

“How to Project Customer Retention” Revisited: The Role of Duration Dependence

Peter S. Fader
Bruce G. S. Hardie
Yuzhou Liu
Joseph Davin
Thomas Steenburgh[†]

January 2018

[†]Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at The Wharton School of the University of Pennsylvania (address: 771 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340; phone: (215) 898-1132; email: faderp@wharton.upenn.edu; web: www.petefader.com). Bruce G. S. Hardie is Professor of Marketing at the London Business School (email: bhardie@london.edu; web: www.brucehardie.com). Yuzhou Liu is a Quantitative Researcher at Man Numeric. Joseph Davin is Head of Data Science, Two Six Capital. Thomas Steenburgh is the Bank of America Research Professor of Business Administration at the Darden Graduate School of Business, University of Virginia.

Abstract

“How to Project Customer Retention” Revisited: The Role of Duration Dependence

Cohort-level retention rates typically increase over time, and the beta-geometric (BG) distribution has proven to be a robust model for capturing and projecting these patterns into the future. According to this model, the phenomenon of increasing cohort-level retention rates is purely due to cross-sectional heterogeneity; an individual customer’s propensity to churn does not change over time. In this paper we present the beta-discrete-Weibull (BdW) distribution as an extension to the BG model, one that allows individual-level churn probabilities to increase or decrease over time. In addition to capturing the phenomenon of increasing cohort-level retention rates, this new model can also accommodate situations in which there is an initial dip in retention rates before they increase (i.e., a U-shaped cohort-level retention curve). A key finding is that even when aggregate retention rates are monotonically increasing, the individual-level churn probabilities are unlikely to be declining over time, as conventional wisdom would suggest. We carefully explore these connections between heterogeneity, duration dependence, and the shape of the retention curve, and draw some managerially relevant conclusions, e.g., that accounting for cross-sectional heterogeneity is more important than accounting for any individual-level dynamics in churn propensities.

Keywords: beta-geometric (BG) distribution, beta-discrete-Weibull (BdW) distribution, retention rate dynamics.

1 Introduction

Any researcher working with data from a business that has a “contractual” relationship with its customers (e.g., one with a subscription-based business model) will want models for projecting customer retention (or equivalently, tenure) as part of their toolkit. For example, estimates of the length of a customer’s relationship with the firm lie at the heart of any attempt to compute customer lifetime value (CLV). Similarly, such models are useful when evaluating the relative performance of different acquisition channels.

Fader and Hardie (2007), hereafter FH, presented the beta-geometric (BG) distribution as a simple probability model for projecting customer retention. This model is based on an easy-to-understand “story” of customer behavior, is simple to implement (e.g., can be done so in Excel), and its estimates of customer retention over a longitudinal holdout period have proven to be surprisingly accurate and robust.

According to this model, the widely observed phenomenon of increasing cohort-level retention rates (Reichheld 1996) is purely due to cross-sectional heterogeneity, with individual customers having a constant propensity to churn. Cohort-level retention rates increase because those customers with high churn propensities drop out early on, leaving an ever-increasing proportion of customers who have low propensities to churn. This assumption that an individual customer’s propensity to churn does not change over time flies in the face of conventional wisdom, which assumes that a customer’s propensity to churn decreases the longer their tenure with the firm.

While cohort-level retention rates are typically monotonically increasing with tenure, we sometimes observe an initial dip before they increase (e.g., Israel 2005, Nitzan et al. 2011, and one of the FH datasets), a phenomenon that the BG model cannot capture. In this paper we develop a generalization of the BG model, one that both relaxes the assumption of time-invariant individual-level propensities to churn and is sufficiently flexible to capture the phenomenon of non-monotonically increasing cohort-level retention rates. Surprisingly, we find that when the assumption of constant individual-level propensities to churn is violated, it is more likely that these propensities *increase* with tenure (rather than decrease, as conventional wisdom would suggest).

This paper is organized as follows. In the next section we re-examine the work of FH,

reviewing the BG model and its empirical performance. We then present our generalization of the BG model, the beta-discrete-Weibull (BdW) distribution, and examine its performance using the two datasets presented in FH. This is followed by an investigation of the properties of the cohort-level retention rates associated with the BdW model. We then investigate the robustness of our results by exploring some alternative model specifications, and conclude with a brief discussion of the implications of this work.

2 A Brief Review of the BG Model

FH propose a simple probability model for characterizing and forecasting the length of a customer's relationship with a firm in a contractual setting that is based on the following "as if" story of customer behavior:

- i) At the end of each contract period, an individual decides whether or not to renew their contract by tossing a coin: "heads" they renew their contract, "tails" they cancel it.
- ii) For a given individual, the probability of a coin coming up "tails" does not change over time.
- iii) The probability of a coin coming up "tails" varies across customers. (This implies that the coins are not assumed to be "fair.")

This is formalized in the following manner. Let the random variable T denote the length of an individual's relationship with the firm, and θ denote the probability of a given individual's coin coming up "tails" when tossed. Assumptions (i) and (ii) are equivalent to assuming that T is distributed geometric with survivor function

$$S(t|\theta) = (1 - \theta)^t, \quad 0 < \theta < 1, \quad t = 0, 1, 2, \dots \quad (1)$$

From the analyst's perspective, the unobserved (and unobservable) θ is a realization of the random variable Θ . Given its flexibility and mathematical convenience, the natural distribution for characterizing Θ is the beta distribution:

$$f(\theta | \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}, \quad \gamma, \delta > 0. \quad (2)$$

It follows that for a randomly chosen individual,

$$\begin{aligned} S(t | \gamma, \delta) &= \int_0^1 S(t | \theta) f(\theta | \gamma, \delta) d\theta \\ &= \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)}, \quad t = 0, 1, 2, \dots \end{aligned} \quad (3)$$

This beta mixture of geometrics is called the beta-geometric (BG) distribution.¹ (See FH for model derivations and information on how to estimate the model parameters; also see Fader and Hardie (2014) for an alternative estimation approach.)

To the best of our knowledge, this mixture model was first derived by Pielou (1962), who used it to characterize the runs lengths of species in plant populations. Potter and Parker (1964) were the first to use it as a model for duration-time data, using it to characterize the number of menstrual cycles a woman experiences before she conceives. Within the marketing literature, it was used by Morrison and Perry (1970) as a model of the number of units purchased on a given transaction occasion, and by Buchanan and Morrison (1988) as a model of response to promotional stimuli; also see Fox et al. (1997). FH explored its properties as a model of the length of a customer’s relationship with a firm in a contractual setting, with Fader and Hardie (2010) taking the logical next step and using it as the basis for calculating CLV.

At the heart of the FH paper is a customer metric of great interest to managers and analysts: the retention rate. When computed at the level of the cohort, the period t retention rate is the portion of period t customers (i.e., those who have “survived” to period t) who renew their contracts at the end of that period. This can be computed as

¹FH called this the shifted-beta-geometric (sBG) model; the term “shifted” is used to make the distinction between two versions of the geometric distribution; one with support $0, 1, 2, \dots$, and the other with support $1, 2, 3, \dots$, with the term “shifted” being applied to the second version. (In our contract-duration setting, the first version would apply when T is defined as the number of contract renewals the individual makes before they cancel their contract, rather than the length of the individual’s relationship with the firm (measured in number of contract periods), as is the case above.) However, most applications of this mixture model are in settings in which the support is $1, 2, 3, \dots$, yet the term shifted is not applied. So as to be consistent with this broader literature, we use the label BG (rather than sBG) for this distribution.

$$\begin{aligned}
r(t | \gamma, \delta) &= \frac{S(t | \gamma, \delta)}{S(t-1 | \gamma, \delta)} \\
&= \frac{B(\gamma, \delta + t)}{B(\gamma, \delta + t - 1)} \\
&= \frac{\delta + t - 1}{\gamma + \delta + t - 1}, \quad t = 1, 2, 3, \dots
\end{aligned} \tag{4}$$

Note that, for any values of γ and δ , this is an increasing function of time. It is important to note that there are no underlying time dynamics at the level of the individual customer; see assumption (ii) above. The increasing (aggregate/cohort-level) retention rate is simply due to a sorting effect in a heterogeneous population.

To elaborate on this sorting effect, let $\rho(t)$ denote the individual-level probability that someone who has made $t - 1$ renewals will renew at the next opportunity (i.e., $P(\text{heads})$). Given the assumption of individual-level relationship durations characterized by the geometric distribution, $\rho(t) = S(t | \theta) / S(t - 1 | \theta) = 1 - \theta$. Recall that the unobserved (and unobservable) θ are viewed as realization of the random variable Θ . Similarly, $\rho(t)$ is a realization of $P(t)$.

Since $\rho(t)$ is a function of θ , the distribution of $P(t)$ is a function of the posterior distribution of Θ across the period t customers (i.e., those who have made $t - 1$ contract renewals). Recalling Bayes' theorem,

$$\begin{aligned}
f(\theta | \gamma, \delta; t-1 \text{ renewals}) &= \frac{S(t-1 | \theta) f(\theta | \gamma, \delta)}{S(t-1 | \gamma, \delta)} \\
&= \frac{\theta^{\gamma-1} (1-\theta)^{\delta+t-2}}{B(\gamma, \delta + t - 1)}, \quad t = 1, 2, 3, \dots
\end{aligned} \tag{5}$$

which is a beta distribution with parameters γ and $\delta + t - 1$. Given that $\rho(t) = 1 - \theta$, the distribution of $P(t)$ across period t customers is simply the reflection of this posterior distribution about $\theta = 0.5$, which is a beta distribution with parameters $\delta + t - 1$ and γ :

$$f(\rho(t) | \gamma, \delta) = \frac{\rho(t)^{\delta+t-2} (1-\rho(t))^{\gamma-1}}{B(\delta + t - 1, \gamma)}, \quad t = 1, 2, 3, \dots \tag{6}$$

The mean of this distribution equals the expression for $r(t | \gamma, \delta)$ given in (4), i.e., $r(t) = E[P(t)]$. Dynamics in $r(t)$ are simply due to changes in the nature of the distribution of $P(t)$, which are

simply due to customers with higher churn propensities dropping out, leaving an ever-increasing proportion of customers who have low propensities to churn.

3 Revisiting FH’s Analysis

We start by revisiting the empirical analysis presented in FH. Our objectives are two-fold. First, we wish to highlight the robustness of the BG model. Second, we wish to identify the phenomenon that motivates this work.

