In-season retail sales forecasting using survival models

M Hattingh[*]   DW Uys[†]

Abstract

A large South African retailer (hereafter referred to as the Retailer) faces the problem of selling out inventory within a specified finite time horizon by dynamically adjusting product prices, and simultaneously maximising revenue. Consumer demand for the Retailer's fashion merchandise is uncertain and the identification of products eligible for markdown is therefore problematic. In order to identify products that should be marked down, the Retailer forecasts future sales of new products. With the aim of improving on the Retailer's current sales forecasting method, this study investigates statistical techniques, *viz.* classical time series analysis (Holt's smoothing method) and survival analysis. Forecasts are made early in the product life cycle and results are compared to the Retailer's existing forecasting method. Based on the mean squared errors of predictions resulting from each method, the most accurate of the methods investigated is survival analysis.

Key words: Retail sales forecasting, survival analysis, time series analysis, Holt's smoothing method.

1   Introduction

A large South African retailer that aims to provide affordable merchandise to the lower- and middle-income target market is considered. The Retailer offers a vast selection of different products, ranging from shoes and clothing to cell phones and home decorations. A large proportion of merchandise that the Retailer sells consists of fashion items, of which the demand is dependent on seasonal trends and consumer sentiment. For this reason, it is difficult to estimate future seasonal demand. For seasonal products, new stock is bought from overseas suppliers every season. For this particular Retailer, this is a once-off transaction, and once the stock has been ordered by the Retailer's buyers, no changes can be made, irrespective of the product's sales performance.

---

[*]Business Intelligence and Actuarial Synthesis Team, Metropolitan Retail: CFO Division, Building 6/1, Parc du Cap, Mispel Road, Bellville, 7530.

[†]Corresponding author: Department of Statistics and Actuarial Science, University of Stellenbosch, Private Bag X1, Matieland, 7602, email: dwu@sun.ac.za

Demand for the Retailer's products is price-elastic due to the nature of the target market. Periodic price cuts throughout the season stimulate sales for products that do not sell out satisfactorily. However, the Retailer should avoid marking down products for which consumers are willing to pay full price. It is important to identify which products should be marked down at an early stage because if markdowns occur too late, inefficient occupation of shelf space may lead to a decrease in revenue.

The problem that the Retailer faces corresponds to the widely researched field of markdown optimisation, originally conducted by Kincaid and Darling (1963). Markdown optimisation deals with the problem of maximising expected total revenue by continuously adjusting prices, given that sales may only take place within a finite time horizon (Gallego & Van Ryzin 1994). A brief overview of the literature in this field and a discussion of why a traditional approach to markdown optimisation is inappropriate in the case of the Retailer are given here. Subsequently, a discussion of the methods used in the study is given.

1.1 Stochastic demand models

As an input in the markdown optimisation model, consumer demand needs to be modelled, either deterministically or stochastically. In stochastic models of consumer demand, consumer demand can either be a random variable or a function of a random variable. An example of the former is to model sales of a specific product over time as a Poisson counting process with transition intensity inversely proportional to price (Chatwin 2000). The objective is then to maximise the expected revenue under the assumed distribution. Under the Poisson model, the maximum expected revenue is a non-decreasing, concave function of remaining inventory over time to the end of the season, and the optimal price is continuously decreasing (Chatwin 2000).

Mantrala and Rao (2001) define a model where demand is a function of both deterministic and stochastic variables. The function is given by

$$D_{tj} = \alpha_t M \left(\frac{P_f}{P_{tj}} \right)^{\gamma_t} \varepsilon_t,$$

where

D_{tj}	is consumer demand for product j at time t at price P_{tj},
α_t	is a seasonal factor at time t,
P_f	is the full price of the product, charged at the beginning of the season,
M	is the total seasonal demand at price P_f,
γ_t	is a function of the sensitivity at time t of consumer demand to a change in price, and
ε_t	is a random variable with a continuous time lognormal distribution (*i.e.* the random disturbance component takes the form of geometric Brownian Motion).

The estimation of parameters for a stochastic model requires an extensive amount of data. For the model described above, the sensitivity parameter γ_t varies over time, and can only be estimated if there are sufficient observed values of D_{tj} for all values of t and j.

