UNIVERSITY OF LONDON THESIS

Degree PhD Year 2006 Name of Author HARMGART Heike

Consumer Behaviour, Feedback Information and the Supermarket Industry

Heike Harmgart

A Dissertation submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
University College London

London
September 2006

Supervisors: Professor Richard Blundell and Professor Rachel Griffith

Abstract

The purpose of this thesis is threefold: First, to explain how learning shapes consumer behaviour over a comestible (experience) good. Second, to examine the role of feedback information and information aggregation for consumer choices and market performance in markets for experience goods. Third, to understand how firms react to heterogeneous consumer choices in the supermarket industry when faced with institutional constraints.

To shed light on how learning influences consumer choices over time in particular when new products or new characteristics are introduced, we employ in Chapter 2 a model of reinforcement learning over products as well as characteristics and apply it to yoghurt drink purchases from a large British consumer panel. We find that learning over both, products and characteristics, is important in explaining consumer choices over time.

How consumer choices are influenced by the choice of others is analysed in Chapter 3 which introduces and studies a new model of aggregate information cascades. We find that if only one of two possible actions is observable —say, how many others bought a particular product but not how many chose not to buy it—only one type of cascade arises in equilibrium. Herding only takes place on the observable action.

A different angle on how the provision of information bears on choices is taken in Chapter 4 on learning trust. Here we examine the effect of different forms of feedback information to consumers and sellers in a market with sequential exchange. Experimental evidence shows that both feedback information on sellers' history to consumers but also feedback information about sellers trading history to other sellers improves market efficiency.

How firms optimally react to institutional constraints when consumer choice heterogeneity is important is developed in Chapter 5 in a model of supermarket entry into different store formats and applied to data from the UK. We are interested in estimating the cost of the institutional constraint of restrictive planning regulation. We find that the institutional set-up matters but the impact of restrictive planning regulation on firm profits is small and increases barriers to entry for large supermarkets only.

Contents

List of Tables

List of Figures

Acknowledgements

I am deeply indebted to my supervisors Richard Blundell und Rachel Griffith. Discussions with each one of them have always been a source of inspiration.

They both have made sure that I was looked after at the Institute for Fiscal Studies (IFS) since November 2003. I consider myself privileged to have spent the last years of my dissertation in such a stimulating and friendly research environment as the IFS.

I am extremely grateful for financial support from the ESRC, the Advanced Institute for Management Research (AIM), the Russel Sage Foundation and the Leverhume Trust.

I would also like to express my gratitude towards my co-authors Iris Bohnet, Antonio Guarino, Rachel Griffith, Steffen Huck and Jean-Robert Tyran. During our scientific cooperation I learned an incredible amount. Working with them has been both a pleasure and a challenging intellectual experience.

Special thanks to Ian Crawford, Georg von Weizsäcker, Aviv Nevo, Jerome Adda and Gareth Macartney who provided insightful and encouraging comments along the way.

Most of all thanks to Steffen for being witty and uplifting on all occasions, supportive and helpful beyond limits, and for pointing out the fun of economic research.

Declaration

1. No part of this thesis has been presented to any University for any degree.
2. Chapter 3 was undertaken as joint work with equal contributions from Antonio Guarino, Steffen Huck and myself.
3. Chapter 4 was undertaken as joint work with equal contributions from Iris Bohnet, Steffen Huck, Jean-Robert Tyran and myself.
4. Chapter 5 was undertaken as joint work with equal contributions from Rachel Griffith and myself.

Heike Harmgart

Chapter 1

Introduction

The purpose of this thesis is threefold: First, to explain how learning shapes consumer behaviour over a comestible (experience) good. Second, to examine the role of feedback information and information aggregation for consumer choices and market performance in markets for experience goods. Third, to understand how firms react to heterogeneous consumer choices in the supermarket industry when faced with institutional constraints.

The second chapter focusses on how learning influences consumer choices over time, in particular when new products or new characteristics are introduced. We employ an adaptive learning model — reinforcement learning — that assumes that the behaviour of consumers confronted with a repeated choice problem is shaped by their past experiences and is best described by a dynamic learning rule. The model takes both, learning over products and characteristics, into account and applies it to yoghurt drink purchases from a large British consumer panel. In this context consumers are faced with a simple recurring decision problem — buying a good in the supermarket where the quality is not perfectly known a priori. This uncertainty plays, of course, a particularly important role when a new characteristic is introduced. Additionally, the evaluation of quality might be heterogeneous and imprecise, i.e.. consumers may need a while to find out whether they prefer a particular taste over another. Since each purchasing decision is relatively unimportant to the individual consumer a boundedly rational reinforcement model may explain actual individual choices well. This view is confirmed in the empirical analysis. The results suggest that consumers' past purchasing history plays an important role in their present choices. But not only product familiarity is important, consumers decompose products into characteristics and these characteristics experiences matter when deciding between different products the next time. This plays a particularly important role when consumers decide to switch to a product with a new characteristic: If consumer switch from one product to another they are more likely to switch to a new product that shares some of the characteristics of the previously consumed one. These results are robust to the introduction of price changes of both, purchased and not purchased, goods and disregarding learning altogether leads to an significant upward bias in both of the price elasticities. Including household characteristics into the analysis sheds further light on choice heterogeneity over time.

How consumer choices are influenced by choices of others is analysed in Chapter 3 where we introduce a new model of *aggregate information cascades.* We find that if only one of two possible actions i.e. how many people chose a particular product but not how many chose not to buy it is observable only one type of cascade arises in equilibrium. This information structure arises very naturally in many applications. Consider, for example, a venture capital firm who discusses a project with an inventor who needs capital to develop a new product. The inventor may gladly mention the credits he has already secured with other investors but probably wants to disclose the number of investors who have rejected the project. Or if we take the example of a central agency in the health sector who must decide how to disclose information on the adoption of a new treatment. It has the choice to inform the doctors on how many others have already decided to adopted the new treatment or on how many have considered it but decided not to adopt. Our central result is that only one type of cascade arises in equilibrium, the aggregate "up cascade" on the observable action. Herding never occurs on the unobservable action. This is crucial for the example of the health agency that can decide which information to disseminate. If it chooses to disclose the number of doctors who have adopted the procedure the only cascade that can arise is one where all adopt and vice versa. Depending on the severity of different bad cascades, the central agency may optimally decide to withhold information. So the aggregate information set up that is introduced here has not only several intriguing properties — some in stark contrast to the predictions of the standard model — it also has potentially important policy implications. We then take our theory to the laboratory. The experimental data suggests that our model does fairly well. We find that the main comparative statics go all in the right direction.

A different angle on how the provision of information bears on choices is taken in the chapter on learning trust. Here we examine the effect of different forms of feedback information to consumers and sellers in a market with sequential exchange. These markets are particularly prone to moral hazard and orthodox theory predicts that providing consumers with information about sellers' trading history should help alleviate this problem. Experimental evidence shows that feedback information on sellers' history to consumers does indeed boost market performance. However, the key finding is that two-sided market transparency where both, consumers and sellers, have access to sellers' trading history improves market performance even further which may be important for the design of new electronic markets.

How firms optimally react to institutional constraints when consumer choice heterogeneity is important is developed in a model of supermarket entry into different store formats and applied to data from the UK. Here the interest lies in estimating the cost of the institutional constraint of restrictive planning regulation. The change in the planning regulation mainly in 1996 centralised the planning regulation and changed it to encourage in-town (small) store formats over out-of-town (large) store formats. As a result we would expect it to restrict the number of large stores and increase the number of small stores in equilibrium. Unsurprisingly, after the reforms we do see an increase in town centre stores and an increase in smaller store formats. However, other factors may have also affected this shift towards the smaller store formats. In particular changes in consumer demand patterns or strategy diversifications by several firms in the supermarket industry may have contributed significantly to this increase in smaller store formats. Thus incorporating a model of heterogeneous consumer demand is crucial in isolating the effect of planning regulation on market structure and firm profits. This is implemented by estimating the structural parameters of the profit function and isolating the increase in fixed costs that is associated with more restrictive planning regulation. The results show that the institutional set-up matters but the impact of restrictive planning regulation on firm profits is small and increases barriers to entry for large supermarkets only.

Each Chapter of this thesis can also be read as a stand alone article.

Chapter 2

Learning to Consume - on the Importance of Characteristics

2.1 Introduction

Consumer choices have generally been seen as a sequence of rational decisions. Recent research has pointed out that this might not always be the case. Several recent empirical papers have found violations of the most general rational choice assumption for a significant portion of the population. For instance Blundell, Browning and Crawford (2003) find that on an aggregated level the general indicator of rational choice, the General Axiom of Revealed Preference (GARP), is violated over 10% of the time[1]using a repeated cross-section of the Family Expenditure Survey for Britain. Mattei (1994) using a micro level Swiss consumer panel dataset, shows that half of the households in the panel violate GARP. In a more recent micro level study that investigates consumer rationality Blow, Browning and Crawford (2003) look at milk purchases in a Danish consumer panel . They find GARP violations in 16% of the cases and the purchase behaviour of over 40% of their sample cannot be rationalised with

[1]The GARP rejections are mainly concentrated in the upper tail of the distribution.

a linear characteristics model[2]. An earlier experimental study by Sippel (1997) documents that even in a very limited choice situation GARP had to be rejected in approximately half of the cases.[3] One lesson from these apparent rationality failures may be that consumers use simpler algorithms to make consumption choices and learn about products and their characteristics over time. But this is at its heart an empirical question.

Empirical research in the past has taken two different routes in explaining consumer purchasing decisions over time. One originates in the marketing literature following Guadagni and Little's (1983) first use of scanner data of ground coffee purchases over Erdem (1996, 2004) to Ho and Chong (2003) work on modelling stock keeping unit (SKU) choices, focussing particularly on the effect of marketing mix strategies and shopping experience on consumer brand choice. This research uses scanner data or data that is collected at the point of sale in combination with loyalty card information, so that the researcher observes very detailed information about the product but no information about the household. While Guadagni and Little are mainly concerned with brand loyalty[4], Erdem decomposes the product into its characteristics (modelling brand as one of them) and investigates whether the similarity between different characteristics explain product choice. Ho and Chong additionally differentiate between consumption and shopping experience where the latter applies to all familiar and available characteristic levels and products that the consumer could be purchasing. None of them investigate the effect of consumer characteristics on the individual shopping behaviour, nor have they looked at the introduction of a new product or characteristic.

The other route evolved from Gorman's work on pure product characteristics (1980) to estimating non-parametric revealed preferences that are consistent with these characteristics models.[5] This literature places the consumer choice problem in a fully rational and static

[2]See e.g. Gorman (1980), Lancaster (1966) or Heckman and Scheinkman (1987).

[3]A more recent similar experimental study by Mattei (2000) confirmed Sippel's results with GARP violations between 25% and 44%.

[4]There is a large marketing literature in brand loyality: some of the main most recent papers are Akçura et al. (2004), Erdem (1998), Osselaer (2000) and Villas-Boas (1999, 2004).

[5]See e.g. Blow, Browning and Crawford (2003).

setting while additionally assuming a time invariant utility function. There are some very recent extensions by Crawford and Rickman (2006) that look at product purchases in a dynamic framework splitting consumption into a rational and addictive part over whole products. Our model is closely related to a product characteristic version of their model that gives zero weight to the rational part of consumption and makes some more structural assumptions on the "addiction" mechanism.

This paper builds on both of these strands of literature and extends the analysis to a dynamic framework of product characteristics. We incorporate consumer learning as well as demographic characteristics to explain - at least partially - choice heterogeneity over time and estimate our model using a large British consumer panel dataset, that covers all grocery purchases for a three year period.

But, what kind of consumer learning? In this context agents are faced with repeated decision problems. The consumer is going on a regular basis to a supermarket to buy groceries, where the goods are typically experience goods. One has to try them out before the quality is known. Also the evaluation of this quality might be heterogeneous and imprecise for example whether I liked the taste of a particular characteristic. Adaptive learning models - like reinforcement learning models - attempt to describe exactly this, the behaviour of agents confronted with repeated decision problems assuming they use simple learning rules. Additionally experimental evidence has shown that these reinforcement learning models describe actual human behaviour quite well.[6] Also because each purchase decision is relatively unimportant to the individual consumer a boundedly-rational behaviour may explain actual choices well.[7]

In the past one major obstacle for empirically analysing micro-level consumer choice behaviour was the absence of micro-level data that combined detailed purchasing and household information over time. The latest availability of micro-level consumer choice data enables us

[6] See Erev and Roth (1998).

[7] In a recent theory paper Hopkins (2006) studied the effects of this type of adaptive consumer learning over two types of product quality on firm pricing.

to take a closer look at individual shopping behaviour in conjunction with the demographic characteristics of the shopper. In this paper we use data from the TNS British consumer panel focussing on probiotic yoghurt drink purchases.

Why probiotic yoghurt drinks? With a comparatively low number of characteristics probiotic yoghurt drinks are rather simple, and are additionally very homogeneous within their characteristic space.[8] They are also increasingly popular hence we have a reasonably big heterogeneous consumer sample. The market for probiotic yoghurt drinks has also seen a large number of successful new characteristic introductions during or just before our sample period. The implementation of a new products makes learning a paritcular important issue. Looking at a successful product introduction also helps us to understand learning from the very beginning so we don't have to make assumptions about initial conditions. Thus it can serve as an ideal benchmark of understanding consumer preferences and learning over products and characteristics. Concerns about separability from other consumption or labour market decisions may be not as acute with yoghurt drinks as with other product innovations that, more often than not, are new durables like consumer electronics. An additional advantage of this data set is that it not only contains barcode level product, characteristic and price information it also entails detailed information on the purchasing household. Thus we can incorporate a large amount of observable consumer heterogeneity into our estimation.

Our results suggest that consumers purchasing history plays an important part in their present product choice. But not only product familiarity is important, consumers decompose products into characteristics and these characteristic experiences matter when deciding between different products. This is particularly crucial when consumers decide whether to switch to a product with a new characteristic.

The paper is organised as follows: The reinforcement model is discussed in Section 2, Section 3 introduces the estimation strategy, Section 4 gives an overview over the data and

[8] See Bajari and Benkard (2001) for a model if product characteristics are unobserved.

the results are shown in Section 5, Section 6 concludes.

2.2 The model

The goal of the model is to predict which product j the consumer i will choose on a purchase occasion given her purchase history, the price of the product and her personal characteristics.

The model assumes that in each time period t (purchase occasion) the consumer chooses from J different products. Let U_{ijt} denote the utility to consumer i of purchasing product j at time t consisting of an observed component V_{ijn} and an unobserved component ε_{ijt}. The utility may depend on observed and unobserved characteristics of products, observed and unobserved characteristics of consumers, including their purchasing history.

$$\begin{aligned} U_{ijt} &= V_{ijt} + \varepsilon_{ijt} && (2.1) \\ &= \alpha_j Z_{ijt} + \gamma_j P_{jt} + \beta_j X_{it} + \varepsilon_{ijt} && (2.2) \end{aligned}$$

Individual household $i's$ utility is the sum of an intrinsic valuation Z_{ijt} and the associated price of the product and a vector of household characteristics X_{it}. The error structure captures serial correlation in characteristic-level and product-specific utilities. The additive form has been used by most prior models and is adopted here for simplicity.

The intrinsic value Z_{ijt} of the product j consists of both product-specific (C_{ijt}) and characteristic-level (C_{inlt}) experiences. Formally,

$$Z_{ijt} = \sum_{n=1}^{N} \sum_{l=1}^{L} = C_{inlt} \cdot I_{jnl} + C_{ijt} \tag{2.3}$$

C_{jnlt} is a vector of observed characteristics-level experiences of individual i of characterist level l in characteristic n, it is multiplied by an indicator variable I_{jnl} is 1 if product j

has characteristic level l in characteristic n and 0 otherwise, this is added to the overall product-specific experience C_{ijt} of individual i of product j at time t - or in other words the ideosynchratic likening of the product beyond it's characteristics. Over time consumers accumulate product and characteristic experiences. For example consumer *Smith's* cumulative product experience or in other words her cumulative attraction to product 2={original, light, yakult} is a sum of her cumulative experience for Yakult, original flavours and light products.

$$Z_{Smith,2,t} = C_{Smith,original,t} + C_{Smith,light,t} + C_{Smith,yakult,t} + C_{Smith,2,t}$$

These cumulative experiences or attractions for the characteristics as well as the product as a whole are updated over time as follows:

$$C_{inlt} = \phi_k \cdot C_{inl(t-1)} + I_{inlt} \tag{2.4}$$

$$C_{ijt} = \phi_p \cdot C_{ij(t-1)} + I_{ijt} \tag{2.5}$$

where ϕ_k and ϕ_p are decay factors. This updating rule will be called reinforcement learning, upon receiving a payoff the agent updates the respective choice propensities. The I_{inlt} describes an incremental reinforcement that consumer i derives form level l in characteristic n in time t. Additionally she receives an incremental reinforcement from the product j as a whole. The distinction between the two incremental reinforcements is that characteristic-level reinforcement affects the intrinsic value of all products that have similar characteristic levels while the product-level reinforcement does not. This captures on the one hand the products uniqueness as well as the shared characteristics with other products.

Characteristic level incremental reinforcement for a characteristic level l in time period depends on whether it was chosen in time period $t-1$.

$$I_{inlt} = \begin{cases} 1 & \text{if level } l \text{ in characteristic } n \text{ was chosen in } t-1 \\ 0 & \text{otherwise} \end{cases} \tag{2.6}$$

In addition to characteristic-level reinforcement the product-level reinforcement captures the consumer's idiosyncratic liking for a product beyond the shared characteristic levels. This will therefore only affect the product's own cumulative attraction and will not influence the attractions of other products. So the incremental reinforcement depends on whether the product j was chosen in time period $t-1$.

$$I_{ijt} = \begin{cases} 1 & \text{if product } j \text{ was chosen in } t-1 \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

Thus the more often a product or a characteristic has been chosen in the past, the higher the cumulative product experience and the higher the intrinsic valuation of this product. This leads to an increased utility associated with this product and therefore a higher probability of choosing this product at the next purchasing occasion.

Supplementary to the assumption that consumer characteristics influence their choices over time this model is based on the behavioural proposition that the consumer accumulates characteristic level and product-level experiences and her consumption experience depends on their familiarity. In the psychological literature Alba and Hutchinson (2000) argue that since the grocery-shopping environment is highly complex, consumers often rely on recall to recognise products. So, when consumers look at the grocery display without preconceptions, characteristic level and product familiarity are likely to influence how easily a product is recognised on the shelf.

Switching between different products or different characteristics can be interpreted as variety-seeking behaviour.[9] A consumer seeks variety if the conditional probability of choosing a product on an occasion (given that the same product was chosen on any prior occasion)

[9] For a comprehensive review on variety-seeking behaviour see Kahn (1998).

is smaller than the unconditional probability of choosing the product. In our model variety-seeking on a product level would occur if $\alpha_j < 0$, this might be due to satiation of a particular characteristic or the product as a whole. Switching to an unfamiliar product characteristic may be caused by satiation on the part of the consumer or it may be the reaction to a change in the relative price (i.e. a change in P_{jt} due to a promotion).

2.3 The Estimation Strategy

To estimate our product choice model we use the multinomial logit estimator dating back to McFaddens (1974) early work on random utility models that emulates our choice rule particularly well. We assume that the error ε_{ijt} is extreme value independent over i, j and t. We include the consumer behaviour over time as the accumulated attractions in the exogenous variable, prices and household characteristics as specified above. So if individuals are reinforcement learners over characteristics we would expect the estimated α_j coefficients to be positive.[10] If the marginal consumption experience decreases with characteristic familiarity, we would expect satiation to occur. This can be one of the reasons for variety-seeking between characteristics. In this case we would expect the α_j coefficients to be negative.[11] The choice probabilities have the following structure:

$$\Pr_{ijt} = \frac{\exp(V_{ijt})}{\sum_{j'} \exp(V_{ij't})} \tag{2.8}$$

or more specifically:

$$\Pr_{ijt} = \frac{\exp(\alpha_j Z_{ijt} + \gamma_j P_{jt} + \beta_j X_{it})}{\sum_{j'} \exp(\alpha_{j'} Z_{ij't} + \gamma_{j'} P_{j't} + \beta_j X_{it})} \tag{2.9}$$

[10] There are several possibilities of entering past consumption in the utility function. Adamowicz (1994) enteres the number of times a particular product is chosen, rather than simply a dummy for the immediately previous choice. Erdem (1996) looks at the number of times a particular attribute was chosen in the past without reinforcing the product as a whole.

[11] For empirical evidence on this type of variety seeking behaviour see e.g. McAlister (1982) or Erdem (1996).

where $\Pr_{ijt}$ is the probability of household i choosing product j at time t. The V_{ijt} include the product and characteristic level familiarity (Z_{ijt}), the price components (P_{jt}) and the household characteristics component (X_{it}) that might all effect household's choice behaviour. The choice probabilities sum to one. In order to identify the model we must normalise one of the coefficients to zero. We use the incumbent product as the baseline outcome and thus all estimates are relative to this baseline outcome. To estimate the multinomial logit model we maximise the log-likelihood function using the Newton-Raphson algorithm. The log likelihood can be derived by defining an indicator variable for each household i, and product j

$$I_{ij} = \begin{cases} 1 & \text{if product } j \text{ was chosen} \\ 0 & \text{otherwise} \end{cases} \tag{2.10}$$

where for each i at a particular point in time t only one I_{ij} is equal to one:

$$\ln L = \sum_{i=1}^{I} \sum_{j=1}^{J} I_{ij} \ln \Pr_{ij} \tag{2.11}$$

The inclusion of the lagged dependent variable - the cumulative product and characteristic attractions - still gives us unbiased estimates as long as it is uncorrelated with the current error and the errors are independent over time. In order to account for the dependence of observation within households we cluster on households.

The estimated α, β and γ coefficients can only be directly interpreted in terms of their sign and their significance level. For example a positive sign on a particular α coefficient makes it more likely that this product is chosen with increasing $Z's$ (product & characteristic familiarity). Consumers are not yet satiated with this product. A negative sign on the α coefficient on the other hand makes it more likely that consumers switch away from this product - seek product variety- with increasing $Z's$. To interpret the magnitude of the effect we calculate marginal effects for the different α, β and γ.

One drawback of using the multinomial logit model is its restrictive substitution pattern. It can only capture proportional substitution across alternatives, if one characteristic becomes more attractive i.e. due to a promotion the increase in this characteristic's probability necessarily means a proportional decrease in the probability of the other characteristics. In other words the multinomial logit model has the property of independence of irrelevant alternatives (IIA). This might be particularly crucial if new choices become available. In our case this would be e.g. new characteristics like strawberry flavour. We undertook two robustness checks to test the severity of the IIA restriction in our setting. First we estimated the basic characteristic reinforcement attraction model without price or consumer characteristics with a small random sample of our population using a multinomial probit which can represent any substitution pattern. Because the probit probabilities do not have a closed-form expression we approximate the choice probabilities numerically using the GHK [12]simulator[13]. As a second test we excluded the newly introduced characteristics (strawberry and multifruit flavour) and reestimated the multinomial logit of the basic model (only characteristic level reinforcement). The Hausman and McFadden (1984) test was used to test whether the two estimates were significantly different from each other.

2.4 The data

We us the TNS Consumer Panel from January 2002-December 2004. This unbalanced panel consists of around 50,000 households over the period of 3 years, their yearly socioeconomic characteristics all fast moving consumer good purchases (these include all groceries but also cleaning products etc.). We extracted all probiotic yoghurt drink purchases. There is some attrition, so we observe households on average for 650 days. 6,308 of those households

[12] Geweke, 1989; Hajivassiliou and McFadden, 1998; Keane, 1994.

[13] A summary of probit simulators is given in Hajivassiliou et al.(1996). In a comparison they found the GHK to be the most accurate in the setting they examined.

purchase probiotic yoghurt drinks at least once within the 3 years.[14] Since we are interested in repeated purchasing behaviour over time we constrained the sample to individuals that have bought a probiotic yoghurt drink at least ten times during the three years. This leaves us with 1883 households that bought on average 26 probiotic yoghurt drinks in 3 years. This amounts to 50,082 yoghurt drink purchases. The purchasing data contains information on date of purchase, store type, store postcode, extensive item characteristics on a barcode level, price, quantity and offer description between January 2002 and December 2004. The demographic data comprises yearly information on age and employment status of the main shopper, household size, social class, number and age of children, household durable equipment (cars, appliances), most frequent shopping method, hours watched TV per week and total spend on fast moving consumer goods. The purchasing variables are collected continuously every day and each product has a unique identifier on the bar code level. Equally each household can be identified by a unique number and is updated every year. The data only covers home consumption.

2.4.1 The Product Category

We chose probiotic yoghurt drinks because it is a new product that is only storable for a short period of time so we do not have to consider bulk purchases or other shopping patterns that are only induced by the longevity of a product. Additionally probiotic yoghurt drinks have a rather small number of characteristic and are homogeneous within their characteristic space. One major advantage is that there have been several successful product introductions in this category in and just before the period of observation. One of the three major brands - Müller - introduced their probiotic yoghurt drink in March 2002 and even though the official introduction of the current market leader Danone already happened in 2001 a large number of households tried it for the first time in the middle of 2002. The third major brand that has

[14] TNS estimated that their purchasing data reflects approx. 80% of total food consumption and their sample is representative of the total UK population.

been in the market the longest is Yakult. On top of new brand introduction the companies also introduced two new flavours during our observation period, strawberry in early 2003 and multifruit in the middle of 2003. This allows us to observe the purchase history from the very beginning and see exactly how characteristic and product familiarity influence consumer choices. We restrict the brands to the three main competitors due to estimatability.[15]

Probiotic yoghurt drinks have three different characteristics flavour, fat content and brand, where fat content has only two levels (light, normal), flavour has six levels (original[16], orange, peach, raspberry, strawberry and multifruit) and we observe three major brands (Danone, Yakult, Müller) in our sample. Since not all brands come in all the flavours we do not have 3x2x6 but only 10 different products.

We observe a considerable amount of heterogeneity between purchasing behaviour. The number of individual yoghurt drink purchases ranges between 10 and 210 times within 3 years with a mean of 26 times. Households purchase their yoghurt drinks on average every 21 days. The average household shops at 3 different stores and 2 different storetypes[17] over the 3 year period. Table 2.1 shows the market shares of the 10 different products within our sample, the biggest is still the incumbent Yakult original with slightly over 15% but aggregating over the different characteristics Danone accounts for more than 50% of all purchases within our sample.

There is significant price heterogeneity, both, between products and characteristics but also over time. Figure 2.2 shows the kernel density distribution of unit price per liter and unit price per pack the difference between the two is mainly due to Yakult being packaged in 65 milliliter bottles whereas the other two brands are in 100 milliliter bottles. Since households choose between these prepackaged sizes their price comparison will more likely be on the pack than on the milliliter base, particularly since one bottle is advertised as one daily

[15]These three brands amount to over 90% of our sample.

