UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Three Essays on Marketing Dynamics

Autor:
Gokhan Yildirim

Directores:
Mercedes Esteban-Bravo
José Manuel Vidal-Sanz

DEPARTAMENTO DE ECONOMIA DE LA EMPRESA

Getafe, May 2012

TESIS DOCTORAL

THREE ESSAYS ON MARKETING DYNAMICS

Autor: Gokhan Yildirim

Directores: Mercedes Esteban-Bravo
José Manuel Vidal-Sanz

Firma del Tribunal Calificador:

Firma

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Calificación:

Getafe, de de

Acknowledgements

First and foremost, I am deeply grateful to my advisors, Mercedes Esteban Bravo and Jose Manuel Vidal Sanz. This thesis would not have been written without their help and guidance. Their untiring efforts and painstaking attention to any single detail yielded this dissertation. Their enthusiasm illuminated my inspiration towards analytical marketing. I benefited very much from their expertise on marketing modeling which was essential for building my research skills and helped me prepare for future challenges. I see them not only as my advisors but also members of my family.

I owe my deepest gratitude to Koen H. Pauwels whose help planning my academic career has been invaluable. He has shown me important skills for a qualified researcher. I have been inspired and motivated very much by his research since the very beginning. Throughout my Ph.D. work, I participated in some of his research projects through which I learned a lot and gained unique experience as to how to approach and solve managerially relevant marketing problems. I cannot find words to describe his invaluable support and contributions to my professional career.

I also wish to thank Marnik Dekimpe, Barbara Deleersnyder, Robert Rooderkerk and Ernst Osinga for their support and hospitality during the time I spent in the Marketing department at Tilburg University as a visiting PhD student. I am also grateful to them for some of the comments on my work.

I am also indebted to Nora Lado for both her wise advice on my research agenda and giving me the opportunity to join her research project that provided financial support to this thesis. Also my sincere thanks go to James Nelson who taught me the course of marketing research in the Ph.D. program, for being one of my referees. I also thank Fabrizio Cesaroni, Pablo Ruiz-Verdu, Marco Giarratana, Lola Duque and Samira Dias Dos Reis for giving me their useful recommendations on a variety of topics.

I warmly thank my colleagues, Andre Santos and Vardan Avagyan, with whom I had many helpful discussions. Among other friends and PhD mates, my special thanks go to Dilan, Argyro, Agata, Juliana, Ana Maria, Kenedy, Patricia, Emanuele, Irina, Jose, Maud, Peter, Sofi, Sara and Vincenzo. I had a wonderful time in Madrid with them.

Lastly, but most importantly, I am grateful to my wife, Asli, for her endless support, tolerance and fruitful talks on real-life marketing problems. I especially thank her for sacrificing many of our evenings and weekends over the past two years. I dedicate this thesis to all members of my family. I would like to express my deepest appreciation to all of them for all their love and encouragement.

I greatly acknowledge the financial support from both the department of Business Administration at Universidad Carlos III de Madrid and the Ministry of Science and Innovation in Spain (research grant ECO2011-30198) by which this research was made possible.

Gracias.

Abstract

This thesis comprises three essays on marketing in a dynamic context. The first essay focuses on optimal dynamic marketing budget allocation problem and proposes a novel decomposition algorithm that allows forward-looking firms to optimize their customer lifetime value (CLV) using simultaneously customized and mass marketing interventions. In addition to showing the performance of the algorithm by using numerical simulations, we provide an empirical application for a manufacturer of kitchen appliances. The second essay deals with sales uncertainty problem from a strategic marketing point of view. We model and forecast time-varying retail sales and marketing mix volatility. In particular, we examine within-brand and between-brand effects and trace the impact of marketing mix actions on sales growth volatility through volatility impulse-response functions. For the analysis, we use the data from six fast moving consumer product categories sold by Dominick's Finer Foods. The third essay centers on the analysis of trends in advertising media channels in the US. Using country-level annual time series data, we investigate whether there is a long-run equilibrium relationship among ten different advertising media channels.

The main contributions of this thesis can be summarized as follows:

- In the first essay, the contributions of our study are three-fold: (i) we solve the high dimensional stochastic dynamic programming (SDP) problem for a large number of customers through our decomposition algorithm, (ii) our model accommodates customized and mass marketing interventions simultaneously, (iii) we treat the CLV as an output rather than a tool for optimal dynamic marketing budget allocation planning. The simulation results as well as the empirical application of the model show that the proposed decomposition algorithm works well as the number of customers increases in the model.

Thus, marketing managers can use the model to determine the level of the price, how much money they should spend on each customer and on general advertising to maximize their CLV even if they have a high number of customers.

- In the second essay, our contribution is that we investigate time varying volatility which has been ignored in the sales response literature. The focus in the marketing literature has been given on the expected sales, but on volatility which is not a desired outcome. Using multivariate time series methodology, we find that lower price and promotional growth rates lead to less volatility in sales growth. Brand managers can use price and promotional actions as useful tools to curb sales volatility, and thus smoothing out the bullwhip effect at the retail level.

- Academic research on advertising at the country-level is less extensive compared to the company-level studies. In the third essay, we empirically investigate whether the entries of new advertising media (TV, yellow pages, cable and internet) affect the incumbents' expenditure level in the form of creating fundamental change in the long-run evolution. We model the dynamic interrelationship among ten different advertising media channels in the U.S. by using multivariate time series econometrics. Our results show that internet and cable media cause substantive shift only on the evolution of newspapers and outdoor, respectively whereas TV and yellow pages entries create fundamental change in the spending levels of all incumbents,except for direct mail. We also find that the long-run elasticity between total advertising expenditures and the GDP is negative implying that total advertising has countercyclical behavior. Furthermore, in the long-run, an increase in internet investment results in a decrease in newspapers as well as magazines investment.

Resumen

Esta tesis está compuesta por tres ensayos sobre marketing en contexto dinámico. El primero está enfocado en el problema de la asignación óptima y de forma dinámica del presupuesto en marketing. En él se propone un algoritmo novedoso que permite a las firmas con visión de futuro, optimizar su valor para el cliente de por vida (CLV) utilizando simultáneamente intervenciones personalizadas y masivas de marketing. Además de mostrar el desempeño de dicho algoritmo por medio de simulaciones numéricas, aportamos una aplicación empírica relativa a un productor de aparatos de cocina. El segundo ensayo trata el problema de la incertidumbre de las ventas desde el punto de vista del marketing estratégico. Desarrollamos un modelo que permite predecir en el tiempo las ventas al por menor y la volatilidad del marketing mix. En particular, examinamos los efectos en las marcas y entre marcas e indagamos el impacto de las acciones relacionadas al marketing mix en la volatilidad del crecimiento de las ventas a través de funciones de impulso-respuesta. Para el análisis usamos datos sobre seis categorías de productos de consumo envasados que se venden con rapidez y a relativamente bajo costo, en el detallista Dominick's Finer Foods. El tercer ensayo se centra en el análisis de las tendencias presentes en los medios de comunicación publicitaria en los Estados Unidos de América. Utilizando datos anuales, investigamos si existe una relación de equilibrio de largo plazo entre diez medios de comunicación publicitaria diferentes.

Las contribuciones principales de la tesis pueden resumirse así:

- En el primer ensayo las contribuciones giran alrededor de tres ejes: (i) nuestro algoritmo resuelve un problema de gran dimensionalidad al considerar un elevado número de clientes en un contexto de programación dinámica estocástica, (ii) nuestro modelo considera simultáneamente intervenciones masivas y personalizadas de marketing, (iii) tratamos el CLV como un resultado y no como una herramienta para la planificación óptima y dinámica del presupuesto en marketing. Tanto los resultados de la simulación como la aplicación empírica del modelo demuestran que

el algoritmo de descomposición propuesto funciona bien, incluso cuando el número de clientes aumenta.

Con esta propuesta, los gerentes de marketing pueden usar el modelo para determinar el precio, cuánto dinero deberían gastar en cada cliente y cuánto en publicidad general para maximizar su CVL aún cuando el número de clientes es elevado.

- En el segundo ensayo, nuestra contribución se centra en el análisis de la volatilidad, elemento ignorado en la literatura sobre la sensibilidad de las ventas a las variables de marketing. La literatura de marketing ha otorgado un papel importante a las ventas esperadas pero no se ha centrado en el los efectos adversos de la volatilidad de las ventas. Por medio del análisis de series temporales multivariante encontramos que tasas de crecimiento de precio y de promoción más bajas conllevan menor volatilidad en el crecimiento de las ventas. Los gerentes de marca pueden usar el precio y las acciones promocionales como herramientas útiles para reducir la volatilidad de las ventas y, por tanto, suavizar el efecto látigo a nivel de los minoristas.

- Hay pocos trabajos de investigación en relación a la inversión de publicidad por países comparado con el numero de estudios a nivel de empresa. En el tercer trabajo de esta tesis, empíricamente investigamos si las entradas de nuevos medios de comunicación (TV, páginas amarillas, cable e internet) afecta el nivel de inversión de los canales de publicidad tradicional, de forma radical en su evolución a largo plazo. En este trabajo, considerados un modelo de interrelación dinámica entre 10 medios de comunicación diferentes en los EE.UU. mediante el uso de la econometría de series temporales multivariantes. Nuestros resultados muestran que los medios de comunicación de Internet y de la TV por cable ocasionan un cambio estructural en la evolución de la inversión en periódicos y en medios al aire libre, respectivamente, mientras que el comienzo de la TV y las Páginas Amarillas crean un cambio estructural en los niveles de gasto de todos los otros medios, a excepción de la Publicidad Directa. También encontramos que la elasticidad de largo plazo entre los gastos de publicidad totales y el PIB es negativo lo que implica que la publicidad total tiene un comportamiento anticíclico. Por otra parte, en el largo plazo, un aumento en los resultados de la inversión en Internet implica en una disminución de inversión en periódicos, así como en revistas.

Contents

Chapter 1

Introduction

1.1 Marketing Models

Marketing modeling has become one of the main areas of marketing (Wierenga, 2008), and also important tools for managers in many industries (Leeflang, 2000). Prior to the development of these models, most marketing decisions were mainly based on heuristic notions, feeling or rules of thumb. When making their decisions, marketing managers were usually using their judgement and experience rather than quantitative method.

Marketing models, also known as analytical marketing or marketing science, began to emerge in the sixties with the purpose of helping companies support their marketing decisions by adopting mathematical models (Eliashberg and Lilien, 1993). Initially, three books were highly influential: Bass et al. (1961), Frank et al. (1962) and Buzzel (1964). They were the first authors developing and explaining how to implement marketing models, useing them to support managers making marketing decisions such as advertising, media planning, pricing, sales force allocation, forecasting and inventory control (Wierenga, 2008).

Since then, the journey of the marketing models has witnessed many developments. In the late sixties, Operations Research (OR) techniques became popular for optimizing marketing mix decisions (Montgomery and Urban, 1969). In subsequent years, market response models arose to examine the relationship between advertising and sales using different functional forms such as linear, concave or S-shape (Clarke, 1976; Little, 1979). Fitting data with nonlinear and dynamic models demanded more sophisticated statistical tools. As a consequence, modern econometric developments came into play to estimate these models. Starting from the nineties, dynamic models were on the rise. In particular, multivariate time series techniques came to the fore with the emphasis given to persistence modeling for making both short- and long-term inferences (e.g. Dekimpe and Hanssens, 1999; Pauwels, 2004). In recent years, customer-focused approach led to Customer Lifetime Value (CLV) models at the individual level as each customer was seen as the lifeblood of the company (e.g. Gupta, Lehmann and Stuart, 2004). Among the current CLV models, dynamic programming approach stood out as it considers customers' reactions to firm's decisions over time and provides better understanding of how a firm should set its marketing mix to maximize its CLV (e.g. Lewis, 2005a; 2005b).

Both advances in computer science and data availability also created many new opportunities that allowed marketing modelers to investigate different phenomena in the field. For instance, companies built huge databases within the Customer Relationship Management (CRM) systems to keep track of each of their customers. This enabled many researchers and firms to develop customer-based models using the latest computer technology (e.g. Reinartz, Thomas and Kumar, 2005). In addition, store level scanner data became widely available which triggered many of the studies on sales promotions (e.g. Dekimpe et al. 2005; Pauwels, 2007). Also, with the advent of the internet, online marketing data such as paid search, social media, visits, clicks, etc. became very important to measure the effectiveness of online marketing instruments (e.g. Rutz, Trusov and Bucklin, 2011; Wiesel, Pauwels and Arts, 2011).

All in all, these developments in the field of marketing models give us many opportunities to explore intriguing marketing problems in a dynamic context. As noted by Hanssens, Parsons and Shultz (2001), marketing responses rarely takes place in a static environment since customers and competitors anticipate and react to the firm's actions.

For example, Leeflang et al. (2009) point out that the effects of marketing efforts do not necessarily end when an advertising campaign is over. Therefore, dynamic models are particularly relevant as they allow us to study the evolution of sales, and the short and long term effectiveness of marketing actions. In this thesis, relying on particularly OR techniques and time series econometrics, we attempt to shed some light on three different marketing issues in a dynamic context.

1.2 Thesis Structure

In this section, we provide an overall perspective of the context of the thesis.

First article: CLV Maximization

In the first chapter, using the state-of-the-art optimization tools we solve the CLV maximization problem, by which marketing managers can allocate their CRM budget in a dynamic context.

Recently, marketing field has witnessed the shift from product-centered view to customer-focused approach (Jain and Singh, 2002; Villanueva and Hanssens, 2007). According to customer-focused approach customers are regarded as the primary assets of the firm (Gupta and Lehmann, 2003). Therefore, firms spend huge amount of money in order to create sustainable and profitable long-term relationships with their customers. The challenge that marketing managers face is to show the return on their CRM expenditures in order to justify their decisions and to make marketing financially accountable (Gupta and Zeithaml, 2006). Researchers develop marketing metrics and models to help marketers manage and measure the success of their actions (Gupta et al., 2006). One of the metrics that has gained much importance and been increasingly used in the literature is CLV which is defined as the expected net present value of future cash-flows obtained from a customer (Berger and Nasr, 1998). The CLV metric is often used to select customers and to guide the marketing budget allocation across those customers (Reinartz and Kumar, 2003; Rust, Lemon and Zeithaml, 2004; Venkatesan and Kumar, 2004).

Typically, in CRM programs, firms spend their advertising budget targeted to both each customer (individual advertising) and the whole customer base (general advertising). Therefore, in the first chapter of this thesis, we aim to answer the following question: How much should a firm invest simultaneously in individual advertising and general advertising so as to maximize its CLV for a large number of customers in its database? This is a large scale dimensional Stochastic Dynamic Programming (SDP) problem which cannot be solved easily because of the curse of dimensionality, i.e. the computational burden is very high as the dimension of the state variable increases. We develop a fully personalized model and propose a novel decomposition algorithm to solve this large scale SDP problem. Our model allows forward looking firms to design their optimal marketing budget allocation through which the optimal CLV is achieved. Simulation results show that standard algorithms, i.e. policy iteration and value iteration, cannot solve the problem for more than 3 customers whereas our proposed decomposition algorithm does not suffer from curse of dimensionality and reaches until 100 customers without any computational burden. The empirical application of the model is also successful for a medium-size of kitchen appliances firm which operates in international markets.

Second article: Sales Uncertainty

The second chapter utilizes multivariate time series econometrics and incorporates time varying volatility into the sales response models so as to examine sales uncertainty problem for marketing management.

Marketing literature has also paid a great deal of attention to measuring the marketing effectiveness of products or brands in the marketplace (e.g., Dekimpe and Hanssens 1999; Jedidi, Mela and Gupta, 1999; Dekimpe, Hanssens and Silva-Risso, 1999). In particular, the main focus has been given on the effects of price (e.g., Nijs et al., 2001), promotions (e.g., Dekimpe et al, 2005), advertising (e.g., Dekimpe and Hanssens, 1995b) and new products (e.g., Pauwels et al., 2004) on top- line performance (sales) or bottom-line performance (profits). Although the effects of these marketing actions on expected sales or profits are well documented in the literature, little is known about the effects of marketing mix on volatility which can be deemed as a second performance measure and not a desired outcome. In the second chapter of this thesis, we address this issue: sales uncertainty problem for marketing management.

Due to sales volatility, firms are exposed to some costs. For example, in the case of demand peaks, firm has to invest in its inventory in order to avoid stock-out situation. In the opposite scenario, when there is a fall in the demand, excessive inventory occurs implying again an extra cost to the firm. Apart from the operational costs, failure to manage sales volatility also gives rise to poor customer relationship, lower loyalty and often distrust among supply chain members. Sales volatility can be even more dangerous in the long supply-chain streams in which demand variability is amplified as one moves up the chain, known as bullwhip effect (Lee, Padmanabhan and Whang, 1997). As a result, in the second chapter, we seek an answer to the following question: Can a marketing manager mitigate sales volatility by using his marketing mix decisions?

Measuring the impact of marketing mix decisions on performance outcome metrics such as brand sales, profits and market shares requires time. As noted by Hanssens, Parsons and Shultz (2001) marketing responses rarely take place in a static environment although in some cases customers can react immediately to advertising and the effect may disappear while advertising is running. To capture this type of dynamic effects, time series analysis is suitable as it allows for modeling response behavior over time with the emphasis on the lag structures of variables and disturbances. Thus, for the above research question, we use multivariate time series methodology so as to examine the impact of marketing mix variables (price and promotions) not only on expected sales but also on sales growth volatility. Using the store-level scanner data from the Dominick's Finer Foods, provided by the James M. Kilts Center, University of Chicago, our results show that there is significant dependence in all analyzed fast moving consumer product categories (cheese, refrigerated juice, laundry detergent, toilet tissue, paper towel and toothpaste) for most of the brands in either mean, variance or both. Furthermore, lower price and promotional growth rates lead to less sales volatility. Hence, marketing mix can be helpful to curb sales volatility, and therefore reducing the bullwhip effect at the retail level.

Third article: Advertising Trends

As with the second chapter, the third chapter employs multivariate time series methodology in order to investigate the common trends in the US media channels.

Marketing managers observe closely the media channel expenditures to foresee long-term trends in the advertising industry by following the reports of external organizations. This information affects their long-run strategies on advertising budget decisions. On the other hand, advertising agencies try to convince the firms to advertise on their channels. With the new media entries such as internet, the advocates of old traditional media are concerned more about the possible future movements in the industry and about the effect of the new media entry on their channels. Academic research is less extensive on the country level advertising research (Tellis and Tellis, 2009) and more research is called for. Thus, understanding how the new players in the market affect the incumbents' key marketing metrics as well as the underlying trends in advertising expenditures of different media channels is important for advertising agencies, marketing managers and academics. Thus, in the third paper, we aim to explore the impact of the new media introduction on the long run equilibrium of the advertising industry. In particular, we address the following questions: Is there any long-run equilibrium relationship among all media channels? If any, which channel(s) responds more and faster to a deviation from this long-run relationship? How sensitive is the total advertising expenditures to the economic conditions in the long-run? More importantly, how are the old media affected by the introduction of new media according to the historical evidence?

We answer these questions by using annual time series data that cover the period of 1935-2007 for ten different media channels. Relying on the Vector Error Correction (VEC) modeling with multiple structural breaks, our results show that TV and yellow pages entries create fundamental (structural) change in the spending levels of the incumbents, while internet and cable cause substantive shift only in newspapers and outdoor, respectively. Moreover, the long-run elasticity between total advertising expenditures and the GDP is negative showing that total advertising has counter-cyclical behavior. Furthermore, crossed long-run elasticities show that an increase in internet investment results in a decrease in newspapers as well as magazines investment.

The structure of the thesis is as follows: In the second chapter, we focus on the essay entitled “Valuing Customer Portfolios with Endogenous Mass and Direct Marketing Interventions Using a Stochastic Dynamic Programming Approach”. The third chapter deals with the essay entitled “Can We Curb Retail Sales Volatility Through Marketing Mix Actions?”. Finally, in the fourth chapter, we present the essay entitled “US Advertising Expenditure Trends: Long-run Effects and Structural Changes with New Media Introductions”.

1.3 References

Bass, Frank M., Robert D. Buzzel, Mark R. Greene, William Lazer, Edgar A. Pessemier, Donald L. Shawver, Abraham Shuchman, Chris A. Theodore, and George W. Wilson, eds. (1961), *Mathematical Models and Methods in Marketing*, Homewood, IL: Richard D. Irwin.

Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12(1), 17-29.

Buzzel, Robert D. (1964), *Mathematical Models and Marketing Management*, Harvard University, Division of Research, Boston MA.

Clarke, Darral G. (1976), "Economic Measurement of the Duration of Advertising Effects on Sales," *Journal of Marketing Research*, 18(4), 345-357.

Dekimpe, Marnik G. and Dominique M. Hanssens (1995b), "The Persistence of Marketing Effects on Sales," *Marketing Science*, 14(1), 1-21.

Dekimpe, Marnik G. and Dominique M. Hanssens (1999), "Sustained Spending and Persistent Response: A New Look at Long-term Marketing Profitability," *Journal of Marketing Research*, 36(November), 397-412.

Dekimpe, Marnik G., Dominique M. Hanssens and Jorge M. Silva-Risso (1999), "Long-Run Effects of Price Promotions in Scanner Markets," *Journal of Econometrics*, 89, 269-291.

Dekimpe, Marnik G., Dominique M. Hanssens, Vincent R. Nijs and Jan-Benedict E.M. Steenkamp (2005), "Measuring Short- and Long-run Promotional Effectiveness on Scanner Data Using Persistence Modeling," *Applied Stochastic Models in Business and Industry*, 421-422.

Eliashberg, Jehoshua and Gary L. Lilien (1993), *Handbooks in Operations Research and Management Science*, Volume 5: Marketing, Elsevier Science Publishers, Amsterdam.

Frank, Ronald E., Alfred A. Kuehn and William F. Massy (1962), *Quantitative Techniques in Marketing Analysis*, Homewood, IL: Richard D. Irwin.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker and S. Sriram (2006), “Modeling Customer Lifetime Value,” *Journal of Service Research*, 9(2), November, 139-155.

Gupta, Sunil and Donald R. Lehmann (2003), “Customers as Assets,” *Journal of Interactive Marketing*, 17(1), 9-24.

Gupta, Sunil, Donald R. Lehmann and Jennifer Ames Stuart (2004), “Valuing Customers,” *Journal of Marketing Research*, 41(1), 7-18.

Gupta, Sunil and Valarie Zeithaml (2006), “Customer Metrics and Their Impact on Financial Performance,” *Marketing Science*, 25(6), 718-739.

Hanssens, Dominique M., Leonard J. Parsons and Randall L. Schultz (2001), *Market Response Models: Econometric and Time Series Analysis*, Second Edition, Boston: Kluwer Academic Publishers.

Jain, Dipak and Siddhartha S. Singh (2002), “Customer Lifetime Value Research in Marketing: A Review and Future Directions,” *Journal of Interactive Marketing*, 16(2), 34-46.

Jedidi, Kamel, Carl F. Mela, and Sunil Gupta (1999), “Managing Advertising and Promotion for Long-run Profitability,” *Marketing Science*, 18(1), 1-22.

Lee, Hau L., V. Padmanabhan and Seungjin Whang (1997). “Information Distortion in a Supply Chain: The Bullwhip Effect,” *Management Science*, 43 (4), 546-558.

Leeflang, Peter S.H., Dick R. Wittink. Michel Wedel and Philippe A. Naert (2000), *Building Models for Marketing Decisions*, Kluwer Academic Publishers, Boston MA.

Leeflang, Peter S.H., Tammo H.A. Bijmolt, Jenny van Doorn, Dominique M. Hanssens, Harald J. van Heerde, Peter C. Verhoef and Jaap E. Wierenga (2009), "Creating Lift versus Building the Base: Current Trends in Marketing Dynamics," *International Journal of Research in Marketing*, 26, 13-20.

Lewis, Michael (2005a), “Research Note: A Dynamic Programming Approach to Customer Relationship Pricing,” *Management Science*, 51(6), 986-994.

Lewis, Michael (2005b), "Incorporating Strategic Consumer Behavior into Customer Valuation," *Journal of Marketing*, 69(October), 230-238.

Little, John D.C. (1979), "Aggregate Advertising Models: The State of the Art," *Operations Research*, 27(4), 629-667.

Montgomery, David B., Glen L. Urban (1969), *Management Science in Marketing*, Prentice Hall, Englewood Cliffs, NJ.

Nijs, Vincent, Marnik G. Dekimpe, Jan-Benedict E.M. Steenkamp, and Dominique M. Hanssens (2001), "The Category-Demand Effects of Price Promotions," *Marketing Science*, 20(1), 1-22.

Pauwels, Koen (2004), "How Dynamic Consumer Response, Competitor Response, Company Support, and Company Inertia Shape Long-Term Marketing Effectiveness," *Marketing Science*, 23(4), 596-610.

Pauwels, Koen, Jorge Silva-Rosso, Shuba Srinivasan and Dominique M. Hanssens (2004), "New Products, Sales Promotions and Firm Value: The Case of the Automobile Industry," *Journal of Marketing*, 68(October), 142-156.

Pauwels, Koen (2007), "How Retailer and Competitor Decisions Drive the Long-term Effectiveness of Manufacturer Promotions for Fast Moving Consumer Goods," *Journal of Retailing*, 83, 297-308.

Reinartz, Werner J. and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67(January),77-99.

Reinartz, Werner, Jacquelyn S. Thomas and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability ," *Journal of Marketing*, 69(January), 63-79.

Rust, Roland T., Katherine N. Lemon and Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68(1), 109-127.

Rutz, Oliver, Michael Trusov and Randolph E. Bucklin (2011), "Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?," *Marketing Science*, 30(4), 646-665.

Tellis, Gerard J. and Kethan Tellis (2009), "A Critical Review and Synthesis of Research on Advertising in a Recession," *Journal of Advertising Research,* 49(September), 304-327.

Venkatesan, Rajkumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68(October), 106-125.

Villanueva, Julian and Dominique M. Hanssens (2007), "Customer Equity: Measurement, Management and Research Opportunities," *Foundations and Trends in Marketing*, 1(1), 1-95.

Wierenga, Berend (2008), *The Past, the Present, and the Future of Marketing Decision Models*, In B. Wierenga (ed), Handbook of Marketing Decision Models, New York: Springer, 3-20.

Wiesel, Thorsten, Koen Pauwels, and Joep Arts (2011), "Marketing's Profit Impact: Quantifying Online and Offline Funnel Progression," *Marketing Science*, 30(4), 604-611.

Chapter 2

Valuing customer portfolios with endogenous mass-and-direct marketing interventions using a stochastic dynamic programming decomposition

2.1 Introduction

Customers are central assets of the firm, and marketing departments increasingly adopt Customer Relationship Management (CRM) schedules to improve customer acquisition, expenditure and retention. Essentially, CRM involves a systematic allocation of differential resources to customers, based on the their individual value to the business. The resources allocated to each customer can be channelled through a mix of alternative interventions, and complemented by mass actions. Traditionally, marketing resource allocation was based on heuristic rules (see Mantrala, 2002). But the benefits of CRM policies are nowadays justified by their impact on firms' return (Rust, Lemon and Zeithaml, 2004). In order to plan the allocation of resources, managers should maximize the value of its customer base. This concept is ideally measured by the summa of Customer Lifetime Values (CLV), that is, the sum of the net present values of discounted cash flows between a customer and the firm (Gupta, Lehmann and Stuart 2004, Gupta and Lehmann 2006). The assessment of customers' values, and the effectiveness of a marketing intervention is typically based on the econometric analysis of large customer databases.

