Illustrating the Performance of the NBD as a Benchmark Model for Customer-Base Analysis

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie†
www.brucehardie.com

May 2004

1. Introduction

Several models for customer-base analysis, such as the Pareto/NBD model developed by Schmittlein, Morrison, and Colombo (1987) and the BG/NBD model developed by Fader, Hardie, and Lee (2004), extend the basic NBD model (Morrison and Schmittlein 1988) by allowing for a customer "dropout" process. Missing from the various empirical applications of these models is an explicit examination of the value of such a "dropout" process. In other words, what is the performance of these models compared to that of the basic NBD model? We undertake such an examination in this note, documenting the steps associated with the implementation of the NBD in a spreadsheet environment.

There are four key stages to the implementation of the NBD benchmark model: (i) estimating the model parameters, (ii) creating the expected frequency distribution of transactions given these parameter estimates, (iii) generating the aggregate sales forecast, and (iv) predicting a particular customer's future purchasing, given information about his past behavior and the parameter estimates. The specific steps are outlined in sections 3–6 below. Section 2 briefly describes the nature of the data used for model calibration. All these sections should be read in conjunction with the Excel spreadsheet `nbd_benchmark.xls`.

2. Data

The worksheet `Raw Data` contains data for a sample of 2357 CDNOW customers who made their first purchase at the web site during the first quarter of 1997. We have information on their repeat purchasing behavior up to the end of week 39 of 1997. We have two pieces of information about each customer's purchasing behaviour: his "frequency" (how many repeat transactions he made in a specified time period), and the length of time over which we have had an opportunity to observe any repeat purchasing behavior. The notation used to represent this information is $(X = x, T)$, where x is the number of transactions observed in the time period $(0, T]$.

While the basic unit of time is one week, we recognize that transactions can occur on each day of the week. Therefore consider customer 0001 (row 2). The number of days (in weeks) during which *repeat* transactions could have occurred is $T = 38.86$, which implies this customer made his first-ever purchase at CDNOW on the first day of the first week of 1997. Over this time period, this customer made $x = 2$ repeat transactions.

3. Parameter Estimation

Let $X(t)$ be the number of events occurring in the interval $(0, t]$. The NBD model is based on the following two assumptions:

- $X(t)$ is distributed Poisson distribution with mean λt, and
- transaction rates (λ) are distributed across the population according to a gamma distribution with parameters r and α.

It follows that

$$\begin{aligned} P(X(t) = x) &= \int_0^\infty \underbrace{\frac{(\lambda t)^x e^{-\lambda t}}{x!}}_{\text{Poisson}} \overbrace{\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}}^{\text{gamma}} d\lambda \\ &= \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \end{aligned} \tag{1}$$

So as to facilitate the comparison of model log-likelihood function values, we fit the timing-model equivalent of the basic NBD model (Gupta and Morrison 1991) to the CDNOW dataset. The associated likelihood function for a randomly-chosen customer with transaction history $(X = x, T)$ is

$$L(r, \alpha \mid X = x, T) = \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{\alpha}{\alpha+T}\right)^r \left(\frac{1}{\alpha+T}\right)^x \tag{2}$$

(Note that (1) and (2) only differ by a factor of $t^x/x!$, which is independent of r and α.)

Suppose we have a sample of N customers, where customer i had $X_i = x_i$ transactions in the interval $(0, T_i]$. The sample log-likelihood function is

$$LL(r, \alpha \,|\, \text{data}) = \sum_{i=1}^{N} \ln \big[L(r, \alpha \,|\, X_i = x_i, T_i)\big] \tag{3}$$

Our task is to "code up" this expression in an Excel worksheet and find the maximum likelihood estimates of the model parameters r and α by using the Solver add-in to find the maximum of this function.

We start by making a copy of the `Raw Data` worksheet — let's call it `"NBD" Estimation` — and inserting four rows at the top of the worksheet. The first thing we need to do is to create an expression for $\ln\big[L(r, \alpha \,|\, X = x, T)\big]$ for each of the 2357 customers in the sample. In order to create the corresponding expression in the spreadsheet without an error message appearing (e.g., `#NUM!` or `#DIV/0!`), we need some so-called starting values for r and α. Provided they are within the defined bounds $(r, \alpha > 0)$, the exact values do not matter. We start with 1.0 for both parameters and locate these values in cells `B1:B2`.

