Engineering and Physical Sciences Research Council

Industrially Focused Mathematical Modelling

EPSRC Centre for Doctoral Training in Industrially Focused Mathematical Modelling

Demand Transfer and Halo Effect on a Small Range of Products

Yu Tian

Contents

1 Introduction **1**

Background 1

Glossary of terms 1

2 Models for sales **2**

Availability of products 2

Additive model for sales 2

Multiplicative model for sales 3

3 Results **3**

Fit in each store 3

Fit in all stores 4

Cross validation 5

4 Discussion, conclusions, & recommendations **5**

5 Potential impact **6**

References **6**

1 Introduction

Background

How should retailers manipulate the range of products on display? How will the demands of products change if they remove or add some of the products? Will customer buy more pasta if pasta sauce has a discount? If so, how much influence will it have? The answer to these questions is not trivial, but of significant interest to retailers and relevant researchers.

Tesco is the UK's largest grocery chain with over 27% of market share in 2018. The sector has a huge turnover but relative small margins which together make the profitability very sensitive to small changes of products. If a potential customer visits a store and finds that it has a very limited range, they are likely to shop elsewhere, which leads to the belief that stores should stock the greatest diversity of products possible. However, the stock has to be sourced, distributed, and handled in stores, which increases the costs. A more optimal approach requires a balance between increasing the range and increasing the costs, and this is where understanding the fundamental relationships between products plays an important role.

Halo effect and demand transfer describe relationships between products

Two important relationships are the *Halo effect* and *demand transfer*. In the retail context, the term Halo effect describes situations where increasing the sales of a particular item, for example via a promotion, will translate into increasing sales of other items, due to their effective dependence and associations. For example, if we reduce the price of a brand of cheese, we expect to sell more cheese, however, we may also sell more bread and crackers. The term demand transfer describes cases when sales of other products increase following an item being removed, or significantly increasing in price, as customers consciously or unconsciously choose a substitution for that item.

Tesco provided anonymised data for four products spanning a two-year period, along with other items customers bought in the basket. These products are chosen in the same *category*, are *ranged* separately, and have low *seasonal effects*.

Our aim is to develop simple mathematical models for sales of products that incorporate both the halo effect and demand transfer, in order to provide insights into explaining and predicting variations in the sales. Moreover, we aim to uncover intrinsic relationships between products, and provide insights on the effects of adding or removing a product.

Glossary of terms

- **halo effect:** increasing sales of one product, for example via a promotion, translating into increasing sales of other products, due to their dependence or association.
- **demand transfer:** sales of one product increasing following an item being removed, or significantly increasing in price, because customers consciously or unconsciously choose a substitution for it.
- **category:** a type of product, such as eggs, which can have different sizes, e.g. "large" and "medium", and different brands, e.g. "The Happy Egg Company" and "Tesco".
- **range:** the products are in range, when they are in stock. The term "ranged differently" means that the products are not always in stock at the same time.
- **seasonal effect:** the change of sales of a product with respect to seasons, such as ice cream which sells much more in the summer than in the other seasons.
- **homogeneous Poisson processes:** a frequently used model for customer arrival time, because of its mathematical tractability.
- **regimes:** cases when different combination of products are available, or in store.
- **stationarity:** an important property of a time series, where the expected value and variation from this value do not change with time

2 Models for sales

Availability of products

A key issue facing the stores is the availability of products. However, given we do not have exact information about whether the products are available in store, we will use information about customer purchases to infer information about the availability of products, and build a flag that indicates when the product is available.

We are interested in weekly data, and so we aggregate the sales data into weekly bins. We first assume our products are either available in a week or not, and leave the case when products become unavailable during the week to future analysis. Thus, we assume that products are definitely available in the weeks with positive sales, and our problem is to determine whether zero sales of a product occur because of unavailability or because no customer makes a purchase during the period.

