Short Communication

A note on an integrated model of customer buying behavior

Peter S. Fader [a], Bruce G.S. Hardie [b,*]

[a] *The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6371, USA*
[b] *London Business School, Sussex Place, Regent's Park, London NW1 4SA, UK*

Abstract

We propose a simple benchmark model for the integrated stochastic model of buying behavior developed in the article "Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior" [Eur. J. Oper. Res. 127(1) (2000) 109–119]. Re-examining the previously analyzed data covering the purchasing of tea, we find that the new benchmark model – which involves merely three parameters and can be estimated entirely within a standard spreadsheet environment – outperforms the original integrated model and provides clearer, more complete answers to the managerial questions posed at the outset of the earlier paper. © 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

Wu and Chen (2000) develop an integrated stochastic model of buying behavior which simultaneously captures the effects of regularity in interpurchase times, in-store buying decisions, and repeat buying behavior. One characteristic of this model is that it acknowledges that there may be a group of customers who can be classified as one-time buyers. The model is applied to customer purchase data for tea, and is shown to perform better than a set of benchmark models.

In this paper we develop a simple alternative benchmark model that explicitly allows for one-time buyers. We find that it fits the Wu–Chen tea dataset very well; in fact, the fit is better than that associated with the original integrated model. We then use this model to answer some relevant managerial questions raised by Wu and Chen at the outset of their paper. We conclude with a few brief comments about the guiding role of benchmark models in the model-development process.

2. The NBD with one-time buyers

In some situations, we are faced with a dataset in which there appears to be a larger-than-expected proportion of people making one and only one purchase. A natural inference that can be

* Corresponding author. Tel.: +44-20-7706-6795; fax: +44-20-7724-1145.

E-mail address: bhardie@london.edu (B.G.S. Hardie).

drawn from such a dataset is that an unknown proportion of these people can be viewed as one-time buyers. That is, they make a purchase in the sampled period but will never appear in the market again. As Wu and Chen point out, some of the buyers may be tourists who will never visit a given store again. Alternatively, some of the buyers may be merely responding to a particular promotion but have no intent of ever repurchasing the product. Clearly it is inappropriate to model the purchasing behavior of such people using a standard stochastic model of buyer behavior, such as the negative binomial distribution (NBD), which does not allow for buyers to come in and out of the sample in this manner.

When faced with such data, it may be more appropriate to use a model based on the following assumptions. Let us assume that an unknown proportion, ω, of the customers in the sample are one-time buyers (i.e., they disappear from the market completely after making one purchase). If we assume that the remaining proportion, $1 - \omega$, of the sample behave in a Poisson manner with the rate parameter distributed according to a gamma distribution (i.e., NBD), the aggregate distribution of purchases is given by

$$P(X = x) = \omega\delta_{x,1} + (1 - \omega)P_{\mathrm{NBD}}(X = x), \quad x$$
$$= 0, 1, 2, \ldots, \tag{1}$$

where $\delta_{x,1}$ is the Kronecker delta ($\delta_{x,1} = 1$ if $x = 1$, 0 otherwise) and $P_{\mathrm{NBD}}(X = x)$ is the NBD probability mass function,

$$P_{\mathrm{NBD}}(X = x) = \frac{\Gamma(r + x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha + 1}\right)^r\left(\frac{1}{\alpha + 1}\right)^x.$$

We call this the "NBD with one-time buyers" (NBD/OTB) model. [1] The mean and variance of the NBD/OTB are given by

$$E(X) = \omega + (1 - \omega)\frac{r}{\alpha}, \tag{2}$$

and

$$\mathrm{var}(X) = (1 - \omega)\left[\frac{r}{\alpha} + \frac{r}{\alpha^2}\right] + \omega(1 - \omega)\frac{(r - \alpha)^2}{\alpha^2}. \tag{3}$$

(When $\omega = 0$, (2) and (3) collapse to their NBD counterparts. Similarly, in the degenerate case of $\omega = 1$, $E(X) = 1$ and $\mathrm{var}(X) = 0$.)

