Probability Models for Customer-Base Analysis

Peter S. Fader
University of Pennsylvania

Bruce G. S. Hardie
London Business School

27th Annual Advanced Research Techniques Forum
June 26–29, 2016

Motivating Problems

Motivating Problem 1

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

ID	Year 1	Year 2	Year 3	Year 4	Year 5
0001	1	1	0	0	0
0002	1	0	0	0	0
0003	1	1	1	0	0
0004	1	1	0	0	0
0005	1	1	1	1	1
0006	1	0	0	0	0
⋮		⋮		⋮	
0998	1	0	0	0	0
0999	1	1	1	0	0
1000	1	0	0	0	0
	1000	631	468	382	326

Motivating Problem 1

Year 1	Year 2	Year 3	Year 4	Year 5
1000	631	468	382	326

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 Problem 1

Q1a: Assuming our current prospect pool has the same characteristics of that from which these customers were acquired, what is the maximum amount you would be willing to spend to acquire a customer?

Q1b: What is the expected *residual* value of this group of 326 customers at the end of Year 5?

Spending on Customer Acquisition

	Year 1	Year 2	Year 3	Year 4	Year 5
P(still a customer)	1.000	0.631	0.468	0.382	0.326
Net CF	\$100	\$100	\$100	\$100	\$100
discount	1	$\frac{1}{1.1}$	$\frac{1}{(1.1)^2}$	$\frac{1}{(1.1)^3}$	$\frac{1}{(1.1)^4}$

Spending on Customer Acquisition

Standing at the beginning of Year 1, the discounted expected value of a customer is

$$\$100 + \$100 \times \frac{0.631}{1.1} + \$100 \times \frac{0.468}{(1.1)^2} + \$100 \times \frac{0.382}{(1.1)^3} + \$100 \times \frac{0.326}{(1.1)^4} = \$247$$

$\Longrightarrow$ We can justify spending up to \$247 to acquire a new customer (based on expected "profitability" over the five year period).

What's wrong with this analysis?

Spending on Customer Acquisition

Problem:

- We are ignoring any cashflow the customer could possibly generate after Year 5.
- To get a true sense of (expected) customer lifetime value, we need to know the probability that the customer is still a customer in year 6, year 7, and so on.

Spending on Customer Acquisition

Solution:

- We know that $326/382 \times 100\% = 85.3\%$ of the Year 4 customers renewed at the end of Year 4.
- Let us assume that this renewal rate will hold ad infinitum.

$$\begin{aligned} P(\text{still a customer in Year 6}) &= 0.326 \times 0.853 \\ &= 0.278 \\ P(\text{still a customer in Year 7}) &= 0.326 \times (0.853)^2 \\ &= 0.237 \end{aligned}$$

Spending on Customer Acquisition

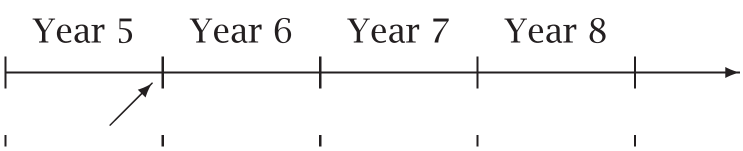
Standing at the beginning of Year 1, the discounted expected value of a customer is

$$\$100 + \$100 \times \frac{0.631}{1.1} + \$100 \times \frac{0.468}{(1.1)^2} + \$100 \times \frac{0.382}{(1.1)^3}$$
$$+ \$100 \times \frac{0.326}{(1.1)^4} + \$100 \times \frac{0.326 \times 0.853}{(1.1)^5}$$
$$+ \$100 \times \frac{0.326 \times (0.853)^2}{(1.1)^6} + \cdots = \$324$$

$\Longrightarrow$ By looking beyond Year 5, we can justify spending up an additional \$77 to acquire a new customer.

Residual Value of the Customer Base

Q1b: What is the expected *residual* value of this group of 326 customers at the end of Year 5?



	Year 5	Year 6	Year 7	Year 8	
P(still a customer)		0.853	$(0.853)^2$	$(0.853)^3$	$\cdots$
Net CF		\$100	\$100	\$100	$\cdots$
discount		1	$\frac{1}{(1.1)}$	$\frac{1}{(1.1)^2}$	$\cdots$

Residual Value of the Customer Base

Standing at the end of Year 5, the discounted expected (residual) value of a Year 5 customer is

$$
\begin{aligned}
&\$100 \times 0.853 + \$100 \times \frac{0.853^2}{1.1} + \ldots \\
&= \$100 \times \sum_{t=1}^{\infty} \frac{0.853^t}{(1.1)^{t-1}} \\
&= \$381
\end{aligned}
$$

$\Longrightarrow$ The expected residual value of the group of customers at the end of Year 5 is $326 \times \$381 = \$124{,}206$.

What's wrong with this analysis?

Motivating Problem 1

- In order to compute the probability that someone is still a customer in Years 6, 7, ..., we have assumed a constant retention rate from Year 4 onwards.
- However retention rates are increasing over time:

$$\text{Year 1:} \quad \frac{631}{1000} = 0.631 \qquad \text{Year 2:} \quad \frac{468}{631} = 0.742$$

$$\text{Year 3:} \quad \frac{382}{468} = 0.816 \qquad \text{Year 4:} \quad \frac{326}{382} = 0.853$$

$\Longrightarrow$ We need a tool for forecasting survival that captures the phenomenon of increasing retention rates.

Motivating Problem 2

- A charity located in the Midwestern United States that is funded in large part by donations from individual supporters.
- We have data for the 11,104 people first-time supporters acquired in 1995.

Assume:

- The value of the average donation is $50 (which is received at the beginning of the year).
- A 10% discount rate.

ID	1995	1996	1997	1998	1999	2000	2001
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
	11104	5652	4674	4019	3552	3555	3163

Motivating Problem 2

Q2a: Assuming our current prospect pool has the same characteristics of that from which these donors were acquired, what is the maximum amount you would spend to acquire a new donor?

Q2b: Given their donation behavior to date, in how many of the subsequent five years can we expect a supporter to make a donation?

What about 100004 (who made repeat donations in four years with the last occurring in 2001) versus 100009 (who made repeat donations in five years with the last occurring in 2000)?

Spending on Customer Acquisition

	1995	1996	1997		2001
P(donate)	1.0	$\frac{5{,}652}{11{,}104}$	$\frac{4{,}674}{11{,}104}$	$\cdots$	$\frac{3{,}163}{11{,}104}$
Net CF	\$50	\$50	\$50	$\cdots$	\$50
discount	1	$\frac{1}{(1.1)}$	$\frac{1}{(1.1)^2}$	$\cdots$	$\frac{1}{(1.1)^6}$

Spending on Customer Acquisition

The discounted expected value of a donor is

$$\begin{aligned}\$50 + \$50 \times \frac{0.509}{1.1} + \$50 \times \frac{0.421}{(1.1)^2} \\ + \$50 \times \frac{0.362}{(1.1)^3} + \$50 \times \frac{0.320}{(1.1)^4} \\ + \$50 \times \frac{0.320}{(1.1)^5} + \$50 \times \frac{0.285}{(1.1)^6} = \$133\end{aligned}$$

$\Longrightarrow$ We can justify spending up to \$133 to acquire a new customer (based on expected donation behavior over the seven year period).