The availability of products is obtained using a homogeneous Poisson process model for customer arrival time

We further assume our customers are independent of each other and do not buy products at the same time, and we use a homogeneous Poisson process [1] to model the time when each customer visits the store. We further assume each customer who visits the store will make a purchase. Fitting the process with the transaction data provided, we then compute the probability that no customer visits during the zero sales period. We label a product as unavailable in a particular week if the product has zero sales in the week, and the probability that no customer arrives during the zero sales period including this week is less than a critical value (we chose 0.05 here); otherwise, we label the product as available in this week. The final version of our availability flag for the four particular products is shown as coloured triangles in Figure 1; we see that it is consistent with the sales data.

We split the data into regimes, defined to separate the weeks when different combination of products are available. For example regime 1110 represents the case when products A,B,C are available but not product D, and weeks 1 to 7 in Figure 1 all belong to this regime. It is clear in Figure 1 that weeks of the same regime are not always contiguous, for

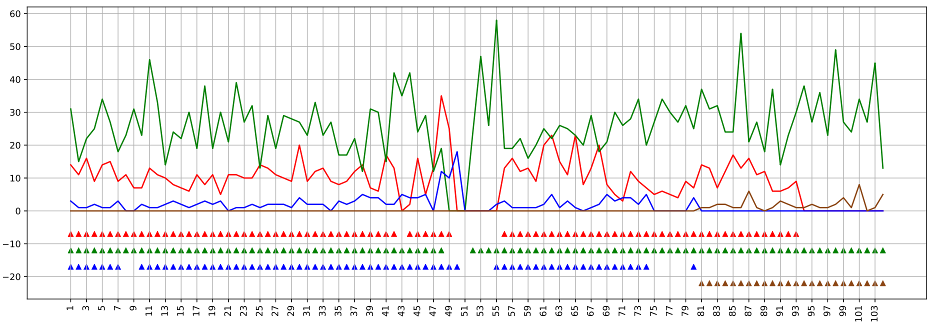


Figure 1 – Graph showing the sales data for four products in store 1 for 103 weeks, with products A,B,C,D shown in red, green, blue and brown respectively, the broken lines in different colours indicating the weekly sales of the corresponding products, and the triangles in different colours showing the availability of the corresponding products

example, product C is not available in weeks 8 and 9, but the previous regime continues from week 10. However, for each store, if we consider each regime separately, there is evidence that the total sales of all available products is *stationary* if the overall length of the regime (i.e. the total number of weeks the regime appears) is longer than 21 weeks, in our two-years-worth of data. Hence, the regime, or the availability of products, is a reasonable factor to use in our models, and the problem now is how this variable works to influence the sales. so as

Additive model for sales

Our first approach is to model the effect of introducing a previously unavailable product, or removing a previously available product, by adding or subtracting the corresponding quantities to the sales of other products. For example, if a store introduces product D, the sales of product A will increase/decrease by some quantity in our model. If the sales of product A increase, then there is a potential halo effect between products A and D, while if the sales decrease, demand transfer is potentially occurring between them.

We assume that, at a weekly level, changes in the expected values of sales of products can be explained by the availability of other products, and other variations can be explained by noise with a *Gaussian distribution* (a common probability distribution that fits the data best). Such changes are known as *structure changes*, and we use our additive model to explain them. We also assume linear dependencies between the sales of one product on the availability of the other products.

Since our model represents the intrinsic relationship between products, it should be independent of which store we consider. Hence, for each product in each store, we estimate the average sales, and scale the sales of the product by the estimated average, in order to remove the store information from the data.

In our model, the expected sales of each product, A,B,C,D, is given by the sum of the baseline demand when only this product is available, and the increase (or decrease) of the demand of this product when the remaining three products become available separately. All the demands estimated by the data are coefficients in our model and the variables are the indicators of the availability of the products, which take the value 0 or 1. Hence, we predict the effect of introducing a product on the sales of other products by changing the corresponding indicator to 1 in our model.