The log of the expression given in equation (2) is

$$\ln\big[\Gamma(r + x)\big] - \ln\big[\Gamma(r)\big] + r\ln(\alpha) - (r + x)\ln(\alpha + T)$$

The formula in cell `D6` is simply the expression of this for the first customer:

```
=GAMMALN($B$1+B6)-GAMMALN($B$1)
    +$B$1*LN($B$2)-($B$1+B6)*LN($B$2+C6)
```

We then copy this down to cell `D2362`.

The sum of these 2357 values is found in cell `B3`; this is the value of the log-likelihood function, equation (3), given the values for the two model parameters in cells `B1:B2`. (With starting values of 1.0 for both parameters, $LL = -14924.92$.)

Given these sample data, we find the maximum likelihood estimates of r and α by maximizing the log-likelihood function. We do this using the Excel add-in Solver, available under the "Tools" menu. The *target cell* is the value of the log-likelihood function (cell `B3`); we wish to *maximize* this by *changing* cells `B1:B2`. The *constraints* we place on the parameters are that both r and α are greater than 0. As Solver only offers us a "greater than or equal to" constraint, we *add* the constraint that cells `B1:B2` are $\geq$ a small positive number (e.g., 0.00001). Clicking the *Solve* button, Solver finds the values of r and α that maximize the log-likelihood function; these are the maximum likelihood estimates of the model parameters. The maximum value of the log-likelihood function is -9763.66, associated with $r = 0.385$ and $\alpha = 12.072$.

As noted in Fader et al. (2004), the log-likelihood function values for the Pareto/NBD and BG/NBD models are -9595.0 and -9582.4, respectively.

On the basis of the BIC model selection criterion, we conclude that both the BG/NBD and Pareto/NBD models provide a far better fit to the data than that of the basic NBD model, indicating the value of including a dropout process.

As an aside, let us also estimate the parameters of the NBD model using its standard counting form (as opposed to the timing equivalent considered above) — see the `NBD Estimation` worksheet.

For sample of N customers, where customer i had $X_i = x_i$ transactions in the period $(0, T_i]$,the log-likelihood function given by is

$$LL(r, \alpha \,|\, \text{data}) = \sum_{i=1}^{N} \ln \left[P(X(T_i) = x_i \,|\, r, \alpha) \right] \tag{4}$$

We first compute $P(X(t) = x)$, equation (1), for each customer; the formula in cell `D6` is simply this expression for the first customer:

```
=EXP(GAMMALN($B$1+B6)-GAMMALN($B$1))/FACT(B6)
    *($B$2/($B$2+C6))^$B$1*(C6/($B$2+C6))^B6
```

We then take the log of the probability in cell `E6`, and copy these two cells down to row 2362. The sum of cells `E6:E2362` is found in cell `B3`; this is the value of the log-likelihood function, equation (4), given the values for the two model parameters in cells `B1:B2`. (With starting values of 1.0 for both parameters, $LL = -8354.32$.) We find the maximum likelihood estimates of r and α by maximizing the log-likelihood function using Solver. The maximum value of the log-likelihood function is -3193.06, associated with $r = 0.385$ and $\alpha = 12.072$.

4. Predicted Distribution of Transactions

Another way of assessing the fit of a model is to compare the actual frequency distribution of transaction counts (how many people made 0, 1, 2, ... repeat transactions) with that predicted by the model.

Let f_x denote the number of people making x repeat transactions in the 39-week model calibration period (x = 0, 1, 2, ...). The actual frequency distribution of repeat transaction counts can easily be determined using the "pivot tables" feature in Excel. Going back to the `Raw Data` worksheet, we select the *PivotTable and PivotChart ...* under the *Data* menu. We use x as the *row field* and use ID as the *data item.* The resulting table is reported in the `Actual Frequency Distribution` worksheet. Using this full distribution, we create a right-censored distribution in which counts greater than 7 are collapsed into a 7+ bin.

