Essays in Industrial Organization

by

Jonathan Lewis Graves

B.Sc. (hons), The University of Victoria, 2008
M.A., The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Economics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2017

Abstract

Chapter 2 examines the role of sales (temporary price reductions) in the pricing of perishable products. When products can be stored, periodic sales are explained using inventories: the ability to store lets consumers wait for better prices. When consumers differ in their ability to wait, firms keep prices regularly high, using sales to target the low prices to the most patient consumers. This explanation is not reasonable for perishable goods, since they cannot be stored. Using a retail dataset, I show that nonetheless a cyclic pattern of sales is a major feature of how perishable products are priced. I explain this pattern using a dynamic model of loss leadership. I then test my model using grocery store data.

Chapter 3 studies large contributions to crowdfunding projects and their impact on project success. I find large contributions display a preference for being effective in helping projects succeed: they are often pivotal in the success of a project. These findings match predictions from a consumer choice explanation of how large contributions are made. I then examine the role large contributions play in project success. Using an instrumental variables approach, I show the ability of a project to attract large contributors is important: a project is 40-60% more likely to succeed if they can attract a large contributor. This inverts the logic of crowdfunding: the crowd may be important, but the success of many projects is driven by large contributors.

Chapter 4 develops a method for determining whether a given observation is a sale or not in the context of a sequence of prices for a retail product. This classification, based on a hidden Markov model framework has the advantage of using all the information available for classifying sales. I develop identification requirements for this method, and illustrate its utility in directly testing questions of correlation for sales and other variables: allowing models to be evaluated without reduced-form analysis. I perform simulations, demonstrating the method's accuracy method in classifying sales and understanding correlations. This chapter adds to the toolbox industrial economists have for studying sales, with advantages over existing methods of sales classification.

Lay Summary

This essay studies several topics in industrial organization. The second and fourth chapters consider why perishable products are placed on sale (sold at a discount): most models of sales in economics cannot explain these kinds of discounts. In Chapter 2, I show sales are explainable when firms use a strategy whereby one product is put on sale to encourage consumers to buy related products. In Chapter 4, I develop a method to determine which prices are sales which improves on some of the statistical issues faced by researchers in this area. In Chapter 3, I study the role of large contributions in crowdfunding, a form of fund-raising used by entrepreneurs. I show that large contributions are both very important for the success of projects and sophisticated in when they contribute.

Preface

Chapter 2 is calculated (or Derived) based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business, data copyright resides with Nielsen. I was responsible for the design of the research program, the analysis of the data, and the performance of the research. This work is the unpublished, original and independent work of the author.

Chapter 3 is based on data collected by Kicktraq Inc, and shared under a data-distribution agreement. My supervisor, Ralph A. Winter was a signatory to the data-distribution agreement. I was responsible for the design of the research program, the analysis of the data, and the performance of the research. This work is also the unpublished, original and independent work of the author.

Chapter 4 is the unpublished, original and independent work of the author.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I would like to thank my advisor Ralph A. Winter for his superlative supervision and encouragement. This thesis is much improved for his suggestions, ideas, and insight. I would also like to thank Thomas Lemieux, Vadim Marmer, and Michael Peters for their assistance and encouragement in the development of this thesis. Additionally, I would like to thank the faculty and staff at the Vancouver School of Economics, particularly Nancy Gallini, Matilde Bombardini, and the members of the empirical lunch for their feedback. I would also be remiss not to mention the help of the current and past students, who were my colleagues and friends during this thesis, particularly Jacob Schwartz, Anujit Chakrabotry, and Hugo Jales.

Professionally, I would like to thank Kicktraq Inc for their assistance with the data for Chapter 3, and the Marketing Data Center at The University of Chicago Booth School of Business Kilts Center for Marketing for their provision of the data for Chapter 2. I would also like to thank the participants at the 2016 CEA Conference who saw early versions of some of these chapters. Their feedback was invaluable. I would also like to recognize the assistance of the SSHRC committee for providing fellowship support for this thesis, and the University of British Columbia for financial and logistical support, particularly with the University-Industry Liaison Office.

Finally, I would like to thank my parents, family, and friends for their support and encouragement during this process. This would not have been possible without all of you. A special note of gratitude must be conveyed to my biggest supporter, my wife, Giuliana, to whom this thesis is dedicated.

Dedication

To my beloved wife, Giuliana

Chapter 1

Introduction

Industrial organization studies the way businesses interact with other individuals in the economy. This dissertation focuses on the interaction between consumers and firms, studying the strategies firms use to maximize profit and compete for the attention and support of consumers in different environments. Using a mixture of theoretical modelling and empirical analysis, this dissertation studies this important economic intersection, shedding new light on the actions of firms providing insight into the implications these have for consumers, policy-makers, and the economy in general.

Chapter 2 studies a unique but critically important market: grocery stores. One of the hallmarks of grocery store pricing in the last century has been the "sale" - a temporary reduction in the price of certain products. While different explanations have been put forward for why firms would want to hold sales, the answers provided are far from comprehensive. This chapter looks to fill a gap in the literature, studying the role of sales in the pricing of perishable products using a combination of theoretical modelling and empirical analysis. Models provide the structure to make predictions about the real world, which can then be analysed with data, providing support for the theory.

The original contributions of Chapter 2 are three-fold. First, I establish the fact that periodic sales are a major part of grocery store pricing for highly perishable products. This is done using the largest existing dataset of grocery store pricing data, the Nielsen-Kilts SCANTRACK database, which provides weekly price and quantity information for a wide range of stores in the United States. Second, I demonstrate that the observed patterns of sales are difficult to reconcile with existing models of sales, which tend to rely on inventory-based explanations. To explain the pattern, I develop a new model based on an intuitive idea: that of loss-leadership. In this context, loss-leadership is when a store discounts a particular product in order to attract consumers to their store, who then subsequently purchase more of other products.

In my model, consumers purchase baskets of goods which contain different mixtures of perishable and storable products. Shopping costs induce consumers to buy their entire

basket at a single store. Because these baskets differ, firms can use them to discriminate between different groups of consumers. Specifically, firms want to attract perishable-buying consumers when they are also purchasing relatively many storable goods. In order to do this, they offer these consumers a better total price for their whole basket. It is optimal to do this by lowering just the price of the perishable good, since this targets the price reduction to the group the firm wishes to attract, keeping the total basket price high for other consumers. Firms time these price reductions to allow the target group of consumers to run down their inventory, which entails trading off present and future profit in an intertemporal optimization problem. The link between storable good inventories and perishable products creates, in equilibrium, periodic sales on the perishable good.

Finally, I test my model by linking the retail data to consumer choice data, a process which requires the development of a data-driven method of classifying prices into sales. The results validate the central prediction of the loss leadership model: when consumers buy perishables on sale they also buy more of other products, particularly storable ones, relative to their purchasing in non-sale periods. These findings highlight the role multi-product competition has on pricing dynamics, and rationalize an empirical finding which is difficult to explain with most models of sales, and demonstrate the important role the connection between firms and consumers plays in this market.

Next, in Chapter 3, I study the behaviour of firms in an emerging market: crowdfunding. Since its emergence in early 2010's, crowdfunding has become a major source of financing for individuals, and small or medium sized businesses. Characterized by large numbers of consumers who pledge money for projects in return for a reward, consumer crowdfunding is an area in which economists are only beginning to comprehensively understand the key driving forces. This chapter examines one aspect of crowdfunding which has been previously overlooked in the literature: the presence and role of large contributors to projects. By their very nature, most contributors to crowdfunding are small in nature: on the order of less than $100 (US). However, as I show, many projects include individuals who pledge many hundreds of times the typical amount, for no obvious benefit. This paper studies the ways these large contributors behave and estimates their impact on the success of projects in reaching their fundraising goals. I do this using a primarily empirical framework, but rely on consumer modelling to provide structure for the empirical work.

This chapter makes three substantial contributions. First, I systematically document the existence of large contributors in this market. With these fact, I then examine the

driving forces behind large contributions, finding that they display an apparent preference for being effective in helping projects succeed; indeed, a substantial proportion of large contributions occur simultaneously with projects reaching their goal, often being pivotal in the success of a project. Second, I develop a consumer choice explanation of how large contributions are made, which creates predictions from a theoretical model to explain the observed facts. Finally, I further examine the role large contributions play in project success using an instrumental variables approach. I find that the ability of a project to attract large contributors is important: a project is approximately 40-60% more likely to succeed if they can attract a large contributor. Large contributions also appear to be disproportionately effective relative to their size, indicating they not only provide support in the amount needed, but also when it is needed.

Finally, Chapter 4 returns to the problem of sales, and address an elementary question in this area: how do we know when a product is on sale? Many studies, such as Berck et al. (2008), Pesendorfer (2002), and my own work in Chapter 2 need to know whether or not a given product is on sale at a particular time, before they can move on to address questions of more specific interest. Generally, this is based on very little data: a sequence of prices, and a handful of other potentially useful or interesting variables. Papers in this area adopt a heuristic rule to tell whether a product is on sale, based on the observed pattern of prices. This problem of *price classification* is fundamental to research in this area, and every author has their own specific method, based on their understanding of the product and environment their data is generated in. For example, Chapter 2 uses a clustering and fitted mixture model to tell whether or not a given product is a sale. Nielsen-Kilts suggests calling a sale any price more than 5% below the average for a product. There are many such methods including a fixed discounts from the previous price, rolling averages, etc. In this chapter, I develop a new method to classify sales, which explicitly takes advantage of the structure of how sales arise in the data. This allows me to not only use more information to classify sales, and provide statistical assessments of how certain a classification is, but also helps link the classification step and the reduced-form analysis many papers seek to carry out. For instance, my method is able to directly test whether certain variables are correlated with sales, a central element of interest for many models of sales. This is advantageous because it circumvents any question of whether anomalies in the classification method could create spurious correlations detected by reduced-form education.

My method of sales classification is based on the hidden Markov model framework. After reviewing the basics of this model, I develop explicit identification results for left-to-

right Markov models which can be applied to sales data. I also demonstrate how to include covariates and lagged state variables into the model, allowing for a variety of different patterns for sales to be classified and studied. I then perform Monte Carlo simulation of sales, and test the simulated data against my classification method. The performance is remarkably good, demonstrating that with good initialization of a sales model, sales can be classified with a high degree of accuracy using this method. I also show that this is robust to even "difficult" or complex sales environments, and that the direct testing of covariances is highly feasible with this method. I also investigate the small-sample performance of this method, to provide guidance for situations where data may be scarce. This chapter develops a new and powerful method of researchers in industrial organization to study, grounded in structural considerations of how sales come about and serves as a natural alternative to heuristics, or situations where researchers may be unsure of which classification method is appropriate for their sales data.

Chapter 2

Sales and Perishable Products

2.1 Overview

Sales are everywhere in the retail landscape: a visit to nearly any store will find products being sold at a discount, regardless of whether that store sells power tools or heads of lettuce. Products with very different properties for consumers and retailers all display similar pricing strategies, characterized by periodic sales (temporary price reductions). Why is this the case? Economists understand the phenomenon of sales in the case of storable products. In most models of periodic sales, temporary price reductions occur because firms want to charge different prices to different types of consumers. When consumers can store, they can wait for lower prices, which allows firms to discriminate between them based on their ability to wait; the more patient the consumer, the more elastic the demand. Periodic price reductions keep the price high for most consumers, while offering a low price to only the most elastic (patient) group (see Berck et al. (2008); Pesendorfer (2002) for examples). In this chapter, I analyse sales on *perishable products,* which are particularly puzzling because they are not explained by conventional models for sales: since perishable goods cannot be stored, inventory-based explanations cannot be applied. These kinds of products are also important, since many staple goods such as meat and vegetables fall into this category. In this chapter, I develop and evaluate a new model to try to explain this phenomenon.

The first step is to determine whether sales are an important part of the pricing strategy used by firms for perishable goods. Previous papers such as Hosken and Reiffen (2004) have found some evidence of this, while others such as Berck et al. (2008) have found suggestive results for semi-perishable goods such as orange juice. This literature tends to focus on models in which sales occur periodically, as most studies strongly reject models such as Varian (1980) in which sales are uncorrelated over time. I focus on an extreme case of perishability, looking at *highly* perishable goods: specifically, packaged meat. Packaged meat is an excellent case study because (1) it is uniform across stores and time, (2) it is extremely perishable, lasting less than a week at home, and (3) it is commonly purchased by

consumers[1]. My dataset is the packaged meat segment from the Nielsen-Kilts SCANTRACK database from 2010-2013 including 17,708,731 observations across 9,605 stores and 461 products.

The results validate the findings suggested by other studies: using a linear probability approach, I find that the strongest effect is cyclic in nature. As time elapses after a sale, the marginal likelihood of a new sale rises from -3.4% to +2.5%, as we move from one to four weeks after the last sale. This kind of periodicity is usually held up as evidence to support an inventory model for sales. However, as described, this is implausible for my dataset. One alternative possibility I examine concerns products expiring: I conclude that while stores are responsive to expiring inventory, this does not play a major role in the occurrence of sales. This is sensible, since we should not expect it to be profitable for stores to discard large amounts of product routinely, especially in the packaged meat segment.

In order to carry out this analysis, similar to other papers, I need to adopt a method to determine whether or not a given price is a sale. My technique is based on a simple observation: all heuristics which determine sales try to separate the data into high (regular) and low (sale) prices over some time frame; for example, based on difference from the modal price in a specific window[2]. I do this explicitly, by estimating the price sequence as a mixture of regular (high) and sale (low) prices. Rather than choosing a single time frame, I break the series of prices into sections using a clustering algorithm to intelligently separate the sequence based on within-group variation. My mixture model is then estimated on each of these sections in turn. This is necessary because many products over a long time frame makes any single existing method unlikely to perform well; for instance, the suggested method in the SCANTRACK database is inadequate. This method can be given explicit economic context in the idea of two unobserved states; a slowly varying pricing "regime" and a quickly varying "sale." Under regularity conditions that I develop, the long run "regime" can be recovered and observations classified into different pricing regimes. These conditions amount to a limitation on the price variability of products; prices cannot change "too much" or "too little" over time. This method also allows more detailed examination of whether points or sales are not, and makes explicit the conditions for the effective classification of the heuristic. In Chapter 4, I demonstrate how my method can be

[1]The consumers in my dataset are American and, based on the NHANES survey, consume an average of about 5.5 oz (130g) of meat per day, a rate that is one of the highest in the world (Daniel et al. (2011)).

[2]Examples include techniques like a sale is a price "5% below the average price" (as in SCANTRACK's method) or "10% below the R3 month average price" .

extended into a structural model, using a technique known as hidden Markov modelling.

Having verified that sales are an important part of the pricing strategy for perishable goods, I develop a model of dynamic loss leadership to explain these results. In my model, there is a single firm selling two goods, a perishable and a storable, to a large number of consumers. There are two kinds of consumers: singles and families. These consumers differ in (1) their tastes for goods and (2) their ability to store. Families consume perishables (which are not storable) and also have the ability to inventory the storable good. Singles do not store, and also do not consume the perishable good. The ability to store on the part of the families means that, over time, their stock of the storable good changes. When families are running out of stock, they would like to purchase both the storable and perishable goods; when they are in stock, they would like to purchase just the perishable good. Both types of consumers shop only once per period, and have the option of shopping at the firm or at a local source who sells only the perishable good at cost.

I assume that consumers are distributed in space, with some families located closer to the firm; consequentially, the distant families face a higher cost of shopping at the firm. The firm then faces a problem; they cannot sell to all three groups (the singles, the local families, and the distant families) at their reservation prices for their baskets, since shopping costs drive a wedge between the different groups. In a situation where the distant families have high inventories (and low demand for the storable), the firm would prefer to sell to the local families and the singles at their willingness to pay. In order to sell to the distant families the firm must provide a discount (a sale) on the bundle; which means lowering the price of one of the two goods. If the firm wants to have a sale, it is optimal to lower the price of the perishable good under the condition that singles buy more in total than the distant families. However, the monopolist wants to have a sale only if the new market it attracts (the distant families) are more valuable than the profit it gives up by discounting the perishable good for the existing market (the local families). If distant family inventories are high, they would prefer not to have a sale; if they are low, a sale becomes more attractive. The monopolist also does not want to keep the price of the perishable good permanently low since they lose money by competing in this fashion.

The firm makes this decision inter-temporally, understanding that if it delays a sale it benefits because inventories will drop over time, increasing the profit from the distant families. However, they also understand that when they hold a sale, the out of stock families refill their inventories and become in-stock again, resulting in an inter-temporal trade-off. Nonetheless, a sale will eventually occur, because as time elapses eventually everyone runs

out of stock. The exact timing of the sale depends on the parameters of the model, but it will reoccur periodically and will be at least one period after the last sale, resulting in a temporary price change (not a permanent one). Families have rational expectations about prices, and sort over time into periods where their bundle is most affordable. The central prediction of this model is typical of loss leadership: periods with sales should be associated with higher-than-normal purchasing of other goods, particularly storables, relative to non-sale periods. I also demonstrate a number of extensions, illustrating that this model is fairly robust to different assumptions about consumer behaviour and firm expectations.

Finally, I try to evaluate my model by going beyond the retail data by using the large-scale nature of my classification methodology to forge a link to the Nielsen-Kilts HMS consumer dataset, matching products at the product-store level. This highlights the usefulness of my heuristic and a large retail sample; in this data, since I need to look at just the perishable-buying consumers, I end up being able to focus on only 26,862 trips. Looking closely at the consumer data, I find that direct inventory explanations (the freezing of meat products) are largely ruled out but the central prediction of loss leadership holds up. The behaviour of consumers demonstrates that they spend more in total (across all goods) when they buy products on sale, and in particular on long-lasting (and highly profitable, from the store's perspective) sundry products such as detergent, soap, and toilet paper. I also find that there is heterogeneity among different types of consumers, with loss-leadership being targeted to certain subsets of the consumers; indicating a substantial degree of sophistication on the part of stores in how they price products.

This chapter makes three main contributions. First, it shows explicitly that regular, periodic sales are an important part of the story surrounding the pricing of perishable products. This is important because it means that when we talk about sales, or model them, our explanations need to either take into account the special nature of these kinds of products or acknowledge that a "one size fits all" approach to explaining sales is not going to work. Second, I illustrate how model sales for these products by developing a model which explains the periodic nature of perishable sales. As the discussion will show, I also explain how this model fits into the set of possible models and frameworks. I further show that this model is plausible given the data by showing that the central causal connection necessary is supported by consumer choice data. This demonstrates the third contribution of this paper: showing how the connection of retail and consumer choice data can help evaluate different types of theoretical models. This necessitates the development of more flexible tools to determine when sales occur, and focuses attention on the economic content

of these kinds of heuristic decisions empirical researchers make.

The remainder of this chapter is organized as follows: in section 2.2, I review some relevant background, while in section 2.3 I discuss the data and the details of the heuristic method I develop to classify sales. Section 2.4 contains my reduced form results for the importance of sales in the pricing of perishable products. Section 2.5 develops my theoretical model of sales, while section 2.6 evaluates my model using the linked retail-consumer dataset. Finally section 2.7 summarizes and concludes, while Appendix A contains proofs, extensions of the model, and extra material. A further chapter of this thesis, Chapter 4 develops the structural version of the heuristic method presented in section 2.3.

2.2 Background

The role sales play in the economy has been the subject of several different lines of analysis, both theoretical and empirical. Sales are important because they are fundamental to prices in the real world, with implications for issues as varied as competition policy and macroeconomic price stickiness. They also are puzzling, because as Chevalier et al. (2000) describe the evidence is firmly in favour of the causal relationship that high demand periods lead to low prices, instead of vice versa; sales occur in periods when demand is highest, in a causal sense. Furthermore, this occurs even in relatively competitive markets, such as grocery or retail environments. Sales also appear to occur with regularity, ruling out supply-side problems on the part of the firm[3]. Understanding the motivation behind sales is further complicated by the fact that most markets are complex in several dimensions at once, making it difficult to pin down any one explanation. For example, pricing a single product in a single grocery store may involve (1) a brand-name product versus substitutable generic products, (2) an upstream/downstream relationship between retailer and wholesaler, (3) store and chain pricing strategies, (4) inter-temporal and geographic competition, (5) consumer inventories, search behaviour, and expectations, (6) store-level managerial and performance incentives, and (7) many, many others. Given this range of factors to consider, many papers therefore find mixed (e.g. Berck et al. (2008); Hosken and Reiffen (2004)) support for different motivations for sales.

On the theoretical side, most papers focus on a single mechanism which can create sales, generally for simplicity of exposition and to provide insight into the economic mech-

[3]For example, a firm purchases too much product and needs to reduce inventory or sell product before it expires.

anisms behind sales. Most models rely on asymmetry between consumers, which results in some way for firms to discriminate between the consumers and maximize profits. This is often price discrimination, which can be, as Varian (1989) summarizes (based on Stigler (1987)) "[be] present when two or more similar goods are sold at prices that are in different ratios to marginal costs." In many cases it is difficult to call the competition a form of price discrimination specifically. For example, models of loss leadership like Chevalier et al. (2000), or advertising-based models like Lal (1990); Lal and Matutes (1994) the asymmetry between consumers is more subtle, induced, or non-existent. Generally, these kinds of models will rely on consumer limitations, such as imperfect information, choice frictions, or travel costs, which firms exploit to maximize profit; with the result looking like sales.

One class of models can be described as search-based. Salop and Stiglitz (1977) uses consumers who are unsure of where the low prices are in the environment, and spend time searching. A heterogeneous search cost leads to differentiation between monopolisitically competitive firms which looks cross-sectionally like sales. Salop and Stiglitz (1982) extends this, moving the heterogeneity to an inventory level held by consumers instead, producing similar results while endogenizing the reason for consumer variation. One drawback of these models is that they imply cross-sectional variation in prices across firms, and not necessarily variation within a given firm; this is largely a consequence of their static nature, and different prices over time must be rationalized as "rotations" of the low prices through the economy. The chief problem with this is that such a "rotation" requires the equilibrium to change from period to period, without any model-based motivation for it to do so; such a pattern is acceptable, but there is no incentive creating it, nor ruling out any other pattern of equilibria over time.

To deal with this Varian (1980) improved upon this model by changing consumer behaviour. Consumers make a decision about whether or not to perform costly search then approach the store with the lowest price. This produces groups of searching and non-searching consumers which firms compete over using price. The equilibrium outcome of this model is a mixed strategy on the part of the firms, rationally supported by the expectations of the consumers about the pricing distribution. This naturally creates both cross-sectional and inter-temporal variation in the pricing, which looks like sales. The major drawback of this model is that, for mixing to be effective, it must be uncorrelated over time; truly random, as pointed out in many empirical studies on the topic (e.g.Hosken and Reiffen (2004); Berck et al. (2008); Pesendorfer (2002)). This is problematic because the

cyclic nature of sales is considered to be a major feature of most environments, and this randomization does not generally stand up to economic scrutiny.

One of the first papers to directly address this cyclicality was Conlisk et al. (1984). In this model, sales are caused by a monopolist responding to varying willingness to pay of long-lived consumers in the market. Over time, consumers with low willingnesses to pay accumulate, until they reach a critical mass and the firm lowers prices to capture them all at once. Sales are created in a cyclic way by this dynamic pattern of arrivals. The question of why willingness to pay would change over time is not explicitly endogenized, but other models have tried to adapt this: this forms a second class of models with the basic idea being some form of inventory management.

For example, in Blattberg et al. (1981), sales are the result of long-lived consumers and firms facing differential storage costs. From the firm's point of view, it is immaterial when a consumer buys a product - they will buy it sooner or later, so really their decision is when to move the product between the store and the household. Variation in stock-holding costs between the two economic agents leads to periodic price reductions: a sort of "warehouse clear-out" model. Pesendorfer (2002) explicitly builds a consumer inventory management model to explain the dynamics of pricing he observes in his data, as do Hendel and Nevo (2013). On a more macroeconomic note, Hendel and Nevo (2006) build a model focusing on consumers, and use it to estimate the elasticities of products which are likely to be inventoried, with implications. Because inventories form a kind of buffer for firms, and a type of savings for households, this has economy-wide implications: for example, Wong (2016) finds that consumer inventory stockpiles are a significant form of savings for low to middle-income households.

A third branch of analysis tries to reconcile sales in light of the multi-product nature of most retailers, and the (often very complicated) multi-level firm-brand-chain structure many markets display. Many of these models show a sharp divide between marketing, business, and economic explanations. For example, in Lal (1990), national brands use retailers as proxies in a competition with local brands over a pool of price-sensitive consumers. It is profit maximizing for the national brands to vary their price randomly over time to capture part of the market for the price-sensitive consumers. This relies on the ability of national brands and chains to create and maintain brand equity to create "loyalist" consumers, something most economic models don't try to address. Lal and Villas-Boas (1998) builds on this model by extending it to include multi-product retailers along with the competing national brands. Competition remains over groups of price sensitive consumers,

with equilibrium outcomes interrelated in a similar way, but now across substitutable or complementary goods. Shelegia (2012) extends the idea of loyalists to multi-product retailing, demonstrating optimal differential pricing strategies based on the cross-elasticities of the goods in question.

Other models use spatial differentiation. Shilony (1977) describes how monopolisitically competitive firms compete over prices using their locations for monopoly power. Sales pricing is the result of the a mixed strategy Nash equilibrium where competition is over the marginal consumer. DeGraba (2006) uses a similar spatial model to show sales are the results of price discriminating between different kinds of consumers who buy different bundles of products, explicitly using the multi-product nature of the firms to create what he describes as "loss-leadership." Behavioural models also abound, examining sales pricing as part of a mix of potential marketing decisions (e.g. Yoo et al. (2000)), with a complicated relationship between pricing, quality, brand premia and sales (see, for example, considerations in Aaker (1996)).

The proliferation of so many different theoretical models, the macroeconomic implications of price variability, and the obvious importance of the problem of how to administer sales from a management point of view has lead to a rich empirical literature studying sales. The papers closest in nature to this paper essentially follow a "differentiation" program of research; they note that in general, the different theoretical models have make different predictions for how products, firms, and situations should appear in the patterns of sales which we observe. They try to evaluate these different explanations in light of the data, and possibly provide their own model for which results seem the most plausible. This has been enabled by the dramatic increase in the availability of scanner data from supermarkets, and the increasing affordability of both computational and data storage resources.

For example, Hosken and Reiffen (2004) looking very broadly at the monthly data used by the United States Bureau of Labour Statistics to create the Consumer Price Index used to measure inflation (among other things). They find evidence against many of the mixed strategy models, such as Varian (1980), but have difficulty pinning down a single alternative model as the definitive winner. This is largely due to the fact that most of the inventory models they consider fail to properly explain the patterns related to perishable goods in their data, coupled with the low granularity of the data available to them. On the other hand, Pesendorfer (2002) uses a study of ketchup bottles to develop and support an inventory model, again finding mixed results for some models of sales, but also looking also

at the role of chains explicitly. In one of the broadest studies to date, Berck et al. (2008) uses orange juice (fresh and frozen) to test a wide variety of sales models, again finding mixed results for different kinds of sales, but relying explicitly on the time series nature of the data under examination. Hendel and Nevo (2013) develops an inventory model of sales based on soda pop data, similar to Salop and Stiglitz (1982); Blattberg et al. (1981). This study is capable of quantifying the effects of sales, finding they are substantially profit improving for firms.

In general, the cyclic pattern of sales has proven difficult to explain; most mixed strategy most explicitly must rule it out, in order for the equilibrium to be supported, or rely on very narrow and difficult-to-imagine equilibrium arrangements to support such a pattern. Accordingly, inventory models, which create an endogenous variation in the willingness to pay for consumers are an appealing alternative. However, these are difficult to consider when it comes to inventory-less products like perishables; you cannot have a variation in inventory driving sales if no inventory can be held. Hosken and Reiffen (2004) points this out explicitly, noting that the same inventory model appears to apply to all of their products, even when it is difficult to explain why. This paper examines the same point in a different context, and presents a model which connects inventories and perishable product through the multi-product nature of consumer buying.

2.3 Data and Sales Classification

2.3.1 Data and Sample Selection

This project uses four years of the Kilts-Nielsen SCANTRACK dataset, 2010-2013 inclusive. These data are a collection of weekly store-level sales data for every UPC[4] processed by a large number of stores located in the United States (see Bronnenberg et al. (2009, 2012); Danaher et al. (2008) for examples of this dataset). For each store in each week, we observe the price of the product, the volume, along with several other covariates which will be discussed later. The products also have a number of hierarchical details, such as their product category, description, and for selected products additional information specific to the UPC description; the detail given is structured so that it is associated with the UPC

[4]UPC stands for Universal Product Code, a barcode system widely used in English-speaking countries for tracking products of a uniform type across stores. It facilitates inventory tracking, checkout, and store management at both the retail and wholesale level and is connected to the company registration number used for firms. The term UPC code is equivalent, albeit somewhat redundant.

registration. In the United States, the registration and use of UPCs is designed primarily to facilitate the differentiation of different products by packaging type (size, number, etc.) rather than by qualitative features relevant to the consumer. Accordingly, the SCANTRACK dataset has a similar structure for UPCs. Similarly, the data contain information on the stores, including chain membership and locations; however, precise identification of the stores (or chains) is not available or recoverable under the terms of the data-sharing agreement.

In this paper, I focus on perishable products; as explained in Hendel and Nevo (2013); Hosken and Reiffen (2004) the pricing patterns of perishable products are difficult to explain with inventory-based models. Specifically, I look at a specific subset of perishable goods: all UPC-coded packaged fresh meat products[5]. I focus on packaged fresh meat products (as opposed to other kinds of grocery store perishables) for three reasons: first, packaged meat is highly perishable, typically lasting for less than a week after it is packaged. This makes it an excellent example of a perishable good, since it is very difficult to inventory or store. Equally useful is the fact that the shelf-life is less than the data collection period in the SCANTRACK dataset, which means stores seek to sell-through their weekly inventory within the period observed. This means we have a close connection between the life of the product and the periods we observe it for. Second, unlike many perishable products (lettuce, apples, etc.) packaged meat is highly uniform in appearance and quality, making it practical to compare products across stores and across time. Finally, it is important: for most Americans perishable meat products like the ones studied form an important part of their weekly or daily diet. This makes them not only inherently economically interesting, but also ensures that consumers are likely to include them in their consumption bundles. This features will prove very important when we turn our attention to SCANTRACK's sister database at the consumer level, which I discuss in section 2.6.

One challenge with these data is that, as in Hendel and Nevo (2013), the collection method of SCANTRACK dataset can cause certain problems for inferring pricing: specifically, because SCANTRACK is collected on a weekly basis, they decide to report prices as the volume-weighted average of the prices sold during the week. When the pricing period for a product agrees with SCANTRACK's window, this is not a problem. However, if (as some stores do) the period in which prices change occurs in the middle of the week,

[5]For example, this means including only individually packaged 16 oz units of lean ground beef, and not random weight ground beef sold through the butcher's counter or packaged in store, or 16 oz packages of ground beef sold 3-for-1 as a "multi-pack."

the observed pricing becomes very noisy and difficult to impute. However, this is easy to detect in the data: stores which change their pricing regularly during the week have a very large number of unique prices relative to observations for each product. We exclude such stores based on a fraction criterion: stores with more than 1/3rd of their observations demonstrating unique prices are deemed to be those which have changes during the week. Given that products with higher frequencies would involve changing all pricing and signage every other week, this is a reasonable restriction. Additionally, in order to make observations more similar, I additionally include only grocery stores (channel "F") in the data; this principally excludes mass merchandise stores (such as Walmart, Costco, Sam's Club, etc.) which sell products other than food, and focus on products with at least three years of data.

Since individual stores may have distinct pricing schemes (based on regional demand, location, etc.) for each product, my unit of observation is the UPC-Store combination. After the preceding restrictions have been made to the data, I am left with 17,708,731 observations across 9,605 stores and 461 different UPCs, for a combination of 92,465 unique UPC-Store level observations. These are dispersed across 66 different chains of grocery stores located in 173 distinct counties in 49 states. I report the numerical descriptive statistics for my sample in Table 2.1. One interesting feature of the SCANTRACK dataset is that Nielsen has collected information on some products using store-level audits. A "feature" is an occurrence of whether or not a product was featured in advertising for the week. A "display" indicated whether or not a product was put on display in the store in a given week. Unfortunately, this is only carried out on a random sample of the population of stores; in my data, about 19.8% feature this information.

However, this information alone is not sufficient to study sales; critically, the incompleteness of the "feature" and "display" metrics means we need to adopt a method to classify which observations are on sale from the time series of prices. The administrators of the SCANTRACK dataset make the suggestion that a sale is any price 5% below the average price for that product in a given store. However, this is incorrect when we consider most products and especially over a long time frame. Accordingly, I develop an alternative method in the following section, and apply it to the sample of data. Due to the large scale of the dataset, this is impossible to monitor manually; this is partly the purpose behind the method I develop. I apply two low-level diagnostic tools to eliminate poor fits: first, we reject any product for which the estimation of regular and sale prices could not converge within a reasonable number of repetitions. Second, we apply a "warning" label to

any observations for which the procedure failed to converge in 1000 iterations. The first is clearly a serious mistake and should be excluded; the second is less obviously serious, since it merely implied that the clustering fit is not "optimal" for at least one repetition of the clustering method. After estimating, I see about 0.07% fail critically, and exclude them from the reduced form results.

2.3.2 Sales Classification

When is a product on sale? If we want to understand the decisions behind why firms put products on sale, we need to have some sense of when a product is actually on sale or not. This seems simple, but is actually not straightforward at all. Generally, a researcher will only observe a time series of prices (and perhaps quantities) for a particular product in a particular store[6]. Some of these prices may be high, some may be low, but classifying these prices into sale or regular prices is complicated by two features: (1) long-term price fluctuations driven by structural change in the business (like wholesaler price increases or market-wide shocks to demand) and (2) idiosyncratic fluctuations driven by store-level unobservables, such as store price changes. To make these concerns concrete, consider Figure 2.1; this depicts the price pattern associated with 20 oz packages of a single meat product in a single grocery store over several years. As we can see, there is a repeated pattern of high, then low, prices; what we would normally call "sales." There is also a long run change in the price, increasing over time for both the high and low prices. Additionally, we can see that in certain weeks, prices are "near" the usual low price but not exactly; are these sales or not? The problem of determining which observations are sales is not straightforward; it requires a method to classify them.

Accordingly, most studies rely on a heuristic methodology based on particular characteristics of the dataset. For example, Hosken and Reiffen (2004) look at the difference below the modal price, while Hendel and Nevo (2013) uses a fixed deviation from a typical price point. The basic idea is that we can study sales in these environments by choos-

[6]Even the creation of variables indicating whether or not a product is on sale is suspect. For example, even if you collected information on sales from flyers at a chain level, the likelihood that a given store would be in perfect compliance is low. Store-specific sales could occur, or sales could run out of the sale-marked product. Even store-level data is suspect; without retrospective analysis of prices, even apparent sales can be (intentionally) misleading. This is particularly acute in retail environments where "price anchoring" and misrepresentation of regular product prices skirt closely to the legal limitations on such practices (see, for example, Manjoo (2010); Tuttle (2014)).

ing a product we understand well, with desirable features[7], then use that information to determine what is, or is not, a sale. This has the advantage of being motivated by the economist's understanding of the product to inform the definition of a sale, but also has attendant disadvantages. To be clear, this approach is generally a good one, but is intractable in situations where we have large amounts of data on relatively heterogeneous products or no clear priors for how sales should appear in the dataset. It's hard to give a clear answer to the question of "what is a sale" when presented with an arbitrary product, and a "one size fits all approach" may not be appropriate in many situation.

A natural way to understand all of these heuristics is that they separate the prices into two groups: high "regular" prices and low "sale" prices. The difference between these two prices is then detected by some kind of averaging or filtering process which tries to determine which price is likely to be the "regular" price then classifies sales on this basis. For example, SCANTRACK suggests that the mean price is close to the "regular" price, and the filter which screens for sale prices is a 5% margin cut-off. I take this general notion and make it explicit: I assume that the observed series of prices is a mixture of regular prices and sale prices, plus some noise. I then proceed to estimate a mixture model to determine the regular and sale prices from the data; if the price sequence was relatively stable, this would be straightforward. However, as we can see in Figure 2.1, the sale and regular prices change over time. In order to deal with this, I break the time series in sections, and estimate the mixture model separately for each sections. This separating process is depicted in Figure 2.2.

However, rather than simply fix a single rule about how to divide the time series into sections, I adopt a more flexible method based on k-means clustering. To understand this method, it is worthwhile to understand the economic intuition behind the idea of both sales and long run variation in pricing. The perspective I take is that stores make weekly decisions about prices which are governed by a long-term pricing strategy (called a "regime"). Returning to Figure 2.1, the idea is that in any given period a store chooses between two prices: a regular price, or a sale price. For many models of sales (including that of section 2.5) this is the natural pricing decision. This also agrees with the idea of a "mixture" of two price levels, developed above. The value of these prices are part of a *pricing regime* which can change over time, causing both the regular and the sale price to also change. For example, we can see that in the first three months of 2010, the price of

[7]For example, most products used have some form of up-stream brand connection to anchor prices, such as the standard price of a 2L bottle of Coca-Cola.

this product varied by about 70 cents, repeating in a cyclic pattern of prices. Then, in April, the price variation began alternating between higher levels; eventually, by September this variation to about 50 cents. This kind of variation is a common occurrence. There are similar examples of such patterns in our dataset, and in other research on the topic (see Pesendorfer (2002); Hendel and Nevo (2013) for example). These changes in the repeating patterns of high and low prices are changes in what I call a pricing regime; the change from high to low price within the regime is the sale. Small variations in these prices which are non-systematic, such as those in February 2013 are example of the potential idiosyncratic features which may arise.

This means we are defining, informally speaking, a sale as a temporary reduction in the price of a good (to the *sale price*), relative to a *regular price* which occurs for a more substantial period of time. This agrees with the "time test"[8] legal definition of a sale, which one of the main legal ways a sale is regarded. As mentioned, if prices are observed for long enough, this is complicated by the fact that the ordinary price may also change over time. For example, most companies will perform systematic price changes for all their products at least once or twice a year, which are passed through to the retail channel. Similarly, most grocery stores will re-evaluate how they price products at least annually. We have already seen examples of this in Figure 2.1. To make this definition more formal, I now begin to impose some structure on the data we observe.

Consider a representative product being sold at a single grocery store over time. The researcher observes the price Y_t and a set of covariates (such as amount sold, product department, weight, etc.) X_t for each (discrete) time period $t = 1, 2, ..., T$. We imagine at each period, the grocery store has several different options to choose for their price, $Y_t \in P$. These prices correspond to the pricing strategy the grocery store has settled on using at that time period; we leave this unmodeled, but we could imagine the pricing options are the equilibrium strategies of a more complicated game played against other grocery stores. In any case, the set P is the pricing set, and is defined as follows:

[8] In Canadian law (a similar definition holds in US or EU law), the Canadian Competition Bureau summarizes this (under Subsections 74.01(2) and 74.01(3) of the Competition Act):

> [The Act] prohibit[s] the making, or the permitting of the making, of any materially false or misleading representation, to the public, as to the ordinary selling price of a product, in any form whatever. The ordinary selling price is determined by using one of two tests: either a substantial volume of the product was sold at that price or a higher price, within a reasonable period of time (volume test); or the product was offered for sale, in good faith, for a substantial period of time at that price or a higher price (time test).

Definition 2.3.1. A *pricing set,* denoted P is a set of prices which consist of (1) a regular price p and (2) k_p discounts from the regular price $\delta_1, \delta_2, \ldots, \delta_{k_p} > 0$. The associated sale price is $s_k \equiv p - \delta_k$. Similarly, a *sale* (at discount k) is defined as the event that $Y_t = s_k$ and is written S_t^k. The regular price can then be written as $Y_t = p - \sum_{k=1}^{k_p} S_t^k \delta_k$. The set of all pricing sets is $\mathbf{P}$.

These pricing sets are the options individual stores have to choose from; a pricing *regime* is the particular set that a store has adopted, which persists for a continguous time. Since this is not directly observable, this is modelled as a state variable which we index with the integers; we represent this indexing with a function $R : \mathbb{Z}^+ \to \mathbf{P}$ which maps the integers into the set of all pricing sets.[9]

Definition 2.3.2. A *pricing regime*, denoted R_t, is a state variable which indexes the available pricing set for a store at time t and lasts for a contiguous period of time.[10] That is, if $R_t = z$ then $P = R(z)$ is the available set of prices. We denote the associated prices and events by association with R_t: $p(R_t) > 0$, $\delta(R_t) = (\delta_1(R_t), \delta_2(R_t), \ldots, \delta_{k_p}(R_t)) \in \mathbb{R}_{++}^{k_p}$, $s_k(R_t) \equiv p(R_t) - \delta_k(R_t)$. If we define S_t^k to be the event that a sale of type k occurs, and $S_t = (S_t^1, S_t^2, \ldots, S_t^{k_p}) \in I_{k_p}$, then the observed price can be written as $Y_t \equiv p(R_t) - \delta(R_t) \cdot S_t'$; to agree with the notion of sales being relatively infrequent, we require that $P(S_t > 0) < 0.5$.

In general, I will make the simplifying assumption that $k_p = 1$, except when explicitly noted. Given that the number of regimes is not *ex ante* fixed, this is not overly restrictive as long as different sale prices occur in distinct time intervals. Note, in general when we discuss a single pricing regime, I will omit the R_t parameter when there is no confusion. Notice, this implies that there are really two kinds of variation in the pricing: high-frequency variation driven by changes in sales (S_t) and low-frequency variation driven by changes in the pricing regime (R_t). Economically, we can imagine that sales are a short-run strategic decisions being made by the store or chain, while pricing regimes are long/medium-run decisions about the overall pricing of a product or group of products. This distinction has been carefully studied in the macroeconomic literature on price stickiness (e.g. Kashyap (1994); Bils and Klenow (2002)) due to the importance it has for general equilibrium models. This is illustrated in Figure 2.1, but similar figures can be found in most other

[9] Since the number of possible date/firm combinations is discrete, the use of the integers as an index set is without loss of generality.

[10] That is to say if $R_t = R$ and $R_{t+k} = R$ then for all $s \in [t, t+k]$, $R_s = R$

empirical studies on this topic. As we can see, the price tends to jump up and down in a fairly cyclic pattern, two weeks of sales followed by two to four weeks of regular prices. However, the "regular" prices and the sale prices evolve over time, as does the frequency and structure of sales; for instance, some periods show far fewer sales than other periods. In terms of the model, we would call the long-run evolution of the pricing the "regime" R_t while the week-to-week discounting of the price is the "sale" S_t.

If we believe that our data is generated according to a series of pricing regimes, this means that in a given regime R_t, the distribution of prices must follow the noise-free mixture model:

$$f_0(y|R_t) = \begin{cases} P(S_t = 1|R_t) & \text{if } y = p - \delta \\ P(S_t = 0|R_t) & \text{if } y = p \\ 0 & \text{otherwise} \end{cases} \tag{2.3.1}$$

I illustrate a method to simulate this model, with error, in the appendix. Unless otherwise mentioned, the counterfactual data is simulated from the appendix model unless being used to illustrate the econometric features of the model. Notice that for our purposes, we are only interested in the parameters p and δ, since they allow us to tell when a product is on sale or not; the parameter $P(S_t = 1|R_t) = (1 - P(S_t = 0|R_t))$ is not of primary interested. This is useful, because we would like to be able to identify p and δ without imposing a particular structure on sales; they could be independent, they could be serially correlated, etc.

Lemma 2.3.1. In the model defined by equation 2.3.1, p and δ are identified, provided that $1 > P(S_t = 1|R_t) > 0$ and $n > 2$.

Proof. This follows from the fact that the distribution is very stark; if $Y \sim f_0(y|R_t)$ and $1 > P(S_t = 1|R_t) > 0$ then $p - \delta = \min(Y)$ and $p = \max(Y)$. Therefore, p is identified directly, and $\delta = \max(Y) - \min(Y)$. □

The condition that $P(S_t = 1|R_t) \in (0, 1)$ is the assumption that both sales and regular prices occur within the regime. This is not restrictive, since if we are interested in examining sales, it is a sensible requirement to assume that sales actually occur. Notice that this result does not rely on imposing any structure on when sales occur; while this term determines the shape of the distribution, the parameters of interest for our purposes

	Obs	Mean	Std Dev	Min	Max
Price	All	5.13	2.57	0.01	27.36
Units	All	23.82	46.06	1	6956
Feature	19.81%	0.045	0.207	0	1
Display	19.80%	0.008	0.088	0	1

Table 2.1: Descriptive Statistics

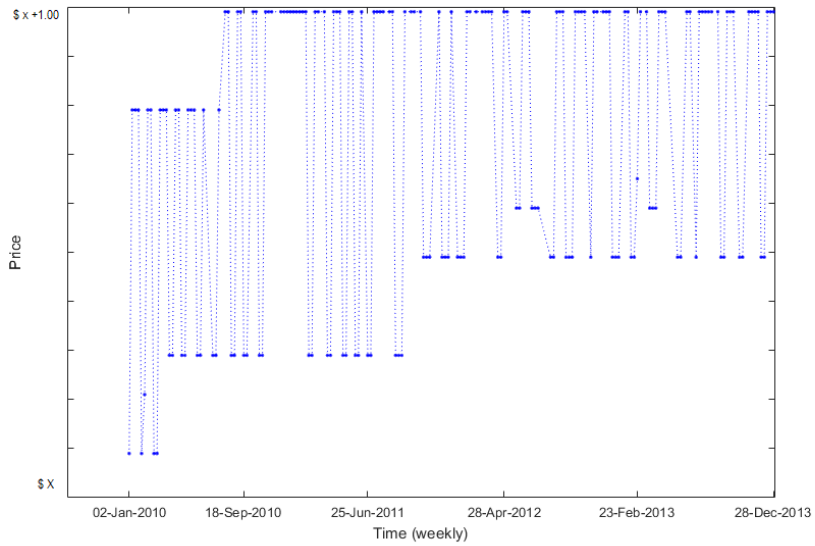


Figure 2.1: A typical time series (20 oz packaged meat), levels omitted for privacy

are governed by the *bounds* of the distribution, which are not affected by the frequency or structure of sales.

There are two problems with actually using this model. First, we know that there are small variations in the prices (errors, idiosyncratic noise, etc.) which cannot be captured by this model. Second, due to the fact that variations from this discrete distribution cannot be captured, estimating the model can be very difficult. This is because the likelihood function, in terms of p and δ is flat almost everywhere (i.e. except on a set of measure zero). This is not amenable to any estimation method which does not explicitly use the sample analogue of the identification procedure above, which is not feasible due to the first problem outlined above.

I order to work around these difficulties, I make a make a simplifying assumption by adding an error term: namely, that the observed values are $Y_i = p - \delta S_t + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma)$. This structure implies that the structure of the simplified model is a two-distribution mixture of Gaussians, with the restriction that the covariance matrix is diagonal and symmetric. I choose a Gaussian noise term for two reasons: first, it has complete support, so even a badly-specified two-point distribution will put non-zero probability on even very different observations. Second, the Gaussian distribution is amenable to a number of numerical optimizations, which make it extremely fast to evaluate, allowing rapid estimation of such a model. The main drawback is that the distribution is non-skewed and can take on both positive and negative values, something we do not really see in the data; this is the motivation for the model in the appendix.

In general, mixture models suffer from generic non-identification, and require some normalization to be identified. However, this model naturally nests both assumptions required for identification: first, the condition that $\delta > 0$ illustrated is equivalent to an assumption on the labels of the mixture. Second, if we maintain the assumption that $1 > P(S_t = 1|R_t) > 0$ (that sales actually occur), we have identification of model as explained in Titterington (1985)[11]. If this second condition fails, we cannot identify p or δ. However, since we (by definition) believe sales are infrequent ($P(S_t = 1|R_t) < \frac{1}{2}$), this implies we can still identify p, the regular prices. Since the objective here is to *classify* sales, being unable to do so in regimes which do not show sales is not a serious problem. It is worth noting that, as in the simpler model above, the potential serial correlation of

[11]The basic requirements for identification of a Gaussian mixture distribution are (1) that the mixtures have non-zero proportion, (2) the means of the component distributions are different, and (3) that the labels of the components are not exchangeable.

different components is not material; although from period to period the probability of being a sale may depend on the previous period, if we consider the model *ex post* and just consider $P(S_t = 1|R_t)$ as an average over all possible histories, this mixture model applies, and identification of p and δ follows, as before.

However, this requires that we know which points are associated with a given pricing regime. Most heuristic methods essentially amount to assuming that all points are grouped into a single regime, then using some method to produce the mixture division made explicitly before. The problem then becomes how to separate the time series into separate regimes in a way which is tractable for large amounts of data. I do this by focusing on the data as a two dimensional object (Y_t, t) and then using a clustering method while overweighting t to preserve the sequencing of the data. Since pricing regimes are contiguous and the timing is evenly spaced, this implies that the centroid (the arithmetic mean of each data point) of each regime is along located the midpoint of the time duration of the regime. The height of this point is governed by the average price; in a clustering method, points are assigned to which of these centres is closer. The essential character of this method is to break the sequence of points into blocks which have minimal within-block dispersion. Under conditions developed in the Appendix under Assumption A.2.1, this process not only separates the data "correctly" but also aligns the blocks with the economic structure developed above.

However, this structural interpretation requires some conditions. First of all, since we are interested in determining whether a product is on sales (and not which regime it is in) it is without loss of generality if we imagine all regimes are of the same duration; if they aren't, we can break the larger ones into two smaller parts. This is required because mechanically larger regimes would have more dispersion even if the price variation was the same. Once we have done this, correct clustering into regime amounts to determining whether or not points will be closest to their regime's centroid, as opposed to another. These conditions can be explicitly stated, as I develop in the Appendix in Assumption A.2.1. They amount to the requirement that prices cannot change "too much" or "too little" from regime to regime. Basically, a new regime should not have points too close to the centroid of the old regime.

The intuition for this is straightforward: if we had a regimes with a regular price p_1 then a new regime with a sale price $s_2 = p_1$, if these regimes border one another and regime 2 starts with a sale, it is impossible to tell whether this point is a sale from regime 2 or a regular price in regime 1. This corresponds to the new regime having prices which

have changed "too much". Similarly, if for regime 2 $p_2 = p_1$ the prices have changed "too little", and we cannot tell when one regime starts and the other ends. This is not usually a problem, however, since the classification of sales from these two regimes would be the same. In general, increasing sequences of pricing regimes (provided they are not extreme) will be well-classified by this method. The most serious problem can arise when the price falls and a regular price of a new regime is close to the median price of the old regime. In this case, these prices will be miss-classified. In the case where there a sales, this is generally fine, since the Gaussian error structure will interpret this failure of identification as noise. However, if there are few sales (or none) in the regime, this may create erroneous sales, which are actually just two regimes without sales. This classification is not completely absurd; nothing, in the basic model, restricts this kind of sales behaviour. Nonetheless, it does not agree with what we would normally think of as a sale.

This is undesirable, but difficult to control for directly within the classification method. Fortunately, because we know the classification failure must occur in a particular way, the data should display "runs" of sales which are not likely if they were truly generated by the mixture process the model implies. This can be detected using standard runs detecting techniques; for instance, the Wald-Wolfowitz non-parametric runs test (Wald and Wolfowitz (1940)). I use a slightly more sophisticated method, in which I used the fitted *ex post* probability of sales to determine the likelihood of a sequence of runs, then reject sequences which are less than 10% likely.

