Customer Focused Price Optimisation

Problem presented by

Tim Butler

Tesco

ESGI100 was jointly hosted by
Smith Institute for Industrial Mathematics and System Engineering
and the University of Oxford

with additional financial support from
Engineering and Physical Sciences Research Council
European Journal of Applied Mathematics
Oxford Centre for Collaborative Applied Mathematics
Warwick Complexity Centre

Report author

Laura Caravenna (University of Oxford)
Jeff Dewynne (University of Oxford)
Chris Farmer (University of Oxford)
Owen Jones (Smith Institute)
Christian Overgaard Lund (University of Cambridge)
Sergey Melnik (University of Limerick)
Bogna Pawlowska (University of Exeter)
Eddie Wilson (University of Bristol)

Executive Summary

Tesco want to better understand how to set online prices for their general merchandise (*i.e.* not groceries or clothes) in the UK. Because customers can easily compare prices from different retailers we expect they will be very sensitive to price, so it is important to get it right. There are four aspects of the problem.

- *Forecasting:* Estimating the customer demand as a function of the price chosen (especially hard for products with no sales history or infrequent sales).
- *Objective function:* What exactly should Tesco aim to optimise? Sales volume? Profit? Profit margin? Conversion rates?
- *Optimisation:* How to choose prices for many related products to optimise the chosen objective function.
- *Evalution:* How to demonstrate that the chosen prices are optimal, especially to people without a mathematical background.

Aggregate sales data was provided for about 400 products over about 2 years so that quantitive approaches could be tested. For some products competitors' prices were also provided.

Version 1.0
July 1, 2014
iv+39 pages

Contributors

Laura Caravenna (University of Oxford)
Pierluigi Cesana (University of Oxford)
Esther Daus (Vienna Institute of Technology)
Jeff Dewynne (University of Oxford)
Chris Farmer (University of Oxford)
Franca Hoffman (University of Cambridge)
Owen Jones (Smith Institute)
Grzegorz Krzyżanowski (Wrocław University of Technology)
David Leppinen (University of Birmingham)
Christian Overgaard Lund (University of Cambridge)
Sergey Melnik (University of Limerick)
Bogna Pawlowska (University of Exeter)
Zubin Siganporia (University of Oxford)
Damian Smug (Wrocław University of Technology)
Eddie Wilson (University of Bristol)

Contents

1 Introduction

(1.1) Tesco are the world's third largest retailer, with annual sales of £72 billion and 520,000 employees. In this project we will consider only online sales in the UK of products in the general merchandise category. This excludes groceries and clothes. Competitors vary depending on the product but they include Amazon, Argos and Ebay.

1.1 Problem statement

(1.2) "Within prevailing market conditions, how can we determine the best price for our customers?"

(1.3) For each product legal and commercial requirements will give an upper and lower bound for the price. Because it is so easy to compare prices from different retailers when shopping online we expect customers will be very sensitive to price, so it is important to get it right.

(1.4) Tesco highlighted four different aspects of the problem.

- Forecasting
- Objective Function
- Optimisation
- Evaluation

(1.5) *Forecasting* means estimating the volume of sales as a function of the price. This is particularly challenging for products with little or no sales history, or products with infrequent sales. This will of course depend on other factors, such as competitors' prices and the weather, which can be observed but not controlled.

(1.6) The *objective function* is what Tesco should aim to maximise to deliver the best price for their customers. For example, they might use profit, sales, margin or conversion rate.

(1.7) *Optimisation* is the process of choosing prices in order to make the objective function as large as possible. This could be done on a product by product basis, or it could be done for all the products at once, or anywhere in between.

(1.8) *Evaluation* is the process of deciding how advantageous a particular approach to pricing is. This happens both before and after a new approach is adopted. Often the evidence that a new approach is advantageous needs to be understandable to executives who do not have a background in mathematics.

1.2 Overview of provided data

(1.9) Tesco provided two sets of data for analysis during the study group. They were based on real data but with various identifiers anonymised and with some of the numbers changed to protect commercially sensitive information.

(1.10) The first set was Tesco aggregate sales data for about 400 products over about 2 years. It was a CSV file where each row contained data about the number of sales of a particular product on a particular day. Each row contained the following information.

- **Date**: the date in question
- **Sales_qty**: the number of items sold on that date
- **Product_id**: a unique identifying number for the product
- **Selling_price**: the price of one unit of the product to the customer
- **Cost_price**: the cost to purchase one unit of the product from the supplier

(1.11) The second set of data was prices for the same products from competing retailers. Each row contained the following information.

- **Date**: date price was collected
- **Product_id**: identifies product
- **Competitor_id**: identifies the retailer
- **Competitor_price**: price collected from the retailer's website

2 Data analysis

2.1 General observations and problems with the data set

(2.1) In general, there are very few sales per week per product in the data set, and there are many products with no sales for multiple weeks (see Fig. 1).

(2.2) Additionally, the cost data is not realistic as demonstrated in Fig. 2.

(2.3) It is not possible to obtain a sensible price-demand relationship based on available data. For example, in Fig. 3 (right, top panel) we show a typical dependence of sales rate on price that we observed for one of the products. The points are scattered, and lower prices do not seem to correspond to higher sales rates. A fundamental problem in determining the price-demand relationship is the separation of several confounding effects including the current prices of competitors, seasonality/popularity, and availability of alternative products.

(2.4) Competitor price data only starts at the end of 2013 and is sparse (see the price panels in Fig. 4). Furthermore, it exist only for some products (see Fig. 3, left).

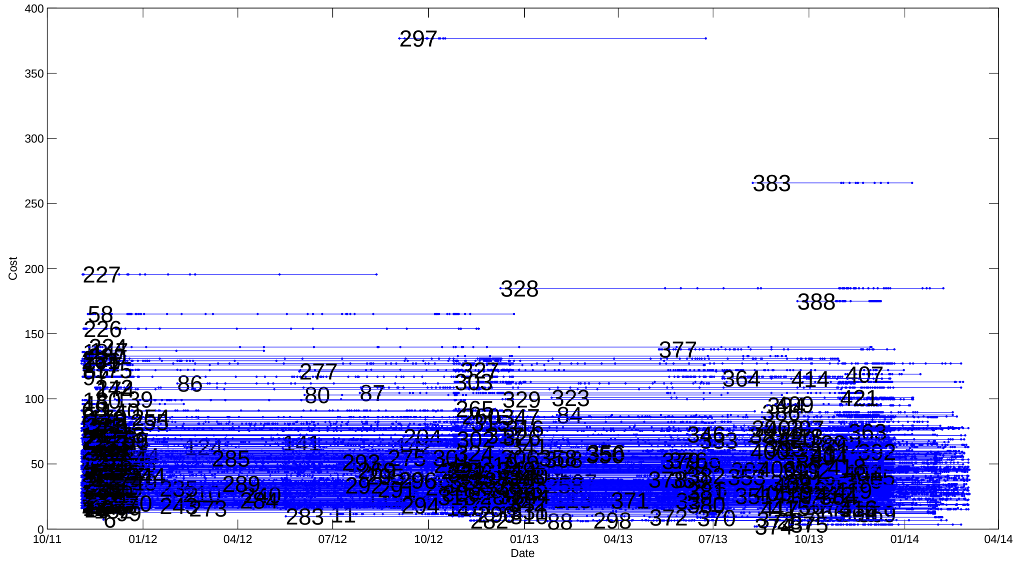

Figure 1: Cost time series for all 421 products in the data set. Each point represents product cost on a day when a product had sales (note that cost never changes throughout the period of observation). Product_ids are shown at the start of each time series. Observe that first-sale dates for many products are distributed throughout the period of observation, with notable surges towards the end of each year.

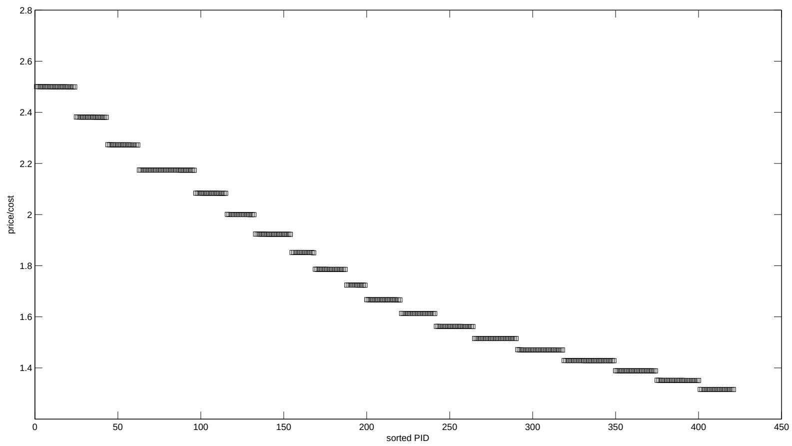

Figure 2: Price-to-cost ratio for all 421 products sorted in descending order. The clearly visible bands suggest that the product cost data is probably artificial, thus we should not rely on it in our analysis.

(2.5) Examples of product-specific time evolution of prices, sales, and profit are shown in Fig. 4.

- Product 10: Peaks in sales during Christmas periods 2012 and 2013, and in the summer 2013. The summer sales may be caused by the decreased price and seem to result in profit loss.
- Product 321: Sales clearly respond to price variation, which probably

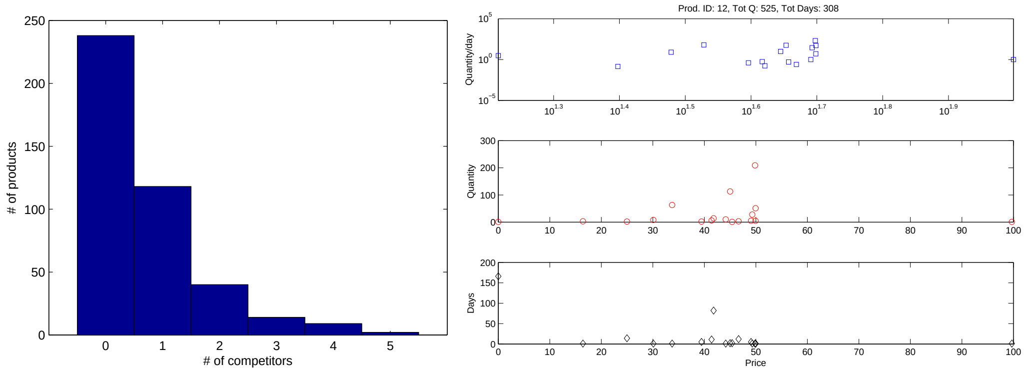

Figure 3: **Left:** Histogram showing the number of products vs the number of competitors for these products in the data set. For most products (approx. 240) there is no competitor price data in the data set. For some products (approx. 115) there is only one competitor. Very few products have price data from more than 1 competitor. **Right:** Top panel: Typical observed dependence of daily product sales on its price. Middle and bottom panels: the total sales at a particular price and the number of days when the price was current.

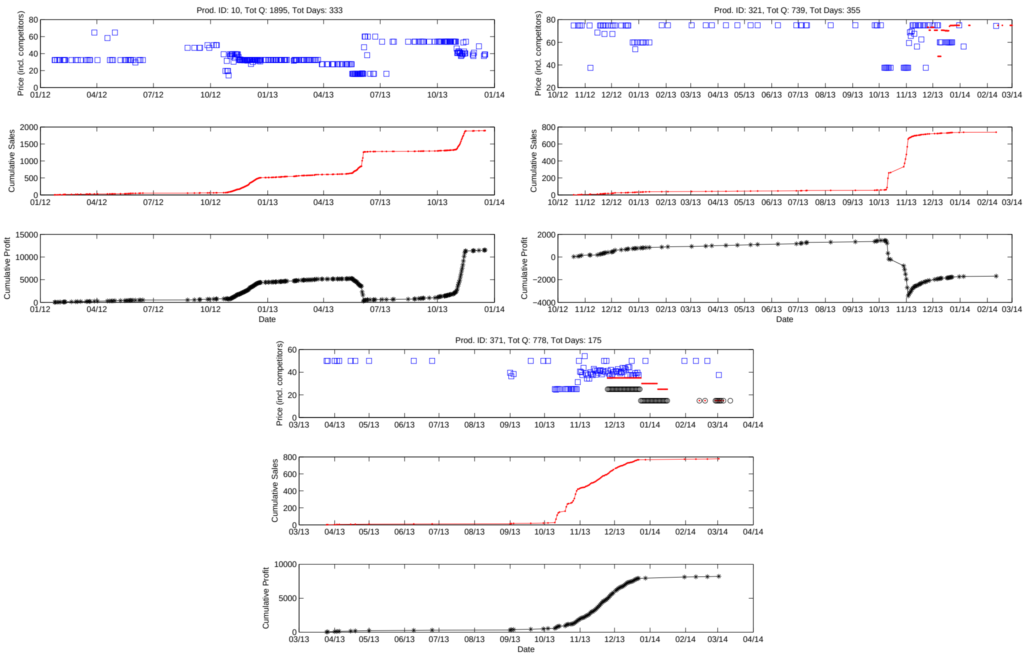

Figure 4: Plots showing time evolution of price (each competitor's price, if available, is indicated by different symbols), cumulative number of items sold, and cumulative profit [*i.e.* (price-cost)×quantity] for 3 different products (IDs: 10, 321, and 371).

was a half-price offer in October-November 2013. Also there is a gap in sales in the mid-October'13, which may indicate that the product was out of stock.