Data available to the particular Retailer investigated in this study only provides information on the effect of late price changes (if any) on consumer demand. In other words, there are

- many observed values of D_{tj} if $P_{tj} = P_f$ for all values of t,
- very few observed values of D_{tj} if $P_f \neq P_{tj}$ for large values of t, or
- no observed values of D_{tj} of $P_f \neq P_{tj}$ for small values of t.

Since model fitting requires a sufficient number of observed values for D_{tj} for all values of t and j, γ_t cannot be estimated accurately based on the available data. If a stochastic model was to be used to optimise markdowns, subjective assumptions would be required about the form of γ_t. These assumptions may be inaccurate, and resulting markdown decisions may potentially lead to losses in revenue.

A further disadvantage of the stochastic approach is that it often requires the assumption of independence of sales quantities in consecutive weeks, which is unlikely to be valid (Lobel & Perakis 2010).

Given the limited nature of the available data for the Retailer investigated, it is not feasible to apply markdown optimisation in its traditional sense to the Retailer's markdown decision problem. However, the data may be useful for predicting what demand will be assuming that the price remains constant. Using the previous notation, a model for D_{tj} can be developed, assuming a constant price, since there are many observed values of D_{tj} where $P_{tj} = P_f$. Even though such a model would not be useful for determining the optimal time and magnitude of markdowns, it may nevertheless help in identifying which products should be marked down.

Therefore, instead of investigating ways of optimising markdowns, this study focuses on the question of *whether* markdowns for particular new products are necessary at all. The identification of products eligible for markdown is done by means of in-season sales forecasting of newly launched products, assuming no price change. Sales forecasts provides information as to whether the products considered will sell out within the specified time horizon if the price remains the same. If not, the product is flagged for markdown.

1.2 The Retailer's approach to markdown identification

To assist with decisions regarding the markdown of products, an early indication of likely future sales performance is needed. The Retailer therefore predicts the remaining shelf life of products shortly after the commencement of sales based on a simple heuristic method, which is hereafter referred to as the forward cover method. The concept of forward cover, also known as "weeks of supply", is widely used in different forms across the retail industry (Meckin 2007). The forward cover is defined as a measure of the number of weeks' worth of inventory in stock at any particular time (Chase *et al.* 2008). The variation of forward cover used by the Retailer is based on the assumption that sales will remain constant over the entire remaining shelf life of the product. The constant rate of future sales is assumed to be an average of the previous 5 weeks of sales.

The forward cover calculated in week n is defined by

$$F_n = \frac{C_n}{\frac{1}{5} \sum_{i=0}^{4} S_{n-i}},$$

where

C_i is the closing inventory for week i, and
S_i is the quantity of products sold during week i.

This calculation is done on a weekly basis, starting as soon as sufficient sales data are available. Products that are not expected to sell out within the allowed time horizon are then identified as being eligible for markdown.

In this study, two alternative forecasting methods (described in §1.4.) are investigated with the aim of improving on the accuracy of the forward cover method. Ideally, the remaining future shelf life of products should be forecasted in a methodical, quantitative manner. Furthermore, the forecasting model should be capable of:

1. producing forecasts of future sales on a weekly basis,
2. using very little data as the basis for the forecasts,
3. using as much as possible of the information underlying the available data, and
4. using knowledge of trends in past data (on sales of similar products) as a reference to estimate sales of a new product.

1.3   Forecasting methods

Two forecasting models are proposed to predict future sales, namely time series analysis and survival analysis.

1.3.1   Time series analysis

A number of time series techniques have been used for sales forecasting, including Autoregressive Integrated Moving Average (ARIMA) models, Bayesian forecasting models and exponential smoothing models. Most of these methods require estimation of several parameters. In this study, an early indicator of future product success is needed. Forecasts are required after only eight weeks of initial sales. Since only eight data points are available on which to base forecasts, a model with as few as possible required parameter estimates is needed.

Holt's smoothing method for exponential trend was used in this study, since there was no significant seasonality over the short time period observed and inventory typically diminishes faster than straight-line decay. Forecasts are obtained for the weekly closing inventory percentages. To obtain an estimate for the forward cover, the number of weeks until the forecasted inventory is less than 1% of total inventory is established.

1.3.2   Survival analysis

The theory of survival analysis considers the time to occurrence of a particular event. Possible events include the time of death in clinical trials, the length of stay in hospital until

discharge, or even how long it takes before a light bulb fuses. In the past, practical applications of survival analysis principles have mainly been in the actuarial field, but survival analysis has increasingly been used in non-traditional fields, including the manufacturing industry (Berry 2009).

