Using Multi-market Information

to Improve Understanding of Firm and

Consumer Behavior

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in Management

by

Paulo Antonio Costa Albuquerque

2006
The dissertation of Paulo Antonio Costa Albuquerque is approved

___________________________________
Daniel Ackerberg

___________________________________
Andrew Ainslie

___________________________________
Charles Corbett

___________________________________
Bart J. Bronnenberg, Committee Chair

University of California, Los Angeles

2006
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>ix</td>
</tr>
<tr>
<td>Vita, Publications and Presentations</td>
<td>x</td>
</tr>
<tr>
<td>Abstract</td>
<td>xi</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2: A Spatio-Temporal Analysis of the Global Diffusion of ISO 9000 and ISO 14000</td>
<td>4</td>
</tr>
<tr>
<td>2.1. Introduction</td>
<td>4</td>
</tr>
<tr>
<td>2.2. Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>2.3. Research questions</td>
<td>9</td>
</tr>
<tr>
<td>2.4. Modeling the global diffusion process</td>
<td>13</td>
</tr>
<tr>
<td>2.4.1. Features of the diffusion model</td>
<td>13</td>
</tr>
<tr>
<td>2.4.2. The multi-country diffusion model</td>
<td>14</td>
</tr>
<tr>
<td>2.4.3. Cross-country influence</td>
<td>15</td>
</tr>
<tr>
<td>2.4.4. Country heterogeneity</td>
<td>19</td>
</tr>
<tr>
<td>2.4.5. Omitted variables</td>
<td>20</td>
</tr>
<tr>
<td>2.5. Data</td>
<td>21</td>
</tr>
<tr>
<td>2.6. Empirical Analysis</td>
<td>22</td>
</tr>
<tr>
<td>2.6.1. Estimation</td>
<td>22</td>
</tr>
<tr>
<td>2.6.2. Model selection</td>
<td>24</td>
</tr>
<tr>
<td>2.7. Results</td>
<td>25</td>
</tr>
<tr>
<td>2.7.1. The diffusion parameters of ISO 9000 and ISO 14000 certification</td>
<td>25</td>
</tr>
<tr>
<td>2.7.2. The research questions revisited</td>
<td>27</td>
</tr>
<tr>
<td>2.7.3. Cross-country influence and susceptibility</td>
<td>30</td>
</tr>
<tr>
<td>2.8. Conclusion</td>
<td>31</td>
</tr>
<tr>
<td>Chapter 3: Measuring Consumer Switching to a New Brand across Local Markets</td>
<td>40</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>3.1. Introduction</td>
<td>40</td>
</tr>
<tr>
<td>3.2. Contribution and place in the literature</td>
<td>44</td>
</tr>
<tr>
<td>3.2.1. Measuring switching</td>
<td>44</td>
</tr>
<tr>
<td>3.2.2. The need for more information to measure heterogeneity correctly</td>
<td>46</td>
</tr>
<tr>
<td>3.3. Model</td>
<td>50</td>
</tr>
<tr>
<td>3.4. The frozen pizza category</td>
<td>53</td>
</tr>
<tr>
<td>3.4.1. Competitive background</td>
<td>53</td>
</tr>
<tr>
<td>3.4.2. Data description</td>
<td>56</td>
</tr>
<tr>
<td>3.4.3. Outside good</td>
<td>57</td>
</tr>
<tr>
<td>3.5. Estimation</td>
<td>59</td>
</tr>
<tr>
<td>3.5.1. Independence between structural unobservables and prices</td>
<td>59</td>
</tr>
<tr>
<td>3.5.2. Summary statistics of the purchase set size</td>
<td>61</td>
</tr>
<tr>
<td>3.5.3. Objective function</td>
<td>62</td>
</tr>
<tr>
<td>3.5.4. Computing the switching</td>
<td>63</td>
</tr>
<tr>
<td>3.6. Results</td>
<td>64</td>
</tr>
<tr>
<td>3.6.1. Impact of additional moments on capturing heterogeneity</td>
<td>64</td>
</tr>
<tr>
<td>3.6.2. Impact of additional moments on brand switching</td>
<td>66</td>
</tr>
<tr>
<td>3.6.3. Multimarket findings</td>
<td>69</td>
</tr>
<tr>
<td>3.7. Conclusion</td>
<td>72</td>
</tr>
<tr>
<td>Chapter 4: Dealer Competition in the Auto Industry</td>
<td>82</td>
</tr>
<tr>
<td>4.1. Introduction</td>
<td>82</td>
</tr>
<tr>
<td>4.2. Literature review</td>
<td>84</td>
</tr>
<tr>
<td>4.3. Model</td>
<td>87</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>4.3.1. Demand</td>
<td>88</td>
</tr>
<tr>
<td>4.3.2. Supply</td>
<td>90</td>
</tr>
<tr>
<td>4.4. Data</td>
<td>91</td>
</tr>
<tr>
<td>4.5. Estimation</td>
<td>95</td>
</tr>
<tr>
<td>4.5.1. Demand parameters</td>
<td>95</td>
</tr>
<tr>
<td>4.5.2. Demand instruments</td>
<td>96</td>
</tr>
<tr>
<td>4.5.3. Supply parameters</td>
<td>98</td>
</tr>
<tr>
<td>4.5.4. Supply instruments</td>
<td>99</td>
</tr>
<tr>
<td>4.6. Results</td>
<td>99</td>
</tr>
<tr>
<td>4.6.1. Demand</td>
<td>99</td>
</tr>
<tr>
<td>4.6.2. Supply and dealer competition</td>
<td>101</td>
</tr>
<tr>
<td>4.6.3. Optimal location of a new dealership</td>
<td>103</td>
</tr>
<tr>
<td>4.7. Conclusion and Future Research</td>
<td>105</td>
</tr>
<tr>
<td>References</td>
<td>115</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 2.1: Map of the spatial sample       36
Figure 2.2: Yearly ISO 9000 and ISO 14000 certification counts for 4 countries       36
Figure 2.3: Actual and fitted new ISO 9000 certifications for 1997 and actual and predicted new ISO 9000 certifications for 2000       37
Figure 2.4: ISO 9000 and ISO 14000 certifications resulting from various diffusion sources       38
Figure 2.5: Comparison of the diffusion coefficients between ISO 9000 and ISO 14000 and between early- and later-adopting countries       38
Figure 2.6: Influence and susceptibility values for a sample of 32 countries       39
Figure 3.1: Choice probabilities for brands in two different scenarios: (a) strong and heterogeneous preferences; (b) tight and weak brand preferences       78
Figure 3.2: Evolution of market shares of main brands in Sacramento and Salt Lake City at the time of introduction of DiGiorno       79
Figure 3.3: Perceptual map for models with and without additional information about the purchase set size, for the Los Angeles market       80
Figure 3.4: Boxplot graphs of (a) switching to DiGiorno and (b) pre-entry share of incumbent brands, for 64 markets       81
Figure 4.1: Histogram of the traveled distance by consumers to buy a car across three segments, luxury SUV, lower midsize and compact SUV       108
Figure 4.2: Geographical distribution of Toyota dealers that sell compact SUVs and fullsize SUVs in California 109
Figure 4.3: Actual and estimated shares for the segments luxury SUV and lower midsize cars, for 2004 110
Figure 4.4: Example of the variability of own-price elasticity across distance for two Honda dealers, in two different markets 111
Figure 4.5: Cost and mark-up of dealers in the lower midsize segment, classified by brands, in 2003 112
Figure 4.6: Cross-elasticities between two Hyundai dealers and the remaining retailers, in the lower midsize segment 113
Figure 4.7: Estimated profit of a new Hyundai dealership, selling in the miniSUV and lower midsize segments, for a sample of zipcodes in the Los Angeles 114
LIST OF TABLES

Table 2.1: Log Marginal Likelihoods for all model specifications in the ISO 9000 and ISO 14000 data sets
33

Table 2.2: Mean Squared Errors for holdout predictions across all model specifications in the ISO 9000 and ISO 14000 data sets
34

Table 2.3: Parameter estimates for the ISO 9000 and ISO 14000 data sets
35

Table 3.1: Evolution of average shares for some brands in the frozen pizza category in each of the years in the data set
74

Table 3.2: Results for generated data for models without added information about purchase set size (unrestricted model) and with added information (restricted model)
75

Table 3.3: Results for the Los Angeles market using the model with traditional moments and with additional moments from purchase set size
76

Table 3.4: Results for the linear regression, with dependent variable the switching measured as a percentage of DiGiorno's share
77

Table 4.1: Average price between 2001 and 2004, for each car segment
106

Table 4.2: Segment size (quantities sold) between 2001 and 2004, for each car segment
106

Table 4.3: Number of dealers in California between 2001 and 2004, for each car segment
107

Table 4.4: Parameters' mean and standard deviation for the year 2003
107
ACKNOWLEDGEMENTS

I am extremely grateful to everyone who has helped me throughout my studies at UCLA Anderson School of Management. First, I would like to thank my committee members, Dan Ackerberg, Andrew Ainslie and Charles Corbett, for all the valuable comments and suggestions on my dissertation.

Many thanks are due to all the other professors who helped me during my doctoral education, seminars and presentations, and provided encouragement and a supporting environment. I would also like to mention the valuable discussions and precious hours spent in the company of my fellow doctoral students, Garrett Sonnier and Zainab Jamal.

Funding from the Portuguese government, through the Foundation for Science and Technology and the European Social Fund, and the UCLA Center for International Business Education and Research are gratefully acknowledged.

The first application paper included in this dissertation benefits from the contribution of Professors Bart Bronnenberg and Charles Corbett, while Bart Bronnenberg is a co-author in the remaining two.

A very special thank you to my advisor, Bart Bronnenberg for his guidance and support. I could not have asked for a better mentor.

I would also like thank my parents and my brother. Although they are a continent and an ocean away, it seems that they have always been with me, carrying me forward.

Finally, I am immensely grateful to my wife, Nancy Hsieh, who through these past five years has had the patience to put up with my busy hours while making my life so full and complete.
VITA

July 2, 1975    Born, Lisbon, Portugal

1998    B.A., Management
        Portuguese Catholic University
        Lisbon, Portugal

1998-1999    Trade Marketing Manager,
              Sales Department
              Warner Lambert, Portugal

1998-2001    Teaching Assistant
              Business and Economics Department
              Portuguese Catholic University

PUBLICATIONS PRESENTATIONS


Albuquerque, Paulo and Bart J. Bronnenberg, “Measuring Consumer Switching to a New Brand across Local Markets,” paper presented at 2005 Summer Marketing Educators’ Conference
ABSTRACT OF THE DISSERTATION

Using Multi-market Information to Improve Understanding of Firm and Consumer Behavior

by

Paulo Antonio Costa Albuquerque

Doctor of Philosophy in Management

University of California, Los Angeles, 2006

Professor Bart J. Bronnenberg, Chair

Recently, datasets about managerial variables with an associated spatial dimension have become available to researchers. Accessibility to such data has improved our understanding of how marketing phenomena are distributed and potentially correlated across space. This dissertation presents three applications of such occurrences and shows that the use of spatial data that spans across multi-markets (countries, cities, zip codes) is essential to a complete analysis of consumer and manager behavior.
Our first project focuses on global diffusion of ISO 9000 and ISO 14000 certification using diffusion models that incorporate the spatial and temporal characteristics of cross-country diffusion processes in a unified framework. The central questions in our analysis are (1) whether diffusion of certification is subject to cross-country influences, (2) if so what the nature of that influence is, and (3) how the diffusion process of ISO 9000 differs from that of ISO 14000. We propose a diffusion model which includes several possible cross-country contagion effects and then use Bayesian methods for estimation and model selection.

The second application deals with the introduction of a new brand in a mature category and the growth of its demand in a mature category of consumer package goods. Specifically, we are interested in decomposing a new brand's sales into that derived from cannibalization, competitive draw, and category expansion. We use data at the local market level to estimate a random coefficients logit demand model.

Finally, in our third undertaking, we study the case of spatial competition among car dealerships in Southern California. The main objective of this paper is to present a method that identifies the set of most direct competitors of a car dealer, across locations (zip codes) and across different segments in which it competes. We do so by computing and analyzing own- and cross-elasticities of price across distance, across segments and across brands.

These three applications show that ignoring or not accounting for the impact of decisions of consumers and managers located in other markets will bias and limit the understanding of important managerial phenomena. In future research, given the
increasing availability of such kind of data, it is imperative that researchers take into account multi-market effects in order to better capture managerial reality.
1 Introduction

This dissertation includes three main projects concerning the influence of multi-market information on managers and consumers. In the first chapter, we study the global diffusion of ISO 9000 and ISO 14000 certification using diffusion models that incorporate the spatial and temporal characteristics of cross-country diffusion processes in a unified framework. The central questions in our analysis are (1) whether diffusion of certification is subject to cross-country influences, (2) if so what the nature of that influence is, and (3) how the diffusion process of ISO 9000 differs from that of ISO 14000. We propose a diffusion model which includes several possible cross-country contagion effects and then use Bayesian methods for estimation and model selection. Using country by year data for 56 countries and 9 years, we find that accounting for cross-country influences improves both the fit and the prediction accuracy of our models. The inferred cross-country contagion mechanism is different across the two standards. Diffusion of ISO 9000 is driven primarily by geography and bilateral trade relations, whereas that of ISO 14000 is driven primarily by geography and cultural similarity. We compare the diffusion parameters across ISO 9000 and ISO 14000 and across early- and later-adopting countries and discuss the results in the context of existing work on global diffusion. We show how our approach can be used to evaluate the relative importance of each country in driving global diffusion.

The second chapter studies how a new brand generates demand in a mature
category. Specifically, we are interested in decomposing a new brand’s sales into that derived from cannibalization, competitive draw, and category expansion. We use data at the local market level to estimate a random coefficients logit demand model. Next we use counterfactuals to decompose brand switching in each market. The decomposition is identified from aggregate data by imposing that various quantities at the individual level match the observed market level data across both consumers and time. A novel aspect of our method, and one which helps in estimating the heterogeneity of consumer tastes, is that the estimation imposes that the inferred consumer purchase set size matches easy to obtain summary statistics of the purchase set sizes. Using data across 64 geographically distributed markets in the United States, we focus on the introduction of a new product, DiGiorno, in the Frozen Pizza Category. we find that adding information about the purchase set size improves considerably the ability of the model to capture preference heterogeneity correctly. At a more substantive level, the results are indicative that DiGiorno was successful at targeting consumers from outside of the category, who represented half of DiGiorno’s share. we also find that local pre-entry share, marketing strategies and the brand name explain a large part of the variation of the switching across markets.

Finally, in the third chapter, we study the case of spatial competition among car dealerships in Southern California. The main objective of this paper is to present a method that identifies the set of most direct competitors of a car dealer across different segments in which it competes. We do so by computing and analyzing own-
and cross-elasticities of price across distance, across segments and across brands. The elasticities are obtained from a structural model of demand and supply that accounts for spatial competition and price endogeneity. We make use of a rich transactional dataset that contains valuable information about the location of dealers and consumers and about wholesale prices that dealers pay to the manufacturer’s and retail prices that they charge to the consumer. We find that high-end segments display the least elastic demand, while low-end segments have very high price elasticities. The findings show that the most direct competitive "threat" comes from dealers carrying the same brand or brands that are perceived as very close competitors, even when these dealers are located at longer distances than other competitors. That is, consumers are willing to travel considerable distances to obtain a better price or a specific brand. This gives rise to true spatial competition between dealers.

With our approach, we are able to provide guidance on the optimal location of a new dealership.

These three applications show that ignoring or not accounting for the impact of decisions of consumers and managers located in other markets will bias and limit the understanding of important managerial phenomena. In future research, given the increasing availability of such kind of data, it is imperative that researchers take into account multi-market effects in order to better capture managerial reality.
2 A SPATIO-TEMPORAL ANALYSIS OF THE GLOBAL DIFFUSION OF ISO 9000 AND ISO 14000

2.1 Introduction

Managers in many leading firms are increasingly concerned about practices in place at their suppliers and other trading partners, for several reasons. Poor management practices at suppliers can lead to poor quality of incoming products, which in turn will cause problems for the firm and its own downstream customers. Firms are also concerned about unreliable shipments from suppliers with poor internal practices, and about the potential damage to a firm’s reputation if its suppliers do not follow appropriate environmental or social practices.

These concerns, combined with the difficulty involved in verifying suppliers’ internal practices, led to the emergence of, at first, the ISO 9000 series of quality management systems standards, later followed by various comparable standards, and by standards for issues that are not immediately related to quality, such as the ISO 14000 environmental management systems standard.

These management standards are intended to be adopted globally but, partly due to their relative youth, little is known about how they diffuse across countries.

---

1 Some variations of the ISO 9000 series include the QS 9000 standard for the automotive industry, TL 9000 for the telecommunications industry, and AS 9100 for the aerospace industry. In the food industry, the HACCP (Hazard Analysis and Critical Control Point) system is increasingly becoming a requirement, imposed by governments or customers. The UN recently backed development of the ISO 26000 guideline for social responsibility.
Specifically, little is known formally about the mechanisms underlying cross-country diffusion, about how global diffusion of ISO 9000 compares to that of ISO 14000, and about how diffusion differs between early and later-adopting countries. This paper aims to fill this gap by studying the spatial and temporal mechanisms involved in the global adoption of management standards. We propose and estimate a diffusion model, in which cross-country influences can follow: (1) geography, where adoption spreads to neighboring countries, (2) trade, where adoption spreads to exporting countries, and (3) culture, where adoption spreads to culturally similar countries, or (4) any combination of these. We estimate the model using Bayesian methods on data tracking the diffusion of ISO 9000 and ISO 14000 certification across countries and years. The results indicate that cross-country influence is very strong and important for ISO 9000 and ISO 14000 certification. Interestingly, however, the underlying mechanisms are different. Diffusion of ISO 9000 follows bilateral trade flows and geographic proximity, while for ISO 14000 certification, cultural similarity is also important. We also find that ISO 14000 diffuses faster than ISO 9000 and that both standards diffuse faster in later-adopting countries.

This paper aims to make three contributions. First, allowing for multiple diffusion mechanisms adds to our understanding of the spatial and temporal paths by which management practices diffuse across countries. Second, using an integrated model enables us to compare global diffusion of ISO 9000 and ISO 14000 within and across countries. It also allows us to determine the relative importance of each country in driving global diffusion. Third, the extension of the diffusion model to
account for unobserved spatial factors offers an econometric approach to the estimation of global diffusion processes for management practices, which helps deal with the inevitable data limitations, such as short time series and missing data on early certification.

In Section 2, we review relevant literature on management standards and global diffusion. Section 3 formulates research questions on diffusion of ISO 9000 and ISO 14000. Section 4 introduces the temporal and spatial aspects of the model; the data are presented in Section 5. Section 6 covers estimation and model selection. Section 7 focuses on the results and section 8 concludes.

2.2 Literature Review

ISO 9000 and ISO 14000 

Most studies of diffusion of ISO 9000 and ISO 14000 seek to explain firm-level adoption decisions or country-level cumulative certification levels as a function of various firm- or country-level characteristics. Using a cross-sectional logit analysis, Anderson et al. (1999) find that the main reasons for US firms to seek ISO 9000 certification are government requirements, export considerations, quality improvements, and cost reduction. Guler et al. (2002) use an 85-country, 6-year dataset of ISO 9000 certificates and find that countries with higher certification counts receive more investment from abroad and have closer trade links to other countries with high numbers of certifications. Mixing the firm and country perspectives, using a 9-country survey, Corbett (2005) finds that, in later-adopting countries, firms with higher exports adopt earlier.

data on 142 countries to find that certifications per capita are positively correlated with, among others, foreign direct investment and exports to Europe and Japan. Using survey data, Christmann and Taylor (2001) find that exports to more developed countries are associated with higher ISO 14000 certification levels among Chinese firms. Several other studies, not listed here for brevity, adopt variations on these methods. The cross-sectional regression in Corbett and Kirsch (2001) suggests that countries’ export propensity and relative number of ISO 9000 certificates are positively associated with ISO 14000 certification levels.

**Diffusion processes** Mansfield (1961) is perhaps the first to explicitly model the process of technology diffusion, using the well-known logistic function and corresponding S-shaped growth curve. Teece (1980) shows that Mansfield’s (1961) model also describes the spread of an administrative innovation, hence making it applicable to ISO standards. Bass (1969) shows that essentially the same model applies to the diffusion of consumer goods and others have integrated the effects of management action into this model (e.g., Horsky and Simon 1983).

Part of the diffusion literature focuses on global diffusion with two main themes: measuring (1) heterogeneity in country-specific diffusion rates and (2) cross-country contagion. In the former category, Gatignon et al. (1989) study national adoption rates of six consumer durables as a function of underlying country characteristics, such as cosmopolitanism and women in the labor force. Talukdar et al. (2002) investigate diffusion of six durable products across 31 countries, using a Bayesian hierarchical model to allow information from other countries and products to help
predict market potential and coefficients of imitation and innovation in countries in the early stages of diffusion or with a lack of information. In the latter category, Takada and Jain (1991) estimate models for each country separately and find that diffusion is faster in later-adopting countries. Ganesh et al. (1997) model the learning effect between lead and lag markets finding that geographical proximity between them has no effect while economic (cultural) similarity usually (always) strengthens the learning. Neelamegham and Chintagunta (1999) use past adoption in a lead country to improve forecasts for adoption in a lag country. Putsis et al. (1997) find strong support for a “mixing” model in which the diffusion of consumer durable goods in one country is influenced by popularity in other countries. Kumar and Krishnan (2002) combine the lead-lag and mixing models, allowing the cross-country influences to affect both innovation and imitation rates.

