Ehrenberg-Bass Institute Working Paper

Forthcoming in the *Journal of Consumer Behaviour*

"Extended Conditional Trend Analysis: Predicting triple period buyer flows with a tri-variate NBD model"

Authors:

Assoc. Prof Giang Trinh - Ehrenberg-Bass Institute
Prof John Dawes - Ehrenberg-Bass Institute
Prof Malcolm Wright - Ehrenberg-Bass Institute
Dr Nick Danenberg - Ehrenberg-Bass Institute
Prof Byron Sharp - Ehrenberg-Bass Institute

Extended Conditional Trend Analysis: Predicting triple period buyer flows with a tri-variate NBD model

Giang Trinh[a], John Dawes[1,a] , Malcolm J. Wright[2,b], Nick Danenberg[a], Byron Sharp[a]

[a] Ehrenberg-Bass Institute for Marketing Science, University of South Australia, Adelaide, Australia
[b] School of Communication Journalism and Marketing, Massey University, Auckland, New Zealand

Author Biographies

Giang Trinh is Associate Professor at Ehrenberg-Bass Institute, University of South Australia. His research focuses on quantitative method development and applications in the areas of consumer purchasing behaviour and marketing dynamics.

John Dawes is a Professor of Marketing and an Associate Director of the Ehrenberg-Bass Institute. His main research interests are buyer behaviour, brand loyalty, the impact of price and price changes, and developing generalisable marketing knowledge.

Malcolm Wright is Professor and MSA Charitable Trust Chair in Marketing at Massey University, New Zealand. He applies empirical principles to marketing problems and has made interrelated discoveries about brand loyalty, forecasting, survey research and optimising the advertising budget.

Nick Danenberg is a founding researcher at the Ehrenberg-Bass Institute for Marketing Science. Dr Danenberg is committed to producing and disseminating scientific knowledge about marketing to help grow brands. His research focuses on advertising and media effectiveness, targeting and loyalty.

Byron Sharp is the Director of the Ehrenberg-Bass Institute at the University of South Australia, Professor Sharp is ranked in the top 1% of academics publishing on advertising and is the author of the "How Brands Grow" (Oxford University Press 2010), published in more than a dozen languages, and the research-based textbook "Marketing: theory, evidence, practice".

Acknowledgements

Data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.

---

[1]Contact for data queries, john.dawes@unisa.edu.au
[2]Corresponding author, m.j.wright@massey.ac.nz

Extended Conditional Trend Analysis: Predicting triple period buyer flows with a tri-variate NBD model

Abstract

Marketers are interested in the loyalty of their customer base. Increasingly this includes examining behavioural loyalty inferred from the frequency or weight of purchase. A typical approach is to divide the customer base into arbitrary segments based on weight of purchase and then attempt to move customers from lighter to heavier segments, rather than have them reduce purchasing or cease buying altogether. Effects can be monitored by examining how purchasing by groups of individuals evolves over successive periods. However, much of the flow between segments represents random fluctuations in period-to-period purchasing rather than true change to underlying loyalty. Accurate analysis requires true change to be separated from these stochastic changes, for example through benchmarks derived from Conditional Trend Analysis (CTA). While CTA considers the two-period case, it provides no guidance for changes seen across three-periods. The three-period case is nonetheless regularly reported by panel companies and relied on by managers. We therefore develop the three-period CTA, using a tri-variate NBD, to allow the analysis of buyer flow across three successive periods. We provide an empirical illustration and demonstrate fresh insights into the evolution of consumer loyalty. The findings allay oft-raised concerns about supposedly 'lost' buyers, as perceived customer loss is often simply regression to the mean of the buying rates. Accordingly, the three-period CTA shows predictable proportions of buyers who move between different buying-weight segments, including first-year buyers who were apparently 'lost' in the second year but return to buy the brand in the third year.

Keywords: customer loyalty; repeat-purchase; NBD; Dirichlet model; buyer flow

University of South Australia | Ehrenberg-Bass Institute for Marketing Science
Working Paper

Extended Conditional Trend Analysis: Predicting triple period buyer flows with a tri-variate NBD model

1. Introduction

Marketing managers are increasingly concerned to examine and grow the loyalty of their customer base. One such analytical approach involves commissioning consumer panel companies to conduct 'buyer-flow analysis' (e.g. InfoScout, 2016) that describes the movement over time of cohorts of buyers based on their weights of purchase. The managerial goal is typically to increase the loyalty of the customer base by moving customers from lighter to heavier brand purchasing weights, rather than have customers reduce the frequency of their purchasing or cease to buy the brand altogether.

Buyer-flow analysis begins by arbitrarily dividing the customer base into several purchasing segments. A typical classification would involve behavioural segmentation of customers into Light-, Medium- and Heavy-buying groups, along with the group of non-buyers. There are no fixed criteria for segment classification, with weights of purchasing used for segment cut-offs being somewhat idiosyncratic. The analysis relies on determining the extent to which individuals change their weight of purchasing, allowing a manager or analyst to track customer migration between different purchasing segments over consecutive periods; for example, first identifying buyers classified as medium-weight in an initial period and then assessing what proportion subsequently became heavier buyers (or lighter, or did not buy at all) in the successive period.

Typical expectations might include that heavy buying would persist from period to period. That is, that there would be a large component of habitually heavy buyers, whose individual purchase requirements are such that they repeatedly buy at a heavy rate, and always therefore remain in the heaviest buying segment. Further expectations might include that greater numbers of heavy buyers would indicate a healthier brand; and that migration from any other buyer group to the group of non-buyers would indicate a loss of customers—an example of the so-called 'leaky bucket' (Ehrenberg, 1988), that would require subsequent recruitment of new-buyers to replace the lost buyers. Yet it is not clear that any of these expectations are justified, as it is difficult to interpret buyer flow analysis without some formal understanding of the normal levels of the ebb and flow between the different cohorts of buyers that would have occurred anyway, even under conditions where a brand's period-to-period penetration remains stationary. Managers are also regularly exhorted to move their consumers up 'the loyalty ladder' (e.g. Power Up Your Marketing, 2017), yet often do not know the extent to which movement to a heavier level of buying would have occurred even without special inducement given to influence this.