The general survival function $S(t)$ is defined as

$$S(t) = P(T > t).$$

It is the probability that a response variable $T \geq 0$ exceeds time t. The survival function denotes the probability that a subject will survive for a minimum period of t time units. In the case where the probability of survival of a subject is dependent on the age of the subject, the survival function may be written as a function of age. The probability that a subject currently aged x will survive for a minimum period of t time units is expressed as

$$p_{tx} = S_x(t) = P(T_x \geq t).$$

The hazard function, $\mu(t)$, is defined as the instantaneous rate at which deaths occur, conditional on no previous deaths occurring. The hazard function is given by

$$\mu(t) = \lim_{\Delta t \to 0} \frac{P(t < T_x < t + \Delta t \mid T_x > t)}{\Delta t} = \frac{f_x(t)}{S_x(t)},$$

where $f_x(t)$ is the probability density function of the future lifetime, T_x, of a subject aged x (Cox & Oakes 1984). The survival function, p_{tx}, is hence defined as a function of the integrated hazard function and is given by

$$p_{tx} = \exp\left(-\int_x^{x+t} \mu(s)\, ds\right).$$

In the context of sales forecasting, the future shelf life of products is considered, *i.e.* the response variable T_x is defined as the time until sale of a product given that the product has been on the shelf for x weeks. The probability of a particular product being sold in any given week is considered analogous to the probability of death. The probability that a product is sold between week x and $x+1$, given that it is not sold by week x, is represented by

$$q_x = 1 - p_{1x} = 1 - \exp\left(-\int_x^{x+1} \mu(s)\, ds\right).$$

The application of survival models in retail sales forecasting is potentially useful because:

1. No distributional assumptions are needed in the model. The model relies solely on data and the result is an empirically derived set of mortality rates that capture all information contained in past data without the need for parametric formulae.
2. The results of the model depend on a mix between information obtained from past data and the latest sales data of new products.

1.4   Cross validation

In order to test the validity of the suggested models, a subset of 11 products was left out of the data analysis and used as test observations. Afterwards, all models were applied to the chosen 11 products. This allowed direct comparison of the two forecasting techniques with the forward cover method.

2   Survival analysis methodology

A diagrammatic outline of the methodology followed to obtain forecasts of future sales is given in Figure 1. These three steps in the methodology is discussed in the following sections.
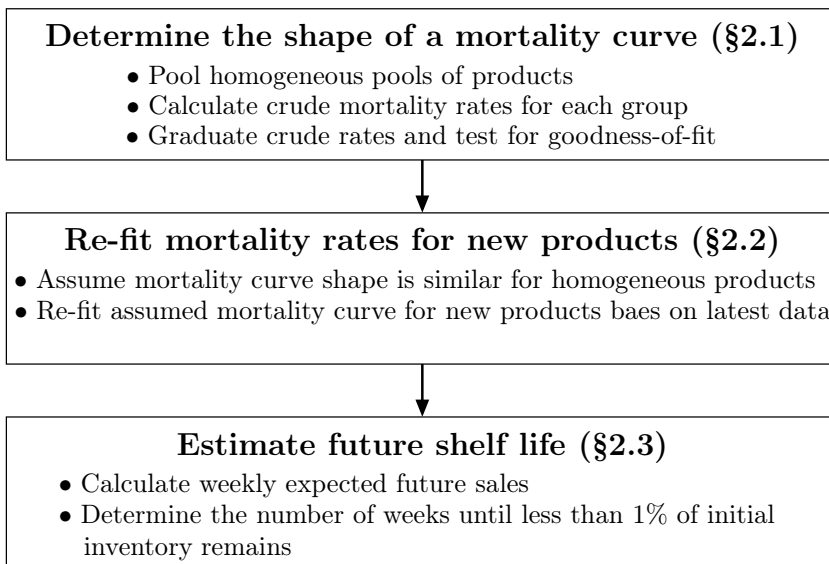
Determine the shape of a mortality curve (§2.1)
- Pool homogeneous pools of products
- Calculate crude mortality rates for each group
- Graduate crude rates and test for goodness-of-fit

Re-fit mortality rates for new products (§2.2)
- Assume mortality curve shape is similar for homogeneous products
- Re-fit assumed mortality curve for new products baes on latest data

Estimate future shelf life (§2.3)
- Calculate weekly expected future sales
- Determine the number of weeks until less than 1% of initial inventory remains

Figure 1:   *An outline of the steps followed in the survival analysis methodology.*

2.1   Determining the shape of a mortality curve

A number of factors influencing the nature of a mortality curve are considered.