Diffusion models are used in these papers to capture installed-base effects on subsequent adoption. Surprisingly, the existing literature on ISO 9000 and ISO 14000 rarely refers to this body of literature, despite the fact that the spread of management standards is probably subject to installed-base effects. In addition, using a diffusion model approach rather than cross-sectional or panel data methods has the following advantages. First, a diffusion model typically focuses on yearly differences in certification counts rather than on their levels. In the ISO 9000 and ISO 14000 certification context, this is appropriate because the time and money involved in obtaining the initial certification is sunk and large compared to the costs of renewing certification. The installed base will therefore predominantly affect the
timing of the initial certification decision, not that of the renewal decision. Second, if most initial certifications are renewed (as is the case for ISO 9000 and ISO 14000), the yearly certification levels are cumulative variables. Newbold and Granger (1974) show that regressions among cumulative variables often yield statistically significant results where none are present. This critique would apply to regressions between certification levels in two countries. The standard solution to this spurious regression effect is to difference the time series, i.e., focus on certification growth rather than levels.

2.3 Research questions

In this section we formulate our research questions on how diffusion of ISO 9000 and ISO 14000 differ from each other and across countries. Existing literature sometimes leads to conflicting predictions, preventing us from formulating unambiguous hypotheses. We start with two questions comparing ISO 9000 and ISO 14000, then a question comparing early- and later-adopting countries.

1. Does ISO 14000 diffuse faster than ISO 9000? To date, the two standards are rarely studied jointly. A reasonable hypothesis would be that ISO 14000 diffuses faster than ISO 9000, as ISO 9000 was the first global management systems standard. As a consequence, firms were uncertain about its value and relevance. ISO 14000, despite its substantially different scope, was no longer an entirely novel concept, and hence might be expected to exhibit higher innovation and imitation rates. Corbett and Kirsch (2001) provide anecdotal examples of how the success of ISO 9000 led government agencies in Japan and Taiwan to be more aggressive about promoting
ISO 14000.

The evidence offered so far by studies of multiple innovations (successive or not) is mixed but on balance favors the prediction that later innovations diffuse faster than earlier ones, consistent with the statement above. Mahajan and Muller (1994) find that the imitation parameter is higher for diffusion in a unified (European Union) market than in individual countries. The globalization that has occurred between the introductions of ISO 9000 (1986) and ISO 14000 (1996) would suggest that ISO 14000 should have a higher imitation rate than ISO 9000. For innovations within the auto industry, Grubler (1991) shows that imitation rates are increasing over time. Based on inspection of 31 electrical household durables in the US over a 74-year period, Van den Bulte (2000) finds that diffusion of new products is accelerating over time. By contrast, Islam and Meade (1997) find that coefficients of innovation and imitation do change across generations of the same technology, but not in a consistent manner: newer generations of cell phones diffuse faster, while newer generations of IBM mainframes diffuse less fast.

2. Does cultural similarity affect cross-country diffusion of ISO 14000 more than that of ISO 9000? ISO 9000, with its focus on quality, has little relevance to other stakeholders than the certified firm, its competitors, and its customer(s). ISO 14000, by contrast, is relevant to communities, NGOs, regulators, and other parties, that need not have any business links with the certified firm. In that sense, ISO 14000 affects a broader set of stakeholders, and hence could reflect a country’s cultural values more strongly than ISO 9000. If so, then countries with similar cultures
would be more likely to influence each other with respect to diffusion of ISO 14000 than of ISO 9000.

The literature supports this view, though often implicitly: studies of ISO 9000 tend to focus more on economic factors, while studies of ISO 14000 tend to include a broader range of non-economic factors. For instance, Guler et al. (2002) and Neumayer and Perkins (2005) consider only economic factors in their studies of diffusion of ISO 9000. Terlaak and King (2005) study the effects of ISO 9000 certification from the perspective of the standard’s ability to signal improved operational performance. By contrast, Neumayer and Perkins (2004) include the number of environmental NGOs in their explanation of diffusion of ISO 14000; Corbett et al. (2003) report that firms adopting ISO 14000 are more motivated by relations with authorities and communities than firms adopting ISO 9000; and Bansal and Hunter (2003) find that early adopters of ISO 14000 in the US already had high environmental legitimacy prior to adopting.

3. *Do ISO 9000 and ISO 14000 diffuse faster in later-adopting countries?* The literature is mixed, but, on balance, suggests that diffusion in later-adopting countries is faster than in early-adopting countries, i.e., that later-adopting countries catch up with earlier-adopting countries.

Takada and Jain (1991) hypothesize and find that the imitation coefficient will be greater for a country in which the product is introduced later. Grubler (1991) and Nakicenovic (1991) report the same occurring with the development of railroad networks in Europe: while in the UK, 100 years passed between the beginning of
railroad construction and attaining maximum network size, in other areas such as Scandinavia, only 50 years passed between those two milestones, indicating that later-adopting countries do catch up with lead countries. Comin and Hobijn (2004) compare diffusion patterns across countries of 25 products spanning 250 years, and find that while economic leaders tend to adopt first, the rate at which lagging countries catch up is accelerating. By contrast, Lücke (1993) estimates separate diffusion models for the spread of process innovations in the textile and steel industries in multiple countries and finds that while economically less developed countries tend to adopt later, the speed of diffusion is similar to more developed countries. Helsen et al. (1993) find that while a longer time lapse between lead and lag country adoption sometimes increases the lag country’s rate of innovation, it always decreases its imitation rate.

A related question concerns the cross-country contagion effects. Later-adopting countries may (or may not) have lower innovation and own-country imitation rates, but they will be more influenced by prior adoption in other countries and hence could exhibit higher cross-country imitation rates. This effect could be exacerbated by the observation that later-adopting countries tend to be less economically developed (Lücke 1993, Comin and Hobijn 2004), more likely to be further upstream in supply chains and hence subject to pressure from a larger number of downstream parties. Ganesh et al. (1997), building on Takada and Jain (1991), predict that a greater lag time between lead and lag country will lead to a stronger learning effect, which could translate to a higher cross-country imitation effect in our model. Corbett
(2005) finds that early-adopting firms in later-adopting countries are more heavily motivated by export considerations, which would also be consistent with a higher cross-country imitation rate.

2.4 Modeling the global diffusion process

2.4.1 Features of the diffusion model

We focus on ISO 9000 and ISO 14000 certification levels by country and by year, particularly on the temporal and cross-sectional aspects of diffusion. Our model has four distinguishing features, each of which can be operationalized in several ways, discussed in more detail below. First, it accounts for differing degrees of cross-country influence, defined as the effect of past certifications in one country on current certification in another. We estimate two versions of the model, where only recent or all past certifications influence current new certifications.

Second, the model allows for alternative views of which countries influence each other. We consider four definitions of “influence sets” of nations based on geography, trade relations, cultural similarity, or a combination of these.

Third, the model includes an econometric control for omitted variables by allowing for contemporaneous correlation of unobserved factors across geography (see e.g. Anselin 1988). For instance, some relevant factors that contribute to the diffusion of certifications may be at the level of economic regions such as the EU or NAFTA. Any such factor that creates a multi-country “trend” in certification will cause contemporaneous correlation when omitted. We specify two models, with and without contemporaneously correlated errors.
Finally, to ensure that what we measure as cross-country influence is not simply a form of unobserved heterogeneity, we allow for random (country-specific) effects. We specify two models, one flexible with random effects and one more restrictive with non-random effects.

These four dimensions of model specification lead to a total design of $2 \times 4 \times 2 \times 2 = 32$ models, each of which is estimated for both ISO 9000 and ISO 14000.\(^2\)

The following subsections focus on each of these four model design dimensions. We present the model in the context of ISO 9000; the models for ISO 14000 are analogous.

### 2.4.2 The multi-country diffusion model

We start by defining a multi-country Bass model where the cross-country imitation effects are based on the cumulative number of certifications in other countries. Here, firms are influenced by other firms, at home and abroad, that have received certification in past periods. We model the number of new certifications $c_{kt}$ in country $k$ at time $t$ as:

$$c_{kt} = \left( p_k + \sum_{k'=1}^{K} q_{kk'} \frac{C_{k',t-1}}{M_{k'}} \right) (M_k - C_{k,t-1}) + e_{kt}, \quad (2.1)$$

where $C_{kt}$ ($c_{kt}$) is the cumulative (incremental) number of certifications in country $k = 1, ..., K$ at time $t = 1, ..., T$. $p_k$ is the coefficient of innovation, $q_{kk'}$ is the contagion effect of past adoption in country $k'$ on current adoption in country $k$, $M_k$ is the potential for the number of certifications and $e_{kt}$ the error term. To allow for heteroskedastic errors, we assume the $e_{kt}$ are normally distributed with mean

\(^2\)Using all combinations of random-effects (yes/no) with omitted-variables control (yes/no), we additionally estimate four specifications without cross-country effects.
zero and variance proportional to the previous year’s growth of certifications.\textsuperscript{3}

\[ e_{kt} \sim N \left( 0, \sigma_e^2 \cdot |c_{k,t-1}| \right) . \]  

(2.2)

The parameters to be estimated are \( p_k, q_{kk'}, M_k \) and the error variance \( \sigma_e^2 \).

An alternative assumption is that a firm only exerts influence on firms in other countries soon after its own certification process. Cross-country contagion is then driven only by recent certifications in other countries. This would occur if managers in different countries are more likely to transmit information about recent or ongoing projects than about older events. Therefore, we also examine a model where the cross-country effects depend on recent rather than cumulative certifications in other countries:

\[ c_{kt} = \left( p_k + q_{kk} \frac{C_{k,t-1}}{M_k} + \sum_{k'=1,...,K; k' \neq k} q_{kk'} \frac{C_{k',t-1}}{M_{k'}} \right) (M_k - C_{k,t-1}) + e_{kt}. \]  

(2.3)

Note that the within-country contagion effects do still depend on cumulative adoption, in order to maintain the structure and spirit of the traditional S-shaped diffusion curve.

### 2.4.3 Cross-country influence

In this section, we define the cross-country effects \( q_{kk'} \), which depend on a country’s “neighbor sets”, i.e., the set of countries that are hypothesized to affect adoption in that country. We define neighbor sets based on geography, trade, and culture.\textsuperscript{4}

\textsuperscript{3}We set \( c_{k,t-1} = 1 \) for the few cases where \( c_{k,t-1} = 0 \). This additive and heteroskedastic residual term is a simple and parsimonious way to account for the empirical patterns in our data: (1) the residuals in countries with more certifications have higher variance, (2) the absolute errors tend to increase over time as certification levels increase, and (3) in very rare occasions, reflecting measurement error and occasional decertification, \( c_{kt} \) can be 0 or negative in the ISO data.

\textsuperscript{4}For a clarifying example of different neighbor sets across these definitions, see the end of this subsection.
Geographic distance. This notion of influence is appropriate if ISO 9000 diffusion follows geographical patterns, for instance starting in Western Europe, then spreading West, South and East from there. For each country $k$, we define the neighbor set $G_k(n)$ as the $n$ geographically closest countries, where distance between countries is measured as the surface distance between country capitals.\(^5\) The contagion effects in equations (2.1) and (2.3) are specified accordingly:

$$q_{k'k}^G = \begin{cases} 
\lambda_k^G & \text{if } k' = k \\
\gamma_k^G & \text{if } k' \in G_k(n) \\
0 & \text{otherwise}
\end{cases}$$

(2.4)

where $k' \in G_k(n)$ means that country $k'$ is among the $n$ closest geographical neighbors of $k$. $\lambda_k^G$ is the usual coefficient of own-country imitation, while $\gamma_k^G$ is the coefficient that measures the strength by which firms in country $k$ imitate firms located in countries $k' \in G_k(n)$. Testing for geographical cross-country influence is equivalent to testing that $\gamma_k^G > 0$. An appropriate value for $n$ is determined empirically on the basis of model fit.

Bilateral trade. If a larger share of country $k$’s exports go to country $k'$, then country $k'$ likely has more influence on adoption in country $k$. We define the share of exports, $BT_{kk'}$, as follows:

$$BT_{kk'} = \frac{\text{Exports}_{kk'}}{\sum_{j=1}^{J} \text{Exports}_{kj}}$$

(2.5)

Country $k$’s neighbor set $B_k(n)$ consists of its $n$ largest export markets, i.e., the $n$

\(^5\)We also tested a continuous distance specification, with $q_{k'k} = \exp(-\gamma_k D_{k'k}^{-1})$, where $D_{k'k}$ is the geographical distance between country $k$ and $k'$. This specification produced worse results in terms of fit.
countries with highest $BT_{kk'}$. The cross-country effects are then operationalized as:

$$q_{kk'}^B = \begin{cases} 
\lambda_k^B & \text{if } k' = k \\
\gamma_k^B & \text{if } k' \in B_k(n) \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (2.6)

Again, $n$ is specified empirically. This definition is not symmetric: for instance, the US is a key export market for many countries which are not major export markets for the US.

**Cultural dimensions** To numerically represent the culture of a given country, we use Hofstede’s (2001) four cultural dimensions: (1) power distance, (2) individualism, (3) masculinity, and (4) uncertainty avoidance. These cultural dimensions have been used in other work on comparing innovativeness across countries. For instance, Shane (1993) finds a link between culture and the number of trademarks per country, while Van Everdingen and Waarts (2003) find that culture affects country-level adoption rates of enterprise resource planning systems. Each country $k$ is represented by four scores $s_{c_{ks}}$, with $s = 1, ..., 4$, one for each of Hofstede’s dimensions. We define the cultural distance between countries $k$ and $k'$ using the distance measure

$$H_{kk'} = \sqrt{\sum_{s=1}^{4} (s_{c_{ks}} - s_{c_{k's}})^2}$$  \hspace{1cm} (2.7)

For each country $k$, $H_k(n)$ contains the $n$ countries culturally closest to $k$. The

---

6These dimensions mean the following: (1) “Power Distance” focuses on the degree of equality vs. inequality between different people, in terms of power and wealth; (2) “Individualism” focuses on the importance given to the individual vs. the collective, in terms of achievements and relationships; (3) “Masculinity” deals with the traditional role played by the man in terms of control, power and achievement; (4) “Uncertainty Avoidance” regards the level of tolerance for uncertainty and risk.
cross-country effects are operationalized as:

\[
q_{kk'}^H = \begin{cases} 
\lambda_k^H & \text{if } k' = k \\
\gamma_k^H & \text{if } k' \in H_k(n) \\
0 & \text{otherwise}
\end{cases}
\]  
\tag{2.8}

with \( n \) determined empirically.

**Combining the neighbor sets**

Finally, neighbor sets can be defined as any combination of the three previous definitions, e.g., the union:

\[
q_{kk'}^A = \begin{cases} 
\lambda_k^A & \text{if } k' = k \\
q_{kk'}^G + q_{kk'}^B + q_{kk'}^H & \text{if } k' \neq k
\end{cases}
\]  
\tag{2.9}

where \( q_{kk'}^G, q_{kk'}^B, \) and \( q_{kk'}^H \) are as defined previously in equations (2.4), (2.6), and (2.8). These cross-country effects contain the parameters \( \gamma_k^G, \gamma_k^B, \) and \( \gamma_k^H \), which are estimated concurrently. The \( \gamma \)'s are identified because for any given \( n \), the sets \( G_k(n), B_k(n) \) and \( H_k(n) \) are not identical. Under the union of these sets, each country \( k \) may thus have more than \( n \) neighbors.

**An example: India**

The countries in our sample that are considered India’s neighbors, assuming \( n = 5 \), under each of the above definitions are:

- geographical distance: Pakistan, United Arab Emirates, Iran, Thailand, Saudi Arabia
- bilateral trade: USA, Japan, United Kingdom, Hong Kong, United Arab Emirates
- cultural: Egypt, Jordan, Saudi Arabia, United Arab Emirates, Kenya
- combined: the union of all 12 countries listed above.

The countries are listed in order of proximity under each measure; it is clear that the neighbor set and ordering of neighbors varies substantially across the different measures.
2.4.4 Country heterogeneity

We use two alternative methods to account for heterogeneity in the country diffusion rates. The first uses country-specific covariates (see e.g. Gatignon et al. 1989; Putsis et al. 1997). We make the parameters $p_k$, $M_k$, $\lambda_k$ and $\gamma_k$ a linear function of country characteristics, $x_k$:

\begin{align*}
M_k &= x_{Mk} \beta_M \\
p_k &= x_{pk} \beta_p \\
\lambda_k &= x_{\lambda k} \beta_\lambda \\
\gamma_k &= x_{\gamma k} \beta_\gamma
\end{align*}

(2.10)\quad(2.11)\quad(2.12)\quad(2.13)

where $x_{pk}$, $x_{Mk}$, $x_{\lambda k}$ and $x_{\gamma k}$ are factors such as population, urbanization, and illiteracy ratings, for each parameter. These expressions can be substituted in equation (2.1) to get a non-random effects model of country heterogeneity. The resulting model captures observed heterogeneity across countries.

An alternative approach accounts for unobserved as well as observed differences across countries through a random coefficients model (e.g., Talukdar et al. 2002), where a hierarchical structure is placed on the parameters, which have a distribution of the following form:

\begin{align*}
M_k &\sim N \left( x_{Mk} \beta_M, \sigma_M^2 \right) \\
p_k &\sim N \left( x_{pk} \beta_p, \sigma_p^2 \right) \\
\lambda_k &\sim N \left( x_{\lambda k} \beta_\lambda, \sigma_\lambda^2 \right) \\
\gamma_k &\sim N \left( x_{\gamma k} \beta_\gamma, \sigma_\gamma^2 \right)
\end{align*}

(2.14)\quad(2.15)\quad(2.16)\quad(2.17)

The first model produces a distribution for the parameter vectors $\beta = [\beta_p; \beta_M; \beta_\lambda; \beta_\gamma]$. The country-specific parameter values then result from equations (2.10) to (2.13). The second model provides distributions of the final parameters $M_k$, $p_k$, $\lambda_k$ and $\gamma_k$ for each country, capturing heterogeneity more flexibly.

We use the following covariates. For the potential number of certifications $M_k$, we choose population and GDP per capita as the main source of heterogeneity.
Larger countries generally have more firms, while richer countries have a higher proportion of firms that can support the costs of ISO 9000 certification. Specifically for ISO 14000, we include the sum of the “social and institutional capacity” and “global stewardship” components of the environmental sustainability index (World Economic Forum, 2002), as a covariate (this measure consists of 35 variables, making it much broader than the measure based on number of environmental treaties used in earlier work).

We use literacy rates as an alternative measure of development. Studies in urban economics (Calem and Carlino, 1991) show that urban centers offer better infrastructure and consequently facilitate faster diffusion of information about adoption of innovations, so we also include percentage of urban population. Empirically, variable selection for each model in (2.10) – (2.17) is based on model likelihood (corrected for overfitting) and prediction.

2.4.5 Omitted variables

A final feature of our model is that it allows for contemporaneous correlation in the residuals $e_{kt}$ across countries. This helps control for those omitted variables that cause multi-country trends in certification, such as business cycles that are common to a neighbor set of countries.

A parsimonious model of contemporaneous correlation can be specified as a spatial correlation on the residuals of equation (2.1). Recall that these residuals are heteroskedastic with variance $c_{kt-1} \cdot \sigma^2_{\epsilon}$. For ease of notation, define the transformed residuals $\tilde{e}_{kt} = e_{kt} / \sqrt{c_{kt-1}}$. Instead of assuming that the error component $\tilde{e}_{kt}$ is independent across countries, we allow for a more general autocorrelated error process on the $[K \times 1]$ error vector $\tilde{e}_t$. That is,

$$\tilde{e}_t = \mu W \tilde{e}_t + \eta_t$$

(2.18)

where $W$ is a $K \times K$ matrix whose elements $W_{kk'}$ are $1/n$ for all $k' \in G_k(n)$ and 0 otherwise (recall $n$ is the number of neighbors). The interpretation of equation (2.18) is that $\tilde{e}_{kt}$ is allowed to be a function of the average $\tilde{e}_{k't}$ in the spatial neighbor set,
where $\mu$ can be interpreted as a spatial autoregressive coefficient and the non-spatial structure component $\eta_{kt}$ is distributed as $\eta_{kt} \sim N (0, \sigma^2_{\eta})$.

To summarize, we have presented a very general model of diffusion of ISO 9000 (and ISO 14000) certification that (1) accounts for cross-country and own-country imitation effects, (2) operationalizes different definitions of cross-country effects, (3) allows for observed and unobserved differences across countries, and (4) accounts for omitted variables with a spatial structure.