The data presented in Table 1, drawn from Berry and Linoff (2004), document the year-on-year renewals for two segments of customers (“Regular” and “High End”) of an unspecified firm in a contractual setting. (See FH for further details.) For a nominal sample of 1000 customers acquired at the beginning of Year 1, we observe their pattern of renewals over 12 consecutive (annual) renewal opportunities. For example, 631 individuals in the Regular dataset renew their contract at the end of the first year, and are therefore customers in Year 2. Of these 631 individuals who have a contractual relationship with the firm in Year 2, 468 renew their contract at the end of the year and are therefore customers in Year 3. And so on.

Year	# Customers	
	Regular	High End
1	1000	1000
2	631	869
3	468	743
4	382	653
5	326	593
6	289	551
7	262	517
8	241	491
9	223	468
10	207	445
11	194	427
12	183	409
13	173	394

Table 1: Pattern of year-on-year renewals for a cohort of 1000 customers from two segments (Regular and High End) acquired at the beginning of Year 1

FH undertake an analysis in which the model is calibrated using the first eight years of data (seven renewal opportunities) and its predictive performance assessed over the remaining

five years of data (five renewal opportunities). The estimation results are presented in Table 2 (columns 2 and 3) and the model-based estimates of survival and retention are compared against the actual numbers for both datasets in Figure 1.²

	Eight-year Calibration Period		Five-year Calibration Period	
	Regular	High End	Regular	High End
γ	0.704	0.668	0.764	1.281
δ	1.182	3.806	1.296	7.790
LL	-1680.3	-1611.2	-1401.6	-1225.1

Table 2: BG model estimation results

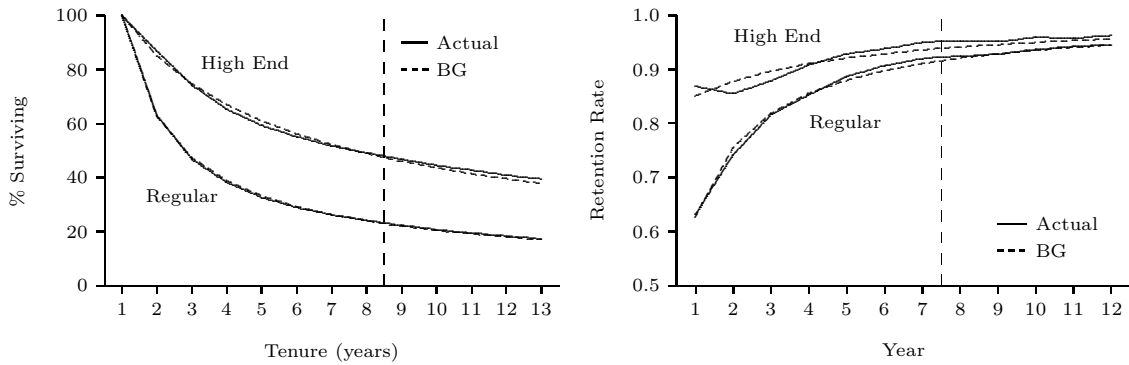


Figure 1: Comparing actual and BG-model-based estimates of survival (LHS) and retention (RHS) given an eight-year model calibration period. (The model-based numbers to the right of the vertical dashed line are projections given the parameter values estimated using the data to the left of this line.)

We note that this simple probability model does an excellent job of predicting survival (and therefore retention) in the Regular dataset. The prediction of survival in the High End dataset is good but not quite as impressive on retention as for the Regular dataset. FH noted that “[d]espite the existence of certain unexplained “blips” as in Year 2 for the High End dataset, the tracking/prediction plot for $r(t)$ is very impressive through Year 12,” and made no further comment.

We now “stress test” the BG model by shortening the calibration period to five years (four renewal opportunities), thereby lengthening the validation period to eight years. The estimation results are presented in Table 2 (columns 4 and 5). Note that for the Regular cohort the parameter estimates are quite similar. In fact, evaluating the eight-year calibration period like-

²A copy of the spreadsheet containing the analyses presented in this paper can be found at <insert URL>.

likelihood function using the five-year calibration period parameter estimates yields a log-likelihood of -1680.6 . This stability in parameter estimates suggests that the BG model is an excellent characterization of the true data-generating process. This conclusion is supported when we compare the model-based estimates of survival and retention to the actual numbers (top half of Figure 2). While not as good as for an eight-year calibration period, the performance is still very impressive when we consider that we are predicting behavior across a holdout period that is twice as long as the calibration period.

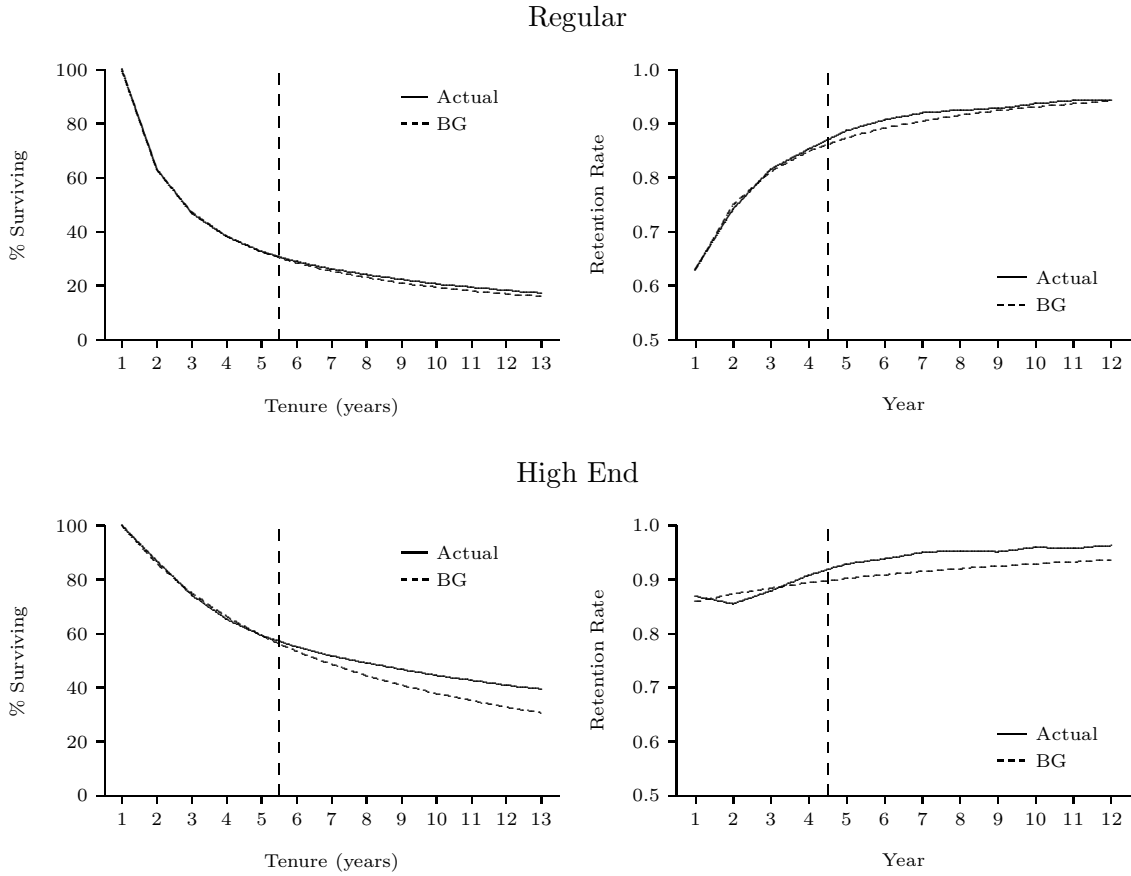


Figure 2: Comparing actual and BG-model-based estimates of survival (LHS) and retention (RHS) for the Regular (top) and High End (bottom) datasets given a five-year model calibration period. (The model-based numbers to the right of the vertical dashed line are projections given the parameter values estimated using the data to the left of this line.)

The results for the High End dataset are a completely different story. We note from Table 2 that the parameter estimates are very sensitive to the length of the calibration period, suggesting that the BG model is a not good characterization of the true data-generating process for this

dataset. This problem is even more evident when we compare the model-based estimates of survival and retention to the actual numbers (bottom half of Figure 2). While the model appears to be tracking actual survival in the calibration period, it progressively under-predicts survival with the passage of time in the validation period. This is reflected in the failure of the model to capture the dynamics in the retention rates observed in this dataset.

For an eight-year calibration period, the dip in the retention rate observed in Year 2 is effectively treated as an outlier that has little impact on model estimation; the overall trend of increasing retention rates is adequately captured (as observed in Figure 1). However, it has a far greater influence on model estimation when we shorten the model calibration period to five years (as observed in the bottom-right plot in Figure 2). Whereas it seemed acceptable for FH to brush aside the Year 2 dip, the shorter calibration period shows that we cannot ignore it.

Cohort-level retention rates are predominantly monotonically increasing (as in the Regular dataset), and the BG model is a robust way to characterize such behaviour. However, it is not a robust model when faced with the type of cohort-level retention rate pattern observed in the High End dataset. We have observed such a dip in several other datasets (e.g., Israel 2005, Nitzan et al. 2011). This suggests the need for an alternative, more flexible, model for characterizing and forecasting the length of a customer’s relationship with the firm in a contractual setting.

4 The BdW Model

When a model doesn’t “work,” we question its underlying assumptions. Reflecting on the “as if” story of buyer behavior underpinning the BG model, a number of people struggle with the assumption that, for a given individual, the probability of a coin coming up “tails” does not change over time. They expect it to become more “headsy” over time (i.e., the individual is expected to become more “loyal” the longer they remain a customer).³

To accommodate this in a continuous-time environment, the natural starting point would be to replace the exponential distribution (the continuous-time equivalent of the geometric distribution) with the Weibull distribution, which allows for an individual’s risk of canceling

³There are two standard explanations for such an expectation. The first is based on an evolution of customer satisfaction argument (e.g., Bolton 1998), while the second is based on an increasing switching costs argument (e.g., Burnham et al. 2003).

their contract to increase or decrease as the length of the relationship with the firm increases (Murthy et al. 2004, Rinne 2009). Working within a discrete-time contractual setting, the natural starting point is to use a discrete Weibull distribution.