Much of the movement, or 'flow' of customers between the weight-of-buying segments over time does in fact represent nothing more than fluctuations in period-to-period purchasing, rather than true changes in the underlying loyalty (long-term buying propensities) of customers. An assumption of the Dirichlet model (Goodhardt et al., 1984) is that buying propensities are fixed over the long-term; an assumption which has substantial empirical support (Ehrenberg et al., 2004; Sharp, 2010). Although the long-term buying propensities of individuals are fixed, each purchase event is probabilistic in nature, occurring as-if randomly but nonetheless in accordance with an underlying mean rate, specified variance and particular probability distribution (such as the Poisson distribution). Accordingly, when individuals' purchasing is examined over successive time periods many changes in weight of purchase (either higher or lower, or no purchase at all) are apparent, despite there being no change to the individuals' underlying, long-term brand buying propensities. Thus, these are *stochastic* changes in individuals' weights of purchase which are evinced despite there being no *true* change to buying propensities (i.e., loyalty). Accurate analysis therefore requires true changes to be separated from these stochastic changes that occur even under stationary (at the aggregate level) conditions.

Goodhardt and Ehrenberg (1967) provided the required stationary market benchmarks for buyer flow between two successive time periods. They drew on the Negative Binomial Distribution (NBD), which even by 1967 had long been shown to accurately describe the discrete distribution of the frequency of purchases made by the customer base within a single period. Their key insight was that for any particular purchase class (say, those purchasing twice in a period), their purchases in the next period would also be NBD, but distributed about the mean level of purchasing of the first period. Combining these NBDs allows detailed calculations for the stationary market case of the expected buyer flow or churn between the two periods. The difference of the *actual* from the *expected* stationary market buyer flow is the dynamic component of brand sales. And it is this dynamic component alone that is tractable to further analysis by panel companies or brand managers. Goodhardt and Ehrenberg illustrated the use of their method with several cases illustrating how the dynamic effects of seasonality, price promotions and persistent sales increases could be teased out. Goodhardt and Ehrenberg called their method of investigating purchase frequencies over two successive time periods Conditional Trend Analysis, or 'CTA' (Goodhardt & Ehrenberg, 1967).

CTA is based on the bi-variate (two-period) NBD model that allows benchmarking and predictions of buying behaviour in two successive time periods and therefore can be used to benchmark 'new' and 'lost' customers of a brand from period 1 to period 2. CTA is regarded as an important analytical approach that yields significant managerial implications; having been used and tested for decades by many researchers in the field (Lenk et al., 1993; Morrison, 1969; Morrison & Schmittlein, 1988; Trinh et al., 2019; Wilkinson et al., 2016). However, due to the two-period nature of the model, a further analysis to investigate the prevalence of 'lost' customers coming back or 'new' customers dropping out is not possible with the current CTA.

In practice, managers are typically interested in buyer acquisition, migration and maintenance over longer periods of time. And managerial observation over three annual time periods is now commonplace. However, to date this has necessarily been carried out without any formal model of expected behaviour for the three-period scenario. For example, if only less than half of the buyers who were apparently 'lost' from period 1 to period 2 were to come back to buy the brand in period 3, would this be a concern, or is it normal and to be expected? The current CTA model is not able to answer such a question. It is therefore important that CTA be extended to support analysis of the dynamic component of buyer flow over more than two periods and at least initially to three periods. For example, of those who are buyers in period 1, what is the probability that they become non-buyers in period 2, and likewise the probability that they *stay* as non-buyers, or alternatively revert to being buyers, in period 3? Likewise, of those who are non-buyers in period 1, what is the probability that they become light-buyers in period 2, and likewise the probability that they stay as light-buyers, or alternatively revert to being non-buyers, in period 3? In this paper, we develop a tri-variate NBD model to extend the CTA to predict buyer flows over three successive periods, providing benchmarks for buying behaviour against which the effects of marketing actions can be assessed. In plain terms, given a buyer's purchase rate in period 1, we want to predict not only their purchase rate in period 2 but also subsequently in period 3.

Application of the three-period CTA will allow researchers to develop greater understanding of the consistency of buyer behaviour, and hence the accuracy and utility of classifying buyers based on single or even double periods. Such fundamental research is sorely needed as technology is allowing marketers far greater ability and ease with which to identify and target cohorts of consumers within their buyer base. However, without a clear understanding of the normal evolution of the customer base over multiple periods, targeting efforts may be based on misconceptions about consumer behaviour.

We next review prior work on the NBD and CTA, develop the general need for an extended CTA, specifically expound the case of the three-period CTA, demonstrate the application of the model, and then compare the subsequent predictions with observed data. Finally, we outline some new research directions opened up by the three-period CTA, including dynamic analysis of the sources of brand growth and decline.

2. Literature Review

2.1. The NBD Model

The NBD has been widely used in marketing to model purchase counts, particularly in the packaged goods context. The model was first applied to category purchasing behaviour by Ehrenberg (1959). Since then, the widespread fit of the NBD to purchase rates has been considered a law-like empirical generalisation (Uncles & Wright, 2004). The NBD is based on two assumptions:

(1) Purchases of a given consumer in successive time periods follow a Poisson distribution. This implies that the variance of purchases within individual consumers is 'as if' random over time (i.e., a Poisson process). For example

Consumer A bought 2 times in year 1, 4 in year 2, and 0 in year 3

Consumer B also bought 2 times in year 1, only 1 in year 2, and 6 in year 3

For each consumer, their purchases are not fixed but 'as if' random over time.

(2) The mean rates of purchasing of different consumers in the long run differ, according to a Gamma distribution. Using the above example

Consumer A's average purchase rate per year is 2

Consumer B's average purchase rate per year is 3

Comparing these two customers, their average purchase rates are different.