### 2.5 Data

Our data for the number of ISO 9000 and ISO 14000 certifications were obtained from ISO, which took over the original Mobil survey of global certification data that started in 1992. Altogether, 12 “cycles” of the global certification survey have been released during 1992-2003 (ISO 2003). Though annual since 1995, the earliest cycles were released at irregular intervals. To transform the early data to an annual grid, we used cubic spline interpolation between sample points. In the year 2000, the major revision ISO 9000:2000 was released. A significant number of firms that were previously certified to the earlier version of the standard did not seek re-certification to the new version, although re-certification is required every three years. Other firms consolidated multiple site-level certifications into single firm-level certification, which was not possible under the earlier versions of the standard. Because neither disadoption nor consolidation is the focus of our study, we discarded the data on ISO 9000 certification after 2000, leaving 9 annual observations per country (1992-2000) for ISO 9000. ISO 14000 was published in September 1996, and did not undergo any major revision. We have 8 annual observations for 1995-2002 (some firms received certification before the final standard was published).

The initial sample includes all countries which have at least one ISO 9000 certification by the year 2002, in total 169 countries. It is impossible to meaningfully estimate diffusion patterns with severely limited data so we restrict our analysis to

---

$^7$Anselin (1988) shows that the joint distribution of the spatial error vector $\tilde{e}$ is normal with mean 0 and variance-covariance matrix equal to $\sigma^2_{\eta} (\mu W' (I - \mu W)^{-1}$
the 59 countries with at least 200 cumulative certifications. We lack information about bilateral trade or country characteristics for Taiwan, Lithuania and Latvia, so our final sample includes 56 countries. For consistency, the same sample was used for ISO 14000. Figure 2.1 gives a graphical representation of the spatial sample.

Of the 56 countries, 31 had less than 20 ISO 9000 certifications in 1993, so in most countries widespread adoption had not yet started. In contrast, for the last year in our analysis ($T = 2000$ for ISO 9000, $T = 2002$ for ISO 14000), the mean cumulative number of ISO 9000 and ISO 14000 certifications is respectively 7,170 and 851, with a standard deviation of 11,664 and 1,622. Figure 2.2 shows that the diffusion curves of selected but typical countries are either S-shaped or convex, consistent with the outcome of a diffusion process.

Economic and demographic information was obtained from the TableBase database (Dialog, 2003), from the CIA World Factbook (CIA, 2003) and from the Census Bureau. We mostly used values for 1997, except for population, which is from the year 2000, and bilateral trade, from 1996, the latter from the World Trade Flows, 1980-1997 (Feenstra, 2000).\(^8\) To make the covariates of consistent scale, they were standardized prior to using them in the diffusion model. Geographic distance was based on the latitude and longitude of the countries’ capitals. The cultural data were taken from Hofstede (2001), where the Arab countries in our sample (Egypt, Jordan, Saudi Arabia and UAE) are treated as one region. For Cyprus, we used Greece’s cultural scores; for Kenya, we applied the scores for Eastern Africa.

### 2.6 Empirical analysis

#### 2.6.1 Estimation

We use Bayesian methods to estimate the parameters’ mean and standard deviation. Bayesian methods allow us to estimate all proposed models, from non-random effects to hierarchical random coefficients within a consistent framework. Our time series are short: for ISO 9000 (ISO 14000), we have only 9 (8) observations per

---

\(^8\)The top exporters do not vary much from year to year, circumventing the need to make this measure time-specific over the observation period.
country. We focus on annual differences in certification rather than levels, so we “lose” one observation. Another observation is needed for initialization in equation (2.3). Finally, we use the last available observation (year 2000 for ISO 9000; 2002 for ISO 14000) as holdout to evaluate prediction. This leaves us with 6 observations per country for model estimation for ISO 9000 and 5 for ISO 14000. Such short time series are not uncommon when studying diffusion processes that are measured infrequently, as is the case of ISO 9000 and ISO 14000. The use of Bayesian methods allows us to make the most of the country-specific data to estimate country diffusion rates. However, for those countries where the within-country data are uninformative of the diffusion process, the method will “shrink” the estimates towards the hierarchical mean, which is based on pooled data across all countries. The amount of shrinkage is determined by the informativeness of the within-country data relative to the pooled across-country data.9

We carried out preliminary tests with different values for neighbor set size \( n \) for each distance measure. The best results in terms of fit and prediction were obtained with small set of neighbors, between 3 and 8, with little sensitivity in that range. To compare the different contagion mechanisms with a manageable number of empirical models, we ran all models with \( n = 5 \). Recall that in addition to the 32 models defined earlier, we defined 4 variants of the baseline model without cross-country effects with and without accounting for unobserved heterogeneity and omitted variables. Our model selection strategy for the two data sets compares all 36 models for each standard.

Figure 2.3 shows the actual number of ISO 9000 certifications compared to the fitted \((T = 1997)\) and the predicted \((T = 2000)\) values from the model with geo-

---

9 Using Markov Chain Monte Carlo (MCMC) methods, we sample from the marginal posterior distribution of each parameter in turn, conditional on the current values of other parameters and on the data (Tanner and Wong, 1987; Gelfand and Smith, 1990). All models were run for a total of 20,000 iterations. The first 15,000 observations were used for initialization and the last 5,000 iterations were used for inference. To reduce the autocorrelation in the MCMC sampler, every fifth draw was saved for analysis. The plots of the sampled values for each parameter show that they converge. We tested for differences in parameters’ means across different intervals (Gelman, Carlin, Stern and Rubin, 1995) and found none significant. In most cases, convergence was achieved after 3,000 to 4,000 iterations.
graphic cross-country effects, random effects to account for heterogeneity, and no contemporaneous correlation in errors. The model is generally very effective in tracking the cross-country heterogeneity in ISO 9000 certification.

2.6.2 Model selection

In order to focus on cross-country effects, we first make a rough selection among the many specifications by comparing them based on marginal likelihood and root of the mean squared error (RMSE). For each model, we calculate the marginal likelihood as the harmonic mean of the posterior likelihood values across iterations of the sampler (Newton and Raftery, 1994; Gelfand and Dey, 1994). The marginal likelihood warrants against overfitting the data, as overfitting increases the variability of the likelihood across iterations which in turn reduces the harmonic mean of the likelihoods. Consequently, models with more parameters, such as the “combined” specification, may present worse marginal likelihood values. The RMSE statistics are estimated from the differences between the predicted vs. actual values for the holdout period ($T=2000$ for ISO 9000, $T=2002$ for ISO 14000) at every sweep of the sampler.

Table 2.1 shows the log of marginal likelihoods for the 36 model specifications for ISO 9000 and ISO 14000. The table shows that, in most cases, the models with long-term (cumulative) cross-country influence provide a better fit than those with short-term influence. Table 2.2 shows the prediction performance for the same 36 models. The maximum prediction accuracy with long-term cross-country influence is slightly better or similar to that of the short-term model. Given the superior fit and (mostly) better prediction results, we retain for further analysis the specifications where adoption depends on cumulative certifications.

We see that the random-coefficients models that account for unobserved heterogeneity produce better fit and substantially better predictions than the non–random effects specifications. This is an important benefit of our Bayesian estimation ap-

\footnote{Consistent with equation (2.2), these differences are scaled by $c_{k,t-1}$. This scaling prevents model selection from being dominated by one or two countries with the largest number of certifications.}
With our short time series, it is difficult to meaningfully account for country-specific diffusion parameters using classical methods. However, our results show that the information in the time series helps make much better predictions. In view of their fit and prediction accuracy, we focus on the model specifications that account for unobserved country heterogeneity through random effects.

Tables 1 and 2 show limited support for the models with contemporaneously correlated errors to account for omitted variables, especially for the random coefficients models. After accounting for unobserved heterogeneity, there appears to be little evidence for contemporaneous correlation in the residuals across countries, so we retain the more parsimonious model without such effects for further analysis.

In sum, the subset of models on which we focus for subsequent analysis of cross-country effects have in common that they (1) are based on cumulative certification levels in other countries, (2) account for unobserved heterogeneity, and (3) have independent errors across countries. In Tables 2.1 and 2.2, the corresponding fit and prediction statistics are underlined for clarity. This subset consists of the best specifications in terms of fit and prediction, or when there is no systematic difference in performance, the specifications that are more parsimonious. The substantive conclusions of our paper do also apply outside this set.

2.7 Results

2.7.1 The diffusion parameters of ISO 9000 and ISO 14000 certification

We now discuss the diffusion parameters of ISO 9000 and ISO 14000 and relate them to what is known about these parameters in other contexts. In the specifications (2.14) to (2.17), the country-specific effects $M_k$, $p_k$, $\lambda_k$ and $\gamma_k$ are modeled as functions of underlying covariates. To prevent overfitting, we base the inclusion of covariates on improvement of the marginal likelihood. For the random coefficients models,\(^1\) the only significant covariate was population (for $M_k$). This is not sur-

---

\(^{11}\) We focus on the random coefficients models because of their fit and prediction superiority. Tests with the non-random coefficients models revealed that GDP per capita and (in the case of ISO 14000) the environmental sustainability index (ESI) were insignificant. While the insignificance of the ESI seems at odds with Neumayer and Perkins (2004), we analyse a different dependent variable...
prising, as the random coefficients model itself already accounts for country-specific effects.

Table 2.3 reports parameter estimates for selected random effects models for both data sets. The potential number of certifications \( M_k \) in country \( k \) is positively related to population size. The innovation coefficient is around 0.04 for ISO 9000 and around 0.05 for ISO 14000. These values are consistent with those found in meta-analyses on diffusion parameters for durable and high technology consumer products. Pooling across more than 200 different applications of consumer durables, Sultan et al. (1990) find an average value for \( p \) of 0.03. This means that the diffusion process for management standards (in essence an industrial durable “good”) has a similar autonomous component as that for a typical durable consumer good.

When cross-country imitation is accounted for, the own-country imitation coefficients are between 0.22 and 0.31 for ISO 9000 and between 0.26 and 0.38 for ISO 14000. Sultan et al. (1990) find an average value of 0.3 for the imitation effect for consumer durables. Note from a comparison of model (1) with models (2)-(6), that the own-country imitation coefficient is inflated by 50\% to 100\% when cross-country imitation is not accounted for. This suggests that any work on multi-country diffusion that does not explicitly consider the cross-country influence may confound within- and cross-country imitation effects and hence overestimate the within-country imitation parameter.

In Table 2.3, the estimated cross-country effect coefficients vary from 0.040 to 0.181. No meta-analysis of cross-national diffusion coefficients exists for us to compare our estimates to. We note that the cross-country diffusion parameters are almost all significant and positive and that the findings of the models are robust to the other empirical specifications from section 4 including those with spatially correlated errors.

To illustrate the differences between countries and to show the flexibility of our model in capturing a rich set of diffusion phenomena, we show the diffusion than they do (certification growth rather than level), and their results indicate that “attitude to the environment” is only weakly significant.
patterns for two countries, the Netherlands and China, and for the two standards. To construct the graph, we use for each standard the model with all cross-country contagion mechanisms, but to avoid a cluttered graph we focus on the contrast between the effects of bilateral trade and cultural similarity.

Figure 2.4 shows that the autonomous component of adoption for both ISO 9000 and ISO 14000 is greater in the Netherlands than in China. In fact, the most developed countries in Europe were the first to adopt the standards and in turn influenced other countries. This is in some part a result of the availability of resources in those countries, which facilitates firms to take risk and be innovators.

In China, adoption of ISO 9000 was initially largely fueled by trade-related pressures. This is to be expected, given that China has become one of the main suppliers of firms in developed countries and downstream pressures are likely to occur. ISO 14000 shows other driving forces beyond trade-relations, as also suggested in the literature (e.g., the strength of relations with communities and other cultural factors). In both China and the Netherlands, the proportion of new ISO 14000 certifications that can be attributed to cultural similarity with other countries is greater than in the case of ISO 9000. The importance of the cultural dimension in the Netherlands for the ISO 14000 standard is explained in part by having environmentally concerned countries such as Denmark, Finland, and Sweden as neighbors.\(^\text{12}\)

2.7.2 The research questions revisited

Our random coefficients procedure yields posterior estimates of all diffusion parameters, i.e., of the market potential \(M\), the innovation effect \(p\), the own-country imitation effect \(\lambda\) and the cross-country imitation effect \(\gamma\), for each country and standard. In Section 3, we speculated how these parameters might be different for ISO 9000 vs. ISO 14000 and for early vs. later-adopting countries. To perform these comparisons, we compute the posterior of the mean of these parameters in a 2 (ISO 9000 vs. ISO 14000) by 2 (early vs. late) design, defining early and late based on

\(^{12}\)For example, these three countries were on the top 10 in the Pilot 2006 Environmental Performance Index (CIESIN, 2006), a study that analyses the success at achieving critical environmental goals.
a median split in penetration of certification in the initial year of our data for ISO 14000.\textsuperscript{13} Figure 2.5 displays the comparisons between the four groups. \textsuperscript{14}

Research question 1: diffusion rate of ISO 9000 vs. ISO 14000. The diffusion process of ISO 14000 certification presents directionally larger average coefficients of innovation, own-country imitation, and cross-country imitation than that of ISO 9000, consistent with the majority of literature reviewed earlier, that is, the later process diffused considerably faster than the earlier one. This allows for two possible explanations: (1) the globalization that occurred between standard introductions had a significant impact in speeding up the diffusion process, leading to a more unified world and increasing imitation rates (Mahajan and Muller, 1994); (2) firms have learned from the diffusion of ISO 9000. This learning could have resulted in a reduction of uncertainty and risk about the economic value of this type of management standard, or in efficiency gains in the certification process due to the significant overlap in procedures and documentation involved in the two standards. This difference in adoption speed across the two standards is however more pronounced for the late adopters group where significance approaches or is above 90\%. The late adopters, usually poorer, developing countries were very slow to adopt ISO 9000 but less far behind in adopting ISO 14000 (see Research question 3).

Research question 2: diffusion mechanisms of ISO 9000 and ISO 14000. Tables 2.1 and 2.2 show that the specifications without cross-country effects always fit and usually predict worse for ISO 9000 and ISO 14000. The empirical evidence therefore clearly points to the presence of cross-country effects in the spread of both standards.

For ISO 9000, the cross-country effect based on geographic distance produces

\textsuperscript{13}The median split on first year ISO 14000 certification levels separates countries that have ISO 14000 certifications in the first year of its availability from countries that did not. Several alternative segmentations based on (1) different lower and upper percentiles of penetration, and (2) timing of first certification yielded substantively similar results. Defining early and late adoption based on ISO 9000 data is more ambiguous due to lack of data for 1986-1993.

\textsuperscript{14}The pooling of data in this 2 by 2 design immunizes the comparison of parameters across cells from small sample biases present in non-linear models (see Van den Bulte and Lilien 1997), even if for each combination of country and standard, the time series are of unequal length across design cells. By pooling the data, the number of observations quickly becomes large enough to wipe out small sample biases. This was verified by means of several Monte Carlo simulations.
superior fit and good prediction, while that based on bilateral trade links provides good fit and superior prediction. However, the cross-country effect defined by cultural similarity performs the worst in terms of fit and especially prediction. Thus, the data suggest that cross-country contagion of ISO 9000 certification follows geography and export relations rather than cultural similarity.

The strong impact of geographic distance suggests that firms in neighboring countries have a greater tendency to observe and share information about management practices. To the extent that this leads to competitive mimicry (Guler et al., 2002), the geographic effect may be seen as “horizontal” contagion (i.e., involving similar firms). The strength of the bilateral trade link is likely the result of firms requiring suppliers to have ISO 9000 certification, regardless of their location. This contagion is among buyer-seller dyads and can therefore be termed “vertical.” Our findings suggest that both types of contagion contribute to global diffusion of ISO 9000 certification.

In contrast, for ISO 14000, the cultural distance specification produces good fit and superior prediction. The bilateral trade specification fits well but produces worse fit. The geographic specification fits the worst. Comparing the two data sets, the role of cultural similarity in cross-country contagion therefore appears larger for ISO 14000. Though tentative, this finding is consistent with our prediction that ISO 14000 is more culturally driven than ISO 9000.

Research question 3: early vs. later-adopting countries. First, and consistent with Takada and Jain (1991), we find that later-adopting countries tend to catch up with early adopters by having significantly higher within-country imitation rates in both ISO 9000 and ISO 14000 (see also Comin and Hobijn, 2004). Second, and specific to the case of ISO 14000, late adopters have a considerably higher rate of innovation, while there seems to be very little difference in innovation behavior between groups for ISO 9000. Early adopting countries exhibit a largely constant adoption behavior across ISO 9000 and ISO 14000, while the later-adopting countries clearly learned from ISO 9000, resulting in faster adoption of ISO 14000.
2.7.3 Cross-country influence and susceptibility

The empirical results suggest that the diffusion of ISO 9000 and ISO 14000 certification takes place within a global network of firms and managers, organized by geography, trade, and/or culture. Here, we use our model to propose two empirical measures of the importance of countries in stimulating certification across borders. These measures, influence and susceptibility, seek to represent how much a country contributes to foreign certification and how much foreign countries contribute to domestic certification, respectively. They are computed from the estimated cross-country influence matrix \( q_{kj}, \{k, j = 1, \ldots, K\} \) (for further discussion in a similar context see Wasserman and Faust 1994). The influence index \( I_{kt} \) is defined as the summed number of certifications that country \( k \) causes in other countries in year \( t \).

\[
I_{kt} = \sum_{j=1,\ldots,K:j\neq k} q_{jk} \cdot \frac{C_{k,t-1}}{M_k} \cdot (M_j - C_{j,t-1}) \cdot \text{untapped firms in } j
\]  

(2.19)

The index \( I_{kt} \) combines two terms. The first term is simply the multiplication of the coefficient \( q_{jk} \), that captures the strength of influence of \( k \) on \( j \), with the cumulative certification base \( \frac{C_{k,t-1}}{M_k} \) in \( k \). This term serves as a measure of the “pressure” exerted by country \( k \) on \( j \). The second term is the “untapped” firms in country \( j \). Similarly, the susceptibility of country \( k \) is equal to

\[
S_{kt} = \sum_{j=1,\ldots,K:j\neq k} q_{kj} \cdot \frac{C_{j,t-1}}{M_j} \cdot (M_k - C_{k,t-1}) \cdot \text{untapped firms in } k
\]  

(2.20)

This measure equals the number of certified firms in country \( k \) and year \( t \) that the model attributes to foreign pressures. Figure 2.6 presents the summed (across time) relative measures of influence, \( \sum_t I_{kt} / \sum_t \hat{c}_{kt} \)\(^{15} \) and susceptibility, \( \sum_t S_{kt} / \sum_t \hat{c}_{kt} \), using the combined specification (6) in Table 2.3. Note that the relative susceptibility index \( \sum_t S_{kt} / \sum_t \hat{c}_{kt} \) is equal to the share of certifications in \( k \) due to foreign

\(^{15}\)In order to compare across countries, we divide the absolute measures by the estimated total of certifications \( \sum_t \hat{c}_{kt} \) in each country, to obtain a relative measure of influence and susceptibility.
pressures from $j \neq k$ (see equation 2.1).\(^\text{16}\)

For ISO 9000, countries like the UK, Japan, USA, etc., appear to have moderate to low susceptibility to cross-national pressures coupled with a large influence on other countries, i.e., the model estimates that for firms in these countries, certification is not driven by firms in other countries, but rather by those at home. Within the confines of the model, these countries owe their influence to their central place in the bilateral trade matrix and/or geography. On the other side of the spectrum reside countries like China and Korea that have low to moderate influence but are quite susceptible. This distribution of countries is consistent with the existence of downstream pressures among trading nations found by Lücke (1993) and Comin and Hobijn (2004).

Comparing this with the bottom graph, we see that environmentally proactive countries such as Sweden, Denmark, and the Netherlands were more influential in the spread of ISO 14000 than of ISO 9000. Overall though, most countries’ relative influence and susceptibility remains broadly similar across ISO 9000 and ISO 14000, suggesting that influence and susceptibility are country-specific rather than innovation-specific.

Interestingly, many countries have a share of certifications attributed to foreign pressures larger than 0.5. For these countries, the majority of domestic certification originates from cross-country influence. For policy makers interested in encouraging rapid diffusion of management practices, our results suggest that focusing on gaining rapid acceptance in a few key influential countries helps accelerate adoption in many other, more susceptible, countries.

### 2.8 Conclusion

We presented a model of international diffusion of management standards and estimated it on country by year data for ISO 9000 and ISO 14000 certification. Our

\(^{16}\)Because many countries are clustered in the region of low influence, we use a log transform of influence, \(\log \left( \sum_{i} I_{kt} / \sum_{i} \hat{e}_{kt} \right)\), in Figure 2.6 to avoid cluttering. For the same reason, the graph only includes the union of the 20% most and least influential countries of both standards.
diffusion model is based on the Bass (1969) model and adds several new elements to it: cross-country dependence, unobserved country differences, three alternative cross-country contagion mechanisms, and omitted variable correction. We believe the class of diffusion models presented here offers advantages over the cross-sectional or panel data models which have been used before in the analysis of ISO 9000 and ISO 14000 certification.

The paper provides a structured empirical comparison of alternative diffusion model specifications intended to account for a wide array of country-specific features and cross-country influences. We infer that cross-country effects are statistically (Table 2.3) and substantively (Figure 2.4) important in both ISO 9000 and ISO 14000 certification. Our findings are also suggestive of the specific nature of the cross-country influence. ISO 9000 certification follows export flows and geographic proximity while ISO 14000 appears to also diffuse across culturally similar countries.