2.1.1   Pooling of homogeneous groups

Products that were deemed relatively homogeneous were pooled together in groups to form different cohorts of 'lives' to produce mortality rates that would be applicable to all products in that group. Each group consisted of various different product styles from the same department. For example, within the "Ladies' Clothing" cohort, there were a number of different styles of dresses, pants and other attire. The five cohorts investigated are ladies' clothing, shoes, girls' clothing, baby girls' clothing and preschool boys' clothing respectively.

The maximum likelihood estimate, $\hat{q}_x$, of the mortality rate is obtained by dividing the number of products sold by the number of exposure units (Broffitt 1984). This estimate is also referred to as the actuarial estimate and is given by $\hat{q}_x = \frac{d_x}{E_x}$, where $\hat{q}_x$ is the crude initial rate of mortality, d_x is the number of products sold during week x (*i.e.* the number of "deaths"), and E_x is the initial number of units exposed to risk during week x.

2.1.2 Graduation of crude rates

A simple moving average approach was used to graduate crude rates for each of the five cohorts. However, for the first 3–10 weeks (depending on the cohort), crude rates and graduated rates were assumed equal, since the exposure data during the first weeks were sufficient to produce adequately smooth rates. Depending on the volatility of the crude rates, a three- or five-point moving average was taken. The graduated rates are denoted by $\dot{q}$ and the formulae for obtaining graduated rates for three- and five-point moving averages are given by

$$\dot{q}_x = \frac{1}{3}(\hat{q}_{x-1} + \hat{q}_x + \hat{q}_{x+1}) \quad \text{and} \quad \dot{q}_x = \frac{1}{5}(\hat{q}_{x-2} + \hat{q}_{x-1} + \hat{q}_x + \hat{q}_{x+1} + \hat{q}_{x+2}),$$

respectively. As an example, an illustration of crude and graduated mortality rates for Ladies' Slippers is given in Figure 2.
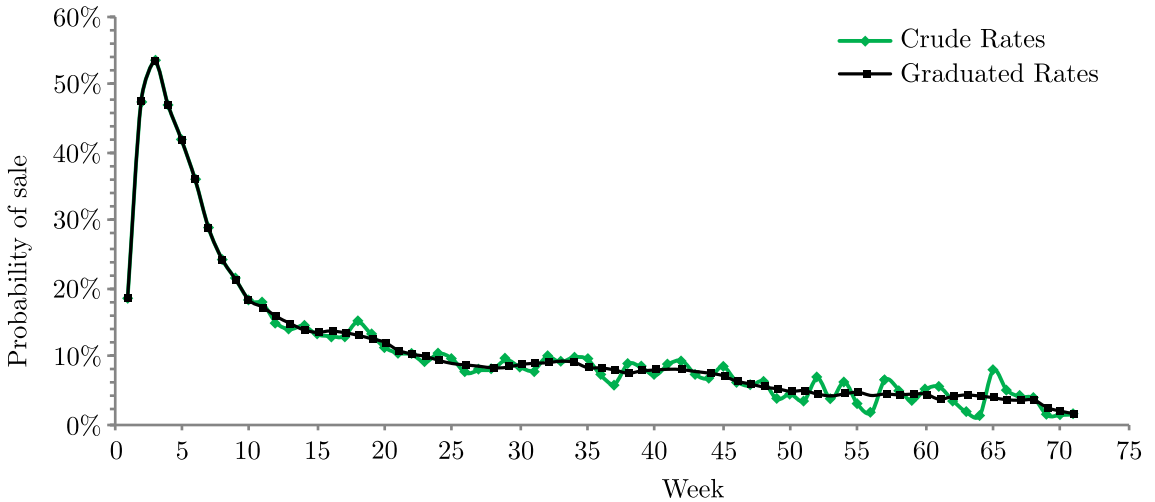

Figure 2: *A graph of the crude and graduated mortality rates for Ladies' Slippers over weeks.*

2.1.3 Graduation tests

The graduation tests below were used to determine whether the graduated rates were significantly biased compared to the data. In each case, the fit is deemed adequate if the null hypothesis is not rejected.

