Probability Models for Customer-Base Analysis

Peter S. Fader
University of Pennsylvania
`www.petefader.com`

Bruce G. S. Hardie
London Business School
`www.brucehardie.com`

19th Annual Advanced Research Techniques Forum
June 15–18, 2008

Agenda

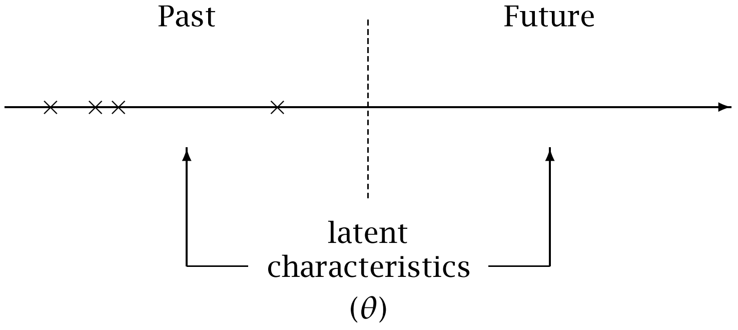
- Introduction to customer-base analysis
- The right way to think about computing CLV
- Review of probability models
- Models for contractual settings
- Models for noncontractual settings
  - The Pareto/NBD model
  - The BG/NBD model
  - The BG/BB model
- Beyond the basic models

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)



Past

Future

latent
characteristics
(θ)

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

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

	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m’ship

Opportunities for Transactions

Type of Relationship With Customers

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 $E(CLV)$

A more correct starting point:

$$E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt$$

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

Review of Probability Models

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*.

Building a Probability Model

(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 λ 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 λ if

$$F(t \mid \lambda) = P(T \le t \mid \lambda) = 1 - e^{-\lambda t}\,,\ t > 0\,.$$

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 p is the probability of success.

> Repeated Bernoulli trials lead to the *geometric* and *binomial* distributions.

Individual-level Building Blocks

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

$$P(X = x \mid p) = p(1-p)^{x-1}\,,\; x = 1, 2, 3, \ldots$$

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 is distributed binomial with parameter p, where

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

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?”

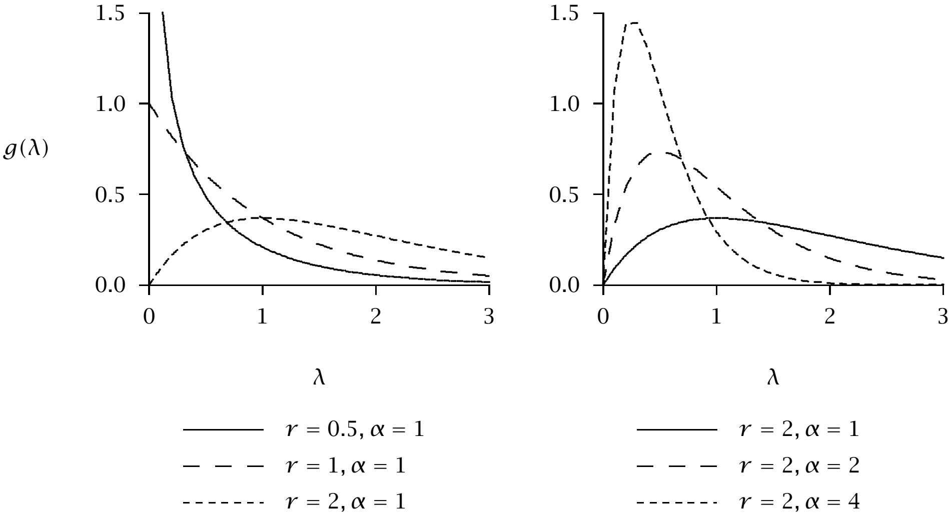
Capturing Heterogeneity in Latent Traits

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 α is the “scale” parameter
- The gamma distribution is a flexible (unimodal) distribution … and is mathematically convenient.

Illustrative Gamma Density Functions



Capturing Heterogeneity in Latent Traits

The beta distribution:

$$g(p \mid \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}, \ 0 < p < 1 .$$

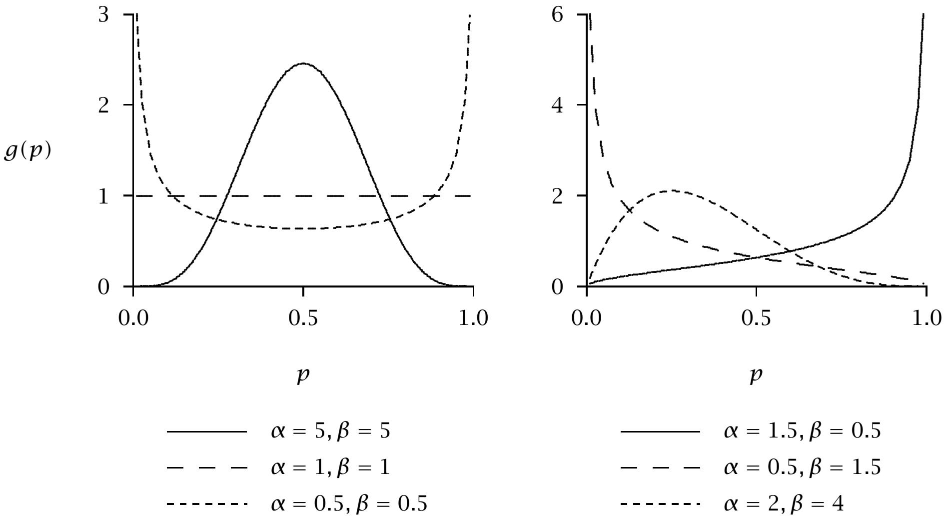
- $B(\alpha,\beta)$ is the beta function, which can be expressed in terms of gamma functions:

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

- The beta distribution is a flexible distribution ... and is mathematically convenient

---

Illustrative Beta Density Functions



The Negative Binomial Distribution (NBD)

- The individual-level behavior of interest can be characterized by the Poisson distribution when the mean λ is known.
- We do not observe an individual's λ but assume it is distributed across the population according to a gamma distribution.

$$\begin{aligned} P(X = x \mid r, \alpha) &= \int_0^\infty P(X = x \mid \lambda) g(\lambda \mid r, \alpha) d\lambda \\ &= \frac{\Gamma(r + x)}{\Gamma(r) x!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x . \end{aligned}$$

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 λ is known.
- We do not observe an individual's λ but assume it is distributed across the population according to a gamma distribution.

$$\begin{aligned} F(t \mid r, \alpha) &= \int_0^\infty F(t \mid \lambda) \, g(\lambda \mid r, \alpha) \, d\lambda \\ &= 1 - \left(\frac{\alpha}{\alpha + t}\right)^r . \end{aligned}$$

The Shifted-Beta-Geometric Model

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

$$\begin{aligned} P(X = x \mid \alpha, \beta) &= \int_0^1 P(X = x \mid p) g(p \mid \alpha, \beta)\, dp \\ &= \frac{B(\alpha + 1, \beta + x - 1)}{B(\alpha, \beta)} \,. \end{aligned}$$

The Beta-Binomial Distribution

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

$$\begin{aligned} P(X = x \mid n, \alpha, \beta) &= \int_0^1 P(X = x \mid n, p) g(p \mid \alpha, \beta)\, dp \\ &= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \,. \end{aligned}$$

Summary of Probability Models

Phenomenon	Individual-level	Heterogeneity	Model
Counting	Poisson	gamma	NBD
Timing	exponential	gamma	EG (Pareto)
Discrete timing (or counting)	shifted-geometric	beta	sBG
Choice	binomial	beta	BB

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

		Stage 2		
		Counting	Timing	Choice
	Counting			
Stage 1	Timing			
	Choice			

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

$$\int_0^1 t^a (1-t)^b (u+vt)^{-c} dt$$
$$= \begin{cases} B(a+1,b+1)u^{-c} \\ \qquad \times {}_2F_1(c,a+1;a+b+2;-\frac{v}{u}), & |v| \leq u \\ B(a+1,b+1)(u+v)^{-c} \\ \qquad \times {}_2F_1(c,b+1;a+b+2;\frac{v}{u+v}), & |v| \geq u \end{cases}$$

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

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

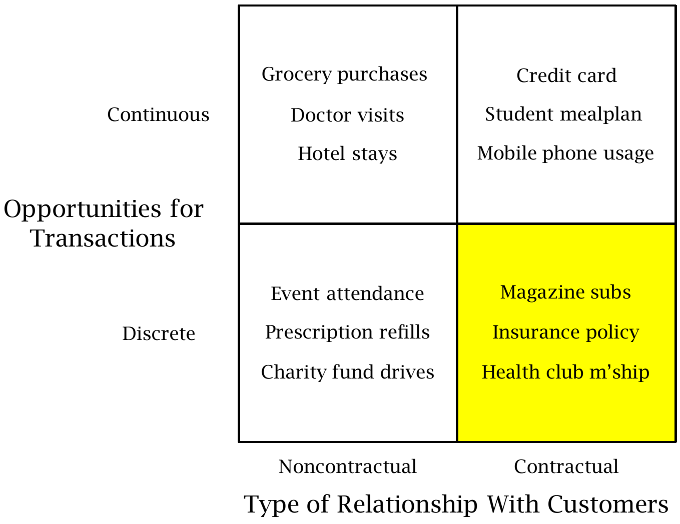
using the recursion

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

where $u_0 = 1$.

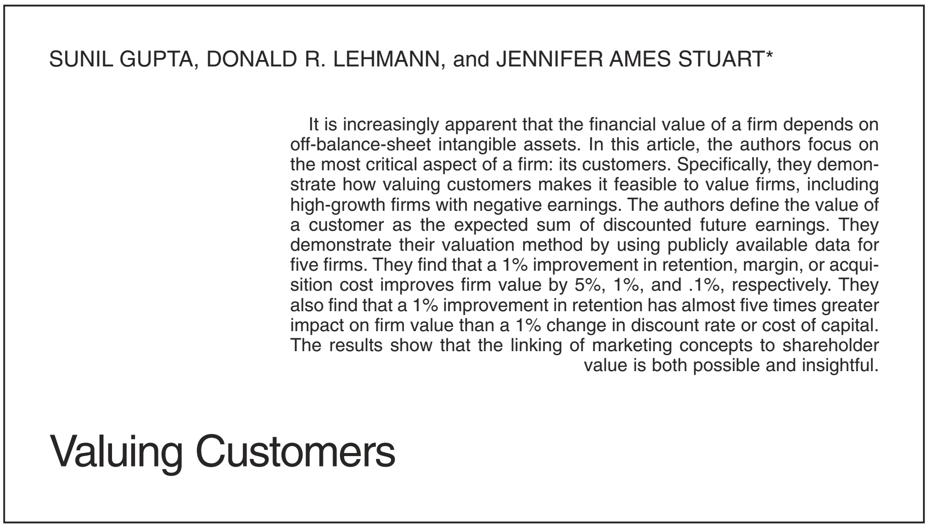
Models for Contractual Settings

Classifying Customer Bases



Opportunities for Transactions	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship

Type of Relationship With Customers



SUNIL GUPTA, DONALD R. LEHMANN, and JENNIFER AMES STUART*

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.

Valuing Customers

Hypothetical Contractual Setting

# Customers	2003	2004	2005	2006	2007
New	10,000	10,000	10,000	10,000	10,000
End of year	10,000	16,334	20,701	23,965	26,569

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:

$$26{,}569 \times \sum_{t=1}^{\infty} \$100 \times \frac{0.691^t}{(1+0.1)^{t-1}} = \$4{,}945{,}049$$

What’s wrong with this analysis?

Hypothetical Contractual Setting

# Customers	2003	2004	2005	2006	2007
New	10,000	10,000	10,000	10,000	10,000
End of year	10,000	16,334	20,701	23,965	26,569

Hypothetical Contractual Setting

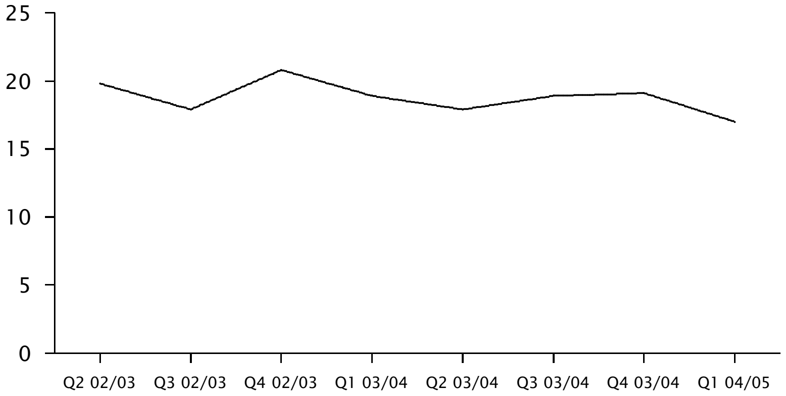
Number of customers who are still alive each year by year-of- acquisition cohort:

2003	2004	2005	2006	2007
10,000	6,334	4,367	3,264	2,604
	10,000	6,334	4,367	3,264
		10,000	6,334	4,367
			10,000	6,334
				10,000
10,000	16,334	20,701	23,965	26,569

Annual Retention Rates by Cohort

2003	2004	2005	2006	2007
--	0.633	0.689	0.747	0.798
	--	0.633	0.689	0.747
		--	0.633	0.689
			--	0.633
				--
--	0.633	0.655	0.675	0.691

Vodafone Germany
Quarterly Annualized Churn Rate (%)



Source: Vodafone Germany "Vodafone Analyst & Investor Day" presentation (2004-09-27)

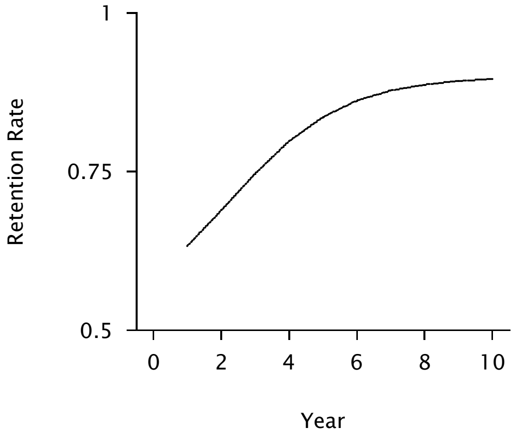
Renewal rates at regional magazines vary; generally 30% of subscribers renew at the end of their original subscription, but that figure jumps to 50% for second-time renewals and all the way to 75% for longtime readers.