- Product 371: fast daily variations in price in November-December 2013. Observe that lower prices lead to higher slopes in cumulative sales. Competitors have lower prices, but Tesco still makes profit, at least initially. There are virtually no sales in 2014. This may suggest that the price was set too high, and that decreasing the price below that of one of the competitors may increase sales and result in greater profit. An alternative explanation for the lack of sales is that this product is only popular for Christmas.

2.2 Seasonal price variations

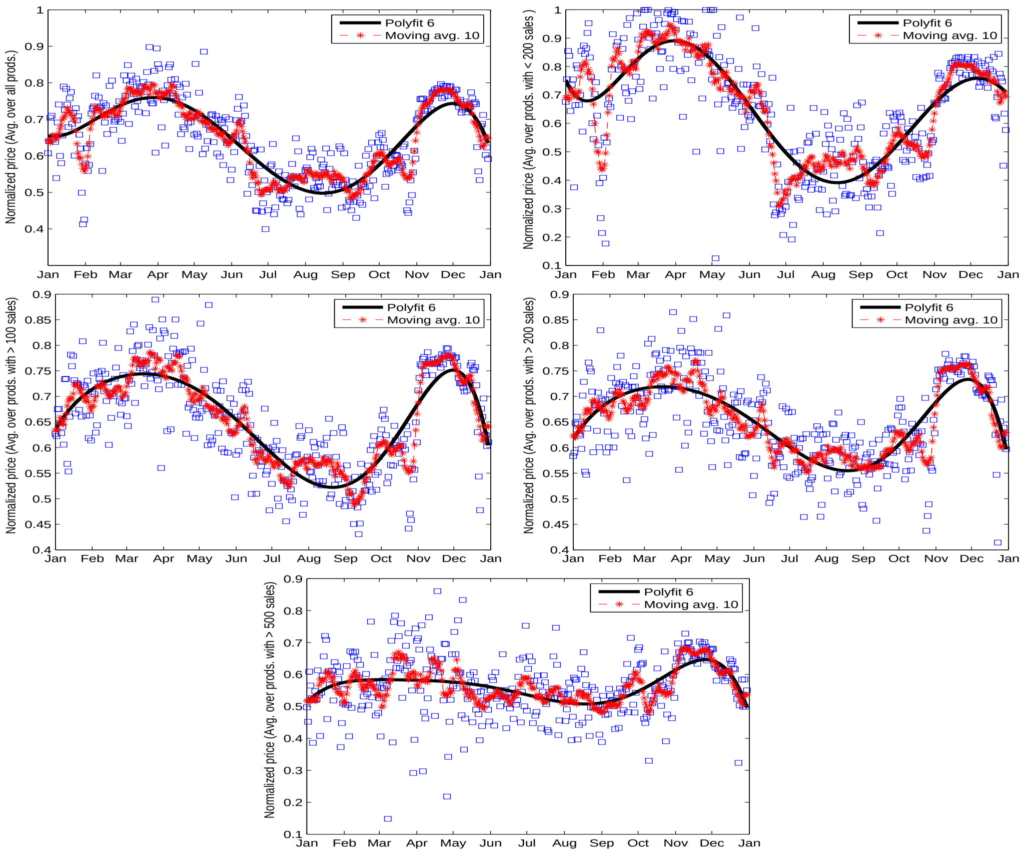

Figure 5: Seasonal variation of product prices during the year. Different panels correspond to averaging over different subsets of products as indicated in the y-axis in each panel. Before averaging, the price of each product is first normalised between 0 and 1 with respect to observed minimum and maximum price for that product. Red asterisks represent moving average over 10 days. Black curve is a fitted degree-6 polynomial function.

(2.6) In Fig. 5, we show the variation of product prices during the year. Different panels correspond to averaging over different subsets of products.

(2.7) When considering all products, the lowest prices are observed in July, August and September. The peak prices are observed in November-December and March.

(2.8) The seasonal variation for infrequently sold products is more pronounced than for products that are sold more frequently. This may be a real effect, or it may be simply due to the different number of samples in each set of products.

(2.9) Note there is a significant price decrease at the end of January for products with <200 sales, but not for products with >200 sales.

2.3 Seasonal variations in sales

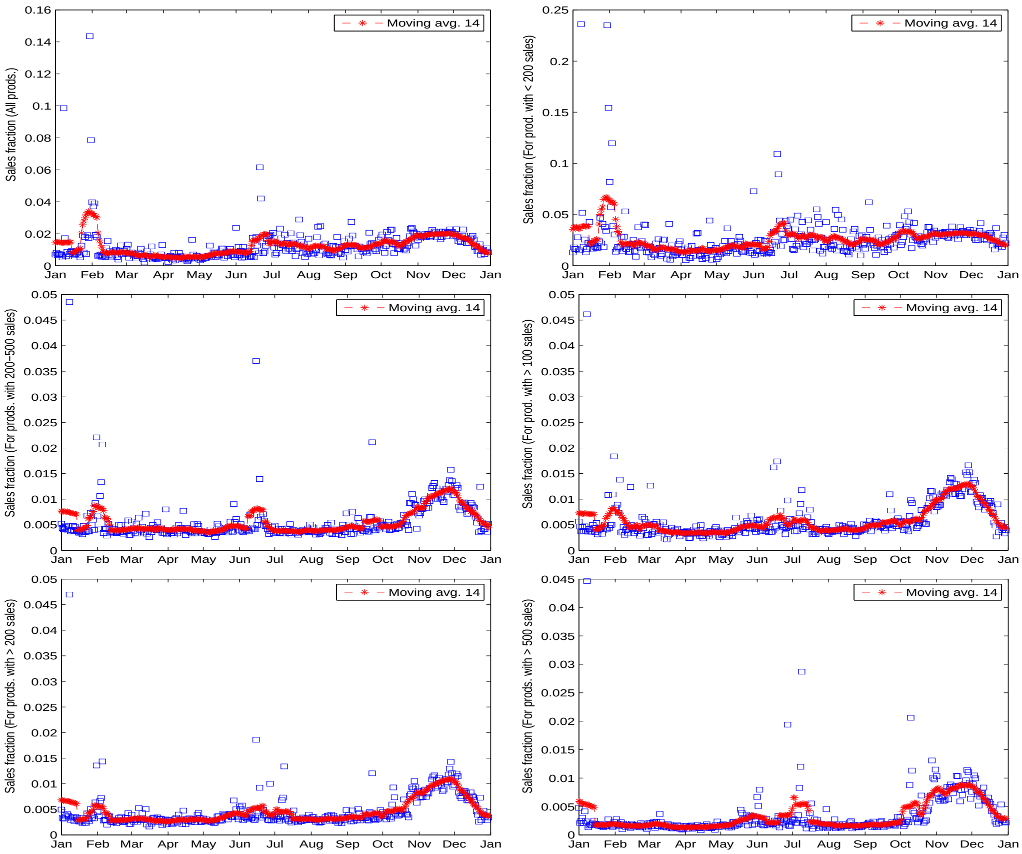

Figure 6: Seasonal variation of product sales during the year. Different panels correspond to averaging over different subsets of products as indicated in the y-axis in each panel. Before averaging, the daily sales of each product is converted to a fraction of its total sales. Red asterisks represent moving average over 14 days.

(2.10) In Fig. 6, we show the variation of product sales during the year. Different panels correspond to different subsets of products.

(2.11) The sales patterns are distinct for infrequently sold products (*e.g.* those with <200 sales) and products that are sold more frequently (*e.g.* those with >200 sales).

(2.12) We expect that most sales happen during Christmas period. This is partly true. Additionally, there are pronounced peaks in sales at the end of January, and some surge of activity in the summer. As illustrated in Fig. 4 (Product 10), there are products with sale spikes in the summer, but possibly not every summer.

(2.13) Interestingly, there is no end-of-January peak in sales for products with > 500 sales, but there are such peaks for other subsets of products. Indeed the seasonal patterns of sales for products with > 500 sales and those with < 200 sales look quite different to each other: the former has a Christmas peak, while the latter peaks at the end of January.

2.4 Seasonal variations in spending

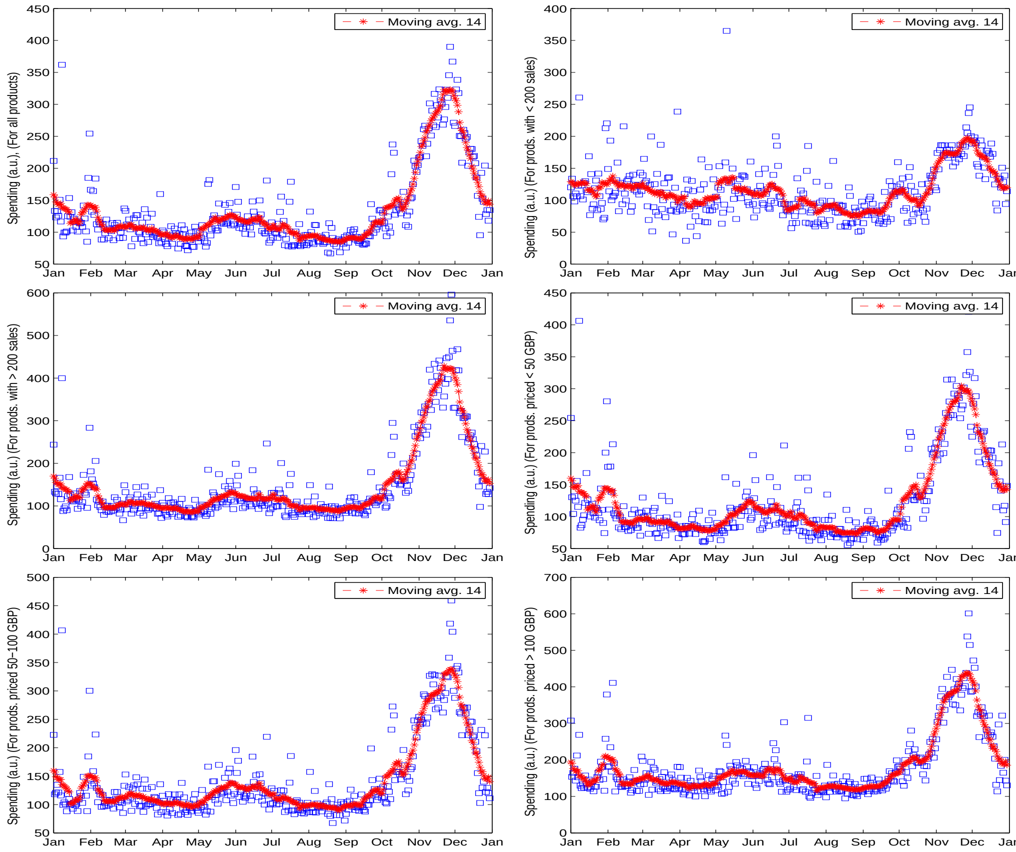

Figure 7: Seasonal variation of customer spending during the year. Different panels correspond to different subsets of products as indicated in the y-axis in each panel. Red asterisks represent moving average over 14 days.

(2.14) In Fig. 7, we show variation in customer spending during the year. Different panels correspond to different subsets of products.

(2.15) We observe increase in spending in (i) the pre-Christmas period (with a maximum at the beginning of December), (ii) the end of January, (iii) June, and (iv) October.

(2.16) The variation in spending seems to be less pronounced for infrequently sold products (*e.g.* those with < 200 sales).

(2.17) Note that for products with < 200 sales, the drop in prices at the end of January in Fig. 5 corresponds to the increase in sales in Fig. 6, but it does not seem to significantly affect the spending in Fig. 7.

(2.18) The spending patterns for expensive (> 100 GBP), intermediate ($50 - 100$ GBP) and cheap (< 50 GBP) products look similar to each other.

(2.19) However, the spending patterns are more distinct if one compares infrequently sold products (*e.g.* those with < 200 sales) with products that are sold more frequently (*e.g.* those with > 200 sales).