[16]The original flavour is similar to a "plain" yoghurt taste

[17]There are eight different storetypes, the four biggest supermarket chains (Asda, Morrison/Safeway, Sainsbury, Tesco), discounter, co-ops, cornershops and the upmarket supermarket chain Waitrose.

portion. Figures 2.3 and 2.4 show the kernel density distributions of price per pack of the different characteristics. We see that the low fat is more expensive than the normal fat and even after adjusting for size Yakult is still the most expensive brand with the smallest price variation. Fruit flavours are less expensive than the original flavour but there is no obvious price difference between new flavour introductions (strawberry, multifruit) and old ones. The price differences between different storetypes are shown in Figure 4 and our empirical analysis indicates that individuals usually compare prices between different products within the same storetype. This is not so surprising since at least for our sample the price ranking between different storetypes stays rather constant over the three year period (see Figure 2.6). Overall prices have fallen very slightly over the three year period mainly due to price reductions from Danone. New flavours were introduced in the middle of the price range with higher than average price variation (see Figure 2.6).

2.4.2 Household data

Every household included in this analysis has purchased probiotic yoghurt drinks at least ten times in the last 3 years and did not have missing household characteristics. Figure 1 shows the distribution over number of different products, brands and flavours purchased by households. Most households purchase two different brands and three different flavours, but around 1% of households have tried all ten products. Further household characteristics included in the analysis are summarised in Table 2, so 80% of the main shoppers are women, the average age is 48 and they watch on average 18 hours television per week. Household size is on average 2.8 and we will look at single households separately to see whether switching behaviour is due to variety seeking or different preferences within households.

2.4.3 Switching

Since we are interested in how past product and characteristic familiarity shape future choices we look at the transition probabilities between different characteristics and different products. We count the transitions from one purchasing occasion to the next and document them in a transition matrix where the rows reflect the initial values, and the columns reflect the final values. We do this separately for singles and larger households and neither unconditional characteristic nor product switching patterns seem significantly different for single households. Tables 2.5 to 2.7 show that brand has the lowest switching probability with more than 80% of the probability mass on the main diagonal. fat content is also a highly persistent characteristic in households choices, but they rather stick with the normal fat content than the light version. Most of the switching takes place between different flavours and in particular between the different fruit flavours (in contrast if a household has consumed the original flavour last period it will do so again this period with 85% probability). Looking at the whole product transition matrix this impression is confirmed with the lowest switching probability for Yakult original (21.5% probability of switching from last purchase) and the highest for Müller peach (71.43 % probability of choosing something different this period) and the main switching within brands across different fruit flavours (see Table 2.9&2.10).

A good example of individual switching behaviour is documented in Figure 7 where we show 5 typical switching patterns of single households. Household A is a good example of someone trying out a different brand once and sticking with it without ever switching back. Households B and C switch between different characteristics of the same brand (fat content for B and flavour for C). The difference is that A has made the transition to the new characteristic after three trials whereas C is still undecided between raspberry and strawberry but the interval between switches decreases. Household D switches between three characteristics of the same brand and does not seem to have settled down on one in particular one and household E's purchases are all over the place switching between all three characteristic types, the rarest of switching patterns.

But what happens to switching over time? If we believe in a learning based model of behaviour one would expect behaviour to settle down after some initial period and the frequency of switches to decrease over time. Figure 2.8 plots the average number of switches across all households over time. On the horizontal axis we use purchase occasions since households have different actual shopping frequencies and on the y-axis we document the number of average switches. We can see that product and flavour switches decrease over time with experience whereas brandswitches stay nearly constant on a significantly lower level. In combination with the fact that price variation is not decreasing over time within or between characteristics this is one indication that a learning based model fits actual consumer purchases quite well.

2.5 The Results

As a benchmark we estimated the simple product-only reinforcement model. The exogenous variables here are only cumulative product attractions but no characteristic attractions, or in other words the $C'_{inlt}s$ are all equal to zero. In this baseline model we also do not include prices or household characteristics or any other covariates. This should give us an indication whether adding characteristic familiarity makes a difference. If it does we would learn that consumers not only accumulate product but also characteristic experiences when they make their purchasing decisions, and that these experiences influence their future purchases. Thus characteristic models are not only a parsimonious way of estimating consumer choices but they are actually the superior behavioural model. Table 2.9 shows the estimated coefficients of the multinomial logit for the different product attractions. We can see that most but not all own product attractions are positive and significant. Table 10 now shows the estimated coefficients for our model of product and characteristic reinforcement. The two main differences are that now all own attractions are positive and significant and that the R^2 increased from 22 to 30 percent. The fact that all own attractions are positive in the characteristics model points out that the probability of choosing a particular product again today does not only depend on my past experience with this exact product but on my past experience with all its characteristics. The overall higher explanatory power of the model confirms our behavioural rational that consumers decompose choices over products into choices over characteristics.

Looking more closely at our estimates from the product and characteristic level reinforcement model reveals that on a characteristic level brand familiarity seems to be more important than flavour or fat content. The same brand attractions are positive a number of times whereas the same fat content or the same flavour i.e. the coefficient for Müller strawberry when looking at Danone strawberry, shows hardly ever a positive and significant coefficient. This is over and beyond the familiarity that is already included in our cumulative attractions. Our cummulative attractions give equal weight to each characteristic and this might be evidence that some characteristics are more important than others. The propensity to choose familiar brands is consistent with the finding of Erdem and Keane (1996) where consumers were found to avoid less familiar brands because they were not certain about the brands' benefit.

Not Surprisingly including unit price per pack as one of the exogenous variables into the model does make a big difference. Table 2.11 shows that the own unitprice per pack coefficients are significantly negative for all products except Yakult light. This is not unexpected since Yakult light exhibits a very similar price pattern than the baseline product of Yakult. But own price is not the only important price variable that influences consumer decisions the price of competitor products matters as well.[18] Here all the coefficients are positive so an increase in the average of competitor product prices within the same storetype leads to

[18] For own price we calculate the unitprice per pack (100ml for danone & muller products and 65ml for yakult). For competitor unit prices we use the unweighted monthly price average of all competitor products within the same storetype. We tried several other price averages across regions or only across products but this one worked best. Which is not completely surprising since one would expect consumers to compare different yoghurt drink prices within the same store rather than across different ones.

an increase purchasing probability of the incumbent product. Again with the exception of Yakult light, here consumers seem not to be influenced by competitor price changes. But even though including prices into the model improves the overall explanatory power it does not change the significance of the reinforcement coefficients.

Table 2.12 shows the estimated coefficients for the full model including household characteristics. Household characteristics that play a significant role in choosing between different products are age and sex of the main shopper.[19] Households with a male main shopper are significantly less likely to shop light products. Older main shopper are less likely to purchase a new brand and more likely to stick with the original unfruity flavour.

The amount of hours a household watches TV per week has also a differential impact on their product purchases. It is solely important for the newest brand - Müller - and increase it's purchasing probability significantly. Changes in weekly TV exposure do not have an impact on Danone or Yakult branded products. We cannot distinguish whether the newest brand advertises more on television or whether households that watch a lot of television are more likely to try out new brands.

To see how strongly our estimates are biased if we do not take learning into account is shown in Table 2.12b. Here we estimate the model with prices and household characteristics only. Not only decreases the overall explanatory power of the model form 0.41 to 0.21, the remaining coefficients are also consistently biased. This is particularly crucial for the own and other price coefficients. All own price coefficients are biased upwards and also the influence of competitor price changes is significantly exaggerated for all but two products. The household characteristic coefficients are also mostly upward biased when learning is not taken into account. Age is still important when choosing a new brand but the coefficients are significantly overestimated, the same is true for the hours of TV exposure per week. The gender of the household head does not come in significant if learning is ignored. This un-

[19] Other household variables that we investigated like employment status, social class or household size do not play a significant role.

derlines the importance of learning when analysing repeated consumer choices of experience goods. It also points at the importance of the panel dimension of the data set. Only if we can observe the same individual making purchases over time and take their learning patterns into account can we correctly assess i.e. own and other price elasticities.

We also estimate the full model for singles only. Table 2.13 shows the results and we can see that singles behave very similarly in terms of product and characteristic level reinforcement learning. They also react similarly to changes in product prices, but age has a slightly different impact for singles. It does not influence the choice of new brands negatively but increasing age makes singles rather less likely to choose light products. Male singles as male shoppers of larger households choose significantly less often the light fat version of a product. The main difference between only singles and the whole sample seems their reaction to TV exposure. TV exposure does not seem to influence purchase decisions for singles at all.

The multinomial logit coefficients are only informative insofar as to whether they are positive or negatively significant. We can not infer the direct economic impact from their size in this nonlinear model. To understand the economic significance we need to calculate marginal effects. One way of understanding the magnitude of the effect of product and characteristic familiarity is to look at the difference in the marginal effects after a change in this variable. We increase the number of a particular product purchase for everyone and calculate the marginal effect before and after this change. The distribution of the difference in marginal effects over all products is shown in Figure 2.9. All differences are positive which shows that even after increasing product and characteristic experience by 10% consumers are not satiated. The positive difference is particularly pronounced for newer characteristics like the strawberry flavour that was only introduced halfway during our sample. Here the impact of an additional 10% of experience leads to a 12% higher probability of consuming it in the future for Danone strawberry and a 5.5% higher probability for Müller strawberry.

But this impact looks rather modest if we compare it to a 10% own price increase. Figure 2.10 shows the impact of a 10% own product price increase and a 10% change in the increase of the average of all the other products in the same storetype. As expected all own price increases lead to a negative difference in marginal effects with the exception of Yakult where consumers seem particularly price inelastic. Müller's products seem particularly price sensitive which might be due to the fact that Müller is the newest brand and has not yet established the same brand familiarity/loyalty. This can be very costly and a 10% own price increase can lead to up to 50% less purchase probability (e.g. for Müller raspberry). An average price increase of all the competitor products by 10% has a comparatively small impact on purchase probability. The main beneficiaries seem to be Danone products that increase their purchase probability up to 20% (for Danone multifruit).

In terms of household characteristics the biggest impact is the gender of the main shopper (Figure 2.11). Increasing the percentage of male shopper by 10% leads to an 11% decrease in light product purchases (more pronounced for Yakult than for Danone). Increasing TV exposure by 10% (an average extra 1.8 hours per week) leads only to a modest increase in Müller (the newest brand) purchases of around 2.3 % in total. Increasing average age by 10% (4.8 years on average) has a very small economic impact it decrease the purchase probability of the newest brand and the new flavour introduction by less than one per cent in total. Hence we can explain a significant portion of choice heterogeneity by including household characteristics.

Concerned about this restrictive property of the IIA assumption we did two robustness checks. We estimated the basic characteristic reinforcement attraction model without price or consumer characteristics with a small random sample of our population using a multinomial probit which can represent any substitution pattern. The results did not look significantly different. As a second test we reestimated the multinomial logit of the basic model (only characteristic level reinforcement) on a subset of choices (leaving out the new characteristics) . The Hausman and McFadden (1984) test indicated that the coefficients of the subset of alternatives are not significantly different from the full model.

2.6 Conclusion and Discussion

In a simple framework of reinforcement learning we have shown that both product and characteristic-level experiences matter for the product choice of the consumer. We find evidence that consumers decompose their consumption choice between products into their characteristics and that characteristic familiarity explains a significant part of product choices over time. Overall we see that our adaptive learning model does explain consumer choices - particularly of new products - well. Our empirical analysis for probiotic yoghurt drinks shows that there is no satiation of any of the products or characteristics yet. All the own product attraction coefficients are positive and significant. We see more within brand switching than across brand switching which is documented in the transition matrix but can also be seen in the positive significant same brand attractions versus often insignificant same flavour or fat content attraction coefficients. This finding is in line with Erdem & Keane (1996) and others who focus on brand loyalty only. Of course our model gives the same weight to all characteristics and it would be an interesting extension to test the relative importance of the different characteristics.

Including prices does improve the model significantly but does not change the significance or sign of the reinforcement coefficients. Both own as well as competitor price changes play an important role. The economic impact of a 10% own price increase can decrease product purchase probability up to 50%.

Household characteristics have a differential impact on product choices. Older households are less likely to switch to the newest brand and male main shoppers are significantly less likely to purchase products with a lower fat content. The amount of television watched by the main shopper only influences their purchase of the newest brand positively. Single households react similarity to product and characteristic familiarity than families. Also gender and age play a similar role for singles the only difference in single household behaviour is that purchasing decisions do not change with the amount of TV exposure. Thus including household characteristics helps us significantly to explain choice heterogeneity over time. To understand how learning is distributed over household is interesting but would go beyond the scope of this paper. Similarly it could be worthwhile investigating whether households can be clustered into meaningful subgroups. For example looking at individuals with lots of observations could help to understand the sofar omitted individual effects.

In the future we would like to allow for more general substitution patterns between different product characteristic as well as allowing unobserved factors to be correlated over time. Since this is not possible within the framework of the multinomial logit model we might want to use more flexible simulation based estimators like the multinomial probit. So far we have only checked the robustness of our multinomial logit estimates with a small random subsample and only for the case without household characteristics and prices. Here the results looked very similar indicating that the IIA property of the multinomial logit does not pose a crucial restriction in this simple case.[20] This might change when looking at the total sample and including other exogenous variables.

[20] Due to computational restrictions we could so far only reestimate the model for this very small random subsample.

2.7 Tables and Figures

Table 2.1: Product market shares

Product	Frequency	Percent
danone multifruit	2,077	4.15
danone orange	7,354	14.68
danone original	6,751	13.48
danone original light	6,993	13.96
danone strawberry	4,671	9.33
muller peach	3,521	7.03
muller raspberry	5,408	10.8
muller strawberry	3,456	6.9
yakult original	7,656	15.29
yakult original light	2,195	4.38
Total	50,082	100

Table 2.2: Descriptive statistics

	mean	std
Stores		
total number of stores	1839.00	
# of stores by HH	3.26	1.90
# of storetypes by HH	2.53	1.18
purchases		
HH tenure	654.89	305.10
Purchase Frequency	21.31	43.90
# of products by HH	3.99	1.85
# of brands by HH	1.81	0.69
# of flavours by HH	3.09	1.45
Learning		
Attraction "light"	20.54	20.01
Attraction "flavour"	17.26	19.34
Attraction "brand"	19.72	20.00
Attraction "product"	15.00	17.00
Attraction "z"	72.37	74.52
households		
HH size	2.83	1.27
% of female shoppers	81.84	
Age	48.14	15.01
TV		
weekly TV hours	18.52	11.04

Table 2.3: Transition matrix light

	all		singles		
	light		light		
Light	0	1	0	1	Total
0	92.19	7.81	90.77	9.23	100
1	34.17	65.83	31.41	68.59	100
Total	81.51	18.49	77.33	22.67	100

Table 2.4: Transition matrix brand

	all brand			singles brand			
Brand	danone	muller	yakult	danone	muller	yakult	Total
Danone	88.77	7.23	4.00	90.00	5.97	4.03	100
Muller	15.46	81.38	3.16	16.69	81.25	2.06	100
Yakult	11.83	4.27	83.91	12.97	2.37	84.66	100
Total	55.57	24.77	19.65	59.27	20.40	20.33	100

Table 2.5: Transition matrix flavour (all)

All Flavour	multifruit	orange	original	peach	raspberry	strawberry	Total
Multifruit	41.84	17.11	13.69	3.21	3.84	20.32	100
Orange	4.80	59.72	17.69	3.10	3.66	11.03	100
Original	1.26	5.53	85.02	1.63	1.98	4.58	100
Peach	1.77	5.69	10.47	28.57	30.98	22.52	100
Raspberry	1.34	4.56	8.00	20.61	46.50	18.98	100
strawberry	5.94	8.90	12.5	8.88	12.71	51.07	100
Total	4.2	14.48	47.13	6.99	10.81	16.39	100

Table 2.6: Transition matrix flavour (singles)

Singles Flavour	multifruit	orange	original	peach	raspberry	strawberry	Total
Multifruit	49.35	13.87	11.94	2.26	2.58	20.00	100
Orange	4.84	69.31	14.42	2.42	2.03	6.97	100
Original	1.19	4.31	87.78	1.40	1.83	3.49	100
Peach	2.90	5.31	13.77	33.57	29.71	14.73	100
Raspberry	1.63	3.42	7.17	20.85	54.23	12.70	100
strawberry	8.39	9.25	13.51	9.82	11.10	47.94	100
Total	4.99	15.59	52.20	6.39	9.61	11.21	100

Table 2.7: Transition matrix product (all)

	all	danone					muller			yakult		
	product	multifruit	Orange	original	light	strawberry	peach	raspberry	strawberry	original	light	Total
danone	multifruit	41.84	17.11	5.08	6.22	17.52	3.21	3.84	2.80	1.56	0.83	100
	orange	4.80	59.72	7.96	4.89	9.30	3.10	3.66	1.73	4.11	0.73	100
	original	1.69	8.18	56.39	19.31	4.85	1.83	2.01	1.25	4.04	0.44	100
	light	1.76	5.87	17.18	62.14	4.51	1.55	2.28	1.18	2.57	0.96	100
	strawberry	8.98	13.78	6.39	6.84	50.05	2.16	4.03	4.39	2.68	0.70	100
muller	peach	1.77	5.69	3.63	2.95	4.13	28.57	30.98	18.40	3.33	0.56	100
	raspberry	1.34	4.56	2.45	2.45	3.52	20.61	46.50	15.46	2.45	0.66	100
	strawberry	1.79	2.22	2.46	1.91	6.07	18.08	24.58	40.41	1.69	0.80	100
yakult	original	0.52	3.77	3.92	2.44	1.79	1.58	1.83	0.93	78.50	4.72	100
	light	0.95	2.37	1.51	3.03	1.80	1.37	1.51	1.14	14.81	71.51	100
	Total	4.20	14.48	13.42	14.06	9.42	6.99	10.81	6.97	15.22	4.43	100

Table 2.8: Transition matrix product (singles)

	singles	danone					muller			Yakult		
	prodcode	multifruit	orange	original	light	strawberry	peach	raspberry	strawberry	original	Light	Total
danone	multifruit	49.35	13.87	4.84	3.87	17.74	2.26	2.58	2.26	0.32	2.90	100
	orange	4.84	69.31	6.97	4.36	6.00	2.42	2.03	0.97	2.42	0.68	100
	original	1.11	7.10	53.66	23.17	2.77	2.88	2.33	1.44	4.77	0.78	100
	light	1.77	4.30	15.18	67.62	2.45	1.43	2.11	1.18	2.78	1.18	100
	strawberry	11.46	14.32	5.97	6.21	49.64	2.39	2.63	3.58	0.95	2.86	100
muller	peach	2.90	5.31	5.80	6.04	2.42	33.57	29.71	2.32	1.93	0.00	100
	raspberry	1.63	3.42	2.28	3.75	1.79	20.85	54.23	0.91	0.81	0.33	100
	strawberry	3.87	1.76	3.87	1.76	5.28	20.77	23.59	4.86	1.41	2.82	100
yakult	original	0.38	2.17	5.09	2.64	1.04	0.38	1.04	0.47	82.38	4.43	100
	light	2.08	3.47	3.47	3.82	6.25	0.35	2.08	1.74	12.85	63.89	100
	Total	4.99	15.59	13.65	18.22	6.82	6.39	9.61	4.39	15.88	4.45	100

Table 2.9: Coefficients from a multinomial logit model with only product reinforcement

All		danone					muller			yakult
		multifruit	orange	original	original light	strawberry	peach	rasberry	strawberry	yakult light
danone	multifruit	0.8510*	0.4674**	0.2496	0.3368*	0.4471**	0.4448	0.3378	0.4484	0.3147
		(0.449)	(0.144)	(0.152)	(0.182)	(0.152)	(0.448)	(0.256)	(0.254)	(0.187)
	orange	0.1013**	0.1537	0.0121	0.0357	0.0873**	0.0331	0.0375	0.0129	0.0142
		(0.033)	(0.323)	(0.026)	(0.034)	(0.033)	(0.033)	(0.032)	(0.035)	(0.034)
	original	0.0162	-0.0114	0.1269**	0.0698**	-0.0146	-0.0223	-0.0389	-0.0299	-0.0422
		(0.031)	(0.026)	(0.028)	(0.029)	(0.028)	(0.038)	(0.027)	(0.029)	(0.041)
	original light	0.1715	0.0554	0.2067**	0.2971**	0.1546**	0.1015*	0.1142*	0.0878	0.0888
		(0.152)	(0.051)	(0.049)	(0.049)	(0.052)	(0.055)	(0.057)	(0.051)	(0.055)
	strawberry	0.2345**	0.1132**	0.0468	0.1071	0.3672**	0.0291	0.0785	0.1539**	0.0759
		(0.061)	(0.059)	(0.062)	(0.065)	(0.061)	(0.069)	(0.064)	(0.061)	(0.083)
muller	peach	0.3273**	0.2299	0.1227	0.1363	0.3024	0.5164**	0.3674**	0.4007**	-0.0487
		(0.088)	(0.871)	(0.092)	(0.089)	(0.486)	(0.085)	(0.086)	(0.087)	(0.119)
	rasberry	0.1307	0.0696	-0.0151	-0.0128	0.1012	0.1998**	0.2923*	0.1967**	0.0051
		(0.146)	(0.044)	(0.090)	(0.071)	(0.065)	(0.043)	(0.163)	(0.042)	(0.071)
	strawberry	0.3921*	0.1260	0.2303	0.2242*	0.4644**	0.4532**	0.4668**	0.6459**	0.0476
		(0.226)	(0.152)	(0.140)	(0.128)	(0.124)	(0.126)	(0.125)	(0.125)	(0.158)
yakult	yakult	-0.1593**	-0.1831**	-0.1511**	-0.1594**	-0.1619**	-0.1731**	-0.1816**	-0.1722**	-0.0608**
		(0.021)	(0.023)	(0.022)	(0.071)	(0.017)	(0.017)	(0.019)	(0.016)	(0.009)
	yakult light	-0.2301**	-0.2348**	-0.3407**	-0.0632	-0.1751*	-0.0573	-0.0929*	-0.0627	0.1557**
		(0.086)	(0.059)	(0.092)	(0.068)	(0.099)	(0.049)	(0.054)	(0.049)	(0.048)
	const.	-1.8788**	0.0785	-0.0848	-0.3692**	-0.8221**	-0.9333**	-0.5857**	-1.0210**	1.3371**
		(0.116)	(0.105)	(0.112)	(0.113)	(0.109)	(0.112)	(0.103)	(0.105)	(0.161)
R2	0.2261									
Obs.	50082									

baseoutcome: yakult original
Standard Errors are reported in brackets and clustered on the household level
* Statistically significant at the 10 percent level, ** statistically significant at the 5 percent level

Table 2.10: Coefficients from a multinomial logit model with product and characteristic reinforcement

All		danone					muller			yakult
		multifruit	orange	original	original light	strawberry	peach	rasberry	strawberry	yakult light
danone	multifruit	0.2995**	0.1535**	0.0695	0.1117*	0.1306**	0.1684**	0.1206*	0.1677**	0.1318*
		(0.061)	(0.061)	(0.063)	(0.061)	(0.064)	(0.062)	(0.066)	(0.065)	(0.074)
	orange	0.0753**	0.0034**	0.0492**	0.0388	0.0492**	-0.0375	-0.0295	-0.0501**	-0.0185
		(0.020)	(0.001)	(0.019)	(0.021)	(0.021)	(0.021)	(0.021)	(0.022)	(0.023)
	original	0.1172**	0.0666**	0.0201**	0.0331	0.1015**	-0.0570	-0.0630**	-0.0625**	-0.0585
		(0.027)	(0.025)	(0.003)	(0.025)	(0.027)	(0.031)	(0.026)	(0.025)	(0.036)
	original light	0.1016**	0.5249**	0.1382**	0.1190**	0.0935**	0.6431**	0.0809**	0.0612**	-0.0052
		(0.036)	(0.031)	(0.036)	(0.032)	(0.037)	(0.034)	(0.033)	(0.030)	(0.037)
	strawberry	-0.0665	-0.0349	-0.0882*	-0.0516	0.0325**	-0.1081**	-0.0785	-0.0881*	0.0009
		(0.045)	(0.047)	(0.048)	(0.047)	(0.004)	(0.048)	(0.046)	(0.045)	(0.055)
muller	peach	0.0500	0.0570	0.0148	0.0270	0.0517	0.1149**	0.0499	0.0502	-0.0248
		(0.044)	(0.044)	(0.049)	(0.044)	(0.427)	(0.042)	(0.042)	(0.043)	(0.056)
	rasberry	-0.0483*	-0.2306**	-0.0542	-0.0475	-0.0489*	-0.0434*	0.1245**	0.0518*	0.0021
		(0.026)	(0.029)	(0.039)	(0.035)	(0.027)	(0.026)	0.003)	(0.026)	(0.042)
	strawberry	0.1157*	0.0225	0.1126	0.0968	0.1165*	0.1373**	0.1389**	0.2169**	0.0229
		(0.065)	(0.076)	(0.071)	(0.067)	(0.064)	(0.066)	(0.065)	(0.065)	(0.086)
yakult	yakult	-0.0478*	-0.0459**	-0.0054	-0.0586**	-0.0517*	-0.0968**	-0.0849**	-0.0973**	-0.0551**
		(0.027)	(0.019)	(0.028)	(0.021)	(0.029)	(0.021)	(0.020)	(0.018)	(0.021)
	yakult light	-0.0551	-0.0454	-0.7566**	-0.0377	-0.0392	0.0162	-0.0055	0.0180	0.0837**
		(0.046)	(0.031)	(0.048)	(0.036)	(0.051)	(0.030)	(0.031)	(0.028)	(0.034)
	const.	-1.8769**	0.0588	-0.0947	-0.3692	-0.8426	-0.8382**	-0.4448**	-1.0201**	-1.3209**
		(0.115)	(0.104)	(0.111)	(0.112)	(0.108)	(0.102)	(0.102)	(0.106)	(0.159)
R2	0.3028									
Obs.	50082									

baseoutcome: yakult original
Standard Errors are reported in brackets and clustered on the household level
* Statistically significant at the 10 percent level, ** statistically significant at the 5 percent level

Table 2.11: Coefficients from a multinomial logit model with product and characteristic reinforcement & price parameters

