Agenda

- Introduction to customer-base analysis
- The right way to think about computing CLV
- Review of probability models
- Models for contractual settings
- Models for noncontractual settings
  - The BG/BB model
  - The Pareto/NBD model
  - The BG/NBD model
- Beyond the basic models
Customer-Base Analysis

• Faced with a customer transaction database, we may wish to determine
  – which customers are most likely to be active in the future,
  – the level of transactions we could expect in future periods from those on the customer list, both individually and collectively, and
  – individual customer lifetime value (CLV).

• Forward-looking/predictive versus descriptive.

Comparison of Modelling Approaches

Traditional approach
future = f(past)

Probability modelling approach
\( \hat{\theta} = f(\text{past}) \rightarrow \text{future} = f(\hat{\theta}) \)
Classifying Analysis Settings

Consider the following two statements regarding the size of a company's customer base:

- Based on numbers presented in a January 2008 press release that reported Vodafone Group Plc's third quarter key performance indicators, we see that Vodafone UK has 7.3 million “pay monthly” customers.

- In his “Q4 2007 Financial Results Conference Call”, the CFO of Amazon made the comment that “[a]ctive customer accounts exceeded 76 million, up 19%” where alive customer accounts represent customers who ordered in the past year.

Classifying Analysis Settings

- It is important to distinguish between contractual and noncontractual settings:
  - In a contractual setting, we observe the time at customers become inactive.
  - In a noncontractual setting, the time at which a customer becomes inactive is unobserved.

- The challenge of noncontractual markets:
  How do we differentiate between those customers who have ended their relationship with the firm versus those who are simply in the midst of a long hiatus between transactions?
Classifying Analysis Settings

Consider the following four specific business settings:

- Airport VIP lounges
- Electrical utilities
- Academic conferences
- Mail-order clothing companies.

Classifying Customer Bases

<table>
<thead>
<tr>
<th>Continuous Opportunities for Transactions</th>
<th>Credit card</th>
<th>Student mealplan</th>
<th>Mobile phone usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery purchases</td>
<td>Doctor visits</td>
<td>Hotel stays</td>
<td>Credit card</td>
</tr>
<tr>
<td>Event attendance</td>
<td>Prescription refills</td>
<td>Charity fund drives</td>
<td>Noncontractual</td>
</tr>
<tr>
<td>Magazine subs</td>
<td>Insurance policy</td>
<td>Health club m’ship</td>
<td>Type of Relationship With Customers</td>
</tr>
</tbody>
</table>
The Right Way to Think About Computing
Customer Lifetime Value

Calculating CLV
Customer lifetime value is the present value of the future cash flows associated with the customer.

- A forward-looking concept
- Not to be confused with (historic) customer profitability
Calculating CLV

Standard classroom formula:

\[
CLV = \sum_{t=0}^{T} m \frac{r^t}{(1 + d)^t}
\]

where

- \( m \) = net cash flow per period (if active)
- \( r \) = retention rate
- \( d \) = discount rate
- \( T \) = horizon for calculation

Calculating \( E(\text{CLV}) \)

A more correct starting point:

\[
E(\text{CLV}) = \int_{0}^{\infty} E[v(t)]S(t)d(t)dt
\]

where

- \( E[v(t)] \) = expected value (or net cashflow) of the customer at time \( t \) (if active)
- \( S(t) \) = the probability that the customer has remained active to at least time \( t \)
- \( d(t) \) = discount factor that reflects the present value of money received at time \( t \)
Calculating $E(CLV)$

- Definitional; of little use by itself.

- We must operationalize $E[v(t)], S(t)$, and $d(t)$ in a specific business setting ... then solve the integral.

- Important distinctions:
  - Expected lifetime value of an as-yet-to-be-acquired customer
  - Expected lifetime value of a just-acquired customer
  - Expected residual lifetime value, $E(RLV)$, of an existing customer

Calculating $E(CLV)$

- The expected lifetime value of an as-yet-to-be-acquired customer is given by
  \[ E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt \]

- Standing at time $T$, the expected residual lifetime value of an existing customer is given by
  \[ E(RLV) = \int_{T}^{\infty} E[v(t)|t > T]S(t - T)d(t - T)dt \]
The Logic of Probability Models

- Many researchers attempt to describe/predict behavior using observed variables.

- However, they still use random components in recognition that not all factors are included in the model.

- We treat behavior as if it were “random” (probabilistic, stochastic).

- We propose a model of individual-level behavior which is “summed” across heterogeneous individuals to obtain a model of aggregate behavior.
Building a Probability Model

(i) Determine the marketing decision problem/information needed.

(ii) Identify the observable individual-level behavior of interest.
    - We denote this by $x$.

(iii) Select a probability distribution that characterizes this individual-level behavior.
    - This is denoted by $f(x|\theta)$.
    - We view the parameters of this distribution as individual-level latent traits.

(iv) Specify a distribution to characterize the distribution of the latent trait variable(s) across the population.
    - We denote this by $g(\theta)$.
    - This is often called the mixing distribution.

(v) Derive the corresponding aggregate or observed distribution for the behavior of interest:

$$f(x) = \int f(x|\theta)g(\theta) \, d\theta$$
Building a Probability Model

(vi) Estimate the parameters (of the mixing distribution) by fitting the aggregate distribution to the observed data.

(vii) Use the model to solve the marketing decision problem/provide the required information.

“Classes” of Models

- We focus on three fundamental behavioral processes:
  - Timing → “when”
  - Counting → “how many”
  - “Choice” → “whether/which”

- Our toolkit contains simple models for each behavioral process.

- More complex behavioral phenomena can be captured by combining models from each of these processes.
Individual-level Building Blocks

Count data arise from asking the question, “How many?”. As such, they are non-negative integers with no upper limit.

Let the random variable $X$ be a count variable:

$X$ is distributed Poisson with mean $\lambda$ if

$$P(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \ x = 0, 1, 2, \ldots$$

Individual-level Building Blocks

Timing (or duration) data are generated by answering “when” and “how long” questions, asked with regards to a specific event of interest.

The models we develop for timing data are also used to model other non-negative continuous quantities (e.g., transaction value).

Let the random variable $T$ be a timing variable:

$T$ is distributed exponential with rate parameter $\lambda$ if

$$F(t \mid \lambda) = P(T \leq t \mid \lambda) = 1 - e^{-\lambda t}, \ t > 0.$$
Individual-level Building Blocks

A Bernoulli trial is a probabilistic experiment in which there are two possible outcomes, 'success' (or '1') and 'failure' (or '0'), where $\theta$ is the probability of success.

Repeated Bernoulli trials lead to the geometric and binomial distributions.

Let the random variable $X$ be the number of independent and identically distributed Bernoulli trials required until the first success:

$X$ is a (shifted) geometric random variable, where

$$P(X = x | \theta) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, 3, \ldots$$

The (shifted) geometric distribution can be used to model either omitted-zero class count data or discrete-time timing data.
Individual-level Building Blocks

Let the random variable $X$ be the total number of successes occurring in $n$ independent and identically distributed Bernoulli trials:

$$X \text{ is distributed binomial with parameter } \theta, \text{ where}$$

$$P(X = x \mid n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \ x = 0, 1, 2, \ldots, n.$$  

We use the binomial distribution to model repeated choice data—answers to the question, “How many times did a particular outcome occur in a fixed number of events?”

Capturing Heterogeneity in Latent Traits

The gamma distribution:

$$g(\lambda \mid r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha \lambda}}{\Gamma(r)}, \ \lambda > 0$$

- $\Gamma(\cdot)$ is the gamma function
- $r$ is the “shape” parameter and $\alpha$ is the “scale” parameter
- The gamma distribution is a flexible (unimodal) distribution ... and is mathematically convenient.
Illustrative Gamma Density Functions

\[ g(\lambda) \]

\[ \lambda \]

\[ r = 0.5, \alpha = 1 \]
\[ r = 1, \alpha = 1 \]
\[ r = 2, \alpha = 1 \]
\[ r = 2, \alpha = 2 \]
\[ r = 2, \alpha = 4 \]

Capturing Heterogeneity in Latent Traits

The beta distribution:

\[ g(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1} (1 - \theta)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < \theta < 1. \]

- \( B(\alpha, \beta) \) is the beta function, which can be expressed in terms of gamma functions:

\[ B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \]

- The beta distribution is a flexible distribution ... and is mathematically convenient
The Negative Binomial Distribution (NBD)

- The individual-level behavior of interest can be characterized by the Poisson distribution when the mean $\lambda$ is known.