Sign Test

The signs test is used to detect whether there is a bias in the graduated rates. Only the sign of the deviation, z_x, is taken into account. The number of positive deviations is

assumed to be binomially distributed with success probability, $p = 0.5$. It is a two-tailed test, since $H_0: p = 0.5; H_1: p \neq 0.5$. A p-value (Benjamin & Pollard 1993) is calculated for

$$p = 2P[X \geq \max\{n_{\mathrm{o}}, p_{\mathrm{o}}\}],$$

where X has a Binomial $(n, 0.5)$ distribution (n is the sample size), n_{o} is the observed number of negative deviations, and p_{o} is the observed number of positive deviations. A significance level of 0.05 was used, *i.e.* a p-value larger than 0.05 resulted in the null hypothesis not being rejected.

Grouping of signs test (Stevens' test)

The grouping of signs test aims to detect long runs or clumps of deviations of the same sign. The number of groups of positive signs, g, is counted. Under the null hypothesis, the probability of having exactly g positive groups is given by

$$P[G = g] = \frac{\binom{n_1}{g-1}\binom{n_2+1}{g}}{\binom{m}{n_1}},$$

where n_1 is the observed number of positive deviations, n_2 is the observed number of negative deviations, g is the number of groups of positive deviations, and m is the sample size (*i.e.* number of weeks over which the graduation was done) (Benjamin & Pollard 1993).

In each case, a p-value was calculated. The p-value,

$$p = \sum_{i=0}^{g} P[G = i] = \sum_{i=0}^{g} \frac{\binom{n_1}{i-1}\binom{n_2+1}{i}}{\binom{m}{n_1}}.$$

is equal to the probability of having a number of positive groups fewer than or equal to that observed.

2.2 Re-fitting of graduated rates for new products

Since the aim is to forecast sales of new products that did not form part of the investigation of mortality rates, the sets of graduated mortality rates cannot be used directly. New products were first classified into groups, and subsequently viewed as new manifestations of the process modelled by past data.

Adjustments were made to the shape of the smoothed mortality rates in light of new information obtained from the first eight weeks of new product sales. The mortality rates assumed for these new products were obtained by taking a linear combination of the smoothed rates, *i.e.*

$$\hat{q}_x = aq_x + b,$$

where $\hat{q}_x$ is the adjusted (new) mortality rate, q_x is the graduated mortality rate based on sales data from previous products, and the parameters a and b were determined using a least squares approach based on observed sales for the first eight weeks, with the constraint that all fitted and predicted mortality rates must be nonnegative. An example of a set of re-fitted mortality rates for a particular new product is given in Figure 3. To illustrate prediction accuracy, the realised actual sales numbers are included in the graph.
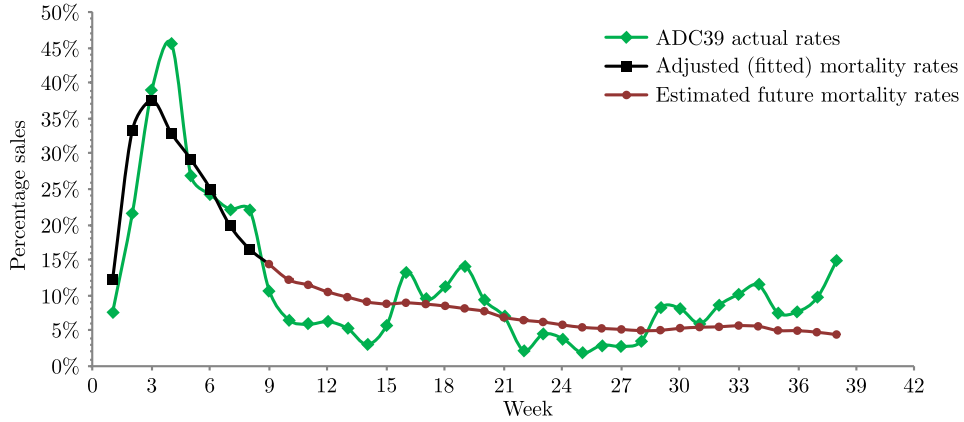
Figure 3: *A graph of the re-fitted product mortality rates.*

2.3 Obtaining forecasts from estimated mortality rates

An estimate of the product's remaining number of weeks on shelves is obtained by calculating the number of products expected to remain after each week by $\ell_{x+t} = \ell_x \times p_{tx}$, where ℓ_x is the number of products expected to remain after x weeks, and p_{tx} is the probability of survival up to week $x + t$, given survival up to week x. This can be computed directly from the estimated mortality rates: $p_{tx} = (1 - q_x)(1 - q_{x+1}) \cdots (1 - q_{x+t})$. The estimate of the complete future shelf life is equal to the smallest value of t for which ℓ_{x+t} is smaller than 1% of the initial inventory level.