All		danone					muller			yakult
		multifruit	orange	original	original light	strawberry	peach	rasberry	strawberry	yakult light
Danone	multifruit	0.1702**	0.0399	0.0447	-0.0048	0.0126	0.0435	-0.0052	0.0386	0.0931*
		(0.049)	(0.049)	(0.052)	(0.051)	(0.052)	(0.056)	(0.061)	(0.061)	(0.053)
	orange	0.0331*	0.0363**	0.0112	-0.0019	-0.0111	0.0019	0.0142	-0.0022	-0.0085
		(0.018)	(0.018)	(0.017)	(0.019)	(0.019)	(0.021)	(0.021)	(0.021)	(0.017)
	original	0.0814**	0.0344	0.0549**	0.0002	0.0655**	-0.0195	-0.0237	-0.0295	-0.0524
		(0.026)	(0.024)	(0.025)	(0.024)	(0.025)	(0.031)	(0.027)	(0.026)	(0.032)
	original light	0.0912**	0.0485**	0.1317**	0.1192**	0.0806**	0.0649*	0.0796**	0.0697**	-0.0034
		(0.034)	(0.021)	(0.034)	(0.031)	(0.035)	(0.035)	(0.034)	(0.032)	(0.033)
	strawberry	-0.0339	-0.0278	-0.0770*	-0.0410	0.0611**	-0.0752*	-0.0530	-0.0651*	0.1000**
		(0.036)	(0.039)	(0.040)	(0.041)	(0.025)	(0.041)	(0.042)	(0.039)	(0.041)
muller	peach	0.0720*	0.0726*	0.0358	0.0496	0.0676	0.1432**	0.0737*	0.0749	-0.0162
		(0.042)	(0.044)	(0.048)	(0.046)	(0.041)	(0.043)	(0.044)	(0.045)	(0.048)
	rasberry	-0.0273	-0.0180	-0.0379	-0.0334	-0.0289	-0.0331	0.0234**	-0.0415*	0.0072
		(0.023)	(0.024)	(0.034)	(0.038)	(0.024)	(0.025)	(0.011)	(0.024)	(0.032)
	strawberry	0.0104	-0.0318	0.0460	0.0301	0.0201	0.0327	0.0403	0.1207**	0.0117
		(0.045)	(0.055)	(0.052)	(0.053)	(0.044)	(0.048)	(0.047)	(0.047)	(0.061)
yakult	yakult	-0.0478*	-0.0246	0.0072	-0.0390*	-0.0319	-0.0699**	-0.0594**	-0.0663**	-0.0548**
		(0.027)	(0.019)	(0.027)	(0.021)	(0.025)	(0.021)	(0.019)	(0.019)	(0.020)
	yakult light	-0.0551	-0.0548	-0.1592**	-0.0535	-0.0411	0.0012	-0.0154	-0.0028	0.0823**
		(0.046)	(0.033)	(0.046)	(0.037)	(0.044)	(0.032)	(0.032)	(0.031)	(0.032)
price	UP own	-0.4536**	-0.4571**	-0.4389**	-0.4370**	-0.4668**	-0.7561**	-0.7318**	-0.7643**	-0.0211
		(0.035)	(0.037)	(0.036)	(0.034)	(0.034)	(0.036)	(0.036)	(0.036)	(0.016)
	UP other	0.1981**	0.0679**	0.0943**	0.0932**	0.0918**	0.0504**	0.0727**	0.0811**	0.0182
		(0.015)	(0.001)	(0.015)	(0.012)	(0.015)	(0.017)	(0.016)	(0.018)	(0.015)
	const.	7.7466**	13.7851**	12.3685**	12.0741**	12.5733**	20.9294**	20.0975**	19.8650**	-1.0034
		(1.243)	(1.317)	(1.299)	(1.231)	(1.261)	(1.337)	(1.324)	(1.363)	(0.705)
R2	0.4134									
Obs.	50082									

baseoutcome: yakult original

Standard Errors are reported in brackets and clustered on the household level

* Statistically significant at the 10 percent level, ** statistically significant at the 5 percent level

Table 2.12a: Coefficients from a multinomial logit model with product and characteristic reinforcement & price parameters & household characteristics

all		danone					muller			yakult
		multifruit	orange	original	original light	strawberry	peach	rasberry	strawberry	yakult light
danone	multifruit	0.1723**	0.0419	-0.0436	-0.0026	0.1596**	0.0398	-0.0040	0.0394	0.0964*
		(0.051)	(0.049)	(0.052)	(0.051)	(0.053)	(0.056)	(0.062)	(0.061)	(0.052)
	orange	0.0322*	0.0368**	-0.0116	-0.0016	-0.0101	0.0038	0.0164	-0.0001	-0.0091
		(0.018)	(0.017)	(0.017)	(0.018)	(0.018)	(0.020)	(0.021)	(0.021)	(0.016)
	original	0.0812**	-0.0344	0.0534**	-0.0014	0.0641**	-0.0169	-0.0205	-0.0253	-0.0522
		(0.026)	(0.025)	(0.025)	(0.024)	(0.025)	(0.031)	(0.027)	(0.026)	(0.032)
	original light	0.0911**	0.0491	0.1319**	0.1190**	0.0801**	0.0668	0.0789**	0.0690**	-0.0046
		(0.034)	(0.031)	(0.034)	(0.031)	(0.035)	(0.035)	(0.033)	(0.032)	(0.033)
	strawberry	-0.0356	-0.0297	-0.0751*	-0.0397	0.0581**	-0.0773	-0.0572	-0.0691	0.0086
		(0.036)	(0.039)	(0.041)	(0.041)	(0.025)	(0.041)	(0.041)	(0.045)	(0.041)
muller	peach	0.0709*	0.0698	0.0304	0.0454	0.0684*	0.1426	0.0740*	0.0762*	-0.0165
		(0.039)	(0.044)	(0.046)	(0.044)	(0.038)	(0.041)	(0.041)	(0.042)	(0.048)
	rasberry	-0.0276	-0.0178	-0.0328	-0.0303	-0.0297	-0.0329	0.0225**	0.0426*	0.0076
		(0.022)	(0.029)	(0.033)	(0.037)	(0.023)	(0.024)	0.002)	(0.026)	(0.031)
	strawberry	0.0104	-0.0307	0.0440	0.0285	0.0189	0.0328	0.0402	0.1199**	0.0121
		(0.044)	(0.076)	(0.051)	(0.052)	(0.043)	(0.046)	(0.046)	(0.046)	(0.059)
yakult	yakult	-0.0261	-0.0233	0.0078	-0.0387	-0.0325	-0.0681	-0.0600**	-0.0678**	-0.0556**
		(0.025)	(0.019)	(0.027)	(0.021)	(0.025)	(0.021)	(0.019)	(0.019)	(0.021)
	yakult light	-0.0595	-0.0568	-0.1606**	-0.0536	-0.0407	-0.0027	-0.0153	-0.0022	0.0835**
		(0.042)	(0.033)	(0.045)	(0.037)	(0.044)	(0.032)	(0.031)	(0.031)	(0.032)
price	UP own	-0.4573**	-0.4615**	-0.4450**	-0.4420**	-0.4683**	-0.7609**	-0.7355**	-0.7677**	-0.0209
		(0.036)	(0.037)	(0.037)	(0.035)	(0.035)	(0.037)	(0.037)	(0.037)	(0.016)
	UP other	0.1980**	0.0677**	0.0933**	0.0919**	0.0929**	0.0511**	0.0751**	0.0849**	0.0179
		(0.015)	(0.001)	(0.016)	(0.015)	(0.015)	(0.017)	(0.016)	(0.018)	(0.015)
TV	TV	0.0042	0.0048	0.0012	-0.0034	0.0019	0.0181**	0.0130**	0.0164**	-0.0060
		(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
demo	ageMS	-0.0002	0.0036	0.0111**	0.0080*	-0.0068	-0.0046**	-0.0104**	-0.0163**	-0.0020
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.005)
	sexMS	-0.2161	-0.1985	-0.1875	-0.3883**	-0.2811*	-0.2730	-0.3262	-0.1768	-0.1152**
		(0.180)	(0.157)	(0.156)	(0.159)	(0.168)	(0.175)	(0.181)	(0.188)	(0.017)
	const.	8.0716**	13.9998**	12.2549**	12.4118**	13.2197**	21.2554**	20.7609**	20.5040**	-0.6811
		(1.273)	(1.345)	(1.333)	(1.267)	(1.295)	(1.385)	(1.375)	(1.417)	(0.776)

R2	0.4155
Obs.	50082

baseoutcome: yakult original

Standard Errors are reported in brackets and clustered on the household level, * Statistically significant at the 10 percent level, ** statistically significant at the 5 percent level

Table 2.12b: Coefficients from a multinomial logit model only price parameters & household characteristics

All		danone					muller			yakult
		multifruit	orange	original	original light	strawberry	peach	rasberry	strawberry	yakult light
price	UP own	-0.5190**	-0.5143**	-0.4859**	-0.4806**	-0.5240**	-0.8246**	-0.8040**	-0.8369**	-0.0272*
		(0.035)	(0.036)	(0.037)	(0.036)	(0.035)	(0.038)	(0.037)	(0.038)	(0.016)
	UP other	0.2373**	0.0671**	0.0768**	0.0840**	0.1415**	0.0945**	0.0965**	0.1393**	0.0308*
		(0.013)	(0.001)	(0.014)	(0.015)	(0.014)	(0.026)	(0.017)	(0.020)	(0.018)
TV	TV	0.0173*	0.0068	0.0027	-0.0053	0.0076	0.0209**	0.0139	0.0179*	-0.0040
		(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.009)	(0.011)
demo	ageMS	-0.0085	-0.0021	0.0080	0.0078	-0.0209**	-0.0123*	-0.0208**	-0.0272**	0.0012
		(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
	sexMS	0.0272	0.1866	-0.1460	-0.3032	-0.0504	-0.0161	-0.2597	0.0816	-0.3437
		(0.244)	(0.245)	(0.241)	(0.255)	(0.236)	(0.280)	(0.266)	(0.274)	(0.296)
	const.	8.9529**	15.1821**	13.8906**	13.8661**	13.9039**	21.8812**	22.6742**	21.2910**	-0.8637
		(1.377)	(1.425)	(1.513)	(1.454)	(1.402)	(1.758)	(1.546)	(1.622)	(0.992)
R2	0.2055									
Obs.	50082									

Table 2.13: Coefficients from the full multinomial logit model (Table12) for singles

Singles		danone					muller			yakult
		multifruit	orange	original	original light	strawberry	peach	rasberry	strawberry	yakult light
danone	multifruit	0.3324**	0.1737	0.1649	0.1252	0.0966	0.2166	0.1291	0.0735**	0.2585
		(0.161)	(0.167)	(0.169)	(0.171)	(0.198)	(0.169)	(0.178)	(0.199)	(0.174)
	orange	0.1257*	0.1921**	0.1261*	0.1403**	0.1825**	0.1549**	0.1751**	0.1768**	0.0661
		(0.063)	(0.059)	(0.065)	(0.058)	(0.067)	(0.062)	(0.063)	(0.063)	(0.059)
	original	-0.0556	-0.0236	0.0592**	0.0341	-0.1167	-0.0212	-0.0136	0.0001	-0.1112
		(0.099)	(0.073)	(0.023)	(0.062)	(0.088)	(0.085)	(0.101)	(0.086)	(0.081)
	original light	0.0702	0.0271	0.1415**	0.1370**	0.1108*	0.0904	0.0605	0.0542	-0.0301
		(0.062)	(0.058)	(0.056)	(0.047)	(0.059)	(0.064)	(0.066)	(0.058)	(0.057)
	strawberry	-0.5183**	-0.4526**	-0.5791**	-0.5317**	0.3747**	-0.5694**	-0.4776**	-0.4617**	-0.1756
		(0.155)	(0.145)	(0.151)	(0.140)	(0.155)	(0.159)	(0.157)	(0.161)	(0.157)
muller	peach	-0.2040**	-0.2093**	-0.3338**	-0.2356**	-0.1755**	0.1298**	0.1937**	0.1941**	-0.0473
		(0.086)	(0.088)	(0.097)	(0.098)	(0.085)	(0.063)	(0.093)	(0.092)	(0.095)
	rasberry	-0.2582**	-0.2201**	-0.2077**	-0.2085**	-0.2353**	0.2784**	0.1779**	0.2447**	-0.1529
		(0.073)	(0.075)	(0.077)	(0.081)	(0.075)	(0.077)	(0.075)	(0.077)	(0.091)
	strawberry	0.8749**	0.7634**	0.9504**	0.8623**	0.8566**	0.8977**	0.8213**	0.9285**	0.3648
		(0.290)	(0.288)	(0.289)	(0.291)	(0.292)	(0.293)	(0.294)	(0294)	(0.313)
yakult	yakult	-0.1551**	-0.1435**	-0.0964**	-0.1429**	-0.1444**	-0.1581**	-0.1736**	-0.1801**	-0.1459**
		(0.038)	(0.049)	(0.042)	(0.037)	(0.037)	(0.041)	(0.042)	(0.039)	(0.041)
	yakult light	0.0938	0.0898	-0.0742	0.0173**	0.1137	0.0851	0.1262*	0.1341*	0.2411**
		(0.063)	(0.079)	(0.078)	(0.064)	(0.070)	(0.075)	(0.068)	(0.069)	(0.078)
price	UP own	-0.3570**	0.3694**	-0.3218**	-0.3427**	-0.3862**	-0.6783**	-0.6562**	-0.6482**	0.0214
		(0.052)	(0.050)	(0.057)	(0.050)	(0.051)	(0.052)	(0.051)	(0.053)	(0.033)
	UP other	0.1971**	0.0916**	0.0799**	0.0881**	0.1080**	0.0738*	0.1084**	0.1794**	0.0353
		(0.039)	(0.003)	(0.036)	(0.031)	(0.037)	(0.041)	(0.038)	(0.044)	(0.030)
TV	TV	-0.0133	0.0010	-0.0143	-0.0228*	0.0221	-0.0022	-0.0145	0.0061	-0.0125
		(0.014)	(0.013)	(0.013)	(0.012)	(0.016)	(0.016)	(0.016)	(0.015)	(0.017)
demo	ageMS	-0.0067	-0.0002	0.0038	-0.0046**	-0.0153	-0.0003	-0.0120	-0.0079	-0.0243**
		(0.011)	(0.009)	(0.010)	(0.002)	(0.011)	(0.011)	(0.012)	(0.011)	(0.012)
	sexMS	-0.1185	-0.3213	-0.3513	-0.8184**	0.2545	-0.2453	-0.5864	-0.0394	-1.0310**
		(0.384)	(0.341)	(0.348)	(0.346)	(0.424)	(0.398)	(0.406)	(0.422)	(0.407)
	const.	4.9768**	10.3484**	9.1920**	10.8635**	8.4325**	17.5973**	17.7213**	12.8095**	-0.7547
		(2.247)	(2.048)	(2.333)	(2.121)	(2.331)	(2.303)	(2.339)	(2.455)	(1.631)
R2	0.4461									
Obs.	6767									

baseoutcome: yakult original

Standard Errors are reported in brackets and clustered on the household level, * Statistically significant at the 10 percent level, ** statistically significant at the 5 percent level

Figure 2.1: Household purchases across products and characteristics

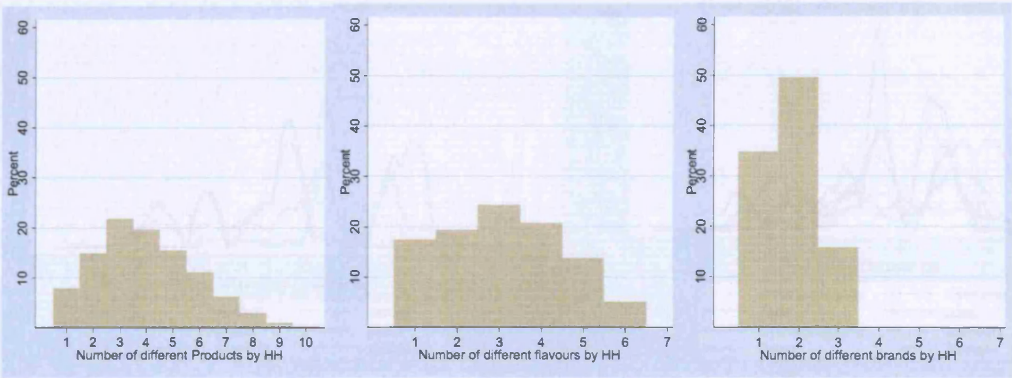


Figure 2.2: Unitprice kernel densities by litre and by pack

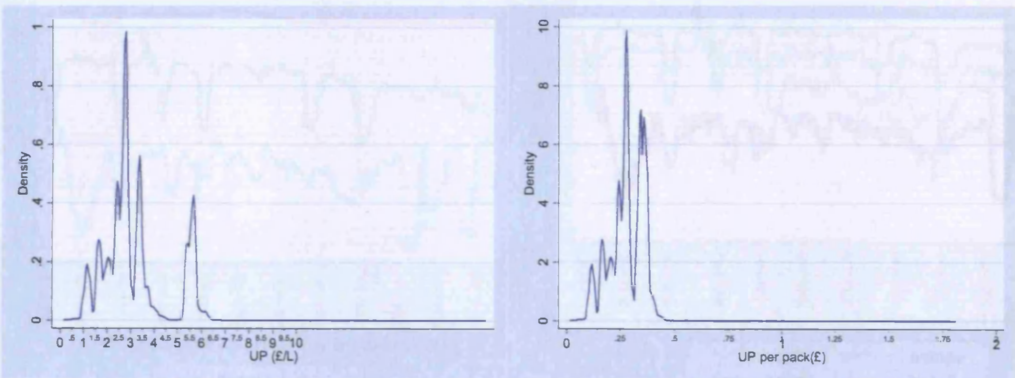


Figure 2.3: Unitprice kernel densities by pack & by fatcontent/brand

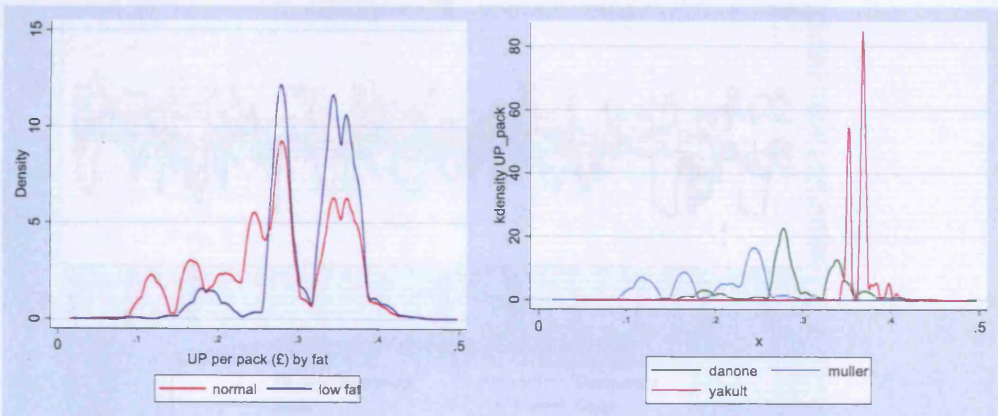


Figure 2.4: Unitprice kernel densities by pack & by flavour/storetype

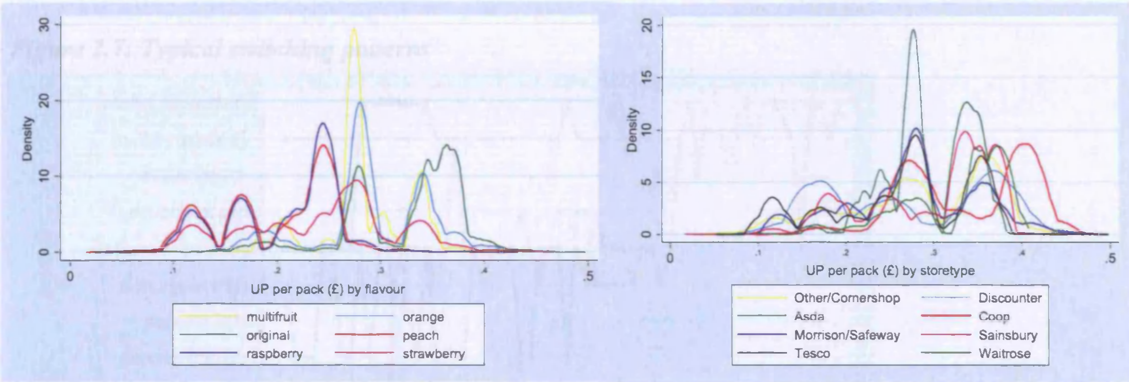


Figure 2.5: Unitprices over time by brand/flavour

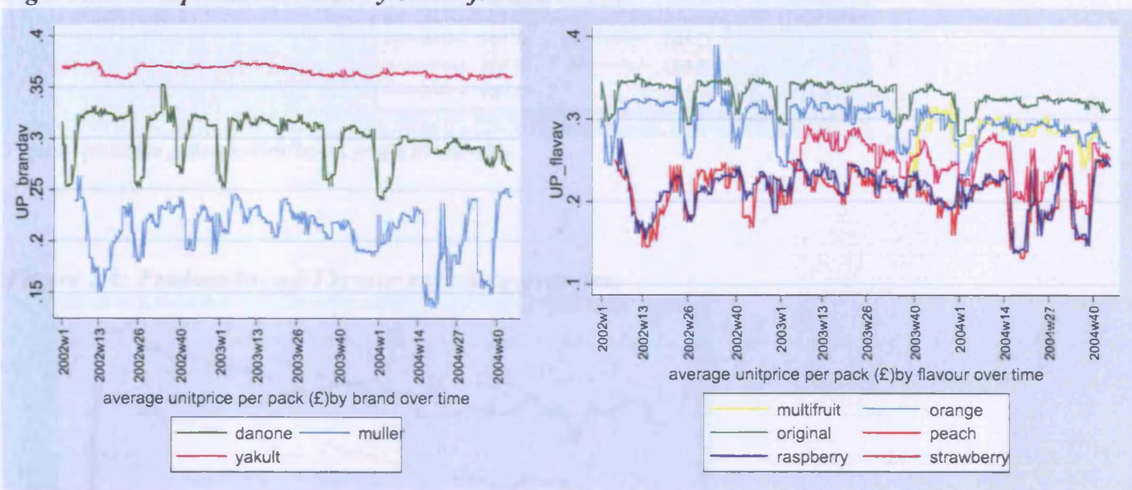


Figure 2.6: Unitprices over time by storetype

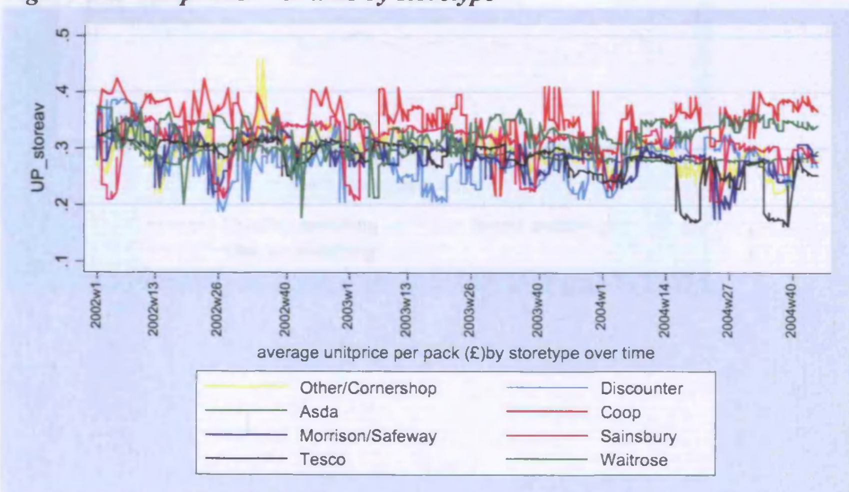
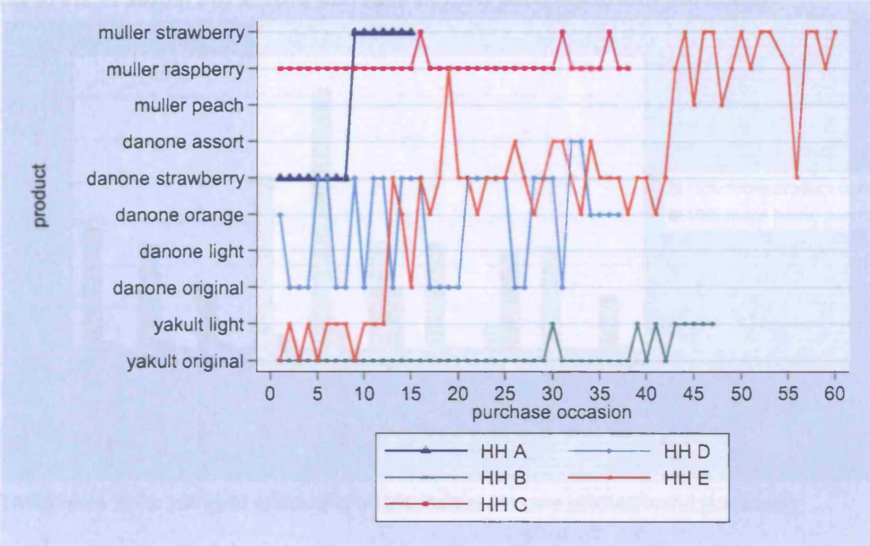