What's wrong with this analysis?

Motivating Problem 2

- We are ignoring any donations we could receive from the donor beyond Year 7.
- What about Q2b?

  In how many of the subsequent five years can we expect a supporter such as 100004, who made repeat donations in four years with the last occurring in 2001, to make a donation? ...

$\Longrightarrow$ We need a model of donation incidence that can be used to predict future behavior, both in the aggregate and conditional on past behavior.

Customer Lifetime Value

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

Two key questions:

- How long will the customer remain “alive”?
- What is the net cashflow per period while “alive”?

Q: How long will the customer remain “alive”?

A: It depends on the business setting . . .

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 year ended 31 March 2014, we see that Vodafone UK had 11.7 million "pay monthly" customers at the end of that period.
- In his "Q2 2014 Earnings Conference Call" the CFO of Amazon made the comment that "[a]ctive customer accounts exceeded 250 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?

Motivating Problem 1

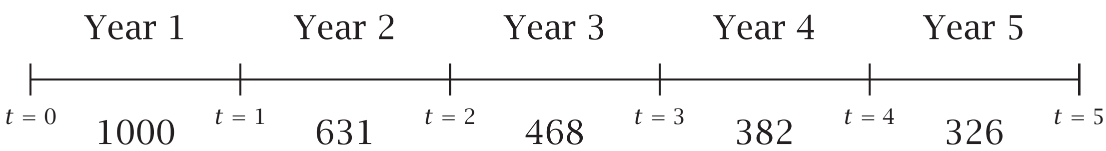
ID	Year 1	Year 2	Year 3	Year 4	Year 5
0001	1	1	0	0	0
0002	1	0	0	0	0
0003	1	1	1	0	0
0004	1	1	0	0	0
0005	1	1	1	1	1
0006	1	0	0	0	0
⋮		⋮		⋮	
0998	1	0	0	0	0
0999	1	1	1	0	0
1000	1	0	0	0	0
	1000	631	468	382	326

Motivating Problem 2

ID	1995	1996	1997	1998	1999	2000	2001
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
⋮		⋮		⋮		⋮	
111102	1	1	1	1	1	1	1
111103	1	0	1	1	0	1	1
111104	1	0	0	0	0	0	0
	11104	5652	4674	4019	3552	3555	3163

Contractual Settings

Notation and Terminology



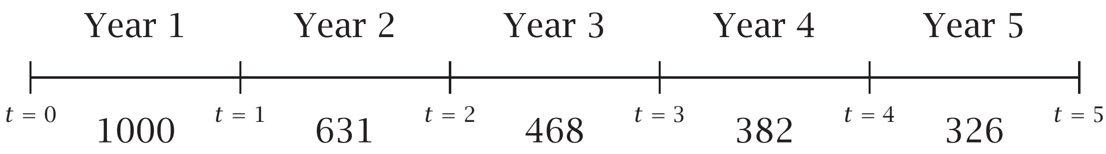
The *survivor function* $S(t)$ is the proportion of the cohort that continue as a customer beyond t.

$$S(0) = ?$$

$$S(1) = ?$$

$$S(2) = ?$$

Notation and Terminology



The *retention rate* is the ratio of customers retained to the number at risk.

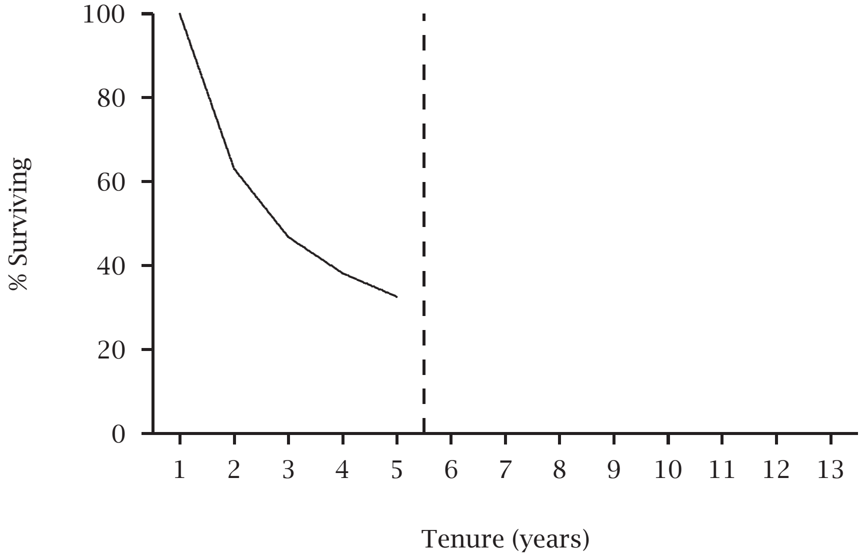
$$r(1) = ?$$

$$r(2) = ?$$

For survivor function $S(t)$, $r(t) = \dfrac{S(t)}{S(t-1)}$.

Modelling Objective

We want to derive a mathematical expression for $S(t)$, which can then be used to generate the desired forecasts.



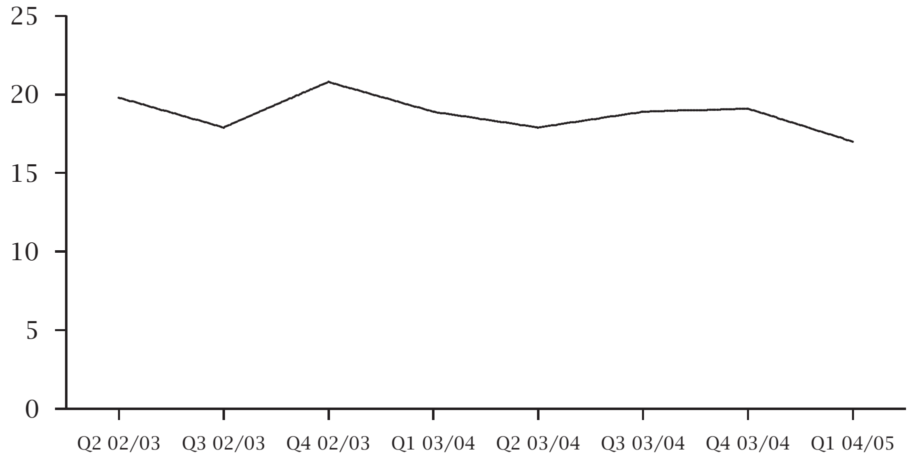
The Phenomenon of Retention Rates

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 (%)

