314*(7081), 646-651.

Middleton, N., Gunnell, D., Whitley, E., Dorling, D., & Frankel, S. (2001). Secular Trends in Antidepressant Prescribing in the UK, 1975-1998. *Journal of Public Health Medicine, 23*(4), 262-267.

Mizik, N., & Jacobson, R. (2004). Are Physicians "Easy Marks"? Quantifying the Effects of Detailing and sampling on New Prescriptions. *Management Science, 50*(12), 1704-1715.

Moe, W. W., & Fader, P. S. (2002). Using Advance Purchase Orders to Forecast New Product Sales. *Marketing Science, 21*(3), 347-364.

Montgomery, D. B., & Silk, A. J. (1972). Estimating Dynamic Effects of Market Communications Expenditures. *Management Science, 18*(10), 485-501.

Moore, G. A. (1992). *Crossing the Chasm*: HarperColins.

Muller, E., Peres, R., & Mahajan, V. (2007). Innovation Diffusion and New Product Growth: Beyond a Theory of Communications.

Nadaraya, E. A. (1970). Remarks on Non-Parametric Estimates for Density Functions and Regression Curves. *Theory of Probability and its Applications, 15*, 134-137.

Naik, P. A., Mantrala, M. K., & Sawyer, A. G. (1998). Planning Media Schedules in the Presence of Dynamic Advertising Quality. *Marketing Science, 17*(3), 214-235.

Naik, P. A., Prasad, A., & Sethi, S. (2008). Building Brand Awareness in Dynamic Oligapoly Markets. *Marketing Science, 54*(1), 129-138.

Naik, P. A., & Raman, K. (2003). Understanding the Impact of Synergy in Multimedia Communications. *Journal of Marketing Research, 40*(4), 375-388.

Narayanan, S., Manchanda, P., & Chintagunta, P. K. (2005). Temporal Differences in the Role of Marketing Communication in New Product Categories. *Journal of Marketing Research, 42*(3), 278-290.

Nestle, M. (1999). Dietary Supplement Advertising: Policies based on Politics, Not Science. *Journal of Nutrition Education, 31*(5), 278-282.

Olfson, M., & Klerman, G. (1993). Trends in the Prescription of Antidepressants by Office-based Psychiatrists. *American Journal of Psychiatry, 150*(4), 571-577.

Palda, K. S. (1964). *The Measurement of Cumulative Advertising Effects*. NJ: Prentice Hall.

Parker, P., & Gatignon, H. (1994). Specifying Competitive Effects In Diffusion Models: An Empirical Analysis. *International Journal of Research in Marketing, 11*(1), 17-39.

Parsons, L. J. (1975). Product Life-Cycle and Time-Varying Advertising Elasticities. *Journal of Marketing Research, 12*(4), 476-480.

Peralta, V., & Cuelta, M. J. (2007). Exploring The Borders of Schizoaffective Spectrum: A Categorical and Dimensional Approach. *Journal of Affective Disorders*, doi:10.1016/j.jad.2007.1009.1009.

Prasad, A. & Sethi, S.P. (2009). Integrated Marketing Communications in Markets with Uncertainty and Competition. *Automatica*, 45(3), 601-610

Prosser, H., Almond, S., & Walley, T. (2003). Influences on GPs' Decision to Prescribe New Drugs--The Importance of Who Says What. *Family Practice, 20*(1), 61-68.

Prosser, H., & Walley, T. (2003). Understanding Why GPs See Pharmaceutical Representatives: A Qualitative Interview Study. *British Journal of General Practice, 53*(489), 305-311.

Putsis, W. P. (1996). Temporal Aggregation in Diffusion Models of First-Time Purchase: Does Choice of Frequency Matter? *Technological Forecasting and Social Change, 51*(3), 265-279.

Rao, R. (1986). Estimating Continuous Time Advertising-Sales Models. *Marketing Science, 5*(2), 125-142.

Ristic, B., Arulampalam, S., & Gordon, N. (2004). *Beyond The Kalman Filter: Particle Filters For Tracking Applications*: Artech House.

Robertson, T. S., & Gatignon, H. (1986). Competitive Effects on Technology Diffusion. *Journal Of Marketing, 50*(3), 1-12.

Rogers, E. (1983). *Diffusion of Innovations* (3 ed.). NY: Free Press.

Rosenthal, M., Berndt, E. R., Donohue, J. M., Epstein, A. M., & Frank, R. G. (2003). Demand Effects of Recent Changes in Prescription Drug Promotion [Electronic Version]. *Forum for Health Economics & Policy*, 6, 1-6.

Rosenthal, M. B., Berndt, E. R., Donohue, J. M., & Frank, R. G. (2002). Promotion of Prescription Drugs to Consumers. *New England Journal of Medicine, 346*(7), 498-505.

Schmittlein, D. C., & Mahajan, V. (1982). Maximum Likelihood Estimation For An Innovation Diffusion Model of New Product Acceptance. *Marketing Science, 1*(1), 57.

Schultz, D. E., Tannenbaum, S. I., & Lauterborn, R. F. (1993). *Integrated Marketing Communications: Putting It Together And Making It Work*: McGraw-Hill.

Shah, S., & Reichman, W. E. (2006). Treatment of Alzheimer's Disease Across the Spectrum of Severity. *Clinical Intervention In Aging, 1*(2), 131-142.