2.5 Product classification

(2.20) Poor data quality may be improved by collecting individual product IDs into classes of similar products. It may be worth trying to cluster products into groups based, for example, on their temporal sales patterns (in contrast or in addition to using Tesco's existing classification of products). [3] may be a good starting point. Such clustering will help alleviate the problem of data sparsity and noise and may lead to a finding hidden relationships among the products.

3 Using time series analysis to forecast demand

(3.1) In order to choose the best pricing strategy or to parameterise the later models, information about customer demand is crucial. Demand forecasting may be done using standard statistical methods. In this section we will consider a time series approach.

(3.2) We have analysed the given set of data containing the product ids, prices and numbers of sold products within a range of time. Among various products, we have noticed one typical behaviour: a significant change in the price (*e.g.* a special offer) is followed by a significant response in the demand. Moreover, the demand stays high for a longer period of time, even if the price increases again. We show this phenomenon in the figure 8.

(3.3) We have checked the correlation between prices and demands at different time points (see figure 9) and we conclude that the time structure gives important information about the demand.

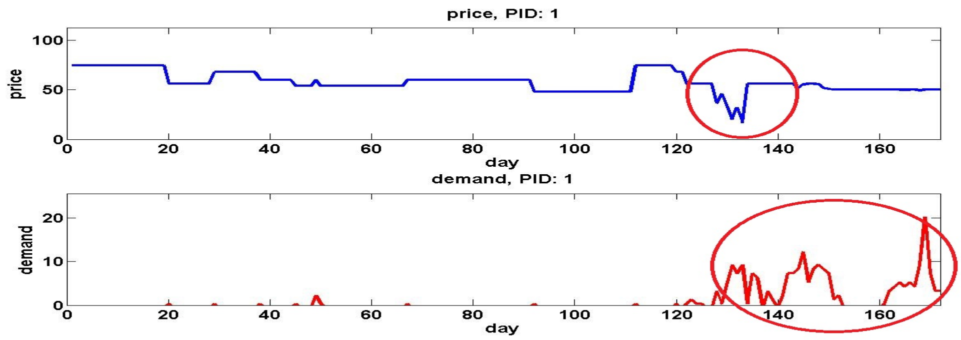

Figure 8: Price and demand: an impulse in price provokes a lagged response in the demand for product id 1.

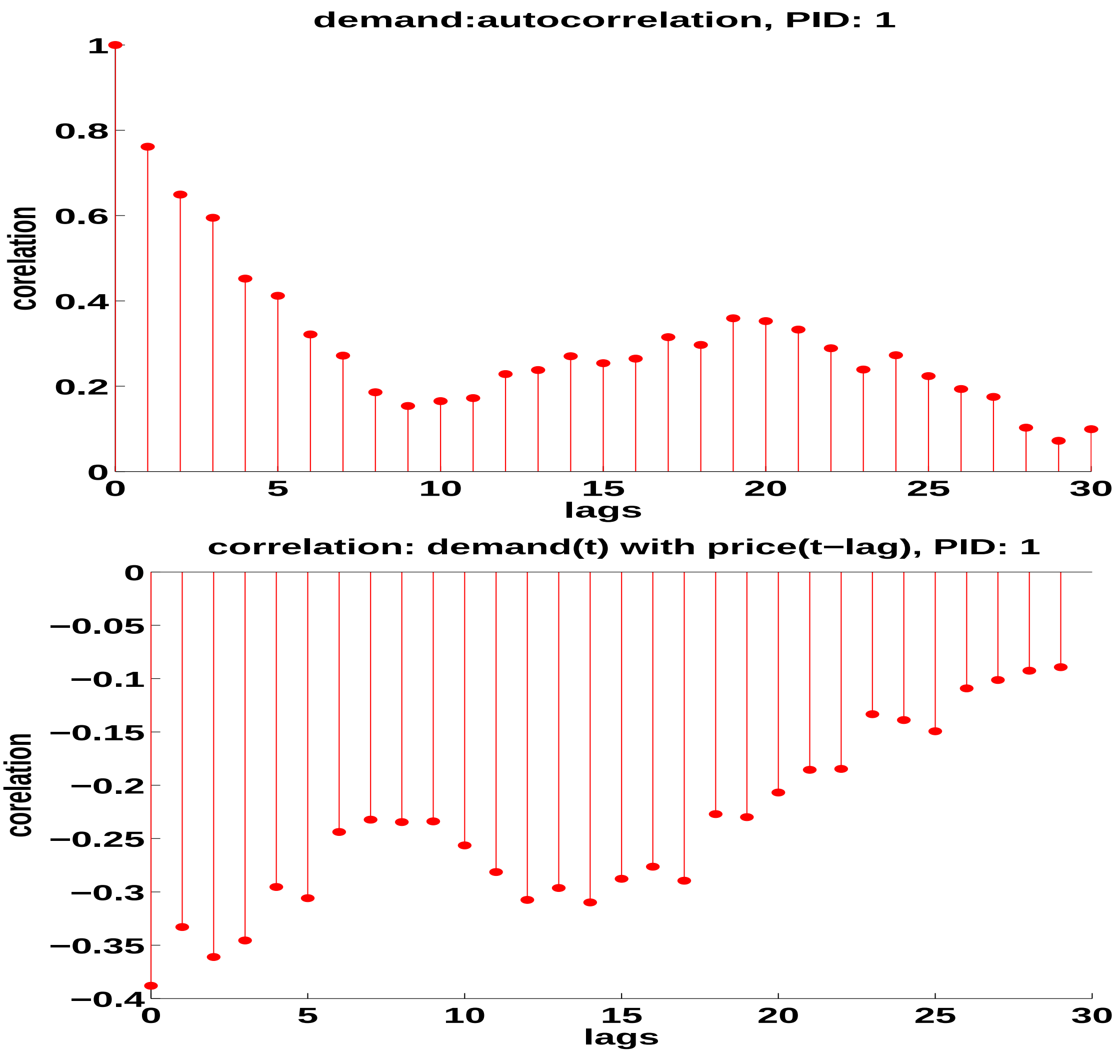

Figure 9: *Top:* Correlation between the current and previous demands for product id 1. *Bottom:* Correlation between the current demand and previous prices for product id 1.

(3.4) Therefore, we propose to forecast it using a time series approach. Given the basic data, we have considered an autoregressive model with exogenous inputs (commonly abbreviated to ARX). That means that we predict the demand at time t, Y_t, using the information about the previous demands (Y_i) and prices (U_j) (including the current price) according to the following formula:

$$\alpha_0 Y_t + \alpha_1 Y_{t-1} + ... + \alpha_n Y_{t-n} = \beta_0 U_t + \beta_1 U_{t-1} + ... + \beta_n U_{t-n} + e_t$$

We calculate the α_i and β_j parameters which make the model fit best on a subset of the observations (the training data), and then see how model predicts the rest of the observations (the test data). The results of this are shown in figure 10 for product id 1, where we have taken $n = 30$ and used the most volatile section of observations as the training data.

(3.5) Before carrying out the statistical analysis we have calculated the mean demand over all observations and subtracted that from the demands so that they have mean 0. However, knowing more details about the price and demand seasonality, a more accurate deseasonalisation procedure may be applied. The presented result is just an example, forecasted for one particular product, having taken a particular set of data. Nevertheless, we outline that a lot of useful information might be contained in the time intervals of volatile price, and a more detailed analysis of price and demand time series might give a good insight into the demand dynamics.

(3.6) Suggested directions for future research:

- Explore how robust the model is.
- Investigate whether products can be segmented by how demand responds to price.
- Incorporate a deterministic model of how demand will vary over time, *e.g.* that demand will decrease after the product is first released, or that it will have a seasonal component, so that the time series is considering only the variation in demand caused by changes in price.
- Include competitors' prices in the model.
- As presented this model assumes a linear relationship between demand and price. It would not be difficult to replace this by a different function, *e.g.* assuming a linear relationship between the log of the demand and the log of the price, as in a standard price elasticity model.

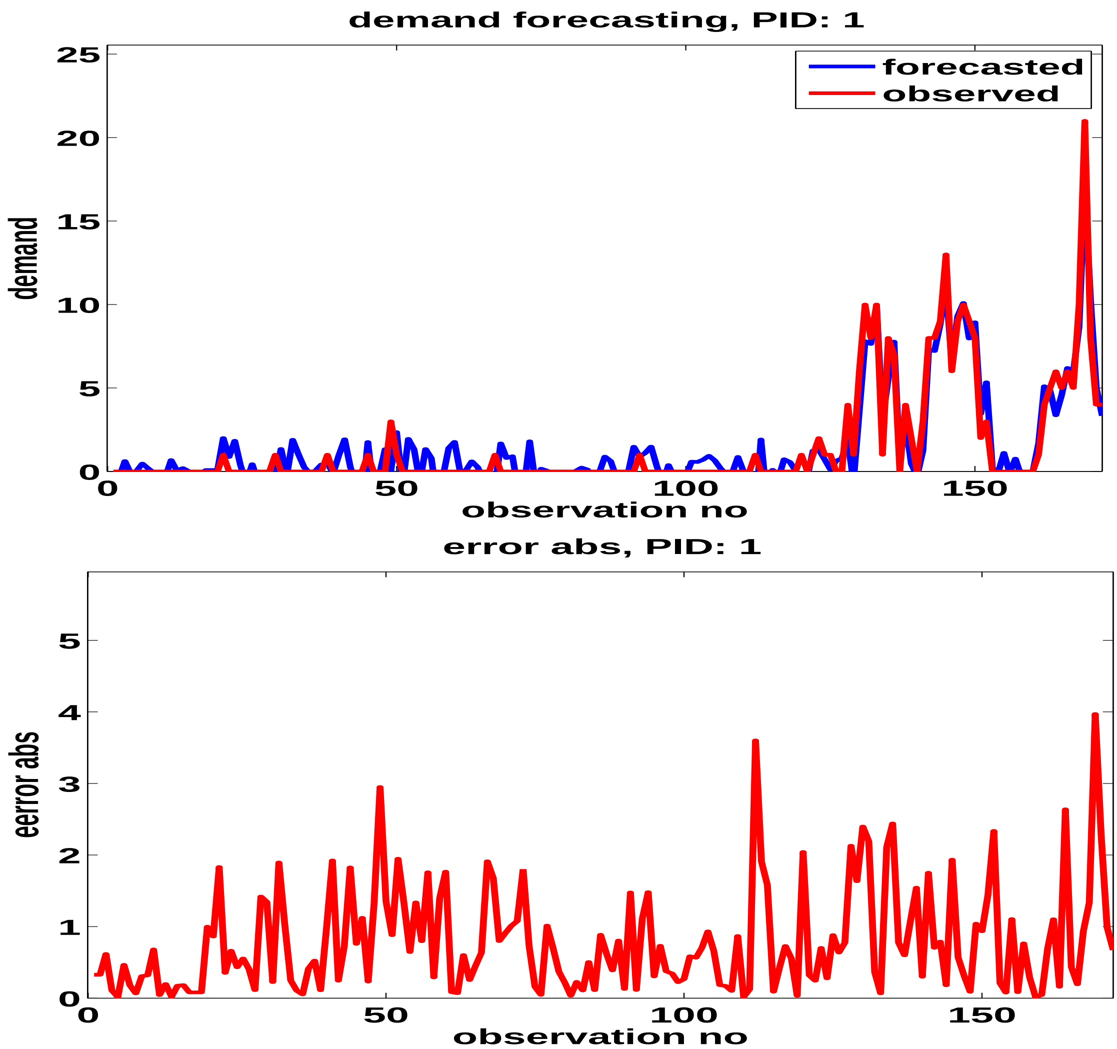

Figure 10: *Top:* The observed demand and the demand predicted by the model for product id 1. *Bottom:* Absolute error between the observed and predicted demand for product id 1.

4 Discrete Choice Models

4.1 Introduction and Basic Set-Up

(4.1) The aim of this approach is to develop a model that, for any one product, will predict the market share of Tesco and each of its competitors, as a function of the prices that they offer for it.

(4.2) Provided competitors' prices can be observed, Tesco may then use this model to set the price of a product, so as (for example) to maximise profit (an optimisation problem) or indeed other objectives that incorporate profit and rules concerning positioning of the brand *etc.*

(4.3) The mathematical set-up is as follows. Suppose there are $i = 1, 2, \ldots, M$ vendors where Tesco is vendor $i = 1$. We suppose that between them, vendors sell products $j = 1, 2, \ldots, N$. (It may be that some vendors do not sell certain products, but this is easily modelled as we shall see shortly.)