Fielding, Michael (2005), "Get Circulation Going: DM Redesign Increases Renewal Rates for Magazines," *Marketing News*, September 1, 9–10.

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.

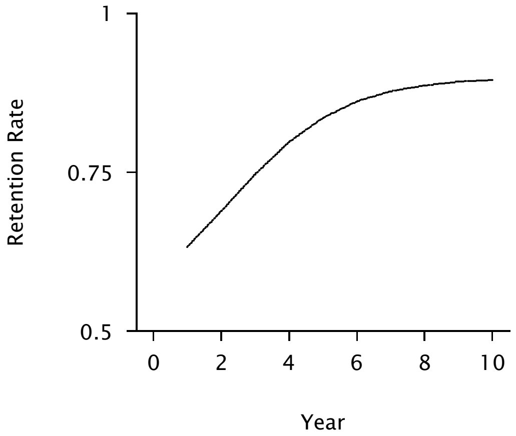
Netflix FY:03 Form 10-K

A Real-World Consideration



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 (e.g., increasing loyalty as the customer gains more experience with the firm).

vs.

A sorting effect in a heterogeneous population.

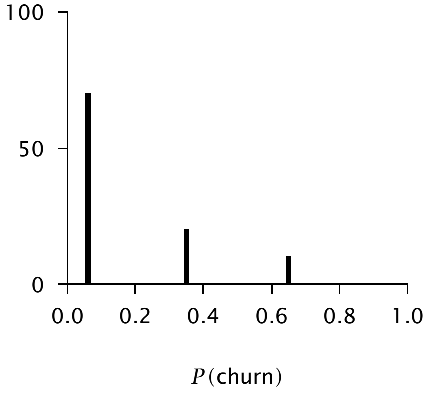
The Role of Heterogeneity

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

Cluster	$P(\text{churn})$	%CB
Low risk	0.06	70
Medium risk	0.35	20
High risk	0.65	10

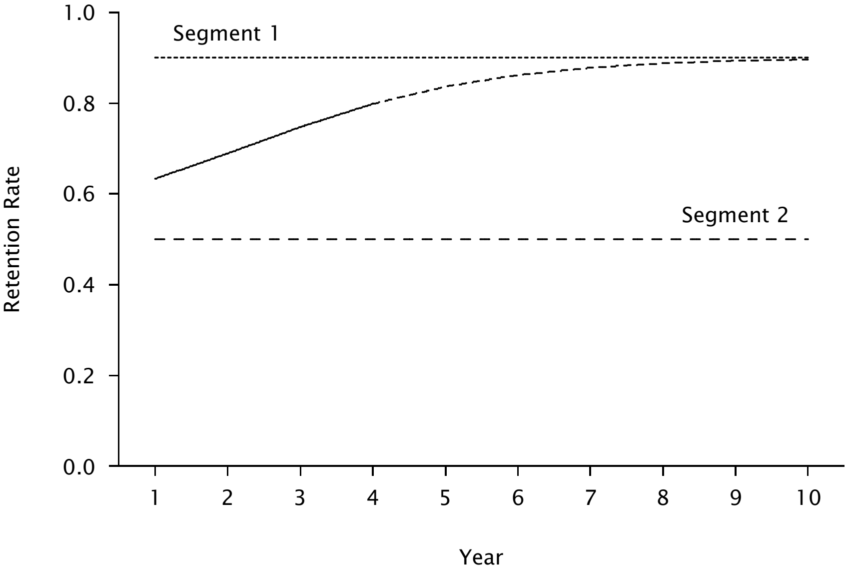


Source: “Vodafone Achievement and Challenges in Italy” presentation (2003-09-12)

The Role of Heterogeneity

	# Customers Still Alive			r_t		
Year	Seg 1	Seg 2	Total	Seg 1	Seg 2	Total
1	3,333	6,667	10,000			
2	3,000	3,334	6,334	0.900	0.500	0.633
3	2,700	1,667	4,367	0.900	0.500	0.689
4	2,430	834	3,264	0.900	0.500	0.747
5	2,187	417	2,604	0.900	0.500	0.798

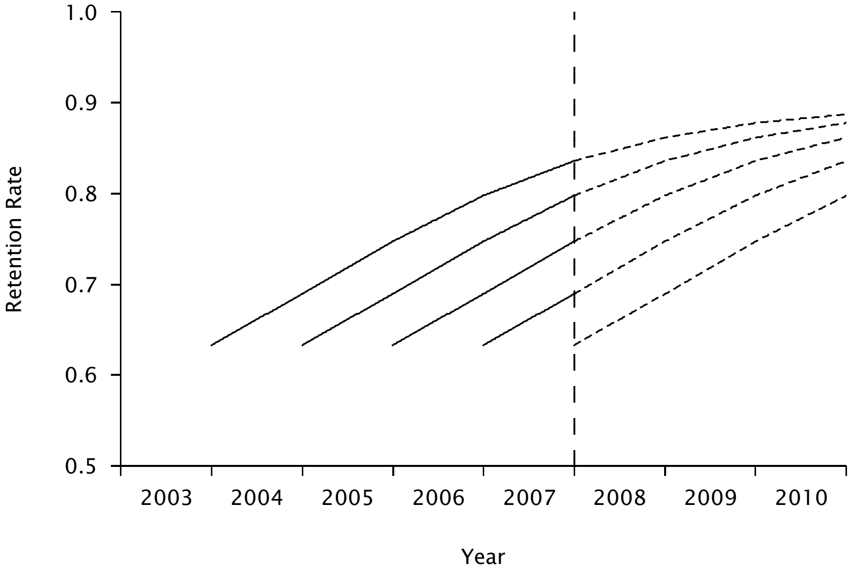
The Role of Heterogeneity



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).

Retention Rates by Cohort



$E(RLV)$ 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:

$$E(RLV) = \sum_{t=1}^{\infty} 100 \times \frac{0.5^t}{(1+0.1)^{t-1}} \\ = \$92$$

$E(RLV)$ of an Active 2003 Cohort Member

According to Bayes' theorem, the probability that this person belongs to segment 1 is

$$\frac{P(\text{renewed contract four times} \mid \text{segment 1}) \times P(\text{segment 1})}{P(\text{renewed contract four times})}$$
$$= \frac{0.9^4 \times 0.333}{0.9^4 \times 0.333 + 0.5^4 \times 0.667}$$
$$= 0.84$$

$$\Longrightarrow E(RLV) = 0.84 \times \$495 + (1 - 0.84) \times \$92 = \$430$$

$P(\text{Seg 1})$ as a Function of Customer "Age"

	# Customers Still Alive			
Year	Seg 1	Seg 2	Total	$P(\text{Seg 1})$
1	3,333	6,667	10,000	0.333
2	3,000	3,334	6,334	0.474
3	2,700	1,667	4,367	0.618
4	2,430	834	3,264	0.745
5	2,187	417	2,604	0.840

Valuing the Existing Customer Base

Recognizing the underlying segments:

Cohort	# Alive in 2007	$P(\text{Seg } 1)$	$E(RLV)$
2007	10,000	0.333	$226
2006	6,334	0.474	$283
2005	4,367	0.618	$341
2004	3,264	0.745	$392
2003	2,604	0.840	$430

Total expected residual value = $7,940,992

Valuing the Existing Customer Base

Cohort	Total RV	Underestimation
Naïve	$4,945,049	38%
Segment (model)	$7,940,992	

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 θ is captured by a beta distribution with pdf

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

A Discrete-Time Model for Contract Duration

- The probability that a customer cancels their contract in period t

$$\begin{aligned} P(T = t \mid \alpha, \beta) &= \int_0^1 P(T = t \mid \theta) f(\theta \mid \alpha, \beta)\, d\theta \\ &= \frac{B(\alpha + 1, \beta + t - 1)}{B(\alpha, \beta)}, \quad t = 1, 2, \ldots \end{aligned}$$

- The aggregate survivor function is

$$\begin{aligned} S(t \mid \alpha, \beta) &= \int_0^1 S(t \mid \theta) f(\theta \mid \alpha, \beta)\, d\theta \\ &= \frac{B(\alpha, \beta + t)}{B(\alpha, \beta)}, \quad t = 1, 2, \ldots \end{aligned}$$

A Discrete-Time Model for Contract Duration

- The (aggregate) retention rate is given by

$$\begin{aligned} r_t &= \frac{S(t)}{S(t-1)} \\ &= \frac{\beta + t - 1}{\alpha + \beta + t - 1}\,. \end{aligned}$$

- This is an increasing function of time, even though the underlying (unobserved) retention rates are constant at the individual-level.

Computing $E(CLV)$

- Recall:

$$E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt\,.$$

- In a contractual setting, assuming an individual's mean value per unit of time is constant ($\bar{v}$),

$$E(CLV) = \bar{v}\int_0^\infty S(t)d(t)dt\,.$$

- Standing at time s, a customer's expected residual lifetime value is

$$E(RLV) = \bar{v}\underbrace{\int_s^\infty S(t \mid t > s)d(t-s)dt}_{\text{discounted expected residual lifetime}}$$

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,

$$\begin{aligned} DERL(d \mid \theta, n-1 \text{ renewals}) &= \sum_{t=n}^{\infty} \frac{S(t \mid t > n-1; \theta)}{(1+d)^{t-n}} \\ &= \frac{(1-\theta)(1+d)}{d+\theta}\,. \end{aligned}$$

- But θ is unobserved ....

Computing DERL

- By Bayes' theorem, the posterior distribution of θ is

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

- It follows that

$$DERL(d \mid \alpha, \beta, n-1 \text{ renewals}) = \left(\frac{\beta+n-1}{\alpha+\beta+n-1}\right) {}_2F_1(1, \beta+n; \alpha+\beta+n; \tfrac{1}{1+d})$$

Computing DERL

Alternative derivation:

$$\begin{aligned} &DERL(d \mid \alpha, \beta, n-1 \text{ renewals}) \\ &= \sum_{t=n}^{\infty} \frac{S(t \mid t > n-1; \alpha, \beta)}{(1+d)^{t-n}} \\ &= \sum_{t=n}^{\infty} \frac{S(t \mid \alpha, \beta)}{S(n-1 \mid \alpha, \beta)} \left(\frac{1}{1+d}\right)^{t-n} \\ &= \sum_{t=n}^{\infty} \frac{B(\alpha, \beta+t)}{B(\alpha, \beta+n-1)} \left(\frac{1}{1+d}\right)^{t-n} \\ &= \left(\frac{\beta+n-1}{\alpha+\beta+n-1}\right) {}_2F_1(1, \beta+n; \alpha+\beta+n; \tfrac{1}{1+d}) \end{aligned}$$

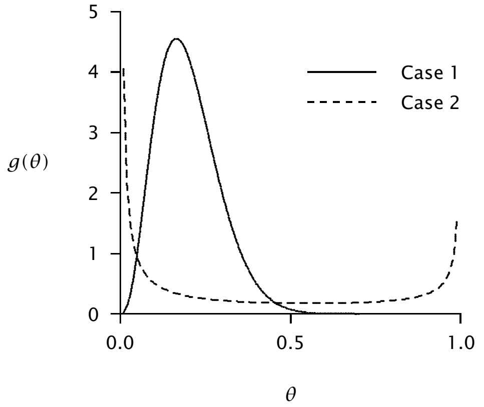
Impact of Heterogeneity on Error

- Assume the following arrival of new customers:

2003	2004	2005	2006	2007
10,000	10,000	10,000	10,000	10,000

- Assume $\bar{v} = \$1$ and a 10% discount rate.
- For given values of α and β, 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	α	β	$E(\theta)$	$S(1)$	$S(2)$	$S(3)$	$S(4)$
1	3.80	15.20	0.20	0.800	0.684	0.531	0.439
2	0.067	0.267	0.20	0.800	0.760	0.738	0.724

Computing DERL Using Excel

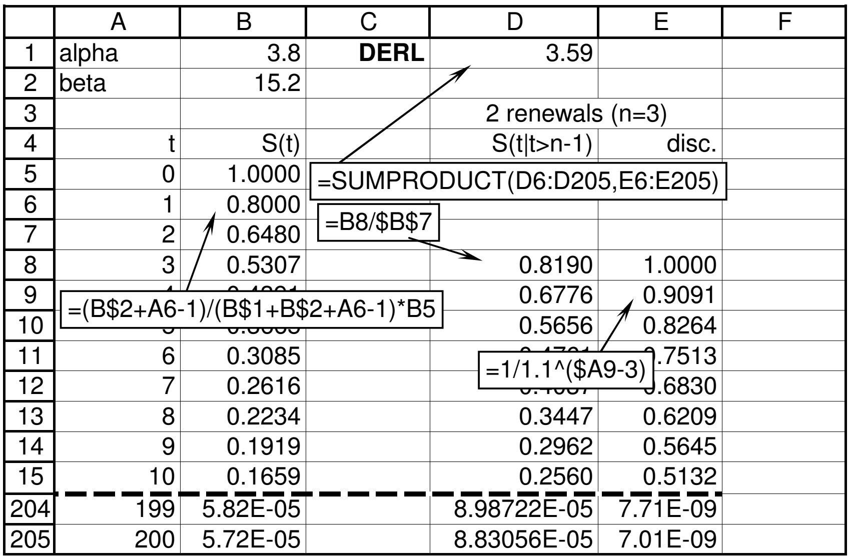
Recall our alternative derivation:

$$
\begin{aligned}
&DERL(d \mid \alpha, \beta, n-1 \text{ renewals}) \\
&\quad = \sum_{t=n}^{\infty} \frac{S(t \mid \alpha, \beta)}{S(n-1 \mid \alpha, \beta)} \left(\frac{1}{1+d}\right)^{t-n}
\end{aligned}
$$

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

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

Calculating DERL



	A	B	C	D	E	F
1	alpha	3.8	**DERL**	3.59		
2	beta	15.2				
3				2 renewals (n=3)		
4	t	S(t)		S(t|t>n-1)	disc.	
5	0	1.0000				
6	1	0.8000				
7	2	0.6480				
8	3	0.5307		0.8190	1.0000	
9	[illegible]	[illegible]		0.6776	0.9091	
10	[illegible]	[illegible]		0.5656	0.8264	
11	6	0.3085		[illegible]	0.7513	
12	7	0.2616		[illegible]	0.6830	
13	8	0.2234		0.3447	0.6209	
14	9	0.1919		0.2962	0.5645	
15	10	0.1659		0.2560	0.5132	
204	199	5.82E-05		8.98722E-05	7.71E-09	
205	200	5.72E-05		8.83056E-05	7.01E-09	

=SUMPRODUCT(D6:D205,E6:E205)

=B8/$B$7

=(B$2+A6-1)/(B$1+B$2+A6-1)*B5

=1/1.1^($A9-3)

Number of Active Customers: Case 1

2003	2004	2005	2006	2007	n	$E(RLV)$
10,000	8,000	6,480	5,307	4,391	5	\$3.84
	10,000	8,000	6,480	5,307	4	\$3.72
		10,000	8,000	6,480	3	\$3.59
			10,000	8,000	2	\$3.45
				10,000	1	\$3.31
10,000	18,000	24,480	29,787	34,178		

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

Impact of Heterogeneity on Error: Case 1

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

$$\begin{aligned}\text{Correct valuation} &= 4{,}391 \times \$3.84 + 5{,}307 \times \$3.72 \\ &\quad + 6{,}480 \times \$3.59 + 8{,}000 \times \$3.45 \\ &\quad + 10{,}000 \times \$3.31 \\ &= \$120{,}543\end{aligned}$$

Naïve underestimates correct by 12%.