CRM requires planning a portfolio of alternative marketing mix interventions. The literature on budget allocation typically considers mass interventions from the marketing mix (advertising promotion and sales force, reference prices and price-promotions, product and production, and distribution channels). For a review see, e.g., Gupta and Steenburgh (2008) and Shankar (2008). The direct marketing literature typically considers a single intervention customized, or at least tailored to small segments. For example, it is common the use of certain pricing decisions (Lewis 2005), catalog mailing (see, e.g., Bitran and Mondschein 1996; Gönül and Shi 1998; Gönül and Ter Hofstede 2006; Simester et al. 2006), couponing decisions (e.g., Bawa and Shoemaker 1987; Rossi et al. 1996), direct mailing (Roberts and Berger 1989) and relationship-oriented magazines (Berry 1995, Bhattacharya and Bolton 1999, McDonald 1998).

Planning the optimal CRM interventions maximizing the global expected CLVs is, by all means, a difficult task. In an attempt to address it, the standard CRM procedure allocates marketing budget to each individual customer, after ranking customers by its CLV value (Reinartz and Kumar 2005, Rust, Lemon and Zeithaml 2004, Venkatesan and Kumar 2004). Assessing new marketing interventions using CLVs computed from historical data is potentially misleading. The planned CRM marketing interventions will change the purchasing behavior of different customers, changing their CLVs, turning upside-down the customers ranking and making our history-based decisions sub-optimal. To cope with this inherent endogeneity, the objective of the allocation marketing models should be a CLV measure computed as the optimal value achieved when the optimal

CRM investment is implemented. The idea is that when the CLV is computed we should take into account how customers will react to the changes in the CRM policies.

To avoid this endogeneity problem, some authors have tried to optimize the expected CLVs. Rust and Verhoef (2005) optimize each individual customer's profitability year by year (a myopic planning). Alternatively, other authors optimize the expected CLV using Stochastic Dynamic Programming (SDP).This is a natural approach to solve this problem, but SDP is affected by the curse of dimensionality (the complexity increases drastically with the size of the problems). Therefore, they consider a partial solution, that consists of ignoring mass interventions (aimed to all the customers) focusing on direct individual interventions, so that the investment decision for each customer is independent, and the standard SDP algorithms can be applied to at low computational cost considering "decoupled" decision problem. Gönül and Shi (1998) and Montoya et al. (2007) study direct marketing problems. Khan et al. (2009) estimate the impact of multiple promotional retail instruments, (discount coupons, free shipping offers, and a loyalty program) on customer behavior, designing a customized promotional schedule solving a different SDP problem for each customer. Yet, how to optimize simultaneously both types of interventions (mass, and direct ones) is an unsolved issue, as the SDP optimization problems are not separable among customers. Maximizing the expected CLVs of a customers portfolio with multiple types of personalized and mass marketing interventions, accounting for long term returns, and solving the endogeneity issue is what Rust and Chung (2006, p. 575) called the "Holy Grail" of CRM.

In this paper, we provide a fully tailored approach for planning policies that maximize the expected CLV of all the customers in the market accounting for the endogeneity issues. Our approach considers that customer behavior follows a Markov model in which sales respond to mass and direct marketing interventions, and marketing expenditures are allocated to maximize the sum of the expected CLVs for all its customers. Because such models can become rather intractable in general, we propose a method to address this problem by splitting it into manageable pieces (subproblems) and by coordinating the solutions of these subproblems. With this approach, we obtain two main computational advantages. First, the subproblems are, by definition, smaller than the original problem and therefore much faster to solve. Second, the uncertainty can be easily handled in each subproblems. To validate the efficiency of the approach, we provide a proof of convergence and have solved several stochastic dynamic CLV models. The numerical results show the effectiveness of the method to solve large-scale problems.

We also present an empirical application. We consider a medium size international wholesale company based in eastern Europe of built-in electric appliances for kitchens. This is a firm with various forms of sales response so its marketing budget allocation strategy involves general marketing investments (mainly advertising and promotions in professional fairs) and personalized customer investments. In this research, we therefore investigate whether these two types of interventions differ across customers. The results show that companies should consider different strategies to different customers to achieve long-term profitability over all of the periods of time.

The paper proceeds as follows. In Section 2.1, we provide a model for dynamically allocating marketing budgets in the context of CRM. The present model considers simultaneously direct marketing interventions tailored to each customer and mass marketing interventions aimed to the customer base. In Section 2.2, we present the proposed decomposition methodology. In section 2.3, we illustrate the performance of the algorithm using numerical simulations, and provide a proof of convergence. In Section 2.4, we present an empirical application to customers of manufacturer of kitchen appliances. Finally, in Section 2.5, we discuss the results and provide some concluding remarks. The Appendix provides technical details about the algorithm implementation.

A Model for optimal dynamic budget allocation in CRM

Planning marketing interventions in CRM requires managers to allocate budget dynamically by maximizing the sum of the expected CLVs from all customers based on historical customer state information. To address the optimal budget allocation problem, the firm must carry out two tasks (see, e.g., Gupta et al. 2009):

Task 1. Estimate the expected CLV by building analytical models to forecast future sales response by customers (Gupta and Lehmann 2003, 2005, Kamakura et al. 2005, Gupta and Zeithaml 2006); and

Task 2. Solve the stochastic dynamic optimization problem including all individual customers (see, e.g., Rust and Verhoef 2005, Rust and Chung 2006).

The first task requires the design of a dynamic panel sales response model. Let $\mathbb{I} = \{1, ..., I\}$ be a finite set of active customers and $t \in \{0, 1, 2, ...\}$ the time index. The firm chooses a sequence of dynamic controls:

- e_{it} is the direct marketing interventions on customer $i \in \mathbb{I}$ at period of time t, such as personalized advertising and directed promotional expenditures. We use the notation $e_t = (e'_{1t}, ..., e'_{It})'$, where e' denotes the transpose of e.
- A_t is the mass marketing interventions at period of time t,
- P_t denotes the prices for the different products.

These controls (A_t, P_t, e_t) are defined on the a control set $\mathcal{A}$, a Borel-measurable subset of the Euclidean space.

The dynamic control variables have an effect on the customer behavior state variables. We will consider the following state model:

- S_{it} is the random vector describing the sales-level state of customer $i \in \mathbb{I}$ at time t, and we use the notation $S_t = (S_{1t}, ..., S_{It})'$. With probability one, S_t takes values on a set of states S a Borel-measurable subset of the Euclidean space.
- We assume that S_t follows a Markovian process with transition probability

$$\begin{aligned} F(s'|s, A, P, e) &= \Pr(S_t \leq s'|S_{t-1} = s, A_{t-1} = A, P_{t-1} = P, e_{t-1} = e) \\ &= \prod_{i \in \mathbb{I}} F_i(s'_i|s_i, A, P, e_i) \end{aligned}$$

The typical example is when the company considers a dynamic panel model where each customer satisfies

$$S_{it} = \rho S_{it-1} + g_i \left(A_{t-1}, P_{t-1}, e_{it-1} \right) + \varepsilon_{it} \tag{2.1}$$

where $|\rho| < 1$, the innovation ε_{it} is a strong white noise independent for each customer with cumulative distribution $H_i(\cdot)$. The functions $g_i(\cdot)$ and $H_i(\cdot)$ are continuous and can vary across customers to allow heterogeneity in the expected responses, so that

$$F_i \left(s_i' | s_i, A, P, e_i \right) = \Pr \left\{ \varepsilon_i \leq s_i' - \rho s_i - g_i \left(A, P, e_i \right) \right\} = H_i \left(s_i' - \rho s_i - g_i \left(A, P, e_i \right) \right).$$

The one-lag memory structure imposed by the Markov dependence assumption can be relaxed by considering p-lags autoregressive models in the space-of-states.

The dynamic model can be estimated using standard econometric techniques for time series cross-section and/or dynamic panels. Firms increasingly store large panel data basis with information about their customers, including social information (such as socio-demographic, geographic information, lifestyle habits) and trade internal data (such as historical transaction records, customers feedback, or Web browsing records), see Bose and Chen (2009). The econometric literature has developed a battery of linear and non-linear models for the dynamic analysis of large data-panels, and the marketing researchers have tailored these models for the prediction of future purchases at customer-level (e.g., Schmittlein and Peterson 1994). Using these tools, company managers often estimate the expected CLV for each customer based on its past behavior, (generally in a *ceteris paribus* context, omitting or fixing the marketing mix variables).

The main contribution of this paper is to propose a methodology for solving Task 2. The firm should choose the CRM policy maximizing the expected sum of its CLVs, constrained to the customer response to feasible marketing policies. This problem is a large dimensional (discounted) SDP problem. In other words, we consider that a rational forward-looking firm has to decide on CRM budget allocation policies over time, drawing profits

$$r\left(S_t, A_t, P_t, e_t\right) := \sum_{i \in \mathbb{I}} r_i\left(S_{it}, A_t, P_t, e_{it}\right) \tag{2.2}$$

at each period of time $t > 0$ from all of their customers[1]. Let $\delta \in (0, 1)$ be a time discount parameter, then we assume that the company maximizes the expected net present value $E_0\left[\sum_{t \geq 0} \delta^t r\left(S_t, A_t, P_t, e_t\right)\right]$.

Marketing budget decisions generally face corporate constraints settled by the interactions between managers, bond holders, and stockholders. We consider that for each state S_{t-1}, there is a non-empty compact set $\mathbb{A}\left(S_{t-1}\right) \subset \mathcal{A}$ of admissible controls at time $t > 0$ which depends upon the previous period sales; i.e. $\left(A_t, P_t, e_t\right) \in \mathbb{A}\left(S_{t-1}\right)$. The admissible state-controls pairs are given by $\mathcal{K} := \{(S, A, P, e) : S \in \mathcal{S}, (A, P, e) \in \mathbb{A}(S)\}$. As usual, we assume that $|r(S, A, P, e)|$ is bounded on $\mathcal{K}$ except for a null probability set.

Problem 1 *Given the initial state S_0, the firm faces the following problem:*

$$\max_{\{(A_t, P_t, e_t) \in \mathbb{A}(S_{t-1})\}_{t>0}} E_0\left[\sum_{t \geq 0} \delta^t r\left(S_t, A_t, P_t, e_t\right)\right] := V\left(S_0\right)$$

As usual, we denote the maximum $V\left(S_0\right)$ as the "value function".

This is a SDP problem in discrete time. Problem 1 is solved by the optimal policy $\left(A^*(s), P^*(s), e^*(s)\right)$, which is a time-invariant function prescribing the best decision for each state s, i.e.

$$V\left(S_0\right) = E_0\left[\sum_{t \geq 0} \delta^t r\left(S_t, A^*\left(S_{t-1}\right), P^*\left(S_{t-1}\right), e^*\left(S_{t-1}\right)\right)\right].$$

Interestingly, for each period of time t, we can interpret $V\left(S_t\right)$ as the expected present discounted value of profits under the current state S_t. Under certain regularity conditions,

[1] We use the standard notation ":=" for definitions.

the optimal policy function $(A^*(s), P^*(s), e^*(s))$ are characterized by the value function $V(\cdot)$ as the solution of the Jacobi-Bellman equation (Bellman, 1955, 1957):

$$V(S_t) = \max_{(A_t, P_t, e_t) \in \mathbb{A}(S_{t-1})} \{r(S_t, A_t, P_t, e_t) + \delta E_t[V(S_{t+1})]\}.$$

Given the optimal policy rule, managers can make optimal decisions on marketing activities $(A^*(s), P^*(s), e^*(s))$, that maximize their expected profits given the sales state s observed in the previous period. Also, $V(S_t)$ gives us the company value derived from the CLVs customer portfolio at time t, provided that the firms are optimally managed. Using optimal policies for solving the SDP problem has several advantages: they are Simple (ease of understanding for managers) and Adaptive (the decisions can be automatically updated as new state-information becomes available). Note that they can be used also for simulation. For each period of time t, given S_t drawn from the conditional distribution $F(s|S_{t-1}, A_{t-1}, P_{t-1}, e_{t-1})$, the values $A_{t+1} = A^*(S_t)$, $P_{t+1} = P^*(S_t)$, $e_{t+1} = e^*(S_t)$ can be used to simulate Monte Carlo scenarios, and then to compute numerically the expected path for the optimal policies $E[A_t]$, $E[P_t]$, $E[e_t]$ and states $E[S_t]$, as well as confidence intervals.

The computation of large SDP remains one of the most challenging optimization problem. Most problems can become intractable as the dimension of the state space increases (the CPU time to calculate a value function increases exponentially in the dimension of the state space), which is the well known "curse of dimensionality" (Bellman, 1961). Due to the curse of dimensionality, SDP problems can be solved numerically for decision problems in which only few state variables are considered. This implies that CRM decision problems with more than 3 customers cannot be solved using the standard approaches: value iteration and policy iteration (see Appendix A for an introduction).

One of the classical strategies to solve large decision problems are the decomposition based approaches. There exists several mathematical programming decomposition algorithms for large optimization problems with an appropriate structure (Danzting-Wolfe and Benders-decomposition in convex problems, and augmented Lagrangian relaxation in nonconvex problems). Some attempts to solve large SDP problems combine traditional decomposition algorithms and statistical sampling techniques. Sampling is used to create a scenario tree that represents the uncertainty (Heitsch and Römisch, 2009). Then the original problem is approximated by a finite deterministic one. The dimension of the tree grows exponentially with the number of states variables, and so does the complexity of the deterministic problem. To tackle this issue, a decomposition method is used such as Benders and Lagrangian schemes (see Birge and Louveaux, 1997), but these methods may converge slowly in practice (see Chun and S.M. Robinson, 1995). In contrast, the current paper first considers the decomposition of the original stochastic problem using the law of iterated expectations, and then, each subproblem is solved either using value-iteration or policy-iteration algorithms. It must be noted that this approach represents a general and versatile tool, as it describes how marketing policies evolve over an infinite number of time periods, and the expected present value of those decisions.

2.2 Solving the SDP using a Bellman-decomposition algorithm

In this section we present the decomposition approach to address large CRM problems. To attain this goal, we first assume,

Condition 2 *There is a random vector* $\overline{S}_t := h(S_t)$ *where* $h(\cdot)$ *is a measurable function from the state space to another Euclidean space of low dimension, such that the expected effect of* S_t *on* $r(S_t, A_t, P_t, e_t)$ *can be summarized in the index* $\overline{S}_t$*, i.e.*

$$E_0\left[r\left(S_t, A_t, P_t, e_t\right)|S_t, A_t, P_t, e_t\right] = E_0\left[r\left(S_t, A_t, P_t, e_t\right)|\overline{S}_t, A_t, P_t, e_t\right], \quad a.e. \qquad (2.3)$$

A relevant example in which this condition is satisfied, is the decision problems in which managers' objectives are given by:

$$\begin{aligned} r\left(S_t, A_t, P_t, e_t\right) &:= \left(P_t - c_0\right) I\overline{S}_t - \sum_{i\in\mathbb{I}} c_i\left(e_{it}\right) - c_m\left(A_t\right), \\ r_i\left(S_{it}, A_t, P_t, e_{it}\right) &:= \left(P_t - c_0\right)\cdot S_{it} - c_i\left(e_{it}\right) - c_m\left(A_t\right)/I, \end{aligned} \tag{2.4}$$

i.e., they optimize the value drawn from a measurement of total sales $\sum_{i\in\mathbb{I}} S_{it} = I\overline{S}_t$, where $(P_t - c_0)$ is the unit margin, $c_m(\cdot) \geq 0$ is the cost of mass advertising interventions (when A_t are monetary units, c_m is the identity function) and $c_i(\cdot) \geq 0$ is the cost of the direct marketing interventions on customer i.

Next, we discuss the transition of the index $\overline{S}_t = h(S_t)$, given by

$$F\left(\overline{s}'|\overline{s}, A, P, e\right) = \Pr\left(\overline{S}_t \leq \overline{s}'|\overline{S}_t = \overline{s}, A, P, e\right) = \int_{\{h(s)\leq\overline{s}'\}} F\left(ds'|\overline{s}, A, P, e\right),$$

where $\overline{s} = h(s)$ and $F\left(s'|\overline{s}, A, P, e\right) = E\left[F\left(s'|S_t, A, P, e\right)|h\left(S_t\right) = \overline{s}, A, P, e\right]$. In practice, the computation of $F\left(\overline{s}'|\overline{s}, A, P, e\right)$ may require the use of numerical methods, but the analysis is particularly simple when we consider dynamic panels as described in (2.1), and $\overline{S}_t = I^{-1}\sum_{i\in\mathbb{I}} S_{it}$ as in (2.4), using that

$$\overline{S}_t = \rho\overline{S}_{t-1} + \overline{g}\left(A_{t-1}, P_{t-1}, e_{t-1}\right) + \overline{\varepsilon}_t,$$

where $\overline{\varepsilon}_t = I^{-1}\sum_{i\in\mathbb{I}} \varepsilon_{it}$ has probability distribution $G_I(\overline{\epsilon}) = G^*(\overline{\epsilon}/I)$ with $G^* = G_1 * ..*G_I$ the convolution of individual shocks' distributions, and $\overline{g}(A, P, e) = \sum_{i\in\mathbb{I}} g_i(A, P, e_i)/I$, so that

$$F\left(\overline{s}'|\overline{s}, A, P, e\right) = G_I\left(\overline{s}' - \rho\overline{s} - \overline{g}\left(A, P, e\right)\right).$$

Finally we assume that admissible prices, mass and direct marketing interventions are bounded by a maximum level which can be adapted to the previous state of sales.

Condition 3 *The non-empty compact set* $\mathbb{A}(S) \subset \mathcal{A}$ *is defined for all* S *as*

$$\begin{aligned}\mathbb{A}(S) := \quad & \Big\{ (A, P, e) \in \mathcal{A} : A^l_{(\overline{S})} \leq A \leq A^u_{(\overline{S})}, \\ & P^l_{(\overline{S})} \leq P \leq P^u_{(\overline{S})},\ e^l_i (S_i) \leq e_i \leq e^u_i (S_i) \Big\},\end{aligned}$$

S_i *is the* i*-th coordinate of* S, *and where* $\overline{S} = h(S)$ *and* $0 \leq A^l \leq A^u$, $0 \leq P^l \leq P^u$, $0 \leq e^l_i \leq e^u_i$ *are bounded continuous functions in* $\mathcal{S}$.

Let us define the subproblems:

$$\begin{aligned}V_i(s_i) \quad &: = \max_{\{e_{it}\}} E_0 \left[\sum_{t \geq 0} \delta^t R_i(S_{it}, e_{it}) \right], \qquad \text{for all } i \in \mathbb{I}, \\ \overline{V}(\overline{s}) \quad &= \max_{\{A_t, P_t\}} E_0 \left[\sum_{t \geq 0} \delta^t R\left(\overline{S}_t, A_t, P_t\right) \right],\end{aligned}$$

where $R_i(S_{it}, e_{it})$ and $R_i(S_{it}, e_{it})$ are conditional expectations

$$\begin{aligned}& R_i(S_{it}, e_{it}) = I \cdot E\left[r_i\left(S_{it}, A^*_t, P^*_t, e_{it}\right) | S_{it}, e_{it} \right], \\ & R\left(\overline{S}_t, A_t, P_t\right) := E\left[r\left(S_t, A_t, P_t, e^*_t\right) | \overline{S}_t, A_t, P_t \right],\end{aligned} \tag{2.5}$$

with A^*_t, P^*_t, e^*_t the optimal decisions for time t.

Notice that any policy function (A, P, e), by the *Law of Iterated Expectations* is satisfied that

$$\begin{aligned}E_0 \left[\sum_{t \geq 0} \sum_{i \in \mathbb{I}} \delta^t r_i(S_{it}, A_t, P_t, e_{it}) \right] &= \sum_{i \in \mathbb{I}} E_0 \left[\sum_{t \geq 0} \delta^t E\left[r_i(S_{it}, A_t, P_t, e_{it}) | S_{it}, e_{it} \right] \right] \\ &= E_0 \left[\sum_{t \geq 0} \delta^t E\left[\sum_{i \in \mathbb{I}} r_i(S_{it}, A_t, P_t, e_{it}) | \overline{S}_t, A_t, P_t \right] \right],\end{aligned}$$

where $A_t = A(S_{t-1}), P_t = P(S_{t-1}), e_t = e(S_{t-1})$; which under conditions (2) and (3) imply that $V(s) = I^{-1} \sum_{i \in \mathbb{I}} V_i(s_i)$ and also that $V(s) = \overline{V}(\overline{s})$ almost everywhere.

Therefore, the subproblems $\{V_i(s_i)\}_{i\in\mathbb{I}}$ and $\overline{V}(\overline{s})$ characterize the value function $V(\cdot)$, the subproblems are, by definition, smaller than the original problem (Problem 1) and therefore much faster to solve. In order to solve the subproblems separately, we need the transition kernel for $\{V_i(s_i) : i \in \mathbb{I}\}$ and $\overline{V}(\overline{s})$ respectively given by

$$\begin{aligned}
\mathbb{F}_i(s_i'|s_i, e_i) &= E\left[F_i\left(s_i|S_{it-1}, A_{t-1}^*, P_{t-1}^*, e_{it-1}\right)|S_{it-1} = s_i, e_{it-1} = e_i\right], \text{for all } i \in \mathbb{I}, \\
\mathbb{F}(\overline{s}'|\overline{s}, A, P) &= E\left[F\left(\overline{s}'|\overline{S}_{t-1}, A_{t-1}, P_{t-1}, e_{t-1}^*\right)|\overline{S}_{t-1} = \overline{s}, A_{t-1} = A, P_{t-1} = P\right].
\end{aligned}$$

and we need also to know $R_i(S_{it}, e_{it})$ and $R\left(\overline{S}_t, A_t, P_t\right)$. The computation of the required conditional probabilities and expectations is unfeasible since the optimal policy function (A^*, P^*, e^*) is unknown.

2.2.1 The algorithm

The general scheme of the algorithm is stated as follows.

`ALGORITHM`

1. `Initialization: Choose a scenario set of states and a starting policy` $\left\{A^k(\overline{s}), P^k(\overline{s}), e^k(s)\right\}$ `with` $e^k(s) = \left(e_1^k(s_1), ..., e_I^k(s_I)\right)$. `Set` $k = 0$.

2. `Repeat:`

   2.1 `Generate recursively` $\left\{S_t^k, A_t^k, P_t^k, e_t^k\right\}_{t=1}^T$ `where` S_t^k `is drawn from`

$$F\left(s|S_{t-1}^k, A^k\left(\overline{S}_{t-1}^k\right), P^k\left(\overline{S}_{t-1}^k\right), e^k\left(S_{t-1}^k\right)\right),$$

   `and compute` $\overline{S}_t^k = h\left(S_t^k\right)$,

   2.2. `With the simulated data compute`

$$\begin{aligned}
R_i^k(S_{it}, e_{it}) &= I \cdot E\left[r_i\left(S_{it}, P_t^k, A_t^k, e_{it}\right)|S_{it}, e_{it}\right], \\
R^k\left(\overline{S}_t, A_t, P_t\right) &= E\left[r\left(S_t, P_t, A_t, e_i^k\right)|\overline{S}_t, A_t, P_t\right].
\end{aligned}$$

and the kernels

$$\begin{aligned}\mathbb{F}^k\left(s_i'|s_i,e_i\right) &= \Pr\left(S_{it}^k\leq s_i'|S_{it-1}^k=s_i,e_{it-1}=e_i\right), \quad i\in\mathbb{I},\\ \mathbb{F}^k\left(\overline{s}'|\overline{s},A,P\right) &= \Pr\left(\overline{S}_t^k\leq \overline{s}'|\overline{S}_{t-1}^k=\overline{s},A_{t-1}=A,P_{t-1}=P\right).\end{aligned}$$

2.3 Solve the SDP subproblems

$$\max_{\{e_{it}\in\mathbb{A}_i(S_{it-1})\}_{t>0}} E\left[\sum_{t\geq 0}\delta^t R_i^k\left(S_{it},e_{it}\right)|S_{i0}=s_i\right]:=V_i^k\left(s_i\right),$$

in $\left\{e_{it}^k\right\}_{t>0}$ for each $i\in\mathbb{I}$, where $\mathbb{A}_i\left(S_{it-1}\right)=\left\{e_i:0\leq e_i\leq\overline{e_i}\left(S_{it-1}\right)\right\}.$

2.4 Solve the SDP subproblem

$$\max_{\{A_t,P_t\}\in\overline{\mathbb{A}}\left(\overline{S}_{t-1}\right)} E\left[\sum_{t\geq 0}\delta^t R^k\left(\overline{S}_t,A_t,P_t\right)|\overline{S}_0=\overline{s}\right]:=\overline{V}^k\left(\overline{s}\right),$$

where

$$\overline{\mathbb{A}}\left(\overline{S}_{t-1}\right)=\left\{(p,A):0\leq A\leq\overline{A}_{\left(\overline{S}_{t-1}\right)},\ 0\leq P\leq\overline{P}_{\left(\overline{S}_{t-1}\right)}\right\}.$$

2.5 Update $\left\{e_{it}^k,A_t^k,P_t^k\right\}$ to $\left\{e_{it}^{k+1},A_t^{k+1},P_t^{k+1}\right\}$, and set $k\longleftarrow k+1.$

3. Until convergence: for some tolerance $\epsilon>0,$ when the stopping criteria are satisfied

- Criterion 1:
$\max\left\{\sup_t\frac{\left|A_t^{k+1}-A_t^k\right|}{1+\left\|A_t^k\right\|_\infty},\ \sup_t\frac{\left|P_t^{k+1}-P_t^k\right|}{1+\left\|P_t^k\right\|_\infty},\ \sup_{t,i}\frac{\left|e_{it}^{k+1}-e_{it}^k\right|}{1+\left\|e_{it}^k\right\|_\infty}\right\}<\epsilon,$
- Criterion 2:
$\sup_{S_0}\left\{\frac{\left|I^{-1}\sum_{i\in\mathbb{I}}V_i^{k+1}(S_{i0})-\overline{V}^{k+1}\left(\overline{S}_0\right)\right|}{1+\left\|I^{-1}\sum_{i\in\mathbb{I}}V_i^{k+1}(S_{i0})\right\|_\infty}:\overline{S}_0=I^{-1}\sum_{i\in\mathbb{I}}S_{i0}\right\}<\epsilon,$

where the superscript ^{k} denotes the current iteration and $\|\cdot\|_\infty$ is the supremum norm.

The algorithm iterates the solution of both types of subproblems. For one set of subproblems, the decision variables are only the direct marketing intervention $\{e_{it}\}_{t>0}$. Once the solutions for these subproblems have been computed, price and mass marketing intervention $\{P_t, A_t\}_{t>0}$ are updated. An economic interpretation of the decomposition draws on this partition of the decision variables into individual and general decisions taken among customers. The convergence of the algorithm is discussed in Appendix B.