A discrete-time equivalent of a continuous distribution can be constructed by treating the discrete lifetime variable as the integer part of the continuous lifetime and discretizing its cdf or, equivalently, survivor function (Lai 2013). Suppose the continuous lifetime random variable X is distributed Weibull with survivor function

$$S(x | \lambda, c) = \exp(-\lambda x^c) = [\exp(-\lambda)]^{x^c}.$$

Letting $\exp(-\lambda) = 1 - \theta$, it follows that the survivor function associated with the discrete lifetime random variable $T = \lfloor X \rfloor$ is

$$S(t | \theta, c) = (1 - \theta)^{t^c}, \quad 0 < \theta < 1, \quad c > 0, \quad t = 0, 1, 2, \dots \quad (7)$$

The associated pmf is given by

$$P(T = t | \theta, c) = (1 - \theta)^{(t-1)^c} - (1 - \theta)^{t^c}, \quad t = 1, 2, 3, \dots \quad (8)$$

This is the discrete Weibull (dW) distribution proposed by Nakagawa and Osaki (1975). While there are other discrete Weibull distributions (Murthy et al. 2004, Chapter 13; Rinne 2009, Section 3.3.1), this one is simple, flexible, and the best analogue of the continuous Weibull distribution (Bracquemond and Gaudoin 2003). We note that the dW collapses to the geometric distribution when $c = 1$, just as the Weibull collapses to the exponential distribution when $c = 1$.⁴

Under the dW distribution, the individual-level probability that someone who has made $t - 1$ renewals will renew at the next opportunity is given by

⁴There are no closed-form expressions for the mean and variance of the dW distribution. Looking at (8), we see that $P(T = 1 | \theta, c) = \theta$, which means the pmf is reverse-J-shaped (i.e., the mode is at $T = 1$) when $\theta > 0.5$. When $c \leq 1$, the pmf is reverse-J-shaped for all values of θ . When $\theta < 0.5$, the pmf has an interior mode (i.e., the cdf is S-shaped) when $c > \ln[\ln(1 - 2\theta)/\ln(1 - \theta)]/\ln(2)$. (In contrast, the geometric distribution pmf is always reverse-J-shaped, which means its cdf is concave.)

$$\begin{aligned}\rho(t|\theta, c) &= \frac{S(t|\theta, c)}{S(t-1|\theta, c)} \\ &= (1-\theta)^{t^c - (t-1)^c}, \quad t = 1, 2, 3, \dots\end{aligned}\tag{9}$$

When $c = 1$, we have a constant individual retention probability, i.e., $P(\text{heads})$ in the “story” underpinning the BG model. When $c > 1$, $t^c - (t-1)^c$ increases with time, which means the proverbial coin becomes less “heads” the longer the individual remains a customer (i.e., the longer they remain a customer, the less likely they are to renew their contract). When $c < 1$, $t^c - (t-1)^c$ decreases with time, which means the coin becomes more “heads” the longer the individual remains a customer (i.e., we have an increasing individual-level retention probability). When $c > 1$, the discrete Weibull is said to exhibit *positive duration dependence* (i.e., an increasing probability of “failure”). Similarly, when $c < 1$, it is said to exhibit *negative duration dependence* (i.e., a decreasing probability of “failure”).

We fit this distribution to our two datasets (using the same five-year calibration period). The estimation results are presented in Table 3, and the associated tracking plots in Figure 3.

	Regular	High End
θ	0.374	0.138
c	0.636	0.910
LL	-1404.0	-1226.5

Table 3: dW model estimation results

As would be expected, our estimate of c is less than 1 (i.e., negative duration dependence) for both datasets, which implies that individual-level retention probabilities increase over time. Comparing the LL values with those associated with the BG model (Table 2), we see that the dW does not fit the data nearly as well as the BG. Looking at the tracking plots in Figure 3, we see that the dW fails to capture the retention rate dynamics observed in both datasets, and therefore fails to track the number of “surviving” customers. Comparing these plots with those in Figure 2, it is clear that, at least for these datasets, a model that explains increases in aggregate retention rates in terms of heterogeneity alone (i.e., the BG, which assumes $c = 1$) does better than one that explains it in term of duration dependence alone. We also note that

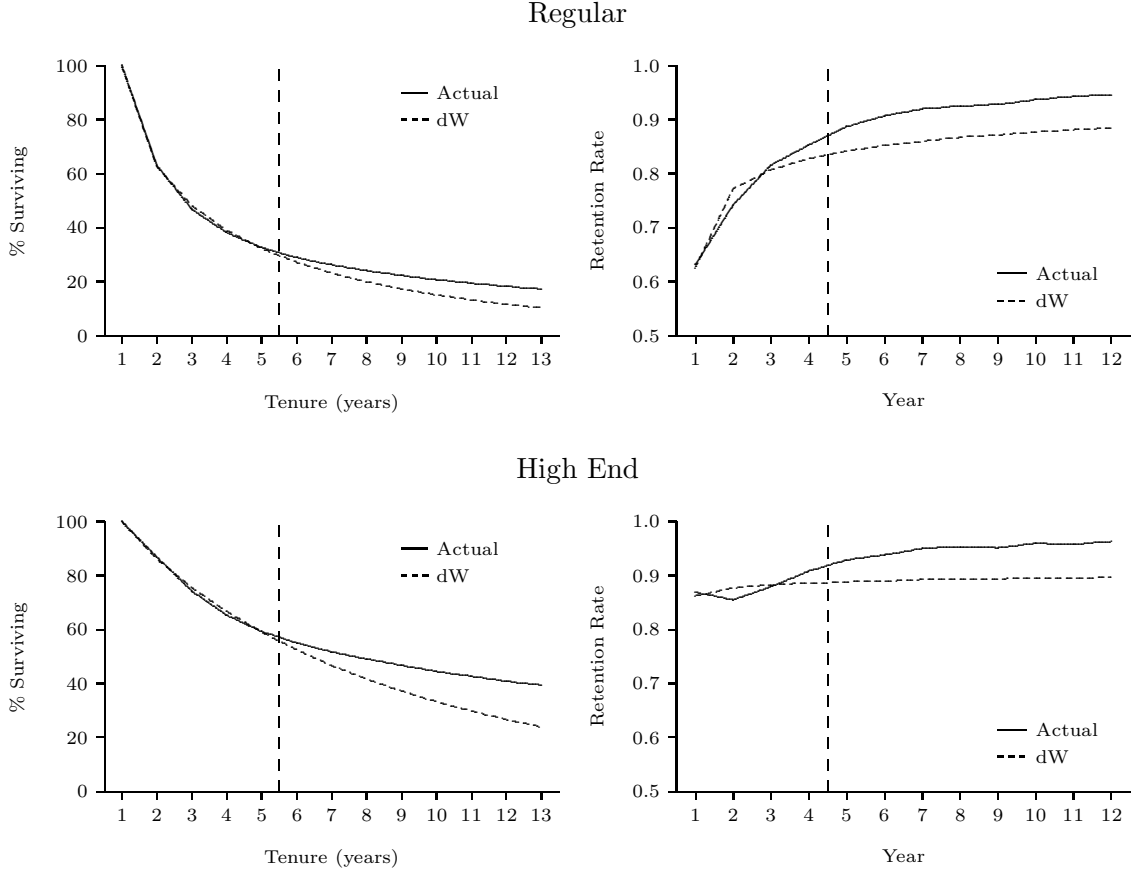


Figure 3: Comparing actual and dW-model-based estimates of survival (LHS) and retention (RHS) for the Regular (top) and High End (bottom) datasets given a five-year model calibration period. (The model-based numbers to the right of the vertical dashed line are projections given the parameter values estimated using the data to the left of this line.)

the dW fails to capture the Year 2 dip in the High End retention rate curve.

So what happens when we allow for both heterogeneity and duration dependence? Assuming cross-sectional heterogeneity in θ is characterized by a beta distribution with parameters (γ, δ) , it follows that for a randomly chosen individual,

$$\begin{aligned}
 S(t | \gamma, \delta, c) &= \int_0^1 S(t | \theta, c) f(\theta | \gamma, \delta) d\theta \\
 &= \frac{B(\gamma, \delta + t^c)}{B(\gamma, \delta)}, \quad t = 0, 1, 2, \dots
 \end{aligned} \tag{10}$$

and

$$\begin{aligned} P(T = t | \gamma, \delta, c) &= S(t-1 | \gamma, \delta, c) - S(t | \gamma, \delta, c) \\ &= \frac{B(\gamma, \delta + (t-1)^c) - B(\gamma, \delta + t^c)}{B(\gamma, \delta)}, \quad t = 1, 2, 3, \dots \end{aligned} \quad (11)$$

We call this parametric mixture model the beta-discrete-Weibull (BdW).

The associated aggregate/cohort-level retention rate is

$$\begin{aligned} r(t | \gamma, \delta, c) &= \frac{S(t | \gamma, \delta, c)}{S(t-1 | \gamma, \delta, c)} \\ &= \frac{B(\gamma, \delta + t^c)}{B(\gamma, \delta + (t-1)^c)} \\ &= \frac{\Gamma(\delta + t^c)}{\Gamma(\delta + (t-1)^c)} \frac{\Gamma(\gamma + \delta + (t-1)^c)}{\Gamma(\gamma + \delta + t^c)}, \quad t = 1, 2, 3, \dots \end{aligned} \quad (12)$$

We explore the shape of the associated retention rate curve below.

The distribution of $P(t)$ (i.e., $P(\text{heads})$) across those customers who have made $t-1$ renewals) is

$$\begin{aligned} f(\rho(t) | \gamma, \delta, c) &= \frac{1}{t^c - (t-1)^c} \frac{1}{\rho(t)} \\ &\times \frac{\left\{1 - \rho(t)^{\frac{1}{t^c - (t-1)^c}}\right\}^{\gamma-1} \left\{\rho(t)^{\frac{1}{t^c - (t-1)^c}}\right\}^{\delta + (t-1)^c}}{B(\gamma, \delta + (t-1)^c)}, \quad t = 1, 2, 3, \dots \end{aligned} \quad (13)$$

The mean of this distribution is, of course, the (aggregate) retention rate $r(t | \gamma, \delta, c)$. (See Appendix A for the derivations.) When $c = 1$ (i.e., $\text{BdW} \rightarrow \text{BG}$), (13) reduces to (6). Similarly, (12) reduces to (4).

We fit this model to both the Regular and High End datasets using a five-year calibration period. (See Appendix B for details of how to estimate the model parameters in Excel.) The estimation results are presented in Table 4. Comparing these results with those for the BG model (Table 2), we see that the improvement in model fit for the Regular dataset is negligible ($\text{LR} = 0.35, p = 0.553$), which means c is not significantly different from 1. On the other hand, we observe a significant improvement in fit for the High End dataset ($\text{LR} = 4.77, p = 0.029$), which means c is significantly different from 1. Comparing these estimates of c to those obtained

when fitting the dW to these datasets (Table 3), we notice that dW estimate is downward biased, reflecting the well-know result that unobserved heterogeneity induces spurious negative duration dependence in models for duration-time data (Kiefer 1988, Proschan 1963, Vaupel and Yashin 1985).