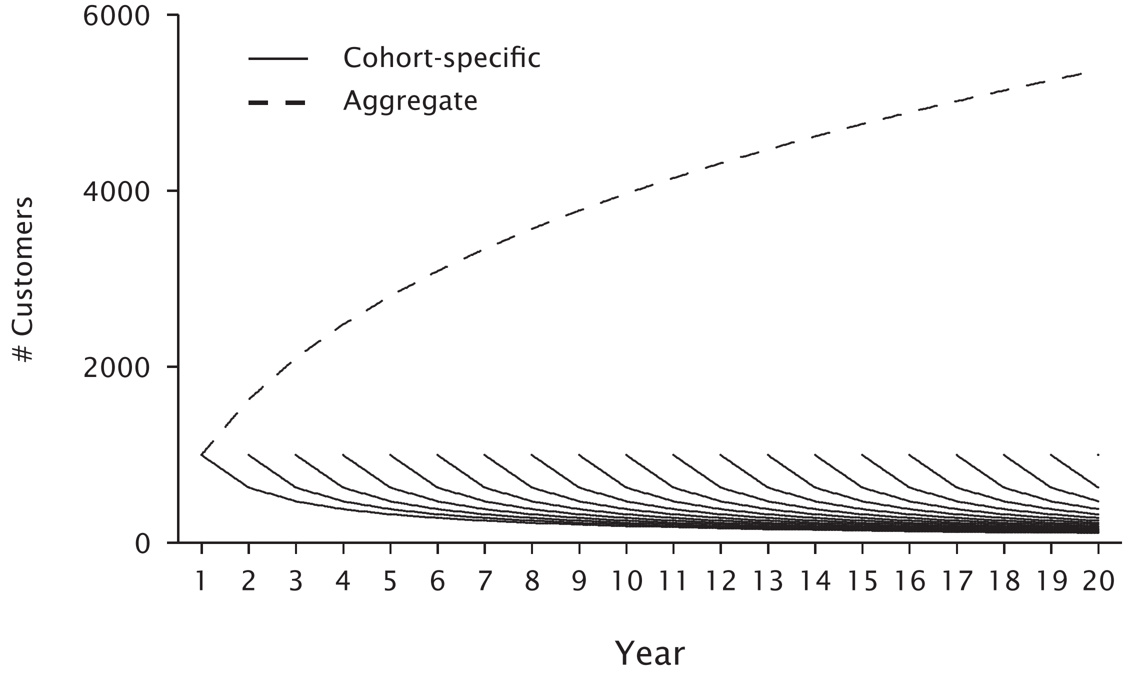


Source: Vodafone Germany "Vodafone Analyst & Investor Day" presentation (2004-09-27)

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



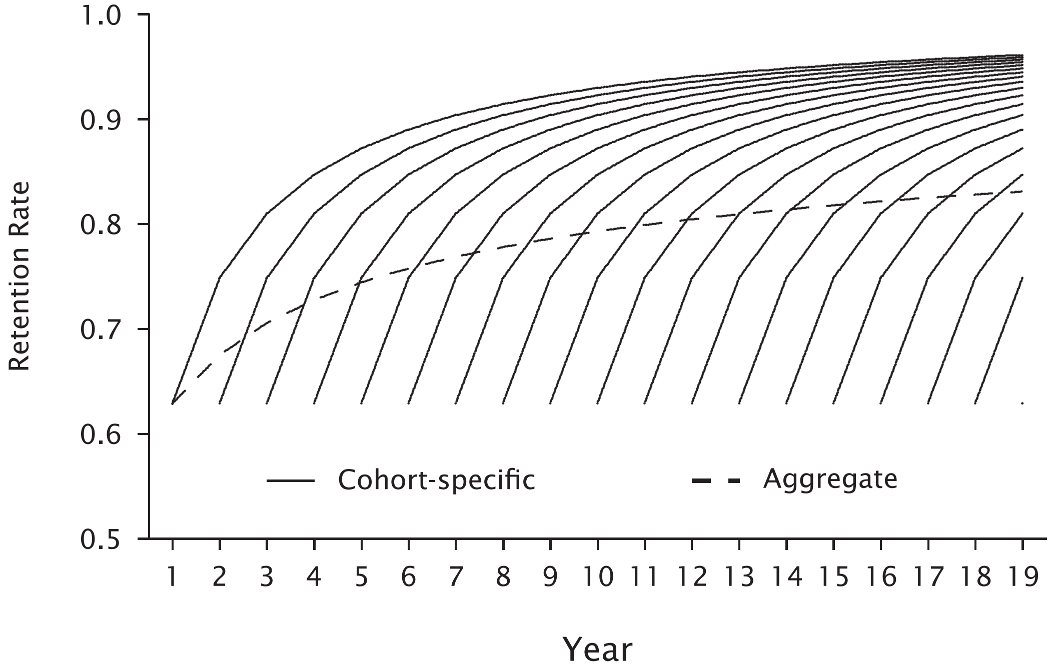
Cohort-level vs. Aggregate Numbers

	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.

The Beta Distribution

- The beta distribution is a flexible (and mathematically convenient) two-parameter distribution bounded between 0 and 1:

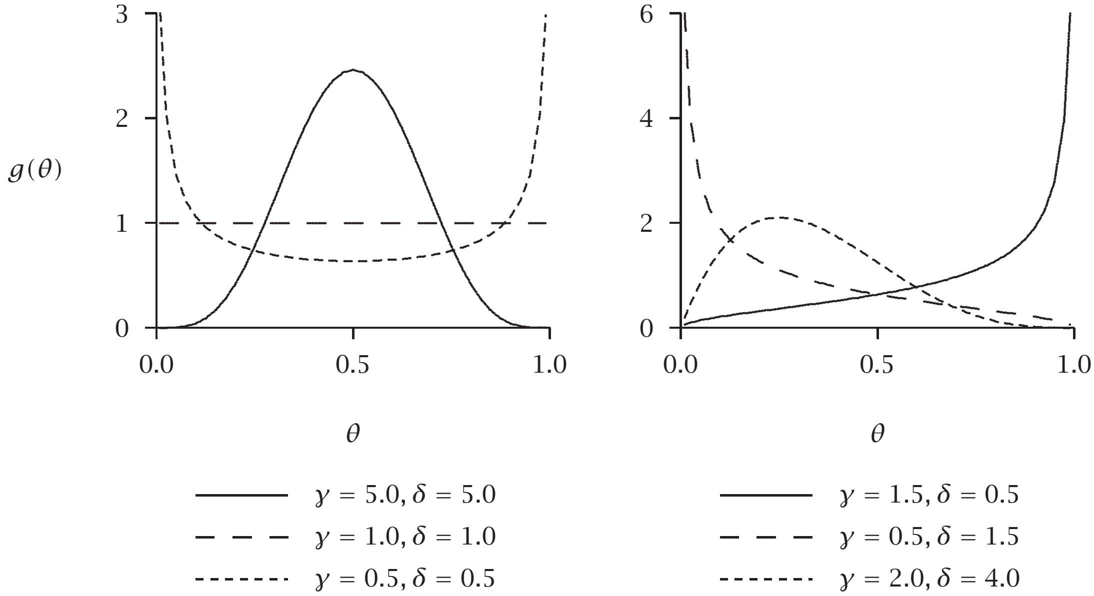
$$g(\theta \mid \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)} \,,$$

where $\gamma, \delta > 0$ and $B(\gamma, \delta)$ is the beta function.