Any classical method to solve SDP such as value iteration or policy iteration can be applied in steps 2.3 and 2.4, since the subproblems are small problems with just one state variable, using the optimal policy computed in the previous iteration of the algorithm as initial point. The specific details are described in Appendix C.

Note that the value function for the original problem $V(S_1, ..., S_I)$ and the associated policy functions $[A, P, e](S_1, ..., S_I)$ cannot be graphically represented for more than two customers due to the dimension. However, graphical figures for these functions would be intuitive user-friendly tools for marketing managers. Interestingly, our algorithm overcomes this problem providing useful and visual tools for managers implementing CRM. After convergence of the algorithm at step k^* to a numerical solution of the original problem, we can depict graphically in the plane the reduced value function $V^{k^*}\left(\overline{S}\right)$ and the associated reduced optimal policy functions $A^{k^*}\left(\overline{S}\right), P^{k^*}\left(\overline{S}\right)$ to provide graphical rules for planning optimally mass advertising and price (provided that the optimal individual e is implemented). Furthermore, we can depict in the plane the reduced value function $V_i^{k^*}(S_i)$ and the associated reduced optimal policy function $e_i^{k^*}(S_i)$ for the $i-th$ customer, which provide a graphical rule for planning optimally the marketing effort on i-individual (provided that the optimal mass advertising and price have been implemented as well as the effort on other individuals).

2.3 Some numerical simulations

Let us consider a dynamic-regression model where sales follow a dynamic panel model

$$S_{it} = \rho S_{it-1} + \alpha_{1i} \ + \beta_{1i}\ e_{it-1} + \beta_{2i}\ A_{t-1} + \beta_{3i}\ P_{t-1}^{\beta_{4i}} + \varepsilon_{it}$$

with $\{\beta_{1i}, \beta_{2i}\} > 0$, and $|\rho| < 1$, where $\{e_{it}\}_{t\geq 1}$ denotes individual marketing efforts, $\{A_t\}_{t\geq 1}$ is the mass marketing effort, $\{P_t\}_{t\geq 1}$ is the price, and $\{\varepsilon_{it}\}_{t\geq 1}$ are independent white noise processes $N(0, \sigma I)$. We assume that $\{e_{it}\}_{t\geq 1}$ and $\{A_t\}_{t\geq 1}$ are given by a cost function $c(x) = \gamma x^{\phi}$, with $\gamma > 0$. Then, given $\delta \in (0,1)$, the firm aims to maximize the expected net present value $E_0 \left[\sum_{t\geq 0} \delta^t r(S_t, A_t, P_t, e_t)\right]$ with

$$r(S_t, A_t, P_t, e_t) := (P_t - c_0) \sum_{i\in\mathbb{I}} S_{it} - c(A_t) - \sum_{i\in\mathbb{I}} c(e_{it}).$$

We have implemented our decomposition algorithm using MATLAB 7.6 on an Intel Core vPro i7 with machine precision 10^{-16}. The algorithm stops whenever $\epsilon = 10^{-8}$.

First, we consider a simplified model in which prices are considered as given, i.e. $\beta_{3i} = 0$ and using a constant exogenous margin m_0 instead of $(P_t - c_0)$. For $m_0 = 50$, $\rho = 0.2$, $\alpha_i = 60$, $\beta_{1i} = 1.2$, $\beta_{2i} = 1.2$, $\sigma = 5$, Table 1 reports the running times (in seconds) until convergence considering different number of customers I, and both policy iteration and value iteration algorithms to solve Steps 2.3 and 2.4 of the algorithm.

Table 1. Properties of the algorithm for different problem sizes in a model without prices

Method	Number of Customers	Stopping Criteria		Number of Iterations	Computational Time (in seconds)
		Criterion 1	Criterion 2		
Policy Iteration	1	0.0000	0.0000	3	3.8922
	5	0.0000	0.0007	4	11.0790
	25	0.0000	0.0008	4	60.5070
	50	0.0000	0.0009	4	166.9300
	100	0.0000	0.0009	4	687.4300
Value Iteration	1	0.0000	0.0000	4	3.5335
	5	0.0000	0.0007	3	8.6160
	25	0.0000	0.0008	4	61.1350
	50	0.0000	0.0009	4	169.0700
	100	0.0000	0.0009	3	545.9600

Then, we extend the basic model to the general case in which prices are considered as a decision variable. For $c_0 = 50$, $\rho = 0.2$, $\alpha_i = 60$, $\beta_{1i} = 1.2$, $\beta_{2i} = 1.2$, $\beta_{4i} = -0.5$, $\beta_{3i} = 0.5$, $\sigma = 5$, Table 2 reports the running times (in seconds) until convergence considering different number of customers I. The results show that the proposed algorithm is capable of solving the problem with many customers in a reasonable amount of computational time.

Table 2.Properties of the algorithm for different problem sizes in a model with prices

Method	Number of Customers	Stopping Criteria		Number of Iterations	Computational Time (in seconds)
		Criterion 1	Criterion 2		
Policy Iteration	1	0.0000	0.0527	3	5.9819
	5	0.0000	0.0202	4	29.5548
	25	0.0000	0.0202	4	150.4374
	50	0.0000	0.0202	4	404.8369
	100	0.0000	0.0202	2	873.8870
Value Iteration	1	0.0000	0.0115	6	11.2360
	5	0.0000	0.0202	2	16.4847
	25	0.0000	0.0202	2	83.7036
	50	0.0000	0.0324	2	189.4250
	100	0.0000	0.0202	2	663.1727

These results suggest that the proposed methodology is an effective and useful tool for solving this type of problems as it breaks down a high-dimensional problem into many low-dimensional ones, hence reducing the curse of dimensionality. It is remarkable that the standard policy iteration approach cannot solve a problem of more than 3 customers.

2.4 An empirical application of a manufacturer of kitchen appliances

In this section we provide an application of the method. We consider a medium size international wholesale company based in eastern Europe. This company distributes and also manufactures a large range of built-in kitchen appliances for kitchens (such as cookers, ovens and hobs, cooker and chimney hoods, external motors, microwaves, dishwashers, washing machines, refrigerators, and related accessories). The company invests in general marketing effort (mainly advertising and promotions in professional fairs) and personalized investments in their customer relationships management. We do not provide additional company information by confidentiality requests of the company managers.

We use a monthly customer-panel from this company spanning from January 2005 to December 2008. The panel is unbalance, although the vast majority of the clients purchases practically every month within the sample period. As the company sells a wide range of products with different sales to each client, they aggregated their data providing us with the monthly net-profit drawn from each client. Therefore, in this section $Y_{i,t}$ is regarded as the financial value obtained from client i at time t, the individual marketing effort on this customer is denoted by $e_{i,t}$, and the general marketing effort is A_t. The basic Markovian model is a dynamic-panel specification

$$Y_{i,t} = \rho Y_{i,t-1} + \beta_1 \ln \ A_{t-1} + \beta_2 \ln e_{i,t-1} + (\eta_i + u_{it}),$$

$$E[u_{i,t} X_{i,t}] = 0, \ E[u_{i,t}] = 0, \text{for all } i,t,$$

where $|\rho| < 1$, $u_{i,t}$ is white noise and η_i is a zero mean random coefficient accounting for individual heterogeneity in customer profitability levels. The noise $v_{it} = \eta_i + u_{i,t}$ is autocorrelated due to the stability of η_i, and therefore the OLS and the Within-Group estimators are both inconsistent (as $Y_{i,t-1}$ is a regressor). Taking first differences in the model, we eliminate the specific group effects

$$\triangle Y_{i,t} = \rho \triangle Y_{i,t-1} + \triangle X'_{i,t-1}\beta + \triangle u_{i,t}, \qquad t = 2, ..., T,$$

where $X'_{i,t-1} = (\ln \ A_{t-1}, \ln e_{i,t-1})'$. The errors $\{\triangle u_{it}\}$ are no longer independent but follow a non invertible $MA(1)$. This equation can be estimated by Instrumental Variables (IV), as proposed by Andersen and Hsiao (1982). It is convenient to use lags of the variable in levels $Y_{i,t-1}$ as instrument, as well as lags of other exogenous regressors. Nonetheless, the IV estimator is not efficient due to the fact that only a few moment conditions are used.

Arellano and Bond (1991) proposed a GMM estimator dealing with this problem. The Arellano and Bond (1991) estimators can perform poorly in certain cases, and the method was refined by Blundell and Bond (1998) who included additional moment conditions (building on previous work by Arellano and Bover, 1995). The model was estimated in STATA using the Blundell-Bond refinement. Table 2 reports the estimators of this model. The Wald global significance test is 169.73 distributed as a χ^2_3 with a p-value 0.0000.

Table 3. Main coefficients in the dynamic-panel model for customer profitability.

Y_{t-1}	Coef.	Std. Err	z	P>|z|
$Y_{i,t-1}$	.024	0.011	2.15	0.031
A_{t-1}	821.52	235.244	3.49	0.000
$e_{i,t-1}$	1175.05	172.395	6.82	0.000

In order to improve the heterogeneity analysis, we have decided to include additional information, classifying clients by continental location (4 large regions with dummies $\{D_{ki}\}_{k=1}^{4}$), and a customers' strategic classification by the company (3 levels with dummies $\{d_{ji}\}_{j=1}^{3}$), so that we have 12 basic segments. Therefore, we introduce heterogeneity in the response to marketing effort as

$$\begin{aligned} Y_{i,t} &= \rho Y_{i,t-1} + \beta_1 \ln\ A_{t-1} + \beta_2 \ln e_{i,t-1} + \\ &+ \sum\nolimits_{j=1}^{3} \gamma_j \left(d_{ji} \times \ln A_{t-1}\right) + \sum\nolimits_{j=1}^{3} \gamma_j' \left(d_{ji} \times \ln e_{i,t-1}\right) \\ &+ \sum\nolimits_{k=1}^{4} \alpha_k \left(D_{ki} \times \ln A_{t-1}\right) + \sum\nolimits_{j=1}^{4} \alpha_k' \left(D_{ki} \times \ln e_{i,t-1}\right) + \left(\eta_i + u_{it}\right). \end{aligned}$$

To ensure identification, we impose that the dummy coefficients sum up to zero by classification factors. Substituting these parametric constraints in the mode, we obtain that

$$
\begin{aligned}
Y_{i,t} \;\; = \;\; & \rho Y_{i,t-1} + \beta_1 \ln \; A_{t-1} + \beta_2 \ln e_{i,t-1} + \\
& + \sum\nolimits_{j=1}^{2} \gamma_j \left(d_{ji} - d_{3i}\right) \ln A_{t-1} + \sum\nolimits_{j=1}^{2} \gamma_j' \left(d_{ji} - d_{3i}\right) \ln e_{i,t-1} \\
& + \sum\nolimits_{k=1}^{3} \alpha_k \left(D_{ki} - D_{4i}\right) \ln A_{t-1} + \sum\nolimits_{j=1}^{3} \alpha_k' \left(D_{ki} - D_{4i}\right) \ln e_{i,t-1} + \left(\eta_i + u_{it}\right)
\end{aligned}
$$

with $\gamma_3 = -\sum_{j=1}^{2} \gamma_j$, $\gamma_3' = -\sum_{j=1}^{2} \gamma_j'$, $\alpha_4 = -\sum_{k=1}^{3} \alpha_k$, and $\alpha_4' = -\sum_{k=1}^{3} \alpha_k'$. The final model was estimated in STATA using the Blundell-Bond refinement. We used $6,728$ observations with 260 customers, and $1.1e{+}03$ instruments. The Wald global significance test is 195.43 distributed as a χ^2_{11} with a p-value 0.0000. The individual marketing effort has a significant impact, as well as the general advertising. The dummy coefficients $\left\{\gamma_j'\right\}_{j=1}^{2}$ are non significant, and set them equal to zero in the optimization part. All the other types of dummy coefficients are significant. After the model coefficients have been estimated, since T is large, we can consistently estimate each specific intercept η_i. For each customer we need to take time-means on the panel regression equations, then replace $\sum_{t=1}^{T} u_{it}/T$ by zero (the expected value), and finally getting the estimator of η_i.

Next, consider a SDP problem for the returns function

$$
r\left(Y_t, A_t, e_t\right) = \sum_{i=1}^{I} Y_{it} - A_t - \sum_{i=1}^{I} e_{it},
$$

where the state variable $\{Y_{it}\}$ are returns drawn from the i-th customer. The transition equations for all customers in one of the identified segments are identical, but there are relevant different across segments.

We have computed the optimal general advertising and marketing effort policies for a stylized version of the model with 12 representative customers, applying the proposed decomposition method. The collocation algorithm was run using a state discretization with 10 scenarios (sales levels, disguised by company request) for each individual-sales variable and 20 equidistant knots for each control, applying policy iteration for each subproblem. It takes 7 iterations (about 11 minutes) of the full decomposition method for the algorithm to converge. Figures 1 and 2 show $\{V_i\,(s_i)\}_{i=1}^{12}$ and $\overline{V}\,(\overline{s})$, the individual and mean reduced value functions respectively.

Figure 1. Individual reduced value functions

(Customer value associated with its sales state)

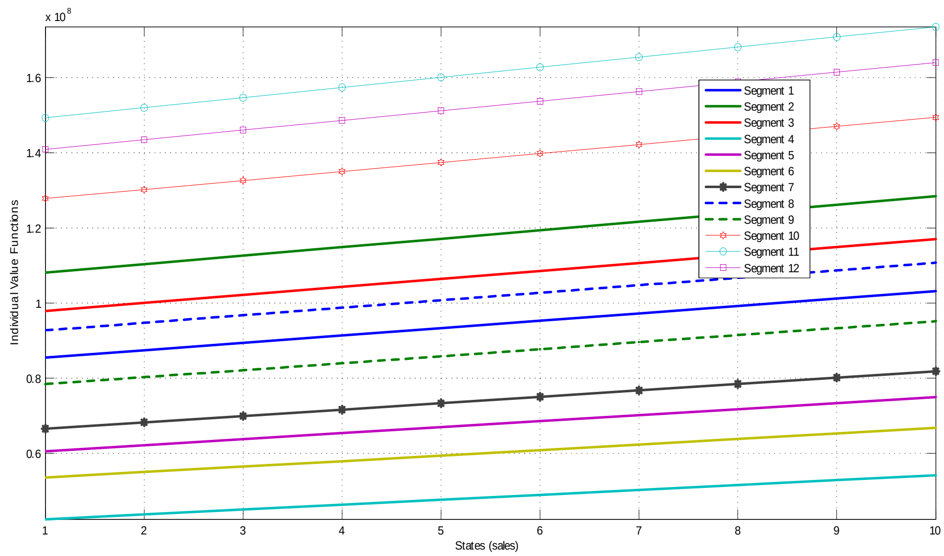
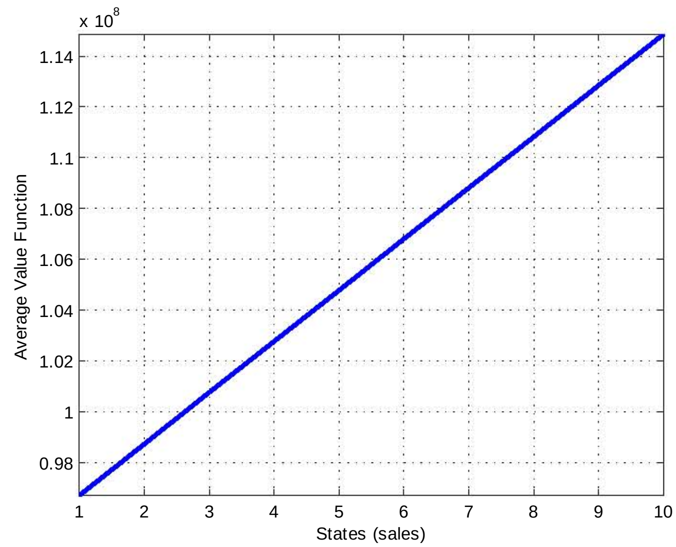


Figure 2. Mean reduced value function

(Total value ssociated with mean sales states)



Figures 3 and 4 show $\{e_i(s_i)\}_{i=1}^{12}$ and $A(\overline{s})$, the optimal individual and general marketing effort reduced policy functions respectively.

Figure 3. Individual marketing effort reduced policy functions

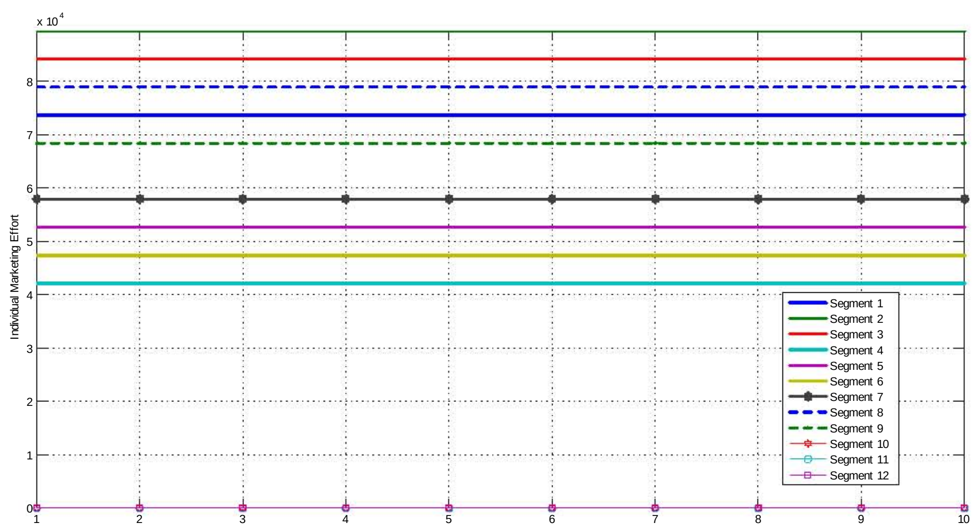
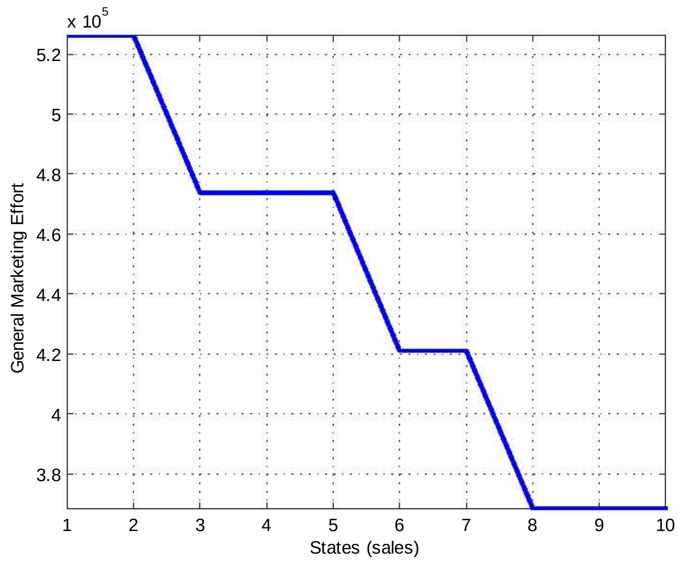


Figure 4. General marketing effort reduced policy functions



These results show that the optimal budget allocated to mass marketing is decreasing with respect to total sales. By contrast, the individual effort is hold constant with sales but, the level is different for each segment. In particular for the 3 segments which have a negligible individual marketing coefficient, the optimal solution prescribes not to invest at all on them. Furthermore, notice that the ranking of individual effort investments by segments does not follow exactly the pattern given by individual reduced value functions. This is not a surprising result as the optimal solution takes into account not just differences in profitability but also different sensibilities of the segments to the marketing mix.

2.5 Conclusions

There is a growing interest for firms to customize their marketing activities to smaller and smaller units —individual stores, customers and transactions" (Buckling et al., 1998), implying an enormous number of decisions. This scale requires Decision Automation tools based on dynamic optimization of small unit panels.

In this paper, we make a computational contribution for solving SDP problems, which allows forward-looking firms to allocate the marketing budget optimizing the CLV of their customer base, simultaneously using customized and mass marketing interventions. The solvability of these models suffers from the curse of dimensionality, which limits practitioners from the modelling standpoint. In this sense, we have introduced a novelty decomposition methodology for the computation of solutions of CRM problems. The proposed approach deflates the dimensionality of the models by breaking the problem into a set of smaller independent subproblems. The numerical results have revealed the efficiency of the methodology in terms of computing time and accuracy, concluding that the proposed approach is promising for application in many marketing problems with similar structure.

We have shown the decomposition method works very well in practice. The methodology has been successfully applied to value more than 260 customers of a medium size international wholesale company. We have presented a customer profitability analysis of the company considering the effect of direct marketing and mass marketing interventions at the customer level, simultaneously.

Since often CRM databases do not involve panel data across several competitors, no competitive effects have been considered in this article. To include competition, we should consider a behavioral model for several firms competing for the same customers with mass and customized marketing actions, and the equilibrium would be given by the Markov perfect equilibrium (see Dubé et al. 2005). The computational effort to solve this problem is formidable, and the decomposition algorithm presented in this article could be a useful tool to address it. We leave this problem for future research.

2.6 References

Anderson, T.W. and Cheng Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47-82.

Arellano, Manuel and Stephen Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277-297.

Arellano, Manuel and Olympia Bover (1995), "Another Look at the Instrumental Variable Estimation of Error-components Models," *Journal of Econometrics*, 68, 29-51.

Bawa, Kapil and Robert W. Shoemaker (1987), "The Effects of a Direct Mail Coupon on Brand Choice Behavior, " *Journal of Marketing Research*, 24(4) 370–376.

Bellman, Richard (1955), "Functional Equations in the Theory of Dynamic Programming: Positivity and Quasilinearity, " *Proceedings of the National Academy of Sciences*, 41, 743–746.

Bellman, Richard (1957), *Dynamic Programming*, Princeton University Press, Princeton, N.J.

Bellman, Richard (1961), *Adaptive Control Processes: A Guided Tour*, Princeton University Press.

Berry, Leonard L. (1995), "Relationship Marketing in Services: Growing Interest, Emerging perspectives," *Journal of the Academy of Marketing Science*, 23(4), 236–245.

Bhattacharya, C. B. and Ruth N. Bolton (1999), *Relationship Marketing in Mass Markets*, in Jagdish N. Sheth, Atul Parvatiyar, eds. Handbook of Relationship Marketing. Sage Publications, Thousand Oaks, CA, 327–354.

Birge, John R. and Francois Louveaux (1997), *Introduction to Stochastic Programming*, Springer, New York.

Bitran, Gabriel R. and Susana V. Mondschein (1996), "Mailing Decisions in the Catalog Sales Industry," *Management Science*, 42(9), 1364-1381.

Blackwell, David (1965), "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 36 (1), 226-235.

Blundell, Richard and Stephen Bond (1998), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.

Bose, Idranil and Xi Chen (2009), "Quantitative Models for Direct Marketing: A Review from Systems Perspective," *European Journal of Operational Research*, 195(1), 1-16.

Bucklin, Randolph E., Donald R. Lehmann, and Little, J.D.C. (1998), "From Decision Support to Decision Automation: A 2020 Vision," *Marketing Letters*, 9(3), 235–246.

Chen, Xiaohong, Oliver Linton and Peter M. Robinson (2001) “The Estimation of Conditional Densities” by *Asymptotic in Statistics and Probability*, Festschrift for George Roussas (M.L. Puri, ed.), VSP International Science Publishers, the Netherlands, 71-84.

Chun, Bock Jin and Stephen M. Robinson (1995) "Scenario Analysis via Bundle Decomposition", *Annals of Operations Research*, 56, 39-63.

Denardo, Eric V. (1967), "Contraction Mappings in the Theory Underlying Dynamic Programming," *SIAM Review*, 9, 165-177.

Dube, Jean-Pierre, K. Sudhir, Andrew Ching, Gregory S. Crawford, Michaela Draganska, Jeremy T. Fox, Wesley Hartmann, Gunter J. Hitsch, V. Biran Viard, Miguel Villas-Boas and

Gonul, Fusun and Meng Ze Shi (1998), "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science,* 44(9), 1249-1262.

Gonul, Fusun and Frenkel Ter Hofstede (2006), "How to Compute Optimal Catalog Mailing Decisions," *Marketing Science*, 25 (1), 65-74.

Gupta, Sunil and Donald R. Lehmann (2003), "Customers as Assets," *Journal of Interactive Marketing*, 17, 1(Winter), 9-24.

Gupta, Sunil, Donald R. Lehmann and Jennifer A. Stuart (2004), "Valuing Customers," *Journal of Marketing Research*, 41(1), 7-18.

Gupta, Sunil and Donald R. Lehmann (2005), *Managing Customers as Investments: The Strategic Value of Customers in the Long Run*, Wharton School Publishing, Upper Saddle River, New Jersey.

Gupta, Sunil and Donald R. Lehmann (2006), "Customer Lifetime Value and Firm Valuation," *Journal of Relationship Marketing*, 5, nos, 2/3 (October), 87-110.

Gupta, Sunil and Valarie Zeithaml (2006), "Customer Metrics and Their Impact on Financial Performance," *Marketing Science*, 25 (6), (November-December), 718-739.

Gupta, Sunil and Thomas J. Steenburgh (2008), "Allocating Marketing Resources," Working Paper, Harvard University, http://hbswk.hbs.edu/item/5868.html.

Gupta, Sunil, Carl F. Mela and Jose M. Vidal-Sanz (2009), "The Value of a "Free" Customer," *University Carlos III de Madrid Working Paper*, Business Economics Series, No. 09-29-03. Available at http://hdl.handle.net/10016/3883 (an 2006 version is available as *Harvard Business School Working Paper*, No. 07-035).

Heitsch, Holger and Werner Romisch (2009) "Scenario Tree Modeling for Multistage Stochastic Programs," *Mathematical Programming*, 118(2), 371- 406.

Howard, Ronald A. (1960), *Dynamic Programming and Markov Processes*, The MIT Press.

Judd, Kenneth (1998), *Numerical Methods in Economics*, MIT Press.

Kamakura, Wagner, Carl F. Mela, Asim Ansari, Anand Bodapati, Pete Fader, Raghuram Iyengar, Prasad Naik, Scott Neslin, Baohong Sun, Peter Verhoef, Michel Wedel and Ron Wilcox, (2005), "Choice Models and Customer Relationship Management," *Marketing Letters*, 16, 3/4, 279-291.

Khan, Romana, Michael Lewis and Vishal Singh (2009), "Dynamic Customer Management and the Value of One-to-One Marketing," *Marketing Science*, 28 (6), 1063–1079.

Lewis, Michael (2005), "A Dynamic Programming Approach to Customer Relationship Pricing," *Management Science*, 51(6), 981-994.

Mantrala, Murali (2002). "Allocating Marketing Resources," In: Barton A. Weitz, Robin Wensley, eds. Handbook of Marketing. Sage Publications, Thousand Oaks, CA, 409-435.

McDonald, William J. (1998). *Direct Marketing: An Integrated Approach.* Irwin-McGraw-Hill, Boston, MA.

Montoya, Ricardo, Oded Netzer and Kamel Jedidi (2007), "Dynamic Marketing Mix Allocation for Long-Term Profitability," *Working Paper*, Columbia University.