	Regular	High End
γ	0.523	0.259
δ	0.894	1.722
c	1.197	1.584
LL	-1401.4	-1222.7

Table 4: BdW model estimation results

In Figure 4 the model-based estimates of survival and retention are compared against the actual numbers for both datasets. The figures speak for themselves; the performance of the BdW model is impressive.⁵

At first glance, it is not clear why the model captures the downward blip in the aggregate retention rate curve we observe in the High End dataset. We note that $\hat{c} = 1.584$, indicating that at the level of the individual, customers become *less* likely to renew their contracts the longer they remain a customer. However, after some time, the sorting effect of heterogeneity (which causes the aggregate retention rate to increase over time) dominates the individual-level decline and the aggregate retention rate starts rising. We explore this further in Section 5 below.

How do our inferences about the underlying level of cross-sectional heterogeneity change when we allow for individual-level duration dependence? The parameters of the beta distribution can be characterized in terms of the mean $E(\Theta) = \gamma/(\gamma + \delta)$ and polarization index $\phi = 1/(\gamma + \delta + 1)$. The logic behind the polarization index is as follows: as $\gamma, \delta \rightarrow 0$ (thus $\phi \rightarrow 1$), the values of θ are concentrated near $\theta = 0$ and $\theta = 1$ and we can think of the values of θ as being very different, or “highly polarized.” As $\gamma, \delta \rightarrow \infty$ (thus $\phi \rightarrow 0$), the beta distribution becomes a spike at its mean; there is no “polarization” in the values of θ . Given the five-year calibration period parameter

⁵Fitting the BdW model to the Regular dataset using an eight-year calibration period yields the following parameter estimates: $\hat{\gamma} = 0.456$, $\hat{\delta} = 0.779$, and $\hat{c} = 1.284$. While these are slightly different from those associated with a five-year calibration period (Table 4), inserting these eight-year estimates in the log-likelihood function associated with a five-year calibration period yields the same value of -1401.4 . Using an eight-year calibration period for the High End dataset yields the following parameter estimates: $\hat{\gamma} = 0.214$, $\hat{\delta} = 1.427$, and $\hat{c} = 1.724$. Inserting these eight-year estimates in the log-likelihood function associated with a five-year calibration period yields a value of -1222.9 (versus -1222.7 in Table 4). This relative insensitivity to the length of the model calibration period suggests that the BdW provides a good characterization of the true data-generating process.

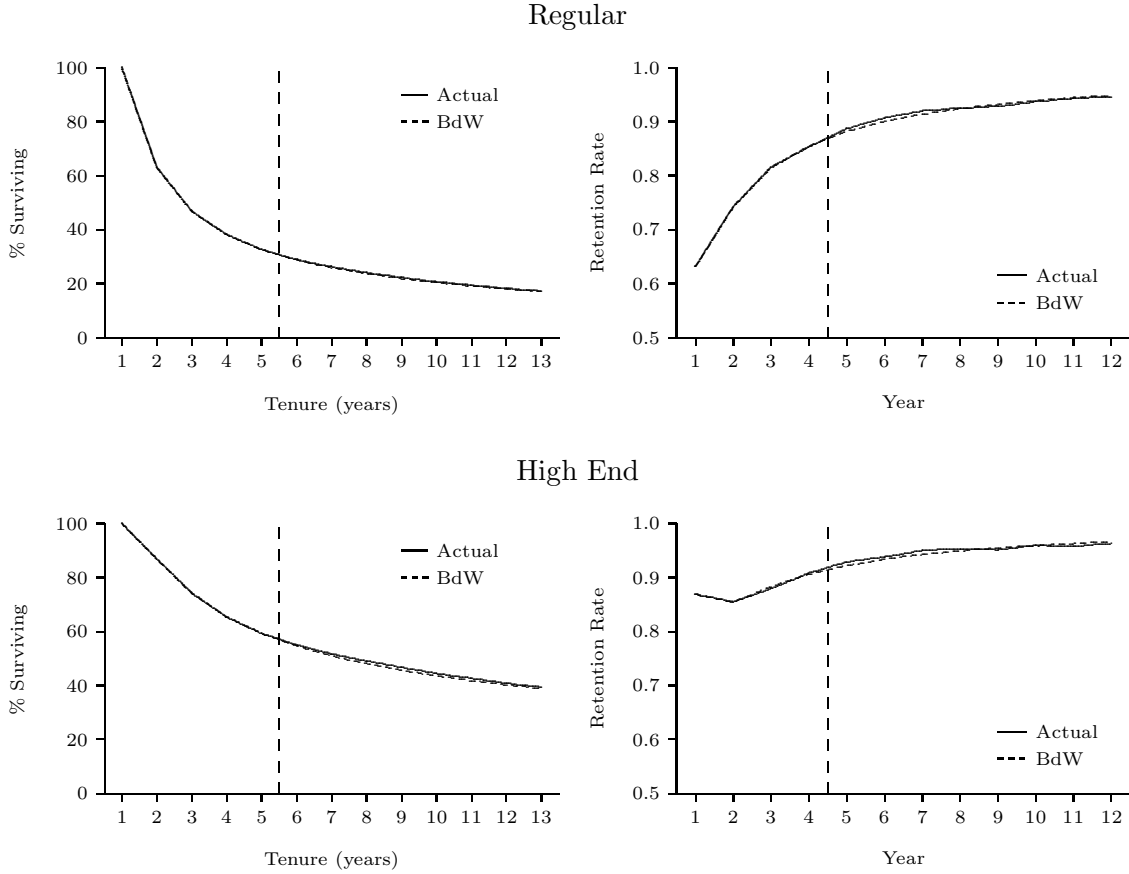


Figure 4: Comparing actual and BdW-model-based estimates of survival (LHS) and retention (RHS) for the Regular (top) and High End (bottom) datasets given an five-year model calibration period. (The model-based numbers to the right of the vertical dashed line are projections given the parameter values estimated using the data to the left of this line.)

estimates for the High End dataset in Table 2, $\hat{\phi}_{BG} = 0.099$; given the parameter estimates in Table 4, $\hat{\phi}_{BdW} = 0.335$. We observe that there is greater heterogeneity in the presence of the positive duration dependence to capture the dominant pattern of increasing aggregate retention rates observed in the data.

5 Exploring the Shape of $r(t)$

Our analysis of the High End dataset demonstrates that the BdW model can capture an early dip in the aggregate retention rate curve, something that is not immediately obvious given the underlying model assumptions. The explanation given is one of the “battle” between individual-

level duration dependence and cross-sectional heterogeneity. We now undertake a more thorough investigation of the shape of $r(t)$ under the BdW model.

Let us start by considering three scenarios: Case 1 ($\gamma = 4.75$ and $\delta = 14.25$), Case 2 ($\gamma = 0.5$ and $\delta = 1.5$), and Case 3 ($\gamma = 0.083$ and $\delta = 0.250$). While the associated distributions of Θ have the same mean ($E(\Theta) = 0.25$), they take on quite different shapes (Figure 5). In Case 1, the distribution of Θ is relatively homogeneous ($\phi = 0.05$) with an interior mode. In Case 2, there is quite a bit of heterogeneity ($\phi = 0.33$) in the distribution of Θ , with the majority of individuals having lowish values of θ . The heterogeneity in Case 3 ($\phi = 0.75$) is extreme; this U-shaped distribution indicates that some of the acquired customers have a high value of θ (which maps to a low probability of renewal), while a larger number of customers have small values of θ .

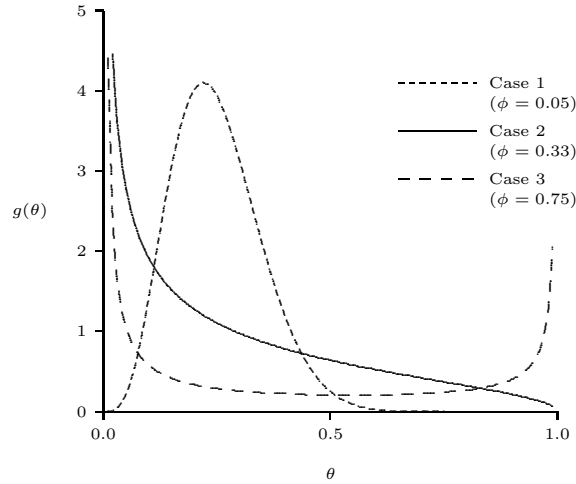


Figure 5: Shape of the beta distribution for Cases 1–3.

When $c < 1$, the cohort-level retention rates always increase over time; see Figure 6a, which shows the cohort-level retention rates when $c = 0.75$ for the completely homogeneous case as well as for Cases 1–3. Referring back to Figure 5, the distribution of Θ for Case 3 is very polarized. This means we have a group of people with high values of θ and they churn almost immediately, leaving us with the group of customers who have very low values of θ (and, because $c < 1$, their churn probabilities only get smaller over time). Thus $r(t)$ jumps and then levels off very quickly. In Case 1, where the distribution of Θ is relatively more homogeneous, the leveling-off process is slower, and is much closer to the homogeneous case. Case 2 lies between these extremes.