Since, in general, the number of regimes is not known *ex ante* the most direct way to do this is by performing k-means clustering for several candidate values of k, then using a cluster comparison measure to decide on the "best" value of k. In order to try to ensure we have similar-sized clusters, and to capture approximately two price changes a year, we check $k = 10$ to 20. This also agrees with the cut-off for number of prices, which we will discuss in Section 3. I show the results of this technique on four simulated datasets created via the data generating procedure in Figure 2.2. It turns out that the most effective cluster comparison metric is the Davies-Bouldin measure (as explained in Davies and Bouldin (1979)), based on a comparison of several different metrics manually on both simulated and actual data. As we can see, the clustering technique does a generally good job, with the exception of the last graph, which incorrectly chooses too few clusters. This mainly occurs when the regime lengths are of different lengths but similar values; for instance, the same data is depicted in the third graph with a different time arrangement.

2.3.3 Estimation and Classification

Once the data has been divided into different sections, the remaining task is to classify the points within each regime into sale and ordinary prices. I do this is by directly estimating the mixture model given above. Since the number of mixture components is small (two), this can be done directly using maximum likelihood estimation on the Gaussian mixture model, since the associated likelihood function is simple:

$$Pr(y|q, p, \sigma, \delta) = (1 - q)\phi(\frac{y - p}{\sigma}) + q\phi(\frac{y - p + \delta}{\sigma})$$

I demonstrate the performance of this in Figure 2.3, which illustrates the fitted values of the two levels in the model by black lines. As we can see, in the simulated data this fit is very good, largely due to the large sample size and well-defined regimes. However, as we can see, the actual data performance is relatively good as well. There are a few errors, mainly coincident with the difficulties the model had with certain regimes, but even with these included the average fit is relatively good. I tested my methodology on a random sub-sample of the projects, and performance is generally typical of the illustration; agreeing with what we would normally call "sales."

To next classify the points, I take the fitted values and form a p-statistic: for each regime, I calculate the probability of a given observation being drawn from the regular price distribution for that regime. I then classify points based on this measure: point with relatively high p-sales are classified as regular prices, while points with relatively low p-values are classified as sales. We can see this performance in Figure 2.4, where the colour of the point indicates a high probability of being a sale price (with blue being regular and yellow being a sale). As we can see, this performs well; at the standard ($p = 0.9$) level of significance, the data correctly classifies all points in the actual data. The method does tend to over-fit errors as sales, but this can be corrected by performing the opposite test: instead of forming a statistic based on regular prices, estimate it also based on sales prices. Then, classify sales only for those points which pass both tests. However, given that in the actual data the definition of what is an "error" and what is a small sale is not transparent, either method can be preferred given the application being considered. This is especially true when we recognize that some regimes may not have any sales in them. I illustrate points which pass zero, one and both tests for the data in Figure 2.5; this generally agrees with the intuition explained above. Using both tests provides a more conservative benchmark, and is definitely more robust to errors in the data. However, it also tends to remove large sections

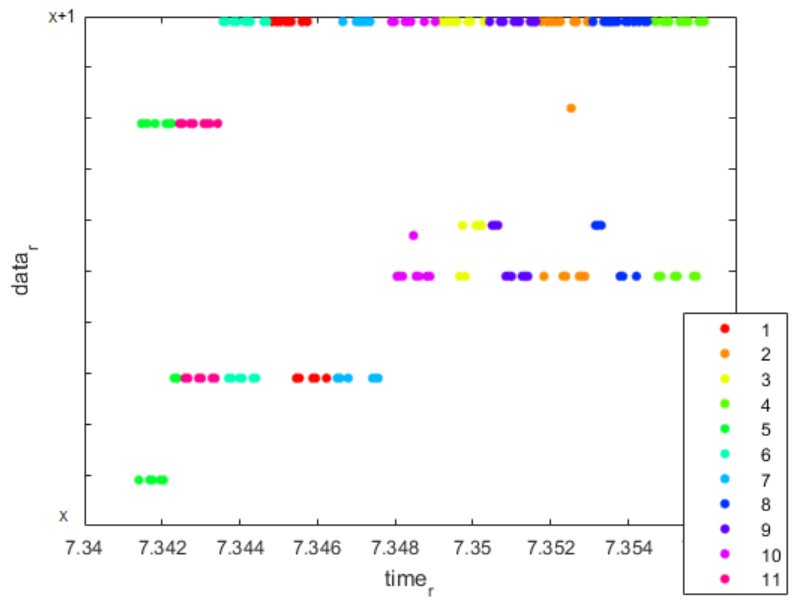


Figure 2.2: Clustering on simulated and actual pricing data

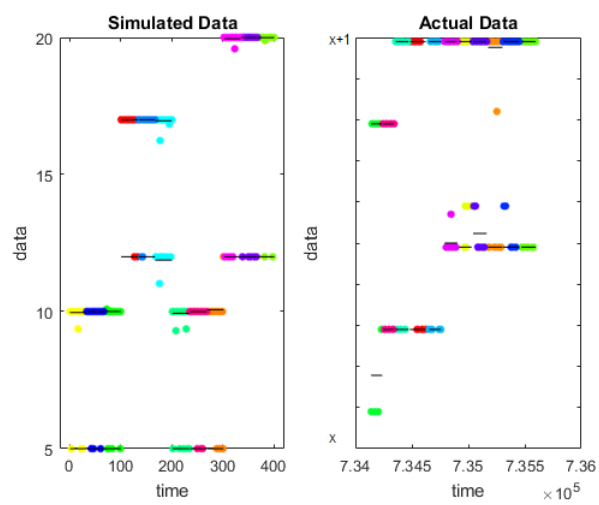


Figure 2.3: Fitted values of model

of the data from classification, considering it ambiguous. The performance on the latter actual sample indicates that for our application here, the one-test standard is probably appropriate, especially given the (previously mentioned) ambiguity. This ability to provide probabilistic assessments of the likelihood of a point being a sale is an improvement upon most heuristic methods, which adopt "cut-off" rule decided *ex ante* instead of using the actual incidence of sales in the data.

Once we have classified the data to an appropriate level, we can then use it to answer questions about the dynamics and drivers of sales in the data being observed. The essence of this method was that the pricing regime provides an economically motivated way of dividing the time series into sections. Once this is done, we can separate the prices into two groups, regular and sale prices. From here, we can then estimate the probability that a given observation is a sale. This method has several advantages, when compared to other kinds of heuristics. First, it allows us to flexibly categorize many different types of products, relying on the clustering method to group or separate the data rather than choosing a time frame or ignoring the variation entirely. Secondly, it explicitly provides probabilities for observations to be sales, which gives us better ways of testing when our method is not working properly. Finally, it makes explicit the economic meaning behind the variation we observe and allows us to develop conditions under which we can be sure it is working properly. This is especially beneficial since, as discussed in section 2.3.4, this can be extended to an explicit structural model.In the next section, we introduce the large-scale data analysis, and illustrate the performance of the classification scheme. We then provide reduced form evidence for the sales detected by the classification scheme in the data, illustrating different patterns and trends.

2.3.4 Comment on Structural Models

This paper follows the literature by taking the following analytical approach: classify data into sales, then analyse it. This is typical in this area (c.f. Berck et al. (2008); Pesendorfer (2002)) and is the preferred methodology for two reasons. First, it is flexible (you can examine many different explanations easily), and second, it is suitable for large amounts of data. However, as we can see, it requires some conditions which can be difficult to test explicitly. Many of these conditions are economically motivated; for example, the idea of using a runs test to infer misclassification of sales demonstrates that there is some structure to the environment we are not using. The way to address this is to incorporate it explicitly

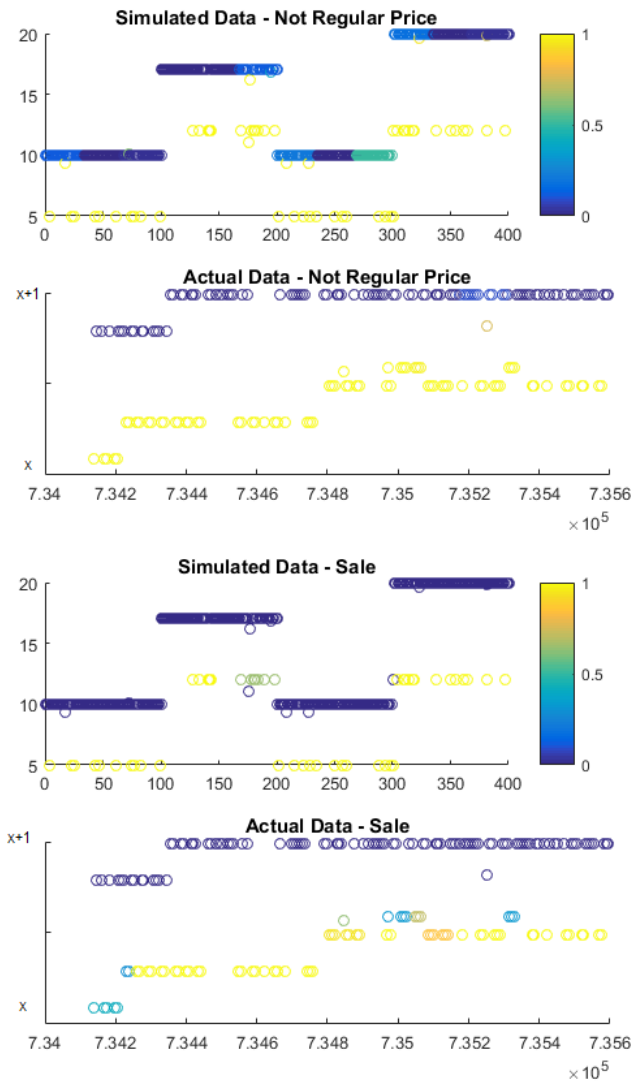


Figure 2.4: Classification of sales from data

in a structural model. Essentially, this is a different analytical approach: jointly classify data and analyse it. This has the advantage of doing away with the classification step, but has several disadvantages. Primarily, we need to have a good idea of what we are looking for to produce a model which can capture the dimensions we are interested in. Secondly, it is also difficult for structural models to be used on large datasets, since they can be very resource intensive; as is the case here. This approach is certainly valuable, and because of this value, I explicitly develop a structural model of sales which can be used to address the same reduced form questions I tackle in this paper. However, because it is not the main empirical focus (and adds nothing to the loss-leadership explanation) I develop this model explicitly in Chapter 4 This model is based on a hidden Markov model, adapted for this environment, and provides explicit identification and estimation techniques for use in sales data. This structural model is suitable for situations where two things are desired: first, a detailed and robust connection of sales regime and sales behaviour. For example, if an economist was primarily interested in examining the dynamics of sales regime changes, or how frequently a grocery store changes its pricing strategy, the structural model would be preferred. Second, it is suitable for smaller-scale data sets, where models can be neatly focused on a few variables of interest and where the number of observations is smaller. For example, this would be useful to economists performing detailed analysis on a few stores, or trying to analyse competition in a duopoly or oligopolistic market.

2.4 Reduced Form Results: What Drives Sales?

2.4.1 Model

Once we have classified the sales observations, we can then ask which features of the data drive sales. My benchmark specification is a linear probability model:

$$Y_{it} = X_i'\beta_2 + W_{it}'\beta_1 + \epsilon_{it} \tag{2.4.1}$$

The dependent variation is an indicator for whether or not the given observation is a sale or not. The variable X_i contains the different level fixed effects, while W_{it} contains the time-varying values. I present the different estimation specifications in Table 2.2. I try to capture the different effects the literature, and theory, suggest for these kinds of products. I include terms related to whether or not a product has been selling below average for

the last several (or rolling) weeks. These terms are included, as I will discuss below, to account for expiring or stale-dated inventory in the stores. Unfortunately, in this dataset I do not observe explicit store inventories of products, so I attempt to measure this using the likely product left on the shelf in a given week instead. I look at this over several time frames, to get an idea of the both the incidence and the timing of supply management problems within a store. I also include the time since the last sale, which is the standard way to capture inventory features (as in Pesendorfer (2002)). When no sale has been observed, I use instead the censored time since last sale as a separated variable. These are not expected to have positive coefficients for fresh products, but since meat products can be frozen, this might play some effect. Motivated by the patterns I see in the data, I also include a variety of terms related to the periodicity of sales. The simplest is to just count the time since the last sale; this is the "time since" measure for periods of 2,3,4, and 6 weeks. I also provide an alternative which is is the whether modulus of the time since term is zero; this capture whether it's "multiples of the period"[12]. The main difference between the "time since" and "period" measure is based on an interpretation of how sales occur: time since is effectively like a hazard rate of a sale increasing, which has a parallel in the models discussed in Section 2.2 as accumulating demand. Periodic terms are more general, and can include many kinds of periodic explanations such as advertising timing or predictability in sale timing. I also use a variety of different controls, especially for time, state, chain, and UPC code. When possible, I use heterogeneity robust standard errors.

2.4.2 Results

Due to the size of the dataset (17 million observations), which causes problems with many estimation procedures purely due to size, I take a random 10% sub-sample of the store-UPC combinations across their entire time period. This results in 1,370,179 observations remaining; the results presented here appear to be relatively stable, having similar coefficients for (different) 1,5, and 7% sub-samples as well. I present the results for the different specifications in Table 2.3 and Table 2.4. Many of the results are ones we should expect from perishable products, based on previous studies.

First, we can see that perishable products are responsive to sales. The number of units sold is a significant predictor of whether or not a product will be on sale or not; for

[12] I also exclude zero; formally, $0 \mod k = 0$ but I omit this possibility since this is lead to a strong effect without much meaning, since days in which there is a sale will be in this situation.

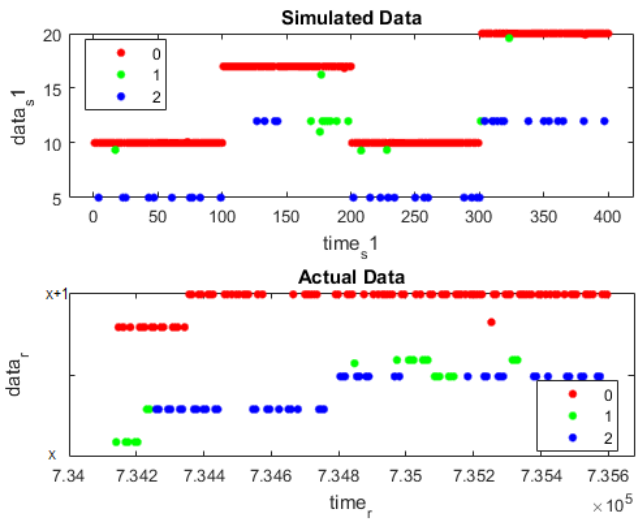


Figure 2.5: Classification tests, two methods, $p = 0.9$

	Specification			
	(1)	(2)	(3)	(4)
Method	OLS	OLS	OLS	OLS
Below average lags	1,2,R3	1,2,R3	R2,R3	R2,R3
Time since last sale	X	X	X	X
Periodicity	2,3,4,6	2,3,4,6	2,3,4,6	2,3,4,6
Period Measure	Time since	Time since	Modulus	Time since
State Dummies	X	X	X	X
UPC F.E.	X	-	X	X
Chain F.E.	X	X	X	X
Month F.E.	X	X	X	X
Year F.E.	X	X	X	X
Errors	robust	robust	robust	robust

Table 2.2: Estimation Specifications

every 100 units sold, the likelihood the associated data being a sale increases by about 1%. Most models of sales, such as Varian (1980) or Pesendorfer (2002) predict such a relationship. Models which do not, such as DeGraba (2006), generally do not make this prediction because of a focus on multi-product competition (and the size of the sale good is a normalization).

One possible reason for sales on perishable products is based on the idea of expiry. Previous studies of perishable products, such as Sweeting (2012), have shown that even relatively unsophisticated (or, at least, individuals without access to analytical resources) are capable of correctly pricing products which have a diminishing utility over time. Therefore, we should expect that grocery stores, with their experienced managers and staff who are incentivized to reduce losses, would also be responsive to expiring product. Because of the patterns of retail delivery, an increase in stale-dated[13] product would be coincident with selling less relative to the average for a given store's sales. The natural response would be to lower prices in a bid to remove the product from the shelves faster, reducing the stale inventory. We see this in the data; if a store undersells by 10 units in a given week, they're about 0.1% more like have a sale in the next week. However, over longer periods of time this pattern reverses; persistently low sales over three weeks shows an decrease in the likelihood of a sale. These types of transient sales would be the most difficult to detect, since they are likely to occur towards the end of a week, and thus would results in only a small average price change for the whole week. With this fact in mind, we can consider this something of a lower bound on this effect. Notice, this also belies a point made in Hosken and Reiffen (2004), which suggests perishable products should not display sales; this is not generally true, and is certainty not a point against the model suggested (that of Varian (1980)). However, a limited effect of expiring inventory is also some we should expect: stores generally receive several shipments of goods during a week, and would be able to reduce their intake of product if they are not selling to expectations during the first part of a week. Managers are also strongly incentivised to avoid throwing away or reducing margin on products, resulting in strong incentives to both accurately predict demand and then manage inventory carefully during a given week. A dramatic effect of expiry, especially which would reoccur indefinitely, indicates substantial potential improvements in supply chain management by stores, and is at odds with profit maximization.

We can also see that the suggestion that the duration since the last sale is unimportant

[13] Product which is no longer suitable for retail sale; perhaps not strictly expired, but either has minor quality problems or would pose a spoilage risk on the shelf.

for perishable products is upheld by the results. The time since the last sale (whether censored or otherwise) is a small, negligible negative predictor of whether or not a product is on sale; for most specifications, the effect is on the order of declining by 0.5 to 1% for 5 weeks without a sale. This is what we would expect from perishable products; there is no inventory motive, as in Pesendorfer (2002) for delaying sales, so they either should happen infrequently and do with underselling (thus being closely related if the underselling problem is not corrected in time) or simply never on sale at all. This is like the case; some UPCs show strong, negative coefficients on the likelihood of a sale. This makes sense for our data here; it also replicates a finding of Berck et al. (2008) who found with orange similar results - despite the perishablility of orange juice being something of a question.[14]

On the other hand, we do find good evidence of periodicity in the data. Specification (4), the modulus, indicates that a "x-weeks" rationale is less strong; but a simple assessment of the time since indicates that sales are 2-3% more likely 2 or 3 weeks after the last sale, then less likely in other periods (-3%). This indicates one reason why the time since the previous sale was insignificant in this model; it was already captured by lower-frequency variables with stronger effects. This is evidence for a inventory effect of some kind; which is not what we would expect for perishable products. There are clear rationales for why a retailer might want periodic (and predictable) sales: in terms of search costs, it is more efficient for the consumer, while it also allows them to accumulate inventory. However, there equally good reasons why a predictable sale period might be undesirable, since in many models it would allow opposing retailers to undercut them (as in Varian (1980)). This finding; that sales not only display temporal dependence, but actual periodicity is difficult to reconcile with the shelf life of the product in question. Is it possible that we are simply picking up the freezability of products? Or is there something more sophisticated at work here? In order to answer this question, and to distinguish between different explanations, I first develop a model of why we might see inventory-like behaviour in perishable products. I then use this model, with another consumer-level dataset, to see which explanations appear to hold up; this is developed in section 4.

[14]The author notes he is not aware of anyone actually disposing of spoiled orange juice. It is perhaps fortunate that orange juice is certainly inventoriable, lasting upwards of 9 months in the pantry and 10+ days in the fridge. Frozen juice lasts indefinitely.

VARIABLES	(1) Sale indicator	(2) Sale indicator	(3) Sale indicator	(4) Sale indicator
Units (100s)	0.120***	0.076***	0.078***	0.076***
	(0.003)	(0.002)	(0.002)	(0.002)
Below Average 1 (100s)	-0.013***	-0.017***		
	(0.002)	(0.002)		
Below Average 2 (100s)	-0.001	0.001		
	(0.002)	(0.001)		
Below Rolling 3 (100s)	0.038***	0.030***	0.021***	0.033***
	(0.004)	(0.003)	(0.003)	(0.003)
Time since last sale (cens)	-0.001***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Time since last sale	-0.002***	-0.003***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Time since, 2 periods	-0.034***	-0.017***		-0.019***
	(0.001)	(0.001)		(0.001)
Time since, 3 periods	0.025***	0.038***		0.038***
	(0.001)	(0.001)		(0.001)
Time since, 4 periods	0.024***	0.030***		0.030***
	(0.002)	(0.002)		(0.002)
Time since, 6 periods	-0.028***	-0.028***		-0.028***
	(0.002)	(0.002)		(0.002)
Below Rolling 2 (100s)			0.001	-0.018***
			(0.002)	(0.002)
Time since, mod 2			-0.064***	
			(0.001)	
Time since, mod 3			-0.039***	
			(0.001)	
Time since, mod 4			0.004***	
			(0.001)	
Time since, mod 6			-0.050***	
			(0.001)	
Observations	1,370,179	1,370,179	1,370,179	1,370,179
R-squared	0.101	0.069	0.078	0.069
UPC FE	X		X	X
Time FE	X	X	X	X
State FE	X	X	X	X
Chain FE	X	X	X	X
Errors	Robust	Robust	Robust	Robust

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10, † p<0.20
Sampling weights used, outliers excluded

Table 2.3: Results for Estimation Specifications

VARIABLES	(1) Sale indicator	(2) Sale indicator	(3) Sale indicator	(4) Sale indicator
Below Average 1 (100s)	-0.018***	-0.021***		
	(0.002)	(0.002)		
Below Average 2 (100s)	0.005***	0.006***		
	(0.001)	(0.002)		
Below Rolling 3 (100s)	-0.001	-0.008***	-0.017***	-0.005*
	(0.003)	(0.003)	(0.003)	(0.003)
Time since last sale (cens)	-0.001***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Time since last sale	-0.002***	-0.003***	-0.002***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Time since, 2 periods	-0.035***	-0.018***		-0.021***
	(0.001)	(0.001)		(0.001)
Time since, 3 periods	0.025***	0.038***		0.037***
	(0.001)	(0.001)		(0.001)
Time since, 4 periods	0.027***	0.033***		0.033***
	(0.002)	(0.002)		(0.002)
Time since, 6 periods	-0.026***	-0.026***		-0.027***
	(0.002)	(0.002)		(0.002)
Below Rolling 2 (100s)			0.001	-0.017***
			(0.003)	(0.003)
Time since, mod 2			-0.064***	
			(0.001)	
Time since, mod 3			-0.038***	
			(0.001)	
Time since, mod 4			0.006***	
			(0.001)	
Time since, mod 6			-0.049***	
			(0.001)	
Observations	1,370,179	1,370,179	1,370,179	1,370,179
R-squared	0.085	0.060	0.069	0.060
UPC FE	X	X	X	X
Time FE	X	X	X	X
State FE	X	X	X	X
Chain FE	X	X	X	X
Errors	Robust	Robust	Robust	Robust

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10, † p<0.20
Sampling weights used, outliers excluded

Table 2.4: Results for Estimation Specifications (No Volumes)

2.5 A Dynamic Model of Loss Leadership

My model is based on the notion of loss leadership: consumers do not purchase products in isolation, but rather shop for bundles of different goods. Consumers bear a cost to shopping trips, and therefore prefer to buy all goods at the same location, which means a low price on one good can make the entire *bundle* of goods more attractive. In my model, there is a single firm selling two goods, a perishable and a storable, to a large number of consumers. There are two kinds of consumers: singles and families. These consumers differ in (1) their tastes for goods (and willingness to pay) and (2) their ability to store. I assume that consumers are distributed in space, with some families (and the singles) located closer to the firm; consequentially, distant families face a higher cost of shopping at the firm[15]. Families have the ability to inventory the storable good and also consume perishables (which are not storable). Singles do not store, and also do not consume the perishable good. The ability to store on the part of the families means that, over time, their stock of the storable good changes. When families are running out of stock, they would like to purchase both the storable and perishable goods; when they are in stock, they would like to purchase just the perishable good. Both types of consumers shop only once per period, and have the option of shopping at the firm or at a local source who sells only the perishable good at cost.

The firm then faces a problem: they cannot sell to all three groups (the singles, the local families, and the distant families) at their reservation prices for their baskets, since shopping costs drive a wedge between the different groups. In a situation where the distant families have high inventories (and low demands for the storable), the monopolist would prefer to sell to the local families and the singles at their willingness to pay. This follows, since in order to sell to the distant families the firm must provide a discount (a sale) on the bundle; which means lowering the price of one of the two goods. If the firm wants to have a sale, it is optimal to lower the price of the perishable good under the condition that the singles are more numerous than the distant families. However, the firm only wants to have a sale if the new market it attracts (the distant families) are more valuable than the profit it gives up by discounting the perishable good for the existing market (the local families). If distant family inventories are high, the firm would prefer not to have a sale; if they are low, a sale becomes more attractive. It is also unattractive to keep the price of the

[15]The distribution of the singles is not important for the analysis; for simplicity, I assume they are all located near the firm.

perishable good permanently low since the firm loses money by competing in this fashion.

This decision is made inter-temporally, with the firm understanding that delaying a sale is beneficial because inventories will fall over time, increasing the profit from the distant families. However, they also understand that when they hold a sale, the out of stock families refill their inventories and become in-stock again, resulting in an inter-temporal trade-off. Nonetheless, a sale will eventually occur, because as time elapses eventually everyone runs out-of-stock. The exact timing of the sale depends on the parameters of the model, but it will reoccur periodically and will be at least one period after the last sale, resulting in a temporary price change (not a permanent one). Families have rational expectations about prices, and sort over time into periods where their bundle is most affordable. The central prediction of this model is typical of loss leadership: sale prices should be associated with higher-than-normal purchasing of other goods, particularly storables. I also demonstrate a number of extensions, illustrating that this model is fairly robust to different assumptions about consumer behaviour and firm expectations.

Formally, we can model this as a dynamic game of complete information as follows. The first player is a monopolistic firm which produces and sells two goods $j \in \{s, p\}$ at constant marginal costs c_j. The good s is a storable good, while good p is a perishable good; these properties will be given a precise meaning shortly. The game takes place in a discrete time infinite horizon setting, in which all players make decisions in time periods $t = 0, 1, 2, \ldots, \infty$. Each period, the firm sets a single price P_j for a unit of good j with the objective of maximizing their total, inter-temporal, profit from the whole game. The firm discounts profit in time period t by rate δ^t. In competition with the monopolistic firm, there is also a "local" firm which sells only good p at cost; that is there is always an outside option to buy the perishable good at price c_p.[16]

The second set of players is a large number of consumers who occur in two types $i \in \{F, S\}$ with masses m_i. These consumers differ in both their tastes for the goods, and their ability to store. Each consumer is constrained to shop at exactly one store per period; they must do all their shopping at a single store, if they wish to buy anything. I make this assumption to focus on the consumer's purchasing decision, rather than purchase frequency. Consumers of type S do not store and consume only good s[17], while consumers

[16]This reflects the fact that the perishable good is widely available and a low-margin product, from the point of view of the monopolistic firm.

[17]This is equivalent to assuming that their valuation for the good is less than c_P and so the store would never sell to them

of type F store k units of the good and may consume both types of goods. Furthermore, suppose that the families are distributed spatially in the economy; a fraction $\lambda \in (0,1)$ are located at a distance from the firm, while the remainder are local to the firm. This implies that for the distant families, they face a shopping cost $\tau > 0$, while the local families and singles have a cost normalized to zero.[18] I will refer to these two sub-types as distant families (DF) and local families (LF).

At this point, I must make a decision about how consumers behave inter-temporally. Following Pesendorfer (2002), and in the interests of simplicity, I assume that consumers follow a reservation pricing strategy: specifically, they have reservation prices v_j^i for the goods. This implies that if, in a given period, a one-good buying consumer of type i demands good j and $P_j < v_j^i$ then the consumer buys the good. Similarly, if a consumer is buying more than one good, then the sum of the prices must be less than the sum of the reservation prices. Consumers of type F consume both types of goods, but not every period; they only want to buy good s in periods when their inventory is not completely full. These reservation prices can equivalently be thought of as choke prices for the goods, with the consumers having unit demand for the goods. As I will show, in equilibrium this strategy is rational; the lack of explicit inter-temporal optimization does not restrict the analysis.

Finally, to ground the movement of consumer inventory, notice that the type F consumers can be in one of k states, corresponding to the amount of inventory left. Let the distribution of these consumers across the states at period t be denoted x_t. Furthermore, suppose that x_t evolves according to a Markov chain M with the properties that (M1) M is strictly left-to-right[19] and (M2) M has exactly one absorbing state at $x^0 \equiv (0,0,...,0,1)$.[20] These conditions on the Markov chain amount to assuming that if the consumer has an inventory level I_t at period t, that $I_{t+1} < I_t$ and that the $\lim_{t\to\infty} I_t = 0$. In other words, consumer inventories can only decrease over time, and that they eventually run out of inventory. The important assumption here is that this process is Markovian, in the sense that only the mass of consumers in the state matters, not their tenure in that state, and that they eventually all run out of stock. Finally, I assume that consumers shop at the lowest possible cost, subject to the reservation prices. If consumers are indifferent between the

[18] We could make the equivalent assumption about the singles, without loss of generality; however, as we will see, distant singles would not add anything to the analysis.

[19] That is to say, that $M(i,j) > 0 \iff j \geq i$ and $M(i,i) = 0$. This strictness can be relaxed, but at the expense of considerable additional complexity and little added to the model.

[20] That is to say, the only stationary point of the distribution is with all consumers stocked out

monopolist and the local store, they choose the monopolist (since they have a wider selection of products; a consideration I do not model explicitly). Two particular states will be useful to define: $x^1 \equiv (1, 0, ..., 0)$, the state in which all consumers have filled up their inventories completely, and $x^2 \equiv Mx^1$, the state immediately following the fully-stocked state.

For the results, some assumptions are necessary beyond the modelling set-up itself:

- (A1) Assume that $v_s^S = v_s^F = v_s$; both types of individuals individuals have the same reservation price for the storable good.[21]
- (A2) Assume that $v_s > c_s + \frac{\tau}{\lambda k}$; that it is possible for the firm to make a profit from attracting the distant families. Note, this also implies that both goods are attractive for the firm to sell.
- (A3) Assume that $v_p = v_p^F = c_p$; that the families choose reservation price at marginal cost, which is rational since the local source offers at this price.[22]
- (A4) Assume that $m_S > k(1-\lambda)m_F$; the singles buy more than the local families.

Under assumptions (A1)-(A4), the model can be analysed as follows.

2.5.1 Analysis

The first step to analysing this model is to note that from the firm's point of view, they have three different markets. The local families with mass $(1-\lambda)m_F$, the distant families with mass λm_F, and the singles with mass m_S. Suppose a distant family has an inventory of I_t; then, they would like to buy up to $k_t \equiv k - I_t$ units of the good. They will buy both items at the firm when:

$$v_p + k_t v_s + \tau \leq P_p + k_t P_s$$

If this condition is not met, they will forgo buying the storable good, and instead buy just the perishable good locally at a cost of c_p. The firm will never set a price $P_p > c_p$,

[21] This is primarily to simplify the analysis; a difference in willingness to pay leads to essentially the same conclusions. Note that the distant families effectively must have a lower willingness to pay, since they also face a shopping cost τ.

[22] This assumption only for notational and expositional convenience. If we relax assumption (A3), the key change is that assumptions about the profitability of the storable good, become statements about overall profitability; all the intuition and results are the same.

since no one will buy perishables from them at this price; this also implies that there is no situation in which the distant families will buy only the storable good; if they're buying the storable, they will buy the perishable as well, since it must be at least as inexpensive as the outside option.

Now, to begin the analysis, consider a situation in which the firm is only interested in attracting the local families and the singles. It is optimal for the firm to set $P_p = c_p$ and $P_s = v_s$, since this means that both singles and local families buy from the firm and obtain no surplus; call these "regular prices." At the regular price, both consumers are priced at exactly their marginal cost. Define $k^*(x)$ to be the average of x, weighted by the inventory; this is the average inventory demanded by families in state x; the total demand will be $m_f k^*$. This yields a total profit of (under (A3)):

$$\pi^R(x_{LF}, x_{DF}) = m_S(v_s - c_s) + (1 - \lambda) m_F k^*(x_{LF})(v_s - c_s)$$

Now, notice that the firm cannot charge the same prices and attract the distant families, since $\tau > 0$. This means there is a trade-off between which consumers the firm attracts; in order to attract the distant families, the firm must offer a total discount (relative to the regular prices) of $\Delta = \tau$. This is discount is precisely why this is a model of "sales" rather than general price-setting. In order in order to sell to the entire group DF, the firm must set either (a)$P_p = c_p - \Delta$ or (b) $P_s = v_s - \Delta$. That is, they can discount either the price of the perishable product, or the price of the storable product; the linearity in utility means they would never use a mixture of discounts, and similarly would never increase the price of one good, since it only makes attracting this group more difficult. Notice that the discount on the storable is higher than necessary for most consumers; a consumer with level k^* will be attracted when $P_s = v_s - \frac{\Delta}{k^*}$; the firm could sell to just some of them by setting a lower discount. The firm is trying to decide which of these possible strategies to use; the following Lemma shows that under assumptions (A1)-(A4), is it always better to use (a).

Lemma 2.5.1. Suppose (A1)-(A4) hold. Then, the firm prefers to discount the perishable product, rather than any discount on the storable good.

Proof. From the monopolist's point of view, the "best" possible discount would be to set $P_s = v_s - \frac{\Delta}{k}$, since this is the smallest possible amount they could discount the bundle and still attract consumers. Notice, similarly, that the best possible situation would be to attract

DF families all buying k; the profit from selling to this (maximal purchasing) group at this discount is the highest possible for any discount level. It is important to note that this is an upper bound on the profit from a strategy of type (b); this cannot actually occur, because not all consumers would be buying k goods and so would want a higher discount. In any state other than the one outlined, the profit from the lowest discount would be lower. You also cannot do better selling to any smaller group than this, since they must obtain the same total discount but buy strictly less of the storable good. In this situation, the profit from strategy (a) is:

$$\pi_a = \lambda m_F(v_p - \Delta - c_p + k(v_s - c_s)) + (1-\lambda)m_F(v_p - \Delta - c_p + k^*(x_{LF})(v_s - c_s)) + m_S(v_s - c_s)$$

Similarly, the upper bound on the profit from (b) is:

$$\pi_b = \lambda m_F(v_p - c_p + k(v_s - c_s) - \Delta) + (1-\lambda)m_F(v_p - c_p + k^*(x_{LF})(v_s - c_s) - \frac{k^*(x_L)}{k}\Delta) + m_S(v_s - c_s - \frac{\Delta}{k})$$

Substituting in (A3), and comparing the terms we see that:

$$\pi_a > \pi_b \iff -(1-\lambda)m_F\Delta + (1-\lambda)m_F\frac{k^*(x_{LF})}{k}\Delta + m_S\frac{\Delta}{k} > 0$$

$$\iff \frac{m_S}{k} > (1-\lambda)m_F(1 - \frac{k^*(x_{LF})}{k})$$

Since the smallest $k^*(x_{LF})$ can be is zero, under assumption (A4), the result follows. □

The key intuition is that the two strategies essentially target different groups. If you discount the perishable good, you attract the distant families but lose profit on the existing local families who are buying the perishable good in any case. If you discount the sundry, you lose profit from both the local families and the singles. The best possible situation is when the local families aren't buying the storable in that period anyway, but the profit from the singles is still foregone. Consequentially, if the singles (relative to the best-case purchasing from the distant families) are more numerous than the local families, the trade-off falls squarely on the side of the perishable good. This is the requirement that (A4) states explicitly.

With this established, the firm's problem becomes more straightforward; we have essentially eliminated one method to attract the distant families. This also has a very important consequence: for both groups of families, if they shop at the firm's store, they always completely refill their inventory regardless of their inventory position. This is important, so I will state it below:

Corollary 2.5.1. If a family shops at the firm, they will always refill their inventory of storable goods.

Proof. Notice that given Lemma 2.5.1, there are only two possible prices regimes for the monopolist over time, and in both situations the price of the storable good is always $P_s = v_s$. In other words, given the shopping decision, the consumer always earns zero surplus from the storable good. Therefore, it is strictly better for them to buy a maximal quantity of the storable good whenever they purchase it, since they will never see a better price in the future and may run out. □

This fact dramatically reduces the complication of the firm's problem, since in a given period the family demand for sundries is exactly their outstanding inventory position. That is to say, if the average stock position is k^* then families will buy exactly k^* in aggregate, scaled by their mass. There is no inter-temporal dynamics for consumers, given the price sequence; they simply refill their inventories when it is profitable to do so. This similarly implies that the local families will only even be in state x^0 or x^1; they have either just refilled their inventory, or will be refilling it this period; they never accumulate losses of products. In effect, they act exactly like the singles. Notice, additionally, that since families are infinitesimal in size, there is no sense in which they would like to buy "less" than a full inventory in order to try and induce the firm to change its pricing pattern; their individual demand is simply too small to matter.

Turning attention back to the distant families, if the firm wants to attract them, it must hold a sale on the perishable good. The question is, would the firm want to attract them in the first place? For the firm, if they focus on just the local market, they obtain profit $\pi^R(x_{LF}, x_{DF})$, as defined above. However, if they choose to hold a sale, Lemma 2.5.1 implies that they earn a profit of (under (A3)) of :

$$\pi^{\Delta}(x_{LF}, x_{DF}) = m_S(v_s - c_s) + (1-\lambda)m_F k^*(x_{LF})(v_s - c_s) + \lambda m_F k^*(x_{DF})(v_s - c_s) - \Delta m_F$$

Comparing these two terms, we see that

$$\pi^R(x_{LF}, x_{DF}) < \pi^\Delta(x_{LF}, x_{DF}) \iff \Delta < \lambda k^*(x_{DF})(v_s - c_s)$$

Notice, that since $\Delta = \tau$ that when $k^*(x_{DF}) = k$ we have that $\tau < \lambda k(v_s - c_s) \iff v_s > c_s + \frac{\tau}{\lambda k}$ which is assumption (A2). In other words, assumption A2 ensures that there is some state x^*_{DF} such that the firm would prefer to hold a sale in this period. Similarly, for all states $x' = M^y x^*_{DF}$ for $y > 1$ the firm would also want to hold a sale, since the evolution of inventories follows a left-to-right Markov chain (condition (M1)). Additionally, we can note that starting from any inventory position eventually such a state must be attained because condition (M2) implies that:

$$\lim_{t\to\infty} k(x_t) = k$$

That is, since eventually all consumers will run out of inventory, it is eventually optimal for the firm to hold a sale. Finally, we can complete the characterization of the behaviour of the model in a temporal setting by noting that after a sale the families all move to state x^0. This implies that the model is stationary, in the sense that if it optimal to have a sale T periods after x^0, then it is optimal to have a sale T periods after T as well.

The preceding analysis simplifies the problem and helps illustrate the characteristics of the problem facing the firm. However, the firm's problem is not static; they face an intertemporal decision and understand that when they hold a sale, the model will move back to state x^0. Essentially, they face a trade-off; by holding a sale now, they refill inventories and obtain the profit from the distant families. However, by delaying a sale, inventories run down more and the marginal profit (relative to the regular prices) increases. The firm solves such a problem in a dynamic setting; however, the preceding work allows us to formulate this as a straightforward dynamic programming problem on the part of the firm, which results in the following theorem.

Theorem 2.5.1. There exists a unique pricing path in which firms place the perishable good on sale.

Proof. Denote $X_t = (x_{DF,t} x_{LF,t})$ and recall that $x_{LF,t} \in \{x^0, x^1\}$. Letting the sale decision be $a_t \in \{0, 1\}$, we can write this problem now as a well-defined value function $V(X_t) = \max_{a_t \in \{0,1\}} a_t(\pi^R(X_t) + \delta V(X_{t+1}(0))) + (1 - a_t)(\pi^\Delta(X_t) + \delta V(X_{t+1}(1)))$, where the transition function is given by:

$$X_{t+1}(a) = \begin{cases} (Mx_{DF,t}, x^1) & \text{if } a = 0 \text{ and } x_{LF,t} = x^0 \\ (Mx_{DF,t}, x^0) & \text{if } a = 0 \text{ and } x_{LF,t} = x^1 \\ (x^0, x^0) & \text{if } a = 1 \text{ and } x_{LF,t} = x^1 \\ (x^0, x^1) & \text{if } a = 0 \text{ and } x_{LF,t} = x^0 \end{cases}$$

The dynamics on x_{LF} are included for completeness, but do not affect the decision of the firm. We can immediately note that this is a bounded function since π^R and π^Δ are bounded and by the note above, the time before a sale must be finite. Next, we can note that this mapping $T(V(X_t)) \equiv \max\{\pi^R(X_t) + \delta V(X_{t+1}(0)), \pi^\Delta(X_t) + \delta V(X_{t+1}(1))$ meets Blackwell's Sufficient Conditions: first, it is monotone since if $W \geq V$ for all x then $T(W) = \max\{\pi^R(X_t) + \delta W(X_{t+1}(0)), \pi^\Delta(X_t) + \delta W(X_{t+1}(1)) \geq \max\{\pi^R(X_t) + \delta V(X_{t+1}(0)), \pi^\Delta(X_t) + \delta V(X_{t+1}(1)) = T(V)$. It is discounted since for any $A \in \mathbb{R}$, $T(V + A) = \max\{\pi^R(X_t) + \delta V(X_{t+1}(0)) + \delta A, \pi^\Delta(X_t) + \delta V(X_{t+1}(1) + \delta A) \leq T(V) + \delta A$. Therefore, it is a contraction mapping and has a fixed point; that is, there is a unique optimal value function which satisfies the above. The existence of a policy sequence $\{a_t\}$ which admits sales follows from the existence of such a value function and the fact that sales must occur, as established above. □

Beyond existence, I am also able to use the facts we established earlier about the properties of the equilibrium to make the following observations.

Corollary 2.5.2. (Properties of Equilibrium) The unique equilibrium of the model has the following properties: (1) sales will occur (in the sense of a temporary price reduction). (2) Sales will occur periodically and indefinitely. (3) When sales occur, they occur on the perishable good (not the storable). (4) Consumers who buy the perishable good on sale also buy more of the storable good than those who buy the perishable good at the regular price. (5) There are exactly two prices charged in equilibrium for the sale (perishable) good, low and high.

Proof. The first property follows immediately from the fact that a_t is not identically zero, since eventually sales must occur. The second property is more subtle, but holds because once a sale occurs, the key state variable in the model $x_{DF} = x^0$. This implies that after every period in which a sale happens, the state of the economy is the same, and as the comment above mentions, this model is stationary.. This implies that sales will occur periodically. Property 3 follows from Lemma 2.5.1, as explained earlier. Property 4 follows

from the fact that the distant families buy k^* units of the sundry good during sale periods, while the local families buy at most $k^*(x^1)$, which can be (functionally) imagined as a normalization. If it is worth it to wait a little while to hold sales, as the empirical results show, the demand will be higher. Property 5 follows from the fact that the price of perishable good is either c_P or $c_P - \Delta$, a distribution of prices characterized by a (high) regular price and a (low) sale price. □

Most of these properties are what we would expect from the evidence presented in Section 4; they capture many of the stylized facts about perishable pricing. However, the key idea of loss leadership is not easy to test at the retail level, since it is a statement about consumer behaviour, not firm behaviour. In particular, property 4 is not testable using the retail data but makes a clear prediction: consumers who buy the perishable good on sale should also buy more of other goods, in particular of profitable storable products. I look at this suggestion, along with other considerations, in Section 6.

2.5.2 Discussion

This model explains many features of the sale pricing of perishable goods, with a sharp characterization of both the type of good placed on sale and the manner in which it is sold. Some of the precision of these conclusions is the result of the stark nature of some aspects of the model; these assumptions were made in order to focus on the intuition and dynamics, rather to present a "realistic" model of store decision-making in a multi-product environment. In effect, the goal of the model is to present one "piece" of the store's complete pricing puzzle, in which they face many more problems but also have many more tools and dimensions with which to solve them. In the model, I focus on the family/single distinction, but this is just a framing mechanism for the intuition; in reality, grocery stores, using consumer loyalty cards and purchasing information, are likely able to come up with far more subtle connections between consumer demands. In fact, while it is useful to give the single type of consumer a frame to guide to intuition, in the model they play the role of "all other sundry buying consumers" that the grocery store is seeking to sell to. This also helps motivate the assumption that their demand is large, relative to the local families; I am not explicitly making a statement about two demographic groups, but instead saying that the group which the firm seeks to distinguish using loss leadership is small, relative to all other consumers. This is a way of attempting to disentangle the

firm's problem and consider just one class of goods, instead of considering the problem as a whole.

This discussion also highlights the important role of the assumption that perishable purchasing perfectly separates the two consumer types. If singles bought some of the perishable product, the relative sizes of the two groups m_i would matter more directly, since the firms trade-off between in pricing the products becomes more entangled. Firms would still need to jointly price each bundle, but now they would face the problem where they cannot charge completely independent prices for the different bundles. This is related to the discussion over how firms select certain products to be loss-leaders; in this model, and that of DeGraba (2006), a product is selected for loss leadership on the basis of an association with another, profitable product. My model, with its explicit treatment of shopping costs is closest to Lal and Matutes (1994) where loss-leaders are selected based on consumer willingness to pay, but trades off dynamic pricing and inventory for explicit modelling of expectations and advertising. This is similar to the motivation in Chen and Rey (2012), where firms use consumer bundles to attract shoppers to their store. However, it differs with explanations like Johnson (2014), which have to do with the salience of the product being offered. Nonetheless, these two models make similar predictions; if we believe that perishable stock is well-understood by consumers[23], this should be selected in the model of Johnson (2014) as well as my model. Overall, this makes it difficult to rule out a behavioural component; the two motivations are complementary, from a firm's point of view. A more subtle point has to be why we expect the sundry/perishable relationship to be focused on by firms; after all, there are many possible relationships which could be exploited, using different types of consumer demands. The behavioural complementarity is one possibility; an intuitive alternative is that sundry goods are highly profitable, so if a firm was focused on trying to increase volume in its profitable segments, they would focus on sundries which would (if the model holds) lead them to discount perishables. This would be in contrast, or alongside, sales on other goods with desirable affiliations; a model which is not developed in this paper.

The assumption about the inventory process of the singles being very stark (no storage, unit demand) is also not essential; as I mentioned, this group of consumers is really a representation of "all other" demand for the sundry good not related to the perishable product. In this case, unit demand could be easily related. Storage, however, would create

[23] A good test is as follows: how much chicken do you have in the fridge at home? How many weeks of paper towels? Which of these questions was easier to answer?

a dynamic feature to both the families and singles which would change the condition for within-period optimality of a sale on perishable to be $m_S^t \geq m_F$ which may or may not hold. As mentioned in the analysis, the local families act "like" singles in a sense; giving the singles inventory dynamics would make them even more similar. In this case, sales on sundries could occur in periods in which the firm considers the demand to be relatively low. However, there would (as in the original model) be inter-temporal considerations necessary which complicate straightforward analysis. Similarly, the notion that the completely out-of-stock consumers buy no storables when they run out is really just a simplification; we can imagine them buying from the local source some "hold-over" amount at an unattractive price to keep them going while they wait to refill their inventory. This variation on the model does not change the implications, but is omitted since it adds complication at little explanatory benefit.

The Markovian nature of the consumer inventory is also a restriction, primarily for simplicity. Families will always consume some inventory, but the amount they consume is only dependent on their existing stock, not their prior consumption. In other words, at the micro-foundation level the consumers are being hit with independent demand shocks at the end of each period. In aggregate, these effectively average out since there are very many consumers. However, a more robust process is also feasible at the expense of considerable additional complication on the part of the model. I omit this to focus on the connection between products, rather than trying to perfectly model how consumers deplete their inventory of a storable good.

A final general comment is about the reservation price assumption used; this is a limitation in the set of equilibria we are considering. A more complicated model would explicitly model a "waiting cost" to being out of stock, then have consumers inter-temporally sort based on their expectations and understanding of the state. However, within the model, the restrictions our assumptions place are minimal. For instance, the assumption that $v_p^F = c_p$ is rational; consumers in equilibrium face either a price $P_p = c_p$ or $c_p - \Delta$ both of which meet the reservation price restriction; the reservation price is consistent with consumer expectations. The other reservation price, v_s is also rational since only a single price (the expected price) will be charged in equilibrium. Notice that it is also rational for consumers to fully refill their inventory when they are out of stock and see a sale; since sales are always the same and always repeat, there is no benefit to delaying since the optimally lowest price has already been achieved. These features demonstrate that our equilibrium, while of a particular sort, includes both forward looking and fully rational consumers in a

framework which I believe is a good approximation of how consumers might make these kinds of choices in the real world.

Competition, Duopoly, and Monopoly

In this model, the firm serves effectively as a monopoly; they are the only one setting prices. However, the existence of the outside option makes this not a pure monopoly environment; consumers have the opportunity to buy the perishable good elsewhere, if necessary. If the outside option did not exist, there would be no trade off on the part of the families; they *have* to go to the store every period no matter what. This would allow the monopolist to always charge regular prices, and the model would not capture the dynamics we are interested in. Alternatively, we would have to model forgoing the perishable, which greatly complicates the model by adding more (undesirable) inter-temporal dynamics on the part of the families. It is not necessary for the outside option's price point to be c_p; this is primarily for notational and expositional reasons, as explained in the model set-up. However, it does has the (appealing) feature of capturing the fact that perishables are relatively low margins with many substitutes, while storables are not.

In order to analyse the role of competition more directly, I also develop in Appendix A a general model of loss leadership which nests both the duopoly and competitive alternatives. The set up is similar, in which we have two groups of consumers: bachelors and families. These two types of consumers have different lifestyles: bachelors live alone, in small apartments, while families live together in larger houses. This manifests in two related ways: first of all, families have more storage capacity for the sundries than bachelors, and secondly, the families consume more perishables per capita from the grocery store, since the bachelors find it difficult to cook and prepare for one (and so eat out more, instead). These basic differences between the consumers will be ultimately result in sales, since inventory dynamics on the part of the families lead them to occasionally purchase more than the bachelors. The fact that they also purchase meat gives the stores a way to price discriminate between the two consumers, offering a more attractive total bundle price to the families when they are buying large amounts of sundries. High purchases of sundries occur in a cyclic fashion alongside low prices of the perishable good, resulting in a cyclic pattern for sales.

The key difference between the two models is that in the duopoly setting, the consumers are more limited in their decision making and sophistication, which results in a

closer connection between the inventory process and the sale pattern. In our model, this is moderated through the reservation price, inventory dynamics, and inter-temporal sorting. The conclusions from both models are largely similar, with sales driven by the connection between consumer inventories and tastes. However, the existing model makes sharper predictions and admits a better representation of the consumer decision makers, which makes it my focus in this paper. On the other hand, the conditions required of the inventory process (specifically, its periodicity) are more clearly spelled out in the duopoly setting, as are the implications of competition. As in the pure monopoly case discussed above, competition eliminates sales, but this time as firms compete away their profits; it is clear that in all three settings, the role of profitability of the different consumer groups to the firm is very important. It is also important to note that the conclusion of this model seem largely robust to variations in the set-up of the model; similar conclusions can be reached through a variety of modelling approaches, which illustrates that the underlying economic intuition is robust.

Quantity Discounts and Other Competitive Strategies

As mentioned earlier, the strategy of discounting perishable products to attract the consumers who buy more of other (profitable) products is just one tool available to firms. A natural alternative is the use the of quantity discounts. Note that in the model the single individuals always purchase unit demands of the storable good, while the families purchase up to their k units. It would be sensible that an alternative way for the firms to compete would be in terms of a quantity discount for the package. This is certainly possible, and could be an alternative pricing strategy, since it would allow them to effectively charge two prices to the two different sets of consumers. However, this relies closely on the assumption of unitary demand on the part of the single individuals. As explained earlier, this is not a fundamental part of the model; it is just a frame of reference, to provide a comparison with the families.