- The mean of the beta distribution is

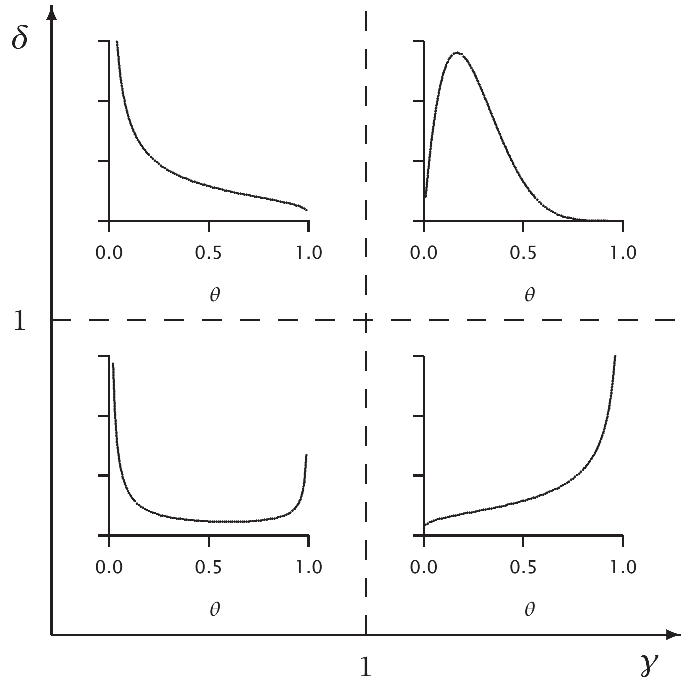
$$E(\Theta) = \frac{\gamma}{\gamma + \delta} \,.$$

Illustrative Beta Distributions



43

Five General Shapes of the Beta Distribution



The Beta Function

- The beta function $B(\gamma, \delta)$ is defined by the integral

$$B(\gamma, \delta) = \int_0^1 t^{\gamma-1}(1-t)^{\delta-1}dt, \;\; \gamma > 0, \delta > 0\,,$$

and can be expressed in terms of gamma functions:

$$B(\gamma, \delta) = \frac{\Gamma(\gamma)\Gamma(\delta)}{\Gamma(\gamma+\delta)}\,.$$

- The gamma function $\Gamma(\gamma)$ is a generalized factorial, which has the recursive property $\Gamma(\gamma + 1) = \gamma\Gamma(\gamma)$. Since $\Gamma(0) = 1$, $\Gamma(n) = (n-1)!$ for positive integer n.

Numerical Evaluation of the Beta Function

- Not all computing environments have a `beta` function (or even a `gamma` function).
- However, we typically have a function that evaluates $\ln\big(\Gamma(\cdot)\big)$, e.g., `gammaln`.
- In Excel,

$$\begin{aligned}\Gamma(\gamma) &= \texttt{exp(gammaln(}\gamma\texttt{))}\\ B(\gamma,\delta) &= \texttt{exp(gammaln(}\gamma\texttt{)+gammaln(}\delta\texttt{)}\\ &\qquad\qquad \texttt{-gammaln(}\gamma\texttt{+}\delta\texttt{))}\end{aligned}$$

A Discrete-Time Model for Contract Duration

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

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

- The aggregate survivor function is

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

A Discrete-Time Model for Contract Duration

- The (aggregate) retention rate is given by

$$\begin{aligned} r(t) &= \frac{S(t)}{S(t-1)} \\ &= \frac{\delta + t - 1}{\gamma + \delta + t - 1}. \end{aligned}$$

- 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 $P(T = 1)$:

$$P(T = t \mid \gamma, \delta) = \begin{cases} \dfrac{\gamma}{\gamma + \delta} & t = 1 \\ \\ \dfrac{\delta + t - 2}{\gamma + \delta + t - 1} \times P(T = t - 1) & t = 2, 3, \ldots \end{cases}$$

A Discrete-Time Model for Contract Duration

- The relationship between $r(t)$ and $S(t)$ implies that, given knowledge of $r(t)$, we can compute $S(t)$ using the *forward recursion*:

$$S(t) = \begin{cases} 1 & \text{if } t = 0 \\ r(t) \times S(t-1) & \text{if } t = 1, 2, 3, \ldots \end{cases}$$

- We can compute the BG survivor function using the following forward-recursion formula from $S(0)$:

$$S(t \mid \gamma, \delta) = \begin{cases} 1 & \text{if } t = 0 \\ \dfrac{\delta + t - 1}{\gamma + \delta + t - 1} \times S(t-1) & \text{if } t = 1, 2, 3, \ldots \end{cases}$$

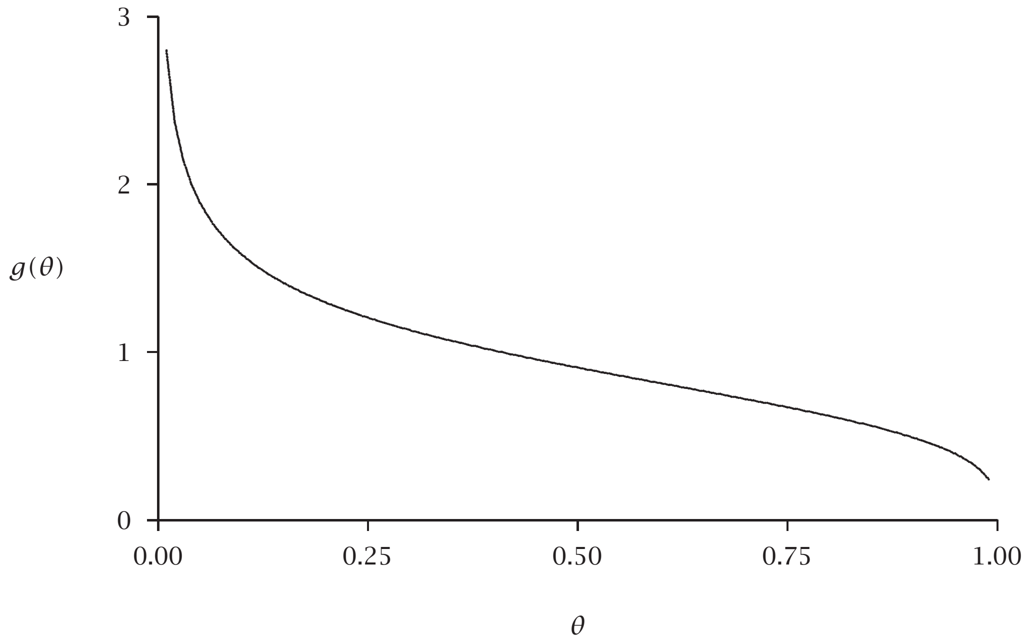
Estimating Model Parameters

	A	B	C	D	E	F
1	gamma	1.000				
2	delta	1.000				
3	LL	-1454.0				
4						
5	t	# Cust.	# Lost	P(T=t)	S(t)	
6	0	1000			1.0000	
7	1		9	0.5000	0.5000	-255.77
8	2	468	163	0.1667	0.3333	-292.06
9	3	382	86	0.0833	0.2500	-213.70
10					.2000	-167.76
11						-524.68

=B1/(B1+B2)

=D7*($B$2+A8-2)/($B$1+$B$2+A8-1)

Estimated Distribution of Churn Probabilities



$$\hat{\gamma} = 0.764, \hat{\delta} = 1.296, \widehat{E(\Theta)} = 0.371$$

Back to Motivating Problem 1

Year 1	Year 2	Year 3	Year 4	Year 5
1000	631	468	382	326

Q1a: Assuming our current prospect pool has the same characteristics of that from which these customers were acquired, what is the maximum amount you would be willing to spend to acquire a customer?