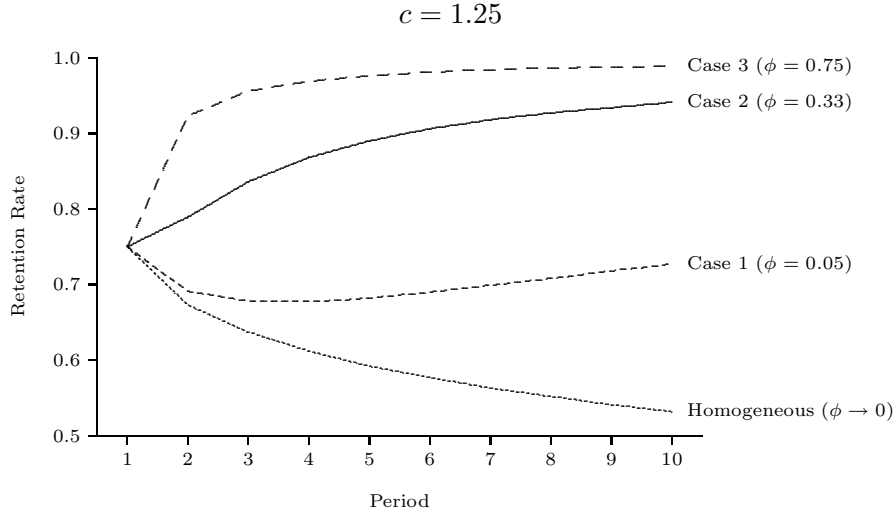
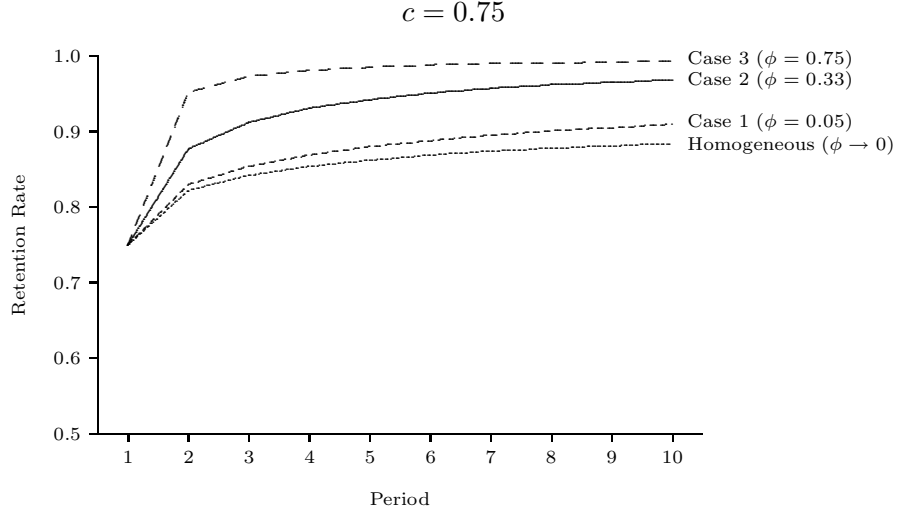


Figure 6: Shape of the beta-discrete-Weibull retention curve for different levels of heterogeneity in Θ and different values of c (with $E(\Theta) = 0.25$ in all cases).

When $c > 1$, an individual's retention probability decreases over time. But what is the countervailing effect of heterogeneity? Consider Figure 6b which shows the cohort-level retention rates when $c = 1.25$ for differing levels of heterogeneity in the distribution of Θ across the cohort members. In the completely homogeneous case, we observe a monotonically decreasing retention rate. But when any heterogeneity is present, we see the “ruse of heterogeneity” (Vaupel and

Yashin 1985): even though the individual-level customer retention probabilities are decreasing over time, the effect of a moderate amount of heterogeneity in the distribution of Θ across cohort members is to cause the retention rate to start increasing, either immediately or (if ϕ is low) after a few periods. We note that model can accommodate a one-period dip in retention rates (as observed in Figure 4) or a more prolonged dip (as observed in Case 1). As the level of heterogeneity increases, as in Cases 2 and 3, we find that the cohort-level retention rate monotonically increases over time (even though the individual-level retention probability is a decreasing function of time).

To further explore the relationship between heterogeneity and dynamics, we look at the shape of the BdW retention curve as a function of c and ϕ for three different values of $E(\Theta)$ (0.10, 0.25, 0.40)—see Figure 7. The monotonically decreasing curve appears only in the degenerate case where there is no heterogeneity ($\phi \rightarrow 0$), so we ignore this special case to focus on the other two general shapes that can occur. When $c < 1$, the aggregate retention curve must increase monotonically. For this range of the shape parameter c , each individual’s renewal probability increases over time. The sorting effect of heterogeneity also pushes the curve up. Since both forces work in the same direction, the resulting aggregate retention curve must rise over time. When $c > 1$, either shape can arise depending on the strength of duration dependence and the level of heterogeneity in the underlying distribution of Θ . A U-shaped curve requires customers to have an increasing propensity to churn in order to generate the initial dip. The latter part of the U results from the sorting effect of heterogeneity. However, as the level of heterogeneity increases, the sorting effect of heterogeneity dominates individual-level duration dependence and we have a monotonically increasing retention curve. Finally, as seen by the gradual “fanning out” of the curved line towards the top of the figure, we note that the shape of the BdW retention curve is surprisingly insensitive to variations in $E(\Theta)$.

To gain additional insight into the U-shaped versus monotonically increasing shape of $r(t)$, let us further investigate the shape of the BdW retention curve for different levels of heterogeneity in θ , with $E(\Theta) = 0.25$ and $c = 1.25$ in all cases. We see in Figure 8 that the location of the minimum of $r(t)$ increases as the level of heterogeneity in θ decreases, with its location going to infinity as $\phi \rightarrow 0$. As the level of heterogeneity increases, the location of the dip shifts to the left and it becomes less deep. We then get to the point where $r(1) = r(2)$ after which $r(t)$

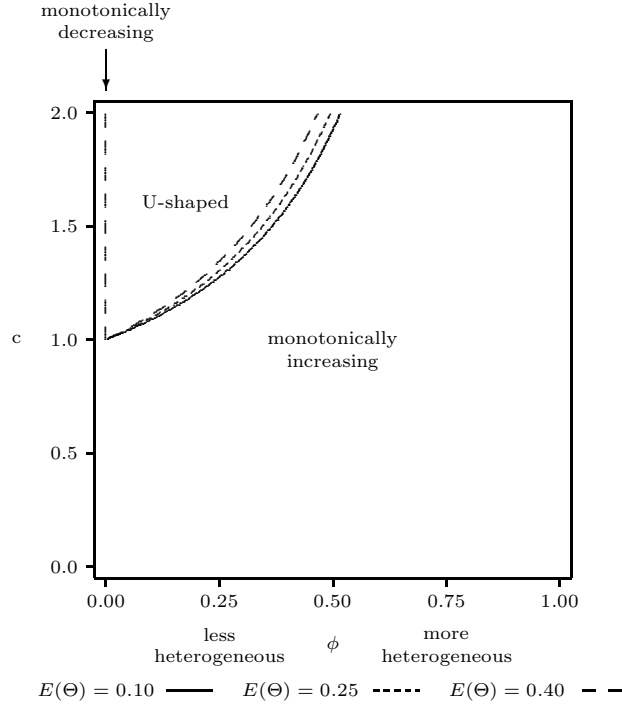


Figure 7: Shape of the beta-discrete-Weibull retention curve as a function of c and ϕ for $E(\Theta) = \{0.10, 0.25, 0.40\}$.

becomes S-shaped and then concave.

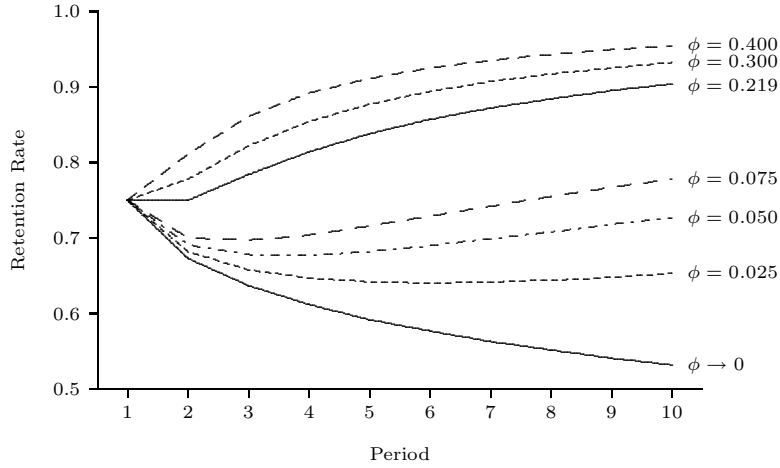


Figure 8: Evolution of the shape of the beta-discrete-Weibull retention curve as the level of heterogeneity in Θ increases (with $E(\Theta) = 0.25$ and $c = 1.25$ in all cases).

Clearly $c > 1$ when $r(t)$ is U-shaped. Is there anything we can learn about the sign of c from the shape of a monotonically increasing $r(t)$? The answer is yes. Given $r(1)$, $r(2)$, and $r(3)$, we can (numerically) solve for the three BdW model parameters (γ, δ, c) . We present in

Figure 9 a plot that indicates whether c is greater than or less than 1 as a function of $r(2)$ and $r(3)$ given $r(1) = 0.60$ and $r(1) = 0.75$. (We limit the analysis to the region of interest above the dashed line where retention rates increase monotonically ($r(1) \leq r(2) \leq r(3)$)). We note that the area where the aggregate retention curve is monotonically increasing while each individual's renewal probability increases over time (i.e., $c < 1$) is very small. In order for c to be less than 1, $r(2)$ must be larger than $r(1)$ but $r(3)$ cannot be much larger than $r(2)$. In other words, the jump in the aggregate retention rate between periods 2 and 3 is crucial to determining whether there is positive or negative dependence at the level of the individual customer: the bigger the difference between $r(2)$ and $r(3)$, the more positive duration dependence there is in the data (i.e., individual customers' propensities to churn *increase* over time). (We observe this in Figure 6 (where $r(1) = 0.75$). For the Case 3 retention curves, $r(2) = 0.952$ and $r(3) = 0.973$ in Figure 6a, while the corresponding numbers in Figure 6b are 0.923 and 0.956. The bigger gap is observed in Figure 6b, which corresponds to the case of $c > 1$.)