In the discussion of Lemma 2.5.1, it was clear that different consumers with different demands would need different marginal discounts. Offering a range of quantity discounts would be a way to implement this, resulting in perfect discrimination between consumers and no incentive to put sales on. However, if packages don't fully cover the range of demands, this would not be an effective strategy; for instance, perhaps the wholesale producers of the goods do not make all package sizes, or perhaps there are costs to offering

many products. The heart of the trade-off facing the firm is the cost of lowering the price of the sundry versus the price of the perishable; package size is one way of making this trade-off more granular, and therefore enact a more profitable strategy. The firm seeks, fundamentally, to discriminate between different types of consumers; different characteristics of demand, including both bundles composition and package size can do this.

This highlights the important role that viewing sales on perishables as part of a pricing *strategy* on the part of the firm plays. Not all pricing or competitive features are captured by this model, and the existence of other kinds of strategies are not evidence against the model. Instead, they show the complex manner in which stores can price products, and the overlapping considerations different strategies must take into account. In particular, there is a close relationship between sales and quantity discounts in this model; this is made even more explicit in Section A.3 which outlines a sense in which quantity discounts can be weakly dominated by sales.

2.6 Going Behind Sales: Consumer Choice and Sales Pricing

Most of the predictions of our theoretical model occur at the retail level, and capture the stylized facts I presented in Section 2.4. However, the prediction that perishable-buying consumers also purchase more storable goods in sale periods when compared to non-sale periods is not testable using retail level. In order to test this prediction (and rule out other explanations), in this section I go beyond the retail level data by extending my data set using a household panel dataset, the Nielsen-Kilts HOMESCAN or HMS survey.

2.6.1 The Nielsen-Kilts HMS Survey

In addition to the retail-level data used in section 2.4 to both impute when sales have occurred, then to determine the different factors which are related to when they occur, the Nielsen-Kilts dataset also collects a companion dataset knowns as the HMS. This is a large panel of consumer scanner data; households, selected by Nielsen on the basis of their factors, household demographics, and other factors are tracked over several years as they buy products from different stores. The intent of the dataset is to be representative of the United States population as a whole; thus they employ a particular sampling methodology to select and retain households in the survey. Each household is given a household scanner, with which they record the amount and price of products they purchased at retail shopping

"trips." For each trip, the products and their prices are recorded, as well as information about which store the goods were purchased from. This information allows a link to be made between the consumer level and the retail level, since Nielsen uses the same UPC and same store code at both levels.

The scale of this dataset is very large, although much smaller than the retail dataset used earlier. The dataset is also divided into two sub-groups, referred to as "magnet" and "non-magnet." The magnet households also record spending on loose (non-UPC) coded items. The non-magnet households do not record such data. Unfortunately, the sampling methodology used by Nielsen means that the two groups are not comparable; analysis must be done with either the magnet or non-magnet households. I restrict my sample to just the non-magnet households for three reasons: first, the data is more reliable, since magnet data needs to be entered manually. Second, the magnet households are much smaller in number than then non-magnet group. Third, the number of matches I can make between my retail data and the households is larger for the non-magnet subset. I show some preliminary statistics for this dataset in Table 2.5. Each year contains around 51 million different products across 5.5 million trips, with around 60,000 households each spending approximately $75 per trip. However, there is substantial variation in the spend; some households spending nothing on a given trip, while others spending thousands of dollars.

In order to match my retail-level data, I use the years 2010-2013 of the consumer panel dataset. In order to cross-reference with my sales data from the preceding step, I match the UPC-store combinations from the retail level with the UPC-store combinations by week. One problem arises in that the consumer panel is recorded daily (or, at least whenever a consumer makes a trip) while the retail data is aggregated to the weekly level. In order to match them up, I re-assign the consumer panel observations to the week in which they would have been recorded at the retail level. Effectively, this means that they are recorded as being the Saturday of the week in which they were originally recorded.

Since the goal is to examine the bundles individuals buy when they are, or are not, on sale, I collapse the dataset down to the trip-level. For trips which do not include any of the UPC codes in the linked retail-level dataset, I omit these from consideration. The resulting dataset is described in Table 2.6. The consolidated dataset contains 28,474 unique trips, of which about 23% include an item purchased on sale. Typical spending is about $69.68 per trip; this is lower than the complete panel outlined in Table 2.5 primarily because this subset restricts attention to only grocery stores, while the complete panel includes mass-

merchandisers (like Walmart) or pharmacies which have higher cost items in general. I also break out the non-linked spending by sale or non-sale; in general, it is slightly lower for sales at $62.75. The average trip is made by a household of 2 adults and 1 or 2 children, who live in a single family residence, and make between $50,000-$60,000 a year. They are generally married or cohabiting, 50-54 years old, with high-school or college education and work full-time or are retired and are white, with non-Hispanic origins. In what follows, these individuals are weighted using the proportional sampling weights provided by Nielsen.

2.6.2 Empirical Model and Results

To begin looking behind the data, I first try to see what the characteristics of individual bundles are once they are chosen by consumers. Specifically, I denoted the linked items, from which I can tell whether or not they are on sale, as targeted items (as in, can be targeted for being on sale). I then calculate the amount spent by consumers on products other than those targeted, and compare how this reacts. My general specification will be to use a regression of the form:

$$Y_{ij} = S_{ij} \times H_{ij}\theta + B_j\beta_1 + C_{ij}\beta_2 + \epsilon_{ij}$$

where i is the trip and j is the household. The dependent variable, Y_{ij} is the spending on other goods, S_{ij} is an indicator for whether the targeted good is on sale, B_j and C_{ij} are controls which deal with the different household characteristics and trip characteristics (for instance, how large their bundle is, or their income level). The term H_{ij} is an interaction to capture different effects across consumers and goods; in different specifications it has different definitions.

For example, in the first specification, I regress this total all-other sales variable on an indicator for whether the targeted product was on sale, and a full set of time, UPC, and demographic controls. I also include the average spending by household, to try an correct for any unobserved heterogeneity in spending patterns. Finally, because not all products are likely to react similarly to sales, I include interactions between the sale indicators and the UPC, creating a "UPC-specific" sales variable. These results are reported in Table 2.7; because the interaction variables are very numerous, I break these out in the table separately. Of the 285 iteration terms which are not naturally coded, 90 are omitted due to insufficient variation. Of the remaining 195, at the 10% level approximately 38 are

	2010	2011	2012	2013
Products purchased	51.67m	54.13m	50.94m	50.93m
Trips	5.52m	5.76m	5.39m	5.30m
Households	60,423	61,824	60,315	60,916
Mean spending per trip	$73.65	$76.20	$77.56	$79.40

Table 2.5: Statistics for the HMS Survey

VARIABLE	Obs	Mean	Std. Dev.	Min	Max
Spending on Linked Product	28,474	6.57	4.99	0.59	232.29
Spending on Sale	28,474	1.39	3.23	0.00	107.73
Total Spending	28,474	69.68	62.74	1.38	1086.99
Average non-sale spending	28,474	63.63	51.28	0.00	679.44
Average sale spending	28,474	62.75	51.72	0.00	690.44
Spending on Non-Linked Items	28,474	63.16	61.99	0.00	1080.12
On Sale Indicator	28,474	0.231	0.421	0	1
Categorical: Higher implies more for numerical variables					
Income	28,474	21.38	5.73	3	27
Household Size	28,474	2.41	1.17	1	9
Residence Type	28,474	1.71	1.65	1	7
Composition	28,474	2.22	2.07	1	8
Age	28,474	7.75	2.51	1	9
Female Head Age	28,474	6.61	2.53	0	9
Male Head Age	28,474	5.75	3.31	0	9
Male Head Employment	28,474	3.74	3.23	0	9
Female Head Employment	28,474	4.81	3.44	0	9
Male Education	28,474	3.40	2.03	0	6
Female Education	28,474	4.03	1.53	0	6
Male Occupation	28,474	5.69	4.79	0	12
Female Occupation	26,953	6.48	4.93	0	12
Martial Status	28,474	1.65	1.08	1	4
Race	28,474	1.27	0.69	1	4
Hispanic Origin	28,474	1.96	0.20	1	2
TV Items	28,310	2.22	0.84	1	3
Internet	28,474	1.09	0.28	1	2

Table 2.6: Summary Statistics for Consolidated Trip-Level Data

significant while the remainder cannot be distinguished from zero. I show these, along with the negative of the level of the sale variable in the lower panel; the bounds indicate an 90% confidence interval. Wald tests on the majority of these coefficients indicate that the sum of these two terms is also different from zero at the 10% level.

Dependent variable: A/O Sales		
Variable	Coeff	SE
Average Spending	0.939***	0.011
Sale Indicator	-12.699	9.376
UPC Controls	X	
Time Controls	X	
Demographic Controls	X	
UPC X Sale Interactions		
Total Number	285	
Omitted	90	
Signifigant ($p = 0.10$)	38	
Insignifigant	157	

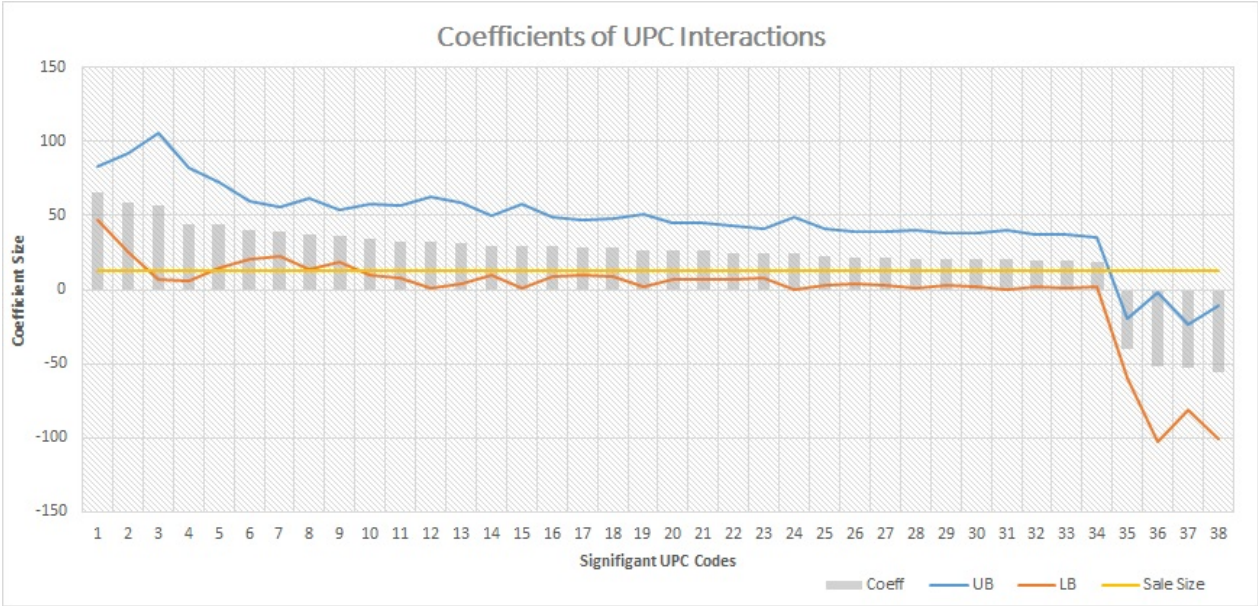


Table 2.7: Regression of A/O Spending

This indicates that for some products, the purchases of all-other goods increase when the project goes on sale, although the effect is mixed and weak. This makes me suspect that the characteristics of the consumer's bundle matter; this is explicitly the case in the theoretical model. So, in order to control for the bundle of products being purchased, I divide goods into ten categories, corresponding with the retail departments in a typical grocery store. These categories are Health and Beauty, Non-food Grocery[24], Alcohol, General Merchandise[25], Dry Grocery[26] items, Dairy, Deli, Frozen foods, Packaged meat (the target category), and Fresh produce. The first four are referred to as `Sundries`; when this is extended this to include Dry Grocery, I call the group `Sundries+`. In Table 2.8, I report the regression of the full set of controls on a sale indicator and the number of categories; this is the baseline result. In specification (2), I add in the average spend variable. Specification (3) removes this variables, as well as the UPC level controls. Specification (4) and (5) repeat (1) and (2) focusing just on `Sundries`. The final specification repeats the baseline for the `Sundries+` variable instead. I focus on sundries because my model makes the prediction that these kinds of inventoriable goods should be closely related. Moreover, sundries are known to be one of the more profitable and high volume segments of the retail grocery market. As in my model, and based on DeGraba (2006) and Pesendorfer (2002) these desirable (profitable) segments are likely to be bundled by sale-buying consumer.

[24]Paper towels, toilet paper, plastic wrapping, etc.
[25]Glue, pots and pans, camping supplies, etc.
[26]Cookies, crackers, canned soups, condiments, etc.

VARIABLES	(1) A/O spending	(2) A/O spending	(3) A/O spending	(4) Sundries	(5) Sundries	(6) Sundries+
Average A/O spending		0.72***			0.08***	
		(0.01)			(0.01)	
Sale indicator	5.29*	6.51***	3.83†	3.42***	3.56***	3.50*
	(2.73)	(2.13)	(2.64)	(0.83)	(0.80)	(2.03)
Num of Departments	22.30***	11.78***	22.47***	4.93***	3.74***	13.84***
	(0.31)	(0.30)	(0.31)	(0.11)	(0.11)	(0.23)
Sale × num departments	-2.01***	-1.74***	-1.90***	-0.83***	-0.80***	-1.42***
	(0.62)	(0.48)	(0.61)	(0.19)	(0.18)	(0.46)
Observations	26,862	26,862	26,862	26,862	26,862	26,862
R-squared	0.54	0.74	0.51	0.34	0.37	0.47
Demographic Controls	X	X	X	X	X	X
UPC FE	X	X		X	X	X
Time FE	X	X	X	X	X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10, † p<0.20
Sampling weights used, outliers excluded

Table 2.8: Regressions on number of departments and sales

First, we can see immediately from the results that sales increase the amount spent on other goods; between $3.8 and $6.5. This indicates that consumers are not substituting away from goods in favour of the sale product, which would be undesirable from the store's point of view. However, we can also see that this effect is not the same for all bundles of goods. Consumers purchasing in more departments (having "larger baskets") mechanically spend more, but consumers buying a larger basket have a diminishing effect on sales. This is plotted in the first panel of Figure 2.6; we can see the break-even for all-other goods is somewhere between two and three departments, but it is difficult to rule out no effect for higher values as well. This means that consumers who buy larger baskets of products spend incrementally less on all-other goods when there is a sale on, eventually becoming negligible or zero. We see similar, but stronger, results when we look at sundries specifically. Despite sundries on average making up on average 12.3% of all-other spending, we can see that the coefficient on incremental sale spending is 66% as large as the equivalent result for total spending. In fact, it is statistically indistinguishable at most significance levels. This indicates that the incremental spending we see is being driven primarily by the high-margin sundries departments. We can also see from columns (5) and (6) that the coefficient is much more stable, indicating that the main effect is captured by the sundries departments, and not unobserved factors or dry goods. The same U-shaped trade-off between number of categories and the impact of sales is also evident for sundries. This is depicted in the second panel of Figure 2.6; the crossing point, and which sales no longer add spending occurs somewhere between 4 and 6 departments, indicating a much more robust effect; only consumer buying the most diverse bundles of products are likely to not spend more when buying a sale product. This is potentially because they were already planning to purchase a robust number of items when arriving at the store, and the sale does not affect them for this reason. One other possibility is that multiple goods could be on sale at the same time; although I do not have sale variables for the other products, I am able to compare average prices. Looking at the spending-weighted average prices of sundry goods across the consumer who buy the perishable one sale versus those that do not, we find that the price is $5.29 at regular prices versus $5.06 on sale, a difference of less than a quarter, and statistically not significant. This implies that the spending variations for sundries is not driven by own-price sales, since the implied elasticity would be implausibly large.

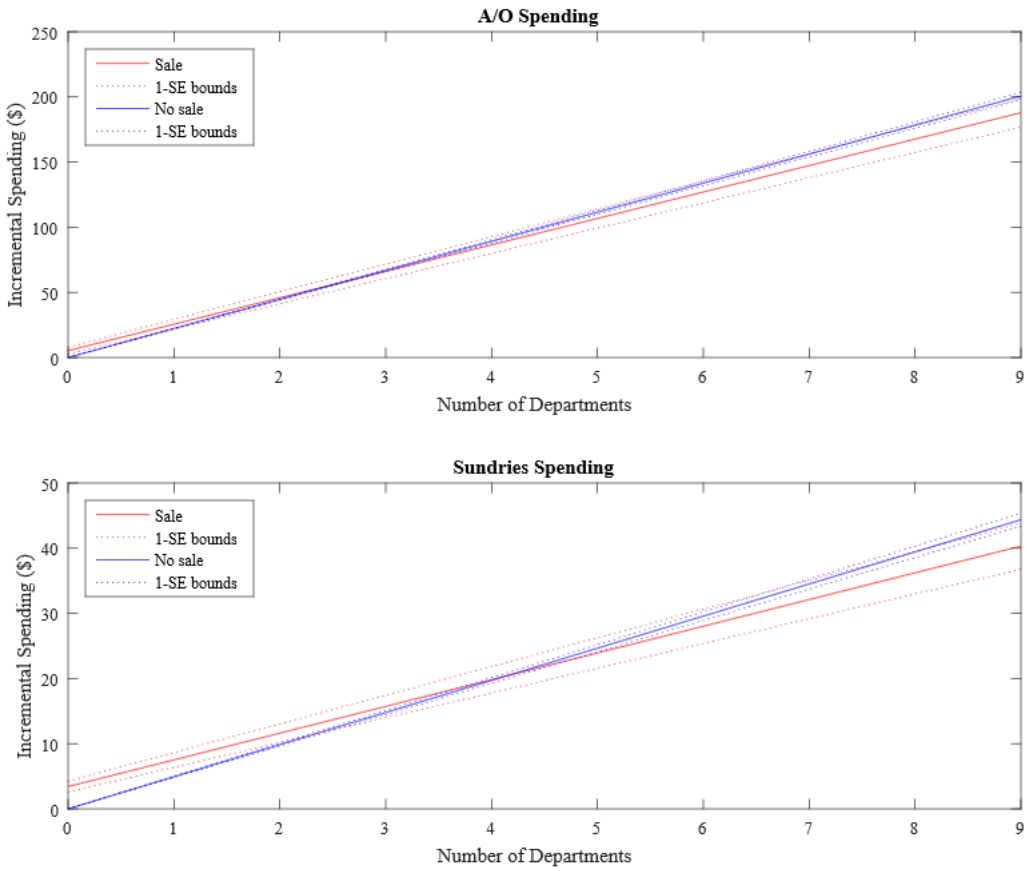


Figure 2.6: Plot of crossing points in Table 2.8

VARIABLES	(1) Total spending	(2) Total spending	(3) Total spending	(4) Sundries	(5) Sundries	(6) Sundries
Num of Departments	22.27***	11.64***	22.40***	5.29***	4.09***	5.35***
	(0.32)	(0.30)	(0.31)	(0.12)	(0.12)	(0.13)
Sale × num departments	-2.04***	-1.77***	-1.92***	-0.90***	-0.88***	-0.89***
	(0.62)	(0.49)	(0.62)	(0.21)	(0.20)	(0.22)
Sale indicator	4.46†	5.70***	3.17	3.87***	4.04***	3.29***
	(2.77)	(2.16)	(2.66)	(0.98)	(0.94)	(0.97)
Single item shopper				6.59***	6.48***	6.44***
				(0.42)	(0.40)	(0.39)
Sale × SI shopper				-1.19*	-1.29**	-1.42**
				(0.68)	(0.64)	(0.67)
Average A/O spending		0.72***			0.08***	
		(0.01)			(0.01)	
Observations	26,862	26,862	26,862	26,862	26,862	26,862
R-squared	0.54	0.74	0.51	0.34	0.38	0.31
Demographic Controls	X	X	X	X	X	X
UPC FE	X	X		X	X	
Time FE	X	X	X	X	X	X

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10, † p<0.20
Sampling weights used, outliers excluded

Table 2.9: Robustness Results I

In order to look more closely at the income and substitution effects of sales, I also perform the same regressions as in Table 2.8, but this time for all spending. These are reported in columns (1)-(3) of Table 2.9. In a standard, two period consumer choice model, we would expect a sale to change the quantity demanded of both products; however, the aggregate spending would remain the same. We can see this is not the case; for consumers with with smaller basket sizes they actually spend more *in total* across all goods. This is eventually reversed, but only for who buy relatively more categories of goods (4+). In either case, there is clearly something more going on than a simple substitution and income effect. One issue might be these results might be highly leveraged by consumers buying a very small number of products; particularly, the targeted or sale products. To test this, I create an indicator variable `Single Item Shopper` which represents whether or not a consumer spends 50% or more of their spending on the targeted product. These consumers go into the store and primarily purchase just the targeted product. These results are presented in columns (4)-(6) of Table 2.9, which replicated the `Sundries` results from the main specification, but with this variable (and interactions) added. Interestingly, we can see that single item shoppers spend substantially more on sundries when they shop. This is likely because we also control for the number of departments being shopped in, and thus it is likely they would buy only one or two other groups, which (apparently) happen to be sundries. The main results are not changed by this specification; the magnitude and direction of the sale coefficient is the same, as is the interaction with the number of departments shopped in. One factor, however, is that single item shoppers tend to buy less of everything else when buying on sale; this makes sense, since these are likely to be the most frugal of the consumers. Interestingly, the magnitude of this effect is quite small, indicating it is not a major consideration; for instance, the sum of the coefficients on this interaction and the other related variables is positive, indicating they are still a desirable group since they spend more even during a sale period.

As an additional test of the robustness of this result, I also look at the substitutability of the target products. One issue which complicates the analysis is the ability of certain meat products to be frozen, then consumed later. Frozen meat products take more time to prepare than fresh, and also suffer from quality deterioration. In particular, even with ideal freezing and thawing procedures to minimize quality loss, meat will lose moisture during the thawing process. The protein structure of the meat is also affected, as is the colour, pH balance, tenderness and microbial balance Leygonie et al. (2012). These are exacerbated by long periods in the freezer or non-ideal freezing or thawing methods. Most

	(1)		(2)	
VARIABLES	Frozen goods	Sundries	Frozen goods	Sundries
Average A/O spending			0.09***	0.08***
			(0.00)	(0.00)
Sale indicator	1.57***	4.09***	1.58***	4.10***
	(0.47)	(0.60)	(0.44)	(0.59)
Num of Departments	2.97***	5.05***	1.53***	3.81***
	(0.04)	(0.05)	(0.05)	(0.06)
Sale × num departments	-0.38***	-0.98***	-0.31***	-0.92***
	(0.09)	(0.11)	(0.08)	(0.11)
Observations	28,381	28,381	28,381	28,381
R-squared	0.27	0.33	0.36	0.36
Demographic Controls	X	X	X	X
UPC FE	X	X	X	X
Time FE	X	X	X	X

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10, † p<0.20
Analytical weights used, outliers excluded, education and occupation controls excluded

Table 2.10: Frozen goods

chefs would agree: frozen meat is generally inferior to fresh. Nonetheless, the fact that individuals are still able to store fresh meat with a quality decline would result in a sort of inventory-like behaviour, which would explain the patterns found in the data. The consumer level data allows us to test this too, albeit only indirectly. Specifically, since in order to preserve meat[27] it must be frozen there is a close substitutability between the fresh (to be frozen) meat and extant frozen meat products and other substitutes. Thus would imply that frozen products should have a much greater degree of substitutability than other products with respect to frozen products. Specifically, we would expect the sales of frozen products to be much smaller than those of the sundries. I test this using a general linear framework, repeating the benchmark regression with two dependent variables: sundries, and spending on frozen products. These results are reported in Table 2.10; in specification (1), I use the baseline model, while in (2) I include average A/O spending. The results indicate that we have similar effects for frozen foods as we do for sundries, albeit at a different scale. In particular, we note that the coefficients move in the same direction, which indicates that substitution between frozen products and meat products on sale is not a major consideration.

These results together paint a very interesting picture of how grocery stores and consumers are behaving with respect to sales. If we believe supermarkets are behaving rationally, the results indicate that they use sales in order to target profitable (or at least, differentially profitable) sub-segments of the consumer market. These correspond to the consumers who are purchasing different sized bundles. About 50% of the population lies in a region where a sale induces them to buy more overall; in a general sense, this agrees with the idea of loss leadership. Grocery stores place goods at a lower price, so that consumers will come into the store and "buy more" of other goods with a higher profit margin. This story, while appealing at first, requires careful consideration in light of the results we see. Specifically, there appears to be a trade-off between inducing some consumers to buy more while other consumers essentially get a "free lunch" and wind up spending less. Accordingly, in order to explain this, I develop a model of loss leadership, which is elaborated in section A.3.

[27]For most households; we omit the possibility that consumers could be smoking or curing grocery store product in any substantial amount

2.7 Conclusion

This chapter makes three main contributions. First, it shows explicitly that regular, periodic sales are an important part of the story surrounding the pricing of perishable products. This is important because it means that when we talk about sales, or model them, our explanations need to either take into account the special nature of these kinds of products or acknowledge that a "one size fits all" approach to explaining sales is probably not going to work. This also requires us to carefully consider what kinds of products our models apply to, and why.

Second, I illustrate how to do this by developing a model which explains the periodic nature of perishable sales. As the discussion in Section 5 shows, I explain how this fits into the set of possible models and frameworks. I further show that this model is plausible given then data by showing that the central causal connection necessary is supported by consumer choice data. This demonstrates the third contribution of this paper: showing how the connection of retail and consumer choice data can help evaluate different types of theoretical models. This necessitates the development of more flexible tools to determine when sales occur, and focuses attention on the economic content of these kinds of heuristic decisions empirical researchers make. Simple heuristics may provide straightforward answers, but they often neglect or overlook potentially important aspects of the data.

This work also leaves room for further development in several areas. First, as developed in Chapter 4 the classification method can be extended to an explicit structural model. Preliminary empirical work on this shows similar results to that reported here, but the structural relationships are potentially more detailed and reflect an alternative analytical framework which would be interesting to compare with the existing literature. Second, the method used in this chapter can be readily applied to other products, particularly sundries. This would allows the correlation of sales on different products to be explicitly analysed, resulting in a more holistic picture of how stores use sales as a pricing strategy. Finally, the model in question leaves substantial room for refinement or sophistication. In particular, profitability and the role of chain-brand competition is not addressed in the existing framework but is known to be an important feature of store decision making. These points illustrate that, despite several decades of attention to the questions and problems sales pose, new techniques and data continue to offer opportunities to learn and improve our understanding of this important area.

Chapter 3

Large Contributions and Crowdfunding Success

3.1 Overview

> In recent years, small businesses and startup companies have struggled to raise capital. The traditional methods of raising capital have become increasingly out of reach for many startups and small businesses. [...] Low-dollar investments from ordinary Americans may help fill the void, providing a new avenue of funding to the small businesses that are the engine of job creation. [...] The promise of crowdfunding is that investments in small amounts, made through transparent online forums, can allow the "wisdom of the crowd" to provide funding for small, innovative companies. It allows ordinary Americans to get in on the ground floor of the next big idea. It is American entrepreneurism at its best, which is why it has the support of the President and many in the business community. (Senator Jeff Merkley, Congressional Record, 112th Congress, December 8th, 2011)

Crowdfunding is a new form of financing by which small, unsophisticated individuals or businesses with an idea for a project or event can "crowd-source" the financing for their project and bring it to market. It has been a dramatic success, funding hundreds of thousands of projects and raising billions of dollars (Cantillon (2014)). The combination of entrepreneurship, consumer interest, and social media which drives crowdfunding has drawn the attention of the media, business owners, and investment experts. While the reception has been mixed (Agrawal et al. (2013)) among experts, there is little doubt that crowdfunding is emerging as a legitimate and commonplace method of financing. As the Wall Street Journal put it:

> Crowdfunding has the potential to revolutionize the financing of small business,

> transforming millions of users of social media such as Facebook into overnight venture capitalists, and giving life to valuable business ideas that might otherwise go unfunded. (Gubler (2013))

It has drawn specific attention from the United States government, who have provided it with legal grounding and protection, clarifying its role as an alternative to venture capital or traditional (bank-based) financing. Nonetheless, we are still just beginning understand the fundamental forces behind crowdfunding. There remain many important questions which we are only starting to answer: why do some projects succeed and others fail? How do you design a project for success? What kinds of things make one project a good fit for crowdfunding, and another a bad fit?

In the rhetoric surrounding crowdfunding, this method of funding takes on something of a Gulliver versus the Lilliputians flavour: many small individuals, motivated by their support for an idea or project, combine their efforts to help it succeed despite it being an insurmountable obstacle to each person individually. This is compounded by the sense that crowdfunding is an alternative to the "big guys:" banks, developers, or studios which are unwilling to support the project. This gives crowdfunding a very populist, individualistic character: the elite, monied interests are out of touch with what the common person cares about, so the little guys will pool their effort and do it anyways.

However, how true is this description of crowdfunding? Are many small contributors really the driving force behind project success? In this chapter, I look at a specific aspect of this question: how important is it to attract large contributors to a project? Does this determine success in a meaningful way? What features of projects lead to large contributions? More importantly, can I say anything causal about the connection between project success and large contributions? In order to do this, I turn to a novel panel dataset on crowdfunding projects spanning three years (2012-2014). This dataset comes from the largest crowdfunding platform (Kickstarter), and allows me to look behind the cross-sectional success or failure of a project to examine the different sizes of contributions.

I find that large contributions are driven by a unique economic process, with an understandable motivation. Large contributors appear to care about helping projects succeed, providing more when projects are in need, and when their contributions are likely to be impactful. This is especially apparent on the "day of success": the pivotal day on which projects succeed in which large contributions frequently push projects over their goal threshold, and into success. This behaviour is rationalizable using a consumer choice

model of crowdfunding, in which individuals care about the benefit of a project, but also understand their impact on the success of a project. Large contributions can be regarded as groups of individuals affiliated with the project, or fans of the projects: both groups will typically will want to donate when a projects needs a "final push" to get it over the goal, conditional on the goal, their willingness to pay, and the state of the project. This agrees with a strand of analysis in the literature on how crowdfunding often serves as a way of formalizing previously informal funding networks among family members or friends (Agrawal et al. (2013)). Which characterization is at play also has important consequences for policy makers and regulators.

I find that large contributions appear to be important for success: using a linear probability model on the data cross-sectionally, I find that projects which can attract at least one large contribution are 30% more likely to succeed, after controlling for a variety of time-effects and covariates. This is robust to a number of specifications, including discrete choice models. However, there are several sources of endogenity, not the least of which is the fact that the causality discussed above can be reversed; the success event is endogenous to large contributions. In order to address this, and assess the causal nature of the relationship, I use an instrumental variables approach. My instrument exploits the fact that the underlying data is not cross-sectional, but an unbalanced panel: I choose a set of secular and religious holidays which are associated with giving, purchasing or income shocks (i.e. Christmas, Black Friday, etc.). By exploiting the fact that project start times and lengths (called "tenures") vary, and controlling for time fixed effects, I can use the number of elapsed holidays to instrument for the presence of large contributions. The key identification assumption is that the tenure decision is not sufficiently sophisticated so to capture these dates at a level more granular than the month; given the fact that crowdfunding is specifically designed and used by unsophisticated individuals, I believe this is reasonable. I try to verify this assumption in several ways, showing that holidays are not proportionally more likely to be included in a project's tenure. This IV approach shows that the underlying results are in fact an underestimate: the IV results are 10-30% larger than the baseline multiple regression framework used. However, because of local average treatment effects, based on the instrument used, it is difficult to say with certainty that this is typical of all projects, and not specifically those affected by the instrument. Some robustness checks demonstrate that this might be the case, but certain explanations cannot be ruled out.

I also dispel a mechanical interpretation of the results: one might expect that since

large contributions are larger than average, they might be helpful simply due to their size alone. In order to examine this, I perform a counterfactual in which we pretend a large contribution occur as simply more individual small contributions. I find that only about 1/3 to 1/5 of the effect of a large contribution is driven by size alone; the majority of the impact is caused by other factors, such as timing or learning by agents. In other words, I find that large contributions do not simply provide more money, but they provide more money when it is most instrumental for a project's success in a number of ways. In addition, I also perform several naive counterfactuals, demonstrating that the IV estimates likely reflect this "instrumentality" of large contributions better than the ordinary least squares estimation.

This chapter builds on a new, but growing, literature on the determinants of crowdfunding success. The problem of predicting what features of projects determine success, and consequently how best to structure a project if you are a project-owner has already been the subject of some attention (see Xu et al. (2014); Zvilichovsky et al. (2013); Belleflamme et al. (2013); Lambert and Schwienbacher (2010); Mollick and Kuppuswamy (2014); Mollick (2014) for a few examples). However, because of the nature of most of this literature (cross-sectional data on projects) the role of large contributions has been difficult to assess. The use of a panel dataset in this study allows me to consider this aspect of the project, and additionally to try to correct for endogeneity using an instrumental variables approach. I find that the standard crowdfunding narrative does not paint a complete picture: while the "little guys" might be important, people with relatively more money than the others are also quite important.

These findings are influential in two ways: first, for policy-makers concerned with the administration and legislative structure of crowdfunding, it is important that contribution size is carefully considered and managed. For example, in the debate leading up the the amendments of the 2012 JOBS Act concerning equity crowdfunding, there was substantial discussion regarding contribution limits and total caps; this research indicates that these concerns were appropriate, and should probably be considered for other (non-equity) forms of crowdfunding. Second, this paper informs future modeling of crowdfunding. There is a small, but growing, literature which seeks to model the individual choice decisions involved in crowdfunding (see Marwell (2016); Chang (2016) for examples); however, most ignore the "intensive" margin of contribution size, instead focusing on the number of contributors and assuming the amount they give is fixed. This research indicates that this is a serious omission and conclusions drawn from such structural models do not

reflect a substantial fraction of the actual process governing crowdfunding research.

The remainder of this chapter is structured as follows. In Section 3.2, I give some background on crowdfunding and large contributions, specifically in the context of this paper and with an eye to establish how this new form of financing "archetypically" works. In Section 3.3, I explain the data source and show some stylized facts about large contributions and crowdfunding; these motivate the analysis explained above. In Section 3.4, I provide a model to make predictions and explain the results, and then describe modelling approach by presenting a simple consumer choice model of crowdfunding. I then, in Sections 3.5 and 3.6, outline my empirical specifications and present my results; section 3.7 discusses these results, and presents some robustness checks. Finally, Section 3.8 concludes, while Appendix B provides additional detail.

3.2 Background

3.2.1 What is Crowdfunding?

Crowdfunding is a new form of financing for projects and businesses, which reached a wide audience beginning in 2010-2012. Since then, it has grown into a billion dollar industry, and become a major source of funding for individuals and small-to-medium sized businesses. This is particularly the case for those in creative or artistic industries like music, comic books, art, or video game design. The fundamental characteristic of crowdfunding is that of the "crowd": through the use of an internet-based platform, many individuals are solicited for small contributions towards a project. Generally, these are paired with a reward which is also usually conditional on certain fund-raising thresholds beings met. Consider the following example, which was the motivation behind the largest crowdfunding platform, Kickstarter:

> Imagine you are a promoter trying to put on a concert. You have an agreement with the musicians, and you have arranged with the venue for a location: the only issues that remains are the financing and the audience. Who pays for the down-payment for all these individuals? If you put up the money yourself, and the audience is small, you will lose money. Worse, you may simply not have the funding to go forward in the first place, even if you are are sure the audience will materialize. The concert does not go ahead, in spite of the fact that if everyone could have cooperated it would have been a success.

This was the situation the founders of Kickstarter saw themselves in: valuable projects they wanted to pursue were failing to materialize because of a combination of demand uncertainty, the speculative nature of the project, and simple credit constraints. Their insight was to try to reach out to the potential consumers to provide funding for the project:

> I thought: "What if people could go to a site and pledge to buy tickets for a show? And if enough money was pledged they would be charged and the show would happen. If not, it wouldn't."(Kickstarter (2014))

This kind of arrangement, in which project-owners tap the unrealized demand or underlying support for an idea or project to secure start-up funding, is the basic innovation which has driven the success of crowdfunding. This feature also differentiates it from the most common alternatives open to individuals seeking funding:

- Unlike traditional debt funding (from a bank or large investor) crowdfunding raises money through the prospective delivery of a product in the future, not a promise to repay a loan or give up collateral. While in most cases the product is a concrete good (like a consumer durable or consumable), crowdfunding is also used to fund projects which (at best) produce only an artistic or public good, which banks are typically unwilling to support.

- Unlike venture capital funding, crowdfunding raises money from a large number of small, non-expert investors interested only in the final project being created. In contrast, venture capitalists provide most or all of the funding for a new start-up, as well as expert advice, in exchange for an equity or debt position in the business. They generally have little actual interest in the product being created beyond its marketability. In most jurisdictions, crowdfunding backers cannot legally be given a financial stake in the company.

- Unlike a traditional pre-sale, the delivery of any reward is not guaranteed, instead it is explicitly contingent on meeting a pre-arranged funding threshold. While some pre-orders function in a similar way (e.g. John Deer tractors), the explicit and central nature of the contingent funding threshold is not present.

With this said, there are several ways in which the details of crowdfunding can be relaxed. The most prominent examples have to do with the reward and the contingent nature of the goal: crowdfunding traditionally deals with *contributions* not *donations*, in the

sense that there is a conditional reward offered. However, there are several large platforms sometimes referred to as "crowdfunding" which deal exclusively in straightforward donations, such as GoFundMe. In these sites, the goal is not contingent, and there is no reward promised. Individuals simply use these websites to donate money to causes they feel are worthy of support. This is not crowdfunding as this paper and most of the literature defines it. It lacks any new features which make it stand out from simply internet-based charitable donations, which have existed since the early 1990's. If anything, the crowdfunding label is likely to confuse visitors to these websites, since the goals are non-binding, and are just "wish lists" for how much a cause seeks to raise.[28].

However, this is complicated by the fact that some websites include both charitable and crowdfunding options for a single project. For example, the website IndieGoGo offers "flexible" and "fixed" funding options for projects. Fixed funding is traditional crowdfunding, as described above, with a contingent goal and reward. However, flexible funding is non-contingent, much like GoFundMe. The pledges made are simply donations, and the reward is open to re-negotiation if the overall goal is not met. The lines are further blurred by the fact that many "crowdfunding" projects, especially highly artistic ones, may have very abstract rewards. Is a conceptual art piece, with no actual event, carried out via flexible funding actually different than simply a charitable cause? The difference is not immediately obvious. However, even in this situation, the way these projects are presented to consumers is different. Consider Figure 3.1, which displays several projects featured on the front page of IndieGoGo and GoFundMe during April 2017. A qualitative study of the way these websites, and projects, present themselves demonstrates clear differences. Even flexibly funded projects emphasize the entrepreneurial nature of the undertaking, focusing on "helping a project to succeed" or "achieving our dream" if any charitable aspect is mentioned. On the other hand, explicitly charitable projects focus directly on the need being addressed by the donations: funding for treating a disease, covering unexpected expenses, or delivering aid to people in need. While the line may be blurred from a technical point of view, there are real and tangible differences between crowdfunding and charitable donations which make them different fields of interest.

[28]To put this another way, if these websites are crowdfunding, then most churches have been crowdfunding to fix their leaky roofs for hundreds of years, and the term has lost any useful meaning

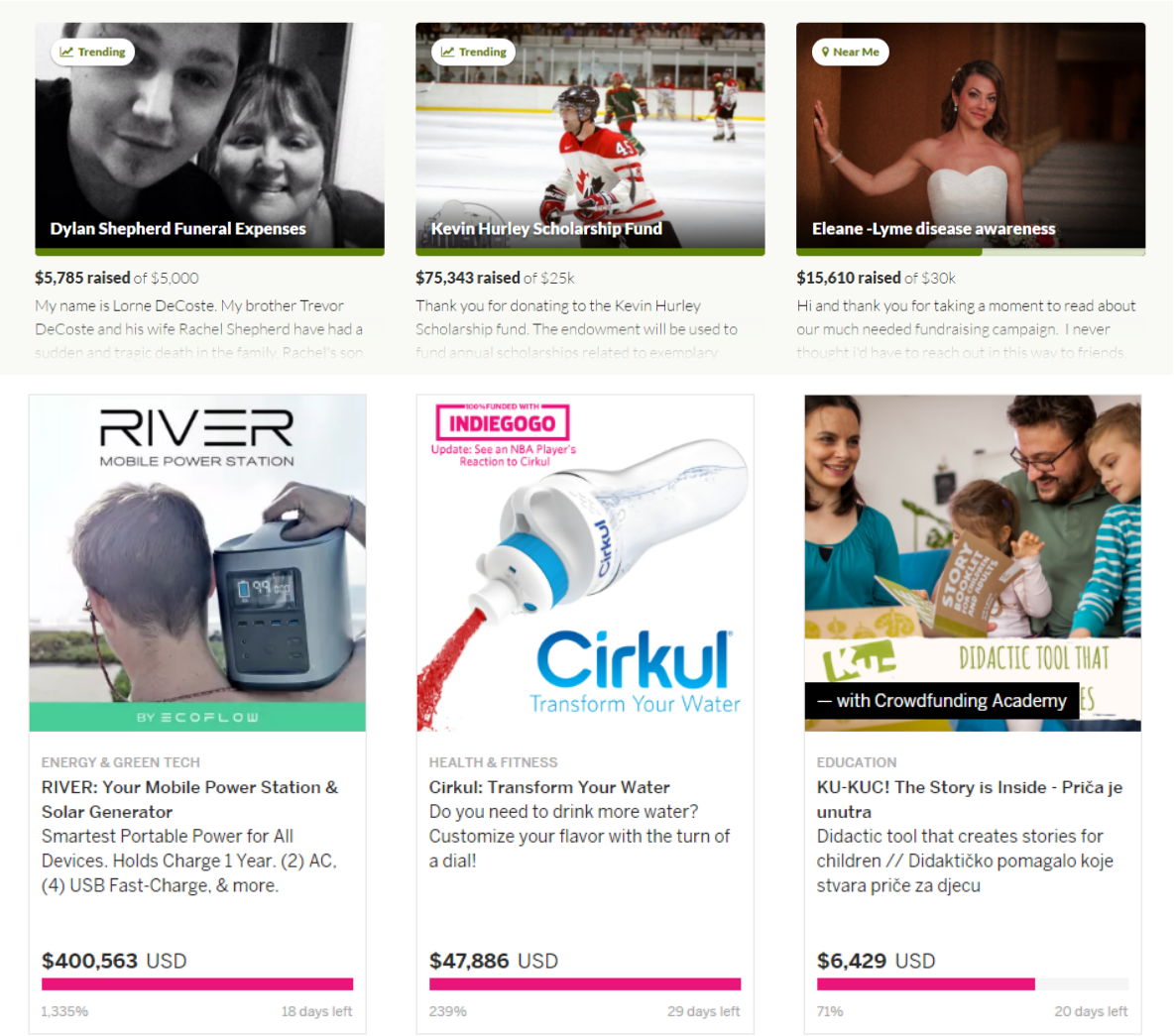


Figure 3.1: Comparison of Featured Projects on GoFundMe (top panel) and IndieGoGo (bottom panel) as of April 24, 2017

Contemporary crowdfunding is also part of a continuum of similar fundraising mechanisms. For example, agricultural farm shares often share a similar arrangement. In these, farmers sell shares in the upcoming harvest to individual consumers, conditional on reaching sufficient funding to plant the fields. When harvest time comes, the consumers are given the produce grown in the fields. This helps farmers with the start-up costs, such as purchasing seeds and fertilizer, and preparing the fields during the winter and spring. In the charitable donation literature, the study of so-called "kickers," or fundraising boosts at given thresholds has been studied extensively. A closer example from Canadian history, featured prominently in Leacock (2016), is a "whirlwind campaign." In this type of campaign, individuals pledge money to a cause, conditional on reaching certain fundraising goals. Although this is purely charitable, the conditional nature of the pledges is central to the mechanism (and Leacock's story). When looked at with the right perspective, many historical and traditional forms of fundraising have a crowdfunding-like aspect to them. Crowdfunding, as carried out by Kickstarter and discussed in the literature, seems to have been a relatively minor innovation. Why has it suddenly become so successful?

The answer lies in two intersecting forces: (1) the rise of social media and the interactive world wide web (Web 2.0) and (2) the technical infrastructure to guarantee and process transactions easily. Both of these innovations occur through the creation of crowdfunding platforms like Kickstarter or IndieGoGo. Large numbers of disaggregated consumers is central to the way crowdfunding operates: if there is no crowd, there can be no crowdfunding. Crowdfunding platforms provide this by allowing for a single, central repository for projects. Rather than having an individual site for each project, interested consumer can peruse thousands of projects at once. This also compels project owners to standardize what they offer, how they explain their project, and how they market it to consumers. This makes comparison "shopping" or project discovery much easier for consumers. This would be impossible without the interactive nature of contemporary websites. Interactive, user-generated content is fundamental to crowdfunding. Platforms do not manage or create project pages. They merely provide the technical framework for project-owners to do the work themselves, much like YouTube, Blogger, or other user-content sites. The social nature of the modern web also facilitates this. Rather than relying on the platforms themselves to promote their project, or relying on random discovery, project-owners can promote their ideas and products using social networks. The often tightly-knit communities surrounding niche products can be immediately leveraged through forums like Reddit or Twitter, sharing projects consumers might be interested in. Friends can quickly share and promote

interesting projects through sites like Facebook, linking directly to the crowdfunding platform. All of these features allow project owners to amplify the reach of their message in a way that was much more difficult, or impossible, to do before these technologies became wide-spread.

Most crowdfunding sites, especially Kickstarter, weave social media into every part of the crowdfunding process. Both backers and project owners create and maintain personal profiles (much like on Facebook or Google Plus), complete with photographs and personal vignettes. The number of projects supported, or created, is prominently displayed and discussion between individuals is encouraged through both comment features and private messaging. More traditional forms of social media are closely integrated, allowing individuals to promote projects they are interested seamlessly through Twitter, Facebook, and many other social media platforms. The modern crowdfunding site seeks to incubate an environment in which there is an easy back-and-forth between investor and investee, between the business network and social network. This provides both a critical communal aspect to crowdfunding but also furthers the "marketplace" notion which drives the process: when every supporter can reach out to their network with a single click to promote a project, the number of people exposed becomes larger and larger. This "viral" type of exposure is often the driving force behind very large projects which end up attracting an audience of backers far larger than a typical project could hope to be exposed to.

The second force, the existence of robust and sophisticated transaction processing, matters because it allows crowdfunding platforms to exist. Platforms are beneficial from a project-owner point of view, because they provide a single location to host projects. This leads to increased visibility and discovery. They also are designed to be highly accessible and easy-to-use: no web development skills are necessary, making hosting a project as easy as writing a news article or making a post on Facebook. However, platforms also serve as a guarantor in several ways. Most importantly, they process payments and guarantee the contigent nature of the crowdfunding contract. When an individual pledges money to a project, it is the platform which ensures they are charged if and only if the project reaches its goal. This effectively makes them a single digital storefront for a large number of individuals who, alone, would have difficulty being trusted with consumer's money or credit information. Platforms also provide security by enforcing truthfulness, setting standards for disclosure, and being the first line of recourse if consumers feel misled by a project. For example, Kickstarter mandates that projects disclose potential risks to consumers, and prevents them from using misleading images, such as photorealistic render of a non-existent

final product. Kickstarter also restricts items which do not meet community standards, could be dangerous, or could be illegal to own or produce, such as weapons, drugs or drug paraphernalia, or live animals. Projects which violate these rules are subject to shutdown by the platform, forfeiting any funds raised.

This allows platforms to provide security and guarantees to individuals taking part in crowdfunding. Individual projects may have problems or challenges, but there is a clear minimal standard being met and enforced. This lets consumers focus on evaluating the project on offer, rather than worrying about whether their payment system is secure or if there's some fine print which could lead to them being scammed or otherwise misled. This role is clearly front of mind for platform operators, as many of the rules have evolved over time in response to controversies surrounding certain projects. For instance, Kickstarter's rule disallowing photorealistic renders of products was likely a response to doubts surrounding the Ouya, an Android-based video game console, and the Pebble, a smartwatch, in which digital renders were easy to mistake for prototypes. Platforms make their money by taking a small percentage (usually in the range of 5%) of the total amount raised. It is in their best interests to build and maintain a reputation for honesty by setting rules and enforcing them. This makes the platform more appealing to consumers, which in turn makes it more appealing for project-owners.

These innovations, coupled with a unique fundraising structure, have been central to the success of crowdfunding. While still evolving, it is clear that crowdfunding plays an important role in the financing of many types of projects, particularly when traditional forms of financing may be difficult to come by. It is also clear that while crowdfunding is part of a rich and lengthy history of similar kinds of financing, it is also a markedly contemporary enterprise, with much of its success due to the modern internet and social media landscape. This innately digital nature is also useful to researchers, since it also means that data is much easier to collect and study. The centralized nature of crowdfunding platforms also makes the underlying structure much more consistent, and helps study the topic more broadly than taking a single project-by-project perspective.

3.2.2 Consumer Crowdfunding on Kickstarter

Given the wide, and often flexible, forms crowdfunding can take, it is useful to explicitly summarize crowdfunding as it is studied in this paper. I call this "consumer crowdfunding" since it deals primarily with the financing of consumer goods. It is also constitutes an

archetypical description of crowdfunding in general, since it describes crowdfunding as it was originally carried out on the first major crowdfunding platform Kickstarter, circa 2010.[29] As discussed in Section 3.2.1, this is not a complete description of all crowdfunding projects or platforms, but it is accurate for the data used in this project, and serves as a guide to the generic structure of other platforms.

As discussed earlier, consumer crowdfunding is associated with two main features (1) a contingent goal and (2) a specific, tangible deliverable. To illustrate the process in detail, consider the following example, motivated by Kickstarter's founding story: you (the *project-owner*) are a promoter who wants to put on an exciting new concert. However, the financing is expensive, and funds are limited; worse, if you put up your own funding and people don't attend, you stand to lose a great deal of money. Accordingly, you decide to use crowdfunding; to do this, you:

1. Put together a *pitch* which explains who you are (a successful concert promoter), what you're going to do (put on a concert), and how it's going to work (we've arranged for this venue, these bands, and these dates).

2. Set a funding *goal* and *deadline* ($6,000 by July 31st) which is the amount your crowdfunding campaign seeks to raise by the time chosen.

3. Describe the *rewards* supporters will receive if they support the project and it succeeds (a ticket to the concert).

4. Associate each reward with a *price* which must be pledged in order to receive the reward (if you pledge $20 you receive a ticket to the concert).

5. Post the pitch, along with the other information on a crowdfunding platform (Kickstarter).

After your pitch is posted on the website and becomes visible, individuals can view it on the crowdfunding website and decide whether or not they wish to support your campaign. An individual who decides to support the project is called a *backer*. If enough money is pledged to reach the goal by the deadline, the project is a *success* and the money pledged is paid to

[29] The degree of influence Kickstarter has had on crowdfunding as the first major player is difficult to understate. As an illustrative point, crowdfunding campaigns are sometimes referred to as "Kickstarters" regardless of where they are being carried out; in a similar manner like Kleenex has so defined tissue paper, where the brand becomes a byword for object itself.

the project owner by the backers, while the project owner is then obligated to deliver the rewards promised. If the goal is not reached by the deadline, the project is a *failure* and no money is exchanged, but also the rewards are not produced; the contract dissolves. The progress of a project towards its goal, the number of backers, and comments/updates from backers or the owners are all publicly visible on the crowdfunding platform for potential backers to see.