The parameters of the NBD/OTB can be estimated using the method of maximum likelihood. Let us assume we have data of the form x_i, $i = 1, \ldots, I$, where x_i is the number of observed counts for individual i in an observation period of unit length. By definition, the likelihood function is the joint density of the observed data. Assuming the observations comprise an independent random sample, this is the product of the NBD/OTB probabilities for each x_i. Equivalently, the log-likelihood function is given by

$$\mathrm{LL}(r, \alpha, \omega \,|\, \mathrm{data})$$
$$= \sum_{x=0}^{x^*} f_x \ln[P(X = x \,|\, r, \alpha, \omega)], \tag{4}$$

where $x^* = \max(x_1, x_2, \ldots, x_I)$ and f_x is the number of individuals with an observed count of x purchases. Using standard numerical optimization routines, we find the values of r, α, and ω that maximize this log-likelihood function, subject to the obvious constraint that $0 \leqslant \omega \leqslant 1$. These are the maximum likelihood estimates of the three NBD/OTB model parameters. Standard errors can be computed in the usual manner.

If the distribution of purchases in Period 1 (of unit length) for a sample of customers is characterized by the NBD/OTB model, where the proportion of one-time buyers is ω, the distribution of their purchases in a non-overlapping Period 2 (also of unit length) will be characterized by the "NBD with spike-at-zero" model (Morrison, 1969), where ω is the size of the "spike-at-zero". The excess number of zeros in Period 2 follows from the fact that, by definition, none of the one-time buyers from Period 1 will ever make another purchase.

Conditional expectations such as $E(X_2 \,|\, X_1 = x)$ (i.e., the expected number of purchases in Period 2

---

[1] A behaviorally richer model would result by assuming that, in any period, the market can be characterized by two segments, "natives" and "tourists", of size π and $1 - \pi$, respectively. The buying behavior of the "natives" can be represented by the NBD, while the "tourists" either buy (with probability γ) or do not buy (with probability $1 - \gamma$) during the period in which they are in the market. We have found that such a model is empirically indistinguishable from the NBD/OTB model (i.e., $\hat{\gamma} = 1$).

given that x purchases were made in Period 1) are perhaps the most managerially relevant property of simple stochastic models such as the NBD (Morrison and Schmittlein, 1981).

Let us consider a setting in which the NBD/OTB has been applied. Suppose we observe an individual who made x purchases in Period 1; how many purchases do we expect this person to make in Period 2? If $x \neq 1$, the person is clearly not one of the one-time buyers and therefore the standard NBD conditional expectation (Morrison and Schmittlein, 1981) applies (i.e., $E(X_2|X_1 = x) = (r + x)/(\alpha + 1)$). On the other hand, if $x = 1$, the person could either be a one-time buyer (in which case $E(X_2|X_1 = 1) = 0$) or a regular buyer who just happened to make exactly one purchase during the observation period (in which case $E(X_2|X_1 = 1)$ would follow the standard NBD conditional expectation). Therefore $E(X_2|X_1 = 1)$ is $(r + 1)/(\alpha + 1) \times P$(customer is not a one-time buyer). By Bayes' theorem, the probability that a given customer with $x = 1$ is a one-time buyer is

$$P(\text{OTB}) = \frac{\omega}{\omega + (1 - \omega)P_{\text{NBD}}(X = 1)}.$$

Therefore, the conditional expectation for the NBD/OTB model is given by

$$E(X_2|X_1 = x)$$
$$= \frac{r + x}{\alpha + 1}\left[1 - \frac{\omega\delta_{x,1}}{\omega + (1 - \omega)[r/(\alpha + 1)][\alpha/(\alpha + 1)]^r}\right].$$
$$(5)$$

To put this new model in its proper context, we should emphasize several relevant aspects of it and the associated dataset. First, the Wu–Chen model is far richer (and far more complex) because it attempts to accommodate deeper behavioral issues such as the regularity of purchase timing, "learning effects" in repeat buying, customer departure (i.e., attrition), and in-store marketing conditions. Thus there is no reason to expect, a priori, that the three-parameter model posited here would perform nearly as well as the Wu–Chen specification. We limit ourselves to our simple model not only because it serves as a natural benchmark to the Wu–Chen framework, but also due to the fact that the histogram they provide (Table 1 of their paper)

does not include any of the covariate data or other information that would be required to construct (or replicate) a model that resembles theirs. Although we are somewhat limited for this reason, it is useful to formulate a model that requires nothing more than the histogram data, since many other actual case studies also share a similar limitation in terms of data availability. A "histogram-only" model is a useful contribution, especially if it can perform sufficiently well compared to a more detailed model such as that of Wu and Chen.