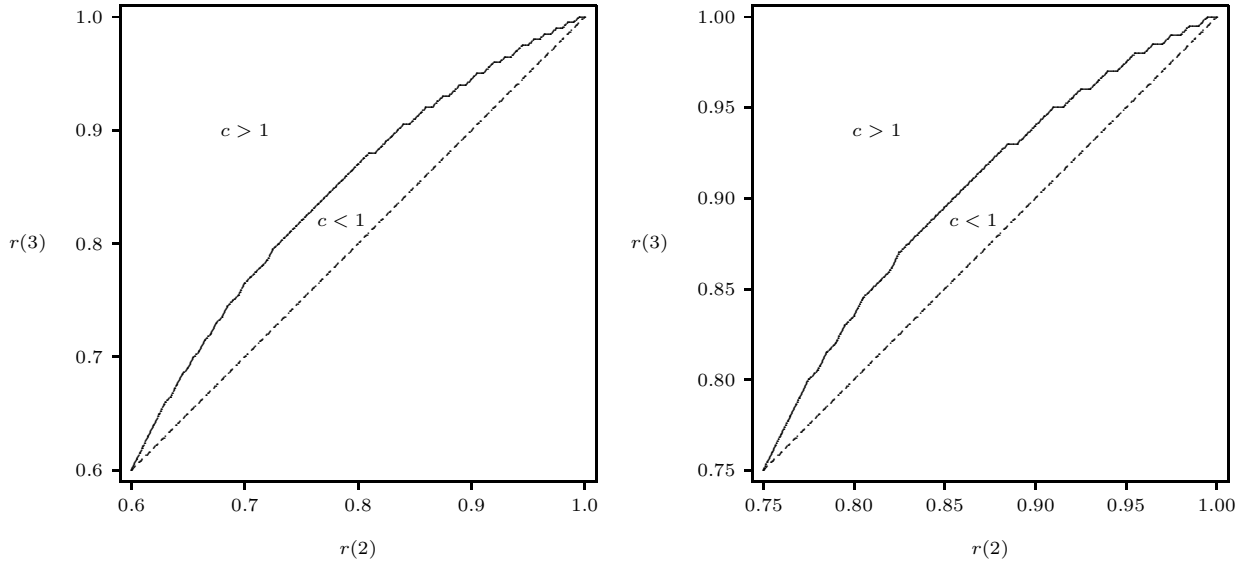


Figure 9: Nature of individual-level duration dependence as a function of $r(2)$ and $r(3)$ given $r(1) = 0.60$ (LHS) and $r(1) = 0.75$ (RHS).

6 Alternative Model Specifications

The conclusions drawn above about the nature of retention dynamics are based on the assumption that the BdW is a valid characterization of the true data-generating process. In this section,

we investigate the robustness of these results by exploring some alternative model specifications. We first consider a model based on an alternative distribution that allows for retention dynamics at the level of the individual customer, and find that our results hold. This analysis still suffers from a key assumption associated with our BdW-based analysis, that of homogeneity in the parameter that determines whether individual-level retention probabilities are decreasing, constant, or increasing over time. We therefore undertake some analysis in which we allow the c parameter of the discrete-Weibull distribution to vary across customers. We do not find any strong evidence to support such an effect.

6.1 Changing the Underlying Weibull Distribution

In Section 4, we noted that the natural starting point for accommodating individual-level retention probability dynamics in a continuous-time environment would be to replace the exponential distribution (the continuous-time equivalent of the geometric distribution) with the Weibull distribution (which reduces to the exponential when $c = 1$). Alternatively, we could assume that the continuous lifetime random variable X is distributed gamma, with pdf

$$f(x | \lambda, s) = \frac{\lambda^s x^{s-1} e^{-\lambda x}}{\Gamma(s)}.$$

where $\lambda > 0$ is the rate (or scale) parameter and $s > 0$ is the shape parameter. The gamma distribution exhibits negative duration dependence when $s < 1$ and positive duration dependence when $s > 1$. This corresponds to increasing and decreasing retention rates, respectively. When $s = 1$, the gamma distribution collapses to the exponential distribution. While there is no closed-form expression for the associated survivor function, it can be written in terms of the incomplete gamma function:

$$S(x | \lambda, s) = 1 - \frac{\gamma(s, \lambda x)}{\Gamma(s)}.$$

Each customer's λ is unobserved (and unobservable) and is treated as a realization of the random variable Λ , which we assume to be distributed gamma with pdf

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

Assuming the value of s is constant across the population, it follows that, for a randomly chosen individual,

$$\begin{aligned} S(x | r, \alpha, s) &= \int_0^\infty S(x | \lambda, s) g(\lambda | r, \alpha) d\lambda \\ &= 1 - \frac{1}{sB(r, s)} \left(\frac{\alpha}{\alpha + x} \right)^r \left(\frac{x}{\alpha + x} \right)^s {}_2F_1(r + s, 1; s + 1; \frac{x}{\alpha + x}), \end{aligned}$$

where ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function. This gamma mixture of gammas is sometimes known as the “beta of second kind” (B2) distribution. The implied retention curve is monotonically increasing when $s \leq 1$ and U-shaped when $s > 1$.^{6,7}

The B2 is a continuous-time distribution. As noted in Section 4, the discrete-time equivalent of a continuous distribution can be constructed by treating the discrete lifetime variable as the integer part of the continuous lifetime and discretizing its survivor function. For $T = \lfloor X \rfloor$,

$$P(T = t | r, \alpha, s) = S(t - 1 | r, \alpha, s) - S(t | r, \alpha, s), t = 1, 2, 3, \dots$$

Fitting this model to the High End dataset using a five-year model calibration period yields the following parameter estimates: $\hat{r} = 0.483$, $\hat{\alpha} = 0.562$, and $\hat{s} = 2.721$; the associated value of the log-likelihood function is -1222.8 .⁸ The model-based estimates of survival and retention are compared against the actual numbers in Figure 10.

In terms of fit and forecasting performance (for both survival and retention), the results are almost identical to those associated with the BdW model. Our estimate of s is greater than 1, which implies that individual churn probabilities increase with tenure, even though the aggregate retention rate increases once we are past the initial dip. This demonstrates that the conclusions drawn using the BdW model are robust to changes in the underlying model specification.

⁶Strictly speaking, the B2 has what is known in the survival analysis literature as an upside-down bathtub-shaped failure-rate/hazard function (Glaser 1980) when $s > 1$; this maps to a U-shaped retention curve.

⁷When $s = 1$, this collapses to the Pareto distribution of the second kind, which is the continuous-time equivalent of the BG model (Fader et al. 2017).

⁸The fact that the B2 survivor function contains the Gaussian hypergeometric function means that it is impractical to estimate the model parameters in Excel. Our maximum likelihood estimates of the model parameters were obtained using MATLAB.

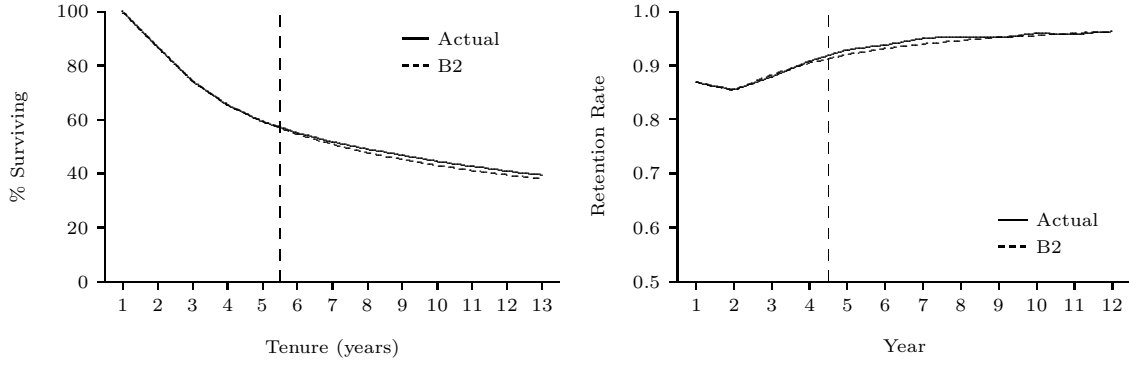


Figure 10: Comparing actual and B2-model-based estimates of survival (LHS) and retention (RHS) for the High End dataset given an five-year model calibration period. (The model-based numbers to the right of the vertical dashed line are projections given the parameter values estimated using the data to the left of this line.)

6.2 Relaxing the Assumption of Common c

The dW distribution has two parameters: θ and c . When deriving the BdW distribution, we allowed θ to vary across customers but assumed that everyone has the same c . To what extent could our counterintuitive results regarding individual-level retention dynamics be a consequence of this assumption? In other words, what happens if we allow c to vary across customers?

To the best of our knowledge, there is no continuous distribution $f(c)$ that can be used to characterize heterogeneity in c that results in a closed-form solution to the integral

$$\int_0^\infty (1 - \theta)^{t^c} f(c) dc.$$

We will therefore capture heterogeneity in c using a discrete mixing distribution.

Let us consider the two-component dW distribution with survivor function

$$S(t | \theta_1, \theta_2, c_1, c_2, \pi) = \pi S(t | \theta_1, c_1) + (1 - \pi) S(t | \theta_2, c_2), 0 < \pi < 1.$$

The five model parameters are not identified if we use the five-year model calibration period (as we only observe four renewal opportunities). We will therefore use the whole dataset, which contains 12 renewal opportunities, in our investigations of heterogeneity in c . The estimation results are reported in Table 5, along with the results for all the nested models that “switch off”

Model Specification:				
– heterogeneity in θ	✓	✓	✗	✗
– heterogeneity in c	✓	✗	✓	✗
θ_1	0.068	0.019		
θ_2	0.288	0.302		
θ			0.135	0.160
c_1	0.853		0.668	
c_2	1.386		2.230	
c		1.232		0.688
π	0.710	0.599	0.840	
LL	−2004.3	−2004.5	−2005.3	−2027.7
AIC	4018.8	4017.0	4018.6	4059.4
BIC	4043.2	4036.6	4038.2	4069.2
Evidence ratio	4.5	1.9	4.4	3×10^9

Table 5: Estimation results for the two-component dW distribution and its nested variants.

the heterogeneity (i.e., $\theta_1 = \theta_2 = \theta$ and/or $c_1 = c_2 = c$).⁹

When we allow for heterogeneity in both θ and c , we see that one segment of the customer base exhibits negative duration dependence while the other exhibits positive duration dependence. At first glance, this would suggest that our assumption of homogeneity in c is not supported. However, we must first compare the fit of this specification to that of its nested variants. On the basis of both AIC and BIC, the specification with heterogeneity in θ and homogeneity in c is the best model. Estimating the BdW model using the full dataset yields the following results: $\hat{\gamma} = 0.250$, $\hat{\delta} = 1.654$, $\hat{c} = 1.597$, $LL = -2004.8$. The associated values of AIC and BIC are 4015.6 and 4030.4, respectively. This means the BdW is the “best” model among those examined in this analysis.

We also report in Table 5 the evidence ratio (Anderson 2008, Burnham and Anderson 2002), which tells us the strength of the empirical support for the model with the minimum AIC relative to the other candidate models.¹⁰ The evidence for the BdW is 4.5 times that for the

⁹What about the fit of a three-component model? The value of the log-likelihood function is -2004.1 . This improvement of 0.2 (relative to the log-likelihood associated with the two-component model) comes at a cost of three additional model parameters, and so the three-component model is clearly dominated by the two-component model.

