

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

23rd Annual Advanced Research Techniques Forum
June 24-27, 2012

©2012 Peter S. Fader and Bruce G. S. Hardie

1

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

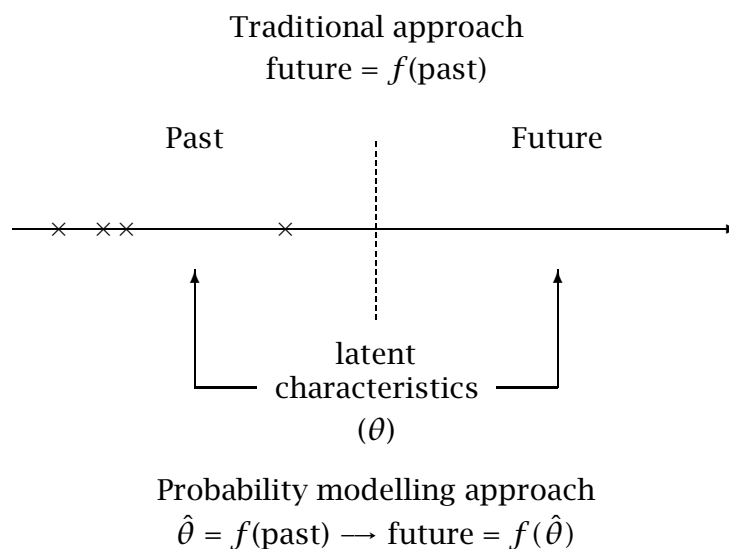
2

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.

3

Comparison of Modelling Approaches



4

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.

5

Classifying Analysis Settings

- It is important to distinguish between contractual and noncontractual settings:
 - In a *contractual* setting, we observe the time at which a customer ended their relationship with the firm.
 - In a *noncontractual* setting, the time at which a customer "dies" 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?

6

Classifying Analysis Settings

Consider the following four specific business settings:

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

7

Classifying Customer Bases

Opportunities for Transactions	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
		Noncontractual	Contractual
Type of Relationship With Customers			

8

The Right Way to Think About Computing Customer Lifetime Value

9

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 alive)

r = retention rate

d = discount rate

T = horizon for calculation

11

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 alive)

$S(t)$ = the probability that the customer is alive beyond time t

$d(t)$ = discount factor that reflects the present value of money received at time t

12

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

13

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

$$E(RLV) = \int_T^{\infty} E[v(t)]S(t | t > T)d(t - T)dt$$

14

Review of Probability Models

15

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.

16

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*.

17

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:

$$f(x) = \int f(x|\theta)g(\theta) d\theta$$

18

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.

19

“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.

20

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

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

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

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

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 θ is the probability of success.

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

23

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

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

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

24

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 θ , where

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

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?”

25

Capturing Heterogeneity in Latent Traits

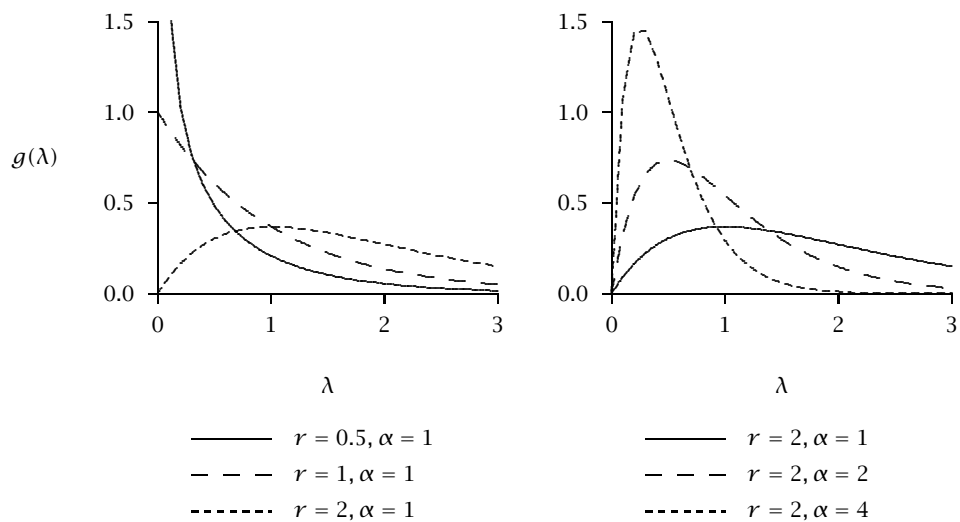
The gamma distribution:

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

- $\Gamma(\cdot)$ 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.

26

Illustrative Gamma Density Functions



27

Capturing Heterogeneity in Latent Traits

The beta distribution:

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

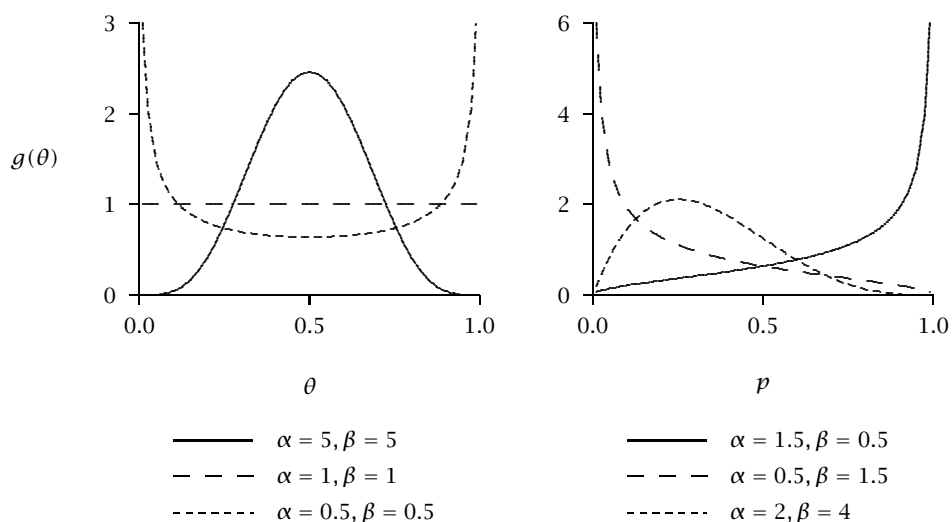
- $B(\alpha, \beta)$ is the beta function, which can be expressed in terms of gamma functions:

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

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

28

Illustrative Beta Density Functions



29

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.

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

30

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 λ , then $X(t)$ has a Poisson distribution with mean λt .
- Assuming λ is distributed across the population according to a gamma distribution,

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

31

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.

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

32

The Shifted-Beta-Geometric Model

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

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

33

The Beta-Binomial Distribution

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

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

34

Summary of Probability Models

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

35

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)

36

A Template for Integrated Models

		Stage 2		
		Counting	Timing	Choice
Stage 1	Counting			
	Timing			
	Choice			

37

Integrated Models

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

$$f(x | \theta_1, \theta_2) = f(f_1(x_1 | \theta_1), f_2(x_2 | \theta_2)).$$

- Solving the integral

$$f(x) = \iint f(x | \theta_1, \theta_2) g_1(\theta_1) g_2(\theta_2) d\theta_1 d\theta_2$$

often results in an intermediate result of the form

$$= \text{constant} \times \int_0^1 t^a (1-t)^b (u+vt)^{-c} dt$$

38

The “Trick” for Integrated Models

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

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

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

39

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

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

where $u_0 = 1$.

40

Models for Contractual Settings

41

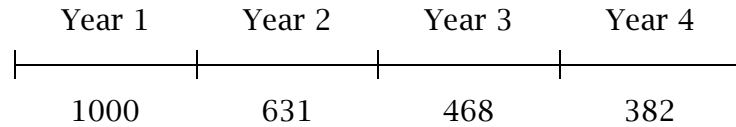
Classifying Customer Bases

Opportunities for Transactions	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
		Noncontractual	Contractual
Type of Relationship With Customers			

42

Illustrative Contractual Setting

1000 customers are acquired at the beginning of Year 1 with the following pattern of renewals:



Assume:

- Each contract is annual, starting on January 1 and expiring at 11:59pm on December 31
- An average net cashflow of \$100/year, which is “booked” at the beginning of the contract period
- A 10% discount rate

43

Motivating Questions

- What is the maximum amount you would spend to acquire a customer?
- What is the expected residual value of this group of customers at the end of Year 4?

44

Spending on Customer Acquisition

	Year 1	Year 2	Year 3	Year 4
Net CF	\$100	\$100	\$100	\$100
<i>P</i> (alive)	1.000	0.631	0.468	0.382
discount	1.000	$(1 + 0.1)$	$(1 + 0.1)^2$	$(1 + 0.1)^3$

45

Spending on Customer Acquisition

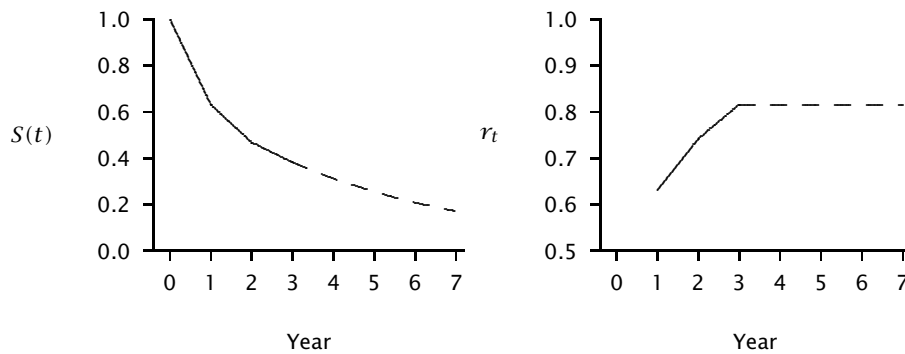
For a randomly chosen customer,

$$\begin{aligned}
 E(\text{CLV}) &= \$100 + \$100 \times \frac{0.631}{1.1} \\
 &\quad + \$100 \times \frac{0.468}{1.21} + \$100 \times \frac{0.382}{1.331} \\
 &= \$225
 \end{aligned}$$

⇒ We can justify spending up to \$225 to acquire a new customer (based on expected “profitability” over the four year period).

46

Looking to the Future



The retention rate for period t (r_t) is the proportion of customers who were alive in period t who renewed their contract at the next opportunity.

$$r_t = S(t)/S(t - 1) \iff S(t) = S(t - 1) \times r_t$$

47

Spending on Customer Acquisition

$$\begin{aligned}
 E(\text{CLV}) &= \$100 + \$100 \times \frac{0.631}{1.1} \\
 &\quad + \$100 \times \frac{0.468}{(1.1)^2} + \$100 \times \frac{0.382}{(1.1)^3} \\
 &\quad + \$100 \times \frac{0.382 \times 0.816}{(1.1)^4} \\
 &\quad + \$100 \times \frac{0.382 \times 0.816^2}{(1.1)^5} + \dots \\
 &= \$307
 \end{aligned}$$

\Rightarrow Looking beyond Year 4, we can justify spending up to \$82 more to acquire a customer.

48

Residual Value of the Customer Base

	Year 4	Year 5	Year 6	Year 7	
Net CF	\$100	\$100	\$100	\$100	...
<i>P</i> (alive)	0.816	0.816 ²	0.816 ³	0.816 ⁴	...
discount	1.000	(1 + 0.1)	(1 + 0.1) ²	(1 + 0.1) ³	...

49

Residual Value of the Customer Base

For a randomly chosen customer,

$$\begin{aligned}
 E(RLV) &= \$100 \times 0.816 + \$100 \times \frac{0.816^2}{(1 + 0.1)} + \dots \\
 &= \$100 \times \sum_{t=1}^{\infty} \frac{0.816^t}{(1 + 0.1)^{t-1}} \\
 &= \$316
 \end{aligned}$$

⇒ The expected residual value of the group of customers at the end of Year 4 is $382 \times \$316 = \$120,712$.

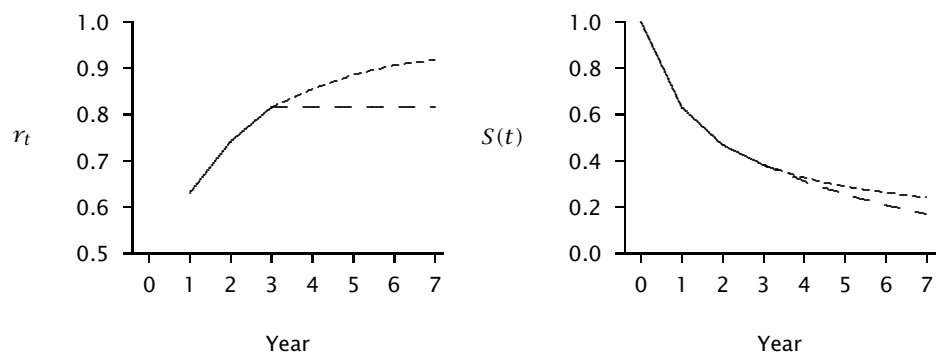
50

What's wrong with this analysis?

51

Projecting Retention

How valid is the assumption of a constant retention rate beyond the observed data?



At the cohort level, we (almost) always observe increasing retention rates (and a flattening survival curve).