The diffusion parameters estimated are fully consistent and strikingly close to the values obtained from meta-analyses in the diffusion literature (e.g., Sultan et al. 1990), further supporting the use of diffusion models in the context of management practices. The flexibility of our model allows us to show how the nature of diffusion differs across countries and over time and which countries have more or less influence on others. Future work can use our modeling approach to test our findings with other emerging standards.
Table 2.1: Log Marginal Likelihoods for all model specifications in the ISO 9000 and ISO 14000 data sets.

<table>
<thead>
<tr>
<th>ISO 9000</th>
<th>Non-Random Effects</th>
<th>Random Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No omitted</td>
<td>Omitted var.</td>
</tr>
<tr>
<td>No cross-country eff.</td>
<td>-1695.3</td>
<td>-1697.7</td>
</tr>
<tr>
<td>Cross-country effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-term: geography</td>
<td>-1685.5</td>
<td>-1682.6</td>
</tr>
<tr>
<td>Short-term: bilat. trade</td>
<td>-1695.4</td>
<td>-1689.9</td>
</tr>
<tr>
<td>Short-term: culture</td>
<td>-1688.2</td>
<td>-1679.0</td>
</tr>
<tr>
<td>Short-term: combined</td>
<td>-1721.4</td>
<td>-1682.8</td>
</tr>
<tr>
<td>Long-term: geography</td>
<td>-1683.2</td>
<td>-1676.9</td>
</tr>
<tr>
<td>Long-term: bilat. trade</td>
<td>-1693.5</td>
<td>-1689.3</td>
</tr>
<tr>
<td>Long-term: culture</td>
<td>-1680.5</td>
<td>-1687.0</td>
</tr>
<tr>
<td>Long-term: combined</td>
<td>-1734.8</td>
<td>-1689.5</td>
</tr>
</tbody>
</table>

| ISO 14000 | No cross-country effects | -1183.1 | -1182.2 | -1048.7 | -1047.9 |
| Cross-country effects | | | | | |
| Short-term: geography | -1185.2 | -1184.8 | -1024.0 | -1018.3 |
| Short-term: bilat. trade | -1184.8 | -1183.6 | -1015.3 | -1015.2 |
| Short-term: culture | -1185.5 | -1183.6 | -1020.8 | -1017.8 |
| Short-term: combined | -1189.0 | -1188.3 | -1008.3 | -1039.6 |
| Long-term: geography | -1181.2 | -1183.4 | -1014.7 | -1018.2 |
| Long-term: bilat. trade | -1182.5 | -1181.5 | -998.2 | -997.2 |
| Long-term: culture | -1184.5 | -1195.6 | -999.1 | -997.0 |
| Long-term: combined | -1197.1 | -1190.9 | -998.7 | -1032.0 |

<sup>a</sup>The underlined statistics correspond to the models selected for further analysis.
### Table 2.2: Mean Squared Errors for holdout predictions across all model specifications in the ISO 9000 and ISO 14000 data sets.

<table>
<thead>
<tr>
<th></th>
<th>ISO 9000</th>
<th>Non-Random Effects</th>
<th>Random Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No omitted</td>
<td>Omitted var.</td>
<td>No omitted</td>
</tr>
<tr>
<td><strong>No cross-country eff.</strong></td>
<td>68.88</td>
<td>65.87</td>
<td>60.01</td>
</tr>
<tr>
<td><strong>Cross-country effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-term: geography</td>
<td>72.59</td>
<td>72.39</td>
<td>56.51</td>
</tr>
<tr>
<td>Short-term: bilat. trade</td>
<td>68.39</td>
<td>65.84</td>
<td>56.66</td>
</tr>
<tr>
<td>Short-term: culture</td>
<td>64.84</td>
<td>74.41</td>
<td>60.47</td>
</tr>
<tr>
<td>Short term: combined</td>
<td>65.70</td>
<td>65.38</td>
<td>56.26</td>
</tr>
<tr>
<td>Long-term: geography</td>
<td>79.51</td>
<td>78.80</td>
<td>56.57&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Long-term: bilat. trade</td>
<td>68.01</td>
<td>66.63</td>
<td>54.75</td>
</tr>
<tr>
<td>Long-term: culture</td>
<td>70.14</td>
<td>67.04</td>
<td>66.05</td>
</tr>
<tr>
<td>Long-term: combined</td>
<td>69.62</td>
<td>67.21</td>
<td>55.96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ISO 14000</th>
<th>Non-Random Effects</th>
<th>Random Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No omitted</td>
<td>Omitted var.</td>
<td>No omitted</td>
</tr>
<tr>
<td><strong>No cross-country effects</strong></td>
<td>25.4</td>
<td>26.1</td>
<td>21.7</td>
</tr>
<tr>
<td><strong>Cross-country effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-term: geography</td>
<td>24.4</td>
<td>25.1</td>
<td>21.3</td>
</tr>
<tr>
<td>Short-term: bilat. trade</td>
<td>25.3</td>
<td>26.3</td>
<td>21.8</td>
</tr>
<tr>
<td>Short-term: culture</td>
<td>24.6</td>
<td>25.7</td>
<td>21.8</td>
</tr>
<tr>
<td>Short term: combined</td>
<td>29.1</td>
<td>24.6</td>
<td>21.7</td>
</tr>
<tr>
<td>Long-term: geography</td>
<td>26.8</td>
<td>26.3</td>
<td>21.5</td>
</tr>
<tr>
<td>Long-term: bilat. trade</td>
<td>25.7</td>
<td>26.5</td>
<td>22.1</td>
</tr>
<tr>
<td>Long-term: culture</td>
<td>24.9</td>
<td>30.2</td>
<td>19.4</td>
</tr>
<tr>
<td>Long-term: combined</td>
<td>30.6</td>
<td>29.4</td>
<td>23.0</td>
</tr>
</tbody>
</table>

<sup>a</sup>The underlined statistics correspond to the models selected for further analysis.
<table>
<thead>
<tr>
<th>ISO 9000</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market potential: intercept</td>
<td>3.670</td>
<td>5.198*</td>
<td>3.998</td>
<td>9.452*</td>
<td>8.009*</td>
<td>7.481*</td>
</tr>
<tr>
<td></td>
<td>(3.630)</td>
<td>(2.426)</td>
<td>(2.731)</td>
<td>(1.680)</td>
<td>(1.625)</td>
<td>(1.629)</td>
</tr>
<tr>
<td>Market potential: population</td>
<td>4.765*</td>
<td>1.888*</td>
<td>2.492*</td>
<td>3.152*</td>
<td>2.052*</td>
<td>1.877*</td>
</tr>
<tr>
<td></td>
<td>(2.160)</td>
<td>(0.940)</td>
<td>(1.109)</td>
<td>(1.284)</td>
<td>(1.003)</td>
<td>(0.899)</td>
</tr>
<tr>
<td>Innovation coefficient</td>
<td>0.043</td>
<td>0.042</td>
<td>0.044</td>
<td>0.045</td>
<td>0.044</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.027)</td>
<td>(0.028)</td>
<td>(0.027)</td>
<td>(0.027)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Imitation coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own-country</td>
<td>0.478*</td>
<td>0.278*</td>
<td>0.306*</td>
<td>0.291*</td>
<td>0.246*</td>
<td>0.218*</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.052)</td>
<td>(0.055)</td>
<td>(0.051)</td>
<td>(0.046)</td>
<td>(0.043)</td>
</tr>
<tr>
<td>Cross-country: geography</td>
<td>0.095*</td>
<td>0.076*</td>
<td>0.059*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-country: bilat. trade</td>
<td>0.061*</td>
<td>0.047</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-country: culture</td>
<td></td>
<td></td>
<td></td>
<td>0.105*</td>
<td>0.066*</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.034)</td>
<td>(0.029)</td>
<td></td>
</tr>
<tr>
<td>Error variance</td>
<td>722.1</td>
<td>456.5</td>
<td>477.3</td>
<td>489.3</td>
<td>436.5</td>
<td>427.66</td>
</tr>
<tr>
<td></td>
<td>(64)</td>
<td>(40.1)</td>
<td>(41.5)</td>
<td>(43.8)</td>
<td>(39.3)</td>
<td>(39.5)</td>
</tr>
<tr>
<td>ISO 14000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market potential: intercept</td>
<td>1.74*</td>
<td>1.301*</td>
<td>1.136*</td>
<td>1.322*</td>
<td>1.044*</td>
<td>0.923*</td>
</tr>
<tr>
<td></td>
<td>(0.399)</td>
<td>(0.244)</td>
<td>(0.238)</td>
<td>(0.244)</td>
<td>(0.214)</td>
<td>(0.212)</td>
</tr>
<tr>
<td>Market potential: population</td>
<td>0.549</td>
<td>0.235</td>
<td>0.260</td>
<td>0.372*</td>
<td>0.221</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(0.405)</td>
<td>(0.165)</td>
<td>(0.148)</td>
<td>(0.176)</td>
<td>(0.113)</td>
<td>(0.108)</td>
</tr>
<tr>
<td>Innovation coefficient</td>
<td>0.040</td>
<td>0.045</td>
<td>0.065*</td>
<td>0.046</td>
<td>0.055</td>
<td>0.064*</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.028)</td>
<td>(0.030)</td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Imitation coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Own-country</td>
<td>0.79*</td>
<td>0.370*</td>
<td>0.379*</td>
<td>0.342*</td>
<td>0.282*</td>
<td>0.260*</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.063)</td>
<td>(0.070)</td>
<td>(0.058)</td>
<td>(0.053)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>Cross-country: geography</td>
<td>0.181*</td>
<td></td>
<td></td>
<td>0.098*</td>
<td>0.085*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td></td>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>Cross-country: bilat. trade</td>
<td>0.118*</td>
<td></td>
<td></td>
<td></td>
<td>0.090*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td></td>
<td></td>
<td></td>
<td>(0.031)</td>
<td></td>
</tr>
<tr>
<td>Cross-country: culture</td>
<td></td>
<td></td>
<td></td>
<td>0.133*</td>
<td>0.108*</td>
<td>0.098*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.035)</td>
<td>(0.032)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Error variance</td>
<td>100.75</td>
<td>75.91</td>
<td>67.85</td>
<td>69.22</td>
<td>65.43</td>
<td>62.31</td>
</tr>
<tr>
<td></td>
<td>(9.75)</td>
<td>(7.37)</td>
<td>(7.02)</td>
<td>(7.33)</td>
<td>(7.40)</td>
<td>(6.12)</td>
</tr>
</tbody>
</table>

Table 2.3: Parameter estimates for the ISO 9000 and ISO 14000 data sets with standard deviations in parentheses. An asterisk indicates that ratio of the parameter and its standard deviation exceeds 2. Model (1) has no cross-country effects. Model (2) – Model (4) contain cross-country effects based on geography, bilateral trade and culture, respectively. Model (5) combines the best fitting two versions of cross-country effects from Model (2) – Model (4). Model (6) combines all definitions of cross-country effects.
Figure 2.1: Map of the spatial sample

Figure 2.2: Yearly ISO 9000 and ISO 14000 certification counts for 4 countries.
Figure 2.3: (a) actual and fitted new ISO 9000 certifications for 1997, (b) actual and predicted new ISO 9000 certifications for 2000
Figure 2.4: ISO 9000 and ISO 14000 certifications resulting from various diffusion sources.

Figure 2.5: Comparison of the diffusion coefficients between ISO 9000 and ISO 14000 and between early- and later-adopting countries
Figure 2.6: Influence and susceptibility values for a sample of 32 countries
3 Measuring Consumer Switching to a New Brand across Local Markets

3.1 Introduction

The effects of marketing actions on brand switching have been studied in detail in past literature. Specifically, evidence has been found of asymmetric switching, i.e., when the effects of a brand’s marketing activities are distributed across competitors unproportionally to their market shares (e.g., Blattberg and Wisniewski, 1989). These asymmetries in the effects from marketing actions are a result of two main factors: brand positioning, which includes unique features of the brand that differentiates it from competitors\(^1\) and different sensitivities from the brands to marketing tools (Carpenter et al., 1988). This paper focuses on a less developed aspect of brand switching, more specifically, switching that occurs due to the introduction of a new brand in mature categories.

New products are pervasive in consumer package goods (CPG) industries (Kahn and McAlister, 1997) and there is little disagreement on the strategic importance of new product development in a competitive environment for mature CPG industries that are dominated by price-promotions. Indeed, whereas price and promotion investments have predominantly short term consequences (Dekimpe and Hanssens, 1999), resources spent on product development and product entry constitute long term investments. Yet surprisingly little is known empirically about consumer switching to a new CPG brand.\(^2\)

\(^1\)These features usually include physical characteristics, but they may also encompass perceived characteristics, sometimes driven by advertising and/or brand name.

\(^2\)Perhaps because the infrequent nature of (researchers) observing or (managers) deciding on new product launch, relative to say price discounts, there has been less observations (i.e., data) as well as less attention given to substitution patterns for new products in CPG industries.
This paper seeks to address this gap in the literature by studying two related research issues. The first is how the sales of a new brand decomposes into (1) cannibalization, when consumers that switch to the new brand are those from the firm’s other incumbent brands; (2) competitive draw, when consumers come from competitors’ brands, and (3) category expansion, when consumers are attracted from outside of the category. The second issue is how these quantities differ across local markets with different initial conditions, marketing strategies, and consumers. This information can ultimately be used during national roll-outs to make predictions about switching patterns in unentered markets. We propose a random coefficients logit model approach to measure the switching to a new brand while accounting for (1) temporary changes in other marketing activities such as price changes or promotional activities, (2) the heterogeneous preferences of consumers, and (3) the endogenous nature of prices.

Our method uses market-level or census divisional level aggregate data. The intuition for our identification strategy is best explained using an $I$ (consumers) by $T$ (time periods) panel data set. The common approach to demand estimation (e.g., Nevo 2001), is to use aggregates across the first dimension of $I$ consumers, i.e., use the $[T \times 1]$ market level time series of shares to infer the mean and the dispersion (heterogeneity) in consumer tastes for a product. This method works well to identify mean preferences, but we show that it has considerable difficulty in recovering the dispersion of preferences across the consumer population (Bodapati and Gupta, 2004; Petrin, 2002). Incorrect identification of preference heterogeneity poses problems in studies of substitution behavior (to a new brand or otherwise) since this behavior is directly driven by the similarity or dissimilarity
of preferences across brands.

Our approach aims to redress this by adding aggregates across the $T$ time periods to better estimate the dispersion of preferences. While in its most complete form, this could lead to the burden of collecting an extra set of individual level data, we show that easily obtained summary statistics of consumer types are successful at identifying heterogeneity in preferences. Specifically, we add to our estimation information about the distribution of the annual purchase set size $S_i$ (number of unique brands purchased by consumer $i$). We next impose that the logit demand system matches the correct distribution of purchase sets across consumers. This cross-sectional information is very informative about the dispersion of consumer preferences, because large purchase sets imply preferences that are tightly distributed, while smaller purchase sets imply more widely spread preferences for the brands.

The combination of two different sources of information to improve the ability of a model to match reality is not new. Berry et al. (1998) combined market level data with individual level information about the alternative foregone to improve estimates about the brand switching patterns. Alternatively, Petrin (2002) uses census information about demographic aspects of consumers that adopted the minivan, when it was launched. Both these studies involve a durable good, where purchase and consequently switching are likely to occur only once during the observable period. The repeat purchase characteristic of CPG leads to situations where consumers may buy one brand in one occasion, but due to changes in factors such as in marketing variables, change their choice in posterior occasions. Correctly modeling the switching that occurs over time and the strength of
brand preferences is essential to obtaining an accurate evaluation of adoption and repeat purchase patterns of a new brand.

The random effects logit demand system is used to estimate the distribution of consumer preferences and of their price sensitivities. In turn, these can be used to evaluate the exact impact of the launch of a new brand in each geographic market separately. Once we have the switching computed for each market, we explain variation in the relative importance of the alternative switching sources across geographic markets using brand factors, initial market conditions, variables describing marketing strategies and variables describing consumer characteristics. Surprisingly, we find that brand name and marketing mix variables, such as price and display, explain very little of the cross-market variation, with local aspects such as pre-entry share appearing as more important.

Our analysis covers volume sales data, promotions and prices for 64 IRI markets in the United States across 260 weeks from January 1995 to December 1999, for the frozen pizza category. These data also cover the total amount of dollars scanned in each of the markets. We have AC Nielsen market penetration data for the category in each of the 9 census divisions. Also from AC Nielsen, we obtained the local distribution of the purchase set size, i.e., the percentage of individuals that bought 0, 1, 2, 3+ unique brands of pizza over the period of a year, for each of the Census subregions. From the census 2000, we have demographic information for each market, such as average income, age, minorities importance in the population and family size.

Our findings are as follows: (1) methodologically, we show that heterogeneity in the consumer preferences is better captured when the additional information about the pur-
chase set size is introduced. More specifically, a model without this additional information would lead to overestimation of the amount of switching that occurs in the market. The purchase size data improves the ability of the random coefficients logit model to capture the correct dispersion of brands as they are perceived by the consumers; (2) substantively, we find that Kraft was very successful at attracting new consumers from outside of the frozen pizza category in the introduction of DiGiorno, with about half of its share originating from category expansion; and (3) in terms of multimarket findings, we show that brand name, marketing mix variables and most importantly pre-entry share explain a significant portion of the cross-market variation in the switching patterns to DiGiorno.

In the next section, the paper continues by presenting its contribution and place in the literature. Section 3 presents the random coefficients model. We describe in more detail the category used to empirically test the model in section 4. Section 5 describes the estimation algorithm and section 6 the results. We conclude in section 7.

3.2 Contribution and place in the literature

3.2.1 Measuring switching

New product introduction is considered as one of the most relevant and consequential marketing actions by both managers and researchers. Past literature has quantified how much share a new brand should obtain, given factors such as order of entry (e.g., Kalyanaram and Urban, 1992) and observable brand attributes (e.g., Hardie and Fader, 1996), while other studies have focused on measuring the impact of the a new product on consumer behavior, namely in terms of consumers’ sensitivity to price and changes in the brand perceptions (Van Heerde et al., 2004).
An important aspect inherent to new product introduction is brand switching, since at least some consumers will be convinced that the new brand is a better alternative when compared to incumbent competitors, therefore changing their choice patterns. However, marketing literature has focused mainly on switching caused by other activities, such as changes in prices and promotional actions. Asymmetry in brand switching caused by price variation has been observed and explained due to brand tiers (Blattberg and Wisniewski, 1989). In mature markets, switching has been explained as a function of price, advertising and brand attributes (Carpenter and Lehmann, 1985), with some brands with strong names being to some extent insulated from the impact of promotions from competitors.

We direct our efforts to the less studied but still relevant aspect of new product introductions: quantifying and explaining brand switching that occurs when the new brand is introduced across a multitude of markets. More specifically, we identify the proportion of consumers that shift consumption from (1) brands from competitors, (2) incumbent brands that belong to the same firm that owns the new brand or (3) alternative options from outside of the category that satisfy the same need. Quantifying these three sources of share is essential to evaluate the success of a new brand, its contribution to the firm’s profits while accounting for cannibalization and its impact on the category performance. We evaluate the switching patterns to the new brand separately for a number of geographic markets and with these results for all markets, we proceed to explain differences and similarities across geography.

The multimarket aspect of the paper is essential, since in most industries, market shares and switching patterns observed in a specific market may not be generalized to the
entire population, given that consumer characteristics, brand history and firms’ marketing behavior is likely to be significantly different across markets (Bronnenberg et. al, 2005). In addition, in the last few years, multimarket aggregate level data sets have become increasingly available, leading to studies that cover multiple areas and investigate both differences and similarities across geography (Bronnenberg and Sismeiro, 2002).

3.2.2 The need for more information to measure heterogeneity correctly

Two different types of information can be used to analyze the research problem at hand: household panel data and aggregate market-level data. Historically, household panel data has been used to study individual choice history and brand switching, while aggregate level data has been usually applied to monitor category and brand sales evolution, evaluate market shares and quantify the impact of marketing actions (Bodapati and Gupta, 2004). Both cases have their own advantages and disadvantages. Although the proposed type of analysis could be done using individual level data, this option would involve information that is, at present, very sparse. Specifically, since we are concerned with the introduction of a new brand across the United States, which has usually attached a roll-out that lasts several years, panel data studies would be required for a sufficient number of stores in each geographic area with different time windows to observe the launch in each market. This kind of information is hard to obtain for both managers and researchers, making us choose aggregate level data. Another reason that led us to opt for the use of aggregate level data derives from the fact that it is information that is in most cases readily available to managers and usually purchased at a comparatively lower cost than individually data.3

3For a more detail comparison of benefits and limitations of the two types of information, please see Bodapati and Gupta, 2004.
Models which define utility at the individual level but use solely aggregate level data for estimation base their identification strategies on the fact that consumers leave a trace of their preferences and actions on the observed aggregate shares. The strength and level of identification of these individual signals using just aggregate level data has generated some debate in the past. The fact is that, on one hand, stands the optimal level of information, i.e., very rich data on every individual (or at least large samples of individuals) for all markets/stores under analysis, but in most cases impossible to obtain. On the other hand, we have the market level data, easiest and cheaper to obtain, but less informative about the individual behavior, as details are lost due to aggregation. We believe that the solution to handling this contrast is to be positioned somewhere in the middle of this scale: use market-level data, more accessible and less sparse, and simultaneously use some additional information about the consumer behavior that gives rise to the results observed at the aggregate level.

