# Incorporating Time-Invariant Covariates into the Pareto/NBD and BG/NBD Models

Peter S. Fader
www.petefader.com

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

August 2007

## 1 Introduction

This note documents how to incorporate the effects of **time-invariant** covariates into the Pareto/NBD and BG/NBD models. It is assumed that the reader is familiar with the derivations of these models, as presented in Fader and Hardie (2005) and Fader et al. (2005).

Let $\mathbf{z}_1$ be the vector of time-invariant covariates that are assumed to explain some of the cross-sectional heterogeneity in the purchasing process, and $\mathbf{z}_2$ be the vector of time-invariant covariates that are assumed to explain some of the cross-sectional heterogeneity in the dropout process. (While it will typically be the case that $\mathbf{z}_1 = \mathbf{z}_2$, this equality is not assumed in these derivations.) Note that since these covariates are specific to each individual, both $\mathbf{z}_1$ and $\mathbf{z}_2$ should have an individual-specific subscript $i$; this has been suppressed for notational convenience.

The basic results are as follows:

- For the Pareto/NBD model, we simply replace $\alpha$ and $\beta$ with

$$\begin{aligned}\alpha &= \alpha_0 \exp\left(-\boldsymbol{\gamma}_1'\mathbf{z}_1\right)\\ \beta &= \beta_0 \exp\left(-\boldsymbol{\gamma}_2'\mathbf{z}_2\right)\end{aligned}$$

  where $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ capture the effects of these two vectors of covariates; $r$ and $s$ remain unchanged.

- For the BG/NBD model, we simply replace $\alpha$, $a$, and $b$ with

$$\begin{aligned}\alpha &= \alpha_0 \exp\left(-\boldsymbol{\gamma}_1'\mathbf{z}_1\right)\\ a &= a_0 \exp\left(\boldsymbol{\gamma}_2'\mathbf{z}_2\right)\\ b &= b_0 \exp\left(\boldsymbol{\gamma}_3'\mathbf{z}_2\right)\end{aligned}$$

where $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ and $\boldsymbol{\gamma}_3$ capture the effects of these two vectors of covariates; $r$ remains unchanged.

## 2 Preliminaries

A popular, easily interpretable method for incorporating the effects of exogenous covariates in event-time models is the proportional hazards approach. In this framework, the covariates have a multiplicative effect on the hazard rate. More specifically, let $F_0(t|\theta)$ be the so-called "baseline" cdf for the distribution of an individual's interpurchase times, and $f_0(t|\theta)$ and $h_0(t|\theta)$ the associated pdf and hazard rate function. The most common formulation of the proportional hazards specification states that

$$h(t|\theta, \boldsymbol{\gamma}, \boldsymbol{z}_i) = h_0(t|\theta)\exp(\boldsymbol{\gamma}'\boldsymbol{z})$$

where $\boldsymbol{z}$ denotes the vector of time-invariant covariates and $\boldsymbol{\gamma}$ denotes the effects of these covariates. (Note that $\boldsymbol{z}$ *must not* include an intercept term.)

When the baseline is distributed exponential with rate parameter $\theta$,

$$h(t|\theta, \boldsymbol{\gamma}, \boldsymbol{z}) = \theta\exp(\boldsymbol{\gamma}'\boldsymbol{z})\,.$$

It follows from the definition of the hazard rate function,

$$h(t) = \frac{f(t)}{1 - F(t)}$$

that, since $F(0) = 0$,

$$F(t) = 1 - \exp\left(-\int_0^t h(u)du\right)\,.$$

Therefore,

$$F(t|\theta, \boldsymbol{\gamma}, \boldsymbol{z}) = 1 - e^{-\theta\exp(\boldsymbol{\gamma}'\boldsymbol{z})t}$$

and

$$f(t|\theta, \boldsymbol{\gamma}, \boldsymbol{z}) = \theta\exp(\boldsymbol{\gamma}'\boldsymbol{z})e^{-\theta\exp(\boldsymbol{\gamma}'\boldsymbol{z})t}\,.$$

## 3 The Case of the Pareto/NBD Model

The Pareto/NBD model is based on the following assumptions:

i. Customers go through two stages in their "lifetime" with a specific firm: they are "alive" for some period of time, then become permanently inactive.