Multiplicative model for sales

Our second approach is to model the influence of introducing a previously unavailable product or removing a previously available product on the sales of another product by adding or subtracting a certain portion of that product's current sales, which is achieved by multiplying the present sales level by some factor. For example, if a store introduces product D, we will expect the sales of product A to be changed by some factor of its original sales level. We assume that these factors are independent of the availabilities of the products, and of the stores. Further, if the factor is greater than 1, then a halo effect probably exists between the two products, and if the factor is less than 1, there is potential a demand transfer between them.

We assume that, at the weekly level, the structure change can be explained by the availability of the other products, and other variations can be explained using a *Poisson distribution* (a common distribution for counting data like sales) for the sales of the product under consideration. Our multiplicative model explains the structure change.

Specifically, we use a log function to translate the multiplication into addition, and leave all the local information of each store in the baseline demand when only the product of interest is available. All the factors and the baseline demand, which will be estimated by the data, are coefficients in the model, and the variables are again the availability indicators which take the values 0 or 1.

3 Results

Before testing the model, we note that in some stores some products are available or unavailable all the time, and hence the availability of these products is not significant in explaining the sales of other products and should not be included in the model when fitting the stores. Hence, we select the group of variables in the model by minimising how much information is lost by the model using the *Akaike Information Criterion* [2].

Fit in each store

When fitted with data from a particular store, the models both correctly predict regime changes

We first fit our model using the data from each store individually, and compare our fitted values to the mean values in each regime, to see whether the assumption that the dependence of the sales of one products on the the availability of another product is independent of the availability of the other products is violated.

It is clear from the data presented in Figure 2 that both our models successfully capture the variations of the mean sales of the product (black line) when the availability of other products change. We see that the multiplicative model performs slightly better than the additive model.

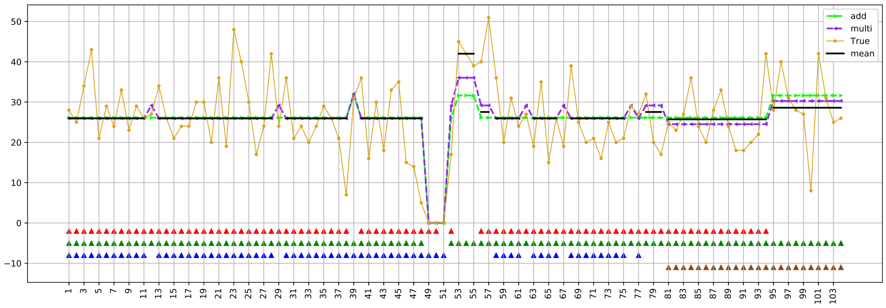
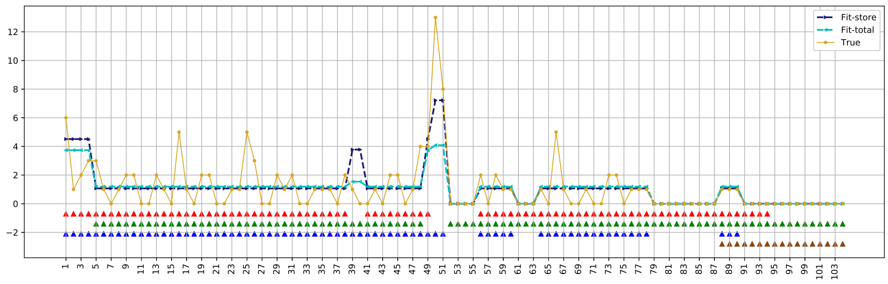