Two other issues have become increasingly important when estimating demand models using aggregate data: endogeneity (usually of prices) and consumer heterogeneity (Chintagunta, 2001). A feasible approach to solving the endogeneity problem has been suggested by Berry (1994), with the use of the contraction mapping to identify the structural unobservable that is correlated with prices. Identifying heterogeneity with aggregate data in the consumers is still in some cases a hard problem to solve. According to Bodapati and Gupta (2004), heterogeneity is in practice hard to identify because the only available information to do so is the result of discrepancies in observed shares movements and the expected movements predicted by a homogeneous model. In cases where the discrepancies
are small, it will be hard to identify heterogeneity parameters. Two reasons are mentioned as causes for the inexistence of large discrepancies. First, the homogeneous model may be flexible enough to capture any complex movements in the data. Second, the sample maybe be too small. Petrin (2002) also mentions the same problem, defending that heterogeneity is identified only if some unusual substitution patterns not captured by the homogeneous model do occur and/or a change in the choice set is present, namely the introduction of a new brand.

There have been some past attempts to close the gap between the aggregate and the individual level data and overcome the mentioned limitations, by adding relevant information, such as information about consumer behavior (Berry et al., 1998) and census information (Petrin, 2002). Our paper follows the same line of thought, focusing specifically on correctly identifying consumer heterogeneity and consequent switching patterns. This constitutes a major distinction between our contribution and the two mentioned papers. Both Berry et al.(1998) and Petrin (2002) focus on a durable goods category, the auto industry, where switching over time is almost null. In the case of CPG, an important characteristic of consumer behavior is repeat purchase. Consumers return after certain time to purchase from the same category, which has important consequences such as the emergence of loyalty to one specific brand or frequent switching among brands. If we are interested in measuring the switching patterns to a new brand in the presence of repeat purchase, it is important to identify correctly consumer heterogeneity in brand preferences.

Consumers are likely to be significantly different in their taste for brands. The strength of their preferences for one or more brands ultimately defines the amount of switching
observed over time. In a situation where most consumers have a strong preference for one of the brands, switching will be minimal and purchase sets include only the brand to which consumers are loyal. However, if preferences are somewhat diffuse and consumers are indifferent among brands, switching will occur much more often.

We now present a simple example that illustrates two possible scenarios regarding preference heterogeneity and that shows how hard it is to distinguish between them when only market-level shares are available. Assume for simplicity that a market has only two consumers, $a$ and $b$, which may opt to buy either brand $j$ or brand $k$. Suppose also that prices are the same for both brands and do not change over time. In a first scenario, presented in Figure 3.1 (a), consumers have strong preferences for one of the brands. Consumer $a$ prefers brand $j$, and over several purchase occasions, she has a high choice probability and will always buy that brand. On the other hand, consumer $b$ enjoys brand $k$ considerably more and is loyal to that brand. In the second case, Figure 3.1 (b), consumers are almost indifferent between the two brands. In this case, choice probabilities are almost the same, around 0.5 for each brand, representing indifference, which leads consumers to switch randomly between the two brands. To summarize, the first scenario is characterized by strong brand preferences for one of the brands, lower switching and consequently a smaller purchase set that includes one brand only, while the second situation shows indifference between the brands, resulting in frequent switching and a larger purchase set of two brands.

These two scenarios provide evidence that in some cases, aggregate level information solely about market shares is not enough to identify consumer heterogeneity and switching
patterns. In the both cases, we roughly obtain market shares of 50% for each brand. However, as described previously, consumer preferences are quite dissimilar. In order to correct for this limitation that results from aggregation over individuals, we propose to add a simple and easy to obtain piece of information that will helps us disentangle the two situations: the marginal distribution of the purchase set size. This additional data will distinguish the two scenarios in the following way: in case (a), 100% of individuals buy one brand, whereas 100% of individual purchase two brands in case (b).

In sum, our methodology to study the switching patterns is as follows. We rely on two particular pieces of aggregate level information to determine correctly the switching patterns to the new brand: first, we observe shares and marketing variables over a long period of time, including the introduction and development of the new brand across a larger number of markets; second, we add aggregate information about the behavior of consumers across time, namely summary statistics of the purchase set size, i.e., the percentage of consumers that bought only one and two unique brands and the percentage of consumers that did not buy at all from the category during a year period. This extra information provides a measure of how frequent consumers substitute across brands over time and consequently constitutes a measure of the level of heterogeneity present in the category.

3.3 Model

We now present the model that will enable us to compute the switching to new brand. We treat each market separately at this time. Our methodology makes use of the way that random coefficients models incorporate consumer heterogeneity. In general terms, heterogeneity is included in these models by taking draws, usually from Normal distribu-
tions, that represent a sample of simulated individuals. The parameters to be estimated include the mean and variance-covariance matrix of these distributions. Typically, at each time period, estimated shares are a result of aggregation over this sample of individuals and parameters are obtained through some optimization routine.\textsuperscript{4} Our model will add an extra dimension to the estimation, making use of information relative to the behavior of consumers over time. Once the parameters are estimated, we can compute the implied choice probability history of each simulated individual. Brand switching is obtained by comparing choice probabilities of two scenarios, the actual case, where the new brand is introduced in the market, and a counterfactual situation, where the new brand is never introduced. The difference in choice probabilities and consequent market shares between these two situations provides us with an accurate measure of switching from each incumbent brand to the introduced alternative.

Assume we observe $t = 1, ..., T$ time periods, in $m = 1, ..., M$ markets, with $i = 1, ..., I$ consumers. In each market $m$ (the subscript for market is removed for simplicity reasons), at each occasion $t$, the utility of brand $j$ for consumer $i$ is given by the following expression:

$$u_{ijt} = \alpha_{ij} + x_{jt}\beta_i + \xi_{jt} + \epsilon_{ijt},$$

(3.1)

$$i = 1, ..., I, \quad j = 1, ..., J, \quad t = 1, ..., T$$

where $\alpha_{ij}$ is individual $i$’s unique perception of brand $j$, $x_{jt}$ is a $K$-dimensional row vector of observed marketing mix variables, $\beta_i$ is a $K$-dimensional vector of individual specific marketing mix coefficients and $\xi_{jt}$ includes structural disturbances that are unobserved.

\textsuperscript{4}The optimization routine may be set to minimize the distance between actual and estimated shares (a minimum squared error approach), satisfy some moment conditions (a Method of Moments approach, which is used here in this paper) or any other plausible objective.
by the econometrician but considered by consumers in their decisions. Finally, $\epsilon_{ijt}$ is an i.i.d. mean-zero stochastic term with type I extreme value distribution.

Consumers are allowed to be heterogenous in their preferences for brands and in their sensitivities to marketing mix variables. The vector of brand intercepts $\alpha_i$ has the following expression:

$$\alpha_i = \bar{\alpha} + z_i, z_i \sim N(0, \Sigma)$$

(3.2)

The matrix $\Sigma$ can be estimated directly or using a factor structure (Chintagunta et al., 2002). We choose this last alternative, as it presents the advantage of reducing the number of parameters required to estimate the heterogeneity matrix and at the same time, when the number of factors is two, this structure also allows the construction of a perceptual map, a tool that most managers are familiar with. The factor structure is implemented by making the matrix $\Sigma$ a function of 2 (or more) perceived attributes $L$:

$$\Sigma = L w L', w \sim N(0, I)$$

(3.3)

where $L$ is a $[J \times 2]$ matrix of perceived attributes. This formulation provides $z_i = Lw_i$.$^5$

Heterogeneity in response to marketing mix variables is modeled in the following way:

$$\beta_i = \bar{\beta} + \lambda' v_i, v_i \sim N(0, I)$$

(3.4)

As it is usual in these models, consumers are allowed not to buy from the category. This alternative will include the no-purchase option as well as choice of substitute options from

$^5$Some restrictions must be made for identification of this model. The intercept mean of the outside good is set to zero, $\bar{\alpha}_0 = 0$, both perceived attributes of the outside good are set to 1, $L_{1out} = L_{2out} = 1$ and finally one of the attributes of the fringe is also set to 1, $L_{ fringe} = 1$. 

52
outside of the category and its utility is fixed at zero for identification purposes. Given the extreme value distribution of $\epsilon_{ijt}$, the probability of household $i$ purchasing brand $j$ at time $t$ is given by:

$$\Pr (j_t \mid x_{jt}, i) = \frac{\exp \left( \alpha_{ij} + x_{jt}\beta_i + \xi_{jt} \right)}{1 + \sum_{k=1}^J \exp \left( \alpha_{ik} + x_{kt}\beta_i + \xi_{kt} \right)}$$  \hspace{1cm} (3.5)$$

To obtain market level shares $s_j$, we integrate over all $i$, which indexes both brand tastes and marketing mix variables sensitivities:

$$\hat{s}_{jt} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \Pr (j_t \mid x_{jt}, i) \phi (v) \phi (z) \, dv \, dz \hspace{1cm} (3.6)$$

$$i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T$$

where $\phi(.)$ is the pdf of the normal distribution. With this set up, the parameters to be estimated are $\theta = [\bar{\alpha}, \bar{\beta}, L, \lambda]$. The integral shown in equation 3.6 lacks an analytical form. We use simulation methods to numerically estimate the shares, as in Nevo (2001), which we described in more detail in the estimation section.

### 3.4 The frozen pizza category

In this section, we describe the category and the data set on which we will test our model empirically. Our main objective here is to provide a description of the competitive environment and give some details about the introduction of a new brand and its impact on the category performance.

#### 3.4.1 Competitive background

Frozen pizza has become one of the most important categories among frozen food, accounting for about 19% of the its sales (Van Heerde et al., 2004). According to managers,
it represents almost 20% of the total pizza business, with delivery pizza being its main competitor outside of the category (Pizza Marketing Quarterly). During 1993-1995, the years preceding our analysis, the category was viewed as stable and characterized by slow growth, with dollar sales marginally increasing from $1.6 to $1.7 billion. 1995 brought a new brand into the market, DiGiorno, followed by Freschetta in 1996. These brands included a new feature, self-rising crust, which was considered a major development in the category. Combined with strong advertising, the new brand introductions led to a fast increase in sales of frozen pizza. In fact, annually growth rates achieved an average of approximately 12% through 1999 (Holcomb, 2000).

Two companies, Kraft and Schwan Food Company, have a dominant place in the category, with a portfolio of multiple brands. Kraft’s brands include DiGiorno, Tombstone and Jack’s while Schwan owns Tony’s, Red Baron and Freschetta. All of these brands except Jack’s are available in most locations in the United States. Jack’s distribution is limited to markets in the northwest area of the country. Our analysis will focus mainly on the brands from Kraft and Schwan’s.6 Another competitor, Pillsbury, owns one major brand named Totino’s, which will also be consider in our analysis, since it has an average share of about 10%. All other brands, in most case local brands, are consolidated into a single quantity that we call the fringe.

In categories where firms launch a new brand that directly competes with its own established brands, cannibalization becomes one of the main concerns of managers, specially in cases where the incumbent brands are dominant players in the market. In order

---

6Schwan’s also owns Freschetta, but in 63 out of 64 markets, it was introduced later than Digiorno and captured a very small share of the market.
to avoid this negative effect and taking advantage of the higher quality associated with rising-crust, Kraft positioned DiGiorno as a direct competitor of delivery pizza, trying to attract consumers from the outside of the category. Its strong advertising made consumers aware of the phrase "It’s not delivery, it’s DiGiorno". A year after the roll-out was complete, the brand had captured an average of about 13% of the frozen pizza market.

Average annual shares of brands from 1995 to 1999 are presented in Table 3.1. In the first three years, DiGiorno went through its introduction and growth phase, gaining 10% of the market. During this period, almost all alternatives (except Totinos) lost a very small percentage of share, usually 1% or 2%, being hard to identify where most of DiGiorno’s share came from or which brand suffered the most with DiGiorno’s entry.

Markets responded differently to the introduction of the new brand. In Figure 3.2, we show the evolution of market shares in Sacramento and Salt Lake City, around the time of DiGiorno’s introduction. In Sacramento, it is clear that Tombstone suffered a heaviest negative impact in the first weeks of DiGiorno’s presence in the market. On the other hand, Red Baron appears untouched by DiGiorno and even gained some share during the same period. Salt Lake City presents almost the opposite situation. Here, it is Red Baron that suffers the most negative consequences, with its average market share falling from about 27% to 20%. The other main brands, Tombstone and Tony’s, also loose some share during the introductory period in this market. In terms of category growth, we can also describe examples of this disparities across markets. For instance, in Houston, the frozen pizza category grew by about 20% during the first two years after DiGiorno was introduced, while Denver showed a very slow growth of 3%.
3.4.2 Data description

Our data consists of four main parts. The first contains market-level information on volume and value sales and marketing mix variables for the frozen pizza category. It includes 260 weeks of data, across 64 IRI markets, from January 1995 to December 1999 and it is the result of aggregation over a sample of stores in each market. The markets are basically defined as metropolitan areas and are distributed across all the United States. They are geographically located with considerable distance between them, making arbitrage very unlikely. The volume sales is used to compute the market shares. In average, there are about 10 brands of frozen pizza present in each market. Included in the marketing mix variables, we have prices, percentage of sales sold under display and featured, distribution levels and share of voice for each brand.

The second set of data consists of summary statistics about the purchase set size and penetration levels. We have obtained from AC Nielsen the percentage of consumers that buy 0, 1, 2, 3+ unique brands over the duration of a year, for each of the 9 census divisions, for 2004. This data was collected by AC Nielsen using a wand panel data, where a sample of consumers scans the purchase brand at home on their own. The penetration levels are at the census divisional level, for the years 2000 to 2003, measuring the percentage of people that have purchased pizza at least once during each year.

The third set consists of information useful to estimate the size of the outside good. We have the yearly average penetration level of frozen pizza, also at the census divisional level. We also have the total amount of dollars spent specifically in frozen pizza and the total amount of dollars scanned for the total basket, for each week at the market level.
We also have the average interpurchase time, obtained from the Frozen Pizza Factbook.

Finally, from the Census 2000, we have obtained detailed demographic information for each market, which includes population size, average income, age, minorities importance in the population and family size.

### 3.4.3 Outside good

One of the limitations of past studies using aggregate level data concerns estimating the relative size of the category under study, as a proportion of the outside good. An example of approximations used in the literature is the number of households present in a certain market, as in Nevo (2001). Even if information is available about the penetration rate, in most cases, there is limited information about the evolution of the category size.

Another possibility is to look at the growth of the category, combined with the penetration level. However, this also poses some problems. Market level data is usually the result of an aggregation of a sample of stores in each geographical area. Frequently, during the process of data collection, we observe both attrition and recruitment, that is, some stores are added to the sample, while others may leave. This change in the sample of stores makes the information about the total volume of sales in the category unusable.

Our proposed alternative is to combine several pieces of information to produce accurate estimates of the outside good for each week and market in our data set. First, we start by looking at the yearly penetration rates from 2000 to 2003 for each of the Census subregions. We extrapolate the penetration levels for 1999, using cubic spline extrapolation, obtaining the yearly mean size of the inside good as a percentage of the potential market for that year, $P_{m1999}$. 

57
However, in order to evaluate the impact of a new brand on size of the category, we need a more dynamic measure, adequately quantified at the weekly level. To do so, we must go from an annual measure of penetration, $P_{m1999}$, to a weekly measure $p_{m1999}$. Since we are in a repeat purchase category, we cannot just simply divide the yearly penetration by 52 weeks. Alternatively, we make use of data about the national average interpurchase time, $\eta$ weeks. Since this value is an average across all consumers, including those that only buy very spontaneously and those that buy very frequently, the interpurchase time represents the average length of a cycle, after which the same consumers will return, for about $52/\eta$ times during the year. Consequently, looking at the penetration level for the year is the same as looking at the penetration level for the $\eta$ weeks, since in average it is always the same consumers that come back. Then, the annual penetration rate is in average equal to the penetration rate for a period of $\eta$ weeks. Assuming that consumers arrive at the store homogeneously through out these $\eta$ weeks, this makes the average weekly penetration level in 1999 to be equal to $p_{m1999} = P_{m1999}/\eta$.

To obtain the weekly evolution of the penetration level, we combine the average weekly penetration rate $p_{m1999}$ with the weekly expenditures on frozen pizzas, $h_{tm}$, measured as a percentage of the total value of the shopping basket, in a certain market. This information is available for all markets and for the 260 weeks in our data set. We compute the mean value of the percentage of expenditures on frozen pizza for 1999, $\bar{h}_{m1999} = 1/52 \cdot \sum_{t \in 1999} h_{mt}$. This quantity is useful, since we need to obtain a correct correspondence between weekly penetration and weekly expenditures, which is given by the ratio of the average values:
Finally, the last step is to translate the expenditures on frozen pizza to weekly penetration levels

\[ p_{tm} = \omega \cdot e_{tm} \]  

\[ \omega = \frac{\bar{P}_{m1999}}{e_{m1999}} \]  

### 3.5 Estimation

We estimate the model using the Generalized Methods of Moments (Hansen, 1982), making use of different sets of information to improve the accuracy of our estimation (Imbens and Lancaster, 1994; Petrin, 2002). We have two sets of moments. The first set of moments is similar to the one used in BLP (1995) and Nevo (2001), where the structural error term is assumed to be independent from the instrumental variables used for price. The second set makes use of summary statistics on the purchase set size of consumers. We now give an outline of the estimation procedures. For further details, refer to Berry (1994), BLP (1995), Imbens and Lancaster (1994), and Petrin (2002).

#### 3.5.1 Independence between structural unobservables and prices

Endogeneity has become one of the main concerns when estimating discrete choice models, since it may cause estimation bias due to correlation between prices and unobserved attributes, \( \xi_{jt} \). Past literature has provided evidence of this bias when using store-level data (Chintagunta, 2001; Villas-Boas and Winer, 1999). BLP (1995) presents an optimization algorithm to estimate random coefficients models that minimizes this correlation, making use of the generalized method of moments (GMM).
We start by taking $N$ draws from the normal distributions $N(0, \Sigma)$ and $N(0, I)$, to obtain a sample of $v_i$ and $z_i$, which, given a certain set of parameters, will be used to obtain the brand tastes and marketing mix variables sensitivities of $N$ simulated individuals, $\alpha_i$ and $\beta_i$ using equations 3.2 and 3.4 respectively. Given these values, each run of the optimization algorithm has two main stages. In the first stage, the main concern involves obtaining the structural unobservables $\xi_{jt}$, using the contraction mapping describe in BLP (1995). Start by defining the mean utility $\delta_{jt}$ for brand $j$ at time $t$ as

$$\delta_{jt} = \alpha_{ij} + x_{jt}\beta_i + \xi_{jt}$$  \hspace{1cm} (3.9)

This method requires that, starting with an initial value for $\delta_{jt}$, and given a set of parameter values, we iterate the following expression until it converges\(^7\):

$$\delta_{jt}^{n+1} = \delta_{jt}^n + \hat{s}_{jt} \left( \xi_{jt}, L, \mu, \lambda \right) - s_{jt}$$  \hspace{1cm} (3.10)

where $\hat{s}_{jt}$ is the estimated share for brand $j$ at time $t$ given all data and parameter values in that run, $s_{jt}$ is the respective actual share and $n$ refers to iteration in the contraction mapping. Once $\delta_{jt}$ are computed, $\xi_{jt}$ is obtained using $\xi_{jt} = \delta_{jt} - \alpha_{ij} - x_{jt}\beta_i$.

The second stage uses the just calculated $\xi_{jt}$ and instrumental variables to take endogeneity into account, using the generalized method of moments. Prices are most likely correlated with $\xi_{jt}$, since the structural disturbances reflect demand shocks that are unobservable to the researcher but impact the utility of the consumer. Retailers have information about these disturbances and how they affect utility and shares and consequently

\(^7\)Convergence is obtained $|\delta_{jt}^{n+1} - \delta_{jt}^n| < \varepsilon$, for $\forall \delta_{jt}$, with $\varepsilon$ very small.
include those factors in their pricing decisions.

Define these instruments as \( Z = [z_1, z_2, ..., z_l] \). The moment can formally be described as:

\[
G_1(\theta) = E\left[ \xi_{jt}(\theta) \mid Z \right] = 0
\]  

(3.11)

3.5.2 Summary statistics of the purchase set size

Our model will add an extra dimension to the estimation, making use of information relative to the behavior of consumers over time. We use the summary statistics about the purchase set size to improve the model’s ability to measure consumer heterogeneity in the brand preferences. Name the purchase set size for individual \( i \) as \( S_i \). \( S_i \) can take the values of 0, 1 or 2+, where the 2+ will include all cases when the purchase set includes two or more different brands.