ii. While alive, the number of transactions made by a customer follows a Poisson process with transaction rate $\lambda$. This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate $\lambda$,

$$f(t_j - t_{j-1} \,|\, \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, \quad t_j > t_{j-1} > 0\,,$$

where $t_j$ is the time of the $j$th purchase.

iii. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter $r$ and scale parameter $\alpha$:

$$g(\lambda \,|\, r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}\,.$$

iv. A customer's unobserved "lifetime" of length $\omega$ (after which he is viewed as being inactive) is exponentially distributed with dropout rate $\mu$:

$$f(\omega \,|\, \mu) = \mu e^{-\mu\omega}\,.$$

(Note: Previous discussions of the Pareto/NBD have used $\tau$ to denote the time at which the customer becomes inactive. Because of notational conflicts, we now use $\omega$ to denote this quantity.)

v. Heterogeneity in dropout rates across customers follows a gamma distribution with shape parameter $s$ and scale parameter $\beta$.

$$g(\mu \,|\, s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)}\,.$$

vi. The transaction rate $\lambda$ and the dropout rate $\mu$ vary independently across customers.

We now assume that interpurchase times are distributed according to the with-covariates exponential distribution

$$f(t_j - t_{j-1}|\lambda_0, \boldsymbol{\gamma}_1, \boldsymbol{z}_1) = \lambda_0 \exp(\boldsymbol{\gamma}_1'\boldsymbol{z}_1) e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1'\boldsymbol{z}_1)(t_j - t_{j-1})} \text{ and}$$
$$F(t_j - t_{j-1}|\lambda_0, \boldsymbol{\gamma}_1, \boldsymbol{z}_1) = 1 - e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1'\boldsymbol{z}_1)(t_j - t_{j-1})}\,,$$

with the unobserved heterogeneity in $\lambda_0$ captured by a gamma distribution with shape parameter $r$ and scale parameter $\alpha_0$:

$$g(\lambda_0 \,|\, r, \alpha_0) = \frac{\alpha_0^r \lambda_0^{r-1} e^{-\lambda_0 \alpha_0}}{\Gamma(r)} \,.$$

We also assume that lifetimes are distributed according to the with-covariates exponential distribution

$$\begin{aligned} f(\omega|\mu_0, \boldsymbol{\gamma}_2, \boldsymbol{z}_2) &= \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2) e^{-\mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)\omega} \text{ and} \\ F(\omega|\mu_0, \boldsymbol{\gamma}_2, \boldsymbol{z}_2) &= 1 - e^{-\mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)\omega} \,, \end{aligned}$$

with the unobserved heterogeneity in $\mu_0$ captured by a gamma distribution with shape parameter $s$ and scale parameter $\beta_0$:

$$g(\mu_0 \,|\, s, \beta_0) = \frac{\beta_0^s \mu_0^{s-1} e^{-\mu_0 \beta_0}}{\Gamma(s)} \,.$$

Following the logic of the derivation presented in Fader and Hardie (2005), it follows that

$$L(\lambda_0, \boldsymbol{\gamma}_1 \,|\, \boldsymbol{z}_1, t_1, \ldots, t_x, T, \omega > T) = \lambda_0^x \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) T}$$

and

$$\begin{aligned} &L(\lambda_0, \boldsymbol{\gamma}_1 \,|\, \boldsymbol{z}_1, t_1, \ldots, t_x, T, \text{inactive at } \omega \in (t_x, T]) \\ &\qquad = \lambda_0^x \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)\omega} \,. \end{aligned}$$

Removing the conditioning on $\omega$ yields the following expression for the individual-level likelihood function:

$$\begin{aligned} &L(\lambda_0, \mu_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \,|\, \boldsymbol{z}_1, \boldsymbol{z}_2, x, t_x, T) \\ &\quad = L(\lambda_0, \boldsymbol{\gamma}_1 \,|\, \boldsymbol{z}_1, x, T, \omega > T) P(\omega > T \,|\, \mu_0, \boldsymbol{\gamma}_2, \boldsymbol{z}_2) \\ &\qquad + \int_{t_x}^{T} \Big\{ L(\lambda_0, \boldsymbol{\gamma}_1 \,|\, \boldsymbol{z}_1, x, t_x, T, \text{inactive at } \omega \in (t_x, T]) \\ &\qquad\qquad \times f(\omega \,|\, \mu_0, \boldsymbol{\gamma}_2, \boldsymbol{z}_2) \Big\} \, d\omega \\ &\quad = \frac{\lambda_0^x \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)}{\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)} e^{-(\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)) t_x} \\ &\qquad + \frac{\lambda_0^{x+1} \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^{x+1}}{\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)} e^{-(\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)) T} \,. \qquad (1) \end{aligned}$$