- We do not observe an individual’s $\lambda$ but assume it is distributed across the population according to a gamma distribution.

\[
P(X = x \mid r, \alpha) = \int_0^\infty P(X = x \mid \lambda) g(\lambda \mid r, \alpha) d\lambda
\]

\[
= \frac{\Gamma(r + x)}{\Gamma(r) x!} \left( \frac{\alpha}{\alpha + 1} \right)^r \left( \frac{1}{\alpha + 1} \right)^x.
\]
The Negative Binomial Distribution (NBD)

- Let the random variable \( X(t) \) be the count of events occurring in the interval \((0, t]\).
- If \( X(1) \) is distributed Poisson with mean \( \lambda \), then \( X(t) \) has a Poisson distribution with mean \( \lambda t \).
- Assuming \( \lambda \) is distributed across the population according to a gamma distribution,

\[
P(X(t) = x \mid r, \alpha) = \int_0^\infty P(X(t) = x \mid \lambda) g(\lambda \mid r, \alpha) d\lambda \\
= \frac{\Gamma(r + x)}{\Gamma(r)x!} \left( \frac{\alpha}{\alpha + t} \right)^r \left( \frac{t}{\alpha + t} \right)^x.
\]

The Exponential-Gamma Model (Pareto Distribution of the Second Kind)

- The individual-level behavior of interest can be characterized by the exponential distribution when the rate parameter \( \lambda \) is known.
- We do not observe an individual’s \( \lambda \) but assume it is distributed across the population according to a gamma distribution.

\[
F(t \mid r, \alpha) = \int_0^\infty F(t \mid \lambda) g(\lambda \mid r, \alpha) d\lambda \\
= 1 - \left( \frac{\alpha}{\alpha + t} \right)^r.
\]
The Shifted-Beta-Geometric Model

- The individual-level behavior of interest can be characterized by the (shifted) geometric distribution when the parameter $\theta$ is known.
- We do not observe an individual’s $\theta$ but assume it is distributed across the population according to a beta distribution.

\[
P(X = x \mid \alpha, \beta) = \int_0^1 P(X = x \mid \theta) g(\theta \mid \alpha, \beta) \, d\theta
= \frac{B(\alpha + 1, \beta + x - 1)}{B(\alpha, \beta)}.\]

The Beta-Binomial Distribution

- The individual-level behavior of interest can be characterized by the binomial distribution when the parameter $\theta$ is known.
- We do not observe an individual’s $\theta$ but assume it is distributed across the population according to a beta distribution.

\[
P(X = x \mid n, \alpha, \beta) = \int_0^1 P(X = x \mid n, \theta) g(\theta \mid \alpha, \beta) \, d\theta
= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)}.\]
### Summary of Probability Models

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>Individual-level</th>
<th>Heterogeneity</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counting</td>
<td>Poisson</td>
<td>gamma</td>
<td>NBD</td>
</tr>
<tr>
<td>Timing</td>
<td>exponential</td>
<td>gamma</td>
<td>EG (Pareto)</td>
</tr>
<tr>
<td>Discrete timing</td>
<td>shifted-geometric</td>
<td>beta</td>
<td>sBG</td>
</tr>
<tr>
<td>Choice</td>
<td>binomial</td>
<td>beta</td>
<td>BB</td>
</tr>
</tbody>
</table>

### Integrated Models

- **Counting + Timing**
  - catalog purchases (purchasing | “alive” & “death” process)
  - “stickiness” (# visits & duration/visit)

- **Counting + Counting**
  - purchase volume (# transactions & units/transaction)
  - page views/month (# visits & pages/visit)

- **Counting + Choice**
  - brand purchasing (category purchasing & brand choice)
  - “conversion” behavior (# visits & buy/not-buy)
A Template for Integrated Models

<table>
<thead>
<tr>
<th>Counting</th>
<th>Timing</th>
<th>Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>Timing</td>
<td></td>
</tr>
<tr>
<td>Choice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Integrated Models

- The observed behavior is a function of sub-processes that are typically unobserved:

\[ f(x \mid \theta_1, \theta_2) = g(f_1(x_1 \mid \theta_1), f_2(x_2 \mid \theta_2)) \]

- Solving the integral

\[ f(x) = \iint f(x \mid \theta_1, \theta_2) g_1(\theta_1) g_2(\theta_2) \, d\theta_1 \, d\theta_2 \]

often results in an intermediate result of the form

\[ = \text{constant} \times \int_0^1 t^a (1 - t)^b (u + vt)^{-c} \, dt \]
The “Trick” for Integrated Models

Using Euler’s integral representation of the Gaussian hypergeometric function, we can show that

\[
\int_0^1 t^a (1 - t)^b (u + vt)^{-c} \, dt = \begin{cases} 
B(a + 1, b + 1) u^{-c} \\
\times_2 F_1(c, a + 1; a + b + 2; -\frac{v}{u}), & |v| \leq u \\
B(a + 1, b + 1) (u + v)^{-c} \\
\times_2 F_1(c, b + 1; a + b + 2; \frac{v}{u + v}), & |v| \geq u
\end{cases}
\]

where \( _2F_1(\cdot) \) is the Gaussian hypergeometric function.

The Gaussian Hypergeometric Function

\[
_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a + j)\Gamma(b + j)}{\Gamma(c + j)} \frac{z^j}{j!}
\]

Easy to compute, albeit tedious, in Excel as

\[
_2F_1(a, b; c; z) = \sum_{j=0}^{\infty} u_j
\]

using the recursion

\[
\frac{u_j}{u_{j-1}} = \frac{(a + j - 1)(b + j - 1)}{(c + j - 1)j} z, \quad j = 1, 2, 3, \ldots
\]

where \( u_0 = 1 \).
Models for Contractual Settings

Classifying Customer Bases

<table>
<thead>
<tr>
<th>Continuous Opportunities for Transactions</th>
<th>Contractual Type of Relationship With Customers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery purchases</td>
<td>Credit card</td>
</tr>
<tr>
<td>Doctor visits</td>
<td>Student mealplan</td>
</tr>
<tr>
<td>Hotel stays</td>
<td>Mobile phone usage</td>
</tr>
<tr>
<td>Event attendance</td>
<td>Magazine subs</td>
</tr>
<tr>
<td>Prescription refills</td>
<td>Insurance policy</td>
</tr>
<tr>
<td>Charity fund drives</td>
<td>Health club m’ship</td>
</tr>
</tbody>
</table>

Noncontractual

41

42
It is increasingly apparent that the financial value of a firm depends on off-balance-sheet intangible assets. In this article, the authors focus on the most critical aspect of a firm: its customers. Specifically, they demonstrate how valuing customers makes it feasible to value firms, including high-growth firms with negative earnings. The authors define the value of a customer as the expected sum of discounted future earnings. They demonstrate their valuation method by using publicly available data for five firms. They find that a 1% improvement in retention, margin, or acquisition cost improves firm value by 5%, 1%, and .1%, respectively. They also find that a 1% improvement in retention has almost five times greater impact on firm value than a 1% change in discount rate or cost of capital. The results show that the linking of marketing concepts to shareholder value is both possible and insightful.

### Hypothetical Contractual Setting

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>End of year</td>
<td>10,000</td>
<td>16,334</td>
<td>20,701</td>
<td>23,965</td>
<td>26,569</td>
</tr>
</tbody>
</table>
Hypothetical Contractual Setting

Assume

- Each contract is annual, starting on January 1 and expiring at 11:59pm on December 31.
- An average net cashflow of $100/year.
- A 10% discount rate

What is the expected residual value of the customer base at December 31, 2007?

Hypothetical Contractual Setting

The aggregate retention rate is the fraction of 2006 customers who renewed their contracts at the beginning of 2007:

\[ \frac{26,569 - 10,000}{23,965} = 0.691 \]

Expected residual value of the customer base at December 31, 2007:

\[ 26,569 \times \sum_{t=1}^{\infty} \frac{100 \times 0.691^t}{(1 + 0.1)^{t-1}} = 4,945,049 \]
What’s wrong with this analysis?

<table>
<thead>
<tr>
<th># Customers</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
<tr>
<td>End of year</td>
<td>10,000</td>
<td>16,334</td>
<td>20,701</td>
<td>23,965</td>
<td>26,569</td>
</tr>
</tbody>
</table>
Hypothetical Contractual Setting

Number of customers who are still alive each year by year-of-acquisition cohort:

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10,000</td>
<td>6,334</td>
<td>4,367</td>
<td>3,264</td>
<td>2,604</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>6,334</td>
<td>4,367</td>
<td>3,264</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>6,334</td>
<td>4,367</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>6,334</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>6,334</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Annual Retention Rates by Cohort

<table>
<thead>
<tr>
<th></th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>--</td>
<td>0.633</td>
<td>0.689</td>
<td>0.747</td>
<td>0.798</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>0.633</td>
<td>0.689</td>
<td>0.747</td>
<td></td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>0.633</td>
<td>0.689</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>0.633</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

49

50
Renewal rates at regional magazines vary; generally 30% of subscribers renew at the end of their original subscription, but that figure jumps to 50% for second-time renewals and all the way to 75% for longtime readers.


New subscribers are actually more likely to cancel their subscriptions than older subscribes, and therefore, an increase in subscriber age tends to lead to reductions in subscriber churn.

Netflix FY:03 Form 10-K
At the cohort level, we (almost) always observe increasing retention rates.

Why Do Retention Rates Increase Over Time?
Why Do Retention Rates Increase Over Time?

Individual-level time dynamics:

- increasing loyalty as the customer gains more experience with the firm, and/or
- increasing switching costs with the passage of time.

vs.

A sorting effect in a heterogeneous population.