Figure 2.7: Typical switching patterns



Typicall purchase patterns of different single households

Figure 2.8: Product/Brand/Flavour switching over time

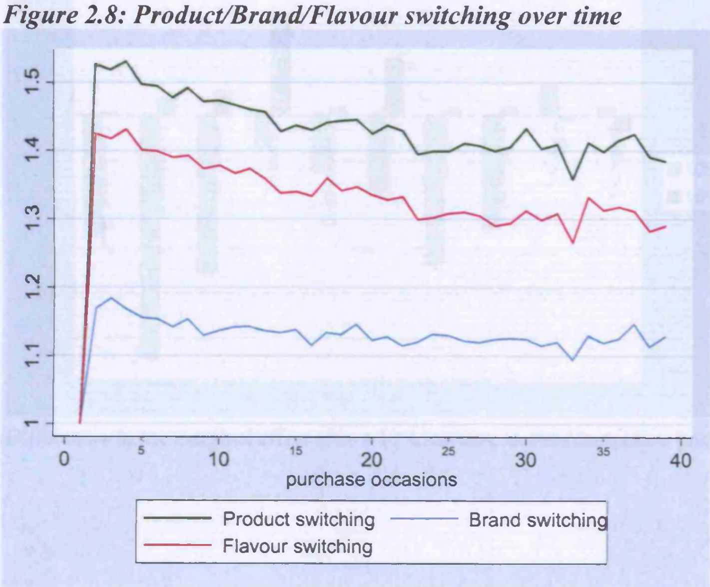
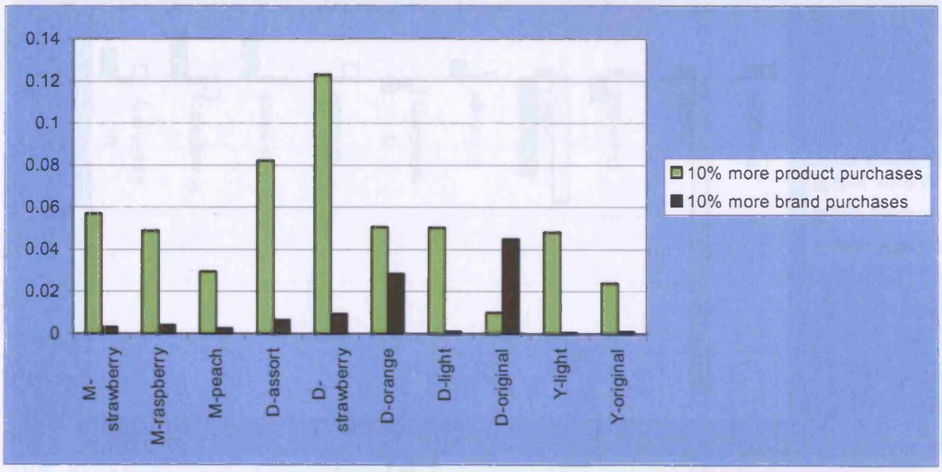
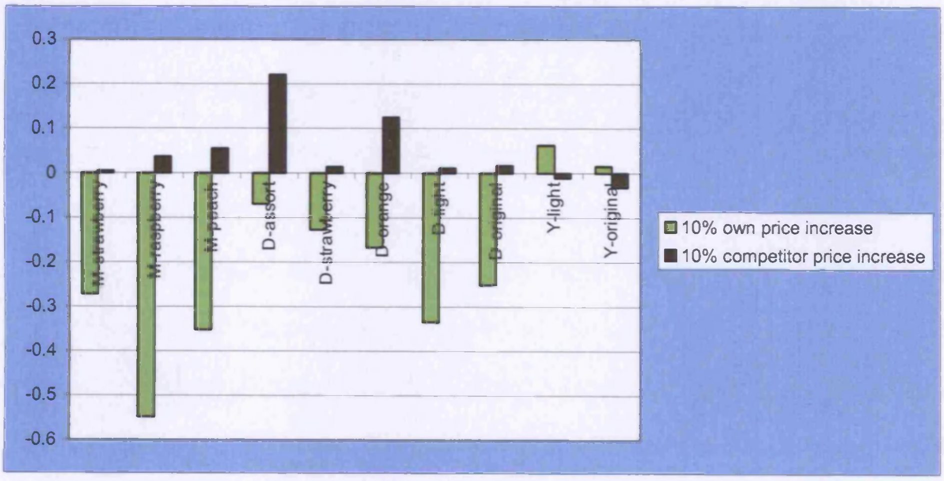


Figure 2.9: Impact of a 10% increase in own product/brand purchases



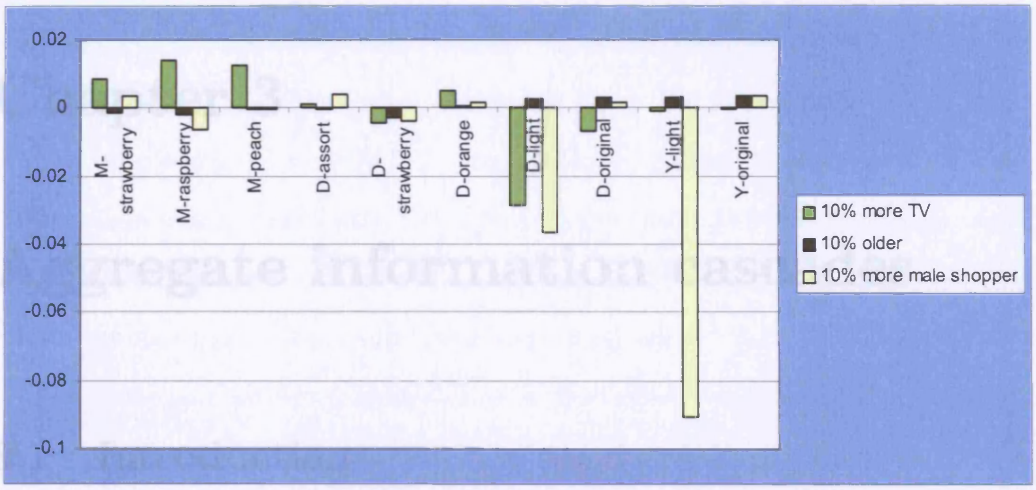
(Differences in the marginal effect after a 10% increase in own product/brand purchases)

Figure 2.10: Impact of a 10% increase in own/competitor prices



(Differences in the marginal effect after a 10% increase in own/competitor prices)

Figure 2.11: Impact of a 10% increase in TVexposure/age/% male shopper



(Differences in the marginal effect after a 10% increase in TVexposure/age/%male shopper)

Chapter 3

Aggregate information cascades

3.1 Introduction

A central agency in health policy must decide how to disclose information on the adoption of a new treatment. One possibility is to inform the doctors on how many others have already decided to adopt the new treatment. Another is to inform them on how many have considered doing it but have judged that it is preferable to stick to the old practice. Another possibility is to reveal both, the number of doctors in favor of the new practice and the number of physicians in favor of the old one. Can the way the information is disclosed make a difference for the diffusion of the new treatment? Suppose the agency is uncertain on the effects of the new treatment and considers as the worst case scenario the situation in which the new treatment is widely adopted while ultimately resulting in worse health outcomes than the old one, for instance because of side effects. Which disclosure policy should the agency employ?

Intuitively, one would think that the disclosure of *all* available information should maximize social welfare. However, this only hold if there are no externalities and a doctor's decision about a treatment for a patient is obviously a case with huge externalities. While the life of a patient may matter a lot to the physician it surely matters more to the patient, at least in most cases.

Bickchandani et al. have argued in their seminal paper on information cascades that the adoption of medical procedures is often based on fairly weak information and that in many cases doctors tend to imitate others. As an example they cite the widespread use of tonsillectomies in the sixties and seventies and argue that it was essentially an information cascade (where the shere the fact that the majority of physiscians employed the procedure overrode any private information individual doctors might have had access to). And in this case it was a "down cascade"—a cascade that generated the worst outcome since it eventually turned out that tonsillectomies did more harm than good.

One of the questions we raise in this paper is whether a central agency that can influence the transmission information may want to decide to *withhold* some information. And indeed it turns out that this might be the case. This is due to the following simple result: If agents have access to information about both, how many have others adopted and how many have not, both types of cascades are possible—cascades where everybody adopts and cascades where everybody does not adopt. (This is, of course, what we know from the literature already.) However, if the agency decides only to inform about how many have adopted then *there is only one type of cascade*—one where evrybody adopts. And if the agency only reports about those who have decided *not* to adopt, there is also only one type of cascade, one where everybody decides *not* to adopt. Thus, if the really bad outcome is the one where everybody herds on the new treatment while the new treatment is, in fact, worse than the old, the agency can avoid that this happens—by withholding the information about how many others so far have decided to adopt.

While information about medical procedures appears to be a particularly appealing policy-relevant example for the theory we develop here, there are, in fact, many other applications. At the core of our paper is simply a model of social learning where agents have to make binary decisions, like adopting a new treatment or not, making an investment or not. Our crucial assumption (where we deviate from the previous literature) is that when an agent makes his decision, he can only observe the total number of others who have already taken *one* of the two available actions, for example, the total number of others who have already decided to adopt the new treatment while he cannot observe the number of those who have previously contempleted the choice but opted otherwise.

This structure appears very natural for many examples that have previously been discussed in the literature on social learning. A restaurant goer who must decide whether or not he wants dinner at a particular restaurant he stands in fron of, may be able to peer through the window to see how many others decide to have dinner there but he can only speculate about how many others tood before the door and decided to pass. A similar example is that of a manager of a venture capital firm who discusses a project with an inventor who needs capital to develop a new product. Say, the inventor has already secured funds from two other venture capital firms. Then we may expect that the inventor gladly mentions this to the manager with whom he negotiates. The manager who will have some private information about the viablity of the project will, of course, also extract some information from knowing that there were two others who obviously had information telling them that the project is good. On the other hand, since we can safely assume that the inventor will not tell the manager about how often he was turned down, the manager can only guess how many other firms had information telling them that the project was bad. Of course, knowing the number of venture capital firms that would, in principle, finance such a project will help him to update his beliefs about this.

In all these examples agents who have to decide between two options have only aggregate information about one of the two options while in all models present in the literature so far, a decision maker has access to information about the *individual* choices of others who decided before him. The standard model of social learning (Banerjee, 1992, and Bikhchandani et al., 1992), for instance, contemplates a sequence of binary decisions which are all observable. Agent n knows whether each predecessor in the sequence, from agent 1 to agent $n-1$, decided in favor of one option or the other. We find that our set up captures quite naturally some situations frequenlty arising in social interactions. Like in the case of the restaurant goer or the venture capital firm in many circumstances, a decision maker can gather some aggregate information (how many agents have already adopted, invested, chosen a restaurant), but he can rarely observe all the individual decisions. Clearly, if the decision is binary, knowing the number of agents who have made a certain decision helps to update on the number of agents who have already made the opposite decision. But this is not equivalent to knowing it. And we can show that it makes an important difference. In particular, we can show, as we have mentioned already above, that with aggregate information there can only be one type of information cascade.

An information cascade occurs when agents rationally neglect their own private information, i.e., when they choose the same action independently of the signal they recieve (for instance following their predecessors). In the standard sequential model of Banerjee (1992) and Bikhchandani et al. (1992) different types of cascades can arise. If the decision is binary, say, between investing and not, there can be cascades where, from a certain point onwards, *all* decision makers decide *to invest*, as well as cascades where, from a certain point onwards, all decisions makers decide *not to invest.* At a first glance, one could think that this is the case in our set up, too. If a restaurant goer sees many people in a restaurant, he could cascade and join the crowd; if he sees the restaurant empty, he could cascade and decide to pass on the restaurant too. However, we can prove that, on the contrary, only the first cascade is possible. In equilibrium it cannot be that there is a cascade on the unobservable action, e.g., that the restaurant remains empty altough some people have read good reviews on it.

In many cases, one of the available actions arises naturally as the observable action. In one of the examples above it is the number of investment decisions (as opposed to the number of declined investments). In another it is the number of decisions to dine in a particular restaurant (as opposed to the number of passes). There are, however, important cases where *third parties* may have the power to decide what kind of information is provided to agents.

This is the case from which we started: the disclosure policy of a health agency. The worst case scenario for the agency is that in which agents start to herd on the wrong action. This induces the biggest welfare loss. The central agency may not know what the wrong action is (if it did, it could simply announce it) but often the welfare losses will be asymmetric: the case where the new treatment is indeed more effective but everybody sticks to the old one may be better or worse than the opposite case, where the old treatment is more effective but everybody switches to the new. Since the central agency can *choose* which action to make observable, it can rule out one of the two erroneous cascades—by withholding information. In summary, we like to argue that the aggregate information set up that we introduce here has not only several intriguing properties—some of which in stark contrast to the predictions of the standard model—it also has a potenatially important policy implications.

We extend the literature on social learning not only theoretically but also experimentally. After establishing our theoretical results, we have also implemented our set up in a laboratory experiment. The aim of this experiment is to have a first "reality check," since we believe that a theory like ours is more appealing if it is not completely off the mark. Previous models of social learning, starting with the standard sequential model with all observable actions, have been extensively tested in the laboratory, with results that are sometimes positive and sometimes less favorable. We find that our theory finds support in the laboratory, although some interesting anomalies emerge.

Two papers close in spirit to ours are Celen and Kariv (2004 and 2005). Similarly to our approach, Celen and Kariv are interested in understanding what happens when we remove the strong assumption that agents can observe the entire history of individual decisions. Celen and Kariv (2004) extend the standard model of sequential social learning by allowing each agent to observe the decision of his immediate predecessor only. Celen and Kariv (2005) test this model in the laboratory. The theoretical predicion of these authors is that, when each agent can only observe his immediate predecessor's decision, behavior does not settle on a single action. Long periods of herding can be oserved, but switches to the other action occur. As time passes, the periods of herding become loger and longer, and the switches more and more rare. However, their predictions find limited support in the laboratory.

Our model, in contrast, does fairly well in the laboratory. In just two simple treatments we find that the main comparative statics go all in the right directions. Moreover, most of the deviations we observe pretty much mirror those that have dosumented in the previous experimental literature on standard cascade models. There is, however, one interesting anomaly that deserves to be mentioned here. With a fifty-fifty prior we find some "reluctance to go first." Agents who observe the empty restaurant but have a good signal (and who should therefore decide to enter the restaurant) do so in only 56% of all cases. We conjecture that this reluctance to go first may stem from the fact that in many other environments, in particular in those with endogenous seqeuncing, one would indeed prefer others to do the first costly experiment. We also conjecture that this behavioural effect mirrors the non-triviality of our theoretical result.

The remainder of the paper is organized as follows. In Section 2 we introduce the formal model. We present its equilibrium analysis in Section 3 and Section 4 contains an example. Section 5 describes the experiment and Section 6 concludes.

3.2 The Model

In our economy there are n agents who have to decide in sequence whether or not to take up a certain option. For convencience, we shall refer to this choice as the decision on whether or not to *invest*. Time is discrete and indexed by $t = 1, 2, ..., n$. Each agent makes his choice only once in the sequence. Agent i's $(i = 1, 2, ..., n)$ action space is given by $\{0, 1\}$, where 1 is interpreted as investment. Player i's action is denoted by $I_i \in \{0, 1\}$. An agent's payoff π_i depends on his choice and on the true state of the world $\omega \in \{0, 1\}$. The prior probability of $\omega = 1$ is $r \in (0, 1)$. If $\omega = 1$ agent i receives a payoff of 1 if he choses to invest, and a payoff of zero otherwise; vice versa if $\omega = 0$. That is,

$$\pi_i = \omega I_i + (1-\omega)(1-I_i).$$

The sequence in which agents make their choices is randomly determined before the first agent makes a decision, and agents are, w.l.o.g., (re-)numbered according to their positions: agent i chooses at time i only. All sequences are equally likely. The agents, however, are *not* informed about which sequence has been chosen. Furthermore, they do not know their own position in the sequence. When called upon, agent i is only informed about the total number of agents before him who have decided to invest. In other words, the decision to invest is assumed to be the only *observable* action. This means that, while the aggregate number of investments is observable, each individual decision to invest or not is not publicly known. We denote the total number of agents who have invested before agent i by T_i, i.e., agent i is informed about $T_i = \sum_{j=1}^{i-1} I_j$. In addition to observing T_i, each agent i receives a private signal $\sigma_i \in \{0,1\}$ that is correlated with the true state ω. In particular, we assume that each agent receives a symmetric binary signal distributed as follows:

$$\Pr(\sigma_i = 1 \mid \omega = 1) = \Pr(\sigma_i = 0 \mid \omega = 0) \equiv q.$$

Note that, conditional on the state of the world, the signals are i.i.d.. We shall refer to $\omega = 1$ as the "good state" and to $\omega = 0$ as the "bad state." A signal pointing in the direction of the good state ($\sigma_i = 1$) shall be called "good signal" and a signal pointing in the opposite direction ($\sigma_i = 0$) "bad signal." We assume that $1 > q > r$ and that $r + q > 1$. These conditions ensure that, in the one-agent case, an agent would invest after a good signal and would not invest after a bad signal, which renders the problem interesting. Note that these two conditions also imply that $q > \frac{1}{2}$, i.e., that the signal respects the monotone likelihood ratio property. Finally, the signal is not perfectly informative, which makes social learning possible and relevant.

Agent i's information set is, therefore, represented by the couple (T_i, σ_i). An agent's strategy $\mathfrak{I}_i$ maps (T_i, σ_i) into an action, i.e.,

$$\mathfrak{I}_i : \{0, 1, 2, ..., n-1\} \times \{0, 1\} \rightarrow \{0, 1\}.$$

An agent's mixed strategy induces, for each (T_i, σ_i), a probability with which the agent invests. We denote the probability with which agent i invests after observing (T_i, σ_i) by $\mathcal{I}_i(T_i, \sigma_i)$.

To conclude the description of our model, it is useful to introduce the notion of an *aggregate information cascade.* The definition is virtually identical to the standard definition of information cascade, with the characteristic that histories are summarized by the aggregate statistic T_i.

Definition 1 *An aggregate information cascade (AIC) occurs when, along the equilibrium path, there is a decision after which all agents choose an action independently of their signal. In particular:*

In an aggregate up cascade (AUC) there is a critical trigger value T^{UP} such that if $T_k = T^{UP}$ all agents from k onwards choose to invest regardless of their private signals. Consequently, there is some k such that $T_{k+j} = T_k + j$ for all $j = 1, ..., n-k$.

In an aggregate down cascade (ADC) there is a critical trigger value T^{DOWN} such that if $T_k = T^{DOWN}$ all agents from k onwards choose not to invest regardless of their private signals. Consequently, in an ADC there is some k such that $T_{k+j} = T_k$ for all $j = 1, ..., n-k$.

We are now ready to start analysing the equilibrium decisions in our economy.

3.3 Equilibrium Analysis

The ultimate goal of our analysis is to understand the social learning process that occurs in our economy. Each agent can learn about the true state of the world from the aggregate information that he receives about other agents' choices. This can lead to better decisions. On the other hand, it may be that also in our economy, as in the canonical model of social learning of Banerjee (1992) and Bickchandani et al. (1992), there is room for information cascades, i.e., for situations in which agents take the same decision independently of their private information. In such a case, the process of information aggregation will not be efficient. We will show that, indeed, "up cascades" of investments are possible even in our set up, as they are in the canonical model. In contast, "down cascades" of non-investments never occur in equilibrium.

We restrict the entire analysis to symmetric Bayesian Nash equilibria. Our economy is represented by a symmetric game and there is nothing in the environment that could help agents to coordinate on an asymmetric outcome. Therefore, the restriction to symmetric equilibira is very natural.

To start our analysis, it is convenient to focus first on the case of $T_i = 0$, in which an agent observes that no one has invested before him. At a first glance, the decision problem in such a situation appears to be fairly complicated. If the agent knew that $T_i = 0$ simply because he is the first decision maker, then he should certainly follow his private signal, since that is the only information available. If, instead, he knew that he is not the first decision maker, then he may decide to ignore the signal and not invest independently of it, as other agents have already chosen the non-investment option. Intuitively, one might think that $T_i = 0$ is pretty bad information if there are many players. Suppose that n is very large and you observe that nobody has invested before you. But at the same time you own private signal is good. Would you trust your own signal? Of course, this will depend on the other agents' strategy choices. While the problem is made hard due to the fact that the agent does not know his position in the sequence it is made easier due to the fact that the only thing that matters about other agents' strategies is what these specify for the very same case of $T = 0$.

To attack the problem, let us start with the following definition:

Definition 2 *An initially-pure equilibrium (IPE) is a Nash equilibrium that prescribes pure actions for $T_i = 0$ and both possible signal realizations $\sigma_i = 0$ and $\sigma_i = 1$.*

Note that there can be mixing in an IPE after observing $T_i > 0$. The definition of an IPE just excludes the cases in which an agent mixes after observing $T_i = 0$. We are able to establish some results that focus on $T_i = 0$. First, we prove that in any IPE agents must follow their signals after observing $T_i = 0$: there cannot exist IPEs in which an agent plays independently of his signal or play against it.

Lemma 3 *In any IPE, an agent follows his own signal if he observes that nobody has invested so far, i.e., $\mathcal{I}_i(0, \sigma_i) = \sigma_i$ for all i.*

Proof. We prove this by contradiction. Suppose that for $T_i = 0$ agents choose either to invest always or never (independently of the private signal). Consider the latter possibility first, i.e., consider a pure-strategy equilibrium with $\mathcal{I}_i(0, 0) = \mathcal{I}_i(0, 1) = 0$. Then, along the equilibrium path, nobody ever invests and, for all agents $j = 1, ..., n$, $T_j = 0$. Hence, $T_i = 0$ does not reveal any information on the true state of the world. Since the posterior probability that $\omega = 1$ is still r, agent i is better off by following his informative signal σ_i. Next, consider the case of investment after $T_i = 0$, i.e., an equilibrium with $\mathcal{I}_i(0, 0) = \mathcal{I}_i(0, 1) = 1$. In this case, along the equilibrium path, only the first agent in the sequence observes that nobody else has invested before. That is, $T_i = 0$ if and only if $i = 1$. Hence, after observing $T_i = 0$ agent i knows that he is the first agent in the sequence and, thus, should follow his signal. Finally, suppose that for $T_i = 0$ agents choose to play against their private information, i.e., consider a pure-strategy equilibrium with $\mathcal{I}_i(0, \sigma_i) = 1 - \sigma_i$. Then, along the equilibrium path, after observing $T_i = 0$, agent i knows that he is either the first in the sequence or other agents before him have received the good signal. In both cases, he should follow his signal. ■

While we have shown that in any IPE an agent who observes zero investments should follow his signal, it remains unclear whether such equilibria exist. The next lemma identifies a necessary and sufficient condition under which an IPE does indeed exist.

Lemma 4 *An IPE exists if and only if* $r \geq \frac{1-q^n}{2-(1-q)^n-q^n}$.

Proof. We first prove that it is indeed optimal for an agent i to follow his own signal after $T_i = 0$ provided that everybody else does so and that the condition stated in the lemma holds. (Notice that what another agent j does for $T_j > 0$ is irrelevant for agent i's optimal choice of $\mathcal{I}_i(0, \sigma_i)$.) Assuming such behaviour of others, an agent i who observes $T_i = 0$ and $\sigma_i = 1$ attaches to the good state a posterior of

$$\begin{aligned}\Pr(\omega &= 1 \mid T_i = 0, \sigma_i = 1) = \\ &\frac{rq\sum_{j=1}^{n}(1-q)^{j-1}}{rq\sum_{j=1}^{n}(1-q)^{j-1} + (1-r)(1-q)\sum_{j=1}^{n} q^{j-1}}.\end{aligned}$$

He will follow his good signal if this posterior is at least $1/2$, i.e., if

$$rq\sum_{j=1}^{n}(1-q)^{j-1} \geq (1-r)(1-q)\sum_{j=1}^{n} q^{j-1}.$$

Solving for the sums and rearranging the terms, we get the condition in the lemma. To complete the proof we have to show that an agent i who assumes that the others play according to the rules stated in the lemma and who observes $T_i = 0$ and $\sigma_i = 0$

does *not* invest, i.e., we need that

$$\begin{aligned}\Pr(\omega &= 1 \mid T_i = 0, \sigma_i = 0) = \\ &\frac{r(1-q)\sum_{j=1}^{n}(1-q)^{j-1}}{r(1-q)\sum_{j=1}^{n}(1-q)^{j-1} + (1-r)q\sum_{j=1}^{n}q^{j-1}} \\ &< \frac{1}{2}\end{aligned}$$

or

$$r(1-q)\sum_{j=1}^{n}(1-q)^{j-1} < (1-r)q\sum_{j=1}^{n}q^{j-1},$$

which can be written as

$$\frac{r}{(1-r)} < \frac{q^2}{(1-q)^2}\frac{1-q^n}{1-(1-q)^n}. \tag{3.1}$$

Since $r < q$ we also have $\frac{r}{1-r} < \frac{q}{1-q}$. Hence, inequality (3.1) holds if $\frac{q}{1-q}\frac{1-q^n}{1-(1-q)^n} > 1$. This can be rewritten as $2q > 1 + q^{n+1} - (1-q)^{n+1}$ which is obviously true for $q > 1/2$.

■

Notice that the condition imposed in the lemma is always fulfilled if $r > 1/2$, i.e., when the good state is initially more likely than the bad state, an IPE always exists.

We now turn our attention to Nash equilibria that are not initially pure. The next lemma trivially follows from Bayesian updating. We state it formally because we shall need it later on. The lemma after that shows that in an equilibrium that is not an IPE agents who observe $T_i = 0$ never invest if their signal is bad, but will invest with some positive probability if their signal is good.

Lemma 5 *(i) In any equilibrium, $\mathcal{I}_i(T_i, 1) \geq \mathcal{I}_i(T_i, 0)$ for all T_i, with the inequality being strict if $\mathcal{I}_i(T_i, 0) < 1$. (ii) In any equilibrium, $0 < \mathcal{I}_i(T_i, 0) < 1 \Rightarrow \mathcal{I}_i(T_i, 1) = 1$ and $0 < \mathcal{I}_i(T_i, 1) < 1 \Rightarrow \mathcal{I}_i(T_i, 0) = 0$ for all T_i.*

Proof. In equilibrium, each agent will infer the same information from observing a particular value of T_i. Whatever the posterior induced by just observing T_i, it follows immediately from Bayes' rule that an agent who has an additional good signal will be more optimistic about the good state than an agent with a bad signal. Hence, the first part follows. Note that the inequality is weak only when $\mathcal{I}_i(T_i, 0) = 1$. The first part of the proposition results from these considerations and from expected payoff maximization. The second part follows from the same argument and the additional observation that mixing requires indifference. ■

Lemma 6 *In any equilibrium that is not an IPE,* $\mathcal{I}_i(0,0) = 0$ *and* $0 < \mathcal{I}_i(0,1) < 1$ *for all* i.

Proof. Given Lemma 5 we just need to rule out an equilibrium with $0 < \mathcal{I}_i(0,0) < 1$ and $\mathcal{I}_i(0,1) = 1$. For an agent to be indifferent between investing and not after $T_i = 0$ and $\sigma_i = 0$ we need $\Pr(\omega = 1 \mid T_i = 0, \sigma_i = 0) = 1/2$. Using Bayes' rule this can be re-written as

$$r \Pr(T_i = 0, \sigma_i = 0 \mid \omega = 1) = (1 - r) \Pr(T_i = 0, \sigma_i = 0 \mid \omega = 0),$$

or

$$r \sum_{j=1}^{n} (1-q)^j (1-p)^{j-1} = (1-r) \sum_{j=1}^{n} q^j (1-p)^{j-1},$$

where p denotes the probability with which all other agents who see $T_i = 0$ and $\sigma = 0$ invest. Rewriting this as

$$\sum_{j=1}^{n} \left[\left(r(1-q)^j - (1-r) q^j \right) (1-p)^{j-1} \right] = 0$$

makes it obvious that there is no $p > 0$ that solves the equation: since we have $q > \max\left\{\frac{1}{2}, r\right\}$the left-hand side is strictly negative for any positive p. ■

Having characterized equilibria that are not initially pure, we must discuss whether they exist. The next lemma introduces a necessary and sufficient condition for such mixed-strategy equilibria to exist.