This contract, if successful, forms a legally binding agreement between the project-owner and the backers; the project-owner must make a good-faith effort to deliver what has been promised, and similarly cannot misrepresent their skills or the deliverables promised. The legalities surrounding crowdfunding in this respect were formalized in 2012; provisions of the 2012 U.S. JOBS act were specifically enacted to enable forms of crowdfunding, while clarifying that it was not designed for use with investment funds or to supplant the role of traditional venture capital. In the best study to date, Mollick and Kuppuswamy (2014) found that while a large majority of projects eventually delivered on their promises to backers, there were typically some delays or changes to the final products. Additionally, while a large majority of backers were ultimately satisfied by the outcome of the project, a minority were unhappy and a small minority of projects failed to ever produce a usable good or refund the backers' money.

3.2.3 Literature Review

The literature studying crowdfunding is still in its infancy, and is complicated by the ongoing discussion over what, exactly constitutes "crowdfunding" as discussed in Section 3.2.1. Two basic programs have emerged however: in the first, researchers try to study a particular crowdfunding market, looking in detail at what makes projects successful and the consequences this has for consumers and project-owners. This program also tries to place crowdfunding, as a funding mechanism, within the family of other alternatives. This highlights the role it plays as a source of funds, and why it might be attractive to certain project-owners. The second program capitalizes on the different structures different markets have, using them to try to analyse consumer behaviour in crowdfunding markets. With this said, there remain many crowdfunding platforms and styles which remain largely unstudied.

Within the first set of studies, understanding the dynamics of backing projects is important. Cross-sectional study of project is highly limiting, because crowdfunding is an explicitly dynamic process. It is also the case that the dynamics of crowdfunding shed light

directly on the valuations of individual backers, which can then be connected to the success of the project. The question of to what extent pro-social or altruistic motivations play in crowdfunding is very important to understanding this market. For example, cross-sectional studies have looked at what makes a project successful. There is evidence of complex dynamics at work for backers; for example Kuppuswamy and Bayus (2013) and Belleflamme et al. (2013), motivated by evidence from their data, and qualitative surveys, suggest that there is a behavioural social benefit to many projects. Indeed, many theoretical papers explicitly require an social benefit to crowdfunding in order to get the qualitative predictions they make to agree with the data. These types models are the subject of some of the best empirical papers on the topic (see also Rao et al. (2014) for example). This is often couched in a "common value" or "social value" framework; in terms of how crowdfunding platforms present themselves, and much of the early discussion on the subject, this remains an influential views of how this form of financing operates.

However, another important aspect of crowdfunding concerns the role of information, such as demand uncertainty or asymmetric information. These can arise on both the consumer and project-owner side of the market. One prominent example has been the suggestion of social learning and herding within a crowdfunding project. As individuals back a project, their action is publicly observed by other individuals. When viewed in this way, crowdfunding becomes a very complicated information setting for backers. The role of the goal becomes an event in which individuals can condition their decision making. For example, imagine a project which requires a dozen backers to reach its is either quite valuable or worthless (if funded), and suppose a single potential backer has some information that indicates the project is worthless. However, suppose they know (also) that their information is of relatively poor quality, and following them, every other backer interested has very good information (say, perfect). Should they back the project? Of course! The structure of the crowdfunding contract is such that they will only actually pay the money when the goal is crossed, which is only the case if the other individuals know that reached project is quite valuable. Thus, they condition the expected value of the project on the event that it is funded: which in this case is equivalent to conditioning on the knowledge that the project is of high quality.

This threshold-conditioning effect is closest, in the existing literature to papers which study sequential voting: for instance Ali and Kartik (2006); Callander (2007). Indeed, with the right perspective, we can imagine that crowdfunding is sort of like a vote: a decision to back is a "vote" in favour of the project being of high quality. However, unlike in sequential

voting environments, the decisions matters directly and not just instrumentally through outcome. This herding-type (or bandwagon) concerns inform much of the skepticism about crowdfunding in the light of asymmetric information. When coupled with the viral nature of many crowdfunding campaigns, the ability for private information to be swamped by the crowd appears to be a relevant concern. One aspect that has largely avoided discussion, however, is the conditioning point raised: herding, in the sense of Banerjee (1992) relies on the existence of a history of decisions. In crowdfunding, the structure of the contract leads to conditional information: a *prospective* herd. In the example given, for instance, the individual herds despite being the first person to make a decision.

This can lead to a piling-on effect, which is difficult to study empirically but has been discussed in several papers (see Agrawal et al. (2013); Kuppuswamy and Bayus (2013)) who find some suggestive evidence. Kuppuswamy and Bayus (2013) has found some evidence of social learning and herding in previous crowdfunding data. This is usually observed as accelerating co-movements of backing decisions, especially after the threshold has been crossed. In general, though, the previous evidence on how backers behave is mixed; this is not their dominant explanation for the empirical patterns in the data, which relies on the social model of valuation described above. There are (certainly) time fixed effects: the arrival rate of individuals at a project changes over the lifespan of the project, which make it difficult to disentangle the different dynamic predictions of a model. For example, a U-shaped pattern of supporters could potentially be a function of the simple fact that new (or ending) projects attract more attention but it can also speak to some of the social aspects which underlie crowdfunding (as in Hekman and Brussee (2013)).

On the other side of the market, Cimon (2017) studies the role uncertainty plays in the decision to undertake crowdfunding. Theoretically, he shows that crowdfunding is particularly valuable for individual investors who face demand uncertainty. This finding is in line with what qualitative surveys such as Mollick and Kuppuswamy (2014) have found: many individuals use crowdfunding not just as a fundraising tool, but also to help them assess the potential market for their product. It can also serve as a form of free advertising, using the reach of social media and the excitement generated by a new project to interact directly with their consumer base.

These more qualitative aspects of crowdfunding are also highly important, and have been the attention of a literature based mainly in marketing. For example, Agrawal et al. (2013) studies a wide-ranging set of explanations and motivations, but makes the intriguing suggestion that crowdfunding often plays the role of formalizing previous informal

fundraising arrangements between friends and family. While generally crowdfunding platforms provide security as middlemen to ensure that the terms of agreements are honoured, and perform a coordination role allowing backers to meet project owners, they also can be a way of explicitly "contractualizing" previous non-contract based loans or gifts. This is possible due to the fact that all the coordination is handled by the platform, making crowdfunding projects, in terms of financing methods, extremely simple to set up and accessible to the non-expert. They are entirely self-driven: there are no onerous capital or collateral requirements, there is no interview with a bank, there is not even a formal legal prospectus required. Only the basic structure of the pitch is specified: the level of detail, expertise, etc. are all left up to the individual. This gives crowdfunding extremely low barriers to entry, making it attractive for this formalizing role.

As mentioned in Section 3.2.1, an important part of crowdfunding deals with what happens after a project reaches its funding threshold. Primarily because data cuts off at this point, this has been largely unstudied. Mollick and Kuppuswamy (2014) takes the first attempt at this, by manually reaching out the supporters of a selection of successful projects to find out what the ultimate results were. His findings largely supported the notion that most projects were carried out in good faith, with a large majority delivering their promised rewards, and backers being largely satisfied with what they received. However, he also highlights that many projects are late with their delivery; over-promising on their delivery time-lines.

This ultimate outcome can legally be in a gray area: while crowdfunding is novel, it is not fundamentally different in terms of its obligations to its consumer or backers. When a project goes ahead, they are legally obligated to fulfill their promises to backers to the best of their abilities. So, while a project may go into production and ultimately fail this is not *prima facie* a legal issue: the context in which the failure occurs matters. A project in which the owners simply abscond with the money, never make a serious effort to produce the goods promised, or blatantly lied about the qualities or qualifications involved in the project are open to class action lawsuits by backers. On the other hand, a project simply overtaken by anticipated challenges or hurdles despite a good faith effort by the owners is legally blameless. For the majority of projects, which lie somewhere in-between these two poles, the protection for the backers and liability of the project owners can be somewhat up in the air. A major part of the 2012 US JOBS Act dealt with clarifying class-action liability for crowdfunding projects, helping to provide a legal framework to address these kinds of concerns.

The second set of papers studies build on work such as Agrawal et al. (2013) and study the choice of platform and funding structure. As mentioned every platform is slightly different, with different rules. Kickstarter, IndieGoGo, and GoFundMe were discussed at length in Section 3.2.1, and are typical of many platforms. Other platforms (MicroVentures, SeedUp) relax the idea of the project, and provide equity stakes (akin to "micro-venture" capital) in entire companies, but at a level much smaller than traditional venture funding levels in jurisdictions which allow this. Finally, others (Patreon) ask for repeated weekly or monthly contributions to support non-discrete ventures, such as an ongoing comic or video series. Each of these different structures has a different implication for how crowdfunding takes place. This also means that the differences in structure can shed light on consumer behaviour or on the decisions of project-owners. For example, Chang (2016) theoretically studies the implications of different fundraising structures for consumers, project-owners, and overall social welfare. Marwell (2016) exploits the difference between IndieGoGo and Kickstarter to determine how individuals behave and the consequences these kind of structures have for firms in an empirical context.

However, implicitly or explicitly, the attention in this literature has primarily focused on small individuals, mirroring the rhetoric (with Agrawal et al. (2013) the notable exception) of the business side of crowdfunding. That is to say, the focus has been on understanding the (extensive) decision of whether or not to back, rather than understanding the (intensive) choice of how much to pledge. This is usually because most projects have a well-defined reward subjects would want, and additional spending does not commensurately provide better deliverables. However, it is a feature of many digital markets to feature transaction data that has a similar pattern to that traditionally seen in a casino: most of the individuals are small time players, contributing small amounts. However, there are a small number of very large players, who spend orders of magnitude more money than the typical player. Adopting the language from casino managers, game developers and analysts refer to these individuals as "whales" (due to their large size) (Kokkonen et al. (2014)). Many digital environments are designed to encourage these kinds of players to pick up, and continue using their software (Lescop and Lescop (2014); Drachen et al. (2012)). The nature of such large contributors to the total revenue makes capturing and retaining their patronage extremely important. Analysts associated with some of the largest names in the video game industry acknowledge that motivations behind these individuals remain opaque, but appear to be at least in part motivated by rational concerns about achievement and tangible benefits (Sinclair (2014)).

I present early evidence that this is also the case for crowdfunding data. When we look at the pattern of contributions, they are generally similar on a per-backer basis day over day. This is consistent with the fact that, for most projects, there is a well-defined deliverable with a well-defined price which most backers would desire (e.g. a ticket to the concert). However, there are also a small number of much larger backers which appear in the data as much higher than normal per-capita contribution levels. These large contributors constitute a small fraction of the total number of backers, but just as in the microtransactions literature, they are disproportionately large sources of income. In the remainder of this chapter, I discuss the data on which these conclusions are reached, and present some stylized facts supporting this point of view. I also demonstrate that large contributions are predictable, rational, and not driven by statistical variation in contribution size.

3.3 Data and Facts About Large Contributions

The data for this chapter come from a novel sample of projects on the Kickstarter crowdfunding platform over a three year period. Each project was observed at regular intervals (daily, with some exceptions) and a variety of variables were captured, resulting in over 3 million observations; this forms a kind of unbalanced panel. In addition, project-level covariates were collected including the deadline for the project, the fundraising goal, and the category, which constituted a (fairly granular) self-selected indicator for an approximate "subject" such as "Christian Music" or "Board Games." The time-varying variables collected include the number of backers, the amount raised, and the number of comments and updates. Comments were publicly visible feedback left for project-owners by backers, while updates were notes appended to the pitch by the project owner. They consist of 128,172 projects with an average of 25.86 observations. Within the sample, projects raised an average of $8,621, with a wide standard deviation: a large number of projects raised nothing, while the largest projects raised tens of millions of dollars. These projects were backed by an average of 115 supporters, with the largest projects bringing in over a hundred thousand supporters. The average project had a goal of $34,320, which was reached an average of 38% of the time; about 1.5% of projects succeeded on their first day. Projects, on average, were 32[30] days in length, with a strong peak at 30 days, but also

[30] A reviewer pointed out the discrepancy between the average of 25.86 observations per panel, and the length being 32. This is primarily due to two features of the data. First, for projects at the beginning or end

VARIABLE	MEAN	STD DEV	MIN	MAX
Total raised to date ($1000s)	5.134	63.936	0	13283.28
Total backers to date (100s)	0.701	6.946	0	1058.52
Total comments to date	12.169	441.679	0	120,498
Total updates to date	1.288	2.861	0	147
Project funding goal ($10k)	0.343	8.766	0.000	1000
Total backers (100s)	1.152	9.713	0	1,058.52
Total funding raised ($1000s)	8.621	8.780	0	13,283.28
Duration of project	31.981	10.760	1	60
# backers today	4.524	67.554	0	37968
Amount raised today ($1000s)	0.338	5.418	0	2351.92
Avg contribution size	82.672	247.329	0	39,840
Is large contribution?	0.008	0.088	0	1
Amount of goal left ($10k)	3.315	99.186	-1,323.328	10,000
% of goal reached, capped at 1	0.263	0.350	0	1
Project successful?	0.377	0.484	0	1
Size of donation relative to mean	1	1.108	0	32.453
Fraction with large contribution	0.198	0.396	0	1
OBSERVATIONS	2,774,993	(Panel: 109,411)		

Table 3.1: Summary statistics

skewed to the right (with a spike at 60 days, the maximum possible). On a typical day, a project raised $347 from 4.5 backers. However, these numbers vary greatly as the project elapses. For a comparison of how this data set fits with other longitudinal studies of crowdfunding, see Appendix B. After cleaning the data for consistency in this project (excluding projects with incomplete data or data that was collected potentially erroneously[31]), I am left with 109,411 projects with a total of 2.7 million observations, whose characteristics are reported in Table 3.1.

This study focuses primarily on so-called "large" contributions; this poses a challenge, because the data is at the project-day level. This means that we cannot observe individual backer choices directly; either in terms of the decision to back, or the decision about how

of the panel, not all dates are captured; a form of either right or left censorsing. Second, due to idiosyncracies in the day projects are captured some observations correspond to more than one date. For example, the most common collection error is where a date was missing, and two days collected at once. Dropping this data, or directly correcting for this does not change the results of this paper.

[31]The overwhelming majority of these exclusions have to do with the data collection mechanism: through mechanical error, a project-day may be accidentally skipped but recorded, resulting in "phantom" days in which the total number of backers declines substantially then rebounds.

much to contribute. Ideally, this would be observable to the econometrician but we are limited by the data collection process. So, to infer when large contributions occur, I use the following method to identify a "large contribution day" (LCD) instead. The method is straightforward: for each project j on a given day $t = 1, 2, ..., T$ I compute the average per-backer contribution (C_{jt}) by taking the amount raised on that day (R_{jt}) and dividing by the number of backers (B_{jt}), so $C_{jt} = R_{jt}/B_{jt}$. I exclude days in which no one donated, or in which contributions were cancelled. I then, by project, compute a project "price" by taking the average of the day-level per-capita contributions, and also compute their standard deviation: $C_j = \frac{1}{T}\sum_t C_{jt}$, $\sigma_j = \sqrt{\frac{1}{T}\sum_t (C_{jt} - C_j)^2}$. I then denote a large contribution day to be a day in which the per-backer contribution exceeds the price by more than three standard deviations: $LCD_{jt} = \mathbb{I}\{C_{jt} > C_j + 3\sigma_j\}$.

For an illustration see Figure 3.2; in this example, I would conclude that day 11 was a large-contribution day for this project. Under the assumption that valuations are independently drawn from a single (Gaussian) population, this implies that "large" donations lie above the 99.7-percentile of the distribution. For a statistical discussion of this definition, and its statistical properties, see Appendix B. In terms of economics, however, there are several things to note. Firstly, a LCD has a straightforward interpretation as a day on which there was at least one backer who contributed more than three standard deviations more than average; there might be more, but by the pigeonhole principle there is at least a single large contributor. Unfortunately, this is as much as we can say given the data available; in a sense, this is a conservative way to determine large contributions, because many small contributions may "wash out" a large contribution making it difficult to detect. Thus, LCD indicate truly large amounts of money being pledged to project; likely a single large contribution laying much further than three standard deviations from the average. Second, it is important to note that this is a project-level definition of what is a "large" contribution; each project, based on its average contribution defines what makes a contribution large. This means that for some project, heuristically "large" amounts of money per-backer may not qualify as a large contribution. I made this decision primarily because projects do not, in Kickstarter, actually have a well-defined "price." Instead, they have may have several pledge levels and it is difficult *a priori* to infer which of these is relevant for backers; accordingly, I remain agnostic here about what is the "correct" price individuals believe they need to pay to support a project and instead infer it from the average. This definition is relatively robust to variations in the number of deviations or the use of the mean versus the median to infer "price."

Finally, I focus on the size of contributions relative to other contributions and not the goal or or absolute size. This is because many projects have very small or very large goals, making it difficult to compare what is large or small in terms of a contribution across projects. For example, the largest project asked for more than $100m[32] while the smallest asked only for a single dollar. A metric like "20% of the goal" for example would yield the perplexing conclusion that a shiny quarter given to the latter would be considered large, while $10m given to the former would not. To avoid these kinds of complexities I focus on the behaviour of individuals relative to one another, and not to the goal, which is driven by a decision on the project-owner's side. This has the drawback that for many projects which attract very few backers that qualitatively "large" contributions are not considered large since the "price" for support is inferred to be large; this is probably reasonable, because we could imagine that such a small project would have a very high "premium" for support from it backers. Alternatively, this means we're eliminating "non-credible" large donations, where large sums are given to projects with no hope of succeeding (since the donations is contigent on success).

In the actual data, approximately 0.78% of the days are characterized by large contributions; an indication that the process is not simply a function of variation in the average contribution size, since this is much higher than the frequency we would expect under a normal distribution. Furthermore, large contributions do not occur randomly across the life-cycle of projects. Approximately 71% of large contributions occur before a project meets its goal. As Fig 3.3 shows, it is also the case that large contributions occur towards the end of a project's life-cycle. Large contributions tend to linearly increase in likelihood up until the project is completed. However, we can also see that large contributions tend to fall off as projects are close to reaching their goal, becoming most frequent when the project is further from success. The interaction of these two terms is unclear from this figure, but it implies there may be different motivations behind large contributions which occur at different times. More interestingly, fully 15% of large contributions occur co-incidentally with the project's success; within the data, there is a single day on which projects cross their threshold. Disproportionately many of these observations occur on this date. A straightforward interpretation of this would be that these are "pivotal" donations: this is the date on which the project succeeds because it is a date which attracts a large

[32]This may seem like an absurd amount of money, but it is actually feasible for some projects. To date, two crowdfunding projects (*Star Citizen* and *The DAO*) have raised more than this goal, although not through Kickstarter directly.

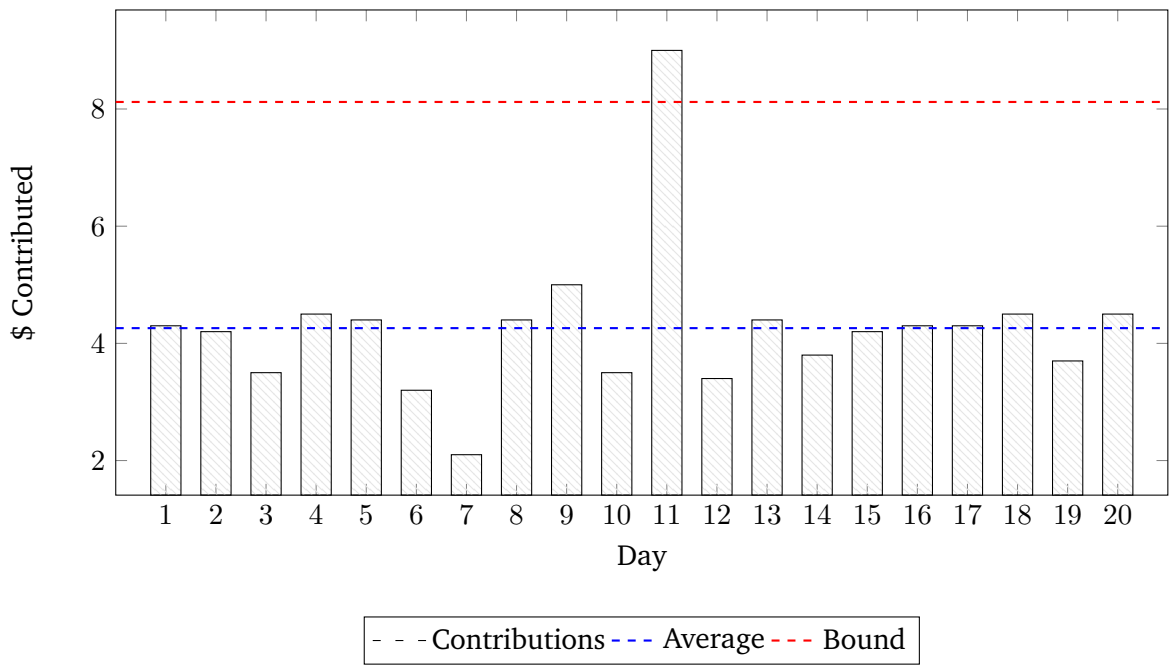


Figure 3.2: Illustration of Inferring Large Contribution-Days

contribution.

To look closer at these motivations, I consider the "excess size" of a large contribution. To get a sense of this, I take the total raised and subtract off the amount that "would have been raised" if the donations had been typical for the project. This gives an idea of how much a large contribution must have contributed on that day. In terms of a percentage of the average contribution, the typical large contribution day raised 6 times more than the average day. However, this is strongly non-normal: the 80th-percentile contributed almost ten times more, and the 95th-percentile contributed in excess of twenty times. The tail of the distribution is even more extreme. To see how these kinds of contributions are distributed between different projects, we can compare the excess size of these contributions to the percentage of the threshold remaining at the start of the day. On average, a large contribution closes about 38% of the gap remaining. However, this is even more striking when we restrict attention to large contributions on the "pivotal" days. As illustrated in Fig 3.4, there appears to be good evidence that large contributions are instrumental in closing the gap, with large numbers closing in excess of 70% of the gap remaining, and many closing far more.

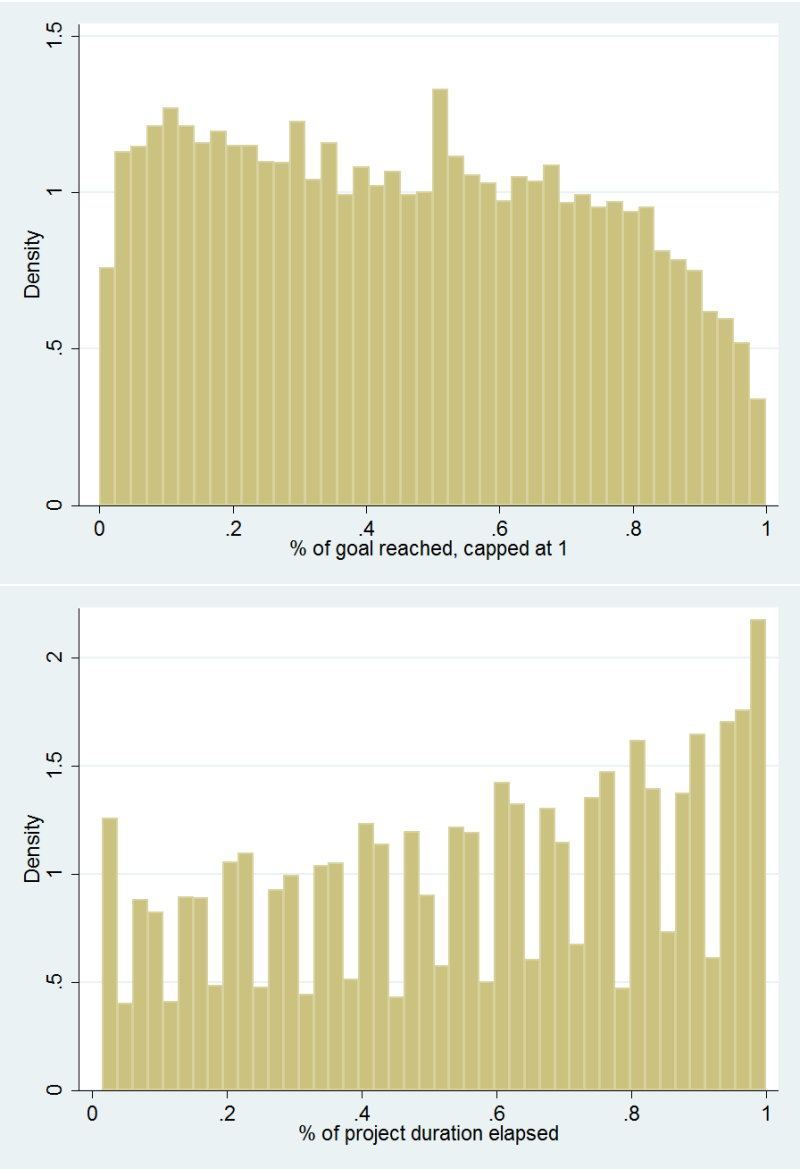


Figure 3.3: Histogram of Large Contributions versus % Elapsed and % of Goal

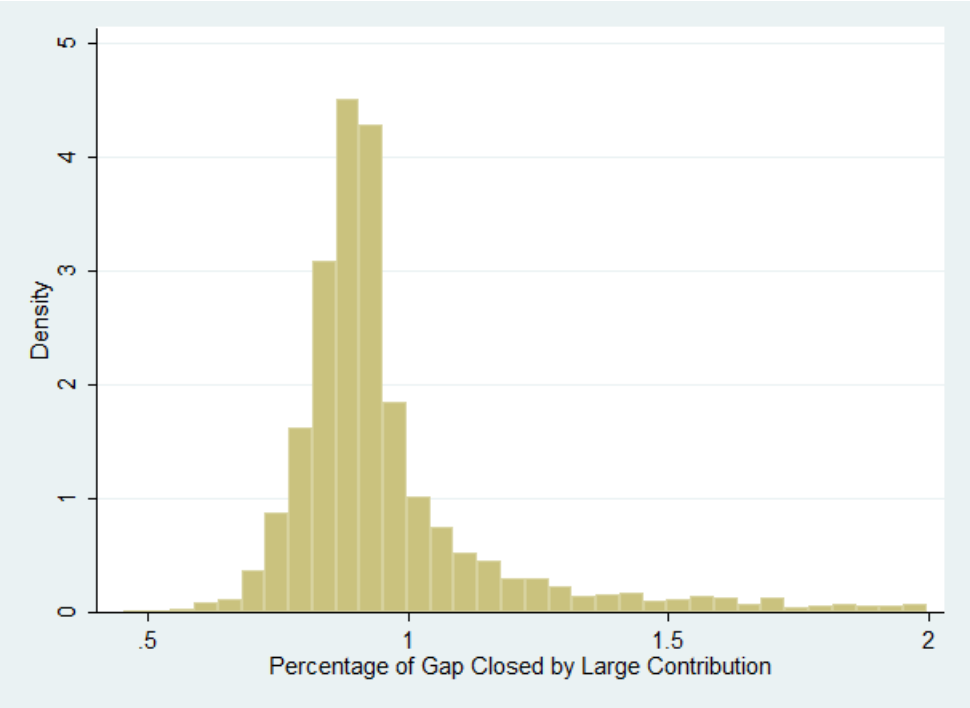


Figure 3.4: Pivotality of Contributions

However, we can also see this is not the entire story. When comparing projects which succeed with those that fail, we can see that as a percentage of the goal, contributions to ultimately successful projects tend to be larger. This implies that backers are concerned with the effectiveness of their contributions: they are more likely to give large sums to projects which are on the road to success and where they have a meaningful effect on the outcome. This provokes the question: do projects which have large contributions prove more likely to succeed? Indeed, projects which have at least one large contribution have a 79% probability of succeeding, while those that do not have only a 27% chance of succeeding. To see whether or not these large contributions are pivotal in a direct manner, I perform a naive counterfactual: for each project, I re-calculate the total raised, assuming that the large contributions were of average size instead. In fact, 1 in 3 of the successful projects no longer reach their goal following this calculation: that is, they would have failed if not for the direct role of the large contributions. However, because we cannot determine the dynamic effects of these large donation, nor can we exclude endogeneity, it is difficult to assess the meaningfulness or causality of this conclusion. Furthermore, it also poses something of a puzzle: based on this thought experiment, approximately 36% of the effectiveness of large contributions is driven by factors other than the size, potentially including the timing and efficiency of the contributions made.

However, these facts together provide evidence that large contributions play an important direct role in how crowdfunding projects are funded. The ability to attract these "whales" to a project can be important indicator of success, both because they donate more than average and because they provide funding when necessary. While pointing towards a conclusion, these stylized facts are difficult to analyse in isolation. In order to try and understand the motivations which might explain the kinds of behaviour we are seeing, I develop a consumer choice model of this kind of decision making, to help motivate the estimation and provide a framework for the results observed. This provides some predictions about how large contributions should behave, which I then test against the data, which is presented in the subsequent section.

3.4 A Model of Large Contributions

I model large contributions in a non-standard consumer choice framework. In order to focus on the intensive margin, I consider the decision about how much to contribute, rather

than whether or not to contribute. This means that utility is the benefit an individual receives, conditional on backing, after contributing an amount b above the minimum to a representative project at time t. That is, these individuals are going to back the project but they have not decided whether to contribute a small or large amount. They derive additional utility from increasing the likelihood of a project succeeding, but prefer to do so at minimum cost. I assume their utility takes the form $U(b) = P(b; s_t)[v - b]$ where P is a function relating the state of the project (s_t) and the amount donated (b) to the probability of the project succeeding, while and v is the value of the project (if successful) to the consumer. Note that this captures the contingent nature of crowdfunding contracts, since the individual only receives the good (and pays b) when the project succeeds. In other words, they decide on contribution size based on their expected utility from a given contribution, understanding that more money affects the likelihood of project success. Because b enters both terms of the expression, it is simpler to consider the following "unrestricted" utility function $U(p, b) = p[v - b]$ where an individual chooses both b and p. Similar to how in production theory not all combinations of output are possible, I assume that this is restricted by the "probability curve" $p = P(b; s_t)$. This allows us to depict this situation in a standard consumer choice diagram; the utility function corresponds to indifference curves where the highest is selected by the consumer subject to the probability curve restraint. However, the shape of this "budget" constraint is governed by the shape of the P function in the given state, and may result in the feasible set not being convex. For simplicity of comparison, I normalize the goal of a project to 100, so b is in terms of the percentage of the goal, and I abstract away from explicit financial constraints; they would arise as a limit on the amount of b possible.

It is important to note that I model the backing decision as a "snap-shot" of the dynamic process of crowdfunding. I assume that a single backer arrives at the project at time t, observes the state of the project s_t and the associated probability curve, and makes a single irrevocable decision about how much to contribute. The probability curve captures the consumer's expectations about how their contributions will affect the success of the project, taking into account expected future dynamics. The shape of this curve and its behaviour reflect both the changing state of the project and the consumer's beliefs about how that state relates to future success.

Notice first that the unrestricted utility function is a hyperbolic paraboloid (in b, p space) with asymptote at v, the project value. The indifference curves are hyperbolas in terms of b. This means that as v gets smaller, the indifference curves become steeper and never

cross the line $v = b$, which corresponds to the situation in which the individual no longer values the good at the contribution level. One interpretation of this is as the way in which the overall budget constraint enters the model; this is one dimension we are viewing, entering through v. As v increases, the curves become relatively flatter, and take on a modest u-shaped aspect. This form of the utility function $U(p, b)$ is the simplest possible; we could imagine that if individuals incorporate altruistic, warm-glow, or other behavioural preferences into their decisions, the shape of the indifference curves can be different.

As described, the probability curve captures consumer expectations about the future success of a project, given their contributions. This is an expectation-based model, but it is reasonable that it should agree with the observed distribution of project success, if consumers are rational. As described earlier, this tends to be strongly bimodal: projects tend to raise either all of their funding, or very little of it. This means that the probability curve takes on an S-shape which generally rises and flattens out as a project gets closer to its goal (holding time constant). Projects with different amounts contributed to date have different *effective* goals g^*, as well as different amounts of time remaining, backer arrival rates, and other time-invariant covariates having to do with the project. The shape of the curve is motivated by individual backer expectations about how likely a project is to succeed given the current state and the goal outstanding. However, notice that this is time-dependent; as projects run out of time, this likelihood declines, resulting in a deformation of the probability curve over time. I model this using a logistic function, in which both the lower asymptote and curvature change as the project evolves over time.

We can see these features of the model illustrated in Figure 3.5. In Panel (A), I depict a pair of typical consumers with different willingnesses to pay facing a project that is just beginning (notice that the effective goal $g^* = g$; the project is only sure to succeed if it raises g from the individuals). Both consumers have a maximum utility (achieved at Point A) where they contribute nothing in excess but the project succeeds with certainty; as indicated, utility increases in the direction of this point. The probability curve is indicated in bright red, and represents the trade-off for both individuals between contributions and the probability of project success. Since this project is beginning, I depict this as a logistic curve with a small lower asymptote and a steep curvature to reflect the observed distribution of projects. The first individual (solid lines) has high willingness to pay for the project; consequentially, they are optimally decide to contribute a large amount to see the project succeed (Point B). This point satisfies the standard tangency condition for an interior solution $\frac{\partial U(p,b)}{\partial b} / \frac{\partial U(p,b)}{\partial p} = -\frac{dP(b)}{db}$. The second individual (dashed lines) has lower

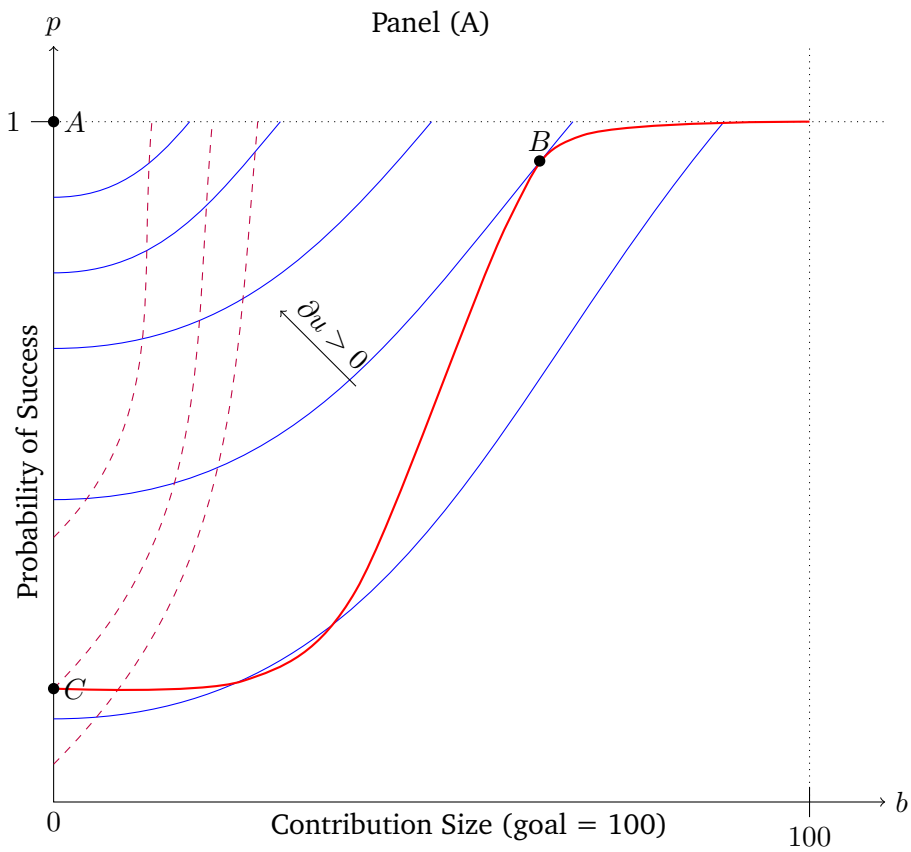


Figure 3.5: Consumer Choice Models (Panel A)

willingness to pay for the project. The slope of their indifference curve is steeper than the average slope of the probability curve, and consequentially they adopt a corner solution, contributing nothing (marginally) to the project. These two individuals correspond to large contributors and small contributors, respectively. The nature of a "large" contribution is such that it forms an interior solution to the consumers (constrained) optimization problem[33]; this fact will be useful later.

In Panel (B), we can see how the probability curve changes shape as time elapses. While this is based on the beliefs of the consumers, I make some reasonable inferences about how such a curve behaves based on the funding patterns and fund-raising goals.

[33]With the exception where a consumer has infinite willingness to pay, in which case it forms a corner solution on the opposite boundary to the other consumer.

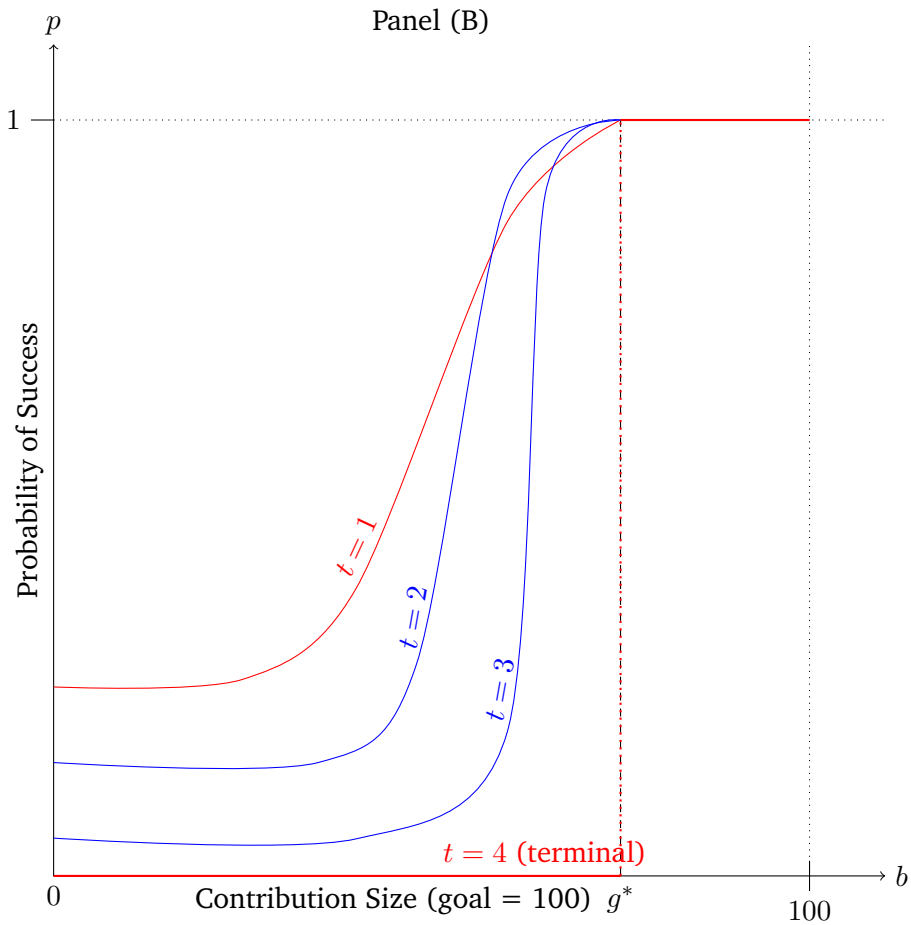


Figure 3.6: Consumer Choice Models (Panel B)

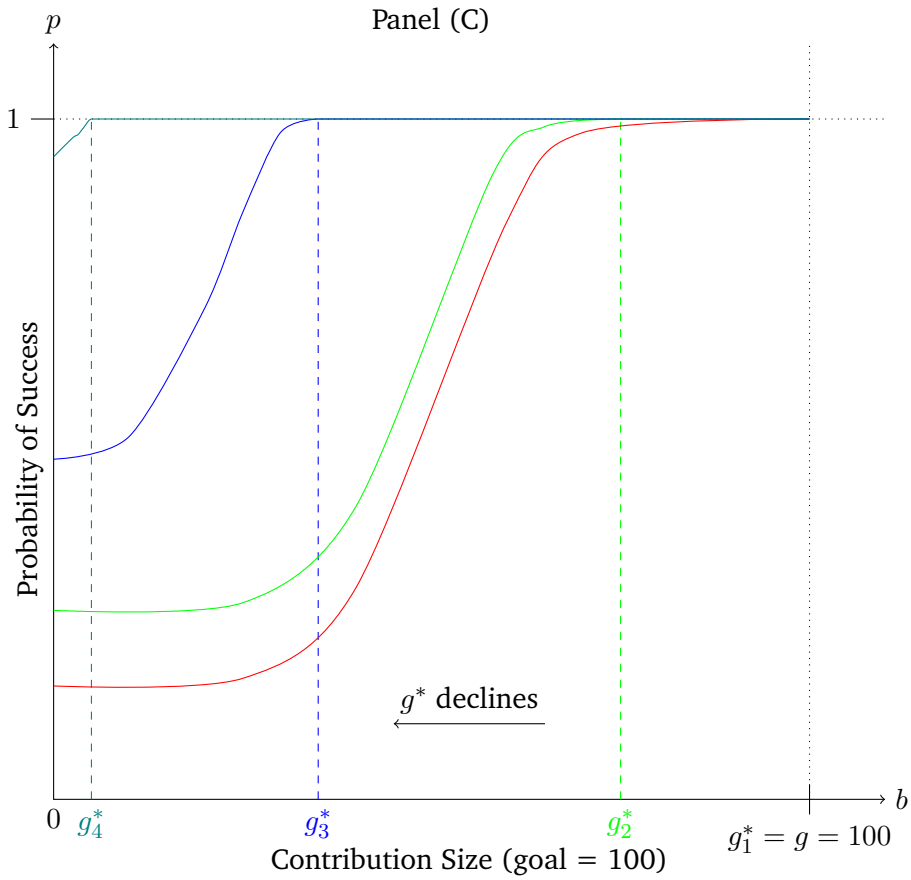


Figure 3.7: Consumer Choice Models (Panel C)

In this illustration, individuals can only arrive at four time periods, $t = 1, 2, 3, T \equiv 4$ and we hold constant the amount donated to the project, so the outstanding goal g^* is fixed. I also assume this function is smooth, mainly for simplicity of analysis. As time elapses, the probability of success for low contributions will fall, since the amount of time remaining for a project to success is running out. Notice that in general one might expect that for a given contribution level success should fall monotonically as time elapses, but this is not necessarily true (pictured by $t = 1$ versus $t = 2, 3$) in some situations. For example, individuals may have different expectations about how certain types of projects are inclined to raise funding based on their current fund-raising; imagine a project which is known to attract attention if it is believed to be "behind" by consumers. Nonetheless, restricting attention to smooth probability curves, we know that the limiting behaviour is such that success for levels of contribution less than g^* falls, while success for level at g^* is 100%; this necessitates a "steepening" of the logistic-shaped curve into a step-function in the limit. That is to say, the slope of the function becomes locally very close to 0 about $b = 0$ and g^*; $m(g^*) = m(0) \approx 0$. Furthermore, as time elapses, these flat regions become closer together in terms of the b-axis, but remain separate in terms of p. That is to say that for small ϵ at $t \approx T$, $P(g^* - \epsilon; s_t) \approx 0$ and $P(g^*) = 1$, while $m(g^* - \epsilon; s_t) \approx m(g^*) = 0$. But, then by smoothness of the function P, it must be the case that the slope become arbitrarily steep as time runs out: $\lim_{t \to T} \frac{dP(b)}{db} = \infty$. Essentially, as time run out, fewer backers are available to support a project in the future, which both makes success less likely and increases the importance of a large contribution in terms of helping a project reach success. In the limit if you are the last possible contributor, the project will be successful if and only if you contribute the outstanding funds. An opposite situation occurs when we hold the amount of time remaining constant, but vary the goal outstanding, g^*. As illustrated in Panel (C), as the goal outstanding falls, the S-shaped probability set shifts upward and to the left, generally flattening in a similar fashion to the previous example and $g^* \to 0$. The upper and lower asymptotes draw closer together as the project's success becomes more certain, and the curve flattens out.

3.4.1 Simulation and Predictions

In order to investigate the model more precisely, and verify our intuition, I perform a numerical simulation of the solutions to the model. Explicit theoretical solution is difficult because of a partial failure of Berge's maximum theorem: the "budget constraint" set cre-

ated by the S-shaped probability curve is non-convex. This means that the set of solutions is not convex as the budget curve changes and the solution can "jump" between different points. In what follows, I assume that consumers have expected utility represented by a hyperbolic parabolic $U(p, b) = p[v - b]$ where v is the consumer's value for the project. For simplicity of illustration, I make the change (relative to the theoretical model) that the contribution is already scaled in terms of % remaining. That is, a contribution of $b = 1$ corresponds to the total goal remaining, not the total goal. In other words, I adjust the x-axis so that it is in terms of g^* rather than g. This has no implications for the solutions, save that the interpretation of v in the budget constraint is relative to g^* not g; since g^* is fixed in each example, this is without loss of generality.

To form the probability set, I parametrize my S-shaped curve as the generalized logistic function

$$P(b) = a + \frac{1 - a}{1 + \exp(-k(b - m))}$$

The parameter a governs the lower asymptote; in this case, the likelihood a project with g^* outstanding would succeed without a large contribution from the donor. m governs the location of the inflection point; for illustration, I initially set $m = \frac{1}{2}$ in what follows. Naturally, in specific situations other locations might be more reasonable; I explore this later in this section. The parameter k governs the slope of the curve; higher values are more aggressively sloped. In what follows, I choose $k = 20$ except where noted.

The consumer's problem is to maximize $U(p, b)$ subject to the constraint that $p \leq P(b)$. Since utility is strictly increasing in p and decreasing in b, this implies that this constraint should hold with equality at an optimum. First, I illustrate the two key solutions pointed out theoretically developed earlier: small contributions, which arise as a corner solution, and large contributions which arise as interior solutions. This is depicted for $v = 0.8$ in Figure 3.8. As we can see, when $a = 0.2$, the initial likelihood of success is 20%, the individual chooses to give a large contribution to the project, in excess of 60% of the outstanding goal. However, when $a = 0.3$, the individual does not choose to give a large contribution. Essentially, when the individual values the project highly, if the probability of success is not high enough, they are willing to give a large amount to make it happen. However, if the project is in "good shape" from their point of view, they do not feel it necessary to do so. This is the central insight of the model illustrated in this environment.

I explore this further in Figure 3.9, which illustrates the way the solution to the model

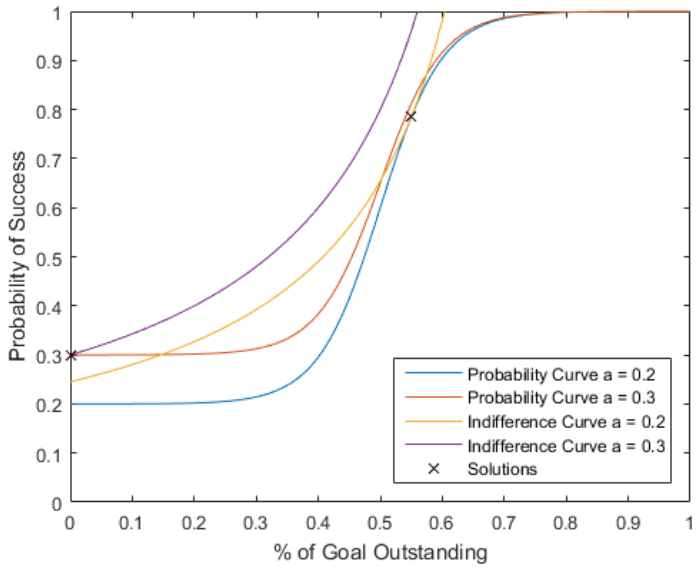


Figure 3.8: Illustration of Numerical Solutions to Model v=0.8

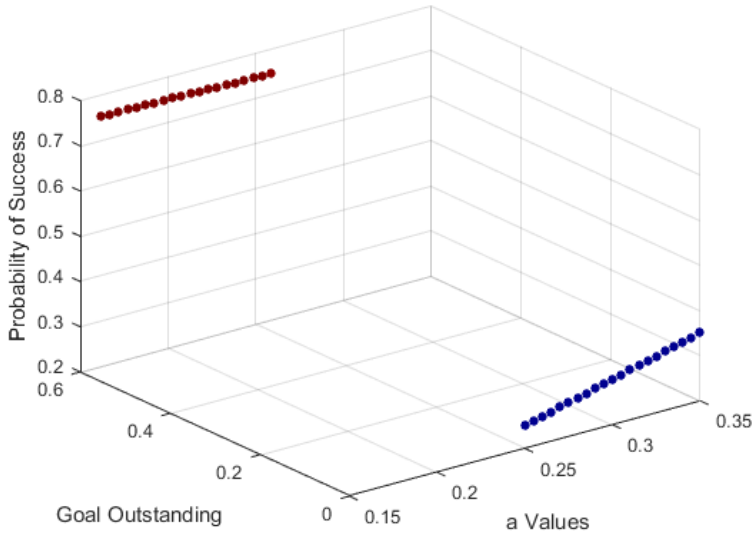


Figure 3.9: Evolution of Numerical Solution to Model v=0.8, Varying a values

evolves as a changes. We can see there is a sharp non-convexity in the solution set, which occurs at $a = 0.25$. Prior to this value, all of the contributions are large. After this value, all of the contributions are small. The significance of $a = 0.25$ is that it happens to be the value of a at which this particular consumer is indifferent between providing a large and a small contribution to the project. One interpretation of this is as an illustration of a consumer's decision as the time remaining falls. Projects with less time remaining for a given outstanding goal have a lower probability of success, which can be modelled as a falling a-value.

Figure 3.10 also illustrates one additional consequence of the model: the inflection point matters. In this figure, I depict the evolution of the numerical solution as the inflection point evolves from 30% to 60% of the goal outstanding. As we can see, this leads to a change in the intra-marginal contribution decision. When a large contribution is relatively effective at creating success individuals are likely to give *less*. In terms of the model, when the inflection point is relatively low, individuals feel the need to give less. As the inflection point moves away, the individuals are likely to give more, up to a point. Eventually, as the inflection point moves far enough away from the origin, individual decide that a large contribution is no longer useful; it is simply too costly to support the provide at this point, since they are not effective enough at helping it succeed. This also has a natural analogy with the model: when projects are running out of time, the inflection point move away from the origin, as depicted in Figure 3.6. As the simulation show, conditional on a large contribution occurring, the contribution size is likely to be large. This is also impacted by the value of k, the curvature, which likely also increases along the same dimensions. However, without a movement in the inflection point this is not a major determinant of contribution size.