As described before, we can calculate the choice probabilities for each individual \( i \) at week \( t \) using equation 3.5. The estimated probability of individual \( i \) purchasing from the outside good during \( \tau \) weeks (\( \tau \) equals 52 for a year) is given by the following expression:

\[
\Pr(S_{i\tau} = 0 \mid x_{jt}, i) = \prod_{t=1}^{\tau} \Pr(j_t = 0 \mid x_{jt}, i)
\]  

(3.12)

The case of \( S_i = 1 \) includes two cases: (1) the case where consumer \( i \) buys the same inside brand in every week \( t = 1, ..., \tau \); (2) the consumers switches between one brand and the outside good, given by the expression:
\[
Pr(S_{ir} = 1 \mid x_{jt}, i) = \left[ \sum_{k=1}^{j} \prod_{t=1}^{j} Pr(j_t = 0 \lor j_t = k \mid x_{jt}, i) \right] - Pr(S_{ir} = 0 \mid x_{jt}, i)
\]

(3.13)

The first terms includes the alternatives where consumers buy one brand, one brand and the outside good and the outside good uniquely. Since we do not want to include the cases where the consumer just buys the outside good, we must subtract the second term. Finally, \( Pr(S_{ir} = 2+ \mid x_{jt}, i) = 1 - Pr(S_{ir} = 0 \mid x_{jt}, i) - Pr(S_{ir} = 1 \mid x_{jt}, i) \).

We then aggregate over consumers’

\[
\widehat{S}_0 = \frac{1}{I} \sum_i Pr(S_{ir} = 0 \mid x_{jt}, i)
\]

(3.14)

and likewise for \( \widehat{S}_1 \) and \( \widehat{S}_{2+} \). The moment condition that enters the optimization algorithm will match estimated and actual summary statistics of the purchase set size:

\[
G_2(\theta) = \left[ \begin{array}{c}
\widehat{S}_0 - S_0 \\
\widehat{S}_1 - S_1 \\
\widehat{S}_{2+} - S_{2+}
\end{array} \right] = 0
\]

(3.15)

### 3.5.3 Objective function

The objective function combines the two sets of moments previously described:

\[
E[G(\theta)] = E\left[ \begin{array}{c}
G_1(\theta) \\
G_2(\theta)
\end{array} \right] = 0
\]

(3.16)

The two-step GMM estimator (Hansen, 1982; Petrin, 2002) has the following formulation:

\[
\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{G}(\theta)' w(\hat{\theta})' w(\hat{\theta}) \hat{G}(\theta)
\]

(3.17)
where \( \hat{G}(\theta) \) is the sample analogue of \( G(\theta) \) and \( w(\tilde{\theta}) \) is a consistent estimate of the "square root" of the inverse of the variance-covariance matrix of the moments, obtained using \( \tilde{\theta} \), a preliminary consistent estimate of \( \tilde{\theta} \). For the first set of moments, \( G_1(\theta) \), its weight is given by

\[
w_1(\theta) = \left[ \frac{1}{T} \sum_{i=1}^{T} g_i(\theta) \cdot g_i(\theta)' \right]^{-1}
\]

(3.18)

where \( g_i(\theta) \) are the moment conditions for each time period. The second moment, \( G_2(\theta) \), is a result of an average of individual probabilities. Its weight is equal to the inverse of the variance, i.e.

\[
w_2(\theta) = \left[ \frac{\text{var} \left( \Pr \left( S_{ir} = i \mid x_{jt}, i \right) \right)}{N} \right]^{-1}, i = 0, 1, 2+
\]

(3.19)

### 3.5.4 Computing the switching

In order to obtain the switching from incumbent brands to DiGiorno, we compare two scenarios. The actual scenario, where DiGiorno was introduced in the market, and an alternative counterfactual case, where we remove DiGiorno from the market. We then compare the difference of shares of incumbent brands in the two scenarios. The idea behind this method is to identify who would have kept the share that was transferred to the introduced brand.

For each time period and incumbent brand (including the outside good option), this difference is computed using the following expression:
\[ \Delta \hat{t}_{jt} = \frac{1}{N} \sum_{i=1}^{N} [\Pr (j_t \mid x_{jt}, i, \text{digiorno inst}) - \Pr (j_t \mid x_{jt}, i, \text{digiorno out})] \]  

\[ j = 1, \ldots, J, j \neq \text{DiGiorno} \quad \forall t \text{ after DiGiornos entry} \]

3.6 Results
3.6.1 Impact of additional moments on capturing heterogeneity

In order to measure the impact of the additional information about the purchase set size statistics, we started by testing our model using a pre-determined set of parameters and verifying its ability to recover these values. The chosen parameter values are shown in first column of Table 3.2. Price and display parameters were set to \(-0.5\) and \(2\) respectively. Brand intercept values were chosen to create shares that resemble what we observe in the real data, with a significant larger share given to the outside option.

For simplicity and clarity in the results, we impose some restrictions in our model. The matrix \(\Sigma\) is assumed to be diagonal and given by:

\[ \Sigma_{J \times J} = \rho \cdot \begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \ldots & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \]  

Another imposed restriction is on display heterogeneity, which is fixed at zero. From the normal distributions \(N(0, 1)\) and \(N(0, \Sigma)\), we draw a sample of \(N\) values for \(v_i\) and \(z_i\) respectively. We combine these draws with real data from the pizza category in the Los Angeles market, namely prices and display, to generate hypothetical market shares for each brand, given the chosen parameter set. The value of \(\rho = 1.5\) was chosen to be high enough to make some consumers loyal to one brand during the course of a year, in this
case about 12% ($S_1 = 0.12$). Given this formulation, we compute market shares (we show the average for weeks pre- and pos-Digiorno’s entry) and switching from brands competing in Los Angeles to Digiorno.

We estimate the parameters using two alternatives, one with the traditional set of moments (information poor model), $G_1 (\theta) = E \left[ \xi_j (\theta) \mid Z \right] = 0$, and another with both set of moments (information rich model), $G_1 (\theta)$ and $G_2 (\theta)$, which makes use of the additional information about the purchase set size. Prices in other markets are used as instrumental variables in our GMM estimation. We opted for prices in markets distanced more than 400 miles from the market being estimated as instruments, to eliminate possible correlation between neighboring prices and structural unobservables. Results are presented in Table 3.28. We start by noting that both models perform very similarly in terms of fit, with estimated shares coming out very close to the generated shares, evidence that both sets of parameters are feasible values to explain share variation. It is important to refer that the method of moments does not minimize the distance between generated and estimated shares, but the correlation between structural unobservables and instrumental variables and under this criteria, the information poor model does marginally better in minimizing this correlation, since it does not have to satisfy the second set of moments.

Although the fit is in fact very similar between the two models, parameter values and consequent consumer behavior is significantly different. In the information poor model, we observe significant biases. Brand intercept heterogeneity is underestimated, which results in a misfit in both brand intercept and marketing mix variables coefficients. Price

---

8In the estimation, we draw a smaller sample of values for $z_i$ and $v_i$ ($N = 200$), due to computational limitations. To avoid small sample biases, we estimated the model with 20 different samples and present the mean values for the estimated parameters.
and display coefficients are slightly underestimated while brand intercepts, although with correct ordering, also show some bias. When we calculate the purchase set size summary statistics implied by these values using equation 3.12 and compare them with the original values, we observe a larger percentage of consumers would buy 2 brands or more (0.97 vs. 0.83) and a smaller percentage would be estimated as loyal to just one brand (0.02 vs. 0.12). These results would lead to a misconception of the amount of switching that occurs across time and of variety consumers search for in a category that involves repeat-purchase. These deviations are corrected in the information rich model. Brand intercept heterogeneity is estimated correctly. Other biases in the parameters means are also reduced significantly and implied purchase set size statistics are significantly closer to the generated consumer behavior.

The incorrect set of parameters in the information poor model does not seem to have an impact on the switching to the new brand, with both models performing considerably well in estimating the different sources of share of the new brand (DiGiorno). This is a result of the restriction imposed on the correlation matrix $\Sigma$ to be diagonal, which is not flexible enough to allow for considerable asymmetric switching. The next section will deal with that issue.

### 3.6.2 Impact of additional moments on brand switching

We now turn our attention to the real data and to the impact of added moments when we used a complete heterogeneity matrix $\Sigma$ with a factor structure. We estimated the model with the data from 64 markets separately and in this section we present the estimation results for the Los Angeles market in Table 3.3 (brand intercepts and marketing mix variables
coefficients) and Figure 3.3 (perceptual map with the location of the brands). We start by noting that, once again, preferences heterogeneity continues to be incorrectly estimated in the model without additional information about the purchase set size (information poor model). In the Los Angeles market, actual data tells us that about 38% of the consumers did not buy from the pizza category and 22% bought exclusively one brand, during a full year. The information poor model would predict 0% and 2%, completely underestimating heterogeneity and consequently overestimating the amount of brand switching over time. Additional moments bring this percentages to much more reasonable values of 34% and 21%, mimicking consumer behavior with much more accuracy.

There are two justifications for this estimations: the first one is the value of price heterogeneity, unreasonably low in the information poor model; the second is the position of the brands in the perceptual map. In Figure 3.3, we depict brand positions for the two alternative models. It is noticeable that the relative positions, although somewhat similar in the two cases, are considerably more disperse in the information rich model. Considering the outside good as the center of the map at (1,1), there is a significant larger distance between the brands in the model that includes the additional information about the purchase set size, which causes a decrease in the amount of brand switching over 52 weeks predicted by this model and matching reality more accurately.

If we continue to look at this perceptual map, we can also find a justification for the differences in the proportions that each brand contributes to DiGiorno’s share, shown in the last five rows of Table 3.3. The main difference comes from the fact Tombstone becomes more important as a source of share in the information rich model (1% vs. 6%
in relative terms), while the fringe and the outside good lose importance. This change is a direct response to the different positioning of the brands in the two scenarios. In the case where no added moments are included, DiGiorno appears much closer to the fringe and the outside good, with all other brands positioned on the opposite area of the perceptual map. When we add moment, Tombstone becomes one of the closest alternatives to DiGiorno, which justifies the higher switching from its consumers to DiGiorno. A managerial consequence of this correction is a more accurate evaluation of cannibalization between the different brands of a firm, in this case of Kraft, essential to a better evaluation of the impact of the new brand in the profits of the firm.

Regarding more substantive issues, we should mention two important findings: first, Kraft was successful at attracting consumers from outside of the category in Los Angeles, avoiding much of the cannibalization that might have emerged. In fact, looking at Figure 3.3, you can see that DiGiorno is positioned in an empty area of the perceptual map, most likely due to its unique feature, the rising-crust. Also, its closest alternative in the mind of the consumers is the outside good, which includes not only the no-buy option, but also the delivery pizza. This relative position of DiGiorno and the outside good is also observed in several other markets and it is indicative that the advertising campaign ran by Kraft that targeted directly consumers from outside of the category may have had results in terms of positioning the brand in the marketplace.

The second finding concerns the fact that two Schwan’s brands, Tony’s and Red Baron, appear close together in the perceptual map in a multitude of markets. This is an unde-

---

9It is important to note that Tombstone presents a much smaller share in the Los Angeles market than both the fringe and the outside, and therefore even though it is closer to DiGiorno, the switching still reflects partly the size of the brands in the marketplace.
sirable positioning for the two Schwan’s brands, as they are considered by consumers as close substitutes and are likely to cannibalize each other. After contacts with managers at Schwan’s, we learned that the firm resorts to pulsing, alternating the timing of promotional activities between these two brands. It seems that consumers are aware of this fact, also switching consumption between the two brands and perceiving them as close substitutes.

3.6.3 Multimarket findings

Once we have calculated brand switching to DiGiorno in all markets, we can proceed to its analysis across geography. We focus on a relative measure of switching, i.e., the share lost by each brand as a proportion of DiGiorno’s share, more than on the absolute value of share each brand lost to DiGiorno, since we are interested in decomposition of the new brand’s share. We start our multimarket analysis by presenting two boxplot graphs in Figure 3.4. The first graph shows how much variation exists in the switching from each brand, while the second shows the variation in pre-entry share. By comparing the two graphs, it is clear that the random coefficients models allows unproportional draw, i.e., the share lost is not dependent exclusively on pre-entry share.

In average, about 50% of DiGiorno’s share came from outside of the category consumers. Although we cannot disentangle the no-buy option from the delivery pizza option, this high proportion of share is indicative that the advertising campaign used to target outside of the category consumers, such as consumers from delivery pizza, had some

---

10 Note that these two measures can have very different implications. For example, a small loss of share by Tombstone, lets say 3%, in a market where DiGiorno got also a low share, e.g. 4%, will represent a large proportion of its share (75%).

11 The campaign uses the phrase "It’s not delivery, it’s DiGiorno".
impact on the performance of the brand.

In terms of share originated from other frozen pizza brands, Tombstone contributes, in average, to about 10% of DiGiorno’s share, with the level of cannibalization reaching 12% when we add the other Kraft’s brand, Jack’s. The two brands from the main competitor, Tony’s and Red Baron from Schwan’s, represent about 7% and 8% respectively. The proportion drawn from the fringe, is in average about 17%, but with a much higher variance than the other alternatives, probably due to the fact that it is composed of a variety of local brands. Although it is acceptable to think that DiGiorno has stolen some shelf space and resources from Tombstone and Jack’s, it appears that managers at Kraft were successful at reducing the negative impact of the new brand on these two brands, drawing on a much larger scale from competitors.

In order to evaluate the reasonability of these results, we take a direct look at the data, evaluating model free evidence. We choose two markets where the estimated percentage of share of DiGiorno coming from the outside good is extremely different: in Chicago, the model estimates that 93% of DiGiorno’s share came from the outside good, whereas in Denver the same quantity reached only 13%. Given these model estimates, we expect to see a significant growth in frozen pizza consumption in Chicago, consequence of new consumers coming into the category, and a slow growth in Denver, as consumers switched from incumbent brands to DiGiorno. This is in fact so. The data shows us a growth in the category of about 16% in Chicago, more than enough to accommodate DiGiorno’s share in that market of about 8% (share among inside goods). In Denver, the category had a very small positive variation of 3%, which is not enough to fuel the 14% of DiGiorno’s
share and so as the model estimated, its share must have come in large part from the incumbent brands in this market.

To explain the cross-market variation, we collect the switching for all incumbent brands (not including the outside good) and markets, measured as a percentage of DiGiorno’s share to form our dependent variable. As explanatory variables, we include brand factors, such as brand intercepts, which are common across markets, and market specific factors, such as: (1) local pre-entry conditions, such as brand pre-entry shares; (2) firm strategies, which include price and promotion differences between incumbent brands and DiGiorno; and (3) local consumer characteristics, such as population size, percentage of households that are car owners, etc. Results are presented in Table 3.4.

For the brands intercepts, we chose the fringe as base brand. We obtain negative coefficients for the brand intercepts of national brands, showing that their brand name is stronger, enough to reduce the switching to DiGiorno to some extent, when compared to the local brands that form the fringe. Pre-entry share has a positive coefficient, which implies that brands with larger pre-entry share contribute more to DiGiorno’s share. This is a reasonable result, since it is likely that some relationship between size and draw still remains. This variable by itself explains the larger percentage of the variation across markets. In terms of marketing strategies, it seems that they do not have much influence in explaining the cross-market variation of switching. However, it is important to notice that actions such as having higher advertising levels (negative coefficient) may be useful ways to reduce the loss of share to a new brand. Finally, demographics explain a very small proportion of the cross-market variation, with age being the only significant factor.
Markets with older population tend to have higher levels of switching to DiGiorno.

### 3.7 Conclusion

Our paper presents several contributions. We combine into one estimation procedure different dimensions of the aggregate data to identify heterogeneity in preferences from aggregate level data. Our demand description can be used to identify the three sources of market demand, cannibalization, competitive draw and category expansion, through the use of easy to obtain market level data.

The launch of DiGiorno, which was very successful, was accompanied by the motto "It’s not delivery, it’s DiGiorno". This suggests that Kraft was focusing its attention on the delivery market as one of their main targets. Our estimates of the introduction of DiGiorno suggest that indeed the outside good, to which the delivery market belongs, was the main source of DiGiorno’s share. However, there are large differences across markets of how successful DiGiorno was at stealing share from the outside good, and we have shown model free evidence that concurs with the existence of these differences. Of the inside goods, DiGiorno stole most demand from Tombstone, but the total Kraft share in the category went up considerably. Similar to what we observe in the outside good, switching from the inside goods is also very different across markets.

This paper demonstrates that besides market shares, the substitution among brands is not a national phenomenon. It may well be that in one market Tombstone dominates Red Baron and in another the inverse occurs, but both brands may still be viewed (or not) as close substitutes to the new brand. Of the factors that could explain the variation in substitution patters even only among the inside goods, we find that pre-entry share is the
most important. Poignantly, brand specific effects explain little of substitution patterns as do demographics.

In practical terms, we believe our model is helpful to managers in evaluating the impact of new product introductions in local markets. To the extent that this impact is spatially dependent across markets, our model can be used in phased national roll-outs, such as the one used by Kraft to launch DiGiorno, to forecast new product switching, at a pre-entry stage, based on post-entry data from nearby markets.
Table 3.1: Evolution of average shares for some brands in the frozen pizza category in each of the years in the data set.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DiGiorno</td>
<td>0.01</td>
<td>0.04</td>
<td>0.10</td>
<td>0.13</td>
<td>0.13</td>
</tr>
<tr>
<td>Jack’s</td>
<td>0.09</td>
<td>0.09</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
</tr>
<tr>
<td>Red Baron</td>
<td>0.12</td>
<td>0.12</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td>Tombstone</td>
<td>0.22</td>
<td>0.21</td>
<td>0.19</td>
<td>0.18</td>
<td>0.17</td>
</tr>
<tr>
<td>Tony’s</td>
<td>0.12</td>
<td>0.12</td>
<td>0.10</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Totinos</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.10</td>
</tr>
<tr>
<td>Fringe</td>
<td>0.33</td>
<td>0.31</td>
<td>0.31</td>
<td>0.30</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Generated Values</td>
<td>Information Poor Model</td>
<td>Information Rich Model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>-0.50</td>
<td>-0.39</td>
<td>-0.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Display</td>
<td>2.00</td>
<td>1.32</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Brand Intercepts</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digiorno</td>
<td>-1.00</td>
<td>-0.36</td>
<td>-1.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red Baron</td>
<td>-1.50</td>
<td>-0.65</td>
<td>-1.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tombstone</td>
<td>-2.00</td>
<td>-0.99</td>
<td>-2.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tony’s</td>
<td>-2.50</td>
<td>-1.34</td>
<td>-2.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fringe</td>
<td>-1.30</td>
<td>-0.54</td>
<td>-1.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price Heterogeneity</td>
<td>0.50</td>
<td>0.42</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept Heterogeneity</td>
<td>1.50</td>
<td>0.38</td>
<td>1.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pre-Entry Shares</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red Baron</td>
<td>0.15</td>
<td>0.15</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tombstone</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tony’s</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fringe</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside Good</td>
<td>0.56</td>
<td>0.56</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Post-Entry Shares</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DiGiorno</td>
<td>0.13</td>
<td>0.13</td>
<td>0.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red Baron</td>
<td>0.11</td>
<td>0.10</td>
<td>0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tombstone</td>
<td>0.08</td>
<td>0.08</td>
<td>0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tony’s</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fringe</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside Good</td>
<td>0.52</td>
<td>0.53</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Summary Statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_0$</td>
<td>0.05</td>
<td>0.01</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_1$</td>
<td>0.12</td>
<td>0.02</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_2$</td>
<td>0.83</td>
<td>0.97</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Switching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red Baron</td>
<td>0.17</td>
<td>0.17</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tombstone</td>
<td>0.14</td>
<td>0.12</td>
<td>0.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tony’s</td>
<td>0.10</td>
<td>0.09</td>
<td>0.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fringe</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside Good</td>
<td>0.43</td>
<td>0.46</td>
<td>0.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Results for generated data for models without added information about purchase set size (unrestricted model) and with added information (restricted model)
### Table 3.3: Results for the Los Angeles market using the model with traditional moments (information poor model) and with additional moments from purchase set size (information rich model)

<table>
<thead>
<tr>
<th></th>
<th>Information Poor Model</th>
<th></th>
<th>Information Rich Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>St. Deviation</td>
<td>Parameter</td>
<td>St. Deviation</td>
</tr>
<tr>
<td>Price</td>
<td>-0.20</td>
<td>0.05</td>
<td>-0.98</td>
<td>0.09</td>
</tr>
<tr>
<td>Display</td>
<td>2.67</td>
<td>0.15</td>
<td>4.31</td>
<td>0.25</td>
</tr>
<tr>
<td>Brand Intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digiorno</td>
<td>-4.27</td>
<td>1.75</td>
<td>-6.35</td>
<td>0.97</td>
</tr>
<tr>
<td>Red Baron</td>
<td>-4.33</td>
<td>1.29</td>
<td>-5.55</td>
<td>0.44</td>
</tr>
<tr>
<td>Tombstone</td>
<td>-3.51</td>
<td>0.96</td>
<td>-3.41</td>
<td>0.55</td>
</tr>
<tr>
<td>Tony’s</td>
<td>-3.72</td>
<td>1.09</td>
<td>-3.88</td>
<td>0.29</td>
</tr>
<tr>
<td>Fringe</td>
<td>-1.54</td>
<td>0.44</td>
<td>-1.18</td>
<td>0.57</td>
</tr>
<tr>
<td>Price Heterogeneity</td>
<td>0.02</td>
<td>0.15</td>
<td>0.47</td>
<td>0.05</td>
</tr>
<tr>
<td>Summary Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_0(actual = 0.38)$</td>
<td>0.00</td>
<td></td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>$S_1(actual = 0.22)$</td>
<td>0.02</td>
<td></td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>$S_2(actual = 0.40)$</td>
<td>0.98</td>
<td></td>
<td>0.45</td>
<td></td>
</tr>
<tr>
<td>Switching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Red Baron</td>
<td>0.01</td>
<td></td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Tombstone</td>
<td>0.01</td>
<td></td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Tony’s</td>
<td>0.01</td>
<td></td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Fringe</td>
<td>0.15</td>
<td></td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>Outside Good</td>
<td>0.82</td>
<td></td>
<td>0.80</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4: Results for the linear regression, with dependent variable the switching measured as a percentage of DiGiorno’s share