Naufel, Vilcassim (2005), "Recent Advances in Structural Econometric. Modeling: Dynamics, Product Positioning and Entry," *Marketing Letters*, 16, (3-4), 209-224.

Reinartz, Werner, Jacquelyn S. Thomas and V. Kumar (2005), "Balancing Acquisition and Retention Resources to Maximize Customer Profitability," *Journal of Marketing*, 69, 1 (January), 63-79.

Rincon-Zapatero, Juan Pablo and Carlos Rodríguez-Palmero (2003), "Existence and Uniqueness of Solutions to the Bellman Equation in the Unbounded Case," *Econometrica*, 71, 1519-1555. There is a corrigendum in *Econometrica*, 2009, vol. 77, issue 1, pages 317-318.

Roberts, Mary Lou and Paul D. Berger (1989), *Direct Marketing Management*, Prentice-Hall, Upper Saddle River, NJ.

Rossi, Peter E., Robert E. McCulloch and Greg M. Allenby (1996), " The Value of Purchase History Data in Target Marketing," *Marketing Science*, 15(4) 321–340.

Rust, Roland T., Katherine N. Lemon and Valarie Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68, 1(January), 109-126.

Rust, Roland T. and Peter C. Verhoef (2005), "Optimizing the Marketing Interventions Mix in Intermediate-Term CRM," *Marketing Science*, 24(3), 477–489.

Rust, Roland. T. and Tuck Siong Chung (2006), "Marketing Models of Service and Relatioships, " *Marketing Science*, 25(6), 560-580.

Roussas, George (1967), Nonparametric Estimation in Markov Processes. *Ann. Inst. Statist. Math.*, 21, 73-87.

Roussas, George (1969). Nonparametric Estimation of the Transition Distribution Function of a Markov Process. *Ann. Math. Stat.*, 40, 1386-1400.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science* 13(1), 40–67.

Shankar, Venkatesh (2008), "Strategic Allocation of Marketing Resources: Methods and Managerial Insights," *MSI Working Paper*, http://ssrn.com/abstract=1270804.

Simester, Duncan I., Peng Sun and John N. Tsitsiklis (2006). "Dynamic Catalog Mailing Policies," *Management Science*, 52 (5), 683–696.

Venkatesan, Rajkumar, and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68, 4(November), 106-125.

2.7 Appendix A: Value iteration and Policy iteration for continuous problems

Continuous SDP problems are usually solved combining the ideas of value iteration and policy iteration with collocation methods. The basic idea of Collocation methods is to consider a sequence of functions $\{\phi_k\}_{k\geq 1} \subset B_\infty$ such that any function $v \in B_\infty$ can be expressed asymptotically as a linear combination of these functions, or more formally for all $v \in B_\infty$

$$\inf_{\{\theta_k\}_{k=1}^K} \left\| v(s) - \sum_{k=1}^{K} \theta_k \ \phi_k(s) \right\|_\infty \overset{K\to\infty}{\longrightarrow} 0,$$

and therefore we can express $V(s) \approx \sum_{k=1}^{K} \theta_k \ \phi_k(s)$ for some coefficients $\{\theta_k\}$ and a large enough K. Several classes of functions that can be used for the approximation (e.g., Chebyshev polynomial, splines, Neural Networks, etc.). When the state variable is multidimensional, the base functions are generally obtained by tensor products on univariate basis. The integer K is exponentially increased with the dimension to obtain a good approximation (this is one type of the *curse of dimensionality*). Notice that the continuous SDP problem can be approximated by another one with finite states (just considering a finite partition $\{\mathcal{S}_k\}$ of the Euclidean state's space $\mathcal{S}$, we can approximating v by simple functions $\sum_{k=1}^{K} \theta_k \ I(s \in \mathcal{S}_k)$, choosing a representative scenario s_k for each element of the partition and interpreting $\theta_k = v(s_k)$).

The coefficients $\{\theta_k\}_{k=1}^K$ are unknown, the collocation method approximates a functional equation in such a way that the approximated function fits exactly at the pre-specified points of the domain. Then, Bellman's Equation becomes

$$\sum_{k=1}^{K} \theta_k \ \phi_k(s) = \max_{(A,P,e)\in\mathbb{A}(s)} \left\{ r(s,A,P,e) + \delta \ \sum_{k=1}^{K} \theta_k \int \phi_k(s') \, F(ds'|s,A,P,e) \right\}. \quad (2.6)$$

Next, we evaluate the linear equation at K grid-points $\{s_1,...,s_K\} \subset \mathcal{S}$ and solve the system in $\{\theta_k\}_{k=1}^{K}$. The system (2.6) can be expressed in matrix notation as

$$\Phi\theta = \Gamma(\theta) \tag{2.7}$$

where $K \times K$ matrix Φ has element $\Phi_{mk} = \phi_k(s_m)$ and the $K \times 1$ vector $\Gamma(\theta)$ has $m-th$ element

$$\Gamma_m(\theta) = \max_{(A,P,e)\in\mathbb{A}(s_m)} \left\{ r(s_m, A, P, e) + \delta \sum_{k=1}^{K} \theta_k \int \phi_k(s') F(ds'|s_m, A, P, e) . \right\}$$

The solution of this system is not trivial, first we need to evaluate the expectations

$$\int \phi_k(s') F(ds'|s_m, A, P, e), \tag{2.8}$$

for $m = 1, ..., K$; often using a numerical integration method or a Monte Carlo approach. When the integral is replaced by an average over a finite set of sampled points, the number of required points required to have a good approach increases exponentially with the dimension of the state variables (this is another type of *curse of dimensionality*). After computing these expectations, it is generally impossible to attain closed form solution to the collocation system (2.7), and some computational algorithm is required.

- The *Value iteration* method considers the system $\theta = \Phi^{-1}\Gamma(\theta)$, and iterates the following:

  $$\theta \longleftarrow \Phi^{-1}\Gamma(\theta)$$

  from an initial point θ^0. It was initially proposed by Bellman (1955, 1957) for discrete problems.

- The *Policy iteration* method uses the Newton iterative updating,

  $$\theta \longleftarrow \theta - \left[\Phi - \Gamma'(\theta)\right]^{-1} \left[\Phi\theta - \Gamma(\theta)\right]$$

where $\Gamma'(\theta)$ is the Jacobian of the collocation function Γ at θ that can be computed *by* applying the *Envelope Theorem* to the optimization problem in the definition of $\Gamma(\theta)$, so that

$$\Gamma'_{mj}(\theta) = \delta \int \phi_j(s') F(ds'|s_m, A, P, e)$$

This method was initially proposed by Howard (1960).

Notice that when the approximation method is based on simple functions, then Φ is the identity function, and we can omit this factor. Each time that the operator $\Gamma(\theta)$ is applied we must solve the maximization problem in $\Gamma_m(\theta)$ for all states $s_m \in \{s_1, ..., s_K\}$. This can be done, e.g., using a global optimization algorithm. In many applications, the maximization is carried out discretizing the decision space $\mathbb{A}(s_m)$. Once we have converged, $V(s) = \sum_{k=1}^{K} \theta_k \ \phi_k(s)$, and the optimal policy is computed at each state $s_m \in \{s_1, ..., s_K\}$, as the maximizing decision taken at $\Gamma_m(\theta)$ for the last iteration and the function is computed interpolating these points. The main problem with the all previous techniques is the *curse of dimensionality* (Bellman, 1961). So far, researchers can solve numerically only SDP problems with very few state variables.

2.8 Appendix B: Convergence Analysis

In this section we discuss the convergence of the algorithm. We first introduce some basic notation. The convergence of classical Value Iteration method is based on central ideas from functional analysis. Define the operator

$$\Gamma(v) = \max_{(A,P,e)\in\mathbb{A}(s)} \left\{ r(s, A, P, e) + \delta \int v(s') F(ds'|s, A, P, e) \right\}$$

transforming a function of the state variables $v(s)$ into another function $\Gamma(v)(s)$. Obviously that value function is a fixed point of Γ, i.e. an element v^* such that $\Gamma(v^*) = v^*$. The value iteration algorithm considers an arbitrary function v_0, and compute recursively $v_j = \Gamma(v_{j-1})$. Under regularity conditions, the sequence $\{v_j\}_{j\geq 1}$ converges to a limit which is the value function v^*.

The argument uses basic concepts of functional analysis. Convergence can be ensured, provided that Γ is a contractive operator in a complete metric space. If B is a complete[2] metric space, an operator $\Gamma : B \to B$ is called contractive if $d\left(\Gamma(v), \Gamma(v')\right) \leq cd(v, v')$ for all $v, v' \in B$ with parameter $c \in (0,1)$. Any contractive operator in a complete metric space has a unique fixed point v^*, and satisfies that $v^* = \lim_{j\to\infty} \Gamma^j(v^0)$ for any initial point $v^0 \in B$, so that the sequence $v^j = \Gamma(v^{j-1}) = \Gamma^j(v^0)$ converges to the fixed point. In particular we consider the Banach[3] space B_∞ of bounded and Borel-measurable real valued functions defined on the Euclidean state's space S, and endowed with the supremum norm $\|v\|_\infty = \sup_y |v(y)|$. If the function $|r(s, A, P, e)|$ is bounded on $\mathcal{K}$, then it is easy to prove that $\Gamma(v)$ is a contractive operator on B_∞ with parameter $\delta \in (0,1)$, and the fixed point $V = \Gamma(V)$ solves the SDP[4], see e.g. Denardo (1967), and Blackwell (1965). Under stronger conditions on the SDP problem, the value function V can be proved to be continuous, Lipschitz, once/twice continuously differentiable.

Unfortunately, the implementation of the algorithms is unfeasible with more than 3-4 state variables, as the computation of $\Gamma(v)$ requires approximation of the numerical integral $\int v(s') F(ds'|s, A, P, e)$ by an average at selected points, and the number of required points to provide an accurate estimate increases exponentially with the dimension of the state variables.

Next we discuss the convergence of the presented algorithm.

[2]A metric space B is complete if it is equal to its closure

[3]A Banach space is a normed linear space, which is complete with respect to the distance $d(v, v') = \|v - v'\|$ defined from its norm.

[4]There are also extensions for the case where $r(s, A, P, e)$ is bounded on compact subsets, by using other distances (see Rincón-Zapatero and Rodríguez-Palmero, 2003).

Recall that $V(S_0) = I^{-1}\sum_{i\in\mathbb{I}} V_i(S_{0i}) = \overline{V}\left(\overline{S}_0\right)$, where

$$\begin{aligned}
\overline{V}(\overline{s}) &= \max_{\{A_t,P_t\}} E_0\left[\sum_{t\geq 0}\delta^t R\left(\overline{S}_t, A_t, P_t\right)|\overline{S}_0 = \overline{s}\right], \\
V_i(s_i) &= \max_{\{e_t\}} E_0\left[\sum_{t\geq 0}\delta^t R_i(S_{it}, e_{it})|S_{i0} = s_i\right].
\end{aligned}$$

Consider the operators:

$$\begin{aligned}
\Upsilon_i(V_i, A, P)(s_i) &= \max_{\{e_i\in\mathbb{A}_i(s_i)\}}\left\{R_i(S_{it}, e_{it}) + \delta\int V_i(s_i')\,F^{A,P}(s_i'|s_i, e_i)\right\}, \\
\Phi\left(\overline{V}, e\right)(\overline{s}) &= \max_{\left\{A,P\in\overline{\mathbb{A}}(\overline{s})\right\}}\left\{R\left(\overline{S}_t, A_t, P_t\right) + \delta\int\overline{V}(\overline{s}')\,F^e(d\overline{s}'|\overline{s}, A, P)\right\}.
\end{aligned}$$

where $F^{A,P}(s_i'|s_i, e_i)$, $F^e(d\overline{s}'|\overline{s}, A, P)$ are defined as in the algorithm steps (2.1) and (2.3). The arguments that maximize these two problems are $\{e_i(s_i)\}_{i=1}^I$ and $(A(\overline{s}), P(\overline{s}))$, respectively. The convergence of the decomposition algorithm can be deduced similarly to the proof of convergence of the policy iteration method, using the following arguments:

1) The solution to the functional equation system

$$\begin{aligned}
\Upsilon_i(V_i, A, P)(s_i) &= V_i(s_i), \qquad i = 1..., n \\
\Phi\left(\overline{V}, e\right)(\overline{s}) &= \overline{V}(\overline{s})
\end{aligned}$$

satisfies by construction that $V(s) = I^{-1}\sum_{i=1}^I V_i(s_i, A(\overline{s}), P(\overline{s})) = \overline{V}(\overline{s}, \{e_i(s_i)\})$ a.e., where $V(s)$ is the value function of the original SDP problem.

2) The algorithm can be considered as a recursion defined by a contractive operator. Consider some initial value $V(s) \in B_\infty$, then we can write $V = \frac{1}{I}\sum_{i=1}^I V_i$ for a vector $(V_1, ..., V_I)$ with coordinates $V_i = \Pi_i V(s)$, where the operator Π_i is defined as:

$$\begin{aligned}
\Pi_i v(s) &= E\left[\sum_{t\geq 0}\delta^t R_{iv}(S_{it}, e_{iv}(S_{it}))|S_{i0} = s_i\right], \\
R_{iv}(S_{it}, e_{iv}(S_{it})) &= E\left[I\cdot r_i(S_{it}, e_{iv}(S_{it}), P_v(S_t), A_v(S_{it}))|S_{it}\right]
\end{aligned}$$

and $A_v(s), P_v(s), e_v(S)$ are the policies rendering the value function $v(s)$. These operators satisfy $\|\Pi_i\ (v)\|_\infty \leq \|v\|_\infty$.

The algorithm can be regarded as a sequence obtained alternating the operators $(\beta_1, ..., \beta_I)$ from $B_\infty \rightarrow B_\infty^I$ defined by $\beta_i = \Upsilon_i \circ \Pi_i V$, with the operator Φ. In other words, it is a recursion defined by the operator $\Delta = \left(\Phi \circ \frac{1}{I} \sum_{i=1}^I \beta_i\right)$ from $B_\infty \rightarrow B_\infty$. The operator Δ is a contractive operator on B_∞, since Φ and Υ_i are Bellman operators (contractive with parameter δ),

$$\begin{aligned}
\|\Delta\ (v)\|_\infty &= \left\|\Phi \circ \left(\frac{1}{I}\sum_{i=1}^{I}\beta_i\right)\ (v)\right\|_\infty \leq \delta \left\|\frac{1}{I}\sum_{i=1}^{I}\beta_i\ (v)\right\|_\infty \leq \delta \frac{1}{I}\sum_{i=1}^{I}\|\Upsilon_i \circ \Pi_i\ (v)\|_\infty \\
&\leq \delta^2 \frac{1}{I}\sum_{i=1}^{I}\|\Pi_i\ (v)\|_\infty \leq \delta^2 \|v\|_\infty
\end{aligned}$$

and we can apply a fixed point theorem to the alternating operator Δ to prove convergence to a fixed point satisfying the conditions in 1).

2.9 Appendix C: Algorithm Implementation

The first step follows the discretization technique. Mainly, we consider a grid of controls, $\{\mathbf{A}, \mathbf{P}, \mathbf{e}_1, ..., \mathbf{e}_I\}$, containing a discretization of the feasible decision set. In particular we consider relatively large finite intervals for each decision, and introduce N equidistant points for each decision.

The second step is the definition of the scenario nodes and transition probabilities across scenario states. The unconditional distribution can be used to define a grid of representative state values, and the conditional distribution to compute the transition matrix across the elements of the grid. In particular, when we consider the model 2.1

$S_{it} = \rho S_{it-1} + g_i + \varepsilon_{it}$ where $\varepsilon_{it} = \eta_i + u_{it} \sim N\left(0, \sigma^2_{\varepsilon_i}\right)$ with $\sigma^2_\varepsilon = \sigma^2_\eta + \sigma^2_u$, and

$$\begin{aligned} S_{it}|S_{it-1}, A, P, e &\sim N\left(\rho S_{it-1} + g_i, \sigma^2_\varepsilon\right), \\ \overline{S}_{It}|\overline{S}_{It-1}, A, P, e &\sim N\left(\rho \overline{S}_{It-1} + \overline{g}, \frac{\sigma^2_\varepsilon}{I}\right), \end{aligned}$$

with $g_i = g_i\left(A, P, e_i\right), \overline{g}\left(A, P, e\right) = \sum_i g_i\left(A, P, e_i\right)/I$. The stationary marginal distribution of S_{ti} and $\overline{S}_t$ are $N\left(\frac{g_i(A,P,e_i)}{(1-\rho)}, \frac{\sigma^2_\varepsilon}{(1-\rho^2)}\right)$ and $N\left(\frac{\overline{g}(A,P,e_i)}{(1-\rho)}, \frac{\sigma^2_\varepsilon}{I\cdot(1-\rho^2)}\right)$, respectively. For the i-th customer, we set scenarios in the interval $\left[S^l_i, S^u_i\right]$, where

$$\begin{aligned} S^l_i &= \min_{A,P,e_i} \frac{g_i\left(A, P, e_i\right)}{(1-\rho)} - 5\sqrt{\frac{\sigma^2_\varepsilon}{(1-\rho^2)}}, \\ S^u_i &= \max_{A,P,e_i} \frac{g_i\left(A, P, e_i\right)}{(1-\rho)} + 5\sqrt{\frac{\sigma^2_\varepsilon}{(1-\rho^2)}} \end{aligned}$$

Therefore, we cover 5 times the standard deviation from the most extreme mean values. After checking that $\max\left\{S^l_i, 0\right\} < S^u_i$ we generate N scenarios distributed uniformly as

$$\begin{aligned} s_{i1} &= \max\left\{S^l_i, 0\right\}, \\ s_{iN} &= S^u_i, \\ s_{in} &= s_{i1} + \left(\frac{s_{iN} - s_{i1}}{N-1}\right)(n-1), \qquad n = 2, 3, .., N-1. \end{aligned}$$

Then we define the product space of states $\mathcal{S}^I = \prod_{i=1}^I \left\{s_{i1}, ..., s_{iN}\right\}$. The discrete scenario grid $\mathcal{S}^I$ be used to compute the Bellman problem, defining the value functions and the policy functions as mappings defined on $\mathcal{S}^I$.

However, in our context it is convenient to think of an augmented space of states including mean sales. Consider the mean interval $\left[S^l, S^u\right]$, with $S^l = \sum_{i\in\mathbb{I}} S^l_i/I$ and $S^u = \sum_{i\in\mathbb{I}} S^u_i/I$, and generate N scenarios $\left\{\overline{s}_1, ..., \overline{s}_N\right\}$ distributed uniformly in $\max\left\{S^l, 0\right\} <$

S^u. Therefore, we can define the augmented space as

$$\mathcal{S}^{I+1} = \left\{ (s, \overline{s}) : s = (s_1, ..., s_I)' \in \mathcal{S}^I, \overline{s} \simeq \frac{1}{I} \sum_{i \in \mathbb{I}} s_i \right\},$$

where $\simeq$ means that $\overline{s}$ is the scenario in $\{\overline{s}_1, ..., \overline{s}_N\}$ closest to $\sum_{i \in \mathbb{I}} s_i / I$. Thus a specific realization of the random vector $\left(S_t, \overline{S}_t\right)$ will be approached by a vector $(s, \overline{s}) \in \mathcal{S}^{I+1}$. Given the structure of the problem, we can define the policy functions $\left(A^k, P^k, e^k\right)$ in the augmented space as a mapping

$$\left(A^k, P^k, e^k\right) : \mathcal{S}^{I+1} \ni s \rightarrow \left(A^k\left(\overline{s}\right), P^k\left(\overline{s}\right), e_1^k\left(s_1\right), ..., e_I^k\left(s_I\right)\right) \in \{\mathbf{A}, \mathbf{P}, \mathbf{e}_1, ..., \mathbf{e}_I\}.$$

The value function can be approximated in $\mathcal{S}^{I+1}$ by a simple function,

$$v\left(s, \overline{s}\right) = \sum_{n_1, ..., n_I, n_{I+1}} \theta_{n_1, ..., n_I, n_{I+1}} \cdot \left\{ \prod_{i=1}^{I} I\left(b_{n_i - 1} < s_i \leq b_{n_i}\right) \cdot I\left(b_{n_{I+1}-1} < \overline{s} \leq b_{n_{I+1}}\right) \right\}.$$

An smooth functional basis could be considered instead of simple functions, e.g. replacing the bracket in the previous expression by a tensor product of orthonormal polynomials.

We need to compute $\mathbb{F}^k\left(s_i' | s_i, e_i\right)$ and $\mathbb{F}^k\left(\overline{s}' | \overline{s}, A, P\right)$ in Step 2.2. In order to marginalize the effect of some policy controls over the transition probabilities, we apply the Monte Carlo method. First, given the policy $\left(A^k, P^k, e^k\right)$ we generate recursively a sample $\left\{S_t^k, A_t^k, P_t^k, e_t^k\right\}_{t=1}^T$ as

$$\begin{aligned} S_{it}^k &= \rho S_{it-1}^k + g_i\left(A_{t-1}^k, P_{t-1}^k, e_{it-1}^k\right) + \varepsilon_{it}, \qquad i \in \mathbb{I} \\ \overline{S}_{t-1}^k &= I^{-1} \sum_{i \in \mathbb{I}} S_{it}^k \end{aligned}$$

with $\varepsilon_i \sim N\left(0, \sigma_{\varepsilon_i}^2 I_T\right)$ and $S_{i0}^k = 0$, and compute recursively the associated controls as follows:

$$\begin{aligned}
A_t^k &= \sum_{n=1}^{N} A^k\left(\overline{s}_n\right) I\left(b_{n-1} < \overline{S}_{t-1}^k \leq b_n\right) \\
P_t^k &= \sum_{n=1}^{N} P^k\left(\overline{s}_n\right) I\left(b_{n-1} < \overline{S}_{t-1}^k \leq b_n\right) \\
e_{it}^k &= \sum_{n=1}^{N} e_i^k\left(s_{in}\right) I\left(b_{i,n-1} < S_{i,t-1}^k \leq b_{i,n}\right), \qquad i \in \mathbb{I},
\end{aligned}$$

where $b_n = \left(\overline{s}_{n+1} + \overline{s}_n\right)/2$ and $b_{i,n-1} = \left(s_{i,n+1} + s_{i,n}\right)/2$ for $n = 1, ..., N-1$, and we set $b_0 = b_{i,0} = -\infty$ and $b_N = b_{i,N} = +\infty$. The last expressions are used due to the fact that the policy functions are defined for discrete scenarios, for example we set $A_t^k = A^k\left(\overline{s}_n\right)$ whenever $\overline{S}_{t-1}^k \in \left(b_{n-1}, b_n\right]$ which is the interval centered in $\overline{s}_n$. We trow away the first 100 observations to remove the effect of the initial data, and continue to generate a large sample with at least $T = 3000$ observations, but this figure could be doubled when the diameter of the feasible decision set or N increases.

In order to define properly the objective function for each subproblem, we compute certain conditional expectations and transition kernels using the simulated sample $\left\{S_t^k, A_t^k, P_t^k, e_t^k\right\}_{t=1}^T$. First, for all $i \in \mathbb{I}$ we compute the conditional expectations $P_{in}^k = E\left[P_t^k | S_{it}^k = s_{in}\right]$, $C_{in}^k = E\left[c_m\left(A_t^k\right) | S_{it}^k = s_{in}\right]$, at the discrete scenarios $\left\{s_{in}\right\}_{n=1}^N$ and $c_{in}^k = E\left[c_i\left(e_{it}^k\right) | \overline{S}_t^k = \overline{s}_n\right]$ at the scenarios $\left\{\overline{s}_n\right\}_{n=1}^N$. Then we compute an approximation of the subproblem objective functions (2.5) evaluated at the discrete scenarios as

$$\begin{aligned}
R_i^k\left(s_{in}, e_{it}\right) &= I \cdot \left(\left(P_{in}^k - c_0\right) \cdot s_{in} - c_i\left(e_{it}\right) - I^{-1} C_{in}^k\right), \\
R^k\left(\overline{s}_n, A_t, P_t\right) &= \left(P_t - c_0\right) \cdot I \cdot \overline{s}_n - \sum_{i \in \mathbb{I}} c_{in}^k - c_m\left(A_t\right).
\end{aligned}$$

The fastest method to compute the conditional expectations is based on a simple parametric regression model (e.g., specifying $E\left[P_t^k | S_{it}^k = s_i\right] = p\left(s_i, \beta\right)$). The model is estimated by a least squares method (e.g., minimizing $\sum_{t=1}^T \left(P_t^k - p\left(S_{it}^k, \beta\right)\right)^2$) for direct use (setting $P_{in}^{k}=p\left(s_{in},\widehat{\beta}^{K}\right)$ for each discrete scenario s_{in}). The parametric approach works well in our application. Alternatively we can use a nonparametric estimator. For example the Nadaraya-Watson estimator of $E\left[P_{t}^{k}|S_{it}^{k}=s_{in}\right]$, is given by

$$E\left[P_{t}^{k}|S_{it}^{k}=s_{in}\right]=\frac{\sum_{t=1}^{T}P_{t}^{k}\ K_{h_{T}}\left(S_{it}^{k}-s_{in}\right)}{\sum_{t=1}^{T}K_{h_{T}}\left(S_{it}^{k}-s_{in}\right)}$$

where $K_{h_{T}}(u)=h_{T}^{-1}K(u/h_{T})$ for an arbitrary kernel density $K(\cdot)$ (e.g. a standard normal density), and a sequence of positive smoothing parameters h_{T} such that $h_{T}+(Th_{T})^{-1}\rightarrow 0$. This approach avoids specification assumptions, but it requires larger sample sizes T than the parametric approach. Besides, an optimal selection of the smoothing parameter is crucial, which is time consuming. However, it might be convenient in some applications.

