Probability Models for Customer-Base Analysis

Peter S. Fader
University of Pennsylvania

Bruce G. S. Hardie
London Business School

24th Annual Advanced Research Techniques Forum
June 9–12, 2013

Agenda

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

Customer-Base Analysis

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

Traditional Modelling Approach

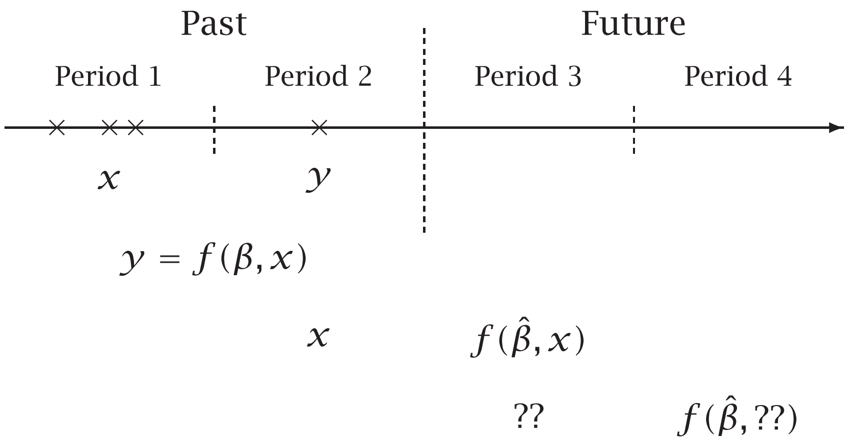
The transaction data are split into two consecutive periods:

- Data from the second period are used to create the dependent variable of interest (e.g., buy/not-buy, number of transactions, total spend).
- Data from the first period are used to create the predictor variables.
- Period 1 behavior is frequently summarized in terms of the customer's "RFM" characteristics: *recency* (time of most recent purchase), *frequency* (number of purchases), and *monetary value* (average spend per transaction).

Comparison of Modelling Approaches

Traditional approach

future = f(past)



Past

Future

Period 1

Period 2

Period 3

Period 4

x

y

$y = f(\beta, x)$

x

$f(\hat{\beta}, x)$

??

$f(\hat{\beta}, ??)$

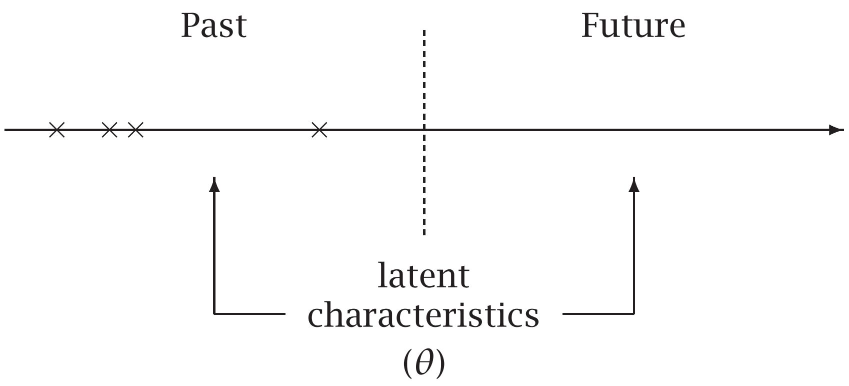
Comparison of Modelling Approaches

In addition to the problem of having to predict Period 3 behavior in order to predict Period 4 behavior (and so on), the traditional approach has other limitations:

- The regression-type models are ad hoc in nature; there is no well-established theory. (Why use RFM? Is the fact that "it works" a good enough reason?)
- The observed behavioral variables (e.g., RFM) are only imperfect indicators of underlying behavioral characteristics. Different "slices" of the data will yield different values of the variables and therefore different parameter estimates . . . and different forecasts.

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 Business Settings

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

- Based on numbers presented in a news release that reported Vodafone Group Plc’s results for the six months ended 30 September 2012, we see that Vodafone UK has 10.8 million “pay monthly” customers.
- In his “Q3 2012 Earnings Conference Call,” the CFO of Amazon made the comment that “[a]ctive customer accounts exceeded 188 million,” where customers are considered active when they have placed an order during the preceding twelve-month period.

Classifying Business 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 (i.e., attrition is latent).
- 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 Business Settings

Consider the following four specific business settings:

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

Classifying Customer Bases

Opportunities for Transactions

	Noncontractual	Contractual
Continuous	Grocery purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies “Friends” schemes

Type of Relationship With Customers

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

Classifying Customer Bases

Further dimensions of possible interest in contractual settings:

- Is usage while under contract observed or unobserved?
- Is the revenue associated with the contract known in advance or not known in advance?

These factors may be determined by technology and the firm’s pricing policies.

Adapted from: Ascarza, Eva (2009), *Modeling Customer Behavior in Contratual Settings*, unpublished Ph.D thesis, University of London.

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 alive)

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 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

Calculating $E(CLV)$

- Definitional; of little use by itself.
- We must operationalize $E[v(t)]$, $S(t)$, and $d(t)$ in a specific business setting … then solve the integral.
- Important distinctions:
  - Expected lifetime value of an as-yet-to-be-acquired customer
  - Expected lifetime value of a just-acquired customer
  - Expected *residual* lifetime value, $E(RLV)$, of an existing customer

Calculating $E(CLV)$

- The expected lifetime value of an as-yet-to-be-acquired customer is given by

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

- Standing at time T, the expected residual lifetime value of an existing customer is given by

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

Review of Probability Models

"Winwood Reade is good upon the subject," said Holmes. "He remarks that, while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to."

Sir Arthur Conan Doyle, *The Sign of Four*.

The Logic of Probability Models

- The actual data-generating process that lies behind any given data on buyer behavior embodies a huge number of factors.
- Even if the actual process were completely deterministic, it would be impossible to measure all the variables that determine an individual's buying behavior in any setting.

$\Longrightarrow$ Any account of buyer behavior must be expressed in probabilistic/random/stochastic terms so as to account for our ignorance regarding (and/or lack of data on) all the determinants.

The Logic of Probability Models

- Rather than try to tease out the effects of various marketing, personal, and situational variables, we embrace the notion of randomness and view the behavior of interest as the outcome of some probabilistic process.
- We propose a model of individual-level behavior that is "summed" across individuals (taking individual differences into account) to obtain a model of aggregate behavior.

Applications of Probability Models

- Summarize and interpret patterns of market-level behavior
- Predict behavior in future periods, be it in the aggregate or at a more granular level (e.g., conditional on past behavior)
- Make inferences about behavior given summary measures
- Profile behavioral propensities of individuals
- Generate benchmarks/norms

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 / how long”
  - 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 θ 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 geometric random variable, where

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

The 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 θ, where

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

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

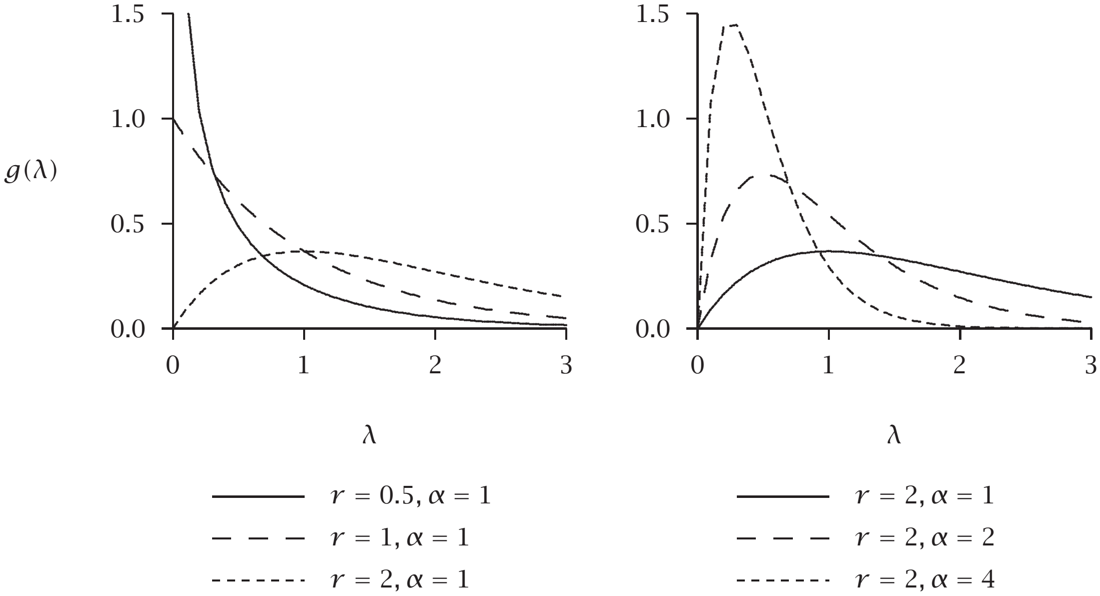
Capturing Heterogeneity in Latent Traits

The gamma distribution:

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

- $\Gamma(\cdot)$ is the gamma function
- r is the “shape” parameter and 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:

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

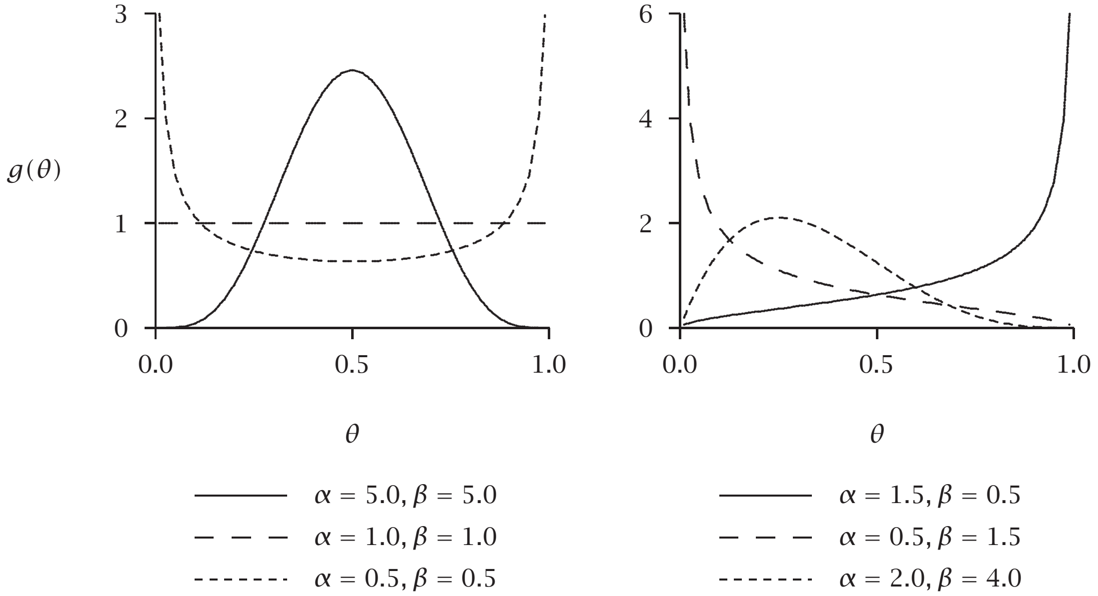
- $B(\alpha,\beta)$ is the beta function, which can be expressed in terms of gamma functions:

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

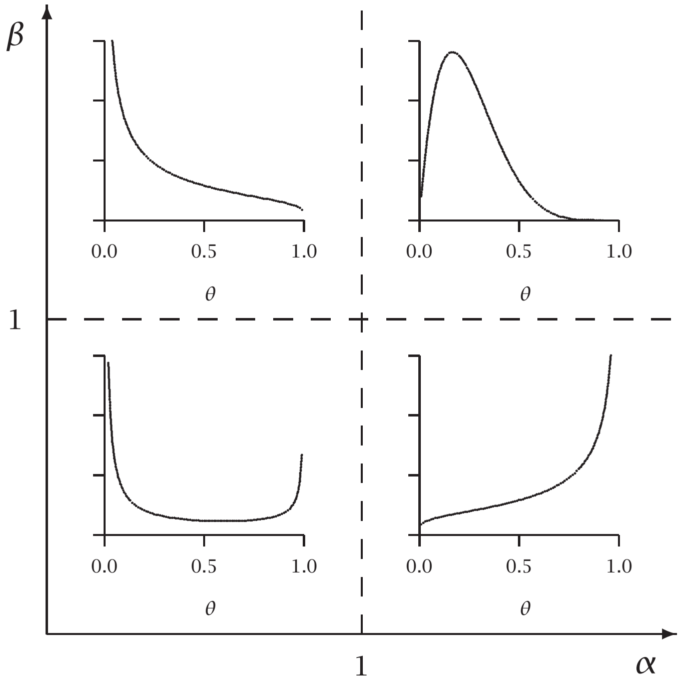
- The beta distribution is a flexible distribution . . . and is mathematically convenient

---

Illustrative Beta Distributions



Five General Shapes of the Beta Distribution



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 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 Beta-Geometric Model

- The individual-level behavior of interest can be characterized by the 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.

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

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.

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

Summary of Probability Models

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

Illustrative 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:

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

- 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

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

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

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

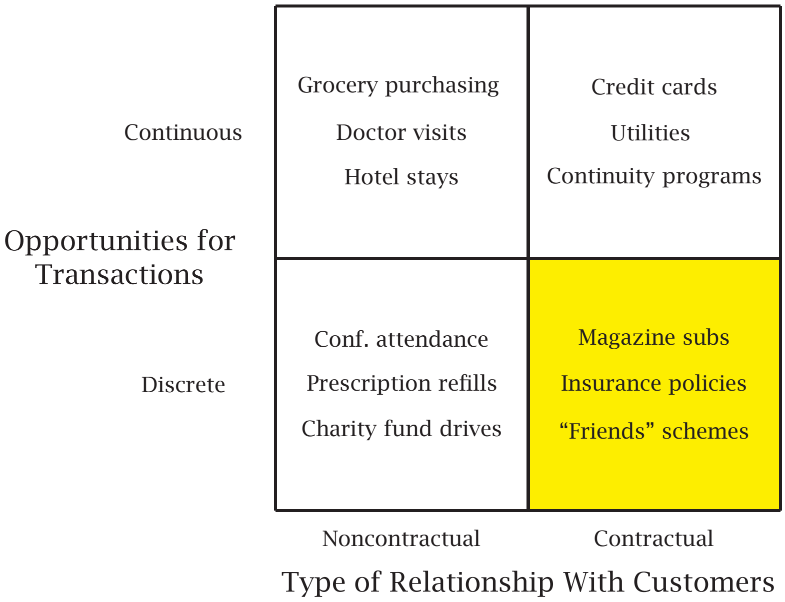
using the recursion

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

where $u_0 = 1$.

Models for Contractual Settings

Classifying Customer Bases



Opportunities for Transactions	Noncontractual	Contractual
Continuous	Grocery purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies “Friends” schemes

Type of Relationship With Customers

Illustrative Contractual Setting

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

Year 1	Year 2	Year 3	Year 4
1000	631	468	382

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

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?

Spending on Customer Acquisition

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

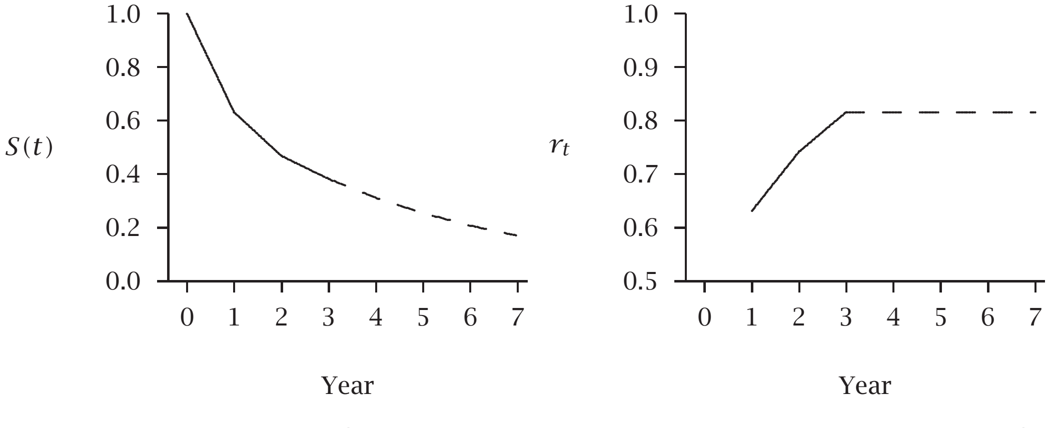
Spending on Customer Acquisition

For a randomly chosen customer,

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

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

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.