52

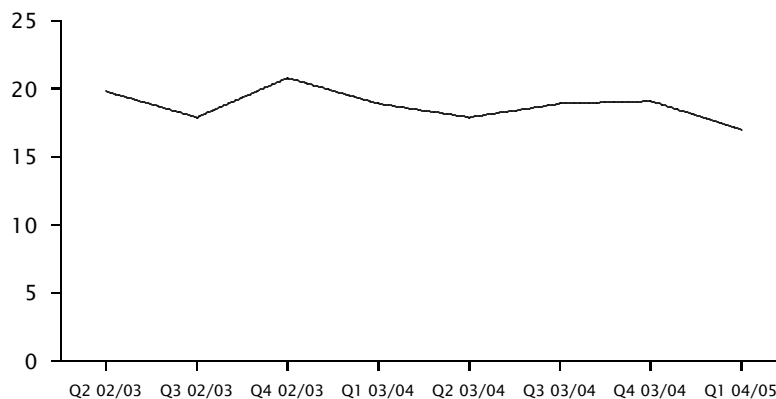
Renewal rates at regional magazines vary; generally 30% of subscribers renew at the end of their original subscription, but that figure jumps to 50% for second-time renewals and all the way to 75% for longtime readers.

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 subscribers, and therefore, an increase in subscriber age helps overall reductions in churn.

Netflix (10-K for the fiscal year ended December 31, 2005)

Vodafone Germany Quarterly Annualized Churn Rate (%)

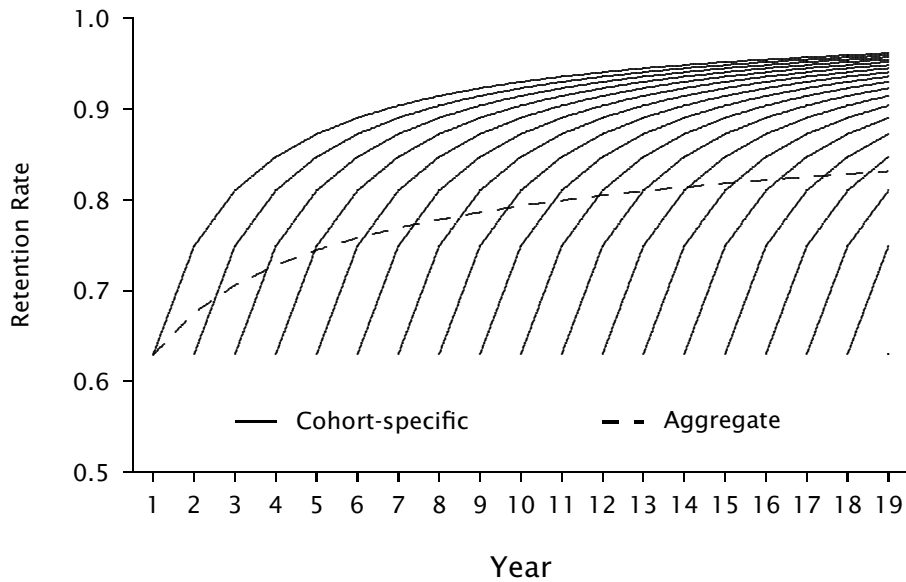


Source: Vodafone Germany "Vodafone Analyst & Investor Day" presentation (2004-09-27)

Cohort-level vs. Aggregate Numbers

	Yr 01	Yr 02	Yr 03	Yr 04	Yr 05	Yr 06	Yr 07	Yr 08	Yr 09	Yr 10	Yr 11	Yr 12	Yr 13	Yr 14	Yr 15	Yr 16	Yr 17	Yr 18	Yr 19	Yr 20
Yr 01	1000	629	471	382	324	283	252	228	208	192	179	167	157	148	140	133	127	122	117	112
Yr 02		1000	629	471	382	324	283	252	228	208	192	179	167	157	148	140	133	127	122	117
Yr 03			1000	629	471	382	324	283	252	228	208	192	179	167	157	148	140	133	127	122
Yr 04				1000	629	471	382	324	283	252	228	208	192	179	167	157	148	140	133	127
Yr 05					1000	629	471	382	324	283	252	228	208	192	179	167	157	148	140	133
Yr 06						1000	629	471	382	324	283	252	228	208	192	179	167	157	148	140
Yr 07							1000	629	471	382	324	283	252	228	208	192	179	167	157	148
Yr 08								1000	629	471	382	324	283	252	228	208	192	179	167	157
Yr 09									1000	629	471	382	324	283	252	228	208	192	179	167
Yr 10										1000	629	471	382	324	283	252	228	208	192	179
Yr 11											1000	629	471	382	324	283	252	228	208	192
Yr 12												1000	629	471	382	324	283	252	228	208
Yr 13													1000	629	471	382	324	283	252	228
Yr 14														1000	629	471	382	324	283	252
Yr 15															1000	629	471	382	324	283
Yr 16																1000	629	471	382	324
Yr 17																	1000	629	471	382
Yr 18																		1000	629	471
Yr 19																			1000	629
Yr 20																				1000
Total	1000	1629	2100	2482	2806	3089	3341	3569	3777	3969	4148	4315	4472	4620	4760	4893	5020	5142	5259	5371

Cohort-level vs. Aggregate Numbers



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.

57

A Discrete-Time Model for Contract Duration

- 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:

$$S(t | \theta) = (1 - \theta)^t, \quad t = 1, 2, 3, \dots$$

- Heterogeneity in θ is captured by a beta distribution with pdf

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

58

A Discrete-Time Model for Contract Duration

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

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

- The aggregate survivor function is

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

59

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.

60

A Discrete-Time Model for Contract Duration

We can compute sBG probabilities using the following forward-recursion formula from $P(T = 1)$:

$$P(T = t | \alpha, \beta) = \begin{cases} \frac{\alpha}{\alpha + \beta} & t = 1 \\ \frac{\beta + t - 2}{\alpha + \beta + t - 1} P(T = t - 1) & t = 2, 3, \dots \end{cases}$$

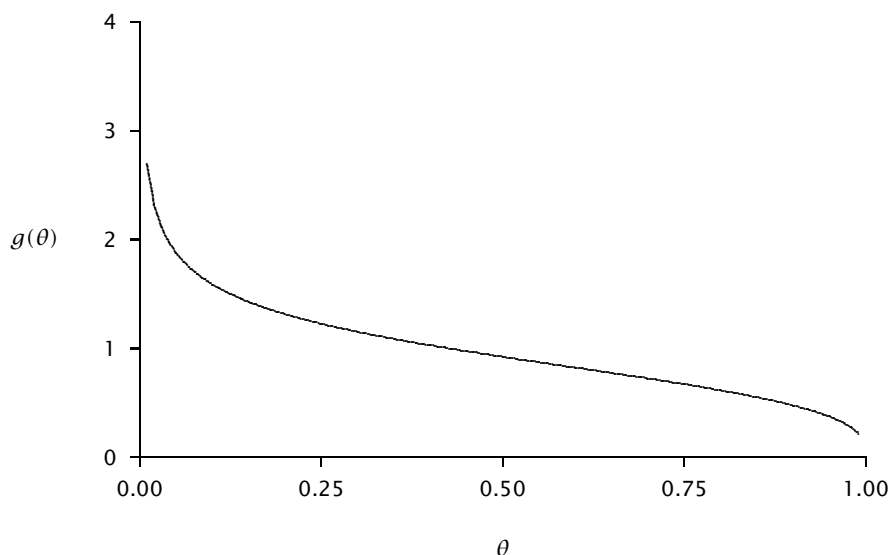
We can compute the sBG survivor function using the retention rates:

$$S(t | \alpha, \beta) = \prod_{i=1}^t r_i, t = 1, 2, 3, \dots, \text{ where } S(0) = 1.$$

Fitting the sBG Model

	A	B	C	D	E
1	alpha	0.784			
2	beta	1.333			
3	LL	-1242.31			
4					
5	Year	P(T=t)	# Cust.	# Lost	
6	0		1000		
7	1	0.3702	631	369	-366.64
8	2	0.1584	468	163	-300.39
9	3	0.0897	382	86	-207.33
10					-367.95

Distribution of Churn Probabilities



63

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}),

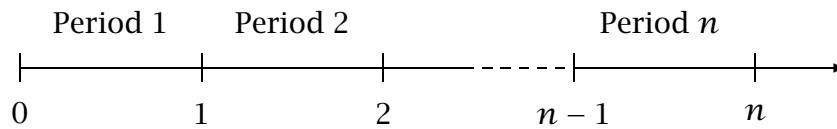
$$E(CLV) = \bar{v} \underbrace{\int_0^{\infty} S(t)d(t)dt}_{\text{discounted expected lifetime}} .$$

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

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

64

Computing DEL



- Standing at time 0 (i.e., before the customer is acquired),

$$\begin{aligned}
 DEL(d | \theta) &= \sum_{t=0}^{\infty} \frac{S(t)}{(1+d)^t} \\
 &= \frac{1+d}{d+\theta}.
 \end{aligned}$$

- But θ is unobserved ...

65

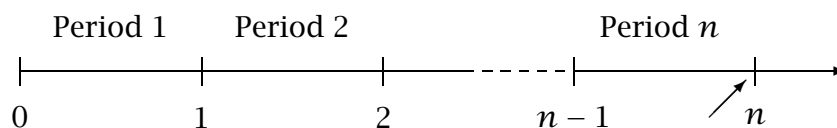
Computing DEL

Integrating over the distribution of θ :

$$\begin{aligned}
 DEL(d | \alpha, \beta) &= \int_0^1 DEL(d | \theta) g(\theta | \alpha, \beta) d\theta \\
 &= \int_0^1 \left(\frac{1+d}{d+\theta} \right) \frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{B(\alpha, \beta)} d\theta \\
 &= \frac{1+d}{B(\alpha, \beta)} \int_0^1 \theta^{\alpha-1} (1-\theta)^{\beta-1} (d+\theta)^{-1} d\theta \\
 &= {}_2F_1\left(1, \beta; \alpha + \beta; \frac{1}{1+d}\right).
 \end{aligned}$$

66

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,

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

- But θ is unobserved

67

Computing DERL

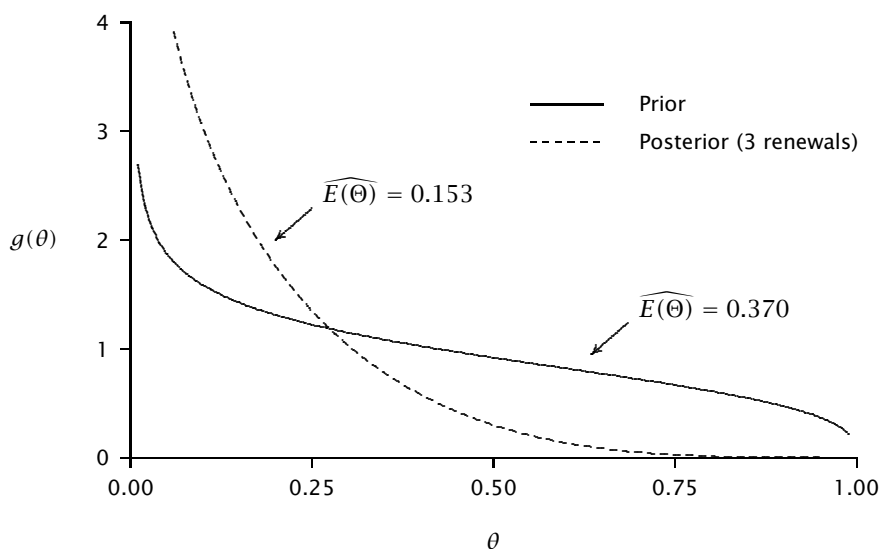
By Bayes' Theorem, the posterior distribution of θ is

$$\begin{aligned}
 g(\theta | \alpha, \beta, n - 1 \text{ renewals}) &= \frac{S(n - 1 | \theta)g(\theta | \alpha, \beta)}{S(n - 1 | \alpha, \beta)} \\
 &= \frac{\theta^{\alpha-1}(1 - \theta)^{\beta+n-2}}{B(\alpha, \beta + n - 1)},
 \end{aligned}$$

which is a beta distribution with parameters α and $\beta + n - 1$.