Following these assumptions, the frequency of consumers making 0, 1, 2, 3, … *x* purchases in a given time period can be modelled by the NBD. Ehrenberg (1959) shows the earliest published example of the NBD model fit to purchasing data in consumer packaged goods, whereby the model-estimated values closely matched the actual data (Table 1). This demonstrates that consumer purchasing follows a predictable pattern, which can provide a useful baseline for managers to evaluate their marketing activities (Sharp, 2010). In practice, if one knows the proportion of people in a population who purchase a product at all in a time period—such as within a year—and the average rate at which the product is purchased (e.g., 5 occasions in a year), then the NBD will successfully estimate the number of people who purchase once, twice, three times, and so on in time periods of any length.

Table 1. The fit of the NBD to actual data

	Number of people	
Number of purchases	*Actual*	*Theoretical (NBD)*
0	1612	1612
1	164	157
2	71	74
3	47	44
4	28	29
5	17	20
6	12	15
7	12	11
8	5	8
9	7	6
10	6	5
11+	19	19

Source: Ehrenberg (1959)

Since the original marketing study by Ehrenberg (1959), the NBD model has been applied to numerous contexts including brands, categories, time periods and countries (Chatfield et al., 1966; Ehrenberg et al., 2004; Goodhardt et al., 1984), as well as gambling (Lam & Mizerski, 2009; Mizerski et al., 2004), consumption of mobile phone services (Lee et al., 2011), radio listening (Lees & Wright, 2009), cigarette purchasing (Dawes, 2014), industrial buying (Wilkinson et al., 2016), first time buying (Bogomolova et al., 2019), physical activities (Gruneklee et al., 2016), attendance at cultural venues and events (Trinh & Lam, 2016), sport attendance (Trinh, 2018), country of origin purchasing (Trinh et al., 2019) and even the retrieval of brand associations from memory (Romaniuk, 2013; Romaniuk & Stocchi, 2009). Due to its robustness, the NBD model has long been credited as one of the most useful models for modelling brand and product purchases (Schmittlein et al., 1985). Yet, the most theoretical appealing aspect of the NBD is not modelling purchase frequency but predicting future purchases based on past purchases. For example, of those people who buy *x* times in period 1, what is their probability of their buying *x* times again in period 2, what is their probability of down-weighting their purchase to less than *x* times or upward to more than *x* times? We will discuss this important aspect of the NBD model in the next section.

2.2 Conditional Trend Analysis

CTA is an extension of the NBD purchasing model to analysis of consumer behaviour over two sequential time periods. It was developed by Goodhardt and Ehrenberg (1967) to benchmark future sales based on past performance. As mentioned, CTA takes the first-period mean of each buyer class and then applies the NBD to generate purchase frequency distributions for each of these buyer classes in the second period. These second-period counts provide the expectation for buyer flow from the first- to second-period. In aggregate, CTA allows researchers to identify whether the increased sales in a second period are accounted for by previous non-buyers, light-buyers or heavy-buyers of the brand who bought even more, i.e. it decomposes any growth and shows where it came from, and whether the source of the growth was as expected (according to the CTA model predictions), or unusual.

Suppose that in the first period, buyers of the brand are divided into five different purchase frequency segments (buyers who bought 0, 1, 2, 3, and 4+ times). These purchase frequencies could be further arbitrarily assigned to classes of non- (0), light- (1), medium- (2 or 3) and heavy- (4+) buyers, thereby enabling examination of an organisation's customer base in a way that is relevant to many practical marketing decisions. Table 2 shows an example of a stationary market case from Goodhardt and Ehrenberg (1967) in which there is no significant change (in aggregate) from period 1 to period 2.

Table 2: Average purchases in period 2 by buyer classes in period 1

	Buyer class in period 1				
Average purchases in Period 2	0	1	2	3	4+
- Observed	.06	.8	1.7	2.5	4.6
- Predicted	.07	.8	1.7	2.6	5.5

Table 3 gives an example of CTA for a non-stationary brand that <u>does</u> experience an increase in total sales, going from 343 purchases in period 1 to 438 purchases in period 2. The question arises, where did the sales increase come from, i.e. from which buyer group(s)? This question is answered by CTA.

Table 3: An example of CTA

Number of purchases	0	1	2	3	4+
Total buyers in period 1	880	53	24	13	29
Total purchases in period 1 (Obs.)	0	53	48	42	200
Total purchases in period 2 (Obs.)	185	57	45	31	165
CTA predicted total purchases in period 2 (Th.)	53	48	41	35	166
Deviations (Obs - Th)	132	9	4	-4	-1

The first row of Table 3 shows the number of consumers, and the second row shows the number of purchases, for each buyer class in period 1. The third row extends the observations to period 2, showing the number of purchases made in period 2 by each of the period 1 buyer classes. But how is the third row to be interpreted? A naïve comparison between purchases made by period 1 buyer classes in period 2 might suggest a decrease of 35 purchases from the heavy buyers (200-165) and an increase of 185 purchases from the non-buyers (0-185). However, CTA predicts that under stationary conditions, period 1 heavy buyers would normally contribute 166 purchases in period 2, while period 1 zero buyers would normally contribute 53 purchases in period 2.

By comparing actual purchases to the CTA benchmarks in period 2, managers can determine which segment of buyers have caused the increased sales. For the example given above, the difference from the CTA benchmark for period 2 is only one less purchase for period 1 heavy buyers (166-165) but 132 extra purchases from period 1 non-buyers (185-53). Hence, the brand is not actually losing sales from heavy buyers (any more than expected), and practically all the sales growth is accounted for by consumers who were non-buyers in period 1. CTA can be used to determine if there are higher or lower levels of purchasing, compared to expectations, for each buyer group, giving further insights into both growth and decline in purchasing.

The existence of the patterns modelled by CTA also tells us that much of what has been described as a "leaky bucket" is merely regression to the mean. Many buyers may just happen to buy the brand within a particular year, perhaps only once, when their long-term propensity is to buy every two years. So we naturally expect to see them as buyers in some years and not in others, but their absence in a given year does not mean that they have been 'lost'. And, indeed the same effect occurs at the heavier end of the purchase distribution: one year's heavy buyers are, on average, not so heavy the following year — for whatever stochastic reasons, many of these heavy buyers simply had an unusually heavy-buying year.