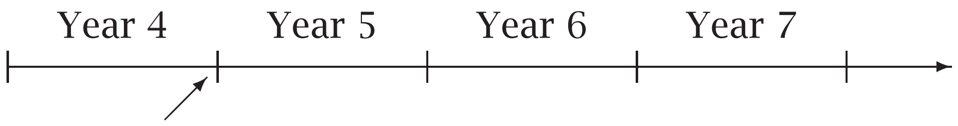
$$r_t = S(t)/S(t-1) \iff S(t) = S(t-1) \times r_t$$

Spending on Customer Acquisition

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

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

Residual Value of the Customer Base



Net CF		$100	$100	$100	⋯
P(alive)		0.816	0.816^2	0.816^3	⋯
discount		1.000	$(1 + 0.1)$	$(1 + 0.1)^2$	⋯

Residual Value of the Customer Base

For a randomly chosen customer,

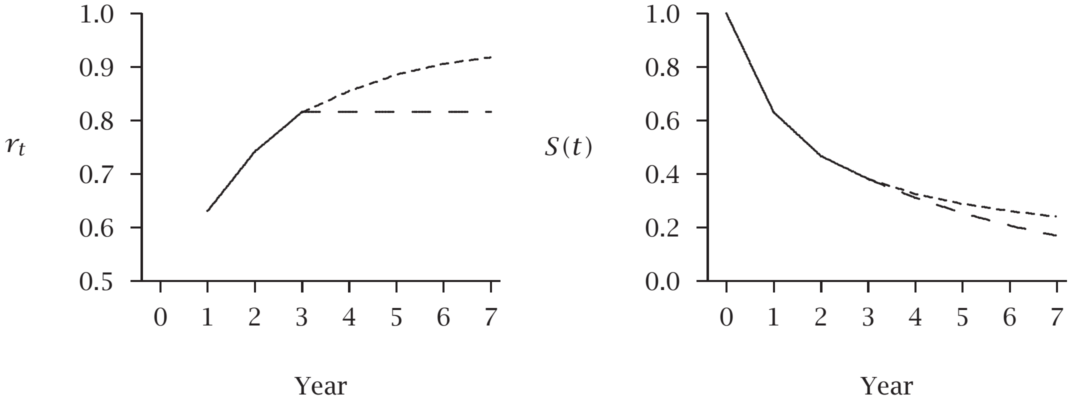
$$
\begin{aligned}
E(RLV) &= \$100 \times 0.816 + \$100 \times \frac{0.816^2}{(1+0.1)} + \dots \\
&= \$100 \times \sum_{t=1}^{\infty} \frac{0.816^t}{(1+0.1)^{t-1}} \\
&= \$316
\end{aligned}
$$

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

What's wrong with this analysis?

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).

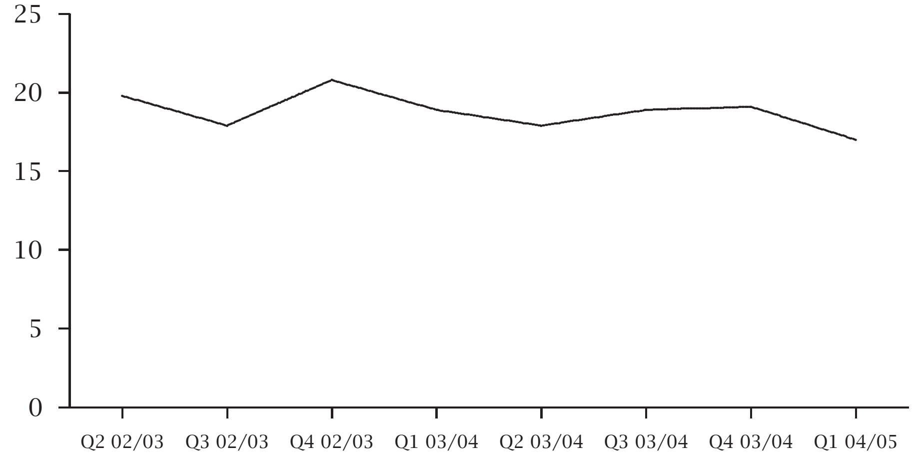
> Renewal rates at regional magazines vary; generally 30% of subscribers renew at the end of their original subscription, but that figure jumps to 50% for second-time renewals and all the way to 75% for longtime readers.

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)

I am happy to report that 41% of new members who joined in 2011 renewed their membership in 2012, and that ION has an overall retention of 78%.

ION Newsletter, Winter 2011–2012.

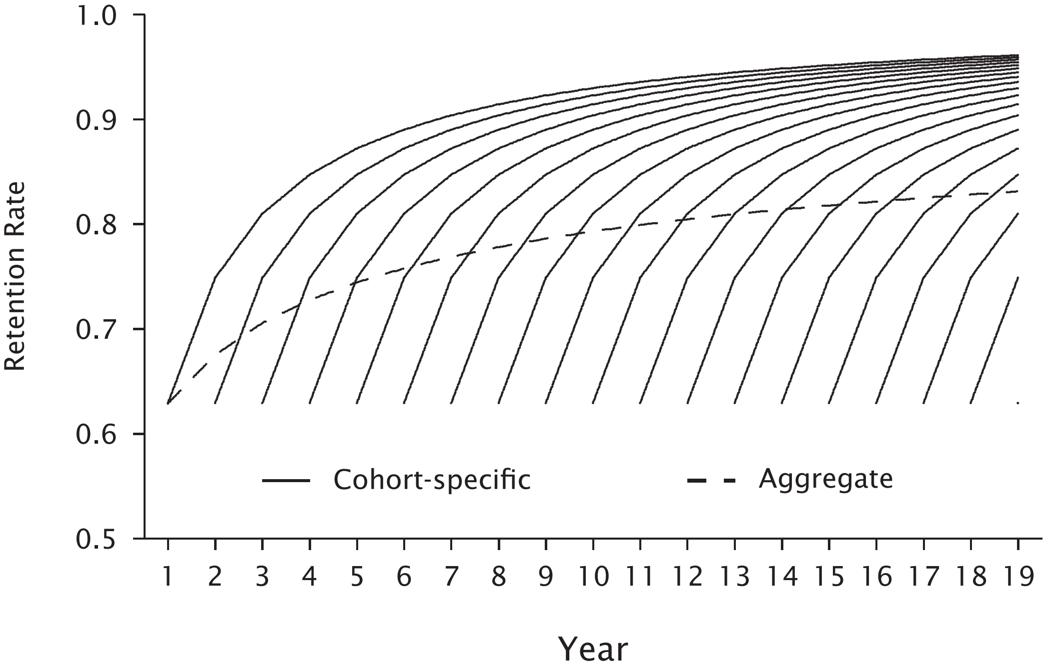
Cohort-level vs. Aggregate Numbers

	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

	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	--	0.629	0.749	0.811	0.848	0.873	0.890	0.905	0.912	0.923	0.932	0.933	0.940	0.943	0.946	0.950	0.955	0.961	0.959	0.957
Yr 02		--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94	0.943	0.946	0.95	0.955	0.961	0.959
Yr 03			--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94	0.943	0.946	0.95	0.955	0.961
Yr 04				--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94	0.943	0.946	0.95	0.955
Yr 05					--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94	0.943	0.946	0.95
Yr 06						--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94	0.943	0.946
Yr 07							--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94	0.943
Yr 08								--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933	0.94
Yr 09									--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932	0.933
Yr 10										--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923	0.932
Yr 11											--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912	0.923
Yr 12												--	0.629	0.749	0.811	0.848	0.873	0.89	0.905	0.912
Yr 13													--	0.629	0.749	0.811	0.848	0.873	0.89	0.905
Yr 14														--	0.629	0.749	0.811	0.848	0.873	0.89
Yr 15															--	0.629	0.749	0.811	0.848	0.873
Yr 16																--	0.629	0.749	0.811	0.848
Yr 17																	--	0.629	0.749	0.811
Yr 18																		--	0.629	0.749
Yr 19																			--	0.629
Yr 20																				--
Aggregate	--	0.629	0.675	0.706	0.728	0.744	0.758	0.769	0.778	0.786	0.793	0.799	0.805	0.809	0.814	0.818	0.822	0.825	0.828	0.831

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.

A Discrete-Time Model for Contract Duration

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

$\longrightarrow$ the duration of the customer's relationship with the firm is characterized by the 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

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

A Discrete-Time Model for Contract Duration

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

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

- The aggregate survivor function is

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

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.

A Discrete-Time Model for Contract Duration

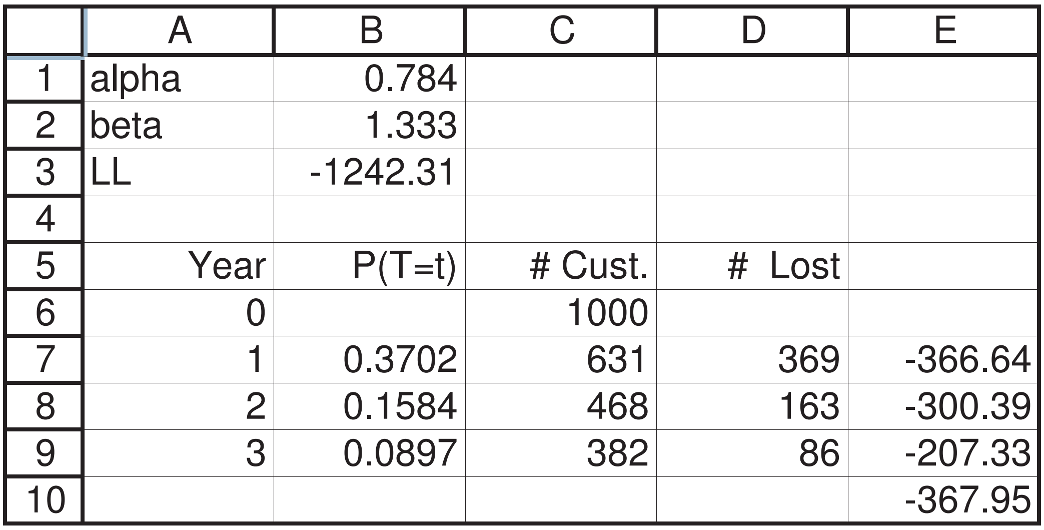
We can compute BG probabilities using the following forward-recursion formula from $P(T = 1)$:

$$P(T = t \mid \alpha, \beta) = \begin{cases} \dfrac{\alpha}{\alpha + \beta} & t = 1 \\[2ex] \dfrac{\beta + t - 2}{\alpha + \beta + t - 1} P(T = t - 1) & t = 2, 3, \ldots \end{cases}$$

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

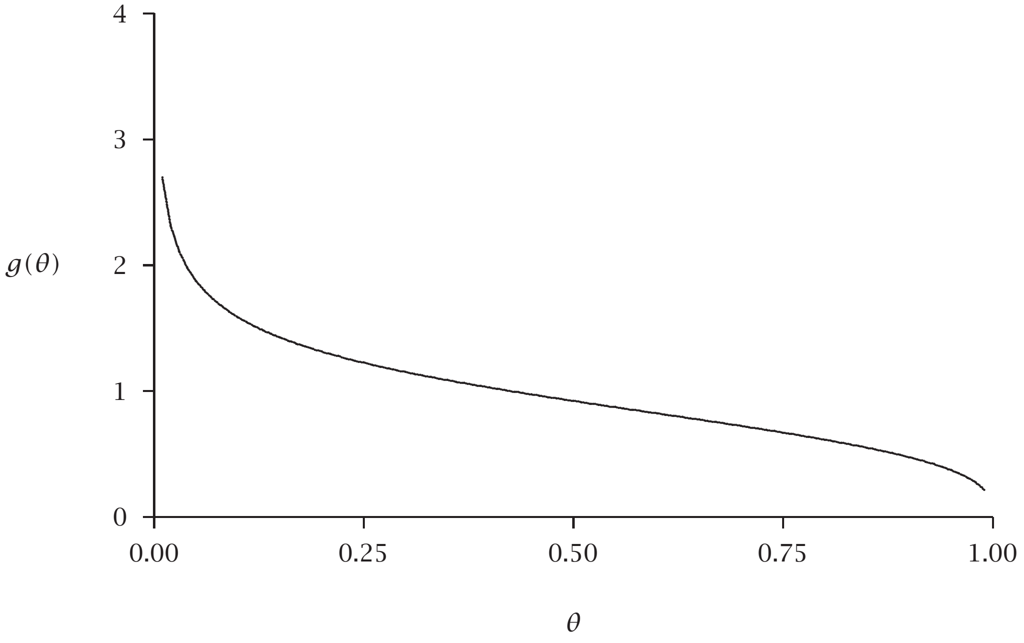
$$S(t \mid \alpha, \beta) = \prod_{i=1}^{t} r_i\,, t = 1, 2, 3, \ldots,\ \text{where } S(0) = 1\,.$$

Fitting the BG 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



Computing $E(CLV)$

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

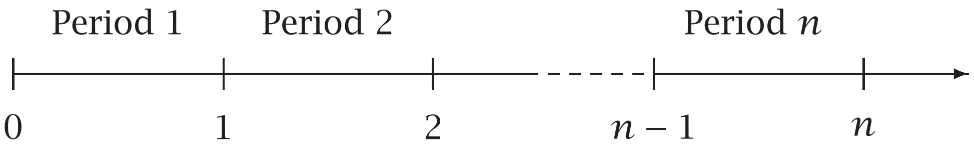
- In a contractual setting, assuming an individual's mean value per unit of time is constant ($\bar{v}$),

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

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

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

Computing DEL



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

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

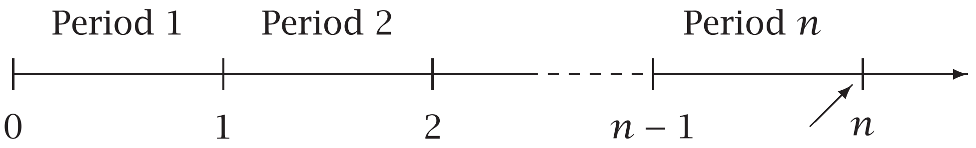
- But θ is unobserved ...

Computing DEL

Integrating over the distribution of θ:

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

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 ....