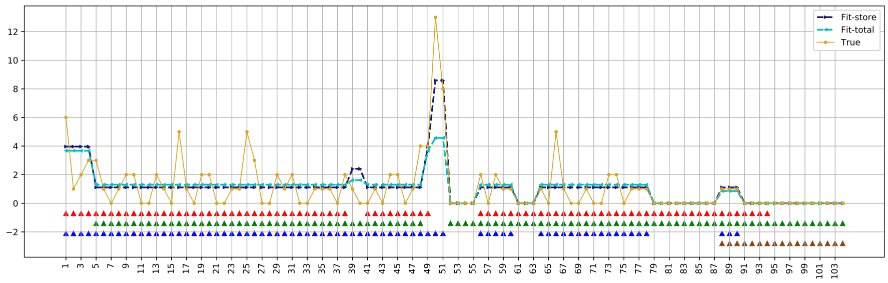
Figure 2 – Graph showing the sales data for product B in store 3 for 103 weeks. The dotted line is the weekly sales data, black line is the mean in each regime, the green right- and purple left-pointed triangle-connected lines are the predictions from the additive and multiplicative models respectively. The up-triangles in red, green, blue and brown represent the availability of products A,B,C,D respectively.

Fit in all stores

We aggregate all the store data, and fit the models to the aggregated data.



(a) additive model



(b) multiplicative model

Figure 3 – Graphs showing the sales data for product C in store 8 in 103 weeks and predictions from the two models fitted in all stores (left-pointed triangle-connected line in cyan) versus in store 8 (right-pointed triangle-connected line in dark blue). The dotted line is the weekly sales data, and the up-triangles in red, green, blue and brown represent the availability of the products A,B,C,D respectively.

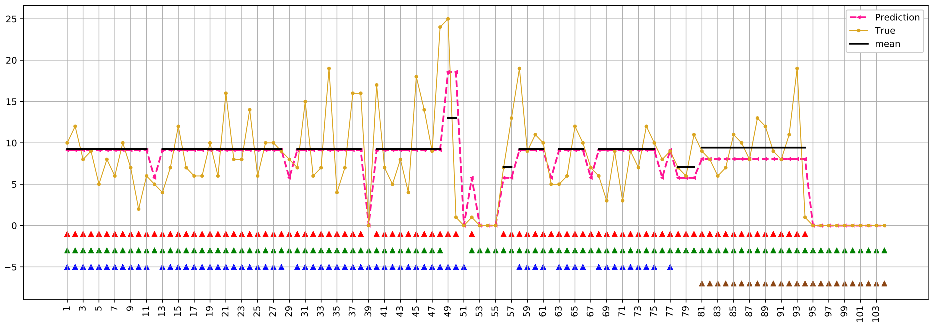
If we fit the models using the data from all stores, they correctly predict the behaviour of individual stores

We then compare the fitted coefficient values to those from the model fitted in each store, to see whether the assumption that the dependence of sales of one product on the availability of another product is independent of stores is violated. We see from Figure 3 that both our models manage to explain the variations that occur when the availability of other products change. We also find that the values of the parameters found by fitting our model to the data from all the stores are very similar to the values found when we fit our model to the data from one store only. Again, the multiplicative model performs slightly better than the additive model.

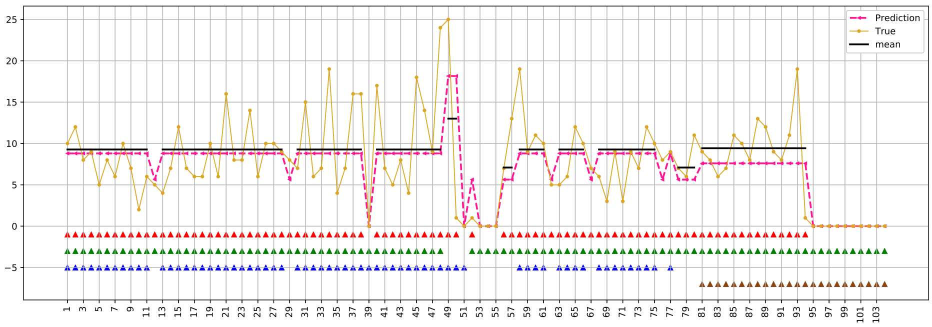
Both of our models have a good predictive abilities, even when they have not seen the store data before

Cross validation

We now compare the performance of our two models in predicting the sales at stores that they have not seen before. Specifically, we fit our two models to the amalgamated data from stores 1 and 5, and use them to predict the sales data for all the other stores, and results for store 3 is shown in Figure 4. As shown in Figure 4, both of our models reproduce the actual changes, when the availability of other products changes, even when they have not seen the store before, and both models have very similar behaviour. Hence, our models are able to provide insights into the changes in sales if we remove or add some of the products for sale.