68

Distributions of Churn Probabilities



69

Computing DERL

Integrating over the posterior distribution of θ :

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

$$\begin{aligned}
 &= \int_0^1 \left\{ \text{DERL}(d \mid \theta, n - 1 \text{ renewals}) \right. \\
 &\quad \left. \times g(\theta \mid \alpha, \beta, n - 1 \text{ renewals}) \right\} d\theta \\
 &= \int_0^1 \frac{(1 - \theta)(1 + d)}{d + \theta} \frac{\theta^{\alpha-1}(1 - \theta)^{\beta+n-2}}{B(\alpha, \beta + n - 1)} d\theta \\
 &= \frac{1 + d}{B(\alpha, \beta + n - 1)} \int_0^1 \theta^{\alpha-1}(1 - \theta)^{\beta+n-1}(d + \theta)^{-1} d\theta \\
 &= \left(\frac{\beta + n - 1}{\alpha + \beta + n - 1} \right) {}_2F_1\left(1, \beta + n; \alpha + \beta + n; \frac{1}{1+d}\right)
 \end{aligned}$$

70

Alternative Derivation of Expression for DEL

$$\begin{aligned}
 DEL(d | \alpha, \beta) &= \sum_{t=0}^{\infty} \frac{S(t | \alpha, \beta)}{(1+d)^t} \\
 &= \sum_{t=0}^{\infty} \frac{B(\alpha, \beta + t)}{B(\alpha, \beta)} \left(\frac{1}{1+d}\right)^t \\
 &= {}_2F_1\left(1, \beta; \alpha + \beta; \frac{1}{1+d}\right).
 \end{aligned}$$

71

Alternative Derivation of Expression for DERL

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

72

Computing DE(R)L Using Excel

- From the alternative derivations:

$$DEL(d | \alpha, \beta) = \sum_{t=0}^{\infty} \frac{S(t | \alpha, \beta)}{(1 + d)^t}$$

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

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

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

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

73

Computing DE(R)L Using Excel

	A	B	C	D	E	F	G	H
1	alpha	0.784		DEL	3.597		DERL	5.171
2	beta	1.333		=SUMPRODUCT(C6:C206,E6:E206)				
3	d	10%						
4							3 renewals (n=4)	
5	t	r_t	S(t)		disc.		S(t t>n-1)	disc.
6	0		1.0000		1.0000			
7	1	0.6298	0.6298		0.9091		=1/(1+\$B\$3)^A7	
8	2	0.7485	0.4714		0.8264			
9		=(B\$2+A7-1)/(B\$1+B\$2+A7-1)			0.7513			
10	4	0.8468	0.3232		=C10/\$C\$9		0.8468	1.0000
11	5	0.8719	0.2818		0.6209		0.7383	0.9091
12	6	0.8899	0.2508		0.5645		0.6570	0.8264
13	7	0.9034	0.2265		0.5132		0.5936	0.7513
14	8	0.9140	0.2074		0.4665		0.5426	0.6830
15	9	0.9225	0.19	=C12*B13			0.5005	0.6209
16	10	0.9295	0.1776		0.3855		0.4652	0.5645
202	196	0.9960	0.0188		7.71E-09		0.0492	1.13E-08
203	197	0.9960	0.0187		7.01E-09		0.0490	1.03E-08
204	198	0.9961	0.0186		6.37E-09		0.0488	9.33E-09
205	199	0.9961	0.0186		5.79E-09		0.0486	8.48E-09
206	200	0.9961	0.0185		5.27E-09		0.0485	7.71E-09

74

Comparing Approaches

	Model	Naïve	Underestimation
$E(CLV)$	\$360	\$307	15%
$E(RLV)$	\$517	\$316	64%

- The naïve estimates will always be lower than those of the sBG model.
- The driving factor is the degree of heterogeneity — see Fader and Hardie (2010).
- The error is especially problematic when computing $E(RLV)$ (and therefore when valuing a customer base).

75

Validating the sBG-based CLV Estimates

We actually have 12 years of renewal data.

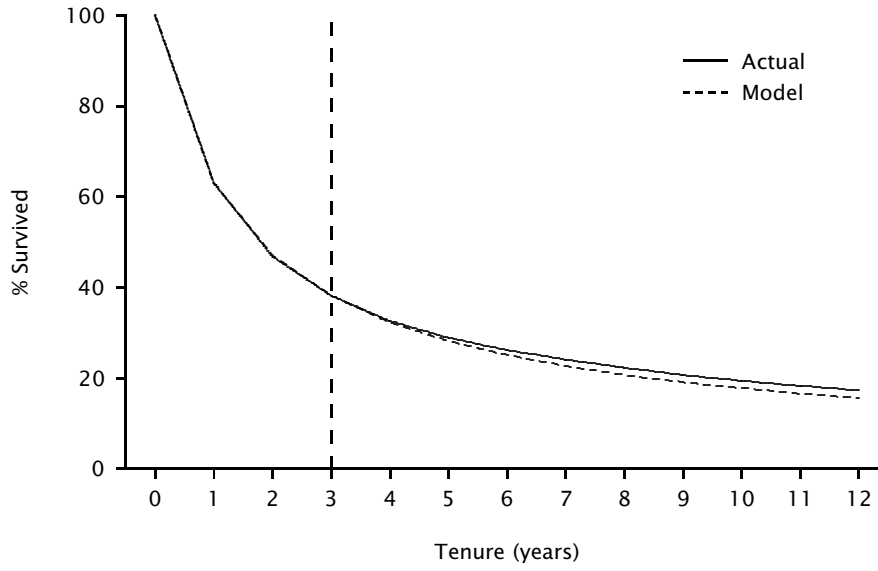
- Examine the predictive performance of the underlying sBG model.
- Compare the naïve and model-based estimates of expected “lifetime” value against the actual average values.

$$E(CLV) = \$100 \times \sum_{t=0}^{12} \frac{S(t)}{(1.1)^t}$$

$$R(CLV) = \$100 \times \sum_{t=4}^{12} \frac{S(t|t > 3)}{(1.1)^{t-4}}$$

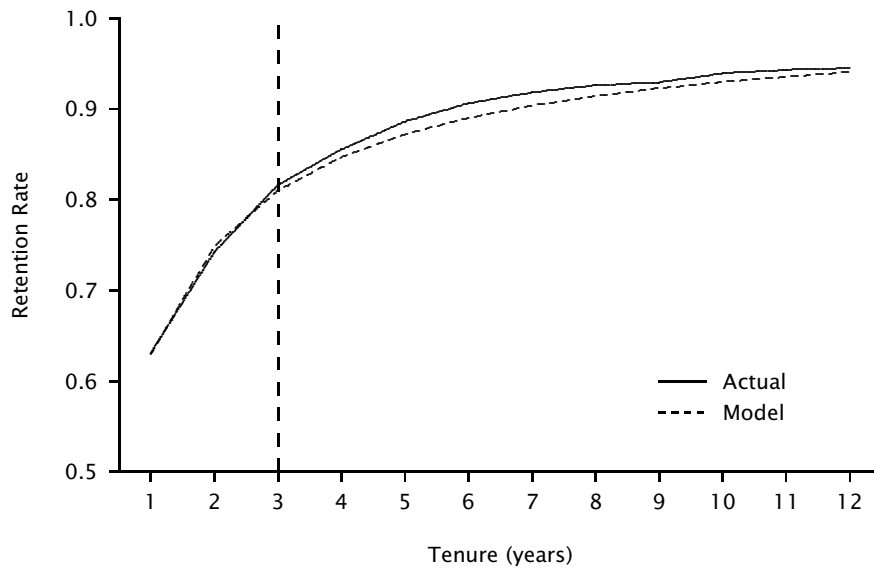
76

Survival Curve Projection



77

Projecting Retention Rates



78

Comparing Approaches

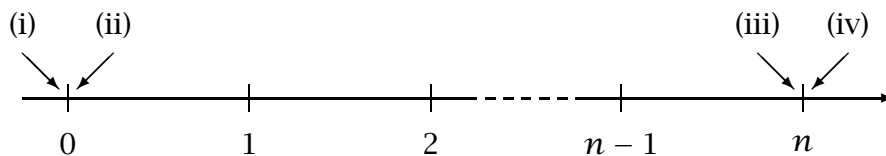
	Infinite Horizon		12 Year Horizon		
	Model	Naïve	Actual	Model	Naïve
$E(CLV)$	\$360	\$307	\$331	\$326	\$302
$E(RLV)$	\$517	\$316	\$407	\$387	\$295

- The model-based estimates are very close to the actual numbers ... while making use of only three renewal observations!
- The undervaluation associated with the naïve model becomes increasingly severe over a longer time horizon.

79

Expressions for DE(R)L

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



80

Expressions for DE(R)L

Case (i):

$$DEL(d | \alpha, \beta) = {}_2F_1(1, \beta; \alpha + \beta; \frac{1}{1+d})$$

Case (ii):

$$\begin{aligned} DERL(d | \alpha, \beta) \\ = \frac{\beta}{(\alpha + \beta)(1 + d)} {}_2F_1(1, \beta + 1; \alpha + \beta + 1; \frac{1}{1+d}) \end{aligned}$$

81

Expressions for DE(R)L

Case (iii):

$$\begin{aligned} DERL(d | \alpha, \beta, \text{ alive for } n \text{ periods}) \\ = \frac{\beta + n - 1}{\alpha + \beta + n - 1} {}_2F_1(1, \beta + n; \alpha + \beta + n; \frac{1}{1+d}) \end{aligned}$$

Case (iv):

$$\begin{aligned} DERL(d | \alpha, \beta, n \text{ contract renewals}) &= \frac{\beta + n}{(\alpha + \beta + n)(1 + d)} \\ &\times {}_2F_1(1, \beta + n + 1; \alpha + \beta + n + 1; \frac{1}{1+d}) \end{aligned}$$

82

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 (2010), “Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity,” *Marketing Science*, 29 (January–February), 85–93. <<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

Opportunities for Transactions	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
		Noncontractual	Contractual
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,

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

- ii. Heterogeneity in λ follows a gamma distribution with pdf

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

Contract Duration in Continuous-Time

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

$$f(t | r, \alpha) = \int_0^{\infty} f(t | \lambda) g(\lambda | r, \alpha) d\lambda$$
$$= \frac{r}{\alpha} \left(\frac{\alpha}{\alpha + t} \right)^{r+1}$$

$$S(t | r, \alpha) = \int_0^{\infty} S(t | \lambda) g(\lambda | r, \alpha) d\lambda$$
$$= \left(\frac{\alpha}{\alpha + t} \right)^r$$

The Hazard Function

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

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

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

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

87

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.

88

Computing DERL

- Standing at time s ,

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

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

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

- But λ is unobserved

89

Computing DERL

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

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

90

Computing DERL

It follows that

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

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

Models for Noncontractual Settings

Classifying Customer Bases

Opportunities for Transactions	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
		Noncontractual	Contractual
Type of Relationship With Customers			

93

Setting

- A charity located in the Midwestern United States that is funded in large part by donations from individual supporters.
- 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?

94

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
100001	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
100004	1	0	1	0	1	1	1	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
100008	1	1	1	1	1	1	1	?	?	?	?	?
100009	1	1	1	1	1	1	0	?	?	?	?	?
100010	1	0	0	0	0	0	0	?	?	?	?	?
⋮			⋮			⋮			⋮			⋮
111102	1	1	1	1	1	1	1	?	?	?	?	?
111103	1	0	1	1	0	1	1	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

Modelling the Transaction Stream

- Each year a listener decides whether or not to support the charity 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 :

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

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

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

97

Modelling the Transaction Stream

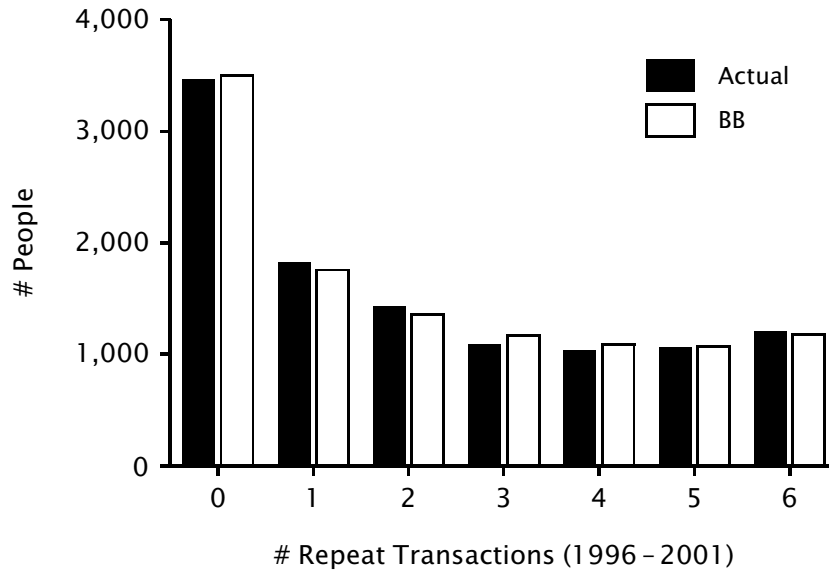
The distribution of transactions for a randomly chosen individual is given by

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

which is the beta-binomial (BB) distribution.