However, these facts, formalised as part of the NBD / CTA model, might still fall short of providing management reassurance about their brand's buyer base due to a general lack of awareness of the NBD / CTA patterns. Consider that the idea of the "leaky bucket" is still regularly written up in practitioner publications, which confuse the concept of customer loss with infrequent purchasing. For instance, "*fewer than half of a brand's customers come back in this period, which suggests the hole [in the bucket] is very large indeed*". (Thompson, 2020).

and

"*To grow, companies have to not only attract new shoppers each year, they also need to replace last year's shoppers with new shoppers. …And of course, that brands lose their highly loyal customers at an alarming rate.*" (Yu, 2013).

The NBD/CTA accurately distinguishes the 'leaky bucket' from expected regression to the mean and so is potentially highly useful to managers and analysts. Recognising the importance of the method, the NBD/CTA approach has been utilised and generalised by other researchers (Chatfield & Goodhardt, 1973; Lenk et al., 1993; Morrison, 1969; Schmittlein et al., 1985; Schmittlein & Morrison, 1983; Trinh et al., 2014), The majority of proposed extensions have dealt with the statistical distributions of the model.

For example, Morrison (1969) extended the NBD/CTA model to include hardcore non-buyers (buyers who never buy the product). Morrison's rationale is that hardcore non-buyers might cause model bias, as these buyers are not appropriate for the gamma distribution. Chatfield and Goodhardt (1973) developed the Condensed NBD model and Schmittlein and Morrison (1983) applied the Condensed NBD to the two-period case. Schmittlein et al. (1985) developed beta binomial (BB)/NBD CTA model as the beta binomial distribution is shown to be reasonable for modelling purchasing for a particular brand. Schmittlein and Morrison (1988) considered alternative distributional assumptions as the Erlang 2 distribution allows the interpurchase times to be more regular than the Poisson distribution. Recently Trinh et al. (2014) replaced the gamma distribution in NBD/CTA with the lognormal distribution as it has more theoretical appeal than the gamma distribution.

Although the extended models are theoretically sound, the literature (Ehrenberg, 1988; Fader & Hardie, 2002; Morrison & Schmittlein, 1988; Schmittlein et al., 1985) empirically indicates that the simple, unextended NBD/CTA model is very effective for predicting future behaviour. As a result, the simple NBD/CTA model seems to be preferred in more recent applications of predicting future purchase such as industrial purchasing (Wilkinson et al., 2016), purchasing behaviour among different ethnic groups (Trinh et al., 2020), and country of origin purchasing (Trinh et al., 2019).

However, despite the considerable interests in the developments and applications of the NBD and CTA, previous research is limited to predicting future purchases only in two successive periods, as the model concerning more than two periods has not been developed. A multi-period analysis is important as a manager may be concerned to know what subsequently happens to those period 1 buyers who down-weighted in period 2. Do they continue to buy less of the brand, or do they perhaps not buy it at all? Or, did they upgrade their purchasing in the third period? In other words, do they 'come back'? A tri-variate NBD model and three-period CTA would be extremely informative in this regard, as it could estimate how many of those supposed 'lost' buyers from period 2 do come back of their own accord the next period.

3. Three-period CTA

3.1 Tri-variate NBD

We now develop the tri-variate NBD. Goodhardt and Ehrenberg (1967) started their NBD/CTA approach with three independent NBD models of r purchases in period 1, s purchases in period 2 and total (r+s) purchases in both periods. They then developed a joint NBD model of r purchases in period 1 and s purchases in period 2. Finally, they derived the conditional purchases in period 2 of those who made r purchases in period 1 by dividing the joint NBD model of (r,s) by the NBD model of r purchases in period 1.

Using the same approach as Goodhardt and Ehrenberg (1967), we extend to a third period of z purchases and develop a joint NBD model of r purchases in period 1, s purchases in period 2 and z purchases in period 3 or a tri-variate NBD. The steps are below.

Consider a NBD of purchase frequency for each of the 3 equal periods with same parameters k and a, where k is the exponent and ak is the mean purchases in each period

$$f(r) = \frac{(1+a)^{-k}\Gamma(r+k)}{r!\Gamma(k)} (\frac{a}{1+a})^{r} \tag{1.1}$$

$$f(s) = \frac{(1+a)^{-k}\Gamma(s+k)}{s!\Gamma(k)} (\frac{a}{1+a})^{s} \tag{1.2}$$

$$f(z) = \frac{(1+a)^{-k}\Gamma(z+k)}{z!\Gamma(k)} (\frac{a}{1+a})^{z} \tag{1.3}$$

Then under stationarity, the mean purchases in the triple period will be $3ak$.

Consider a buyer of r + s + z purchases in the triple period. According to the Poisson component of the NBD model, these purchases can be regarded as r + s + z independent events, and the probability of each one occurring in the first period rather than the second or third is 1/3. Therefore, the probability that the r + s + z purchases split r in the first period, s in the second and z in the third is given by the multinomial term:

$$\frac{(r+s+z)!}{r!s!z!} (\frac{1}{3})^{r+s+z} \tag{1}$$

The proportion of the population making r + s + z purchases in the triple period is:

$$f(r+s+z) = \frac{(1+3a)^{-k}\Gamma(r+s+z+k)}{(r+s+z)!\Gamma(k)} (\frac{3a}{1+3a})^{r+s+z} \tag{2}$$

The proportion buying r in the first period, s in the second and z in the third is the product of (1) and (2):

$$f(r,s,z) = \frac{(r+s+z)!}{r!s!z!} \left(\frac{1}{3}\right)^{r+s+z} \frac{(1+3a)^{-k}\Gamma(r+s+z+k)}{(r+s+z)!\Gamma(k)} \left(\frac{3a}{1+3a}\right)^{r+s+z} \tag{3}$$