Number of Active Customers: Case 2

2003	2004	2005	2006	2007	n	$E(RLV)$
10,000	8,000	7,600	7,383	7,235	5	\$10.19
	10,000	8,000	7,600	7,383	4	\$10.06
		10,000	8,000	7,600	3	\$9.86
			10,000	8,000	2	\$9.46
				10,000	1	\$7.68
10,000	18,000	25,600	32,983	40,218		

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

Impact of Heterogeneity on Error: Case 2

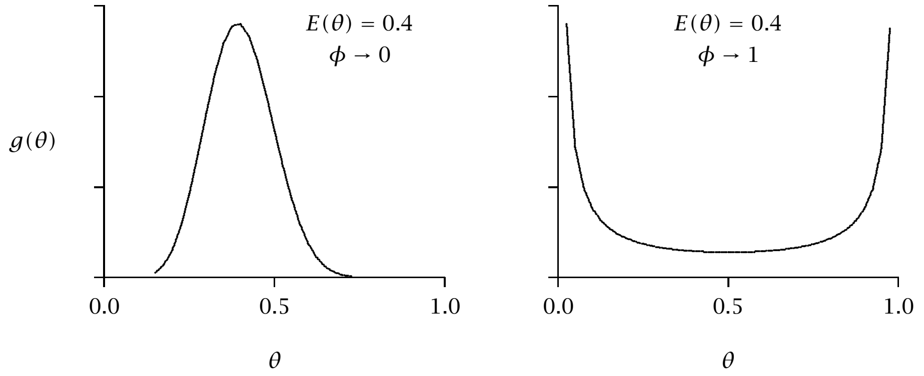
$$\begin{aligned}\text{Naïve valuation} &= 40{,}218 \times \sum_{t=1}^{\infty} \frac{0.92^t}{(1+0.1)^{t-1}} \\ &= \$220{,}488\end{aligned}$$

$$\begin{aligned}\text{Correct valuation} &= 7{,}235 \times \$10.19 + 7{,}383 \times \$10.06 \\ &\quad + 7{,}600 \times \$9.86 + 8{,}000 \times \$9.46 \\ &\quad + 10{,}000 \times \$7.68 \\ &= \$375{,}437\end{aligned}$$

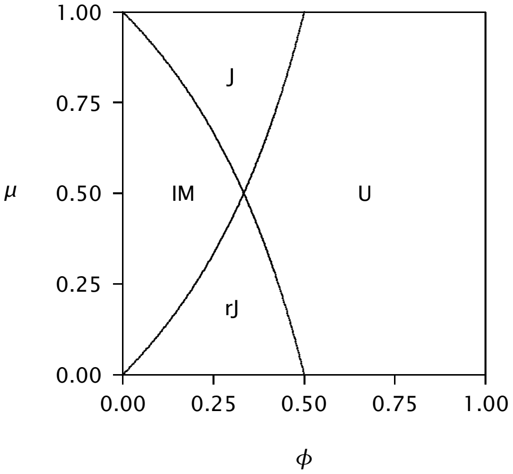
Naïve underestimates correct by 41%.

Interpreting the Beta Distribution Parameters

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

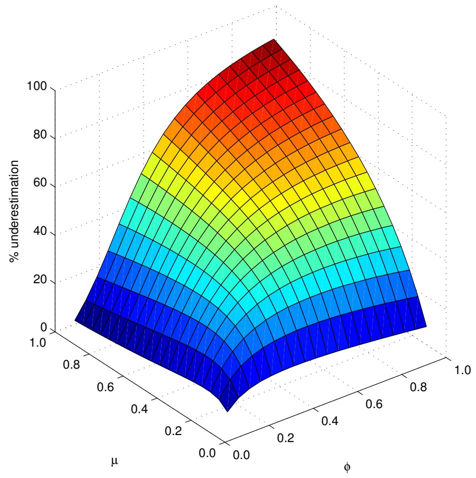


Shape of the Beta Distribution

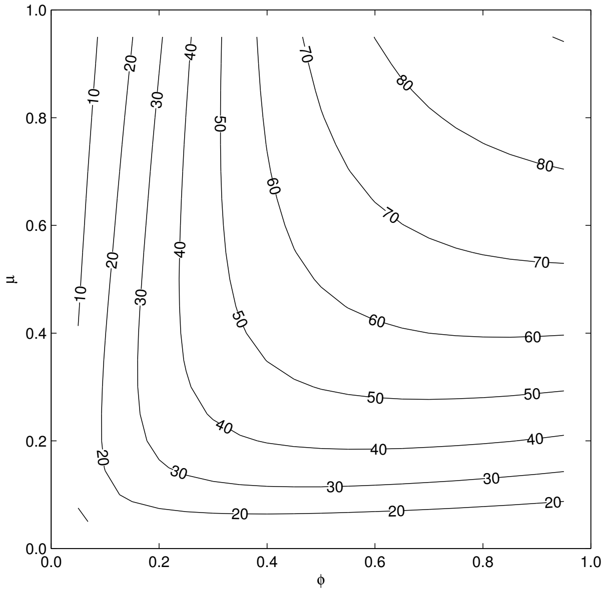


Error as a Function of μ and ϕ

For a fine grid of points in the (μ, ϕ) space, we determine the corresponding values of (α, β) and compute % underestimation:



Error as a Function of μ and ϕ



Re-analysis Using (r_1, r_2)

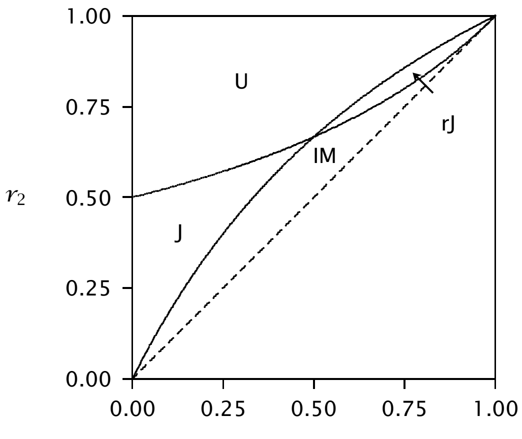
- μ and ϕ are not quantities that most managers or analysts think about; retention rates are easier to comprehend.
- Since the period 1 and 2 retention rates are, respectively,

$$r_1 = \frac{\beta}{\alpha + \beta} \text{ and } r_2 = \frac{\beta + 1}{\alpha + \beta + 1},$$

it follows that

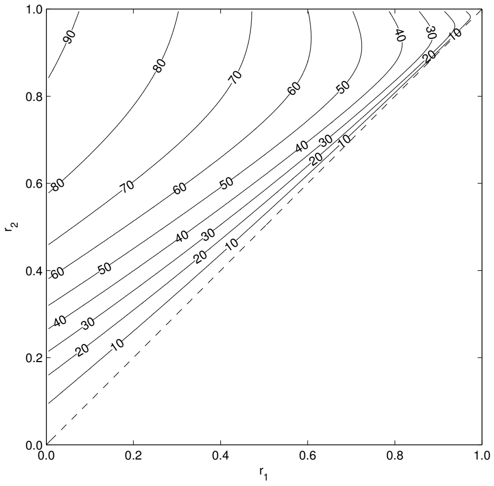
$$\alpha = \frac{(1 - r_1)(1 - r_2)}{r_2 - r_1} \text{ and } \beta = \frac{r_1(1 - r_2)}{r_2 - r_1}.$$

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 (α, β) and compute % underestimation:



Further Reading

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 (2008), “Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity.” <`http://brucehardie.com/papers/022/`>

Fader, Peter S. and Bruce G. S. Hardie (2007), “Computing DERL for the sBG Model Using Excel.” <`http://brucehardie.com/notes/018/`>

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

Opportunities for Transactions

	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship

Type of Relationship With Customers

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,

$$f(t \mid \lambda) = \lambda e^{-\lambda t}$$
$$S(t \mid \lambda) = e^{-\lambda t}$$

ii. Heterogeneity in λ follows a gamma distribution with pdf

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

Contract Duration in Continuous-Time

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

$$\begin{aligned} f(t \mid r, \alpha) &= \int_0^\infty f(t \mid \lambda) g(\lambda \mid r, \alpha)\, d\lambda \\ &= \frac{r}{\alpha}\left(\frac{\alpha}{\alpha + t}\right)^{r+1} \end{aligned}$$

$$\begin{aligned} S(t \mid r, \alpha) &= \int_0^\infty S(t \mid \lambda) g(\lambda \mid r, \alpha)\, d\lambda \\ &= \left(\frac{\alpha}{\alpha + t}\right)^{r} \end{aligned}$$

The Hazard Function

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

$$\begin{aligned} h(t) &= \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t | T > t)}{\Delta t} \\ &= \frac{f(t)}{1 - F(t)} \end{aligned}$$

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

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

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 δ,

$$\begin{aligned} DERL(\delta \mid \lambda, \text{tenure of at least } s) &= \int_s^\infty e^{-\lambda(t-s)} e^{-\delta(t-s)} dt \\ &= \frac{1}{\lambda + \delta} \end{aligned}$$

- But λ is unobserved ....

Computing DERL

By Bayes' theorem, the posterior distribution of λ for an individual with tenure of at least s,

$$\begin{aligned} g(\lambda \mid r, \alpha, \text{tenure of at least } s) &= \frac{S(s \mid \lambda) g(\lambda \mid r, \alpha)}{S(s \mid r, \alpha)} \\ &= \frac{(\alpha + s)^r \lambda^{r-1} e^{-\lambda(\alpha+s)}}{\Gamma(r)} \end{aligned}$$

Computing DERL

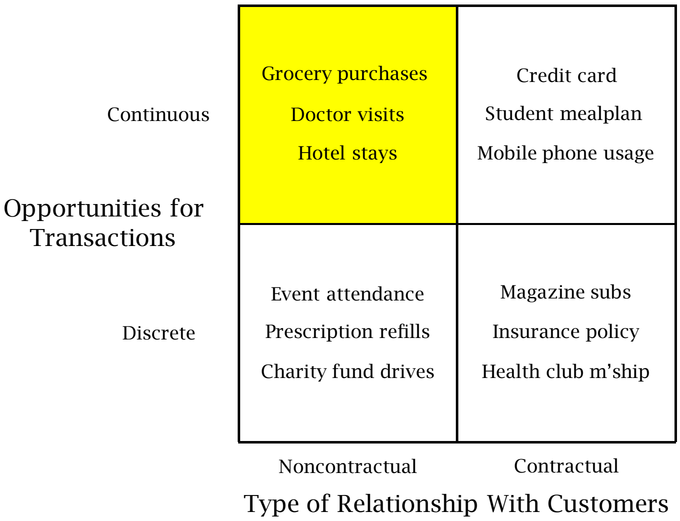
It follows that

$$\begin{aligned} &DERL(\delta \mid r, \alpha, \text{tenure of at least } s) \\ &\qquad = \int_0^\infty \Big\{ DERL(\delta \mid \lambda, \text{tenure of at least } s) \\ &\qquad\qquad \times g(\lambda \mid r, \alpha, \text{tenure of at least } s) \Big\} d\lambda \\ &\qquad = (\alpha + s)^r \delta^{r-1} \Psi(r, r; (\alpha + s)\delta) \end{aligned}$$

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

Models for Noncontractual Settings

Classifying Customer Bases



Opportunities for Transactions

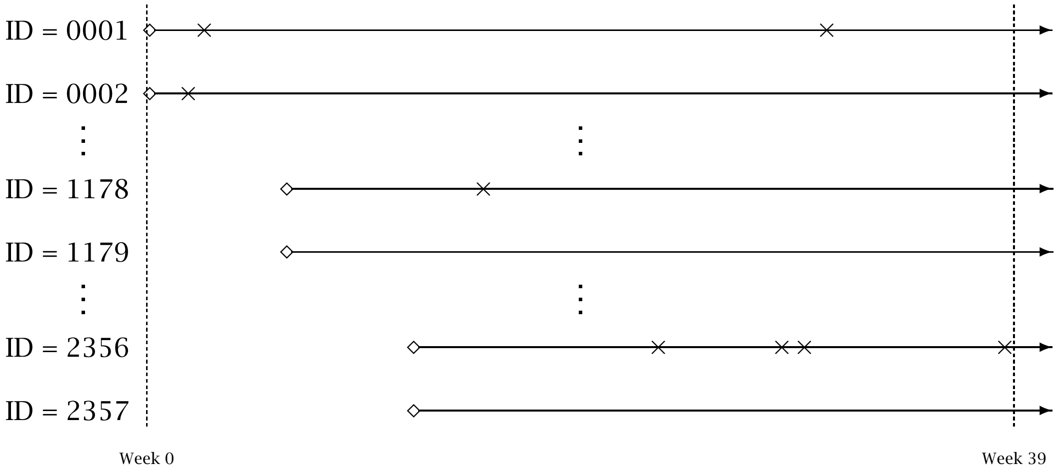
	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship