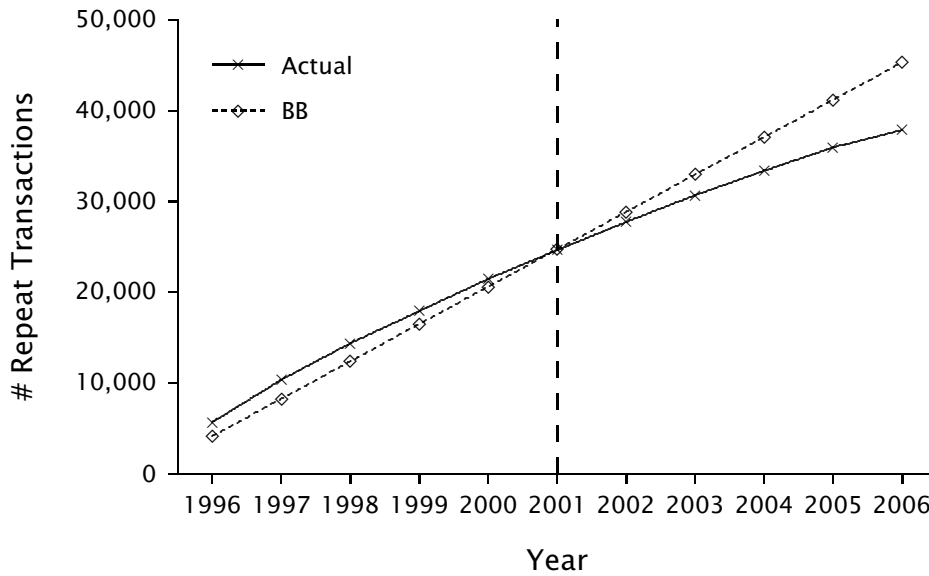
98

Fit of the BB Model



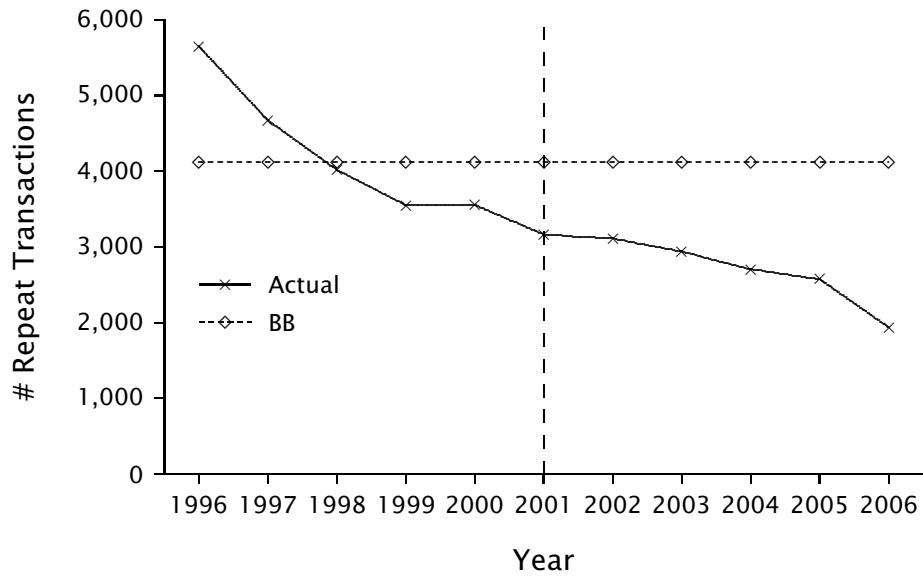
99

Tracking Cumulative Repeat Transactions



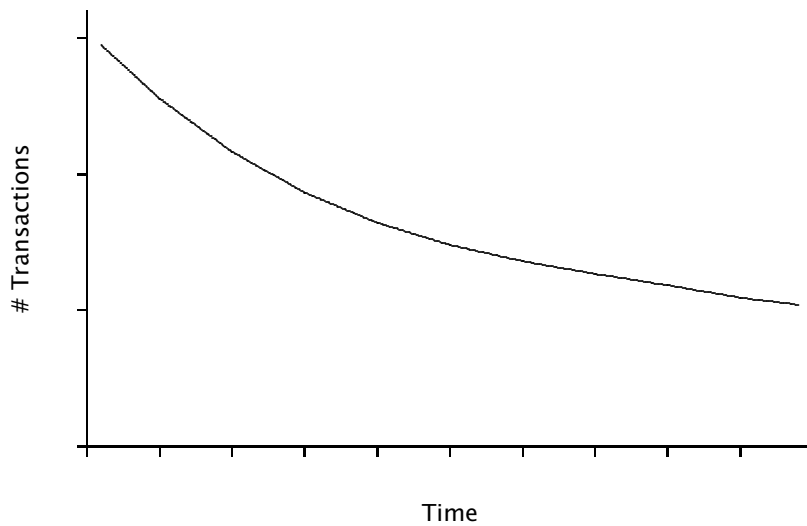
100

Tracking Annual Repeat Transactions



101

Towards a More Realistic Model



102

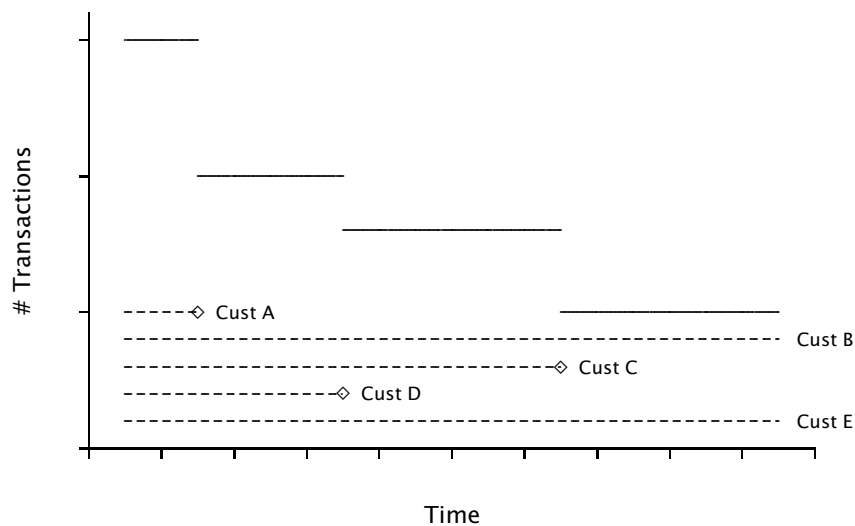
Towards a More Realistic Model

The “leaky bucket” phenomenon:

A harsh reality for any marketer is that regardless of how wonderful their product or service is, or how creative their marketing activities are, the customer base of any company can be viewed as a leaky bucket whose contents are continually dripping away. Customer needs and tastes change as their personal circumstances change over time, which leads them to stop purchasing from a given firm or even stop buying in the product category all together. In the end, they literally die.

103

Towards a More Realistic Model



104

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,” which is a function of their dropout propensity
- Dropout propensities vary across customers

105

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 θ

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

106

Model Development

Consider the following transaction pattern:

1996	1997	1998	1999	2000	2001
1	0	0	1	0	0

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

1996	1997	1998	1999	2000	2001
A	A	A	A	D	D
A	A	A	A	A	D
A	A	A	A	A	A

107

Model Development

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

$$\begin{aligned}
 f(100100 | p, \theta) &= p(1-p)(1-p)p \underbrace{(1-\theta)^4 \theta}_{P(\text{AAAADD})} \\
 &\quad + p(1-p)(1-p)p(1-p) \underbrace{(1-\theta)^5 \theta}_{P(\text{AAAAAD})} \\
 &\quad + \underbrace{p(1-p)(1-p)p(1-p)(1-p)}_{P(Y_1=1, Y_2=0, Y_3=0, Y_4=1)} \underbrace{(1-\theta)^6}_{P(\text{AAAAAA})}
 \end{aligned}$$

108

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,

$$f(100100 | p, \theta) = f(001100 | p, \theta) = f(010100 | p, \theta)$$

- *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

Summarizing Repeat Transaction Behavior

	1996	1997	1998	1999	2000	2001	\longrightarrow	x	t_x	n	# Donors
1	1	1	1	1	1	1		6	6	6	1203
2	1	1	1	1	1	0		5	6	6	728
3	1	1	1	1	0	1		5	5	6	335
4	1	1	1	1	0	0		4	6	6	512
5	1	1	1	0	1	1		4	5	6	284
6	1	1	1	0	1	0		4	4	6	240
7	1	1	1	0	0	1		3	6	6	357
								3	5	6	225
								3	4	6	181
								3	3	6	322
								2	6	6	234
								2	5	6	173
								2	4	6	155
								2	3	6	255
								2	2	6	613
								1	6	6	129
								1	5	6	119
								1	4	6	79
								1	3	6	129
								1	2	6	277
62	0	0	0	0	1	0		1	1	6	1091
63	0	0	0	0	0	1		0	0	6	3464
64	0	0	0	0	0	0					11104

Model Development

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

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

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

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

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

111

Model Development

Removing the conditioning on the latent traits p and θ ,

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

$$= \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta + n)}{B(\gamma, \delta)} + \sum_{i=0}^{n-t_x-1} \frac{B(\alpha + x, \beta + t_x - x + i)}{B(\alpha, \beta)} \frac{B(\gamma + 1, \delta + t_x + i)}{B(\gamma, \delta)}$$

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

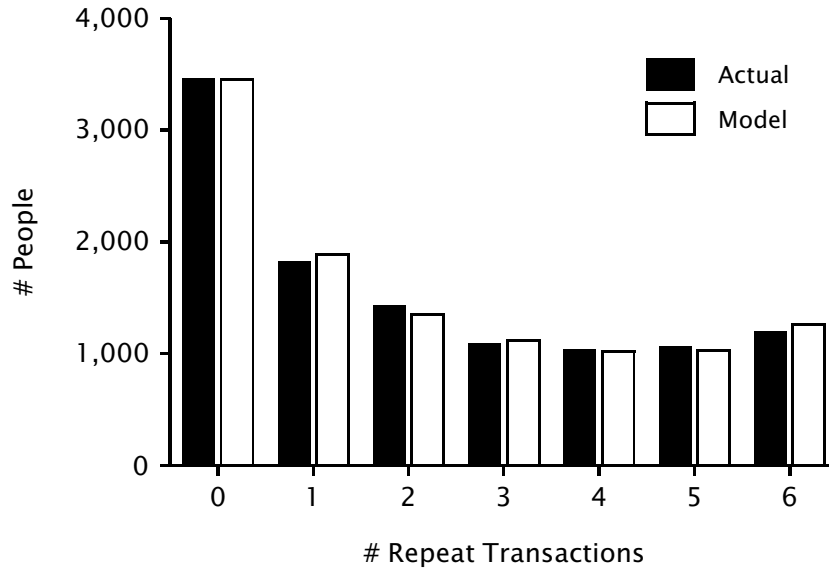
112

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	alpha	1.204	B(alpha,beta)		1.146	=EXP(GAMMALN(B1)+GAMMALN(B2)-GAMMALN(B1+B2))								
2	beta	0.750												
3	gamma	0.657	B(gamma,delta)		0.729									
4	delta	2.783												
5														
6	LL	-33225.6	=SUM(E9:E30)			=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								
7														
8	x	t_x	n	# donors	L(. X=x,t_x,n)	n-t_x-1			0	1	2	3	4	5
9	6	6	6	1203	-2624.6	0.1129	-1	0.1129	0	0	0	0	0	0
10	5	6	6	728	-3126.7	0.0136	-1	0.0136	0	0	0	0	0	0
11	4	6	6	512					=IF(\$B-\$G9,EXP(GAMMALN(\$B\$1+A9)+GAMMALN(\$B\$2+\$B9-\$A9+\$B8)-GAMMALN(\$B\$1+\$B\$2+\$B9+\$B8))/E\$1*EXP(GAMMALN(\$B\$3+1)+GAMMALN(\$B\$4+\$B9+\$B8)-GAMMALN(\$B\$3+\$B\$4+\$B9+\$B8+1))/E\$3,0)					
12	3	6	6	357					0	0	0	0	0	0
13	2	6	6	234	-1322.5	0.0035	-1	0.0035	0	0	0	0	0	0
14	1	6	6	129	-630.0	0.0076	-1	0.0076	0	0	0	0	0	0
15	5	5	6	335	-124	=C15-B15-1		0	0.0136	0.0107	0	0	0	0
16	4	5	6	284	-1447.1	0.0061		0	0.0046	0.0015	0	0	0	0
17	3	5	6		=D19*LN(F19)	63.5	0.0036	0	0.0030	0.0006	0	0	0	0
18	2	5	6	173	-52.6	0.0041	0	0.0035	0.0005	0	0	0	0	0
19	1	5	6	119	-567.3	0.0085			=SUM(H19:N19)					
20	4	4	6	240	-923.6	0.0213	1	0.0046	0.0152	0.0015	0	0	0	0
21	3	4	6	181	-915.7	0.0063	1	0.0030	0.0027	0.0006	0	0	0	0
22	2	4	6	155	-805.3	0.0055	1	0.0035	0.0015	0.0005	0	0	0	0
23	1	4	6	78	-356.5	0.0104	1	0.0076	0.0018	0.0009	0	0	0	0
24	3	3	6	322	-1135.8	0.0294	2	0.0030	0.0230	0.0027	0.0006	0	0	0
25	2	3	6	255	-1151.6	0.0109	2	0.0035	0.0054	0.0015	0.0005	0	0	0
26	1	3	6	129	-545.0	0.0146	2	0.0076	0.0043	0.0018	0.0009	0	0	0
27	2	2	6	613	-1846.4	0.0492	3	0.0035	0.0383	0.0054	0.0015	0.0005	0	0
28	1	2	6	277	-993.9	0.0276	3	0.0076	0.0130	0.0043	0.0018	0.0009	0	0
29	1	1	6	1091	-2497.1	0.1014	4	0.0076	0.0737	0.0130	0.0043	0.0018	0.0009	0
30	0	0	6	3464	-4044.3	0.3111	5	0.0362	0.1909	0.0459	0.0189	0.0098	0.0058	0.0037

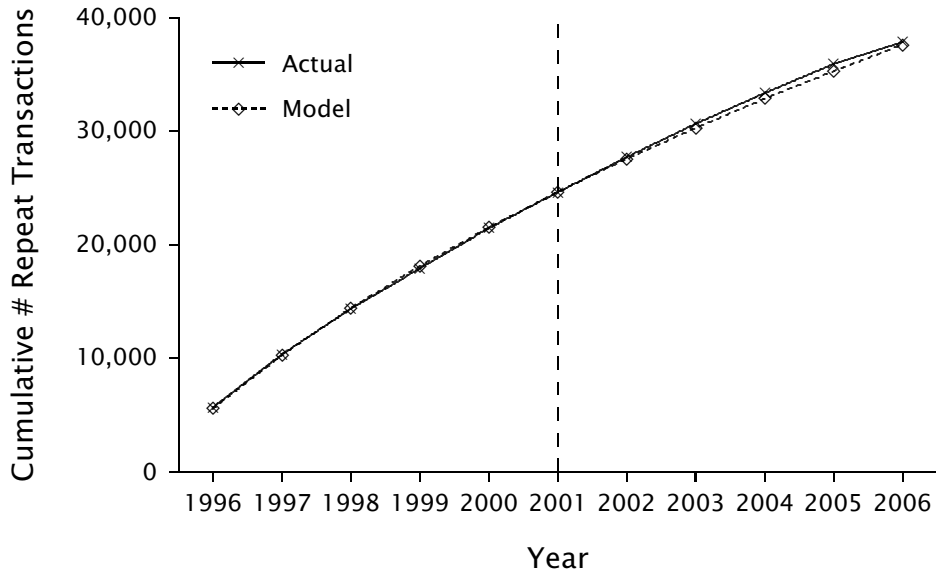
Parameter Estimates (1995 Cohort)