After some simplifications (3) becomes, which is a tri-variate NBD

$$f(r,s,z) = \frac{(1+3a)^{-k}\Gamma(r+s+z+k)}{r!s!z!\Gamma(k)} \left(\frac{a}{1+3a}\right)^{r+s+z} \tag{4}$$

3.2 The Conditional Distribution

To obtain the conditional distribution of purchases in the third period made by buyers of r in the first period and s in the second, it is necessary to divide (4) by the proportion of the population who bought r in the first period and s in the second. According to Goodhardt and Ehrenberg (1967), this proportion is:

$$f(r,s) = \frac{(1+2a)^{-k}\Gamma(r+s+k)}{r!s!\Gamma(k)}\left(\frac{a}{1+2a}\right)^{r+s} \qquad (4.GE)$$

Divide (4) by (4.GE) we obtain:

$$f(z|r,s) = \left(\frac{1+3a}{1+2a}\right)^{-(k+r+s)}\frac{\Gamma(r+s+z+k)}{z!\Gamma(r+s+k)}\left(\frac{a}{1+3a}\right)^{z} \qquad (5)$$

One theoretical aspect of the NBD model is that for any consumer groups who made r purchases in period 1 and s purchases in period 2, their purchase distribution in period 3 is also a simple NBD model. The parameters of this simple NBD are related to a, k, r, and s as shown below.

Let

$$\frac{a}{1+2a} = a'$$

and

$$k + r + s = k'$$

substituting *a'* and *k'* into (5) the conditional distribution becomes:

$$f(z|r,s) = (1+a')^{-k'}\frac{\Gamma(z+k')}{z!\Gamma(k')}\left(\frac{a'}{1+a'}\right)^{z} \qquad (6)$$

which is an NBD and with parameters *a' and k'*.

Having derived the three-period CTA, we now turn to an empirical demonstration.

4. Empirical analysis

For an illustrative empirical analysis we apply the three-period CTA to data for a US brand of breakfast cereal. We set out to examine how closely the tri-variate model predicts actual buyer flow of zero, light-, medium- and heavy-buyer classes across three annual periods. For instance, how many *heavy* buyers in year-1 migrate to be *medium* in year-2 then back to *heavy* again in year-3?

For this purpose we use Nielsen panel data (Kilts Centre for Marketing, 2016) provided via the University of Chicago Booth School of Business. Approximately 60,000 households across 74 US geographic markets report their purchasing to Nielsen via hand-held scanning equipment. We use Nielsen data for three calendar years 2013-2015 for this brand. The brand has stable overall sales for the three years (7.4%, 7.4% and 7.3% market share for the three years, respectively).

We first fit the base NBD model to year 1 data, using equation 1.1 and maximum likelihood estimation, to derive values for the parameters a and k. We then apply these parameter estimates to equation 6 to estimate the three period buyer flows. As per East and Hammond (1996), we aim for light, medium and heavy breakdown of 50%, 30%, and 20% with the 50% of purchasers with the lowest purchase frequencies in the first year classified as light, the top 20% of purchasers with the highest purchase frequencies classified as heavy and therefore, those between the lightest 50% and heaviest 20% classified as medium. However, as purchase frequencies advance in integers, the actual breakdown is (45%, 31% and 24%), where light buyers are those who made 1-2 purchases, medium buyers 3-6 purchases and heavy buyers 7+ purchases in the first year.

Figure 1 shows the fit of the NBD model to year 1 data. As can be seen, the NBD model estimates buyer classes very well. The observed proportions of non-, light-, medium- and heavy-buyers and the theoretical proportions are almost identical and the weighted mean absolute percentage errors (MAPE) for these four comparisons is only 0.4%. Weighted MAPEs were used to measure the accurate prediction of the whole sample, taking into account the importance of each group relatively to their size. This measure overcomes the weakness of MAD or MAPE, where small group size could distort the overall result due to small denominator values (Ludwichowska et al., 2017).

Figure 1: NBD model fits to year 1 data

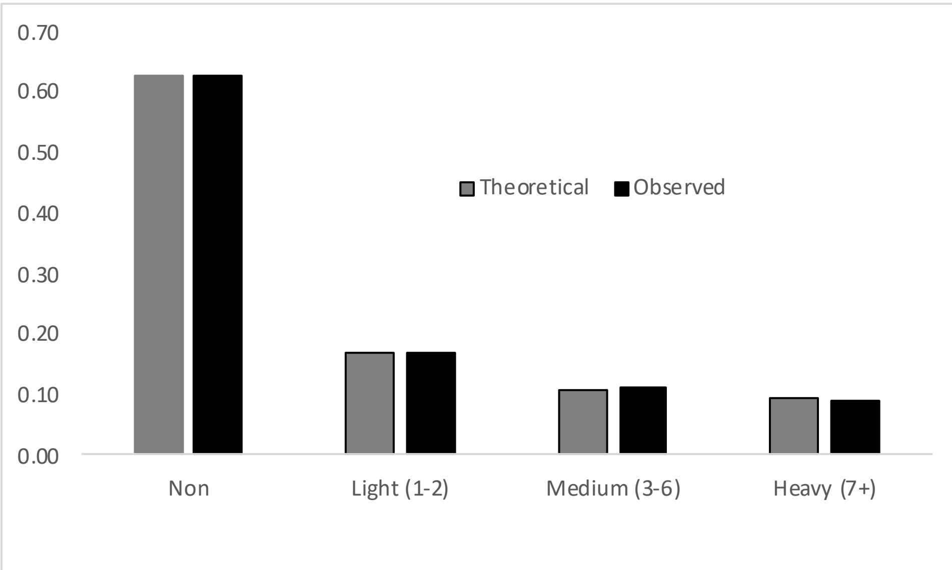
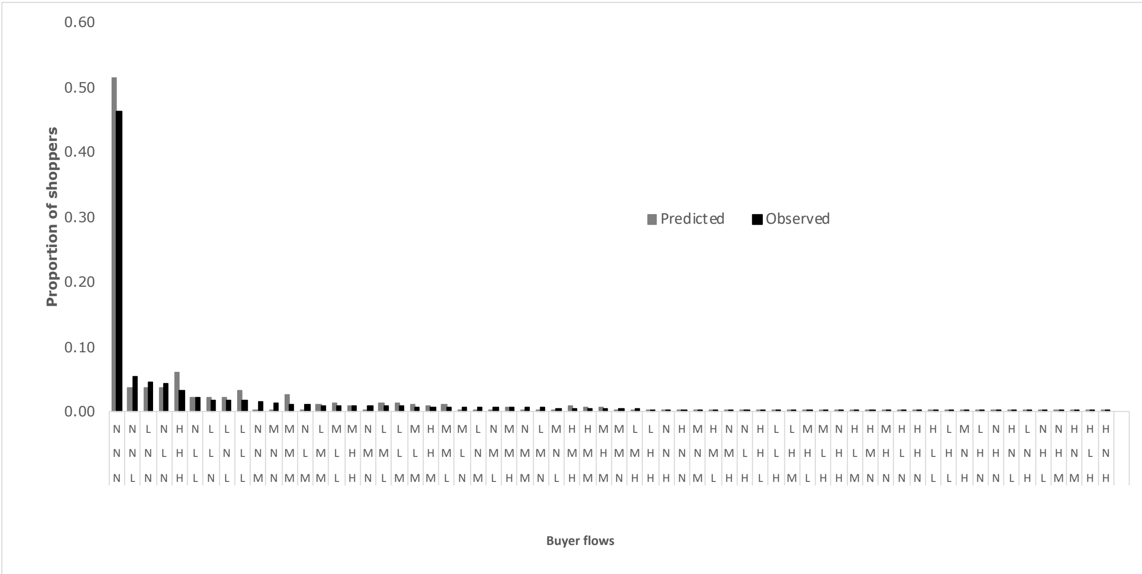


