Simplified Methods of Fitting the Truncated Negative Binomial Distribution: A Model that Allows for Non-Users.

Abayomi A. Akomolafe[1] and Atinuke Akinyele[2].

[1]Department of Mathematical Sciences, Ondo State University of Science and Technology, PMB 353, Okitipupa, Nigeria.
[2]Department of Mathematical Sciences, West African Union University, Cotonou, Republic of Benin.

E-mail: akomolafe0@yahoo.com[1]
tinukemii@yahoo.com[2]

ABSTRACT

Retailers monitor customer buying-behavior as a measure of their stores' success. However, summary measures such as the total buying-behavior provides little insight about individual-level shopping behavior. Additionally, behavior may evolve over time, especially in a changing environment like the Internet.

This research developed a useful stochastic model for analyzing period to period fluctuations in sales thereby generalizing the model proposed by Goodhardt and Ehrenberg to allow for non-buyers of the product category. So as the composition of the customer population changes (e.g., as customers mature or as large numbers of new and inexperienced Internet shoppers enter the market), the overall degree of buyer heterogeneity that each store faces may change.

A systematic bias in their simple negative binomial distribution [NBD] model is demonstrated. If the proportion of non-buyers is large, the simple model will be wrong. As a result, frequent buyers often comprise the preferred target segment. We find evidence supporting the fact that people who visit a store more frequently are more likely to buy. We also gives explicit formula and directions that allow a moderately analyst to perform his own conditional trend analysis.

(Keywords: buying behavior, negative binomial distribution, heterogeneity, conditional trend analysis)

INTRODUCTION

In the recent research, Goodhardt and Ehrenberg extensively used the bivariate negative binomial distribution in analyzing period-to- period consumer behavior. The consumers are segmented by the number of purchases they make in a second period. Using this approach, Goodhardt and Ehrenberg determined the source of a seasoned increased and the effect of a deal (there were more new buyers than expected; previous buyers increased their rate of consumption).

The purpose of this research is to generalize the research propounded by Goodhardt and Ehrenberg model to allow for hard core nonusers. We will show that the lack of an explicit inclusion of these consumers can results in systematically biased findings. Companies and establishments of profit making venture and essential services already purchasing panel data will be able to perform these analysis for virtually no extra computing cost. However, the man-hours required to learn this kind of analysis may be quite costly.

METHODOLOGY

Conditional trend analysis used the quantity $E(R_2/r_1 \text{ in } t_1)$ which is the expected number of purchases an individual will make in period 2 given that he made r_1 purchases in t_1. In calculating $E(R_2/r_1 \text{ in } t_1)$ we assume that the individual's purchase probabilities are distributed according to a unimodal distribution. In each time period, each household can purchase the brand 0, 1, 2……….n times. Each household may also buy more than one brand in a given purchase occasion, although, this is not considered here as we concentrate on a single brand.

The model for the data is that each household has a fixed propensity to buy the brand. Buying behavior is modeled as a Poisson random variable (number of purchases over a time period) with the parameter λ denoting propensity to buy. Each household may have different λ and these parameters are modeled as being distributed as a Gamma random variable with parameters α and β. This can be interpreted as a Bayesian model of a Poisson random variable with a Gamma conjugate prior giving a Negative Binomial Posterior distribution. This is not however, mentioned the book repeat buying and in many other treatments of the negative binomial in the marketing literature.

The main driver of the probabilistic foundation of Ehrenberg's model is the Negative Binomial distribution called '' NBD'', average purchase rate in period 2 remain the same as those in period 1. We do this to isolate (by comparing the observed values and the estimated no-trend norms) those consumers responsible for the trend in the overall average buying rate or sales level. Obviously, the critical point in this kind of analysis is the calculation of E $(R_2/r_1 \text{ in } t_2)$. For this we need to know how R_2, the number of purchases in period 2 is affected by r_1, the actual number of purchases can be described by a negative binomial distribution (NBD), these calculations are straight forward.