3. Empirical analysis

As just noted, we examine the NBD/OTB model using the dataset reported in Table 1 of Wu and Chen (2000), which gives the frequency of purchasing of Ten Ren tea by a sample of 1366 customers drawn from a specialty store. We estimate the parameters of the NBD/OTB model by the method of maximum likelihood using (4); we also estimate the parameters of the NBD model by constraining ω to zero. The maximum values of the log-likelihood functions and the corresponding parameter estimates are as follows: [2]

	NBD/OTB	NBD
r	0.507	0.590
α	0.122	0.168
ω	0.203	0
LL	−3077.7	−3194.0

On the basis of log-likelihood, the NBD/OTB offers a significantly better fit than the NBD. From Fig. 1, it is clear that the NBD/OTB provides a good fit across the full range of the histogram. More formally, this fit is confirmed by a standard chi-square goodness-of-fit test ($\chi^2 = 24.8$, $\chi^2_{0.05,16} = 26.3$). It is interesting to note that, on the basis of this same test, Wu and Chen's integrated model does *not* adequately fit the observed data

---

[2] A copy of the spreadsheet in which this analysis was performed is available at http://brucehardie.com/pmnotes.html.
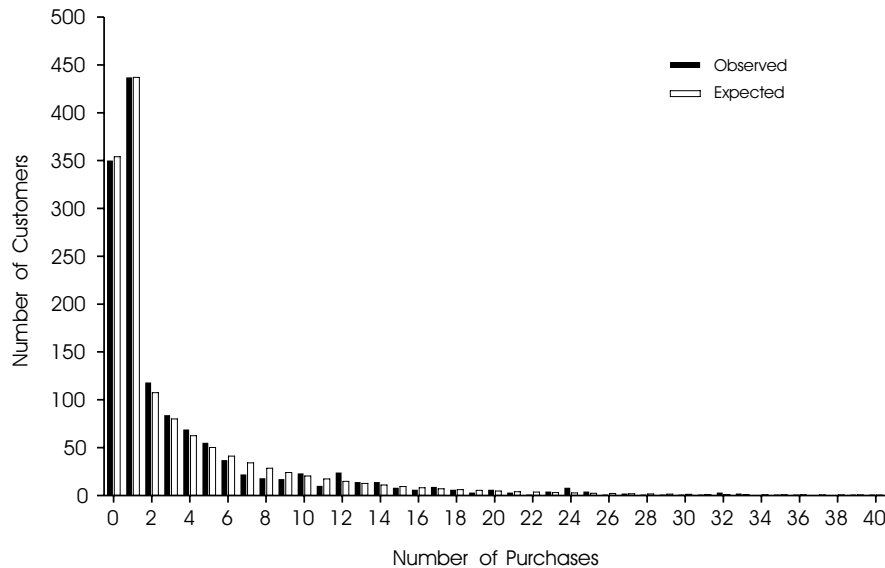
Fig. 1. Observed versus expected frequency of purchases for the NBD/OTB model.

($\chi^2 = 51.6$, $\chi^2_{0.05,2} = 6.0$). [3] The better fit associated with the NBD/OTB is also reflected in Theil's U: $U = 0.0465$ versus $U = 0.0649$ for the integrated model.

4. Discussion

In light of our earlier observation about the relative size and richness of the Wu–Chen model compared to our model, these top-line results are nothing short of remarkable. The added complexity of the behavioral process put forth by those authors seems difficult to justify when such a capable alternative model is readily available.