Figure 2 shows the estimates of triple period buyers flows compared with observed data. The bottom row of the X axis is buyer classes in year 1 where N is non buyers, L is light buyers, M is medium buyers and H is heavy buyers. The second row from the bottom of the X axis is buyer class in year 2, and the third row from the bottom is buyer classes in year 3. As we can see, the three-period CTA predicts buyer flows well. For example, the model predicts 51% of shoppers will stay non-buyers from year 1 to year 2 and year 3, while the observed proportion is 46%. Likewise, for the case of non-buyers in year 1 who go on to being light-buyers in year 2 and then return to being non-buyers in year 3, the model predicts 3.8% while the observed proportion is 4.4 %. Overall, the three-period CTA gives a weighted MAPE of 12% over the 64 possible combinations of buyer flow which is a rather good fit.

Figure 2: Triple period buyer flows



5. Discussion: practical applications and future insights

5.1 Contribution

CTA has been recognised as one of the most important methods in stochastic modelling of consumer buying behaviour. Many researchers have applied this method in different contexts but only investigated stochastic changes over two consecutive time periods. In this article we extend the method to multi-period analysis, through conceptualisation and derivation of the three-period CTA. The availability of the three-period CTA provides a substantive methodological contribution for researchers interested in buyer behaviour, but also considerable practical help for applied researchers grappling with problems faced by particular brands. Using the three-period CTA, applied researchers now not only can predict normal changes such as ‘new’ and ‘lost’ buyers (Ehrenberg, 1988), but also how many ‘new’ buyers remain buyers and how many ‘lost’ buyers come back under stationary market conditions. Likewise, researchers not only can predict how many heavy buyers in period 1 remain heavy buyers in period 2 (Romaniuk & Wight, 2015), but can also predict how many will also remain heavy in period 3. A three period analysis also sheds light on the sources of brand growth or decline in a longer term as opposed to short term, e.g. one year (Trinh et al., 2014).

5.2 Managerial implications

One practical application is to analyse the effects and after-effects of a marketing campaign. For example, Goodhardt & Ehrenberg (1967) show how the CTA can be used to analyse the sales uplift produced by a promotion. Without CTA benchmarks, the sales uplift seemed to be nearly entirely due to the on-promotion buying of those who were non-buyers in the period immediately before the promotion. Whereas the CTA showed that in fact 40% of sales were due to period 1 buyers buying more than expected in the subsequent on-deal period. But what of after-effects? the three-period CTA can breakdown the return-to-baseline sales by each buying sub-class. For example, it might show that the promotion (via purchase and consumption) had nudged some of the new promotion period buyers to now include the brand in their repertoire at a higher frequency (or not).

Another example of the usefulness of the three-period CTA, arising from the authors' provision of industry advice, was from a manufacturer of relatively infrequently purchased cleaning products who had launched a new environmentally-sensitive product in their range. Their knowledge of marketing fundamentals was sufficient for them to realise that the product's commercial viability depended not so much on there being a small group of hardcore 'green consumers' but rather on most consumers occasionally buying 'green'. Their sales forecasting model expected that few 'regular consumers' would adopt the product as the dominant brand in their repertoire, but that this shortfall would be made up by a small segment of 'green consumers' who would. There was considerable concern, especially amongst the company's chemists, that too many 'regular' consumers might find the lower cleaning efficacy unacceptable and hence reject the product after tyring it. The CTA showed that repeat rates from period 1 to 2 were close to normal, however management worried that rejection was being masked by the lower rejection rates of hardcore green consumers. This supposition was confirmed by the three-period CTA which showed that fewer second-period non-purchasers returned in the third period than would normally be expected.

The three-period CTA also provides fresh insights into the long-running marketing question of what buyer 'loss' over time is normal. For example, studies by Catalina/CMO council were reported in AdAge in both 2009 and 2011 (Neff, 2011; Pointer Media Network, 2009) each announcing that leading consumer goods brands lost 'half of their most loyal buyers', in that only half stayed in the top loyalty group in the subsequent year[1]. This claim was rightly rebuffed as a manifestation of regression to the mean, i.e. a mis-understanding that taking a snapshot of buyers at a point in time when they happen to be heavy will misclassify them as long-run loyals (Sharp, 2011). Yet, managers might be forgiven for thinking, *but what happens to the half of year 1 loyals that didn't 'stay loyal'?* The three-period CTA adds additional understanding, and reassurance, that many of these heavy buyers who were apparently 'lost' or who down weighted their brand purchasing in year 2, come back in very predictable fashion in year 3, that the sales shortfall from them down weighting in year 2 is counterbalanced by predictable levels of increased purchasing by non-, light-, and medium-buyers, and that the true dynamic component of buyer flow can be identified by comparison against these theoretical norms.