Generating the NBD

The theory of Negative binomial distribution assumes that the proportion of people who purchase 0, 1, 2……units in a fixed time t_1 follows the NBD. This can occur if each person has a constant probability $\lambda\,\Delta t$, and the purchase intensity λ is distributed according to gamma distribution among individual consumers. Less formally this means that a given individual has the same propensity to purchase the product during each separate units of time (e.g. one week).

The gamma assumption implies. The gamma distribution is flexible and a good approximation to almost any reasonable unimodal distribution. We use the upper case R_2 to indicate that we are dealing with a random variable; we use the lower case r_1 to indicate the numerical outcome of the random variable R_1, the number of purchases in period 1. In a more precise formulation, we assume that:

(a) Each person's buying behavior is a Poisson process with intensity λ and

(b) λIs distributed according to the gamma distribution among individuals in the population.

Conditional Trend Analysis

That Ehrenberg and his colleagues have shown to fit purchase data from numerous product categories. In a more precise formulation, we assume that: (a) each person's buying behavior is a Poisson process with intensity λ and (b) $\lambda$is distributed according to the gamma distribution among individuals in the population, that is:

$$f_Y\ (\lambda / r't')=)\frac{e^{-\lambda t'}(\lambda t')^{r'-1}}{(r'-1)!}$$

$$E(\lambda) = \frac{r'}{t'}$$

$$\text{Var } (\lambda)= \frac{r'}{(t')2}$$

When conditions (a) and (b) are satisfied, the satisfied, the proportion of individuals who purchase 0, 1,2,3…. Units has the NBD,

$$p\ (x) = (r' + \frac{x-1}{x}\ \left[\frac{t'}{1+t'}\right] r' \left[\frac{1}{1+t'}\right] x \qquad x = 0,1,2, \ldots\ldots$$

The main and variance of this distribution are:

$$\text{E[X]}=\frac{r'}{t'}$$

$$\text{Var [X]}= r'\ (\frac{1+t'}{(t')^2})$$

Calculating $E[R_2 / r_1 \text{ in } t_1]$.

For this simple model, it is easy to calculate the expected number of purchase period 2, given an individual made r_1 purchase in period 1. A person who made a high number of purchases in period 1 is likely to be consumer with a high purchase intensity (a high value of λ). We can update our knowledge on his λ value by the usual Bayesian methods. Assuming that period 1 is one unit in time (which is necessary when estimating r' and t') we obtain:

(1) $$E[R_2/r_1 \text{ in } t_1]= \frac{r'+r_1}{t'+1}$$

If Period 2 is the same length as period 1. If Period 2 is of different length we obtain,

(1t) $E[R_2/r_1 \text{ in } t_1] = \frac{r' + r_1 t_2}{t' + 1t_1}$

Moreover, the conditional random variable R_2 is also a negative binomial random variable with parameters

$$r'' = r' + r_1$$
$$t'' = (t' + 1)t_1$$

Estimating r' and t'.

If the manager can estimate the two [parameters of the model, he can perform his own conditional trend analysis with Equation 1t. this can be easily done. To estimate r' we solve the Equation (2). Where $\underset{x}{\leftarrow}$ is the sample proportion of non-buyers in period 1. The second parameter is formed by using the solution to (2) and formula:

(2) $t' = \frac{r'}{x}$

There are other methods for estimating the parameter of the NBD. However, the preceding method is efficient and is the one used by Ehrenberg and his colleagues in their extensive original work consumer purchase patterns.

EFFECT OF HARD CORE NON-BUYERS

One obvious generalization of the model as presented so far will make it more realistic. The gamma distribution on the purchase intensities is appropriate for those consumers that ever buy the product. However, in many situations the non-buyers in period 1 will include consumers who just happened not to buy and those consumers who have never bought and will never buy the product. These later buyers we will call hard core non-buyers.

In some product categories, it may be possible to separate these two consumer classes, but in others separation will not be feasible. We will now construct a model that allows for hard core non buyers. The development of this model is justified by the following considerations.

1. When there is a group of hard core non buyers, the simple negative binomial conditional trend model can give misleading results (to be demonstrated later).