Q1b: What is the expected *residual* value of this group of 326 customers at the end of Year 5?

Motivating Problem 1 (Q1a)

	Year 1	Year 2	…	Year n
	$t=0$ to $t=1$	$t=1$ to $t=2$	…	$t=n-1$ to $t=n$
P(still a customer)	$S(0)$	$S(1)$		$S(n-1)$
Net CF	$v(0)$	$v(1)$		$v(n-1)$
discount	1	$\frac{1}{(1+d)}$		$\frac{1}{(1+d)^{n-1}}$

$$E(CLV) = \sum_{t=0}^{\infty} \frac{v(t)\,S(t)}{(1+d)^t}$$

$$= \bar{v} \sum_{t=0}^{\infty} \frac{S(t)}{(1+d)^t}$$

Motivating Problem 1 (Q1a)

	A	B	C	D	E	F	G
1	gamma	0.764			**E(CLV)**	$362	
2	delta	1.296					
3	d	0.1					
4	v bar	$100					
5							
6	Year	t	r(t)	S(t)		disc.	
7	1	0		1.0000		1.0000	
8	2	1	0.6292	0.6292		0.9091	
9	3	2	0.7504	0.4721		0.8264	
10	4	[illegible]	[illegible]	[illegible]		[illegible]	
11	5	4	0.8491	0.3255		0.6830	
12	6	5	0.8740	0.2845		0.6209	
13	7	6	0.8918	0.2537		0.5645	
14	8	7	0.9052	0.2296		0.5132	
15	9	8	0.9157	0.210		0.4665	
16	10	9	0.9241	0.194		0.4241	
17	11	10	0.9309	0.1809		0.3855	
105	99	98	0.9923	0.0342		0.0001	
106	100	99	0.9924	0.0339		0.0001	

=B4*SUMPRODUCT(D7:D106,F7:F106)

=($B$2+B8-1)/($B$1+$B$2+B8-1)

=1/(1+$B$3)^B8

=D13*C14

Motivating Problem 1 (Q1b)

Standing at the end of Year n, what is the expected residual lifetime value of a customer?

	Year n	Year $n+1$	Year $n+2$
	$t=n$	$t=n+1$	$t=n+2$
P(still a customer)		$S(n \mid T > n-1)$	$S(n+1 \mid T > n-1)$
Net CF		$v(n)$	$v(n+1)$
discount		1	$\dfrac{1}{(1+d)}$

$$E(RLV \mid n-1 \text{ renewals}) = \sum_{t=n}^{\infty} \frac{v(t)\,S(t \mid T > n-1)}{(1+d)^{t-n}}$$

$$= \bar{v} \sum_{t=n}^{\infty} \frac{S(t)/S(n-1)}{(1+d)^{t-n}}$$

Technical Aside

$S(t \mid T > n-1)$ is the probability the customer survives beyond t given they have survived beyond $n-1$, $t = n$, $n+1, n+2, \ldots$.

$$\begin{aligned} S(t \mid T > n-1) &= r(n) \times r(n+1) \times \cdots \times r(t) \\ &= \prod_{i=n}^{i=t} r(i) \\ &= \prod_{i=1}^{i=t} r(i) \Big/ \prod_{i=1}^{i=n-1} r(i) \\ &= S(t)/S(n-1) \end{aligned}$$

(We can also derive this result using Bayes' theorem.)

Motivating Problem 1 (Q1b)

	A	B	C	D	E	F	G
1	gamma	0.764			**E(RLV)**	$568	
2	delta	1.296					
3	d	0.1					
4	v bar	$100					
5					Given 4 renewals		
6	Year	t	r(t)		S(t|t>4)	disc.	
7	1	0					
8	2	1	0.6292				
9	3	2	0.7504				
10	4	3	0.8119				
11	5	4	0.8491				
12	6	5	0.8740		0.8740	1.0000	
13	7	6	0.8918		0.7794	0.9091	
14	8	7	0.9052		0.7056	0.8264	
15	9	8	0.9157		0.6461	0.7513	
16	10	9	0.9241		[illegible]	[illegible]	
17	11	10	0.9309		0.5558	0.6209	
105	99	98	0.9923		0.1050	0.0001	
106	100	99	0.9924		0.1042	0.0001	

=B4*SUMPRODUCT(E12:E106,F12:F106)

=C12

=E12*C13

=1/(1+$B$3)^(B13-5)

$\Longrightarrow$ expected residual value of the group of customers at the end of Year 5 is $326 \times \$568 = \$185{,}168$.

Comparing Approaches

	Model	Naïve	Underestimation
$E(CLV)$	\$362	\$324	10%
$E(RLV)$	\$568	\$381	33%

- 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

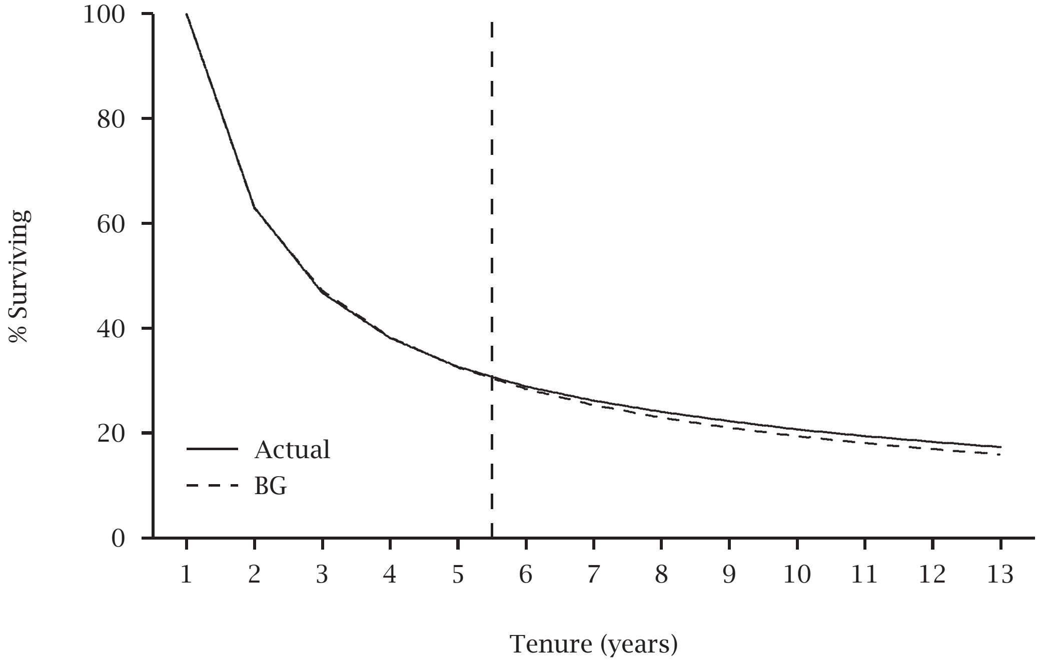
We actually have 12 years of renewal data.

- 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.