(4.4) We let $p_i^{(j)}$ be the price that vendor i charges for one unit of product j, and we let $f_i^{(j)}$ ($0 \leq f_i^{(j)} \leq 1$) be the respective market share, as a proportion of the total sales amongst these vendors. It follows that

$$\sum_{i=1}^{M} f_i^{(j)} = 1 \quad \text{for each } j. \tag{1}$$

(4.5) For each product j, we want to model the market shares $f_i^{(j)}$, $i = 1, 2, \ldots, M$ (outputs) as a function of the prices $p_i^{(j)}$ $i = 1, 2, \ldots, M$ (inputs). We would expect that one's market share increases (resp. decreases) as one's price decreases (resp. increases), and one's market share increases (resp. decreases) if a competitor's price increases (resp. decreases). In other words

$$\frac{\partial f_i^{(j)}}{\partial p_i^{(j)}} \leq 0 \qquad \text{and} \qquad \frac{\partial f_i^{(j)}}{\partial p_k^{(j)}} \geq 0 \quad \text{for } i \neq k. \tag{2}$$

(4.6) There is a well-established topic in behavioural economics known as *discrete choice theory*, which models the proportions of users that make alternative choices as a function of the utility of those choices – in this case, there is a disutility (price) associated with each choice, where the choice is which vendor an individual consumer decides to purchase from.

(4.7) A common approach that produces $f_i^{(j)}$ that automatically satisfy equations (1,2) is to define functions $g_i^{(j)}(p_i^{(j)})$ with

$$\frac{\partial g_i^{(j)}}{\partial p_i^{(j)}} \leq 0, \tag{3}$$

and then define

$$f_i^{(j)}(\mathbf{p}^{(j)}) = \frac{g_i^{(j)}(p_i^{(j)})}{\sum_{k=1}^{n} g_k^{(j)}(p_k^{(j)})}, \tag{4}$$

where $\mathbf{p}^{(j)} := (p_1^{(j)}, p_2^{(j)}, \ldots, p_M^{(j)})$ is the vector of vendors' prices for product j.

(4.8) There are many different functional forms used for $g_i^{(j)}$ in the literature. For illustration in this report, we use the *logit* model where

$$g_i^{(j)}(p_i^{(j)}) = c_i^{(j)} \exp\left(-a_i^{(j)} p_i^{(j)}\right), \tag{5}$$

where $a_i^{(j)} \geq 0$, $c_i^{(j)} \geq 0$ are constant parameters. It follows that for this choice of g, the market shares are given by

$$f_i^{(j)} = \frac{c_i^{(j)} \exp\left(-a_i^{(j)} p_i^{(j)}\right)}{\sum_{k=1}^{M} c_k^{(j)} \exp\left(-a_k^{(j)} p_k^{(j)}\right)}. \tag{6}$$

(4.9) In practice, for each product j there are thus $2M$ parameters $c_i^{(j)}$, $a_i^{(j)}$, for $i = 1, 2, \ldots, M$, but because the model is invariant to scaling all the $c_i^{(j)}$ together, we may assume without loss of generality that

$$\sum_{i=1}^{M} c_i^{(j)} = 1, \tag{7}$$

so that there $2M - 1$ degrees of freedom in the parameters.

(4.10) Note if a vendor i does not offer product j, this might be captured by setting $c_i^{(j)} = 0$, or equivalently by setting $p_i^{(j)} = +\infty$.

(4.11) In layman's terms, the constants $c_i^{(j)}$ capture (something like) the baseline market share for each vendor – this idea will be refined shortly – whereas the $a_i^{(j)}$ model the price sensitivity. For example, Tesco would probably tend to have a large $c_1^{(j)}$ based on its prominent standing and large market share, but perhaps also a relatively high $a_1^{(j)}$, representing that Tesco's customers expect its prices to be very competitive and will divert to other vendors if they do not. In contrast, another vendor i, (say) John Lewis, might have a relatively small baseline market share $c_i^{(j)}$, but also a small $a_i^{(j)}$ – representing that its customers have very strong brand loyalty and are prepared to pay a small amount extra to purchase from what they consider to be a premium service vendor.

(4.12) The challenge that we now have is that there are too many parameters. If each product has a distinct set of $2M - 1$ parameters, then we would have to wait until that particular product launches to observe data for prices and market shares, before estimating those parameters. This is not particularly useful and rather we seek an approach that will *predict* market shares, ahead of a product launch, and so help Tesco set its initial price.

(4.13) To achieve this predictive capability, we shall suppose that all similar products within a given market sector will tend to have parameters that relate to each other – which can thus be estimated from sales of past products. (The estimation of these parameters themselves is dealt with in detail in Section 4.2.)

(4.14) However, even within the same sector (*e.g.* luxury televisions), the baseline (*i.e.* typical) prices of a pair of different products may be enormously different. So it is not sufficient to use the same $\mathbf{c}^{(j)} = (c_1^{(j)}, c_2^{(j)}, \ldots, c_M^{(j)})$ and $\mathbf{a}^{(j)} = (a_1^{(j)}, a_2^{(j)}, \ldots, a_M^{(j)})$ for each product j. Rather, one should use parameters that scale appropriately with baseline prices.

(4.15) Our idea is to suppose that each product j in a given sector has a 'typical' price $p_*^{(j)}$, and then set

$$g_i^{(j)}(p_i^{(j)}) = c_i \exp\left[-a_i \left(\frac{p_i^{(j)} - p_*^{(j)}}{p_*^{(j)}}\right)\right], \tag{8}$$

where the c_i and a_i are now common to *all* products in the given sector, so that they need only be estimated once (from past data).

(4.16) The point of this approach is that we may take two products with very different baseline prices $p_*^{(j)}$, yet model (8) will lead to market shares that change in the same way when the vendors' prices are changed in the same *proportional* way.

(4.17) Note that with this approach, the c_i are exactly the market shares that each vendor secures if all the vendors sell at the baseline price.

(4.18) But how should we define the baseline price? We have in mind that it is a typical price for a given product that is something like the mean price over all vendors. However, since that in turn would depend on vendors' strategy, we propose instead that we define $p_*^{(j)}$ (non-subjectively) to be the base price that Tesco *pays* for the given good. The fact that this price is somewhat lower than that offered to consumers will not matter provided the rule is used consistently over all products.

(4.19) Strictly speaking, note that (8) makes a modelling assumption about consumer behaviour – that their sensitivity is to proportional differences rather than absolute price differences. This assumption requires further testing, but at present is the only way we have of generalising parameters over many products.

4.2 Parameter estimation from data

(4.20) We are now going to explain how to fit the parameters of the discrete choice model using data of the kind that Tesco presented at the study group, plus a couple of supplementary data items/labels that will be explained.

(4.21) Specifically, we suppose that there is (1) a large day-by-day record of Tesco's sales volume for each of a large range of products and the price at which these products were sold at on each day; and (2) there are tables of data that for the same products, give the day-by-day prices of competing vendors, but *not* their individual sales volumes.

(4.22) Note that the data we were presented with at the study group was impaired in that there were relatively few products and a very confined time frame in which tables (1) and (2) overlapped.

(4.23) In consequence, work at the study group focussed on the theoretical development of this parametrisation idea, and if discrete choice modelling is to be put into practice, there will need to be a concerted effort and lead time where more substantial vendor price data is collected (*i.e.* mined from the internet).

(4.24) Supposing that such a large collection of data has been achieved, we will require that the products in it have been classified according to market sector, *e.g.* 1. luxury televisions; 2. large white goods; 3. small kitchen electricals (toasters, kettles *etc.*) – for example. This classification could be performed manually, or via NLP (natural language processing) methods. A thorough study should examine the effect of choosing different product classifications and balancing larger class sizes vs smaller class sizes. In short, large classes reduce statistical sampling error in the parameter estimation process, but may inappropriately group together products where the consumer behaviour is not very similar.

(4.25) From now on, we describe the parametrisation procedure as it applies to a single such class, containing products $j = 1, 2, \ldots, N$ with base prices $p_*^{(j)}$ that, for example, are set equal to Tesco's own base costs $b^{(j)}$ as described above. The parameter estimation problem is to estimate c_i, a_i for $i = 1, 2, \ldots, M$ common to this class.

(4.26) When inspecting the sales data presented at the study group, we noted a strong day-of-week pattern in the sales volume of many products. For the purposes of the method described here, it is necessary to use input data where the overall demand for each product is time-independent. This means that the day-of-week pattern may be dealt with either by parametrising different models for different days of the week – or more simply, we may aggregate sales volumes to weekly totals.

(4.27) A significant problem is that although we observe Tesco's own sales volume for each product, what we really wish to know is what market share this corresponds to. This does not require knowledge of the other vendors' individual sales totals (impossible to obtain) – but it does require us to have an estimate $d^{(j)}$ of the total market size for each product j, measured in the number of sales of units per the given (aggregated) time interval (*e.g.* one week). At the study group, Tesco's representative indicated that they acquire market intelligence data that provides them with this kind of statistic, at least for the major product lines.

(4.28) Tesco's sales volume for any one product j, even aggregated over one week, can be rather small. It is therefore essential to model these sales as a random process in time, with a rate parameter equal to $d^{(j)} f_1^{(j)}$. For any one sales record, $f_1^{(j)}$ is constructed via formulae (4,8), and thus incorporates known parameters that relate to that record (base price $p_*^{(j)} = b^{(j)}$, and prices $p_i^{(j)}$, $i = 1, 2, \ldots, M$ in force at the time of that record) as well as the parameters a_i, c_i that we are trying to estimate.

(4.29) In short, the sales volumes are Poisson distributed with parameter $\lambda = d^{(j)} f_1^{(j)}$, and so the likelihood of any particular record with Tesco sales volume s is

$$L = P(s \,|\, \mathbf{a}, \mathbf{c}) = \frac{\lambda^s}{s} \mathrm{e}^{-\lambda}, \tag{9}$$

where

$$\lambda = d^{(j)} f_1(\mathbf{a}, \mathbf{c};\, \mathbf{p}^{(j)}, p_*^{(j)}). \tag{10}$$

(4.30) The point is that we can now mix together a large number of observations of sales, both over time and over products $j = 1, \cdots, P$ in the same class. Let their likelihoods be L_1, L_2, ..., L_P be computed according to the prescription (9,10) – note that the input parameters $\mathbf{p}^{(j)}$, $p_*^{(j)}$ and $d^{(j)}$ will generally be different for each record, but the parameters $\mathbf{a}$ and $\mathbf{c}$ that we seek are common to all of the records.

(4.31) Assuming independence, we are thus trying to maximise the likelihood $L = L_1 L_2 \cdots L_P$ of all of the observations. In practice, this is most easily achieved by maximising the log likelihood. That is, the best fit parameters are the $\mathbf{a}$ and $\mathbf{c}$ that maximise

$$F := \log L = \sum_{k=1}^{P} \log L_k. \tag{11}$$

This problem can be solved using a numerical black-box optimiser. Finally, note the optimisation should be performed subject to the constraint $\sum_{i=1}^{M} c_i = 1$ so as to eliminate the linear scaling symmetry.

(4.32) Finally, note that confidence estimates of the $\mathbf{a}$ and $\mathbf{c}$ can most easily be obtained by bootstrapping the data. This means that an ensemble of new artifical datasets are created by resampling the real data with replacement. The parametrisation process can then be performed for each element of the ensemble and thus an idea of the sensitivity of **a** and **c** obtained. This can be a computationally expensive procedure but is a fairly standard one in modern Applied Statistics practice.

4.3 Optimisation of profit

(4.33) At this point, we suppose that the parameters **a** and **c** have been estimated for a given product class and that a new product in this class has just launched. How should Tesco set its price in response to the prices of the other vendors that it continuously observes?

(4.34) For simplicity, we suppose that Tesco's core objective is to maximise profit, that is the product of market share and the margin per unit sale. (Perhaps a more realistic objective is to maximise sustainable profit – but this is a complex concept involving human factors such as sustaining brand reputation *etc.*)

(4.35) Built into our problem formulation is the implicit assumption that the total market size $d^{(j)}$ is fixed and if Tesco (say) cuts its price, then it increases its sales by taking market share from other vendors, rather than by increasing $d^{(j)}$. Of course, in practice, it is probable that aggressive price cutting by a major vendor will increase market size. We did not model this effect at the study group and it is beyond the scope of this report; suffice to say, it might be possible to embed this effect in a refined hierarchical discrete choice model – as is done successfully in the transport modelling community where modelling the decision to travel/not to travel sits hierarchically above the choice of travel mode, route *etc.*

(4.36) At the study group, the Tesco representative pressed upon us that they have many corporate agendas other than maxmising profit – for example, they do not want to be the very cheapest vendor, and they wish to avoid driving down overall market prices too far. As these agendas were expressed to us, it seemed that they took the form of constraints upon the optimisation process described here, rather than changes to the objective function. It is trivial to modify what follows by introducing constraints. However, it would also be relatively simple to modify the objective function to incorporate more complex factors in Tesco's corporate agenda.