We remove the conditioning on unobserved latent variables $\lambda_0$ and $\mu_0$ by taking the expectation of (1) over the distributions of $\lambda_0$ and $\mu_0$. To facilitate this calculation, we perform the change of variables $\lambda = \lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)$ and $\mu = \mu_0 \exp(\boldsymbol{\gamma}_2' \boldsymbol{z}_2)$, which implies

$$g(\lambda \,|\, r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}\,, \text{ where } \alpha = \alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)\,, \text{ and}$$
$$g(\mu \,|\, s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)}\,, \text{ where } \beta = \beta_0 \exp(-\boldsymbol{\gamma}_2' \boldsymbol{z}_2)\,.$$

**Note the addition of the addition of minus signs in the exponential terms as we go from multipliers of $\lambda_0, \mu_0$ to multipliers of $\alpha_0, \beta_0$**

This gives us

$$\begin{aligned} L(r, \alpha_0, &s, \beta_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \,|\, \boldsymbol{z}_1, \boldsymbol{z}_2, x, t_x, T) \\ &= \int_0^\infty \int_0^\infty \bigg\{ \bigg( \frac{\lambda^x \mu}{\lambda + \mu} e^{-(\lambda+\mu)t_x} + \frac{\lambda^{x+1}}{\lambda + \mu} e^{-(\lambda+\mu)T} \bigg) \\ &\qquad\qquad \times g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) \bigg\} \, d\lambda d\mu \,. \end{aligned}$$

We know from Fader and Hardie (2005) that the solution to this is

$$\begin{aligned} L(r, \alpha_0, &s, \beta_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \,|\, \boldsymbol{z}_1, \boldsymbol{z}_2, x, t_x, T) \\ &= \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)} \bigg\{ \bigg( \frac{s}{r+s+x} \bigg) \mathsf{A}_1 + \bigg( \frac{r+x}{r+s+x} \bigg) \mathsf{A}_2 \bigg\} \end{aligned} \tag{2}$$

where

$$\mathsf{A}_1 = \begin{cases} \dfrac{{}_2F_1\big(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}\big)}{(\alpha+t_x)^{r+s+x}} & \text{if } \alpha \geq \beta \\ \dfrac{{}_2F_1\big(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}\big)}{(\beta+t_x)^{r+s+x}} & \text{if } \alpha \leq \beta \end{cases} \tag{3}$$

and

$$\mathsf{A}_2 = \begin{cases} \dfrac{{}_2F_1\big(r+s+x, s; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}\big)}{(\alpha+T)^{r+s+x}} & \text{if } \alpha \geq \beta \\ \dfrac{{}_2F_1\big(r+s+x, r+x+1; r+s+x+1; \frac{\beta-\alpha}{\beta+T}\big)}{(\beta+T)^{r+s+x}} & \text{if } \alpha \leq \beta \end{cases} \tag{4}$$

In other words, the likelihood function for the time-invariant-covariates version of the Pareto/NBD model is the likelihood function associated with the basic model where $\alpha$ and $\beta$ are replaced by $\alpha = \alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)$ and $\beta = \beta_0 \exp(-\boldsymbol{\gamma}_2' \boldsymbol{z}_2)$; $r$ and $s$ remain unchanged.