3 Empirical results

A comparison of the actual vs. predicted shelf life of all three forecasting methods is given in Figure 4.

The predictions arising from both the forward cover methodology and time series analysis are underestimated, since forecasts are unanimously below the actual values. The survival analysis predictions seem to be the most accurate, and are not consistently biased. To formalise this conclusion, a comparison between the prediction errors and resulting mean squared error (MSE) for each of the models is given below in Table 1.

3.1 Signs test

Resulting p-values for each group is given below in Table 2. Since all p-values are greater than 0.1, the null hypothesis for each cohort is not rejected at the 10% significance level. There is thus no significant bias, and the graduation fits the data adequately.

3.2 Grouping of signs test

A summary of results of the grouping of signs test is given in Table 3. Since all p-values are greater than 0.1, the null hypothesis is not rejected at the 10% significance level, and it can be concluded that the graduation fits the data adequately.
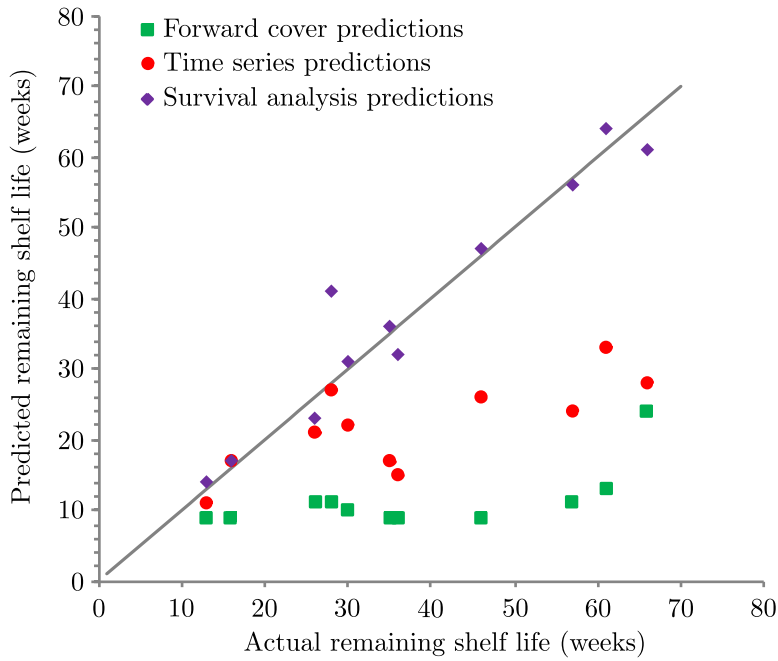
Figure 4: *A plot of the actual* vs. *predicted number of remaining weeks for the three models.*

| Product | Prediction errors | | |
	Forward cover method	Time series analysis	Survival method
Ladies' pencil skirt	−42	−38	−5
Ladies' bengaline pants	−48	−28	3
Ladies' slippers (style A)	−4	−2	1
Ladies' slippers (style B)	−15	−5	−3
Ladies' slippers (style C)	−26	−18	1
Girls' T-shirt (style A)	−46	−33	−1
Girls' T-shirt (style B)	−7	1	1
Baby girls' fancy top	−17	−1	13
Baby girls' leggings	−37	−20	1
Preschool boys' yarn dye T-shirt	−27	−21	−4
Preschool boys' printed T-shirt	−20	−8	1
MSE	903.4	416.1	21.3

Table 1: *Prediction errors and resulting MSE for the three forecasting methods.*

Ladies	Shoes	Girls	Baby	Girls	Preschool Boys
Sample size	84	50	69	45	48
Number of positive deviations	49	21	35	21	21
Proportion of positive deviations	58.3%	42.0%	50.7%	46.7%	43.8%
p-value	0.101	0.161	0.810	0.766	0.471

Table 2: *Prediction errors and resulting MSE for the three survival analysis.*

A comparison of the three methods is given in Table 4.