2. The generalized model is more realistic and just as easy to understand and use.

<u>Generalized Model, Including Non-users</u>

This new model is defined by the following:

1. A proportion, of consumers is hard core non buyers.

2. The remaining proportion, 1-α, of consumers purchase randomly according to a Poisson process and the purchase intensities are distributed gamma with parameter r' and t'.

This model is then identical to the previous model except for the inclusion of hard core non buyers. It is fully defined once we have values for α, r' and t'.

<u>Properties of the Model</u>

The number of purchases has no well-known distribution. However, it is extremely easy to calculate this probability distribution. The potential buyers will have an NBD with parameters r' and t'. Letting x equal the number of purchases in period 1 we get:

$$P(x) = (1-\alpha)P_{NB}\ (r', t'),\ x = 1,2,3 \ldots\ldots$$

$$P(0) = \alpha + (1-\alpha)P_{NB}\ (x = 0/r', t'),$$

Where

$P_{NB}\ (x = 0/r', t')$, is the negative binomial probability of x purchases given the parameter values r' and t'. let $p_{0,\alpha \text{ and } \sigma^2}$ be the proportion of consumers who make zero purchases, the mean number of purchases and variance of the number of purchases respectively. For our new model we have:

(4) $\rho_{0=\alpha+(1-\alpha)(\frac{t'}{1+t'})^{r'}}$

(5) $\mu = (1-\alpha)\frac{r'}{t'}$

(6) $\sigma^2 = (1-\alpha)\frac{r'}{(t')^2}$ (αr' +1+t')

Conditional Trend Analysis for the Model

Obviously, if a consumer has made one or more purchases in period 1, he is a member of the 1-α proportion of buyers. Hence, the conditional trend analysis is the same as before. However, if a consumer has made zero purchases, he can be either a hard core non-buyer or a potential buyer who just happened not to buy. The conditional trend formulas that we obtain (assuming Period 1 is of unit length) are:

(7) $E[R_2/r_1 \text{ in } t_1] = \frac{r'+r_1 t_2}{t'+1 t_1}, \quad r_1 = 1,2,3 \ldots\ldots\ldots)$

$E[R_2/r_1 = 0 \text{ in } t_1]$

(7') $= \frac{r'}{t'+1}\left[1 - \frac{\alpha}{\alpha+(1-\alpha)\ \left(\frac{t'}{1+t'}\right)^{r'}}\right]\frac{t_2}{t_1}$

These two formulas are all we need to perform the conditional trend analysis once we estimate the three parameters.

Estimating α, r' and t'

Anscombe proposes five methods for estimating the two parameters of the NBD [1, section 3]. The two methods most thoroughly examined are simple and efficient. The first consists of setting the sample proportion of zeros equal to the theoretical proportion of zeroes and setting the sample mean equal to the theoretical mean. This yields the procedure discussed earlier for estimating r' and t' for the simple negative binomial model.

The second method sets the sample mean and variance equal to the theoretical mean and variance. We can combine these two procedures and obtain three estimating equations for our three parameters. From the properties of our model (Equations 4, 5, and 6) and letting $p_0,\ x \text{ and } s^2$ equal the sample proportion of zeroes, sample mean and sample variance respectively:

(8) $\rho_0 = \alpha + (1-\alpha)\left(\frac{t'}{1+r'}\right)^{r'}$

(9) $x = (1-\alpha)\frac{r'}{t'}$

(10) $s^2 = (1-\alpha)\left(\frac{r'}{(t')^2}\right)$ (αr' +1+ t')

Combining (9) and (10) we have:

(11) $r' = \frac{x^2}{s^2\ x + \alpha\,(x - s^2 - x^2)}$

Given a value for α we obtain r', which in turn yields,

(12) $t' = \frac{(1-\alpha)r'}{x}$

Since α is restricted to $0 \leqq \alpha \leqq 1$, we can try different values of α until we satisfy (8), the sample proportion of zeroes equation. We could use research techniques to solve this set of equations, but the problem does not really warrant that degree of sophistication. We could write a short computer program and in matter of seconds try (for example) α= .01, .02, ………, .98, .99 and see which value fits (8)

If we wish to solve the system of equations by hand, we would select a value of α, obtain r' and t' and put these three values into (8). If the right-hand side is greater than p_0, we will next try a smaller value for α and vice versa. After four or five attempts we will have a good answer. Once the three parameters are estimated, we perform the conditional trend analysis with Equation 7 and 7'.