Suppose we have transaction data for a sample of $N$ customers, where customer $i$ made $x_i$ purchases in the period $(0, T_i]$ (with the last transaction occurring at $t_{x_i}$), and covariate vectors $\boldsymbol{z}_{1i}$ and $\boldsymbol{z}_{2i}$. The sample log-likelihood function is given by

$$LL(r, \alpha_0, s, \beta_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \sum_{i=1}^{N} \ln \left[ L(r, \alpha_0, s, \beta_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \,|\, \boldsymbol{z}_{1i}, \boldsymbol{z}_{2i}, x_i, t_{x_i}, T_i) \right].$$

In numerically evaluating this function, we note that the $\alpha \geq / \leq \beta$ condition associated with (2)–(4) must now be checked for each $i$. For any given $\alpha_0, \beta_0$, it may be the case that $\alpha > \beta$ for person $i$ while $\alpha < \beta$ for person $j$ because of the values of $\boldsymbol{z}_{1i}, \boldsymbol{z}_{2i}$ and $\boldsymbol{z}_{1j}, \boldsymbol{z}_{2j}$ (and $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2$).

For all related expressions (e.g., $E[X(t)], E(Y(t) \,|\, x, t_x, T)$), it is easy to show that we simply replace $\alpha$ and $\beta$ with $\alpha = \alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)$ and $\beta = \beta_0 \exp(-\boldsymbol{\gamma}_2' \boldsymbol{z}_2)$ to arrive at their with-time-invariant-covariates equivalents.

# 4 The Case of the BG/NBD Model

The BG/NBD model is based on the following assumptions (the first three of which are identical to the corresponding Pareto/NBD assumptions):

i. Customers go through two stages in their "lifetime" with a specific firm: they are "alive" for some period of time, then become permanently inactive.

ii. While alive, the number of transactions made by a customer follows a Poisson process with transaction rate $\lambda$. This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate $\lambda$,

$$f(t_j - t_{j-1} \,|\, \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}, \quad t_j > t_{j-1} > 0,$$

where $t_j$ is the time of the $j$th purchase.

iii. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter $r$ and scale parameter $\alpha$:

$$g(\lambda \,|\, r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda \alpha}}{\Gamma(r)}.$$

iv. After any transaction, a customer becomes inactive with probability $p$. Therefore the point at which the customer "drops out" is distributed across transactions according to a (shifted) geometric distribution with pmf

$$\begin{aligned} &P(\text{inactive immediately after } j\text{th transaction}) \\ &\qquad = p(1-p)^{j-1}, \quad j = 1, 2, 3, \ldots. \end{aligned}$$

v. Heterogeneity in dropout probabilities follows a beta distribution with parameters $a$ and $b$:

$$f(p \mid a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}, \quad 0 \leq p \leq 1 .$$

vi. The transaction rate $\lambda$ and the dropout probability $p$ vary independently across customers.

We now assume that interpurchase times are distributed according to the with-covariates exponential distribution

$$f(t_j - t_{j-1} | \lambda_0, \boldsymbol{\gamma}_1, \boldsymbol{z}_1) = \lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)(t_j - t_{j-1})}$$

and that the unobserved heterogeneity in $\lambda_0$ is distributed gamma with shape parameter $r$ and scale parameter $\alpha_0$:

$$g(\lambda_0 \mid r, \alpha_0) = \frac{\alpha_0^r \lambda_0^{r-1} e^{-\lambda_0 \alpha_0}}{\Gamma(r)} .$$

Following the logic of the derivation presented in Fader et al. (2005),

$$\begin{aligned} L(\lambda_0, p, \boldsymbol{\gamma}_1 \mid \boldsymbol{z}_1, x, t_x, T) = (1-p)^x \lambda_0^x \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) T} \\ + \delta_{x>0} p (1-p)^{x-1} \lambda_0^x \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x e^{-\lambda_0 \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) t_x} . \end{aligned}$$

Taking the expectation of this over the distribution of $\lambda_0$ gives us

$$\begin{aligned} &L(r, \alpha_0, p, \boldsymbol{\gamma}_1 \mid \boldsymbol{z}_1, x, t_x, T) \\ &\qquad = (1-p)^x \frac{\Gamma(r+x) \alpha_0^r \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x}{\Gamma(r)(\alpha_0 + \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) T)^{r+x}} \\ &\qquad\quad + \delta_{x>0} p (1-p)^{x-1} \frac{\Gamma(r+x) \alpha_0^r \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1)^x}{\Gamma(r)(\alpha_0 + \exp(\boldsymbol{\gamma}_1' \boldsymbol{z}_1) t_x)^{r+x}} \\ &\qquad = (1-p)^x \frac{\Gamma(r+x) [\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)]^r}{\Gamma(r)(\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + T)^{r+x}} \\ &\qquad\quad + \delta_{x>0} p (1-p)^{x-1} \frac{\Gamma(r+x) [\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)]^r}{\Gamma(r)(\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + t_x)^{r+x}} . \end{aligned} \tag{5}$$