Second we compute the marginal transition kernels $\mathbb{F}^{k}(s_{i}'|s_{i},e_{i})$ and $\mathbb{F}^{k}(\overline{s}'|\overline{s},A,P)$. There are several possibilities: parametric methods, semiparametric, and nonparametric. The fastest method is based on a parametric model, postulating regression model, $E\left[S_{it}^{k}|S_{it-1}^{k},e_{it-1}^{k}\right]=m_{i}\left(S_{it-1}^{k},e_{it-1}^{k},\beta_{i}\right)$, $E\left[\overline{S}_{t}^{k}|\overline{S}_{t-1}^{k},A_{t-1}^{k},P_{t-1}^{k}\right]=\overline{m}\left(\overline{S}_{t-1}^{k},e_{it-1}^{k},\theta\right)$, estimating the model by a ordinary/nonlinear least squares method. In our applications we consider this method for a linear in parameters model without intercept where first regressor is in levels and the controls are in logarithms. Assume that the errors are conditionally independent of the state variables, we can use the residuals

$$\begin{aligned}\widehat{u}_{it} &= S_{it}^{k}-m_{i}\left(S_{it-1}^{k},e_{it-1}^{k},\widehat{\beta}_{i}\right)\\ \widehat{\overline{u}}_{t} &= \overline{S}_{t}^{k}-\overline{m}\left(\overline{S}_{t-1}^{k},A_{t-1}^{k},P_{t-1}^{k},\widehat{\theta}\right)\end{aligned}$$

to estimate the error densities $g_{i}(u_{it})$, $\overline{g}(\overline{u}_{t})$. In particular we have assumed Gaussian distributions $N\left(0,\sigma_{u_{i}}^{2}\right)$ and $N\left(0,\sigma_{\overline{u}}^{2}\right)$ respectively, estimating the variances $\sigma_{u_{i}}^{2}$ and $\sigma_{\overline{u}_{t}}^{2}$

with the mean squared residuals, we get

$$\begin{aligned}\mathbb{F}_i\left(s_i'|s_i,e_i\right) &= \frac{1}{\widehat{\sigma}_{u_i}}\int_{-\infty}^{s_i'}\phi\left(\frac{z-m_i\left(s_i,e_i,\widehat{\beta}_i\right)}{\widehat{\sigma}_{u_i}}\right)dz=\Phi\left(\frac{s_i'-m_i\left(s_i,e_i,\widehat{\beta}_i\right)}{\widehat{\sigma}_{u_i}}\right),\\ \mathbb{F}\left(\overline{s}'|\overline{s},A,P\right) &= \frac{1}{\widehat{\sigma}_{\overline{u}}}\int_{-\infty}^{\overline{s}'}\phi\left(\frac{z-\overline{m}\left(\overline{s},A,P,\widehat{\theta}\right)}{\widehat{\sigma}_{\overline{u}}}\right)dz=\Phi\left(\frac{\widehat{\overline{u}}_t-\left(\overline{s}'-\overline{m}\left(\overline{s},A,P,\widehat{\theta}\right)\right)}{\widehat{\sigma}_{\overline{u}}}\right)\end{aligned}$$

Notice that if it is difficult to determine the residuals distribution, we could estimate $g_i\left(u_{it}\right)$, $\overline{g}\left(\overline{u}_t\right)$ nonparametrically. For example, integrating the Rosenblatt-Parzen kernel density estimator we obtain a cumulative conditional distribution

$$\begin{aligned}\mathbb{F}_i\left(s_i'|s_i,e_i\right) &= \int_{-\infty}^{s_i'}\left(\frac{1}{T-2}\sum_{t=2}^{T}K_{h_T}\left(\widehat{u}_{it}-\left(z-m_i\left(s_i,e_i,\widehat{\beta}_i\right)\right)\right)\right)dz,\\ \mathbb{F}\left(\overline{s}'|\overline{s},A,P\right) &= \int_{-\infty}^{\overline{s}'}\left(\frac{1}{T-2}\sum_{t=2}^{T}K_{h_T}\left(\widehat{\overline{u}}_t-\left(z-\overline{m}\left(\overline{s},A,P,\widehat{\theta}\right)\right)\right)\right)dz,\end{aligned}$$

where $K_{h_T}\left(u\right)=h_T^{-1}K\left(u/h_T\right)$. This semiparametric method slows down the algorithm compared with the parametric case. The last alternative is a fully nonparametric estimator such as the cumulated integral of the conditional density estimator by Roussas (1967, 1969) and Chen, Linton and Robinson (2001),

$$\begin{aligned}\mathbb{F}_i\left(s_i'|s_i,e_i\right) &= \int_{-\infty}^{s_i'}\frac{\sum_{t=2}^{T}K_{h_T}\left(S_{it}^k-z\right)K_{h_T}\left(S_{it-1}^k-s_i\right)K_{h_T}\left(e_{it-1}^k-e_i\right)}{\sum_{t=2}^{T}K_{h_T}\left(S_{it-1}^k-s_i\right)K_{h_T}\left(e_{it-1}^k-e_i\right)}dz\\ \mathbb{F}\left(\overline{s}'|\overline{s},A,P\right) &= \int_{-\infty}^{\overline{s}'}\frac{\sum_{t=2}^{T}K_{h_T}\left(\overline{S}_t^k-z\right)K_{h_T}\left(\overline{S}_{t-1}^k-\overline{s}\right)K_{h_T}\left(A_{t-1}^k-A\right)K_{h_T}\left(P_{t-1}^k-P\right)}{\sum_{t=2}^{T}K_{h_T}\left(\overline{S}_{t-1}^k-\overline{s}\right)K_{h_T}\left(A_{t-1}^k-A\right)K_{h_T}\left(P_{t-1}^k-P\right)}dz\end{aligned}$$

This method requires very large simulated samples, and it is quite sensitive to the selection of the smoothing number that must be optimally determined. In general we do not recommend it for this algorithm, but it might be useful in some applications.

To apply the collocation method for the Bellman equation associated to each subproblem we have to integrate the basis functions with respect to $\mathbb{F}_i\left(s_i'|s_i,e_i\right)$ and $\mathbb{F}\left(\overline{s}'|\overline{s},A,P\right)$, which requires a numerical integration method. We use the Tauchen's method (1986) to approximate the continuous transition kernel $\mathbb{F}^k\left(s_i'|s_i,e_i\right)$ and $\mathbb{F}^k\left(\overline{s}'|\overline{s},A,P\right)$ by analogous finite-state transition matrix on the states grid $\{s_1,...,s_N\}$, considering for all $n,m=1,...,N$ the transition from s_n to s_m

$$\begin{aligned}\Pi^i_{nm}\left(e_i\right) &= \mathbb{F}_i\left(b_{i,m}|s_{in},e_i\right)-\mathbb{F}_i\left(b_{i,m-1}|s_{in},e_i\right),\\ \Pi^{mean}_{nm}\left(A,P\right) &= \mathbb{F}\left(b_m|s_n,A,P\right)-\mathbb{F}\left(b_{m-1}|s_n,A,P\right),\end{aligned}$$

where $b_{i,m}=\left(s_{i,m+1}+s_{i,m}\right)/2$, $b_m=\left(\overline{s}_{m+1}+\overline{s}_m\right)/2$ for $m=1,...,N-1$, and we set $b_{i,0}=b_0=-\infty$ and $b_{i,N}=b_N=+\infty$ so that $\Pi^i_{n1}\left(A,P,e\right)=\mathbb{F}_i\left(b_1|s_n,e_i\right),\Pi^i_{nN}\left(A,P,e\right)=1-\mathbb{F}\left(b_{N-1}|s_n,A,P\right)$, and similarly for $\Pi^{mean}_{n1}\left(A,P\right)$ and $\Pi^{mean}_{nN}\left(A,P\right)$. In order to apply the collocation value iteration, or policy iteration method, the continuous-state expectations of the basis functions (2.8) for each subproblem, namely $\int\phi_k\left(s'\right)\mathbb{F}_i\left(ds'|s_m,e_i\right)$ and $\int\phi_k\left(s'\right)\mathbb{F}\left(ds'|s_m,A,P\right)$, are approximated by the expected values in the analogous discrete Markov chain $N^{-1}\sum_{n=1}^N\phi_k\left(s_n'\right)\Pi^i_{nm}\left(e_i\right)$ and $N^{-1}\sum_{n=1}^N\phi_k\left(s_n'\right)\Pi^{mean}_{nm}\left(A,P\right)$ respectively.

Chapter 3

Can we curb retail sales volatility through marketing mix actions?

3.1 Introduction

Do marketing managers care about sales volatility? Admittedly, they tend to focus on sales trends and overlook short-term fluctuations. But, retailers and suppliers agree that the volatility of sales leads to significant added costs at the retail and supplier level (Cachon, Randall, Schmidt 2007) through, for example, the requirement of inventory investments for avoiding stock-out in case of demand peaks (see Holt et al. 1960, Bo 2001), stock that will be redundant with demand falls, leading to excessive inventory and inefficient production with a row of financial costs. The demand variability also expands costs in human resources due to hiring, training, dismissals and possible payoffs of employees.

Recently, Capgemini consulting company has conducted an annual global supply chain study that covers 300 leading companies from various sectors across Europe, the US and Canada, Asia-Pacific and Latin America. The 2011 study reveals that 40% of respondents answer that demand volatility is the number one business driver. Additionally, a

2010 online survey prepared by Edge Research for IBM-Sterling Commerce shows that managing sales volatility and risk is one of the top priorities for the majority of the respondents (the survey is based on a screened panel of 301 sales, IT and supply chain corporate decision makers). Today, companies spend billions of dollars annually on software, personnel and consulting fees to achieve accurate demand forecasts (Aiyer and Ledesma 2004). The analysis of real time demand is the most prominent information "black hole" among companies.

In practice, managers usually rely on estimations of future expected (mean) sales conditional on historical data, but the magnitude of deviations (oscillations) around that mean can evolve over time and finally become even a direr threat itself. The oscillating (volatile) demand is often magnified when the product is brought to customers through long *distribution channel* and *supply chain* streams. Down-stream echelons (retail) increase (reduce or stop) orders under high (low) demand adapting their inventory buffer. But, moving up-stream the demand oscillations are magnified as up-stream echelons (manufacturers) try to fulfill the demand of its predecessor in the chain (the *Bullwhip effect*). Literature first recognizes the volatility of demand for Procter and Gamble's diapers, whose end consumer demand was reasonably stable (Lee, Padmanabhan, and Whang, 1997, 2004). Since then, many authors suggest several managerial practices to mitigate the volatility amplification, but even at retail level volatility is a challenging threat.

Various issues related to volatility in marketing have been studied in the literature. For example, Srivastava, Shervani, and Fahey (1998) study how marketing drivers (such as retention, customer satisfaction, and loyalty) can reduce volatility by the direct stabilization of revenues and cash-flow of the firm. Surprisingly, there has been little research that considers the effects of marketing actions on both expected (mean) sales and its volatility. Raju (1992) examines the effect of discounts and price on variability in weekly category sales and finds that the magnitude of discounts is positively associated with sales volatility. Vakratsas (2008) studies the effects of marketing actions (advertising, price and distribution) on market share volatilities and their relevance to firm decision making. Fischer, Shin, Hanssens (2012) show how marketing spending policies have an impact on the level as well as the volatility of revenues and cash flows.

Unfortunately, to our knowledge, the marketing sales response literature has not paid attention to volatility measured as a conditional variance. In this paper, we study retail sales and marketing mix dynamics considering both conditional mean and covariance of sales and marketing mix within brands and across brands. Modeling volatility of sales in a particular brand seems important, but the co-movements of sales and marketing mix decisions over different brands is relevant from a strategic perspective. This is why, in this paper, we consider multivariate co-volatility models including all marketed brands simultaneously.

The article is organized as follows: First, we explain the reasons for focusing on sales conditional mean and variance. Next, we present the models that we will consider for the empirical problem at hand. Then, we apply the model to six product categories sold by *Dominick's Finer Foods*. The paper concludes with some strategic and managerial implications for brand management.

3.2 Reasons for focusing on sales conditional mean and variance (volatility)

Let us consider a stationary process $\{X_t\}_{t \in \mathbb{Z}}$ where X_t is a random vector in $\mathbb{R}^d$, with first moments $\mu = E[X_t]$, $H = Var[X_t]$ positive definite. The vector X_t contains sales of complementary/substitutive brands sold in the same market and the marketing mix variables related to the studied brands (all variables in logarithmic differences to ensure stationarity). Denote by $\Im_t$ the past information available up to time t. As the observations are generally dependent, the use of information in $\Im_t$ can improve the quality of the forecast of X_t. Let us denote the conditional models by $\mu_t = E[X_t|\Im_t]$, and $H_t = Var[X_t|\Im_t]$. The classical time series models (e.g., VAR, VARMA, VARMAX, and their extensions for integrated processes) are focused on the specification and estimation of μ_t, assuming that $H_t = H$ for all periods of time, which is appropriate when there is no volatility. But if that is not the case, H_t provides a much better intuition about sales fluctuations than H. To show this, recall that from the variance equation analysis, $H = E[H_t] + Var[\mu_t]$, implying that $H \geq E[H_t]$ (i.e., $(H - E[H_t])$ is positive definite) so that the risk is on average overrated if we use a marginal or static variance, forcing companies to have oversized safety stocks regularly. Moreover, $H \geq E[H_t]$ is compatible with reverse situations $H \leq H_t$ in some scenarios, suggesting that the safety stocks determined from H can be occasionally too short for insurance against the stock-out risk.

Note also that fluctuations in μ_t are relatively easy to forecast. They can be caused by seasonality (modeled with deterministic dummies, or modeled by seasonal unit roots leading to more realistic stochastic approaches), or they can be caused by business cycles (that can be modeled with sinusoidal deterministic trends, or a more flexible linear ARMA type model with complex roots leading to stochastic cycles). Expected sales fluctuations can be anticipated, and companies can adapt the production-inventory policies to fit the forecast. The actual risk is caused by deviations $(X_t - \mu_t)$ that cannot be anticipated previously, but its average magnitude is measured by H on the whole, and by H_t for each specific period of time. Our central objective is the study of H_t. As the vector X_t includes sales and marketing mix variables for several competitors in the same market, by estimating H_t we study the crossed co-volatilities of marketing-mix on retail sales.

3.3 The Model

To model and forecast time-varying retail sales and marketing activities volatility and their crossed effects within brand and between competitive brands, we combine features from classical time series models for the analysis of conditional means with recent models for conditional variances. There is an extensive literature on modeling volatility in financial time series since the introduction of the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) univariate model by Engle (1982) and Bollerslev (1986). These models have been extended to multivariate time series.

In particular, we consider that sales and marketing mix variables define a stochastic $\mathbb{R}^d$-vector process $\{X_t\}_{t\in\mathbb{Z}}$ satisfying that

$$\begin{aligned} X_t &= \mu_t + u_t, \\ u_t &= H_t^{1/2}\varepsilon_t, \end{aligned}$$

where $E[u_t|\Im_t] = E[\varepsilon_t|\Im_t] = 0$, $Var[\varepsilon_t|\Im_t] = I$. Typically $\mu_t = E[X_t|\Im_t]$ is defined by a VAR or a vector ARMA model (or a more sophisticated model dealing with features such as seasonality, unit roots, etc.) This is widely used in marketing (see Dekimpe and Hanssens, 2000), and we do not delve into the analysis of the conditional mean. For the conditional heteroskedasticity, we assume that $H_t = Var[X_t|\Im_t]$ follows a $BEKK(p,q,1)$ model (where BEKK stands for Baba-Engle-Kraft-Kroner) given by

$$H_t = \widetilde{\omega}\widetilde{\omega}' + \sum_{j=1}^{q} \widetilde{A}_j u_{t-j} u_{t-j}' \widetilde{A}_j' + \sum_{j=1}^{p} \widetilde{B}_j H_{t-j} \widetilde{B}_j', \tag{3.1}$$

where $\widetilde{\omega}$ is a lower triangular $\mathbb{R}^{d\times d}$ matrix and $\widetilde{A}_j$, $\widetilde{B}_j$ are $\mathbb{R}^{d\times d}$ matrices. The model was introduced in Baba, Engle Kraft and Kroner (1991), and Engle and Kroner (1995). This model is a multivariate generalization of the $GARCH$ process guarantying that H_t is positive definite. When the matrices $\widetilde{A}_j$, $\widetilde{B}_j$ are diagonal, and the only nonzero elements are those associated to square elements $\left\{u_{jt-1}^2\right\}$, then we have a process X_t which conditional covariances are constant, and only the conditional variances evolve. When this is not the case, it means that X_t has covolatility (the level of volatility in one element X_{jt} affects volatility in other element X_{it}).

Denote by vec the operator that stacks the column of a matrix, and $vech$ the vector-half operator which stacks the lower triangular portion of a matrix (on and below the main diagonal). The vec operator satisfies that $vec\,(ABC) = (C' \otimes A)\ \ vec\,(B)$, where $\otimes$ denotes the Kronecker product of two matrices. In order to handle the BEKK model, we can rewrite (3.1) using the vec operator as follows

$$vec\,(H_t) = (\widetilde{\omega} \otimes \widetilde{\omega})\ \ vec\,(\ I) + \sum_{j=1}^{q} \left(\widetilde{A}_j \otimes \widetilde{A}_j\right)\ \ vec\left(u_{t-j}u'_{t-j}\right) + \sum_{j=1}^{p} \left(\widetilde{B}_j \otimes \widetilde{B}_j\right)\ \ vec\,(H_{t-j})\,.$$

The dimension of $vec\,(H_t)$ is d^2. Since the matrices involved in this representation are symmetric, we can reduce the dimension. Using the vector-half operator $vech$ we rewrite (3.1) as

$$\begin{aligned} h_t &= w + \sum_{j=1}^{q} A_j\ vech\left(u_{t-j}u'_{t-j}\right) + \sum_{j=1}^{p} B_j\ h_{t-j} \\ &= w + \alpha\,(L)\ \ vech\,(u_t u'_t) + \beta\,(L)\ \ h_t. \end{aligned} \tag{3.2}$$

where $h_t = vech\,(H_t)$ has dimension $d\,(d+1)\,/2$, and we have used the matrices $w = D_d^+\,(\widetilde{\omega} \otimes \widetilde{\omega})\,D_d vec\,(\ I)$, $A_j = D_d^+ \left(\widetilde{A}_j \otimes \widetilde{A}_j\right) D_d$, $B_j = D_d^+ \left(\widetilde{B}_j \otimes \widetilde{B}_j\right) D_d$, with D_d the $d^2 \times d\,(d+1)\,/2$ duplication matrix defined by the property $vec(H) = D_d\ vech\,(H)$ for any symmetric $d \times d$ matrix H (i.e., D_d contains some columns from the identity matrix $I_{d^2 \times d^2}$ extracting elements in $vec\,(H)$ coming from the lower triangle of H) and D_d^+ denotes its Moore Penrose inverse. For the last equation (3.2) we have used a compact notation with matrix polynomials $\alpha\,(L) = \sum_{j=1}^{q} A_j L^j$ and $\beta\,(L) = \sum_{j=1}^{p} B_j L^j$ in the lag operator L. In this paper we examine the marketing strategic implications that can be used to curb volatility whilst increasing expected levels in μ_t. We also analyze crossed effects among competitors, via mean and variance, and how this can be used as a competitive advantage.

3.4 Empirical setting

We use store-level scanner data made available by the James M. Kilts Center, University of Chicago, from *Dominick's Finer Foods*, the largest grocery retailer in the Chicago market. The database includes all weekly sales, shelf price, possible presence of sales promotions (coupons, bulk buy, or a special sale), retail margin, and daily store traffic, by individual item (referenced by UPC) for more than 25 product categories, and collected for 96 stores operated in the Chicago area over a period of more than seven years from 1989 to 1997. For the analysis, we aggregate the weekly sales data across stores, computing also the average price. We also compute a continuous promotion variable defined as the percentage of stores implementing any sales promotion. We perform our empirical analysis using six different 'fast moving consumer product categories' (products are sold quickly and at relatively low cost): cheese, refrigerated juice, laundry detergent, toilet tissue, paper towel and toothpaste. As can be seen in Table 1, for cheese and refrigerated juice categories, we consider two brands with the highest market share, forming 80% and 82% of the total category volume respectively whereas for laundry detergent, toilet tissue, paper towel and toothpaste categories we focus on the top three selling brands constituting 70%, 66%, 60% and 73% of the market, respectively.

Table 1. Description of six categories used in the application

Category	Number of analyzed brands	Total number of brands in the category	Market Share of the analyzed brands
Cheese	2	12	80%
Refrigerated juice	2	7	82%
LaundryDetergent	3	14	70%
Toilet tissue	3	10	66%
Paper towel	3	13	60%
Toothpaste	3	13	73%

Table 2 provides more details on the analyzed brands in each category. In the **cheese** category, the two competing brands Dominick's and Kraft do not differ much since their average prices, promotions and sales are very close to each other. Also, the standard deviations of those variables show very small difference. In the **refrigerated juice** category, the two brands, Minute Maid and Tropicana, are similar in terms of average prices. The prices of those brands do not vary too much, i.e. the standard deviations are very close to each other. However, the sales of Tropicana almost doubles that of Minute Maid. For the **laundry detergent** we observe that brand's promotion intensity, are close to each other on average as well as their variabilities. The prices differ across the three brands. This difference may be perceived as signals of quality. This difference may make the brands differentiate them from the competitors. Similarly, in the **toilet tissue** category, brands' average promotions as well as the variability of the brand's promotions do not differ much, but prices are different and have different volatilities. Regarding the **paper towel** category, the average prices differ across the brands Bounty, Scott and Dominick's. Scott does more promotion on average, but we see almost no difference in the promotions variability. In the **toothpaste** category, the average prices of the three brands are very close to each other. Aquafresh has the highest price variability and the lowest average sales, while Crest has the moderate price variability, but the highest average sales. The average promotion differs across brands, but the level of the promotion is low.

Table 2. Descriptive statistics of the analyzed brands

Category	Variable	Mean	Median	Maximum	Minimum	Std. Dev.	Observations
Cheese	*Sales Dominick's*	104123.20	101624.00	264441.00	53745.00	26753.70	392
	Price Dominick's	2.10	2.11	2.71	1.26	0.21	392
	Prom Dominick's	92.24	97.67	100.00	0.00	20.09	392
	Sales Kraft	138445.50	122791.00	543061.00	83550.00	53543.57	392
	Price Kraft	1.99	1.95	4.10	0.90	0.28	392
	Prom Kraft	93.81	97.67	100.00	0.00	15.31	392
Refrigerated Juice	Sales Minute Maid	38690.04	23664.00	263612.00	13651.00	35673.52	396
	Price Minute Maid	2.14	2.17	3.49	1.06	0.35	396
	Prom Minute Maid	69.93	94.12	100.00	0.00	40.51	396
	Sales Tropicana	65726.83	48877.50	271965.00	26883.00	40931.44	396
	Price Tropicana	2.47	2.55	3.44	1.21	0.35	396
	Prom Tropicana	82.64	97.42	100.00	0.00	31.10	396
Laundry detergent	*Sales Wisk*	7339.28	4841.50	52357.00	1967.00	7061.49	396
	Price Wisk	5.31	5.24	8.88	2.81	0.87	396
	Prom Wisk	65.11	92.96	100.00	0.00	42.59	396
	Sales All	8332.82	5368.50	133703.00	2265.00	12171.34	396
	Price All	4.51	4.55	7.05	2.48	0.56	396
	Prom All	67.96	94.12	100.00	0.00	40.92	396
	Sales Tide	26318.69	20320.50	135839.00	10586.00	19925.67	396
	Price Tide	6.20	6.29	9.22	3.43	0.85	396
	Prom Tide	68.04	94.12	100.00	0.00	41.12	396
Toilet Tissue	*Sales Scott*	86464.14	56745.50	2062849.00	26163.00	155969.67	384
	Price Scott	0.70	0.66	1.88	0.25	0.20	384
	Prom Scott	47.67	56.89	100.00	0.00	44.17	384
	Sales Charmin	37143.97	19470.50	478101.00	9436.00	58301.06	384
	Price Charmin	2.11	2.13	3.14	0.70	0.51	384
	Prom Charmin	40.30	4.22	100.00	0.00	44.37	384
	Sales Northern	31854.61	18358.00	314957.00	10125.00	38232.68	384
	Price Northern	1.71	1.67	3.45	0.82	0.39	384
	Prom Northern	49.55	61.76	100.00	0.00	46.12	384
Paper Towel	*Sales Bounty*	34452.86	29840.50	163198.00	17202.00	17121.08	388
	Price Bounty	1.42	1.44	2.68	0.75	0.24	388
	Prom Bounty	39.72	2.33	100.00	0.00	44.20	388
	Sales Scott	22764.20	20314.00	112534.00	6115.00	15312.97	388
	Price Scott	1.35	1.28	2.28	0.76	0.30	388
	Prom Scott	61.15	86.04	100.00	0.00	42.96	388
	Sales Dominick's	23822.99	18030.50	208968.00	1346.00	24810.48	388
	Price Dominick's	0.76	0.71	2.31	0.34	0.21	388
	Prom Dominick's	40.29	35.92	100.00	0.00	40.49	388
Toothpaste	*Sales Aquafresh*	3746.22	3159.00	20727.00	1642.00	2171.50	398
	Price Aquafresh	2.24	2.31	3.24	1.05	0.50	398
	Prom Aquafresh	36.43	1.18	100.00	0.00	43.25	398
	Sales Colgate	9082.61	8297.00	25196.00	4926.00	3383.23	398
	Price Colgate	2.21	2.25	2.60	1.56	0.24	398
	Prom Colgate	49.72	48.54	100.00	0.00	44.31	398
	Sales Crest	12176.40	11528.00	49820.00	6769.00	4144.69	398
	Price Crest	2.30	2.37	2.80	1.18	0.32	398
	Prom Crest	43.10	29.48	100.00	0.00	43.99	398

3.5 Analysis of conditional sales mean and variance

So far, we have considered the BEKK model in sales and marketing activities X_t for each period of time. However, several steps should be considered to specify the model: In step 1, we perform preliminary analysis that includes exploratory data analysis, and the analysis of the time series' levels. In step 2, we estimate consistently a model for the conditional mean (typically a VAR model parameters by OLS). Then, in step 3, we explore the existence of volatility in the data and specify a BEKK model for the residuals using a Maximum Likelihood Estimation (MLE) method. In step 4, we compute preliminary estimations for the parameters of the BEKK model. Finally, in step 5, we improve the efficiency of the estimators. We simultaneously estimate the VAR and BEKK parameters by full Gaussian Maximum Likelihood, using the consistent estimates from step 2 and 4 as initial values for the Newton Method (this choice is crucial given the high dimension of the problem). In Step 6, we analyze the estimation output and perform specific tests of independence and Granger causality.

Step 1. Preliminary Analysis:

We perform the standard exploratory data analysis. Then, we study the usual properties such as stationarity and cointegration involving inspection of data graphs, Auto Correlation Functions (ACF), crossed, and partial autocorrelations. We decide to take natural logarithm for all variables. For most of the brands in all categories, we observe that the ACFs decay very slowly which is typical of a nonstationary time series. This conclusion is also consistent with the results of the Augmented Dickey-Fuller, Phillips-Perron and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root tests. Given the evidence, we opt for taking one difference of all variables in logarithms, which can be interpreted as growth ratios of the original series from period to period. Next, we conduct Johansen's cointegration test to study whether the integrated variables are cointegrated (i.e., if they have a long-run equilibrium in levels). Cointegration would imply the specification of a Vector Error Correction (VEC) model instead of a VAR model for variables in differences. For all categories, we fail to reject the null hypothesis that the variables are not cointegrated. Note that in general these tests (unit root and cointegration tests) do not take into account conditional heteroskedasticity, and the output is somewhat exploratory, but it confirms the graphical analysis suggestions. Therefore, we proceed to estimate a VAR model for variables in logarithmic first differences. More complex models could be used if the inspection of the data shows evidence of other alternative specifications such as VARMA or VECM.