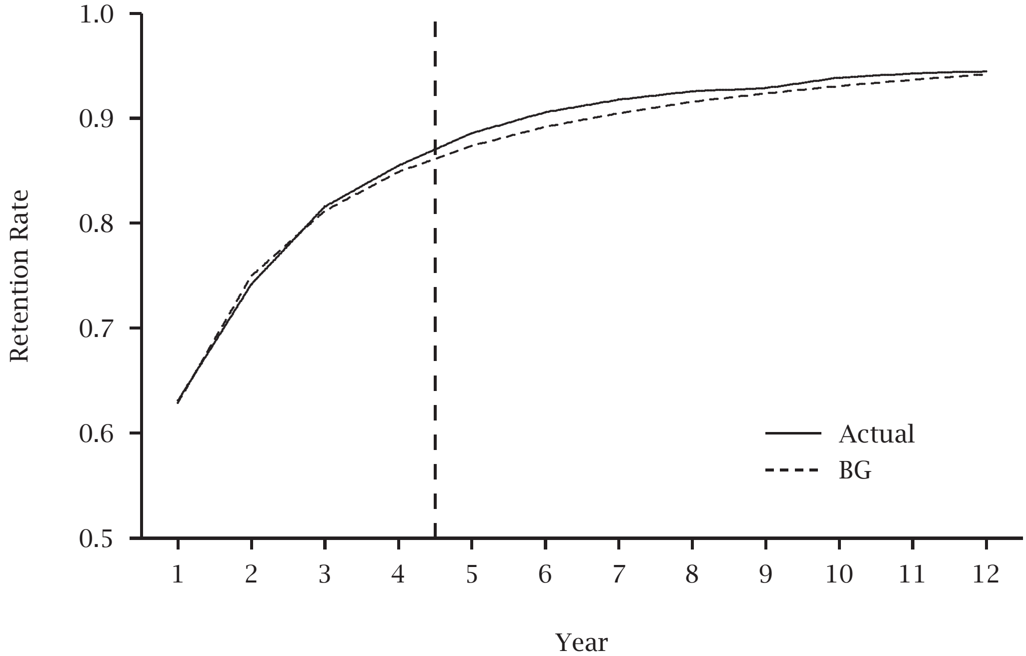
$$E(CLV) = \$100 \times \sum_{t=0}^{12} \frac{S(t)}{(1.1)^t}$$

$$R(CLV) = \$100 \times \sum_{t=5}^{12} \frac{S(t|T>4)}{(1.1)^{t-5}}$$

Survival Curve Projection



Projecting Retention Rates



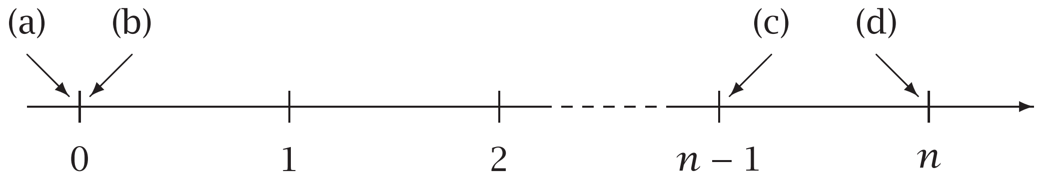
Comparing Approaches

	Infinite Horizon		12 Year Horizon		
	Model	Naïve	Actual	Model	Naïve
$E(CLV)$	\$362	\$324	\$331	\$327	\$314
$E(RLV)$	\$568	\$381	\$413	\$395	\$331

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

Expressions for $E(CLV)$ / $E(RLV)$

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



(a) An "as-yet-to-be-acquired" customer

(b) A "just-acquired" customer

(c) At the beginning of their nth period as a customer

(d) At the end of their nth period as a customer

Lessons Learnt

- How to compute CLV in contractual settings
  - Appreciating the need to project customer survival beyond the observed data.
  - Appreciating the distinction between the value of a new versus existing customer.
- How to use a probability model to forecast customer survival.
- Understanding the phenomenon of retention rate dynamics.

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 (2014), "A Spreadsheet-Literate Non-Statistician's Guide to the Beta-Geometric Model." <`http://brucehardie.com/notes/032/`>

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 (2014), "What's Wrong With This CLV Formula?" <`http://brucehardie.com/notes/033/`>

Fader, Peter S. and Bruce G. S. Hardie (2012), "Reconciling and Clarifying CLV Formulas." <`http://brucehardie.com/notes/024/`>

Noncontractual Settings

Motivating Problem 2

ID	1995	1996	1997	1998	1999	2000	2001
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
	11104	5652	4674	4019	3552	3555	3163

Motivating Problem 2

Q2a: Assuming our current prospect pool has the same characteristics of that from which these donors were acquired, what is the maximum amount you would spend to acquire a new donor?

Q2b: Given their donation behavior to date, in how many of the subsequent five years can we expect a supporter to make a donation?

What about 100004 (who made repeat donations in four years with the last occurring in 2001) versus 100009 (who made repeat donations in five years with the last occurring in 2000)?

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	?	?	?	?	?

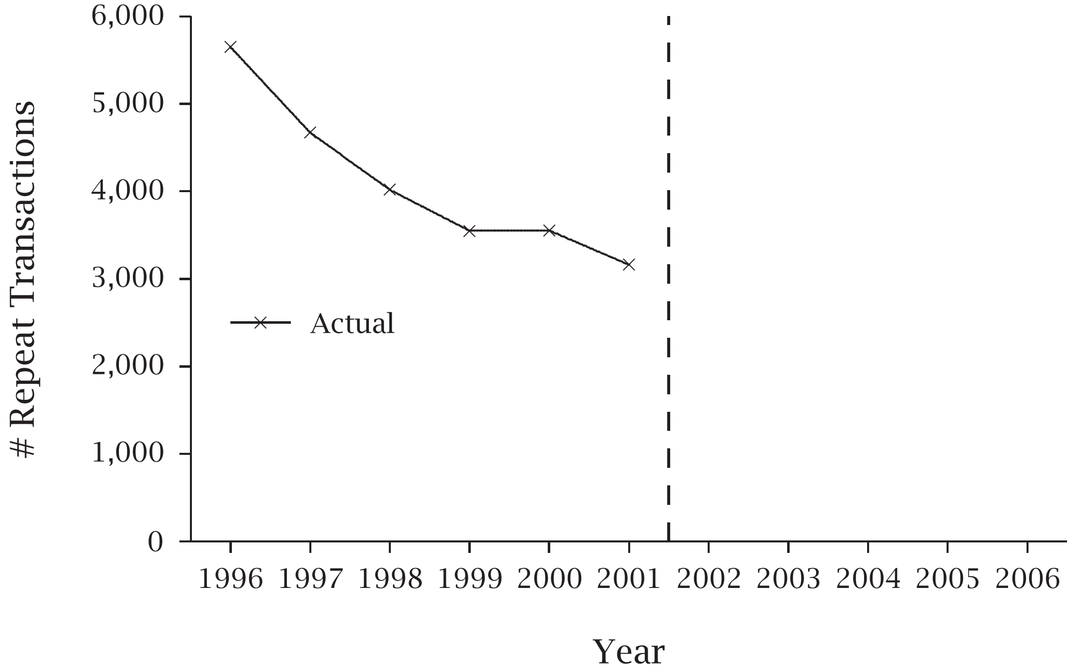
Notation

- A customer's transaction history can be expressed as a binary string, where $y(t) = 1$ if a transaction occurred at or during the tth transaction opportunity, and 0 otherwise.
- Let the random variable

$$X(n) = \sum_{t=1}^{n} Y(t)$$

denote the number of transactions occurring across the n transaction opportunities in the interval $\{1, 2, \ldots, n\}$.