(a) additive model



(b) multiplicative model

Figure 4 – Graphs showing the sales data for product A in store 3 in 103 weeks and predictions from our two models fitted in store 1 and 5 (left-pointed triangle-connected line in pink). The dotted line is the weekly sales data, and the black line is the mean sales in each regime, and the up-triangles in red, green, blue and brown represent the product A,B,C,D are available in the weeks respectively.

4 Discussion, conclusions, & recommendations

In the era of "big data", the data collected from customers has increased exponentially, while the development of efficient techniques to analyse it has not matched the pace. Specifically for the halo effect and demand transfer, most research has focused on qualitative, rather than quantitative, understanding. We have proposed two simple mathematical models, with the halo effect and demand transfer built into the coefficients, to explain changes in the sales of products. Each model captures a different mechanism for how the availability of other products affects the sales, either by adding a certain amount, or by multiplying by a certain factor. Both models have been fitted to data from individual stores, data from all stores and data from some stores, and they successfully capture variations when availability of other products changes in most cases.

Product availability plays an essential role in both our additive and multiplicative models, since the indicators of the availability of products are the only variables, and our models perform well in predicting mean sales. The availability flag of products is obtained by modelling the customer arrival time using a homogeneous Poisson process, and only two cases are considered here, either the product is available during the whole week or not.

We will further develop this work by introducing more features of the data

There are also various avenues for future work. Firstly, some violation of the homogeneous Poisson process assumption occurs, which means that a better model for the customer arrival time should be incorporated. Further, the case when products become unavailable during a week should be treated differently from the case when they are fully available, and the estimated number of products sold when certain combinations of products are available should be included.

The time-invariant feature of the mean of the total sales in each regime should be further explored, and we expect some correlation between the sales. We note that incorporating in our mode the sales quantities of some products to explain sales of another product does not significantly improve the predictions. This may be caused by the coarse graining of the data into weekly base. In the future, we will reformulate the model in terms of daily, rather than weekly, data. Also, since there is no reason why the relationship between sales should be linear, we will explore nonlinear relationships driven by the sales data.

There are also cases where our models have significant discrepancy from the mean level of the sales data in some regimes, and we spot that this only occurs when the corresponding regime-change happens only in one particular store. Hence there is evidence that some coefficients of our model may depend on the regimes the products are in, or the availability information about other products, representing a complex dynamics of products, and we will build networks to explain these relationships in the future.

5 Potential impact

Our mathematical models provide simple and straightforward ways to predict changes in the mean sales of a product when adding or removing another product. The models provide insight into the Halo effect and demand transfer, and have predictive power. Hence, they can be used by stores to further manipulate the range of products on sale, and by the supply chain to determine the products to provide, in order to optimise the total profit.

Dr Alisdair Wallis, Data Science Manager at Tesco said: "*Yu's work on developing a simple mathematical model of demand transfer has given us a great initial step forward in better understanding the relationships between products. The problem of demand transfer is difficult and developing a quantitative, data-driven understanding of it is key for any retailer. We are therefore very excited to take this project to the next stage*"

References

[1] J. F. C. Kingman. *Poisson Processes*. Clarendon Press, 12 1992.

[2] Jie Ding, Vahid Tarokh, and Yuhong Yang. Model selection techniques: An overview. *Signal Processing*, pages 16–34, 2018. doi: 10.1109/MSP.2018.2867638. URL https://doi.org/10.1109/MSP.2018.2867638.