Step 2. Conditional mean analysis:

Let us denote X_t the vector of log-differenced variables. Here we focus on the analysis of $\mu_t = E\left[X_t|\Im_t\right]$. We model the dynamic interactions among the variables through a VAR model (including all variables as endogenous). We choose the optimal lag length of the VAR model to be 1 based on the visual inspection of the ACFs of the first differenced log series. We also compute the information criteria (commonly used in the marketing literature, see Dekimpe and Hanssens, 1999; Pauwels et al. 2004). Schwarz information criterion (SIC) suggests one lag for all categories. As a result, we specify a VAR(1) model $\mu_t = \Pi X_{t-1}$. We estimate $\widehat{\Pi}$ by OLS, minimizing

$$Q(\Pi) = tr\left\{\sum_{t=2}^{T}(X_t - \Pi X_{t-1})'(X_t - \Pi X_{t-1})\right\} \tag{3.3}$$

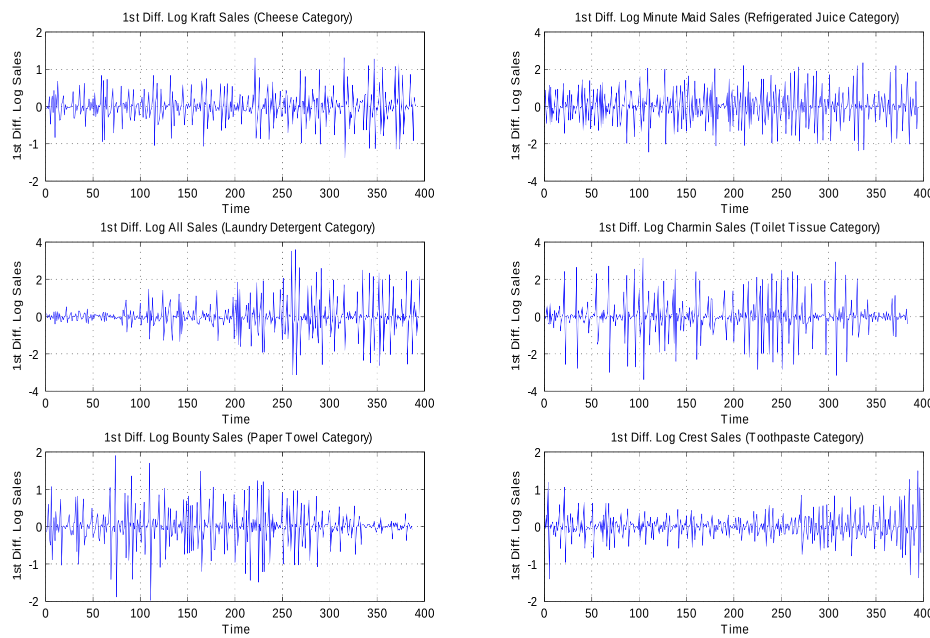
where tr denotes the trace. The solution is $\widehat{\Pi}' = \left(\sum_{t=2}^{T} X_{t-1}' X_{t-1}\right)^{-1} \sum_{t=2}^{T} X_{t-1}' X_t$. We also obtain the residuals $\widehat{u}_t = X_t - \widehat{\Pi} X_{t-1}$ to be used as a preliminary tool for volatility analysis.

Step 3. Volatility Modeling:

Before carrying out our volatility model estimation, we explore the presence of volatility in our data. We first study the volatility of the residual series independently (univariate analysis), and then study the appropriate multivariate BEKK model:

(i) *visual inspection of the sales plots:* As an example, Figure 1 shows the first differenced logarithm of one brand for each category. The plots show that in general the volatility is higher at some periods than others indicating that the conditional variance is not constant over time.

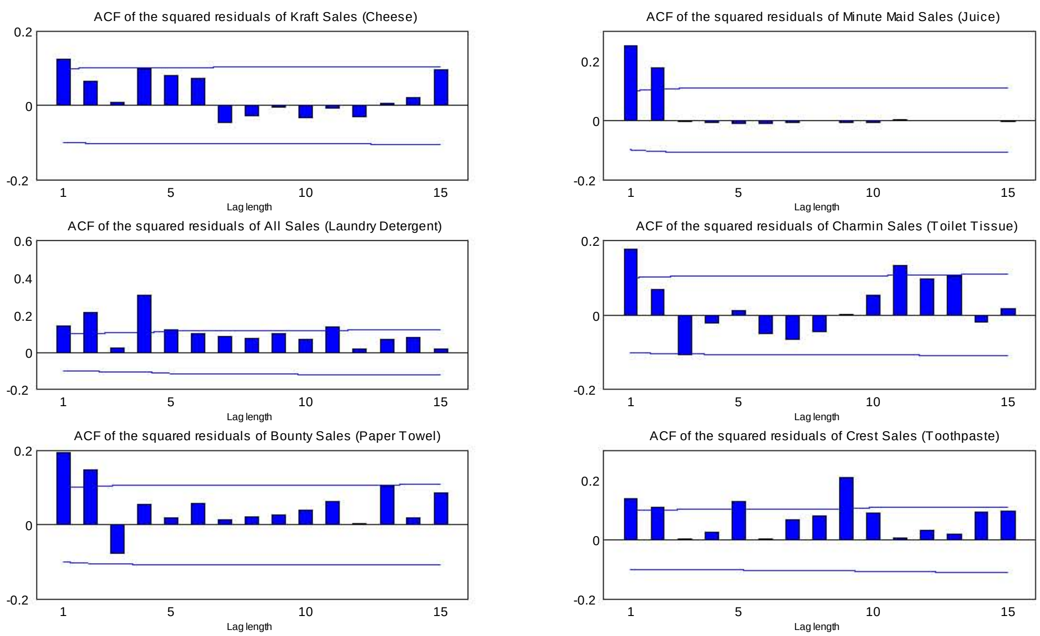
Figure 1. First differenced logged sales for some brands



(ii) *ACFs of the squared OLS residuals from the VAR(1) model:* We find substantial evidence of Autoregressive Conditional Heteroscedasticity (ARCH) effects as judged by the autocorrelations of the squared residuals. As can be seen from Figure 2, even though the magnitude of the autocorrelations sometimes are small after lag 1 or 2, the ACF plots of the squared residuals of sales variables show the presence of autocorrelation patterns. This suggests the existence of (Generalized) Auto Regressive Conditional

Heteroskedasticity (ARCH/GARCH models).

Figure 2. ACFs of the squared residuals of sales for some brands



(iii) *ARCH test:* We test formally the hypothesis of conditional heteroskedasticity applying Engle's (1982) ARCH test. The null hypothesis is that there is no autocorrelation in the squared residuals (and therefore no ARCH effect). For all brands (with two exceptions: brand Wisk in the laundry detergent category and brand Colgate in the toothpaste category), we reject the no-ARCH hypotheses, supporting our findings in (i) and (ii).

Finally, we study multivariate co-volatility relationships and specify a full BEKK model. In order to decide how many lags to be included, we use the ACFs and partial ACFs of the squared residuals. If we define $\eta_t = vech\,(u_t u_t') - h_t$, then we can write model (3.2) as a $VARMA(r,p)$ with $r = \max(p,q)$, given by

$$vech\,(u_t u_t') = w + (\alpha(L) + \beta(L))\;\, vech\,(u_t u_t') - \beta(L)\,\eta_t + \eta_t,$$

where η_t are martingale differences, and if X_t has four order moments $E[\eta_t \eta_t'] = Var[vech(X_t X_t')] - E[h_t h_t']$. The model is covariance stationary if and only the roots of $|\alpha(L) + \beta(L)| = 1$ lie outside the unit circle, which usually occurs when $(\alpha(1) + \beta(1))$ has eigenvalues with modulus smaller than one. We also assume that p, q are as small as possible given that the matrices A_r and B_p have full rank, and the polynomials $(I - (\alpha(L) + \beta(L)))$ and $(I - \beta(L))$ have neither unit roots nor common left factors other than unimodular ones. The $VARMA(r, p)$ representation shows that we can identify p, q with the classical tools. If we estimate μ_t with standard time series methods (i.e. without taking care of the heteroskedasticity), and we can use the residuals $\widehat{u}_t = (X_t - \widehat{\mu}_t)$ to estimate autocorrelation functions for $vech(\widehat{u}_t \widehat{u}_t')$, which can be used to determine an appropriate p, q orders. In our case, inspection of sample autocorrelations for $vech(\widehat{u}_t \widehat{u}_t')$, subsequent estimation of the identified models, and implementation of a diagnosis process, leaded us to accept that a BEKK(1,1,1) model is an appropriate choice for all product categories.

Step 4. Preliminary BEKK model estimation:

The estimation of the volatility model, similar to that of a univariate GARCH. We denote by θ the parameter vector of the model, the matrices $w = w(\theta)$, $A_j = A_j(\theta)$, $B_j = B_j(\theta)$ are functions of θ (in practice the components of θ are precisely the entries in these matrices). The parameters θ can be estimated by conditional pseudo maximum likelihood, i.e. minimizing

$$-2T \cdot L(\theta) = \sum_{t=1}^{T} \left(\ln |H_{t,\theta}| + (X_t - \mu_t)' \; H_{t,\theta}^{-1} \; (X_t - \mu_t) \right).$$

Results for the asymptotic properties of the estimator have been studied by Jeantheau (1998) and Comte and Lieberman (2003). In order to simplify its computation, once we have estimated μ_t using (3.3), we replace μ_t by $\widehat{\mu}_t$ in the likelihood function. This estimation is consistent, but inefficient as it is based on inefficient OLS estimations for the VAR model.

Step 5. Simultaneous estimation of the VAR and BEKK parameters to improve efficiency:

We consider the estimated parameters from VAR(1) and BEKK(1,1) models as initial values, and use them in the full likelihood function to estimate all parameters together in order to achieve asymptotic efficiency. Therefore, including the parameters in μ_t in the vector θ we minimize

$$-2T \cdot L(\theta) = \sum_{t=1}^{T} \Big(\ln |H_{t,\theta}| + \left(X_t - \mu_{t,\theta}\right)' \; H_{t,\theta}^{-1} \; \left(X_t - \mu_{t,\theta}\right) \Big),$$

using the Newton-Raphson method from the preliminary estimators. Using Step 2 and 4 estimations as initial point is crucial for ensuring convergence given the high computational effort caused by the large dimension of the parameters.

Step 6. Inference analysis:

We applied the analysis described above to the full vector of sales and marketing mix actions (price, promotion) for all the selected leader brands on each of the six categories. In all cases, the estimations are globally significant, and their signs and magnitudes are as expected.

The dimension of the tables with the estimators is too large, and we do not report them in detail (they can be provided from the authors upon request). The dynamic structure of the volatility can be visualized using appropriate impulse response functions. Notice that we can expand (3.2) as

$$h_t = (I - \beta(1))^{-1} w + \Psi_{p,q}(L) \;\; vech\left(u_t u_t'\right).$$

where

$$\Psi_{p,q}(L) = (I - \beta(L))^{-1} (\alpha(L)) = \sum_{j=1}^{\infty} \Psi_j L^j,$$

the coefficients $\{\Psi_j\}$ can easily computed, they can be interpreted as an impulse-response function explaining the effect of previous unexpected changes of $vech\left(u_t u_t'\right)$ over current covolatility levels h_t. In the particular case of a BEKK(1,1,1) we have

$$\begin{aligned} h_t &= (I - B_1)^{-1} w + \Psi_{1,1}(L)\ \ vech\left(u_t u_t'\right) \\ \Psi_{1,1}(L) &= (I - B_1 L)^{-1} A_1 L = \left(\sum_{j=0}^{\infty} B_1^j L^j \right) A_1 L = \sum_{j=0}^{\infty} B_1^j A_1 L^{j+1}, \end{aligned}$$

so that

$$\begin{aligned} \Psi_0 &= 0, \Psi_1 = A_1,\ \Psi_2 = B_1 A_1 \\ \Psi_j &= B_1^{j-1} A_1 = B_1 \Psi_{j-1}, \quad j \geq 2. \end{aligned}$$

Inversion of the *vech* operator leads to an infinite BEKK expansion. Figures depicting coefficients in the matrices Ψ_j provide a visual description of the volatility (or covolatility) transmission of random shocks. Some of these graphs are shown in the main results section.

Furthermore, we can obtain much more insightful features from the conditional maximum likelihood estimations by testing conditional independence and Granger causality hypotheses. Consider a partition of X_t two groups of variables X_1 and X_2, then we can study the crossed effects between the different parts. In particular, we study the exogeneity and the independence of marketing mix (price and promotions) and sales within the context of a brand. A similar analysis is carried out for several competitors (e.g. the crossed relationship between sales of a brand and marketing mix of a competitor). From the VAR model, $X_t = \Pi X_{t-1} + u_t$, we partition in two blocks:

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix}.$$

If we only consider **mean-dependence**, it is sufficient to test some of the following hypotheses:

- If $\Pi_{12} = 0$ holds (i.e., Π is block-triangular) with Π_{21} significant, then there is

Granger causality from X_{1t} to X_{2t}.

- If $\Pi_{12} = 0$, $\Pi_{21} = 0$ holds, (i.e., Π is block-diagonal), then X_{1t} and X_{2t} are independent conditionally on the past.

In order to test $H_0 : \Pi_{12} = 0$, (e.g. to test Granger causality of the marketing mix of a brand on the sales of the same/other brand) one can consider a Wald test

$$T \cdot vec(\widehat{\Pi}_{12}) \quad \left[Var(vec(\widehat{\Pi}_{12}))\right]^{-1} \quad vec(\widehat{\Pi}_{12}) \rightarrow \chi^2_k$$

where $Var(vec(\widehat{\Pi}_{12}))$ is the block- component (1,2) of the Maximum Likelihood Estimator covariance matrix (we estimate this matrix using a standard HAC estimator), and k is the number of tested parameters in Π_{12}. Tests for conditional independence are analogous, using estimators for Π_{12} and Π_{21}, and it can be applied for example to test independence between brands. The standard causality tests, including the presented Wald test consider just in-mean effects. Note that with conditional heteroskedasticity, the standard Granger causality tests cannot be used as the concept involves causality in both mean (VAR) and variance (BEKK) equations.

Note that any test based on the parameters of Π ignores the conditional variance dependencies. Under volatility patterns, we should also pay attention to the conditional covariance model, testing if the appropriate parameters in $\widetilde{A}_j$ and $\widetilde{B}_j$ in (3.1) are zero. Consider, for example, the matrix $\widetilde{A}_1$. If the sub-matrix $\widetilde{A}_{12} = 0$ the conditional variance of X_{1t} does not depend on X_{2t} which is a requirement for exogeneity. This is obvious computing the symmetric matrix

$$\begin{pmatrix} \widetilde{A}_{11} & 0 \\ \widetilde{A}_{21} & \widetilde{A}_{22} \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ u_{2,t-1} \end{pmatrix} \begin{pmatrix} u_{1,t-1} \\ u_{2,t-1} \end{pmatrix}' \begin{pmatrix} \widetilde{A}_{11} & 0 \\ \widetilde{A}_{21} & \widetilde{A}_{22} \end{pmatrix}'$$

and noticing that the element $\widetilde{A}_{11} u_{1t-1} u'_{1t-1} \widetilde{A}'_{11}$ does not depend on X_{2t}, and analogously for the coefficients in the matrices $\widetilde{B}_j$. Therefore, an exogeneity Wald test in this context is given by,

$$T\left(vec(\widehat{\Pi}_{12}), vec(\widehat{\widetilde{A}}_{12}), vec(\widehat{\widetilde{B}}_{12})\right)' \left[Var\left(vec(\widehat{\Pi}_{12}), vec(\widehat{\widetilde{A}}_{12}), vec(\widehat{\widetilde{B}}_{12})\right)\right]^{-1}$$
$$\left(vec(\widehat{\Pi}_{12}), vec(\widehat{\widetilde{A}}_{12}), vec(\widehat{\widetilde{B}}_{12})\right).$$

If both $\widetilde{A}_{12}$ and $\widetilde{A}_{21}$ are zero there is block independence between the conditional variance of X_{1t} and X_{2t}. Therefore, for testing full conditional independence the Wald test is a quadratic form including the estimators

$$\left(vec(\widehat{\Pi}_{12}), vec(\widehat{\Pi}_{21}), vec(\widehat{\widetilde{A}}_{12}), vec(\widehat{\widetilde{A}}_{21}), vec(\widehat{\widetilde{B}}_{12}), vec(\widehat{\widetilde{B}}_{21})\right).$$

If we are not interested in mean effects, but just in co-volatility, we would compute a Wald test with the estimators $\left(vec(\widehat{\widetilde{A}}_{12}), vec(\widehat{\widetilde{A}}_{21}), vec(\widehat{\widetilde{B}}_{12}), vec(\widehat{\widetilde{B}}_{21})\right)$.

Summarizing, we consider total independence (exogeneity) test for all the parameters, a **mean**-independence (exogeneity) test using the VAR parameters, and a **variance**-independence (exogeneity) test using the BEKK parameters.

3.6 Mean-variance crossed-effects of marketing mix and sales

Upon our conditional maximum likelihood estimation for the complete model with conditional mean (VAR model) and variance (BEKK model), we compute the Wald tests, and discuss the results of Granger exogeneity[1] for the marketing mix, and the independence[2] tests of marketing mix variables and sales (all measured as logarithmic growth rates).

[1]Exogeneity test example: we test if marketing mix of brand A is independent of its sales (in our context conditional mean and variance do not depend on sales), and not vice versa. Put it differently, causality is one-directional that goes from marketing mix to sales.

[2]Independence test example: we test the block independence between marketing mix and sales of brand A. In other words, neither of them affects the other through expectations or variances.

First, we discuss the results of the exogeneity and independence tests of marketing mix variables and sales for each brand separately, and then across brands. Note that given the observed volatility, all standard parametric inferences based on VAR model would be erroneous, as their usual tests do not account for conditional heteroskedasticity.

3.6.1 Within-brand Analysis

For each brand, we particularly test the exogeneity of the marketing mix of the brand from its sales, and the independence between the sales and the marketing mix of the same brand. The results show that, for all brands in all categories (laundry detergent, toilet tissue, toothpaste, paper towel, cheese and refrigerated juice), **we do reject** the exogeneity of marketing mix hypotheses and we **do** also **reject** the strongest conditional independence hypotheses with a 95% of confidence (meaning that for each brand, the empirical evidence supports that sales means and variances depend on previous sales and marketing mix actions, and vice versa marketing mix actions are set based on previous sales and marketing actions). Table 3 contains a summary of these tests.

Table 3. Within-brand Wald tests analysis

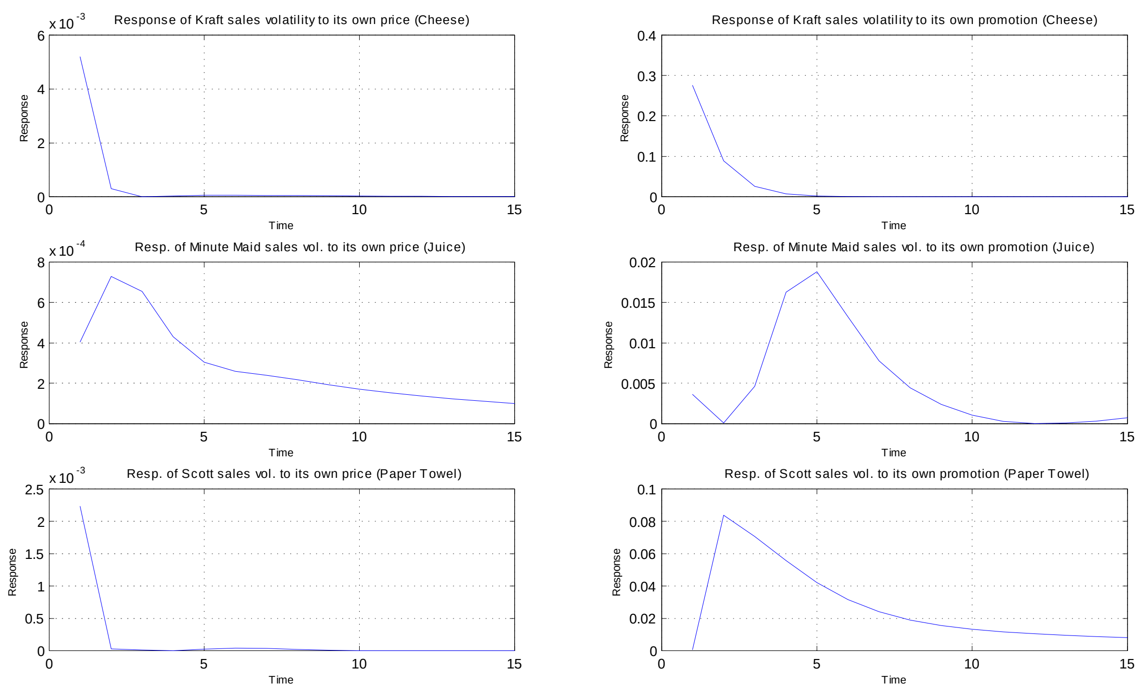
Category	Brand	Null hypothesis	Wald Test	d.f.	Chi-square Critical Value at 5%
Cheese	Dominicks	Marketing mix exogeneity	565,0	6	12,59
		Conditional independence	3992,7	12	21,03
	Kraft	Marketing mix exogeneity	441,0	6	12,59
		Conditional independence	5309,2	12	21,03
Refrigerated juice	Minute Maid	Marketing mix exogeneity	7383,7	6	12,59
		Conditional independence	29949,0	12	21,03
	Tropicana	Marketing mix exogeneity	7556,7	6	12,59
		Conditional independence	114770,0	12	21,03
Laundry Detergent	Wisk	Marketing mix exogeneity	1096,0	6	12,59
		Conditional independence	5409,2	12	21,03
	All	Marketing mix exogeneity	1088,0	6	12,59
		Conditional independence	12103,0	12	21,03
	Tide	Marketing mix exogeneity	3064,9	6	12,59
		Conditional independence	10246,0	12	21,03
Toilet Tissue	Scott	Marketing mix exogeneity	513,1	6	12,59
		Conditional independence	2970,7	12	21,03
	Charmin	Marketing mix exogeneity	90,0	6	12,59
		Conditional independence	910,6	12	21,03
	Northern	Marketing mix exogeneity	262,0	6	12,59
		Conditional independence	7584,2	12	21,03
Paper Towel	Bounty	Marketing mix exogeneity	1316,7	6	12,59
		Conditional independence	3574,5	12	21,03
	Scott	Marketing mix exogeneity	2456,4	6	12,59
		Conditional independence	15463,0	12	21,03
	Dominicks	Marketing mix exogeneity	593,3	6	12,59
		Conditional independence	3892,9	12	21,03
Toothpaste	Aquafresh	Marketing mix exogeneity	8689,0	6	12,59
		Conditional independence	20994,0	12	21,03
	Colgate	Marketing mix exogeneity	885,1	6	12,59
		Conditional independence	3335,7	12	21,03
	Crest	Marketing mix exogeneity	965,5	6	12,59
		Conditional independence	2462,6	12	21,03

We have also performed a narrowed version of the analysis to test the conditional independence and exogeneity (*particularized for the conditional mean* and for *the conditional variance* separately). In this setting, all but one null independence hypotheses are rejected when they are carried our just for VAR and for BEKK parameters, in line with the joint tests reported in Table 3, with just an exception: when we focus just on the VAR parameters, we accept the mean-independence between the marketing mix and sales of Northern in the Toilet Tissue category [4.8 (0.308)][3]. Nevertheless, focusing on the BEKK parameters the null hypotheses of volatility independence of Northern's sales and marketing mix is rejected.

[3]The first value is the Wald test statistic and the second value in parenthesis is the corresponding p-value. From now on, we will show the results of the rejected hypotheses in this this format.

In order to display estimation results and to show the impact of a unit shock to a marketing mix element on sales volatility over time, we use the impulse-response analysis. Because of the space limitation we do not provide all of them. As an example, Figure 3 shows the volatility impulse-response function (VIRF) plots for cheese, refrigerated juice and paper towel categories. Notice that for Kraft brand in the Cheese category, increasing price and promotion growth rate has a positive impact on the sales growth volatility although the effect decays in few periods. For Minute Maid brand in the juice category, the effect is longer for prices than for promotions, whereas for Scott paper towel promotions have a longer effect. Recall that all variables are in logarithmic differences, meaning that for the in-levels series the impact is permanent.

Figure 3. Within-brand volatility impulse response functions



3.6.2 Between-brand Analysis

For all categories, we test the exogeneity of the focal brand's marketing mix from the competitors' sales, the independence between the focal brand's marketing mix and the competitors' sales, the exogeneity of the focal brand's marketing mix from the competitors' marketing mix, the independence between the focal brand's marketing mix and the competitor's marketing mix, the exogeneity of the focal brand's sales from the competitors' sales, and the independence between the focal brand' sales and the competitors' sales.

When we consider jointly the VAR and BEKK model parameters in the Wald test, we find **significant crossed effects** for all brands in the all categories. We reject the conditional independence between the sales of all competitors. We also reject the block conditional dependence between sales and marketing mix for all pairs of competitors, see Table 4. If we consider just exogeneity (unidirectional effects), the results are analogous with a few exceptions. For example, in the Cheese category we accept that Dominick's sales are independent from Kraft's sales [4.2 (0.2407)], but the opposite effect is rejected suggesting that Dominick's is a leader and Kraft is a follower in this market regardless of the fact that Kraft average sales are slightly larger (see Table 2). Both use their marketing mix as a competitive tool, since the block-independence between their marketing mix is rejected [6055.3 (0.0001)]. Also, the exogeneity is rejected for any of them. Similar conclusions can be drawn for the other product categories.