Bias in using Simple Negative Binomial Distribution Model

Suppose that our model with the hard core non buyers is actually true, but we use the simple NBD model. That is, we estimate the parameters from (2) and (3) and then use (1) to perform the conditional trend analysis. What errors will we make?

Equation 2 does not yield an explicit solution (we must use trial and error or a series of approximations). Hence, we cannot obtain an analytical answer to this question. Therefore, we tried over 1,000 different combinations of α, r' and t; to determine what errors would result from incorrectly assuming the simple NBD model is true. We varied r' and t' from 0.1 to 5.0 (when r'=1, the gamma distribution on λ becomes exponential), and the α ranged from 0.1 to 0.9. In every case the same bias occurred. If the purchase patterns remained stationary, using the simple NBD model made it appear that light buyers (except [possibly the non-buyers) were buying more and heavy buyers were buying less. This bias becomes larger as α , the proportion of hard core non-buyers becomes larger as some examples will illustrate.

Assume that Period 2 is the same length as Period 1. The first column of the table is the ''true'' expected number of purchases din period 2 given r_1 purchases in period 1. The second column is the ''mistaken expected number of purchases. These are arrived at by using the true proportion of zeroes, (4), and the true mean (5) and inserting these numbers into the estimation Equations (2) and (3). This yields mistaken estimates of r' and t' (α is of course assumed to be zero), and then (1) is used to construct the mistaken expected number of purchases in period 2, assuming the purchase patterns remain stationary.

It would be incorrect to take these labels of ''true'' and ''mistaken too seriously. They only apply to the abstract model defined previously with characteristics given by Equation 4, 5 and 6. Empirically it is not obvious which, if either, the true model is. Therefore, in the table we will label the column '' Hardcore'' and ''Simple NBD''. In the first example α, the proportion of hard core non-buyers is one half. Some fairly large biases result. In the second example only one tenth of the consumers are hard core non-buyers; again there are biases, but not nearly as large.

The relevancy of these numerical examples can be demonstrated. The author was involved in a study on consumption changes of foil wraps over a two-year period using MRCA's 7,500s member National Consumer Panel. Of these 7,500, only 3,494 families purchased any foil over the two-year period. Clearly, there exists a sizable proportion (approximately half) of hard core non users.

Any model used to analyze quarter-to-quarter or year-to-year trends in consumption must incorporate these hard core non-users. For a product, such as foil there is no way to identify hard core non-buyers on the basis of demographic characteristics. Further, if no previous purchase histories for these families exists (as in this study), there is no valid means for screening families.

Obviously, any kind of conditional trend model must be explicitly including these hard core non-buyers. For many products (e.g., dog food) it will be possible to identify hard core non-buyers. However, there may still be some dog owners who only feed their dog's table scraps. Hence, the best procedure would seem to be eliminated of as many non-buyers as possible (we do not want the zero class to influence unduly the parameter estimates), and then apply the more general model. It is no more difficult to apply than the simple NBD. Moreover, if we have not been successful in screening out hard core non-buyers, biased results are avoided.

Implications of the Model

If there really exists a hard core of non-buyers, the estimates of α should be independent of the time period used. This criterion can be used as one test of the model. However, if we are dealing with particular brands within a product category, the α's associated with each brand need not to be similar. Clearly, a consumer can be a hard core non-user of Brand A and a user of Brand B. hence, the α for any particular brand should be independent of the time period used, but the α's across brands will be expected to vary. For the foil study mentioned earlier, we estimated α using both a six- month and a one-year period. For six months we obtained α=0.47; for one year, α= 0.45, of course, this is only one result for one product category; nevertheless, it is encouraging that the hard core remained fairly constant for two different time periods.