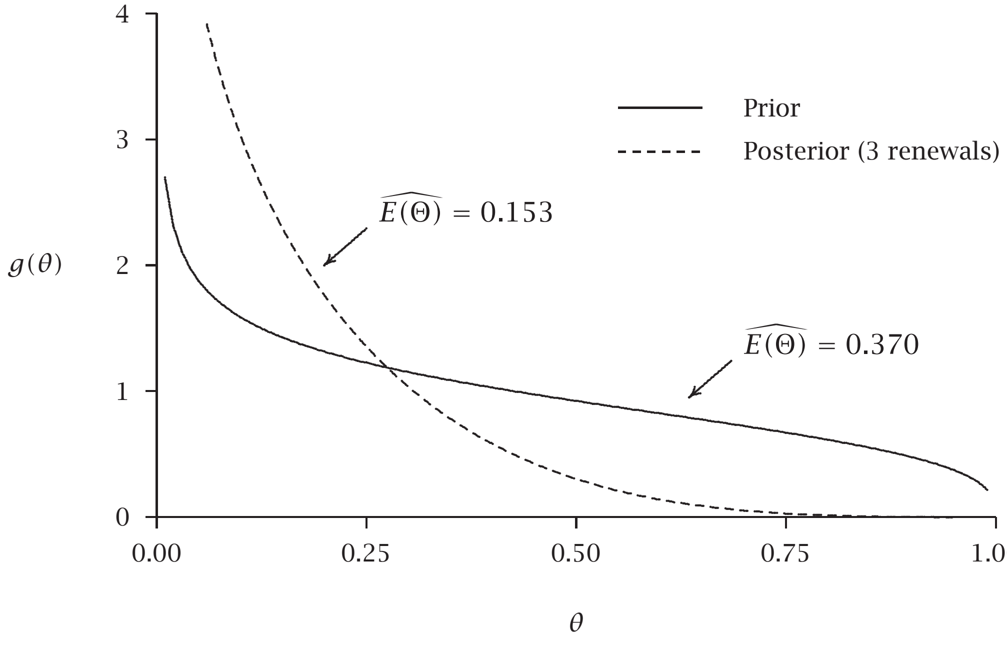
Computing DERL

By Bayes' Theorem, the posterior distribution of θ is

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

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

Distributions of Churn Probabilities



Computing DERL

Integrating over the posterior distribution of θ:

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

Alternative Derivation of Expression for DEL

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

Alternative Derivation of Expression for DERL

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

Computing DE(R)L Using Excel

- From the alternative derivations:

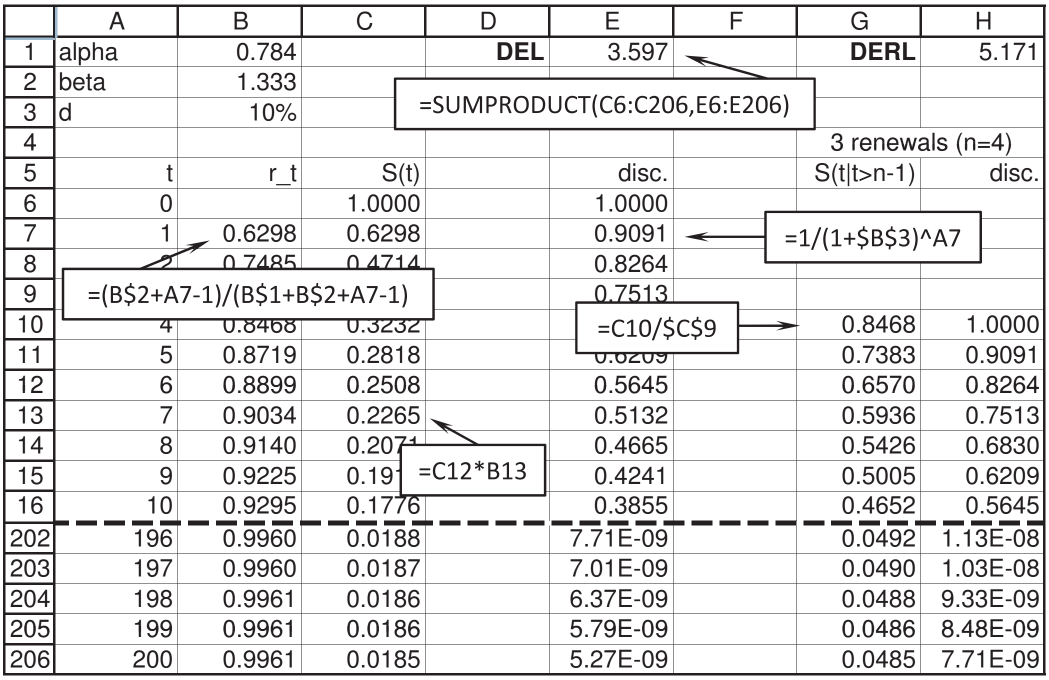
$$
DEL(d \mid \alpha, \beta) = \sum_{t=0}^{\infty} \frac{S(t \mid \alpha, \beta)}{(1+d)^t}
$$

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

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

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

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						
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			
8	2	0.7485	0.4714		0.8264			
9					0.7513			
10	4	0.8468	0.3232				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	0.2071		0.4665		0.5426	0.6830
15	9	0.9225	0.191		0.4241		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

=SUMPRODUCT(C6:C206,E6:E206)

=1/(1+$B$3)^A7

=(B$2+A7-1)/(B$1+B$2+A7-1)

=C10/$C$9

=C12*B13

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 BG 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).

Validating the BG-based CLV Estimates

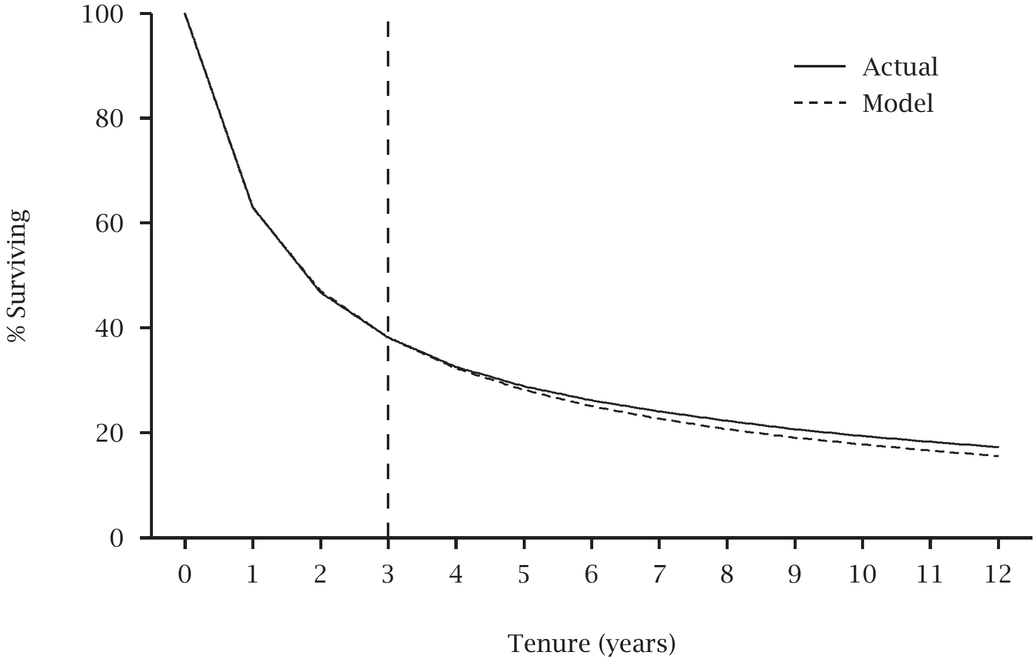
We actually have 12 years of renewal data.

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

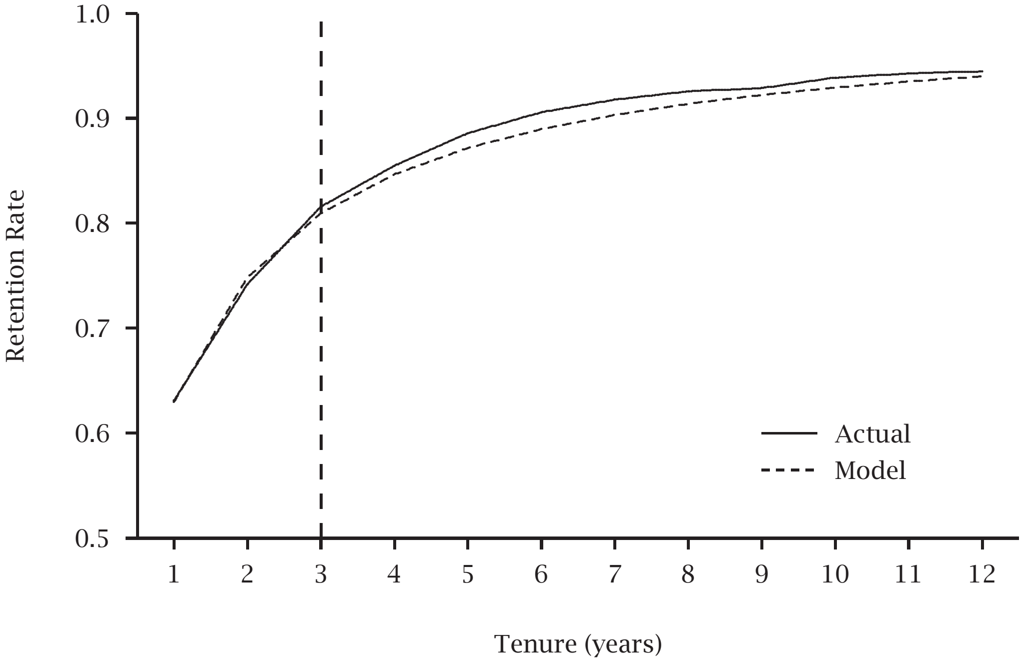
$$E(CLV) = \$100 \times \sum_{t=0}^{12} \frac{S(t)}{(1.1)^t}$$

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

Survival Curve Projection



Projecting Retention Rates



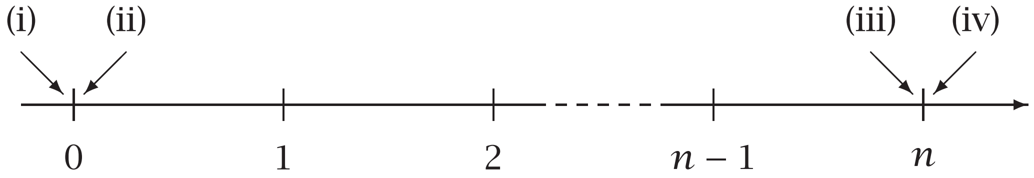
Comparing Approaches

	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.

Expressions for DE(R)L

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



Expressions for DE(R)L

Case (i):

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

Case (ii):

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

Expressions for DE(R)L

Case (iii):

$$DERL(d \mid \alpha, \beta, \text{alive for } n \text{ periods}) = \frac{\beta + n - 1}{\alpha + \beta + n - 1}\, {}_2F_1\big(1, \beta + n; \alpha + \beta + n; \tfrac{1}{1+d}\big)$$

Case (iv):

$$DERL(d \mid \alpha, \beta, n \text{ contract renewals}) = \frac{\beta + n}{(\alpha + \beta + n)(1 + d)} \times {}_2F_1\big(1, \beta + n + 1; \alpha + \beta + n + 1; \tfrac{1}{1+d}\big)$$

Further Reading

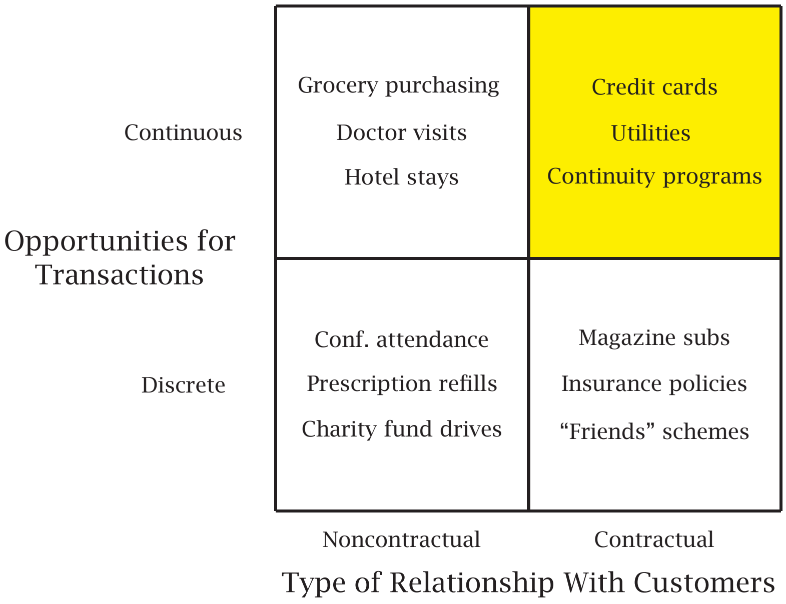
Fader, Peter S. and Bruce G. S. Hardie (2010), "Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity," *Marketing Science*, **29** (January–February), 85–93. <`http://brucehardie.com/papers/022/`>

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

From Discrete to Continuous Time

- We have considered a setting where the discrete contract period is annual.
- In some cases, there is a quarterly contract period, others monthly.
- In a number of cases, the contract is effectively “renewed” on a daily basis $\Longrightarrow$ “continuous” time.

Classifying Customer Bases



	Noncontractual	Contractual
Continuous	Grocery purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies “Friends” schemes

Opportunities for Transactions

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 Pareto Type II 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

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

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

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

The Hazard Function

- For the exponential distribution,

$$h(t|\lambda) = \lambda$$

- For the Pareto Type II model,

$$h(t|r, \alpha) = \frac{r}{\alpha + t}$$

- In applying the Pareto Type II 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

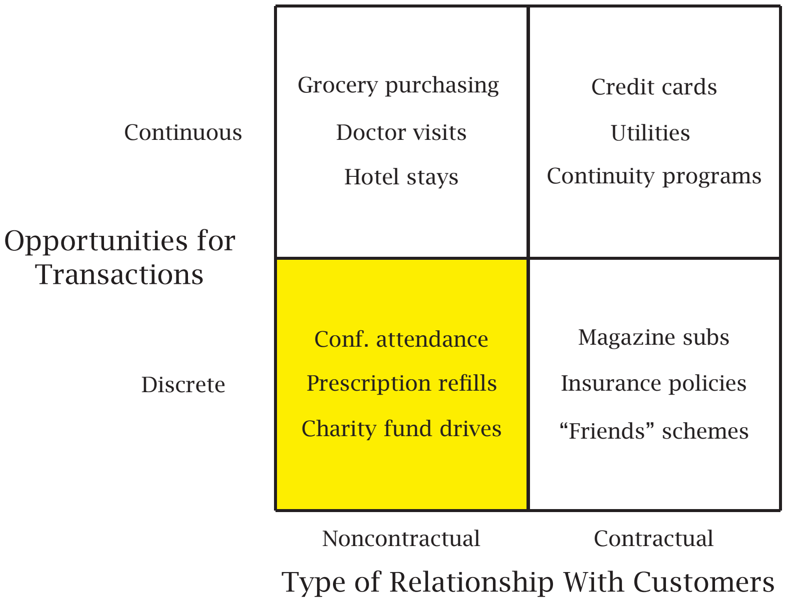
It follows that

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

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 purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies “Friends” schemes

Type of Relationship With Customers

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?

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 an individual decides whether or not to support the charity by tossing a coin:

$$\mathbb{H} \rightarrow \text{donate}$$

$$\mathbb{T} \rightarrow \text{don't donate}$$

1996	1997	1998	1999	2000	2001
1	0	1	1	0	0
$\mathbb{H}$	$\mathbb{T}$	$\mathbb{H}$	$\mathbb{H}$	$\mathbb{T}$	$\mathbb{T}$

- An individual tosses the same coin each year.
- $P(\mathbb{H})$ varies across individuals.

Modelling the Transaction Stream

- The number of transactions in the interval $\{1, 2, \ldots, n\}$ is distributed binomial,

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

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

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

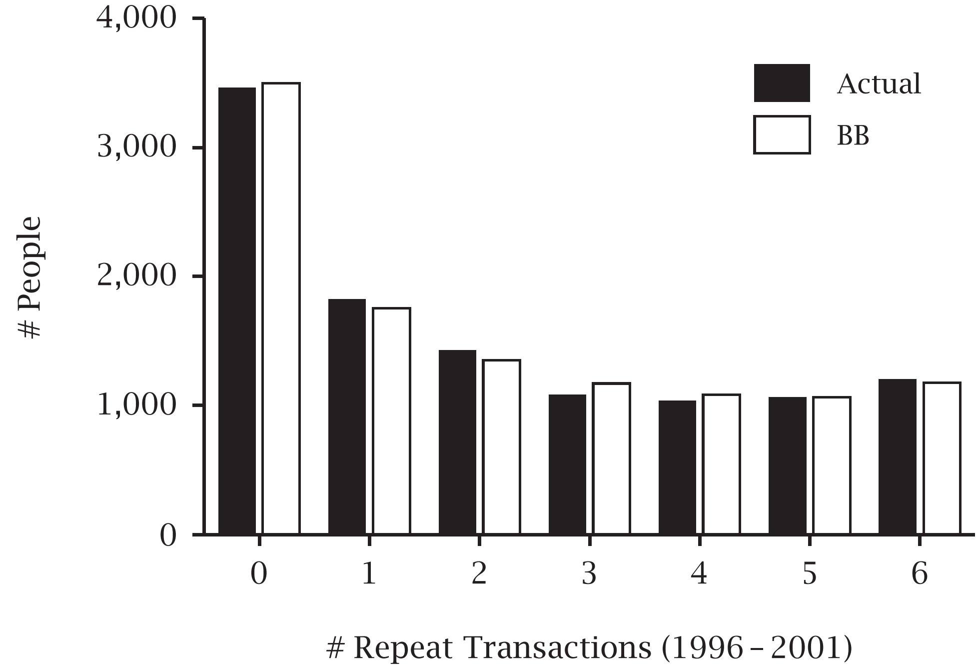
Modelling the Transaction Stream

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

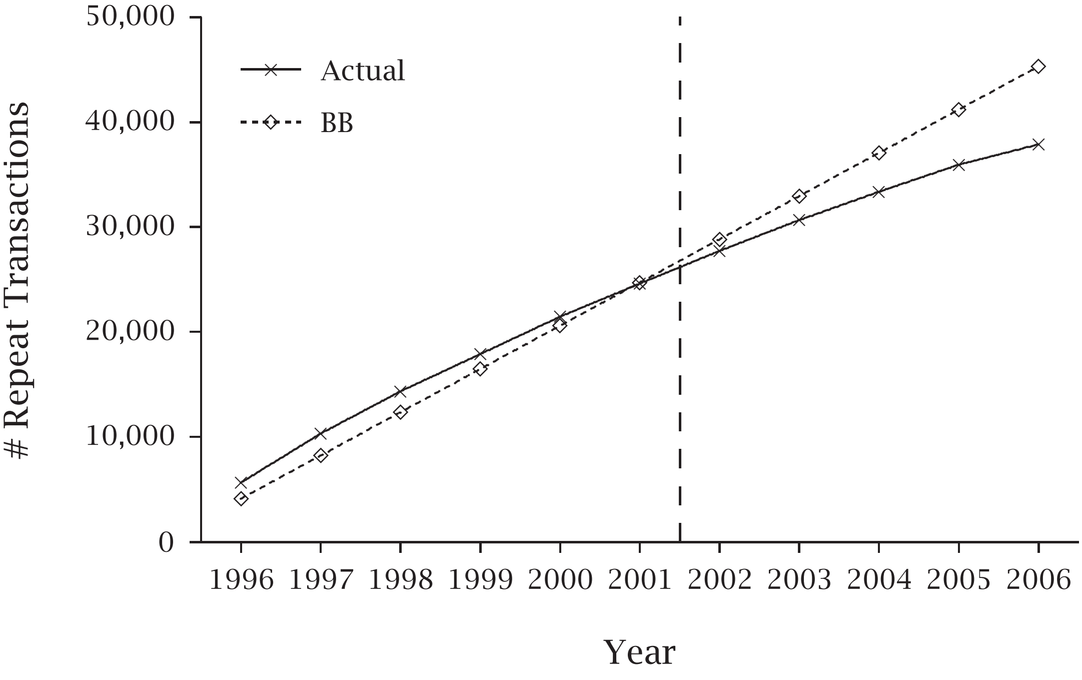
$$\begin{aligned} P(X(n) = x \mid \alpha, \beta) &= \int_0^1 P(X(n) = x \mid p) g(p \mid \alpha, \beta)\, dp \\ &= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \,, \end{aligned}$$

which is the beta-binomial (BB) distribution.