¹⁰The evidence ratio is the relative likelihood of a pair of models. The evidence ratio for the best model (i.e., the one with the lowest AIC) versus model i is computed as $E_i = \exp((AIC_i - AIC_{\min})/2)$. An evidence ratio of E_i means that the probability that the model with the lowest AIC is the K-L best model is E_i times that of model i . By K-L best, we mean that the model has the smallest estimated Kullback-Leibler distance (i.e., it

full two-component dW model and 4.4 times that for the model that allows for heterogeneity in c but not in θ . The evidence for the BdW is only 1.9 times that for the two-component model with homogeneous c . The conclusion we draw from this is that there is no reason to reject our modeling assumption of homogeneity in c .

We find similar results when we bring in the beta distribution for θ and use a discrete mixture for c alone. Even though there is intuitive appeal for having heterogeneity in c , there is very little empirical support for it. Most of the “action” is in the heterogeneity in the baseline churn propensities across customers; once this factor is accommodated, there is virtually nothing left over to be explained by a more elaborate model specification.

7 Discussion

Generally speaking, aggregate/cohort-level retention rates increase over time, and the BG model has proven to be a robust tool for projecting retention (and therefore survival) into the future. According to the BG model, this phenomenon is entirely due to heterogeneity; individual-level propensities to churn are assumed to be constant. Despite the performance of the model, a number of people struggle with this assumption, contending that increasing cohort-level retention rates are the result of individual-level propensities to churn decreasing over time.

Occasionally we observe an initial dip in the cohort-level retention rates before they increase. This phenomenon cannot be captured by the BG model. We have presented the BdW model as an extension to the BG model, one that relaxes the assumption of time-invariant individual-level propensities to churn. This model is sufficiently flexible to capture the phenomenon of non-monotonically increasing cohort-level retention rates.¹¹ If the aggregate retention curve is U-shaped then individual-level churn probabilities must increase over time (i.e., $c > 1$). However, $c > 1$ does not guarantee that the aggregate retention curve is U-shaped; the effect of heterogeneity can swamp the individual-level positive duration dependence to yield a monotonically increasing aggregate retention curve. Our analysis suggests that when the assumption of

has the lowest information lost among the models used to approximate the true data-generating process).

¹¹As previously noted, Fader and Hardie (2010) explore how to use the BG model as the basis for calculating CLV. We show in Appendix C how such calculations can be performed assuming lifetimes are characterized by the BdW model.

constant individual-level propensities to churn is violated in a setting where aggregate retention rates are increasing, it is most likely that these individual propensities *increase* with tenure (rather than decrease, as conventional wisdom would suggest).

This surprising result that, if not constant, individual-level churn propensities are expected to increase over time is supported by other researchers. In an analysis of health-club membership data, Giudicati et al. (2013) find a negative correlation between length of membership and the probability of a member renewing their contract. Lemmens and Croux (2006) find that a customer's churn probability is positively correlated with the length of time they have owned their current phone. Jamal and Bucklin (2006) use a latent-class Weibull model in their analysis of churn among customers of a direct-to-home satellite TV provider. In all three segments, the estimate of c is greater than 1. Schweidel et al. (2008) use a Weibull-gamma model (a continuous-time analogue of the BdW) to analyze churn among customers of a telecommunications provider and find that $c > 1$. In a non-marketing context, Morrison and Schmittlein (1980) observe the same result in some analyses of job duration data.

What are some potential causes for such an effect? Possible explanations include the novelty of the new service wearing off or boredom setting-in over time, increasing competitive pressures, and changes in consumer preferences that are not matched by changes in the firm's offerings. In addition to exploring such possible causes, future research should explore the robustness of this result across other product categories. Given the increasing role of networks in society, it is important that we understand the role of network externalities. Of interest is how changes in network structure (and the customer's centrality in the network) affect retention rates.

Appendix A: Deriving the Distribution of $P(t)$

Recall from (9) that, conditional on θ and c , the individual-level probability that someone who has made $t - 1$ renewals will renew at the next opportunity is

$$\rho(t | \theta, c) = (1 - \theta)^{t^c - (t-1)^c}, \quad t = 1, 2, 3, \dots \quad (\text{A1})$$

The distribution of Θ across those individuals who have made $t - 1$ renewals is simply the posterior distribution of Θ for the BdW model:

$$\begin{aligned} f(\theta | \gamma, \delta, c; t - 1 \text{ renewals}) &= \frac{S(t - 1 \text{ renewals} | \theta, c) f(\theta | \gamma, \delta)}{S(t - 1 \text{ renewals} | \gamma, \delta, c)} \\ &= \frac{(1 - \theta)^{(t-1)^c} \frac{\theta^{\gamma-1} (1 - \theta)^{\delta-1}}{B(\gamma, \delta)}}{\frac{B(\gamma, \delta + (t-1)^c)}{B(\gamma, \delta)}} \\ &= \frac{\theta^{\gamma-1} (1 - \theta)^{\delta + (t-1)^c - 1}}{B(\gamma, \delta + (t-1)^c)}, \quad t = 1, 2, 3, \dots \end{aligned}$$

We derive the distribution of $P(t)$ using the basic result for deriving the distribution of the function $Y = g(X)$ of random variable X :

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)). \quad (\text{A2})$$

Rewriting (A1) as $\rho(t) = g(\theta)$, we have

$$\begin{aligned} g^{-1}(\rho(t)) &= 1 - \rho(t)^{\frac{1}{t^c - (t-1)^c}} \text{ and} \\ \frac{d}{d\rho(t)} g^{-1}(\rho(t)) &= -\frac{1}{t^c - (t-1)^c} \rho(t)^{\frac{1}{t^c - (t-1)^c} - 1}. \end{aligned}$$

It follows from (A2) that

$$\begin{aligned} f(\rho(t) | \gamma, \delta, c) &= \frac{1}{t^c - (t-1)^c} \rho(t)^{\frac{1}{t^c - (t-1)^c} - 1} \frac{\left\{ 1 - \rho(t)^{\frac{1}{t^c - (t-1)^c}} \right\}^{\gamma-1} \left\{ \rho(t)^{\frac{1}{t^c - (t-1)^c}} \right\}^{\delta + (t-1)^c - 1}}{B(\gamma, \delta + (t-1)^c)} \end{aligned}$$

$$= \frac{1}{t^c - (t-1)^c} \frac{1}{\rho(t)} \frac{\left\{1 - \rho(t)^{\frac{1}{t^c - (t-1)^c}}\right\}^{\gamma-1} \left\{\rho(t)^{\frac{1}{t^c - (t-1)^c}}\right\}^{\delta + (t-1)^c}}{B(\gamma, \delta + (t-1)^c)}, \quad t = 1, 2, 3, \dots$$

The mean of this distribution is

$$E[P(t) \mid \gamma, \delta, c] = \int_0^1 \frac{1}{t^c - (t-1)^c} \frac{\left\{1 - \rho(t)^{\frac{1}{t^c - (t-1)^c}}\right\}^{\gamma-1} \left\{\rho(t)^{\frac{1}{t^c - (t-1)^c}}\right\}^{\delta + (t-1)^c}}{B(\gamma, \delta + (t-1)^c)} d\rho(t)$$

which, letting $z = \rho(t)^{\frac{1}{t^c - (t-1)^c}}$,

$$\begin{aligned} &= \frac{1}{B(\gamma, \delta + (t-1)^c)} \int_0^1 (1-z)^{\gamma-1} z^{\delta + t^c - 1} dz \\ &= \frac{B(\gamma, \delta + t^c)}{B(\gamma, \delta + (t-1)^c)}, \quad t = 1, 2, 3, \dots, \end{aligned}$$

which is $r(t)$ under the BdW model.

Appendix B: Implementing the BdW Model in Excel

We briefly describe how to estimate the BdW model parameters using Microsoft Excel. It is assumed that the reader is familiar with the basics of estimating the parameters of the BG model, as covered in FH, Appendix B.

With reference to Figure B1, we “code up” our expression for $S(t)$ (cells F7:F11) and compute $P(T = t)$ as $S(t-1) - S(t)$ (cells E8:E11). So as to reduce the chances of any numerical precision problems, we compute $S(t | \gamma, \delta, c)$ as $\exp(\ln(B(\gamma, \delta + t^c)) - \ln(B(\gamma, \delta)))$. The elements of the log-likelihood function are computed in cells G8:G12, and their sum computed in cell B4.

	A	B	C	D	E	F	G	H	I
1	gamma	1.000		ln(B(gamma, delta))		0.000			
2	delta	1.000							
3	c	1.000		=GAMMALN(B1)+GAMMALN(B2)- GAMMALN(B1+B2)					
4	LL	-1674.3							
5									
6	Year	t	# Cust.	# Lost	P(T=t)	S(t)			
7	1	0	1000			1.0000			
8	2	1	869	131	0.5000	0.5000	-90.80		
9	3	2	74	16	0.1667	0.3333	-79.76		
10	4	3	65	90	=EXP(GAMMALN(\$B\$1)+GAMMALN(\$B\$2+A8^\$B\$3) -GAMMALN(\$B\$1+\$B\$2+A8^\$B\$3)-\$E\$1)				
11	5	4	593	60					
12							-954.40		

Figure B1: Screenshot of the Excel worksheet for parameter estimation (High End dataset)

The values of γ, δ, c that maximize the value of the log-likelihood function are found using the Excel add-in Solver—see Figure B2. (In this particular case, the log-likelihood function is quite flat near its maximum. The parameter estimates reported in Table 4 are obtained by running Solver twice, with the solution from the first “run” serving as starting values for the second “run”.) With reference to Figure B3, we can now project $S(t)$ by copying cell F11 down to F19, and compute the associated retention rates as $S(t)/S(t-1)$ (cells H8:H19).