	α	β	γ	δ	LL
BB	0.487	0.826			-35,516.1
BG/BB	1.204	0.750	0.657	2.783	-33,225.6

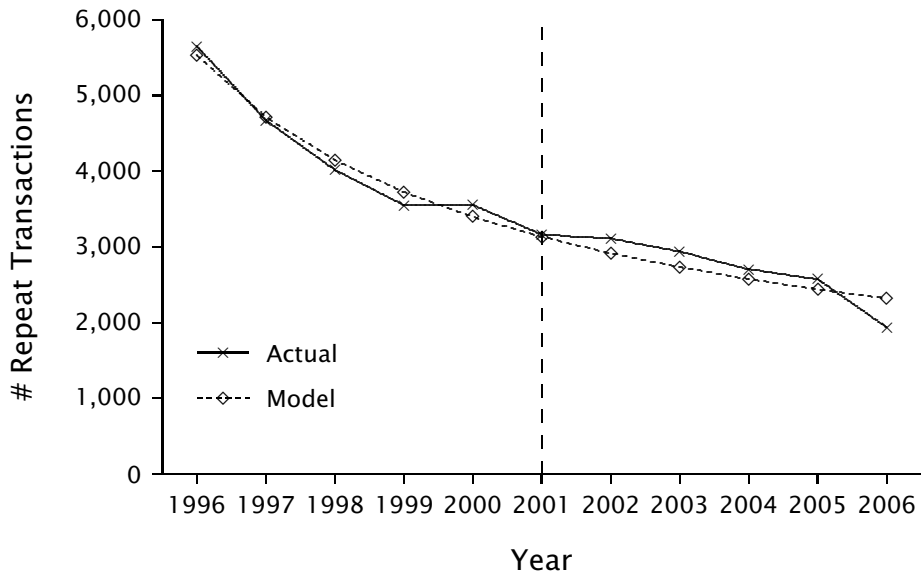
Fit of the BG/BB Model



Tracking Cumulative Repeat Transactions

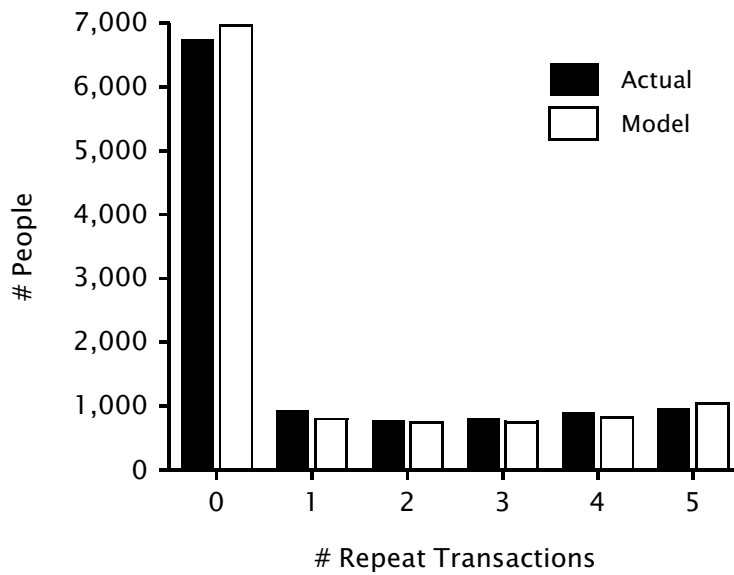


Tracking Annual Repeat Transactions



117

Repeat Transactions in 2002 - 2006



118

Key Results

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

- $P(\text{alive in period } n + 1 \mid x, t_x, n)$
The probability he will be “alive” in the next period.
- $P(X(n, n + n^*) = x^* \mid x, t_x, n)$
The probability he will make x^* transactions across the next n^* transaction opportunities.
- $E[X(n, n + n^*) \mid x, t_x, n]$
The expected number of transactions across the next n^* transaction opportunities.
- $DETR(d \mid x, t_x, n)$
The discounted expected residual transactions.

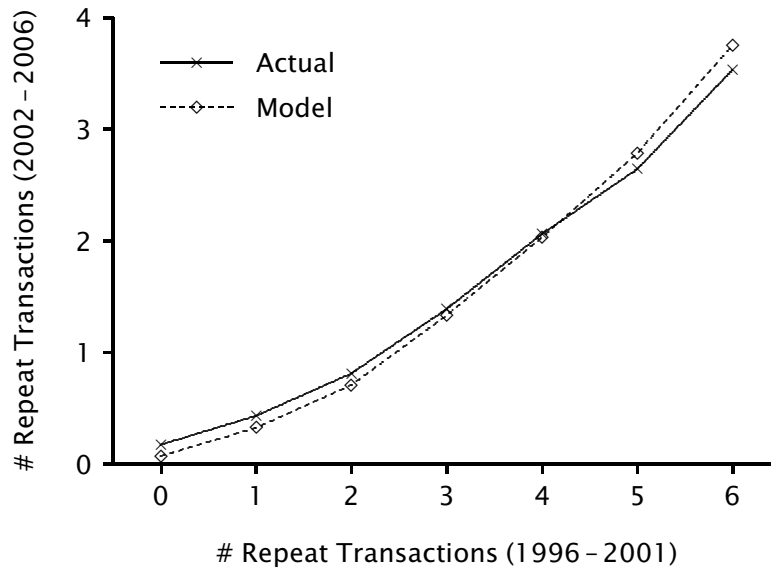
ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
100001	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
100004	1	0	1	0	1	1	1	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
100008	1	1	1	1	1	1	1	?	?	?	?	?
100009	1	1	1	1	1	1	0	?	?	?	?	?
100010	1	0	0	0	0	0	0	?	?	?	?	?
⋮			⋮			⋮			⋮			⋮
111102	1	1	1	1	1	1	1	?	?	?	?	?
111103	1	0	1	1	0	1	1	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

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

# Rpt Trans. (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.07						
1		0.09	0.31	0.59	0.84	1.02	1.15
2			0.12	0.54	1.06	1.44	1.67
3				0.22	1.03	1.80	2.19
4					0.58	2.03	2.71
5						1.81	3.23
6							3.75

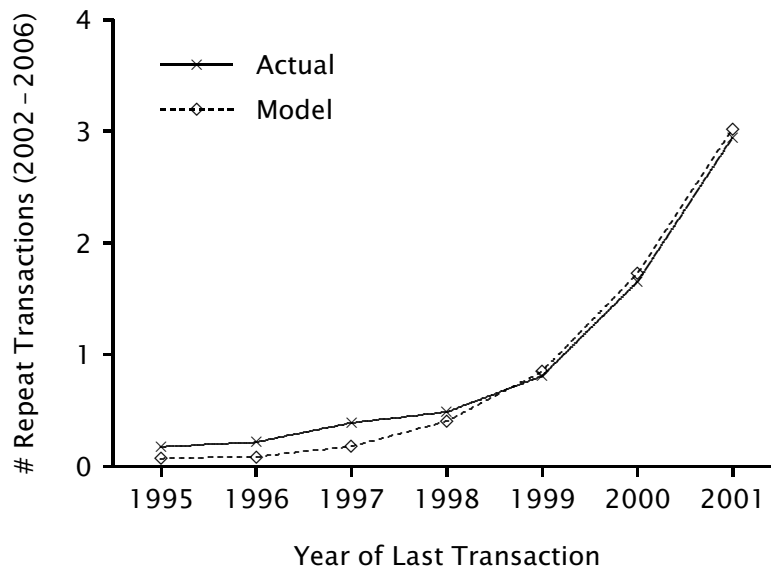
121

Conditional Expectations by Frequency



122

Conditional Expectations by Recency



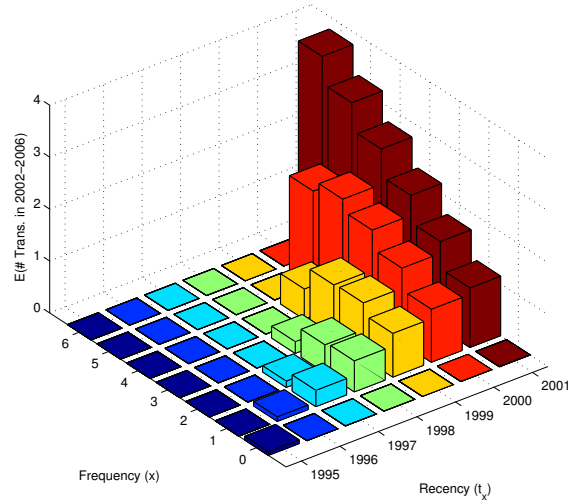
123

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

# Rpt Trans. (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.07						
1		0.09	0.31	0.59	0.84	1.02	1.15
2			0.12	0.54	1.06	1.44	1.67
3				0.22	1.03	1.80	2.19
4					0.58	2.03	2.71
5						1.81	3.23
6							3.75

124

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



125

P(alive in 2001) as a Function of Recency and Frequency

# Rpt Trans. (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.12						
1		0.07	0.27	0.52	0.73	0.89	1.00
2			0.07	0.32	0.63	0.86	1.00
3				0.10	0.47	0.82	1.00
4					0.22	0.75	1.00
5						0.56	1.00
6							1.00

126

Posterior Mean of P as a Function of Recency and Frequency

# Rpt Trans. (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.49						
1		0.66	0.44	0.34	0.30	0.28	0.28
2			0.75	0.54	0.44	0.41	0.40
3				0.80	0.61	0.54	0.53
4					0.82	0.68	0.65
5						0.83	0.78
6							0.91

127

Moving Beyond a Single Cohort

Cohort	Size
1995	11,104
1996	10,057
1997	9,043
1998	8,175
1999	8,977
2000	9,491

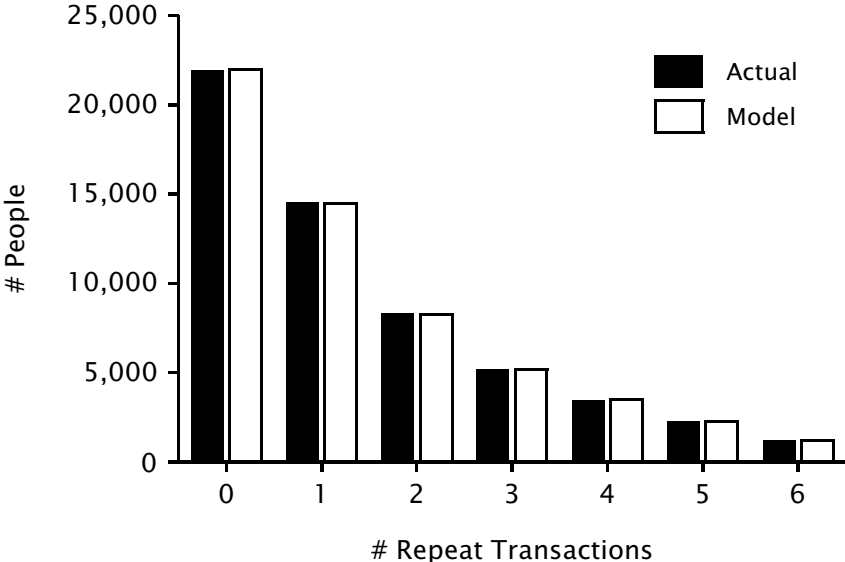
- Pooled calibration using the repeat transaction data for these 56,847 people across 1996 – 2001
- Hold-out validation period: 2002 – 2006

128

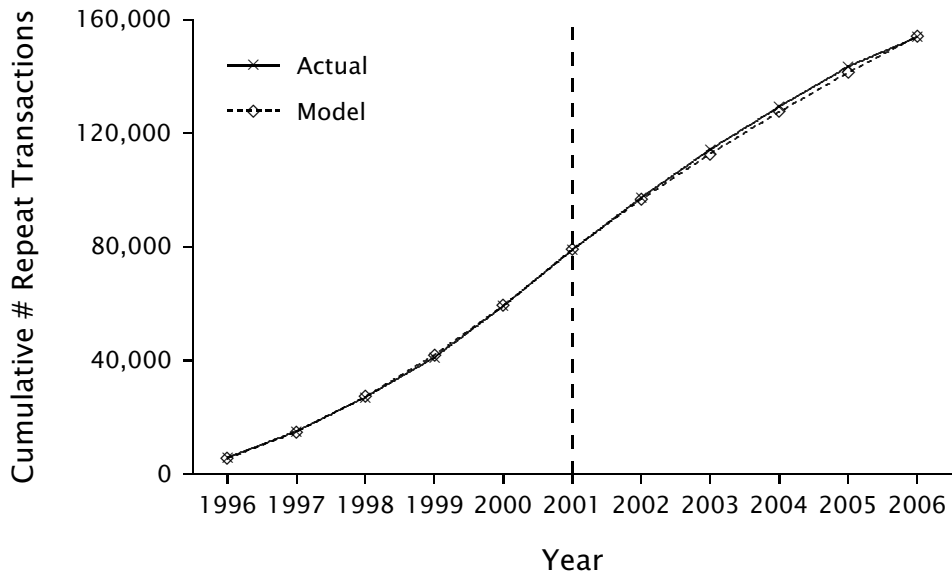
Parameter Estimates (Pooled)