Type of Relationship With Customers

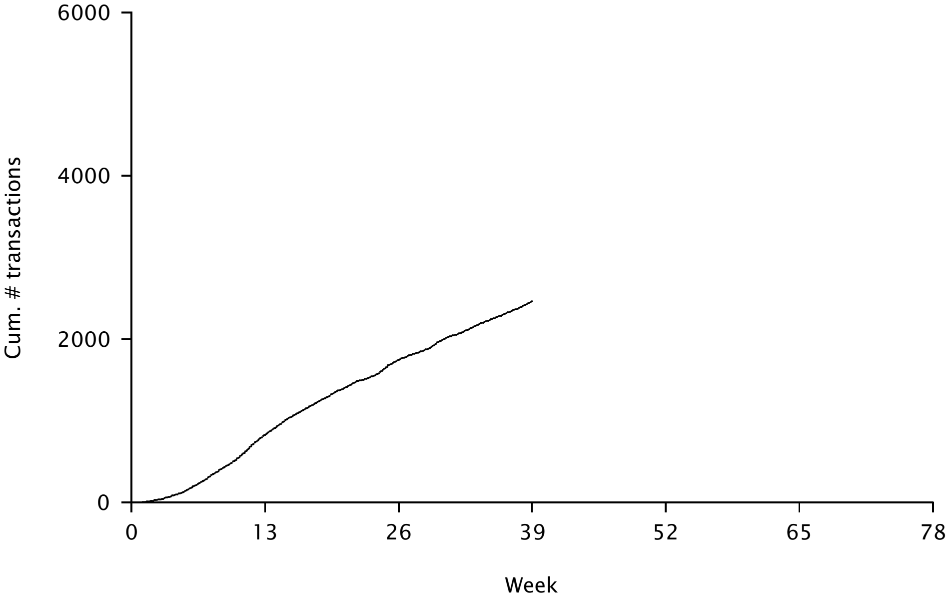
Setting

- 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



Cumulative Repeat Transactions



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

- A customer purchases “randomly” with an average transaction rate λ
- Transaction rates vary across customers

Modelling the Transaction Stream

- Let the random variable $X(t)$ denote the number of transactions in a period of length t time units.
- At the individual-level, $X(t)$ is assumed to be distributed Poisson with mean λt:

$$P(X(t) = x|\lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

- Transaction rates (λ) are distributed across the population according to a gamma distribution:

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

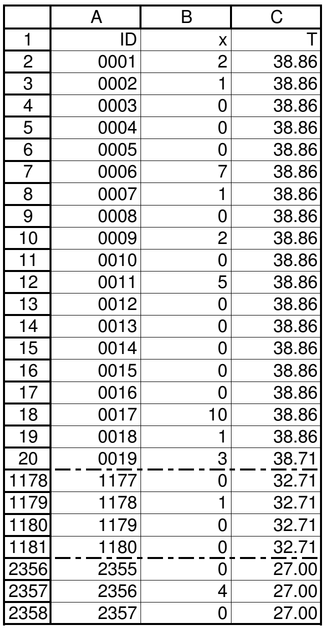
Modelling the Transaction Stream

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

$$\begin{aligned} P(X(t) = x \mid r, \alpha) &= \int_0^\infty P(X(t) = x \mid \lambda)\, g(\lambda)\, d\lambda \\ &= \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha + t}\right)^r \left(\frac{t}{\alpha + t}\right)^x , \end{aligned}$$

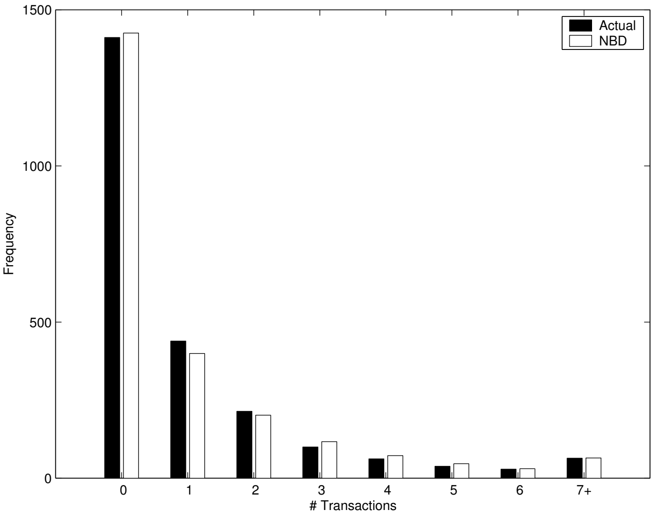
which is the negative binomial distribution (NBD).

Raw Data



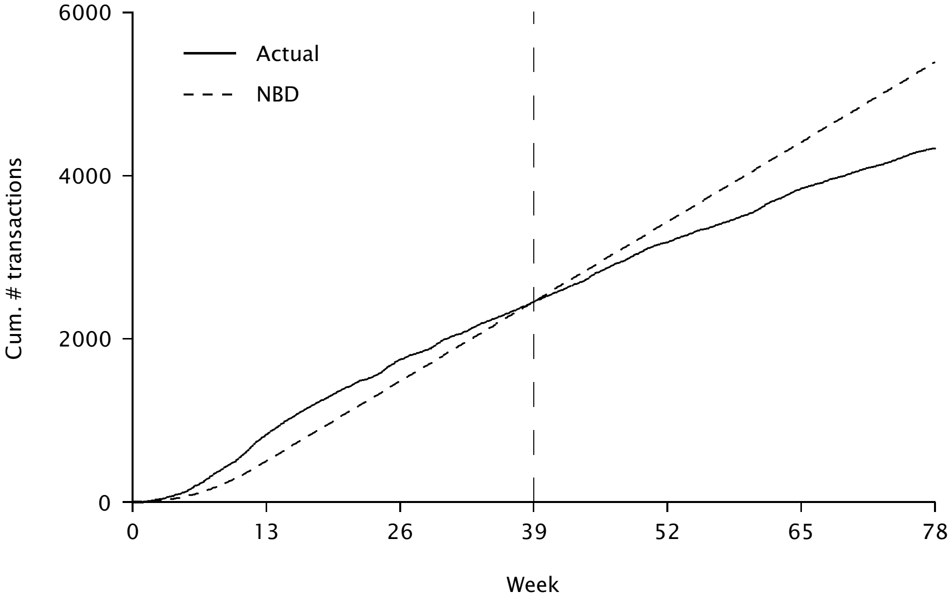
	A	B	C
1	ID	x	T
2	0001	2	38.86
3	0002	1	38.86
4	0003	0	38.86
5	0004	0	38.86
6	0005	0	38.86
7	0006	7	38.86
8	0007	1	38.86
9	0008	0	38.86
10	0009	2	38.86
11	0010	0	38.86
12	0011	5	38.86
13	0012	0	38.86
14	0013	0	38.86
15	0014	0	38.86
16	0015	0	38.86
17	0016	0	38.86
18	0017	10	38.86
19	0018	1	38.86
20	0019	3	38.71
1178	1177	0	32.71
1179	1178	1	32.71
1180	1179	0	32.71
1181	1180	0	32.71
2356	2355	0	27.00
2357	2356	4	27.00
2358	2357	0	27.00

Frequency of Repeat Transactions

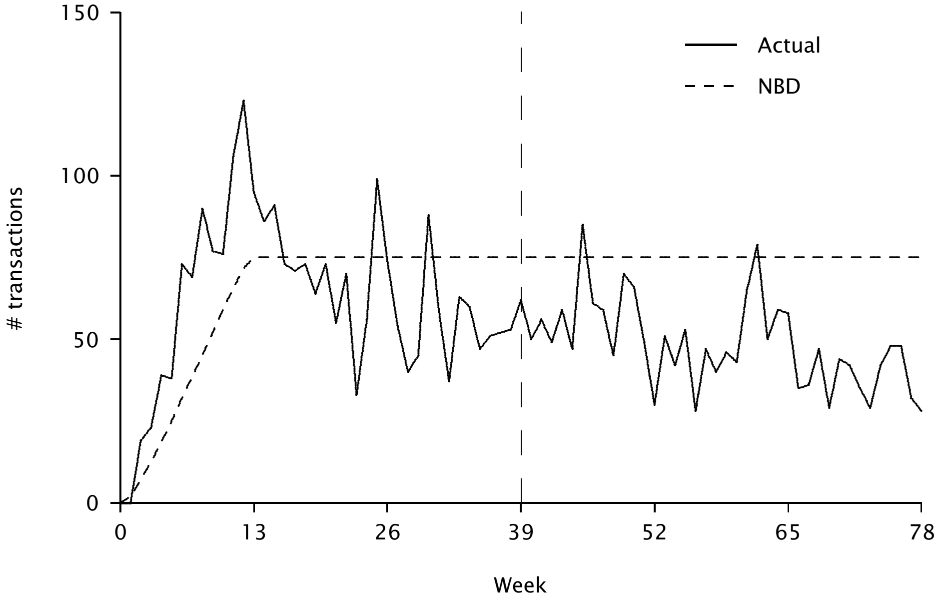


$$\hat{r} = 0.385, \hat{\alpha} = 12.072, LL = -9763.7$$

Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions

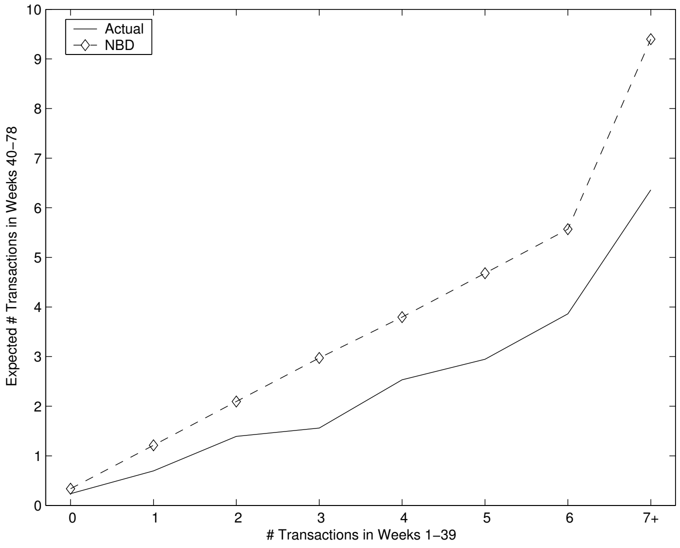


Conditional Expectations

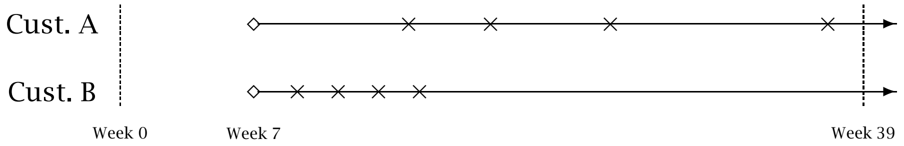
- We are interested in computing $E[X(T, T+t)|\text{data}]$, the expected number of transactions in an adjacent period $(T, T+t]$, conditional on the observed purchase history.
- For the NBD, a straight-forward application of Bayes' theorem gives us

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

Conditional Expectations



Conditional Expectations

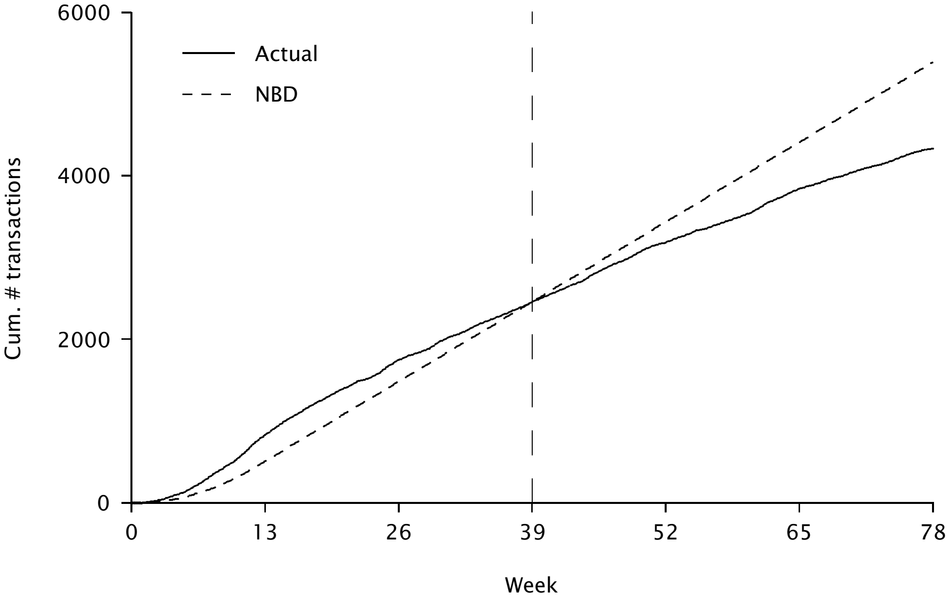


According to the NBD model:

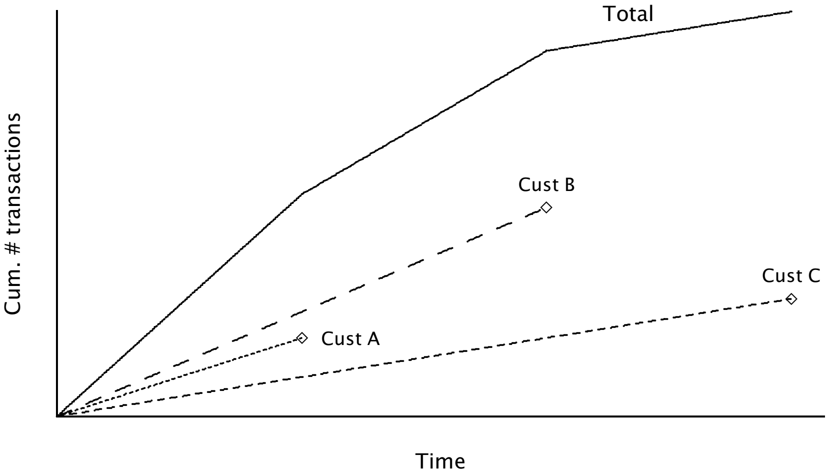
Cust. A: $E[X(39, 78) \mid x = 4, T = 32] = 3.88$

Cust. B: $E[X(39, 78) \mid x = 4, T = 32] = ?$

Tracking Cumulative Repeat Transactions



Towards a More Realistic Model



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 λ.
- Heterogeneity in transaction rates across customers is distributed gamma(r, α).