(4.37) In what follows we constrain the discussion to that of choosing the optimal price for a single product, and maximising the profit derived from that product. Hence we may drop the (j) superscript in order to simplify the notation. The profit (per unit time) derived from this product is thus

$$\mathcal{F} = d(p_1 - b) f_1(\mathbf{p}), \tag{12}$$

being market size times margin per unit sale times Tesco's market share – where b is Tesco's cost price per unit and f_1 is put together via formulae (4,8) – using the parameters $\mathbf{a}$ and $\mathbf{c}$ that have already been estimated from historical sales of products in the same class, and a typical price p_* for this particular product where we recommend the choice $p_* = b$.

(4.38) We seek the optimum choice for p_1 so as to maximise $\mathcal{F}$. This is in fact a very simple optimisation of a scalar function with respect to a single scalar variable. Moreover, it turns out that $\mathcal{F}$ is unimodal in p_1, so it is sufficient to find the single turning point of $\mathcal{F}$ by solving

$$\frac{\partial \mathcal{F}}{\partial p_1} = 0, \tag{13}$$

and this turning point will automatically be the global maximum of $\mathcal{F}$. This problem could be given directly to a numerical optimisation/root finding package, but in fact some analytical progress is possible.

(4.39) Note

$$\frac{1}{d}\frac{\partial \mathcal{F}}{\partial p_1} = f_1(\mathbf{p}) + (p_1 - b)\frac{\partial f_1(\mathbf{p})}{\partial p_1}, \tag{14}$$

and using (4), we have

$$\frac{\partial f_1(\mathbf{p})}{\partial p_1} = \frac{\partial}{\partial p_1}\left(\frac{g_1(p_1)}{\sum_{i=1}^{M} g_i(p_i)}\right), \tag{15}$$

$$= \frac{g_1'(p_1)\sum_{i=2}^{M} g_i(p_i)}{\left(\sum_{i=1}^{M} g_i(p_i)\right)^2}. \tag{16}$$

But using (8), we see that

$$g_1'(p_1) = -\frac{a_1}{p_*} g_1(p_1). \tag{17}$$

(4.40) To solve (13), we can thus combine equations (8,13,14,16,17) and on some re-arrangement obtain

$$\left[\frac{a_1\left(\sum_{i=2}^{M} g_i(p_i)\right)}{p_*}\right] p_1 - c_1 \exp\left[-a_1\left(\frac{p_1 - p_*}{p_*}\right)\right] = \left(1 + \frac{a_1 b}{p_*}\right)\left(\sum_{i=2}^{M} g_i(p_i)\right). \tag{18}$$

Note that the recommended choice $p_* = b$ tidies up the right-hand side a little. But the main point is that $\sum_{i=2}^{M} g_i(p_i)$ does not involve p_1, so it is effectively a constant.

(4.41) Equation (18) thus takes the form

$$\alpha x - \exp(-\beta x) = \gamma, \tag{19}$$

where α, β and γ are constants, and (19) can be solved simply for the optimal price $x := p_1$ by using either numerical methods or Lambert functions if these are built in to one's chosen mathematical package.

5 Integrated forecasting and pricing using a Bayesian surface model

(5.1) In this section we will discuss a Bayesian method for conditioning a sales-rate surface model, where the surface is a function of price and time, that might allow Tesco to:

- 1. Compute a Gaussian probabilistic sales-forecast at the next time for each price.
- 2. Choose which price leads to the maximum expected profit.
- 3. Enable them to use Bayes theorem to update their statistical model of the sales-rate surface.

(5.2) In particular, assume that the rate of sales per unit time, $q(i, n)$, is a matrix indexed by n, a time index, and i, a price index. Increasing n means later in time and increasing i means a higher price.

(5.3) Let $p(q|q_{av}, C, S_n)$ be a multivariate prior Gaussian probability density for the matrix q at the beginning of the sales period n, where q_{av} is the posterior mean vector of q (for all times and prices), C is the posterior covariance matrix of q conditioned on all of the observed sales, S_n, up to, and including, sales at time n. In practice if one chooses the prior covariance so that it has a sparse inverse (known as the precision matrix) and one observes particular sales rates with a known zero-mean error, then the precision matrix remains sparse. This has the advantage that one does not need to invert large matrices (which is usually impracticable) and instead one may use a sparse linear solver, such as the pre-conditioned conjugate gradient method to calculate the updated posterior mean sales-rate surface after each sales period.

(5.4) Suppose that at each time period n a decision is to be made to sell over the period $(n + 1)$ at the price $i(n)$. This could be chosen so that at the next period the expected profit from the sales is a maximum. As the model of the sales-rate surface is Gaussian then the forecast sales probability density will also be Gaussian, this then implies that one chooses the price that maximises the expected profit.

(5.5) After the time period the sales figures are supposed known with a small error. Then, using standard Bayesian updates (similar to a Kalman filter), the probability density over all times and prices can be updated, so that the probability density function is conditioned on the new sales figures too.

The posterior precision matrix is updated by adding contributions at the relevant price, to the diagonal that are the precisions (inverse variance) of the sales measurement errors.

(5.6) At the start, the prior mean could be set so that it includes intuitions about seasonal fluctuations and the idea that generally a higher price means fewer sales (but could lead to a higher profit). See Fig. 11 for an example of an estimated sales-rate surface which includes fluctuations on a yearly time scale.

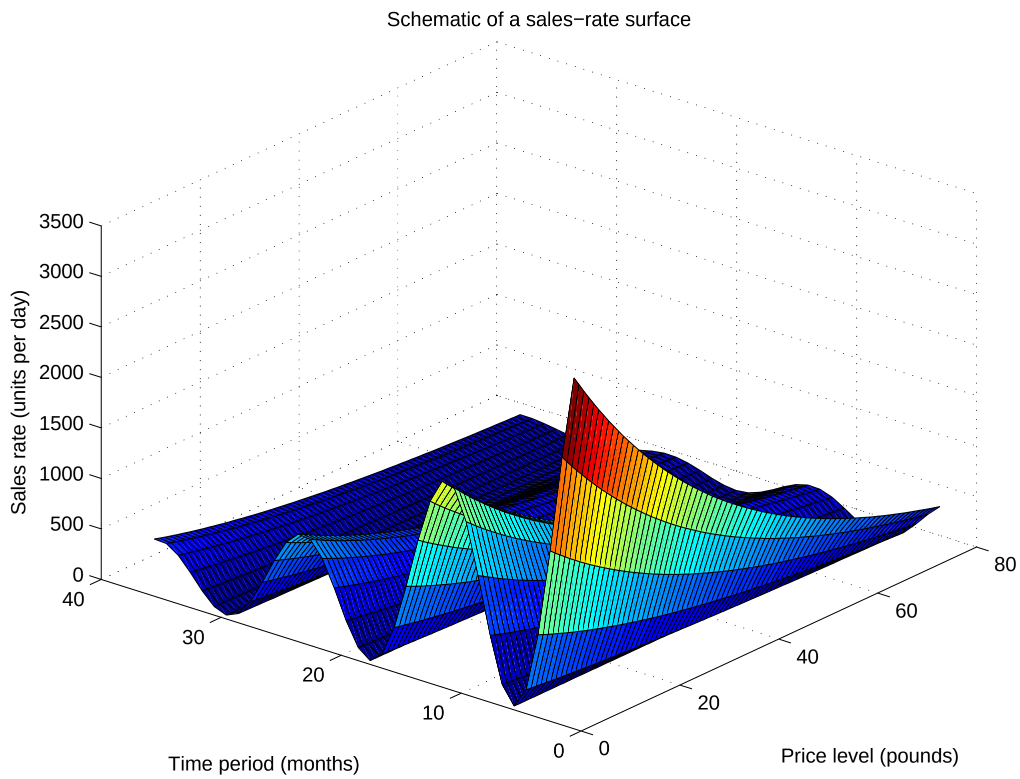

Figure 11: An example of an estimated sales-rate surface.

(5.7) For further background see [6]. For the decision theory background see [4].

6 Stochastic sales rate model

(6.1) In this model we suppose that the rate of sales of a given product is a stochastic process, X_t, which evolves according to the stochastic differential equation

$$dX_t = \alpha(X_t, p_t, t)\,dt + \beta(X_t, t)\,dW_t, \tag{20}$$

where t denotes time, p_t is the price of the product at time t, $\alpha(r, p, t)$ is a function which determines a general trend in sales rate, $\beta(r, t) > 0$ is a function which describes the strength of noise in the sales rate and W_t is a standard Brownian motion which models the noise. We assume that α depends on the price at which we sell the product and that increasing the price will tend to result in a lower rate of sales,

$$\frac{\partial \alpha}{\partial p}(X, p, t) < 0.$$

Allowing α and β to depend on time, t, means that in principle we can build in cyclic and seasonal effects. The function α may be calibrated either by using historical data or by experimentation while the function β is probably best determined from historical data. A full discussion of calibrating the model is outside the scope of this report, however. The chief advantage of this model is that it does not require a knowledge of competitors' costs or sales. The chief disadvantage is that it assumes the sales rate can be modelled by the SDE (20).

(6.2) Over the period $[t, t + dt)$ the profit per unit of product is

$$(p_t - c_t),$$

where c_t is the unit cost (to the retailer) of the product.[1] If the product sells at rate X_t then the total sales over this period is $X_t\,dt$ and the realised profit is

$$(p_t - c_t)\,X_t\,dt.$$

This profit is received in the future so its value should be discounted to obtain the present value, *i.e.*

$$e^{-rt}(p_t - c_t)\,X_t\,dt,$$

where $r > 0$ is the appropriate interest rate which, for simplicity, we take to be a positive constant. The total profit over the product's sales horizon, $T > 0$, is

$$\int_0^T e^{-rt}(p_t - c_t)\,X_t\,dt.$$

As the sales rate X_t and profit are stochastic, the natural quantity to work with is the expected profit,

$$\mathbb{E}\big[\int_0^T e^{-rt}(p_t - c_t)\,X_t\,dt\,\big].$$

[1] In principle, the cost could depend on both time and cumulative sales, $c_t = c(Y_t, t)$ where $Y_t = \int_0^t X_s\,ds$ which would allow us to model economy of scales effects by setting $\partial c/\partial Y < 0$, for instance.

(6.3) In this model the price is regarded as a control variable and is used to maximise the expected profit over the product's sales horizon,

$$\max_{p_t} \mathbb{E}\big[\int_0^T e^{-rt}(p_t - c_t)\, X_t\, dt \big].$$

The simplest case to analyse is where the product's sales horizon is infinite,[2] $T = \infty$, in which case the optimisation problem is

$$\max_{p_t \in P} \mathbb{E}\big[\int_0^\infty e^{-rt}(p_t - c_t)\, X_t\, dt \big], \tag{21}$$

where P is a set of admissible prices, for example if there are price bounds then $P = \{p_{\min} \leq p_t \leq p_{\max}\}$, and the optimal price is

$$p_t^* = \arg\max_{p_t \in P} \mathbb{E}\big[\int_0^\infty e^{-rt}(p_t - c_t)\, X_t\, dt \big]. \tag{22}$$

6.1 Bellman equation

(6.4) A standard approach to this sort of stochastic optimisation problem is to introduce a value function,

$$v(x,t) = \max_{p_t \in P} \mathbb{E}\big[\int_t^\infty e^{-r(s-t)}(p_s - c_s)\, X_s\, ds \,\big|\, X_t = x \big], \tag{23}$$

split the integral and use the tower law of expectations to show that

$$\begin{aligned} v(x,t) &\\ &= \max_{p \in P} \mathbb{E}\Big[\Big(\int_t^{t+dt} + \int_{t+dt}^\infty\Big)\big(e^{-r(s-t)}\,(p_s - c_s)\, X_s\, ds\big) \,\Big|\, X_t = x\Big] \\ &= \max_{p \in P} \mathbb{E}\Big[(p_t - c_t)\, X_t\, dt + e^{-rdt}\int_{t+dt}^\infty e^{-r(s-t-dt)}(p_s - c_s) X_s ds \,\Big|\, X_t = x\Big] \\ &= \max_{p \in P} \mathbb{E}\big[\, (p_t - c_t)\, x\, dt + e^{-rdt}\, v(x + dX_t, t + dt)\,\big] \\ &= \max_{p_t \in P}\Big((p_t - c_t)\, x\, dt + (1 - r\, dt)\, \mathbb{E}\big[\, v(x + dX_t, t + dt)\,\big]\Big). \end{aligned} \tag{24}$$

[2]The finite sales horizon problem requires a terminal condition at T, for example some sort of penalty for unsold inventory or a stopping time condition where we continuing selling until the inventory is exhausted, whereas this sort of complication does not arise in the infinite horizon problem. For reasons mentioned in the following section, using an infinite horizon is not necessarily as restrictive as it may seem at first sight.