	α	β	γ	δ	LL
BB	0.501	0.753			-115,615.0
BG/BB	1.188	0.749	0.626	2.331	-110,521.0

Fit of the BG/BB Model

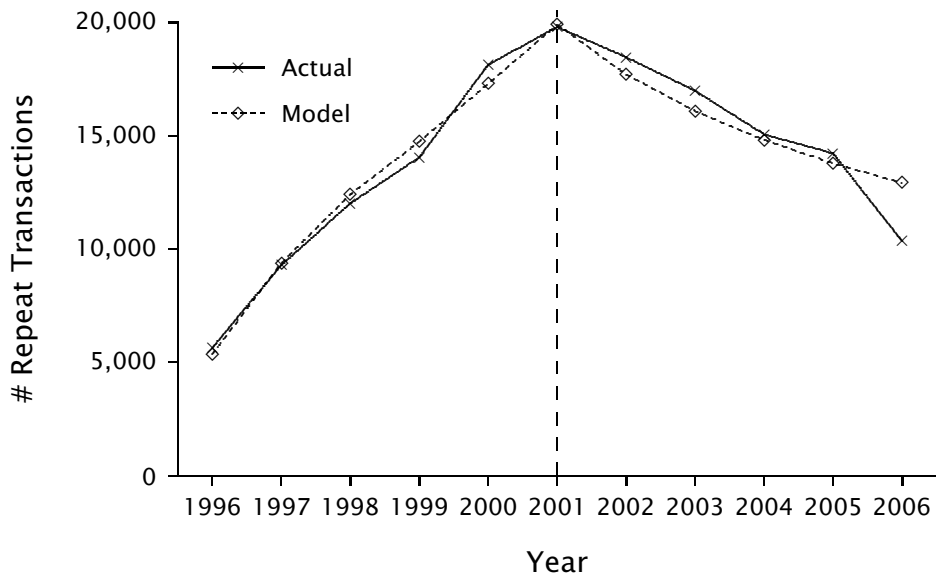


Tracking Cumulative Repeat Transactions



131

Tracking Annual Repeat Transactions



132

Computing $E(CLV)$

- Recall:

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

- 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

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

133

Computing $E(RLV)$

- Standing at time T ,

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

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

134

Computing DERT

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

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

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

135

Computing DERT

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

136

DERT as a Function of Recency and Frequency ($d = 0.10$)

# Rpt Trans. (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.11						
1		0.13	0.49	0.94	1.32	1.61	1.81
2			0.19	0.84	1.67	2.27	2.63
3				0.35	1.63	2.84	3.45
4					0.92	3.20	4.27
5						2.86	5.09
6							5.91

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," *Marketing Science*, **29** (November-December), 1086-1108. <<http://brucehardie.com/papers/020/>>

Fader, Peter S. and Bruce G. S. Hardie (2011), "Implementing the BG/BB Model for Customer-Base Analysis in Excel." <<http://brucehardie.com/notes/010/>>

Fader, Peter S. and Bruce G. S. Hardie (2011), "Implementing the S_{BB} -G/B Model in MATLAB." <<http://brucehardie.com/notes/023/>>

“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.

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

139

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?

140

From Discrete to Continuous Time

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

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

141

Classifying Customer Bases

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

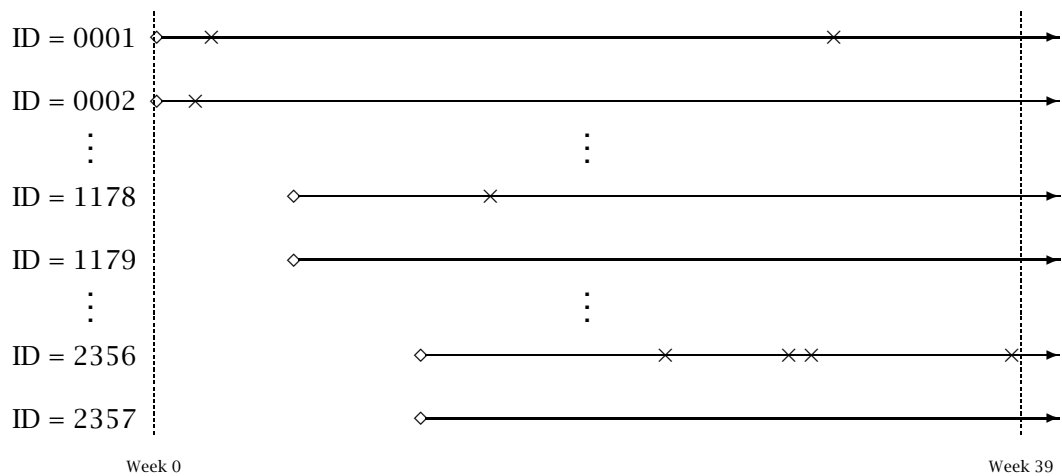
142

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

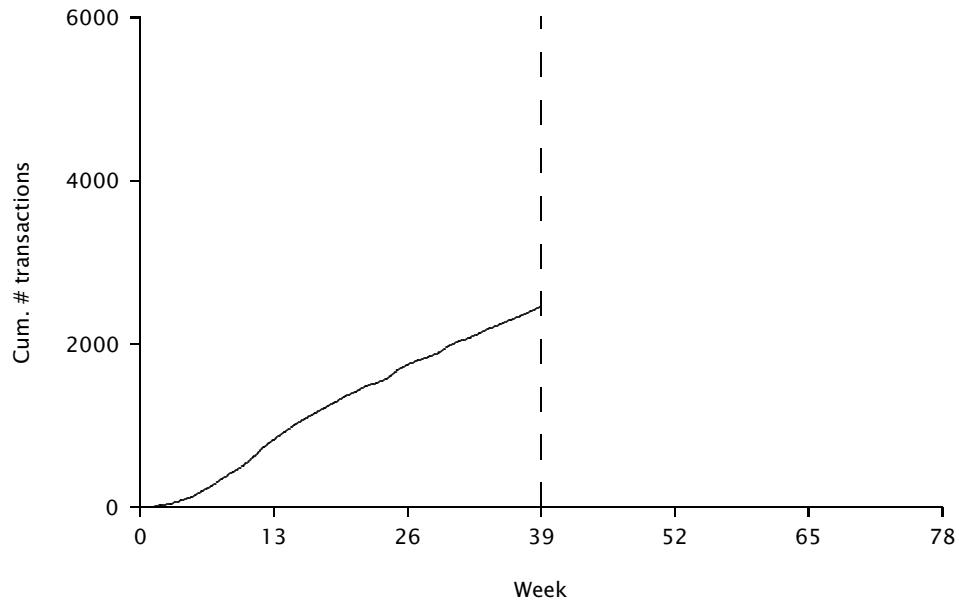
143

Purchase Histories



144

Cumulative Repeat Transactions



145

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.

146

Modelling the Transaction Stream

Transaction Process:

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

Dropout Process:

- Each customer has an unobserved “lifetime,” which is a function of their dropout propensity
- Dropout propensities vary across customers

147

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

Transaction Process:

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

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 $\text{gamma}(s, \beta)$.

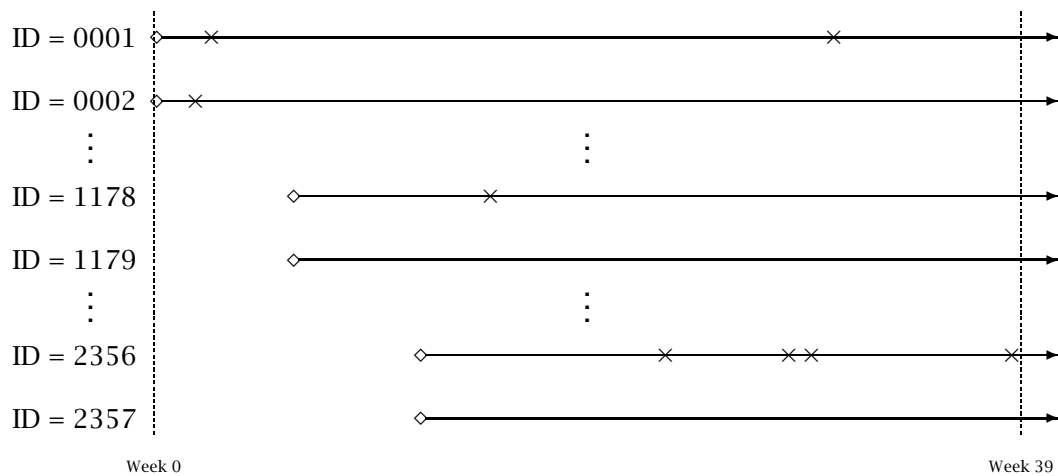
148

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.

149

Purchase Histories



150

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

$$L(r, \alpha, s, \beta | x, t_x, T)$$

$$= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)} \left\{ \left(\frac{s}{r+s+x} \right) \frac{{}_2F_1(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x})}{(\alpha+t_x)^{r+s+x}} \right. \\ \left. + \left(\frac{r+x}{r+s+x} \right) \frac{{}_2F_1(r+s+x, s; r+s+x+1; \frac{\alpha-\beta}{\alpha+T})}{(\alpha+T)^{r+s+x}} \right\}, \text{ if } \alpha \geq \beta$$

$$L(r, \alpha, s, \beta | x, t_x, T)$$

$$= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)} \left\{ \left(\frac{s}{r+s+x} \right) \frac{{}_2F_1(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x})}{(\beta+t_x)^{r+s+x}} \right. \\ \left. + \left(\frac{r+x}{r+s+x} \right) \frac{{}_2F_1(r+s+x, r+x+1; r+s+x+1; \frac{\beta-\alpha}{\beta+T})}{(\beta+T)^{r+s+x}} \right\}, \text{ if } \alpha \leq \beta$$

Key Results

$$E[X(t)]$$

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

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

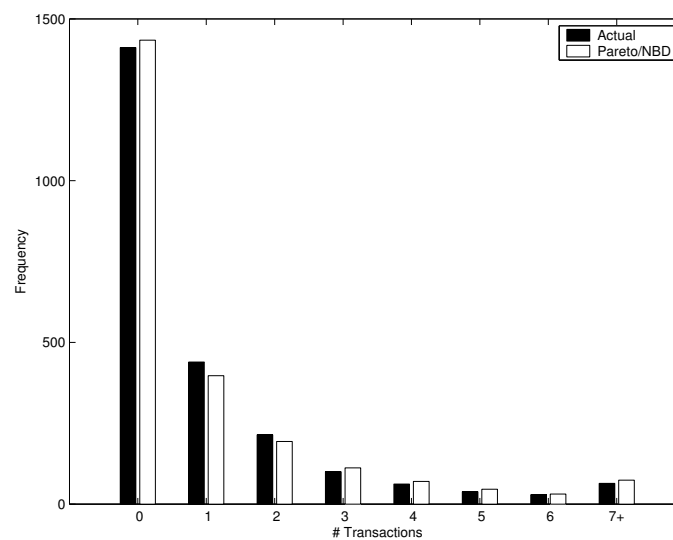
The probability that an individual with observed behavior (x, t_x, T) is still alive at time T .