The Role of Heterogeneity

Suppose we track a cohort of 10,000 customers, comprising two underlying segments:

- Segment 1 comprises one-third of the customers, each with a time-invariant annual retention probability of 0.9.
- Segment 2 comprises two-thirds of the customers, each with a time-invariant annual retention probability of 0.5.
Vodafone Italia
Churn Clusters

<table>
<thead>
<tr>
<th>Cluster</th>
<th>P(churn)</th>
<th>% CB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low risk</td>
<td>0.06</td>
<td>70</td>
</tr>
<tr>
<td>Medium risk</td>
<td>0.35</td>
<td>20</td>
</tr>
<tr>
<td>High risk</td>
<td>0.65</td>
<td>10</td>
</tr>
</tbody>
</table>

Source: “Vodafone Achievement and Challenges in Italy” presentation (2003-09-12)

The Role of Heterogeneity

<table>
<thead>
<tr>
<th>Year</th>
<th>Seg 1</th>
<th>Seg 2</th>
<th>Total</th>
<th>Seg 1</th>
<th>Seg 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3,333</td>
<td>6,667</td>
<td>10,000</td>
<td>0.900</td>
<td>0.500</td>
<td>0.633</td>
</tr>
<tr>
<td>2</td>
<td>3,000</td>
<td>3,334</td>
<td>6,334</td>
<td>0.900</td>
<td>0.500</td>
<td>0.689</td>
</tr>
<tr>
<td>3</td>
<td>2,700</td>
<td>1,667</td>
<td>4,367</td>
<td>0.900</td>
<td>0.500</td>
<td>0.747</td>
</tr>
<tr>
<td>4</td>
<td>2,430</td>
<td>834</td>
<td>3,264</td>
<td>0.900</td>
<td>0.500</td>
<td>0.798</td>
</tr>
<tr>
<td>5</td>
<td>2,187</td>
<td>417</td>
<td>2,604</td>
<td>0.900</td>
<td>0.500</td>
<td>0.798</td>
</tr>
</tbody>
</table>
Implications for Valuing a Customer Base

- Not only do we need to project retention beyond the set of observed retention rates . . .
- We also need to recognize inter-cohort differences (at any point in time).
E(RLV) by Segment

- If this person belongs to segment 1:
  \[ E(RLV) = \sum_{t=1}^{\infty} 100 \times \frac{0.9^t}{(1 + 0.1)^{t-1}} \]
  \[ = \$495 \]

- If this person belongs to segment 2:
  \[ E(RLV) = \sum_{t=1}^{\infty} 100 \times \frac{0.5^t}{(1 + 0.1)^{t-1}} \]
  \[ = \$92 \]
\( E(\text{RLV}) \) of an Active 2003 Cohort Member

According to Bayes’ Theorem, the probability that this person belongs to segment 1 is

\[
P(\text{renewed contract four times} \mid \text{segment 1}) \times P(\text{segment 1})
\]

\[
= \frac{0.9^4 \times 0.333}{0.9^4 \times 0.333 + 0.5^4 \times 0.667}
\]

\[
= 0.84
\]

\[\Rightarrow E(\text{RLV}) = 0.84 \times \$495 + (1 - 0.84) \times \$92 = \$430\]

\[\text{P(Seg 1) as a Function of Customer “Age”}\]

<table>
<thead>
<tr>
<th>Year</th>
<th># Customers Still Alive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Seg 1</td>
</tr>
<tr>
<td>1</td>
<td>3,333</td>
</tr>
<tr>
<td>2</td>
<td>3,000</td>
</tr>
<tr>
<td>3</td>
<td>2,700</td>
</tr>
<tr>
<td>4</td>
<td>2,430</td>
</tr>
<tr>
<td>5</td>
<td>2,187</td>
</tr>
</tbody>
</table>
Valuing the Existing Customer Base

Recognizing the underlying segments:

<table>
<thead>
<tr>
<th>Cohort</th>
<th># Alive in 2007</th>
<th>P(Seg 1)</th>
<th>E(RLV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>10,000</td>
<td>0.333</td>
<td>$226</td>
</tr>
<tr>
<td>2006</td>
<td>6,334</td>
<td>0.474</td>
<td>$283</td>
</tr>
<tr>
<td>2005</td>
<td>4,367</td>
<td>0.618</td>
<td>$341</td>
</tr>
<tr>
<td>2004</td>
<td>3,264</td>
<td>0.745</td>
<td>$392</td>
</tr>
<tr>
<td>2003</td>
<td>2,604</td>
<td>0.840</td>
<td>$430</td>
</tr>
</tbody>
</table>

Total expected residual value = $7,940,992

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Total RV</th>
<th>Underestimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve</td>
<td>$4,945,049</td>
<td>38%</td>
</tr>
<tr>
<td>Segment (model)</td>
<td>$7,940,992</td>
<td></td>
</tr>
</tbody>
</table>
Exploring the Magnitude of the Error

- Systematically vary heterogeneity in retention rates
- First need to specify a flexible model of contract duration

A Discrete-Time Model for Contract Duration

i. An individual remains a customer of the firm with constant retention probability $1 - \theta$

- the duration of the customer’s relationship with the firm is characterized by the (shifted) geometric distribution:

$$S(t \mid \theta) = (1 - \theta)^t, \quad t = 1, 2, 3, \ldots$$

ii. Heterogeneity in $\theta$ is captured by a beta distribution with pdf

$$g(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1 - \theta)^{\beta-1}}{B(\alpha, \beta)}.$$
A Discrete-Time Model for Contract Duration

• The probability that a customer cancels their contract in period $t$

$$P(T = t \mid \alpha, \beta) = \int_{0}^{1} P(T = t \mid \theta) g(\theta \mid \alpha, \beta) d\theta$$

$$= \frac{B(\alpha + 1, \beta + t - 1)}{B(\alpha, \beta)}, \quad t = 1, 2, \ldots$$

• The aggregate survivor function is

$$S(t \mid \alpha, \beta) = \int_{0}^{1} S(t \mid \theta) g(\theta \mid \alpha, \beta) d\theta$$

$$= \frac{B(\alpha, \beta + t)}{B(\alpha, \beta)}, \quad t = 1, 2, \ldots$$

A Discrete-Time Model for Contract Duration

• The (aggregate) retention rate is given by

$$r_t = \frac{S(t)}{S(t-1)}$$

$$= \frac{\beta + t - 1}{\alpha + \beta + t - 1}.$$

• This is an increasing function of time, even though the underlying (unobserved) retention rates are constant at the individual-level.
Computing $E(\text{CLV})$

- Recall:
  \[ E(\text{CLV}) = \int_0^{\infty} E[v(t)]S(t)d(t)dt. \]

- In a contractual setting, assuming an individual’s mean value per unit of time is constant ($\bar{v}$),
  \[ E(\text{CLV}) = \bar{v} \int_0^{\infty} S(t)d(t)dt. \]

- Standing at time $s$, a customer’s expected residual lifetime value is
  \[ E(\text{RLV}) = \bar{v} \int_s^{\infty} S(t \mid t > s)d(t - s)dt \]
  discounted expected residual lifetime

Computing DERL

- Standing at the end of period $n$, just prior to the point in time at which the customer makes her contract renewal decision,
  \[ \text{DERL}(d \mid \theta, n - 1 \text{ renewals}) = \sum_{t=n}^{\infty} \frac{S(t \mid t > n - 1; \theta)}{(1 + d)^{t-n}} \]
  \[ = \frac{(1 - \theta)(1 + d)}{d + \theta}. \]

- But $\theta$ is unobserved ....
Computing DERL

By Bayes' Theorem, the posterior distribution of $\theta$ is

$$g(\theta \mid \alpha, \beta, n - 1 \text{ renewals}) = \frac{S(n - 1 \mid \theta)g(\theta \mid \alpha, \beta)}{S(n - 1 \mid \alpha, \beta)} = \frac{\theta^{\alpha-1}(1 - \theta)^{\beta + n - 2}}{B(\alpha, \beta + n - 1)}$$

$$\Rightarrow DERL(d \mid \alpha, \beta, n - 1 \text{ renewals})$$

$$= \int_0^1 \left\{ \text{DERL}(d \mid \theta, n - 1 \text{ renewals}) \times g(\theta \mid \alpha, \beta, n - 1 \text{ renewals}) \right\} d\theta$$

$$= \left( \frac{\beta + n - 1}{\alpha + \beta + n - 1} \right)_{2F_1}(1, \beta + n; \alpha + \beta + n; \frac{1}{1+d})$$

Computing DERL

Alternative derivation:

$$\text{DERL}(d \mid \alpha, \beta, n - 1 \text{ renewals})$$

$$= \sum_{t=n}^{\infty} \frac{S(t \mid t > n - 1; \alpha, \beta)}{(1 + d)^{t-n}}$$

$$= \sum_{t=n}^{\infty} \frac{S(t \mid \alpha, \beta)}{S(n - 1 \mid \alpha, \beta)} \left( \frac{1}{1 + d} \right)^{t-n}$$

$$= \sum_{t=n}^{\infty} \frac{B(\alpha, \beta + t)}{B(\alpha, \beta + n - 1)} \left( \frac{1}{1 + d} \right)^{t-n}$$

$$= \left( \frac{\beta + n - 1}{\alpha + \beta + n - 1} \right)_{2F_1}(1, \beta + n; \alpha + \beta + n; \frac{1}{1+d})$$
Impact of Heterogeneity on Error

- Assume the following arrival of new customers:

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

- Assume $\bar{v} = 1$ and a 10% discount rate.

- For given values of $\alpha$ and $\beta$, determine the error associated with computing the residual value of the existing customer base using the naïve approach (a constant aggregate retention rate) compared with the "correct" model-based approach.