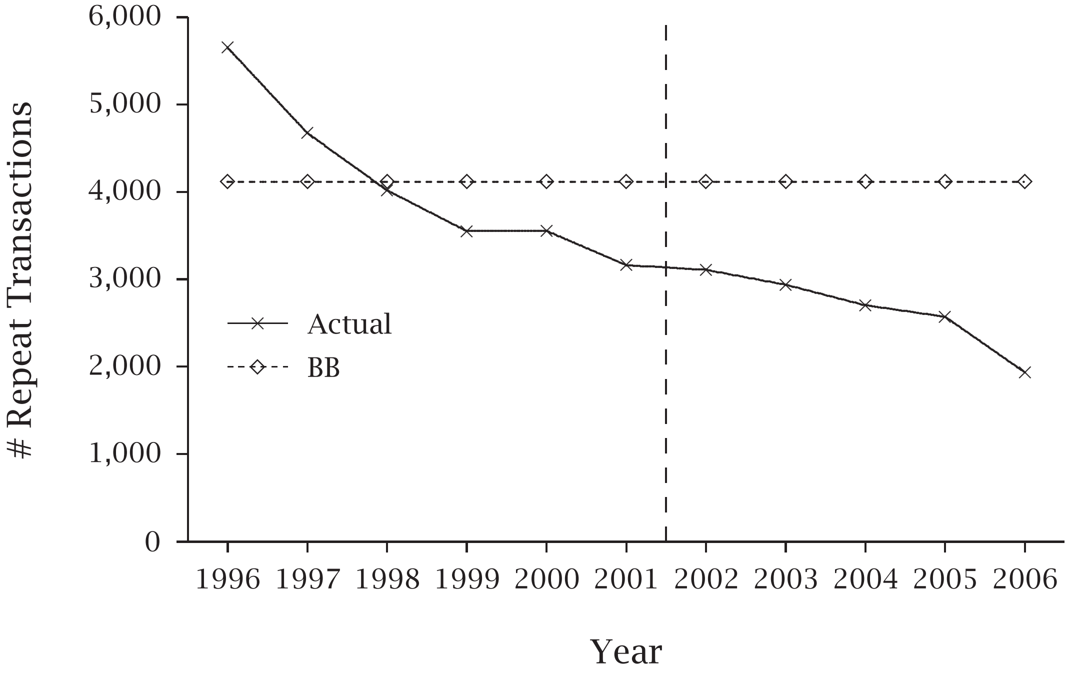
Fit of the BB Model



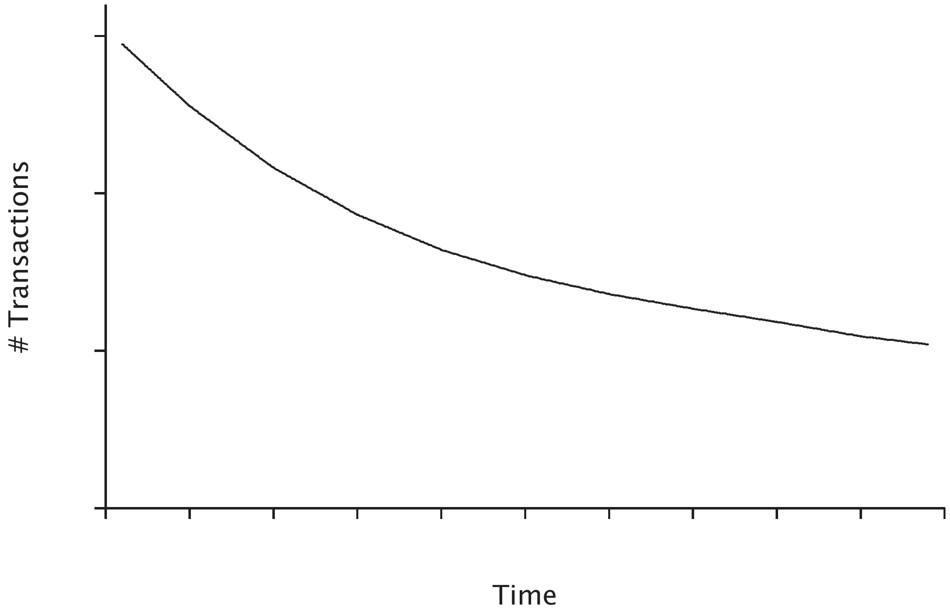
Tracking Cumulative Repeat Transactions



Tracking Annual Repeat Transactions



Towards a More Realistic Model

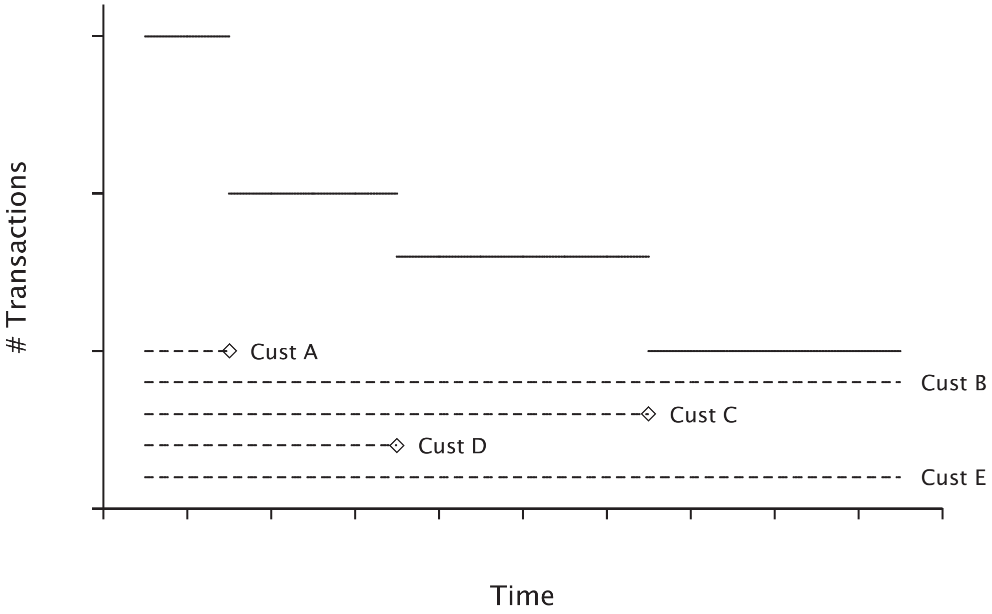


Towards a More Realistic Model

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.

Towards a More Realistic Model



Modelling the Transaction Stream

A customer's relationship with a firm has two phases: they are "alive" for an unobserved period of time, then "dead."

Transaction Process:

- While "alive," a customer makes a transaction at any given transaction opportunity following a "coin flip" process.
- Transaction probabilities vary across customers.

Latent Attrition Process:

- A "living" customer "dies" at the beginning of a transaction opportunity following a "coin flip" process.
- "Death" probabilities vary across customers.

Model Development

A customer's relationship with a firm has two phases: they are "alive" (A) then "dead" (D).

- While "alive," the customer makes a transaction at any given transaction opportunity with probability p:

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

- A "living" customer "dies" at the beginning of a transaction opportunity with probability θ

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

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

Model Development

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

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

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,

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

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

  ⟹ 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
1	1	1	1	1	1	1
2	1	1	1	1	1	0
3	1	1	1	1	0	1
4	1	1	1	1	0	0
5	1	1	1	0	1	1
6	1	1	1	0	1	0
7	1	1	1	0	0	1
		⋮			⋮	
		⋮			⋮	
		⋮			⋮	
62	0	0	0	0	1	0
63	0	0	0	0	0	1
64	0	0	0	0	0	0

⟶

x	t_x	n	# Donors
6	6	6	1203
5	6	6	728
5	5	6	335
4	6	6	512
4	5	6	284
4	4	6	240
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
1	1	6	1091
0	0	6	3464
			11104

Model Development

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

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

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

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

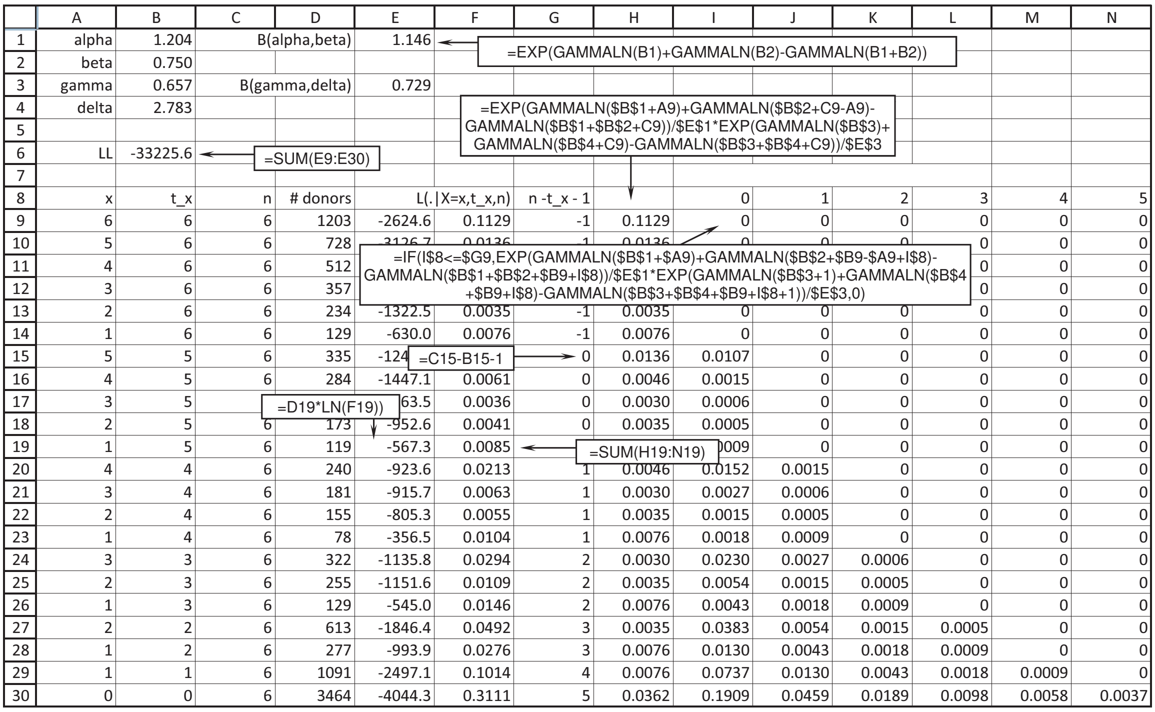
$$g(\theta \mid \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}$$

Model Development

Removing the conditioning on the latent traits p and θ,

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

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



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	alpha	1.204		B(alpha,beta)	1.146									
2	beta	0.750												
3	gamma	0.657		B(gamma,delta)	0.729									
4	delta	2.783												
5														
6	LL	-33225.6												
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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
12	3	6	6	357	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	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[illegible]	[illegible]	0	0.0136	0.0107	0	0	0	0	0
16	4	5	6	284	-1447.1	0.0061	0	0.0046	0.0015	0	0	0	0	0
17	3	5	[illegible]	[illegible]	[illegible]63.5	0.0036	0	0.0030	0.0006	0	0	0	0	0
18	2	5	6	173	-952.6	0.0041	0	0.0035	0.0005	0	0	0	0	0
19	1	5	6	119	-567.3	0.0085	[illegible]	[illegible]	[illegible]009	0	0	0	0	0
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

=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:E30)

=IF(I$8<=$G9,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+I$8)-GAMMALN($B$1+$B$2+$B9+I$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+I$8)-GAMMALN($B$3+$B$4+$B9+I$8+1))/$E$3,0)

=C15-B15-1

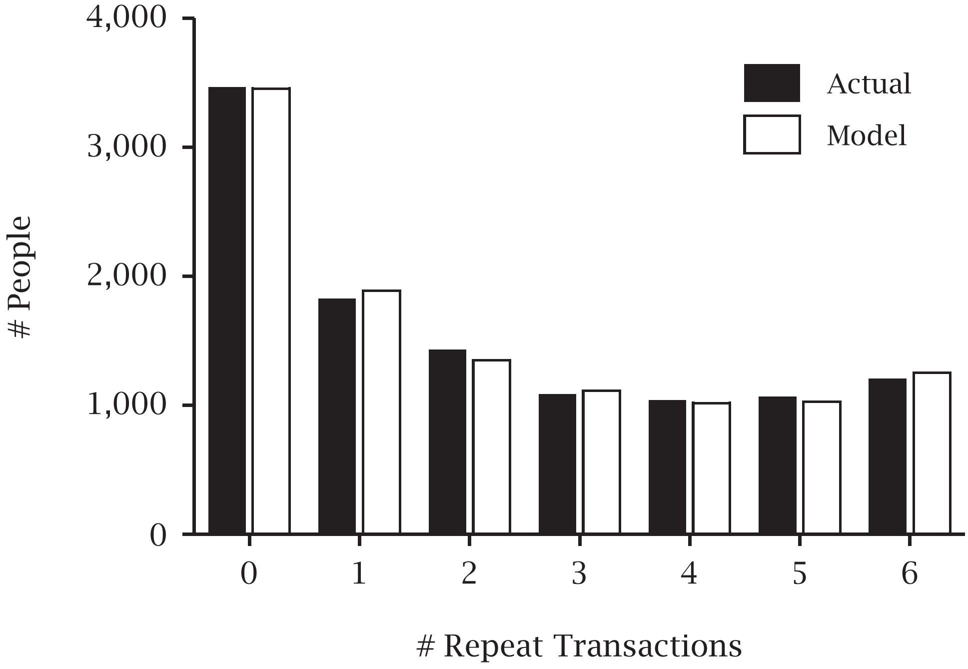
=D19*LN(F19))

=SUM(H19:N19)