Based on the numerical investigations of the model, and the theoretical set up, we can formulate some predictions the model will make about when projects should see large contributions. These can be based both on the simulations and on extrapolations based on the theoretical model:

- **Prediction 1**: Large contributions should be more common as time remaining drops, conditional on goal outstanding. Theoretically, this prediction is supported by noting that a large contribution occurs as an interior solution to the consumer optimization problem described above. In particular, it requires that tangency exists between their indifference curves and the slope of the probability curve. In this model, this is more

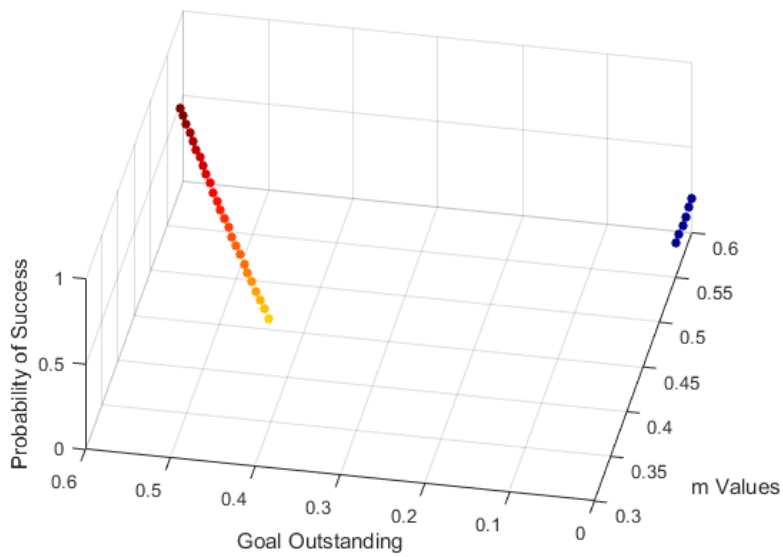


Figure 3.10: Evolution of Numerical Solution to Model v=0.8, Varying m values

likely as projects run out of time, since the slope of the budget line evolves from being very flat to very steep. The main limitation on this is that the region of steepness must occur before v, the maximum willingness to pay; otherwise the point is not feasible for the consumer to purchase. This can be interpreted as the notion that individuals with high value for the project are more likely to contribute large sums of money towards the end of the project, either because they have high income (and therefore low marginal utility of wealth) or they value the good substantially more than other consumers. We can see direct support for this in our simulated results, in which time running out can be viewed as a fall in the value of a. This prediction is illustrated in Figure 3.9.

- **Prediction 2**: If backer arrivals change over time, large contributions are more likely in period in which there is a relatively higher mass of arrivals, all else equal. In particular, if project attract more attention early in their lifespan, large contributions are more likely early. This prediction is a straightforward extrapolation from Prediction 1; if more individuals occur, conditional on the model set up, more large contributions should occur relative to a lower arrival period.

- **Prediction 3**: Large contributions close to a project's deadline should be more likely to be pivotal than those earlier in the project's lifespan. This follows from the same reasoning as Prediction 1, noting that not only are they more likely to occur, but also the point at which the contribution is placed is closer to g^* the outstanding goal for most WTPs (slopes). In the numerical simulations, this follows from the investigation of the change in m. Pivotality, in the numerical model, is described by large values of b. The numerical simulations predict that shorter time frames imply larger m, which mean that conditional on wanting to provide a large contribution, the size should be large. This is illustrated in Figure 3.10.

- **Prediction 4**: Large contributions should be much less likely after success. Theoretically, this follows from the fact that large contributions are not necessary once a project has succeeded; the optimum point for all agents can be achieved at zero additional contribution to the project. The numerical model also captures this, since a successful project is a point on the extreme upper left of the feasible set, which implies it must be a corner solution (and a small contribution).

Some of these predictions have already been spoken to by some of the preliminary facts addressed in Section 3.3, but there is a problem; primarily, there is a great deal of heterogeneity in projects, with many different dimensions changing at the same time. In order to address this challenge, in the next section I describe a linear probability model to examine when large contributions occur. The linear regression framework used allows me to control for the different covariates of interest, analysing them in a single framework,

3.5 Why Do Individuals Provide Large Contributions?

This chapter first looks at what features of projects are related to the occurrence of large contributions; what determines whether or not a large contribution occurs on a given day? As described in Section 3.4, this is a complicated issue, with many, often-contradictory parts. I use a multiple regression framework in order to separate the effects of different parts, while still addressing them together. Specifically, I regress an indicator variable for whether or not a contribution day was "large" on a variety of covariates, both time-varying and static. This implies that the baseline specification is a linear probability model with equation:

$$LCD_{jt} = X_i'\beta_1 + W_{jt}'\beta_2 + \epsilon_{jt}$$

where ϵ_{jt} is the error term, $LCD_{jt} \in \{0, 1\}$ is the presence of a large contribution for project j at time t, X_j is a set of project-specific covariates, while W_{jt} is a set of time-varying covariates.

As an alternative, and in order to control for potential project-level effects not captured by the set of covariates X_i, I also adopt a fixed-effects model in which:

$$LCD_{jt} = \beta_j + X_j'\beta_1 + W_{jt}'\beta_2 + \epsilon_{jt} \iff \bar{LCD}_{jt} = \bar{W_{jt}}'\beta_2 + \bar{\epsilon_{jt}}$$

where the bar notation indicates the within-panel transformation. The drawback of this method is that, because of the fixed effects assumption, only the time varying components can be included.

It is important to note that the relationships in this model are correlations, *not* causal effects; many of the regressors evolve endogenously with the project, and even in the absence of this factor the identification of the model is probably implausible to assume, given the unobserved heterogeneity which is likely in the data. However, these do allow us to see which components of the projects move together, and assess how they balance against the different behavioural explanations for large contributions. I find that, even in the absence of a causal effect, the regression framework supports the story being told by the descriptive evidence: large contributions care about project success, and try to help projects succeed. We are also able to give an assessment of the predictions made by our theoretical model.

The project level variables used are the length of the project, its goal, its average contribution size, and a detailed indicator variable for the category of the project. The categories comprise 164 different groups selected by the project owner as a description of their project. For example, categories include Video Games, Publishing, Product Design, Sculpture, Performance Art and Classical Music, to name a few. The baseline omitted category is normalized to be the most populous (generic "Art"). The time-varying covariates included change for different specifications, but in total include: time fixed effects for day of week, year, and month, the goal outstanding, a quadratic term based on the percentage of time elapsed, whether or not the project has succeeded, the number of comments, backers and announcements for a project, and interaction terms relating to the goal outstanding and success, or percentage elapsed. The results are summarized for the baseline specification in

Table 3.2. Standard errors are heteroskedasticity robust, and clustered at the project level when considering fixed effect models. For time variables, I omit January, Sunday, and 2012 as the baseline comparison for fixed effects, when present. I also omit projects which never receive any backers, since they have are generally uninformative for the analysis here.

A first important fact to note is that, contrary to an interpretation of large contributions as simply being a mechanical effect based on the incidence of large values from a population, we can see that both the duration of the project and the average contribution size have very small effects: both on the order of 0.01%; they are significant, but tightly estimated around zero. In fact, the coefficient on longer projects is actually negative, albeit very small. Similarly, the number of total comments has a small, negative effect - indicating that while this may directionally be bad news, the aggregate number of backers who comment does not play a direct role in the incidence of large contributions. This can also be seen directly: considering the number of backers on a given day, the coefficients is very small. This means that large contributions are, as we suspected from the stylized facts, truly being driven by a process distinct from the overall arrival and backing rates of projects.

On the other hand, projects do tend to attract more large contributions when they are reaching their deadline: a project at the end of its lifespan is 0.04% more likely to see a large contribution than one at the beginning, peaking at about 60% completion. This effect is small, over when compared to the primary drivers. First, project post-success are less likely to attract a large contribution: an effect which becomes larger with time, peaking at about -0.3%. In the baseline and RE model, the size of the goal outstanding is non-significant. The largest coefficient belongs to the "pivotal day" indicator; this is the day on which projects succeeded (and is therefore neither fully pre or post success); projects are fully between 1% and 9% more likely to see a large contribution on the day of successful, depending on when that day occurs in the lifespan of the project.

Turning to the fixed effects model, we see the results are generally robust, but do differ slightly. This indicates that unobserved project-level effects are probably at play. Most notably, we now see a statistically significant effect of the outstanding goal; a larger outstanding goal leads to fewer large contributions, indicating a preference for "pivotality" or being effective in the large donation. Next, while projects are still more likely to attract large contributions as they approach the end of their lifespan, now purely increasing over the life of the project. Furthermore, projects are also much less likely to see large contributions after success. These results are also general, if we more beyond the linear probability

model and look at the marginal effects of the outcome of a probit specification. This should not be read that this "pivotal" day causes a large contribution; rather it is likely the opposite, in which large contributions are very likely to be pivotal themselves, pushing a project over the threshold. This provides good evidence that we should believe large contributions are likely to be instrumental in project success.

It is also clear that, on some level, there is a clear story being told about large contributions. Large contributions show a preference for being effective, occurring more frequently in projects which have not yet succeeded, but are nearer their goals. They also show a strong "pivotality" affiliation, being very likely to occur simultaneously with project success, especially if the project is late in its life. A major drawback with this is the fact that this study is observational; due to dynamic effects and endogeneity, it is unlikely to tell us too much about the direct effects of any of these features.

3.5.1 Assessment of Theoretical Predictions

Our theoretical model makes several different predictions about when large contributions should occur, and what they should look like.

- **Evaluation of Prediction 1**: Large contributions should be more common as time remaining drops, conditional on goal outstanding. This hypothesis is confirmed by our analysis; first of all, we see that pivotal contributions, which as the hypothesis points should be more common as time runs out, are strongly related to the time remaining. A pivotal day is about 8% more likely to be a large contribution day towards the end of the project, relative to the beginning. We also see that non-pivotal large contributions are more likely towards the end of the project as well by looking at the time remaining, although this is not a linear relationship and peaks at about 60%; the reason for not seeing a monotone relationship is likely due to the trade-off between pivotality and non-pivotality; projects which are closer to their deadline are likely to have a smaller goal outstanding which implies an equally-sized large contribution is more likely to be pivotal towards the end, causing a hump-shaped relationship as we observe.

- **Evaluation of Prediction 2**: If projects attract more attention in a given period, then large contributions are more likely in this period. We do not find support for this hypothesis. In particular, it appears that although total numbers of backers (or

potential backers, see Marwell (2016) for example) are larger earlier on, this does not translate into large contributors. As we see the number of backers today is not influential on the number of large contributors, with an essentially zero coefficient. This is likely because large contributors are willing to wait and observe a project to assess whether or not it is actually in need of their aid.

- **Evaluation of Prediction 3**: Large contributions close to a project's deadline should be more likely to be pivotal than those earlier in the project's lifespan. This hypothesis is confirmed by our data. As we note, large contributions are substantially more likely to occur on pivotal days (1% more likely) and this effect is increasing over time, to a maximum of 9% at the end of the project's lifespan.

- **Evaluation of Prediction 4**: Large contributions should be much less likely after success. This is somewhat supported by our data; we can see that only about 15% of contributions occur after success, and the coefficient on success has a negative coefficient. However, the fact that some contributions still occur, and the size of the coefficient indicate that this is not conclusive. This is possibly because of individual preferences which are not captured by the model, or alternatively because projects may still ask for additional money after reading their goal; these so-called "stretch goals" are unofficial add-ons which serve as quality improvers, and may motivate additional large contributions.

We can also rule out some alternative explanations. Specifically, we can see that large contributions are not a product of statistical variation in the size of contributions: the precisely estimated zero-order coefficients on length, average contribution size, number of backers, and number of comments left on a project all speak to the fact that large contributions are purposely and involved. This also rules out naive private value explanations, such as a situation where the size of a contribution is solely driven by exogenous variation in the WTP for the reward. Additionally, we can also see that there are project-level fixed effects at play: the inclusion of fixed effect into the model provide similar results, but generally supports the notion of unobserved heterogeneity in the data, indicating this is a concern for further analysis.

In general, we find that large contributions largely match the dynamic behaviour we would expect from the consumer choice model outlined. Large contributions are interested in the reward from a project, but are also interested in influencing the probability of suc-

cess. This interest in being influential speaks to the reason we see large contributions and pivotality being so closely linked in the data; the donation which funds a project is a locus of the behaviour of individuals in the model. However, as Prediction 4 outlines, there is clearly something more going on in some cases; potentially due to aspect of the projects we cannot see, or due to different types of behaviour on the part of consumers we do not model. The overall picture remains clear: large contributions are motivated by improving the probability of project success, and individuals are willing to pledge large sums of money in order to help make these projects work. This leads to the second question: if this is the motivation for large contributions, then how effective are large contributions at achieving this end. In the balance of the paper, I seek to address this question, beginning in section 6.

3.6 How Important Are Large Contributions to Crowdfunding Success?

To assess the impact of these large contributions on the ultimate outcome of a crowdfunding project, I consider the project-level data set, and collapse over the time dimension. I use the baseline linear probability model to regress an indicator variable for whether or not a project was successful (Y_j) on a set of covariates (X_i). The covariates in this model include all of the project-level covariates from the panel data regressions, plus a variety of variables relating the presence of large contributions: an indicator for having a large contribution, the number of large contributions, and the excess size of the contributions over the average. I omit projects which did not see any contributions from which a price could be generated, since these cannot be informative about the impact of large contributions. These results are reported in Table 3.4.

In the baseline model, we can see that the impact of an additional large donation increases the likelihood of a project succeeding by 33.7%; this is tightly estimated. This is an order of magnitude stronger than any other variable on project success. The impact is nonlinear, however: regression on an indicator for any large contributions has approximately the same coefficient. This could imply that it is the *presence* of large contributions, not necessarily their number, which matters. This is also likely the case because most projects only have one or zero large contributions, so the further marginal effect is not very present in the data-set. There is a knock-on effect of large donations outside of the simple monetary

	Baseline Coeff/Std. err.	RE Model Coeff/Std. err.	FE Model Coeff/Std. err.
Duration of project (10s)	-0.0048***	-0.0047***	
	(4.3e-05)	(4.3e-05)	
Project funding goal ($100k)	0.0001	0.0001	
	(0.0002)	(0.0002)	
Number of backers today (100s)	-9.9e-04***	-9.9e-04***	-0.0015***
	(1.9e-04)	(1.9e-04)	(3.4e-04)
Total backers to date (100s)	0.0001***	0.0001***	-0.0007***
	(0.0000)	(0.0000)	(0.0001)
Price (100s)	0.0012***	0.0012***	
	(4.4e-05)	(4.4e-05)	
Amount of goal left ($100k)	-0.0001	-0.00012	-0.0008***
	(0.0002)	(0.0002)	(0.0002)
% of project duration elapsed	0.0028***	0.0027***	0.0023**
	(0.0008)	(0.0008)	(0.0008)
% elapsed squared	-0.0025***	-0.0023**	0.0017*
	(0.0007)	(0.0007)	(0.0007)
Goal Outstd. X % Elapsed	-2.0e-06	-2.0e-06	-9.5e-07
	(1.5e-06)	(1.5e-06)	(5.8e-07)
Total comments to date	-1.2e-06***	-1.2e-06***	-3.1e-07
	(1.7e-07)	(1.7e-07)	(3.9e-07)
Total updates to date (100s)	0.1234***	0.1236***	0.0686***
	(0.0034)	(0.0034)	(0.0066)
Pivotal day indicator	0.0101***	0.0101***	0.0103***
	(0.0024)	(0.0024)	(0.0024)
(Piv Ind)*(Pct elapsed)	0.0837***	0.0837***	0.0858***
	(0.0035)	(0.0035)	(0.0036)
Succeeded	0.0009	0.0009	-0.0065***
	(0.0005)	(0.0005)	(0.0008)
(Succeeded)*(Pct elapsed)	-0.0031***	-0.0032***	-0.0030**
	(0.0007)	(0.0007)	(0.0009)
Category controls	X	X	
Time controls	X		
Project FE			X
Constant	0.0050***	0.0056***	0.0107**
	(0.0004)	(0.0006)	(0.0041)
N	2.5e+06	2.5e+06	2.5e+06
VCE	robust	robust	cluster

Table 3.2: What leads to large contributions?

benefit being provided, which is generally less than the level of impact would indicate.

The central problem is that the large contribution measure might be endogenous for several reasons: trivially, large contributions could be correlated with unobserved project-specific variables which are also correlated with project success. More importantly, however, we explicitly have a situation of reverse causality in which large contributions are affected by success, and vice versa. In order to overcome this, I use an instrumental variables approach. Specifically, because we think large contributions may have something to do with individuals caring about a project in an outsized way, I created a list of 15 religious and secular holidays which are related to gift-giving, charity, or generosity[34]. Because projects vary in terms of when they start and the length of time they cover (the "tenure"), the number of holidays they capture also varies, which introduces useful variation in the data. Since I control for the number of backers, this implies that the instrument can only affect the project through the *amount* given, not the decision about whether or not to back the project. Additionally, since I control for length, month and year fixed effects (this is also robust to more granular time controls, such as week-of-year), the key identification assumption then becomes that the tenure start dates, controlling for month, year and other project covariates, are independent of the success of the project. In particular, this means that we must believe project-owners are not manipulating the start/end dates of their projects to pick off marginal holidays. That is, a decision like "we will start our project in December because of Christmas" is fine but a decision like "we will delay our project from November 23 to 25th to capture Christmas" would bias our instrument. I believe this assumption is credible for several reasons; first of all, the reason most individuals use crowdfunding is because they are relatively unsophisticated, and have trouble accessing traditional capital markets. This unsophistication is also likely to mean they are unable to make very granular decisions about project timing. An additional reason is that project launch date is could be difficult to completely control by the project-owner: there are many stakeholders and moving parts involved, which make the exact launch date difficult to set with precision. This is supported by the fact that, with the exception of the end of the year (around New Years), project start dates are fairly evenly distributed across the year. If this assumption fails, the results are likely to be overestimates; we are accidentally picking up the effect of "sophistication," in the time-manipulation sense, on the likelihood of project success, and it seems plausible that more sophisticated project-owners are more likely to

[34]For example, Christmas, Black Friday, the start of Lent, Boxing Day, Chinese New Year, etc. See appendix for more details.

	Holidays # projects (mean)	Non-Holidays # projects (mean)	95% CI (Difference)	Least Likely Difference (P-value)
All Dates	2766.44	3316.63	[-56.5,1156.5]	$H_a < 0$ (0.96)
2013 Only	3195.23	3248.00	[-158.8,264.83]	$H_a < 0$ (0.68)
Dec 2013 Only	3112.92	3354.05	[-165.5,649.5]	$H_a < 0$ (0.88)

Table 3.3: IV Robustness Checks

succeed.

To try to detect this "time-manipulation" I consider whether or not projects are more likely to pick up holidays during their tenure than other, unrelated dates. In Table 3.3, I calculate the number of active projects on each date, then look at the average by holiday/non-holiday. As we can see, most it appears that projects are less likely to be active during the holiday dates than other days; however, because data collection begins in 2012 and ends in 2014, this might mean that the early/late data might be incomplete (especially in 2012, since crowdfunding was just beginning). Accordingly, restrict my sample to just 2013, the best full year in the sample. We see the same result, and are able to reject the hypothesis that projects are more likely to capture holidays. In the IV estimates, since we are particularly concerned about month controls, I also look at selected months; December 2013 is presented here, but the results are similar for other months. It appears that projects are not more likely to be active on holidays than other dates, supporting the idea that project owners are not trying to select additional holidays. I also check whether or not the projects appear different within a month, versus outside; specifically, I focus on December 2013 and perform the same IV regression as before; the results are similar, indicating that even comparing projects which elapse during December, the IV estimates are stable, indicating that unobserved effects are not driving manipulation which might affect the instrument's validity. Additionally, running the IV regression excluding Christmas, Christmas Eve and Boxing Day (since these are the most "visible" holidays from a US perspective) gives similar results with a point estimate of 0.44 (Std err. 0.09).

The identification assumption, backed by the time controls, and the number of backers, means that the number of holidays are not conditionally correlated with the overall project's success except through their impact on contributions. This allows us to use this as an instrument for large contributions, via the TSLS approach. However, we might be concerned that the instrument effect is very small, causing a weak instrument situation. To check this intuition, I perform a regression of whether or not a day has a large contribution

on the variables in the baseline model, plus an indicator for the holiday variable, has a significant coefficient; it does, at the 5-pecentile. The first stage of the TSLS is similarly strong, with an F-stat of over 40; these facts together provide evidence that the instrumentation is strong enough to be valid.

As we can see, the results are similar but generally stronger when we using instrumental variables. The effect of having large donations is about 10-30% stronger in this case; this implies that for many projects, the endogenous project-level factors are negatively correlated. This is plausible if we imagine that some large contributions are not made seriously (i.e. with the goal of obtaining a reward) and instead made for other reasons, such as costlessly signalling interest in a project or simply for a joke. This means that the causal effect of large contributions in the baseline model is actually probably higher than 30%. This indicates that large donations are important for project success - a project's ability to attract large donors is an important element for their success. This agrees with the naive uncontrolled evidence we saw earlier, in which approximately 20% of projects would have failed if the large donations had been of average size instead.

These results indicate strongly that large contributions play an important role in project success, and are worthy of further study. However, because of the identification methods used, there are interpretation issues surrounding the results. I discuss further attempts to alternatively evaluation these results, and discuss robustness and some further exploration in the next section.

3.7 Discussion

3.7.1 Effectiveness of large contributions relative to size

The preceding section has established that there is a relationship between project success and large contributions which goes beyond unobserved project-level effects or statistical noise. In fact, large contributions appear to be very important for overall success. We could imagine this occurs for different reasons: the first is mechanical. Simply put, large contributions contribute more money than a typical contribution, which goes towards the goal, and therefore is relatively effective at reaching the goal. While simple, this is actually difficult to assess, primarily because we cannot control for the amount raised and still interpret the regression model sensibly (the dependent variable is a linear combination of the amount raised and the goal). As an alternative, I perform a "smoothing" exercise: in

	Model 2: LP Coeff/Std. err.	Model 2: IV Coeff/Std. err.
Price (100s)	0.0190***	0.0126***
	(0.0022)	(0.0035)
Goal (1000s)	-0.1101**	-0.1019*
	(0.0417)	(000397)
Length	-0.0082***	-0.0089***
	(0.0001)	(0.0003)
Number of Large Contribs	0.3374***	0.5298***
	(0.0038)	(0.0920)
Total Comments (100s)	-0.0022***	-0.0017***
	(4.77e-04)	(4.82e-04)
Total updates	0.0256***	0.02014***
	(0.0007)	(0.0027)
Total Backers (100s)	0.3807***	0.3458***
	(0.0724)	(0.0694)
Constant	0.6857***	0.6841***
	(0.0130)	(0.0133)
Category Controls	X	X
Time controls	X	X
N	95,792	95,792
Errors	robust	robust

Table 3.4: How important are large contributions?

	Adjustment of Backers
Change in Variable	29.6
(1) Impact (Baseline)	11.25%
(2) Impact (IV)	10.01%
Net Impact of Large Contributions (Baseline)	22.49%
Net Impact of Large Contributions (IV)	42.96%

Table 3.5: Counterfactual "smoothing" exercise

the data, the average large contribution is $2,098 larger than expected. We could imagine moving this amount around in the model by shifting how it arrived. Basically, we imagine more individuals arrived, and compute, given the median contribution, how many incremental arrivals the typical large contribution "represented". Then, a measure of the direct effect would be a comparison of the change in the probability by trading off this weight. This is depicted in Table 3.5, which illustrates that the effect is about 10% directly. This means that about 2/3s-3/4s of the effect of large contributions is being driven by factors other than the simple amount being donated. The simplest explanation is that large contributions are not merely providing more money: they are providing more money *when* it is needed and in the *amount* needed. This "targeting" of contributions to when they are most urgent would explain the fact that these contributions are essentially about 3-4 times more effective than we would expect. An alternative explanation, which is difficult to analyse in the data, has to do potential informational concerns. Some authors (namely Mollick (2014)) have discussed the potential of information transmission or herding following contributions. It could be possible that some of these effectiveness is being driven by large contribution providing a positive signal to the market, inducing more success by attracting others to the project.

The magnitude of these effects roughly coincides with the more naive assessment reported previously; when we eliminate the large contributions excess, and instead pretend it was a single regular contribution, we see about 2/3rds of projects would have failed without their large contributions. Of course, this does not control for any of the factors discussed previously, but it does agree in magnitude with the total effect reported by the IV estimates, giving us confidence that these results are reasonable in magnitude.

3.7.2 Interpretation and Robustness of IV estimates

We can also try to examine the magnitude of these coefficients by imagining (as discussed earlier) there are two kinds of backers: those interested in the material reward (fans) primarily, and those interested in the project's success (friends of the project). If you are interested in the reward, then a large contribution might be reasonable to make early on, since you have limited liability if the project fails. Otherwise, individuals most concerned with success tend to donate when they are most needed, generally occurring when the project is "pushed" over the edge of its goal. An alternative way to think about this is to imagine these "friends" are directly related to the project owner; then, if most projects are funding by a mixture of personal and crowdfunded capital, these large contributions are a way of relaxing the commitment to the funding goal by using more personal capital. This can be advantageous if the project-owner is unsure of their ability to raise funds from the crowd; however, there is a cost associated since the crowdfunding platform will take a fraction of the contribution, making this akin to a kind of self-insurance.[35] The IV approach should address this issue, as discussed earlier, but we can also analyse this directly by separating the indicator for a large contribution into two parts: before and after success. Looking just at large contributions which occurred prior to the project succeeding, we can repeat the previous analysis with this in place of the large contribution variable. The results are reported in Table 3.6.

As expected, the coefficient is more modest: large contributions prior to success lead to a 21% increase in the likelihood of success. On the other hand, the IV regression carried out before is still valid, and even more dramatic: a large contribution prior to success leads to a 70% higher chance of the project succeeding. This is consistent with our understanding of why these contributions would occur. For individuals who care about helping a project succeed, they seek to spend their money in a manner which is most effective. This means giving large sums to money to projects in such a way that it "pushes" over the cusp of success. This is, in general, not by directly filling the gap, but by closing the gap to such an extent that it becomes very likely that other backers will provide the remainder.

To confirm this intuition, I also include variables for large contributions both preceding and post-success. Now, due to endogeneity, we cannot meaningfully interpret the post-success coefficient, but this allows us to compare the baseline and preceding regression. We can see the results are, as expected, intermediate. The effect of the pre-success con-

[35] I thank Daniel Ershov in particular for this suggestion.

tributions remains much lower than the baseline, but the IV results are still accordingly high. This concurs with the interpretation that pre-success contributions are causal in the manner described: individuals want to "tip" projects over into the success. The low coefficient in the non-IV model is because this isn't always effective, and many projects are lost causes even with large contributions. The high coefficient in the baseline model is created by a combination of this factor (helping success) and individuals who are primarily self-interested claiming large rewards after the project has succeed.

	Model 3: Pre-success Coeff/Std. err.	Model 3: Pre-success (IV) Coeff/Std. err.	Model 3: Pre/Post Coeff/Std. err.	Model 3: Pre/Post (IV) Coeff/Std. err.
Average price (100s)	0.025***	0.012**	0.020***	0.013***
	(0.003)	(0.004)	(0.002)	(0.004)
Goal (1000s)	-0.121**	-0.112*	-0.107**	-0.101*
	(0.045)	(0.044)	(0.040)	(0.039)
Length	-0.008***	-0.009***	-0.008***	-0.009***
	(0.000)	(0.000)	(0.000)	(0.000)
Pre-success Large	0.208***	0.703***	0.266***	0.513***
	(0.004)	(0.135)	(0.004)	(0.113)
Total Comments (100s)	-0.003***	-0.002***	-0.002***	-0.002***
	(5.49e-06)	(4.94e-06)	(4.77e-06)	(4.71e-06)
Total Updates	0.031***	0.022***	0.025***	0.020***
	(0.001)	(0.003)	(0.001)	(0.002)
Total Backers (100s)	0.443***	0.444***	0.347***	0.336***
	(0.080)	(0.078)	(0.069)	(0.067)
Post-success Large			0.521***	0.581***
			(0.004)	(0.028)
Category/Time FE	X	X	X	X
Constant	X	X	X	X
N	95,792	95,792	95,792	95,792
Errors	robust	robust	robust	robust

Table 3.6: Robustness I: Pre/Post Success

One reason why these results may not be generally true for all projects has to do with the IV procedure. Specifically, we know that the coefficient being estimated is not necessarily the effect for the average project in the sample; instead, it is a local average, specifically for those projects which have large contributions influenced by the instrument chosen. As we recall, the instruments I chose were based around holidays which have to do with gift-giving or charity. The idea was to encourage "large" contributions by affiliating the results with days on which individuals feel more like giving large sums of money. Of course, not all individuals feel the same about all types of projects. Backers of a punk rock concert may feel substantially less moved by the onset of Lent[36] than backers of a Christian rock album. Similarly, tech toys and video games may be very appealing on Black Friday, but a performance art event may not. The coefficients measured have to do, specifically, with the pool of individuals (and projects they are interested in) which are affected by the instrument. Thus, the 70% increase in Table 3.6 reflects the marginal effect of large donations on projects for which giving or charitable holidays affect the backer pool. If we think, as is probably reasonable, that projects which have backers with react to charity are slightly more generous or interested in the outcome relative to a generic project, this estimate is overstated.

In order to examine this, I break out my regression into sub-groups, based on an intuitive assessment of which projects might be affected differently by the instrument. These groups are broken out by category, and are presented in Table 3.7. As we can see, different project types clearly react differently. The board games category appears to be strongly influenced by these kinds of support. This is typically because board games are a very niche, fan-oriented product, in which people care a great deal about the project succeeding. On the other hand, video games shows a non-significant but relatively tightly estimated zero. This indicates that these concerns are less relevant for supporters of video games. Fashion has a similar, but less tightly estimated coefficient. Documentary is interesting, this that while the coefficient is small, it shows wide variance despite a larger than normal sample size. This is potentially because projects in the documentary category tend to be long shorts, or fall into the category discussed. This makes the influence of the individuals attempting to help projects succeed less relevant, as it is diluted by individuals supporting "lost causes" or pet projects.

A similar issue could be related to unobserved variables, in particular the subjective

[36]Lent is the Christian religious month of penance and prayer, celebrated mainly by Catholics, Orthodox, and the more traditionalist members of the Anglican Communion.

"quality" of a project. The IV procedure used is robust to variations in project quality, as long as they do not have an impact on the holidays captured by the tenure. However, if we imagine that individuals who are motivated to give to a project are more likely to give to high quality projects, this could pose a problem. In particular, if "quality" is also related to the amount the project raises, it could be the case that high quality projects are both more influenced by the instrument (i.e. more likely to get a large contribution on a holiday) while also needing a lower size of large contribution to create success. These combinations of effects could overstate the impact of a large contribution, since it would mean the projects impacted the most by the IV are also projects which are relatively sensitive to large contributions. Unfortunately, this is difficult to test for explicitly, since "quality" is difficult to ascertain; it is difficult to infer from the covariates, and even if we assume it was capturable by the covariates ("no selection on unobservables"), approaches which try to control for unobserved heterogeneity, such as propensity score matching, fail to account for the reverse causality built into the standard suggestion. Indeed, propensity score estimates of the effect of a large contribution on success indicate a negligible effect; this is actually expected, since we know that large contributions are less likely after success, which means that projects near the success-failure boundary are likely to be influential in this comparison, but then these over-state the rate of failure for large contributions, once matched. In other words, the decision process behind large contributions is endogenous to the (expected) success or failure of a project and therefore serves as a kind of "levelling" process which makes the liminal projects very similar in terms of their success. At best this indicates that unobserved variables might drive some of the variation, but we have to believe that controlling for them fails in the IV procedure.

A further concern has to do with the second potential issue: the fact remains that the reduced form evidence shows that causality travels both directions: large contributions may lead to success, but success may also attract large contributions. As an alternative to the instrumental variables approach, we could consider this directly. Specifically, this implies that the model can be depicted as a system of equations; the probability of a project succeeding and the probability of large contributions can both be viewed a dependent variables. By excluding the number of backers and the comments and updates from the equation determining the likelihood of a large contribution (since these were insignificant and do not control for scale in the baseline model), we can use 3SLS to estimate jointly the probability of a project attracting a large donation, and succeeding. The coefficient on large donations is insignificant; large variation in the coefficient size is possible. The

	Board games	Documentary	Fashion	Video games
	Coeff/Std. err.	Coeff/Std. err.	Coeff/Std. err.	Coeff/Std. err.
Pre success lrg	1.53197*	0.1533902	0.8881502	0.362638
	(0.6236692)	(0.6084708)	(0.7551741)	(0.2613181)
Average price	-0.0001658	0.0002604	0.000205	0.0001396
	(0.0003977)	(0.0001631)	(0.0002982)	(0.0001174)
Goal ($10k)	-.0175291**	-0.0096762***	-0.0057097	-0.0055031***
	(.0064334)	(0.0023282)	(0.0037605)	(0.0009464)
Length	-0.0162298***	-0.006306*	-0.0078877***	-0.0047164***
	(.0031471)	(0.0028043)	(0.0014181)	(0.0008091)
Total comments	-0.0000239***	-0.0019844	0.001435	-3.50e-06
	(5.95e-06)	(.0013042)	(.0010702)	(5.39e-06)
Total updates	0.0169161***	0.0253838	0.028132	0.0234804***
	(0.001739)	(0.0177603)	(0.0240123)	(0.0070042)
Total backers	0.0168696***	0.0438697**	0.0092527*	0.0026646***
	(0.002733)	(0.0142683)	(0.0036724)	(0.0004791)
Constant	0.7003526***	0.5376474***	0.4152503***	0.2178438***
	(0.0890382)	(0.0513129)	(0.0501182)	(0.0468504)
N	2218	5,040	3,479	3,280
vce	robust	robust	robust	robust

Table 3.7: Robustness II - By Category

coefficient on success is 5%, with a 95% confidence interval between 4.9% and 5.2%, indicating that once we control for the simultaneous nature of the project, the effect is stronger than before. This is likely due to the fact that this method separates out "non-credible" large contributions from those actually determined by the success of the project. The main difficulty in evaluating this method is that the exclusion restriction is difficult to establish in this framework; we simply have too much noise to assess the coefficient value accurately. The relates to the propensity score problem earlier; exclusion restrictions are required for identification in both methods, which do not readily arise in either context, leaving the IV estimation as the most credible.

I also perform robustness checks to see whether or not outliers or skewness in the covariates might affect the results. The most skewed and influential variable is the number of backers. I perform several non-linear controls, both in absolute terms and scaled, to check whether or not this matter. I also perform robust regression to control for outliers. These are depicted in Panel (A) of Table 3.8. As we can see, the outliers do not seem to play a large role in the regression; the coefficients are similar to those in the baseline OLS specifications. Similarly, most of the non-linear controls provide similar coefficient signs and magnitudes to the IV estimates found earlier. However, I find that the specification seems sensitive to the log transform; a similar, but weaker result is found with the cube-root specification, although not to the same degree. This is somewhat puzzling since earlier studies (c.f. Kuppuswamy and Bayus (2013); Mollick (2014)) which compared the two specifications did not see a difference. A qualitative assessment of the distribution residuals also does not show large changes; they are relatively well-distributed in both cases. The main difference between this specification and the others is that it makes small and large values "more similar," effectively compressing the distribution of the covariates.

Motivated by this, I examine whether or not variation in the size of project (measured in different ways) has a heterogeneous effect on the IV coefficient; this is depicted in Panel (B) of Table 3.8. As we can see, the standard results are driven by projects slightly above the average number of backers; between 150 and 300; these positive and large significant coefficients in the baseline IV specification drive the results. A similar result is shown when we look at total funding. This implies that when we compress the number of total backers, we essentially make these groups harder to distinguish, washing out the results in the middle which drive the baseline result. In other words, what are detecting is actually heterogeneity on the part of the impact; projects which struggle to attract backers, or attract a very large number of backers are not as impacted by a large contribution. This is

likely because they are in the process of irrevocably failing, or are already nearly sure to succeed. This agrees with our model of how projects should be funded, but it does require use to bear in mind that not all projects are the same.

The major drawback of this assessment is that the specifications have much smaller sample size and are not precisely comparable, due to variable sample composition; this likely understates the problem, since the controls become more likely to be highly predictive as the sample size falls. The result for funding also explains why other literature might not see a difference; many papers trim off projects which are small (typically raising less than $5000) which appears to be influential for these results.

3.7.3 Policy Implications

The implications of large contributions primarily have to do with which interpretation of the motivations you adhere to. Since we cannot directly identify large contributors, this must remain a judgement call on the part of the policy-maker; however, it still gives clear direction regarding what must be taken into account to ensure crowdfunding works as intended. The most benign interpretation is that of the "fans" discussed previously. If large contributions are primarily driven by interested, uninvolved parties, the scope for malfeasance is limited. The main concern revolves around the role of asymmetric information. If consumers are not aware of the prominent role of large contributions, and use the goal (for reasons such as those discussed in Section 3.2) as an indication of the level of support necessary, they may end up being misled into believing a project needs more support (in terms of number of supporters) than it really does. If a policy maker believes that crowdfunding is useful because it helps to overcome limited information on the part of some consumers, this research indicates that large contributions complicate this role. Finally, a common concern regarding crowdfunding surrounds the sophistication of backers; if we believe these large contributing individuals are more sophisticated, or simply more careful, than the average person, the fact that they are so relevant for success can ameliorate these concerns.

This is compounded by the self-insurance role that friends or family might play. If large contributions are a form of limited commitment to the threshold, this again can have negative implications when asymmetric information is involved. The reason is similar, but is compounded by the fact that if the individuals making the contributions are involved in the project, then the goal conveys even less information than in the preceding case.

Panel (A)

	Robust (OLS) Coeff/Std. err.	Cube Root Coeff/Std. err.	Quadratic Coeff/Std. err.	Cubic Coeff/Std. err.	Log Coeff/Std. err.
Price (100s)	0.0593***	0.0175***	0.0132***	0.0136***	0.0144***
	(0.00105)	(0.00367)	(0.00348)	(0.0035)	(.0000316)
Goal	-0.0196***	-0.0001**	-0.0001**	-0.0001**	-0.0001*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Length (10s)	-0.0696***	-0.0673***	-0.0884***	-0.0880***	-0.0422***
	(0.0013)	(0.0046)	(0.0034)	(0.0034)	(0.0064)
Total comments	-0.0260***	-0.0038***	-0.0021***	-0.0016***	-0.0014***
(100s)	(1.51e-04)	(7.49e-04)	(4.27e-04)	(4.58e-04)	(3.37e-04)
Total updates	0.0146***	0.0048***	0.0183***	0.0173***	0.0027**
	(0.0003)	(0.0012)	(0.0026)	(0.0025)	(0.0009)
Total backers	0.1834***		0.0098***	0.0142***	
	(0.0003)		(0.0007)	(0.0010)	
Total backers^2	-0.0080***		-0.0000***	-0.0000***	
	(3.79e-07)		(1.25e-06)	(6.09e-06)	
Large cont.	0.2481***	0.1835	0.5220***	0.5130***	-0.1472
	(0.0033)	(0.1099)	(0.0918)	(0.0919)	(0.1314)
Total backers$^{1/3}$		0.4837***			
		(0.0383)			
Total backers^3				2.88e-08***	
				(6.36e-09)	
$\log$(total backers)					0.1963***
					(.0151268)
Constant	0.5819***	0.3791***	0.6827***	0.6823***	1.005***
	(0.0123)	(0.0262)	(0.0132)	(0.0132)	(.0276231)
Category Controls	X	X	X	X	X
Time Controls	X	X	X	X	X
N	95,792	95,792	95,792	95,792	95,792
VCE		robust	robust	robust	robust

Panel (B)

Backers (100s)	0-0.5	0.5-1	1-1.5	1.5-2	2-3	3-4	4-5	5-6	6+
Coefficient	-0.12	-0.19	0.09	0.59	0.38	-0.41	-0.10	0.33	-0.17
p-value	0.71	0.377	0.64	0.15†	0.24	0.44	0.63	0.63	0.86
Total funding ($10k)	0-0.5	0.5-1	1-2	2-3	3-4	4-5	5-6	6-7	7+
Coefficient	0.29	-0.02	0.07	0.34	0.94	-0.25	0.53	0.30	-0.45
p-value	0.15	0.93	0.7	0.19†	0.83	0.62	0.34	0.72	0.50

Table 3.8: Specification Robustness and Heterogeneity

Especially for very small contributions, this could lead to very misleading beliefs. It also widens the scope for moral hazard issues. Since the main route to address malfeasance on the part of a project owner is through a class action lawsuit, small projects are particularly vulnerable to creators who are not serious about carrying through with their rewards (as the class might be too narrow to profitably certify). Large contributions by involved parties provide a mechanism to fund projects which have no reasonable chance of reaching their goal otherwise, and which are unlikely to deliver their promised rewards.

A second consequence of large contributions has to do with the formalization role discussed in Agrawal et al. (2013). Formalization of previously informal funding arrangements implies some degree of external communication and bargaining; we could imagine this as an explicit equity arrangement. Crowdfunding, because of the all-or-nothing funding arrangement provides much more leverage for interested parties, especially if a project proceeds to go through several rounds of funding. Consider a project seeking to raise $1m, which is $50,000 short close to the deadline. An unscrupulous but interested investor could promise to give the $50,000 in exchange for favourable terms in a later round of equity funding. The nature of crowdfunding gives this offer disproportionately more impact than it would on its own, potentially making these projects vulnerable to predatory investing. From a welfare point of view this is not critical, but it does present a problem for regulators.

3.8 Conclusion

This chapter has shown some of the first evidence that the traditional narrative behind crowdfunding is not the complete story: projects are highly influenced by the presence of large contributors. Specifically, we see that projects are much more likely to be successful in reaching their crowdfunding goals when they are able to attract a large contribution. This is directly relevant for both individuals thinking about carrying out projects, researchers, and policy makers. First of all, if you are considering crowdfunding, consider your sources of funding. Is it likely you have some large supporters out there who will be able to assist you? Are they friends, family, or just individuals passionate about your product? If you cannot answer “yes” to this question, it is unlikely your project will have the resources to succeed. One suggestion indicated by this research is when using a mixture of personal and crowdfunded capital, reserve the personal capital for later in favour of a more aggressive

goal; you can use a large backer holding this personal money later to "rescue" your project if it is in distress.

For academic researchers, in addition to providing some understanding of the crowdfunding process, it also provides salient guidance for modelling. Most structural models of crowdfunding only look at the number of backers, assuming (implicitly or explicitly) that they all pay the same amount. This paper demonstrates that this is a poor assumption; large contributions appear to be at least equally important to understanding the success or failure of campaigns. The degree to which the failure to model this aspect of crowdfunding affects the structural estimates depends on the model under consideration, as well as the speculative reasons for large contributions; for example, in models such as Marwell (2016) an "insurance" interpretation of large contributions would complicate interpretation but not be critical to the analysis. However, this paper also suggests future ways to improve structural models, and other potential ways to build up these models into a more robust explanatory framework. This also suggests future work: analysing the causal determinants of large contributions is an obvious extension. It would also be useful to go behind the data, and link crowdfunding projects and their different pledge levels, to get a sense of how these large contributions are claimed.

Chapter 4

Sales Classification via Hidden Markov Models

4.1 Overview

What is a sale? In everyday life, we think of a sale as a kind of "deal" - some kind of temporary discount relative to a "regular" price. Sales are attractive to many consumers, since they indicate good value for money, and are frequently advertised. To prevent consumers from being misled, the intuitive notion of a sale is given strength by a precise legal definition in many jurisdictions. For example, in Canada, the Competition Act specifies that a sale may only be advertised if the offered price is a "bargain" relative to the ordinary price of the product[37]. However, it is not always straightforward for consumers or academic researchers to tell whether or not a given price is a sale. Frequently, we only observe the price of a given product, and not information about whether not that price is a sale. If these prices are observed for long periods of time, natural price adjustments due to supply-chain variations or changing market conditions may make it very difficult to infer which prices are sales, or otherwise. For instance, when is a price reduction truly a sale, rather than an adjustment in the face of increased competition? The question is rarely straightforward to answer.

The task of deciding whether or not a given price is a sale is called *classification*. For consumers, it is important since it helps them know whether they are getting a bargain or not. For researchers, it is important since sales are an integral part of the retail economy, and it is necessary to know which prices should be regarded as sales or not. However, this is not an easy task. The fundamental problem of sales classification amounts to the observation that prices change both in the short run and in the long run and this may cause confusion. If a retailer lowers their regular price for a product, an observer might

[37]Which can be defined by either duration or volume, as in the *Canadian Competition Act*, Sections 74.01, 74.04

be misled into thinking this new, lower, price is a sale - especially initially. Similarly, if inflation causes all prices to rise, a consumer might be mistaken into thinking that the new, higher sale price might be a regular price.

This is compounded by the fact that direct auditing of prices - say, by visiting a store and observing which products are advertised as "on sale" or not - is generally difficult to carry out. From a practical point of view, visiting a large number of stores on a weekly (or more frequent) basis and carefully collecting information on all the products in the store may be costly, if not impossible. Equally difficult is the fact that many firms deliberately obfuscate the different kinds of pricing in their stores. A given store may offer "sale prices," "everyday low prices," "cardholder prices," "manager specials," "buy large and save" offers to mention only a few of the many schemes available. Which of these are true sales? They are all advertised similarly, with brightly coloured tags calling out their offers, and many even see similar placement in newspaper circulars. This frenzied battle for the attention of the shopping public does little to aid researchers, who might very well make mistakes in judgement about what constitutes a true sale or not.

The solution is, of course, to let the numbers speak for themselves. Faced with the daunting task of trying to infer sales from incomplete and deliberately baffling information, researchers fall back on heuristic definitions based on the observed sequence of prices. These can be recorded accurately and repeatedly with ease at high frequency, generally through the use of retail scanner data (i.e. prices recorded by computer systems at the time of purchase). Since we have a natural definition of a sale as a "lower" price relative to some standard, the researcher's problem then becomes a question of how to make this definition practically useful. There have been many solutions proposed to this problem, almost as many as there are papers written studying sales. Some papers adopt a fixed standard as sale price (i.e. below $\$d$, as in Pesendorfer (2002) or Hendel and Nevo (2013)), while others choose a discount relative to some average standard (e.g. 25% below the mode, as in Berck et al. (2008) or Hosken and Reiffen (2004)). The largest scanner data provider, Nielsen-Kilts, suggests the method of 5% below the (local) average. My own work in Chapter 2 adopted a more flexible method, based on clustering and mixture models. Nonetheless, all of these methods are *ad hoc*: they rely on the researcher's understanding of the products, their prices, and their stability to choose a definition which is correct. Moreover, all of this must be carried out *before* any analysis: prices must be classified into sale and regular prices, after which reduced-form investigations can be carried out. This is typically without any acknowledgement of, let alone adjustment for, the uncertainty

inherent in such a classification.

This chapter develops an alternative model, which directly addresses many of the downsides existing methods of sales classification face. The way consumers typically resolve the fundamental problem of sales classification is by using more information: if a price is stable over a long period of time, it's probably the normal price. If the price goes up after a higher price then stays there, the first price observed was probably a sale. Consumers use information about the price sequence and its dynamics to classify sales. However, in academic research this is often flipped on its head: the rationale is that first one determines what is, or is not, a sale, and then uses that determination to uncover interesting correlations or relationships with other data (like sales dynamics). However, then *by definition* we are not using all of the information available when we perform the classification step! In the standard method, classification is a one-way street, showing relationships between sales (and their dynamics) and covariates but never using them in the determination of sales themselves. This oversight is largely ignored by the literature on sales, mainly because it is innate in the structure of classification itself: most methods result in a clean yes/no cut-off (entering into estimation as a binary indicator variable based on a stark cut-off). If we define a sale as $1.00 less than the modal price, then what if a product is $0.99 off? Why is a dollar discount definitely a sale, while something sold at only a penny more is not? Why don't we include information like the fact that the second-to-last price was also $0.99 more? Although methods that make this kind of choice are backed by careful consideration of the price distribution, there is something unsatisfying about this kind of rule. A better method should help a researcher quantify this statistical uncertainty, either using it making a judgement call about what is a sale or not (e.g. more than 90% likely, as in Chapter 2), or better still incorporating it directly into the study. It should also use the dynamics of prices explicitly, as our intuition from consumer behaviour implies.

This chapter develops such a method, using the hidden Markov model framework to describe the way sales function. At their most basic, hidden Markov models are descriptions of structured but random processes where the observed outcomes are determined indirectly by a hidden state. The key hidden state in the study of sales is the central piece of missing information: is this product on sale at a given time? By considering the observed sequence of prices as a result of an *unobserved* sequence of sales, we can use this model to begin to piece together the driving forces behind price variation. The power of the hidden Markov framework arises through the combination of inference about the unobserved

state of the world and its connection to the observed reality. Hidden Markov models capture closely the intuitive mechanism through which sales become observed prices, resulting in a powerful technique for classification. To overcome the fundamental problem, I model the sale and regular prices in the manner originally developed in Chapter 2: as a changing set of regimes. So, rather than facing a single pair of prices, the researcher regards the observed data as being a sequence of these pairs, changing over time. This long-run pricing evolution can be captured in the same manner as the shorter term sale variation, provided that we impose some conditions on the underlying process governing the states.

To be precise, in this chapter we will adapt the definitions from Chapter 2. Consider a representative product being sold at a single grocery store over time. There is a fundamental price $\hat{Y}_t$ for each (discrete) time period $t = 1, 2, ..., T$. We imagine at each period, the grocery store has several different options to choose for their price, $\hat{Y}_t \in P$. These prices correspond to the pricing strategy the grocery store has settled on using at that time period; in any case, the set P is the pricing set, and is defined as follows:

Definition 4.1.1. A *pricing set* denoted P is a pair of prices which consist of (1) a regular price p and (2) a discount from the regular price $\delta > 0$. The associated sale price is $s \equiv p - \delta$. Similarly, a *sale* is defined as the event that $\hat{Y}_t = s$ and is written $S_t = 1$. The fundamental price can then be written as $\hat{Y}_t = p - S_t\delta$. The set of all pricing sets is $\mathbf{P}$.

These pricing sets are the options individual stores have to choose from; a pricing *regime* is the particular set that a store has adopted, which persists for a contiguous time. Since this is not directly observable, this is modelled as a state variable which we index with the integers; we represent this indexing with a function $R : \mathbb{Z}^+ \to \mathbf{P}$ which maps the integers into the set of all pricing sets.

Definition 4.1.2. A *pricing regime*, denoted R_t, is a state variable which indexes the available pricing set for a store at time t and lasts for a contiguous period of time.[38] That is, if $R_t = z$ then $P = R(z)$ is the available set of prices. We denote the associated prices and events by association with R_t: $p(R_t) > 0$, $\delta(R_t) > 0$, $s(R_t) \equiv p(R_t) - \delta(R_t)$. If we define S_t to be the event that a sale occurs, the observed price can be written as $\hat{Y}_t \equiv p(R_t) - \delta(R_t)S_t$.

This is a simplification of the definition from Chapter 2 for the case where there is only a single discount level being offered, as most sales environments feature. The extension of the methods developed in this chapter to more discounts is straightforward. Notice,

[38] That is to say if $R_t = R$ and $R_{t+k} = R$ then for all $s \in [t, t+k]$, $R_s = R$

also, that this specifies the *fundamental* price, not necessarily the observed price. Most environments with sales will feature noise in the observed price sequence; for white noise, the interpretation is that $\hat{Y}_t$ is the mean of the distribution of prices at a given period.