---

Two Scenarios

Case 1

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(\theta)$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Case 2

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>0.0</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>2.5</th>
<th>3.0</th>
<th>3.5</th>
<th>4.0</th>
<th>4.5</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(\theta)$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

---

For the two scenarios:

<table>
<thead>
<tr>
<th>Case</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$E(\theta)$</th>
<th>$S(1)$</th>
<th>$S(2)$</th>
<th>$S(3)$</th>
<th>$S(4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.80</td>
<td>15.20</td>
<td>0.20</td>
<td>0.800</td>
<td>0.684</td>
<td>0.531</td>
<td>0.439</td>
</tr>
<tr>
<td>2</td>
<td>0.067</td>
<td>0.267</td>
<td>0.20</td>
<td>0.800</td>
<td>0.760</td>
<td>0.738</td>
<td>0.724</td>
</tr>
</tbody>
</table>
Computing DERL Using Excel

Recall our alternative derivation:

\[
DERL(d \mid \alpha, \beta, n - 1 \text{ renewals}) = \sum_{t=n}^{\infty} \frac{S(t \mid \alpha, \beta)}{S(n - 1 \mid \alpha, \beta)} \left( \frac{1}{1 + d} \right)^{t-n}
\]

We compute \( S(t) \) from the sBG retention rates:

\[
S(t) = \prod_{i=1}^{t} r_i \quad \text{where} \quad r_i = \frac{\beta + i - 1}{\alpha + \beta + i - 1}.
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>alpha</td>
<td>3.8</td>
<td></td>
<td>DERL</td>
<td>3.59</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>beta</td>
<td>15.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>t</td>
<td>S(t)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.8000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0.6480</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>0.5307</td>
<td></td>
<td>0.8190</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0.4391</td>
<td></td>
<td>0.6776</td>
<td>0.9091</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0.3665</td>
<td></td>
<td>0.5656</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>0.3085</td>
<td></td>
<td>0.4761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>0.2616</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>0.2234</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>0.1919</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>0.1659</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>196</td>
<td>6.14E-05</td>
<td>9.48E-05</td>
<td>0.5132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>197</td>
<td>6.03E-05</td>
<td>9.31E-05</td>
<td>0.5132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>198</td>
<td>5.93E-05</td>
<td>9.15E-05</td>
<td>0.5132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>199</td>
<td>5.82E-05</td>
<td>8.99E-05</td>
<td>0.5132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>200</td>
<td>5.72E-05</td>
<td>8.83E-05</td>
<td>0.5132</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of Active Customers: Case 1

<table>
<thead>
<tr>
<th>Year</th>
<th>n</th>
<th>E(RLV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>10,000</td>
<td>$3.84</td>
</tr>
<tr>
<td>2004</td>
<td>8,000</td>
<td>$3.72</td>
</tr>
<tr>
<td>2005</td>
<td>6,480</td>
<td>$3.59</td>
</tr>
<tr>
<td>2006</td>
<td>5,307</td>
<td>$3.45</td>
</tr>
<tr>
<td>2007</td>
<td>4,391</td>
<td>$3.31</td>
</tr>
</tbody>
</table>

Aggregate 06–07 retention rate = 24,178/29,787 = 0.81

Impact of Heterogeneity on Error: Case 1

Naïve valuation = $34,178 \times \sum_{t=1}^{\infty} \frac{0.81^t}{(1 + 0.1)^{t-1}} = $105,845

Correct valuation = 4,391 \times $3.84 + 5,307 \times $3.72
+ 6,480 \times $3.59 + 8,000 \times $3.45
+ 10,000 \times $3.31 = $120,543

Naïve underestimates correct by 12%.
Number of Active Customers: Case 2

<table>
<thead>
<tr>
<th>Year</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>n</th>
<th>E(RLV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>10,000</td>
<td>8,000</td>
<td>7,600</td>
<td>7,383</td>
<td>7,235</td>
<td>5</td>
<td>$10.19</td>
</tr>
<tr>
<td>2004</td>
<td>10,000</td>
<td>8,000</td>
<td>7,600</td>
<td>7,383</td>
<td>4</td>
<td>$10.06</td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>10,000</td>
<td>8,000</td>
<td>7,600</td>
<td>3</td>
<td>$9.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>10,000</td>
<td>2</td>
<td>$9.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>10,000</td>
<td>1</td>
<td>$7.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Aggregate 06–07 retention rate = 30,218/32,983 = 0.92

Impact of Heterogeneity on Error: Case 2

Naïve valuation = $220,488

Correct valuation = $375,437

Naïve underestimates correct by 41%.
Interpreting the Beta Distribution Parameters

mean $\mu = \frac{\alpha}{\alpha + \beta}$ and polarization index $\phi = \frac{1}{\alpha + \beta + 1}$

Shape of the Beta Distribution
Churn Rate as a Function of $\mu$ and $\phi$

For a fine grid of points in the $(\mu, \phi)$ space, we determine the corresponding values of $(\alpha, \beta)$ and compute the associated aggregate 2006/07 churn rate:

![Churn Rate Graph]

Naïve Model Valuation as a Function of $\mu$ and $\phi$

Expected residual lifetime value (in $000) of the customer base computed using the aggregate 2006/07 churn rate:

![Naïve Model Valuation Graph]
sBG Model Valuation as a Function of $\mu$ and $\phi$

Expected residual lifetime value (in $000) of the customer base computed using the sBG model:

% Underestimation as a Function of $\mu$ and $\phi$
Re-analysis Using \((r_1, r_2)\)

- \(\mu\) and \(\phi\) are not quantities that most managers or analysts think about; retention rates are easier to comprehend.

- Since the period 1 and 2 retention rates are, respectively,
  \[
  r_1 = \frac{\beta}{\alpha + \beta} \quad \text{and} \quad r_2 = \frac{\beta + 1}{\alpha + \beta + 1},
  \]
  it follows that
  \[
  \alpha = \frac{(1 - r_1)(1 - r_2)}{r_2 - r_1} \quad \text{and} \quad \beta = \frac{r_1(1 - r_2)}{r_2 - r_1}.
  \]

Shape of the Beta Distribution \((r_1, r_2)\)
Error as a Function of \((r_1, r_2)\)

For a fine grid of points in the \((r_1, r_2)\) space, we determine the corresponding values of \((\alpha, \beta)\) and compute % underestimation:

Expressions for DE(R)L

Different points in time at which a customer's discounted expected (residual) lifetime can be computed:

(i) (ii) (iii) (iv)
Expressions for DE(R)L

Case (i):

\[ \text{DEL}(d \mid \alpha, \beta) = \, {}_{2}F_{1} \left( 1, \beta; \alpha + \beta; \frac{1}{1+d} \right) \]

Case (ii):

\[ \text{DERL}(d \mid \alpha, \beta) = \frac{\beta}{(\alpha + \beta)(1 + d)} \, {}_{2}F_{1} \left( 1, \beta + 1; \alpha + \beta + 1; \frac{1}{1+d} \right) \]

Expressions for DE(R)L

Case (iii):

\[ \text{DERL}(d \mid \alpha, \beta, \text{active for } n \text{ periods}) = \frac{\beta + n - 1}{\alpha + \beta + n - 1} \, {}_{2}F_{1} \left( 1, \beta + n; \alpha + \beta + n; \frac{1}{1+d} \right) \]

Case (iv):

\[ \text{DERL}(d \mid \alpha, \beta, \text{ n contract renewals}) = \frac{\beta + n}{(\alpha + \beta + n)(1 + d)} \times {}_{2}F_{1} \left( 1, \beta + n + 1; \alpha + \beta + n + 1; \frac{1}{1+d} \right) \]
Further Reading


<http://brucehardie.com/papers/022/>

Fader, Peter S. and Bruce G. S. Hardie (2007), “Fitting the sBG Model to Multi-Cohort Data.”  
<http://brucehardie.com/notes/017/>

Classifying Customer Bases

<table>
<thead>
<tr>
<th>Continuous Opportunities for Transactions</th>
<th>Discrete Opportunities for Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery purchases</td>
<td>Credit card</td>
</tr>
<tr>
<td>Doctor visits</td>
<td>Student mealplan</td>
</tr>
<tr>
<td>Hotel stays</td>
<td>Mobile phone usage</td>
</tr>
<tr>
<td>Event attendance</td>
<td>Magazine subs</td>
</tr>
<tr>
<td>Prescription refills</td>
<td>Insurance policy</td>
</tr>
<tr>
<td>Charity fund drives</td>
<td>Health club m’ship</td>
</tr>
<tr>
<td>Noncontractual Type of Relationship With Customers</td>
<td>Contractual Type of Relationship With Customers</td>
</tr>
</tbody>
</table>
**Contract Duration in Continuous-Time**

i. The duration of an individual customer’s relationship with the firm is characterized by the exponential distribution with pdf and survivor function,

\[
f(t | \lambda) = \lambda e^{-\lambda t}
\]

\[
S(t | \lambda) = e^{-\lambda t}
\]

ii. Heterogeneity in \( \lambda \) follows a gamma distribution with pdf

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

This gives us the exponential-gamma model with pdf and survivor function

\[
f(t | r, \alpha) = \int_0^\infty f(t | \lambda) g(\lambda | r, \alpha) \, d\lambda
\]

\[
= \frac{r}{\alpha} \left( \frac{\alpha}{\alpha + t} \right)^{r+1}
\]

\[
S(t | r, \alpha) = \int_0^\infty S(t | \lambda) g(\lambda | r, \alpha) \, d\lambda
\]

\[
= \left( \frac{\alpha}{\alpha + t} \right)^r
\]
The Hazard Function

The hazard function, $h(t)$, is defined by

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t}$$

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

and represents the instantaneous rate of “failure” at time $t$ conditional upon “survival” to $t$.