Annual Repeat Transactions

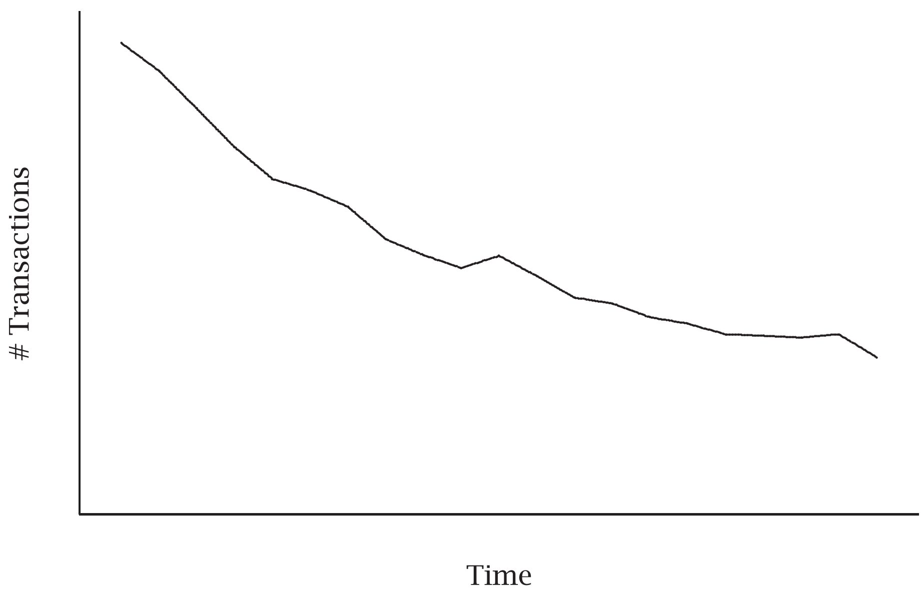


Repeat-buying in a Noncontractual Setting

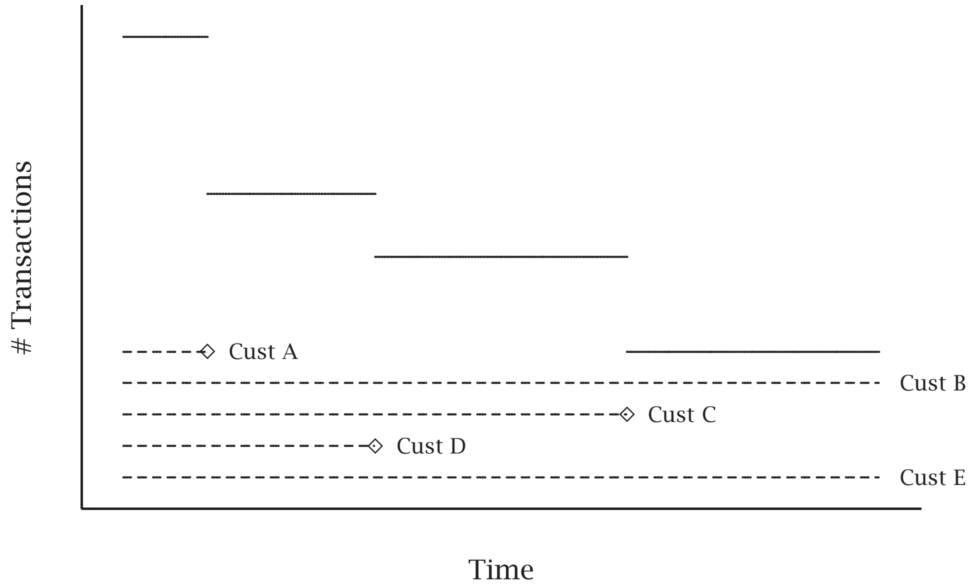
The "leaky bucket" phenomenon:

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.

Repeat-buying in a Noncontractual Setting



Repeat-buying in a Noncontractual Setting



"latent attrition"

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 transaction string conditional on each scenario and multiply it by the probability of that scenario:

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

Model Development

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

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

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

  $\Longrightarrow$ 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 transaction 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.

We call this the BG/BB (beta-geometric/beta-Bernoulli) model.

	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]0009	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)

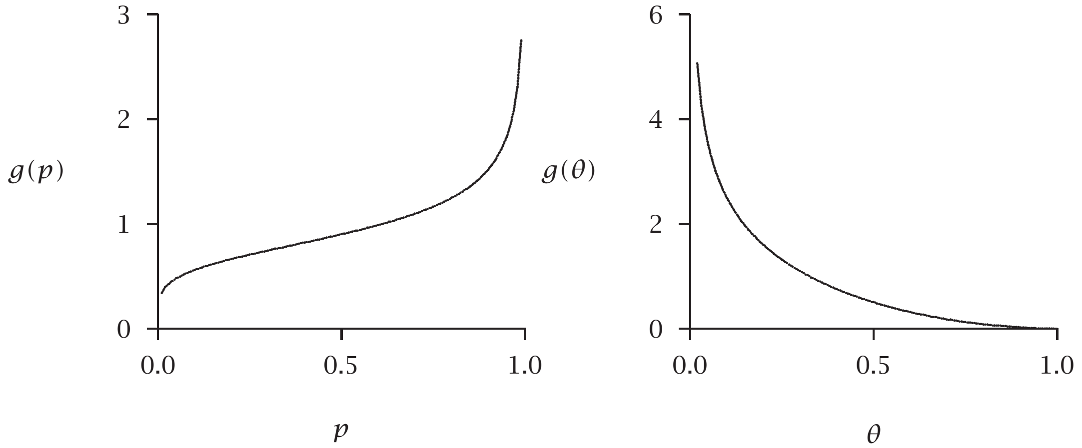
=IF(I$8<=$G9,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+I$8)-GAMMALN($B$1+$B$2+$B9+I$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+I$8)-GAMMALN($B$3+$B$4+$B9+I$8+1))/$E$3,0)

=C15-B15-1

=D19*LN(F19))

=SUM(H19:N19)

Estimated Beta Distributions



$\alpha = 1.204,\ \beta = 0.750$
$E(P) = 0.616$

$\gamma = 0.657,\ \delta = 2.783$
$E(\Theta) = 0.191$

Key Results

We are interested in the probability that a customer makes a transaction at the tth transaction opportunity.

- Recall our model assumptions:
  - $P(Y(t) = 1 \mid p, \text{alive at } t) = p$
  - $P(\text{alive at } t \mid \theta) = S(t) = (1-\theta)^t$
- Therefore, $P(Y(t) = 1 \mid p, \theta) = p(1-\theta)^t$.
- But p and θ are unobserved ...