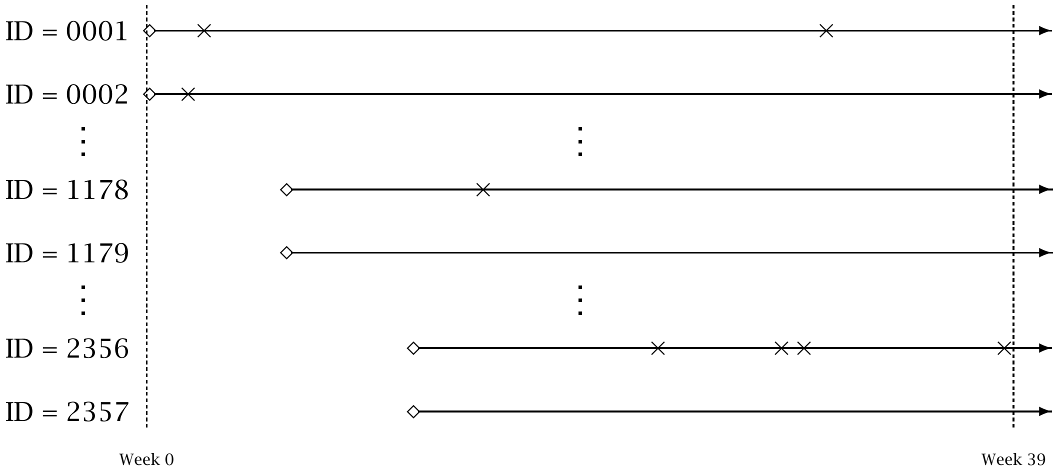
Dropout Process:

- Each customer has an unobserved “lifetime” of length ω, which is distributed exponential with dropout rate μ.
- Heterogeneity in dropout rates across customers is distributed gamma(s, β).

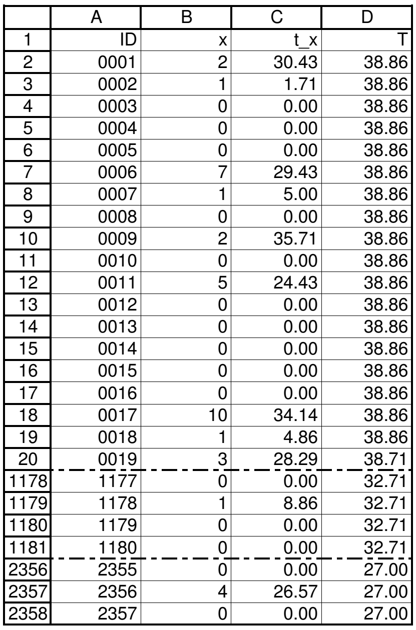
Summarizing Purchase Histories

- 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 *recency* and *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



Raw Data



	A	B	C	D
1	ID	x	t_x	T
2	0001	2	30.43	38.86
3	0002	1	1.71	38.86
4	0003	0	0.00	38.86
5	0004	0	0.00	38.86
6	0005	0	0.00	38.86
7	0006	7	29.43	38.86
8	0007	1	5.00	38.86
9	0008	0	0.00	38.86
10	0009	2	35.71	38.86
11	0010	0	0.00	38.86
12	0011	5	24.43	38.86
13	0012	0	0.00	38.86
14	0013	0	0.00	38.86
15	0014	0	0.00	38.86
16	0015	0	0.00	38.86
17	0016	0	0.00	38.86
18	0017	10	34.14	38.86
19	0018	1	4.86	38.86
20	0019	3	28.29	38.71
1178	1177	0	0.00	32.71
1179	1178	1	8.86	32.71
1180	1179	0	0.00	32.71
1181	1180	0	0.00	32.71
2356	2355	0	0.00	27.00
2357	2356	4	26.57	27.00
2358	2357	0	0.00	27.00

Pareto/NBD Likelihood Function

Removing the conditioning on λ and μ:

$$
\begin{aligned}
&L(r,\alpha,s,\beta \mid x, t_x, T)\\
&\quad= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\left\{\left(\frac{s}{r+s+x}\right)\frac{{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}\right)}{(\alpha+t_x)^{r+s+x}}\right.\\
&\qquad\left.+\left(\frac{r+x}{r+s+x}\right)\frac{{}_2F_1\left(r+s+x, s; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}\right)}{(\alpha+T)^{r+s+x}}\right\},\ \text{if } \alpha \ge \beta
\end{aligned}
$$

$$
\begin{aligned}
&L(r,\alpha,s,\beta \mid x, t_x, T)\\
&\quad= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)}\left\{\left(\frac{s}{r+s+x}\right)\frac{{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}\right)}{(\beta+t_x)^{r+s+x}}\right.\\
&\qquad\left.+\left(\frac{r+x}{r+s+x}\right)\frac{{}_2F_1\left(r+s+x, r+x+1; r+s+x+1; \frac{\beta-\alpha}{\beta+T}\right)}{(\beta+T)^{r+s+x}}\right\},\ \text{if } \alpha \le \beta
\end{aligned}
$$

Key Results

$E[X(t)]$

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

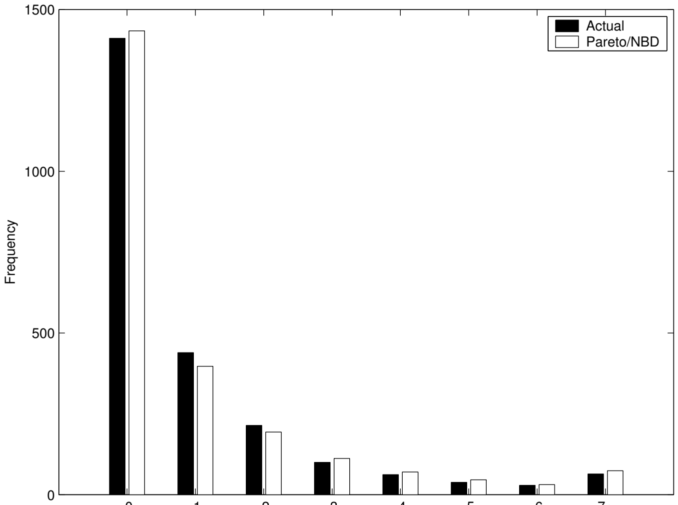
$P(\text{alive} \mid x, t_x, T)$

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

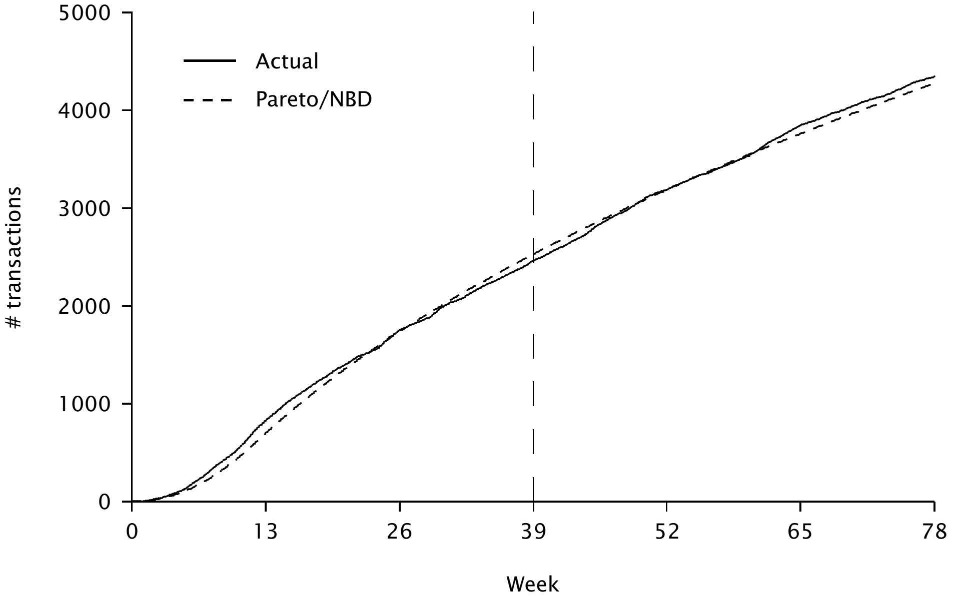
$E[X(T, T + t) \mid x, t_x, T]$

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

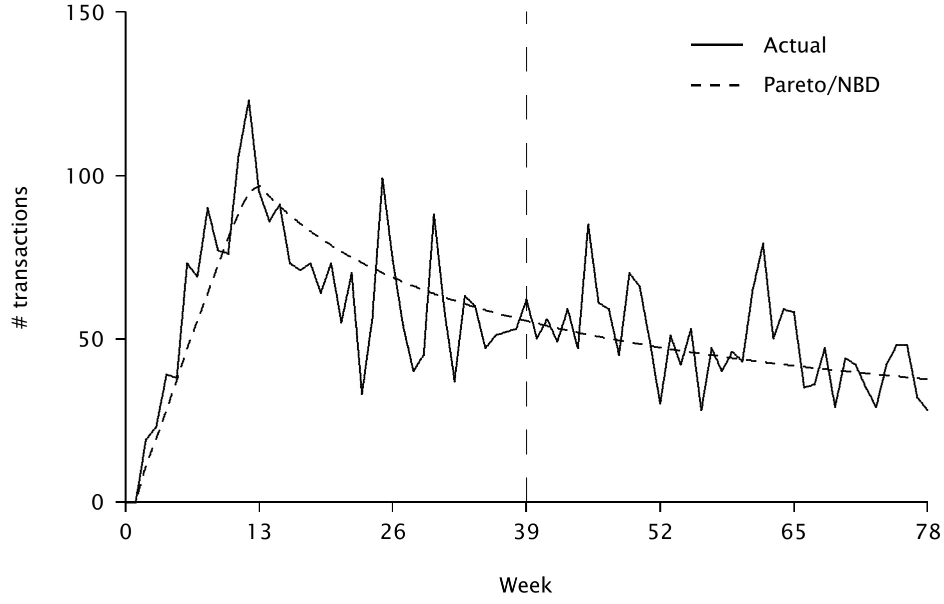
Frequency of Repeat Transactions



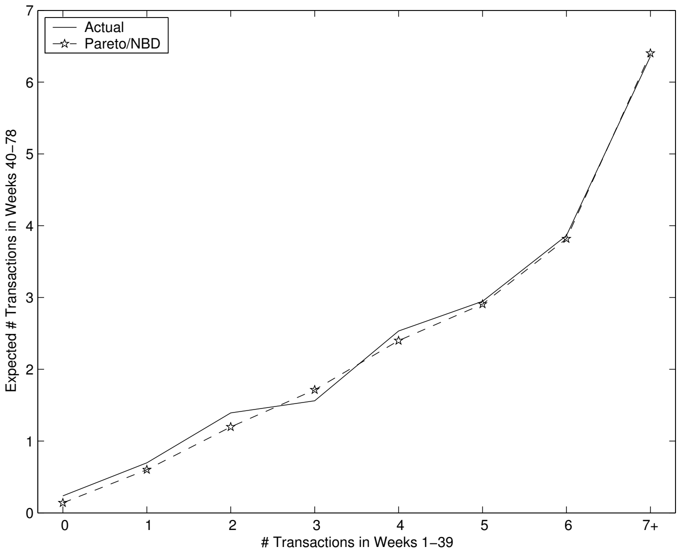
Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions



Conditional Expectations



Computing $E(CLV)$

$$E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt$$

If we assume that an individual's spend per transaction is constant, $v(t)$ = net cashflow / transaction × $t(t)$ (where $t(t)$ is the transaction rate at t) and

$$\begin{aligned} E(CLV) &= E(\text{net cashflow} / \text{transaction}) \\ &\quad \times \int_0^\infty E[t(t)]S(t)d(t)dt\,. \end{aligned}$$

Computing $E(RLV)$

- Standing at time T,

$$E(RLV) = E(\text{net cashflow / transaction}) \times \underbrace{\int_T^\infty E[t(t)]S(t \mid t > T)d(t - T)dt}_{\text{discounted expected residual transactions}}.$$

- The quantity $DERT$, discounted expected residual transactions, is the present value of the expected future transaction stream for a customer with purchase history (x, t_x, T).

Computing DERT

- For Poisson purchasing and exponential lifetimes with continuous compounding at rate of interest δ,

$$\begin{aligned} DERT(\delta \mid \lambda, \mu, \text{alive at } T) &= \int_T^\infty \lambda\left(\frac{e^{-\mu t}}{e^{-\mu T}}\right)e^{-\delta(t-T)}dt \\ &= \int_0^\infty \lambda e^{-\mu s}e^{-\delta s}ds \\ &= \frac{\lambda}{\mu + \delta} \end{aligned}$$

- However,
  - λ and μ are unobserved
  - We do not know whether the customer is alive at T

Computing DERT

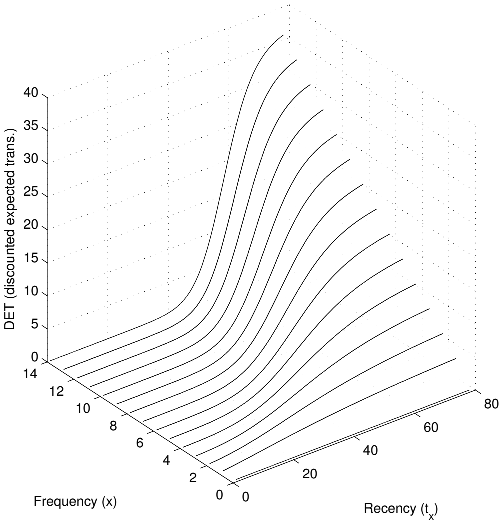
$$
\begin{aligned}
&DERT(\delta \mid r, \alpha, s, \beta, x, t_x, T)\\
&\quad= \int_0^\infty \int_0^\infty \Big\{ DERT(\delta \mid \lambda, \mu, \text{alive at } T)\\
&\qquad\qquad \times P(\text{alive at } T \mid \lambda, \mu, x, t_x, T)\\
&\qquad\qquad\quad \times g(\lambda, \mu \mid r, \alpha, s, \beta, x, t_x, T) \Big\} d\lambda d\mu\\
&\quad= \frac{\alpha^r \beta^s \delta^{s-1} \Gamma(r + x + 1) \Psi\big(s, s; \delta(\beta + T)\big)}{\Gamma(r)(\alpha + T)^{r+x+1} L(r, \alpha, s, \beta \mid x, t_x, T)}
\end{aligned}
$$

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

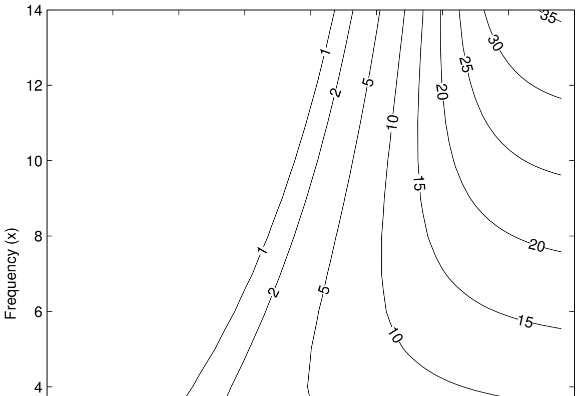
Continuous Compounding

- 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 $\delta = \ln(1 + d)/k$.