If $v(x,t)$ is a sufficiently smooth function then Itô's lemma shows[3]

$$\begin{aligned} v(x+dX_t, t+dt) &= v(x,t) + \frac{\partial v}{\partial t}\,dt + \frac{\partial v}{\partial x}\,dX_t + \frac{\partial^2 v}{\partial x^2}\,dX_t^2 \\ &= v(x,t) + \frac{\partial v}{\partial t}\,dt + \frac{\partial v}{\partial x}\big(\alpha(x,p_t,t)\,dt + \beta(x,t)\,dW_t\big) \\ &\qquad + \tfrac{1}{2}\frac{\partial^2 v}{\partial x^2}\,\beta(x,t)^2\,dt, \end{aligned}$$

where all partial derivatives are evaluated at (x,t). For a Brownian motion we have $\mathbb{E}[dW_t] = 0$, so

$$\mathbb{E}\big[\,v(x+dX_t, t+dt)\,\big] = v(x,t) + \Big(\frac{\partial v}{\partial t} + \tfrac{1}{2}\beta(x,t)^2\,\frac{\partial^2 v}{\partial x^2} + \alpha(x,p_t,t)\,\frac{\partial v}{\partial x}\Big)dt$$

and we find that (24) reduces to the Bellman equation

$$\frac{\partial v}{\partial t} + \tfrac{1}{2}\beta(x,t)^2\,\frac{\partial^2 v}{\partial x^2} - r\,v + \max_{p_t\in P}\Big((p_t - c_t)\,x + \alpha(x,p_t,t)\,\frac{\partial v}{\partial x}\,\Big) = 0. \quad (25)$$

If there are price bounds, say $p_{\min} \le p_t \le p_{\max}$, the problem must be solved numerically to find the optimal price, p_t, and the value function $v(x,t)$.

(6.5) If there are no price bounds then we may proceed by finding[4]

$$\max_{p_t\in P}\Big((p_t - c_t)\,x + \alpha(x,p_t,t)\,\frac{\partial v}{\partial x}(x,t)\,\Big).$$

If we assume the maximum exists and α is smooth enough then we may find it by solving

$$x + \frac{\partial \alpha}{\partial p_t}(x,p_t,t)\,\frac{\partial v}{\partial x}(x,t) = 0, \quad \frac{\partial^2 \alpha}{\partial p_t^2}(x,p_t,t) < 0 \quad (26)$$

to find the optimal price p_t.

6.2 A simple example

(6.6) As an example, consider the simple case where the cost is constant, $c \ge 0$, and the sales rate evolves as

$$\frac{dX_t}{X_t} = -\tfrac{1}{2}\alpha\,p_t^2\,dt + \beta\,dW_t, \quad (27)$$

[3] In practice this amounts to noting that for a Brownian motion $dW_t^2 = dt$ and then performing a Taylor expansion up to and including all terms of $\mathcal{O}(dt)$.

[4] If this function does not have a global maximum, in p_t, we *must* impose price bounds.

where $\alpha > 0$ and $\beta > 0$ are constants. If, as it turns out, the optimal price p_t^* is a positive constant, say $p_t^* = p^* > 0$, then the solution of this SDE is simply

$$X_t = X_0 \exp\bigl(-\tfrac{1}{2}(\alpha p^{*2} + \beta^2)\, t + \beta\, W_t\bigr) \tag{28}$$

and so if $X_0 > 0$ then $X_t > 0$ for all $t > 0$, thus avoiding the possibility of a *negative* rate of sales. The x in $v(x,t)$ means that the expectation is conditioned on $X_t = x$, so it follows that we must take $x > 0$. This will be assumed in all the follows without further comment. The process (28) also has the property that the sales rate will, with probability one, tend to zero as t tends to infinity[5] which means that the product has an effectively finite sales horizon.

(6.7) The optimal price is determined by

$$p_t^* = \arg\max_{p_t} \Bigl((p_t - c)\, x - \tfrac{1}{2}\alpha\, x\, p_t^2\, \frac{\partial v}{\partial x}\Bigr),$$

which has the unique solution[6]

$$p_t^* = \Bigl(\alpha\, \frac{\partial v}{\partial x}\Bigr)^{-1} \tag{29}$$

and the corresponding maximal value is

$$\Bigl(\Bigl(2\alpha\, \frac{\partial v}{\partial x}\Bigr)^{-1} - c\Bigr)\, x.$$

From (25), v satisfies the partial differential equation

$$\frac{\partial v}{\partial t} + \tfrac{1}{2}\beta^2\, x^2\, \frac{\partial^2 v}{\partial x^2} + x\, \Bigl(2\alpha\, \frac{\partial v}{\partial x}\Bigr)^{-1} - r\, v = c\, x.$$

The problem is time-autonomous and has no terminal time, so it is clear that the solution is a function of x alone. When $x = 0$ there are no sales so $v(0) = 0$ and increasing the sales rate x implies increasing value so $v'(x) > 0$. Thus $v = v(x)$ satisfies the problem

$$\tfrac{1}{2}\beta^2\, x^2\, v''(x) + \frac{x}{2\alpha\, v'(x)} - r\, v = c\, x, \quad v(0) = 0, \quad v'(x) > 0. \tag{30}$$

[5]To see this note that we can write $W_t = \sqrt{t}\,\phi$ for some some zero mean unit variance normally distributed $\phi \sim N(0,1)$. For any $\epsilon > 0$ we have

$$\text{prob}(X_t < \epsilon) = \text{prob}\bigl(\log(X_t) < \log(\epsilon) = \text{prob}\bigl(\phi < f(t)\bigr) = \Phi\bigl(f(t)\bigr),$$

where

$$f(t) = \frac{\log(\epsilon/X_0) + \frac{1}{2}(\alpha p^{*2} + \beta^2)t}{\sqrt{\beta^2 t}}, \quad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-q^2/2} dq.$$

For any $\epsilon > 0$, $X_0 > 0$, $\lim_{t\to\infty} f(t) \to \infty$, which gives the result.

[6]Assuming that $\partial v/\partial x > 0$ and $x > 0$. The former is intuitively obvious, the greater the sales rate the greater the expected profits, and may be imposed via boundary conditions. The latter follows from the fact that $x = X_t > 0$.

If we assume a linear solution of the form[7]

$$v(x) = A\,x \tag{31}$$

we find that $v'(x) = A$ satisfies the quadratic equation

$$r\,A^2 + c\,A - \frac{1}{2\alpha} = 0$$

and the positive root of this equation may be written as[8]

$$A = \frac{1}{\left(1 + \sqrt{1+\gamma}\,\right)\alpha\,c}, \tag{32}$$

where

$$\gamma = \frac{2\,r}{\alpha\,c^2} > 0.$$

The optimal price is a constant,

$$p_t^* = \frac{1}{\alpha\,A} = \left(1 + \sqrt{1+\gamma}\,\right)c > 2\,c \tag{33}$$

and, somewhat surprisingly, it is independent of β. With this price and given the current sales rate, X_t, the expected present value of total sales is

$$v(X_t) = A\,X_t.$$

6.3 Simulations

(6.8) It is straighforward to perform a Monte Carlo simulation by using (28) to simulate X_t at a series of (closely spaced) time points and from this to approximate the stochastic integral in (23). Moreover, as we know that with probability one the value of X_t will become arbitrarily small if t is large enough we only have to evaluate the integral out to some large time, rather than infinity. If we do this for many different realisations of X_t we can estimate $v(x,t)$ in (23) and compare the expected profits for various choices of price p.

(6.9) The somewhat arbitrary values of the parameters in the simulations are

$$\alpha = 0.15,\ \beta = 0.20,\ r = 0.02,\ c = 1.00,\ X_0 = 1.00,$$

for which the optimal price and associated expected profit are

$$p^* = 2.1255, \quad v(1) = 3.1366.$$

[7]If we look at the behaviour of solutions of the ODE in (30) in the small x limit, this is the only possible form of the solution consistent with $v(0) = 0$ and so it is unique.

[8]In the case $c = 0$, the positive root is $A = \sqrt{2/\alpha\,r}$ which gives $p_t^* = \sqrt{r/2\alpha}$.

For a given simulation of X_t, the integral is estimated by integrating over $0 \le t \le T$ where $T = 400$, at which point $0 < e^{-rT}X_T < 10^{-10}$ (for all simulations) and continuing further would have negligible effect, using a time step of $dt = 0.05$. The process is repeated for 100,000 realisations of the process X_t and from these simulations we construct an approximation to the probability density function for the profits and also compute the expected profit.

(6.10) The effect of prices p with $c < p < p^*$, $p = p^*$ and $p > p^*$ on the distribution of profits and the expected profit are shown in the following three figures, respectively. Note also that when $p \neq p^*$ not only is the expected profit smaller, but so too is the median profit. The mass in the right tail of the distributions is decreasing in p, as is the spread of the distributions. These effects arise because in this model increasing p results in the sales rate becoming neglible sooner.