The probability of “failing” in the next small interval of time, given “survival” to time $t$, is

$$P(t < T \leq t + \Delta t | T > t) \approx h(t) \times \Delta t$$

The Hazard Function

- For the exponential distribution, $h(t | \lambda) = \lambda$

- For the EG model, $h(t | r, \alpha) = \frac{r}{\alpha + t}$

- In applying the EG model, we are assuming that the increasing retention rates observed in the aggregate data are simply due to heterogeneity and not because of underlying time dynamics at the level of the individual customer.
Computing DERL

• Standing at time $s$,

$$DERL = \int_s^\infty S(t \mid t > s) d(t - s) dt$$

• For exponential lifetimes with continuous compounding at rate of interest $\delta$,

$$DERL(\delta \mid \lambda, \text{tenure of at least } s) = \int_s^\infty e^{-\lambda(t-s)} e^{-\delta(t-s)} dt$$

$$= \frac{1}{\lambda + \delta}$$

• But $\lambda$ is unobserved ....

Computing DERL

By Bayes’ Theorem, the posterior distribution of $\lambda$ for an individual with tenure of at least $s$,

$$g(\lambda \mid r, \alpha, \text{tenure of at least } s) = \frac{S(s \mid \lambda) g(\lambda \mid r, \alpha)}{S(s \mid r, \alpha)}$$

$$= \frac{(\alpha + s)^r \lambda^{r-1} e^{-\lambda(\alpha+s)}}{\Gamma(r)}$$
Computing DERL

It follows that

\[ DERL(\delta \mid r, \alpha, \text{tenure of at least } s) = \int_0^\infty \left\{ DERL(\delta \mid \lambda, \text{tenure of at least } s) \times g(\lambda \mid r, \alpha, \text{tenure of at least } s) \right\} d\lambda = (\alpha + s)^r \delta^{r-1} \Psi(r, r; (\alpha + s)\delta) \]

where \( \Psi(\cdot) \) is the confluent hypergeometric function of the second kind.

Models for Noncontractual Settings
Classifying Customer Bases

<table>
<thead>
<tr>
<th>Continuous</th>
<th>Discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opportunities for Transactions</td>
<td>Type of Relationship With Customers</td>
</tr>
<tr>
<td>Grocery purchases</td>
<td>Noncontractual</td>
</tr>
<tr>
<td>Doctor visits</td>
<td>Event attendance</td>
</tr>
<tr>
<td>Hotel stays</td>
<td>Prescription refills</td>
</tr>
<tr>
<td>Credit card</td>
<td>Charity fund drives</td>
</tr>
<tr>
<td>Student mealplan</td>
<td>Contractual</td>
</tr>
<tr>
<td>Mobile phone usage</td>
<td>Magazine subs</td>
</tr>
<tr>
<td></td>
<td>Insurance policy</td>
</tr>
<tr>
<td></td>
<td>Health club m’ship</td>
</tr>
</tbody>
</table>

Setting

- A major public radio station located in the Midwestern United States.
- Supported in large part by listener contributions.
- Initial focus on 1995 cohort, ignoring donation amount:
  - 11,104 people first-time supporters.
  - This cohort makes a total of 24,615 repeat donations (transactions) over the next 6 years.
  - What level of support (# transactions) can we expect from this cohort in the future?
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100001</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100002</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100003</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100004</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100005</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100006</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100007</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100008</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100009</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100010</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Modelling the Transaction Stream

- Each year a listener decides whether or not to support the station by tossing a coin.
- \( P(\text{heads}) \) varies across listeners.
Modelling the Transaction Stream

- Let random variable $X(n)$ denote the # transactions across $n$ consecutive transaction opportunities.

- The customer buys at any given transaction opportunity with probability $p$:

$$P(X(n) = x \mid p) = \binom{n}{x} p^x (1 - p)^{n-x}.$$ 

- Purchase probabilities ($p$) are distributed across the population according to a beta distribution:

$$g(p \mid \alpha, \beta) = \frac{p^{\alpha-1} (1 - p)^{\beta-1}}{B(\alpha, \beta)}.$$ 

The distribution of transactions for a randomly-chosen individual is given by

$$P(X(n) = x \mid \alpha, \beta) = \int_0^1 P(X(n) = x \mid p) g(p \mid \alpha, \beta) \, dp$$

$$= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)},$$

which is the beta-binomial (BB) distribution.
Fit of the BB Model

Tracking Cumulative Repeat Transactions
Repeat Transactions in 2002 – 2006

Conditional Expectations
Conditional Expectations

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cust. A</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cust. B</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Let $X(n, n + n^*)$ denote the number of transactions in the interval $(n, n + n^*)$.

According to the BB,

Cust. A: $E[X(6, 11) \mid x = 2, n = 6] = 1.70$

Cust. B: $E[X(6, 11) \mid x = 2, n = 6] = ?$

Tracking Cumulative Repeat Transactions

![Graph of cumulative repeat transactions](image-url)
Towards a More Realistic Model

Modelling the Transaction Stream

Transaction Process:
- While "alive", a customer makes a purchase at any given transaction opportunity as-if randomly
- Transaction probabilities vary across customers

Dropout Process:
- Each customer has an unobserved “lifetime”
- Dropout rates vary across customers
Model Development

A customer’s relationship with a firm has two phases: he is “alive” (A) for some period of time, then “dies” (D).

- While “alive”, the customer buys at any given transaction opportunity with probability $p$:
  \[ P(Y_t = 1 \mid p, \text{alive at } t) = p \]

- A “living” customer becomes “dies” at the beginning of a transaction opportunity with probability $\theta$
  \[ \Rightarrow P(\text{alive at } t \mid \theta) = P(AA\ldots A \mid \theta) = (1 - \theta)^t \]

Model Development

Consider the following transaction pattern:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

- The customer must have been alive in 1999 (and therefore in 1996–1998)
- Three scenarios give rise to no purchasing in 2000 and 2001

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>
Model Development

We compute the probability of the purchase string conditional on each scenario and multiply it by the probability of that scenario:

\[ f(100100 \mid p, \theta) = p(1 - p)(1 - p)p \left(1 - \theta\right)^4 \theta \]

\[ + p(1 - p)(1 - p)p(1 - p) \left(1 - \theta\right)^5 \theta \]

\[ + p(1 - p)(1 - p)p(1 - p)(1 - p) \left(1 - \theta\right)^6 \]

Model Development

- Bernoulli purchasing while alive \( \Rightarrow \) the order of a given number of transactions (prior to the last observed transaction) doesn’t matter. For example,

\[ f(100100 \mid p, \theta) = f(001100 \mid p, \theta) = f(010100 \mid p, \theta) \]

- Recency (time of last transaction, \( t_x \)) and frequency (number of transactions, \( x = \sum_{t=1}^{n} y_t \)) are sufficient summary statistics

\[ \Rightarrow \] we do not need the complete binary string representation of a customer’s transaction history
### Model Development

For a customer with purchase history \((x, t_x, n)\),

\[
L(p, \theta \mid x, t_x, n) = p^x (1 - p)^{n-x} (1 - \theta)^n + \sum_{i=0}^{n-t_x-1} p^x (1 - p)^{t_x-x+i} \theta (1 - \theta)^{t_x+i}
\]

We assume that heterogeneity in \(p\) and \(\theta\) across customers is captured by beta distributions:

\[
g(p \mid \alpha, \beta) = \frac{p^{\alpha-1} (1 - p)^{\beta-1}}{B(\alpha, \beta)}
\]

\[
g(\theta \mid \gamma, \delta) = \frac{\theta^{\gamma-1} (1 - \theta)^{\delta-1}}{B(\gamma, \delta)}
\]
Model Development

Removing the conditioning on the latent traits $p$ and $\theta$,

$$L(\alpha, \beta, \gamma, \delta | x, t_x, n) = \int_0^1 \int_0^1 L(p, \theta | x, t_x, n) g(p | \alpha, \beta) g(\theta | \gamma, \delta) dp \, d\theta$$

$$= \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \frac{B(\gamma + t_x - x + i)}{B(\gamma, \delta)} B(\delta + t_x + i)$$

... which is (relatively) easy to code-up in Excel.
<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>N</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Parameter Estimates (1995 Cohort)

<table>
<thead>
<tr>
<th></th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>0.487</td>
<td>0.826</td>
<td>0.729</td>
<td>0.783</td>
<td>−35,516.1</td>
</tr>
<tr>
<td>BG/BB</td>
<td>1.204</td>
<td>0.750</td>
<td>0.657</td>
<td>2.783</td>
<td>−33,225.6</td>
</tr>
</tbody>
</table>
Fit of the BG/BB Model

Tracking Cumulative Repeat Transactions
Tracking Annual Repeat Transactions

Repeat Transactions in 2002 – 2006
**Key Results**

For an individual with observed behavior \((x, t_x, n)\):

- \(P(\text{alive in period } n + 1 | x, t_x, n)\)
  
  The probability he will be “alive” in the next period.

- \(P(X(n, n + n^*) = x^* | x, t_x, n)\)
  
  The probability he will make \(x^*\) transactions across the next \(n^*\) transaction opportunities.

- \(E[X(n, n + n^*) | x, t_x, n]\)
  
  The expected number of transactions across the next \(n^*\) transaction opportunities.