Lemma 7 *Mixed-strategy equilibria with $\mathcal{I}_i(0,0) = 0$ and $0 < \mathcal{I}_i(0,1) < 1$ for all i exist if and only if there is a $p \in (0,1)$ that solves*

$$r[1 - (1 - pq)^n] = (1 - r)[1 - (1 - p(1 - q))^n].$$

Proof. The lemma follows from observing that, if all other agents $j \neq i$ use $I_i(0,0) = 0$ and $I_i(0,1) = p$ agent i's indifference between investing and not investing after $T_i = 0$ and $\sigma_i = 1$ requires $\Pr(\omega = 1 \mid T_i = 0, \sigma = 1) = 1/2$. After applying Bayes' rule and some algebraic manipulation, this gives

$$rq \sum_{j=1}^{n} (1 - pq)^{j-1} = (1 - r)(1 - q) \sum_{j=1}^{n} (1 - p(1 - q))^{j-1},$$

which is equivalent to the equation in the lemma. ■

This lemma completes our characterization of equilibrium decisions after observing $T_i = 0$. In the following proposition we summarize what we have learned so far.

Proposition 8 *If $r \geq 1/2$ agents who observe $T_i = 0$ follow their signal in all equilibria.*

If $\frac{1-q^n}{2-(1-q)^n-q^n} < r < 1/2$ there is an equilibrium where agents who observe $T_i = 0$ follow their signal but there may also be other (mixed-strategy) equilibria where agents who observe $T_i = 0$ follow their signal if it is bad and mix if it is good.

If $r \leq \frac{1-q^n}{2-(1-q)^n-q^n}$ there can only be equilibria where agents who observe $T_i = 0$ follow their signal if it is bad and mix if it is good.[1]

[1] *Notice that the third part of the proposition touches on an existence problem. For obvious reasons we have restricted our analysis to symmetric Bayesian Nash—in case of bad priors these may fail to exist.*

Proof. The proposition follows immediately from the four lemmas and the observation that $\frac{1-q^n}{2-(1-q)^n-q^n} < 1/2$. ■

Our analysis essentially shows that, when facing a situation with no previous investments, an agent should either follow his signal or use a mixed strategy. An agent should never decide independently of his signal, neither should he decide against it. This clearly indicates that we should not observe a "down cascade" where all agent choose not to invest. In other words, a restaurant will not stay empty forever only because it is empty now. While this puts already a lot of structure on the equilibrium solution of our game, we still need to investigate what happens for different values of the aggregate investment T.

To this purpose, we establish in the next step an intuitive monotonicity result, according to which a higher value of T is always (weakly) good news: when an agent observes a higher number of investments made before him, he cannot be less willing to invest himself. Once this monotonicity lemma is established, we will be able to prove two fundamental results about aggregate cascades.

Lemma 9 *In any equilibrium,* $T_i' \leq T_i'' \implies \mathcal{I}_i(T_i', \sigma_i) \leq \mathcal{I}_i(T_i'', \sigma_i)$ *for both* $\sigma_i = 0$ *and* $\sigma_i = 1$.

Proof. From the first part of Lemma 5 we get that $\text{prob}(T_i = T' + 1 \mid \omega = 1, T_{i-1} = T') \geq \text{prob}(T_i = T' + 1 \mid \omega = 0, T_{i-1} = T')$. Hence,

$$\frac{\text{prob}(T_i = T \mid \omega = 1)r}{\text{prob}(T_i = T \mid \omega = 0)(1-r)}$$

is increasing in T, which gives, by Bayes' rule,

$$\text{prob}(\omega = 1 \mid T_i = T'') \geq \text{prob}(\omega = 1 \mid T_i = T') \Leftrightarrow T'' \geq T'.$$

The statement in the lemma follows directly from that and expected payoff maximization. ■

The intuition for the lemma is as follows. Due to the first monotonicity result in Lemma 5 we know that, for any given player, the transition from T to $T+1$ is more likely in the good state than in the bad state. This is simply true because a good signal is more likely in the good state and a good signal makes an agent more likely to invest. Hence, for any expectations about your own position in the sequence the higher T the more likely it is that you are in the good state. In other words, the more people you see in the restaurant the more likely it is that it is a good restaurant.

We are now ready to state our two main propositions that characterize which forms of cascades will or will not arise. In particular, we will see that aggregate down cascades *never* arise, while aggregate up cascades are *always* part of an equilibrium.

Proposition 10 *In any equilibrium, $\mathcal{I}_i(T_i, 1) > 0$ for all T_i, i.e., an agent with a good signal always invests with positive probability and an ADC never occurs in equilibrium.*

Proof. The proposition follows from Proposition 8 and Lemma 9. ■

There are no cascades on the unobservable action. Incidentally, we note that such a result just comes from an equilibrium argument. One could imagine that, when facing a "low" value of T_i, in order to make his decision, agent i should consider all possible sequences and attach a probability to the event that he is the first in the sequence, or the second, etc. After all, a low number of investments may merely come from the fact that only few agents had the opportunity to invest so far, in which case the low value of T_i should be considered "good" news. Or it could arise from many agents having the option of investing but few only using it, in which case the low T_i should be viewed as bad news. All this inference process could be quite complicated. Our analysis solves the problems by just invoking some equilibrium arguments.

The next proposition considers cascades on the observable action and it shows that AUCs do arise—and are, in fact, part of *any* equilibrium.

Proposition 11 *AUCs are part of* any *equilibrium. In particular, in any equilibrium* $\mathcal{I}_i(T_i, \sigma_i) = 1$ *for all* $T_i > \frac{n}{2}$.

Proof. Consider an agent i who observes $T_i > \frac{n}{2}$ and suppose he *knew* that he were the last agent in the sequence. Further suppose there were no AUC. Then, due to Lemma 9, this agent knows that there were at least T_i good signals and no more than $n - T_i - 1$ bad signals. Hence, even if this agent's own signal is bad, he knows that there were altogether more good signals than bad signals and he will decide to invest. Of course, agent i can't be sure that he really is the last agent. But if he isn't, this means that there were *fewer* bad signals so far, while he can still be sure that there were T_i good signals. Hence, an agent who observes $T_i > n/2$ will always invest and, thus, trigger an AUC. ■

The value $\frac{n}{2}$ is just a lower bound for the critical mass of observable choices that triggers an AUC. Depending on the parameters' values, AUCs may be triggered earlier. But AUCs are indeed part of *all* equilibria. Of course, this does not necessarily imply that AUCs will actually be triggered, since there is always the possibility of sufficiently many bad signals occurring such that the critical T_i that triggers an AUC may not be reached.

3.4 An example

It is now the moment to illustrate our theory through a simple example. The example will highlight some properties of the equilibrium analysis and will be the basis for the experiment that follows.

Consider the case in which $n = 3$ and $r \geq 1/2$. From Proposition 8 we know that $I_i(0, \sigma_i) = \sigma_i$ and from Proposition 8 we know that $I_i(2, \sigma_i) = 1$ and that $I_i(1, 1) = 1$. But what do agents do after observing $T_i = 1$ and $\sigma_i = 0$. As is clear from the results illustrated in Table 1, this depends on further conditions on r and q. Let us first check under which conditions agent i rationally follows his bad signal. Recall that we are analysing a symmetric equilibrium, therefore suppose each opther agent j chooses $I_j(1,0) = 0$. Then it is optimal for agent i to do the same if his posterior for the good state is not bigger than 1/2, i.e., if

$$\frac{r[q(1-q)+2q(1-q)^2]}{r[q(1-q)+2q(1-q)^2]+(1-r)[q(1-q)+2q^2(1-q)]} \leq \frac{1}{2}$$

which is equivalent to

$$q \geq 2r - \frac{1}{2} \equiv \underline{q}.$$

Similarly, $I_i(1,0) = 1$ is optimal if

$$r[q(1-q)+q(1-q)^2] \geq (1-r)[q(1-q)+q^2(1-q)].$$

which is equivalent to

$$q \leq 3r - 1 \equiv \overline{q}.$$

Observing that $\underline{q} \leq \overline{q}$, we obtain three equilibrium regions. For $q < \underline{q}$ there is a unique pure-strategy equilibrium in which $I_i(1,0) = 1$ and an AUC starts with $T_i = 1$. For $q > \overline{q}$ there is a unique pure-strategy equilibrium in which $I_i(1,0) = 0$ and an AUC starts only with $T_i = 2$. Finally, for $\underline{q} \leq q \leq \overline{q}$ both the two pure-strategy equilibria exist and there is a mixed-strategy equilibrium as well—with $I_i(1,0) = \frac{1+2q-4r}{q-r}$.

3.5 An experiment

In order to check whether our theory can reasonably be expected to be of any empirical relevance we conducted a laboratory experiment. Basically, we implemented the three-agent example that we analysed above for two sets of parameters. In both treatments the signal precision is $q = 0.7$. But while the prior in treatment A is $r = 1/2$, the prior in treatment B is more optimistic, $r = 3/4$. For both treatments, theory predicts unique equilibria, shown in Table 3.1. The only difference between the two treatments concerns the critical T that triggers an AUC. In treatment A an AUC is only triggered for $T = 2$, while in treatment B an AUC starts already with $T = 1$.

	Treatment A ($r = 1/2$)		Treatment B ($r = 3/4$)	
	$\sigma = 0$	$\sigma = 1$	$\sigma = 0$	$\sigma = 1$
$T = 0$	0	1	0	1
$T = 1$	0	1	1	1
$T = 2$	1	1	1	1

Table 3.1: Equilibrium investments in both treatments

The experiments were fully computerized and run at the ELSE laboratory at UCL in Spring 2005. In each session subject were matched in multiple groups of three at the beginning of the experiment. The experiment consisted of 15 rounds of game playing and groups remained constant over time. We employed the strategy method, i.e., subjects had to indicate for each possible combination of T and σ whether or not they would invest. Specifically, they had to click radio buttons in a table that was precisely structured as the left panel of Table ??. Once everybody had submitted a strategy the computer drew the state of the world, the sequence in which agents had to decide and each agent's signal. Using subjects' strategies the computer then determined the outcome and subjects received feedback about their actual choice that was relevant, i.e., about their σ and T, as well as about the state ω and their resulting payoff. They did not receive any feedback about other subjects' behavior or outcomes. Altogether, we observed 57 subjects who were randomly assigned to the two treatments, with 11 groups in treatment A, and 8 groups in treatment B.[2]

Table 3.2 summarizes the results using an identical format to Table 3.1. For each combination of σ and T the table shows the mean investment rate. A number of observations is in order. For most cells, average investments are actually quite close to equilibrium predictions. Also, standard errors (shown in parentheses) are rather small, so groups did not

[2]One session we had scheduled for treatment B was cancelled due to too many no-shows.

behave wildly different. However, there appear to be two outliers: the investment rates after, both, $T = 0$ and $\sigma = 1$ in treatment A and after $T = 1$ and $\sigma = 0$ in treatment B, are substantially smaller than the equilibrium prediction of full investment. In case of the first deviation there seems to be some reluctance to go first. This reluctance disappears in treatment B where the prior is much more optimistic. The second deviation is different. Here and AUC should be triggered with $T = 1$ but observed investment rates are below $1/2$. Again there is more scepticism in actual behavior than in equilibrium behavior.

	Treatment A ($r = 1/2$)		Treatment B ($r = 3/4$)	
	$\sigma = 0$	$\sigma = 1$	$\sigma = 0$	$\sigma = 1$
$T = 0$	0.15	0.56	0.21	0.86
	(0.03)	(0.03)	(0.09)	(0.05)
$T = 1$	0.24	0.87	0.41	0.94
	(0.04)	(0.03)	(0.06)	(0.03)
$T = 2$	0.68	0.85	0.76	0.94
	(0.07)	(0.03)	(0.09)	(0.04)

Table 3.2: Average investments rates in both treatments.

(Standard errors in parentheses)

However, the theory captures the comparative statics remarkably well. In particular, after $T = 1$ and $\sigma = 0$, the only information set for which the equilibrium predictions differ for the two treatments, we do observe a substantial and highly significant difference between the two treatments. The doubling of the investment rate is significant at the 5% level.

It is interesting to compare our data for $T = 2$ with results from standard information cascade experiments because with $T = 2$ subjects know for sure in which position they are in the sequence and they also know the entire history, namely that everybody before them invested. We observe that the majority of subjects does herd in this case in both treatments. But herding is far from perfect in this case. However, as it turns out this reluctance pretty much mirrors the data from standard cascade experiments (see, for example, the nice overview in Kubler and Weizsacker 2005). With a more optimistic prior the reluctance becomes smaller. This, too, has been observed in the previous literature (and is also documented in Kubler and Weizsacker.)

Overall, we consider the results from our experiment encouraing. While not offering a full-fledged empirical analysis of our model it does lend its credibility some support and should, in fact, encourage further investigations.

3.6 Discussion

We have introduced a new model of aggregate information cascades. The crucial difference between our model and the models already in the literature is that only one action taken by agents is observable to others who, when it is their turn, simply receive aggregate information about how many others before them took the observable action. We argue that this setup is in many cases realistic, for example, when entrepreneurs seek investors they will typically not inform new investors whom they approach about how many others have turned them down but, surely, they will mention who else decided previoulsy to invest in their project. This asymmetry in observability, which in many cases arises naturally, dramatically affects *all* equilibria in such games. Most importantly, there can be no down cascades. If an action is unobservable, there can be never information cascades where agents take this action.

This has an important consequence for applications where a third party can decide which information it is to release. Consider the case of a central health agency that has observes how many doctors choose a new treatment to cure some diseases. There are two potentially grave scenarios: The new treatment is better but everybody sticks to the old. Or the old treatment is better but everybody switches to the new. While both scenarios are bad, from the view point of the central agency one might be much worse than the other. Say, that the worse scenario is the one where the new treatment is better but not adopted. In that case we have shown is that by *withholding* information, i.e., by publishing only the number of doctors who previously decided to switch and by withholding the information about those who decided to stay with the conservative treatment, the central agency can rule out the worst case where the old treatment is administered by all docotors despite the superiority of the new. Of course, from the viewpoint of the agents more information will always be better than less and, if there are no externalities, total welfare will be reduced if some information is withheld. But if there are externalities, say, if this is about life and death and if patients' care more about their own survival than their doctors do then the central agency may indeed have very good reasons to withhold some information it has access to.

Finally, we provide a first experimental test of our model. The theory organizes the data remarkably well. In particular, all major comparative statics are as predicted. However, we also observe an interesting anomaly: a pronounced reluctance to go first in an environment with a fifty-fifty prior. The intuition that an empty restaurant may be bad news is perhaps deeply rooted—rendering our theoretical result perhaps non-trivial and inducing subjects' deviation from its prescription. Nevertheless, the vast majority of decisions does follow the equilibrium prediction which we view as encouraging for our theory.

3.7 Appendix

3.7.1 Instructions

Welcome to our experiment!

Please be quiet during the entire experiment. Do not talk to your neighbours and do not try to look at their screens. Simply concentrate on what you have to do. If you have a question, please raise your hand. We will come to you and answer it privately.

You are participating in an economics experiment in which you interact with two other participants for 15 rounds. There are more participants in this room, but you will interact with only two of them.

Depending on your choices, the other two participants' choices and some luck you can earn a considerable amount of money. You will receive the money immediately after the experiment. Notice that all participants have the same instructions.

The experiment

What you have to do

You will have to decide whether you want to invest in a project or not. The project may be good or not and we will give you some useful information about how the chances are. Additionally, you will also know something about what the other two participants decided to do.

What determines whether the investment is good or not

The computer will decide randomly whether in a given round the investment is good or not. The two possibilities are equally likely. This is equivalent to say that the computer will choose whether the investment is good or not by tossing of a coin.

Note that if in a given round the investment is good, it is good for all three participants. Similarly, if it is bad, it is bad for all three of you.

What you earn if you decide to invest or not

In real life, if you choose a good investment the prize is that you enjoy a good return. And if it is bad, the cost is that you have spent money on something that was not profitable. In our experiment, if the project is good and you choose to invest we give you £3 for the smart decision. If, instead, you decide not to invest, then you get nothing. Similarly, if the project is bad and you decide not to invest, we pay you £3 for the smart decision. And if you decide to invest, we pay you nothing.

How you can make your decision

As we said the computer will choose one of the two projects. Of course, we will not tell you which one has been chosen. But we will give you a piece of information.

If the project is good, then the computer will draw a ball from an urn containing 70 green and 30 red balls. If it is bad, it will draw a ball from an urn containing 70 red and 30 green balls. You (and only you) will be told the colour of this ball. Clearly if the ball is green, it is more likely (but not sure) that the project is good. If it red, it is more likely (but not sure) that it is bad.

Note that the computer will draw a different ball for each participant. It will choose a ball for you and then replace it in the urn. Then a ball for another participant and then replace it. And so on. Therefore, it is well possible that you receive a green ball and another participant a red one, and vice versa.

This is not the only information that you will get. In the next paragraph you will discover why.

When you make your decision

You and the other two participants will make the decision to invest or not in the project in sequence. Therefore, you may be the first, or the second, or the third. Your position in the sequence is assigned to you randomly by the computer. Whether you will be first, or second, or third is equally likely. However, we will not tell you your position. But we will tell you the number of people who have decided to invest before you.

Let us briefly look at the different possibilities that can arise. You might see that two others have decided to invest in which case you know that you must be the last to make a decision. If you see that just one other participant has invested before you the situation is less clear. Obviously, you are not the first in the sequence. You might be the second and the first might have decided to invest. But you might also be the last with one of your predecessors having decided to invest and the other having decided to pass the opportunity. Finally, you might observe that none of the others has decided to invest so far. In that case you might be the first in the sequence. But you might also be second and the first passed the opportunity, or you might be the third and both others decided not to invest.

The procedure

At the beginning of each round the computer will decide whether the project is good or bad. Moreover, it will draw a ball for you and one for each other participant. And it will decide the sequence in which the three of you decide.

Then it comes to your decision. But notice instead of telling you your ball colour and the number of people who have decided to invest before you, we will do something different. We will ask you to make your decision for each possible case. We will ask you to make a decision to invest or not depending on the number of people who have already decided to invest and on the colour of your ball. Specifically, you will see a table like this:

	Green Ball	Red Ball
Nobody has invested before you		
One other has invested before you		
Two others have invested before you		

For each possible combination you will have to decide whether you invest or not. Of course, when we compute your payoff, we will take into account only the decision corresponding to the actual situation (that will be revealed to you afterwards). The other five decisions will not be taken into account. Therefore, you can decide for each case as you if knew the ball colour and the actual number of people who decided to enter before you.

Is all this clear? If not, do not worry, here are two examples.

Example 1

Suppose you make the following decisions:

	Green Ball	Red Ball
Nobody has bought before you	NO	NO
One other has bought before you	NO	INVEST
Two others have bought before you	NO	INVEST

At the end of the round the computer tells you that actually in this round you received the red ball. Moreover, that you were third and another participant decided to invest before you. This is equivalent to the case "One other has invested before you". Therefore, we will compute your payoff considering only your decision to INVEST for the case "One other has INVESTED before you". All the other decisions are irrelevant. Hence, if the project is good, we will pay you £3.

Example 2

Suppose you make the following decisions:

	Green Ball	Red Ball
Nobody has bought before you	NO	NO
One other has bought before you	INVEST	INVEST
Two others have bought before you	INVEST	INVEST

At the end of the round the computer tells you that actually in this round you received the green ball and you were first. This is equivalent to the case "Nobody has invested before you". Therefore, we will compute your payoff considering only your decision NOT to INVEST for the case "Nobody has invested before you". All the other decisions are irrelevant.

Procedures for each round

Remember that the experiment is organized into different rounds and that within each round you will have to make six investment decisions. So, now it is time to summarize what happens within each round.

1. At the beginning of each round the computer randomly chooses whether the project is good or bad. The project is the same for all participants. But you will not be told which project has been chosen.

2. The computer draws a ball from an urn for each participant. The ball is drawn and then replaced, so that the total number of balls in the urn is always the same before the computer makes another draw. The computer draws a ball for participant one and puts it back into the urn. Then it does the same for participant 2. And then for participant

3. If the project is good, the computer draws the ball from an urn containing 70 green and 30 red balls. If it is bad, from an urn containing 70 red and 30 green balls. You will be the only one knowing your ball colour.

4. The computer decides randomly which participant is first, who is second and who is third.

5. You will make your decisions for the six cases illustrated above. Of course, given that you do not know in which case you actually are, you will have to think of the best solution for each of the six cases in which you may be.

Once the round is over, you will be informed of your ball colour. We will also tell how many other participants decided to invest before you. And of course we will tell you whether the project was indeed good or bad and how much you earned.

Then, we will repeat the same procedure for the second round at the beginning of which the computer will choose again a project and so on. Note that at the beginning of each round the computer chooses the project always with an equal chance of being good or bad, independently of what was chosen in previous rounds. We will repeat the same procedures for altogether 15 rounds.

Final payment

For the simple fact that you showed up in time for the experiment you earn £4. The rest of the payment depends on how you perform. The computer will randomly choose one round out of the first 5 rounds, one among the 6th through the 10th and one among the 11th though the 15th. Your payment will depend on how you performed in the selected rounds. We will sum up your payoffs in these three rounds. Your final payment will be equal to this amount plus the £4 for showing up.

Chapter 4

Learning Trust

4.1 Introduction

Reputation building in repeated trust games requires that trustors have some information about trustees' behavior in the past. Consider a buyer-seller framework where sequential exchange induces a moral hazard problem. First, buyers make a decision about whether or not to send some money to a seller who has advertised a good. After having received the money, the seller then decides whether or not to deliver the promised good. In such a market a seller can build up a reputation for being honest if and only if buyers can at least partially observe the seller's trading history.

Thus, providing buyers with information about sellers' past should help to alleviate the moral hazard problem. We shall call such feedback provision to buyers *one-sided market transparency.* The first result that we establish in this paper is that it indeed helps to improve efficiency in laboratory markets that suffer from moral hazard. However, our key finding is that *two-sided market transparency* where both, buyers *and* sellers, have access to sellers' trading history improves market performance even further. From the vantage point of orthodox theory, this is a surprising result. Whether or not sellers can observe other sellers' past should be irrelevant. But as we conclusively show, it is not.

The key to understanding this result is very simple. There are some sellers who, when left to their own devices, simply do not understand the mechanics of reputation building. In markets with one-sided transparency only, they make use of any opportunity to rip off their customers despite the drastic consequences this implies for their reputation. Typically, it does not take long until such sellers establish a firm reputation as cheats and lose all business. This is different in markets with two-sided transparency because here sellers *can learn from other sellers.* In particular, sellers who initially do not understand the incentives for reputation building can now observe others who do. And they can see that those who do, get more business and are soon much better off than they are. Given a second chance, they can now imitate successful reputation building. This process of social learning gave this paper its title.

While the benefits of one-sided market transparency have already been documented in the literature,[1] the interaction of social learning and reputation incentives that makes two-sided transparency superior in our experiment has, to the best of our knowledge, not been demonstrated before.[2] Of course, the often cited example of ebay's feedback mechanism is one that implements two-sided transparency. On ebay, everybody, buyers and sellers, have access to information about sellers' history. However, previous studies have—probably guided by orthodox reasoning—ignored the role of providing sellers with information about each other. Our results suggest that, in fact, two-sided transparency is an important ingredient for the design of well-functioning markets that are prone to moral hazard.

4.2 Experimental design and procedures

In our experiments subjects play the binary-choice trust game shown in Figure 4-1. Payoffs are in pence and strategies and player roles are labelled exactly as in the experiment. As-

[1] See, for example, Keser (2002), Bolton, Katok, and Ockenfels (2004) or Bohnet and Huck (2004).

[2] Studies that show how subjects can learn from other subjects to improve their decision making in other contexts include Offerman and Sonnemans (1998) and Slembeck and Tyran (2004).

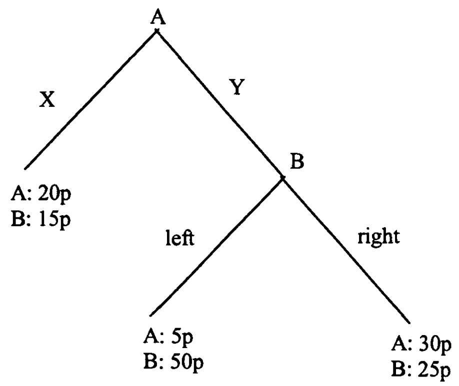


Figure 4-1: The trust game.

suming that players maximise some monotone function of their monetary payoff and that this is common knowledge the game has a unique Nash equilibrium, in which the first mover chooses "X", i.e., not to trust, and the second mover chooses "right", i.e., not to honour trust if being trusted. In the following we will refer to the first mover as the buyer and to the second mover as the seller.

Payoffs are deliberately chosen to be asymmetric in order to make the moral hazard problem as difficult as possible.[3] Subjects play this game in all treatments for 30 periods. Keeping their roles they are randomly rematched at the start of each period. Each matching group consists of four sellers and four buyers.

The treatments differ in what subjects know about the past. In the baseline treatment, NoInfo, subjects have no information about the past. Whenever they are rematched, they are simply told "You have been rematched with a new participant" without knowing anything about this particpant's identity or history. In all other treatments sellers can be identified with labels (B1, B2, B3, and B4). In treatment Reputation, all buyers know all sellers' past. In treatment Imitation all sellers know each other's past. And, finally,

[3]With symmetric payoffs after honored trust (Y, right) subjects find it much easier to achieve efficiency already in one-shot games, see, for example, Bacharach, Guerra, and Zizzo (2001) or Bolton, Katok, and Ockenfels (2004).

		Sellers know sellers' history	
		No	Yes
Buyers know	No	NOINFO	IMITATION
sellers' history	Yes	REPUTATION	TWO-SIDED

Table 4.1: The 2x2 design.

in treatment TWO-SIDED both, sellers and buyers, can observe all sellers' past.[4] This 2x2 design is summarized in Table 4.1.

The experiments were computerised[5] and sellers' history was made available using a simple graphical tool. In the left part of the screen subjects could see four columns consisting of 30 hash signs, each column representing one seller. Each row represented one period. Initially, all hash signs were white. Then, after each period, hash signs in the row representing this period changed their colour. They turned black if the seller had not to make a decision because his buyer did not trust him. They turned green if the seller honoured the buyer's trust. And they turned red if the seller exploited the buyer's trust. This colour coding is, of course, obvious and makes it rather easy to read the comparatively complex history information.

The experiments were conducted at the University of London. For each of the four treatments we conducted six separate sessions, each with eight subjects who had been recruited via emails to the college's entire student body. Altogether, 192 subjects participated in the experiments which lasted on average less than an hour. Average earnings were £11.07 (including a £5 show-up fee).

[4] In each treatment, the information structure is publicly known. For example, in REPUTATION, both, buyers and sellers know that buyers can oberseve sellers' past while sellers cannot.

[5] We used Fischbacher's (1999) z-tree.