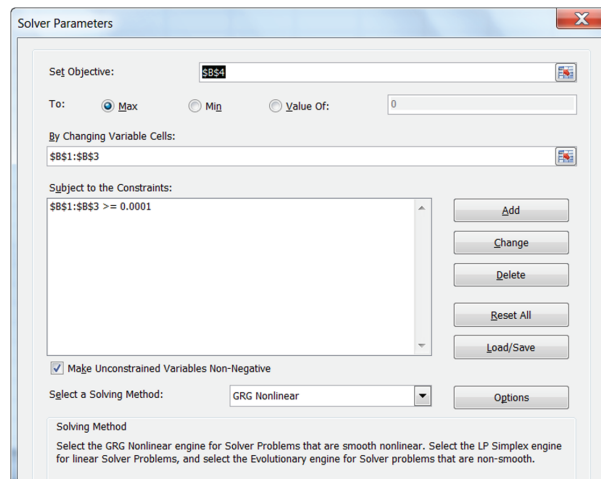


Figure B2: Solver settings

	A	B	C	D	E	F	G	H
1	gamma	0.259		ln(B(gamma, delta))		1.166		
2	delta	1.722						
3	c	1.584						
4	LL	-1222.7						
5								
6	Year	t	# Cust.	# Lost	P(die)	S(t)		r(t)
7	1	0	1000			1.0000		
8	2	1	869	131	0.1309	0.8691	-266.41	0.8691
9	3	2	743	126	0.1272	0.7419	=F8/F7	0.8536
10	4	3	653	90	0.0871	0.6548	-219.02	0.8825
11	5	4	593	60	0.0619	0.5929	-166.98	0.9055
12	6	5				0.5466	-309.94	0.9219
13	7	6				0.5104		0.9337
14	8	7				0.4810		0.9425
15	9	8				0.4567		0.9494
16	10	9				0.4360		0.9548
17	11	10				0.4182		0.9592
18	12	11				0.4027		0.9628
19	13	12				0.3889		0.9658

Figure B3: Screenshot of the Excel worksheet for projecting survival and retention (High End dataset)

Appendix C: Computing CLV under the BdW Model

Fader and Hardie (2010) introduce the idea of *discounted expected lifetime* and *discounted expected residual lifetime*, which they label DEL and DERL, and show how these quantities are of use when computing (expected) customer lifetime value and residual lifetime value in contractual settings. It is more correct to think of the random variables discounted lifetime (DL) and discounted residual lifetime (DRL) and to compute their means, giving us expected discounted lifetime ($E(DL)$) and expected discounted residual lifetime ($E(DRL)$).

For discount rate d , the expected discounted lifetime of an as-yet-to-be-acquired customer is, by definition,

$$E(DL) = \sum_{t=0}^{\infty} \frac{S(t)}{(1+d)^t}, \quad (C1)$$

while expected discounted residual lifetime of customer at the end of period n (i.e., someone who has made $n - 1$ renewals) is

$$E(DRL | \text{active for } n \text{ periods}) = \sum_{t=n}^{\infty} \frac{S(t | T > n - 1)}{(1+d)^{t-n}}. \quad (C2)$$

When the duration of a customer's relationship with the firm is characterized by the BG model, we can substitute (3) in (C1) and (C2), solve the infinite sums, and derive closed-form expressions for $E(DL)$ and $E(DRL)$. This is not the case when lifetimes are characterized by the BdW model. However, we can simply evaluate (C1) and (C2) terminating the series at a point where additional terms are effectively zero. With reference to Figure C1, we first compute $S(t)$ for $t = 0, 1, 2, \dots, 100$ using (10) in column C. Given an annual discount rate of 10% (cell B4, the discount factor is computed in column E. Summing up the product of the two terms (cell E6) gives us the value of $E(DL)$ for an as-yet-to-be-acquired customer.

In order to compute $E(DRL)$, we need to evaluate the conditional survivor function $S(t | T > n - 1)$. Standing at the end of year 5, the customer has made four renewals, and we can compute the conditional survivor function as $S(t)/S(4)$ —see column G. Given the discount factor computed in column H, we sum up the product of the two terms (cell H6), giving us the value of $E(DRL)$.

	A	B	C	D	E	F	G	H	I
1	gamma	0.259		ln(B(gamma, delta))	1.166				
2	delta	1.722							
3	c	1.584	=SUMPRODUCT(C9:C108,E9:E108)						
4	d	0.1				=SUMPRODUCT(G14:G108,H14:H108)			
5									
6				E(DL)	5.99		E(DRL)	7.32	
7							Given 4 renewals		
8	Year	t	S(t)		disc.		S(t T>4)	disc.	
9	1	0	1.0000		1.0000				
10	2	1	0.8691		0.9091				
11	3	2	=1/(1+B\$4)^B10		0.8264		=1/(1+B\$4)^(B14-5)		
12	4	3			0.7513				
13	5	4	0.5929		0.6830				
14	6	5	0.5466		0.6209		0.9219	1.0000	
15	7	6	0.5104		0.5645		0.8608	0.9091	
16	8	7	0.4810		0.5132	=C14/C\$13		0.8113	0.8264
17	9	8	0.4567		0.4665		0.7702	0.7513	
18	10	9	0.4360		0.4241		0.7354	0.6830	
19	11	10	0.4182		0.3855		0.7054	0.6209	
20	12	11	0.4027		0.3505		0.6791	0.5645	
21	13	12	0.3889		0.3186		0.6559	0.5132	
22	14	13	0.3766		0.2897		0.6352	0.4665	
23	15	14	0.3656		0.2633		0.6166	0.4241	
106	98	97	0.1659		0.0001		0.2798	0.0002	
107	99	98	0.1652		0.0001		0.2786	0.0001	
108	100	99	0.1645		0.0001		0.2774	0.0001	

Figure C1: Screenshot of the Excel worksheet for computing $E(DL)$ and $E(DRL)$ under the BdW model

References

- Anderson, David R. (2008), *Model Based Inference in the Life Sciences: A Primer on Evidence*, New York, NY: Springer Science+Business Media, LLC.
- Berry, Michael J. A. and Gordon S. Linoff (2004), *Data Mining Techniques*, 2nd edition, Indianapolis, IN: Wiley Publishing, Inc.
- Bolton, Ruth N. (1998), “A Dynamic Model of the Duration of the Customer’s Relationship with a Continuous Service Provider: The Role of Satisfaction,” *Marketing Science*, **17** (1), 45–65.
- Bracquemond, Cyril and Olivier Gaudoin (2003), “A Survey on Discrete Lifetime Distributions,” *International Journal of Reliability, Quality and Safety Engineering*, **10** (1), 69–98.
- Buchanan, Bruce and Donald G. Morrison (1988), “A Stochastic Model of List Falloff with Implications for Repeat Mailings,” *Journal of Direct Marketing*, **2** (Summer), 7–15.
- Burnham, Kenneth P. and David R. Anderson (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach*, 2nd edition, New York, NY: Springer-Verlag, Inc.
- Burnham, Thomas A., Judy K. Frels, and Vijay Mahajan (2003), “Consumer Switching Costs: A Typology, Antecedents, and Consequences,” *Journal of the Academy of Marketing Science*, **31** (Spring), 109–126.
- Fader, Peter S. and Bruce G. S. Hardie (2007), “How to Project Customer Retention,” *Journal of Interactive Marketing*, **21** (Winter), 76–90.
- Fader, Peter S. and Bruce G. S. Hardie (2010), “Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity,” *Marketing Science*, **29** (January–February), 85–93.
- Fader, Peter S. and Bruce G. S. Hardie (2014), “A Spreadsheet-Literate Non-Statistician’s Guide to the Beta-Geometric Model.” <<http://brucehardie.com/notes/032/>>
- Fader, Peter S., Bruce G. S. Hardie, Daniel McCarthy, and Ramnath Vaidyanathan (2017), “Exploring the Equivalence of Two Common Mixture Models for Duration Data,” unpublished working paper.
- Fox, Richard J., Srinivas K. Reddy, and Bharat Rao (1997), “Modeling Response to Repetitive Promotional Stimuli,” *Journal of the Academy of Marketing Science*, **25** (3), 242–255.
- Giudicati, Gianna, Massimo Riccaboni, and Anna Romiti (2013), “Experience, Socialization and Customer Retention: Lessons From the Dance Floor,” *Marketing Letters*, **24** (December), 409–422.
- Glaser, Ronald E. (1980), “Bathtub and Related Failure Rate Characterizations,” *Journal of the American Statistical Association*, **75** (September), 667–672.
- Israel, Mark (2005), “Tenure Dependence in Consumer-Firm Relationships: An Empirical Analysis of Consumer Departures from Automobile Insurance Firms,” *The RAND Journal of Economics*, **36** (Spring), 165–192.

- Jamal, Zainab and Randolph E. Bucklin (2006), "Improving the Diagnosis and Prediction of Customer Churn: A Heterogeneous Hazard Modeling Approach," *Journal of Interactive Marketing*, **20** (Summer/Autumn), 16–29.
- Kiefer, Nicholas M. (1988), "Economic Duration Data and Hazard Functions," *Journal of Economic Literature*, **26** (June), 646–679.
- Lai, C.D. (2013), "Issues Concerning Constructions of Discrete Lifetime Models," *Quality Technology and Quantitative Management*, **10** (2), 251–262.
- Lemmens, Aurélie and Christophe Croux (2006) "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, 43 (May), 276–286.
- Morrison, Donald and Arnon Perry (1970), "Some Data Based Models for Analyzing Sales Fluctuations," *Decision Sciences*, **1** (3–4), 258–274.
- Morrison, Donald G. and David C. Schmittlein (1980), "Jobs, Strikes, and Wars: Probability Models for Duration," *Organizational Behavior and Human Performance*, 25 (April), 224–251.
- Murthy, D.N. Prabhakar, Min Xie, and Renyan Jiang (2004), *Weibull Models*, Hoboken, NJ: John Wiley & Sons.
- Nakagawa, Toshio and Shunji Osaki (1975), "The Discrete Weibull Distribution," *IEEE Transactions on Reliability*, **24** (December), 300–301.
- Nitzan, Irit, Barak Libai, and Danit Ein-Gar (2011), "The Time of Decreasing Retention: Payments as a Virtual Lock-in," unpublished working paper.
- Pielou, E. C. (1962), "Runs of One Species with Respect to Another in Transects through Plant Populations," *Biometrics*, **18** (December), 579–593.
- Potter, R. G. and M. P. Parker (1964), "Predicting the Time Required to Conceive," *Population Studies*, **18** (July), 99–116.
- Proschan, Frank (1963), "Theoretical Explanation of Observed Decreasing Failure Rate," *Technometrics*, **5** (August), 375–383.
- Reichheld, Frederick F. (1996), *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*, Boston, MA: Harvard Business School Press.
- Rinne, Horst (2009), *The Weibull Distribution: A Handbook*, Boca Raton, FL: Chapman & Hall/CRC Press.
- Schweidel, David A., Peter S. Fader, and Eric Bradlow (2008), "Understanding Service Retention Within and Across Cohorts Using Limited Information," *Journal of Marketing*, **72** (January), 82–94.
- Vaupel, James W. and Anatoli I. Yashin (1985), "Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics," *The American Statistician*, **39** (August), 176–185.