- \(DERT(d | x, t_x, n)\)
  
  The discounted expected residual transactions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100001</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100002</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100003</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100004</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100005</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100006</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100007</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100008</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100009</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>100010</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>111102</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>111103</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>111104</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Expected # Transactions in 2002–2006 as a Function of Recency and Frequency

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Conditional Expectations by Frequency

![Graph showing the comparison between actual and model conditional expectations by frequency](image-url)
Conditional Expectations by Recency

![Graph showing conditional expectations by recency](image)

**Expected # Transactions in 2002–2006 as a Function of Recency and Frequency**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.09</td>
<td>0.31</td>
<td>0.59</td>
<td>0.84</td>
<td>1.02</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
<td>0.54</td>
<td>1.06</td>
<td>1.44</td>
<td>1.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.22</td>
<td>1.03</td>
<td>1.80</td>
<td>2.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.58</td>
<td>2.03</td>
<td>2.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>1.81</td>
<td>3.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.75</td>
<td></td>
</tr>
</tbody>
</table>
Expected # Transactions in 2002 - 2006 as a Function of Recency and Frequency

Posterior Mean of $p$ as a Function of Recency and Frequency

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.66</td>
<td>0.44</td>
<td>0.34</td>
<td>0.30</td>
<td>0.28</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.75</td>
<td>0.54</td>
<td>0.44</td>
<td>0.41</td>
<td>0.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.80</td>
<td>0.61</td>
<td>0.54</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.82</td>
<td>0.68</td>
<td>0.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.83</td>
<td>0.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.91</td>
</tr>
</tbody>
</table>
P(alive in 2002) as a Function of Recency and Frequency

<table>
<thead>
<tr>
<th># Rpt Trans. (1996 - 2001)</th>
<th>Year of Last Transaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.11</td>
</tr>
<tr>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>2</td>
<td>0.07</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Moving Beyond a Single Cohort

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>11,104</td>
</tr>
<tr>
<td>1996</td>
<td>10,057</td>
</tr>
<tr>
<td>1997</td>
<td>9,043</td>
</tr>
<tr>
<td>1998</td>
<td>8,175</td>
</tr>
<tr>
<td>1999</td>
<td>8,977</td>
</tr>
<tr>
<td>2000</td>
<td>9,491</td>
</tr>
</tbody>
</table>

- Pooled calibration using the repeat transaction data for these 56,847 people across 1996 – 2001
Parameter Estimates (Pooled)

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\delta$</th>
<th>LL</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>0.501</td>
<td>0.753</td>
<td></td>
<td></td>
<td>$-115,615.0$</td>
</tr>
<tr>
<td>BG/BB</td>
<td>1.188</td>
<td>0.749</td>
<td>0.626</td>
<td>2.331</td>
<td>$-110,521.0$</td>
</tr>
</tbody>
</table>

Fit of the BG/BB Model
Computing $E(\text{CLV})$

- Recall:
  \[ E(\text{CLV}) = \int_0^\infty E[v(t)]S(t)d(t)dt. \]

- Assuming that an individual’s spend per transaction is constant, $v(t) = \text{net cashflow/transaction} \times t(t)$ (where $t(t)$ is the transaction rate at $t$) and
  \[ E(\text{CLV}) = E(\text{net cashflow/transaction}) \times \int_0^\infty E[t(t)]S(t)d(t)dt. \]

Computing $E(\text{RLV})$

- Standing at time $T$,
  \[ E(\text{RLV}) = E(\text{net cashflow/transaction}) \times \int_T^\infty E[t(t)]S(t \mid t > T)d(t - T)d(t)dt. \]

- The quantity $DERT$, discounted expected residual transactions, is the present value of the expected future transaction stream for a customer with a given purchase history.
Computing DERT

• For a customer with purchase history \((x, t_x, n)\),

\[
DERT(d \mid p, \theta, \text{alive at } n) = \sum_{t=n+1}^{\infty} \frac{P(Y_t = 1 \mid p, \text{alive at } t)P(\text{alive at } t \mid t > n, \theta)}{(1 + d)^{t-n}}
\]

\[
= \frac{p(1 - \theta)}{d + \theta}
\]

• However,
  
  - \(p\) and \(\theta\) are unobserved
  - We do not know whether the customer is alive at \(n\)

\[
DERT(d \mid \alpha, \beta, \gamma, \delta, x, t_x, n)
\]

\[
= \int_0^1 \int_0^1 \left\{DERT(d \mid p, \theta, \text{alive at } n) \times P(\text{alive at } n \mid p, \theta, x, t_x, n) \times g(p, \theta \mid \alpha, \beta, \gamma, \delta, x, t_x, n)\right\} dp \, d\theta
\]

\[
= \frac{B(\alpha + x + 1, \beta + n - x) \cdot B(y, \delta + n + 1)}{B(\alpha, \beta) \cdot B(y, \delta)(1 + d)}
\]

\[
\times \frac{2F_1(1, \delta + n + 1; y + \delta + n + 1; \frac{1}{1+d})}{L(\alpha, \beta, y, \delta \mid x, t_x, n)}
\]
DERT as a Function of Recency and Frequency ($d = 0.10$)

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.13</td>
</tr>
<tr>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>0.21</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

“Discrete-Time” Transaction Data

A transaction opportunity is

- a well-defined point in time at which a transaction either occurs or does not occur, or
- a well-defined time interval during which a (single) transaction either occurs or does not occur.

<table>
<thead>
<tr>
<th>“necessarily discrete”</th>
<th>attendance at sports events</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>attendance at annual arts festival</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“generally discrete”</th>
<th>charity donations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>blood donations</td>
</tr>
</tbody>
</table>

| discretized by recording process | cruise ship vacations |
Further Reading


From Discrete to Continuous Time

- Suppose we have a year of data from Amazon.
- Should we define
  - 12 monthly transaction opportunities?
  - 52 weekly transaction opportunities?
  - 365 daily transaction opportunities?
From Discrete to Continuous Time

As the number of divisions of a given time period $\to \infty$

- binomial $\to$ Poisson
- beta-binomial $\to$ NBD
- geometric $\to$ exponential
- beta-geometric $\to$ exponential-gamma
  (Pareto of the second kind)

Classifying Customer Bases

<table>
<thead>
<tr>
<th>Continuous Opportunities for Transactions</th>
<th>Discrete Opportunities for Transactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery purchases</td>
<td>Credit card</td>
</tr>
<tr>
<td>Doctor visits</td>
<td>Student mealplan</td>
</tr>
<tr>
<td>Hotel stays</td>
<td>Mobile phone usage</td>
</tr>
<tr>
<td>Event attendance</td>
<td>Magazine subs</td>
</tr>
<tr>
<td>Prescription refills</td>
<td>Insurance policy</td>
</tr>
<tr>
<td>Charity fund drives</td>
<td>Health club m'ship</td>
</tr>
<tr>
<td>Noncontractual Type of Relationship With Customers</td>
<td>Contractual Type of Relationship With Customers</td>
</tr>
</tbody>
</table>
**Setting**

- New customers at CDNOW, 1/97–3/97
- Systematic sample (1/10) drawn from panel of 23,570 new customers
- 39-week calibration period
- 39-week forecasting (holdout) period
- Initial focus on transactions

**Purchase Histories**

<table>
<thead>
<tr>
<th>ID</th>
<th>Week 0</th>
<th>Week 39</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2356</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2357</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Modelling Objective

Given this customer database, we wish to determine the level of transactions that should be expected in next period (e.g., 39 weeks) by those on the customer list, both individually and collectively.
Modelling the Transaction Stream

Transaction Process:

- While active, a customer purchases “randomly” around his mean transaction rate
- Transaction rates vary across customers

Dropout Process:

- Each customer has an unobserved “lifetime”
- Dropout rates vary across customers

The Pareto/NBD Model
(Schmittlein, Morrison and Colombo 1987)

Transaction Process:

- While active, # transactions made by a customer follows a Poisson process with transaction rate $\lambda$.
- Heterogeneity in transaction rates across customers is distributed $\text{gamma}(r, \alpha)$.

Dropout Process:

- Each customer has an unobserved “lifetime” of length $\omega$, which is distributed exponential with dropout rate $\mu$.
- Heterogeneity in dropout rates across customers is distributed $\text{gamma}(s, \beta)$. 
Summarizing Purchase Histories

- Given the model assumptions, we do not require information on when each of the \( x \) transactions occurred.

- The only customer-level information required by this model is \textit{recency} and \textit{frequency}.

- The notation used to represent this information is \((x, t_x, T)\), where \( x \) is the number of transactions observed in the time interval \((0, T]\) and \( t_x \) \((0 < t_x \leq T)\) is the time of the last transaction.

Purchase Histories
Pareto/NBD Likelihood Function

\[
L(r, \alpha, s, \beta \mid x, t_x, T) = \frac{\Gamma(r + x) \alpha^r \beta^s}{\Gamma(r)} \left\{ \left( \frac{s}{r + s + x} \right) \frac{2F_1 \left( r + s + x, s + 1; r + s + x + 1; \frac{\alpha - \beta}{\alpha + t_x} \right)}{(\alpha + t_x)^{r+s+x}} \right. \\
+ \left. \left( \frac{r + x}{r + s + x} \right) \frac{2F_1 \left( r + s + x, s; r + s + x + 1; \frac{\alpha - \beta}{\alpha + T} \right)}{(\alpha + T)^{r+s+x}} \right \}, \text{ if } \alpha \geq \beta
\]

\[
L(r, \alpha, s, \beta \mid x, t_x, T) = \frac{\Gamma(r + x) \alpha^r \beta^s}{\Gamma(r)} \left\{ \left( \frac{s}{r + s + x} \right) \frac{2F_1 \left( r + s + x, r + x; r + s + x + 1; \frac{\beta - \alpha}{\beta + t_x} \right)}{(\beta + t_x)^{r+s+x}} \right. \\
+ \left. \left( \frac{r + x}{r + s + x} \right) \frac{2F_1 \left( r + s + x, r + x + 1; r + s + x + 1; \frac{\beta - \alpha}{\beta + T} \right)}{(\beta + T)^{r+s+x}} \right \}, \text{ if } \alpha \leq \beta
\]
Key Results

\( E[X(t)] \)

The expected number of transactions in the time interval \((0, t]\).