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

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

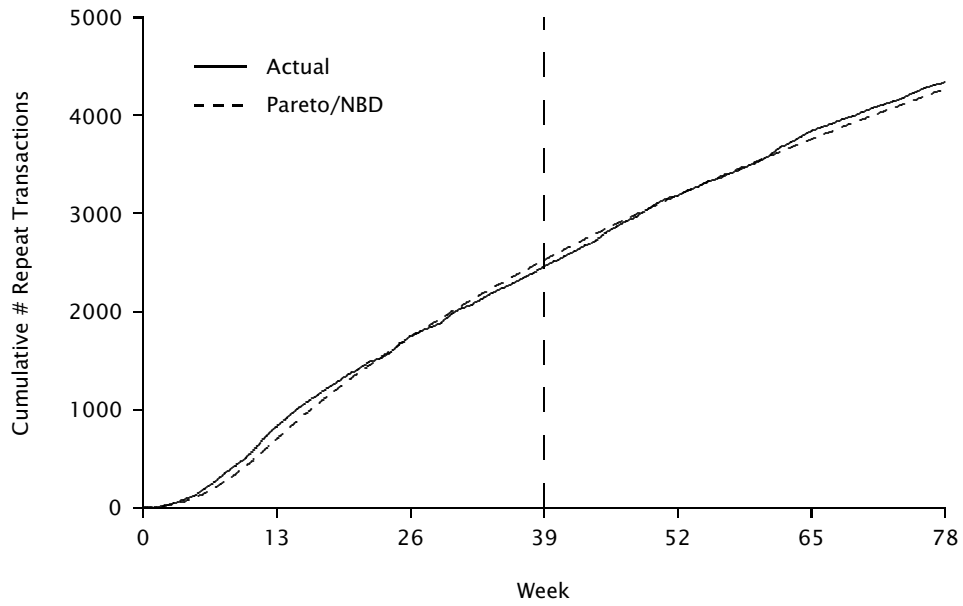
153

Frequency of Repeat Transactions



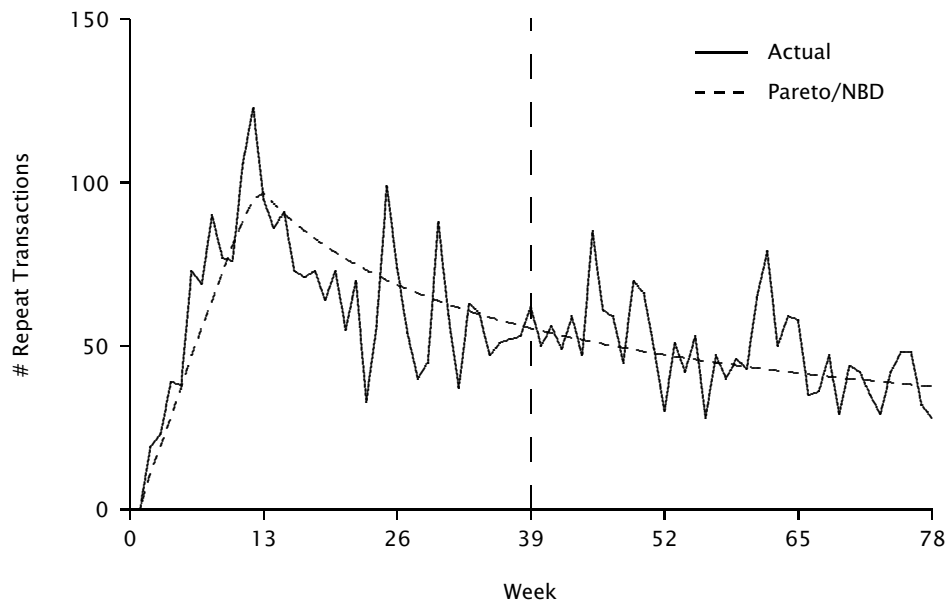
154

Tracking Cumulative Repeat Transactions



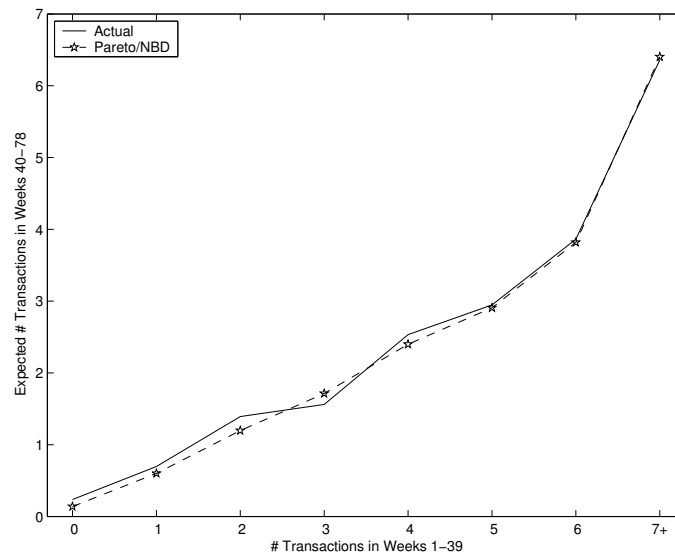
155

Tracking Weekly Repeat Transactions



156

Conditional Expectations



157

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

158

Computing DERT

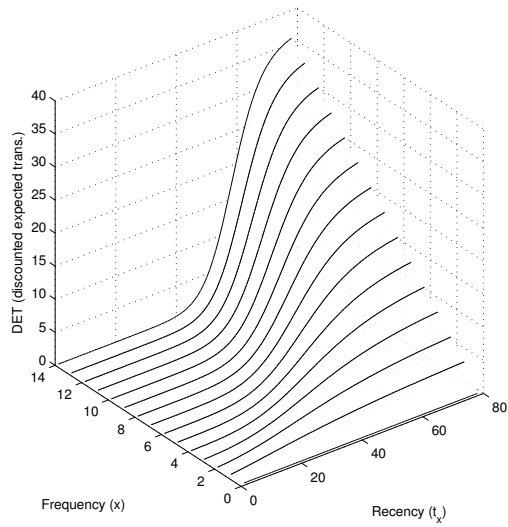
$$\begin{aligned}
 & DERT(\delta | r, \alpha, s, \beta, x, t_x, T) \\
 &= \int_0^\infty \int_0^\infty \left\{ DERT(\delta | \lambda, \mu, \text{alive at } T) \right. \\
 &\quad \times P(\text{alive at } T | \lambda, \mu, x, t_x, T) \\
 &\quad \left. \times g(\lambda, \mu | r, \alpha, s, \beta, x, t_x,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

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

Continuous Compounding

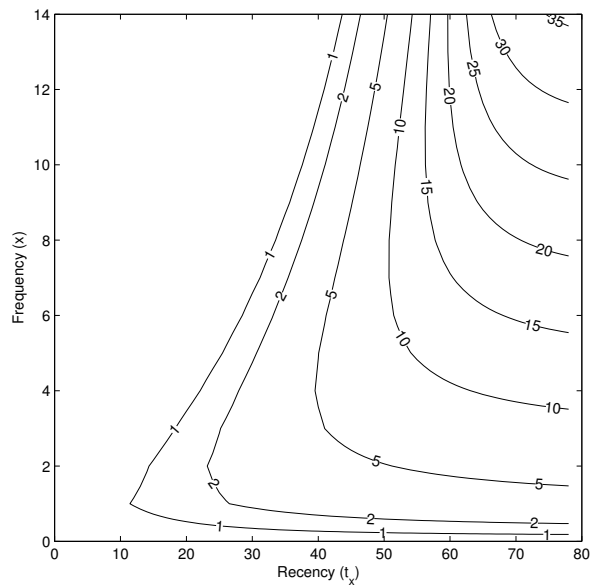
- An annual discount rate of $(100 \times d)\%$ is equivalent to a continuously compounded rate of $\delta = \ln(1 + d)$.
- If the data are recorded in time units such that there are k periods per year ($k = 52$ if the data are recorded in weekly units of time) then the relevant continuously compounded rate is $\delta = \ln(1 + d)/k$.

DERT by Recency and Frequency



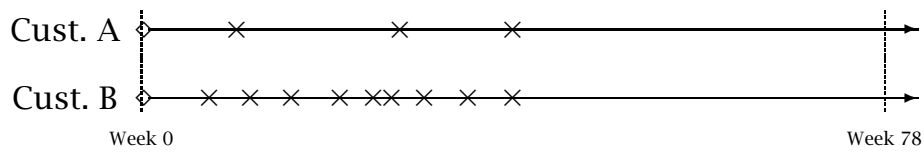
161

Iso-Value Representation of DERT



162

The “Increasing Frequency” Paradox



	DERT
Cust. A	4.6
Cust. B	1.9

163

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):

$$E(RLV) = f(R, F, M) = DERT \times g(F, M)$$

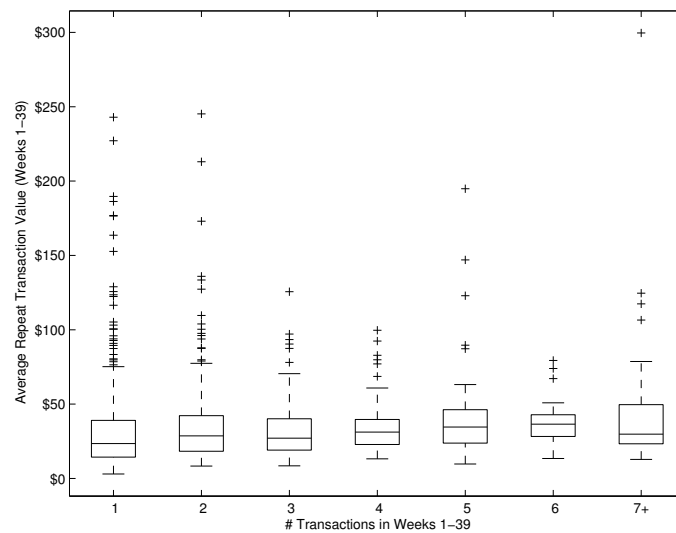
164

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.

165

Independence of the Spend Process



$$\text{corr}(\# \text{ transactions, avg. transaction value}) = 0.06$$

166

Modelling the Spend Process

- For a customer with x transactions, let z_1, z_2, \dots, 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 $E(M|m_x, x)$

167

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

	\$
Minimum	2.99
25th percentile	15.75
Median	27.50
75th percentile	41.80
Maximum	299.63
Mean	35.08
Std. deviation	30.28
Mode	14.96

168

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

169

Modelling the Spend Process

Marginal distribution for m_x :

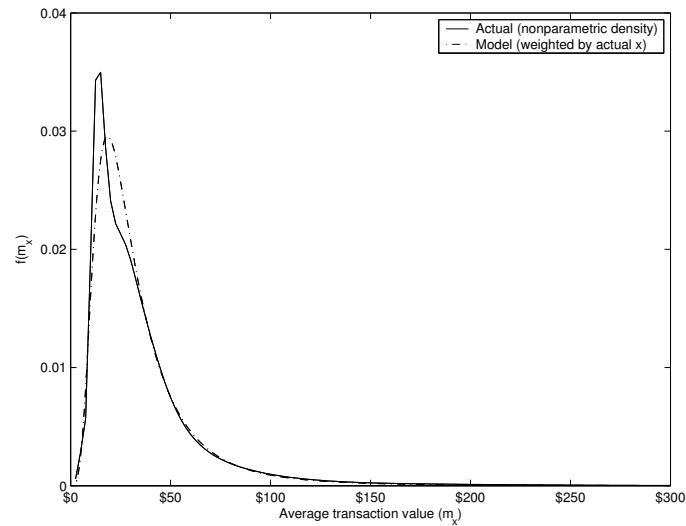
$$f(m_x | p, q, \gamma, x) = \frac{\Gamma(px + q)}{\Gamma(px)\Gamma(q)} \frac{\gamma^q m_x^{px-1} x^{px}}{(\gamma + m_x x)^{px+q}}$$

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

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

170

Distribution of Average Transaction Value



171

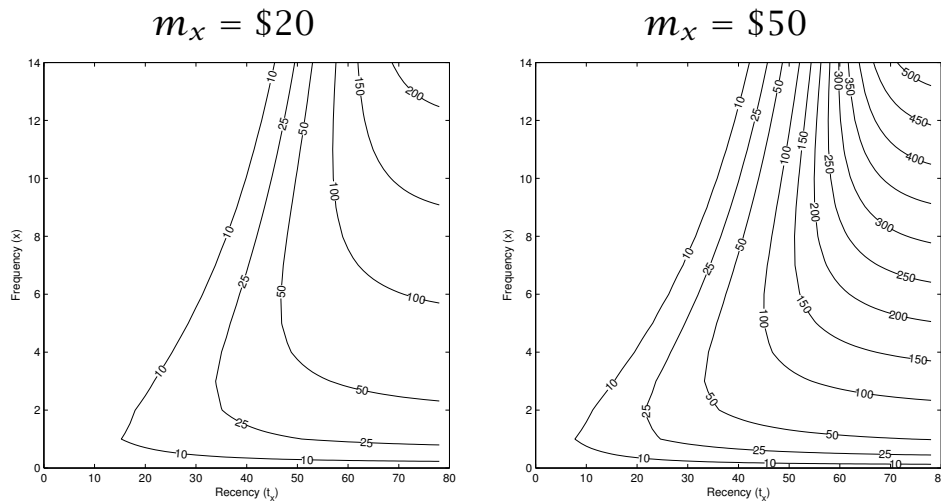
Computing Expected Residual Lifetime Value

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

$$\begin{aligned} E(RLV) &= \text{margin} \times \text{revenue/transaction} \times DERT \\ &= \text{margin} \times E(M|p, q, \gamma, m_x, x) \\ &\quad \times DERT(\delta | r, \alpha, s, \beta, x, t_x, T) \end{aligned}$$

172

Estimates of $E(RLV)$



(Margin = 30%, 15% discount rate)

173

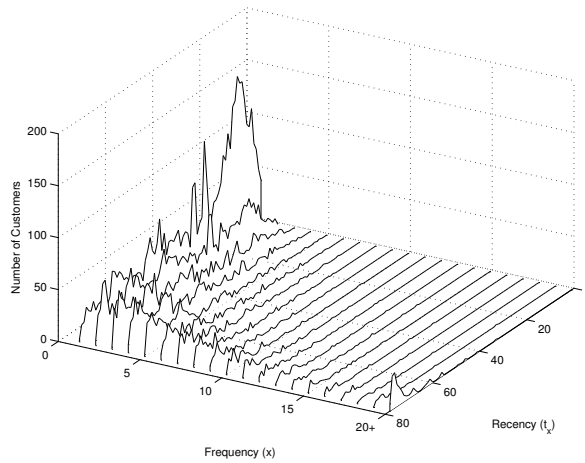
Closing the Loop

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.

174

Distribution of Repeat Customers



(12,054 customers make no repeat purchases)

175

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

176

Total Residual Value by RFM Segment

		Recency			
		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 alive
 tails → dies
 - $P(\text{tails})$ varies across customers

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

Purchase Process:

- While alive, # 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 dies with probability p .
- Heterogeneity in dropout probabilities across customers is distributed beta(a, b).