	NoInfo	Reputation	Imitation	Two-Sided
honour rate	0.19	0.44	0.19	0.62
	(0.11)	(0.26)	(0.16)	(0.23)
Trust rate	0.21	0.31	0.14	0.43
	(0.15)	(0.17)	(0.11)	(0.15)
Efficiency rate	0.05	0.17	0.04	0.29
	(0.05)	(0.14)	(0.04)	(0.18)

Table 4.2: Average honor, trust, and efficiency rates in all four treatments. (Standard deviations in parentheses.)

4.3 Results

4.3.1 A static view

Table 4.2 shows, for each treatment, average *honour rates*, i.e., the average frequency with which sellers honour buyers' trust, average *trust rates*, i.e., the average frequency with which buyers trust sellers, and, finally, average *efficiency rates*, i.e., the average frequency with which subjects play (Y, right) and reach the individually rational efficient outcome (which, from here on, we shall simply call the "efficient outcome" or refer to it as "efficient trade").[6] Figures 4-2, 4-3, and 4-4 show the same information graphically and, in addition, honour, trust, and efficiency rates for, both, the best and the worst session in each treatment. A few observations are in order.

1. In treatments without incentives for reputation building (NoInfo and Imitation) honour rates are very low (below 20%) and so are trust rates. Consequently, there is hardly any efficient trade.

2. Introducing incentives for reputation building in treatment Reputation more than doubles the average honour rate which also boosts the trust rate. As a result the number of efficient trades is more than tripled. However, there is considerable variance between sessions and the overall outcome is far from perfect.

[6] Notice that the sum of payoffs is equal in both end nodes that can be reached after the first mover decision to trust. Again, this is a feature that makes it harder for subjects to cooperate. See also footnote 3.

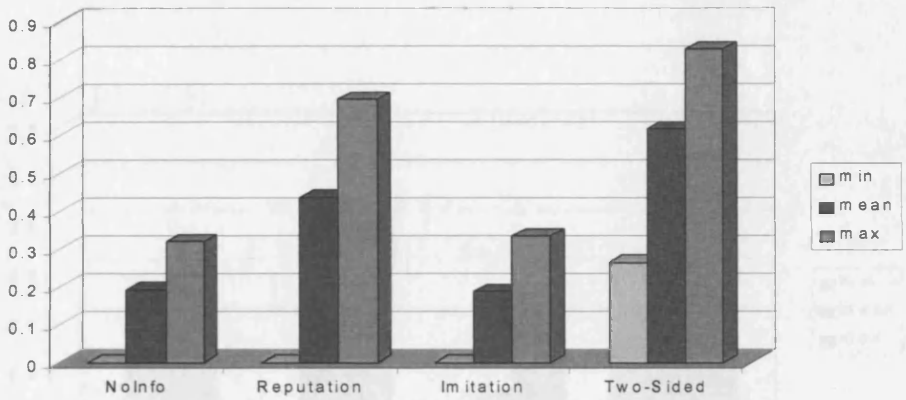


Figure 4-2: Honour rates in all treatments.

3. The reputation effects are considerably enhanced in treatment TWO-SIDED where both, buyers and sellers, have access to sellers' trading history. Again, there are considerable differences between sessions.

Conducting statistical tests[7] reveals that the effects of introducing incentives for reputation building are highly significant.[8] However, despite the consistently higher averages in treatment TWO-SIDED, tests fail to show any significant benefits of two-sided market transparency. Comparing treatments REPUTATION and TWO-SIDED, we find *no* significant differences—neither for honour rates, nor for trust and efficiency rates. Does this mean that there are indeed no effects of added market transparency and that, as orthodox theory predicts, it only matters whether or not buyers can observe sellers? In the next subsection we shall argue that this conclusion would be premature.

[7] We always take one session as one independent observation and then perform pairwise MWU-tests with six against six observations.

[8] Comparing NOINFO with REPUTATION we find that, both, honour and efficiency rates are significantly higher in the latter (one-sided $p = 0.023$ and $p = 0.074$, respectively). Comparing IMITATION and TWO-SIDED reveals that all three rates are higher when buyers know sellers' history ($p = 0.005$ for trust, $p = 0.008$ for honour, and $p = 0.004$ for efficiency).

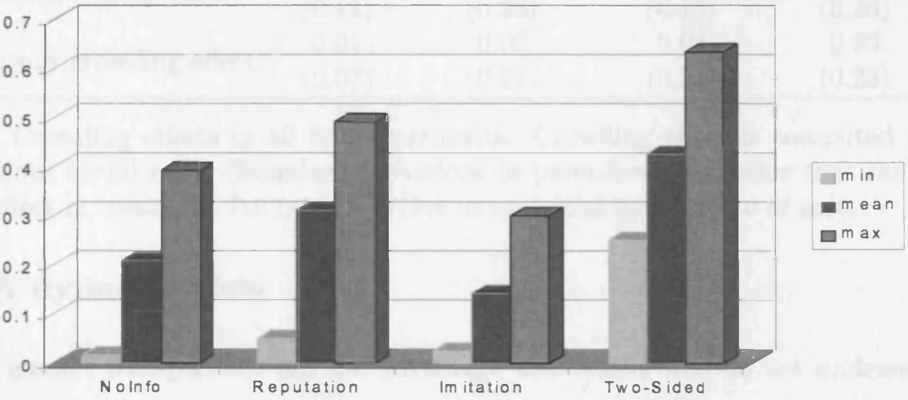


Figure 4-3: Trust rates in all treatments.

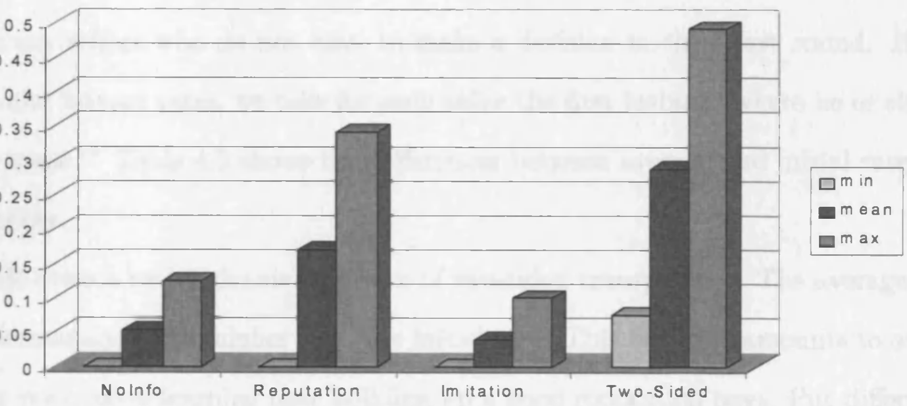


Figure 4-4: Efficiency rates in all treatments.

	NoInfo	Reputation	Imitation	Two-Sided
Honour crowding effect	0.11 (0.17)	0.05 (0.31)	0.19 (0.16)	0.33 (0.49)
Trust crowding effect	−0.04 (0.11)	−.0.06 (0.28)	−0.03 (0.15)	0.22 (0.26)
Efficiency crowding effect	0.01 (0.07)	0.00 (0.21)	0.04 (0.04)	0.20 (0.23)

Table 4.3: Crowding effects in all four treatments. Crowding effect is computed as average rate minus initial rate. (Standard deviations in parentheses.) Notice that the honour crowding effect in treatment Imitation is due to an initial honour rate of zero.

4.3.2 A dynamic view

Two-sided market transparency has the advantage that sellers who do not understand the mechanics of reputation building can *learn* from other sellers who do. If such learning is important one would predict that two-sided transparency *crowds in* honour, trust, and efficiency over time.[9] To test for such dynamic effects we will, therefore, analyse a very simple measure capturing the market dynamics. For each session we shall compute the difference between the average honour (trust/efficiency) rate over time and the *initial* honour (trust/efficiency) rate. In the case of trust and efficiency, these initial rates are simply computed for the first round. This approach does not work for initial honour rates since there are many sellers who do not have to make a decision in their first round. Hence to compute initial honour rates, we take for each seller the first instance where he or she had a decision to make.[10] Table 4.3 shows the differences between average and initial rates for all four treatments.

The table reveals rather dramatic effects of two-sided transparency. The average honour rate is 33 percentage points higher than the initial rate. This basically amounts to one third of the seller population learning that building up a good reputation pays. Put differently, it

[9]Bohnet, Frey, and Huck (2001) provide a theoretical model for crowding in of trustworthiness.

[10]Notice that these measures, while somewhat crude, are extremely clean. In particular, more sophisticated measures for the initial propensities to trust or honour would unavoidably be confounded with learning or other dynamic effects. Also, by taking simply one ratio we avoid making any assumptions about the functional form of the dynamics.

amounts to every third seller turning from a cheat into a reliable trading partner.

In treatment REPUTATION there is also a slight increase in honour rates but, the effect is both, smaller and without consequences for overall market performance. Under two-sided market transparency increasing honour rates translate into increasing trust and, hence, increasing efficiency rates. In treatment TWO-SIDED overall efficiency is 20 percentage points higher than initial efficiency while in all other treatments efficiency does virtually not change over time.

Not surprisingly, these dynamic effects of two-sided market transparency are not only strong in size but also highly significant. A comparison of treatments REPUTATION and TWO-SIDED tests for the *additional* benefit of sellers observing other sellers in the presence of incentives for reputation building. Pairwise tests reveal that all three crowding effects are significantly higher in TWO-SIDED than in REPUTATION.[11] Thus, we see that two-sided market transparency has indeed an important beneficial effect for market performance that could not have been predicted by orthodox theory.

4.4 Conclusion

We examine the effects of different forms of feedback information on the performance of markets that suffer from moral hazard problems due to sequential exchange. We find that, as orthodox theory predicts, providing buyers with information about sellers' trading history boosts market performance. With such one-sided market transparency sellers have an incentive to build up a reputation as reliable trading partners and many sellers use this opportunity, which helps to alleviate moral hazard. This beneficial effect of incentives for reputation building is considerably enhanced if sellers, too, can observe other sellers' trading history. Apparently some sellers do not understand the mechanics of reputation building on

[11] The p-values are $p = 0.055$ for the crowding in of trust and honour and $p = 0.075$ for the crowding in of efficiency (one-sided MWU-tests).

their own. However, with two-sided market transparency these sellers can learn from those who manage to build a good reputation. Thus, there is a systematic learning process turning sellers who initially cheat into reliable trading partners. This dramatically increases market performance over time.

This result adds to the existing literature on feedback information, some of which has been motivated by the success of ebay's celebrated feedback mechanism. It suggests that ebay benefits from having sellers' feedback rating freely available to *both* market sides. While this was perhaps a very natural design choice for ebay where people act as both, buyers and sellers, this might be less obvious for specialised trading and procurement platforms where the two market sides are more separated. Here two-sided transparency might be a less obvious but—as our findings suggest—very recommendable choice.

Chapter 5

Supermarket entry and planning regulation

5.1 Introduction

The UK retail sector has been the focus of policy concern because of its relatively low productivity growth, high prices and concentrated market structure. The wholesale and retail sector account for around one-fifth of the UK's productivity gap, with the US and supermarkets accounting for the single biggest component of this sector.[1] Additionally, the supermarket industry has been repeatedly investigated by the UK competition authorities. The literature has emphasised the importance of entry to productivity growth in US retail, perhaps enabling the more rapid introduction of ICT in newer stores or reallocation from less productive to more productive stores. In contrast, in the UK growth has come mainly from incumbents.[2] One potential reason for this productivity gap that is emphasised by the Competition Commission 2002 report is that most stores in the UK operate below efficient scale.

[1] See, for example, Griffith, Harrison, Haskel and Sako (2003), Figure 3.
[2] For the US see Foster, Haltiwanger and Krizan (2002) and for the UK see Haskel and Khawaja (2003).

Attention has focused on planning regulation as one of the root causes. An influential report by McKinsey (1998) highlighted planning regulations as one of the major policy issues affecting productivity in this sector. And in his recent Pre-Budget Report, on 5 December 2005, the Chancellor and Deputy Prime Minister announced a review of the land use planning system in England. It will consider how land use regulation affects economic growth.

Yet there is relatively little work studying the overall impact or cost of planning regulation. A notable exception is a paper by Bertrand and Kramarz (2002) who show that planning regulation in France created an important barrier to entry for large supermarkets. Evaluation by the Office of the Deputy Prime Minister (ODPM) have suggested that planning regulation has affected entry by new large stores.[3] But this work does not control for other factors that may affect store entry decisions. For example, a report by CB Hillier Parker to the ODPM notes that *"... food retailers are changing their store formats and focusing more on town centre and edge of centre sites, this has been as much due to commercial considerations as to [planning regulations] ..."*

Our interest in this paper is to quantify the impact that planning regulation has had on market structure in the UK supermarket industry, and thus the cost of this regulation. This is of course only part of the story, as these costs then need to be set against any potential benefits. Before 1993 planning regulation in England was decentralised. Over the late 70s and 80s there was rapid growth in large out-of-town store formats. Driven by a desire to stem this trend, which government feared was harming social cohesion in communities, reforms in 1993 and more stringently in 1996 centralised planning regulation and changed it to encouraged in-town (small) store formats and discourage out-of-town (large) store formats.

What impact would we expect planning regulation to have on equilibrium market structure?

We would expect it to restrict the number of large stores and increase the number of small

[3]See, for example, ODPM (2005), para 1.5, *".. emerging evidence suggests that since the mid-1990s national planning policy has had a significant impact in terms of increasing the proportion of retail development locating in town centres..."*

stores. Unsurprisingly, after the reforms we see an increase in town centre stores and an increase in smaller store formats. However, as highlighted by the quote above, other factors may have also affected this shift towards smaller store formats. In particular, recent strategy diversification by several firms in the supermarket industry, for example Tesco moving into the Express and Metro format and Sainsbury moving into the Local format, suggests that at least some of the shift from large to small store format over the late 1990s may be due to changing consumer preferences and store strategy. So incorporating a model of consumer demand is crucial in isolating the effect of planning regulation on market structure and firm profits.

We take the model of Bresnahan and Reiss (1991) as a starting point and estimate the impact of planning regulation on market equilibrium. We use regional variation in planning applications granted and treat these as exogenous. We assume that firms make decisions about entry into the large store format independently of the number of small stores whereas the decision to enter into the small store format takes the number of large stores as given.[4]

The structure of the paper is as follows: Section 2 describes the planning and land use regulations as they apply to grocery stores in England. Section 3 describes a model of supermarket entry and equilibrium market structure and our estimation strategy. Section 4 discusses the data. Our results are shown in Section 5 and a final section concludes.

5.2 Planning regulation

Land use regulation is often cited as an important barrier to entry[5] and in the UK it has been held largely responsible for higher prices and lower labour productivity in retail stores, and particularly in supermarkets.[6] However, land use regulation is rarely a barrier to all forms

[4] Some of the main papers are Bresnahan and Reiss (1991), Berry (1992), Mazzeo (2002).

[5] See, inter alia, Djankov et al (2002), Bertrand and Kramatz (2002).

[6] Bertrand and Kramatz (2002) show that such regulation in France has had a negative impact on employment. See also McKinsey (1996), Haskel and Khawaja (2003), Competition Commission (2000).

of entry, but rather imposes a constraint on certain types of entry. A good example of this is English planning regulation. Rapid growth in the number of out-of-town supermarkets in the late 1980s and early 1990s led to concerns about out-of-centre development and the impact this was having on the vibrancy of city centres, and the possible impact this had on, for example, social exclusion. Changes to regulation sought to encourage, for example, retail development in town centres.

Land use regulation in England changed in 1993 and more importantly in 1996 with the introduction of Planning Policy Guideline 6 (PPG6). It changed in a way that favoured in town developments over out of town. PPG6 put forward the sequential approach which states that out-of-town centres should only be build as a last resort, if there is not viable city centre alternative. PPG6 stated that LPAs should take a 'plan-led' approach to development control. LPAs are required to have a development plan. This should set out the authority's policies for the development and use of land in its area. LPAs should identify which centres should grow and identify sites for specific types of development where a need for additional provision has been identified. City, town and district centres should be the preferred locations for developments that attract many trips. This type of planning regulation therefore favours the opening of certain types of stores (smaller city centre stores) over other types of stores (larger out-of-town stores). What impact will such regulation have on profitability and market structure?

On the one hand, smaller stores may be less efficient (more costly) to run.[7] On the other hand, there is a trade-off between higher fixed costs for larger out-of-town stores versus lower variable costs to run these stores. The over all impact will depend on: the strategic response

[7] The Competition Commission (2000) report into supermarkets surveys the literature on economies of scale and argues that there are economies of scale up to 30,000 sq. m in supermarkets, but not after that. The economies of scale were especially in staff costs, which, the CC shows, are the bulk of value added in retailing. The CC present data on international comparisons of supermarket shop size. Britain has significantly fewer supermarkets and they are much smaller than the US, but they are larger than Continental Europe. In addition, US supermarkets are, on average just over 3,000 sq metres, which is just above the minimum scale required to achieve the highest levels of labour productivity. UK supermarkets are below that, suggesting that this might account for part of our productivity disadvantage.

of firms that would have opened large stores, do they now open more small stores; the entry response of other firms, are other barriers to entry lower for small stores, did the presence of large out-of-town stores raise the entry barriers to small stores; the relative cost efficiency of small and large stores.[8]

5.3 A model of supermarket entry and equilibrium

We take Bresnahan and Reiss (1991), henceforth B&R, as our starting point and consider N symmetric firms entering into a Cournot competition. We allow for two type of store - in town and out of town. As emphasised above, this is an important characteristics of the UK supermarket industry where there are two different types of stores - small city centre stores like Tesco Metro and Sainsbury Local and large often out of town supermarkets which were differentially affected by planning regulation.

We assume that demand for groceries takes the following form: consumers demand the bulk of their groceries in a one-stop shopping trip in big stores; they subsequently top-up with additional items that were forgotten or unexpectedly needed in small stores. Adding two product types to the simultaneous-move entry game of B&R introduces multiple equilibria (some of which might be mixed strategy equilibria).[9] In order to get a unique pure-strategy equilibrium we need some additional assumptions. Mazzeo (1992) and Toivanen and Waterson (1999) model entry with more than one type of entrant and introduce two additional assumptions for uniqueness. First, they introduces Stackelberg competition, with an exogenously defined sequence in which firms make irrevocable choices whether or not to enter, and of which type. This seems an implausible assumption for the supermarket industry where

[8] To the extent that retailing productivity growth is due to firms closing older, low productivity stores, and opening newer, high productivity shops to replace them, this might result in lower productivity growth, which may feed through into higher prices. Recent work has suggested that the adoption and use of ICT has been an important contributor to the US productivity acceleration of the late 1990s.

It is likely that ICT usage is higher and more effective in newer shops, but it is not clear whether a market with fewer large out-of-town stores or more small in-town-stores will lead to more entry and exit.

It may also be that ICT is easier to adopt in larger out-of-town stores.

[9] See Tamer (2002) for empirical strategies of estimating multiple equilibria in discrete games.

there is no natural ordering. The second assumption is that the decrease in profits is larger if the competitor is of the same type, which seems natural.

In this paper we introduce differential fixed costs across store type to reflect the impact of planning regulation. This is easier and more natural to incorporate into a simultaneous move game.[10] More formally, we consider firms to play a three-stage game. In the initial stage, firms decide whether to enter by opening a big store. Once these entry decision have been made, small firms compete for the residual demand. Profits of large stores therefore do not depend on the presence of small stores, but small stores take the number of big stores into account when deciding whether to enter or not. Once firms have made their entry decisions symmetric price competition between each type ensues and payoffs are determined. We assume a free entry equilibrium - firms enter up to the point where profits cover fixed costs. The profits of firms that do not enter are normalised to zero.

In the Nash equilibrium the following inequalities hold (B is the total number of big stores, B_{-i}, S_{-i} is the number of big and small stores of others and C_B, C_S are the fixed costs of a big and a small store):

(1) entry of *Big* if $\widetilde{\Pi}_{iB}(B_{-i}) > C_B$

(2) entry of *Small* if $\widetilde{\Pi}_{iS}(S_{-i}, B) > C_S$

(3) exit from *Big* if $\widetilde{\Pi}_{iB}(B_{-i} - 1) < 0$

(4) exit from *Small* if $\widetilde{\Pi}_{iS}(S_{-i} - 1, B) < 0$

A store enters if profits are sufficient to cover fixed costs, and they exit the market if they make negative net profits. To prove the existence of a pure strategy equilibrium we need to impose some structure on the profit function. One natural restriction is that firm profits decline in rivals' entry so we assume that profits decline in number of big firms in the first and second stage and profits decline in number of small firms on the second stage.

[10] In a sequential game like Mazzeo's (1992) varying fixed costs would effect the entry sequence as well as the product choice, so he focusses on varying variable costs by type.

5.3.1 Market demand and supply

Market demand is characterised in the following way. Consumers demand a composite good "groceries" and thus aggregate demand can be modelled in the form

$$Q = d(X, P)S(Y) \tag{5.1}$$

where Q is total quantity of groceries demanded, $d(X, P)$ is the demand of a representative consumer, which is a function of demographic variables X and the price P of the groceries. $S(Y)$ denotes the number of consumers. So the demand function is homogeneous of degree one. We also assume that demand increases linearly in S.[11]

We can take two different approaches to introducing different store formats, one through the demand function, the other through the cost function.

The dominant model of consumer shopping behaviour is the one-stop shopping model.[12] Different store formats arise because of consumer preferences and shopping habits. Consumers demand groceries. They buy a large portion of their groceries in one shopping trip and then top-up forgotten items of perishables in small trips. For the one-stop shopping trip they prefer a large variety of different goods and therefore prefer a large store format. For the top-up shopping, on the other hand, a small convenient store that is in close proximity is preferred. In this case we can model aggregate quantity demanded from one-stop shopping as Q_1 and top-up shopping by Q_2 as

$$Q = Q_1 + Q_2 = d_1(X, P)S(Y) + d_2(X, P, Q_1)S(Y) \tag{5.2}$$

where $d_1(X, P)$ is the demand of a representative consumer from one-stop shopping and $d_2(X, P, Q_1)$ the demand from top-up shopping. Both are a function of demographic variables X and the price P of the groceries. $S(Y)$ denotes the number of consumers. This demand

[11] This assumption is crucial since non-linear demand in S might lead to multiple equilibria.

[12] See, inter alia, Smith (2004), CC (2000).

function is homogeneous of degree one. We also assume that demand increases linearly in S. Demand from one-stop shopping is unaffected by the demand from top-up shopping, the cross derivatives $\frac{\delta Q_1}{\delta d_2}$is equal to zero, $\frac{\delta Q_2}{\delta d_1}$ can be different from zero since it depends on the residual demand after one-stop shopping of Q_1. This assumption of independent demand for one-stop shopping is fairly strong, but it allows us to estimate the resulting profit functions for big format stores independently from small format stores, to investigate the differential impact of planning regulation, using an ordered probit model. In estimating the resulting profit functions for small format stores we include the number of big stores in line with our residual demand hypothesis.

An alternative approach to introducing different store types is from the supply side. On the supply side firms profits equal revenue minus costs. Firms incur fixed costs of $F(R)$ and have per capita variable costs of $V(W)$ where W are variable cost shifters and R are exogenous variables that affect fixed costs. As mentioned above we assume a free entry equilibrium - firms enter up to the point where profits cover fixed costs. Fixed costs can vary with the number of firm in the market, so the fixed costs for a duopolist can be different than for a triopolist. Thus firms have U-shaped average total costs, declining initially due to fixed costs and rising later due to marginal costs. Consider two supply curves. The first is for firms operating at or above minimum efficient scale (30,000 sq ft)[13], these have high fixed costs and low variable costs, but variable costs rising at some point. The second supply curve is for small firms, with low fixed costs but high variables costs. Where demand is sufficient to cover the fixed costs of a large firm, then they will enter. If excess demand is smaller than the amount needed to cover fixed cost, then a small store will enter.

[13]Definition of the Competition Commission (2000).

5.3.2 Firm profits

Assuming profits to be additively separable in observed and unobserved components we have the following profit function, where subscript n denotes the number of firms (1 monopoly, 2 duopoly etc.), j denotes the area and f denotes the store format (big or small).

$$\Pi_{njf} = S_f(Y_j)V_f(W_j, \alpha, \beta) + F_f(R_j, \gamma_n, \theta) + \varepsilon_{njf} \tag{5.3}$$

where α, β, γ and θ are parameters to be estimated, Y_j are factors that vary across region and affecting market size, W_j are factors that vary across region and affect per capita profits (demand and variable costs) and R_j are factors that shift fixed costs (e.g. regulatory constraints) and the unobserved error ε captures idiosyncratic variation in profits. We assume that ε follows a normal distribution and is independent across markets and from observables, with zero mean and constant variance.

We model firms per capita variable profits as a linear function of the observables W ,

$$V_f(W, \alpha, \beta) = \alpha + \beta_f W_j \tag{5.4}$$

Fixed costs depend on exogenous variables R, including planning regulation and land prices, and a market structure level unobservable, γ_n

$$F_f(R_j, \gamma_n) = \gamma_1 + \theta_f R_j + \sum_{n=2}^{N} \gamma_n. \tag{5.5}$$

The γ_n terms allow for later entrants to have higher fixed costs. If we observe $\gamma_n > 0$ we conclude that later entrants have higher fixed costs.

5.3.3 Empirical model

Under the assumption that the error is i.i.d. in profit specification (5.3), we can write the probabilities of observing a particular market structure for the baseline case of no store type differentiation in the following way. Let $\overline{\Pi}_n$ denote the estimated per firm profits - so the monopolist's profit is $\Pi_1 = \overline{\Pi}_1 + \varepsilon$. We assume that profits can be ordered, so that $\overline{\Pi}_1 > \overline{\Pi}_2 > \overline{\Pi}_3 > .. > \overline{\Pi}_N$, i.e. a monopolist's profit is higher than a duopolist's and higher than a triopolist's. We assume that there is no heterogeneity between firms, we can write any market structure profits in terms of monopoly profits[14]

$$\overline{\Pi}_1 = \overline{\Pi}_n + \mu_{n-1} \tag{5.6}$$

where μ_n is the constant difference between n and $n-1$ profits and $\mu_n > \mu_{n-1}$. Thus the probabilities of observing a particular market structure with Φ the cumulative normal density function can be written as

- Empty market

$$\Pr(y=0) = \Pr(\overline{\Pi}_1 + \varepsilon < 0) = \Pr(\varepsilon < -\overline{\Pi}_1) = 1 - \Phi(\overline{\Pi}_1) \tag{5.7}$$

- Monopoly

$$\begin{aligned} \Pr(y &= 1) = \Pr(\overline{\Pi}_1 + \varepsilon > 0 \text{ and } \overline{\Pi}_2 + \varepsilon < 0) = \Pr(-\overline{\Pi}_1 < \varepsilon < -\overline{\Pi}_2) && (5.8) \\ &= \Phi(\overline{\Pi}_1) - \Phi(\overline{\Pi}_2) = \Phi(\mu_1 - \overline{\Pi}_1) - \Phi(-\overline{\Pi}_1) && (5.9) \end{aligned}$$

[14] In the case of no firm heterogeneity this is an unproblematic assumption, but with firm heterogeneity it would be more crucial.