\( P(\text{alive} \mid x, t_x, T) \)

The probability that an individual with observed behavior \((x, t_x, T)\) is still “active” at time \(T\).

\( E[X(T, T + t) \mid x, t_x, T] \)

The expected number of transactions in the future period \((T, T + t]\) for an individual with observed behavior \((x, t_x, T)\).

Frequency of Repeat Transactions

![Frequency of Repeat Transactions Graph](image-url)
Tracking Cumulative Repeat Transactions

[Graph showing cumulative repeat transactions over weeks with lines for Actual and Pareto/NBD models.]

Week

Cumulative # Repeat Transactions

Tracking Weekly Repeat Transactions

[Graph showing weekly repeat transactions over weeks with lines for Actual and Pareto/NBD models.]

Week
Conditional Expectations

Computing DERT

- For Poisson purchasing and exponential lifetimes with continuous compounding at rate of interest $\delta$,

$$DERT(\delta \mid \lambda, \mu, \text{alive at } T) = \int_{T}^{\infty} \lambda \left(\frac{e^{-\mu t}}{e^{-\mu T}}\right) e^{-\delta(t-T)} dt$$

$$= \int_{0}^{\infty} \lambda e^{-\mu s} e^{-\delta s} ds$$

$$= \frac{\lambda}{\mu + \delta}$$

- However,
  - $\lambda$ and $\mu$ are unobserved
  - We do not know whether the customer is alive at $T$
Computing DERT

\[ \begin{align*}
DERT(\delta \mid r, \alpha, s, \beta, x, t_x, T) &= \int_0^\infty \int_0^\infty \left\{ DERT(\delta \mid \lambda, \mu, \text{alive at } T) \times P(\text{alive at } T \mid \lambda, \mu, x, t_x, T) \times g(\lambda, \mu \mid r, \alpha, s, \beta, x, t_x, T) \right\} d\lambda d\mu \\
&= \frac{\alpha^r \beta^s \delta^{s-1} \Gamma(r + x + 1) \Psi(s, s; \delta(\beta + T))}{\Gamma(r)(\alpha + T)^{r+x+1} L(r, \alpha, s, \beta \mid x, t_x, T)}
\end{align*} \]

where \( \Psi(\cdot) \) is the confluent hypergeometric function of the second kind.

Continuous Compounding

- An annual discount rate of \((100 \times d)\% \) is equivalent to a continuously compounded rate of \( \delta = \ln(1 + d) \).

- If the data are recorded in time units such that there are \( k \) periods per year (\( k = 52 \) if the data are recorded in weekly units of time) then the relevant continuously compounded rate is \( \delta = \ln(1 + d) / k \).
DERT by Recency and Frequency

Iso-Value Representation of DERT
The “Increasing Frequency” Paradox

Cust. A

Cust. B

Week 0

Week 78

DERT

<table>
<thead>
<tr>
<th>Cust. A</th>
<th>4.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cust. B</td>
<td>1.9</td>
</tr>
</tbody>
</table>

Key Contribution

- We are able to generate forward-looking estimates of DERT as a function of recency and frequency in a noncontractual setting:

\[
DERT = f(R, F)
\]

- Adding a sub-model for spend per transaction enables us to generate forward-looking estimates of an individual’s expected residual revenue stream conditional on his observed behavior (RFM):

\[
E(RLV) = f(R, F, M) = DERT \times g(F, M)
\]
Modelling the Spend Process

- The dollar value of a customer's given transaction varies randomly around his average transaction value.
- Average transaction values vary across customers but do not vary over time for any given individual.
- The distribution of average transaction values across customers is independent of the transaction process.

For a customer with $x$ transactions, let $z_1, z_2, \ldots, z_x$ denote the dollar value of each transaction.

The customer's average observed transaction value

$$m_x = \frac{\sum_{i=1}^{x} z_i}{x}$$

is an imperfect estimate of his (unobserved) mean transaction value $E(M)$.

Our goal is to make inferences about $E(M)$ given $m_x$, which we denote as $E(M|m_x, x)$. 
Summary of Average Transaction Value

946 individuals (from the 1/10th sample of the cohort) make at least one repeat purchase in weeks 1–39

<table>
<thead>
<tr>
<th></th>
<th>$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>2.99</td>
</tr>
<tr>
<td>25th percentile</td>
<td>15.75</td>
</tr>
<tr>
<td>Median</td>
<td>27.50</td>
</tr>
<tr>
<td>75th percentile</td>
<td>41.80</td>
</tr>
<tr>
<td>Maximum</td>
<td>299.63</td>
</tr>
<tr>
<td>Mean</td>
<td>35.08</td>
</tr>
<tr>
<td>Std. deviation</td>
<td>30.28</td>
</tr>
<tr>
<td>Mode</td>
<td>14.96</td>
</tr>
</tbody>
</table>

Modelling the Spend Process

- The dollar value of a customer's given transaction is distributed gamma with shape parameter $p$ and scale parameter $\nu$
- Heterogeneity in $\nu$ across customers follows a gamma distribution with shape parameter $q$ and scale parameter $\gamma$
Modelling the Spend Process

Marginal distribution for $m_x$:

$$f(m_x|p, q, \gamma, x) = \frac{\Gamma(px + q)}{\Gamma(px)\Gamma(q)} \frac{\gamma^q m_x^{px-1} x^{px}}{(y + m_x x)^{px+q}}$$

Expected average transaction value for a customer with an average spend of $m_x$ across $x$ transactions:

$$E(M|p, q, \gamma, m_x, x) = \left(\frac{q - 1}{px + q - 1}\right) \frac{\gamma p}{q - 1} + \left(\frac{px}{px + q - 1}\right) m_x$$

Distribution of Average Transaction Value

![Distribution Graph](image-url)
Computing Expected Residual Lifetime Value

We are interested in computing the present value of an individual’s expected residual margin stream conditional on his observed behavior (RFM)

\[
E(RLV) = \text{margin} \times \text{revenue/transaction} \times DERT
\]

\[
= \text{margin} \times E(M|p, q, y, m_x, x) \\
\times DERT(\delta | r, \alpha, s, \beta, x, t_x, T)
\]

Estimates of \(E(RLV)\)

\[
m_x = \$20 \quad \quad m_x = \$50
\]

(Margin = 30%, 15% discount rate)
Closing the Loop

Combine the model-driven RFM-CLV relationship with the actual RFM patterns seen in our dataset to get a sense of the overall value of this cohort of customers:

- Compute each customer’s expected residual lifetime value (conditional on their past behavior).
- Segment the customer base on the basis of RFM terciles (excluding non-repeaters).
- Compute average $E(RLV)$ and total residual value for each segment.

Distribution of Repeat Customers

(12,054 customers make no repeat purchases)
### Average $E(RLV)$ by RFM Segment

<table>
<thead>
<tr>
<th>Recency</th>
<th>Frequency</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M=0</td>
<td>0</td>
<td></td>
<td>$4.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=1</td>
<td>1</td>
<td>$6.39</td>
<td>$20.52</td>
<td>$25.26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$7.30</td>
<td>$31.27</td>
<td>$41.55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$4.54</td>
<td>$48.74</td>
<td>$109.32</td>
<td></td>
</tr>
<tr>
<td>M=2</td>
<td>1</td>
<td>$9.02</td>
<td>$28.90</td>
<td>$34.43</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$9.92</td>
<td>$48.67</td>
<td>$62.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$5.23</td>
<td>$77.85</td>
<td>$208.85</td>
<td></td>
</tr>
<tr>
<td>M=3</td>
<td>1</td>
<td>$16.65</td>
<td>$53.20</td>
<td>$65.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$22.15</td>
<td>$91.09</td>
<td>$120.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$10.28</td>
<td>$140.26</td>
<td>$434.95</td>
<td></td>
</tr>
</tbody>
</table>

### Total Residual Value by RFM Segment

<table>
<thead>
<tr>
<th>Recency</th>
<th>Frequency</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M=0</td>
<td>0</td>
<td>$53,000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M=1</td>
<td>1</td>
<td>$7,700</td>
<td>$9,900</td>
<td>$1,800</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$2,800</td>
<td>$15,300</td>
<td>$17,400</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$300</td>
<td>$12,500</td>
<td>$52,900</td>
<td></td>
</tr>
<tr>
<td>M=2</td>
<td>1</td>
<td>$5,900</td>
<td>$7,600</td>
<td>$2,300</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$3,600</td>
<td>$26,500</td>
<td>$25,800</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$500</td>
<td>$37,200</td>
<td>$203,000</td>
<td></td>
</tr>
<tr>
<td>M=3</td>
<td>1</td>
<td>$11,300</td>
<td>$19,700</td>
<td>$3,700</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$7,300</td>
<td>$45,900</td>
<td>$47,900</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$1,000</td>
<td>$62,700</td>
<td>$414,900</td>
<td></td>
</tr>
</tbody>
</table>
An Alternative to the Pareto/NBD Model

- Estimation of model parameters can be a barrier to Pareto/NBD model implementation
- Recall the dropout process story:
  - Each customer has an unobserved “lifetime”
  - Dropout rates vary across customers
- Let us consider an alternative story:
  - After any transaction, a customer tosses a coin
    heads → remain active
tails → become inactive
  - \( P(\text{tails}) \) varies across customers

The BG/NBD Model
(Fader, Hardie and Lee 2005c)

Purchase Process:

- While active, # transactions made by a customer follows a Poisson process with transaction rate \( \lambda \).
- Heterogeneity in transaction rates across customers is distributed gamma(\( r, \alpha \)).