Shankar, V., Carpenter, G. S., & Krishnamurthi, L. (1998). Late Mover Advantage: How Innovative Late Entrants Outsell Pioneers. *Journal of Marketing Research, 35*(1), 54-70.

Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. In *Monographs on Statistics and Applied Probability* (Vol. 26). London: Chapman & Hall.

Simon, H. (1997). Hysteresis in Marketing--A New Phenomenon? *Sloan Management Review, 38*(3), 39-49.

Simon, J. L., & Arndt, J. (1980). The Shape of the Advertising Response Function. *Journal of Advertising Research, 20*(4), 11-28.

Sloot, L. M., Fok, D., & Verhoef, P. C. (2006). The Short- and Long-Term Impact of an Assortment Reduction on Category Sales. *Journal of Marketing Research, 43*(4), 536 - 548

Srinivasan, V., & Mason, C. H. (1986). Nonlinear Least Squares Estimation of New Product Diffusion Models. *Marketing Science, 5*(2), 169-178.

Stengel, R. F. (1986). *Stochastic Optimal Control: Theory and Application*. NY: John Wiley and Sons.

Stremersch, S., & Lemmens, A. (2009). Sales Growth of New Pharmaceuticals Across the Globe:The Role of Regulatory Regimes. *Marketing Science, Forthcoming*.

Strom, B. L. (2000). *Pharmacoepidemiology*: John Wiley & Sons Ltd.

Sultan, F., Farley, J. U., & Lehmann, D. R. (1990). A Meta-Analysis of Applications of Diffusion-Models. *Journal of Marketing Research, 27*(1), 70-77.

Tellis, G. J. (2004). *Effective Advertising: Understanding When, How, and Why Advertising Works*. London: Sage Publications.

Tellis, G. J., Chandy, R. K., MacInnis, D. J., & Thaivanich, P. (2005). Modeling the microeffects of television advertising: Which ad works, when, where, for how long, and why? *Marketing Science, 24*(3), 359-366.

Tellis, G. J., Chandy, R. K., & Thaivanich, P. (2000). Which ad works, when, where, and how often? Modeling the effects of direct television advertising. *Journal of Marketing Research, 37*(1), 32-46.

Vakratsas, D. (2005). Advertising Response Models with Managerial Impact: An Agenda for the Future. *Applied Stochastic Models In Business And Industry, 21*(4-5), 351-361.

Vakratsas, D., & Ambler, T. (1999). How Advertising Works: What Do We Really Know? *Journal Of Marketing, 63*(1), 26-43.

Vakratsas, D., Feinberg, F. M., Bass, F. M., & Kalyanaram, G. (2004). The Shape of Advertising Response Functions Revisited: A Model of Dynamic Probabilistic Thresholds. *Marketing Science, 23*(1), 109-119.

Vakratsas, D., & Kolsarici, C. (2008). Dual-Market Diffusion for a New Prescription Pharmaceutical. *International Journal of Research in Marketing, 25*(4), 282-293.

Vakratsas, D., & Naik, P. (2007). Essentials of Planning Media Schedules. In T. A. a. G. Tellis (Ed.), *Handbook of Advertising*: Sage Publications.

Van den Bulte, C. (2000). New Product Diffusion Acceleration: Measurement and Analysis. *Marketing Science, 19*(4), 366-380.

Van den Bulte, C., & Joshi, Y. V. (2007). New Product Diffusion with Influentials and Imitators. *Marketing Science, 26*(3), 400-421.

Van den Bulte, C., & Lilien, G. L. (2001). Medical Innovation Revisited: Social Contagion versus Marketing Effort. *American Journal of Sociology, 106*(5), 1409-1435.

Van Heerde, H. J., Leeflang, P. S. H., & Wittink, D. R. (2001). Semiparametric analysis to estimate the deal effect curve. *Journal of Marketing Research, 38*(2), 197-215.

Venkataraman, S., & Stremersch, S. (2007). The Debate on Influencing Doctors' Decisions: Are Drug Characteristics the Missing Link? *Management Science, 53*(11), 1688-1701.

Watson, G. S. (1964). Smooth Regression Analysis. *Sankhya, Series A*(26), 359-372.

Webb, P. H. (1979). Consumer Initial Processing in a Difficult Media Environment. *Journal of Consumer Research, 6*(3), 225-236.

Webb, P. H., & Ray, M. L. (1979). Effects of Tv Clutter. *Journal of Advertising Research, 19*(3), 7-12.

Welch, G., & Bishop, G. (2001). *An Introduction to Kalman Filter*: Siggraph Publications.

Winer, R. S. (1979). Analysis of the Time-Varying Effects of Advertising - Case of Pinkham,Lydia. *Journal of Business, 52*(4), 563-576.

Wosinska, M. (2002). *Just What the Patient Ordered? Direct-to-Consumer Advertising and the Demand for Pharmaceutical Products*.Unpublished manuscript.

Wosinska, M. (2005). Direct-To-Consumer Advertising and Drug Therapy Compliance. *Journal of Marketing Research, 42*(3), 323-332.

Xie, J. H., Song, X. M., Sirbu, M., & Wang, Q. (1997). Kalman Filter Estimation of New Product Diffusion Models. *Journal of Marketing Research, 34*(3), 378-393.

Zuvekas, S. H., & Cohen, J. W. (2007). Prescription Drugs and The Changing Concentration of Health Care Expenditures. *Health Affairs, 26*(1), 249-257.