179

BG/NBD Likelihood Function

We can express the model likelihood function as:

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

where

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

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

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

$$A_4 = \left(\frac{a}{b+x-1}\right)\left(\frac{1}{\alpha+t_x}\right)^{r+x}$$

180

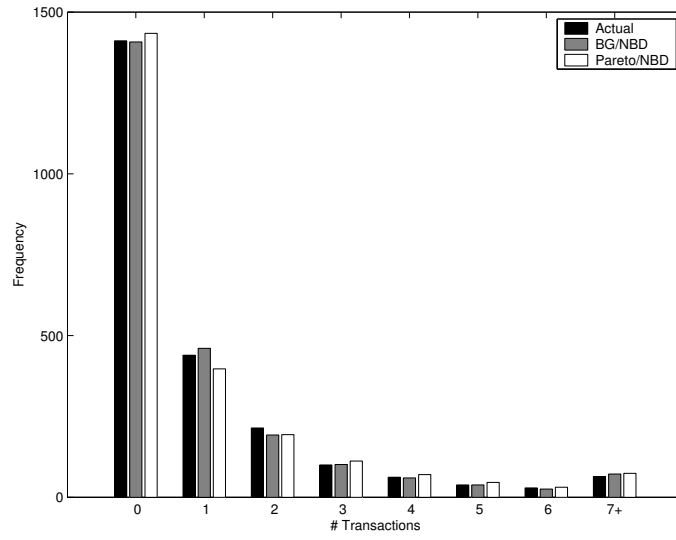
BGNBD Estimation

	A	B	C	D	E	F	G	H	I
1	r	0.243							
2	alpha	4.414	=GAMMALN(B\$1+B8)-			=IF(B8>0, LN(B\$3)-LN(B\$4+B8-1)-			
3	a	0.793	GAMMALN(B\$1)+B\$1*LN(B\$2)			(B\$1+B8)*LN(B\$2+C8),0)			
4	b	2.426							
5	LL	-9582.4					=-(B\$1+B8)*LN(B\$2+D8)		
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	=SUM(E8:E2364)		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	=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))								
13	=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))					=GAMMALN(B\$3+B\$4)+GAMMALN(B\$4+B8)-			
14	0007	1	5.00	38.86		GAMMALN(B\$4)-GAMMALN(B\$3+B\$4+B8)			
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

Model Estimation Results

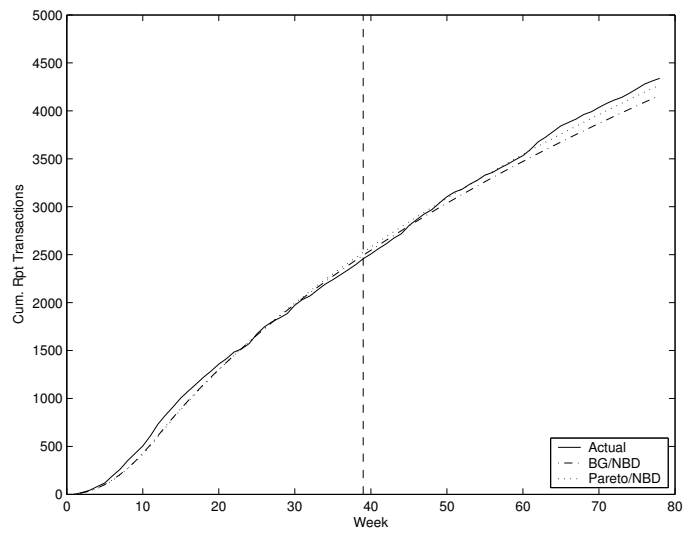
	BG/NBD	Pareto/NBD
r	0.243	0.553
α	4.414	10.578
a	0.793	
b	2.426	
s		0.606
β		11.669
LL	-9582.4	-9595.0

Frequency of Repeat Transactions



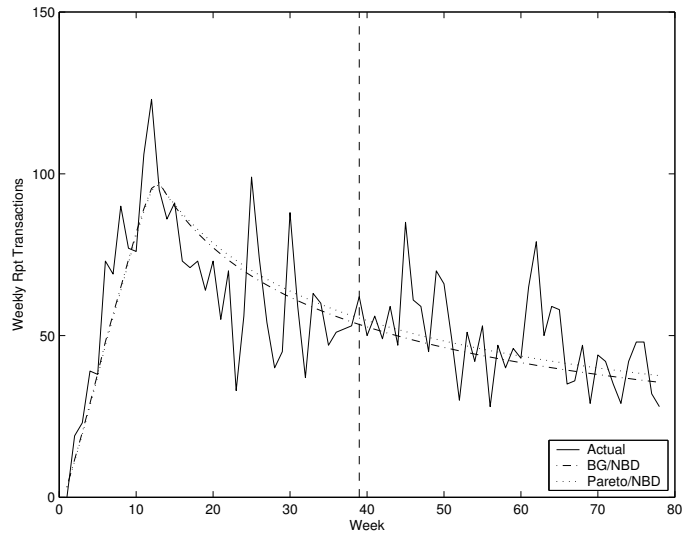
183

Tracking Cumulative Repeat Transactions



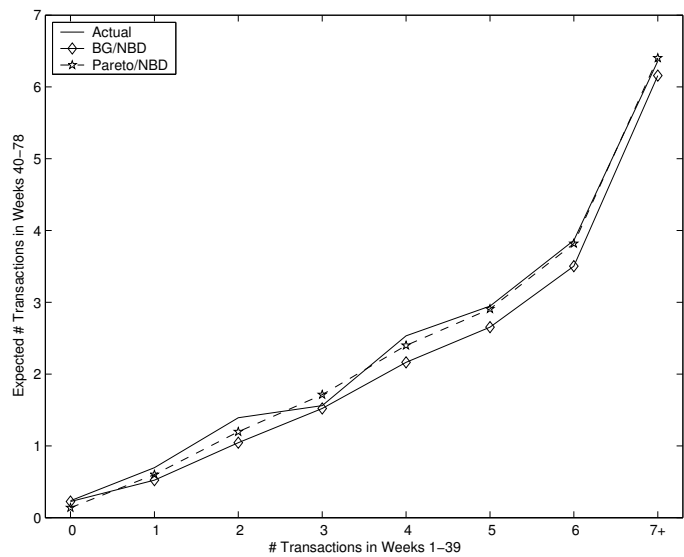
184

Tracking Weekly Repeat Transactions



185

Conditional Expectations



186

Computing DERT for the BG/NBD

- It is very difficult to solve

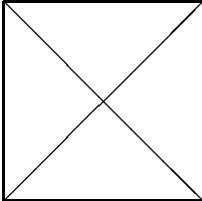
$$DERT = \int_T^{\infty} E[t(t)]S(t | t > T)d(t - T)dt$$

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

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

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

Classifying Differences in Assumed Customer Death Processes

Opportunities for Death	Continuous time		Pareto/NBD
	Discrete time	BG/NBD	PDO
		Transaction time	Calendar time
		Time measure	

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/>>

Further Reading

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. <<http://brucehardie.com/papers/018/>>

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/>>

Further Reading

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, Peter S. Fader, and Bruce G. S. Hardie (2011),
“New Perspectives on Customer ‘Death’ Using a Generalization
of the Pareto/NBD Model,” *Marketing Science*, **30**
(September-October), 866-880.

<<http://brucehardie.com/papers/027/>>

Beyond the Basic Models

Implementation Issues

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

193

Implications of Data Recording Processes (Contractual Settings)

Cohort	Calendar Time →				
1	n_{11}	n_{12}	n_{13}	...	n_{1I}
2		n_{22}	n_{23}	...	n_{2I}
3			n_{33}	...	n_{3I}
⋮				⋱	⋮
I					n_{II}
	$n_{.1}$	$n_{.2}$	$n_{.3}$...	$n_{.I}$

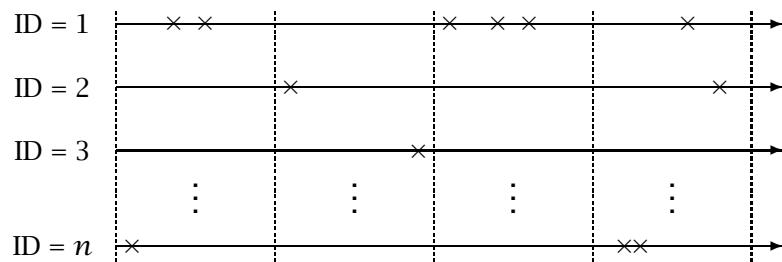
194

Implications of Data Recording Processes (Contractual Settings)

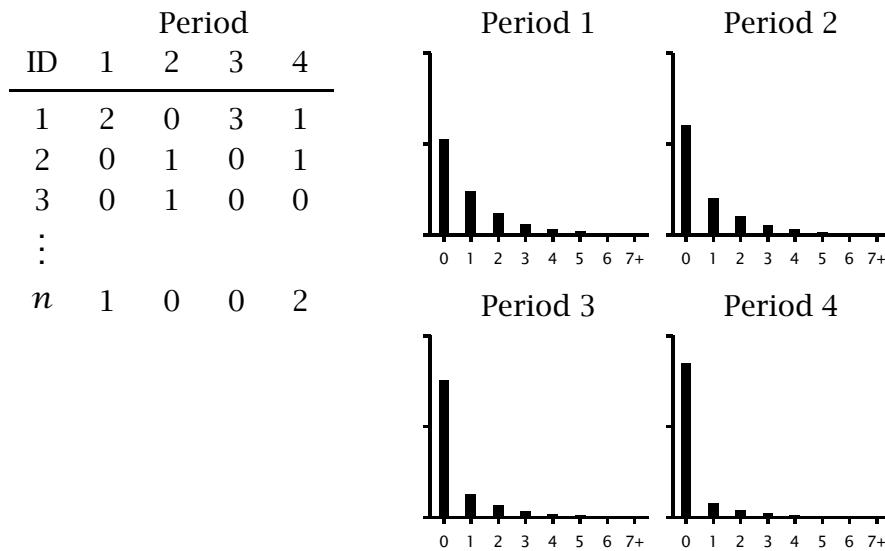
Cohort	Calendar Time →	
1	n_{11}	n_{1I}
2	n_{22}	n_{2I}
⋮		⋮
⋮		⋮
I-1	$n_{I-1,I-1}$	$n_{I-1,I}$
1		n_{II}
	$n_{,1}$	$n_{,I}$

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

Implications of Data Recording Processes (Noncontractual Settings)



Implications of Data Recording Processes (Noncontractual Settings)



197

Implications of Data Recording Processes (Noncontractual Settings)

Match the model likelihood function to the data structure:

- Interval-censored individual-level data

Fader, Peter S. and Bruce G.S. Hardie (2010), "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.

Jerath, Kinshuk, Peter S. Fader, and Bruce G.S. Hardie (2011), "Customer-Base Analysis on a 'Data Diet': Model Inference Using Repeated Cross-Sectional Summary (RCSS) Data."
<<http://brucehardie.com/papers/025/>>

198

Model Extensions

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

199

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 :

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

- This is equivalent to a constant hazard function.

200

The Weibull Distribution

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

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

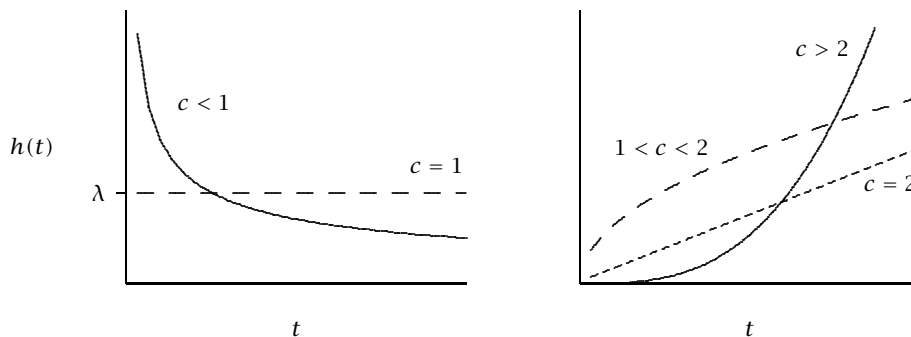
$$h(t) = c\lambda t^{c-1}$$

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 $c > 1$.

201

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 $c > 1$.

202

Individual-Level Duration Dependence

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

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

- 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.

203

Individual-Level Duration Dependence

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

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

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

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

204

Nonstationarity

- “Buy then die” \Leftrightarrow 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.

205

Correlation

We typically assume independence of the latent traits.

- Correlation can sometimes be accommodated using Sarmanov distributions:
 - Park, Young-Hoon and Peter S. Fader (2004), “Modeling Browsing Behavior at Multiple Websites,” *Marketing Science*, 23 (Summer), 280–303.
 - Danaher, Peter J. and Bruce G. S. Hardie (2005), “Bacon With Your Eggs? Applications of a New Bivariate Beta-Binomial Distribution,” *The American Statistician*, 59 (November), 282–286.
- Transformations of multivariate normals more flexible ... but no closed-form solutions.

Fader, Peter S. and Bruce G. S. Hardie (2011), “Implementing the S_{BB} -G/B Model in MATLAB.”

<<http://brucehardie.com/notes/023/>>

206

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

207

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.

208

Philosophy of Model Building

Problem: Managers are not using the “state-of-the-art” models developed by researchers.

Solution: Adopt an evolutionary approach to model building.

- Maximize likelihood of acceptance by starting with a (relatively) simple model that the manager can understand AND that can be implemented at low cost.
- Model deficiencies can be addressed, and more complex (and costly) models can be developed/implemented, if benefits > cost.

209

Philosophy of Model Building

We are specifically interested in kick-starting the evolutionary process:

- Minimize cost of implementation
 - use of readily available software (e.g., Excel)
 - use of data summaries
- Purposively ignore the effects of covariates and other “complexities” at the outset.

Make everything as simple as possible, but not simpler.

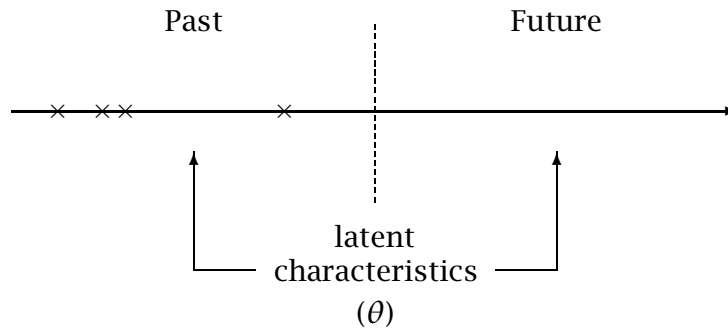
Albert Einstein

210

Central Tenet

Traditional approach

$$\text{future} = f(\text{past})$$



Probability modelling approach

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

211

Classifying Customer Bases

Opportunities for Transactions	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
		Noncontractual	Contractual
Type of Relationship With Customers			

212