Dropout Process:

- After any transaction, a customer becomes inactive with probability \( p \).
- Heterogeneity in dropout probabilities across customers is distributed beta(\( a, b \)).
BG/NBD Likelihood Function

We can express the model likelihood function as:

\[
L(r, \alpha, a, b \mid x, t_x, T) = A_1 \cdot A_2 \cdot (A_3 + \delta_{x>0} A_4)
\]

where

\[
A_1 = \frac{\Gamma(r + x) \alpha^r}{\Gamma(r)}
\]

\[
A_2 = \frac{\Gamma(a + b) \Gamma(b + x)}{\Gamma(b) \Gamma(a + b + x)}
\]

\[
A_3 = \left( \frac{1}{\alpha + T} \right)^{r + x}
\]

\[
A_4 = \left( \frac{a}{b + x - 1} \right) \left( \frac{1}{\alpha + t_x} \right)^{r + x}
\]
Model Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>BG/NBD</th>
<th>Pareto/NBD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>0.243</td>
<td>0.553</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>4.414</td>
<td>10.578</td>
</tr>
<tr>
<td>$a$</td>
<td>0.793</td>
<td></td>
</tr>
<tr>
<td>$b$</td>
<td>2.426</td>
<td></td>
</tr>
<tr>
<td>$s$</td>
<td></td>
<td>0.606</td>
</tr>
<tr>
<td>$\beta$</td>
<td></td>
<td>11.669</td>
</tr>
<tr>
<td>$LL$</td>
<td>$-9582.4$</td>
<td>$-9595.0$</td>
</tr>
</tbody>
</table>

Frequency of Repeat Transactions
Tracking Cumulative Repeat Transactions

Tracking Weekly Repeat Transactions
Conditional Expectations

Computing DERT for the BG/NBD

- It is very difficult to solve

\[ DERT = \int_T^\infty E[t(t)]S(t \mid t > T)d(t - T)dt \]

when the flow of transactions is characterized by the BG/NBD.

- It is easier to compute DERT in the following manner:

\[
DERT = \sum_{i=1}^{\infty} \left( \frac{1}{1+d} \right)^{i-0.5} \left\{ E[X(T, T + i) \mid x, t_x, T] - E[X(T, T + i - 1) \mid x, t_x, T] \right\}
\]
Further Reading


Further Reading


Beyond the Basic Models
Implementation Issues

- Handling multiple cohorts
  - treatment of acquisition
  - consideration of cross-cohort dynamics
- Implication of data recording processes

Implications of Data Recording Processes
(Contractual Settings)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Calendar Time  →</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( n_{11} ) ( n_{12} ) ( n_{13} ) ( \ldots ) ( n_{1I} )</td>
</tr>
<tr>
<td>2</td>
<td>( n_{22} ) ( n_{23} ) ( \ldots ) ( n_{2I} )</td>
</tr>
<tr>
<td>3</td>
<td>( n_{33} ) ( \ldots ) ( n_{3I} )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>I</td>
<td>( n_{II} ) ( n_{1I} ) ( n_{2I} ) ( n_{3I} ) ( \ldots ) ( n_{JI} )</td>
</tr>
</tbody>
</table>
Implications of Data Recording Processes
(Contractual Settings)

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Calendar Time</th>
<th>Cohort</th>
<th>Calendar Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_{11}$</td>
<td>1</td>
<td>$n_{11}$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{22}$</td>
<td>2</td>
<td>$n_{22}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>I−1</td>
<td>$n_{I−1,I−1}$</td>
<td>I−1</td>
<td>$n_{I−1,I−1}$</td>
</tr>
<tr>
<td>1</td>
<td>$n_{1I}$</td>
<td>1</td>
<td>$n_{1I}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Calendar Time</th>
<th>Cohort</th>
<th>Calendar Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_{11}$</td>
<td>1</td>
<td>$n_{11}$</td>
</tr>
<tr>
<td>2</td>
<td>$n_{22}$</td>
<td>2</td>
<td>$n_{22}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>I−1</td>
<td>$n_{I−1,I−1}$</td>
<td>I−1</td>
<td>$n_{I−1,I−1}$</td>
</tr>
<tr>
<td>1</td>
<td>$n_{1I}$</td>
<td>1</td>
<td>$n_{1I}$</td>
</tr>
</tbody>
</table>

Implications of Data Recording Processes
(Noncontractual Settings)

<table>
<thead>
<tr>
<th>ID = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>x x</td>
</tr>
<tr>
<td>x x</td>
</tr>
<tr>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
</tr>
<tr>
<td>x</td>
</tr>
<tr>
<td>. .</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID = 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>. .</td>
</tr>
<tr>
<td>. .</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID = n</th>
</tr>
</thead>
<tbody>
<tr>
<td>. .</td>
</tr>
<tr>
<td>. .</td>
</tr>
<tr>
<td>. .</td>
</tr>
<tr>
<td>. .</td>
</tr>
<tr>
<td>. .</td>
</tr>
<tr>
<td>. .</td>
</tr>
<tr>
<td>x x</td>
</tr>
</tbody>
</table>
Implications of Data Recording Processes
(Noncontractual Settings)

The model likelihood function must match the data structure:

- Interval-censored individual-level data
  
  Fader, Peter S. and Bruce G. S. Hardie (2005), “Implementing the Pareto/NBD Model Given Interval-Censored Data.”
  <http://brucehardie.com/notes/011/>

- Period-by-period histograms (RCSS)
  
Model Extensions

- Duration dependence
  - individual customer lifetimes
  - interpurchase times
- Nonstationarity
- Covariates

Individual-Level Duration Dependence

- The exponential distribution is often characterized as being “memoryless”.

- This means the probability that the event of interest occurs in the interval \((t, t + \Delta t]\) given that it has not occurred by \(t\) is independent of \(t\):

\[
P(t < T \leq t + \Delta t \mid T > t) = 1 - e^{-\lambda \Delta t}.
\]

- This is equivalent to a constant hazard function.
The Weibull Distribution

- A generalization of the exponential distribution that can have an increasing and decreasing hazard function:

\[ F(t) = 1 - e^{-\lambda t^c}, \lambda, c > 0 \]
\[ h(t) = c\lambda t^{c-1} \]

where \( c \) is the “shape” parameter and \( \lambda \) is the “scale” parameter.

- Collapses to the exponential when \( c = 1 \).

- \( F(t) \) is S-shaped for \( c > 1 \).

The Weibull Hazard Function

\[ h(t) = c\lambda t^{c-1} \]

- Decreasing hazard function (negative duration dependence) when \( c < 1 \).

- Increasing hazard function (positive duration dependence) when \( c > 1 \).
Individual-Level Duration Dependence

- Assuming Weibull-distributed individual lifetimes and gamma heterogeneity in $\lambda$ gives us the Weibull-gamma distribution, with survivor function

$$S(t \mid r, \alpha, c) = \left( \frac{\alpha}{\alpha + t^c} \right)^r$$

- DERL for a customer with tenure $s$ is computed by solving

$$\int_s^{\infty} \left( \frac{\alpha + s^c}{\alpha + t^c} \right)^r e^{-\delta(t-s)} dt$$

using standard numerical integration techniques.

215

Individual-Level Duration Dependence

- In a discrete-time setting, we have the discrete Weibull distribution:

$$S(t \mid \theta, c) = (1 - \theta)^{tc}.$$

- Assuming heterogeneity in $\theta$ follows a beta distribution with parameters $(\alpha, \beta)$, we arrive at the beta-discrete-Weibull (BdW) distribution with survivor function:

$$S(t \mid \alpha, \beta, c) = \int_0^1 S(t \mid \theta, c) g(\theta \mid \alpha, \beta) d\theta$$

$$= \frac{B(\alpha, \beta + t^c)}{B(\alpha, \beta)}.$$
Nonstationarity

- “Buy then die” ⇔ latent characteristics governing purchasing are constant then become 0.
- Perhaps more realistic to assume that these latent characteristics can change over time.
- Nonstationarity can be handled using a hidden Markov model
  or a (dynamic) changepoint model

Covariates

- Types of covariates:
  - customer characteristics
  - customer attitudes and behavior
  - marketing activities
- Handling covariate effects:
  - explicit integration (via latent characteristics and hazard functions)
  - used to create segments (and apply no-covariate models)
- Need to be wary of endogeneity bias and sample selection effects
The Cost of Model Extensions

- No closed-form likelihood functions; need to resort to simulation methods.
- Need full datasets; summaries (e.g., RFM) no longer sufficient.

Philosophy of Model Building

- Keep it as simple as possible
- Minimize cost of implementation
  - use of readily available software (e.g., Excel)
  - use of data summaries
- Purposively ignore the effects of covariates (customer descriptors and marketing activities) so as to highlight the important underlying components of buyer behavior.
Central Tenet

Traditional approach
future = f(past)

Probability modelling approach
\( \hat{\theta} = f(\text{past}) \rightarrow \text{future} = f(\hat{\theta}) \)

Classifying Customer Bases

<table>
<thead>
<tr>
<th>Continuous Opportunities for Transactions</th>
<th>Credit card</th>
<th>Student mealplan</th>
<th>Mobile phone usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grocery purchases</td>
<td>Prescription refills</td>
<td>Event attendance</td>
<td>Charity fund drives</td>
</tr>
<tr>
<td>Doctor visits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hotel stays</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discrete Opportunities for Transactions</td>
<td>Magazine subs</td>
<td>Insurance policy</td>
<td>Health club m'ship</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noncontractual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contractual</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of Relationship With Customers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>