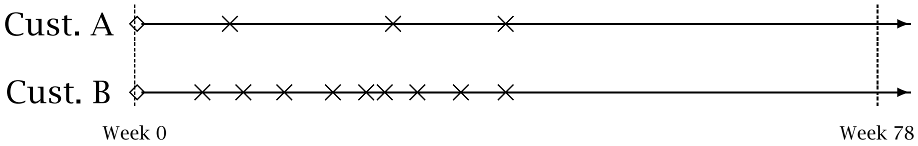
DERT by Recency and Frequency



Iso-Value Representation of DERT



The “Increasing Frequency” Paradox



	DERT
Cust. A	4.6
Cust. B	1.9

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.

Modelling the Spend 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

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

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 $E(M|m_x, x)$

Modelling the Spend Process

- The dollar value of a customer's given transaction is distributed gamma with shape parameter p and scale parameter ν
- Heterogeneity in ν across customers follows a gamma distribution with shape parameter q and scale parameter γ

Modelling the Spend Process

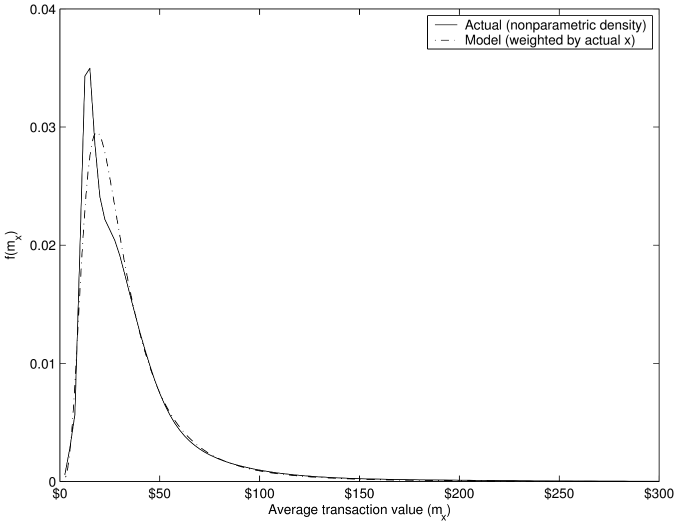
Marginal distribution for m_x:

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

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



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.

Average $E(RLV)$ by RFM Segment

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

Total Residual Value by RFM Segment

		Recency			
	Frequency	0	1	2	3
M=0	0	$53,000			
M=1	1		$7,700	$9,900	$1,800
	2		$2,800	$15,300	$17,400
	3		$300	$12,500	$52,900
M=2	1		$5,900	$7,600	$2,300
	2		$3,600	$26,500	$25,800
	3		$500	$37,200	$203,000
M=3	1		$11,300	$19,700	$3,700
	2		$7,300	$45,900	$47,900
	3		$1,000	$62,700	$414,900

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(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 λ.
- Heterogeneity in transaction rates across customers is distributed gamma(r, α).

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

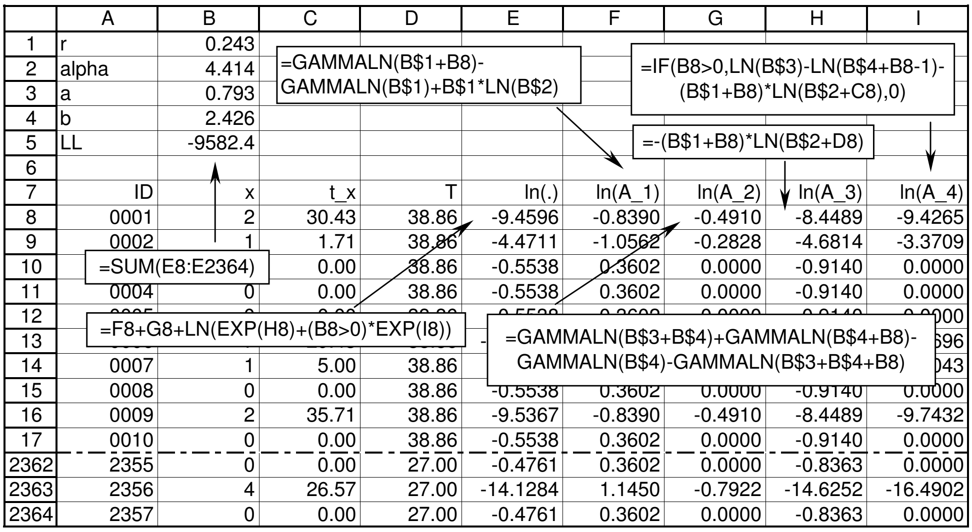
We can express the model likelihood function as:

$$L(r, \alpha, a, b \mid x, t_x, T) = \mathsf{A}_1 \cdot \mathsf{A}_2 \cdot (\mathsf{A}_3 + \delta_{x>0}\,\mathsf{A}_4)$$

where

$$\begin{aligned}
\mathsf{A}_1 &= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)} \\
\mathsf{A}_2 &= \frac{\Gamma(a+b)\Gamma(b+x)}{\Gamma(b)\Gamma(a+b+x)} \\
\mathsf{A}_3 &= \left(\frac{1}{\alpha+T}\right)^{r+x} \\
\mathsf{A}_4 &= \left(\frac{a}{b+x-1}\right)\left(\frac{1}{\alpha+t_x}\right)^{r+x}
\end{aligned}$$

BGNBD Estimation



	A	B	C	D	E	F	G	H	I
1	r	0.243							
2	alpha	4.414							
3	a	0.793							
4	b	2.426							
5	LL	-9582.4							
6									
7	ID	x	t_x	T	ln(.)	ln(A_1)	ln(A_2)	ln(A_3)	ln(A_4)
8	0001	2	30.43	38.86	-9.4596	-0.8390	-0.4910	-8.4489	-9.4265
9	0002	1	1.71	38.86	-4.4711	-1.0562	-0.2828	-4.6814	-3.3709
10			0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
11	0004	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
12									
13									
14	0007	1	5.00	38.86					
15	0008	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
16	0009	2	35.71	38.86	-9.5367	-0.8390	-0.4910	-8.4489	-9.7432
17	0010	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
2362	2355	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000
2363	2356	4	26.57	27.00	-14.1284	1.1450	-0.7922	-14.6252	-16.4902
2364	2357	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000

=GAMMALN(B$1+B8)-GAMMALN(B$1)+B$1*LN(B$2)

=IF(B8>0,LN(B$3)-LN(B$4+B8-1)-(B$1+B8)*LN(B$2+C8),0)

=-(B$1+B8)*LN(B$2+D8)

=SUM(E8:E2364)

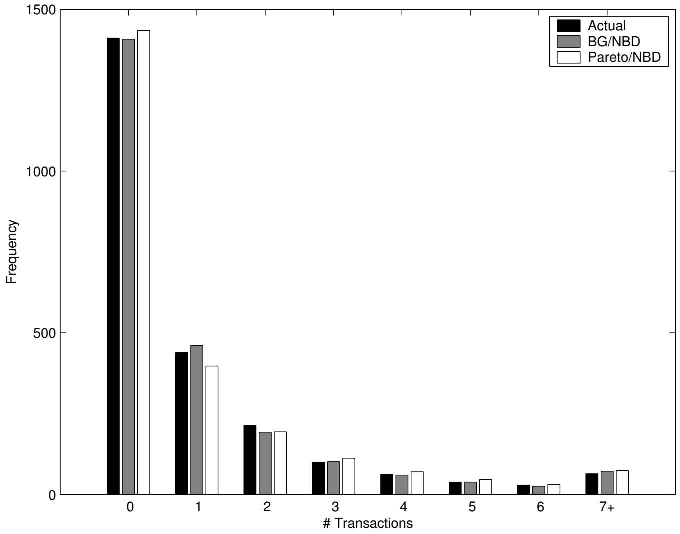
=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))

=GAMMALN(B$3+B$4)+GAMMALN(B$4+B8)-GAMMALN(B$4)-GAMMALN(B$3+B$4+B8)

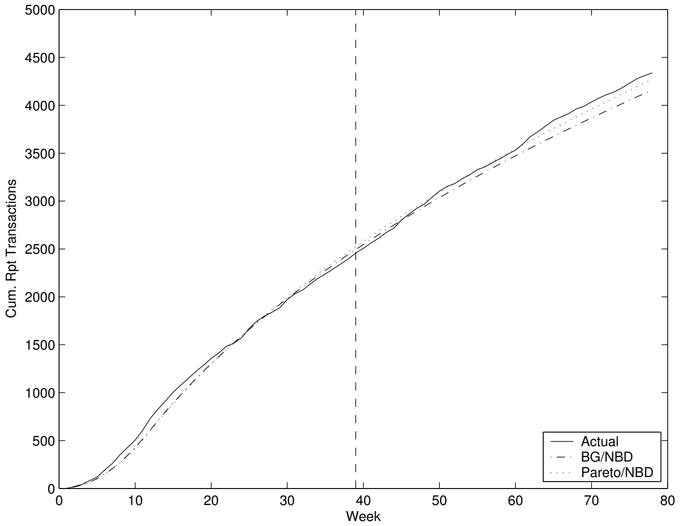
Model Estimation Results

	BG/NBD	Pareto/NBD
r	0.243	0.553
α	4.414	10.578
a	0.793	
b	2.426	
s		0.606
β		11.669
LL	−9582.4	−9595.0

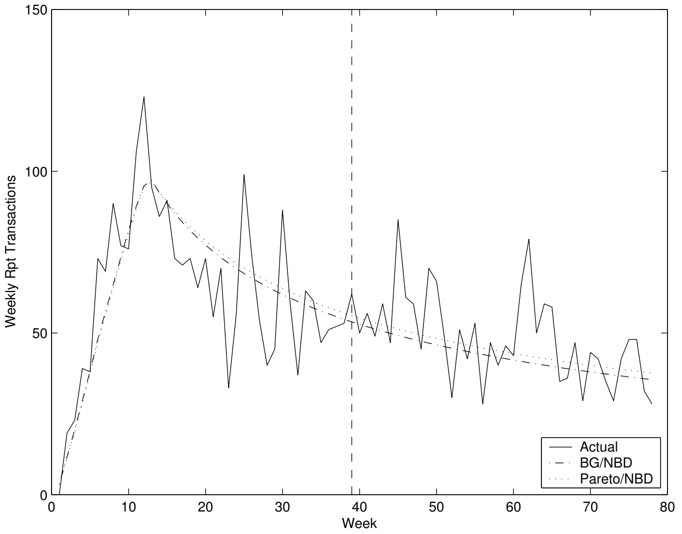
Frequency of Repeat Transactions



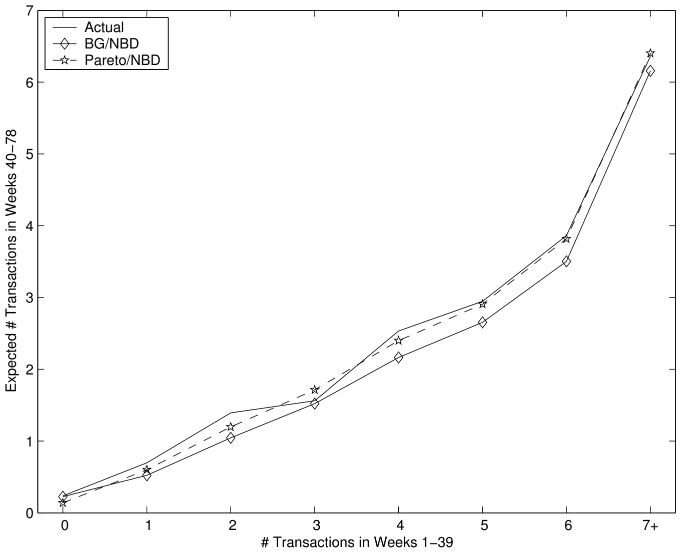
Tracking Cumulative Repeat Transactions



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:

$$DERT = \sum_{i=1}^{\infty} \Big(\frac{1}{1+d}\Big)^{i-0.5} \Big\{E\big[X(T, T+i) \mid x, t_x, T\big] \\ - E\big[X(T, T+i-1) \mid x, t_x, T\big]\Big\}$$

Further Reading

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

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

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005a), “A Note on Implementing the Pareto/NBD Model in MATLAB.” <`http://brucehardie.com/notes/008/`>

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005b), “RFM and CLV: Using Iso-value Curves for Customer Base Analysis,” *Journal of Marketing Research*, **42** (November), 415–430.

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

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005d), “Implementing the BG/NBD Model for Customer Base Analysis in Excel.” <`http://brucehardie.com/notes/004/`>

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2007), “Creating a Fit Histogram for the BG/NBD Model .” <`http://brucehardie.com/notes/014/`>

Fader, Peter S. and Bruce G. S. Hardie (2004), “Illustrating the Performance of the NBD as a Benchmark Model for Customer-Base Analysis.” <`http://brucehardie.com/notes/005/`>

Jerath, Kinshuk, Fader, Peter S., and Bruce G. S. Hardie (2007), “New Perspectives on Customer ‘Death’ Using a Generalization of the Pareto/NBD Model.” <`http://papers.ssrn.com/sol3/papers.cfm?abstract_id=995558`>

Modelling the Transaction Stream

How valid is the assumption of Poisson purchasing?

⟶ can transactions occur at any point in time?

Classifying Customer Bases

Opportunities for Transactions

	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship

Type of Relationship With Customers

“Discrete-Time” Transaction Opportunities

“necessarily discrete”	attendance at sports events attendance at annual arts festival
“generally discrete”	charity donations blood donations
discretized by recording process	cruise ship vacations

“Discrete-Time” Transaction Data

A *transaction opportunity* is

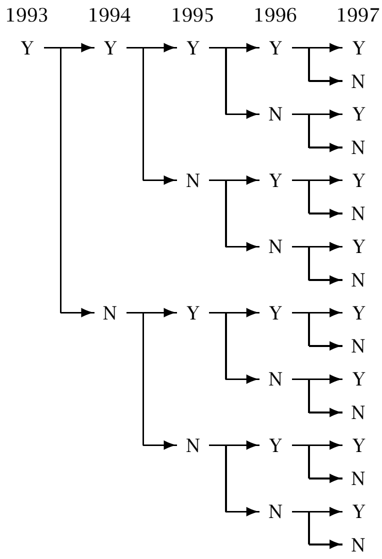
- a well-defined *point in time* at which a transaction either occurs or does not occur, or
- a well-defined *time interval* during which a (single) transaction either occurs or does not occur.