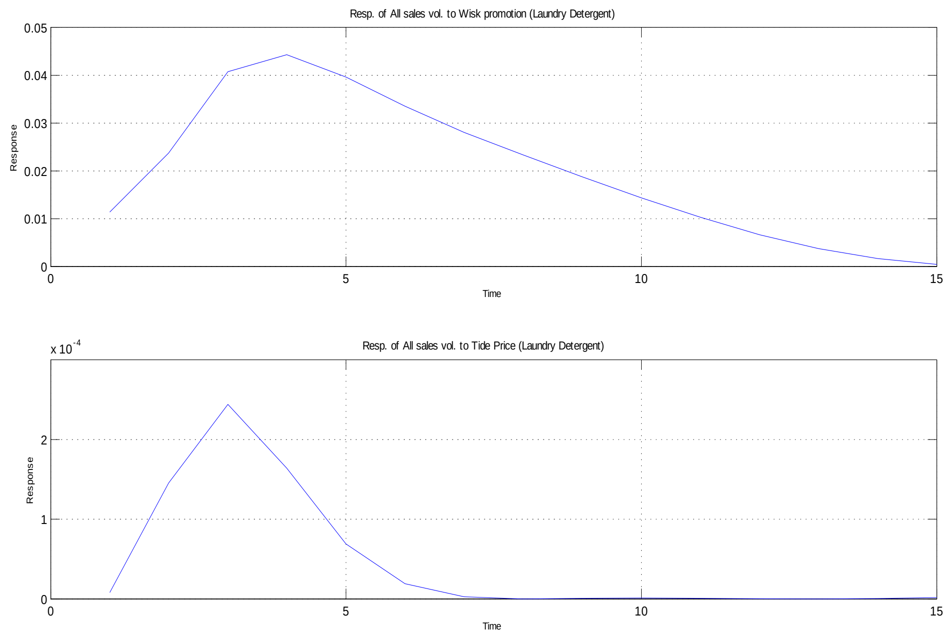
Table 4. Between-brand Wald tests analysis

Category	Null hypothesis (Block conditional independence)	Wald Test	d.f.	Chi-square Critical Value at 5%
Cheese	Between Dominicks marketing mix and Kraft sales	4004,7	12	21,03
	Between Dominicks marketing mix and Kraft marketing mix	6055,3	24	36,42
	Between Dominicks sales and Kraft sales	591,6	6	12,59
	Between Dominicks sales and Kraft marketing mix	2297,7	12	21,03
Refrigerated juice	Between Minute Maid marketing mix and Tropicana sales	7117,9	12	21,03
	Between Minute Maid marketing mix and Tropicana marketing mix	37997,0	24	36,42
	Between Minute Maid sales and Tropicana sales	626,6	6	12,59
	Between Minute Maid sales and Tropicana marketing mix	19314,0	12	21,03
Laundry Detergent	Between Wisk marketing mix and competitors (All and Tide) sales	30939,0	24	36,42
	Between Wisk marketing mix and competitors (All and Tide) marketing mix	88156,0	48	65,17
	Between Wisk sales and competitors (All and Tide) sales	12564,0	12	21,03
	Between Wisk sales and competitors (All and Tide) marketing mix	20854,0	24	36,42
	Between All marketing mix and competitors (Wisk and Tide) sales	610400,0	24	36,42
	Between All marketing mix and competitors (Wisk and Tide) marketing mix	270940,0	48	65,17
	Between All sales and competitors (Wisk and Tide) sales	13794,0	12	21,03
	Between All sales and competitors (Wisk and Tide) marketing mix	31793,0	24	36,42
	Between Tide marketing mix and competitors (Wisk and All) sales	34891,0	24	36,42
	Between Tide marketing mix and competitors (Wisk and All) marketing mix	254000,0	48	65,17
	Between Tide sales and competitors (Wisk and All) sales	6978,8	12	21,03
	Between Tide sales and competitors (Wisk and All) marketing mix	65035,0	24	36,42
Toilet Tissue	Between Scott marketing mix and competitors (Charmin and Northern) sales	14614,0	24	36,42
	Between Scott marketing mix and competitors (Charmin and Northern) marketing mix	100470,0	48	65,17
	Between Scott sales and competitors (Charmin and Northern)sales	3729,0	12	21,03
	Between Scott sales and competitors (Charmin and Northern) marketing mix	14181,0	24	36,42
	Between Charmin marketing mix and competitors (Scott and Northern) sales	533520,0	24	36,42
	Between Charmin marketing mix and competitors (Scott and Northern) marketing mix	88358,0	48	65,17
	Between Charmin sales and competitors (Scott and Northern) sales	2947,3	12	21,03
	Between Charmin sales and competitors (Scott and Northern) marketing mix	15391,0	24	36,42
	Between Northern marketing mix and competitors (Scott and Charmin) sales	12160,0	24	36,42
	Between Northern marketing mix and competitors (Scott and Charmin) marketing mix	119460,0	48	65,17
	Between Northern sales and competitors (Scott and Charmin) sales	918,7	12	21,03
	Between Northern sales and competitors (Scott and Charmin) marketing mix	21128,0	24	36,42
Paper Towel	Between Bounty marketing mix and competitors (Scott and Dominicks) sales	19483,0	24	36,42
	Between Bounty marketing mix and competitors (Scott and Dominicks) marketing mix	104160,0	48	65,17
	Between Bounty sales and competitors (Scott and Dominicks) sales	2310,5	12	21,03
	Between Bounty sales and competitors (Scott and Dominicks) marketing mix	27106,0	24	36,42
	Between Scott marketing mix and competitors (Bounty and Dominicks) sales	33632,0	24	36,42
	Between Scott marketing mix and competitors (Bounty and Dominicks) marketing mix	142900,0	48	65,17
	Between Scott sales and competitors (Bounty and Dominicks) sales	5243,2	12	21,03
	Between Scott sales and competitors (Bounty and Dominicks) marketing mix	15374,0	24	36,42
	Between Dominicks marketing mix and competitors (Bounty and Scott) sales	39155,0	24	36,42
	Between Dominicks marketing mix and competitors (Bounty and Scott) marketing mix	97228,0	48	65,17
	Between Dominicks sales and competitors (Bounty and Scott) sales	4168,0	12	21,03
	Between Dominicks sales and competitors (Bounty and Scott) marketing mix	16443,0	24	36,42
Tooth Paste	Between Aquafresh marketing mix and competitors (Colgate and Crest) sales	10639,0	24	36,42
	Between Aquafresh marketing mix and competitors (Colgate and Crest) marketing mix	42411,0	48	65,17
	Between Aquafresh sales and competitors (Colgate and Crest) sales	5946,1	12	21,03
	Between Aquafresh sales and competitors (Colgate and Crest) marketing mix	34889,0	24	36,42
	Between Colgate marketing mix and competitors (Aquafresh and Crest) sales	135270,0	24	36,42
	Between Colgate marketing mix and competitors (Aquafresh and Crest) marketing mix	56618,0	48	65,17
	Between Colgate sales and competitors (Aquafresh and Crest) sales	3864,9	12	21,03
	Between Colgate sales and competitors (Aquafresh and Crest) marketing mix	9563,3	24	36,42
	Between Crest marketing mix and competitors (Colgate and Aquafresh) sales	20942,0	24	36,42
	Between Crest marketing mix and competitors (Colgate and Aquafresh) marketing mix	72844,0	48	65,17
	Between Crest sales and competitors (Colgate and Aquafresh) sales	5374,4	12	21,03
	Between Crest sales and competitors (Colgate and Aquafresh) marketing mix	18314,0	24	36,42

We have also narrowed the analysis to just mean or just variance dependence. Most conditional independence and exogeneity tests for volatility are rejected in all categories with few exceptions. Reciprocally, if we only focus on the conditional mean parameters, most conditional independence and exogeneity tests are also rejected which is well established in the sales response models literature. The volatitlity analysis can shed some additional insights. For example, in the cheese category (in spite of the fact that we rejected that the sales of brand Kraft are independent from Dominick's sales), if we focus just on the volatility parameters we accept it [2.2 (0.3329)] (we reject for mean [204 (0.0001)]). This indicates that the leadership of Dominick's matters in terms of volatility rather than average patterns. To sum up, the competitive effects are transmitted either through mean or variance, but usually both effects are relevant.

We can depict some between-brands effect using volatility impulse-response functions. For example, Figure 4 shows VIRF plots for laundry detergent category. Notice that a unit shock to the promotion change of the brand Wisk leads to increase in the sales growth volatility of the brand All, and a unit shock to brand Tide's price growth rate generates an increment on the sales growth volatility of All. Since all variables are in logarithmic differences, for the in-levels series the impact is permanent. An emergent conclusion is that promotional actions can be used to increase sales volatility of a competitor, which eventually can lead to a cost increment, and therefore to a competitive advantage. But, aware competitors could apply a similar strategy. This suggests that some commercial wars could be triggered by co-volatilities, rather than by the effects on average sales.

Figure 4. Between-brand co-volatility impulse response functions



3.7 Strategic recommendations

Managers are aware that not all operational decisions work equally well in every setting. What is right for a brand sales depends on its circumstances. We have developed a potential solver framework for understanding the interplay between marketing mix activities and sales response. Managers should be able to identify whether their actions can generate potential logistic risks. Each decision must be assessed for each contribution to company strategy, accounting for the expected results and risks.

Depending on the specific retail database and product category, we can specify different type of VAR and BEKK models. For these models one can compute the IRF. For example, in a VAR(1) $X_t = \Pi X_{t-1} + u_t$ as considered here, the IRF are implicit in the Wold expansion $X_t = \left(\sum_{j=0}^{\infty} \Pi^j L^j\right) u_t$. Also, we have discussed how to compute the

VIRF. Summing up the coefficients of these functions, i.e., evaluating the polynomial lag filters at $L = 1$, we can compute the aggregate impact over time between a specific marketing mix instrument and sales, which is also known as the filter gain in the time series literature.

Looking at the effectiveness of marketing mix to control sales (e.g. looking at the gains but also looking at the economic value of sales gains), we can identify and discuss 4 strategic scenarios:

	Sales mean up	**Sales mean down**
Volatility up	*risk ambivalence*	*worst-case*
Volatility down	*best-case*	*risk smoother*

In other words,

1. *Risk ambivalence*: "To greed, all nature is insufficient" (Seneca), when both sales-mean and volatility are increased by marketing actions. In this no-risk no-gain context, if managers are too greedy in terms of expected sales, they may end up with a high risk at stake.

2. *Risk smoother*: "An honest tale speeds best, being plainly told" (William Shakespeare, Richard III Quote, Act IV, Scene IV), when both are decreased by marketing actions. Plain decent sales might perform better than volatile fancier ones.

3. *Worst-case scenario*: "Abandon all hope ye who enter here" (Inferno, Dante), when mean is decreased and volatility is increased by marketing actions.

4. *Best-case scenario*: "When you want something, all the universe conspires in helping you to achieve it" (The Alchemist, Paulo Coelho), when mean is increased and volatility is decreased by marketing actions.

To succeed firms must cultivate business analytic capabilities to anticipate and adapt quickly to market changes, and to control the random distribution of their market sales.

Whatever the scenario is, managers should optimize their results (or at least minimize their loss), as a general rule, managers should balance the positive effects of higher expected sales, and the negative effects of higher volatility. A reasonable criteria for short-term oriented managers would be maximizing a conditional mean-variance utility function, using the VAR and BEKK models as inputs. For example, if we simply estimate a firm model including firm own sales X_{1t} and $X_{2t} = (p_t, i_t)'$ are marketing instruments price and promotion intensity respectively, then a myopic decision maker maximizes

$$E_t\left[(p_t - c)\, X_{1t} - i_t\right] - \gamma Var_t\left[(p_t - c)\, X_{1t} - i_t\right] = (p_t - c)\, \mu_{1t} - i_t - \gamma\,(p_t - c)^2\, h_{1t},$$

where $\gamma \in (0, 1)$ is the risk penalty parameter. Notice that μ_{1t}, h_{1t} depends on previous sales and marketing interventions p_{t-1}, i_{t-1}. Substituting the model for conditional mean and variance, and setting stable decisions $p_t = p_{t-1}$ and $i_t = i_{t-1}$ the optimal marketing interventions can be computed easily.

3.8 Conclusion

Sales data often have a high level or temporal aggregation which disguises their volatility. The use of relatively short time aggregation windows, such as weekly, daily, and even hourly for internet sales, allows marketers to capture short term fluctuations impacting production and stock management. In turbulent markets, it is possible to find volatility even with data aggregated over larger time windows, such as monthly and quarterly sales. A closer analysis of sales volatility may lead to better management of distribution and supply chain relationships, creating long-term competitive advantages for marketers.

In this paper, we analyze the presence of volatility in weekly retail sales and marketing mix data. We build a VAR model for the conditional mean and a BEKK model for the conditional variance. Based on the estimated parameters, we study conditional independence and exogeneity using Wald tests. We observe significant dependence in all categories for most brands, either in mean, variance or both. The volatility impulse response analysis shows the impact of marketing mix changes (price or promotions) over sales growth volatility, either for own marketing mix or a rival's action. One possibility to alleviate the sales growth variability could be to lower the rate of change in promotional intensity. Also, the retailer may choose more stable price policy because price fluctuations may result in stockpiling behavior of the customers which in turn leads to sales volatility (Lee et al., 1997).

A managerial implication of this research is the fact that marketing mix (at least, price and promotional actions) can be a useful tool for product and brand managers to curb volatility for smoothing out eventually the Bullwhip effect at the retail source level. Lower price and promotional growth rates lead to less volatility in sales growth. Managers should balance the positive effects on expected sales, and the negative effects on volatility. The article complements the work by Hanssens (1998) in which better expected sales data forecasts is proposed as an instrument to handle Bullwhip effects.

3.9 References

Aiyer, Sundi, Gabi Ledesma (2004), "Waste Not, Want Not," *Logistics Today*, 45(4), 35–37.

Baba, Yoshihisa, Robert F. Engle, Dennis F. Kraft and Kenneth F. Kroner (1991), "Multivariate Simultaneous Generalized ARCH," University of California, San Diego: Department of Economics, *Discussion Paper*, 8957.

Bo, Hong (2001), "Volatility of Sales, Expectation Errors, and Inventory Investment: Firm Level Evidence," *International Journal of Production Economics*, 72, 273-283.

Bollerslev, Tim (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307-327.

Cachon, Gerar P., Taylor Randall and Glen M. Schmidt (2007), "In Search of the Bullwhip Effect," *Manufacturing & Service Operations Management*, 9(4), 457-479.

Capgemini Consulting (2011),The 2011 Global Supply Chain Agenda: Market and Demand Volatility Drives the Need for Supply Chain Visibility. http://storage.vuzit.com/public/2larq/The_2011_Global_Supply_Chain_Agenda.pdf, accessed October 5th, 2011.

Comte, F. and O. Lieberman (2003), "Asymptotic Theory for Multivariate GARCH Processes," *Journal of Multivariate Analysis*, 84, 61-84.

Dekimpe, Marnik G. and Dominique M. Hanssens (1999), "Sustained Spending and Persistent Response: A New Look at Long-term Marketing Profitability," *Journal of Marketing Research*, 36, 397-412.

Dekimpe, Marnik G. and Dominique M. Hanssens (2000), "Time-series Models in Marketing: Past, Present and Future," *International Journal of Research in Marketing*, 17, 183-193.

Edge Research Inc. (2010), Manufacturing & Logistics Industry Survey: Supply Chain Volatility, http://www.sterlingcommerce.com/PDF/Sterling_Commerce_Report.pdf, accessed October 4th, 2011.

Engle, Robert F. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflations," *Econometrica*, 50, 987-1007.

Engle, Robert F. and Kenneth F. Kroner (1995), "Multivariate Simultaneous Generalised ARCH," *Econometric Theory*, 11, 122-150.

Fischer, Marc, Hyun Shin, Dominique M. Hanssens, "Marketing Spending and the Volatility of Revenues and Cash Flows", *Working Paper,* UCLA Anderson School of Management, April 2012.

Forrester, Jay Wright (1961), *Industrial Dynamics*, MIT Press.

Hanssens, Dominique M. (1998), "Order Forecasts, Retail Sales, and the Marketing Mix for Consumer Durables, *Journal of Forecasting*, 17, 327-346.

Holt, Charles C., Franco Modigliani, Muth John F. and Simon Herbert A. (1960), *Planning Production, Inventories, and the Work Force*, Prentice-Hall, New Jersey.

Jeantheau, Thierry (1998), "Strong Consistency of Estimators for Multivariate ARCH Models," *Econometric Theory*, 14, 70-86.

Lee, Hau L., V. Padmanabhan and Seungjin Whang (1997), "Information Distortion in a Supply Chain: The Bullwhip Effect," *Management Science*, 43 (4), 546-558.

Lee, Hau L., V. Padmanabhan and Seungjin Whang (2004), "Ten Most Influential Titles of “Management Science’s” First Fifty Years," *Management Science*, 50 (12), 1875-1886.

Pauwels, Koen, Jorge Silva-Rosso, Shuba Srinivasan and Dominique M. Hanssens (2004), "New Products, Sales Promotions and Firm Value: The Case of the Automobile Industry," *Journal of Marketing*, 68(October), 142-156.

Raju, Jagmohan S. (1992), “The Effect of Price Promotions on Variability in Product Category Sales,” *Marketing Science*, 11, 207-220.

Srivastava, Rajendra K., Tasadduq A. Shervani and Liam Fahey (1998), "Market Based Assets and Shareholder Value: A Framework for Analysis," *Journal of Marketing*, 62(1), 2-18.

Sterman, John D. (1989), "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment," *Management Science*, 35 (3), 321-339.

Vakratsas, Demetrios (2008), "The Effects of Advertising, Prices and Distribution on Market Share Volatility," *European Journal of Operational Research*, 187 (1), 283-293.

Chapter 4

US advertising expenditure trends: long-run effects and structural changes with new media introductions

4.1 Introduction

The battle among different media channels severely takes place in the market and trends in advertising media channels is of great interest to both practitioners and researchers. A substantial number of studies on advertising trends are conducted or sponsored by media firms or advertising agencies (Tellis and Tellis, 2009). These agencies try to convince the firms to advertise on their channels. This induces a severe competition among the media channels in the market, particularly between the old traditional such as newspapers, magazines, radio and the new entrants such as internet, cable and yellow pages.

Marketing managers of big-size companies look closely at the media channel expenditures to foresee the long-term trends in the advertising industry by following the reports of key organizations such as the Internet Advertising Bureau (IAB), the Newspaper Association of America (NAA), and research companies such as Kantar Media Intelligence, eMarketer and AC Nielsen. This information influences marketing managers strategies to allocate their advertising budget. For instance, if they expect a global upward trend in internet advertising, they will be inclined to spend more on that channel rather than the one that shows a downward trend, say, newspapers. When the majority of the companies cut their advertising budget, this may result in a considerable drop in the overall aggregated advertising spending (Deleersnyder et al., 2009). There is a vast body of academic research on advertising at the company level, but the research at the aggregated macroeconomic level is less extensive. The aggregated advertising expenditure has a stable ratio with respect to real GDP (see Van der Wurff and Bakker, 2008). Recently, more research has emerged to examine the relationship between business cycles and advertising spending at more disaggregated level, TV, magazine, newspaper and print (see Deleersnyder et al. 2009).

Historically, U.S. marketers have increased exponentially their total advertising expenditure. Simultaneously, they have been rebalancing their budget media shares to match audience shifts when new media were introduced. Perhaps, the most recent example is the growing trend on internet's share which is partially replacing radio, newspapers and magazines ad expenditures. But these phenomena often happened in the past as we can observe in several historical landmarks.

- The first newspaper in US appeared in Boston in 1690. Since them Newspaper growth continued unabated until the first third of the XX century. Between 1890 to 1920, the period known as the "golden age" of print media, William Randolph Hearst, Joseph Pulitzer, and Lord Northcliffe built huge publishing empires. From the 1920s, radio broadcast increasingly forced Newspapers to re-evaluate their business, and the same happened in the 1950s when TV broadcast exploded onto the media scene. During the second part of the XX century Newspaper circulation dropped, and the ad expenditure budget show this impact.

- The first radio broadcast was in 1906, but its golden age in the US spans from the early 1920s when the first broadcast licenses were granted until the 1950s when it was replaced by TV as the primary entertainment media. Initially individual radio programs were sponsored by a single business, but gradually started to sell small time allocations to multiple businesses. Commercial advertising was not generalized until the 1940s.

- The TV business started out in the 30s but household penetration took-off after the Second World War, evolving slowly into an advertising based business when by Procter & Gamble, Unilever and other companies started to develop ads for Soap Opera's. In the 1950s advertisement time was sold to multiple sponsors. From the 1960s big campaigns featured different mass media channels such as TV, radio and magazine extensive ads.

The historical evidence shows that during the first years after introduction, new media generally have little impact on the ad industry. When this occurs, dramatic changes shake the structure of the advertising industry. When we account for the impact of structural changes, can we still find persistent relationships and a global equilibrium?

Marketing researchers have studied how new product/brand entries and exits in a market affects the competitive setting faced by incumbent companies. For example, Nijs et al. (2001) study the new product introductions as a way to expand permanently the category demand. Fok and Franses (2004) analyze marketing mix effectiveness of incumbents resulting from a new brand introduction. Pauwels and Srinivasan (2004) examine how store brand entry structurally changes the performance of and the interactions among all market players. Moreover, Van Heerde et al. (2004) investigate how the innovative product alters the market dynamics structure. Allowing for multiple breaks at unknown points in time, Kornelis et al. (2008) explores to what extent competitive entry creates structural change in incumbents' revenues. Thus, understanding how the new players in the market affect the incumbents' key marketing metrics as well as the underlying trends in advertising expenditures of different media channels is important for advertising agencies, marketing managers and academics.

In this paper, we address the following questions: Is there any long-run equilibrium relationship among all media channels? If any, which channel(s) responds more and faster to a deviation from this long-run relationship? How sensitive is the total advertising expenditures to the economic conditions in the long-run? More importantly, how are the old media affected by the introduction of new media according to the historical evidence? As Kornelis et al. (2008) discuss, competitive entry might not just be a temporal nuisance to incumbents, but could also fundamentally change the latter's performance evolution. All in all, we study the impact of the new media introduction on the long run equilibrium of the advertising industry.

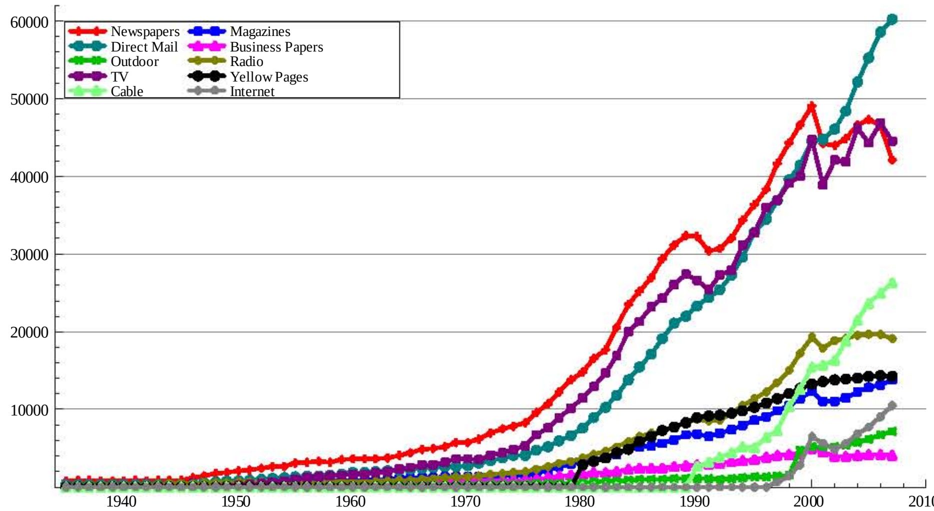
The paper is organized as follows: In the next section, we introduce the data. We study annual time series data ranging from 1935 to 2007 on ten different media channels: newspapers, magazines, direct mail, business papers, outdoor, radio, TV, yellow pages, cable, internet. In section 3, we present a preliminary analysis narrowing down the analysis to the advertising expenditures in newspapers, magazines, direct mail, business papers, outdoor and radio. We model the dynamic interactions among different media channels through Vector Error Correction (VEC) models, capturing the system's gradual adjustment toward a long-run equilibrium (see Dekimpe and Hanssens, 1999). In Section 4, we extend the model to include the advertising expenditures in new media (TV, yellow pages, cable and internet), thus controling for the impact of new media introductions. We take into account the fact that the impact of new entries on a market can produce a persistent structural change (see Kornelis et al. 2008). In section 5, we conclude the paper with the main findings.

4.2 Data

There exists several sources to compile data for the US advertising expenditure. One of the oldest databases is the McCann-Erickson-Magna database. In 1935, L.D.H. Weld, Director of Research for McCann-Erickson and formerly professor of business administration, Sheffield Scientific School, Yale University, published advertising data in the magazine Printers' Ink. Robert J. Coen joined McCann-Erickson in 1948 and two years after Weld died. Coen took up the compilation from 1950 until 2008 when he retired as vice president and forecasting director at the media agency. An early version of this work was published in the Census Bureau (1970, Part II, pages 855-7). The Television Advertising Bureau has made a recent version of Coen's data available online This recent version covers the period from 1948 to 2007. Then, these data were completed by Dr. Douglas A. Galbi, economist at Federal Communications Commission. He added Coen's data to the period from 1919 to 1947. He also included some categories of advertising expenditures for the period spanning from 1919 to 1934. As a result, the final version of the compiled dataset covers the time period of 1919-2007 and contains the advertising expenditures on the following media: newspapers, magazines, direct mail, business papers, billboards, out of home, yellow pages, radio, television, broadcast, cable, internet and total ads. We sum up the ad expenditure on 'out of home' and 'billboards' as the former was the antecedent of the latter, and called the new variable 'outdoor'. We followed the same approach for 'television' and 'broadcast', and called the final variable TV. We also obtained the real GDP variable from the U.S. Department of Commerce, Bureau of Economic Analysis for the period of 1929 and onwards in order to account for the impact of economic crisis and expansions in the advertising industry.

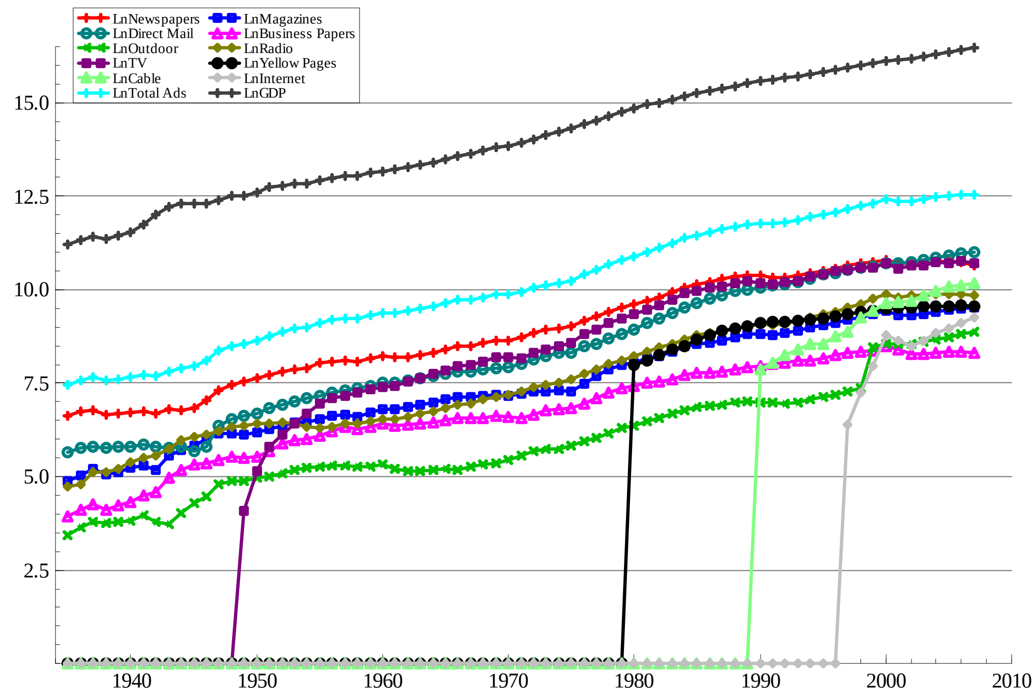
Finally, our dataset is comprised of the following variables: newspapers, magazines, direct mail, business papers, outdoor, radio, TV, yellow pages, cable, internet, total ads and GDP. For the analysis, we choose the time period 1935-2007 so as to have less missing variables in the system of variables. Figure 1 plots the series in their original level. In general, we can observe trends in the series. However, after the year 2000, TV, newspapers and radio advertising spending show a decreasing pattern. By contrast, direct mail, cable and internet advertising spending exhibit an increasing pattern. Outdoor advertising spending shows a sharp increase in 1999 and keeps this increase afterwards.

Figure 1. USA advertising expenditures (in million \$) over time



In this paper, we consider a time series model for $X_t^* = (\ln GDP_t, \ln TA_t, \ln m_t')'$ where TA_t denotes total advertising expenditure, and GDP_t the gross domestic product. The column vector $\ln m_t$ represents logarithms of expenditures on the different media by birth order (newspapers, magazine, direct mail, business papers, outdoor, radio, TV, yellow pages, cable, internet advertising spending in the United States). We use yearly data for the period between 1935 and 2007. Figure 2 shows the variables of the analysis in log-levels. For a given media, observations before the breaking time where the media takes-off are recorded as zeros. We acknowledge that the introduction times can potentially change the cointegration structure.