In this particular example, the task of computing the expected frequency distribution is slightly complicated by the fact that the time period over which repeat transactions could have occurred varies across customers. Let

n_s is the number of customers who made their first purchase at CDNOW on day s of 1997 (and therefore have $t - \frac{s}{7}$ weeks within which to make repeat purchases). It follows that the expected number of people in this cohort of new customers with x repeat transactions is computed in the following manner:

$$E(f_x) = \sum_{s=1}^{84} n_s P(X(t - \tfrac{s}{7}) = x)\,,\ x = 0, 1, 2, \dots \tag{5}$$

The first step is to determine n_s, which we do using Excel's pivot table tool. We start by making a copy of the `Raw Data` worksheet — let's call it `n_s`. Given T, the number of days (in weeks) during which *repeat* transactions could have occurred in the 39-week calibration period, it follows that the "time of first purchase" (column `D`) is simply $39 - T$. Selecting the *PivotTable and PivotChart ...* under the *Data* menu, we use "time of first purchase" as the *row field* and use ID as the *data item*. We see that 18 people made their first-ever purchase at CDNOW on the first day of the first week of 1997, 22 on the second day of the first week of 1997, ..., and 30 on the seventh day of the twelfth week of 1997.

For the NBD applied to a observation period of length t, it is easy to show that we can compute the probabilities using the following forward recursion formula:

$$P(X(t) = x) = \begin{cases} \left(\dfrac{\alpha}{\alpha + t}\right)^r & x = 0 \\[2ex] \dfrac{(r + x - 1)t}{x(\alpha + t)} P(X(t) = x - 1) & x = 1, 2, \dots \end{cases}$$

with $P(X(t) \geq 7) = 1 = \sum_{x=0}^{6} P(X(t) = x)$

In columns `F–CK` of the `Histogram` worksheet, we use this formula to compute $P(X(T) = x)$ ($x = 0, 1, \dots, 7+$) for each of the 84 possible values of T, the time period during which repeat transactions could have occurred. The $E(f_x)$, equation (5), are computed in cells `C6:C13`.

The expected numbers of people with 0, 1, ..., 7+ repeat transactions in the 39-week model calibration period for the NBD, Pareto/NBD, and BG/NBD models are compared with the actual frequency distribution in Figure 1.

A quick visual inspection suggests that the fits of the three models are very close. On the basis of the chi-square goodness-of-fit test, we find that the BG/NBD model provides the best fit to the data ($\chi^2_3 = 4.82$, $p = 0.19$). In contrast to the comparison based on the log-likelihood function values, we find the NBD model fits the data surprisingly well ($\chi^2_5 = 10.27$, $p = 0.07$). Both models dominate the Pareto/NBD model ($\chi^2_3 = 11.99$, $p = 0.007$).

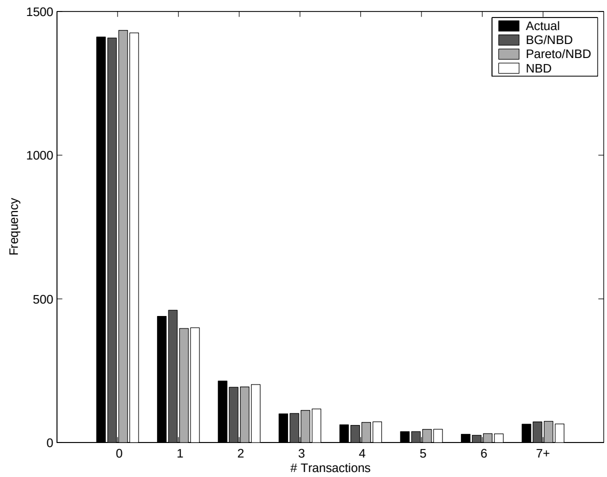

Figure 1: Predicted versus Actual Frequency of Repeat Transactions

5. Aggregate Sales Forecast

Any value in extending the basic NBD model by allowing for customer "dropout" is not immediately apparent from the above analysis. To further examine the NBD's performance as a benchmark model, we consider how well its predictions of total repeat transactions track the actual numbers over time, during both the 39-week calibration period and the subsequent 39-week "forecast" period.