<table>
<thead>
<tr>
<th></th>
<th>(1) Mean</th>
<th>(1) St.D.</th>
<th>(2) Mean</th>
<th>(2) St.D.</th>
<th>(3) Mean</th>
<th>(3) St.D.</th>
<th>(4) Mean</th>
<th>(4) St.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.160</td>
<td>0.017</td>
<td>0.160</td>
<td>0.016</td>
<td>0.130</td>
<td>0.021</td>
<td>-0.265</td>
<td>0.355</td>
</tr>
<tr>
<td>National Posit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jack’s</td>
<td>-0.058</td>
<td>0.034</td>
<td>-0.095</td>
<td>0.032</td>
<td>-0.092</td>
<td>0.036</td>
<td>-0.092</td>
<td>0.031</td>
</tr>
<tr>
<td>Red Baron</td>
<td>-0.099</td>
<td>0.024</td>
<td>-0.101</td>
<td>0.022</td>
<td>-0.069</td>
<td>0.031</td>
<td>-0.100</td>
<td>0.021</td>
</tr>
<tr>
<td>Tombstone</td>
<td>-0.093</td>
<td>0.024</td>
<td>-0.094</td>
<td>0.023</td>
<td>-0.047</td>
<td>0.031</td>
<td>-0.093</td>
<td>0.022</td>
</tr>
<tr>
<td>Tony’s</td>
<td>-0.107</td>
<td>0.024</td>
<td>-0.107</td>
<td>0.023</td>
<td>-0.092</td>
<td>0.029</td>
<td>-0.108</td>
<td>0.023</td>
</tr>
<tr>
<td>Totinos</td>
<td>-0.090</td>
<td>0.0253</td>
<td>-0.097</td>
<td>0.024</td>
<td>-0.116</td>
<td>0.032</td>
<td>-0.097</td>
<td>0.023</td>
</tr>
<tr>
<td>Local Positioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-entry share</td>
<td>1.527</td>
<td>0.241</td>
<td>1.561</td>
<td>0.256</td>
<td>1.561</td>
<td>0.254</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marketing Strat.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price difference</td>
<td>-0.062</td>
<td>0.027</td>
<td>-0.063</td>
<td>0.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature diff.</td>
<td>-0.138</td>
<td>0.106</td>
<td>-0.123</td>
<td>0.105</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Display diff.</td>
<td>-0.021</td>
<td>0.098</td>
<td>-0.017</td>
<td>0.097</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution diff.</td>
<td>0.066</td>
<td>0.074</td>
<td>0.072</td>
<td>0.074</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of Voice</td>
<td>-0.075</td>
<td>0.027</td>
<td>-0.078</td>
<td>0.028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.119</td>
<td>0.231</td>
</tr>
<tr>
<td>% whites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.062</td>
<td>0.651</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.109</td>
<td>0.049</td>
</tr>
<tr>
<td>% Car owners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.075</td>
<td>0.328</td>
</tr>
<tr>
<td>R-Square</td>
<td>0.09</td>
<td>0.21</td>
<td>0.27</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4: Results for the linear regression, with dependent variable the switching measured as a percentage of DiGiorno’s share.
Figure 3.1: Choice probabilities for brands in two different scenarios: (a) strong and heterogeneous preferences; (b) tight and weak brand preferences.
Figure 3.2: Evolution of market shares of main brands in Sacramento and Salt Lake City at the time of introduction of DiGiorno.
Figure 3.3: Perceptual map for models with and without additional information about the purchase set size, for the Los Angeles market.
Figure 3.4: Boxplot graphs of (a) switching to DiGiorno and (b) pre-entry share of incumbent brands, for 64 markets.
4 Dealer Competition in the Auto Industry

4.1 Introduction

One of the key concerns of store managers is to accurately identify stores that directly compete for the same customers and to measure the consequent strength of competition effects on demand. The first issue of identifying potential competitors is by itself a complex problem to solve, especially when a store sells an extended product line and different sets of competitors exist for each product carried by the retailer. A clarifying example can be provided using the automobile industry. In this category, car dealerships face competition from two different sources: other stores that sell the same car brand and competitors that offer a close substitute model but from a different brand. In the first case, product lines are likely to be the same and therefore the two car dealers compete across all consumer segments. In the second case however, the existence of variation in the product lines of each brand limits the competition of dealers to consumer segments where they coincide. For instance, a Lexus dealer selling the Lexus LX model, competes in the Luxury SUV segment with neighboring Lexus dealers but also with Porsche and Hummer retailers. Looking at the high-end sedan segment, dealers selling the Lexus LS will likely worry about the other Lexus retailers as well but less concerned with the Porsche or Hummer dealers, since these brands do not provide a car model for this segment.

Once identification of potential competitor stores has been done, the issue is now measuring the strength of competition on demand, which will naturally have an impact on a variety of managerial decisions, such as setting prices, setting rebates or providing
credit. Factors like the geographical location of stores have a strong impact on both the definition of the competitor set and on the intensity of competition. In some cases, stores that are close neighbors likely vie to attract the same customers and consequently will be direct competitors, while stores located far apart, even though possibly targeting similar consumer segments with similar products, are likely not to see each other as competitors since their target demand is located in different areas. In other cases, consumers have stronger incentives, such as the high price of the good, to travel longer distances in search of a better alternative.

It is within this framework that this paper presents its contribution. Our main objective is to provide a method that can be useful at, on one hand, identifying the competitor set of a store (in our case, of a car dealership), and on another hand, measuring the intensity of competition across time and distance for each segment where the dealership is a participant. Our measure of competition will be price-demand elasticity, that is, we will quantifying the impact of price changes by competitor dealers that vary in location and brand on the demand of a specific dealership.

The information available to us to address this question is a unique dataset covering: (1) data on individual level car transactions, such as price, rebates, date, brand and model of car; (2) data on car retailers, such as location and brand sold; (3) data on consumers, such as zip code location, complemented with census macro data on demographics and income characteristics. Our model performs the estimation of demand parameters from individual level transaction data using maximum likelihood and corrects for potential endogeneity of prices as well as accounts for retailer characteristics that influence the
consumers’ decision but that are unobservable to the researcher. We also provide the estimation of supply parameters, using observed manufacturers’ prices to the retailer and final prices charged by dealers to the customers. Using this methodology, we are also able to make some recommendations on the optimal location of a new car dealer.

We find a number of interesting results. Across segments, we observe that high-end segments present lower price sensitivity, which results in a smaller competitor set and lower levels of competition through price between car dealerships. We also find that consumers appear to have similar sensitivity to distance across all segments, although the price paid for the final car is quite different. In terms of identification of competitor sets, we find in most cases dealers primarily compete with (1) other dealers that offer the same brand and (2) dealers of other brands that are perceived as close substitutes. In most cases, we find that the brand effect dominates the distance effect, observing a stronger impact of price changes by these two types of dealers on demand than the impact of other dealers located closer but selling brands that are not perceived as substitutes.

The paper continues by describing the positioning of the paper in the literature in the next section. In section 3, we present our model approach. Section 4 describes the data. The estimation algorithm is presented in section 5 and the results are analyzed in section 6. Section 7 concludes.

4.2 Literature Review

When positioning this paper in the literature, we will concentrate on two streams of research, the first one concerning spatial competition and a second one focusing on competition in the auto industry.
There has been increasing interest in modeling spatial competition, mainly as a result of recent availability of spatially distributed data. Some recent examples of this literature area are Pinkse et al. (2002) and Chan et al. (2005) with applications in the gasoline market, in the United States and Singapore respectively; Davis (2001) studies competition in movie theaters; and Duan and Mela (2006) focus on developing a map of latent demand patterns for apartments across space. These papers all use aggregate level demand. Because of the unique circumstances of having individual level data, our paper is methodologically closer to Gaynor and Vogt (2003). The latter studies research competition across hospitals, estimating a structural model that uses individual-level data in estimation and accounts for the impact of distance between consumer and hospital locations on the demand for medical services. The main purpose of their paper is to study the impact of a merger on competition of profit vs. non-profit hospitals and on consumer welfare.

Thomadsen (2005) presents another example of spatial competition among retailers, focusing on the fast food industry. He uses price and cost variation, retailer location and demographic data, combined with firms’ first-order conditions to estimate demand and supply parameters without the knowledge of quantities. His results show that consumers have a low willingness to travel in the fast food industry. Estimates show that consumers are willing to travel about 1/3 of a mile to save about one dollar on a meal. The author also uses counterfactuals to show how spatial differentiation has a strong impact on price, showing how the location of a retailer in a mall (a center of attraction of consumers) or very close to a competitor has a positive or negative impact on price and on mark-ups. We
partly study similar questions about spatial competition and differentiation. However, we do so in an industry where consumers are known to be willing to travel longer distances and most importantly where each retailer has a larger product line that targets different sets of consumers that may differ in their price and distance sensitivity. In addition, our model covers other several aspects, both substantive and methodological, not included in previous papers on spatial competition.

Substantively, we propose a method that identifies the set of competitors that have an impact on a specific store demand, and measure the intensity of competition across distance, time and segments. We also deal with the limitation that the only consumer-specific information available to us is the location of the consumer. To overcome the lack of individual level demographic information about consumers, we combine a large micro data set with transactional information with macro information from the US census about income and other demographic data at the location of each consumer.

Regarding the auto industry, Berry et al. (1995 and 1998), study the automotive industry in two complementary methods. The authors start by developing a model that analyzes demand and supply of differentiated products using aggregate-level data (1995) and then expanded their methodology to combine micro and macro data (1998). Among other results, they are able to produce demand elasticity of price and other observed attributes and find considerable variability across types of cars and models. For instance, in the first paper, they find that cheaper cars present a demand that is considerably more elastic than more expensive cars. We observe a somewhat similar result, with high-end segments showing a lower price sensitivity to price than low-end segments. In a related
paper, Petrin (2002) also applies a random coefficients model to the auto-industry to analyze the impact of the introduction of the mini-van on consumer welfare. He shows that combining market level data about shares, prices and other car characteristics with purchaser aggregate information easily available in the Consumer Expenditure Survey helps improve parameter estimates, especially in reducing their variance\(^1\).

The main difference between our paper and the mentioned literature on the car industry is that our approach is adequate when there is individual level availability of transactional data, instead of the aggregate level data. Our data has a deeper level of detail, since it is the individual level data that would in aggregation result in the data set available to Berry et al. (1995) or the main data set on shares of Petrin (2002). Specifically, the individual information about the location of both dealers and consumers allows us add to the understanding of how (1) demand for dealers and brands vary depending on spatial location; and (2) own- and cross-price elasticities vary across dealerships, for different segments in the market. To our knowledge, our model is one of the first to present a method able to recommend the location of a future dealership, based on both demand and supply reactions estimated using data at the individual level.

4.3 Model

Our model is based on the behavior of consumers that search and purchase a new car and retailers that offer a certain brand of vehicle in the automobile category. In this section, we start by describing the primitives of demand (consumers) and follow by presenting the details on the supply side (retailers).

\(^1\)Another example of a similar approach is available in Goldberg (1995).
4.3.1 Demand

We choose to explain competition among inside goods, that is, explain how and why demand for cars varies across different dealers, instead of between dealers and other transport alternatives. Our assumption is basically that consumers have decided to buy a car and are now looking for different alternatives within the category. Consequently, in our model approach, we do not have the need to include the outside good alternative. Consumers continue their search process by deciding which car segment is more adequate to their own needs, based on some unobserved (to the researcher) reasons and consumer characteristics, such as family size, income or safety concerns. These segments are defined a priori, based in a classification from J.D.Power and Associates and are as follows: luxury SUV, fullsize SUV, mini SUV, upper midsize and lower midsize cars. We assume that consumers know what segment is more adequate to them and will look for alternatives and eventually choose which brand to purchase and which dealer to buy it from.\footnote{The main implication of these assumptions is that we do not require to model the choice between inside goods and the outside good and that segments are assumed to be independent. We are currently expanding our study to not only include the outside alternative and but also allow for substitution across segments, using a nested logit approach.}

For all individuals, utility is defined for each car model of a brand at a specific dealer. It depends upon consumer-specific, car model-specific and dealer-specific characteristics (our formulation will include both characteristics that are observable and unobservable to the econometrician).

Our formulation of utility is simplified by two characteristics of the car industry: first, more than 90% of the dealers sell only one car brand. Second, as described in the previous paragraph, we assume that consumers that are searching for a certain car in a determined
segment (such as luxury SUV or upper midsize sedans) are unlikely to change to other car segments at the time of purchasing the vehicle. Given these two assumptions, the our model has the following two characteristics: (1) segments are modeled as independent; (2) given the almost perfect one-to-one relation between dealer and brand, a consumer that chooses a dealer is in reality also choosing a brand.

For individual $i$ ($i = 1, ..., I$), utility of retailer $r$ ($r = 1, ..., R$) at time $t$ ($t = 1, ..., T$) in segment $s$ ($s = 1, ..., S$) is given by the following expression (the subscript $s$ for segment is not included for simplicity):

$$U_{irt} = \alpha_{br} + \lambda \cdot p_{rt} + \beta \cdot x_{rt} + \delta \cdot \exp(-\rho \cdot d_{ir}) + \gamma \cdot p_{rt} \cdot h_{zi} + \xi_{rt} + \varepsilon_{irt} \quad (4.1)$$

where:

$\alpha_{br}$ is a brand specific intercept; each retailer $r$ that sells brand $b$ ($b = 1, ..., B$) will have similar $\alpha_{br}$;

$\lambda \cdot p_{rt}$ measures the impact of price on the utility. $\lambda$ is the price coefficient that measures price sensitivity and $\bar{p}_{rt}$ is the average price of vehicles sold by retailer $r$ within segment $s$ during period $t$;

$\beta \cdot x_{rt}$ measures the influence of other retailer specific characteristics, $x_{rt}$, such as average rebate levels and APR, with coefficient $\beta$;

$\delta \cdot \exp(-\rho \cdot d_{ir})$ quantifies the impact of distance between individual $i$ and retailer $r$ on utility. $d_{ir}$ denotes the distance between individual $i$ and retailer $r$, and the two parameters $\delta$ and $\rho$ measure respectively utility sensitivity to distance and the degree of how fast it declines over distance;
\( \gamma \cdot p_{rt} \cdot h_{zi} \) measures observed heterogeneity to price, quantified by parameter \( \gamma \). 

\( h_{zi} \) are some demographic characteristics, such as income or age (variables \( p_{rt} \) and \( h_i \) are demeaned), at the zip code \( z_i \), where individual \( i \) resides.

\( \xi_{rt} \) includes demand unobservable variables to the econometrician but observable to the retailer who sets the price in period \( t \);

\( \varepsilon_{irt} \) is an error term that is unobserved to both the retailers and the econometrician and is assumed to have the extreme value distribution.

In each segment, consumers choose retailer \( r \) that maximizes their utility. It is important to note that in this framework, earlier literature (McFadden et al. 1977, Hausman and Wise 1978, Berry et al, 1995) have shown that the interaction between consumer specific-characteristics, in our case consumer location and income, and product/dealer characteristics, such as dealer location and price, determines the substitution patterns estimated by the discrete choice models. With larger variation in the consumer-specific characteristics, similar alternatives become closer substitutes (Berry et al. 1998). Consequently, if consumers are sensitivity to distance, our model will estimate that dealers located at closer distances will be stronger substitutes.

4.3.2 Supply

Each dealer is assumed to maximize profits, by selling each of its vehicles at a single price. The retailer buys the vehicle from the manufacturer at wholesale price \( W_r \) and has additional variable costs of \( C_r \). It faces demand \( D_r(\bar{p}) \), where \( \bar{p} \) is the vector of prices charged by all retailers to the consumers. In each segment, at time \( t \), dealer \( r \) has the following profit function:
\[
\pi_{rt} = (p_{rt} - W_r) \cdot D_r (\bar{p}) - C_{rt} (D_r (\bar{p}), A_{rt}, \nu_{rt})
\] (4.2)

where \( A_{rt} \) and \( \nu_{rt} \) are respectively observed and unobserved cost-shifters for retailer \( r \). A dealer that maximizes profit playing a Bertrand-Nash pricing game will have the following first-order condition to define optimal price:

\[
p_r = W_r + \frac{\partial C_r}{\partial D_r} - \frac{D_r}{\partial p_r}
\] (4.3)

In a scenario where segments are not independent and there might be substitution between them, managers will be able to coordinate pricing across segments. Define \( \Theta \) as a \((R \times S) \times (R \times S)\) matrix, where \( R \) and \( S \) are respectively the number of retailers and segments, and \( \Theta_{rq} = 1 \) if retailer \( r \) sells cars in segment \( q \). The pricing equation is defined by:

\[
p_{rs} = W_{rs} + \frac{\partial C_{rs}}{\partial D_{rs}} + \left[ \Theta \otimes \left( \frac{dD_{rs}}{dp_{rs}} \right) \right]^{-1} D_{rs}
\] (4.4)

where \( p_{rs} \) is a \((R \times S) \times 1\) vector that stacks all retailer prices across all segments, \( W_{rs} \) is a \((R \times S) \times 1\) vector that stacks all manufacturers across all segments (and similar for other variables), and \( \otimes \) is the element-by-element Hadamart matrix multiplication.

4.4 Data

We combined several data sets in order to be able to apply the described model. The first dataset was obtained from a marketing information firm which prefers to remain anonymous. It includes details about one million car transactions, from 2001 to 2005:
car brand and model, car price, model year, zip codes of the location of dealer and of consumer (assumed to be the residence location), manufacturer recommended price, date and type of transaction (credit, leasing, cash), duration of credit and APR and how long the car was in stock before being sold to the customer.

Our second dataset is based on information contained in the U.S. Census. It covers demographic characteristics of population in the zip codes included in the first dataset, such as income levels and population density. In seldom cases where there is information missing about a zip code included in our transactional dataset, we assign the data values of the closest zip code with available information.

Finally, we collected information about the location (zip code latitude and longitude) of both retailers and consumers from the website zipinfo.com. We now give more details about these three datasets.

In Table 4.2 and 4.3, we present the relative size and the number of dealers that compete in each of the segments between 2001 and 2004, as well as the evolution of the total number of transactions that occurred in California for these segments. The market expanded from about 150,000 transactions in 2001 to more than 220,000 in 2003. It then suffered a decline to about 200,000 units sold. Part of the positive change that we observe in sales is a direct consequence of the development of the luxury SUV segment, that had a growth in share from 8.2% to 16%, with the number of dealers competing in this segment doubling from 90 to 180. Most of the other segments present a stable evolution. It is important to note that the upper midsize sedans suffered a strong decline in 2004, although the number of dealers participating in the segment has not change significantly.
Table 4.1 displays the average price of the cars sold in each of the segments available in our data. We observe a clear price differentiation across segments, with prices averaging about $45,000 for the high-end segment luxury SUV and about $19,000 for the low-end segment lower midsize sedans, with most firms competing in multiple segments. Only two segments present similar average price, the mini SUV and the upper midsize sedans. However, these two segments are differentiated by the basic "shape" of the car - sedan vs. SUV. Prices seem to be stable across time, with the highest increase showing in the luxury SUV segment with a 10% positive change. Others segments presented a total increase in price of about 5% over these 4 years, while only the lower midsize sedan segment presented a drop in price, of about 3%.

For our utility formulation, we need to compute the average retailer price $p_{rt}$, for all retailers and all segments. We need to use average prices charged by dealers across a period of time because for each transaction, we only have access to the price paid by the consumer at one dealer, but we do not know what price this specific consumer would be charged at competing dealers. We solve this problem by using the average price $p_{rt}$. The average is across all transactions of the same car model over a certain period of time. In our case, we choose an yearly interval of time.

An interesting feature of our data is that we know the location of both the consumer and the dealer for each transaction. On average, consumers were willing to travel about 20 miles to 30 miles to buy a car. See a Figure 4.1 for histograms for three segments. In most segments, we can see a similar distribution of traveled distance, with about 3/5 of the total consumers traveling less than 20 miles. It is worth noting that consumers travel...
slightly more for a high-end car, such a luxury SUV or a full-size SUV, mostly likely due to the fact that a lower number of dealers that carry vehicles in these segments (see table 4.3). This relation between distance and number of dealers is more evident in the luxury SUV segment, where, from 2001 to 2004, the median distance as dropped almost 20%, from 14 miles to 12 miles, when the number of dealers almost doubled.