Figure 2. Log. of USA Advertising and GDP (million \$)



4.3 Preliminary Model

In this section, we consider a preliminary analysis which considers the advertising expenditure (in logarithm) on incumbent media: newspapers, magazines, direct mail, business papers, outdoor and radio, together with GDP and the total advertising expenditure. Within this section, X_t^* will denote this vector of time series. We do not consider yet any potential structural change, but we estimate this simplified model as a first step.

Assume that $\{X_t = X_t^* - \mu_t\}$ is a $\mathbb{R}^k$-valued stochastic time series process, written in deviations from its unconditional mean $\mu_t = E[X_t^*]$, with $X_t^* = 0$ a.s. for $t \leq 0$ (with finite autoregressive models sometimes other specific initial values are considered). The mean $\mu_t \in \mathbb{R}^k$ contains deterministic components (trends, intervention analysis components, etc.). Then, we say that $\{X_t\}$ is **integrated** of order $d \in \{0, 1, 2, ..\}$, also denoted as $I(d)$, if **each coordinate** in $\Delta^d X_t$ follows an invertible stationary linear model, where $\Delta^d = (1-L)^d$ and L is the lag operator ($L^j X_t = X_{t-j}$). Our preliminary analysis suggests that X_t^* is an $I(1)$ process. First, we study the autocorrelation function (ACF) plots as well as the Augmented Dickey Fuller (ADF) unit root tests (for a review of unit root tests in marketing see Dekimpe, Hanssens and Silva-Risso (1999). We observe that the ACFs of all series decay slowly, which is a signal of a nonstationary time series. We perform the ADF tests by adopting two options: (i) only stochastic trend in the series, (ii) both deterministic trend and stochastic trend in the series. We find that tha latter option suits better since the coefficient of the deterministic trend is significant for all series, except for direct mail and outdoor. As can be seen from Table 1, for all variables, we fail to reject the null hypothesis of the ADF test that the series contains a unit root. Thus, the ADF unit root tests support our finding based on correlogram analyis.

Table 1. ADF unit root tests

Variables	*ADF test result (p-value)*		*Conclusion*
	Intercept	*Intercept and trend*	
LnNewspaper	0.7009	0.9584	I(1)
LnMagazines	0.7155	0.5891	I(1)
LnDirect Mail	0.9473	0.1819	I(1)
LnBusiness Papers	0.0993	0.8532	I(1)
LnOutdoor	0.9783	0.7350	I(1)
LnRadio	0.5179	0.7624	I(1)
LnTotal Ads	0.7927	0.8826	I(1)
LnGDP	0.7007	0.5149	I(1)

When $\{X_t^*\}$ is $I(1)$, two possibilities emerge when we look at the whole system: (1) $\{X_t\}$ is **jointly integrated** of order d, that is, it is integrated of order d and $(1-L)^d X_t$ follows an invertible vector Wold process

$$\Delta^d X_t = B(L)\,\varepsilon_t$$

with ε_t white noise (actually ε_t is zero for $t \leq 0$), $B(L) = \sum_{j=0}^{\infty} B_j L^j$ is a matrix-coefficient polynomial with $B_0 = I$ (where invertibility means that the roots of $|B(L)|$ are outside the unit circle, and the process admits a convergent autoregressive representation), or (2) $\{X_t\}$ is **cointegrated** of order d, b with $b \leq d$, and denote it by $C(d, b)$, that is the process is $I(d)$ and there are $r \leq k$ linear combinations defined by the $k \times r$ matrix β such that $\beta' X_t$ is jointly $I(d-b)$. The most important case is $d = b = 1$. The idea goes back to Box and Tiao (1977), but it was popularized by Granger (1981). Cointegrated $C(1,1)$ variables can be expressed with Granger's representation "Vector Error Correction Mechanism" (VECM),

$$\Delta X_t = -\alpha\left(\beta' X_{t-1}\right) + \sum_{j=1}^{\infty} \Gamma_j \Delta X_{t-j} + \gamma + \varepsilon_t,$$

for details see the path-breaking article by Engle and Granger (1987). We include the deterministic trend in the VEC model as $E[\Delta X_t] = \gamma$, based on our preliminary finding from the ADF unit root tests.

To determine the number of cointegrating vectors, we follow Johansen's (1995) approach. More specifically, to determine the rank of the cointegrating matrix β, we adopt the following sequential hypothesis testing. By using STATA-10 and OX version 3.4 (see Doornik, 2001), first, we test the null hypothesis that there is no cointegration against the alternative hypothesis that there is at least one cointegrating vector.

Table 2. Johansen's cointegration test

Maximum Rank	Parameters	Loglikelihood	Eigenvalue	Trace Statistic	Critical Value
0	7	712.0779	.	164.6590	124.2400
1	20	744.1711	0.5900	100.4728	94.1500
2	31	765.4818	0.4468	57.8514*	68.5200
3	40	778.8472	0.3101	31.1205	47.2100
4	47	785.3744	0.1658	18.0662	29.6800
5	52	790.1340	0.1238	8.5470	15.4100
6	55	793.9661	0.1010	0.8827	3.7600
7	56	794.4075	0.0122		

As can be seen from Table 2, we reject the null hypothesis as the trace statistic (164.66) is greater than its critical value (124.24). Next, we test the null hypothesis that there is one cointegrating vector. Again, we reject the null hypothesis since the trace statistic (100.47) is higher than its critical value (94.15). This implies that there might be two cointegrating vectors. Further, we test the null hypothesis that there are two cointegrating vectors. We fail to reject the null hypothesis as the trace statistic (57.85) is less than its critical value (68.52). Therefore, Johansen's cointegration test results show that there are two-cointegrating vectors, i.e. the maximum rank is two. We also tested whether γ is significant using the likelihood ratio test. The statistics is 35.92, distributed under the null hypothesis as a χ^2_8. The corresponding p-value is 0.0000 and the coefficients are found significant.

4.4 Modeling the whole period with structural breaks: Main Results

In this section, we consider the whole dataset including the new media. Notice that TV and yellow pages were introduced in 1949 and 1980 while cable and internet in 1990 and 1997, respectively. As different media channels entered the market at different moments of time, we set the investments on these media before its introduction equal to zero. Therefore, we should consider structural breaks for the whole system, whenever a new media starts to be exploited by the advertising industry. Persistence and cointegration tests can be dramatically affected by the presence of structural breaks. Perron (1989) shows that structural changes can induce apparent unit roots. Structural breaks typically have little effect on the size of the usual cointegration tests, but they affect the power of the tests. There is a large literature studying cointegration under known or unknown structural breaks. Maximum likelihood procedures, as the Johansen test, have better power than the Dickey-Fuller based cointegration tests (see Campos et al. 1996). Johansen test requires modeling the break, but this is less restrictive in our context, where the break time is observed. We use study the impact of the new media introduction on the long run equilibria of the advertising industry.

Let us assume that there are structural changes associated to the introduction of media introductions (TV, yellow pages, cable and internet). Let $T = (T_1, ..., T_k)'$ the time origin 0 and the introduction times of the k different media, and $I(t > 0)$ is the indicator function (equal to 1 if $t > 0$ and zero otherwise). We consider that the introduction times are deterministic (exogenous variables and we condition the process upon their value). The introduction of a new media may cause a permanent structural change in the growth rates of incumbent media (intervention analysis). Therefore, if the system grows at an autonomous vector rate γ until the structural breaks occur, and at a different rate after the entrance of a new media, then we can consider a deterministic component

$\mu_t = E[X_t^*]$ given by

$$\mu_t = \mu_0 + \gamma\ t + \Phi\ \ (tI - T)^+,$$

where $(tI_k - T)^+$ is a vector of shifts with $j-th$ coordinate equal to $\max\{(t - T_j), 0\}$. Notice Φ contain the effect of new media introductions modeled as permanent shifts. Then, for $t \geq 1$,

$$\Delta\mu_t = E[\Delta X_t^*] = \gamma + \Phi\ D_t,$$

where D_t is a deterministic vector of step functions, such that the $j-th$ coordinate is defined as $D_{jt} = I(t \geq T_j)$ taking the value one if $t \geq T_j$ and zero otherwise. We impose some restrictions on the coefficient matrix Φ. It must have a triangular media-structure, as we impose the restrictions that new media introductions in the advertising market does not affect investments on media introduced on the distant future. Therefore, (i) TV introduction cannot cause any structural change in yellow pages since TV enters the market before yellow pages, (ii) TV and yellow pages cannot cause any structural change in cable series because TV and yellow pages enter the market before cable, (iii) TV, yellow pages and cable cannot cause any structural change in internet as internet enter the market after all these channels.

Next, we will follow the Johansen (1991, 1994) framework. Assume that X_t^* follows an integrated $VAR(p)$ vector autoregression

$$\Delta X_t^* = \Pi X_{t-1}^* + \sum_{j=1}^{p} \Gamma_j \Delta X_{t-j}^* + (\gamma + \Phi D_t) + \varepsilon_t, \tag{4.1}$$

where typically $t > p$. The error vectors$\{\varepsilon_t\}$ are assumed to be Gaussian white noise $N(0, \Omega)$. Consider the characteristic lags matrix polynomial

$$A(L) = (1 - L) I_k - \Pi L - \sum_{j=1}^{p} \Gamma_j (1 - L) L^j.$$

If all the roots of the polynomial $|A(L)|$ are outside the unit circle (so that $A(1) = -\Pi$

has full rank), then the process is jointly integrated. However, if there are $(k-r)$ of the roots equal to 1 and the remaining roots are outside the complex unit circle, then $A(1) = -\Pi$ has rank r, and we can express $\Pi = \alpha\beta'$, where α, β are $k \times r$ matrix of rank $r < k$. The VECM representation indicates that the current increment in X_t depends on previous deviations from the long-run equilibrium $\beta' X_{t-1}$, the effect of deterministic components, and previous corrections ΔX_{t-j}

$$\Delta X_t^* = -\alpha \left(\beta' X_{t-1}^*\right) + \sum_{j=1}^{p} \Gamma_j \Delta X_{t-j}^* + (\gamma + \Phi D_t) + \varepsilon_t. \tag{4.2}$$

The parameters $(\alpha, \beta, \Gamma_1, .., \Gamma_p, c, \Phi,\ \Omega)$ are freely varying, but we normalize β to estimate the individual coefficients. The cointegrating rank of the last system is usually determined using Johansen's (1988, 1991, 1995) maximum eigenvalue and trace tests. Johansen also considered the Maximum Likelihood estimators of the full model, and the asymptotic distribution. For details see Johansen et al. (2000) and Hungnes (2010). Pesaran et al. (2000) extend these ideas about deterministic components μ_t to models with exogenous process.

Since we now include all series with zeros before the introduction times, we first update the unit root and cointegration tests with structural changes using Perron (1989) and Johansen's tests. We find that TV, yellow pages, cable and internet variables also contain unit root, and that there are two cointegrating vectors. Therefore, we confirm our preliminary findings. Hence, to capture the short-run dynamics towards the identified long-run equilibrium, we estimate the VEC model with $r = 2$ (two cointegrating vectors) and $p = 1$. The model is estimated by maximum likelihood method using OX version 3.4 and GRaM (see Hungnes, 2005). We run the models up to four lags and compute the AIC and SIC criteria. Both information criteria suggest to use one lag in the final analysis. As in the previous steps of the analysis, we check whether γ is significant. We perform likelihood ratio test. The resulting ratio is 118.99 and distributed as χ^2_{12}. The corresponding p-value is 0.0000 supporting the significance of γ in the final model.

From the estimation output, parameters α are deemed as short-run adjustment parameters whereas parameters β are regarded as long-run equilibrium relationship parameters. Table 3 shows the estimated cointegration vectors (β_1, β_2), and the estimated (α_1, α_2) measuring the response of each variable to deviations from each cointegration equilibrium relationship. Among the old traditional media, the α coefficients for newspapers, direct mail and outdoor are significant at 1%, 1% and 10% level, respectively showing that they respond to the disequilibrium in the industry. Among the new media, TV, yellow pages and cable show significant results at 1%, 1% and 10% level, respectively meaning that they also respond to the disequilibrium in the industry. As can be seen from Table 3, the estimated α for yellow pages in the first cointegration equation (CE1) is higher than that of the rest of the media. Similarly, the estimated α for TV in the second cointegration equation (CE2) is higher than that of the other media. This finding shows us that the advertising investment in these new media is more sensitive to deviations from the long-run equilibrium than those in the other media.

Table 3. Estimated adjustment coefficients α and cointegration parameters β

	CE1	CE2
ΔLnGDP	0.0153	0.0027
	(0.0199)	(0.0098)
ΔLnTotal Ads	0.0015	-0.0446***
	(0.0237)	(0.0117)
ΔLnNewspapers	0.0005	-0.0616***
	(0.0297)	(0.0146)
ΔLnMagazines	0.0036	0.0008
	(0.0359)	(0.0177)
ΔLnDirect Mail	-0.0435	-0.0787***
	(0.0277)	(0.0136)
ΔLnBusiness Papers	0.0558	-0.0036
	(0.0356)	(0.0175)
ΔLnOutdoor	-0.1209*	-0.0111
	(0.0686)	(0.0337)
ΔLnRadio	-0.0003	0.0068
	(0.0290)	(0.0143)
ΔLn TV	0.0271	-0.2160***
	(0.0468)	(0.0230)
ΔLnYellow Pages	-0.1364***	-0.0059
	(0.0081)	(0.0040)
ΔLnCable	0.0006	0.0260*
	(0.0307)	(0.0151)
ΔLnInternet	0.0332	0.0161
	(0.0719)	(0.0354)

Notes: ***, * signs imply that the coefficients are significant at 1% and 10% level, respectively. Standard errors are given in parentheses.

	Vector 1	Vector 2
LnGDP	1.0000	2.8273
LnTotal Ads	-5.3527	1.0000
LnNewspapers	1.3294	-4.6309
LnMagazines	1.6715	-1.9134
LnDirect Mail	0.9422	3.1573
LnBusiness Papers	-0.5296	0.5599
LnOutdoor	1.0112	-0.8769
LnRadio	-0.8504	0.1400
LnTV	0.0994	0.8753
LnYellow Pages	0.4638	0.4887
LnCable	-0.4422	-1.0836
LnInternet	-0.3393	0.9949

In Table 3, the coefficients of vector β_1 can be interpreted as long-run elasticities of advertising expenditures on different media with respect to the GDP. The coefficient of total advertising expenditures is negative (-5.35) implying that in the long-run total ads has a negative elasticity with respect to the economic conditions in the US. Thus, total ads show counter-cyclical behavior, i.e. it rises when the economy is in contraction, and falls when the economy is in expansion. Furthermore, the sensitivity of newspapers and magazines to the GDP are higher than that of all new media, and the other old traditional media. The second vector β_2 is normalized with respect to total advertising, and the ratios between parameters associated to different media can be interpreted as crossed long-run elasticities between investments on different media. For instance, the ratio of $\beta_{2,newspapers}/\beta_{2,internet}$ shows the cross elasticity between newspapers and internet. Focusing on the most recent and the oldest media, from Table 3, we can see that a 1% increase in internet spending results in a drop of 4.7% in newspapers in the long-run. Similarly, a 1% increase in internet spending translates into a 1.9% decrease in the spending level of magazines in the long-run.

Table 4 reports the estimates of Φ, the matrix of of the structural change effects. Focusing on relatively older media, TV and yellow pages, we notice that their entry affects the competitive setting faced by the incumbents inducing a significant level shift for almost all of them (newspapers, magazines, business papers outdoor and radio) and that the sign is negative for all, except for the sign of yellow pages entry on outdoor. Moreover, more recent media, cable and internet, induce significant shift only for one incumbent. Cable causes significant shift for outdoor whereas internet leads to a significant shift only for newspapers. As to the magnitude of the impact, note that the shift caused by TV is larger than that caused by yellow pages, followed by internet and cable.

Table 4. Estimated dummy coefficients

	LnNewspapers	LnMagazines	LnDirect Mail	LnBusiness Paper	LnOutdoor	LnRadio	LnTV	LnYellow Pages	LnCable	LnInternet	LnTotal Ads	LnGDP
D_{TV}	-0.2550***	-0.3377***	-0.0729	-0.1966***	-0.3297***	-0.1728***	3.8079***	0.0000	0.0000	0.0000	-0.1972***	-0.0225
	(0.0416)	(0.0613)	(0.0627)	(0.0698)	(0.0758)	(0.0540)	(0.1035)	(0.0000)	(0.0000)	(0.0000)	(0.0388)	(0.0425)
$D_{Yellow\ Pages}$	-0.1527***	-0.1343**	0.0170	-0.1836***	0.4741***	-0.1004*	-0.1200	7.9000***	0.0000	0.0000	-0.1126***	-0.0149
	(0.0419)	(0.0624)	(0.0634)	(0.0701)	(0.0757)	(0.0540)	(0.1077)	(0.0120)	(0.0000)	(0.0000)	(0.0388)	(0.0425)
D_{Cable}	-0.0008	0.0018	-0.0137	0.0370	-0.1778**	0.0176	-0.1064	0.0221*	7.8635***	0.0000	0.0143	-0.0206
	(0.0443)	(0.0654)	(0.0634)	(0.0735)	(0.0760)	(0.0571)	(0.1077)	(0.0122)	(0.0494)	(0.0000)	(0.0404)	(0.0436)
$D_{Internet}$	0.0793*	0.0967	-0.0394	0.0979	-0.1503	0.0913	-0.1327	0.0169	0.1300*	6.4245***	0.0422	-0.0130
	(0.0440)	(0.0654)	(0.0634)	(0.0736)	(0.1209)	(0.0572)	(0.1082)	(0.0126)	(0.0675)	(0.1601)	(0.0405)	(0.0440)

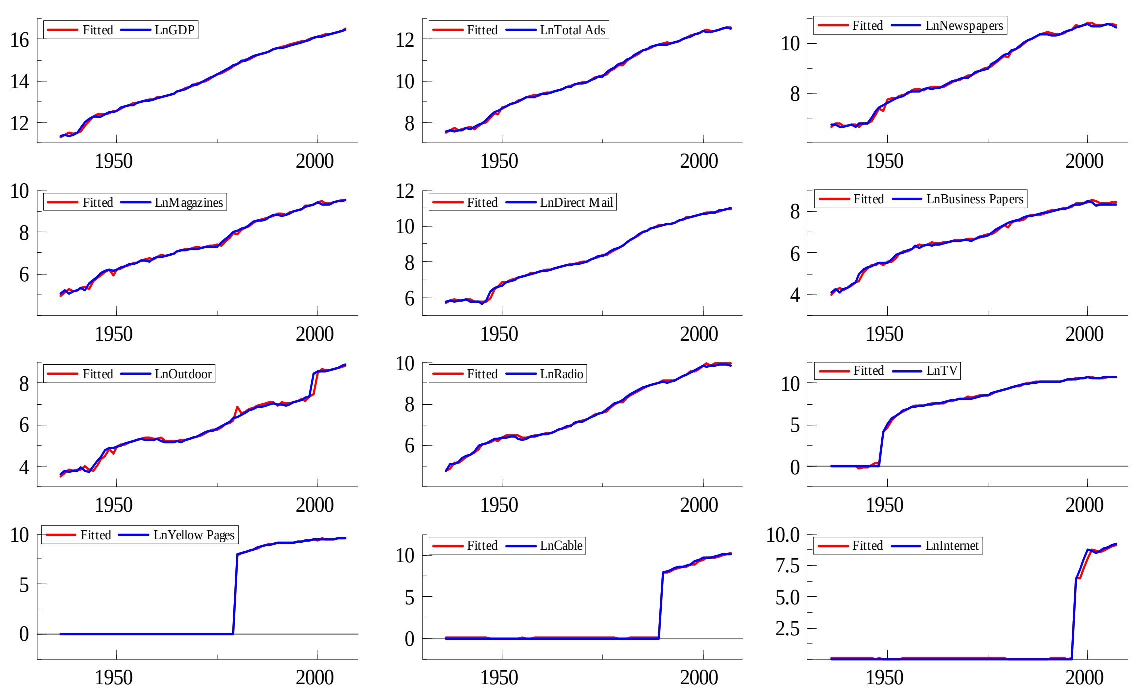
Notes: ***, **, * signs imply that the associated coefficient is significant at 1%, 5% and 10% level, respectively. Standard errors are given in parentheses.

Overall, we can conclude that the incumbents, newspapers, magazines, business papers, outdoor and radio, are vulnerable to the entries of TV and yellow pages. Internet and cable alter fundamentally only the evolution of newspapers and outdoor, respectively.

In addition, we have tested some linear hypothesis on the coefficients of Φ. We have tested if there is a simultaneous level shift structural break in the system after the introduction of the Internet. This requires testing simultaneously if the coefficients of $D_{Internet}$ are equal to zero for all the other variables. After estimating the model with this constraint, we apply likelihood ratio test to test the null hypothesis that there is a level shift in all variables in the system. The resulting ratio is 286.63 and distributed as χ^2_{12}. The corresponding p-value is 0.0000. Therefore, we reject the null hypothesis, i.e. there is a level shift in all variables after the internet enters the industry. This finding shows us how powerful the introduction of the internet is over the incumbents.

As can be seen from Figure 3, our model predicts well the system dynamics.

Figure 3. Actual versus predicted series



4.5 Conclusions

In this paper, we empirically study whether the entries of new advertising media affect the incumbents' expenditure level in the form of creating fundamental change in the long-run evolution. We use the annual time series data on ten different advertising media channels in the U.S. at the aggregate level and build Vector Error Correction (VEC) model with break dummies. Our proposed methodology allows for modeling both short- and long-run dynamics among the variables and takes into account multiple structural breaks that occur at the entry times of TV, yellow pages, cable and internet advertising media.

Our results show that there is a long-run equilibrium among the series identified by the two cointegrating vectors. In addition, recent media, TV and yellow pages, respond faster to the disequilibrium. This implies that they are more reactive towards the disequilibrium more than any other media channel.Furthermore, we find that the long-run elasticity between total advertising expenditures and the GDP is negative. Our finding supports that total ads show counter-cyclical behavior, i.e. it increases when the economy weakens, and decreases when the economy strengthens. This result is in line with the argument that advertising behaves counter-cyclically in countries high on long-term orientation (see Deleersnyder et al. 2009). Moreover, the long-run cross elasticity between the most recent media, internet, and the oldest media, newspapers, is negative suggesting that an increase in internet investment results in a decrease in newspapers investment. The same finding holds for the cross elasticity between internet and magazines.

The main focus of this paper is given on whether the entry of new media changed substantially the spending level of the old traditional media. Our results indicate that the TV and yellow pages entries create fundamental (structural) change in the spending levels of the incumbents and that their effect is negative for almost all incumbents. In other words, they fundamentally changed the incumbents' evolution by making them permanently alter their spending levels. In addition, internet and cable cause substantive shift only on the evolution of newspapers and outdoor, respectively. Moreover, the shift caused by TV is larger than that caused by yellow pages, followed by internet and cable.

By imposing restrictions on the coefficients of internet dummy for all equations at the same time, we also find that internet has a profound impact on the subsequent evolution of advertising media industry in the US.

4.6 References

Box, G.E.P. and G.C. Tiao (1977), "A Canonical Analysis of Multiple Time Series," *Biometrika* 64, 355–365.

Campos, Julia, Neil R. Ericsson and David F. Hendry (1996), "Cointegration Tests in the Presence of Structural Breaks," *Journal of Econometrics*, 70(1), 187-220.

Dekimpe, Marnik G. and Dominique M. Hanssens (1999), "Sustained Spending and Persistent Response: A New Look at Long-term Marketing Profitability," Journal of Marketing Research, 36(4), 397-412.

Dekimpe, Marnik G., Dominique M. Hanssens and Jorge M. Silva-Risso (1999), "Long-run Effects of Price Promotions in Scanner Markets," *Journal of Econometrics*, 89, 269-291.

Deleersnyder, Barbara, Marnik G. Dekimpe, Jan-Benedict E.M. Steenkamp and Peter S.H. Leeflang (2009), "The Role of National Culture in Advertising' s Sensitivity to Business Cycles: An Investigation Across Continents," *Journal of Marketing Research,* 46 (October), 623-636.

Doornik, Jurgen A. (2001), *Object-Oriented Matrix Programming Using Ox*, 4th ed. London: Timberlake Consultants Press and Oxford.

Engle, Robert F. and Clive W. J. Granger (1987), "Co-integration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55(2), 251-276.

Fok, Dennis and Philip Hans Franses (2004), "Analyzing the Effects of a Brand Introduction on Competitive Structure Using a Market Share Attraction Model," *International Journal of Research in Marketing*, 21(2), 159-177.

Granger, Clive W. J. (1981) "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*,16, 121-130.

Hungnes, Håvard (2005), "Identifying the Deterministic Components in Cointegrated VAR Models using GRaM for Ox Professional- User Manual and Documentation," *Working Paper.*

Hungnes, Håvard (2010), "Identifying Structural Breaks in Cointegrated Vector Autoregressive Models," *Oxford Bulletin of Economics and Statistics*,72(4), 551-565.

Johansen, Soren (1988), "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*,12, 231-254.

Johansen, Soren (1991), "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59, 1551-1580.

Johansen,Soren (1994), "The Role of the Constant and Linear Terms in Cointegration Analysis of Nonstationary Variables," *Econometric Reviews*, 13(2), 205-229.

Johansen, Soren (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, New York.

Johansen, Soren and Katarina Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169-210.

Johansen, Soren, Mosconi Rocco and Bent Nielsen (2000), "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend," *Econometrics Journal*, 3, 216–249.

Kornelis, Marcel , Marnik G. Dekimpe, Peter S.H. Leeflang (2008), "Does Competitive Entry Structurally Change Key Marketing Metrics?", *International Journal of Research in Marketing*, 25, 173–182.

Nijs, Vincent R., Marnik G. Dekimpe, Jan-Benedict E. M. Steenkamps, Dominique M. Hanssens (2001), "The Category Demand Effects of Price Promotions," *Marketing Science*, 20(1), 1-22.

Pauwels, Koen and Shuba Srinivasan (2004), "Who benefits from store brand entry?," *Marketing Science*, 23(3), 364-390.

Perron, Pierre (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometrica*, 57, 1361-1401.

Pesaran, M.Hashem, Yongcheol Shin and Richard J. Smith (2000), "Structural Analysis of Vector Error Correction Model with Exogenous I(1) Variables," *Journal of Econometrics*, 97, 293-343.

Tellis, Gerard J. and Kethan Tellis (2009), "A Critical Review and Synthesis of Research on Advertising in a Recession," *Journal of Advertising Research,* 49(September), 304-327.

U.S. Bureau of the Census (1975), Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition, Parts 1 and Part II, (House Document Series, No 93-78), US department of Commerce. Washington, D.C.

Van der Wurff, Richard and Piet Bakker (2008), "Economic Growth and Advertising Expenditures in Different Media in Different Countries," *Journal of Media Economics*, 21, 28–52.

Van Heerde, Harald J., Carl F. Mela and Puneet Manchanda (2004), "The Dynamic Effect of Innovation on Market Structure," *Journal of Marketing Research*, 41(2), 166-183.