- Duopoly

$$\begin{aligned}
\Pr(y &= 2) = \Pr(\overline{\Pi}_2 + \varepsilon > 0 \text{ and } \overline{\Pi}_3 + \varepsilon < 0) = \Pr(-\overline{\Pi}_2 < \varepsilon < -\overline{\Pi}_3) && (5.10)\\
&= \Phi(\overline{\Pi}_2) - \Phi(\overline{\Pi}_3) = \Phi(\mu_2 - \overline{\Pi}_1) - \Phi(\mu_1 - \overline{\Pi}_1) && (5.11)
\end{aligned}$$

- Triopoly

$$\begin{aligned}
\Pr(y &= 3) = \Pr(\overline{\Pi}_3 + \varepsilon > 0) = \Pr(\varepsilon > \overline{\Pi}_3) && (5.12)\\
&= \Phi(\overline{\Pi}_3) = 1 - \Phi(\mu_2 - \overline{\Pi}_1) && (5.13)
\end{aligned}$$

- etc.

The probabilities sum to one.

Under these assumptions we can now estimate the equilibrium market structure using a standard ordered probit model. The estimated coefficients can be directly translated as structural coefficients of the underlying monopoly profit function, and with the estimated constants μ_n we can also recover the profit function coefficients.

The choice probabilities allow identification only up to an arbitrary normalisation. If we make assumptions about the constant coefficients - either μ_0 or β_0 is 0 , we assume $\beta_0 = 0$ -.and the variance of the error term equals unity (as in Bresnahan and Reiss). Using equations (5.3), (5.4) and (5.5) we can rewrite profits as,

$$\begin{aligned}
\Pi_{1j} &= Y_j(\alpha + \beta W_j) - (\gamma_1 + \theta R_j) + \varepsilon_j, && (5.14)\\
\Pi_{2j} &= Y_j(\alpha + \beta W_j) - (\gamma_1 + \theta R_j + \gamma_2) + \varepsilon_j, &&\\
\Pi_{3j} &= Y_j(\alpha + \beta W_j) - (\gamma_1 + \theta R_j + \gamma_2 + \gamma_3) + \varepsilon_j &&\\
&\quad \textit{etc.} && (5.15)
\end{aligned}$$

If we define dummy variables $I2_j = 1$ if there are two or more stores in j, $I3_j = 1$ if there are three of more stores in j, etc., we can collapse these into one expression,

$$\Pi_{nj} = \alpha Y_j + \beta Y_j W_j. - \gamma_1 - \gamma_2 I2_j - \gamma_3 I3_j - ... - \theta R_j + \varepsilon_j. \tag{5.16}$$

The $\gamma' s$ can be interpreted as the thresholds between monopoly and duopoly profits and triopoly and monopoly profits, respectively. For the β and θ we can only directly interpret the sign: a positive sign shifts to higher categories when x increases, a negative sign shifts to lower categories when x increases. In order to interpret the magnitudes we calculate marginal effects for the different β and θ.

5.4 Data

In order to estimate the model described above we need data on: the location and characteristics of existing and new store; planning applications and decisions taken by LEAs; regional demographics. These data are available from several different sources.

5.4.1 Supermarket entry and location

Opening dates and size of each grocery store in England comes from the Institute of Grocery Distribution (IGD) data. The data comprises 12,503 observations of grocery stores in the UK. It covers all large chain stores and Co-ops and around 80% of independent stores. The data record the company name, the fascia, the postcode, the opening and closing dates.

Figure 5.1 shows the number of new store openings over time. The grey shaded firms represent the larger store formats and we can see that at the beginning of the 90s up to 1996 the majority of store openings took place in this large format. After 1996 the number of smaller store format (coded in green colours) openings increased substantially and the number of big store openings decreased at the same time.

The UK supermarket industry is dominated by four main companies whose market shares in terms of grocery sales add to around 75% (these are Asda, Sainsbury, Safeway/Morrisons and Tesco). Eventhough our empirical analysis takes all stores into account Figures 5.2-5.6 use the big four to illustrate the main trend. In Figure 5.2 we see that since 1997 the number of new out-of-town stores opening has declined, while the number of new town-centre stores opening has increased. The CC (2000) report suggests that the minimum efficient scale for supermarkets is around 30,000 sq ft. Figure 5.2 shows the number of new store openings above and below this size, with the vertical bar indicating the introduction of new planning regulation. There has been a marked increase in the number of smaller stores opening, suggesting that this could be a reason for the lower efficiency growth in the UK.

Figure 5.3 shows that much of the new store opening was accounted for by Tesco, and Figure 5.4 shows that this was dominated by the Tesco Express format. As highlighted above, an open question is the extent to which this diversification was driven by planning regulation, or a strategic decision by Tesco to differentiate itself and target a different demographic (or a response to changes in consumer preferences towards this type of format). We hope to be able to disentangle these effects in our empirical analysis.

We identify two types of stores, those above and below 30,000 square feet.[15] We consider the number of stores in a postcode sector (this is the first word plus the first character of the second word, e.g. for the postcode WC1E 7AE the postcode sector is WC1E7).A postcode sector has a median diameter of 2.4 kilometer and an average population of 6700 people. For small stores we use demographic variables within the postcode sector. For large stores we use demographic variables for all postcode sectors within a 10 km radius.[16],[17],[18]

[15]This is the minimum efficient scale defined by Competition Commission in its 2000 report on the supermarket industry. As a robustness check we also consider stores above and below 15,000 square feet to distinguish between different consumer purchasing behaviour by the CC

[16]Precisely, we use information from all postcode sectors for which the mid point of that postcode sector is 10 km or less from the centre of the postcode sector under consideration.

[17]The Competition Commission report suggests that the average distance for a one-stop shopping trip is 8 miles.

[18]We are aware that we have overlapping regions which might bias our estimates. We are investigating the direction of a possible bias and also look into isolated areas to check the robustness of our results.

Table 5.1 shows the number of small stores within postcode sectors and the number of large stores within 10k of a postcode sector midpoint. We have truncated the market outcome for small stores at five and for large stores at 10. There are very few postcode sectors that have 6, 7 and 8 small stores. For large stores the maximum value is 33. As a robustness check we also estimate using the non-truncated variable.

5.4.2 Planning applications

Data on the number of planning applications approved comes from the Office of the Deputy Prime Minister (ODPM). We have information on the number of applications and approval for major retail establishments (use class A1 and A3) and change of use. Class A1 contains all establishments where one can buy food to consume it elsewhere and A3 is the necessary permit for establishment with on-site food consumption. These include not just supermarkets but also other large retail sites and restaurants. The data is provided to the ODPM by the Local Planning Authorities. It includes: number of new use applications made; number of new use applications approved; number of change of use applications made; number of change of use applications approved. Figure 5.6 shows the number of decisions granted, decided and the approval rate over time. To measure the extend to which the regulation poses an entry barrier we use the number of decisions granted. The approval rate seems less good a measure since firms may have reacted to the policy and might have applied less for out-of-town sites. This is in line with the decreasing number of application in Figure 5.6.

5.4.3 Other data

A key contribution of this paper is to control for other demographic variables. The demographics we focus on include the population, number of single person households, number of unemployed people, average distance traveled to work, whether the postcode sector is a metropolitan area (and whether it is in London). In the UK these are available from the

Office of National Statistics at the Output Area /Enumeration District level which is roughly half the size of a postcode sector and there are 16,000 in England.[19] For earlier years we do not have information on distance travelled to work, so we use the total number of cars per household to capture mobility and access to a car. We also use the average rateable value of retail property, available from the ODPM, and the distance to the supply centers of the big four (from IGD).

Descriptive statistics of the variables used appear in Table 5.2.

5.5 Results

Table 5.3 shows the estimated coefficients from an ordered probit on the number of small stores in a postcode sector and the number of large stores within 10 kilometers for cross-sections in 1993 and 2002. We estimated our model for every year from 1991 to 2003 and the results are very similar. In general the estimates accord with our prior expectations. There are more stores in more populated areas. The factors that affect variable profits suggest that these are higher for small stores in areas with a large proportion of single person households. This captures the fact that single people might have a more uncertain demand for groceries and therefore do more of their shopping as top-up shopping in small in-town stores. This is confirmed in the estimates for large stores where the proportion of single people effect variable profits negatively. Variable profits are lower for small stores in areas where people have more access to cars (perhaps picking up an effect that these people are more willing to drive to a large store) The effect on for large stores is also positive but even bigger reflecting the fact that cars are even more important when doing a large one-stop shopping trip. Variable profits also depend on whether a store locates within a metropolitan area or in London which both has a negative effect on variable profits (perhaps reflecting higher

[19] For the earlier years we use information coded as Enumeration Districts which leads to a lower number of overall observations.

variable costs in these areas). Again this effect is more pronounced for small stores. The proportion of unemployed people also effects variable profits in both store formats negative this might proxy for lower income or worse access to alternatives.

In terms of fixed costs we see that planning regulations had a statistically significant impact on small stores in 1993, but not in 2002. This accords with the policy changes described above, up until 1996 planning regulations were on the whole locally determined and did not explicitly target small or large store formats. After 1996 there was a national mandate that prejudiced against development of large out of town stores and in favour of smaller town centre stores. The positive coefficient makes sense - more applications granted means lower fixed costs of entry. Here we see a different picture with respect to large store formats. There is no effect of planning regulation in 1993 and then a significant effect in 2002.[20]Property have a significant negative effect on variable profits of small stores but seem to have little effect on fixed costs of large stores possibly due to cheaper land prices in out-of-town locations. Our residual demand hypothesis is also confirmed since the number of big stores does have a negative and significant effect on the profits of small firms in all years.

The $\gamma' s$ represent unobservable market structure differences and we would expect them to increase with the number of firms in a market. They are significant and increase with number of firms the fact that they do increase overproportionally[21] indicates that entrants into markets with more incumbents have higher fixed costs. But we can't distinguish whether this is true because of lower efficiency or higher entry barriers.

While the sign and statistical significance of the coefficients in Table 5.3 make sense, in order to understand the economic significance we need to calculate the marginal effects. These are shown in Table 5.4 for our main variable of interest - planning regulations. We show the marginal effects for small stores for the outcome of zero stores (an empty market)

[20]When we run the regressions for all years this variable is positive and significant in all years for large store formats after 1996 except 2000 when it is insignificant.

[21]overproportional in the sense of: $\frac{\gamma_n}{n} > \gamma_1$.

and 5 stores (the maximum number) for small stores and 10 stores for large stores.[22] For comparison Table 5.5 shows the marginal effects from a model where we only include planning regulations, and don't condition on any demographics. The marginal effects for large stores are statistically significant from 1997 onwards, though small in magnitude. If we compare the effect with the model without demographics the effects look substantially larger, showing that omitting to control for variation in demographics and market variables gives a misleading picture of the impact of planning regulations and increases its impact tenfold.

To understand the magnitude and the distribution across market types of the effect of planning regulation better we additionally undertake the following policy experiment. We increase the amount of planning regulations by 10% i.e. a 1/3 additional application approved on average in every Local Authority and calculate the marginal effects. The distribution of the difference in marginal effects for large stores is shown in Figure 5.7. The negative difference in an empty market shows that increasing planning approvals by 10% leads to a 0.08% decrease in empty markets and an increase in 0.1 % markets with 10 or more large stores. We get more markets with 5 or more stores the decrease in entry barrier leads to an increase in the number of stores. Is this a big or a small effect in economic terms. To gain more insight into the magnitude of the effect we compare it with a 1% increase in population, both are contrasted in Figure 5.8. This comparison suggests that planning approvals play a significant part in restricting entry of large store formats but the magnitude of the effect is rather small.

5.5.1 A simple version of observable firm heterogeneity

One important extension to our model is to include firm heterogeneity. We need some additional assumptions in order to get existence and uniqueness of a pure strategy equilibrium with firm heterogeneity. We follow Berry (1992) in assuming that firms can be ranked in

[22] Interpreting the marginal effects at intermediate rates is tricky as they capture movement from both lower and higher groups. We consider this explicitely in our policy experiment.

terms of their profitability and that ranking does not change when the number or the set of entering firms change.

We do this by imposing a single index of profitability ω_{jk} (j denotes the area and k denotes the firm) and it is easy to show that a Nash equilibrium exists in each market but it is not a unique one. There are several equilibria which entail the same number of firms which we will focus on. To ensure uniqueness of the number of firms in equilibrium we assume that firm differences in profitability only affect the fixed portion of profit and thus the post-entry pricing game remains symmetric. This is obviously restrictive but more general formulations of firm heterogeneity would lead to multiplicity of equilibria. So the resulting profit functions looks like this

$$\Pi_{njk} = S(Y_j)V(W_j, \alpha_n, \beta) + F(R_j, \gamma_n, \theta) + \omega_{jk} + \varepsilon_{nj} \tag{5.17}$$

In the most general version the equilibrium number of firms in a market j has a multinomial distribution with the $(N_j + 1)$ distinct and mutually exclusive outcomes. Thus the region of integration to obtain the probability of market outcomes given the number of combination of firms that lead to e.g. a quadropoly gets increasingly complicated.[23] In the most simple case we assume that there is no unobserved firm heterogeneity and there is an effectively infinite supply of potential entrants. This allows us to include firm heterogeneity into our ordered probit model.[24]

As described above our first approach to include firm heterogeneity is to focus on observable firm differences. One important observable difference between at least the big four supermarkets in the UK is their supply network. We include the distance to the closest supply center of the big four supermarket chains (Asda, Sainsbury, Safeway/Morrison and Tesco) into our model and assuming that there is no further unobserved heterogeneity we

[23]For a more detailed analysis see Berry (1992).

[24]This places restrictions on possible combinations of entering firms and limits the amount to which firm profitability can differ.

can estimate the profit functions with ordered probit. These results are documented in Table 5.7. Since our main focus is on the impact of planning regulation it is reassuring to see that the effect is robust to including heterogeneity in distance to the different supply networks. The coefficients are jointly significant for large stores and thus do capture some of firm heterogeneity. Distance to the closest supply network seems to decrease profits more for Safeway/Morrison whereas it does not have a significant impact on the other three.

5.6 Conclusions and future work

This paper develops a model of supermarket entry and applies it to data for the UK. We are interested in estimating the cost of restrictive planning regulation. We do this by estimating the parameters of the profit function and backing out the increase in fixed costs that is associated with more restrictive planning regulations. Our estimates suggest that planning regulation did have a statistically significant impact on market equilibrium outcomes and that it has represented an entry barrier. However, the economic magnitude of this effect is substantially overestimated by not controlling for variation in demographic variables and a small number of other variable and fixed cost drivers.

In further work we would like to allow for more flexibility in the profit function, for example, by allowing more general heterogeneity between firms and between types and allowing for multi-store firms.

5.7 Tables and Figures

Table 5.1: Number of Stores

Number of stores	Small stores in postcode sector		Large stores within 10km	
	1992	2002	1992	2002
0	4,536	4,446	1,210	1,025
1	2,000	2,102	796	541
2	893	948	765	538
3	436	387	714	549
4	153	156	662	640
5	82	61	477	631
6			498	460
7			410	362
8			476	286
9			330	325
10			1,762	2,743
Total		8,100	8,100	8,100

Source IGD

Table 5.2: Descriptive statistics

	Mean (std. dev.)	**Mean (std. dev.)**
Market size	**1992**	**2002**
Population in postcode sector (in 1000s)	**7.014**	**6.841**
	(4.001)	(3.924)
Population within 10k (in 1000s)	**608**	**649**
	(736)	(805)
Variable profits		
singles (% all households)	**0.127**	**0.171**
	(0.078)	(0.098)
families (% all households)	**0.192**	**0.169**
	(0.055)	(0.059)
unemployed people (% all people)	**0.031**	**0.034**
	(0.019)	(0.018)
distance traveled to work (km)		**10.34**
		(3.38)
number of cars	**0.950**	
	(0.256)	
metro area	**0.244**	**0.244**
	(0.430)	(0.430)
London	**0.137**	**0.137**
	(0.344)	(0.344)
Fixed costs		
rateable value retail land (£/m^2)	**52.65**	**52.65**
	(89.43)	(89.43)
decisions granted	**4.229**	**3.817**
	(4.866)	(4.751)

Source ONS, ODPM, Experian for 2002

Table 5.3: Coefficients from ordered probit model for number of small and large stores

	Small stores		Large stores	
	1992	2002	1992	2002
Market size				
Population in pc/10k	0.394	0.208	0.234	0.018
	(0.022)	(0.017)	(0.004)	(0.005)
Variable profits				
singles in pc/10k	0.072	0.063	-0.029	-0.034
	(0.481)	(0.031)	(0.001)	(0.011)
unemployed people in	-0.218	-0.116	-0.143	-0.047
pc/10k	(0.002)	(0.129)	(0.021)	(0.027)
families in pc/10k	-0.297	-0.547	0.033	-0.030
	(0.058)	(0.045)	(0.016)	(0.015)
distance traveled to work		0.0004		0.0004
in pc/10k		(0.0006)		(0.0001)
number of cars	-0.156		-0.014	
	(0.015)		(0.016)	
metro area	-0.026	-0.011	-0.002	-0.0014
	(0.005)	(0.005)	(0.0004)	(0.0006)
London	-0.044	-0.042	-0.001	-0.002
	(0.007)	(0.006)	(0.0005)	(0.0006)
Fixed costs				
number of big stores	-0.017	-0.009		
	(0.006)	(0.005)		
decisions granted (t-1)	0.004	0.0008	0.022	0.034
	(0.004)	(0.003)	(0.018)	(0.016)
rateable value retail land	-0.115	-0.679	0.051	0.409
(2000)	(0.016)	(0.140)	(0.418)	(0.252)
γ1	0.790	0.815	0.504	0.801
	(0.051)	(0.055)	(0.126)	(0.096)
γ2	1.601	1.652	1.223	1.433
	(0.053)	(0.057)	(0.160)	(0.104)
γ3	2.180	2.273	1.868	1.995
	(0.056)	(0.060)	(0.198)	(0.116)
γ4	2.729	2.789	2.485	2.555
	(0.065)	(0.065)	(0.233)	(0.137)
γ5	3.179	3.313	3.100	3.201
	(0.073)	(0.077)	(0.275)	(0.156)
γ6			3.617	3.838
			(0.307)	(0.193)
γ7			4.202	4.365
			(0.343)	(0.191)
γ8			4.728	4.845
			(0.383)	(0.198)
γ9			5.418	5.274
			(0.415)	(0.199)
γ10			5.942	5.781
			(0.479)	(0.218)
R2	0.058	0.065	0.391	0.423
Observations	8096	8096	8096	8096

Standard errors are reported in brackets and clustered on Local Authority level

Table 5.4: Marginal effects of decisions granted (t-1) on number of small/ large stores, from Table 5.3

	small		large	
	1992	2002	1992	2002
outcome empty market	-0.0013	0.0003	-0.00009	-0.977e-06
	(0.0015)	(0.0015)	(0.0001)	(0.00001)
outcome 5 stores	0.00005	-0.00001		
	(0.00006)	(0.00005)		
outcome 10 stores			0.0005	0.0068
			(0.0005)	(0.0032)

Table 5.5: Marginal effects of decisions granted (t-1) from regression with only this variable

	small		large	
	1992	2002	1992	2002
outcome empty market	0.0018	0.0044	-0.0092	-0.014
	(0.0038)	(0.0033)	(0.0037)	(0.003)
outcome 5 stores	-0.0001	-0.0002		
	(0.0002)	(0.0002)		
outcome 10 stores			0.012	0.028
			(0.084)	(0.007)

Table 5.6: Differences in supply networks (distance)

	Mean (std. dev.)
Distance to nearest ASDA supplier (km)	42.90
	(35.51)
Distance to nearest Sainsbury supplier (km)	38.18
	(37.06)
Distance to nearest Safeway/Morrison supplier (km)	75.95
	(43.54)
Distance to nearest Tesco supplier (km)	42.54
	(40.81)

Table 5.7: Coefficients from ordered probit model with heterogeneity

	2002 small	2002 large
Market size		
Population in pcs /10k	0.219	0.027
	(0.015)	(0.004)
Variable profits		
singles in pcs /10k	0.045	-0.042
	(0.042)	(0.013)
unemployed people in	-0.139	-0.079
pcs /10k	(0.135)	(0.028)
families in pcs/10k	-0.549	-0.037
	(0.046)	(0.015)
metro area	-0.014	-0.002
	(0.005)	(0.0007)
London	-0.042	-0.135
	(0.006)	(0.047)
Fixed costs		
number big stores	-0.015	
	(0.005)	
decisions granted (t-1)	0.0018	0.0345
	(0.0039)	(0.018)
ASDA dist. Supply net.	-0.043	-0.313
	(0.090)	(0.336)
Tesco dist. Supply net.	0.052	0.471
	(0.060)	(0.241)
Sainsbury dist. Supply	-0.002	-0.042
	(0.094)	(0.381)
Safeway/Morrison dist.	-0.025	-0.351
Supply net.	(0.054)	(0.159)
$\gamma 1$	0.809	0.391
	(0.055)	(0.136)
$\gamma 2$	1.645	1.007
	(0.057)	(0.142)
$\gamma 3$	2.266	1.551
	(0.060)	(0.148)
$\gamma 4$	2.782	2.095
	(0.065)	(0.163)
$\gamma 5$	3.306	2.726
	(0.077)	(0.177)
$\gamma 6$		3.351
		(0.209)
$\gamma 7$		3.868
		(0.197)
$\gamma 8$		4.343
		(0.201)
$\gamma 9$		4.771
		(0.206)
$\gamma 10$		5.283
		(0.229)
R2	0.065	0.418
Observations	8696	8696

Standard errors are reported in brackets and clustered on Local Authority level

Figure 5.1: Supermarkets in the UK - store openings

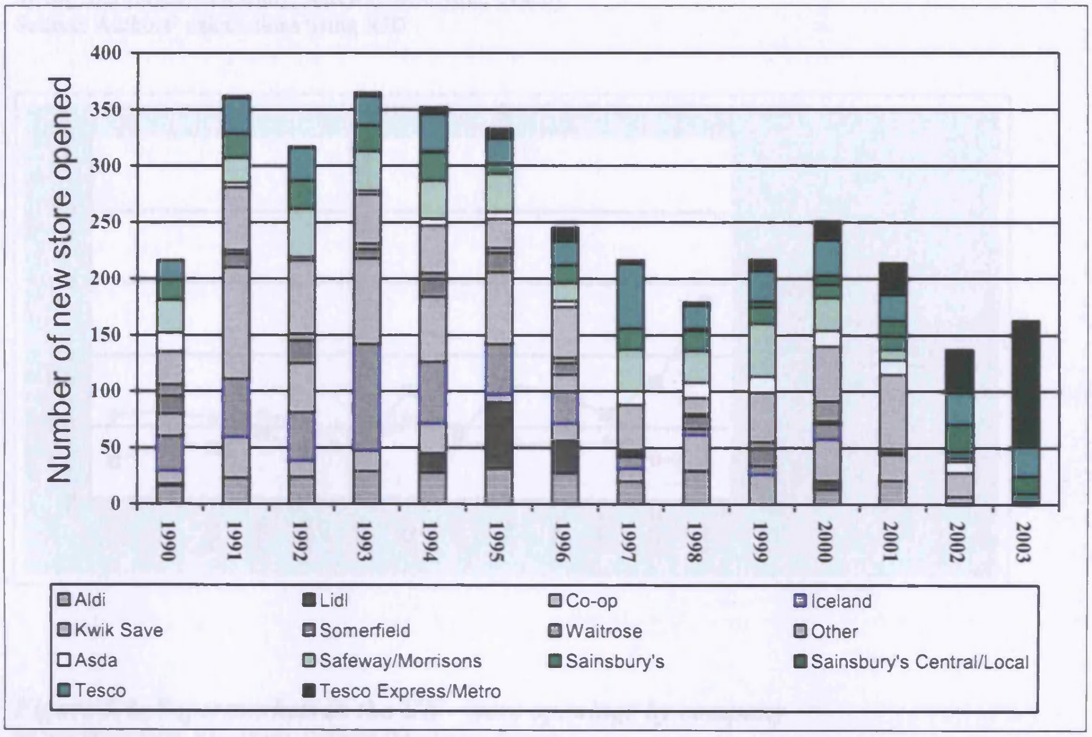


Figure 5.2: Supermarkets in the UK - store openings by location
by big four (Asda, Sainsbury, Safeway/Morrison, Tesco)
Source: Authors' calculations using IGD

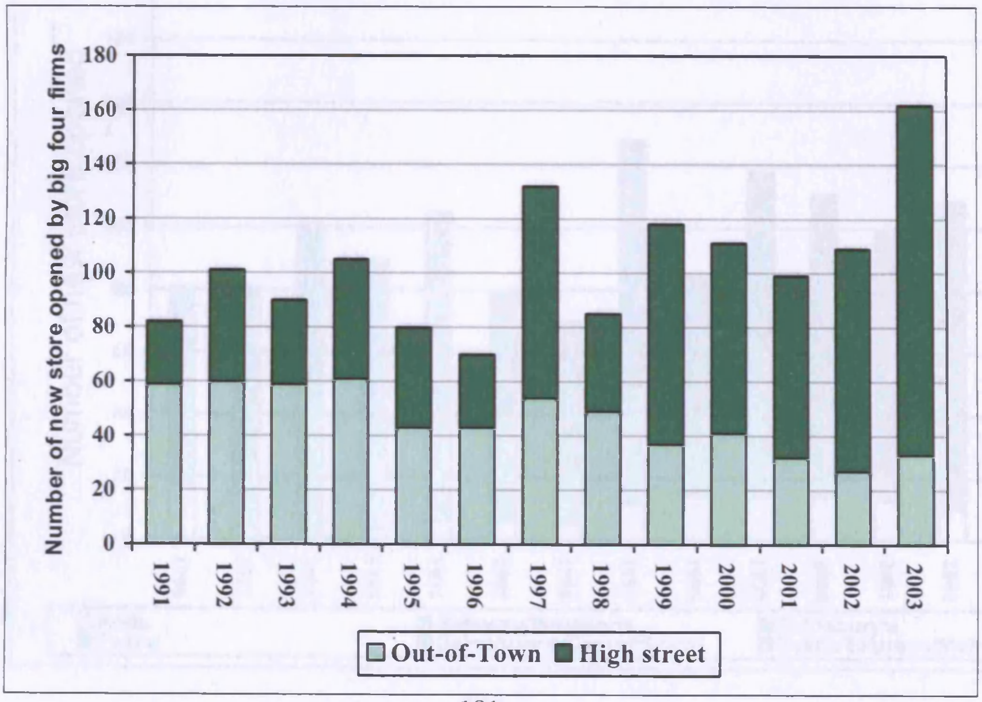


Figure 5.3: Supermarkets in the UK - store openings by size
by big four (Asda, Sainsbury, Safeway/Morrison, Tesco)
Source: Authors' calculations using IGD