The multi-period buyer flow model also offers enhanced understanding of how to evaluate intervention efforts aimed at particular buyer groups. For example, suppose a manager has been able to identify a light-buyer cohort and wishes to target them with a view to move them 'up the loyalty ladder'. This intervention is to be evaluated over a two-year window. Without the understanding garnered by the buyer flow analysis described in this paper, the manager might well be mystified that the intervention worked for some but not others — some period 1 light-buyers will have upweighted their purchasing to become medium and heavy, but then some others will have actually ceased to buy in period 2. Moreover, the manager will notice that the apparently favourable effect of the intervention seen in year 2 dissipated in year 3, with many of those who had moved up the loyalty ladder slid back down again. Buyer flow analysis based on the three-period CTA would provide benchmarks for exactly how many of each buyer type would have anyway moved up or down (stochastically, in the absence of any intervention), in years 2 and 3, thereby allowing a far more informed interpretation.

Another aspect of the multi-period buyer flow is the very significant sales impact of households who are non-buyers in a base period, contributing to brand sales in subsequent periods. The multi-period buyer flow shows that for a stationary brand with an annual penetration of 37% and average purchase frequency of 5.3, to maintain its current market share it will need to be

[1] Gerald Goodhart commented on the first AdAge article "After 'Some brands lost more than a third…… while others held on to more than 60%……' I stopped reading!".

purchased at least once by 12% of households who didn't buy it in year 2 and 12% of households who didn't buy it in year 1, and indeed 6% of households who did not buy it in *either* of the previous years. These outcomes are extremely helpful to understand the rationale for spending on conventional, mass-reach advertising: to merely maintain a stable market share, a brand must be bought by a very significant number of households who have not bought it even once in the last one or two years.

In terms of the "leaky bucket" in year 2, the three-period CTA shows that it is normal that only half of year 1's buyers would come back in year 3 (46%). This is nothing the brand manager should worry about as it is mostly a manifestation of infrequent purchasing coupled with multi-brand buying (such that *any* particular brand will have some buyers buy it in one year and not the next). The model identifies that a predictable proportion of buyers will return in year 3, (and indeed others will come back in the following years). As such, marketing activity to specifically plug the "leaky bucket" seems unnecessary.

5.3 Limitations and future research

Many more insights will likely emerge as marketing scientists employ multi-period CTA analyses across different categories, for different sized brands, for both growing and declining brands. Further research applying stochastic tools, such as the three-period CTA, will provide the opportunity for significant theoretical advances over the static, snapshot analytical approaches popular in marketing, such as classifying consumers as being either a buyer or not, either holding a particular attitude or not. These static approaches provide a rather shallow understanding as they are based on ideas inconsistent with decades of observation of *actual* buyer behaviour. Conversely, stochastic techniques, such as the three-period CTA introduced in this article, provide a richer picture and more informative account of the evolution of repeat buying, and so will hopefully serve as a fruitful tool for new avenues of research into longitudinal consumer behaviour.

For future research, researchers can modify the assumptions of the tri-variate NBD to see if the prediction accuracy can be improved. Modelling researchers could develop tri-variate condensed NBD (Chatfield & Goodhardt, 1973), tri-variate BBD (Morrison & Schmittlein, 1988), or tri-variate Poison lognormal (Trinh et al., 2014) and compare with the tri-variate NBD. Another potential extension is integrating a tri-variate NBD of category purchases with a Dirichlet model of brand choice (Goodhardt et al., 1984) to predict buyer flows across brands. For example, of those brand A loyal/heavy buyers in period 1 that become 'lost' buyers of the category in period 2, but come back in period 3, are they still brand loyal/heavy buyers of brand A or turning to disloyal/light buyers of brand A while becoming loyal/heavy buyers of brand B? A rich research agenda beckons for those studying persistence in buyer behaviour.

Disclaimer

1. Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business.
2. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

6. References

Bogomolova, S., Anesbury, Z., Kapulski, N., Lockshin, L., & Bogomolov, T. (2019, January). Exploring the incidence and antecedents of buying an FMCG brand and UPC for the first time. *Journal of Retailing & Consumer Services, 46*, 121-129. https://doi.org/http://dx.doi.org/10.1016/j.jretconser.2017.05.010

Chatfield, C., Ehrenberg, A. S. C., & Goodhardt, G. J. (1966). Progress on a simplified model of stationary purchasing behaviour. *The Journal of the Royal Statistical Society, 129*(3), 317-367. https://doi.org/10.2307/2343502

Chatfield, C., & Goodhardt, G. J. (1973). A consumer purchasing model with Erlang inter-purchase times. *Journal of the American Statistical Association, 68*(344, December), 828-835.

Dawes, J. (2014). Cigarette brand loyalty and purchase patterns: an examination using US consumer panel data. *Journal of business research, 67*(9), 1933-1943.

East, R., & Hammond, K. (1996). The Erosion of Repeat-Purchase Loyalty. *Marketing Letters, 7*(2), 163-171.

Ehrenberg, A. (1988). *Repeat-buying: Facts, theory and applications*. Oxford University Press.

Ehrenberg, A. S. C. (1959, March). The pattern of consumer purchases. *Applied Statistics, 8*(1), 26-41. https://doi.org/10.2307/2985810

Ehrenberg, A. S. C., Uncles, M. D., & Goodhardt, G. J. (2004, December). Understanding brand performance measures: using dirichlet benchmarks. *Journal of business research, 57*(12), 1307-1325. https://doi.org/10.1016/j.jbusres.2002.11.001

Fader, P. S., & Hardie, B. G. S. (2002). A note on an integrated model of customer buying behavior. *European Journal of Operational Research, 139*(3), 682-687.