As I demonstrate, the correct way to model this kind of a process is through the use of block left-to-right Markov models, a term which we will define precisely in Section 4.3. These kind of Markov models are typically used to model processes which are one directional, like sequences of words in human speech, or developmental processes over time, of which sales are a good example. In sale pricing, there is both short-run and long-run variation; the short-run variation is the sale, while long-run is how sales change over time. The blocks correspond to periods of sales with particular prices, and the movement between blocks corresponds to the long-run variation. These kinds of models violate the basic assumptions for hidden Markov models, as typically used. Accordingly, I develop explicit identification conditions for these models which rely on cross-sectional variation. Next, I show that the basic description of sales as a hidden Markov model falls neatly within the general framework developed. I also illustrate that this can neatly incorporate information about multiple covariates or other variables of interest. This fact not only incorporates more information into the classification of sales, but also allows researchers to explicitly test correlations between sales and variables of interest at the classification step. This removes the necessity for reduced-form analysis, and provides a powerful tool for the evaluation of competing models of sales. I then show how even complex dependencies between sales over time can be modelled using higher-order Markov models, which again fall within the framework under consideration.

I evaluate the performance of my technique on a variety of Monte Carlo simulated scenarios, ranging from simple to complex. I simulate sequences of prices including sales, then apply the model to the simulated data to evaluate performance: because we know the true generating process, we can easily evaluate how well the model recovers the hidden states (sales). The model generally performs very well on both dimensions of interest. First, the estimates of the relationship between sales and observed outcomes are accurate, allowing for correct inference about the true relationships between sales, prices, and other covariates. More importantly, the classification remains highly accurate, even in situations designed to be difficult to make inferences. This implies that this method can be used to classify sequences of observations into sale and regular prices in a number of ways which will be generally correct. This classification method is highly robust and powerful, presenting itself as a natural alternative to other *ad hoc* alternatives. Finally, I investigate

the small-scale properties of the model, in situations where the cross-sectional nature of the data is dubious. I find that the model, provided it is correctly specified and well-instantiated, remains accurate. This demonstrates its utility even for problems where a great deal of data is not available.

The remainder of this chapter is organized as follows: in Section 4.2, I provide some background on Markov chains and hidden Markov models which serves as a (largely) self-contained introduction to the general setting. Section 4.3 extends these results to block left-to-right models, and then demonstrates how sales can be modelled using this framework. Section 4.4 provides evidence of the performance of the method using simulations, and investigates the small-scale properties of the model. Finally, Section 4.5 concludes. Some selected proofs are reserved to Appendix C.

4.2 Background on Hidden Markov Models

In this section, I begin with a general background on hidden Markov models and Markov chains. We will connect these to sales explicitly in Section 4.3; this section serves to define terminology, notation, and as an introduction to the area for those unfamiliar with the subject matter. A hidden Markov model (HMM) is statistical method designed for analysing situations where the observed outcomes depend on an underlying state of the world which is itself not observable. For example, perhaps an economist is trying to understand how their research output varies with their innate productivity. They know that on some days they are in a "productive" mood, while on other days their mood is unproductive. However, they (perhaps due to a lack of objective introspection) cannot observe which mood they're in on a given day, only some measure of their productivity - say, number of words written. Productive days naturally agree with large numbers of words, while unproductive days correspond with smaller amounts written. The economist is interested in two questions. First, how does the number of words written vary with the unobserved mood? For example, what is the average (or expected) number of words they write on a productive day. Second, how do moods evolve or change over time? This would let them understand not only the dynamics of their moods, but also how long they expect to be in a productive mood in a given length of time. Understanding these two questions together gives them a good understanding of their research productivity, and would help them do things like plan when to take holidays (during periods of expected low productivity, for example), and

forecast their research output in the future. In the context of sales, the economist observes a sequence of prices over time. These prices are driven by an underlying pair of states: whether or not there is a sale being offered, and which pricing regime is currently active. As the underlying state changes between sales and non-sale periods, the price rises and falls. Similarly, the level of these prices also changes as the active regime moves over time. Sales classification is the task of trying to recover the underlying state (sale or non-sale) from only the observed information.

As the examples show, hidden Markov models are particularly useful in trying to analyse behaviour in which the underlying, hidden, state is important for a researcher. The researcher would like to understand not only the relationship between the state and the observed variables, but would also like to understand the hidden state itself. These two sources of randomness are why HMMs are a type of *doubly stochastic processes*; the observations are random and the randomness is itself driven by a random process. This doubly-random feature also forms the key complication which makes them different from a traditional model which could be solved by maximum-likelihood: if the hidden state were known, standard techniques would suffice, a fact exploited by Baum et al. (1970). Usually, the way we would solve this would be to "average out" the unobserved state[39]. However, as we will see, this is not feasible in this environment, and would make it difficult to infer the hidden state. This interest in the underlying state itself also means we are considering a specific use of Markov models, where we are interested in the model parameters and structure itself, not merely as a means of forecasting or data mining a process (as summarized in Cappé et al. (2005)) which differs from many applications (particularly in this area).

First explicitly studied by Baum and Petrie (1966), Hidden Markov models form part of a rich set of related processes, including Markov-switching models, Markov-jump processes, among only a few (see Frühwirth-Schnatter (2006); Creal (2012) and Cappé et al. (2005) for examples).

The key assumption which makes hidden Markov models particularly tractable is that underlying state changes in a manner which depends only on its current value, and not on its entire history. This type of process is called *Markovian*: in our setting here, we will assume that it that takes a specific form called a *Markov chain*. In a hidden Markov model, the Markov chain governs the behaviour of the hidden states, while the other parts

[39]For example, if the probability of an observation Y depends on some unobserved X with distribution $g(x)$, we can calculate $P(Y) = \int_X P(Y|s)g(s)ds$, averaging out the X variable

of model govern the way these hidden states link to the observed variables. In our sales context, the Markov chain would define which regime is active and whether there is a sale or not, while the rest of the model would determine how this translates into the observed prices (and other covariates). Therefore, it is useful to carefully define a Markov chain, and some related terms, as they will be used in this chapter, since they are central to the idea of a hidden Markov model. It is important, to note, however, that there are many generalizations and relaxations of the basic structure given here; I will point out particularly useful ones as necessary.

Intuitively, a Markov chain is a sequence of random variables which change between a fixed number of values or states. The next value of the chain depends only on the current value of the chain, and not its entire history. The analogy of a "chain" is used to highlight the fact that each period is "linked" to others only through the single connections formed by the immediate past and future. In the context of sales, the chain models the "true" state of the world; whether or not in a given period the retailer is holding a sale or not and which regime is active. We can mathematically express this definition of a Markov chain, as used in this chapter, as follows:

Definition 4.2.1. A *Markov chain* is a discrete-time stochastic process $\{X_t\}$ which takes on k values in a finite set S, and evolves such that there is a fixed transition probability a_{ij} such that for any time period $t \geq 0$, and states $j, i_0, i_1, \ldots, i_{t-1}$ in S that

$$P(X_{t+1} = i | X_t = j, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = P(X_{t+1} = i | X_t = j) = a_{ij}$$

That is, the conditional distribution of a future state given a history of states depends only on the current state. The value a_{ij} is the probability of the chain moving from state j to i, known as the *transition probability*.

The fact that the transition probabilities do not change over time makes this a *time-homogeneous* Markov chain, as opposed to a situation in which the underlying chain changes over time. It is also worth noting that this is a finite, discrete time Markov chain, in contrast to one which exists in continuous time or has infinite states. The finite, discrete, nature of the states, and the movement of the chain in time, gives a convenient representation of the evolution of the Markov chain in a transition matrix, formed by collecting all of the transition probabilities:

Definition 4.2.2. Suppose X_t is a Markov chain. Then, the k-by-k matrix containing the transition probabilities for all i, j in S

$$A = [a_{ij}]_{i,j \in S}$$

is called the (one-step) *transition matrix* for the Markov chain.

This is also why the states of a Markov chain are typically labelled $1, 2, \ldots, k$ to coincide with the rows and columns of the transition matrix. In many applications these states may be have real meaning, while in others they may simply be a data artefact; akin to the meaningfulness of clusters in cluster estimation (see Cappé et al. (2005) chapter 1 for a discussion). This kind of process is natural in many economic environments, such as consumer and firm decision making. For example, many learning processes such as Banerjee (1992) turn out to be Markovian, in that the history can be summarized in such a way that it only enters through the current state of the world. Implicitly, many models of consumer decision making are similarly structured, in that consumers base their decisions only on the current state of their finances and not on their past consumption choices.

The property that the future state of the Markov chain only depends on the previous state is an assumption; this kind of Markov chain is referred to as *first-order* Markov chain. Higher order chains, where the future state depends on more past values are also possible. However, attention is usually restricted only to first-order chains, since by a suitable re-definition, higher order Markov chains can be transformed into first-order chains in many applications. For example, in the context of sales, we may be interested in sales which are periodic and become more likely after a certain number of periods. This will be discussed in more detail in Section 4.3.3.

We can also introduce some notation for the state of a Markov chain. The distribution of states in a Markov chain evolves over time, according to the transition matrix. This is defined as follows:

Definition 4.2.3. The *state distribution* in a given Markov chain at time t is denoted π_t, where $\sum_k \pi_t(k) = 1$. In particular, the initial distribution is denoted $\pi_1 = \pi$.

Notice that definition, combined with the transition matrix, gives a convenient expression for the dynamics of the states of a Markov chain:

$$\pi_{t+1} = A\pi_t$$

This allows for easy calculation of distributions at arbitrary periods into the future, since by the Markovian condition, $\pi_{t+k} = A^k \pi_t$. This summarizes the basic facts about a Markov chain necessary to introduce a hidden Markov model. Markov chains, by themselves, form an interesting and complex field of study, with many interesting features. I have focused only on the facts and properties necessary for this chapter; for more details, interested readers are referred to Ross (2014), Ching and Ng (2006), or Freedman (2012).

4.2.1 Notation and Basics

We know begin our development of a general hidden Markov model, using the machinery built to understand hidden Markov chains in the previous section. At its most basic level, a hidden Markov model is a series of realizations of a pair of random variables $\{X_t, Y_t\}$ over a period of time $t = 1, 2, \ldots, T$. However, only the variable Y_t is observed at each period, resulting a single series of observations $\{y_1, y_2, \ldots, y_T\}$. In what follows here, I will use the generally accepted notation of Rabiner (1989) and Cappé et al. (2005), where appropriate.

As explained earlier, the connection between Y_t and X_t in a hidden Markov model is that the underlying variable $\{X_t\}$ is a Markov chain which evolves over time according to its transition matrix A, and it produces an observable output in the form of Y_t. This leads Y_t to often be referred to as the *emission* from the underlying state space. This process is sometimes depicted in graphical form as shown in Figure 4.1. In the context of sales, the underlying state of the world determines where or not there is a sale (X_t), while the Markov model determines how this translates into observed prices (Y_t). The mathematical expression of this notion is summarized in the following two assumptions, one on the observed variable and one on the hidden variable:

Assumption 4.2.1. $\{Y_t\}_{t=1}^T$, conditional on $\{X_t\}_{t=1}^T$ is a sequence of independent random variables where the conditional distribution of Y_t depends only on X_t (the current state).

Assumption 4.2.2. The sequence of random variables $\{X_t\}$ forms a Markov chain (with transition matrix A).

In order to give the emissions Y_t some structure, we assume that they come from a known family of distributions and depend on the hidden variable X_t. Each state of X_t selects a member from the family of distributions, which then results in the emission.

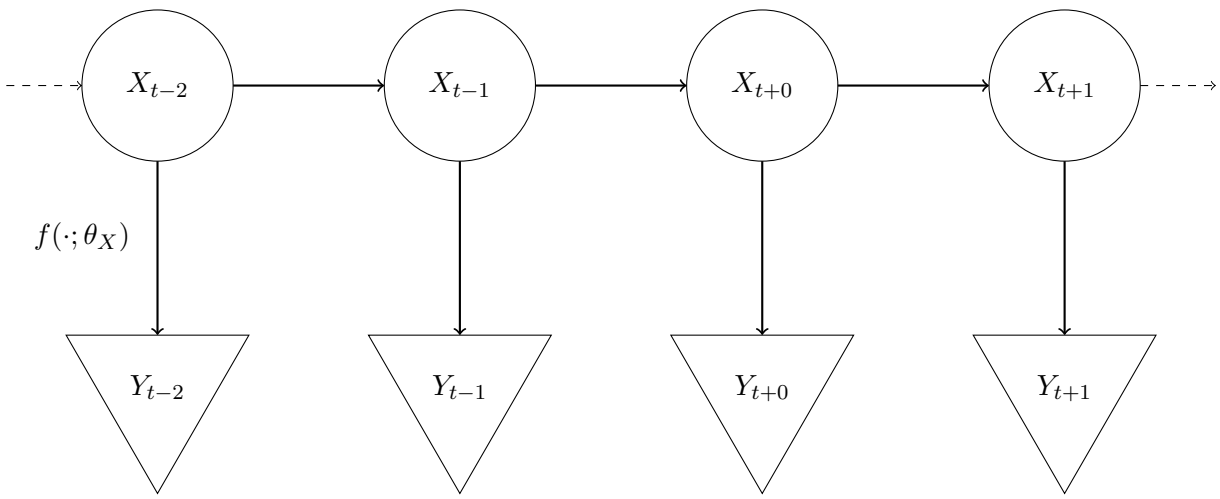


Figure 4.1: A graphical depiction of a first-order hidden Markov model

In this chapter, I consider real-valued continuous random variables which come from a parametric family. We can make this definition precise in the following assumption:

Assumption 4.2.3. Let $\{f(\cdot;\theta)|\theta \in \Theta\}$ be a family of density functions on the real line, parametrized by θ and $\{\theta_1, \theta_2, \ldots, \theta_k\} \subseteq \Theta$. Then, the distribution of Y_t conditional on X_t is given by $f(\cdot;\theta_{X_t})$.

In our new notation, we can write that $\{Y_t\}_{t=1}^T$ is generated as a sequence of draws from $\{f(\cdot;\theta_{X_t})\}_{t=1}^T$ where the sequence of distributions is given by the realization of $\{X_t\}_{t=1}^T$. Each state selects a member of the parametric family θ_{X_t}, which in turn produces a particular emission for that state. In many applications (including sales in situations with very stark pricing), these variables need not be continuous. The study of discrete-valued emissions (known as "symbols") is one of the major applications of hidden Markov models in computer science and signals processing (Callander (2007); Creal (2012); Frühwirth-Schnatter (2006)). For example, in sales, a sale period may translate into a distribution with a low average price, while a non-sale period will have a higher average price.

With the basic ideas pinned down, we can now define a hidden Markov model as a pair of random variables X_t and Y_t which meet the assumptions laid out above. In mathematical terms, I define a hidden Markov model as follows:

Definition 4.2.4. A *hidden Markov model* is a sequence of random variables $\{X_t, Y_t\}$ over a period of time $t = 1, 2, \ldots, T$, in which only $\{Y_t\}_{t=1}^T$ is observed, and which meet the

conditions of Assumptions 4.2.1, 4.2.2, and 4.2.3. The parameters associated with the model are denoted $\Psi \equiv (A, \{\theta_k\}_{k=1}^K, \pi)$

From an applied point of view, if we wish to study a phenomenon using a hidden Markov model, the estimation of the model parameters Ψ is critical, since they fully characterize the HMM. Rabiner (1989) introduces three basic problems for analysing a HMM:

1. Given an observation sequence $y \equiv \{y_1, y_2, \ldots, y_T\}$ and a set of parameters Ψ, can we calculate $P(y|\Psi)$?

2. Given an observation sequence y and a set of parameters Ψ, what is the distribution of hidden states $\{X_1, X_2, \ldots, X_T\}$ which produces the observed data?

3. Given an observation sequence y and a calculation of $P(y|\Psi)$, how do we adjust Ψ to maximize $P(y|\Psi)$?

The first and third problems basically amount to the standard question of maximum likelihood: given a set of observed data, what is the probability of the data given the model, and how do we maximize that probability? The second problem is particular to the hidden nature of a HMM. Inferring the most likely sequence of hidden states is important, first for calculation of the likelihood, but also because in many applications the states are a key object of study. For example, in sales the first and third problem amount to correctly specifying the relationship between sales and prices. The second problem is recovering which periods are sales or not. Unfortunately, direct calculation of the likelihood function (by summing over the unobserved hidden states) involves on the order of $T \cdot K^T$ elementary operations (Maruotti (2007))[40].

To overcome this difficulty, the key innovation in hidden Markov models, developed by Baum et al. (1970), is the *forward-backward* or Baum-Welch algorithm. Essentially, this method calculates the likelihood function in a convenient, inductive manner which also facilitates other calculations. This is explained in Rabiner (1989), specifically for discrete symbols, and I will illustrate it here for completeness in our environment.

We can do this by introducing some new variables which will help us calculate the probability of observed sequences of emissions, and their hidden states. The first variable

[40]That is, for even a small HMM with 4 states observed for only 30 periods, there would be 3.4×10^{19} additions and multiplications required for *each* evaluation of the likelihood function

to be considered is $\alpha_t(k)$, which is defined the probability of the sequence of observations up to time t, if the state at time t is k. Precisely:

$$\alpha_t(k) \equiv P(y_1, y_2, \ldots, y_t, X_t = k|\Psi) \tag{4.2.1}$$

This term can is particularly useful because it can be defined inductively. Intuitively, the first time period, the probability of seeing a value y_1 conditional on being in state k is just the probability of being in state k (given by π, the initial distribution) and the emission distribution created by k. From here, we can simply use the transition matrix, the probabilities from the previous step, and the emission distributions, to iterate forward until time T. That is:

$$\begin{array}{rcl} \text{Basis:} & \alpha_1(k) = \pi(k) f(y_1; \theta_k) & 1 \leq k \leq K \\ \text{Induction:} & \alpha_{t+1}(j) = \left[\sum_{k=1}^{K} \alpha_t(k) a_{kj}\right] f(y_{t+1}; \theta_j) & 1 \leq j \leq K \end{array}$$

The process ends with period T, at which point all of the observations have been exhausted, and the induction step gives the expression for the likelihood of the entire sequence of observations by summing over the possible states:

$$P(y|\Psi) = \sum_{k=1}^{K} \alpha_T(k) \tag{4.2.2}$$

This resolves the first of the questions given above, yielding a simple way to calculate the likelihood function. In order to answer the next question, about the hidden states, we need to define a counterpart to the forward probability α. This "backward" term asks the opposite question: suppose we are in state k at time t. What is the probability of the future observations, conditional on this fact. Specifically, it we define this as:

$$\beta_t(k) \equiv P(y_{t+1}, y_{t+2}, \ldots, y_T|X_t = k, \Psi) \tag{4.2.3}$$

Again, this term is useful due an inductive calculation. We start by *defining* the term at time T as one, since at the last period there is no future states (and so any future path has probabilities 1). We can then iterate backwards in the same manner as we did for α, using

the transition matrix and the future probabilities. This can be carried out as follows:

$$\begin{array}{rcl} \text{Basis:} & \beta_T(k) = 1 & 1 \leq k \leq K \\ \text{Induction:} & \beta_t(k) = \sum_{j=1}^{K} a_{kj} f(y_{t+1}; \theta_j) \beta_{t+1}(j) & 1 \leq j \leq K \end{array}$$

The key use of β is not in calculating the likelihood function, but rather to infer the hidden states in the model. The probability of being in a given state, given the observations, is the most natural way of thinking about this. We call this distribution γ, defined as:

$$\gamma_t(k) = P(X_t = k|y, \Psi) \tag{4.2.4}$$

That is, γ_t is the likelihood of being in a given state in period t, given the observed data and the parameters of the model. Then, by Bayes' rule, we can calculate that:

$$\gamma_t(k) = \frac{\alpha_t(k)\beta_t(k)}{\sum_{i=1}^{K} \alpha_t(i)\beta_t(i)} \tag{4.2.5}$$

This allows for the recovery of the underlying states. The related question of what is the *most likely* set of states can be calculated in several different ways, as discussed in Rabiner (1989) or Cappé et al. (2005). The estimation of the transition matrix can then be recovered from the distribution of γ, by counting up the probability of transitions of states. All of these methods generalize easily for multiple sequences of independently drawn Markov models, as explained in Cappé et al. (2005). Essentially, all of the above terms (α, β, γ)can be calculated sequence-by-sequence, and the aggregate likelihoods calculated as the products of the individual components. Transition matrices can be re-estimated by summing over all transitions for each sequence.

The final problem, that of actually maximizing the likelihood functions calculated above, is more difficult. Essentially any method desired can be used, since given the above discussion, the underlying calculation of the likelihood function is straightforward. In most applications, a version of the EM algorithm is used, which is guaranteed to find a local maximization of the likelihood function (See Bilmes et al. (1998) for a clear illustration of this procedure, or McLachlan and Peel (2004)). Along with proper initialization of the Markov chain, this will coincide with the true optimal set of parameters. The EM algorithm, in the

context of a hidden Markov model is also called the Baum-Welch algorithm, in light of the specific innovations necessary to make computation feasible in this environment.

The existence and uniqueness of a global solution to this problem is a related and important question: that of identification. Unfortunately, because much of the research in this area comes from a practical (engineering, computer science) point of view, many sources do not explicitly make it clear what conditions are required. For most basic applications of a hidden Markov model, researchers are concerned with a single sequence of observations, and the limiting results deal with the process as $T \to \infty$. That is, we observe the Markov chain over an arbitrarily long period of time.

The basic requirements placed on such a process concern the two key elements of the model: the Markov chain and the emissions. We can think about this as follows: in order to determine the transitions and other elements of the Markov chain, we need to observe (in the population) a large number of transitions between states. We wouldn't want a chain to get "stuck" in only a small subset of the states, since then we couldn't learn about the other states. In other words, we want a chain which, as time goes on, visits every state an infinite number of times. We also want to be able to tell the emissions apart; if we're not sure about the state, the observed emission will be a combination of our best estimates of the emissions associated with the hidden states. We want to make sure that these emissions can be teased apart statistically. Mathematically, the way to express these are as follows:

Assumption 4.2.4. The Markov chain X_t is ergodic

Assumption 4.2.5. The distribution formed by a mixture of at most K components of $\{f(\cdot;\theta)|\theta \in \Theta\}$ is identifiable (up to relabelling of the components).

Assumption 4.2.4 for a discrete, finite Markov chain has a convenient representation. Specifically, it is sufficient that two conditions are met: (1) that the Markov chain is irreducible, i.e. that it is possible to reach any state from any state (including itself) and (2) that there is some number N such that it is possible to reach a given state in at most N steps from any other state. In particular, this means that the Markov chain must eventually return to any given state an infinite number of times as $t \to \infty$, as we intuitively desired. Assumption 4.2.5 highlights the close connection between hidden Markov models and finite mixture models (see McLachlan and Peel (2004) for a discussion of these models). As we discussed, in a given period, the probability of a given y_t is a mixture of the $f(\cdot;\theta_k)$ distributions, weighted by γ_t, the distribution over hidden states. Since γ_t itself is a combination of A and the values of θ_k, the assumption of identifiability (up to relabelling), this

implies that the mixture parameters and the weights are recoverable from the observed distribution over time. The structure of the Markov model itself allows for a natural labelling, since the underlying states have meaning and transit in a particular pattern. This allows for both A and θ_k to be recovered, provided that all states are reached; something which follows from Assumption 4.2.4.

This also highlights why it is necessary to look specifically at sales: the assumption of ergodicity is not reasonable for most sale situations. In particular, the long-run variation caused by regime changes means that it is not possible to to return to a past regime from a future one. This means there isn't enough information available to calculate transition probabilities accurately, since each regime transition happens exactly once. In addition, we typically do not imagine that $T \to \infty$ in the study of sales, since the time dimension is usually small. These limitations necessitate an extension in order to use hidden Markov models for sales classification.

Identifiability of a mixture distribution seems like a tall order, but fortunately, there are many families of density functions which meet these conditions. As cited in McLachlan and Peel (2004), Titterington (1985) points out that most finite mixtures for continuous density functions are identifiable in the above sense. One particularly useful category is that of multivariate Gaussian density functions, which not only are identifiable but also are particularly computationally tractable, leading to their natural adoption in many applications, especially where no explicit alternative is obvious.

This summarizes the basic material needed for the following study of the role of hidden Markov models in sales. Next, I show how a generalization of the basic hidden Markov model framework has can solve our problem with ergodicity as a necessary condition, and has natural role in classifying and understanding sales. This requires the development of some conditions for more complicated hidden Markov models, with special attention paid to the role of identification on this extended environment.

4.3 Applications of Hidden Markov Models to Sales

Having established the necessary background and notation, we can now begin to specify our environment to its use in the classification of sales. As we have discussed earlier in this chapter, and in Chapter 2, the typical way sales are studied in industrial organization is a two-stage process. First, observations are classified into sales, then the classified data is

used to perform reduced-form analysis. Unfortunately, most classification techniques focus on either observation-by-observation classification, or small windows about a given observation rather than the entire sequence of prices. They also typically do not use information provided by variables other than the price itself. Many are also unable to provide information about the uncertainty of a given classifier. This means that in many applications we may be unsure of how a classifier performs, or under what conditions it is accurate. As I note in Section 2.1, the classification method suggested by Nielsen-Kilts itself, for their own dataset, does not perform well in many applications. In this section, I demonstrate how hidden Markov models form a natural way of classifying sales which overcomes these difficulties, and also can allow researcher to directly inspect features of interest without the need for an additional reduced form analysis step.

This technique allows the researcher to explicitly capture sales dynamics in the classification itself; either at a lower level, to provide a more "robust" version of the classification technique developed in section 2.3, or at a more structural level to avoid reduced-form inference entirely if the problem is suitable. The drawback of this method is that it involves significantly more computational and analytical overhead than other alternatives. As discussed in Section 4.2, finding an initial starting point for the parameter estimation in a hidden Markov model is very important, since only local maxima are guaranteed by the Baum-Welch algorithm. While the literature suggests general heuristics for finding such a point, in a specific application like this, best practice suggests that a basic classification method must to be performed *first* in order to provide a robust starting point for estimation. Additionally, the technique developed is multiplicatively more cumbersome to use as the number of lagged variables within the model increases, as we will discuss in Section 4.3.3.

In this environment, consider a representative product being sold at a single grocery store over time. We observe the price Y_t and a set of covariates W_t which may vary for each time period $t = 1, 2, ..., T$. I assume that the underlying data-generating process consists of R pricing regimes, $R_t \in \{1, 2, 3, ..., R\} \equiv R$. These reflect the underlying channel prices being offered to the grocery store over time for the product; they may rise, or fall, based on market conditions and relationships with suppliers. However, in particular, I assume that these trends occur at a level "higher" than the individual store; the regime changes are independent of any individual store's sales or performance. I also normalize the regimes such that they occur in order; at time $t = 1$, $R_t = 1$ and if $R_s > R_{s'}$ then $s > s'$.

This is a natural environment for a hidden Markov model. We have an underlying, hidden state of the world, which manifests in observable outcomes Y_t and W_t that are

generated through some process which is not directly observed. If we take this seriously, and suppose the probability of shifting from regime i to j follows a Markov chain with rate q_{ij}, this means that the transition matrix takes the block-diagonal form:

$$A_r = \begin{bmatrix} q_{11} & 1-q_{11} & 0 & \dots & 0 & 0 \\ 0 & q_{22} & 1-q_{22} & \dots & 0 & 0 \\ 0 & 0 & q_{33} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 1-q_{R-2,R-1} & 0 \\ 0 & 0 & 0 & 0 & p_{R-1,R-1} & 1-p_{R-1,R-1} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \tag{4.3.1}$$

However, we have not yet introduced sales. As defined in the introduction, I assume that a regime, in the context of this model, will consist of a pair of prices: one "regular" price $\mu_r \equiv p(r)$ and one potential discount $\delta_r \equiv \delta(r)$. This results in two prices: a regular and sale price. The discount will occur and the sale price change when a product intermittently undergoes a sale.

Given the Markov structure of the pricing regime, we can incorporate sales by considering them as a higher frequency variable layered on top of the slowly changing regime. For exposition, suppose that lagged variables are not important here (we will return to this topic in Section 4.3.3). Then, we can consider the joint movement of regimes and sales as a large Markov chain by assigning to each period a unique state consisting of two elements $X_t = (R_t, S_t)$ where S_t is the hidden variable indicating the presence of a sale in period t. In total there are $2R$ such states. The first would be $(1, 0)$ consisting of regime 1 without a sale. The second would be $(1, 1)$ which consists of regime 1 with a sale, and so forth. This has an associated transition matrix with some restrictions on values, since not all states can

be reached from each value. This is a block diagonal matrix, as depicted in equation 4.3.2:

$$A \equiv \begin{array}{c|ccccccc} & (1,0) & (1,1) & (2,0) & (2,1) & \dots & (R,0) & (R,1) \\ \hline (1,0) & \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} & \dots & 0 & 0 \\ (1,1) & \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} & \dots & 0 & 0 \\ (2,0) & 0 & 0 & \alpha_{33} & \alpha_{34} & \dots & 0 & 0 \\ (2,1) & 0 & 0 & \alpha_{43} & \alpha_{44} & \dots & 0 & 0 \\ \vdots & & & & & & & \vdots \\ (R,0) & 0 & 0 & 0 & 0 & \dots & \alpha_{R-1,R-1} & \alpha_{R-1,R} \\ (R,1) & 0 & 0 & 0 & 0 & \dots & \alpha_{R,R-1} & \alpha_{R,R} \end{array} \tag{4.3.2}$$

Then, with this chain defined, we can specify the price in state $X_t = (R_t, S_t)$ as:

$$Y_t(X_t) = \mu_{R_t} + \delta_{R_t} S_t + \epsilon_t \tag{4.3.3}$$

where ϵ_t is a noise term. Essentially, the model assumes that prices are generated from a hidden Markov model with emissions governed by the distribution of the noise term. The objects of key interest are the parameters of this model, and the distribution of likely states for each observation γ_t.

However, by inspection we can see the regime Markov chain given by Equation 4.3.1 is "left-to-right" (or Bakis, see Rabiner (1989) or Cappé et al. (2005)); that is, we move between regimes at most once. This is generally the case with state space models where the transitions model an evolving process. However, this is a problem: how can one correctly recover the transition probabilities if we only observe them once? Standard hidden Markov models make the assumption that the underlying Markov chain is ergodic (as discussed in Section 4.2), which implies that the long-run distribution of the states in the model matches the stationary distribution. In particular, all ergodic Markov chains are recurrent, and so every state occurs infinitely often in the long run. This is not the case with left-to-right Markov chains.

The necessary approach is to use cross-sectional variation to study these kinds of environments. Essentially, by looking at several different Markov chains, rather than just a single long one, we can infer the desired transition probabilities. However, our model here (Equation 4.3.2) is not explicitly left-to-right , and while some attention has been paid to the estimation of these kinds of models, a careful development of identification and its application to sales is still desirable. In the following section, I do exactly this, highlighting

how these kinds of models can be useful for economists.

4.3.1 Identification of Block Left-to-Right Hidden Markov Models

This section establishes the econometric properties of hidden Markov models (HMM) in a generalization of the environment developed earlier in this chapter, then provides sufficient conditions for parametric identification of the structured Markov model described developed earlier in this section. The notation and set-up is standard, following Section 4.2. However, because we now consider cross-sectional variation as well, we suppose we have a panel $i = 1, 2, ..., N$ of observations Y_{it}, each observed for a series of periods $t = 1, 2, ..., T$. In general, we will imagine that T is fixed, while N may vary. This is the opposite assumption for a standard HMM, which assumes $N = 1$ is fixed, while T is large and may vary. Let the unobserved variable be X_{it}, the elements of a stationary first-order Markov chain with state space $S = \{1, 2, ..., K\}$ and associated transition probability matrix $A = [\alpha_{ij}]_{ij}$. Then, suppose $\{f(\cdot;\theta)|\theta \in \Theta\}$ is a family of density functions on the real line and let $\{\theta_1, \theta_2, ..., \theta_K\}$ be elements of Θ.

To make this a hidden Markov model, we assume that $\{Y_{nt}\}_{t=1}^T$ is generated as a sequence of draws from $\{f(\cdot;\theta_{X_{it}})\}_{t=1}^T$ where the sequence of distributions is given by the realization of $\{X_{it}\}_{t=1}^T$. This process occurs identically and independently for each $i = 1, 2, ..., N$ in the panel. In other words, we assume that each member of the panel is generated as a standard hidden Markov model. Furthermore, for generality, let's assume that the parameters of the model are (possibly) driven by some underlying parametric specification ϕ: $\alpha_{ij}(\phi)$,$\theta_i(\phi)$. In the general case, ϕ is identical with the underlying parameter space.

This underlying specification is useful because we may have some structural parameters in mind for testing, which (through an economic model) can directly connect to the hidden Markov model. For example, suppose that we believe that regime r changes with probability q_r, but across regimes the probability of a sale is the same q_s. Then, starting in regime r the probability of transitioning from a regular price into a sale in regime r is $(1 - q_r)q_s$ and the probability of transitions from a regular price into a sale in regime $r + 1$ is $(1 - q_{r+1})q_s$. These can be linked to the transition probabilities in the underlying model, and the sale probability recovered directly. This allows a researcher to make statements like "the probability of a sale is q_s" in a grounded and specific way[41].

[41]This also makes the communication of the results more transparent, since it avoids a discussion of the

As discussed, we are particularly interested in considering HMMs which are non-ergodic; their transition matrices typically are not recurrent. In this case, the initial starting point of the underlying Markov chain is very important. Denote the distribution of starting states in the population by π. Consider the following related definitions, which will be used to motivate which states can be identified within the model:

Definition 4.3.1. A transition matrix (of size K) is *left-to-right* if it takes the form:

$$\begin{bmatrix} p_{11} & 1-p_{11} & 0 & \dots & 0 \\ 0 & p_{22} & 1-p_{22} & \dots & 0 \\ 0 & 0 & p_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

where $p_{ii} \in (0,1)$ for $i < K$.

Essentially, a left-to-right matrix is one in which the state can only move forward in time; never backwards, and although it may linger in particular states for a substantial period of time it will not remain there indefinitely.

Definition 4.3.2. A transition matrix (of size K) is *block left-to-right* if it takes the form

$$\begin{bmatrix} A_{11} & A_{12} & 0 & \dots & 0 \\ 0 & A_{22} & A_{23} & \dots & 0 \\ 0 & 0 & A_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & F \end{bmatrix}$$

where the A_{ij} are square and have elements in $[0,1)$, and the sum by columns of A_{ij} and $A_{i(j+1)}$ is 1 (i.e. they form a well-defined transition matrix), and F is an absorbing state (i.e. one in which the chain cannot move away from)

A block left-to-right matrix is a generalization of the left-to-right form, in which the "states" in a standard left-to-right matrix are composed of *sets* of states in a larger matrix. Next, I introduce some conditions which will be useful for identification:

Assumption 4.3.1. Identification Conditions:

hidden Markov model framework

- Condition 1: The family of mixtures of at most K components of $\{f(\cdot;\theta)|\theta \in \Theta\}$ is identifiable.
- Condition 2: For each i, j, k the functions $\alpha_{ij}(\phi)$ and $\theta_k(\phi)$ are continuous, 1:1 mappings.
- Condition 3: For each i, the Markov chains are independent and identically generated from an initial distribution π.

Some of these assumptions are familiar from the discussion in Section 4.2. Condition 1 is identical to Assumption 4.2.5 for regular Markov chains, and the rationale is the same: it provides a way to identify the emission parameters and the mixing coefficients, which are key parts of the Markov chain. Condition 3 is what allows us to use cross-sectional variation to capture the behaviour of a left-to-right Markov chain. Condition 2 allows for the identification of lower-level parameters from the structure of the hidden Markov model.

The first result is a lemma, something of a folk theorem, which establishes the identifiability of left-to-right matrices.

Lemma 4.3.1. Under Conditions 1,3, suppose the transition matrix $A \equiv [a_{ij}]$ is left-to-right, $\pi = (1, 0, ..., 0)'$, and $K < T$. Then, we can identify (1) A, and (2) θ_k for all k.

Proof. See appendix. □

This shows that a generic left-right transition matrix creates an identifiable Markov model. However, what if the structure is more flexible, as in our model of a block left-to-right model? In this case, it will depend on the initial vector π and the structure of the chain. We can generalize the basic result using the notion of accessibility, which is defined for our environment as:

Definition 4.3.3. We say that a state j is *accessible by T from state i* if $[a_{ij}]^T > 0$. Similarly, a state is *accessible from π* (a vector) if the state is accessible by a state associated with non-zero member of that vector.

Theorem 4.3.1. Suppose Conditions 1 and 3 hold. Let S' be the set of states which are accessible by T from π. Suppose $K < T$. Then, we can identify: (1) $[a'_{ij}]$, the sub-matrix of $[\alpha_{ij}]$ composed of transitions between states in S', (2) θ_k for $k \in S'$, and (3) the initial state distribution π.

Proof. See appendix. □

This result is very useful, because it generalizes the left-to-right structure. It means that if, during the period of observation (T), there is a chance that a given transition (and state) will be observed, then we can recover both the likelihood of that transition, and the attendant state parameters associated with it. We can also extend this result to the underlying parameter space, if it differs from the hidden Markov model, via the following:

Corollary 4.3.1. Under Condition 2, the underlying parameter vector ϕ is identified for a hidden Markov model which meets the assumptions of Theorem 4.3.1.

In other words, we can recover the underlying parameters which generate a hidden Markov model if the mapping is sufficiently smooth and 1:1. This is a necessary condition, but is certainly not sufficient. Weaker conditions are certainly possible (e.g. rank-order conditions) given details of the relationship between the model and the more primitive parametrization.

We can also apply this result to the block left-to-right environment which we have developed for sales, using the following result:

Corollary 4.3.2. Under Conditions 1, 2 and 3, the parameters underlying a block left-to-right hidden Markov model which begins from block A_{11} are identified if (1) all the states within a given block (r) are accessible from every other state that block within time t_r (2) the sum of all such t_r is less than or equal to T, and (3) $K < T$.

This is a stronger result which depends on the particular structure of a block left-to-right matrix, in a similar way to Lemma 4.3.1. In this situation, since all states in a block are accessible within a time which can be met within T, the block left-to-right model can be identified completely.

Corollary 4.3.2 is the key piece of theory we need to connect hidden Markov models to sales classification. The framework developed earlier in this section is a block left-to-right hidden Markov model which meets the three conditions of the theorem provided the time frame is reasonably long. This means that a hidden Markov model such as the one we have developed for sales is identifiable using cross-sectional data on these kinds of Markov chains. This means that we can determine the parameters which govern our model of sales, thereby recover which periods are most likely to be sales: a classification method. However, what if there are underlying covariates or heterogeneity. For example, maybe we

are considering a particular product, but perhaps at different stores? Additionally, how do we incorporate information about other variables, like volume, into the determination of a sale (and the associated prices)? This question is analysed in the following section.

4.3.2 Heterogeneity and Covariates

So far, we have only considered hidden Markov models which a single observable variable (Y_{it}) and used the cross-sectional variation in i to identify the parameters of interest. However, it is unlikely in some applications to find a large panel of identical draws from a single Markov model, meeting Condition 3. The first possible extension of the the model is to include observable heterogeneity in i. For example, suppose that we can observe that products come from different chains of stores which are observable. In this case, we would like to classify sales understanding that these different stores might offer different prices. This section considers such situation. Denote the vector of observed panel-level variables by W_i. Then, we can admit the possibility that ϕ, the vector of underlying model parameters, may depend on W_i, in the sense that $alpha_{ij}(\phi_i), \theta_k(\phi_i)$ and $\phi_i = \phi(W_i)$.

Then, we can consider the following replacement for Condition 3 in Assumption 4.3.1:

Assumption 4.3.2. (Condition 3a) The observed heterogeneity W_i is independently and identically distributed for all i. Denote the distribution of W_i by $g(\cdot)$.

This assumption means that the heterogeneous component of the variation in W_i is created independently by a process independent of the model. That is, while there may be heterogeneity which informs the manner in which the different components transit the model, this variation is itself exogenous. We immediately have the following result:

Corollary 4.3.3. Under Conditions 1,2, and 3a, the parameters underlying a block left-to-right hidden Markov model which begins from block A_{11} are identified if (1) all the states within a given block (r) are accessible from every other state that block within time t_r for all W_i (2) the sum of all such t_r is less than or equal to T, and (3) $K < T$.

This is the heterogeneous version of Corollary 4.3.2; the proof is identical. The intuition is that since W_i is exogenous, the model with heterogeneity is a combination of the unconditional model in proportions given by $g(\cdot)$. Since W_i is observable, g is known, and the heterogeneity can be separated out and the identification carried out piecewise. Essentially, it is like identify the model separately for each W_i, the putting them together. The

critical assumption is that the observed heterogeneity is independent of the model; this means, in particular, that it is independent of X_{it} and Y_{it}, except through the parameters of the model. In particular, this restricts a result like where two different W_i have different families for their emissions, or transit a different number of states.[42]

A more challenging problem arises if we consider the problem of time-varying covariates W_{it}. For example, we may not only observe pricing information for a given store, but perhaps volume or another set of variables as well. This variable might be informative about the presence of a sale or not in a given time period. As discussed in Section 4.2, this is one of the major drawbacks of most sales classification methods: the inability to use all available information to determine whether something is a sale or not. The use of such other information in classification also assesses the degree to which it is associated with a sale in general, which is a direct way of testing many models of sales.

In the hidden Markov model framework, the way to incorporate covariates is by treating them as a *joint* emission from the underlying Markov chain: $\mathbf{Y_{it}} \equiv \{Y_{it}, W_{it}\}$. Suppose W_{it} consists of $M - 1 > 0$ variables. Then, we define $\{\mathbf{f}(\cdot;\theta)|\theta \in \Theta\}$ to be a family of density functions on $\mathbb{R}^M$ (and $\{\theta_1, \theta_2, \ldots, \theta_K\}$ be elements of Θ, as before). Then, the covariate-robust extension of the original hidden Markov model is to assume that $\{\mathbf{Y_{it}}\}_{t=1}^T$ is generated as a sequence of draws from $\{\mathbf{f}(\cdot;\theta_{X_{it}})\}_{t=1}^T$, where the sequence of distributions is given by $\{X_{it}\}$, the states of the underlying Markov chain.

Essentially, the key difference is that we consider all variables together as a set of joint emissions, and adjust the distributions and other parts of the model accordingly. The assumption which needs to be adjusted is as follows:

Assumption 4.3.3. (Condition 1a) The family of mixtures of at most K components of $\{\mathbf{f}(\cdot;\theta)|\theta \in \Theta\}$ is identifiable.

If we replace Condition 1 with Condition 1a, Corollary 4.3.2 goes through without modification. This assumption restricts the number of families which are acceptable for emissions from the underlying model. However, there are still a large number of choices: most notably, the multivariate Gaussian distribution meets the requirements of Condition 1a. This generalization of the basic model means that the underlying model is identifiable using the joint distribution of all the variables. In terms of sales, this means that sales can be classified using this more complete set of information.

[42]This is particularly restrictive for models like those discussed in Section 4.3.3; if regimes are the only major difference between two models, the covariate with fewer regimes will be over-fitted, resulting in a spurious set of regimes for this model.

This also means, as explained, that we can test the association of sales with different covariates. For example, suppose we were interested in knowing whether or not the volume of product sold increased during a sale (in addition to price), and we chose a multivariate Gaussian specification. Then, we could test whether volume increased during a sale by comparing $\mu_{W,X}$, the average parameter of the Gaussian distribution for volume (W) when X is a sale or not. Under the null hypothesis of no relationship, the means should be the same in both states. The alternative hypothesis would indicate a difference in the means between the states, and would provide evidence for this variable being associated with sales. In general, if there are R regimes, this would require R comparisons. A multiple-testing procedure like the Bonferroni correction can be used to correct the size of the test to take this into account if desired.

While this allows a researcher to robustly test the relationship between underlying states and different variables, it does not necessarily make the causal relationship explicit. For example, it could be the case that a covariate is higher during a sale because the sale causes it to increase. However, it could also be the case that a higher level of that variable led to a sale. This method does not give insight into which of these is necessarily the case, in a similar way that a regression model is silent on the causal direction once the model is specified. The researcher must rely on an underlying economic model or motivation to understand the implications of a model. However, most models provide predictions about correlations between given variables (like sales and volume), which this framework is ideal for testing.

4.3.3 Higher Order Markov Processes

One important situation which does not fall within the framework developed in Sections 4.3.2 or 4.3.1 is where we would like to include some kind of temporal dependence in the Markov chain. For example, in many regression models of sales, researchers will include lagged values of the sale indicator variable. This is usually to try to capture some manner of periodicity in the underlying data generating process which is not recoverable from exogenous timing variables. For example, the model of Chapter 2 explicitly studies this kind of model.

The framework we have developed relies exclusively on first-order, finite, left-to-right Markov chains, which do not admit longer order Markov dependence. However, as mentioned in Section 4.2, this is largely without loss of generality (at least formally). We can

include higher order Markov dependence, and therefore the presence of lagged state variables, by converting the higher order Markov chain into a first order chain. This is done by redefining the state space so that each vector of past states (up to the order of the original chain) is a unique state. This also then implies restrictions on the transition matrix, since then some states are by definition inaccessible from other states.

For example, consider our basic model with a left-to-right structure for the pricing regimes. For exposition, suppose that the only lagged variable of importance is the unobserved sale variable last period. In other words, we are considering a second-order Markov chain. Then, we can imagine the system as a very large Markov chain by assigning a period a unique state consisting of three elements $X_t = (R_t, S_{t-1}, S_t)$; there are in total 2^2R such states. The first would be $(1, 0, 0)$ consisting of regime 1 without a present sale, and without a sale last period. The second would be $(1, 1, 0)$ which consists of regime 1 without a present sale but with a sale last period, and so forth in this manner. This model has an associated transition matrix with restrictions on values, since not all states can be reached from each value. This is again a block diagonal matrix, with block elements of the form:

	$(1,0,0)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$	$(2,0,0)$	$(2,0,1)$	$(2,1,0)$	$(2,1,1)$	$\dots$
$(1,0,0)$	p_{11}	p_{12}	0	0	p_{15}	p_{16}	0	0	
$(1,0,1)$	0	0	p_{23}	p_{24}	0	0	p_{27}	p_{28}	
$(1,1,0)$	p_{31}	p_{32}	0	0	p_{35}	p_{36}	0	0	
$(1,1,1)$	0	0	p_{43}	p_{44}	0	0	p_{47}	p_{48}	
$\vdots$				$\vdots$					$\ddots$

This redefinition of the higher-order Markov chain allows for all of the results of Section 4.3.1 and 4.3.2 to be applied to models with this kind of dependence. The fact that the identification and general method of application is identical to the simpler model is somewhat deceptive: this kind of dependence comes at a cost. Suppose a model has R regimes and we consider a Markov process of order L. Then, the number of states in the model is $R \cdot 2^{(L+1)}$ states and approximately $(L+1)R \cdot 2^{(L+2)}$ transition probabilities. For even relatively small orders of lagged dependence, this can become very large. From an identification point of view, the condition that $T > K$ means that lagged models require exponentially more periods of observation than the basic model. From a practical point of view, even if the identification condition is met, it may be the case that for a given set of

data drawn from a model, very large amounts of data may be required to get accurate estimates of the many parameters of the model. This can lead to difficulty making inferences, and in particular recovering structural parameters or testing hypotheses as in Section 4.3.2.

These kind of problems are particularly acute in the case where the lagged state variable is of a long duration. For sales, this may not be that reasonable, but it depends on the length of the period being considered[43]. However, this is much more reasonable in the case where the researcher believes that regime itself has some kind of internal dynamics. For example, it may follow a fragility process, wherein it is very unlikely to change shortly after creation, but then suddenly becomes exponentially more fragile after a certain point in time is crossed. If this was modelled as straightforward time dependence, this would probably become highly infeasible for most datasets, since a large number of lags would need to be included in the Markov chain.

An alternative is to take advantage of the hidden nature of the Markov model and instead model the regime fragility as a sequence of states itself: one of lower order than the expected duration. In a left-to-right Markov chain, the mean duration in a given state is approximately proportional to the inverse of the transition probability from that state[44]. This means that an appropriately granular underlying set of states with sufficient duration can capture fragility concerns for regimes in a natural way. The method is identical to the transformation of the process for lagged state variables above: merely define a state as the combination of hidden variables.

4.4 Monte Carlo Simulations

In order to investigate the utility of this framework in a practical setting, in this section I report the results of a series of Monte Carlo simulations of sales. Generating simulated data is useful, because it allows me to test the accuracy of the hidden Markov model framework directly. We can compare the estimates from the model to the true underlying parameters. I generate data using a process which is plausible for sales. To be specific, as described in Section 4.3, there is a hidden regime and a hidden sale variable which jointly govern the observed distribution of prices (and covariates). I model the emissions as being

[43]For example, a lag of 10 periods may be unreasonable for weekly sales data, but might be very reasonable for daily sales data.

[44]In a general Markov chain, the expected duration in a state j conditional on starting from state i is defined by $s_{ij} = \delta_{i,j} + \sum_k P_{ik} s_{kj}$ where δ is the indicator delta and P is the matrix of transition probabilities of transient states. This can be solved as in Ross (2014) for the durations.

(multivariate) normal and uncorrelated in terms of noise. As each regime changes, the regular and sale prices in the model change. The central objects of interest are (1) the prices (sale and otherwise) imputed to the different situations and (2) the classification of observations into sales. For simplicity, I focus on models which start in regime 1 at a regular prices (that is, π is known); investigation of alternative simulations indicate that this assumption is not critical to the results.

For the basic model (Model 1), I choose $R = 4$ regimes, and consider only first-order Markov behaviour, resulting in eight total states. I assume a panel size of $N = 500$ and suppose this is observed weekly for four years, resulting in $T = 208$ time periods. Motivated by Chapter 2, I suppose sales happen every four weeks, along with idiosyncratic random sales. The fragility of the regimes is $q = \frac{1}{104}$, resulting in an average regime change every one to two years. Sale prices within a regime are a \$5 discount from the regular price, which follows an inflationary trajectory (25, 28, 30, 32) as the regimes change. It is worth noting that this model meets the assumptions required for Corollary 4.3.2 to hold.

I estimate the model using the techniques developed in section 4.2, implemented for the panel structure of the data. As a starting point, I choose locations near the true parameters. In order to calculate confidence intervals (and standard errors), I use 500 panel bootstrap re-estimates of the sample. These results are robust to other bootstrap sizes. My results for the basic model are presented in Table 4.1. In short, the model does a remarkably good job. The means and deviations of the emissions are all accurate, being close to the true parameters. This means that the price imputation from the model is essentially correct. More importantly, from a classification point of view, the fit of γ, the predicted state for each observation, is essentially perfect - nearly 100% accuracy[45] One point to note is that because the fragility of each regime is $q = \frac{1}{104}$, this means that the expected number of regime changes is two during the 208 time periods of the model. This means that not all of the panels each the fourth regime. This is why we see increasing standard errors for the variables as the regime numbers increase: every panel reaches regime 1 but only a quarter or so reach regime 4.

In order to pressure the framework, I also simulated a more complicated model (Model 2). This model retains the sample and panel sizes ($N = 500$, $T = 208$) but greatly reduces

[45] I calculate accuracy for γ by comparing the true state to the vector of probabilities imputed by the model. For example, if we represent the distribution across states as a vector $v = (v_1, v_2, \ldots, v_N)$ (for N states) and the true state by $t = (t_1, t_2, \ldots, t_N)$, then the error rate is $\frac{1}{2}\sum_{n=1}^{N} |v_n - t_n|$, which can then be averaged over all observations to get a fit. For example, if every predicted state was 75% correct, then the fit would be 75%.