Beyond these empirical results, our simple model also provides straightforward answers to the questions posed by Wu and Chen at the outset of their paper:

- How many customers would make purchases?
- How many customers only purchase once and never come back, and how many will return to make purchases again in the future?

- What level of transaction should be expected for those customers who purchase more than once, both individually and collectively?
- What level of ongoing company sales can be predicted?

These are interesting and important questions, yet Wu and Chen do not provide direct answers to any of them, and it is not immediately clear how their model can be utilized to address them. In contrast, the answers to these questions arise quite naturally from the NBD/OTB model, its parameters, and the conditional expectation formula shown earlier. For this cohort of 1366 customers, these answers can be briefly summarized as follows:

- For a period of length t "time units," where one unit is equal to 48 weeks (i.e., the size of the dataset used for model calibration), the fraction of customers that make at least one purchase can easily be shown to be equal to $\omega + (1 - \omega)[1 - (\alpha/(\alpha + t))^r]$. For instance, over a 2-year period (i.e., 104 weeks, or 2.17 time units), 81.9% of the customers would buy Ten Ren tea at least once; 24.8% of these customers would be one-time buyers.
- In the original dataset (from Table 1 of Wu and Chen), we observe that 437 of the 1366 panelists (i.e., 32.0%) made only one purchase. But the

---

[3] To satisfy the usual requirements of the test (i.e., expected frequency $\geqslant 5$) for both models, the data are right-censored at $x = 19$.

ω parameter from our model indicates that 20.3% of the sample can be classified as "one-time only" buyers. In other words, over a third of the observed one-time buyers are still in the market and would be likely to make future purchases.

- The conditional expectation formula (5) shows how future purchasing is expected to vary as a function of observed past behavior. Given the above parameter estimates, Fig. 2 provides a graphical representation of this equation for $x \leqslant 10$. With the exception of the $x = 1$ group, this is a simple linear relationship. Contrasting it to the dashed 45° line, there is clear evidence of a "regression to the mean" pattern for buyers with extreme numbers of purchases. Future expectations for the $x = 1$ group are obviously pulled down by the fact that most of these group members are expected to disappear from the market after their initial purchase. (Note that, with the exception of the $x = 1$ group, we are seeing regression to the mean of the NBD (i.e., r/α), not the mean of the NBD/OTB.)

- The original dataset contains a total of 4788 purchases for the full set of 1366 customers. According to our simple model, 20.3% (i.e., 277) of these customers will drop out after completing one purchase, but the remaining 79.7% will continue to

purchase at the same rates as before. Therefore, the ongoing level of aggregate sales to this cohort should be $4788 - 277 = 4511$ purchases for each 48-week period in the future, or 94 units per week.

In addition to these questions posed by Wu and Chen, a number of other useful diagnostics can also be obtained from the proposed model. One such example would be the ability to produce a Lorenz curve, which can capture so-called "80:20" customer concentration patterns.

However, in order to fully appreciate these quantitative summaries, it is important to revisit the behavioral aspects of the model. As noted earlier, we do not necessarily believe that our simple model captures the true, underlying behavioral process that actually generated the original data in the first place. But the fact that our model can provide these summary statistics with such ease is a compelling reason to work closely with this type of model as a solid benchmark. Actual data from a future time period can be compared to these model predictions, and the presence/nature of any significant deviations can be used as an effective guide in an iterative model development process. Examining the model's performance in a holdout setting like this would provide a much more rigorous test of its suitability
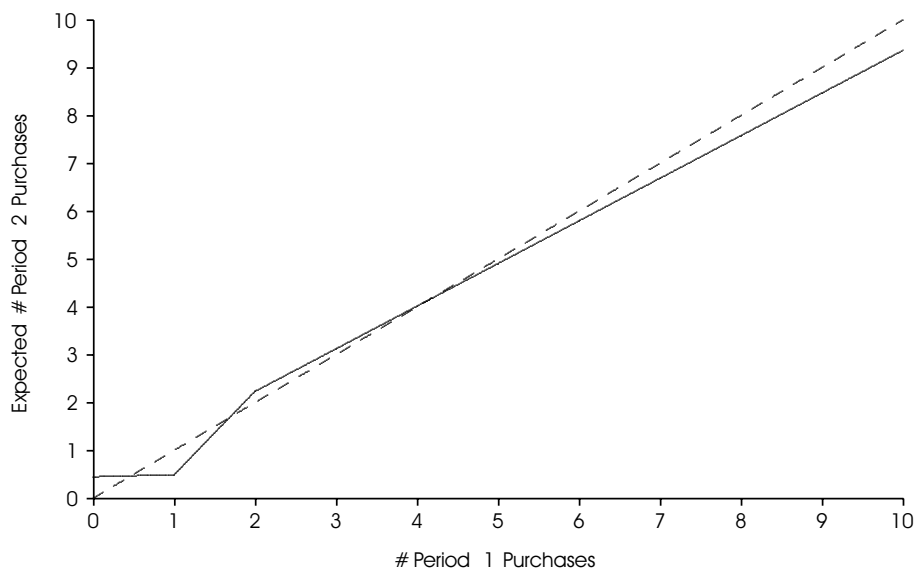