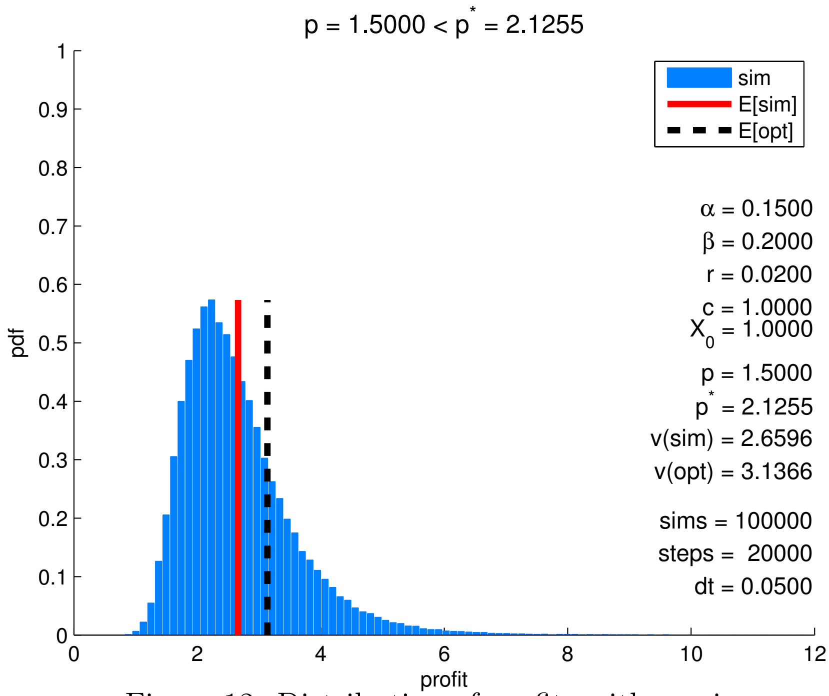

Figure 12: Distribution of profits with a price $c < p < p^*$

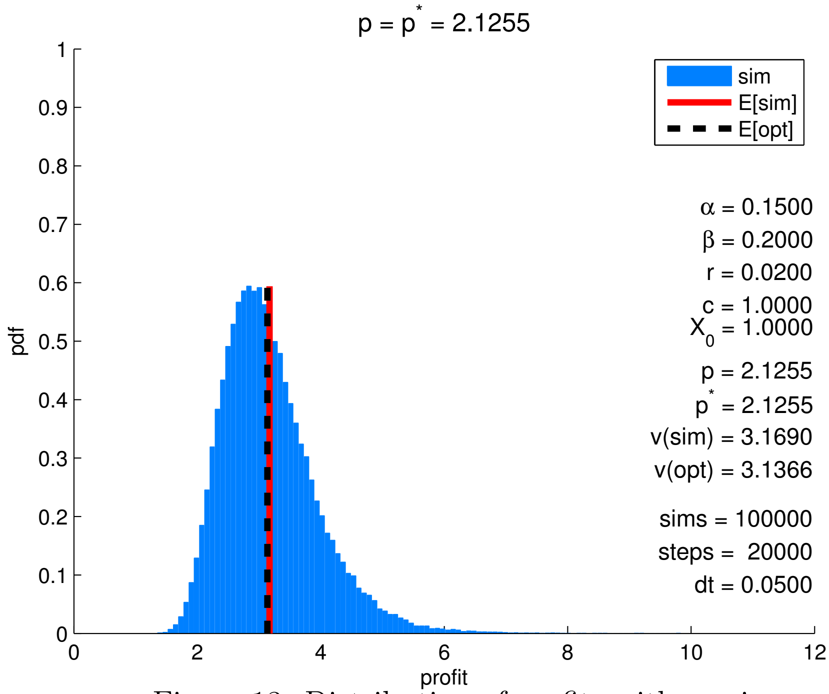

Figure 13: Distribution of profits with a price $p = p^*$

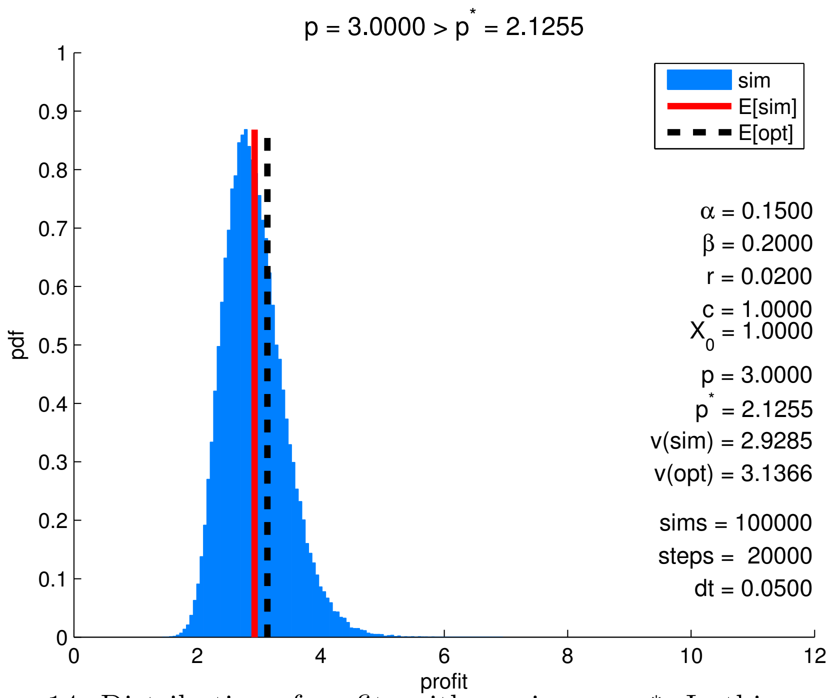

Figure 14: Distribution of profits with a price $p > p^*$. In this case, one might argue that the small loss of expected profit is an acceptable price to pay for the smaller spread in the distribution of profits.

7 Continuum of products model

(7.1) In this deterministic model we consider a set of (possibly related) products indexed by $x \in [0,1]$. For each product x we let

$$\begin{aligned} c(x) \geq 0 &\quad \text{be the cost of the product (to the retailer),} \\ \alpha(x) \geq 0 &\quad \text{be the `intrinsic' demand for the product,} \\ p(x) \geq 0 &\quad \text{be the price of the product,} \\ Q(x,y) &\quad \text{be the impact of the price of } y \text{ on the demand for } x. \end{aligned}$$

The realised demand for product x is

$$q(x) = \alpha(x) - \int_0^1 Q(x,y)\, p(y)\, dy$$

so the 'intrinsic' demand may be taken to be the demand given that all prices are zero. Given that we must have $q(x) \geq 0$, α, p and Q must be chosen so that

$$\alpha(x) \geq \int_0^1 Q(x,y)\, p(y)\, dy \quad \forall\, 0 \leq x \leq 1. \tag{34}$$

We do not assume that $p(x) \geq c(x)$ as product x might be a loss leader. We can assume that the products are ordered in such a way that $c(x)$ is not decreasing in x.

(7.2) The total profit from sales of the product is

$$P = \int_0^1 q(x) \big(p(x) - c(x)\big)\, dx$$

and the aim is to choose the prices $p(x)$ that maximise the profit,

$$P^* = \max_p \int_0^1 q(x) \big(p(x) - c(x)\big)\, dx$$

given the costs, $c(x)$, intrinsic demands, $\alpha(x)$ and price impacts $Q(x,y)$. This may be written as

$$P^* = \max_p \int_0^1 \int_0^1 \big(\alpha(x) - Q(x,y)\, p(y)\big) \big(p(x) - c(x)\big)\, dy\, dx. \tag{35}$$

(7.3) Let us assume there is some optimal price function, $p^*(x)$, which achieves the maximum profit. If so, we can use the following standard calculus of variations argument to find it. Let $\epsilon \ll 1$ and $\epsilon\phi(x)$ be a small perturbation to this optimal solution and write

$$P(\epsilon) = \int_0^1 \int_0^1 \Big(\alpha(x) - Q(x,y)\, \big(p(y) + \epsilon\, \phi(y)\big)\Big) \big(p(x) + \epsilon\phi(x) - c(x)\big)\, dy\, dx.$$

Assuming $P(\epsilon)$ is sufficiently differentiable in ϵ, if $\epsilon = 0$ corresponds to a maximum then we must have

$$P'(0) = 0, \quad P''(0) \leq 0.$$

We find that

$$\begin{aligned} P'(0) &= \int_0^1 \int_0^1 \phi(x)\big(\alpha(x) - Q(x,y)\,p(y)\big)dy\,dx \\ &- \int_0^1 \int_0^1 \phi(y)\,Q(x,y)\big(p(x) - c(x)\big)dy\,dx \\ &= \int_0^1 \int_0^1 \phi(x)\big(\alpha(x) - Q(x,y)\,p(y)\big)dy\,dx \\ &- \int_0^1 \int_0^1 \phi(x)\,Q(y,x)\big(p(y) - c(y)\big)dy\,dx \\ &= 0. \end{aligned}$$

As this is true for any (reasonable) function ϕ, we must have

$$\int_0^1 \Big(\alpha(x) - Q(x,y)\,p(y) - Q(y,x)\big(p(y) - c(y)\big)\Big)dy = 0$$

for each $x \in [0,1]$. This implies that the price function which maximises profits satisfies the relation

$$\alpha(x) = \int_0^1 \Big(Q(x,y)\,p(y) + Q(y,x)\big(p(y) - c(y)\big)dy\Big), \quad x \in [0,1], \tag{36}$$

which may also be written as

$$\int_0^1 \big(Q(x,y) + Q(y,x)\big)p(y)\,dy = \alpha(x) + \int_0^1 Q(y,x)\,c(y)\,dy, \quad x \in [0,1].$$

Setting

$$K(x,y) = Q(x,y) + Q(y,x), \quad f(x) = \alpha(x) + \int_0^1 Q(y,x)\,c(y)\,dy,$$

this becomes a Fredholm integral equation of the first kind for the optimal price function,

$$\int_0^1 K(x,y)\,p(y)\,dy = f(x), \quad x \in [0,1], \tag{37}$$

with a symmetric kernel, $K(x,y) = K(y,x)$.

(7.4) It is easy to see that $P''(\epsilon)$ is independent of ϵ. It is given by

$$P''(\epsilon) = -2 \int_0^1 \int_0^1 \phi(y)\, Q(x,y)\, \phi(x)\, dy\, dx.$$

Therefore, the solution of (37) maximises profits only if the price impact function is positive semi-definite, in the sense that

$$\int_0^1 \int_0^1 Q(x,y)\, \phi(x)\, \phi(y)\, dy\, dx \geq 0 \quad \forall\ \phi. \tag{38}$$

This, of course, means that the symmetric kernel $K(x,y)$ is also positive semi-definite. Together with (34) this implies that

$$\int_0^1 p(x)\, \alpha(x)\, dx \geq \int_0^1 \int_0^1 p(y)\, Q(x,y)\, p(x)\, dy\, dx \geq 0.$$

From (35) and (36) we find that the maximum profit is given by

$$P^* = \int_0^1 \int_0^1 \big(p(y) - c(y)\big) Q(x,y) \big(p(x) - c(x)\big) dy\, dx \tag{39}$$

so if $Q(x,y)$ is *strictly* positive definite then this will always be positive provided that $p(x)$ is not identically equal to $c(x)$. If $p(x) = c(x)$ for all x then (36) implies that

$$\alpha(x) = \int^1 Q(x,y)\, p(y)\, dy.$$

Therefore if the inequality in (36) is strict, *i.e.*

$$\alpha(x) > \int_0^1 Q(x,y)\, p(y)\, dy,$$

and $Q(x,y)$ is strictly positive definite then the maximum profit must be positive. If $Q(x,y)$ is strictly positive definite then, of course, so too is $K(x,y)$.

7.1 Discrete version of the model

(7.5) Divide the interval $[0,1)$ into n equal subintervals

$$[0, 1/n),\ [1/n, 2/n),\ \ldots\ ,\ [1 - 1/n, 1)$$

and set $c(x)$, $\alpha(x)$ and $p(x)$ to be constants on each subinterval, say

$$c(x) = c_i,\ \alpha(x) = \alpha_i,\ p(x) = p_i \quad \text{for } x \in [(i-1)/n, i/n),$$

where $i = 1, 2, \ldots, n$. Similarly, divide the unit square into an $n \times n$ chess board of squares of the form

$$[(i-1)/n, i/n] \times [(j-1)/n, j/n],$$

where i and j range from 1 to n and set

$$Q(x,y) = Q_{ij} \quad \text{for} \quad (x,y) \in [(i-1)/n, i/n] \times [(j-1)/n, j/n].$$

If we set

$$K_{ij} = Q_{ij} + Q_{ji}, \quad f_i = \alpha_i + \sum_{j=1}^{n} Q_{j,i}\, c_j$$

then (37) reduces to the linear system, with symmetric matrix $K_{ij} = K_{ji}$,

$$\sum_{j=1}^{n} K_{ij}\, p_j = f_i, \quad i = 1, 2, \ldots, n \tag{40}$$

and condition (38) becomes the condition that the matrix Q_{ij} is positive semi-definite,

$$\sum_{i,j=1}^{n} y_i\, Q_{ij}\, y_j \geq 0$$

for all n-vectors (y_i), and implies that the symmetric matrix (K_{ij}) has the same property.

8 Designing an objective function and formulating the optimisation problem

8.1 Introduction

(8.1) The aim in this section of the report is to try and formulate a model that optimises the price of a product based on the welfare of Tesco and the welfare of the customers, using an optimal control approach.

(8.2) We will use the following notation.

$$\begin{aligned} x(t) &= \text{ number of sales of the product at time t} \\ p(t) &= \text{ price for the product at time t} \\ c &= \text{ what Tesco paid for the product} \end{aligned}$$

(8.3) For the model to be implemented a number of choices have to be made and a number of factors have to be taken into account. The model needs the following three inputs before it becomes operational.

1. A function of t, $x(t)$ and $p(t)$ which describes the welfare of Tesco and the customers.

2. A function which describes how sales depend on the price of the product.
3. An interval in which the price of the product is allowed to vary.

In the remainder of this introduction we will describe these three inputs to the model in turn.

(8.4) First, the welfare function. Tesco has expressed their interest in taking into account both the welfare of Tesco and the welfare of the customers when choose prices. There is a choice involved in how to measure these two things and how to combine them in a way that reflects what Tesco value most. One suggestion is to proceed as follows: Let W_T denote the welfare of Tesco, W_C the welfare of the customers and $\mathcal{L}$ the combined welfare. For W_T we will use the revenue from selling the product. For W_C we use the sales. If customers buy the product, they must be happy with it. This leads us to the following expressions:

$$W_T(t) = (p(t) - c)\, x(t)$$
$$W_C(t) = x(t)$$

(8.5) For the combined welfare we could multiply them together.

$$\mathcal{L}(x(t), p(t), t) = W_T \cdot W_C = (p(t) - c)\, x(t)^2.$$

(8.6) Note that this is just one among many choices. Which choice one will go with depends on exactly what Tesco wishes to optimise. An alternative welfare function could be

$$\mathcal{L}(x(t), p(t), t) = \alpha \cdot W_T(t) + \beta \cdot W_C(t) \quad \text{for} \ \ \alpha + \beta = 1.$$

where $\alpha, \beta \in \mathbb{R}$ would have to be chosen depending on what Tesco consider more important, W_T or W_C.

(8.7) Now consider the second item on the list, the function relating sales to price. The dynamics of the system is controlled by a function describing changes in sales as a function of time, the current sales and the current prices, that is

$$\dot{x}(t) = f(x(t), p(t), t).$$

The function f depends on the product in question and would have to be estimated by studying the sales history. Carrying out this analysis may very well show that a general function can be chosen for all products, but this can only be concluded after conducting an experiment on a representative sample of the products.

(8.8) Finally we look at the third item on the list, the price interval. This model will find the price $p(t)$ at time t of the product within a predefined price interval $[p_1, p_2]$. This interval is chosen by Tesco based on the following parameters

1. Company policy (online shop cheaper than stores, *etc.*).
2. Legal considerations.
3. Competitors' prices for the same product. For instance, it may be company policy not to be the bottom of the market.