	Regime 1		Regime 2		Regime 3		Regime 4	
	Estimate	Actual	Estimate	Actual	Estimate	Actual	Estimate	Actual
μ_r	24.999	25.000	27.999	28.000	29.999	30.000	31.998	32.000
	(0.0006)	-	(0.0006)	-	(0.0009)	-	(0.0011)	-
μ_s	20.000	20.000	23.000	23.000	25.002	25.000	28.002	28.000
	(0.0009)	-	(0.0011)	-	(0.0013)	-	(0.0016)	-
σ_r	9.9	10	10.1	10	9.9	10	10.4	10
(1000-ths)	(0.080)	-	(0.091)	-	(0.127)	-	(0.158)	-
σ_s	9.8	10	9.9	10	10.2	10	10.3	10
(1000-ths)	(0.125)	-	(0.160)	-	(0.213)	-	(0.254)	-
γ-fit	100.00%							

Table 4.1: Estimates relative to Monte Carlo Simulation: Model 1 (Basic), no Covariates

the separation between the different regime prices. I choose $R = 4$ regimes, with the same fragility of $q = \frac{1}{104}$ as before. However, I deliberately make the regimes more difficult to tell apart. The regular prices are set at $(25, 25, 30, 32)$ while the sale prices are chosen to be $(22, 20, 25, 25)$. This means that regimes 1 and 2 have the same regular prices, and only differ in their sale price (by \$2). Additionally, regimes 2 and 3 have the feature that the sale price of regime 3 is the regular price of regime 3. This kind of switch-over is difficult for many classification methods (as discussed in Chapter 2 to handle). I estimate the model in the same manner as for the basic model, with no change in the procedures.

The results are presented in Table 4.2. As we can see, the results remain broadly similar to the simpler model. The parameter estimates for the means and variances of the emissions in each of the four regimes are very close to their true values. This is despite the fact that these were chosen to be deliberately misleading and difficult to infer from regime to regime. Consequentially, the fit γ is also very good; 99.4% on average. The majority of the errors occurred in either Regime 1 or Regime 2, and were generally small. In general, this implies that the classification of the model (especially in terms of sales) remains very accurate. We will return to this point in the next section, where we look at the inclusion of covariates and the consequences of this extension.

4.4.1 Covariates

In this section, I examine the role covariates can play in the specification of a hidden Markov model. As discussed in Section 4.3, one of the major roles (besides classification)

	Regime 1		Regime 2		Regime 3		Regime 4	
	Estimate	Actual	Estimate	Actual	Estimate	Actual	Estimate	Actual
μ_r	24.9996	25	24.9996	25	29.9994	30	31.9988	32
	(0.0006)	-	(0.0006)	-	0.0009	-	(0.0011)	-
μ_s	22.0003	22	20.0001	20	25.0020	25	25.0026	25
	(0.0009)	-	(0.0011)	-	0.0013	-	(0.0016)	-
σ_r	9.9207	10	9.9032	10	9.9123	10	10.3615	10
(1000-ths)	(0.0798)	-	(0.0907)	-	0.1271	-	(0.1581)	-
σ_s	9.8494	10	10.1299	10	10.1766	10	10.3343	10
(1000-ths)	(0.125)	-	(0.1592)	-	0.2145	-	(0.2585)	-
γ-fit	99.40%							

Table 4.2: Estimates relative to Monte Carlo Simulation: Model 2 (Complex), no Covariates

that hidden Markov models can play in the analysis of sales is by evaluating the association of different covariates and the (hidden) sale variable. This provides additional power, since it lets other associated variables guide the model's definition of a sale, and not just the price.

I first extend Model 1 by including two covariates of interest. The first, X_1, is correlated with sales; we can imagine it as volume. In periods without a sale, it has a mean value of 0.66. In periods with a sale, however, it jumps up to a mean of 1.52. The second variable, X_2 is uncorrelated with sales. In every period, regardless of whether it is a sale period or otherwise, it takes on an average value of 0.86. I estimate this new model in the same manner as before, except that I also include these covariates as emissions from the underlying state variables.

The results of the estimation are presented in Table 4.3. In this table, I omit the estimates of the noise variance; they are also accurate and similar to that of Table 4.1. They confirm the predictions of the model, and agree with the true interpretation of the covariates. As we can see, similar to the result from the basic model, we have extremely good estimates of the average price in each regime and sale state. As we would expect from the previous results, this also leads to a nearly perfect fit of the γ probabilities, indicating that the classification of observation into sales is accurate. The results for the covariates are just as good: accurate predictions of the means and variances of the emissions. Equally importantly, the accuracy of these variables also means that the interpretation of the hidden Markov model agrees with the true interpretation of the variables. A hypothesis test (either individually or jointly) robustly rejects the null hypothesis that $\mu_{X_1,r} = \mu_{X_1,s}$ at standard

confidence levels. This indicates that the X_1 variables significantly co-varies along with the hidden sale variable. Similarly, a hypothesis test of $\mu_{X_2,r} = \mu_{X_2,s}$ is not rejected at most confidence levels. This would imply to a researcher that X_1 is associated with sales while X_2 is not; a conclusion which is correct. We can also see the fact repeated from the basic model that as the regimes increase, variance also rises, consistent with the relatively scarcer data in later regimes.

As with the basic model without covariates, I also estimate the intentionally complicated model (Model 2) with the inclusion of covariates. I include X_1 and X_2, generated precisely as in the basic model, with the same means and variances. I also include another covariate, X_3 which is an indicator variable for the presence of being in the fourth period. The idea of this covariate is to highlight how indicator variables for certain periodic patterns, as in Chapter 2, can be included in the model. The results are presented in Table 4.4. They highlight a few things: first of all, even with the more complicated model, the fit remains very good. The hidden Markov framework remains able to correctly recover the means of the prices and the other of the generating variables from the model. The standard errors of the fits generally increase, mainly due to uncertainty over some of the state fitting in the model. The variances are not presented in this table, but are similarly accurate. The periodic variable, X_3 which was not generated as a multivariate normal emission, also performs well. In particular, it correctly captures the fact that there is a 4-period pattern in sales. Sales are 86% more likely in a sale period than in a non-sale period. Given that there are 3 non-sale periods, each with a 5% chance to have an idiosyncratic sale, this is good assessment of the relative probability.

More importantly, we can see that even within a framework designed to confuse the estimation, the γ fit remains high at 99.5%, meaning that most of the states are accurately imputed by the model. The fact that this is marginally better than using pricing alone indicates that the inclusion of more covariates can help classify sales more accurately than just using price alone, a fact typically overlooked by most classification methods. Looking more in-depth, we can see that the model primarily made one of two mistakes: either confusing Regime 1 and Regime 2 or Regime 3 and Regime 4, as predicted. The first error is much more common, mainly because Regimes 1 and 2 are more frequent, as are regular prices. Importantly, most of the errors also came from the regular prices, with only about 25% of the error rate (0.1%) attributable to the sale periods. This is likely because sales are less frequent; if we correct for this, errors are still less common for sales, albeit not as dramatically (approximately equally likely). The typical size of the mistakes is also not

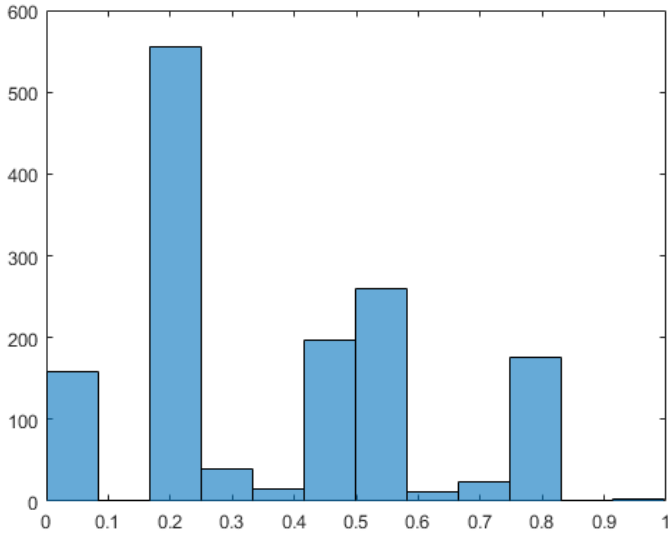


Figure 4.2: Histogram of State Errors, Model 2 (Complex), Covariates

large: on average, errors are not complete failures to impute the state, being only a mistake of about 36.1%. On average, the state imputed is more correct than not, and very few are highly incorrect. I illustrate the distribution of errors by size (1 = 100%) as a histogram in Figure 4.2. This fact implies that most methods, such as the naive γ best-fit method or the Viterbi algorithm (as discussed in Rabiner (1989)) would correctly impute the correct states. It also means that classification on such a basis, such as simply choosing the most likely state, would be very likely to classify the states correctly.

4.4.2 Higher Order Markov Processes

In this section, I consider the performance of hidden Markov models for examining higher-order processes, as developed in Section 4.3.3. The first model is an extension of the basic model for lower order processes. I retain $R = 4$ regimes, but consider third-order Markov behaviour, resulting in sixteen total states. Basically, we assume that sales occur four weeks after the last sale, in addition to a 5% idiosyncratic chance each period. This is different than the absolute reference point (every four weeks) used in the first-order process considered before. I continue to use a panel size of $N = 500$ observed for $T = 208$ time periods. The fragility of the regimes remains $q = \frac{1}{104}$. Sale prices within a regime are

	Regime 1		Regime 2		Regime 3		Regime 4	
	Estimate	Actual	Estimate	Actual	Estimate	Actual	Estimate	Actual
μ_r	24.999	25.000	27.999	28.000	29.999	30.000	31.998	32.000
	(0.0006)	-	(0.0006)	-	(0.0009)	-	(0.0011)	-
μ_s	20.000	20.000	23.000	23.000	25.002	25.000	28.002	28.000
	(0.0009)	-	(0.0011)	-	(0.0013)	-	(0.0016)	-
$\mu_{X1,r}$	0.660	0.66	0.660	0.66	0.659	0.66	0.660	0.66
	(0.0002)	-	(0.0002)	-	(0.0003)	-	(0.0003)	-
$\mu_{X1,s}$	1.519	1.52	1.519	1.52	1.519	1.52	1.519	1.52
	(0.0003)	-	(0.0003)	-	(0.0004)	-	(0.0006)	-
$\mu_{X2,r}$	0.860	0.86	0.860	0.86	0.860	0.86	0.859	0.86
	(0.0002)	-	(0.0002)	-	(0.0003)	-	(0.0003)	-
$\mu_{X2,r}$	0.860	0.86	0.859	0.86	0.859	0.86	0.860	0.86
	(0.0003)	-	(0.0003)	-	(0.0004)	-	(0.0006)	-
γ-fit	100.00%							

Table 4.3: Estimates relative to Monte Carlo Simulation: Model 1 (Basic), Covariates

the same: a \$5 discount from the regular price, which follows an inflationary trajectory (25,28,30,32) as the regimes change. However, there are now three regular price states within a regime, and one sale state. The emissions from the model are Gaussian, and I denote the average of the emission by μ_r, L for regular prices at a given lag L. I also include the two covariates, X_1 and X_2, used before, with the same generating process: X_1 is either 0.66 or 1.52 on average in non-sale and sale periods, while X_2 is always 0.66 on average in every period.

The results are displayed in Table 4.5. As before, I present results for the means of the emission variables; the variances are also similarly accurate. The model once again performs extremely well. As we can see, the model correctly imputes the prices associated with the true values in all of the states, and across regimes. It also correctly estimates the emissions associates with the other covariates, X_1 and X_2 in all of the states. This also means that the fit of the predicted states γ is essentially completely accurate. This is remarkable for two reasons: first, it means that classification, even in much more complicated environment, is just as accurate as in the simplest model. More importantly, however, it also means that the model is able to correctly impute states which are detectable only through their transitions. Note that there is no observable difference, in terms of emissions, between $\mu_r, L1$ and $\mu_r, L1$. The only way to detect these states is through (1) the struc-

		Price		X_1		X_2		X_3
	Mean	Estimate	Actual	Estimate	Actual	Estimate	Actual	Estimate
Regime 1	μ_r	24.9996	25	0.6601	0.66	0.8601	0.86	0.0000
		(0.0006)	-	(0.0002)	-	(0.0002)	-	(0.0000)
	μ_s	22.0003	22	1.5197	1.52	0.8601	0.86	0.8736
		(0.0009)	-	(0.0003)	-	(0.0003)	-	(0.0026)
Regime 2	μ_r	24.9996	25	0.6600	0.66	0.8600	0.86	0.0000
		(0.0006)	-	(0.0002)	-	(0.0002)	-	(0.0000)
	μ_s	20.0001	20	1.5196	1.52	0.8599	0.86	0.8753
		(0.0011)	-	(0.0003)	-	(0.0003)	-	(0.0032)
Regime 3	μ_r	29.9994	30	0.6595	0.66	0.8601	0.86	0.0000
		(0.0009)	-	(0.0003)	-	(0.0003)	-	(0.0000)
	μ_s	25.0020	25	1.5199	1.52	0.8597	0.86	0.8800
		(0.0013)	-	(0.0004)	-	(0.0004)	-	(0.0041)
Regime 4	μ_r	31.9988	32	0.6603	0.66	0.8599	0.86	0.0000
		(0.0011)	-	(0.0003)	-	(0.0003)	-	(0.0000)
	μ_s	25.0026	25	1.5192	1.52	0.8604	0.86	0.8737
		(0.0016)	-	(0.0006)	-	(0.0006)	-	(0.0052)
	γ-fit	99.5%						

Table 4.4: Estimates relative to Monte Carlo Simulation: Model 2 (Complex), Covariates

ture of the left-to-right model and (2) their estimated correlation with emission-correlated states (like sales). This means that even very subtle modelling assumptions for sales can be captured accurately by a hidden Markov model, provided sufficiently good initial values can be supplied to start the model near the global maximum.

In Table 4.6, I present the results for the complex version (Model 2) of the parameters, which are intentionally designed to be difficult to impute sales from. As we can see, the results remains very good. The hidden Markov model correctly infers the means of the price variables, as well as the covariates. The estimates for the variances are also similarly accurate. As a consequence, the fit for the model is also very good, with most errors being less than 50%. This means that, as discussed earlier, methods for imputing the true state (and classifying sales) will typically be correct. Generally, the errors arise due to the model being unsure about which price belong to which regime, which is compounded by the fact that some of the underlying states are hidden (based only on the distance from the most recent sale). Overall, the model performs generally very well, even in the complicated environment with many variables necessary to estimate.

4.4.3 Small Sample Performance

In this section, I examine a question of practical, rather than theoretical, interest: how useful is the hidden Markov framework for classification of sales when a researcher does not have very many observations? The identification results developed in section 4.3.1 relied on the panel structure of the data to get precise predictions, and in the preceding Monte Carlo investigation, I used samples of size $N = 500$ to simulate the model. However, in many applications we may only have a limited amount of data: $N = 1$ is a common situation. In this section, I repeat my Monte Carlo exercise, but looking specifically at small sample sizes.

I consider the complex model (model 2), with $T = 208$ and $N = 1$. I choose only samples which reach the fourth regime (making the specification correct), and I repeat the estimation carried out in the previous section. Of the 150 candidate samples, 5 (3.3%) fail to converge during estimation. This mainly occurs when a regime (typically Regime 4) is too short for estimation to be carried out well; the resulting local maximum is too far from the initial point for estimation to converge. Looking at the remainder of the sample shows the results are still fairly good: in terms of classification, the fit is approximately 94.4% accurate on average. The accuracy varies between about 89% and 98% within the panel, which is depicted in Figure 4.3. This implies that the classification, even with much smaller samples than required, still performs very well on average. However, fit errors, when they occur, tend to be complete: that is, states are mistaken for one another nearly completely. This means that errors, when they occur, are likely to change the classification of the mistaken state. However, most of these occur as expected between regimes and not between the sales themselves.

The fit for the means of the emission variables are similarly accurate. On average, the means of the emission variables have an error rate[46] of 1.03%. They range between 0% and 2% within the samples considered; their distribution is depicted in Figure 4.4. This is largely expected, given the results for the overall classification fit, above. However, this still demonstrates that sale prices can be properly imputed, as can the values of associated covariates. This implies that even for small samples the techniques and inferences developed in Section 4.3.2 can be applied.

I finally consider the small sample performance of the classification method with a higher-order Markov chain. I continue to use model 2, the complex version, this time re-

[46] Calculated as the %-difference between the estimated value and the true value

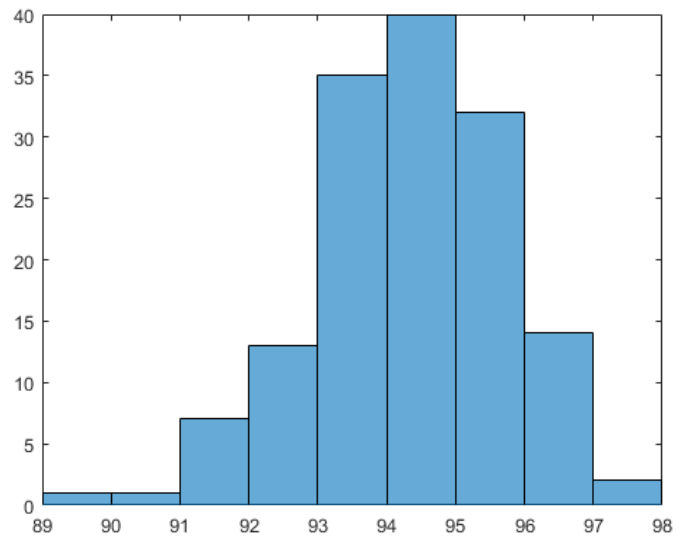


Figure 4.3: Histogram of γ-fit Errors in Small Sample Estimates (Model 2)

peating the analysis in the preceding section for $N = 1$, as above. In this model, there are also 150 candidate samples, of which 5 (3.3%) also fail to converge for similar reasons as in the more basic model. Of the remainder, the γ-fit is 99.2% accurate, and the fits range between 97% and 100% accurate. This indicates that even with the more complicated model driving the data, the fit remains very good. This carries over to the means of the emission variables, which show an average error rate of 1.62%, ranging between approximately zero and 2.5%. The overall fit of the model is comparable to the simpler model, without the lagged covariates.

This demonstrates the utility of the overall framework, even with very small sample sizes. The robustness and strength of the hidden Markov model framework, structured carefully to be used for sales, will perform well even in environments in which it is not guaranteed to perform well. However, it is still very important to correctly specify the underlying model for the data generating process, in order to structure the model properly. It is equally important to provide a good initial condition for the estimation, so that it can both efficiently and correctly locate the local maximum corresponding to the true values of the parameters.

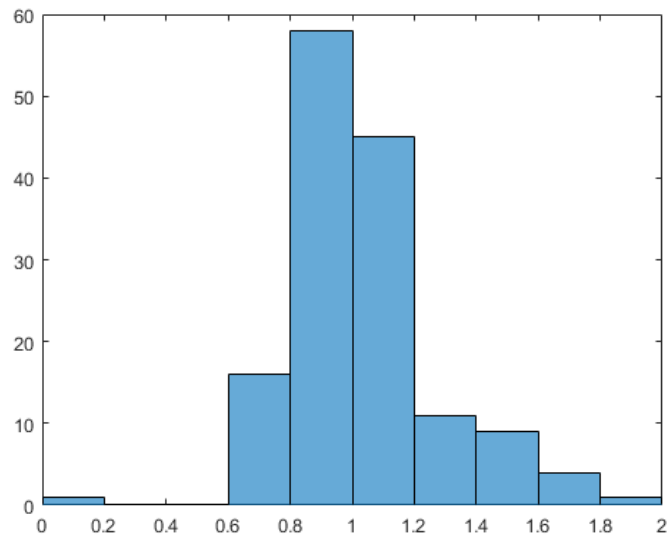


Figure 4.4: Histogram of μ Percentage Errors in Small Sample Estimates (Model 2)

4.5 Conclusion

In this chapter, I have developed and investigated a new method for classifying sales from data. Based on the hidden Markov model framework, this method is both intuitive economically and robust empirically. The underlying structure of the model closely mirrors the way sales intuitively arise in the economy, and conditions necessary for this framework to work are both natural and practical in many applications. The idea that sales manifest as co-movements in prices and other variables in the economy is a natural insight, and one which my method takes full advantage of. By using both price information as well as other variables, my method is able to provide a careful and capable classification procedure for sales, which uses all of the available information to infer whether a given observation is a sale or not. This also allows researchers to side-step the need for reduced-form analysis in many cases. Since many important questions about sales have to do with the correlations between a sale and a given covariate, all that is necessary to examine these relationships is a measure of that covariance. My classification method allows researchers to do this *directly*, jointly estimating both the covariance of a given variable with sales with the classification of sales itself. This eliminates the need for the regression framework used in

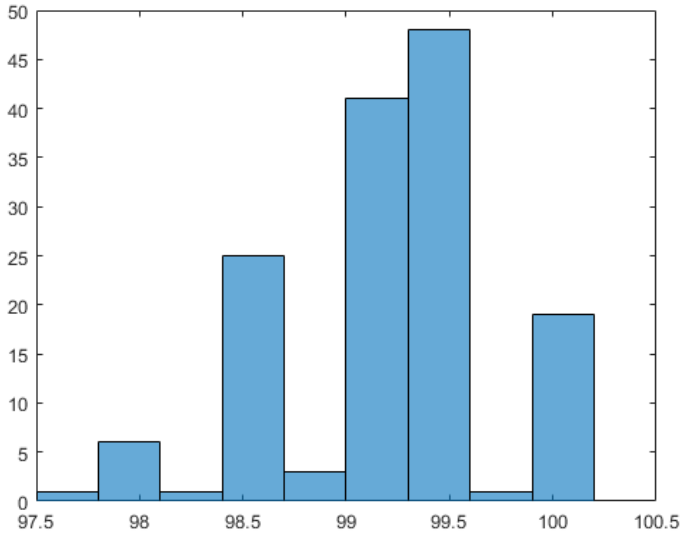


Figure 4.5: Histogram of γ-fit Errors in Small Sample Estimates, Higher Order (Model 2)

many studies, and is a powerful tool for understanding sales. It also provides a natural default to for sales classification, this time based on explicit economic reasoning and with clear identification requirements; it answers the question "which method should we use" by providing a powerful and robust framework which works in many environments.

My technique is particularly well-suited to sales environments typical of retail scanner data. Choosing products which are observed for a long period of time, and have similar pricing structures is a natural fit for this model. For example, products like beer or soft drinks, which have clear competitive structures governing their prices are an excellent product group for study. Alternatively, chains of stores with homogeneous prices are also a natural fit, since then a given product across the stores will help provide a natural panel for study.

As my Monte Carlo investigation shows, the method is highly accurate both at classifying sales from data and recovering the underlying relationships which appear alongside sales. This is true with both complicated environments, including many lagged state variables, and small sample sizes (something not guaranteed by the theory). However, this requires a good initial position for the model to estimate from, including accurate state classification and emission estimates. Many alternatives, such as that developed in Chap-

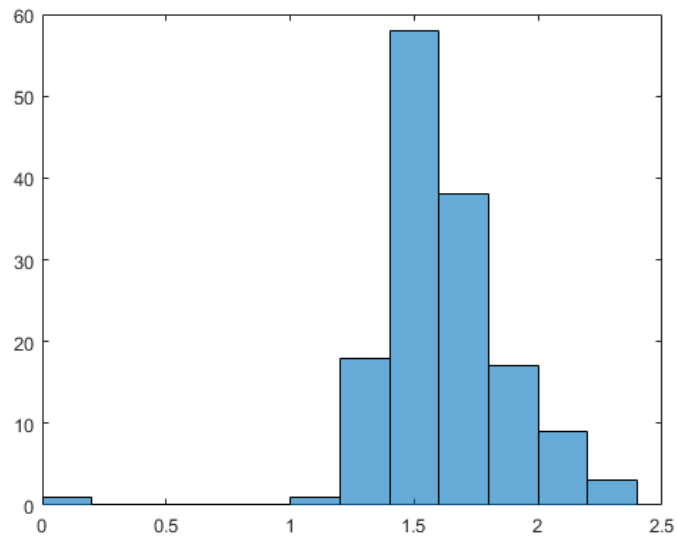


Figure 4.6: Histogram of μ Percentage Errors in Small Sample Estimates, Higher Order (Model 2)

ter 2, provide such a method in certain circumstances. A researcher can run a lower-level classification scheme to provide initial conditions, then run the method developed in this chapter to obtain a final classification. In many situations this is not markedly more complex than running this method alone.

This chapter also provides room for additional research, both theoretically and empirically. Theoretical work on the small scale properties of the estimation remains an open question, as does the consequences of miss-specification of the underlying model. Additionally, the key requirement for the use of panel data was the lack of ergodicity in most sales data. However, it is possible that a transformation of the data could restore this property. This would be advantageous since it would allow the more robust set of tools developed for ergodic hidden Markov models to be applied here, and it would make the technique now widely applicable. Empirically, the application of this to more types of sale environments will help provide more information on robust and overall performance on real-world data. Performing a similar exercise as this with actual sale information gathered first-hand from stores would be an interesting counterpoint to this chapter.

		Price		X_1		X_2	
	Variable	Estimate	Actual	Estimate	Actual	Estimate	Actual
Regime 1	$\mu_r, L1$	25.000	25	0.6606	0.66	0.8602	0.86
		(0.0009)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L2$	25.000	25	0.6600	0.66	0.8600	0.86
		(0.0009)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L3$	24.999	25	0.6598	0.66	0.8602	0.86
		(0.0009)	-	(0.0003)	-	(0.0003)	-
	μ_s	19.999	20	1.5196	1.52	0.8599	0.86
		(0.0009)	-	(0.0003)	-	(0.0003)	-
Regime 2	$\mu_r, L1$	27.998	28	0.6596	0.66	0.8598	0.86
		(0.0011)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L2$	28.001	28	0.6604	0.66	0.8600	0.86
		(0.0012)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L3$	28.000	28	0.6604	0.66	0.8602	0.86
		(0.0012)	-	(0.0004)	-	(0.0003)	-
	μ_s	23.000	23	1.5192	1.52	0.8600	0.86
		(0.0011)	-	(0.0003)	-	(0.0003)	-
Regime 3	$\mu_r, L1$	29.999	30	0.6591	0.66	0.8603	0.86
		(0.0013)	-	(0.0004)	-	(0.0005)	-
	$\mu_r, L2$	30.001	30	0.6595	0.66	0.8592	0.86
		(0.0016)	-	(0.0005)	-	(0.0004)	-
	$\mu_r, L3$	30.000	30	0.6602	0.66	0.8603	0.86
		(0.0016)	-	(0.0005)	-	(0.0004)	-
	μ_s	25.001	25	1.5196	1.52	0.8602	0.86
		(0.0015)	-	(0.0004)	-	(0.0004)	-
Regime 4	$\mu_r, L1$	32.003	32	0.6596	0.66	0.8598	0.86
		(0.0018)	-	(0.0006)	-	(0.0006)	-
	$\mu_r, L2$	31.996	32	0.6605	0.66	0.8605	0.86
		(0.0019)	-	(0.0006)	-	(0.0005)	-
	$\mu_r, L3$	32.001	32	0.6599	0.66	0.8602	0.86
		(0.0023)	-	(0.0006)	-	(0.0006)	-
	μ_s	28.000	28	1.5202	1.52	0.8598	0.86
		(0.0016)	-	(0.0006)	-	(0.0006)	-
	γ-fit	100.00%					

Table 4.5: Higher-Order Markov Chain: Model 1 (Basic), with Covariates

	Variable	Price Estimate	Price Actual	X_1 Estimate	X_1 Actual	X_2 Estimate	X_2 Actual
Regime 1	$\mu_r, L1$	25.0005	25	0.6606	0.66	0.8602	0.86
		(0.0009)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L2$	25.0005	25	0.6600	0.66	0.8600	0.86
		(0.001)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L3$	24.9992	25	0.6599	0.66	0.8602	0.86
		(0.001)	-	(0.0003)	-	(0.0003)	-
	μ_s	21.9990	22	1.5196	1.52	0.8599	0.86
		(0.001)	-	(0.0003)	-	(0.0003)	-
Regime 2	$\mu_r, L1$	24.9974	25	0.6597	0.66	0.8598	0.86
		(0.0011)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L2$	25.0008	25	0.6604	0.66	0.8600	0.86
		(0.0012)	-	(0.0003)	-	(0.0003)	-
	$\mu_r, L3$	25.0003	25	0.6604	0.66	0.8602	0.86
		(0.0012)	-	(0.0004)	-	(0.0003)	-
	μ_s	20.0004	20	1.5192	1.52	0.8600	0.86
		(0.0011)	-	(0.0003)	-	(0.0003)	-
Regime 3	$\mu_r, L1$	29.9994	30	0.6592	0.66	0.8603	0.86
		(0.0013)	-	(0.0004)	-	(0.0005)	-
	$\mu_r, L2$	30.0005	30	0.6595	0.66	0.8592	0.86
		(0.0016)	-	(0.0005)	-	(0.0004)	-
	$\mu_r, L3$	30.0001	30	0.6602	0.66	0.8603	0.86
		(0.0015)	-	(0.0005)	-	(0.0004)	-
	μ_s	25.0005	25	1.5196	1.52	0.8602	0.86
		(0.0015)	-	(0.0004)	-	(0.0004)	-
Regime 4	$\mu_r, L1$	32.0026	32	0.6596	0.66	0.8598	0.86
		(0.0018)	-	(0.0006)	-	(0.0006)	-
	$\mu_r, L2$	31.9965	32	0.6605	0.66	0.8605	0.86
		(0.0019)	-	(0.0006)	-	(0.0005)	-
	$\mu_r, L3$	32.0011	32	0.6599	0.66	0.8602	0.86
		(0.0023)	-	(0.0006)	-	(0.0006)	-
	μ_s	24.9997	25	1.5202	1.52	0.8598	0.86
		(0.0016)	-	(0.0006)	-	(0.0006)	-
	γ-fit	99.48%					

Table 4.6: Higher-Order Markov Chain: Model 2 (Complex), with Covariates

Chapter 5

Conclusion

In this dissertation, I have contributed to the literature on industrial organization in several related ways. First, Chapter 2 looks in a detailed fashion at the role of sales in the pricing of perishable products. It makes several contributions, first by establishing that periodic sales are an important part of the pricing of perishable products. This is the opposite of the predictions several models of sales make for these kinds of products, and necessitates the second contribution: developing a model which can explain the periodic nature of sales for these kinds of products. I also demonstrate how this model fits into the set of other models which are used to study sales, and highlight some of the pros and cons of using these models. Next, I go beyond the theoretical model to show how the central causal connection necessary is supported by consumer choice data. This highlights the important, and largely overlooked, role that joint retail and consumer data can play in the study of sales. This also demonstrates the need for flexible heuristic tools to classify and study sales; a subject I return to in Chapter 4.

I next move from the area of traditional retail sales to the emerging world of selling products (or developing products to sell) through crowdfunding in Chapter 3. After reviewing the history and economic content of crowdfunding, I focus on the specific context of consumer crowdfunding. After defining this unique type of fundraising, I provide some of the first evidence that large contributions play a very important role in the financing of crowdfunding projects. This is interesting because not only does it have important policy consequences for the regulation and supervision of crowdfunding, but also goes against the way crowdfunding portrays itself as largely driven by many small consumers. This chapter demonstrates that large contributors are both very common, but also appear to follow a rational economic motivation: they seek to try and help projects succeed in reaching their goal. I develop an intuitive model of this behaviour, based on a modified consumer choice framework, which succeeds in capturing the main features we observe for large contributions. Next, I try and assess the extent to which large contributions are important in helping projects succeed. Using an instrumental variables framework, I find that they are

highly important; inverting the standard logic of crowdfunding. I also show that this is not purely due to their size, but also their timing, demonstrating sophistication on the part of large contributors.

In Chapter 4, I return to the question of sales, this time to consider their classification and the nature of reduced-form study in this area. I develop a new method for classification based on a hidden Markov model, which has two key advantages over other methods: first, it uses all of the information available, including not just the entire price sequence but also other variables of interest. Second, it allows for estimation of the connection between sales and other covariates *jointly* with the classification of sales. This allows for many interesting questions to be answered without the need for reduced-form estimation and its attendant uncertainties. I also demonstrate empirically that the performance of these kinds of models is very good, even in complicated environments. This shows the utility of my classification method for many areas of study, and highlights a new tool available for the economic study of topics like sales in industrial organization.

Bibliography

Aaker, D. A. (1996). Measuring brand equity across products and markets. *California management review*, 38(3):102–120.

Agrawal, A. K., Catalini, C., and Goldfarb, A. (2013). Some simple economics of crowdfunding. *National Bureau of Economics Working Papers*.

Ali, S. N. and Kartik, N. (2006). A theory of momentum in sequential voting. *UC San Diego, Stanford University*.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, pages 797–817.

Baum, L. E. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, 37(6):1554–1563.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171.

Belleflamme, P., Lambert, T., and Schwienbacher, A. (2013). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*.

Berck, P., Brown, J., Perloff, J. M., and Villas-Boas, S. B. (2008). Sales: tests of theories on causality and timing. *International Journal of Industrial Organization*, 26(6):1257–1273.

Bilmes, J. A. et al. (1998). A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126.

Bils, M. and Klenow, P. J. (2002). Some evidence on the importance of sticky prices. Technical report, National Bureau of Economic Research.

Blattberg, R. C., Eppen, G. D., and Lieberman, J. (1981). A theoretical and empirical evaluation of price deals for consumer nondurables. *The Journal of Marketing*, pages 116–129.

Bronnenberg, B. J., Dhar, S. K., and Dubé, J.-P. H. (2009). Brand history, geography, and the persistence of brand shares. *Journal of political Economy*, 117(1):87–115.

Bronnenberg, B. J., Dubé, J.-P. H., and Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. *The American Economic Review*, pages 2472–2508.

Callander, S. (2007). Bandwagons and momentum in sequential voting. *The Review of Economic Studies*, 74(3):653–684.

Cantillon, R. (2014). Kickstarter entering crowded irish scene. *Irish Times*.

Cappé, O., Rydén, T., and Moulines, E. (2005). *Inference in Hidden Markov Models*. Springer Science & Business Media.

Chang, J.-W. (2016). The economics of crowdfunding. *UCLA Job Market Paper*.

Chen, Z. and Rey, P. (2012). Loss leading as an exploitative practice. *The American Economic Review*, 102(7):3462–3482.

Chevalier, J. A., Kashyap, A. K., and Rossi, P. E. (2000). Why don't prices rise during periods of peak demand? evidence from scanner data. Technical report, National Bureau of Economic Research.

Ching, W. K. and Ng, M. K. (2006). *Markov chains*. Springer.

Cimon, D. A. (2017). Crowdfunding and risk. *CEA Working Paper*.

Conlisk, J., Gerstner, E., and Sobel, J. (1984). Cyclic pricing by a durable goods monopolist. *The Quarterly Journal of Economics*, pages 489–505.

Creal, D. (2012). A survey of sequential monte carlo methods for economics and finance. *Econometric reviews*, 31(3):245–296.

Danaher, P. J., Bonfrer, A., and Dhar, S. (2008). The effect of competitive advertising interference on sales for packaged goods. *Journal of Marketing Research*, 45(2):211–225.

Daniel, C. R., Cross, A. J., Koebnick, C., and Sinha, R. (2011). Trends in meat consumption in the usa. *Public health nutrition*, 14(04):575–583.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2):224–227.

DeGraba, P. (2006). The loss leader is a turkey: Targeted discounts from multi-product competitors. *International journal of industrial organization*, 24(3):613–628.

Drachen, A., Sifa, R., Bauckhage, C., and Thurau, C. (2012). Guns, swords and data: Clustering of player behavior in computer games in the wild. In *Computational Intelligence and Games (CIG), 2012 IEEE Conference on*, pages 163–170. IEEE.

Freedman, D. (2012). *Markov chains*. Springer Science & Business Media.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Graves, J. (2015). Crowdfunding: a structured approach.

Gubler, Z. (2013). Inventive funding deserves creative regulation. *Wall Street Journal*.

Hekman, E. and Brussee, R. (2013). Crowdfunding and online social networks.

Hendel, I. and Nevo, A. (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673.

Hendel, I. and Nevo, A. (2013). Intertemporal price discrimination in storable goods markets. *American Economic Review*, 103(7):2722–51.

Hosken, D. and Reiffen, D. (2004). Patterns of retail price variation. *RAND Journal of Economics*, pages 128–146.

Hotelling, H. (1990). *Stability in competition*. Springer.

Johnson, J. P. (2014). Unplanned purchases and retail competition. *Available at SSRN 2319929*.

Kashyap, A. K. (1994). Sticky prices: New evidence from retail catalogs. Technical report, National Bureau of Economic Research.

Kickstarter (2014). Press - kickstarter.

Kokkonen, T. et al. (2014). Microtransactions in an android game. Master's thesis, Karelia University of Applied Sciences.

Kuppuswamy, V. and Bayus, B. L. (2013). Crowdfunding creative ideas: The dynamics of project backers in kickstarter. *SSRN Electronic Journal*.

Lal, R. (1990). Price promotions: Limiting competitive encroachment. *Marketing Science*, 9(3):247–262.

Lal, R. and Matutes, C. (1994). Retail pricing and advertising strategies. *Journal of Business*, pages 345–370.

Lal, R. and Villas-Boas, J. M. (1998). Price promotions and trade deals with multiproduct retailers. *Management Science*, 44(7):935–949.

Lambert, T. and Schwienbacher, A. (2010). An empirical analysis of crowdfunding. *Social Science Research Network*, 1578175.

Leacock, S. (2016). *Sunshine sketches of a little town*. Read Books Ltd.

Leroux, B. G. (1992). Maximum-likelihood estimation for hidden markov models. *Stochastic processes and their applications*, 40(1):127–143.

Lescop, D. and Lescop, E. (2014). Exploring mobile gaming revenues: the price tag of impatience, stress and release. *Communications & Strategies*, (94):99.

Leygonie, C., Britz, T. J., and Hoffman, L. C. (2012). Impact of freezing and thawing on the quality of meat: Review. *Meat science*, 91(2):93–98.

Manjoo, F. (2010). Online deals for holiday shopping: Buyer beware.

Maruotti, A. (2007). Hidden markov models for longitudinal data.

Marwell, N. B. (2016). Competing fundraising models in crowdfunding markets. *University of Wisconsin-Madison Job Market Paper*.

McLachlan, G. and Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1):1–16.

Mollick, E. and Kuppuswamy, V. (2014). After the campaign: Outcomes of crowdfunding. *Available at SSRN*.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, pages 1417–1426.

Pesendorfer, M. (2002). Retail sales: A study of pricing behavior in supermarkets*. *The Journal of Business*, 75(1):33–66.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rao, H., Xu, A., Yang, X., and Fu, W.-T. (2014). Emerging dynamics in crowdfunding campaigns. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 333–340. Springer.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Salop, S. and Stiglitz, J. (1977). Bargains and ripoffs: A model of monopolistically competitive price dispersion. *The Review of Economic Studies*, 44(3):493–510.

Salop, S. and Stiglitz, J. E. (1982). The theory of sales: A simple model of equilibrium price dispersion with identical agents. *The American Economic Review*, 72(5):1121–1130.

Shelegia, S. (2012). Multiproduct pricing in oligopoly. *International Journal of Industrial Organization*, 30(2):231–242.

Shilony, Y. (1977). Mixed pricing in oligopoly. *Journal of Economic Theory*, 14(2):373–388.

Sinclair, B. (2014). Free-to-play whales more rational than assumed. *GamesIndustry.biz*.

Sweeting, A. (2012). Dynamic pricing behavior in perishable goods markets: Evidence from secondary markets for major league baseball tickets. *Journal of Political Economy*, 120(6):1133–1172.

Teicher, H. (1967). Identifiability of mixtures of product measures. *The Annals of Mathematical Statistics*, 38(4):1300–1302.

Titterington, D. M. (1985). *Statistical analysis of finite mixture distributions*. Wiley.

Tuttle, B. (2014). Why all those holiday deals aren't such good deals. *Time Magazine*.

Varian, H. R. (1980). A model of sales. *The American Economic Review*, 70(4):651–659.

Varian, H. R. (1989). Price discrimination. *Handbook of industrial organization*, 1:597–654.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162.

Wong, A. (2016). Wealth in the pantry: Implications of consumer inventory stockpiling for household savings. Technical report, Northwestern University Working Paper Series.

Xu, A., Yang, X., Rao, H., Fu, W.-T., Huang, S.-W., and Bailey, B. P. (2014). Show me the money!: An analysis of project updates during crowdfunding campaigns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 591–600. ACM.

Yoo, B., Donthu, N., and Lee, S. (2000). An examination of selected marketing mix elements and brand equity. *Journal of the academy of marketing science*, 28(2):195–211.

Zvilichovsky, D., Inbar, Y., and Barzilay, O. (2013). Playing both sides of the market: Success and reciprocity on crowdfunding platforms. *SSRN Working Paper*.

Appendix A

Appendix to "Sales and Perishable Products"

This appendix contains several sections which cover different aspects, extensions, and details for the main material in Chapter 2. Each section is independent of the others, except when noted, and constitute separate extensions of the underlying material. To summarize, in Section A.1 I develop a more realistic error model for price fluctuations than the Gaussian assumption made in the paper to test whether the error pattern is material or not. In Section A.2, I develop Assumption A.2.1, the conditions under which k-means clustering recovering the underlying pricing regimes in sales data. In Section A.3, I develop a duopoly version of the monopolistic model in Section 2.5, which also examines the role of competition and other kinds of discounting. Finally, in Section A.4, I extend the duopoly model to relax some of the assumptions about consumer behaviour, such as myopia or inventory restrictions.

A.1 A More Realistic Error Model

The distribution used in the model is very stark; it does not allow for measurement error, or the possibility of a misspecification of k_p within the regime. Denote $P(S_t = 1|R_t) = q$. Accordingly, I allow for the addition of a uniform error term $\epsilon \sim$ unif$[a, b]$ with $a \leq 0 \leq b$ which can occur with probability m:

$$Y_t|R_t = p - \delta S_t + M_t \epsilon_i$$

Where M_t is the event that an error has occurred. Note that the error term is not necessarily white noise; this is because, in general, pricing mistakes or errors are not equally likely to be above or below a given price point. This implies that we have the following distribution:

$$P(Y_t = y|R_t) = (1-m)f_0(Y_t = y|R_t) + mf_1(y)$$

$$f_1(y) = q \cdot u(a+p-\delta, b+p-\delta) + (1-q)u(a+p, b+p)$$

Where $u(a,b)$ is the PDF of the uniform distribution on $[a,b]$. The associated CDF is:

$$P(Y_t \leq y|R_t) = (1-m)F_0(y|R_t) + mF_1(y)$$

$$F_0 = \begin{cases} 0 & \text{if } y < p-\delta \\ q & \text{if } y \in [p-\delta, p] \\ 1 & \text{if } y \geq p \end{cases}$$

$$F_1 = q \cdot U(a+p-\delta, b+p-\delta) + (1-q)U(a+p, b+p)$$

Where $U(a,b)$ is the CDF of the uniform distribution on $[a,b]$. Let $Q_\tau(Y_t)$ denote the τ-th quantile of the distribution. Notice that:

$$\begin{aligned} Q_0(Y_t|R_t) &= p-\delta+a \\ Q_{0.5}(Y_t|R_t) &= p-\delta q+\frac{m}{2}(a+b) \\ Q_1(Y_t|R_t) &= p+b \end{aligned}$$

Now, in order to identify the distribution more easily, let's (A1) assume that $p+a > p-\delta+b$; that is, the largest negative deviation from the top half of the distribution is still larger than the greatest positive deviation from the bottom half of the distribution. Then, if we definite Q'_τ to be the quantile of the distribution condition on $Y_t \leq p-\delta+b$, then we know the quantiles of this lower distribution are $Q'_\tau = m\{(1-\tau)(a+p-\delta)+\tau(b+p-\delta)\}+(1-m)(p-\delta)$. Then, it is clear we can recover a, b and $c = p-\delta$ by examining the quantiles of this power distribution. But, given these terms, the three overall quantiles above have only three unknowns (q, p (or δ) and m). Thus, this distribution is identifiable from the distribution of prices.

In order to distinguish between different pricing regimes, let's assume that m, a, b are

the same across regimes. Then, $Q_1(Y_t|R_t) - Q_1(Y_t|R_t) = p(R_t) - p(R'_t)$, $Q_0(Y_t|R_t) - Q_0(Y_t|R'_t) = p(R_t) - p(R'_t) - [\delta(R_t) - \delta(R'_t)]$. Since for at least one of these parameters must differ, this implies that by considering the maximum, minimum, and span of the two distributions, they must be distinguishable. It is also convenient to assume that $Q_{0.5}(Y_t|R_t) \neq Q_{0.5}(Y_t|R'_t)$ if two distributions differ.

A.2 k-means Clustering on Pricing Regimes

For simplicity, let's assume that regimes have equal lengths (L), and consider the case of two sequential regimes, R_1 and R_2. Because any regime can be divided into two smaller regimes continuously, this is not a restrictive assumption. Furthermore, as will become apparent, the choice of considering two regimes immediately generalizes, since regimes are separated contiguously in time. Associate with regime i starting point t_i and ending point $T_i = t_i + L$, regular price p_i and discount δ_i. Let the discount occur with rate $\frac{1}{2} > q_i > 0$. Also, let's suppose that the observations have a time delay d; for example, if they are weekly, $d = 7$ while if they are daily, $d = 1$. Then, without loss of generality assume that $t_2 = T_1 + d$, and normalize $t_1 = 0$. Then, in the population the centroid of regime 1 is $C_1 = (\frac{1}{2}L, p_1 - q_1\delta_1)$ and the centroid of regime 2 is $C_2 = (\frac{1}{2}L + L + d, p_2 - q_2\delta_2)$. Now, a point P is classified as in regime 1 if $d(P, C_1) < d(P, C_2)$. Naturally, then ,we need only consider points on $[\frac{1}{2}L, \frac{3}{2}L + d]$ (since points on the far side of either centroid will mechanically be closer to the given centroid than the other option. So, consider $P_L = (L, Y_L)$, the last point in regime 1 or $P_{L+d} = (L + d, Y_{L+d})$ the first point in regime 2; these will be the critical classification points, since if these are classified correctly then all of the other points in the respective groups will also. Let's suppose initially that P_L is a regular price: $P_L = (L, p_1)$. Then, letting $\Delta = p_1 - p_2$

$$\begin{aligned} d(P_L, C_1)^2 &= (\frac{1}{2}L)^2 + (q_1\delta_1)^2 \\ d(P_L, C_2)^2 &= (\frac{1}{2}L + d)^2 + (\Delta + q_2\delta_2)^2 \end{aligned}$$

Then,

$$d(P_L, C_2)^2 - d(P_L, C_1)^2 = d^2 + dL + 2q_2\delta_2\Delta + q_2^2\delta_2^2 - q_1\delta_1^2 + \Delta^2$$

$$\implies d(P_L, C_2)^2 > d(P_L, C_1)^2 \iff d^2 + dL + 2q_2\delta_2\Delta + \Delta^2 > (q_1\delta_1)^2 - (q_2\delta_2)^2 \tag{A.2.1}$$

Thus, this point is classified as long as, relative to the length of the regime and the gap d, the change in the prices relatively larger and positive. Essentially, the change in the prices needs to be larger than the change in the average deviation from the regular price, controlling for d and L. For instance, suppose $\delta_1 = \delta_2$ and $q_1 = q_2$. Then, this condition reduces to:

$$d^2 + dL + 2q_2\delta_2\Delta + \Delta^2 > 0$$

Since,

$$d^2 + dL + 2q_2\delta_2\Delta + \Delta^2 > d^2 + dL + 2q_2\delta_2\Delta$$

it is sufficient for $\Delta > \frac{-d^2-dL}{2q_2\delta_2}$.In other words, if we believe that only the regular price is changing, and everything else is respectively the same, this amount to not allowing the price to fall by more than $\frac{d(d+L)}{2q_2\delta_2}$. For instance, an inflationary set of price regime changes would suffice to meet this condition.

Now, when P_L is a sale price, $P_L = (L, p_1 - \delta_1)$ and:

$$d(P_L, C_1)^2 = (\frac{1}{2}L)^2 + (p_1 - (1 - q_1)\delta_1)^2$$

$$d(P_L, C_2)^2 = (\frac{1}{2}L + d)^2 + (\Delta + q_2\delta_2 - \delta_1)^2$$

$$\implies d(P_L, C_2)^2 > d(P_L, C_1)^2 \iff d^2+dL+(p_1-(1-q_1)\delta_1)^2-(\Delta+q_2\delta_2-\delta_1)^2 > 0 \tag{A.2.2}$$

Similarly, for the other point:

$$(\frac{1}{2}L + d)^2 + (\Delta - q_1\delta_1)^2 > (\frac{1}{2}L)^2 + (q_2\delta_2)^2 \tag{A.2.3}$$

$$(\frac{1}{2}L + d)^2 + (\Delta - q_1\delta_1 + \delta_2)^2 > (\frac{1}{2}L)^2 + ((1 - q_2)\delta_2)^2 \tag{A.2.4}$$

Which can be simplified as the preceding expression. The following assumption then ensures that the k-means clustering algorithm will successfully identify clusters:

Assumption A.2.1. Equations A.2.1-A.2.4 are met for all regimes in the data.

A.3 A Duopoly Model of Loss Leadership

My duopoly model of loss leadership is based on that of DeGraba (2006); in it, two monopolisitically competitive grocery stores compete over a population of consumers who buy different bundles composed of perishables and sundries. To capture the intuition, suppose that these consumers come in two types (j): (1) bachelors ($j = B$) and (2) families ($j = F$). These two types of consumers have different lifestyles: bachelors live alone, in small apartments, while families live together in larger houses. This manifests in two related ways: first of all, families have more storage capacity for the sundries than bachelors, and secondly, the families consume more perishables per capita from the grocery store, since the bachelors find it difficult to cook and prepare for one (and so eat out more, instead). These basic differences between the consumers will be ultimately result in sales, since inventory dynamics on the part of the families lead them to occasionally purchase more than the bachelors. The fact that they also purchase meat gives the stores a way to price discriminate between the two consumers, offering a more attractive total bundle price to the families when they are buying large amounts of sundries. High purchases of sundries occur in a cyclic fashion alongside low prices of the perishable good, resulting in a cyclic pattern for sales.

To make things concrete, consider a discrete time environment $t = 1, 2, 3, ...$, and suppose that bachelors always demand one unit of the sundry good every period $S_{Bt} = 1$ since they cannot store goods very well. Similarly, suppose that that families always demand one unit of the perishable product every period (since it expires, and cannot be stored) while the bachelors never demand the perishable product. The families also buy the sundry good, but do so in an inventory-based fashion; that is, they have an inventory level I_t which is replenished to some (full) level K when the sundry good is purchased. The demand per period is denoted $S_{Ft} \geq 0$ and depends on the number of families buying the sundry in a given period. Both types of consumers are risk-neutral, expected utility maximizers, and suppose they occur in masses m_j.

On the firms' side, I model competition using a standard Hotelling model (as in Hotelling

(1990)) where the stores are located at either end of the unit line. Stores compete in prices over two goods $i = M, S$, and I assume that grocery stores face constant marginal costs of production for both goods, c^j, and the stores are risk-neutral profit maximizers. Consumers are located uniformly along the unit line, and face a marginal cost of transport $k > 0$, which is the degree of differentiation between the firms. The timing of the game is demand realization, price-setting, purchasing, and then consumption.