Note that we have yet to include the effects of $\mathbf{z}_2$, the vector of time-invariant covariates that are assumed to explain some of the cross-sectional heterogeneity in the dropout process. It is not possible to include the effects of covariates into $p$ and then allow for unobserved heterogeneity using the beta distribution.

Using the logic of the beta-logistic model (Heckman and Willis 1977, Rao and Steckel 1995), we incorporate the effects of time-invariant covariates on the dropout process via the parameters of the beta mixing distribution by making $a$ and $b$ functions of $\boldsymbol{z}_2$. That is, we assume that heterogeneity in the dropout probabilities follows a beta distribution with parameters $a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2)$ and $b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2)$:

$$f(p \,|\, a_0, b_0, \gamma_2, \gamma_3; \mathbf{z}_2) = \frac{p^{a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2) - 1} (1-p)^{b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2) - 1}}{B(a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2), b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2))}.$$

Taking the expectation of (5) over this distribution of $p$ gives us

$$\begin{aligned} L(r, \alpha_0, a_0, b_0, &\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3 \,|\, \boldsymbol{z}_1, \boldsymbol{z}_2, x, t_x, T) \\ &= \Bigg\{ \frac{B(a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2), b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2) + x)}{B(a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2), b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2))} \\ &\qquad \times \frac{\Gamma(r+x)[\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)]^r}{\Gamma(r)(\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + T)^{r+x}} \Bigg\} \\ &\quad + \delta_{x>0} \Bigg\{ \frac{B(a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2) + 1, b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2) + x - 1)}{B(a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2), b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2))} \\ &\qquad\qquad \times \frac{\Gamma(r+x)[\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1)]^r}{\Gamma(r)(\alpha_0 \exp(-\boldsymbol{\gamma}_1' \boldsymbol{z}_1) + t_x)^{r+x}} \Bigg\} \end{aligned}$$

In other words, the likelihood function for the time-invariant-covariates version of the BG/NBD model is the likelihood function associated with the basic model where $\alpha$, $a$, and $b$ are replaced by $\alpha = \alpha_0 \exp(-\boldsymbol{\gamma}_1' \mathbf{z}_1)$, $a = a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2)$, and $b = b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2)$; $r$ remains unchanged.[1]

For all related expressions (e.g., $E[X(t)], E(Y(t) \,|\, x, t_x, T)$), it is easy to show that we simply replace $\alpha$, $a$, and $b$ with $\alpha = \alpha_0 \exp(-\boldsymbol{\gamma}_1' \mathbf{z}_1)$, $a = a_0 \exp(\boldsymbol{\gamma}_2' \mathbf{z}_2)$, and $b = b_0 \exp(\boldsymbol{\gamma}_3' \mathbf{z}_2)$ to arrive at their with-time-invariant-covariates equivalents.

[1]We note that the BG/NBD has 50% more covariate parameters than the Pareto/NBD. While this difference could be removed by making only $a$ a function of covariates (i.e., set $\boldsymbol{\gamma}_3$ to $\mathbf{0}$), doing so would constrain the way in which the covariates could influence the shape (e.g., variance) of the beta distribution. Furthermore, it would be inconsistent with the basic beta-logistic formulation.

## References


Fader, Peter S. and Bruce G. S. Hardie (2005), “A Note on Deriving the Pareto/NBD Model and Related Expressions.” <http://brucehardie.com/notes/009/>

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), “"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model,” *Marketing Science*, **24** (Spring), 275–284.

Heckman, James J. and Robert J. Willis (1977), “A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women,” *Journal of Political Economy*, **85** (February), 27–58.

Rao, Vithala R. and Joel H. Steckel (1995), “Selecting, Evaluating, and Updating Prospects in Direct Marketing,” *Journal of Direct Marketing*, 9 (Spring), 20–31.