Fig. 2. Period 2 conditional expectations.

than merely examining in-sample fit statistics, as we (and Wu and Chen) have done with this particular dataset.

One lingering paradox is the idea that such a simple model can actually outperform a model that appears to be far more realistic and flexible. [4] For example, Wu and Chen go to great lengths to relax the typical assumption of memoryless exponential interpurchase times; they indicate that a more regular Erlang-5 process is warranted by the data. [5] Yet our results do not indicate the presence of any problems with the exponential assumption; perhaps the added complexity of the Wu–Chen model forced the Erlang-5 structure to counterbalance the unforeseen effects of another component of their model. There is a fair amount of literature (e.g., Ehrenberg, 1988; Morrison and Schmittlein, 1988) indicating that the basic NBD model (and its rather simplistic assumptions) is very effective (in terms of model fit and conditional predictions) even when there is strong reason to believe that it might not hold true. However, the main point here is not to encourage blind reliance on the NBD and its various extensions, but to use such models to point out which assumptions should be changed and which types of model components should be added.

Wu and Chen deserve credit for raising an excellent set of managerial questions, sharing their data in a public forum, and calling attention to a relatively wide range of different models for comparative testing. Their openness on these issues invites the type of model-building exercise conducted in this paper, and hopefully will increase researchers' interest and willingness to carefully consider the right kinds of benchmark models for any given managerial problem. While it may be unusual to see a simple benchmark model perform so well from a comparative perspective, it is often the case that such models can fruitfully guide the modeling process towards robust, efficient solutions for many different types of problems in actual practice.

Finally, we should emphasize that we are not putting forth the NBD/OTB model as a new framework that can (or should) be broadly applied to other sets of count data. We have not systematically investigated the prevalence of datasets that require (or might benefit from) the addition of a "one-time buyer" component. But whenever an analyst has any reason to believe, a priori, that such a phenomenon might exist, then this model is worth considering as a very logical starting point.

References

Bass, F.M., Krishnan, T.V., Jain, D.C., 1994. Why the Bass model fits without decision variables. Marketing Science 13 (Summer), 203–223.

Ehrenberg, A.S.C., 1988. Repeat Buying, second ed. Griffin, London (reprinted in the Journal of Empirical Generalisations in Marketing Science 5, 392–770, available at the web site http://www.empgens.com/).

Morrison, D.G., 1969. Conditional trend analysis: A model that allows for nonusers. Journal of Marketing Research 6 (August), 342–346.

Morrison, D.G., Schmittlein, D.C., 1981. Predicting future random events based on past performance. Management Science 27 (9), 1006–1023.

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

Wu, C., Chen, H.-L., 2000. Counting your customers: Compounding customer's in-store decisions, interpurchase time, and repurchasing behavior. European Journal of Operational Research 127 (1), 109–119.

---

[4] An interesting parallel can be drawn with the work of Bass et al. (1994) who show, in a diffusion modeling context, that having covariate information does not always improve model fit.

[5] They do note that the improvement in fit over that associated with exponential interpurchase times is small.