The decision framework of the agents depends on the degree of consumer sophistication and the way in which sundries deplete. The simplest benchmark is to suppose that consumers are completely myopic, and follow a heuristic rule to replenish their inventories only when they they have exhausted their supply. Economically, this could correspond to a situation where it is only cost-effective for the families to refill their supply when they run out; for instance, if there is only one package size available it is very "costly" to have excess product spilling out of the inventory. To model the family inventory, suppose each period a representative family receives an independent shock to their expected consumption, $s_j \in \{0, 1, 2, ...K\}$ which is the amount of the sundry they will use up in the upcoming week. If they have inventory level I, then they will need to purchase a new package of sundries when $s_j \geq I$, since otherwise they will run out of the good.

If we consider the inventory level the associated state variable, then the probability a family with inventory I purchases the sundry is $P(I) \equiv \sum_{j=I}^{K} s_j$. The path of their inventory state is determined by a Markov chain, $I_{t+1} = \mathbf{M} I_t$ where the transition matrix $\mathbf{M}$ is given by

$$\mathbf{M} = \begin{bmatrix} s_K & s_{K-1} & s_{K-2} & s_{K-3} & \cdots & s_1 & s_0 \\ 0 & s_0 + s_K & s_{K-1} & s_{K-2} & \cdots & s_2 & s_1 \\ 0 & s_1 & s_0 + s_K & s_{K-1} & \cdots & s_3 & s_2 \\ 0 & s_2 & s_1 & s_0 + s_K & \cdots & s_4 & s_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s_K & s_{K-1} & s_{K-2} & s_{K-3} & \cdots & s_1 & s_0 \end{bmatrix}$$

A critical feature of this matrix is the value of s_0. If $s_0 = 0$, then the associated Markov chain is aperiodic and may not admit a stationary distribution; the number of individuals in each state may never settle down, rotating forever between different values. More importantly, if the Markov chain does have a stationary distribution, then there is no requirement that it is also a limiting distribution. This will be very important for the dynamics of sales.

Denote the distribution of families with a given inventory by the vector x. Then, the probability a randomly selected family purchases the product is $\sum_{i=0}^{K} P(i)x_i$; thus, in aggregate they purchase this many units, multiplied by their mass and K. Now, because consumers are myopic, they do not inter-temporally adjust their purchasing decisions. Thus, no decision on the part of the firms will effect the choices from period-to-period, and firms will play a repeated static game against one another. We therefore have the following results:

Proposition A.3.1. The price of the leading good increases or decreases depending on the value $r(x) \equiv \sum_{i=0}^{K} P(i)x_i$, independently of m_j.

Proof. As in DeGraba (2006), the utility function of firm i is given by:

$$\Pi_i = (P_{Mi} - c_M + S_{Ft}(P_{Si} - c_S))(\frac{1}{2} + P_{M(-i)} - P_{Mi}$$

$$+S_{Ft}(P_{Si} - P_{S(-i)})/2k)m_F + (S_{Bt}(P_{Si} - c_S))(\frac{1}{2} + S_{Bt}(P_{S(-i)} - P_{Si})/2k)m_S$$

which has associated first order conditions

$$P_{Mi} - c_M + S_{Ft}(P_{Si} - c_S) = k$$

$$S_{Bt}(P_{Si} - c_S) = k$$

and imply that:

$$P^*_{Mi} = c_M + k(1 - \frac{S_{Ft}}{S_{Bt}})$$

$$P^*_{Si} = c_S + \frac{k}{S_{Bt}}$$

However, as we established earlier, the average consumption of S by the F-type consumers is

$$S_{Ft} = K \sum_{i=0}^{K} P(i)x_t(i)$$

and similarly

$$S_{Bt} = 1$$

The result immediately follows by inspection since $P^*_{Mi} = c_L + k(1 - S_{Ft}) = c_L + k(1 - K\sum_{i=0}^{K} P(i)x_t(i))$. □

This leads to the conclusion that sales depend varying levels of demand for the *associated* sundry good. It also means that we have the following results which illustrate directly the importance of the periodicity of the inventory process:

Corollary A.3.1. (1) Suppose that the inventory Markov chain $\mathbf{M}$ is periodic and the distribution of consumers do not start in a stationary distribution for the Markov chain. Then, sales on the perishable product occur in a cyclic pattern forever. (2) Suppose that the inventory Markov chain $\mathbf{M}$ is aperiodic; then, sales will occur temporarily, before eventually ceasing.

The condition that the inventory process be periodic may seem restrictive, but it is actually not; for example, it is sufficient that $s_0 = 0$. Many common inventory processes fall into this category; for instance, a deterministic process in which families always consume one unit of the sundry falls is periodic. Similarly, a situation where the family may consume one or two units is periodic as long as the package size is three or higher. If this condition does not hold, then sales may occur temporarily but eventually stop as the the distribution of inventory across consumers becomes stationary. In the stationary distribution, although individuals will go through cycles of buying products, the aggregate population will not which results in permanently lower prices which are stable.

This model also generates some predictions about sales; first of all, they should be cyclic, since they occur alongside the periodicity of the inventory process. However, this is not unique; many different models make predictions that sales should be cyclic. Fortunately, this model also make predictions about the composition of the consumer's bundle: specifically, among the perishable-purchasing consumers, they should also purchase more of other sundry goods when the perishable good is on sale. This prediction is distinct from other models, especially inventory-based explanations for sales, because it tells us that there should be a link between sales and the kinds of goods being purchased in the *whole* bundle (and not just consider the discounted product alone). This prediction is also testable, although not at the retail level; by looking at the consumers' choice bundles, we

should be able to assess whether not consumers purchase more sundries when the perishable is on sale, or not. This is summarized in the proposition below:

Corollary A.3.2. In periods when consumers purchase perishable products on sale the demand for the sundry products should be higher, when compared with periods where consumers purchase the perishable product at its regular price. In terms of the model, then $S_{Ft} > S_{Ft'}$ when the product is on sale at time t and not at time t'.

A.3.1 Discussion

This model produces fairly precise predictions about what conditions are necessary for sales; this precision is largely due to the starkness of the assumptions made on the different parts of the model. This starkness is used to focus on the intuition and economics of the model, which are complicated even in this basic environment. Essentially, consumers differ in a single dimension (the "type"), which is associated with both their ability to store and their desire to purchase perishable products. In the model, I call this the family/bachelor distinction, but this is just a framing mechanism for the intuition; in reality, grocery stores, using consumer loyalty cards and purchasing information, are likely able to come up with far more subtle connections between consumer demands. The basic intuition of the model is that consumers who hold sundry inventory in a periodic fashion will go through periods of higher purchases of the sundries; as consumers run out, their willingness to pay increases, which manifests in an increase in demand. In the myopic model, this variation is quite stark, but the story is similar if this is relaxed. When consumers demand larger bundles, they are more valuable to attract to your store; competition between the two stores creates the incentive for them to lower the price of this bundle to attract these valuable consumers.

However, because firms can choose prices for both the sundry goods and the perishable goods, the problem becomes to decide which price they should lower in order to change the price of the total bundle. The optimum equilibrium behaviour uses the fact that perishable purchasing distinguishes the two types of consumers. To see this, imagine that initially firms just tried to sell to the bachelors; the optimal price would be as given above as the sundry price. Then, supposed they decide to just sell to the families, taking the sundry price as given. The equilibrium price would be as given, again. If a firm chose to lower P_S and raise P_M they would lose profit from the bachelors while keeping the profit for the

families the same[47]. Intuitively, changing P_M has no impact on profit from the bachelors, and so it is optimal to effectively "segregate" the market into two parts; this is similar to how a monopolist would try to discriminate between the two groups of consumers.

As mentioned, the dynamic aspect of sales is generated primarily through the inventory process. In the myopic model, this process is easy to translate from the underlying stochastic demands because there is no inter-temporal decisions made by the consumers, and therefore also by the firms. The inclusion of forward-looking behaviour on the part of consumers changes this demand process significantly, but the real challenge arises in the behaviour of the firms, who now may compete inter-temporally as well. I explore this model in the Appendix, illustrating that while myopia gives a simpler characterization, it is not essential; however, firms need to be long-lived in order to close down inter-temporal competition.

The assumption about the inventory process of the bachelors being very stark (no storage) is also not essential. The essential requirements are the same as in the simpler model, but the statements have to do about the differences in the stationary or periodic distributions of the inventory process. One way to see this intuitively is to pretend that that the assumption that the bachelors always demand a single unit of the sundry is a "normalization" instead of an assumption; then, Proposition 15 becomes a statement of relative, as opposed to absolute demand. This variation of the model is developed further in the appendix.

Related to the above is the question of firm expectations; in the model, we assume that consumers basically commit to making trips to the store, then firms set prices to attract them to their particular store. This is not really realistic, since firms actually are uncertain about demand (which occurs over a period of time) and set prices ahead of time. The assumption that consumer demand is observable is probably not very realistic; this is best understood to be expectations about consumer demand, which are (in equilibrium) correct. The fact that consumers form a mass with infinite density makes this possible. However, how reasonable are these expectations? Would firms understand when demand is high or low? This is difficult to assess directly, but we can imagine that firms have good planning processes and understand (in aggregate) what consumer replenishment of particular products must look like. This could be from features like loyalty cards or club cards. This is the process by which they would understand the perishable/sundry connection in

[47] See DeGraba (2006) for a detailed explanation of this in the static context, which follows the same logic.

the first place, so it would also be reasonable for them to make predictions of consumer demand. Indeed, in equilibrium, these could be based explicitly only on past demand and pricing patterns, since these would re-occur over time.

A.3.2 Quantity Discount versus Sales

A second comment has to do with the use of price discounts on different products versus quantity discounts. Because consumers are myopic, the families refill their inventory entirely; this is equivalent to assuming they purchasing a package of size K at the store. However, they pay a KP_S which is equivalent to buying K units of the bachelor's package size. It would be sensible that an alternative way for the firms to compete would be in terms of a quantity discount for the package. This is certainly possible, and could be an alternative pricing strategy. However, any uniform discount (δ for the larger package) across both firms would result in the same incentive to discount perishable products, not resulting in a change in the model. More importantly, it cannot increase profit since the equilibrium strategy maximizes profit for both types of consumers individually; any segregation on the basis of bundle size rather than bundle composition can at best recover the same profit level. This means this is more of an "alternative" strategy which could be adopted for competitive reasons. In this model, these are equivalent; however, relaxation of the inventory size assumption for the bachelors makes this no longer true; perishable sales become more attractive again.

This can be formally stated in the following set up: suppose that there is a measure m_{SB} of "storing bachelors;" that is, bachelors who have the ability to store the sundry good. These are similar to the agents in the model developed in the appendix where both types of agents can store. Now, consider the limiting value as $m_{SB} \to 0$; for small values of m_{SB} the aggregate per-bachelor demand is $\frac{m_B}{m_B+m_{SB}} + \frac{m_{SB}}{m_B+m_{SB}} S_{SB} \approx 1$ which means the equilibrium is approximately the same as above. Then, we have the following result:

Proposition A.3.2. Suppose $m_{SB} \to 0$. Then, sales on sundries weakly dominate quantity discounts.

Proof. Let the equilibrium prices in the original model be p_m^* and p_s^*. Then, bachelors a total price of p_s^* and families pay a total price of $p_m^* + kp_s^*$. In equilibrium, we know that these two bundle prices constitute an equilibrium. So, consider the alternative $p_{s1} = p_s^*$, $p_m = p = \sup\{p_m^*\}$ (a constant) and $p_{sk} = kp_s^* + p_m^* - p$. Notice that $p_{sk} > p_s^* = p_{s1}$

and $p_m - p_m^* \geq 0$. That is, bachelors do not want to purchase the bundled sundry, and the meat price is fixed (and higher) than the meat price in the sales model. However, notice that the total spending for both consumer bundles is the same as in the original equilibrium. Therefore, sales do as well as quantity discounts. However, for the m_{SB} bachelors who purchase the inventory-sized bundle, quantity discounts do strictly worse, since $p_{sk} < kp_s^*$ which is their spending in the original equilibrium. Therefore, sales are preferred to quantity discounts. □

This is intuitive, because a quantity discount on a product bought by both consumers may attract the smaller customers to "buy up" which is undesirable if an alternative exists; only the stark assumption on the bachelor's inventory makes this possible. Of course, in the real world, the assumption about complete separation of perishable purchasing by consumers is also too stark; a relaxation of both these assumptions likely explains why, in the real world, we see both sales and quantity discounts being used.

A.3.3 Competition and Monopoly

The simplest way to examine the importance of the assumption that competition is monopolistic is by considering the limiting properties of k, the differentiation parameter. As $k \to 0$, firm differentiation is eliminated and the market become Bertrand competition. As in the standard model, equations 1 and 2 should that the prices approach the marginal costs as markets become more competitive. In particular, because prices become constant this eliminates sales both intermittently and periodically. This occurs because loss leadership (and sales in general) are only useful through attracting valuable customers to your store. These customers are attracted by a lower price; however, there is a trade-off between price and location. This leads stores to shade their pricing, which in turn creates sales. When all consumers are equally price sensitive, small deviations in price become more valuable, resulting in a Bertrand-style race to the bottom which also eliminates price variation.

At the other extreme, as $k \to \infty$ the market becomes monopolistic. In this case, the optimal price for a bundle approaches M_j the maximal consumer willingness to pay for a bundle. This result holds in both the myopic and forward-looking models under commitment. Since consumers must buy the product at some point, the monopolist appropriates all the surplus. One issue is that since consumers of different types may have different willingnesses to pay, the monopolist will effectively cross subsidize if $M_f < M_b$, setting

$P_s = M_b$ and $P_m = M_f - kP_s$ as long as $P_m > c_m$. If this does not hold, it is not optimal for the monopolist to offer perishables. However, choice of product line is beyond the scope of this model, so we can set this case aside. Since these prices are fixed, monopoly leads to a similar situation as competition: fixed prices. This eliminates sales of both kinds, just as in the case when $k \to 0$ but for the opposite reason. Sales as useful because they allow stores to compete over consumers: if there is no need to compete, there's no need to offer sales.

A.4 Extensions to the Duopoly Model of Loss Leadership

This appendix introduces several extensions to the duopoly loss leadership model. The basics are essentially the same, but are summarized below for completeness; the key difference is (1) in an allowing the bachelors to store inventory and (2) relaxing the assumption on myopic consumers. These two variations are explored together, in the following sections. To make things more clear, and to highlight the differences in the model, I use the notation L and S to refer to the leading goods (the perishable) and the side good (the sundry); the consumers, who now differ only in their taste for the perishable, are identified with whether or not they consume that good. In the context of the original model, the bachelors would be the S-consumers, while the families would be the L-consumers. The remainder of the model is similar; changes are pointed out where they are relevant.

These stores are monopolisitically competitive, which is represented by a standard Hotelling model where the stores are located at either end of the unit line. Stores compete in prices over two goods indexed by j: the leading good L and the sundry good S. I assume that grocery stores face constant marginal costs of production for both goods, c^j, and the stores are risk-neutral profit maximizers. Consumers are located uniformly along the unit line, and face a marginal cost of transport $\tau > 0$. Each consumer demands S^j units of good j, and is a risk-neutral expected utility maximizer.

I assume that the preceding occurs within a discrete time framework, $t = 1, 2, 3, ...$, and the two goods differ in terms of their durability. The sundry good S is inventoriable; each consumer has an inventory level I which can be replenished to a maximum amount K by purchasing the sundry. Unlike in the baseline model, because both consumers buy the same type of good, I omit package-size or unit considerations; there is only one size (K) available, sold for a single price. I assume there are two kinds of consumers: L-buyers and S-buyers. These consumers differ in their tastes over the two goods. Specifically, the

L-buyers always demand one unit of L every period, while the S-buyers never demand L. However, both types of consumers consume S, albeit in different amounts S_j depending on their decision to buy the sundry good. These consumers occurs in measures m_j, uniformly distributed along the unit line. Within each period, I assume there is a demand realization step, a purchasing step, then a consumption step. So, for instance, a consumer will purchase and add to their inventory before consuming any goods.

The decision framework of the agents depends strongly on how the degree of consumer sophistication, the way in which sundries deplete, and idiosyncratic noise. The simplest benchmark is to imagine that consumers are completely myopic, and follow a heuristic rule to replenish their inventories only when they they have exhausted their supply. I develop this in the following section.

A.4.1 Myopic Consumers, Both with Inventories

In this section, suppose that inventories deplete according to a random walk, which can differ for each type of consumer (I suppress this at the moment, for clarity). Unlike in the baseline model, now both consumers are capable of holding inventory. Each period, a consumer may need to consume j units of the sundry, $j \in \{0, 1, 2, ...K\}$, independently of any other instance of consumption, each with probability s_j. Then, for an individual with inventory I the probability that they purchase the sundry is $P(I) \equiv \sum_{j=I}^{K} s_j$. If we imagine the inventory levels for a consumer are states, the path of their inventory, under optimal decision making, forms a discrete time Markov chain with transition matrix:

$$\mathbf{M} = \begin{bmatrix} s_K & s_{K-1} & s_{K-2} & s_{K-3} & \cdots & s_1 & s_0 \\ 0 & s_0 + s_K & s_{K-1} & s_{K-2} & \cdots & s_2 & s_1 \\ 0 & s_1 & s_0 + s_K & s_{K-1} & \cdots & s_3 & s_2 \\ 0 & s_2 & s_1 & s_0 + s_K & \cdots & s_4 & s_3 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ s_K & s_{K-1} & s_{K-2} & s_{K-3} & \cdots & s_1 & s_0 \end{bmatrix}$$

In general, $\mathbf{M}$ may different for the different types of consumers as $\mathbf{M}_j$. Notice, this Markov chain is recurrent; all states are visited with positive probability provided $s_0, s_K > 0$. Under this condition, this chain is also aperiodic. Then, this Markov chain has an associated stationary distribution π where:

$$\pi_i = \frac{C}{M_i} = \lim_{n\to\infty} \mathbf{M}_{ii}^{(n)}$$

where M_i is the mean recurrence time and C the normalizing constant. If the preceding condition is not met, then the Markov chain is periodic and may not admit a stationary distribution; the number of individuals in each state may never settle down, rotating forever between different values. More importantly, if the Markov chain does have a stationary distribution, then there is no requirement that it is also a limiting distribution.

Suppose the distribution of consumers with a given inventory is given by the vector x. Then, the probability a randomly selected consumer of type j purchases the product is $\sum_{i=0}^{K} P_j(i)x_i$; thus, in aggregate they purchase this many units, multiplied by their mass. Now, because consumers are myopic, they do not inter-temporally adjust their purchasing decisions. Thus, no decision on the part of the firms will effect the choices from period-to-period, and firms will play a repeated static game against one another. We therefore have the following results:

Proposition A.4.1. The price of the leading good increases or decreases depending on the ratio $r(x^L, x^S) \equiv \frac{\sum_{i=0}^{K} P_L(i)x_i^L}{\sum_{i=0}^{K} P_S(i)x_i^S}$, independently of m_j.

Proof. The proof is identical to before; the first-order conditions imply that the prices are:

$$P_{Li}^* = c_L + k(1 - \frac{S_{Lt}}{S_{St}})$$

$$P_{Si}^* = c_S + \frac{k}{S_{St}}$$

However, as we established earlier, the average consumption of S by the L-type consumers is now:

$$S_{Lt} = \sum_{i=0}^{K} P_L(i)x_{Lt}(i)$$

and similarly

$$S_{St} = \sum_{i=0}^{K} P_S(i)x_{St}(i)$$

The result immediately follows by inspection. □

These results yield a variation of the original proposition.

Corollary A.4.1. (1) Suppose $\mathbf{M}_j$ is periodic, x_j does not begin in a stationary distribution of $\mathbf{M}_j$, and either (i)$\mathbf{M}_{-j}$ is periodic and $x_{-j} \neq x_{-j}$ or (ii) $\mathbf{M}_j$ is aperiodic. Then, sales will re-occur infinitely for $j = S$ or $j = L$

(2) If the conditions of (1) do not hold, then sales will either (i) never occur, and prices will be stable or (ii) occur temporarily, diminishing before

We can see the consequences of these different outcomes in Figure A.1. The idea of this result is that the properties of the Markov chain determined by consumer inventory choices translates into statements about sales. Sales, in this context, are reductions in the price of a given product. In this model, this corresponds to periods where the ratio $r(x^L, x^S)$ is smaller than other periods. This depends on the distribution of individuals in each state x^j; as the Markov chain causes these distributions to evolve, so too does the ratio, and therefore the prices. However, if the Markov chain is ergodic, the long-term (and the steady state) distribution of both types of consumers becomes more and more similar which causes the $r(x^L, x^S)$ to approach unity, which means that price variation eventually ceases. This can occur in two ways; demands rising or falling for the two types of consumes; it is the relative size which is critical here.

Again, the key requirement is that the inventory accumulation be aperiodic. Consumer inventories need to rotate, in order to generate cyclic patterns of sales. If they do not, then the Markov chain eventually converges to a steady state; the duration of this period (and the length of time in which there are detectable sales) is governed by the eigenvalues of the inventory process. We can see, as discussed previously that the essential character of the model is unchanged, although the interpretation and some of the predictions are more subtle. In particular, the prediction of when sales should occur is now *relative* to the customer groups; this implies that the original prediction is a sufficient, but not necessary, condition for there to be sales. In other words, sales can still occur even without the central prediction of the baseline model; the behaviour necessary is, however, very difficult to detect from data.

A.4.2 Forward-looking Consumers

In this section, suppose that inventories deplete according to a random walk, as in the myopic model, with consumers at time t requiring c_t units of the sundry. However, instead

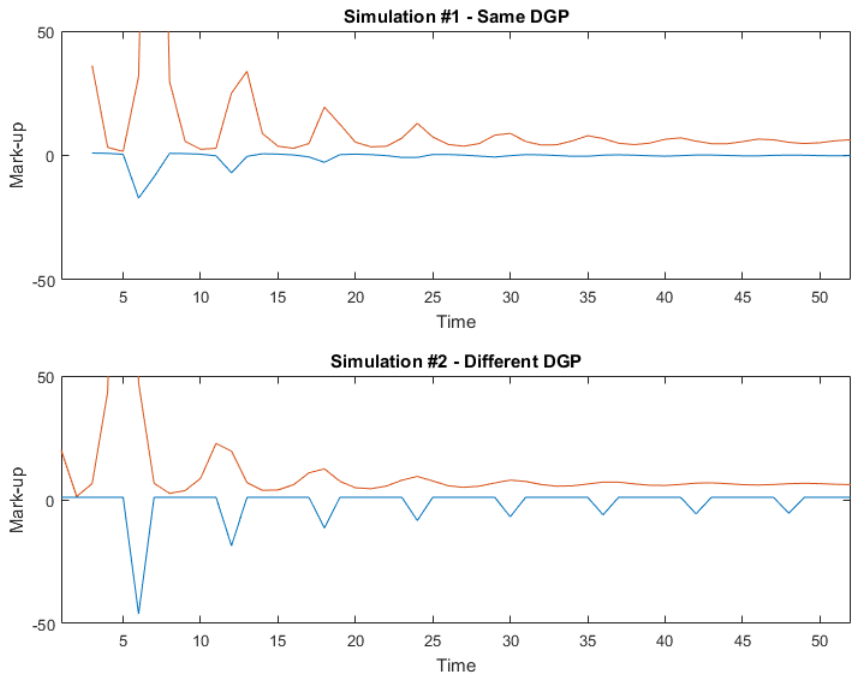


Figure A.1: Different Sales Patterns

of assuming consumers are myopic, suppose that they live for T periods, at which point they die and are replaced with more consumers. Furthermore, let's suppose that consumers who meet their demand for the good in a given period obtain utility u, but are punished with dis-utility v if they run out of inventory in a given period. I denote the price of the bundle by P, understanding that for L-type consumers the utility includes the price of the L-good (which they must purchase each period). All consumers share a common discount factor $\beta < 1$, while firms have a discount rate $\delta \leq 1$. Furthermore, following Hendel and Nevo (2006), let's suppose that consumers believe prices follow a first-order Markov process. Then, consumers solve a dynamic optimization problem with Bellman equation:

$$V_t(I_t, c_t, P_t) = \max_{a_t \in 0,1} u - I(I_t \leq c_t)(1 - a_t)v - P_t a_t + \beta E V_{t+1}(I_{t+1}, c_{t+1}, P_{t+1})$$

$$I_{t+1} = a_t K + (1 - a_t) I_t - c_t$$

$$V_T(I_T, c_T, P_T) = \begin{cases} \max\{u - v, u - P_T\} & \text{if } I_T - c_T \leq 0 \\ u & \text{if } I_T - c_T > 0 \end{cases}$$

The profit of the firm at time T is given by the static solution found above, and is therefore:

$$\Pi^* = \frac{1}{4} m_L + \frac{1}{4} m_S$$

which does not depend on the number of consumers purchasing at time T. Then, we can find an equilibrium as follows:

Proposition A.4.2. Suppose firms are very long-lived, relative to consumers, so $\delta = 1$. Then, there is an rational expectations equilibrium in which firms set prices according to Proposition 14, and consumers behave according to the Bellman equation above.

Proof. Since at time T, the only equilibrium action for the firms is to play according to the Proposition 14, the result holds in this period. But, in this case, there is no change in the last period profit as the inventory varies. Therefore, at time $T - 1$, if the firms set prices according to Proposition 14, there is no difference in obtaining the profit now or in the future since $\delta = 1$. Thus, setting prices according to Proposition 14 is optimal in period

$T-1$. The result follows, since the Bellman equation given above is a contraction mapping, and consumers are rationally correct in their expectations about the pricing process. □

The basic intuition behind this result is that in the terminal period of the consumer's lives, firms do not compete against themselves in the future for the current consumers. This fixes their profit at a level given by the static equilibrium of the Hotelling game. The equilibrium of this game is such that it does not depend on the relative amounts the consumers buy, since the surplus is competed away by the two firms, So, when firms are sufficiently long-lived, they do not care whether they buy from consumers in the current period, later, or earlier. This allows the static equilibrium outcome to persist, because inter-temporal competition does not affect the profitability of the firms. This kind of assumption is similar to Blattberg et al. (1981), in that the management of the inventory of consumers is viewed by firms as being parallel to the inventory in-store: essentially, because firms are aware consumers will buy at some point, goods can be "earmarked" for future consumption, making it immaterial when the good is actually moved from shelf to pantry (unless there are inventory issues on the part of the firm, which is the focus on the paper).

From the Bellman equations, it is possible to show the following result:

Corollary A.4.2. Consider a particular time t. Then, consumers are more likely to buy the sundry good when the price is low at time t.

Proof. This follows from the fact that, since consumer form a unit mass (with measure m_j) an individual consumer decision will not affect the overall distribution of states or demand, both now or in the future. Now, at time T, we can see that $dV_T(I_T, c_T, P_T)/dP_T \leq 0$. Now, at time $T-1$, if the price goes up, fewer consumers will buy now, therefore increasing the number buying at time T. An increase in the number buying at time T lowers the price of the sundry. So,

$$\frac{dV_{T-1}(a_t = 1)}{dP_{T-1}} = -1 + \beta E \frac{dV_T}{dP_{T-1}} = -1 + \beta E \frac{dV_T}{dP_T} \frac{dP_T}{dP_{T-1}} = -1 + \beta(+)(-) < 0$$

This means that an increase in the price lowers the propensity to consumer the good. Repeating this process backwards in time produces the result. □

Appendix B

Appendix to "Large Contributions and Crowdfunding Success"

B.1 Large Contribution Size

Similar to the literature on crowdfunding developed in Graves (2015), a benchmark model of the process of backer arrival is to consider a Poisson process. Suppose that backers arrive at the project on a given date t at a Poisson rate λ_t. Furthermore, suppose each of them has a "most-preferred" option to back the project - which is either low L or high $H > L$, with H much larger than L. Suppose the probability of a person having valuation V is p_V. Then, the arrival rate of L-backers is $p_L\lambda_t$ and the arrival rate of H-backers is $p_H\lambda_t$. Furthermore, suppose that the amount V-backer i contributes is $V+\epsilon_i$, where $\epsilon_i \sim N(0, \sigma_V)$ is a normally distributed white noise process. Then, for a given number of V-backers, the per capita contribution is normally distributed with mean V and variance σ_V. Suppose that $p_L \sim 1$ and λ is bounded for all t. Then, the probability in a discrete number of periods of an H-backer arriving is zero. Thus, under the standard assumption, the average donation should be normally distributed as above. Therefore, the likelihood of a "large" contribution coming from the underlying L distribution is 0.3%. Given that H is large, the Bayesian probability that a large donation comes from the L distribution is:

$$Prob(V_i = L|D_i > L+3\sigma_L) = \frac{Prob(D_i > L + 3\sigma_L|V_i = L)P_L}{Prob(D_i > L + 3\sigma_L|V_i = L)P_L + Prob(D_i > L + 3\sigma_L|V_i = H)P_H}$$

$$\implies Prob(V_i = L|D_i > L + 3\sigma_L) = \frac{0.0003P_L}{0.0003P_L + (1 - \Phi(\frac{L+3\sigma_L-H}{\sigma_H}))(1 - P_L)}$$

$$\sim \frac{0.0003P_L}{1 - 0.9997P_L} \sim \frac{0.0003}{\frac{1}{P_L} - 0.9997}$$

which is very small in general: for $P_L = 0.99$, this is still less than 5%, and for $P_L = 0.995$, it is less than 8%. The break-even point (where H and L are equally likely) occurs at around 0.9997.

B.1.1 Discussion and Size Considerations

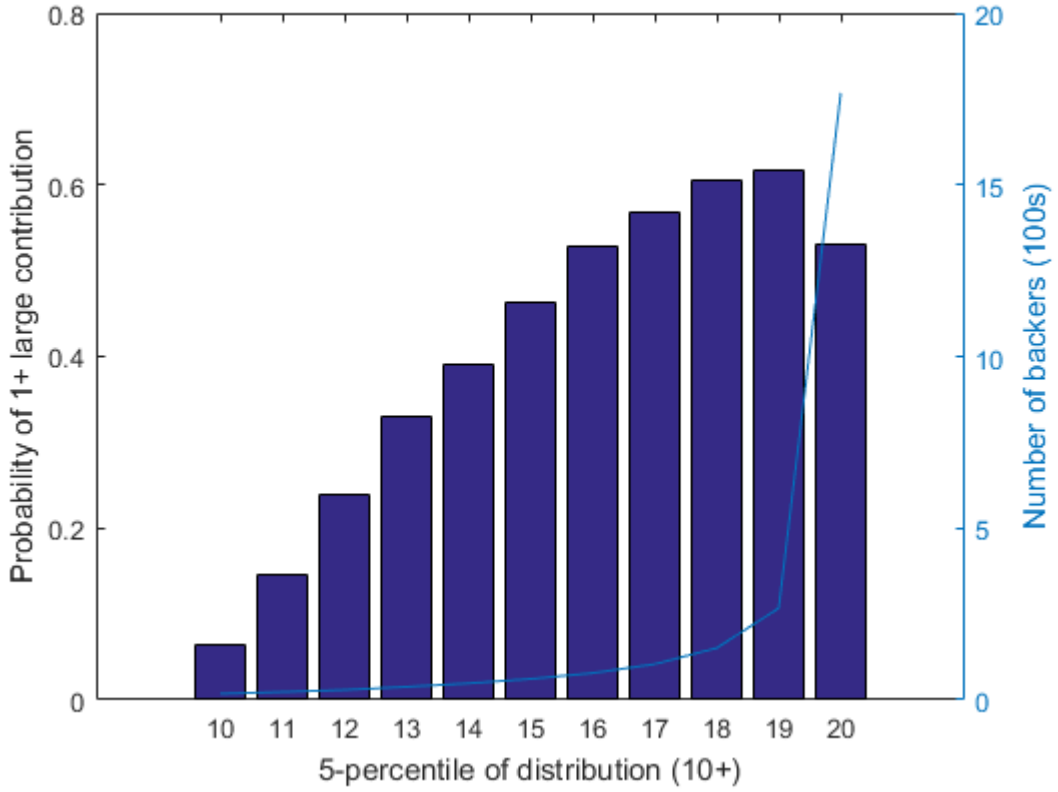
One challenge with interpreting the large contributions in the data is that they are determined (essentially) as being outliers from the mean of the time series of per-capita contributions. Furthermore, this time series is computed based on averages; essentially, they are outliers from an average of averages. This means that, in particular, this indicates that the large contribution measure tends to be an *underestimate* of the number of large contributions, since on a particular day a large number of small backers can "dilute" a large contribution. This mainly affects the interpretation of large contributions - it is likely that since the definition chosen systematically understates the presence in certain kinds of projects, this means our analysis is specific to the kinds of projects which are likely to have large contributions detected; small to mid-sized. However, this seems to be a modest effect. Plotting the 5-percentiles of the dataset for those projects who attact at least one backer, we can see in Figure B.1 the frequency of large contributions is generally monotonic except for the very large projects. However, even for projects in the top 5% of the data, we still detect large contribution in more than 50% of them, indicating that this problem does not fully exclude all large projects or unduly stratify the sample.

B.2 Overview of Crowdfunding Data

This section generally overviews the data used in this chapter for comparison with other papers working in the same literature. We see the same basic patterns most individuals have mentioned, particularly (1) a hump-shaped pattern of backers over time and (2) a clearly bimodal distribution of projects, with most being either very successful or not successful at all.

As Fig B.2 shows, the number of backers raised today (as a fraction of the total) varies in a U-shaped manner as projects approach completion. This is typical of crowdfunding

Figure B.1: Distribution of Large Contributions



data (see, for example Kuppuswamy and Bayus (2013)), although the reasons behind this feature of the data are unclear. The first possible interpretation is simply that projects garner more attention at the beginning and end of the project cycles, which mechanically introduces more backers. However, previous studies have shown that even conditional on this affect, there is still a recoverable U-shape in the data. An alternative suggestion is that a public good effect is at work, which depending on the formulation induces a positive effect at the beginning and end of the project cycle. Another possibility is that individuals select endogenous arrival periods, in a manner which creates the observed variation. Trying to tease apart these different explanations is a major goal of my research.

We can see from Figure B.3 the number of projects over time is growing; this is mainly due to the fact that Kickstarter, as a platform, was still in its early stages, leading to growth over time. The only reason the 2014 bar is not higher is because the data lacks the last

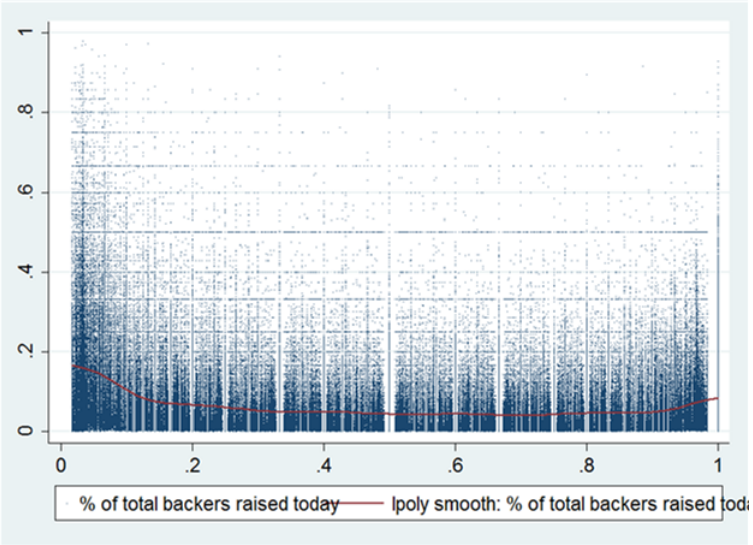


Figure B.2: % of total backers raised today versus % of project-time elapsed

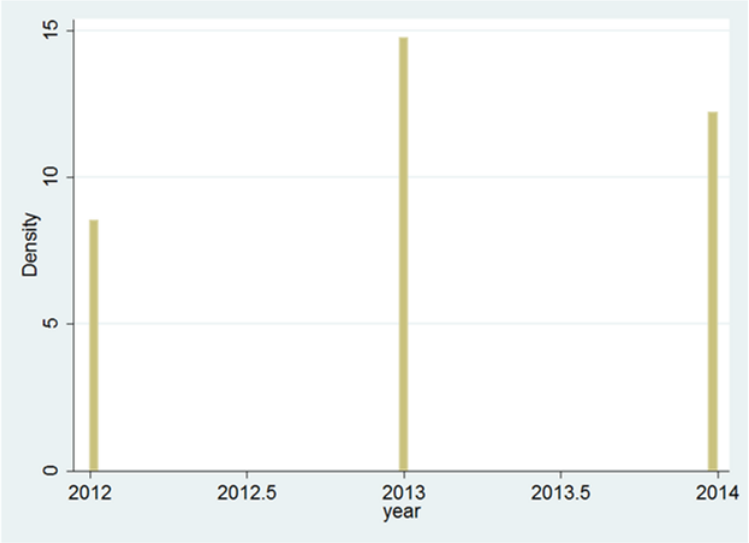


Figure B.3: Number of Projects by Year

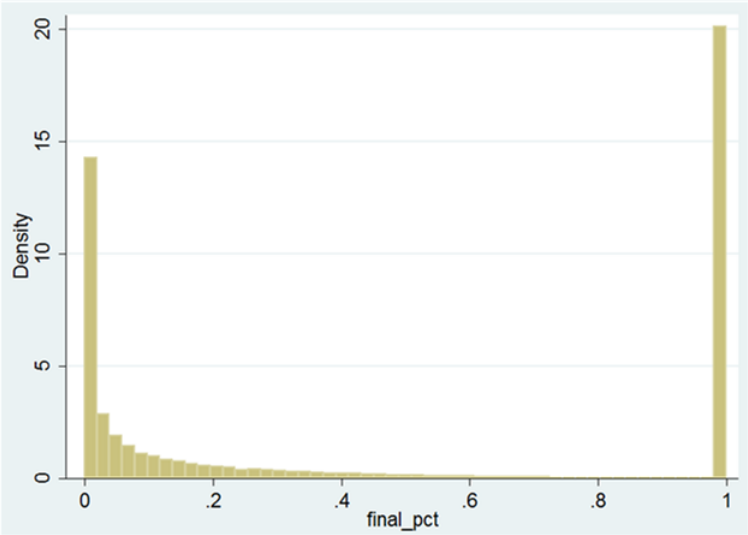


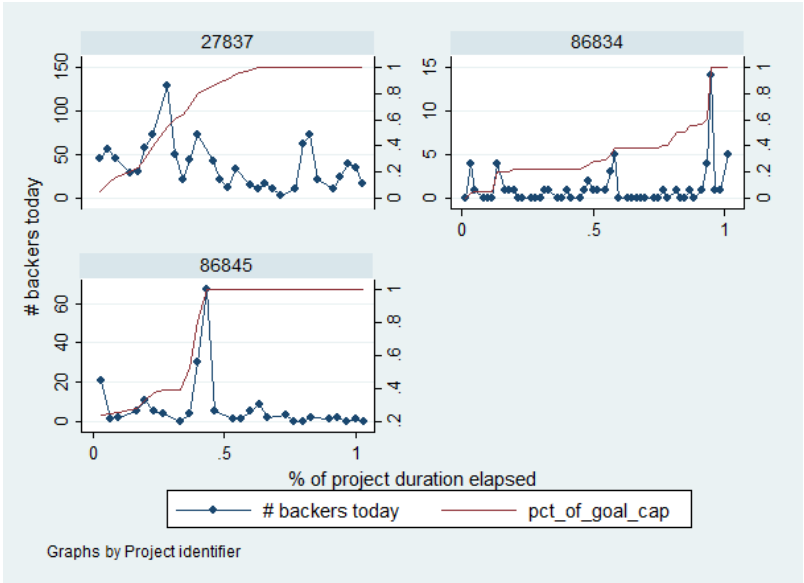
Figure B.4: Project histogram of final % of goal (capped at 1)

three months of 2014; including these, it is likely to be approximately as large. This fact is also well known, since Kickstarter has shown year-over-year growth.

Figure B.4 demonstrates that projects tend to fall into two categories: very successful, or very un-successful. There are very few liminal projects, which get close to their goal but don't quite reach it. In fact, empirically, the probability of being successful contingent on having raised only 40% of your goal is in excess of 90%. This demonstrates a kind of separation in the data; however, what drives this is up in the air. The natural implication, from the previous discussion is that individuals tend to pull projects across the line if they are on the bubble - but there could be other factors at work, such a sorting of projects by goal into feasible and infeasible groups. This speaks to an aspect of crowdfunding which is not well explored - the role project creators play in choosing goals and other aspects of the project.

In the run-up to projects succeeding, the days just prior to success have a greatly increased number of backers. This effect diminishes linearly as we look ahead of the success date. This indicates that as projects are about to succeed, backers become more likely to back the project. We can also see, from a selection of projects, that a large "reaction" is typical of many projects - and while, in general it relates to the success threshold, it may be more general. For example, see Figure B.2.

In order to understand how the different stylized facts are related, I first use a standard



within FE model to estimate the results. I chose the FE model because the data strongly suggests there are idiosyncratic project-level effects which have a serious causal impact on the data. This can be verified using a Hausmann test to demonstrate that a RE model is resoundingly ruled out by the data, but the empirical evidence is suggestive enough on its own. The FE model has two drawbacks: first, it cannot take into account time-invariant project characteristics; these must be captured by the project-level fixed effects, which are being differences out from the panel.

The second issue relates to the issue of endogenity in the model. A number of the variables (such as the pre-success indicators or % of goal) are functions of the lagged values of the dependent variable. Indeed, in one specification presented I directly include the lagged dependent variable. Because of the within transformation, these estimates are known to be biased. However, mitigating this is the fact that because the length of the panel is (on average) 35 days in length, and (as in Nickell (1981)) the bias is known to be of order $O(1/35) \approx O(0.029)$, the Nickell bias is likely to small since the panel is sufficiently long. Secondly, because we know that the bias tends to negatively biased when (as in my model) we expect a positive correlation in values, this implies the results are an underestimate of the values.

Finally, I present two specifications: one including the lagged dependent variable, and one without. The inclusion of the lagged dependent variable is used to try and assess the

extent of herding, and control for an acceleration effect in the size of the project. The specification excluding this variables allows us to more carefully examine the first period of the model, which is known, from the stylized facts, to be very relevant to the project outcomes. These are included in the appendix. The dependent variable is the number of backers arriving on a given day; this is preferred to % of goal or funding-based measure, because it better captures the backer dynamics, and any scale effects (big versus small projects) should be differences out by the within transformation.

Looking first at Table B.1, I include a quadratic term in both % of project elapsed, and in % of goal reached; this is to try to capture the non-stationary time and goal effects noted earlier. Both coefficients demonstrate the expected U-shaped curve; as projects become closer to their deadline (crossing the threshold) they become more likely to be supported. In addition to this, we can see there are large first observation effects. However, these have little in comparison to the pre-success indicator: an increase of (on average) 69 backers on the day just prior to (or including) success occurring. In addition, we can see that the post-success variable has a much smaller, but opposite effect; post-success, backers fall by an average of 7.

Also of note, is that these effects are non-stationary: over time, both diminish; with (for instance) a pre-success occurring in the last 25% of the project's lifespan having only about 25% of the baseline effect. Similarly, the post-success effect drop in the last part of the project is similarly reduced in scale. Communication between project-owners and backers seems to have small effect on the number of backers, but it difficult to argue this is a fully causal effect. Finally, the day of week terms largely agree with the facts presented earlier (although they are not shown here for space reasons.)

Next, Table B.2 indicates that these results are generally consistent when we include a lagged dependent variable term. The coefficient on backers today is positive and substantial. This could indicate some kind of herding effect, or simply a "spread of information" effect, as individuals who back a project share the word about the project. The directional results from before are largely unchanged: however, the magnitude of most of the terms has been reduced. Regardless, there is still a large and substantial pre-success effect, and post-success drop in the number of backers arriving. However, it is difficult to interpret the meaning of these coefficients in a vacuum: first of all, the lagged term necessitates changing the sample (losing one day to the lagging), which happens to exclude the first day, and therefore a large number of projects who succeed on day 1, and the indicator for the first observation. Secondly, it is difficult to interpret the coefficient on the lagged dependent

variable, and thus it is hard to say whether the decline in the size of the coefficients is meaningful here.

These results are generally robust to the inclusion of other time fixed effects, or changing the sample specifications. It is clear that some of the coefficient values are driven by the largest projects; by restricting the sample to projects of less than $200,000 in total funding, we obtain smaller coefficient estimates more in line with the second specification. This indicates the difference between the two specifications is probably mostly driven by large projects who have their first observations dropped on the first day (on which they usually succeeded). Indeed, for large projects, the rate of success on the first day is 28%, compared with 1.5% for the entire sample.

VARIABLE	Coef	Std dev	t	p	%95 Confidence	
First observation	5.911519	0.184623	32.02	0	5.549664	6.273374
Update today?	2.822745	0.122702	23	0	2.582254	3.063237
% of time elapsed	-31.745	0.500512	-63.42	0	-32.7259	-30.764
% elapsed squared	27.28646	0.42752	63.82	0	26.44854	28.12439
Days in capture	0.235139	0.072432	3.25	0.001	0.093174	0.377103
Days in capture sq	0.044555	0.002392	18.62	0	0.039866	0.049244
Pre-success indicator	69.43112	0.639596	108.55	0	68.17753	70.6847
Pre-success X % elapsed	-70.3452	0.770081	-91.35	0	-71.8546	-68.8359
Post-success indicator	-7.44758	0.425197	-17.52	0	-8.28095	-6.6142
Post-success X % elapsed	2.216039	0.410276	5.4	0	1.411912	3.020165
% of goal reached	20.24087	0.892339	22.68	0	18.49191	21.98982
% of goal squared	-12.3724	0.881135	-14.04	0	-14.0994	-10.6454
Constant	6.274616	2.065283	3.04	0.002	2.226735	10.3225
Year indicators	Yes					
Month indicators	Yes					
Day of week indicators	Yes					
N	3,186,363					

Table B.1: FE Regression on Number of Backers - Baseline Specification

B.3 Holidays Used For Instrumentation

This table presents the holidays used to instrument for large contributions. All of the dates are for American holidays in the years listed (when different countries disagree, e.g.

VARIABLE	Coef	Std dev	t	p	%95 Confidence	
Backers Today (t-1)	0.396599	0.000336	1181.24	0	0.395941	0.397257
Update today?	4.00926	0.072452	55.34	0	3.867257	4.151264
% of time elapsed	-12.7182	0.299331	-42.49	0	-13.3049	-12.1315
% elapsed squared	11.25743	0.254383	44.25	0	10.75885	11.75601
Days in capture	-1.29638	0.693902	-1.87	0.062	-2.6564	0.063648
Days in capture sq	0.222935	0.230228	0.97	0.333	-0.2283	0.674174
Pre-success indicator	23.19426	0.413043	56.15	0	22.38472	24.00381
Pre-success X % elapsed	-20.0114	0.493131	-40.58	0	-20.9779	-19.0449
Post-success indicator	-11.7195	0.253527	-46.23	0	-12.2164	-11.2226
Post-success X % elapsed	9.214561	0.243205	37.89	0	8.737888	9.691234
% of goal reached	1.109525	0.574982	1.93	0.054	-0.01742	2.23647
% of goal squared	2.873862	0.552795	5.2	0	1.790402	3.957321
Constant	5.001013	1.327311	3.77	0	2.399531	7.602496
Year indicators	Yes					
Month indicators	Yes					
Day of week indicators	Yes					
N	3,186,363					

Table B.2: FE Regression on Number of Backers - Lagged Dependent Variable

Thanksgiving); for moveable Christian religious feasts such as Easter or Lent, the dates published by the American Council of Catholic Bishops are used. The Islamic calendar used is the official sighting-based *Hijri* calendar produced by Saudi Arabia and used by the Islamic Society of North America (to coincide with the *Hajj* preperations in Saudi Arabia).

HOLIDAY NAME	DATE	FREQUENCY
Black Friday	Friday following Thanksgiving	3
Boxing Day	December 26th	3
Chinese New Years	1st day, Chinese Lunar Calendar	3
Christmas	December 25th	3
Christmas Eve	December 24th	3
Cyber Monday	Monday following Thanksgiving	3
Easter	Complicated (Sunday in March or April)	3
Father's Day	3rd Sunday in June	3
Halloween	October 31st	3
Mother's Day	Second Sunday of May	3
New Year's	January 1st	3
Ramadan Starts	9th Month, Islamic Lunar Calendar	3
Lent Starts (Ash Wednesday)	46 Days Prior to Easter	3
Thanksgiving	Fourth Thursday in November	3
Valentine's Day	February 14th	3

Table B.3: List of Holidays

Appendix C

Appendix to "Sales Classification via Hidden Markov Models"

C.1 Proofs

C.1.1 Proof of Lemma 4.3.1

Proof. We can write the probability of the observations at time t simply, since the Markov chain begins in state 1: $f(y_1) = f(y_1; \theta_1)$. This implies from the distribution of the time 1 values we can recover θ_1 by Condition 1. Next, note that:

$$f(y_2) = a_{11} f(y_2; \theta_1) + (1 - a_{12}) f(y_2; \theta_2)$$

Again, by condition 1, we can identify a_{11}, a_{12} and θ_2. We can then repeat the process, since in general, $f(y_t) = [a_{ij}]^t \pi \cdot [f(y_t|\theta_j)]_j$. Since each of these steps adds exactly one element of θ_k and a_{ij}, this implies we can recover all the parameters as long as the maximal number of steps, T is at least K, which is assumed. □

C.1.2 Proof of Theorem 4.3.1

Proof. The first step is noticing that at the first observation, $t = 1$, the observed variables are Y_{i1} which have a mixture distribution given by

$$f(y_{i1}) = \pi(1) f(y_{i1}; \theta_1) + \pi(2) f(y_{i1}; \theta_2) + ... + \pi(K) f(y_{i1}; \theta_K)$$

Now, this is a mixture distribution with (at most) K components. By Condition 1, this is identifiable, and by Condition 3 we observe a population of this from the data. Thus, we can identify π and θ_k where $\pi(k) > 0$. Then, consider $t = 2$. At time 2, the distribution of the states evolves according to the underlying Markov chain $\pi^2 = [a_{ij}]\pi$. Then, at this

state, the distribution of the observable is given by:

$$f(y_{i2}) = \pi(1)^2 f(y_{i2};\theta_1) + \pi(2)^2 f(y_{i2};\theta_2) + ... + \pi(K)^2 f(y_{i2};\theta_K)$$

and again, by Condition 1, this is identifiable, so we have recovered π^2 and any θ_k where k is accessible by 2 from π. Clearly, this is repeatable for any $t = 1, 2, ..., T$. Therefore, Result (2) follows immediately since $K < T$ and $k \in S'$. We also have the vector π^t for each time period. Now, we can note that for $t < T$, the joint distribution of y_{nt} and $y_{n,t+1}$ is given by:

$$f(y_{nt}, y_{n,t+1}) = \sum_i \sum_j \pi(i)^t a_{ij} f(y_{nt};\theta_i) f(y_{n,t+1};\theta_j) = \sum_i \pi(i)^t \underbrace{\sum_j a_{ij} f(y_{nt};\theta_i) f(y_{n,t+1};\theta_j)}_{\equiv g(a_{ij})}$$

This is exactly a mixture of products with parameters a_{ij}. Then, as in Leroux (1992), we know that this is itself identifiable via a result in Teicher (1967). In particular, this means the mixing coefficients of this mixture are also identifiable, and therefore we can recover a_{ij}. for all i such that $\pi(i)^t > 0$. But, since for some t, any state in S' has that $\pi(i)^t = [a_{ij}]^T \pi > 0$, Result (1) immediately follows. □