⟶ a customer’s transaction history can be expressed as a binary string:

$y_t = 1$ if a transaction occurred at/during the tth transaction opportunity, 0 otherwise.

Repeat Purchasing for Luxury Cruises

(Berger, Weinberg, and Hanna 2003)



1994	1995	1996	1997	# Customers
1	1	1	1	18
1	1	1	0	34
1	1	0	1	36
1	1	0	0	64
1	0	1	1	14
1	0	1	0	62
1	0	0	1	18
1	0	0	0	302
0	1	1	1	16
0	1	1	0	118
0	1	0	1	36
0	1	0	0	342
0	0	1	1	44
0	0	1	0	292
0	0	0	1	216
0	0	0	0	4482

Objectives

- Develop a model of buyer behavior for discrete- time, noncontractual settings.
- Derive expressions for quantities such as
  - the probability that an individual is still “alive”
  - the present value of the expected number of future transactions ($DERT \longrightarrow E(RLV)$ calculations)

  conditional on an individual’s observed behavior.
- Complete implementation within Microsoft Excel.

Model Development

A customer's relationship with a firm has two phases: he is "alive" (A) for some period of time, then becomes permanently inactive ("dies", D).

- While "alive", the customer buys at any given transaction opportunity (i.e., period t) with probability p:

$$P(Y_t = 1 \mid p, \text{alive at } t) = p$$

- A "living" customer becomes inactive at the beginning of a transaction opportunity (i.e., period t) with probability θ

$$\Rightarrow P(\text{alive at } t \mid \theta) = P(\underbrace{AA\ldots A}_{t} \mid \theta) = (1-\theta)^t$$

Model Development

What is $P(Y_1 = 1, Y_2 = 0, Y_3 = 1, Y_4 = 0, Y_5 = 0 \mid p, \theta)$?

- Three scenarios give rise to $Y_4 = 0, Y_5 = 0$:

Alive?

	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
i)	A	A	A	D	D
ii)	A	A	A	A	D
iii)	A	A	A	A	A

- The customer must have been alive for t = 1, 2, 3

Model Development

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

$$\begin{aligned} f(10100 \mid p, \theta) &= p(1-p)p\,\underbrace{(1-\theta)^3\theta}_{P(\text{AAADD})} \\ &\quad + p(1-p)p(1-p)\,\underbrace{(1-\theta)^4\theta}_{P(\text{AAAAD})} \\ &\quad + \underbrace{p(1-p)p}_{P(Y_1=1, Y_2=0, Y_3=1)}(1-p)(1-p)\,\underbrace{(1-\theta)^5}_{P(\text{AAAAA})} \end{aligned}$$

Model Development

- Bernoulli purchasing while alive ⇒ the order of a given number of transactions (prior to the last observed transaction) doesn't matter
- For example, $f(10100 \mid p, \theta) = f(01100 \mid p, \theta)$
- *Recency* (time of last transaction, t_x) and *frequency* (number of transactions, $x = \sum_{t=1}^{n} y_t$) are sufficient summary statistics
  - $\Longrightarrow$ we do not need the complete binary string representation of a customer's transaction history

Repeat Purchasing for Luxury Cruises

1994	1995	1996	1997	# Customers
1	1	1	1	18
1	1	1	0	34
1	1	0	1	36
1	1	0	0	64
1	0	1	1	14
1	0	1	0	62
1	0	0	1	18
1	0	0	0	302
0	1	1	1	16
0	1	1	0	118
0	1	0	1	36
0	1	0	0	342
0	0	1	1	44
0	0	1	0	292
0	0	0	1	216
0	0	0	0	4482

⟶

x	t_x	n	# Customers
4	4	4	18
3	4	4	66
2	4	4	98
1	4	4	216
3	3	4	34
2	3	4	180
1	3	4	292
2	2	4	64
1	2	4	342
1	1	4	302
0	0	4	4482

Model Development

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

$$\begin{aligned} L(p, \theta \mid x, t_x, n) = {} & p^x (1-p)^{n-x} (1-\theta)^n \\ & + \sum_{i=0}^{n-t_x-1} p^x (1-p)^{t_x - x + i} \theta (1-\theta)^{t_x + i} \end{aligned}$$

We assume that heterogeneity in p and θ across customers is captured by beta distributions:

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

$$g(\theta \mid \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}$$

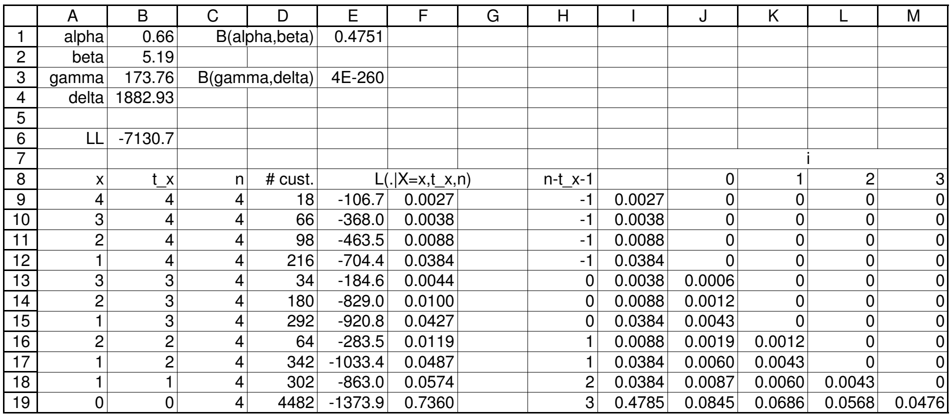
Model Development

Removing the conditioning on p and θ,

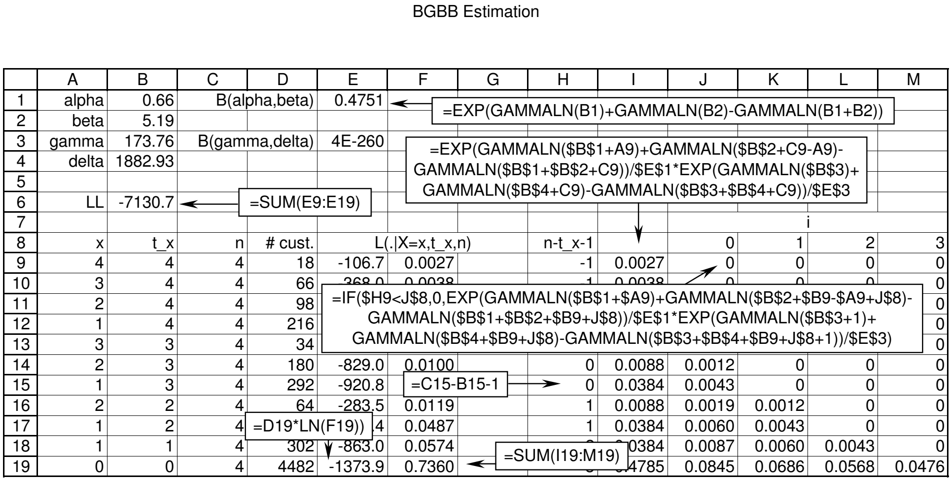
$$\begin{aligned}
&L(\alpha,\beta,\gamma,\delta \mid x, t_x, n) \\
&\quad = \int_0^1 \int_0^1 L(p,\theta \mid x, t_x, n) g(p \mid \alpha,\beta) g(\theta \mid \gamma,\delta)\, dp\, d\theta \\
&\quad = \frac{B(\alpha+x, \beta+n-x)}{B(\alpha,\beta)} \frac{B(\gamma,\delta+n)}{B(\gamma,\delta)} \\
&\qquad + \sum_{i=0}^{n-t_x-1} \frac{B(\alpha+x, \beta+t_x-x+i)}{B(\alpha,\beta)} \frac{B(\gamma+1, \delta+t_x+i)}{B(\gamma,\delta)}
\end{aligned}$$

... which is (relatively) easy to code-up in Excel.

BGBB Estimation



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	alpha	0.66	B(alpha,beta)		0.4751								
2	beta	5.19											
3	gamma	173.76	B(gamma,delta)		4E-260								
4	delta	1882.93											
5													
6	LL	-7130.7											
7										i			
8	x	t_x	n	# cust.	L(.|X=x,t_x,n)			n-t_x-1		0	1	2	3
9	4	4	4	18	-106.7	0.0027		-1	0.0027	0	0	0	0
10	3	4	4	66	-368.0	0.0038		-1	0.0038	0	0	0	0
11	2	4	4	98	-463.5	0.0088		-1	0.0088	0	0	0	0
12	1	4	4	216	-704.4	0.0384		-1	0.0384	0	0	0	0
13	3	3	4	34	-184.6	0.0044		0	0.0038	0.0006	0	0	0
14	2	3	4	180	-829.0	0.0100		0	0.0088	0.0012	0	0	0
15	1	3	4	292	-920.8	0.0427		0	0.0384	0.0043	0	0	0
16	2	2	4	64	-283.5	0.0119		1	0.0088	0.0019	0.0012	0	0
17	1	2	4	342	-1033.4	0.0487		1	0.0384	0.0060	0.0043	0	0
18	1	1	4	302	-863.0	0.0574		2	0.0384	0.0087	0.0060	0.0043	0
19	0	0	4	4482	-1373.9	0.7360		3	0.4785	0.0845	0.0686	0.0568	0.0476



BGBB Estimation

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	alpha	0.66		B(alpha,beta)	0.4751								
2	beta	5.19											
3	gamma	173.76		B(gamma,delta)	4E-260								
4	delta	1882.93											
5													
6	LL	-7130.7											
7										i			
8	x	t_x	n	# cust.	L(.|X=x,t_x,n)			n-t_x-1		0	1	2	3
9	4	4	4	18	-106.7	0.0027		-1	0.0027	0	0	0	0
10	3	4	4	66	-368.0	0.0038		-1	0.0038	0	0	0	0
11	2	4	4	98	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
12	1	4	4	216	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
13	3	3	4	34	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0
14	2	3	4	180	-829.0	0.0100		0	0.0088	0.0012	0	0	0
15	1	3	4	292	-920.8			0	0.0384	0.0043	0	0	0
16	2	2	4	64	-283.5	0.0119		1	0.0088	0.0019	0.0012	0	0
17	1	2	4	[illegible]	[illegible]	0.0487		1	0.0384	0.0060	0.0043	0	0
18	1	1	4	302	-863.0	0.0574		2	0.0384	0.0087	0.0060	0.0043	0
19	0	0	4	4482	-1373.9	0.7360		3	0.4785	0.0845	0.0686	0.0568	0.0476

=EXP(GAMMALN(B1)+GAMMALN(B2)-GAMMALN(B1+B2))

=EXP(GAMMALN($B$1+A9)+GAMMALN($B$2+C9-A9)-GAMMALN($B$1+$B$2+C9))/$E$1*EXP(GAMMALN($B$3)+GAMMALN($B$4+C9)-GAMMALN($B$3+$B$4+C9))/$E$3

=SUM(E9:E19)

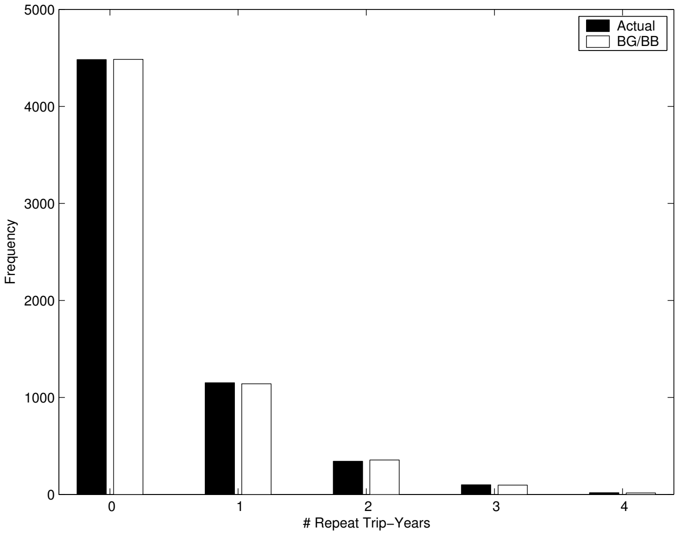
=IF($H9<J$8,0,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+J$8)-GAMMALN($B$1+$B$2+$B9+J$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+J$8)-GAMMALN($B$3+$B$4+$B9+J$8+1))/$E$3)

=C15-B15-1

=D19*LN(F19))

=SUM(I19:M19)

Model Fit



$\hat{\alpha} = 0.66, \hat{\beta} = 5.19, \hat{\gamma} = 173.76, \hat{\delta} = 1882.93, LL = -7130.7$

Key Results

$P(\text{alive in period } n + 1 \mid x, t_x, n)$

The probability that an individual with observed behavior (x, t_x, n) will be “active” in the next period.

$E[X(n + 1, n + n^*) \mid x, t_x, n]$

The expected number of transactions across the next n^* transaction opportunities for an individual with observed behavior (x, t_x, n).

$DERT(d \mid x, t_x, n)$

The discounted expected residual transactions for an individual with observed behavior (x, t_x, n).