CONCLUSION

The conditional trend analysis model proposed by Goodhardt and Ehrenberg is highly useful. Their article [2] gives detailed examples which illustrate the value of the model. We have merely made a generalization to allow for hard core non-buyers. This generalized model is just as easy to use and avoids some biases that result from not explicitly including hard core non-buyers.

Any company already using panel data can apply this model for very little extra cost. To do this they must:

Table 1: Effect of Hard Core Non-buyers on Conditional Trend Analysis.

r	Hardcore expected purchases	Simple NBD expected purchases
50 percent hardcore non-buyers		
0	0.16	0.24
1	1.67	1.08
2	2.33	1.92
3	3.00	2.76
4	3.67	3.61
5	4.33	4.45
6	5.00	5.29
7	5.67	6.13
8	6.33	6.98
9	7.00	7.82
	α= 0.5	$\rho_0 = 0.596$
	r'= 1.5	μ= 1.5
	t'= 0.5	mistaken r'= 0.279
		mistaken t'= 0.186
10 percent hard core non-buyers		
0	0.63	0.72
1	1.67	1.45
2	2.33	2.19
3	3.00	2.92
4	3.67	3.65
5	4.33	4.39
6	5.00	5.12
7	5.67	5.85
8	6.33	6.59
9	7.00	7.3200
	α = 0.1	$\rho_0 = 0.273$
	r'= 1.5	μ= 2.7
	t'= 0.5	mistaken r'= 0.982
		mistaken t'= 0.364

Segment consumers by their number of purchases in Period 1, Calculate $\rho_{0,}\ x\ and\ s^2$,Estimate α, r' and t' from Equations 11, 12, and 8,Calculate the expected number of purchases in Period 2 from Equations 7 and 7'and Compare the actual number of purchases in Period 2 with the expected number of purchases. The analyst will now be in a position to determine which kind of consumer (heavy or light buyer) is causing the fluctuation in the overall sales level.

REFERENCES

1. Anscombe, F.J. 1995. "Sampling Theory of the Negative Binomial and Logarithmic Distribution'.' *Biometrika*. 37:358-382.

2. Brass, W. 1999. "Simplified Methods of Fitting the Truncated NBD". *Biometrika*. 45:59-68.

3. Chatfield, C. 2002. "Some Statistical Models for Buying Behaviour". Ph.D. Dissertation. London University: London, UK.

4. Ehrenberg, A.S.C. and G.J. Goodhardt. 1998. "A Comparison of American and British Repeat-Buying Habits". *Journal of Marketing Research*. 5:29-33.

5. Ehrenberg, A.S.C. and G.J. Goodhardt. 2007. "Conditional Trend Analysis: A Breakdown by Initial Purchasing Level". *Journal of Marketing Research*. 4:155-162.

6. Grahn, G.L. 2004. "The Negative Binomial Distribution Model of Repeat-Purchase Loyalty: An Empirical Investigation". *Journal of Marketing Research*, 5(February 1969):72-78.

7. Morrison, D.G. 2009. "Analysis of Consumer Purchase Data: A Bayesian Approach". *Industrial Management Review*. 9:31-40.

8. Raiffa, H. and R. Schlaifer. 1990. *Applied Statistical Decision Theory*. Division of Research, Harvard University: Boston, MA.

ABOUT THE AUTHORS

Abayomi A. Akomolafe, is a Lecturer in the Deptment of Mathematical Sciences, Ondo State University of Science and Technology, Okitipupa, Nigeria.

Atinuke Akinyele, is a Lecturer in the Department of Mathematical Sciences, West African Union University, Republic of Benin, Cotonou.

SUGGESTED CITATION

Akomolafe, A.A. and A. Akinyele. 2015. "Simplified Methods of Fitting the Truncated Negative Binomial Distribution: A Model that Allows for Non-Users". *Pacific Journal of Science and Technology*. 16(1):110-116.