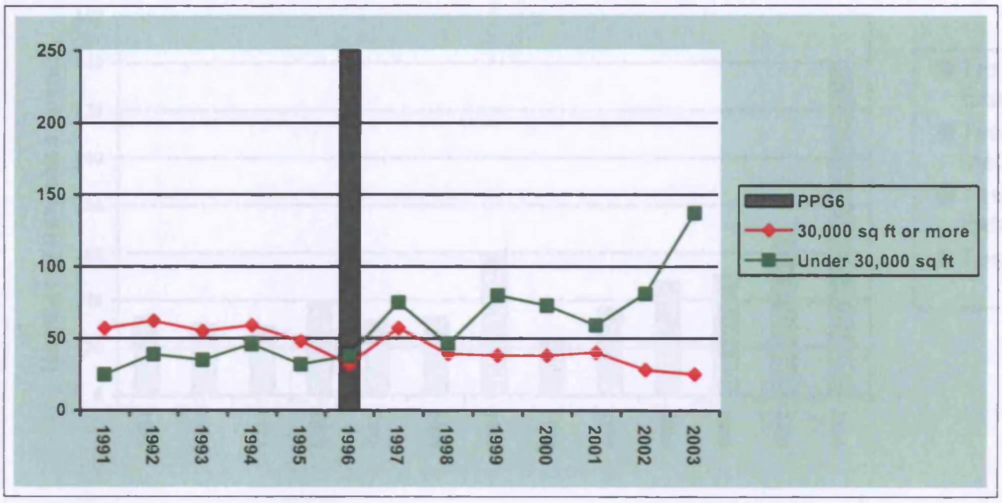


Figure 5.4: Supermarkets in the UK - store openings by company
by big four (Asda, Sainsbury, Safeway/Morrison, Tesco)
Source: Authors' calculations using IGD

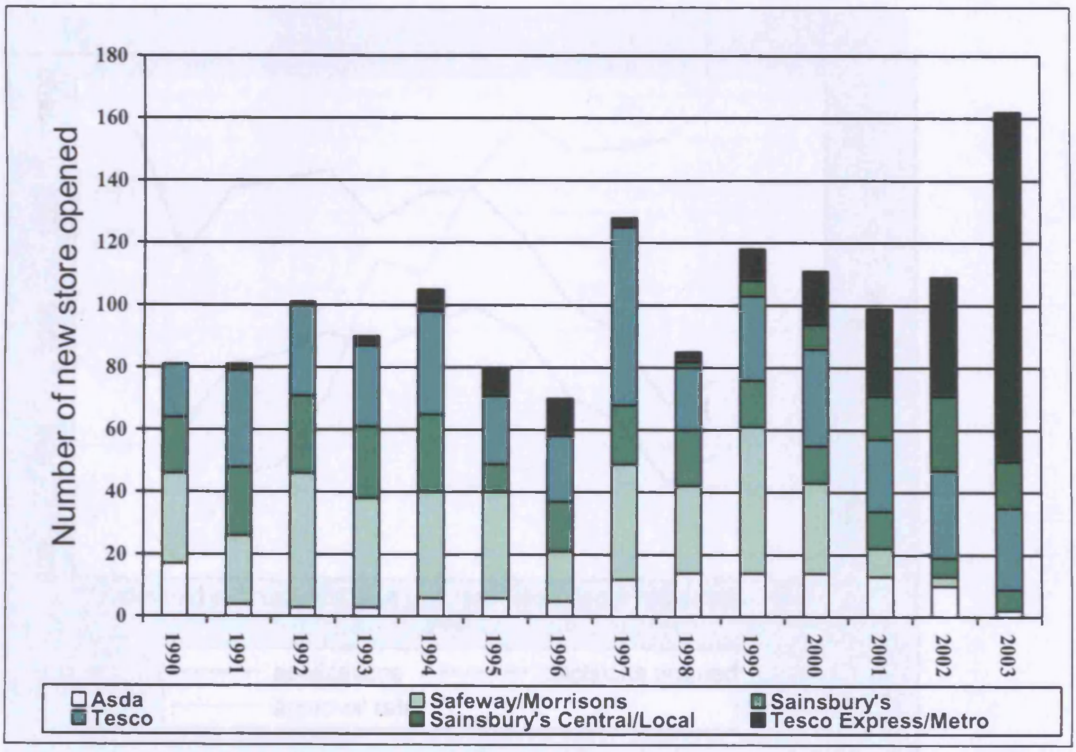


Figure 5.5: Tesco in the UK - store openings by store format
Source: Authors' calculations using IGD

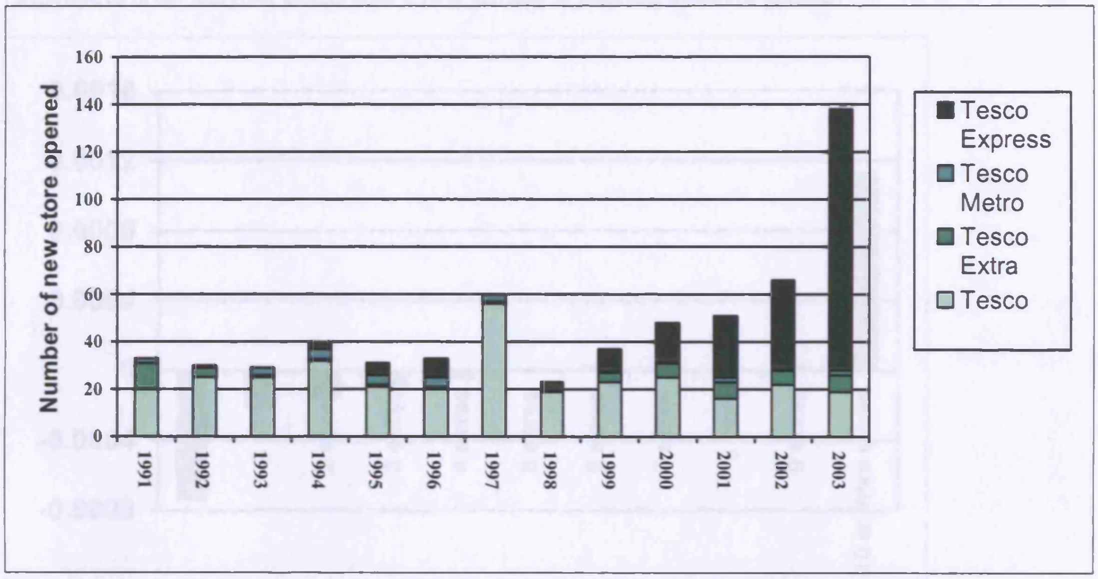


Figure 5.6: Retail applications, decisions granted and approval rate (A1& A3)
Source: Authors' calculations using ODPM

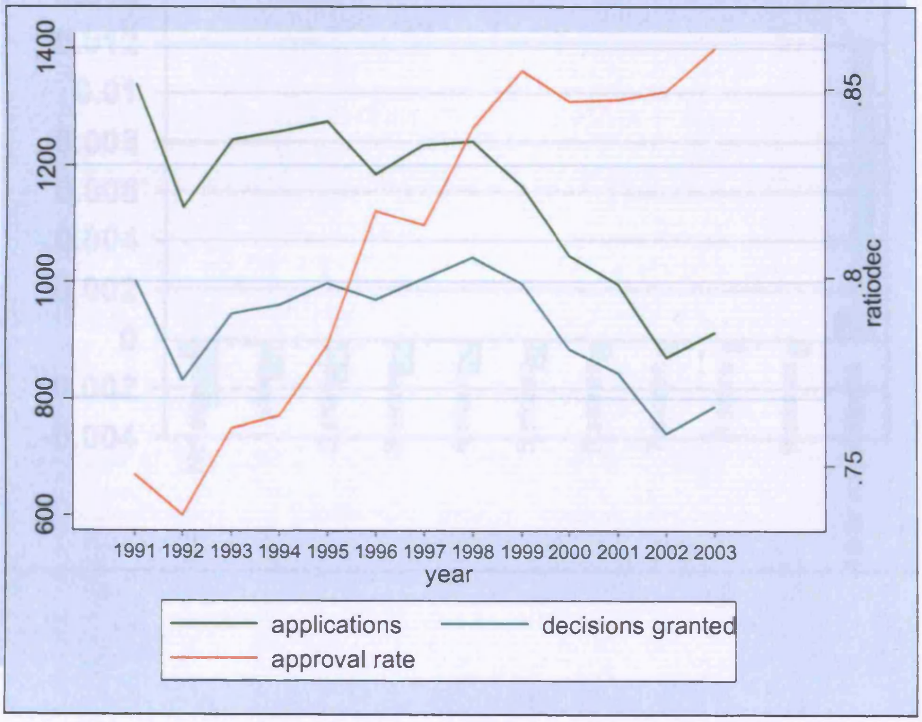


Figure 5.7: Impact of a 10% increase in planning approvals
(Differences in the marginal effects after a 10% increase in planning decisions granted)

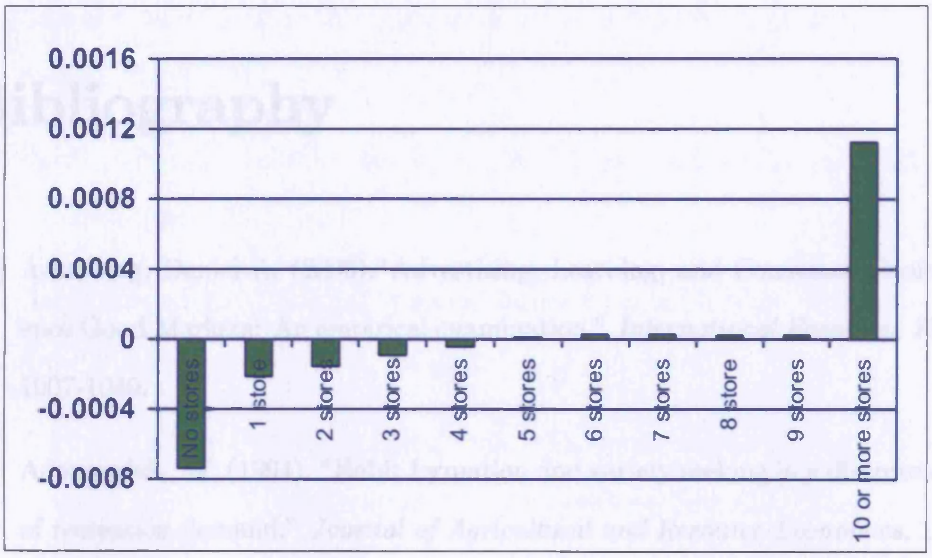
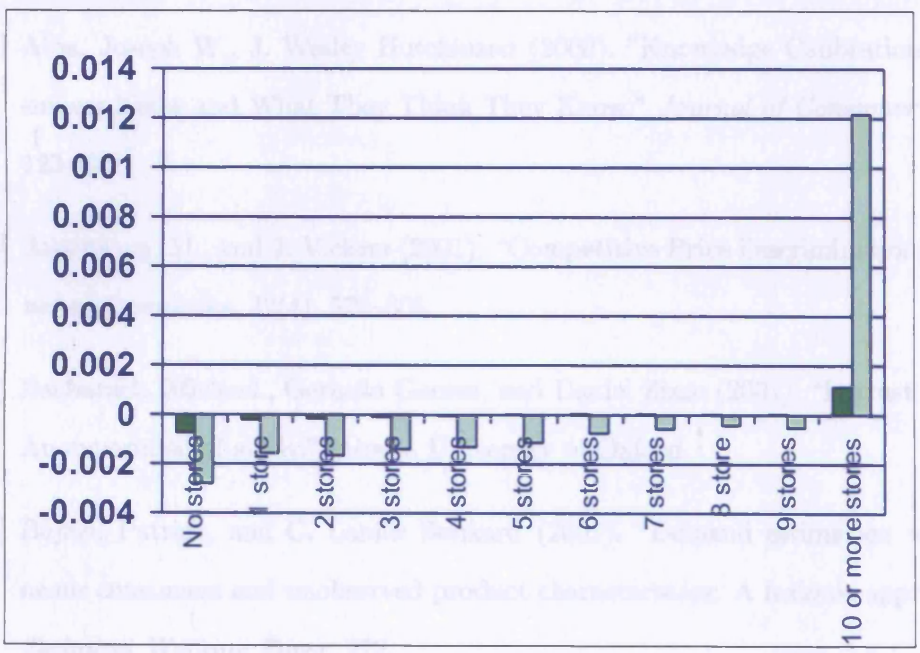


Figure 5. 8: Comparison of the impact of a 10% increase in planning approvals and a 1% increase in population (Differences in the respective marginal effects)



Bibliography

[1] Ackerberg, Daniel A. (2003)."Advertising, Learning, and Consumer Choice in Experience Good Markets: An empirical examination." *International Economic Review*, 44(3), 1007-1040.

[2] Adamowicz , W. (1994). "Habit formation and variety seeking in a discrete choice model of recreation demand." *Journal of Agricultural and Resource Economics*, 19, 19-31.

[3] Akçura, M. Tolga, Füsun F. Gönül, and Elina Petrova (2004). "Consumer Learning and Brand Valuation for Nonprescription Drugs." *Marketing Science*, 23(1), 156-169.

[4] Alba, Joseph W., J. Wesley Hutchinson (2000). "Knowledge Calibration: What Consumers Know and What They Think They Know." *Journal of Consumer Research*, 27, 123-156.

[5] Armstrong, M., and J. Vickers (2001). "Competitive Price Discrimination." *Rand Journal of Economics*, 32(4), 579–605.

[6] Bacharach, Michael., Gerardo Guerra, and Daniel Zizzo (2001). "Is trust self-fulfilling? An experimental study." *mimeo.* University of Oxford.

[7] Bajari, Patrick, and C. Lanier Benkard (2001). "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach." *NBER Technical Working Paper*, 272.

[8] Bajari, Patrick, Han Hong and Stephen Ryan (2004). "Identification and Estimation of Discrete Games of Complete Information." *mimeo.*

[9] Banerjee, A.V. (1992). "A simple model of herd behavior." *Quarterly Journal of Economics* 107, 797-817.

[10] Berry, Steven T. (1992). "Estimation of a Model of Entry in the Airline Industry". *Econometrica,* 60(4), 889-917.

[11] Berry, Steven T. (1994)."Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal*, 25(2), 242-262.

[12] Bertrand, M and F Kramatz (2002). "Does Entry Regulation Hinder Job Creation? Evidence from French Retail Industry." *The Quarterly Journal of Economics*, November, 1369-1413.

[13] Bikhchandani, S., D. Hirshleifer, and I. Welch (1992)." A theory of fads, fashion, custom, and cultural change as informational cascades." *Journal of Political Economy,* 110, 992-1026.

[14] Blow, Laura, Martin Browning, and Ian Crawford (2004)."Nonparametric methods for the consumer characteristics model." *mimeo.*

[15] Blundell, Richard W., Martin Browning, and Ian A. Crawford (2003). "Nonparametric Engel Curves and Revealed Preference." *Econometrica*, 71(1), 205-240.

[16] Bohnet, Iris, Bruno S. Frey, and Steffen Huck (2001). "More order with less law: On contract enforcement, trust, and crowding." *American Political Science Review*, 95, 131-144.

[17] Bohnet, Iris, and Steffen Huck (2004). "Repetition and reputation: Implications for trust and trustworthiness when institutions change." *American Economic Review* (Papers and Proceedings), 94, 362-366.

[18] Bolton, Gary, Elena Katok, and Axel Ockenfels (2004). "How effective are electronic feedback mechanisms? An experimental investigation." *Management Science*, forthcoming.

[19] Bresnahan, Timothy and Peter Reiss (1991). "Empirical models of discrete games." *Journal of Econometrics*, 48, 57-82.

[20] Bresnahan, Timothy and Peter Reiss (1991). "Entry and Competition in Concentrated Markets." *The Journal of Political Economy*, 99(5), 977-1009.

[21] Bresnahan, Timothy, Peter Reiss, Robert Willig and George Stigler (1987)."Do Entry Conditions Vary across Markets?." *Brookings Papers on Economic Activity*, 1987(3), 833-881.

[22] Bulow, Jeremy I., John D. Gaenakoplos, and Paul Klemperer (1985)."Multimarket Oligopoly: Strategic Substitutes and Complements." *The Journal of Political Economy*, 93(3), 488-511.

[23] Celen, B., and S. Kariv (2004)."Observational learning under imperfect information." *Games and Economic Bahvior*, 47, 72-86.

[24] Celen, B., and S. Kariv (2005). "An experimental test of observational learning under imperfect information." *Economic Theory*, 26, 677-699.

[25] Chintagunta, Pradeep K. (1999). "Variety Seeking, Purchase Timing, and the "Lightning Bolt" Brand Choice Model." *Management Science*, 45(4), 486-498.

[26] Clay, Karen, Ronald Goettler, and Eric Wolff (2003). "Consumer Learning about Experience Goods: Evidence from an Online Grocer." *mimeo.*

[27] Cohen, Andrew and Michael Mazzeo (2005). "Market structure and competition among retail depository institutions." *mimeo.*

[28] Competition Commission (2000). Supermarkets.

[29] Cotterill, R and L Haller (1992). "Barrier and Queue Effects: A Study of Leading US Supermarket Chain Entry Patterns." *The Journal of Industrial Economics*, 40(4), 427-440.

[30] Crawford, Ian and Neil Rickman (2006). "Nonparametric methods for the Rational Addiction Model." *mimeo.*

[31] Daunfeld, Sven-Olov, Matilda Orth, and Niklas Rudholm (2005). "Entry into the Swedish Retail Food Market: A Real Options Approach." *mimeo.*

[32] Dixit, A., and R. Pindyck. (1994). *Investment Under Uncertainty.* Princeton University Press, Princeton.

[33] Djankov, La Porta, Lopez-de-Silanes and Shleifer (2002). "The Regulation of Entry." *The Quarterly Journal of Economics*, February 2002'

[34] Dong, Diansheng, Chanjin Chung, and Harry M. Kaiser (2004). "Modelling milk purchasing behaviour with a panel data double-hurdle model." *Applied Economics*, 36, 769-779.

[35] Ellickson, Paul B. (2004). "Supermarkets as a Natural Oligopoly." *mimeo.*

[36] Erdem, Tülin, Michael P. Keane (1996)."Decision-making Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer goods Markets." *Marketing Science*, 15(1), 1-20.

[37] Erdem, Tülin (1996). "A dynamic analysis of market structure based on panel data". *Marketing Science*, 15, 359-378.

[38] Erdem, Tülin (1998)."An Empirical Analysis of Umbrella Branding". *Journal of Marketing Research*, 35, 339-351.

[39] Erdem, Tülin, Michael P. Keane, and Baohong Sun (1999). "Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters." *Journal of Econometrics*, 89, 177-196.

[40] Erdem, Tülin, Michael P. Keane, and Baohong Sun (2004). "The Impact of Advertising on Consumer Price Sensitivity in Experience Goods Markets." *mimeo.*

[41] Erev, Ido and Alvin E. Roth (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria". *American Economic Review*, 88(4), 848-881.

[42] Fader, Peter S., Bruce G. S. Hardie (1996). "Modeling Consumer Choice among SKUs." *Journal of Marketing Research*, 33(4), 442-452.

[43] Fischbacher, Urs (1999): "z-tree. Toolbox for readymade economic experiments." IEW Working Paper 21, University of Zurich.

[44] Fudenberg and Levine (1998) "Theory of Learning in Games".

[45] Geweke, John. (1989). "Bayesian inference in econometric models using Monte Carlo integration". *Econometrica*, 57, 1317-1339.

[46] Geweke, John F., Michael P. Keane, and David E. Runkle (1994). "Alternative computational approaches to inference in the multinomial probit model." *Review of Economics and Statistics*, 76, 609-632.

[47] Geweke, John F., Michael P. Keane, and David E. Runkle (1997). "Statistical Inference in the Multinomial Multiperiod Probit Model." *Journal of Econometrics*, 80, 125-165.

[48] Gorman, W.M. (1980). "A possible procedure for analysing quality differentials in the eggs market." *Review of Economic Studies*, 47, 843-856.

[49] Gönül, Füsun F. (1999). "Estimating price expectations in the OTC medicine market: An application of dynamic stochastic discrete choice models to scanner panel data." *Journal of Econometrics*, 89,41-56.

[50] Griffith, R, R Harrison, J Haskel and M Sako (2003). "The UK Productivity Gap and the Importance of the Service Sectors" *AIM Briefing Note*, December 2003, http://www.aimresearch.org/publications/rgbrief.pdf

[51] Guadagni, Peter M., and John D.C. Little (1983). "A Logit Model of Brand Choice Calibrated on Scanner Data." *Marketing Science*, 2(3), 203-238.

[52] Gul, Faruk and Wolfgang Pesendorfer (2004)."Self-Control and the Theory of Consumption." *Econometrica*, 72, 119-158.

[53] Hajivassiliou, Vassilis, Daniel McFadden, and P. Ruud (1996). "Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results." *Journal of Econometrics*, 72, 85-134.

[54] Haskel, Jonathan and Naveed Khawaja (2003). "Productivity in UK Retailing: Evidence from Micro Data." *CERIBA Working Paper.*

[55] Hausman, Jerry A., and David A. Wise (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences." *Econometrica*, 46(2), 403-426.

[56] Hausman, Jerry A. and Daniel L. McFadden (1984). "Specification tests for the multinomial logit." *Econometrica*, 52, 1219-1240.

[57] Heckman, James and Jose Scheinkman (1987). "The Importance of Bundling in a Gorman-lancaster Model of Earnings." *The Review of Economic Studies*, 54(2), 243-255.

[58] Ho, Teck-Hua, and Juin-Kuan Chong (2004). "A Parsimonious Model of Stock-Keeping Unit Choice." *Journal of Marketing Research*, Vol: XL, 351-365

[59] Hopkins, Ed (2006). "Adaptive Learning Models of Consumer Behavior." *mimeo.*

[60] Kahn, Barbara E. (1998). "Variety: From the Consumer's Perspective", in *Research Advances in Variety Management*, ed. T.H. Ho and C.S. Tang, Bosten MA: Kluwer Academic Publishers, 19-37.

[61] Keane, Michael P. (1997)."Modeling Heterogeneity and State Dependence in Consumer Choice Behavior." *Journal of Business & Economic Statistics*, 15(3), 310-326.

[62] Keser, Claudia (2002). "Trust and reputation building in e-commerce." *mimeo.* Cirano Monetral.

[63] Kubler, D., and G. Weizsacker (2005)."Are longer cascades more stable?" *Journal of the European Economic Association*, 3, 330-339.

[64] Lancaster, Kelvin (1966). "A New Approach to Consumer Theory." *Journal of Political Economy*, 74, 132-157.

[65] Mattei, Aurelio (1994). "La Cohèrence des choix des consuommateurs suisses." *Revue Suisse d'Economie Politique et de Statistique*, 130, 3-20.

[66] Mattei, Aurelio (2000). "Full-scale real tests of consumer behavior using experimental data." *Journal of Economic Behavior & Organization*, 43, 487-497.

[67] Mazzeo, M. (2002). "Product choice and oligopoly market structure." *RAND Journal of Economics*, 33,1-22.

[68] McAlister, Leigh (1982). "A Dynamic Attribute Satiation Model for Choices Made Across Time." *Journal of Consumer Research*, 9(3), 141-150.

[69] McFadden, Daniel L. (1974). "Conditional logit analysis of qualitative choice behavior." In P. Zarembka, ed. *Frontiers in Econometrics*, Academic Press, New York, 105-142.

[70] McFadden, Daniel L., and Kenneth E. Train (1996). "Consumers' Evaluation of New Products: Learning from Self and Others." *The Journal of Political Economy*, 104(4), 683-703.

[71] McKinsey (1998) "Driving productivity and growth in the UK economy" McKinsey Global Institute.

[72] Meyer, Robert J., and Arvind Sathi (1985). "A Multiattribute Model of Consumer Choice during Product Learning." *Marketing Science*, 4(1), 41-61.

[73] Miravete, Eugenio J., and Ignacio Palacios-Huerta (2002). "Learning Temporal Preferences." *CEPR Discussion Paper*, WP-DP3604.

[74] Offerman, Theo, and Joep Sonnemans (1998). "Learning by experience and by imitating successful others." *Journal of Economic Behavior and Organization*, 34, 559-575.

[75] Osselaer, Stijn M.J. van (2000). "Consumer Learning and Brand Equity." *Journal of Consumer Research*, 27, 1-16.

[76] Papatla, Purushottam, and Lakshman Krishnamurthi (1992). "A Probit Model of Choice Dynamics." *mimeo.*

[77] Petrin, A. and K. Tran (2004). "Omitted Product Attributes in Differentiated Product Models." *mimeo.*

[78] Reiss, Peter C. (1996). "Empirical Models of Discrete Strategic Choices." *The American Economic Review,* 86(2), 421-426.

[79] Reis, Ricardo (2006)."Inattentive consumers." *Journal of Monetary Economics*, forthcoming.

[80] Sippel, Reinhard (1997). "An experiment on the pure theory of consumer's behaviour." *Economic Journal*, 107, 1431-1444.

[81] Slembeck, Tilman, and Jean-Robert Tyran (2004). "Do institutions promote rationality? An experimental study of the three-door anomaly." *Journal of Economic Behavior and Organization*, 54, 337-350.

[82] Smith, Howard (2004). "Supermarket Choice and Supermarket Competition in Market Equilibrium." *Review of Economic Studies*, 71, 235-263.

[83] Song, Minjae (2004)."Measuring Consumer Welfare in the CPU Market: an application of the pure characteristics demand model." *mimeo.*

[84] Tamer, Elie (2002). "Empirical Strategies for Estimating Discrete Games with Multiple Equilibria." http://www.olin.wustl.edu/cres/research/calendar/files/ACF1D3A.pdf

[85] Train, Kenneth (2003). *Discrete Choice Methods with Simulation.* New York: Cambridge University Press.

[86] Villas-Boas, J. Miguel, and Russell S. Winer (1999). "Endogeneity in Brand Choice Models." *Management Science*, 45(10), 1324-1338.

[87] Villas-Boas, J. Miguel (2004). "Consumer Learning, Brand Loyalty and Competition." *Marketing Science*, 23(1), 134-15.