P(alive in 1998) as a Function of Recency and Frequency

	Year of Last Cruise				
# Cruise-years	1997	1996	1995	1994	1993
4	0.92				
3	0.92	0.79			
2	0.92	0.81	0.68		
1	0.92	0.82	0.72	0.61	
0					0.60

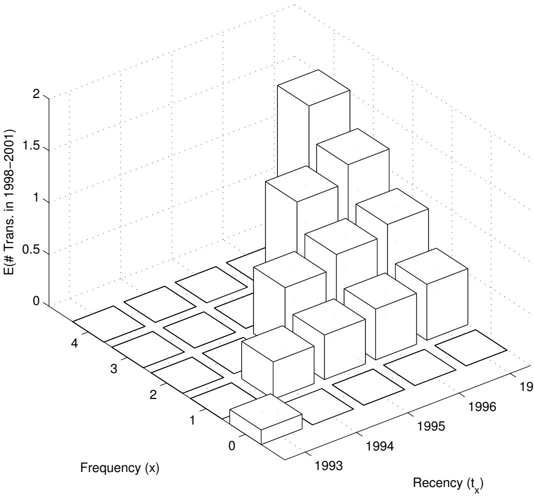
Posterior Mean of p as a Function of Recency and Frequency

	Year of Last Cruise				
# Cruise-years	1997	1996	1995	1994	1993
4	0.47				
3	0.37	0.38			
2	0.27	0.27	0.28		
1	0.17	0.17	0.18	0.19	
0					0.08

Expected # Transactions in 1998–2001 as a Function of Recency and Frequency

	Year of Last Cruise				
# Cruise-years	1997	1996	1995	1994	1993
4	1.52				
3	1.20	1.03			
2	0.87	0.77	0.64		
1	0.54	0.49	0.43	0.36	
0					0.14

Expected # Transactions in 1998–2001 as a Function of Recency and Frequency



Computing DERT

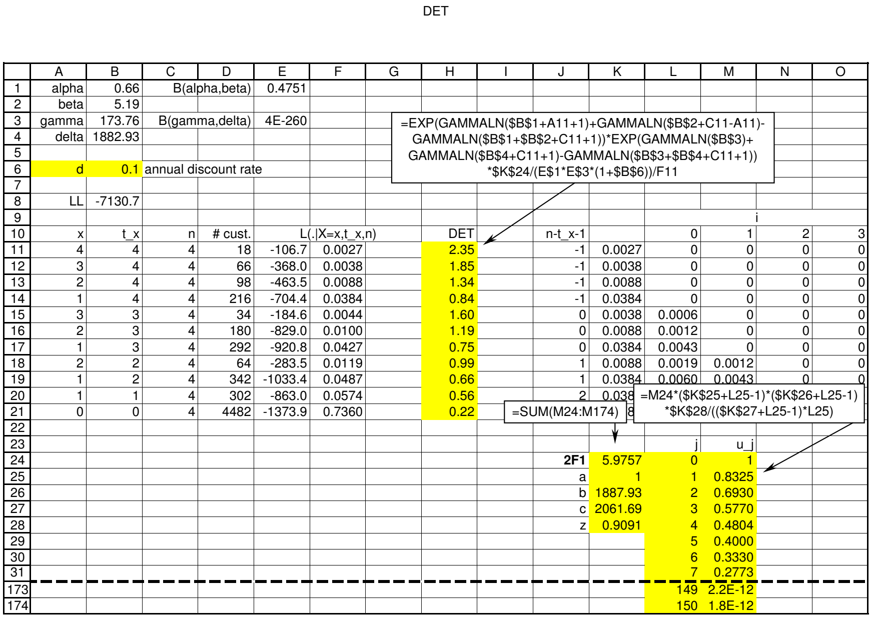
- For a customer with purchase history (x, t_x, n),

$$\begin{aligned} &DERT(d \mid p, \theta, \text{alive at } n) \\ &\quad = \sum_{t=n+1}^{\infty} \frac{P(Y_t = 1 \mid p, \text{alive at } t)P(\text{alive at } t \mid t > n, \theta)}{(1+d)^{t-n}} \\ &\quad = \frac{p(1-\theta)}{d+\theta} \end{aligned}$$

- However,
  - p and θ are unobserved
  - We do not know whether the customer is alive at n

Computing DERT

$$\begin{aligned}
&DERT(d \mid \alpha, \beta, \gamma, \delta, x, t_x, n) \\
&\quad = \int_0^1 \int_0^1 \Big\{ DERT(d \mid p, \theta, \text{alive at } n) \\
&\qquad\qquad \times P(\text{alive at } n \mid p, \theta, x, t_x, n) \\
&\qquad\qquad\quad \times g(p, \theta \mid \alpha, \beta, \gamma, \delta, x, t_x, n) \Big\} dp\, d\theta \\
&\quad = \frac{B(\alpha + x + 1, \beta + n - x) B(\gamma, \delta + n + 1)}{B(\alpha, \beta) B(\gamma, \delta)(1 + d)} \\
&\qquad\quad \times \frac{{}_2F_1(1, \delta + n + 1; \gamma + \delta + n + 1; \frac{1}{1+d})}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}
\end{aligned}$$



DET

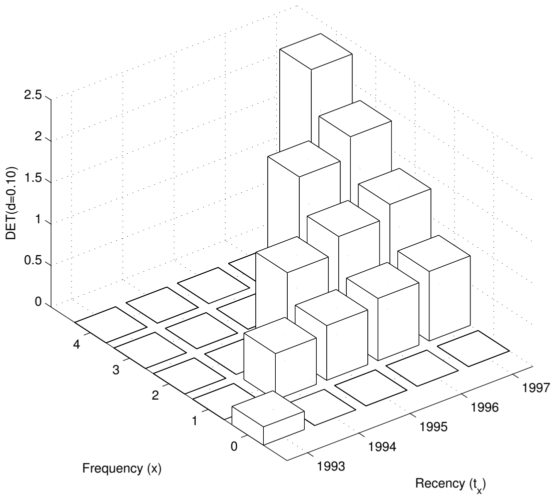
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	alpha	0.66		B(alpha,beta)	0.4751										
2	beta	5.19													
3	gamma	173.76		B(gamma,delta)	4E-260										
4	delta	1882.93													
5															
6	d	0.1	annual discount rate												
7															
8	LL	-7130.7													
9													i		
10	x	t_x	n	# cust.	L(.|X=x,t_x,n)			DET		n-t_x-1		0	1	2	3
11	4	4	4	18	-106.7	0.0027		2.35		-1	0.0027	0	0	0	0
12	3	4	4	66	-368.0	0.0038		1.85		-1	0.0038	0	0	0	0
13	2	4	4	98	-463.5	0.0088		1.34		-1	0.0088	0	0	0	0
14	1	4	4	216	-704.4	0.0384		0.84		-1	0.0384	0	0	0	0
15	3	3	4	34	-184.6	0.0044		1.60		0	0.0038	0.0006	0	0	0
16	2	3	4	180	-829.0	0.0100		1.19		0	0.0088	0.0012	0	0	0
17	1	3	4	292	-920.8	0.0427		0.75		0	0.0384	0.0043	0	0	0
18	2	2	4	64	-283.5	0.0119		0.99		1	0.0088	0.0019	0.0012	0	0
19	1	2	4	342	-1033.4	0.0487		0.66		1	0.0384	0.0060	0.0043	0	0
20	1	1	4	302	-863.0	0.0574		0.56		2	0.038				
21	0	0	4	4482	-1373.9	0.7360		0.22			8				
22															
23												j	u_j		
24										**2F1**	5.9757	0	1		
25										a	1	1	0.8325		
26										b	1887.93	2	0.6930		
27										c	2061.69	3	0.5770		
28										z	0.9091	4	0.4804		
29												5	0.4000		
30												6	0.3330		
31												7	0.2773		
173												149	2.2E-12		
174												150	1.8E-12		

=EXP(GAMMALN($B$1+A11+1)+GAMMALN($B$2+C11-A11)-
GAMMALN($B$1+$B$2+C11+1))*EXP(GAMMALN($B$3)+
GAMMALN($B$4+C11+1)-GAMMALN($B$3+$B$4+C11+1))
*$K$24/(E$1*E$3*(1+$B$6))/F11

=SUM(M24:M174)

=M24*($K$25+L25-1)*($K$26+L25-1)
*$K$28/(($K$27+L25-1)*L25)

DERT as a Function of R & F ($d = 0.10$)



DERT as a Function of R & F ($d = 0.10$)

	Year of Last Cruise				
# Cruise-years	1997	1996	1995	1994	1993
4	2.35				
3	1.85	1.60			
2	1.34	1.19	0.99		
1	0.84	0.75	0.66	0.56	
0					0.22

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Paul D. Berger (2004), "Customer-Base Analysis with Discrete-Time Transaction Data." `<http://brucehardie.com/papers/020/>`

Fader, Peter S., Bruce G. S. Hardie, and Paul D. Berger (2005), "Implementing the BG/BB Model for Customer-Base Analysis in Excel." `<http://brucehardie.com/notes/010/>`

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)

Cohort	Calendar Time →				
1	n_{11}	n_{12}	n_{13}	$\ldots$	n_{1I}
2		n_{22}	n_{23}	$\ldots$	n_{2I}
3			n_{33}	$\ldots$	n_{3I}
$\vdots$				$\ddots$	$\vdots$
I					n_{II}
	$n_{.1}$	$n_{.2}$	$n_{.3}$	$\ldots$	$n_{.I}$

Implications of Data Recording Processes (Contractual Settings)

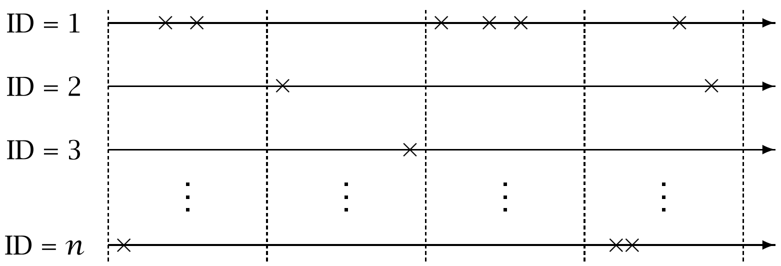
Cohort	Calendar Time →				
1	n_{11}				n_{1I}
2		n_{22}			n_{2I}
⋮					⋮
I−1				$n_{I-1,I-1}$	$n_{I-1,I}$
I					n_{II}

Cohort	Calendar Time →				
1	n_{11}				
2		n_{22}			
⋮					
I−1				$n_{I-1,I-1}$	
I					n_{II}
	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.I-1}$	$n_{.I}$

Cohort	Calendar Time →				
1					n_{1I}
2					n_{2I}
⋮					⋮
I−1					$n_{I-1,I}$
I					n_{II}
	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.I-1}$	$n_{.I}$

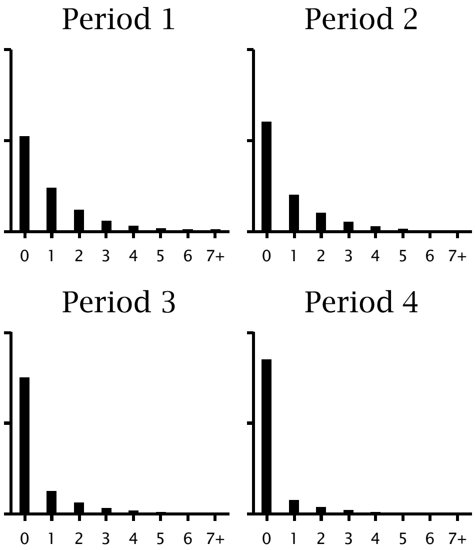
Cohort	Calendar Time →				
1				n_{1I-1}	n_{1I}
2				n_{2I-1}	n_{2I}
⋮				⋮	⋮
I−1				$n_{I-1,I-1}$	$n_{I-1,I}$
I					n_{II}

Implications of Data Recording Processes (Noncontractual Settings)



Implications of Data Recording Processes (Noncontractual Settings)

	Period			
ID	1	2	3	4
1	2	0	3	1
2	0	1	0	1
3	0	1	0	0
⋮				
n	1	0	0	2



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)

  Fader, Peter S., Bruce G. S. Hardie, and Kinshuk Jerath (2007), "Estimating CLV Using Aggregated Data: The *Tuscan Lifestyles* Case Revisited ." *Journal of Interactive Marketing*, **21** (Summer), 55–71.

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

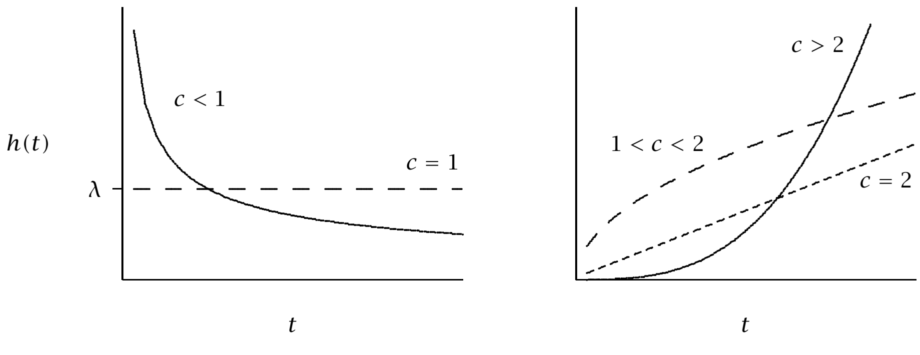
- A generalization of the exponential distribution that can have an increasing and decreasing hazard function:

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

  where c is the "shape" parameter and λ 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 λ 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.

Individual-Level Duration Dependence

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

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

- Assuming heterogeneity in θ follows a beta distribution with parameters (α, β), we arrive at the beta-discrete-Weibull (BdW) distribution with survivor function:

$$\begin{aligned} S(t \mid \alpha, \beta, c) &= \int_0^1 S(t \mid \theta, c) g(\theta \mid \alpha, \beta)\, d\theta \\ &= \frac{B(\alpha, \beta + t^c)}{B(\alpha, \beta)} . \end{aligned}$$

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

  Netzer, Oded, James Lattin, and V. Srinivasan (2008, "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, **27** (March–April), 185–204.

  or a (dynamic) changepoint model

  Fader, Peter S., Bruce G. S. Hardie, and Chun-Yao Huang (2004), "A Dynamic Changepoint Model for New Product Sales Forecasting," *Marketing Science*, **23** (Winter), 50–65.

Covariates

- Types of covariates:
  - customer characteristics
  - customer attitudes and behavior
  - marketing activities
- Handling covariate effects:
  - explicit integration (via latent characteristics and hazard functions)

    Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "Understanding Service Retention Within and Across Cohorts Using Limited Information," *Journal of Marketing*, **72** (January), 82–94.
  - 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.

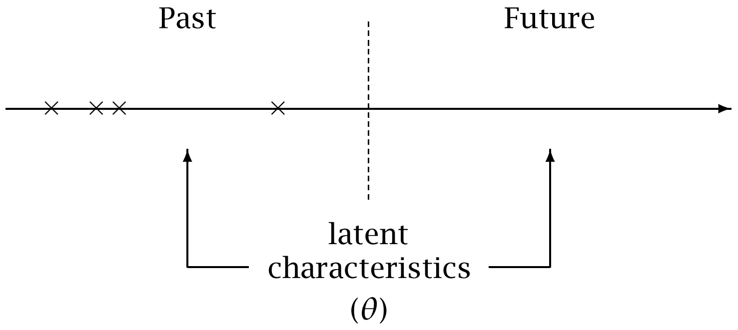
Philosophy of Model Building

- 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

$\hat{\theta} = f(\text{past}) \longrightarrow \text{future} = f(\hat{\theta})$

Classifying Customer Bases

Opportunities for Transactions

	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship

Type of Relationship With Customers