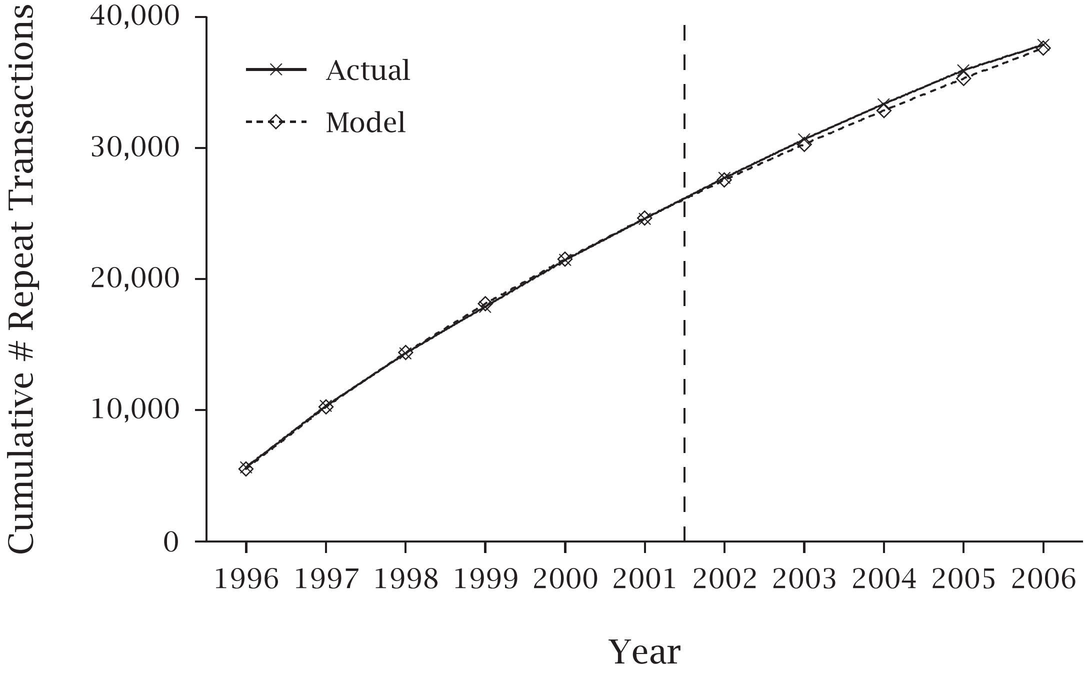
Parameter Estimates (1995 Cohort)

	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

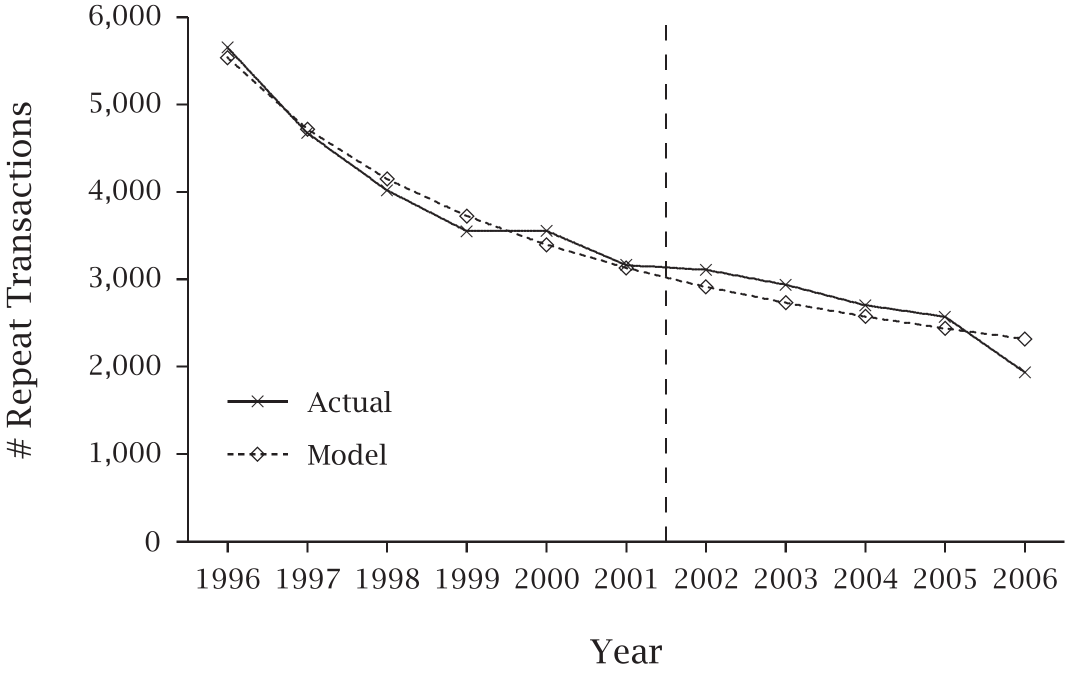
Fit of the BG/BB Model



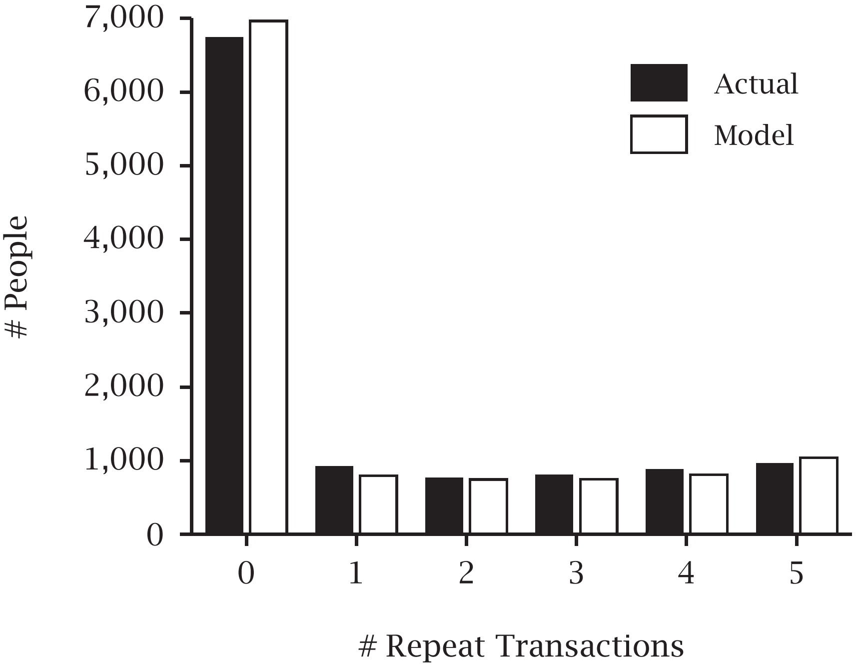
Tracking Cumulative Repeat Transactions



Tracking Annual Repeat Transactions



Repeat Transactions in 2002 – 2006



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.

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

  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	?	?	?	?	?

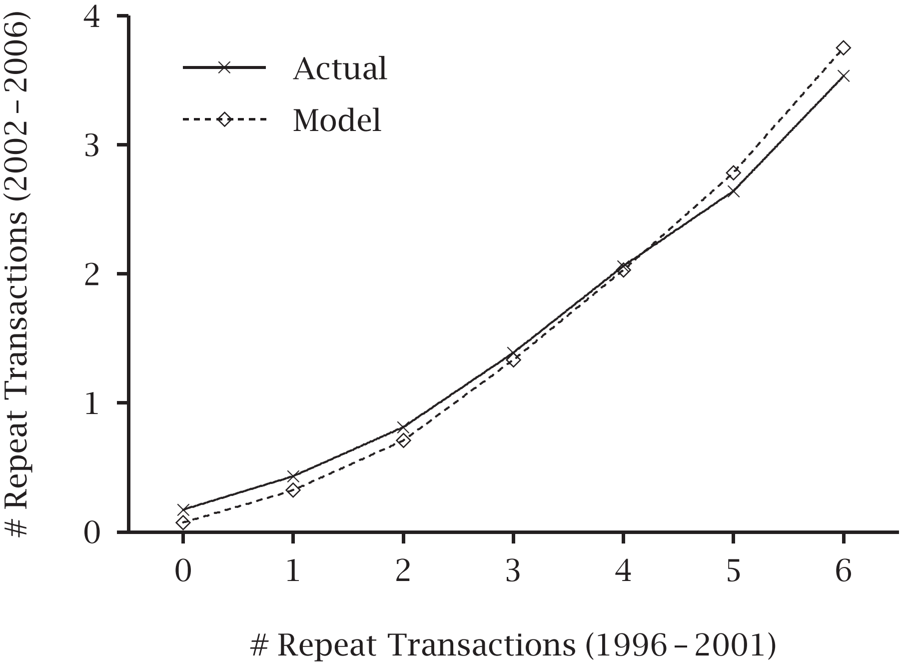
Expected # Transactions in 2002 – 2006 as a Function of Recency and Frequency

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	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

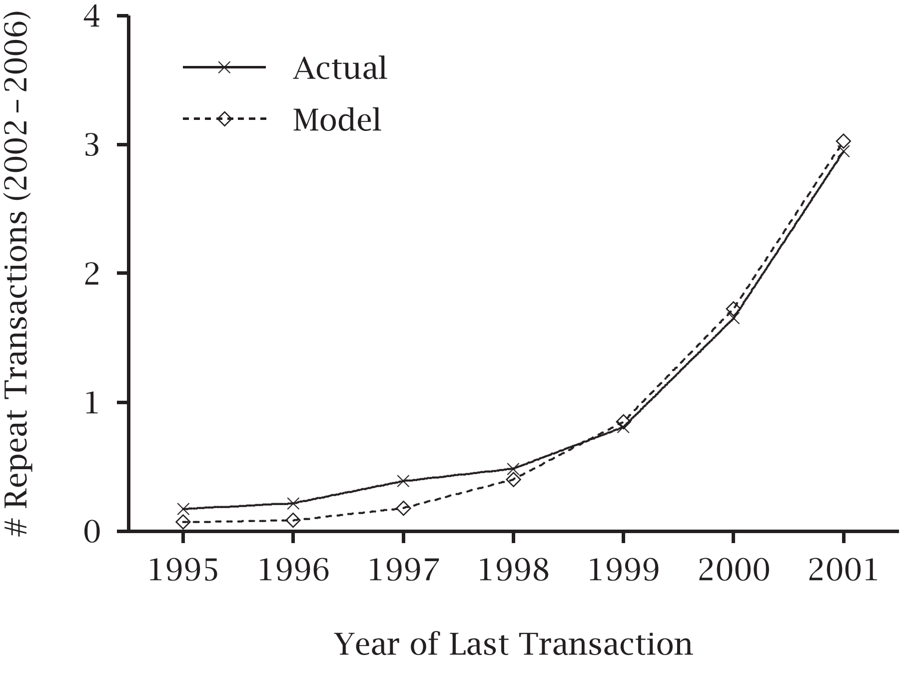
Actual Average # Transactions in 2002 – 2006 as a Function of Recency and Frequency

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	1995	1996	1997	1998	1999	2000	2001
0	0.17						
1		0.22	0.37	0.60	0.56	1.14	1.47
2			0.40	0.46	0.74	1.41	1.89
3				0.46	0.94	1.66	2.29
4					0.84	1.91	2.72
5						1.74	3.06
6							3.53

Conditional Expectations by Frequency



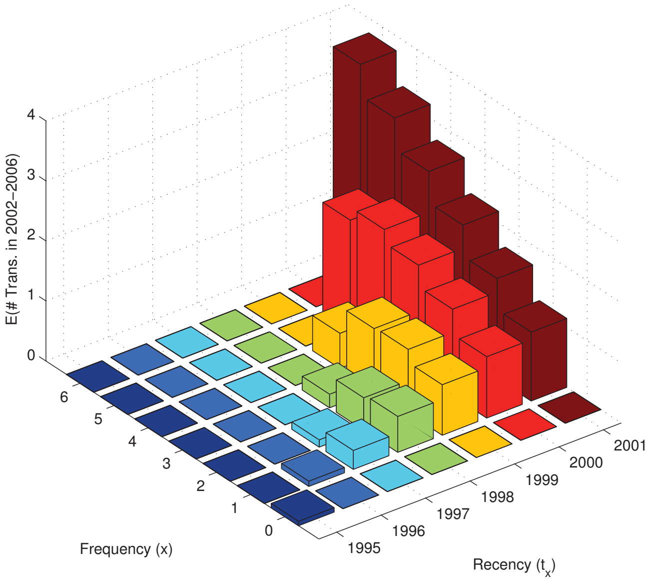
Conditional Expectations by Recency



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

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	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

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



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

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	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

Posterior Mean of *P* as a Function of Recency and Frequency

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	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

Moving Beyond a Single Cohort

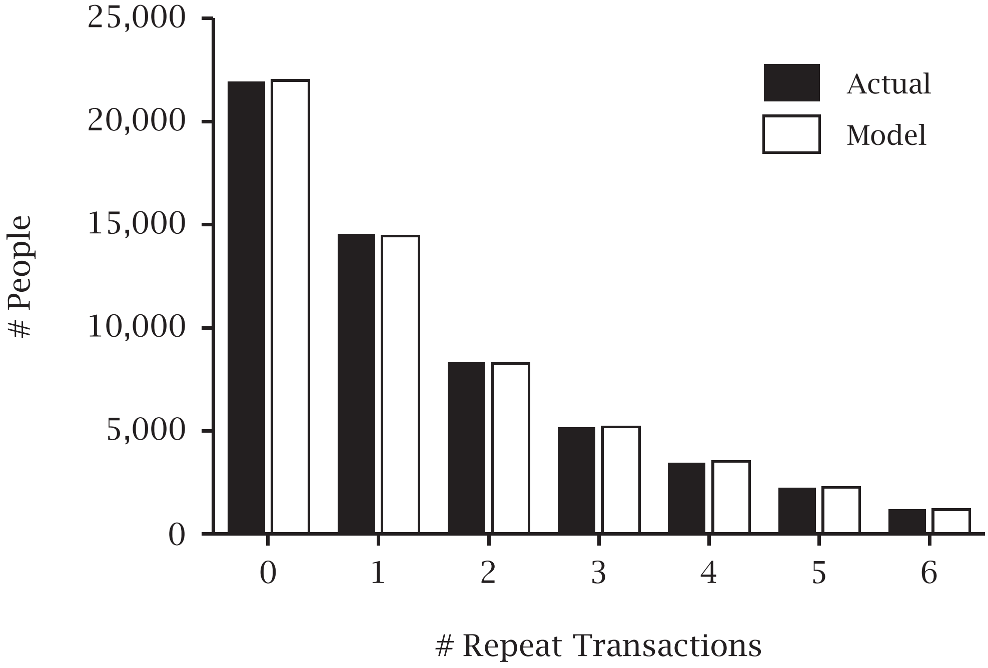
Cohort	Size
1995	11,104
1996	10,057
1997	9,043
1998	8,175
1999	8,977
2000	9,491

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

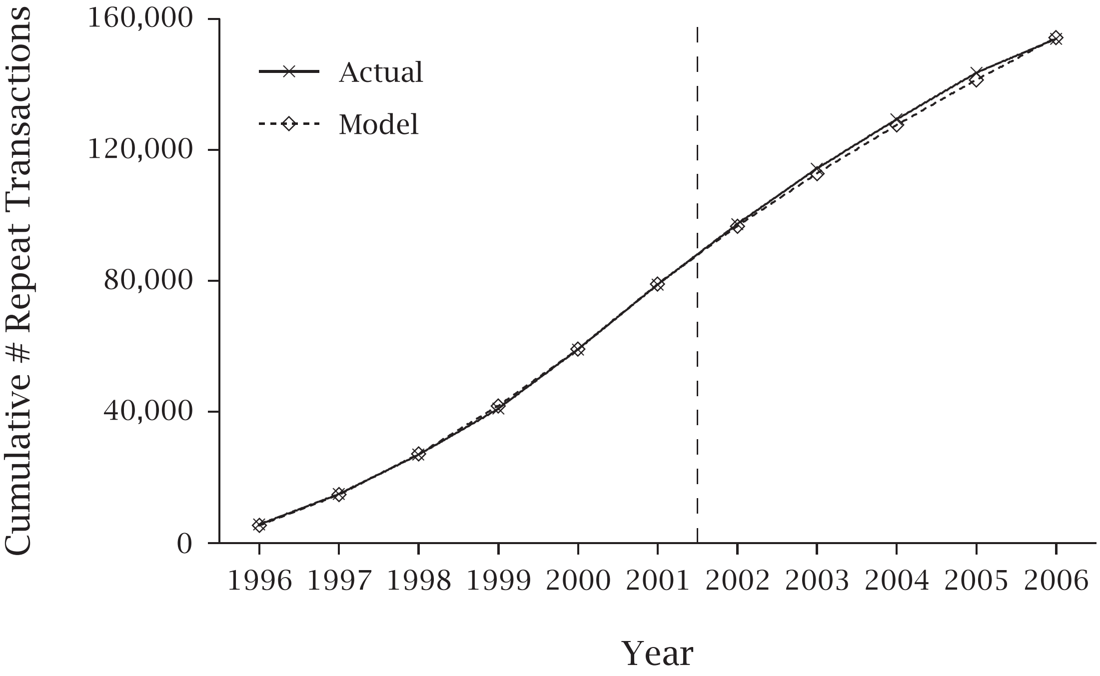
Parameter Estimates (Pooled)