For a randomly-chosen individual, the expected number of transactions in a time period of length t is given by the mean of the NBD,

$$E(X(t) \,|\, r, \alpha) = \frac{rt}{\alpha}\,.$$

However, we are not interested in the expected number of repeat transactions for a randomly-chosen individual; rather we are interested in tracking (and forecasting) the total number of repeat transactions by the cohort of customers. In computing this cohort-level number, we need to control for the fact that different customers made their first purchase at CDNOW at 84 different days during the first quarter of 1997. Therefore, total repeat transactions can be computed as follows:

$$\text{Total Repeat Transactions by } t = \sum_{s=1}^{84} \delta_{(t>\frac{s}{7})} n_s E[X(t-\tfrac{s}{7})]\,, \tag{6}$$

where n_s is the number of customers who made their first purchase at CDNOW on day s of 1997 and $\delta_{(t>\frac{s}{7})} = 1$ if $t > \frac{s}{7}$, 0 otherwise.

Referring to the worksheet `Aggregate Repeat Sales`, columns `J`–`CO` see us computing $E[X(t)]$ for *calendar time* $t = 1/7, 2/7, \ldots, 78$. (The formulae in cells `J6:CO551` shift calendar time to time since first purchase.) The final step in the evaluation of equation (6) is performed in cells `I6:I551`. Cells `C6:C83` report the cumulative number of repeat transactions for each of the 78 weeks (using the `=OFFSET()` function to "pick out" the relevant numbers from column `I`). The expected number of weekly repeat transactions are reported in cells `F6:F83`.

In Figure 2 we see how well the predictions of total repeat transactions from the NBD, Pareto/NBD, and BG/NBD models track the actual numbers over time. We immediately observe that the NBD not only fails to track actual sales in the 39-week calibration period, but also deviates significantly from the actual sales trajectory over the subsequent 39 weeks. By the end of June 1998, the NBD model is over-forecasting by 24%; in contrast, the Pareto/NBD and BG/NBD models under-forecast total cumulative repeat transactions by 2% and 4%, respectively.

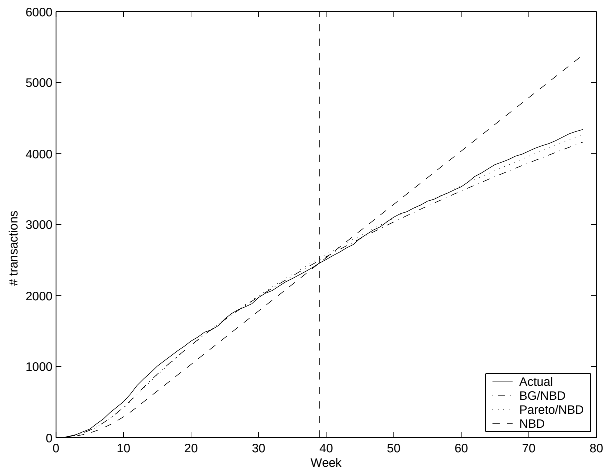

Figure 2: Predicted versus Actual Cumulative Repeat Transactions

The value of extending the basic NBD model by adding a "dropout" process becomes very clear when we examine the corresponding week-by-week repeat-transaction numbers (Figure 3). The sales figures rise through the end of week 12, as new customers continue to enter the cohort, but after that point it is a fixed group of 2357 eligible buyers. The NBD predicts a constant level of repeat transactions; in contrast, the Pareto/NBD and BG/NBD models predict a decline in the level of repeat transaction as customers become inactive. We note that the BG/NBD and Pareto/NBD models track the underlying trend in repeat-buying behavior, albeit with obvious deviations because of promotional activities and the December holiday season.

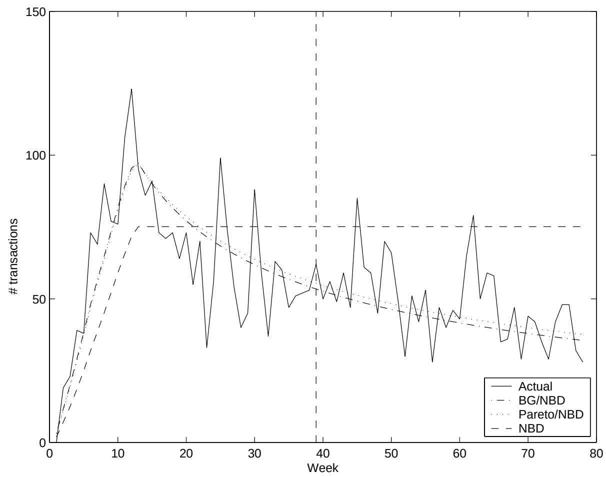

Figure 3: Predicted versus Actual Weekly Repeat Transactions

6. Conditional Expectations

Our final, and perhaps most critical, examination of the NBD's performance as a benchmark model focuses on the quality of individual-level predictions of future behaviour, conditional on the number of observed transactions in the model calibration period.