Goodhardt, G., & Ehrenberg, A. (1967). Conditional trend analysis: A breakdown by initial purchasing level. *Journal of marketing research, 4*(May), 155-161.

Goodhardt, G. J., Ehrenberg, A., & Chatfield, C. (1984, October). The Dirichlet: A comprehensive model of buying behaviour. *Journal of the Royal Statistical Society, 147*(5), 621-643. https://doi.org/10.2307/2981696

Gruneklee, N., Rundle-Thiele, S., & Kubacki, K. (2016). What can social marketing learn from Dirichlet theory patterns in a physical activity context? *Marketing Intelligence & Planning, 34*(1), 41-60. https://doi.org/ https://doi.org/10.1108/MIP-12-2014-0233

InfoScout. (2016, 23 March 2016). *How To - Buyer Flow*. InfoScout. Retrieved 8th April 2021 from https://infoscout.zendesk.com/hc/en-us/articles/207988453-How-To-Buyer-Flow

Lam, D., & Mizerski, R. (2009). An investigation into gambling purchases using the NBD and NBD-Dirichlet models. *Marketing Letters, 20*(3), 263-276.

Lee, R., Rungie, C., & Wright, M. (2011). Regularities in the consumption of a subscription service. *Journal of Product & Brand Management, 20*(3), 182-189.

Lees, G., & Wright, M. (2009). Do the NBD Dirichlet's empirical generalisations hold true for radio listening? Australian and New Zealand Marketing Academy Conference Proceedings,

Lenk, P. J., Rao, A. G., & Tibrewala, V. (1993). Nonstationary Conditional Trend Analysis: An Appplication to Scanner Panel Data. *Journal of marketing research, 30*(August), 288-304.

Ludwichowska, G., Romaniuk, J., & Nenycz-Thiel, M. (2017). Systematic response errors in self-reported category buying frequencies. *European Journal of Marketing, 51*(7/8), 1440-1459.

Mizerski, D., Miller, R., Mizerski, K., & Lam, D. (2004). The stochastic nature of purchasing a State's Lottery products. *Australasian Marketing Journal, 12*(3), 56-69.

Morrison, D. G. (1969). Conditional trend analysis: A model that allows for nonusers. *Journal of marketing research, VI*(August), 342-346.

Morrison, D. G., & Schmittlein, D. C. (1988, April). Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *Journal of Business & Economic Statistics, 6*(2), 145-159.

Neff, J. (2011). Catalina: Major Packaged-Goods Brands Lost 46% of Loyalists: Firm Finds Companies Can Maximize Revenue Growth by Maintaining Brand Faithful. Retrieved 22 December, from https://adage.com/article/news/catalina-major-packaged-goods-brands-lost-46-loyalists/229640

Pointer Media Network. (2009). *Losing loyalty: the consumer defection dilemma*.

Power Up Your Marketing. (2017, 7th July 2017). *Where are your clients on the loyalty ladder?* Power Up Your Marketing. Retrieved 8th April 2021 from https://powerupyourmarketing.com/where-are-your-clients-on-the-loyalty-ladder/

Romaniuk, J. (2013). Modeling mental market share. *Journal of business research, 66*(2), 188-195.

Romaniuk, J., & Stocchi, L. (2009). *Understanding and Predicting Brand Attribute Frequency Distributions* ANZMAC, Melbourne, Australia.

Romaniuk, J., & Wight, S. (2015). The stability and sales contribution of heavy buying households. *Journal of Consumer Behaviour, 14*(1), 13-20. https://doi.org/10.1002/cb.1490

Schmittlein, D. C., Bemmaor, A. C., & Morrison, D. G. (1985). Why does the NBD model work? robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Science, 4*(No. 3, Summer), 255-266.

Schmittlein, D. C., & Morrison, D. G. (1983). Prediction of Future Random Events with the Condensed Negative Binomial Distribution. *Journal of the American Statistical Association, 78*(No. 382), 449-456.

Sharp, B. (2010). *How Brands Grow*. Oxford University Press.

Sharp, B. (2011). US brands are not losing their loyal customers – even more misleading metrics. *Marketing Science: Commentary by Professor Byron Sharp*. https://byronsharp.wordpress.com/2011/09/09/brands-are-not-losing-their-loyal-customers/

Thompson, J. (2020). *What Is Leaking Bucket Marketing?* (Marketing, Issue. https://smallbusiness.chron.com/leaking-bucket-marketing-31002.html

Trinh, G. (2018). The attendance at sporting events: A generalised theory and its implications. *International Journal of Market Research, 60*(3), 232-237.

Trinh, G., Corsi, A. M., & Lockshin, L. (2019). How country of origins of food products compete and grow. *Journal of Retailing and Consumer Services, 49*, 231-241.

Trinh, G., Khan, H., & Lockshin, L. (2020, August). Purchasing behaviour of ethnicities: Are they different? *International Business Review, 29*(4). https://doi.org/10.1016/j.ibusrev.2018.06.002

Trinh, G., & Lam, D. (2016). Understanding the attendance at cultural venues and events with stochastic preference models. *Journal of business research, 69*(9), 3538-3544. https://doi.org/https://doi.org/10.1016/j.jbusres.2016.01.033

Trinh, G., Rungie, C., Wright, M., Driesener, C., & Dawes, J. (2014). Predicting future purchases with the Poisson log-normal model. *Marketing Letters, 25*, 219-234.

Uncles, M. D., & Wright, M. (2004). Editorial: empirical generalisation in marketing [Editorial]. *Australasian Marketing Journal, 12*(3), 5-12.

Wilkinson, J. W., Trinh, G., Lee, R., & Brown, N. (2016). Can the negative binomial distribution predict industrial purchases? *Journal of Business and Industrial Marketing, 31*(4), 543-552.

Yu, J. (2013). Penetration is paramount, but it is also a leaky bucket. Retrieved 2 December, from https://www.kantarworldpanel.com/global/News/Penetration-is-paramount-but-it-is-also-a-leaky-bucket-