	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

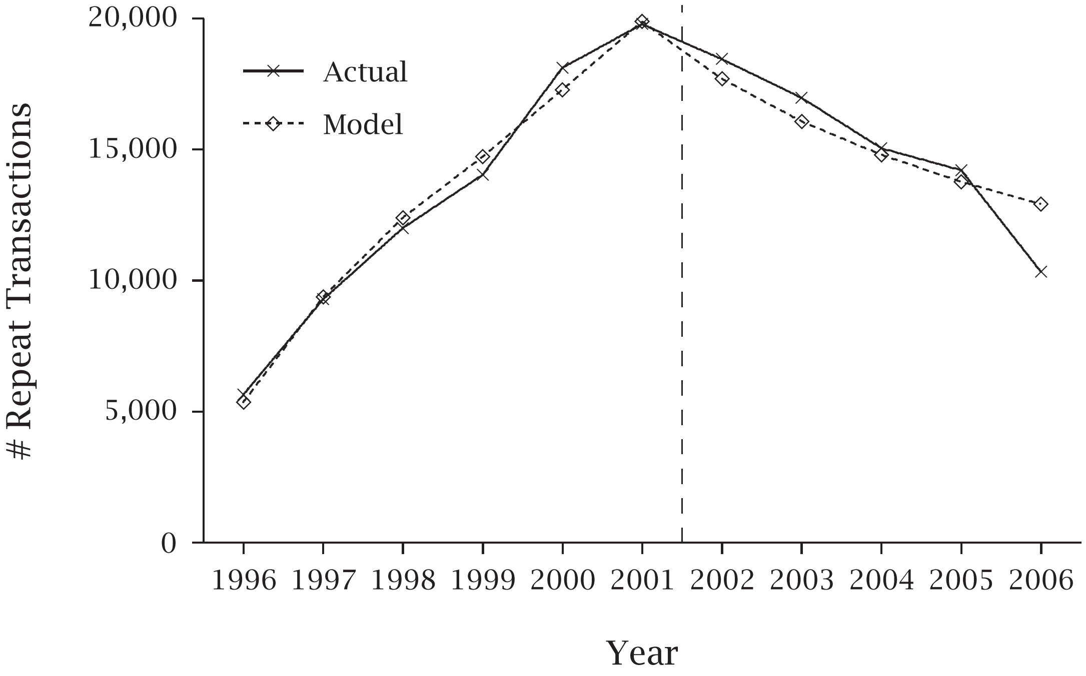
Fit of the BG/BB Model



Tracking Cumulative Repeat Transactions



Tracking Annual Repeat Transactions



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, $v(t)$ = net cashflow / transaction × $t(t)$ (where $t(t)$ is the transaction rate at t) and

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

Computing $E(RLV)$

- Standing at time T,

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

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

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

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

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

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	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

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

# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	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

DERT versus Conditional Expectations

- For any given analysis setting, the DERT numbers differ from the conditional expectations by a constant, independent of the customer's exact purchase history.
- In this empirical setting, DERT = 1.575 × CE.
- As a result, any ranking of customers on the basis of DERT will be exactly the same as that derived using the conditional expectation of purchasing over the next n^* periods.

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/`>

"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

From Discrete to Continuous Time

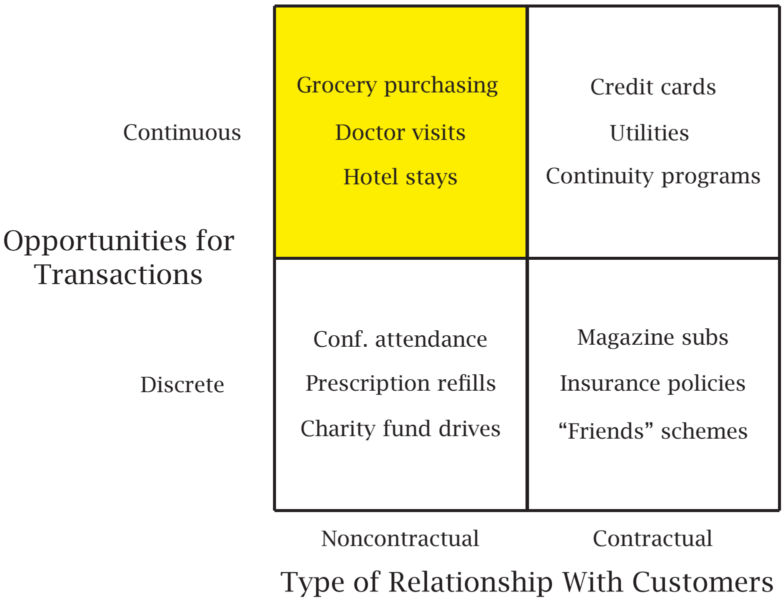
- Suppose we have a year of data from Amazon.
- Should we define
  - 12 monthly transaction opportunities?
  - 52 weekly transaction opportunities?
  - 365 daily transaction opportunities?

From Discrete to Continuous Time

As the number of divisions of a given time period ⟶ ∞

binomial	⟶	Poisson
beta-binomial	⟶	NBD
geometric	⟶	exponential
beta-geometric	⟶	Pareto Type II
BG/BB	⟶	Pareto/NBD

Classifying Customer Bases



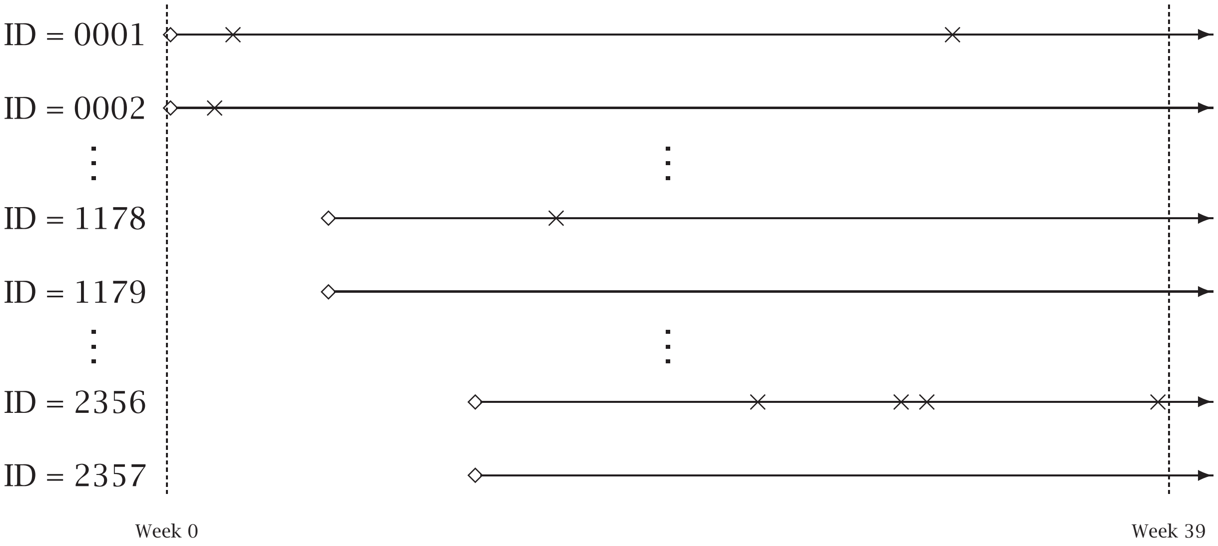
Opportunities for Transactions	Noncontractual	Contractual
Continuous	Grocery purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies “Friends” schemes

Type of Relationship With Customers

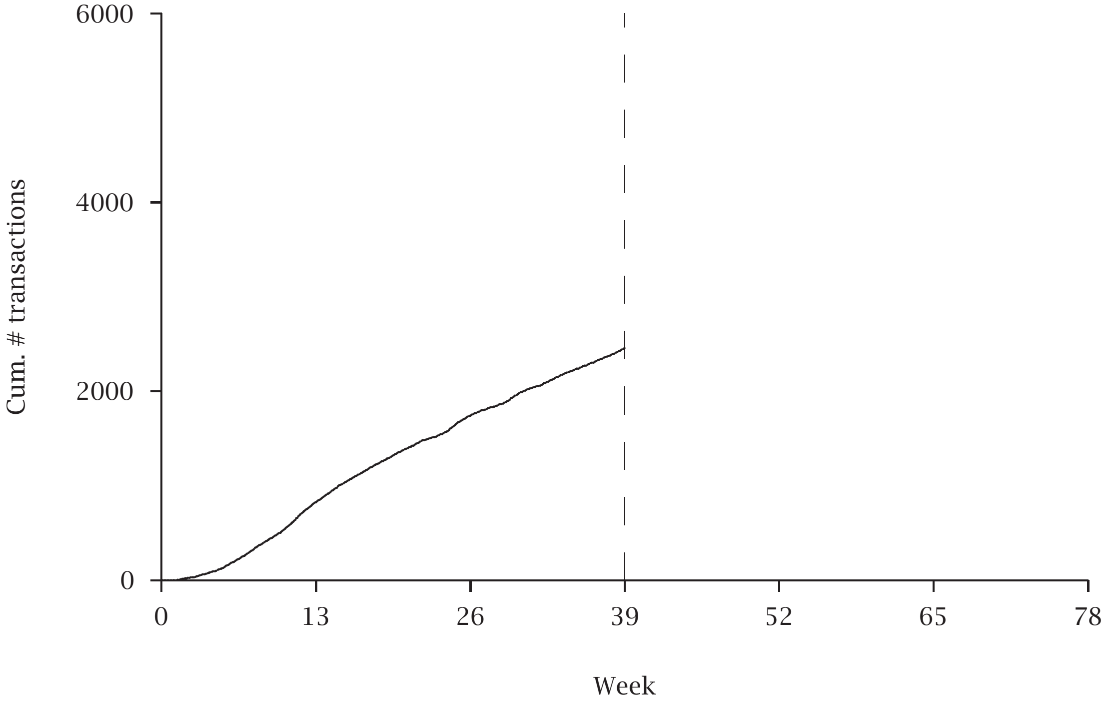
Setting

- 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's relationship with a firm has two phases: they are "alive" for an unobserved period of time, then "dead."

Transaction Process:

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

Latent Attrition Process:

- Each customer has an unobserved "lifetime," which is a function of their death rate.
- Death rates vary across customers.

The Pareto/NBD Model

(Schmittlein, Morrison and Colombo 1987)

Transaction Process:

- While alive, the number of transactions made by a customer follows a Poisson process with mean transaction rate λ.
- Heterogeneity in transaction rates across customers is distributed gamma(r, α).

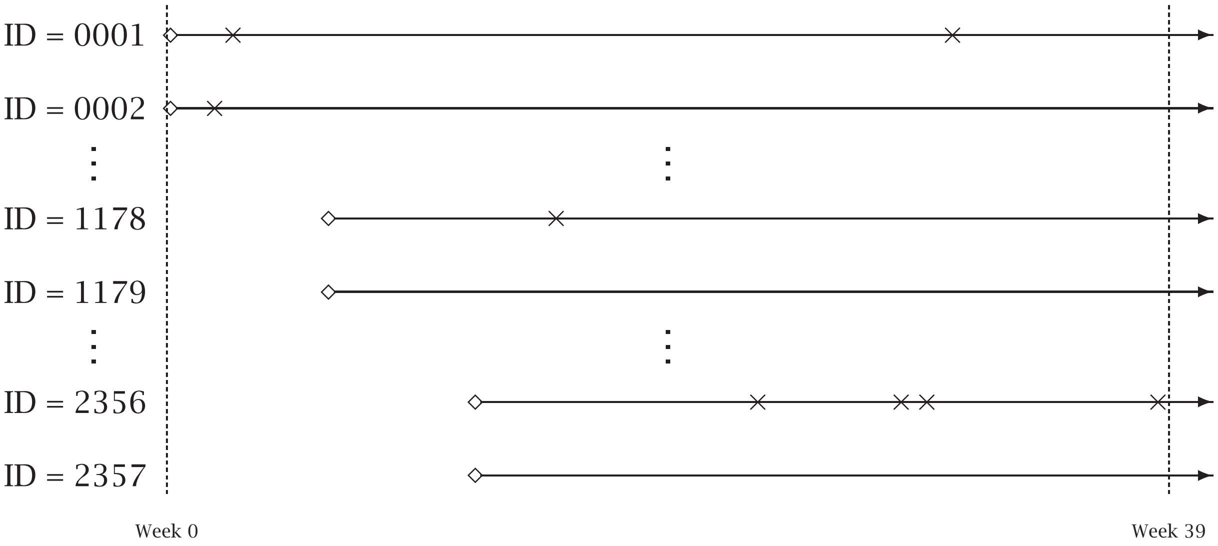
Latent Attrition Process:

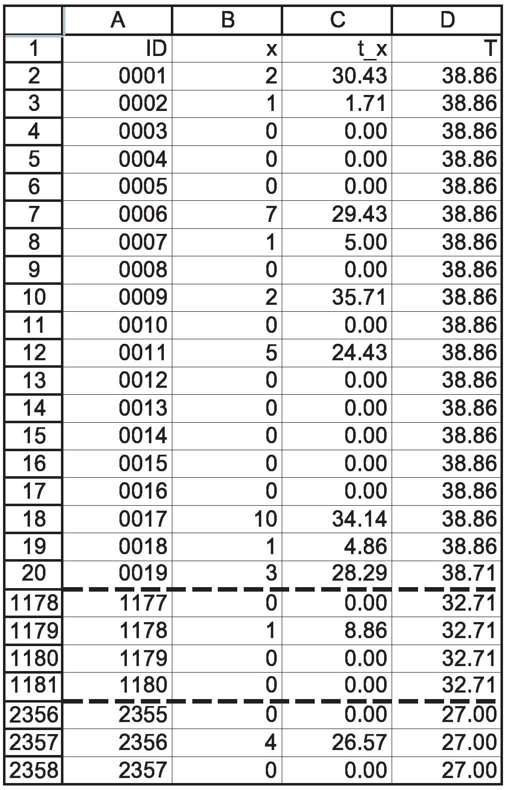
- Each customer has an unobserved "lifetime" of length ω, which is distributed exponential with death rate μ.
- Heterogeneity in death rates across customers is distributed gamma(s, β).