More specifically, we are interested in $E[Y(t) \,|\, \text{data}]$, the expected number of transactions in an adjacent period $(T, T+t]$, conditional on the observed purchase history. We call this the *conditional expectation.*

For the NBD model, a straight-forward application of Bayes' theorem gives us (Morrison and Schmittlein 1988)

$$E\big[Y(t)|X = x, T, r, \alpha\big] = \left(\frac{r+x}{\alpha+T}\right) t\,,$$

where $t = 39$ for our analysis. We compute this conditional expectation for each of the 2357 customers in the worksheet `Conditional Expectation`. Using the "pivot tables" feature, we then compute the average conditional expectation by the number of calibration-period repeat transactions (right-censoring at 7).

These NBD conditional expectations (as well as those associated with the Pareto/NBD and BG/NBD models) are reported in Figure 4, along with the average of the actual number of transactions that took place in the forecast period, broken down by the number of calibration-period repeat transactions. Once again, we see that the NBD model generates poor individual-level predictions of the expected number of transactions in weeks 40–78. Similarly, the BG/NBD and Pareto/NBD models provide excellent predictions.

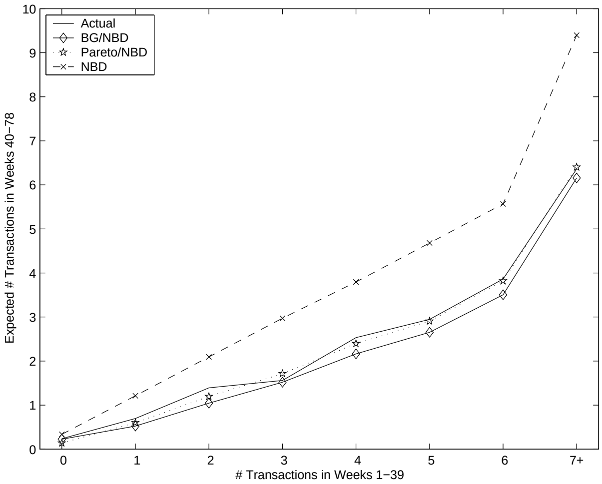

Figure 4: Conditional Expectations

7. Conclusions

In comparing the NBD, Pareto/NBD, and BG/NBD models in a customer-base analysis setting, we have been able to examine the value of extending the basic NBD model by allowing for customer "dropout".

- We conclude that both the BG/NBD and Pareto/NBD models provide a far better fit to the data than that of the basic NBD model, when model fit is assessed on the basis of the value of the log-likelihood function.

- However, the conclusion is not so clear when we examine the fit of the expected frequency distribution of transactions (as computed using each model) to the actual distribution of repeat transaction counts— the fit of the NBD dominates that of the Pareto/NBD.

- When considering the ability of each model to track the total number of repeat transactions (for the whole cohort of customers) over time, during both the 39-week calibration period and the subsequent 39-week "forecast" period, the poor performance of the basic NBD model is clearly illustrated in Figures 2 and 3.

- Turning from aggregate-level predictions to individual-level predictions, we find that the conditional expectations of future purchasing generated using the NBD are consistently higher than the actual numbers, whereas the BG/NBD and Pareto/NBD models provide excellent predictions.

We can therefore conclude that there is value in extending the basic NBD model by allowing for customer "dropout". We would encourage researchers involved in the development of models for customer-base analysis to use models such as the Pareto/NBD and BG/NBD as benchmark models in their analyses. Furthermore, we would encourage them to compute and examine the types of performance measures and managerial diagnostics (i.e., tracking/forecasting aggregate purchasing and individual-level conditional expectations) used in this note.

References

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2004), “"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model,” working paper.

Gupta, Sunil and Donald G. Morrison (1991), “Estimating Heterogeneity in Consumers’ Purchase Rates,” *Marketing Science*, **10** (Summer), 264–269.

Morrison, Donald G. and David C. Schmittlein (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** (April), 145–159.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), “Counting Your Customers: Who They Are and What Will They Do Next?” *Management Science*, **33** (January), 1–24.