Ladies	Shoes	Girls	Baby	Girls	Preschool Boys
Sample size	84	50	69	45	48
Groups of positive signs	23	14	19	10	10
p-value	0.8684	0.8670	0.6402	0.2416	0.1509

Table 3: *Results of grouping of signs test.*

4 Conclusion

The forecasts produced by survival analysis produce the most accurate results of the methods investigated in this paper. It is a computationally expensive method to implement on a large scale.

Since the Retailer's current forecasting method has been shown to produce inaccurate forecasts, which may lead to sub-optimal markdown decisions, it is recommended that further resources be spent to investigate the use of survival analysis as a forecasting options.

Acknowledgements

Method	Factors that fundamentally impede accuracy of the method	Advantages	Disadvantages
Forward cover method (Retailer's current method)	• Inappropriate to assume constant sales over the entire shelf life of the product, since past data confirms that sales usually peak during the first 3–5 weeks, and then decrease as the product ages. • Very sensitive to outliers in the first weeks of sales. This is of particular concern, since sales volumes are usually highly volatile throughout the season.	• Sheer simplicity • Ease of understanding the method • Ease of calculation • Full automation of calculation possible	• Shown to be historically inaccurate (negatively biased) • A consequence of the above is that markdowns usually occur too late, with adverse effect on revenues.
Time seris analysis	• Information on the usual distribution of sales is available through historical sales data, but is not seen by the model, which uses only data from the first 8 weeks of sales of the new product.	• Quickly and easily applied • Ample software tools (e.g. SAS, Statistica) available to enable automation of calculation	• Significant level of added complexity for little improvement in accuracy • Requires some level of expertise to apply the method
Survival analysis	• The model does not take seasonal factors into account, e.g. large sales volumes over the festive season. • Other external factors (e.g. competitors' actions; consumer behaviour) are also not taken into account by the model. These factors should nevertheless form part of the markdown decision process in a qualitative sense.	• Vast improvement in accuracy over forward cover method. • Uses historical sales data, whereas the other methods only use data from the new product being analysed. This implies that the accuracy of method may be improved even further if more data were used (in this study, only a small subset of the Retailer's data was used).	• Application of the method is time consuming and may be difficult to implement on a large scale (however, software tools could be developed in order to overcome this problem).

Table 4: *A comparison of the three methods used in this study.*

References

[1] BENJAMIN B & POLLARD JH, 1993, *The analysis of mortality and other actuarial statistics*, 3rd Edition, The Institute of Actuaries and The Faculty of Actuaries, London.

[2] BERRY MJA, 2009, *The application of survival analysis to customer-centric forecasting*, Proceedings of the 22nd Annual NorthEast SAS User Group Conference, Burlington (VI), pp. 1–10.

[3] BROFFITT JD, 1984, *Maximum likelihood alternatives to actuarial estimators of mortality rates*, Transactions of Society of Actuaries, **36(1)**, pp. 77–142.

[4] CHASE RB, JACOBS FR & AQUILANO NJ, 2006, *Operations management for competitive advantage*, 11th Edition, McGraw-Hill, Monterey (CA).

[5] CHATWIN RE, 2000, *Optimal dynamic pricing of perishable products with stochastic demand and a finite set of prices*, European Journal of Operational Research, **125(1)**, pp. 149–174.

[6] COX DR & OAKES D, 1984, *Analysis of survival data*, Chapman and Hall, London.

[7] GALLEGO G & VAN RYZIN GJ, 1994, *Optimal dynamic pricing of inventories with stochastic demand over finite horizon*, Management Science, **40(8)**, pp. 999–1020.

[8] KINCAID WM & DARLING DA, 1963, *An inventory pricing problem*, Journal of Mathematical Analysis and Applications, **7(2)**, pp. 183–208.

[9] LOBEL R & PERAKIS G, 2013, *Dynamic Pricing through sampling based optimisation*, [Online], [Cited: 14 December 2013], Available from http://ssrn.com/abstract=1748426.

[10] MECKIN D, 2007, *Naked finance: Business finance pure and simple*, Nicholas Brealey Publishing, London.

[11] MANTRALA MK & RAO S, 2001, *A decision-support system that helps retailers decide order quantities and markdowns for fashion goods*, Interfaces, **31(3)**, pp. 146–165.