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





	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

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

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

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

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

Key Results

$E[X(t)]$

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

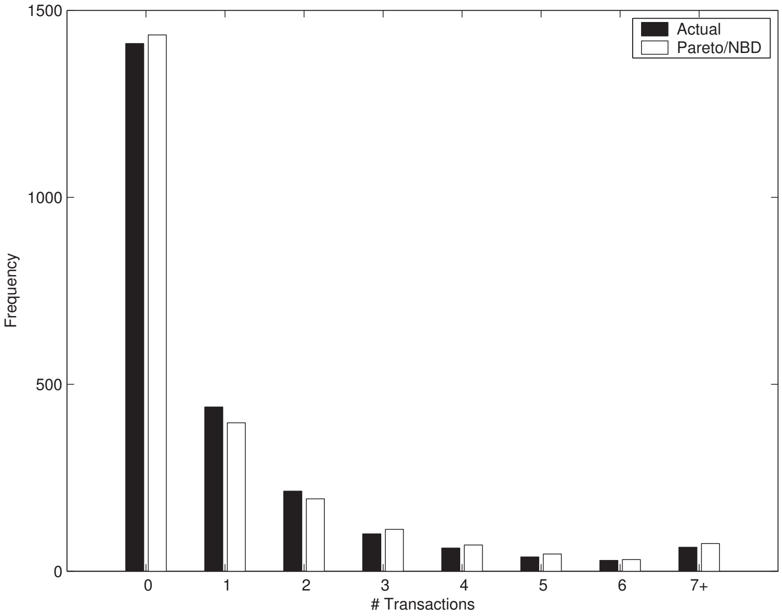
$P(\text{alive} \mid x, t_x, T)$

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

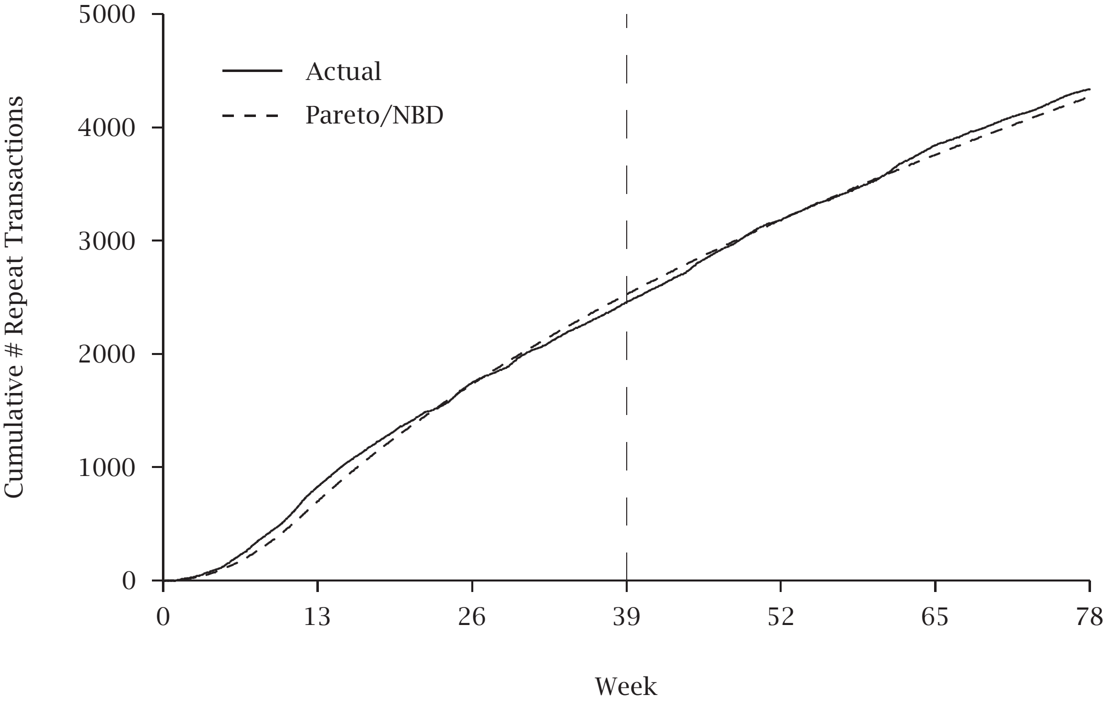
$E[X(T, T+t) \mid x, t_x, T]$

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

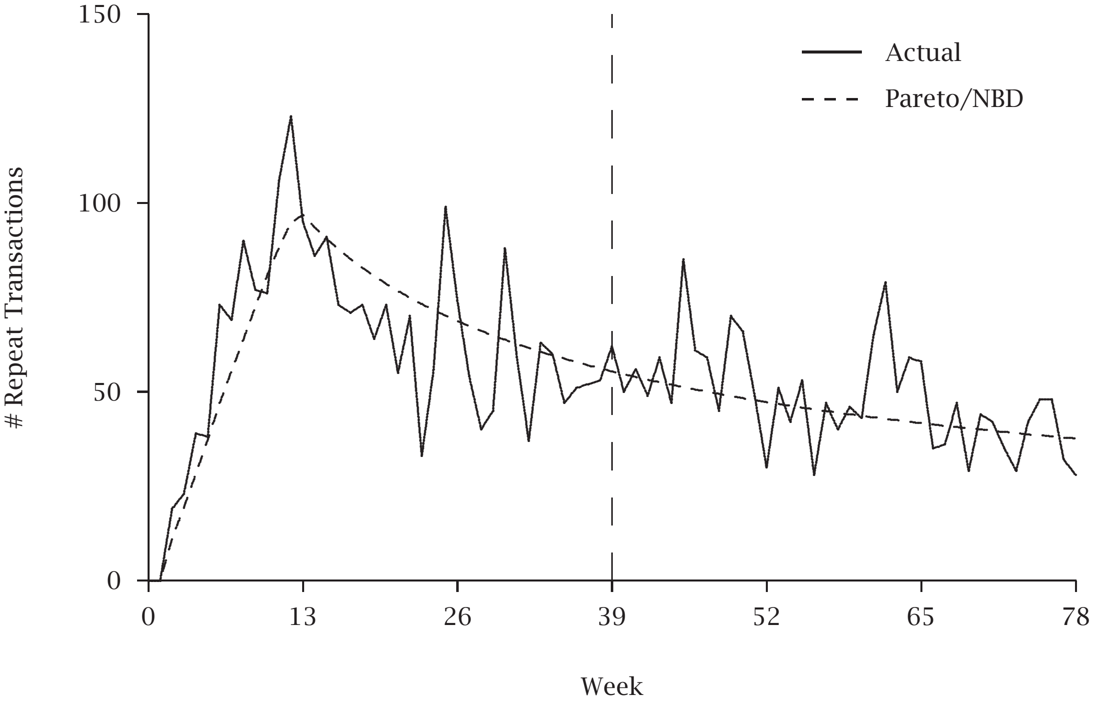
Frequency of Repeat Transactions



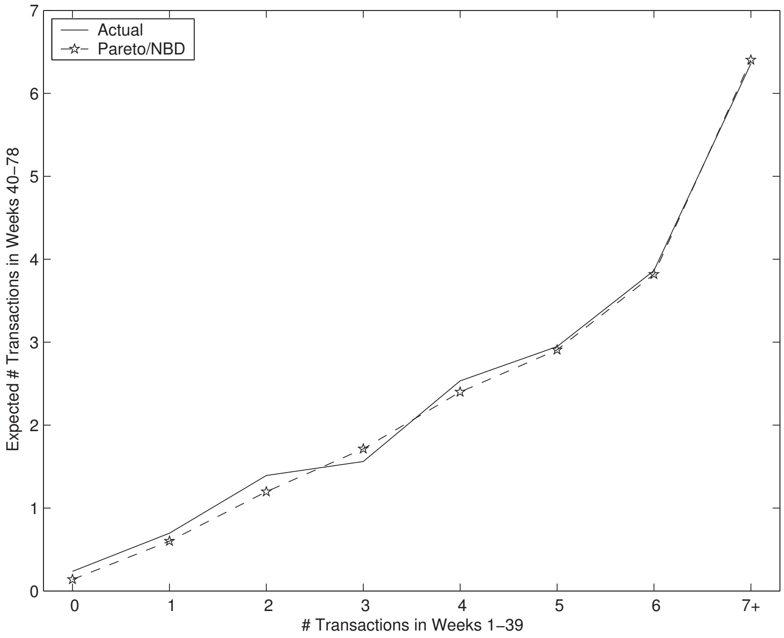
Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions



Conditional Expectations



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

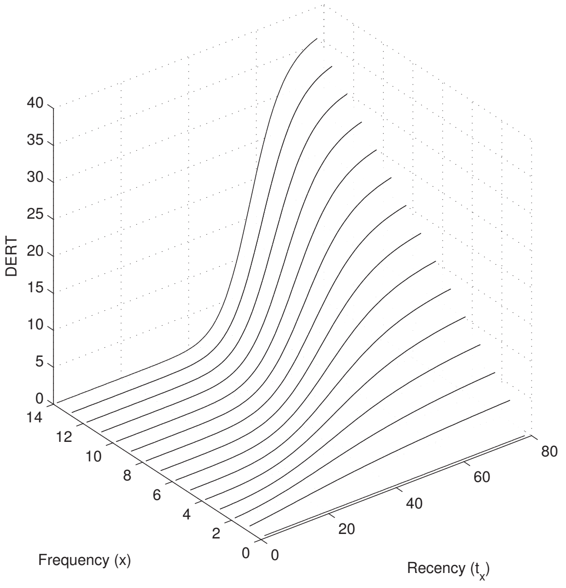
$$\begin{aligned} &DERT(\delta \mid r, \alpha, s, \beta, x, t_x, T) \\ &\quad = \int_0^\infty \int_0^\infty \Big\{ DERT(\delta \mid \lambda, \mu, \text{alive at } T) \\ &\qquad\qquad \times P(\text{alive at } T \mid \lambda, \mu, x, t_x, T) \\ &\qquad\qquad \times g(\lambda, \mu \mid r, \alpha, s, \beta,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

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

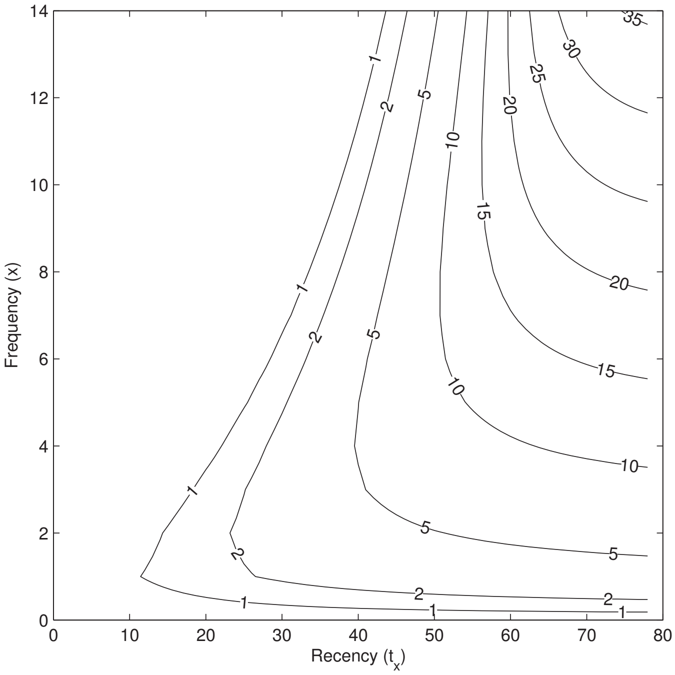
Continuous Compounding

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

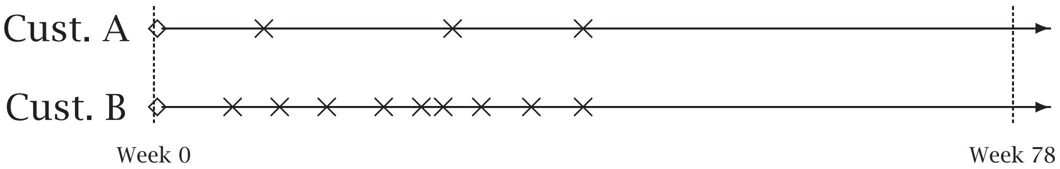
DERT by Recency and Frequency



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:

$$DERT = f(R, F)$$

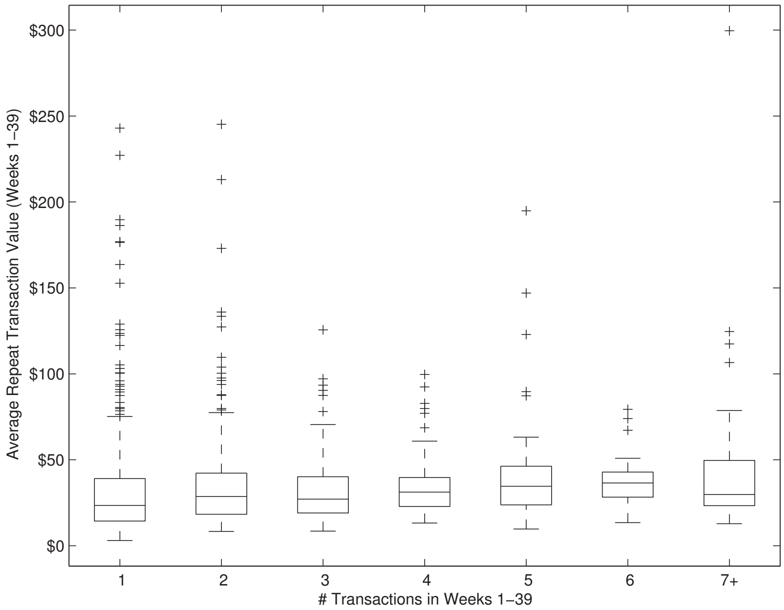
- Adding a sub-model for spend per transaction enables us to generate forward-looking estimates of an individual's expected *residual* revenue stream conditional on his observed behavior (RFM):

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

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.

Independence of the Spend Process



corr(# transactions, avg. transaction value) = 0.06

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 observed average transaction value

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

  is an imperfect estimate of his (unobserved) mean transaction value ζ.
- Our goal is to make inferences about ζ given $\bar{z}$, which we denote as $E(Z \mid \bar{z}, x)$.

Summary of Average Transaction Value

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

	\$
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

Modelling the Spend Process

Given the assumptions

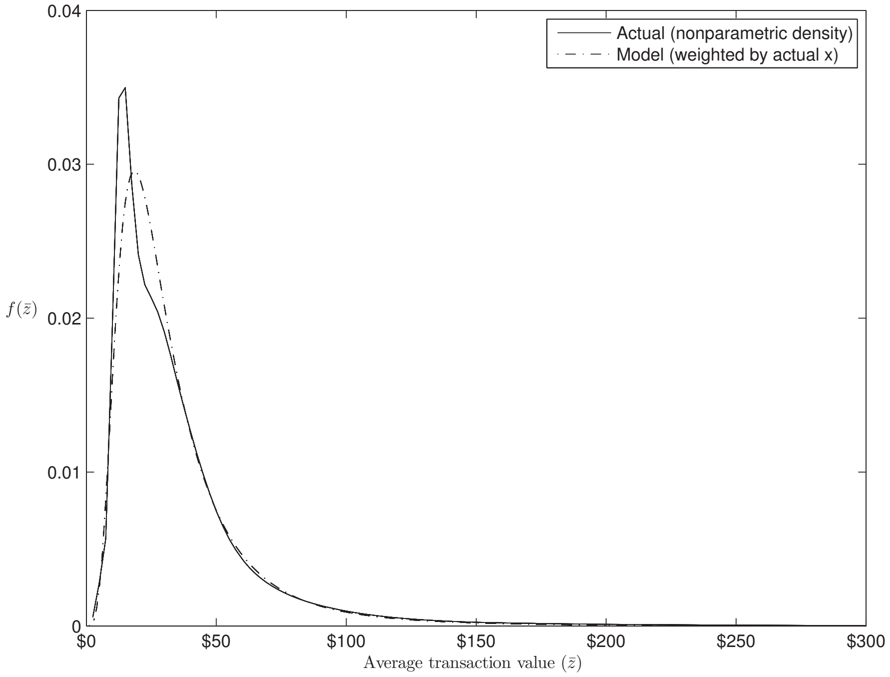
i) The dollar value of a customer's given transaction is distributed gamma with shape parameter p and scale parameter ν (which implies $\zeta = p/\nu$)

ii) Heterogeneity in ν across customers follows a gamma distribution with shape parameter q and scale parameter γ

it follows that the marginal distribution of $\bar{z}$ is

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

Distribution of Average Transaction Value

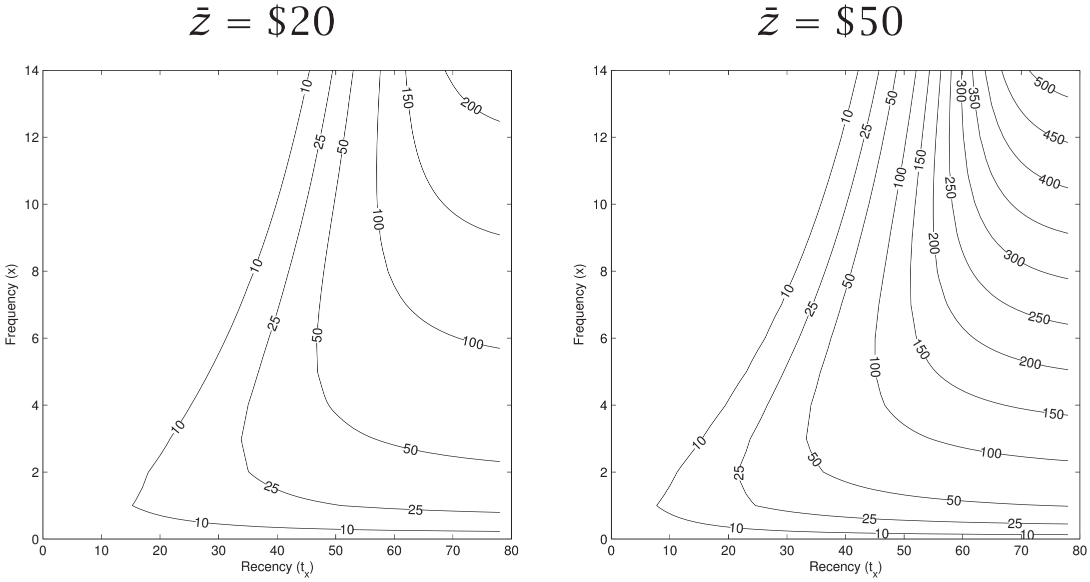


Computing Expected Residual Lifetime Value

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

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

Estimates of $E(RLV)$



$\bar{z} = \$20$ $\bar{z} = \$50$

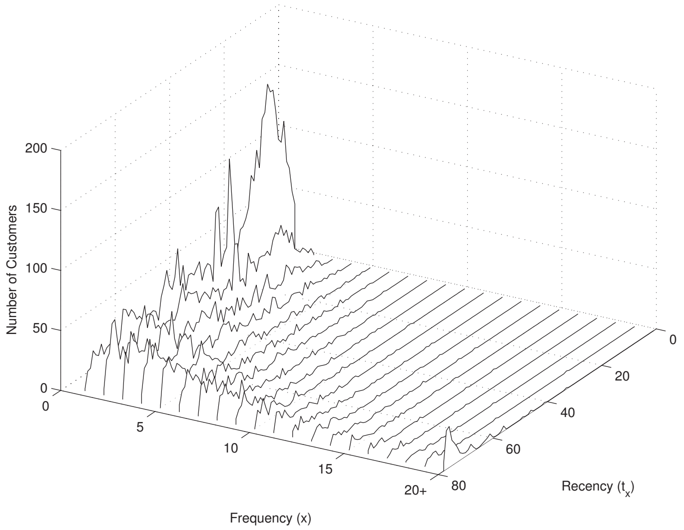
(Margin = 30%, 15% discount rate)

Closing the Loop

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

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

Distribution of Repeat Customers



(12,054 customers make no repeat purchases)

Average $E(RLV)$ by RFM Segment

		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 latent attrition story:
  - Each customer has an unobserved “lifetime”
  - Death rates vary across customers
- Let us consider an alternative story:
  - After any transaction, a customer tosses a coin
    heads → remain alive
    tails → dies
  - P(tails) varies across customers

The BG/NBD Model

(Fader, Hardie and Lee 2005c)

Transaction Process:

- While alive, the number of transactions made by a customer follows a Poisson process with mean transaction rate λ.
- Heterogeneity in transaction rates across customers is distributed gamma(r, α).