In Figure 4.2, we provide examples of the distribution of car dealership for two segments, compact SUVs and mini SUVs, for 2004. The size of the circles is a function of dealer sales. It is natural to observe that dealers tend to concentrate near population centers, such as Los Angeles, San Francisco, Sacramento and San Diego, in both segments. We also observe some differences across these two segments. First, a considerable number of Toyota dealers carry compact SUVs but do not carry the fullsize alternative, which is likely to increase competition levels in the first segment. Second, although we see significant correlation in the sales of both segments across retailers, we also note that some retailers that have large sales in one segment, are smaller competitors in the other. For instance, in San Francisco, we observe about five large dealers in the compact SUV segment, but only one in the mini SUV segment.

Our model is estimated using part of this data. We focus our attention on the southern part of California, by including all transactions completed at dealerships located in zip codes 90000 to 93000. This area includes all the Los Angeles and San Diego areas and comprises slightly more than half of all the observations in our original data set. Although not extremely necessary, limiting the data to these regions reduces the computation demands on our estimation algorithm, while still maintaining a very complete data set on
dealer competition in a smaller geographical area.

4.5 Estimation

4.5.1 Demand parameters

Our choice of estimation approach is a result of some of the objectives of the paper and assumptions described previously. We are concerned with the identification of different competitor sets that a dealer faces in each segment in which it participates. Combined with the assumption that consumers that have need for a car will identify the type of car (segment) in which they are interested and then choose between alternative brands and dealers within that segment, a logical solution is to divide the data into segments and estimate our model for each segment separately. We also divide the data sets into yearly periods. Despite losing some information by not pooling the data together, the main advantage of this approach is that evolution of the parameters across time is completely dominated by the data without being constrained by a parametric form. This approach also constitutes a significant reduction of computational burden.

Our estimation procedure is based on techniques proposed by Berry (1994), Berry, Levinsohn and Pakes (1995) and similar to Gaynor and Vogt (2003). Our algorithm has several stages that we now describe in detail. We start by estimating the demand parameters. In this part of the estimation, we are concerned with possible endogeneity, between the demand unobservables (to the researcher) $\xi_{rt}$, which include retailer-specific quality elements, and the average price that retailer charges, $p_r$.\footnote{If the unobserved demand characteristics are positively correlated with price, estimates of the price parameter will be biased towards zero (Trajtenberg, 1989).} We account for this possibility using the following methodology presented in Gaynor and Vogt (2003). Note
that the utility is defined in equation 4.1:

\[ U_{irt} = \alpha_{br} + \lambda \cdot p_{rt} + \beta \cdot x_{rt} + \delta \cdot \exp(-\rho \cdot d_{irt}) + \gamma \cdot p_{rt} \cdot h_{zi} + \xi_{irt} + \varepsilon_{irt} \]

Define the utility of retailer \( r \) to individual \( i \) in the following way (the time subscript \( t \) is removed for simplicity, since our estimation is done separately for each time period):

\[ U_{ir} = \phi_r + \delta \cdot \exp(-\rho \cdot d_{ir}) + \gamma \cdot p_r \cdot h_i + \varepsilon_{ir} \quad (4.5) \]

Using this formulation of utility, we are able to estimate the consumer-retailer interaction parameters \( \delta, \rho \) and \( \gamma \) and a retailer-specific mean effects parameter \( \phi_r \) using maximum likelihood. \( \phi_r \) includes all factors that are constant across individuals, including price and demand unobservables which give rise to endogeneity. We obtain these parameters by regressing \( \phi_r \) on the observable retailer characteristics, that is:

\[ \phi_r = \lambda \cdot p_r + \beta \cdot x_r + \xi_r \quad (4.6) \]

The unobservable aspects of the utility \( \xi_r \) are now the errors of this regression. The parameters of this regression are estimated using the instrumental variables approach, since we want to control for the possible correlation between the covariate \( p_{rt} \) and the errors \( \xi_r \).

### 4.5.2 Demand instruments

Our objective in choosing a good set of instruments is selecting variables that are correlated with the average price offered by each retailer, but uncorrelated with unobservable
characteristics of the retailers $\xi_{rt}$. These demand unobservables has two main components. First, it can include dealer characteristics that used by consumers as signal of quality, such as consumer service and reputation of the dealer. Second, the choice to assortment of car extras by the retailers is also a distinguishing characteristic that is not observed by the researchers and is naturally correlated with the dealer’s quality image. We assume that the remaining utility components, such as dealer characteristics and taste for distance is mean independent from the demand unobservables.\(^4\)

Formally, let $z_{rt} = [x_{rt}, w_{rt}]$, where $x_{rt}$ includes all observed retailer-specific characteristics except price, and $w_{rt}$ is the set of instruments used. We have information about the wholesale price of the manufacturer, which can be a good candidate as an instrument to prices. However, in our case, both the final price charged to the consumer and the manufacturer wholesale price are a function of the assortment of extras and accessories that are included in each car, which are not observed in our data. This fact prevents us from directly using the average manufacturer price to the dealer as instrument, since it is also correlated with the demand unobservables $\xi_{rt}$. To account for this, we instead compute the average manufacturer wholesale price across all dealers that sell the same car brand $b$. This variable measures the manufacturer price of an average vehicle of brand $b$. This quantity eliminates all retailer differences within a certain brand and while still strongly correlated with average price, correcting for both components of endogeneity, quality signals and extras assortment. Since we include brand intercepts in our utility function, $\xi_{rt}$ is not correlated with costs that vary across brands.

\(^{4}\)This assumption is frequently used. For examples, see Bresnahan (1987), Berry et al. (1995) and Gaynor and Vogt (2003).

97
Besides average manufacturer price, the instruments \(w_{rt}\) also includes the number of days that the car remained in stock at the dealership, which is used as a proxy for inventory cost.

### 4.5.3 Supply parameters

We assume that dealership managers want to maximize their profit. Using available information about prices and costs, we are able to estimate a pricing equation that informs us about the impact of several factors, such as demand and other marginal costs besides the manufacturer's price have on price. For the supply, we base our estimation on the first order condition of equation 4.3:

\[
p_r - \frac{D_r}{\partial p_r} - MP_r = \frac{\partial C_r}{\partial D_r} \tag{4.7}
\]

Both \(p_r\), the average price charged by the dealer, and \(MP_r\), the manufacturer’s price charged to the dealer, are available in our data, which will already give a good approximation to the mark-up. From the demand side, we can obtain \(D_r/\partial p_r\), completing the left-hand side of the equation. For the right-hand side, we use a linear form to approximate to the marginal costs that occur beyond the manufacturer’s price. We obtain:

\[
p_r - \frac{D_r}{\partial p_r} - MP_r = c_0 + c_d \cdot D_r + c_a \cdot A_r + \zeta_r \tag{4.8}
\]

As mentioned previously, \(A_{rt}\) and \(\nu_{rt}\) are respectively observed and unobserved cost-shifters for retailer \(r\).
4.5.4 Supply instruments

We cannot use OLS estimation to obtain the coefficients in equation 4.8, since it is likely that there are some unobservable cost shifters that are correlated with the demand $D_r$, especially since this demand $D_r$ is a function of the price charged by the dealers. For instruments of $D_r$ we follow Gaynor and Vogt (2003), by estimating demand of retailer $r$ without including the impact of price, that is, the utility function is defined as

\[ U_{irt}^{IV} = \alpha_{br} + \beta \cdot x_{rt} + \delta \cdot \exp(-\rho \cdot d_{ir}) + \varepsilon_{irt} \tag{4.9} \]

Given the estimate of parameters obtained initially, the instrument for demand $D_r^{IV}$ is just the aggregation of the logit choice probabilities across all consumers.

4.6 Results

We start by describing the general results of the model and follow by giving more details about similarities and differences across segments, time and brands.

4.6.1 Demand

Our model seems to fit the data well. In Figure 4.3, we show the actual and estimated market shares of the different dealers for two segments, luxury SUV and lower midsize cars. We find that a large portion of the variation in shares across retailers is captured by our formulation of equation 4.1.\textsuperscript{5}

\textsuperscript{5}Note that the estimated shares are computed with equation 4.1 and not equation 4.5. The later equation is only a first step used to estimate the parameters of the first equation. Given the formulation of equation 4.5, shares estimated using this equation almost perfectly fit the actual shares, as it is intended. Berry et al. (1995) contraction mapping methodology has the same result, if one includes the unobservables in the estimation of the shares.
We also computed the hit-rate of correctly identifying the retailer chosen by the individuals. Across segments and years, the hit-rates fluctuate between the 15% to 30% within sample. This result provides a good lift when compared to random prediction, given such a large set of alternatives, with smaller hit rates for segments where a larger number of dealers compete.

Table 4.4 presents the results for the demand parameters for 2003, across five different car segments. We exclude brand intercepts, since each segment has its own set of competing brands. As expected, for every case, the estimated price coefficient is negative and significant in most cases. However, we observe considerably variability across segments, which translates into significantly different own-price semi-elasticities. In higher quality segments, such as the luxury SUV and the fullsize SUV, consumers are less sensitive to price, with an almost inelastic demand in the fullsize SUV segment. Lower priced segments, such as the mini SUV and the lower midsize car segments, present elastic demand, with demand decreasing 6.9% and 8.7% respectively with a change of $1000 dollars in the average price offered by the dealer. We also observe significant price heterogeneity in almost all segments, when price interacts with average income of consumer zip codes, reflecting that higher income zip codes are likely to buy more expensive cars.

Distance between the dealers and consumers play an important part in the decision of choosing where to buy a car. Consumers are very sensitive to distance, without considerable difference across them. This seems reasonable, since we observe that in average

---

6 The choice rule that we used to predict choice was "consumers choose the retailer that provides the maximum utility, as estimated by our model".

7 We follow Berry, Levinsohn and Pakes in the computation of semi-elasticity: instead of 1% change in price, we note the change in demand that results from a $1,000 increase in price. The results of using the earlier option are not significantly different from the ones presented here.
consumers travel similar distances to purchase a car across segments.

We now analyze in more detail the results for own-price elasticity. We start by plotting the own-elasticity measure against distance, as exemplified in Figure 4.4. This figure was created by selecting two dealerships, in this case both selling Honda and both competing in two different segments: the mini SUV and upper midsize segments. Each dot in a graph represents the average own-price semi-elasticity for consumers located at distance intervals of 5 miles. We start at 0-5 miles and extend our analysis until the 145-150 miles range.

The figures show that consumers that are located further away "need" to be more sensitive to price, to accept traveling longer distances, especially in the more price sensitive segment. We also observe that even within the same segment and brand, two dealers have different elasticities, due to the interaction between the dealer’s and consumer characteristics, such as location, price or income.

4.6.2 Supply and dealer competition

In the lower part of table 4.4, we show the estimation parameters for the supply side. The parameter values for brand intercepts are once again not included, but it is important to note that there are significantly different marginal costs among the brands in most of the segments. In the covariates displayed, it is important to note a negative significant coefficient of days in stock on the marginal costs of dealers in the SUV segment. A potential explanation for this is that dealers that have larger spaces are the ones that hold a larger number of vehicles that stay in the stock for longer periods. These dealers can disperse their costs across a larger number of vehicles and present lower marginal costs.
We do not observe the same negative impact in other segments.

Our model is also able to estimate the mark-up of each dealer. In Figure 4.5 we show the total cost and mark-up for all the dealers that compete in the lower midsize sedan, for 2003. Adding these two values together totals the average price charged by that dealer to the consumer. Dealers are identified by the brand they sell, showed on the x-axis. They are ordered from left to right, with the retailer with the smallest margin in the furthest left.\(^8\) With some exceptions, ordering the dealers by mark-up also means ordering them by brand. Volkswagen dealers present consistently the highest margin of this segment, while Hyundai and Kia display the smallest. This ordering of dealer’s margins may be a result of either differences in the costs or the ability to charge a higher price due to brand equity. We find that the it is the former reason that dominates, since there is no significance differences of prices across brands. In fact, in most cases, the price of Volkswagen cars is not significant higher than competitors in this segment, with the cheapest observed price belonging to a Volkswagen dealer.

To analyze dealer competition, we continue by computing the cross-price semi-elasticities between dealers. We present an example in figure 4.6. In this figure, we present the case of two Hyundai dealers. Each dot represents the effect that a price increase of a competitor selling a certain brand has on each of these two dealers. We display that amount against distance between all other competitor dealers and the two retailers being studied. The first dealer is located close to the city limits of Los Angeles, while the second is situated about 90 miles from the city. One of the differences between the two dealers is the number of competitors located within a close radius. The first retailer has about 20 competitors

\(^8\)Mark-ups range from $500 to $1400.
within a 30 miles distance, while the second has only 7 (note that the scale of the x-axis is
different for the two figures). Consequently, the manager of the city dealership has to be
concerned about a larger number of competitors than the one located in the country-side.

If we study the figure about the city dealership in more detail, we can see that the
stronger competition comes from two main groups of dealers: first, dealers that sell the
same brand, located about 7 and 12 miles away; second, dealers from the Volkswagen
brand, in a distance range between 5 and 35 miles. These two brands seem to be perceived
by the consumer as direct substitutes. This contrasts with the effect on Hyundai demand
of price changes by Chevy dealers located 15 miles away, where we observe an almost
negligible impact on demand.

This information can be used as a tool to identify the main competitors of each dealer.
For instance, assume that we consider as competitors dealers that show a cross-price
elasticity above 0.5. In each case, we are able to pin-point who are the dealers that satisfy
that condition, where they are located and the brand they are selling. In our example, for
the Hyundai city dealership, this would mean that its most important competitors would
be six dealers with the highest cross-elasticity, two of them selling the same brand and
the four remaining selling the Volkswagen brand. This information can be used in both
pricing and communication strategies that would emphasize specifically the advantages
that dealer has over other those dealers and of the Hyundai cars over Volkswagen cars.

4.6.3 Optimal location of a new dealership

Our model is based on primitives and considers optimization decisions of both demand
and supply sides of the industry. With this framework, we are able to evaluate a number of
policy changes, using counter-factual experiments on a number of decisions by the dealer managers and measure the consequent impact on consumer choices and pricing decisions of dealers. In this section, we study the specific case of the creation of a new dealer that will sell one of the incumbent brands.

We assume that a dealer manager decides to open a new Hyundai dealership, and that he will carry cars in two segments, the mini SUV and the lower midsize. There are several decisions that he must make. Namely, he has to decide on a location for his dealership and the price he is to charge in the segments where Hyundai has vehicle models.

In order to predict demand for the dealer in all alternative locations, we start by looking at the pricing decision. We assume that the dealer will be able to obtain the cars from the Hyundai manufacturer at the average price (across all Hyundai dealers). We also assume that he has average additional marginal costs. We then input these quantities in the estimated pricing equation (4.8) and are able to compute the dealer’s optimal price.

Once we have the price, we can then plugged it in the demand side. Using our model, we then can estimate demand and prices for all retailers, including the one for which we are simulating entry. The resulting profit for a sample of zip codes in the Los Angeles area is presented in Figure 4.7. We present profit for both segments and total profit. We find that in the lower midsize, there is not much variability in the profits across locations, while in the mini SUV, there are definitely some locations that present themselves as good options. This is a result of the interaction of all aspects of the model, such as consumer and other dealers location and tastes for cars and prices. Several zip codes seem to be

---

9 It is possible to extend this analysis to consider the choice of brand as endogeneous as well. We would need to apply the strategy described here to evaluate profits to all brands and locations (instead of just locations). The dealer would decide to enter in the best brand-location pair.
appropriate locations for the dealership in these area, such as 90034, 90046 or 90065.

4.7 Conclusion and Future Research

This paper presented an approach to the problem that a retailer, in our case a car dealer, has in identifying competitors that sell the same or substitute brands at other locations and in measuring the strength of competition across distance, segments and brands. We show that there is considerable variability in both own- and cross-price elasticity across segments, with high-end segments such as the luxury SUV and the fullsize SUV segments presenting almost inelastic demand to price. Dealers must be aware of this fact when negotiation the final price with the consumers, since some segments respond better to price reductions then others. We also find that the travel distance that a consumer has travelled may be an indicator of how sensitive he is to price.

We present examples of how different dealerships have very different competitor sets, across distance and that in some occasions, dealers that sell other brands are in fact a more direct "threat" than dealers of carrying the same brand.

In terms of future research, the main limitation of this paper is the separation of demand by segments. An alternative approach that we are currently applying is to pool all the data together across segments and apply a nested logit formulations, where consumers will choose initially the segment and then the dealer within that segment. This formulation would have implications on the supply side as well, as retailers would then coordinate strategies across the different car segments.
### Table 4.1: Average price between 2001 and 2004, for each car segment

<table>
<thead>
<tr>
<th>Segment</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury SUV</td>
<td>42,294</td>
<td>44,117</td>
<td>47,559</td>
<td>46,589</td>
</tr>
<tr>
<td>Full Size SUV</td>
<td>36,269</td>
<td>37,077</td>
<td>37,858</td>
<td>38,070</td>
</tr>
<tr>
<td>Compact SUV</td>
<td>27,564</td>
<td>28,111</td>
<td>28,722</td>
<td>28,859</td>
</tr>
<tr>
<td>Mini SUV</td>
<td>21,004</td>
<td>21,891</td>
<td>21,449</td>
<td>21,642</td>
</tr>
<tr>
<td>Upper Midsize</td>
<td>21,859</td>
<td>22,180</td>
<td>22,169</td>
<td>22,756</td>
</tr>
<tr>
<td>Lower Midsize</td>
<td>19,957</td>
<td>19,950</td>
<td>19,170</td>
<td>19,497</td>
</tr>
</tbody>
</table>

### Table 4.2: Segment size (quantities sold) between 2001 and 2004, for each car segment

<table>
<thead>
<tr>
<th>Segment size as % of total transactions</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxury SUV</td>
<td>8.2%</td>
<td>8.1%</td>
<td>9.8%</td>
<td>16.0%</td>
</tr>
<tr>
<td>Full size SUV</td>
<td>16.4%</td>
<td>14.1%</td>
<td>16.7%</td>
<td>16.4%</td>
</tr>
<tr>
<td>Compact SUV</td>
<td>21.4%</td>
<td>23.8%</td>
<td>22.0%</td>
<td>23.4%</td>
</tr>
<tr>
<td>Mini SUV</td>
<td>10.1%</td>
<td>11.9%</td>
<td>12.1%</td>
<td>10.9%</td>
</tr>
<tr>
<td>Upper Midsize</td>
<td>36.7%</td>
<td>36.9%</td>
<td>33.5%</td>
<td>28.8%</td>
</tr>
<tr>
<td>Lower Midsize</td>
<td>7.1%</td>
<td>5.1%</td>
<td>5.9%</td>
<td>4.4%</td>
</tr>
<tr>
<td>Total number of transactions</td>
<td>154,426</td>
<td>200,441</td>
<td>225,824</td>
<td>197,498</td>
</tr>
</tbody>
</table>

Table 4.1: Average price between 2001 and 2004, for each car segment

Table 4.2: Segment size (quantities sold) between 2001 and 2004, for each car segment
<table>
<thead>
<tr>
<th>Retailer</th>
<th>Luxury</th>
<th>Fullsize</th>
<th>Mini</th>
<th>Upper</th>
<th>Lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUV</td>
<td>174</td>
<td>232</td>
<td>72</td>
<td>264</td>
<td>23</td>
</tr>
<tr>
<td>SUV</td>
<td>99</td>
<td>224</td>
<td>385</td>
<td>298</td>
<td>22</td>
</tr>
<tr>
<td>SUV</td>
<td>147</td>
<td>224</td>
<td>224</td>
<td>407</td>
<td>264</td>
</tr>
<tr>
<td>SUV</td>
<td>180</td>
<td>264</td>
<td>232</td>
<td>432</td>
<td>264</td>
</tr>
</tbody>
</table>

Table 4.3: Number of dealers in California between 2001 and 2004, for each car segment

<table>
<thead>
<tr>
<th>Retailer</th>
<th>StD</th>
<th>StD</th>
<th>StD</th>
<th>StD</th>
<th>St.D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>-1.70</td>
<td>0.68</td>
<td>-0.56</td>
<td>0.61</td>
<td>-10.7</td>
</tr>
<tr>
<td>APR</td>
<td>-0.05</td>
<td>0.05</td>
<td>-0.00</td>
<td>0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>Income</td>
<td>-0.35</td>
<td>0.47</td>
<td>-0.07</td>
<td>0.42</td>
<td>1.18</td>
</tr>
<tr>
<td>Area</td>
<td>-0.04</td>
<td>0.08</td>
<td>0.04</td>
<td>0.04</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Table 4.4: Parameters’ mean and standard deviation for the year 2003
Figure 4.1: Histogram of the traveled distance by consumers to buy a car across three segments, luxury SUV, lower midsize and compact SUV (for a sample of 5000 individuals in each segment).
Figure 4.2: Geographical distribution of Toyota dealers that sell compact SUVs and fullsize SUVs in California. Size of circles is a function of dealer sales.
Figure 4.3: Actual and estimated shares for the segments luxury SUV and lower midsize cars, for 2004.
Figure 4.4: Example of the variability of own-price elasticity across distance for two Honda dealers, in two different markets.
Figure 4.5: Cost and mark-up of dealers in the lower midsize segment, classified by brands, in 2003
Figure 4.6: Cross-elasticities between two Hyundai dealers and the remaining retailers, in the lower midsize segment.
Figure 4.7: Estimated profit of a new Hyundai dealership, selling in the miniSUV and lower midsize segments, for a sample of zipcodes in the Los Angeles area.
REFERENCES


119


120


121