8.2 The objective function

(8.9) We are now ready to define the objective function, which measures the welfare of Tesco and the customers over some chosen time interval between now (time zero) and some later time T. Hence the objective function becomes

$$O(x(t), p(t)) := \int_0^T \mathcal{L}(x(t), p(t), t)\, dt.$$

We wish to maximise this function over the timespan $t \in [0, T]$ subject to the constraints

- $\dot{x}(t) = f(x(t), p(t), t)$
- $p(t) \in [p_1, p_2]$
- $x(0) = x_0$ known and $x(T)$ free.

(8.10) We have chosen to formulate the problem this way, because it then becomes an optimal control problem. The advantage with turning the problem into an optimal control problem is that there is a rich theory for how to solve this type of problem. In the following subsection we will summarise how Tesco could actually solve the problem and find functions $x(t)$ and $p(t)$ which would predict sales and choose the optimal price for the product based on the market and the welfare of Tesco and the customers.

8.3 How to optimise the objective function

(8.11) Let us assume that Tesco has found/decided on the following three inputs.

- $\mathcal{L}(x(t), p(t), t)$
- $f(x(t), p(t), t)$
- the price interval $[p_1, p_2]$.

(8.12) Then we are ready to solve the problem. First, we must make two definitions. If $z(t)$ is a continuous function defined on $[0, T]$ then the **Hamiltonian** is

$$H_0(x(t), p(t), z(t), t) = \mathcal{L}(x(t), p(t), t) + z(t) f(x(t), p(t), t).$$

(8.13) Note that the solution depends on the functions $\mathcal{L}$ and f so we cannot be too explicit before they have been chosen.

i) Suppose $(x^*(t), p^*(t))$ is an optimal solution, which we do not currently know. Then we can use Theorem 8.1 to give us conditions on $x^*(t)$ and $p^*(t)$, which allow us to work them out. In some cases we will get explicit analytic expressions for the functions and in other cases we will get numerical solutions.

ii) If the Hamiltonian is concave in the variables $x(t)$ and $p(t)$ then the solution $(x^*(t), p^*(t))$ we have found will maximise the objective function.

(8.14) To actually determine the functions $x^*(t)$ and $p^*(t)$ and to check if they maximise the objective function we will use the following theorem, which is essentially Pontryagin's Maximum Principle, and its corollary.

Theorem 8.1. *Suppose $(x^*(t), p^*(t))$ is an optimal solution to the problem. Then there exists a continuous function $z(t)$ defined on $[0, T]$ such that*

i) $p^*(t)$ optimises the Hamiltonian function, that is

$$H(x^*(t), p^*(t), z(t), t) \geq H(x^*(t), p(t), z(t), t) \quad \forall\ p(t), t$$

Note that p^* need only be piecewise continuous and p^* need not satisfy the inequality at the points of discontinuity.

ii) We have $\frac{\partial H}{\partial x} = -\dot{z}(t)$.

iii) $z(T) = 0$.

Corollary 8.2. *If the Hamiltonian is concave in the variables $x(t)$ and $p(t)$, then the optimal pair $(x^*(t), p^*(t))$ we found using the theorem maximises the objective function.*

8.4 To think about

(8.15) We finish by mentioning a few of the concerns regarding this model and emphasising some important points relating to implementing this model.

1. It is probably a good assumption to let customer welfare, W_C, at time t simply be expressed by $x(t)$, and it is probably a good assumption to let Tesco welfare, W_T, be the profit from sales at time t, $p(t) \cdot x(t) - cx(t) = (p(t) - c)x(t)$. What is important to notice is that this is a choice and not a mathematical problem.
2. The relationship between $\dot{x}(t)$ and $p(t)$ should be some smooth function obtained via some mathematical software derived from analysing the sales history, where you look for a smooth function which expresses the derivative of sales as a function of time, current sales and current prices. Three examples are Matlab, Mathematica and Maple.
3. In this model choosing the right price interval is important, because this is the only part of the model where the rest of the market, legislation, and company policy is being incorporated into the model.

4. If the market conditions change, then Tesco would have to update the price interval (or the function f, but this is less likely to be necessary) and then run the optimisation again.
5. Tesco could consider formulating the model discretely instead of continuously (as done here). This would have the advantage of simplifying the computations but would have the disadvantage of limiting the flexibility and accuracy of the model.
6. A disadvantage with this model is that you have to re-run the optimisation every time the market changes.On the other hand, the optimisation is not computationally expensive and can be automated when the algorithm has been put in place.
7. Another thing to be aware of is that it may require people with a fair bit of mathematical understanding to set up the model properly. Maintaining it should not be very demanding though.

9 An approach involving Optimal Transport

(9.1) In this section we address another model for making pricing decisions. We consider Tesco as a monopolist transacting business with a field of anonymous customers (agents) whose preferences are known only statistically. This model comes from a recently developed Mathematical Theory called 'Optimal Transport', which is extremely active and successful.

(9.2) We believe this model may be useful for Tesco because the data analysis and demand forecast of Sections 2 and 3 suggest ways to get the statistical information needed. Since it is easily computable, we suggest that it is worth being considered and possibly compared with another model. There are more sophisticated variants and adaptations which may be considered: we introduce here only the elementary one. We finally mention the relation with the 'Discrete Choice Method', which is very popular in economics.

9.1 Description of the model

(9.3) We would like to make an optimisation which includes:

- The individuality of customers. Because of loyalty programs or just from the registration, Tesco possesses several sources of data from which it could deduce what an individual customer prefers.
- That each customer takes into account the prices before choosing the product to buy within a category. If the preferred product is not a good buy, they may choose some other similar product in the category which is a better deal.

Although our model does not directly include the influence of competitors, it is included indirectly through the values of parameters which are estimated statistically.

(9.4) The mathematical model is defined by

- The customers: they are modelled with a measure space M^+ on which a probability measure $\mu^+(dx)$ describes the relative frequency of customers of type $x \in M^+$.
- The products: they are modelled with a measure space M^- on which a probability measure $\mu^-(dy)$ describes the relative frequency of products of type $y \in M^-$.
- The prices: they are modelled by a function $p : M^- \to \mathbb{R}$.

One can choose either discrete or continuous probability measures, depending on the preference. The probability μ^- is related to the demand or to stock levels, and it may depend on the cost function; in this first approach we just suppose that it is given, for example by statistics. Other things that the company and the environment should set are:

- An index $b(x, y)$ of the benefit that product $y \in M^-$ produces to customer $x \in M^+$. This can be derived knowing the history of what customer x buys, or with questionnaires, or in a smarter way.
- The cost $c : M^- \to [0, +\infty]$ of producing each product $y \in M^-$.
- The allowed intervals of prices, which are modelled by a linear constraint of the form $L(p(y)) \leq 0$.

Example The benefit that a product yields to a category of customers could be estimated from data coming from the physical stores, ratings, surveys, other models, ...

(9.5) Once the model of Paragraph 9.4 is set, we can formulate a strategy for assigning a price to each product, that will be sold to whichever customer chooses to buy it. The aim is here maximising the profits, but we consider below a variant where we would like to maximise the welfare.

(9.6) **Step 1.** Each customer chooses the products to buy within a category:

$$y_{b,p}(x) = \text{argmax} \left\{ b(x, y) - p(y) \;\middle|\; y \in M^- \right\}$$

will be the product that the customer of type $x \in M^+$ chooses given the price function p and the benefit function b. This expresses that if I would like to buy a pen, and my favorite brand is much more expensive than a second brand that I also like, then I will choose to buy the second brand. We discuss now two possibilities for the second step.

(9.7) **Step 2: Maximising the profits.** From the first step, given a price function we know what customers are going to buy. This allows us to compute the profits that a price function brings, given by the real number

$$\int_{M^+} [p(y_{b,p}) - c(y_{b,p})] \, d\mu^+(x).$$

We can therefore maximise them: we are looking for the price function

$$\operatorname{argmax}\left\{\int_{M^+}[p(y_{b,p})-c(y_{b,p})]\,d\mu^+(x)\;\Big|\;p\;:\;p(y_\emptyset)\text{ fixed}, L(p)\leq 0\right\}.$$

(9.8) We shall also take into account the constraint due to limited stocks: Tesco cannot sell more products then it has, even if the customers would like to have more of it. This can be formulated mathematically with the condition on the push forward measure

$$(y_{b,p})_\sharp\mu^+\leq\mu^-,\qquad\text{which means }\mu^-(B)\geq\mu^+\big((y_{b,p})^{-1}(B)\big)\text{ for }B\subset M^-$$

where $\mu^-(dy)$ describes the availability of the product $y\in M^-$. For simplicity we do not discuss this detail here.

(9.9) **Variant of Step 2: Maximising welfare.** Suppose one does not only want to maximise the profits, but Tesco would like to set the best price for the customer within its margin of freedom. Then one could set a welfare function

$$w(x,p(y_{b,p}(x)))$$

which is a combination of:

- The gain for the producer. We assume that it is proportional to the price of products, it is given by $\int p(y_{b,p}(x))d\mu^+(x)$. This integral is equal to $\int p(y)d\mu^-(y)$ if $\mu^-=(y_{b,p})_\sharp\mu^+$.
- The satisfaction of the client, described by a function $D(x,p(y))$. One should make some choices for these objective functions. Again, they may come from statistical data, surveys, other models, ... We refer to Section 8 for more considerations.

(9.10) As when maximising the profits, even if it is a monopolist market at this stage, a good function $D(x,p(y))$ measuring the satisfaction of the client for a given set of prices would take into consideration the global situation of the market. Unless there are models in the literature discussing how to take this D, this satisfaction function could also be an empirical function coming from a statistical analysis of the data.

(9.11) The goal would then be to maximise

$$\int_{M^+}\left\{w(x,p(y_{b,p}(x)))d\mu^+(x)\Big|\;p\;:\;p(y_\emptyset)\text{ fixed}, L(p)\leq 0\right\}$$

9.2 An important remark

(9.12) The models described above would need a more careful mathematical formulation and analysis. For example, why should an optimal choice exist? There are assumptions on the functions b,c – or w in the case one would like to maximise the welfare – that would ensure it (see [5]).

(9.13) It is not essential that there is a unique maximiser for the optimisations in Paragraphs 7, 9, because each maximiser yields the same profits/welfare and therefore one can just pick one of them. It may be important in Step 1 where a customer chooses what to buy. If there are more optimal choices for this first optimisation, one may introduce a second selection criteria, for example the product among the maximiser which was bought more times in total in the past history of the customer x.

9.3 Computability

(9.14) There is a very interesting analysis of the general models, which can provide also smart computational tools. In the specific case of Tesco, when there are discrete measures describing products and customers, then all the steps reduce to finding the maximum component of suitable vectors or matrices: the size shown by the data provided by Tesco show that even a high level software like Matlab can perform these operations effectively.

9.4 Comparison with the 'Discrete Choice Method'

(9.15) We did not introduce above the dual formulation of the Optimal Transport models that were formulated. References for this are for example [7, 8, 1]. Discrete choice models are a kind of stochastic perturbation of the above optimal transport problem in its dual form and they provide a variational argument for the existence and equilibria of equilibrium prices in such models. For this link, refer to Section 3.2 in [2], attributed to Roberto Cominetti.

Bibliography

[1] A users guide to optimal transport, L. Ambrosio, N. Gigli. Lecture Notes in Mathematics 2013, pp. 1–155, Springer.

[2] Lecture Notes Optimal Transportation and Economic Applications, G. Carlier, IMA, New Mathematical Models in Economics and Finance, June 2010

[3] Clustering longitudinal profiles using P-splines and mixed effects models applied to time-course gene expression data', N. Coffey, J. Hindeb, and E. Holian, Computational Statistics and Data Analysis **71**, 14 (2014)

[4] Making Decisions (2nd Edition). D.V. Lindley, Wiley, 1985.

[5] Five lectures on optimal transportation: geometry, regularity and applications, R. McCann, N. Guillen. In *Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the Seminaire de Mathematiques Superieure (SMS)* Montreal 2011. G. Dafni *et al.*, eds. Providence: Amer. Math. Soc. (2013) 145-180.

[6] Gaussian processes for machine learning. C.E. Rasmussen and C. Williams, MIT Press 2006. This book and other related material can be found at the website `http://www.gaussianprocess.org`.

[7] Topics in Optimal Transportation, C. Villani. Providence, RI American Mathematical Society, AMS, Graduate Studies in Mathematics vol. 58, 2003.

[8] Optimal Transport: Old and New, C. Villani. Volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.