Latent Attrition Process:

- After any transaction, a customer dies with probability p.
- Heterogeneity in death probabilities across customers is distributed beta(a, b).

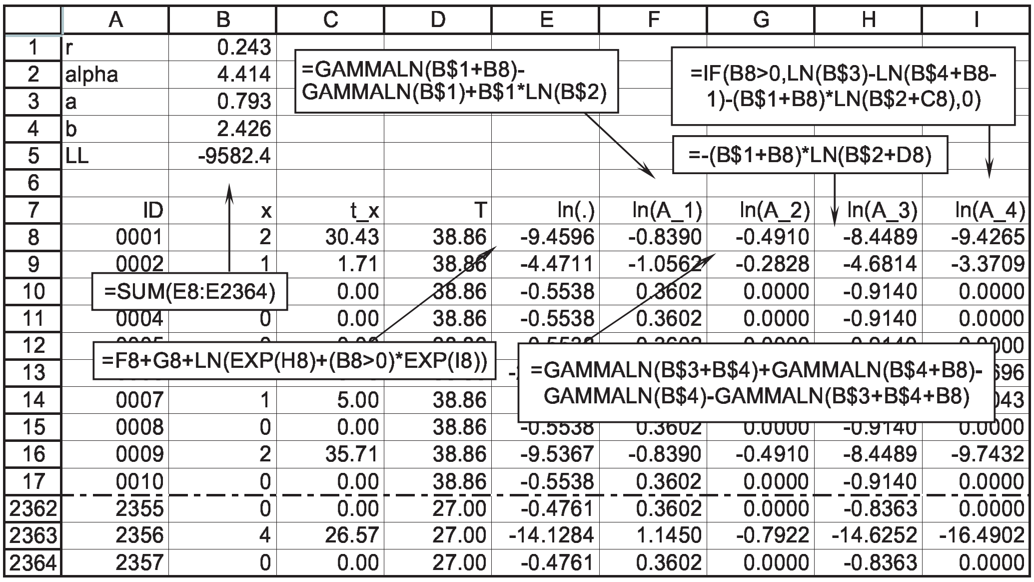
BG/NBD Likelihood Function

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

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



	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									0000
13									596
14	0007	1	5.00	38.86					043
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)

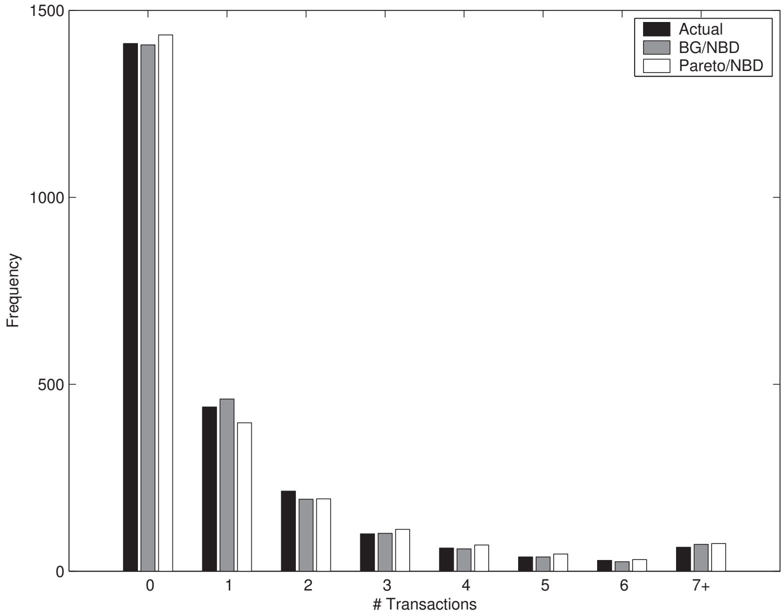
=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))

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

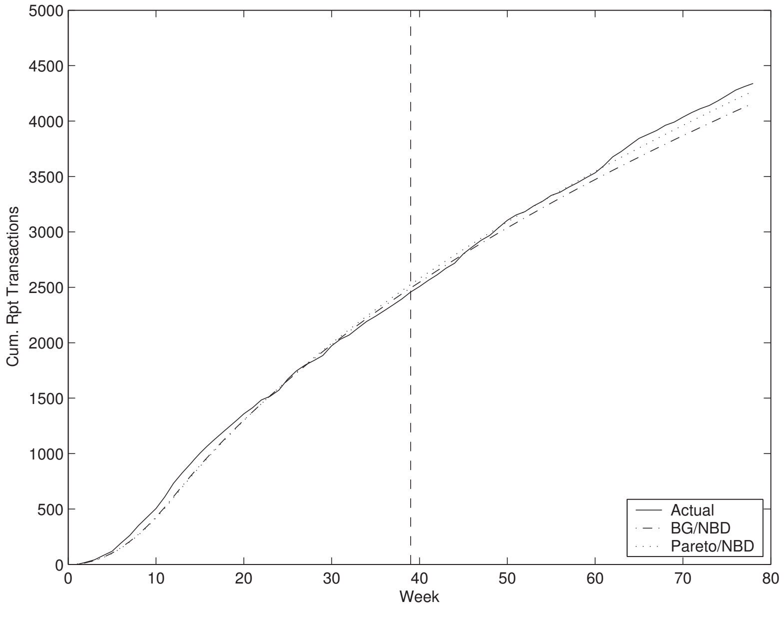
Model Estimation Results

	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

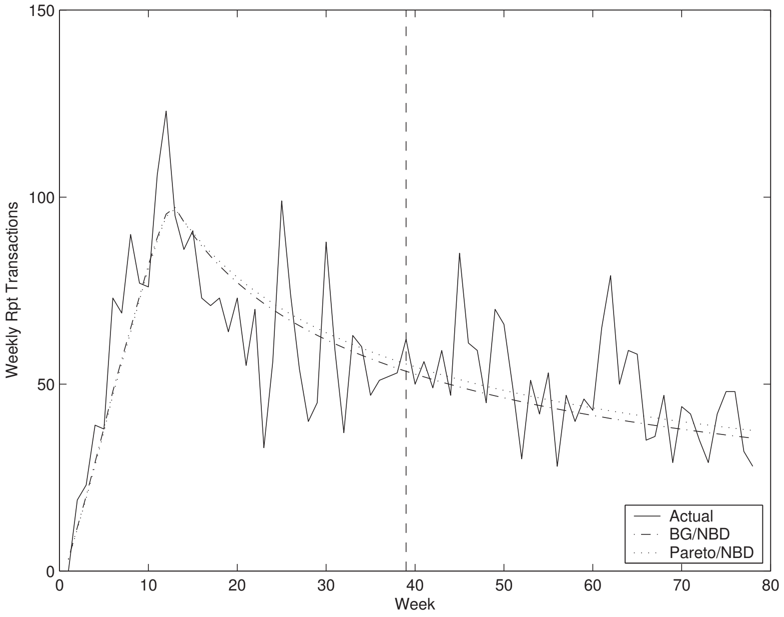
Frequency of Repeat Transactions



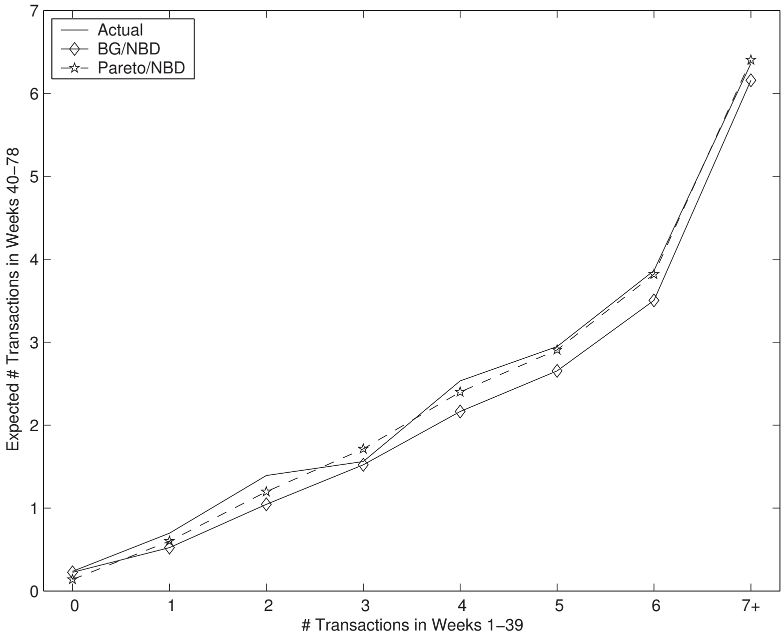
Tracking Cumulative Repeat Transactions



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), “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 (2005b), “A Note on Implementing the Pareto/NBD Model in MATLAB.” `<http://brucehardie.com/notes/008/>`

R Package “BTYD: Implementing Buy ‘Til You Die Models.” `<http://cran.r-project.org/package=BTYD>`

Further Reading

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/`>

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 (2013), “The Gamma-Gamma Model of Monetary Value.” <`http://brucehardie.com/notes/025/`>

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)

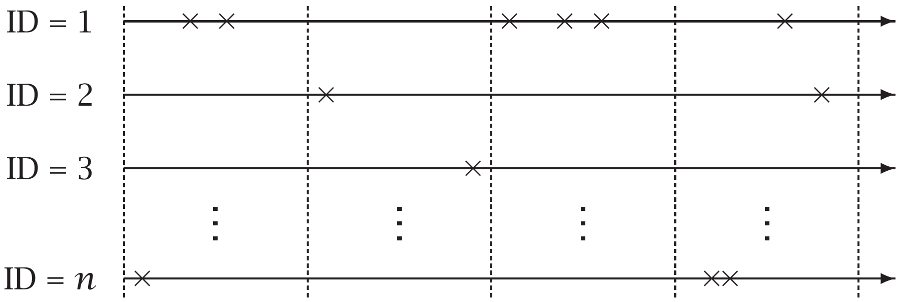
Cohort	Calendar Time →				
1	n_{11}				n_{1I}
2		n_{22}			n_{2I}
⋮					⋮
I−1				$n_{I-1,I-1}$	$n_{I-1,I}$
I					n_{II}

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

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

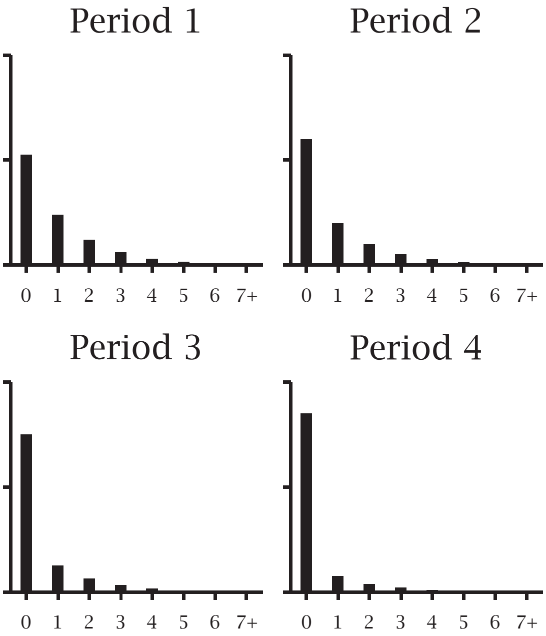
Cohort	Calendar Time →				
1				n_{1I-1}	n_{1I}
2				n_{2I-1}	n_{2I}
⋮				⋮	⋮
I−1				$n_{I-1,I-1}$	$n_{I-1,I}$
I					n_{II}

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)

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 (2013), “Customer-Base Analysis on a ‘Data Diet’: Model Inference Using Repeated Cross-Sectional Summary (RCSS) Data.” <http://brucehardie.com/papers/025/>

Model Extensions

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

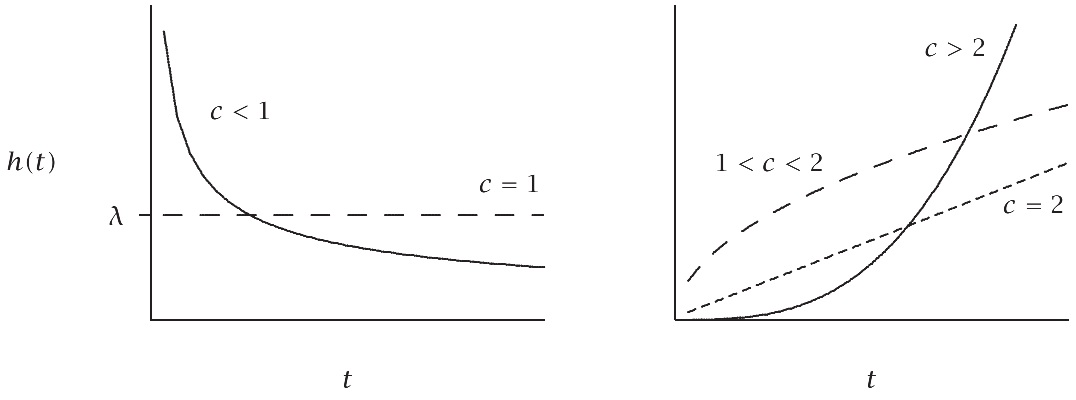
- A generalization of the exponential distribution that can have an increasing and decreasing hazard function:

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

  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 (generalized) Burr Type XII 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.

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 are more flexible … but there are 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/`>

Covariates

- Types of covariates:
  - customer characteristics (e.g., demographics, attitudes)
  - marketing activities
  - competition
  - “macro” factors
- Handling covariate effects:
  - explicit integration (via latent characteristics)
  - 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

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.

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.

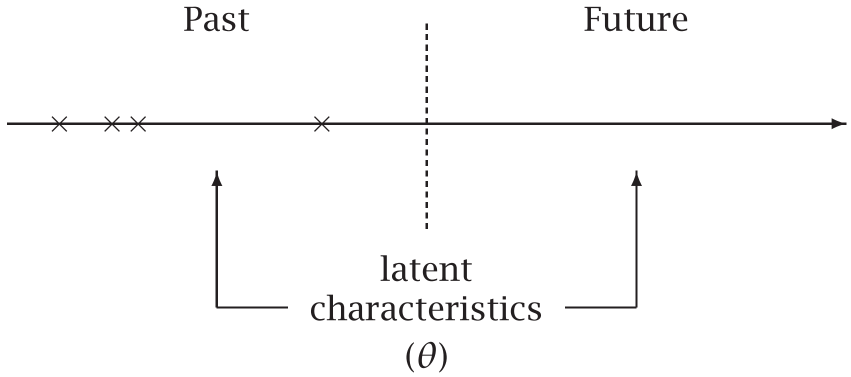
Make everything as simple as possible, but not simpler.

Albert Einstein

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 purchasing Doctor visits Hotel stays	Credit cards Utilities Continuity programs
Discrete	Conf. attendance Prescription refills Charity fund drives	Magazine subs Insurance policies “Friends” schemes

Type of Relationship With Customers