Key Results

Removing the conditioning on p and θ:

$$
\begin{aligned}
P(Y(t) = 1 \mid \alpha, \beta, \gamma, \delta)& \\
&= \int_0^1 \int_0^1 P(Y(t) = 1 \mid p, \theta) g(p \mid \alpha, \beta) g(\theta \mid \gamma, \delta)\, dp\, d\theta \\
&= \left(\frac{\alpha}{\alpha + \beta}\right) \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)},
\end{aligned}
$$

which can computed using the recursion

$$
P(Y(t) = 1 \mid \alpha, \beta, \gamma, \delta) = \begin{cases} \left(\dfrac{\alpha}{\alpha + \beta}\right)\left(\dfrac{\delta}{\gamma + \delta}\right) & t = 1 \\ \dfrac{\delta + t - 1}{\gamma + \delta + t - 1} \\ \qquad \times P(Y(t-1) = 1) & t = 2, 3, \ldots \end{cases}
$$

Key Results

Expected # transactions in $\{1, 2, \ldots, n\}$:

$$
E[X(n) \mid \alpha, \beta, \gamma, \delta] = \left(\frac{\alpha}{\alpha + \beta}\right)\left(\frac{\delta}{\gamma - 1}\right) \times \left\{1 - \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma + \delta + n)} \frac{\Gamma(1 + \delta + n)}{\Gamma(1 + \delta)}\right\}.
$$

Alternatively,

$$
E[X(n) \mid \alpha, \beta, \gamma, \delta] = \sum_{t=1}^{n} P(Y(t) = 1 \mid \alpha, \beta, \gamma, \delta).
$$

Key Results

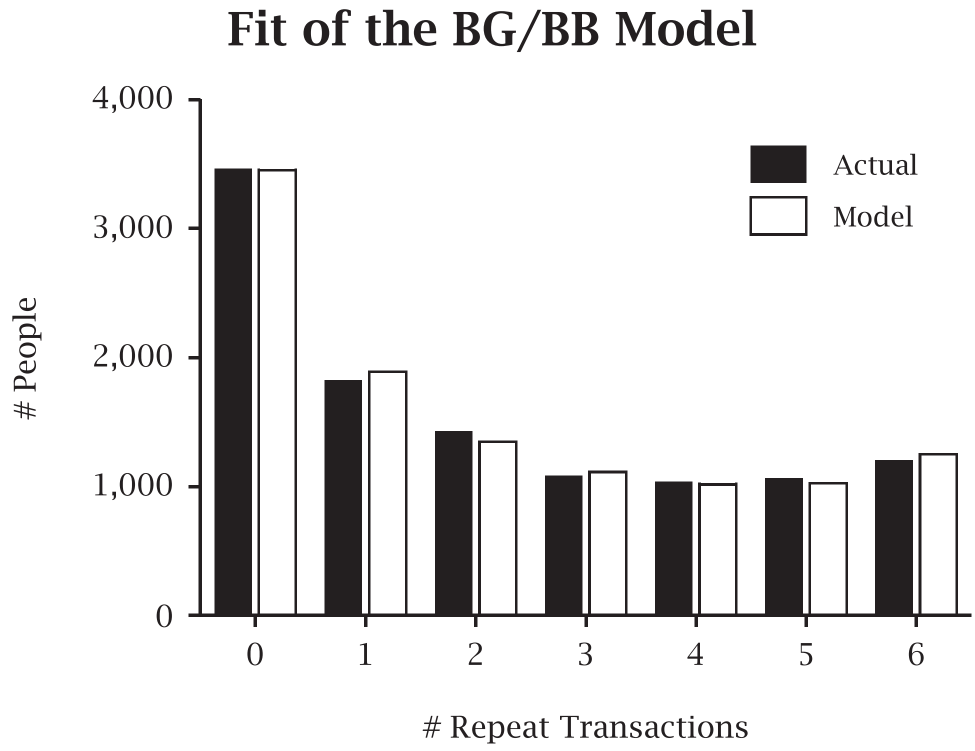
Distribution of transactions in the interval $\{1, 2, \ldots, n\}$:

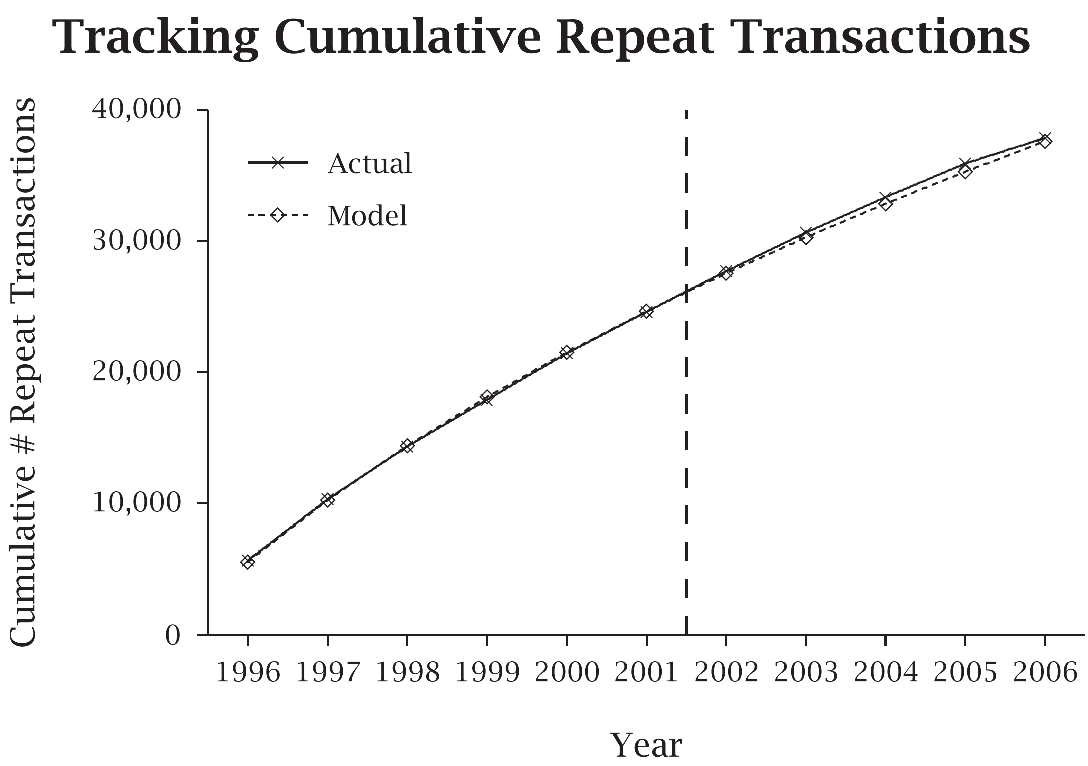
$$
\begin{aligned}
P(X(n) = x \mid \alpha, \beta, \gamma, \delta)
&= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta + n)}{B(\gamma, \delta)} \\
&\quad + \sum_{i=x}^{n-1} \binom{i}{x} \frac{B(\alpha + x, \beta + i - x)}{B(\alpha, \beta)} \frac{B(\gamma + 1, \delta + i)}{B(\gamma, \delta)} .
\end{aligned}
$$

Key Results

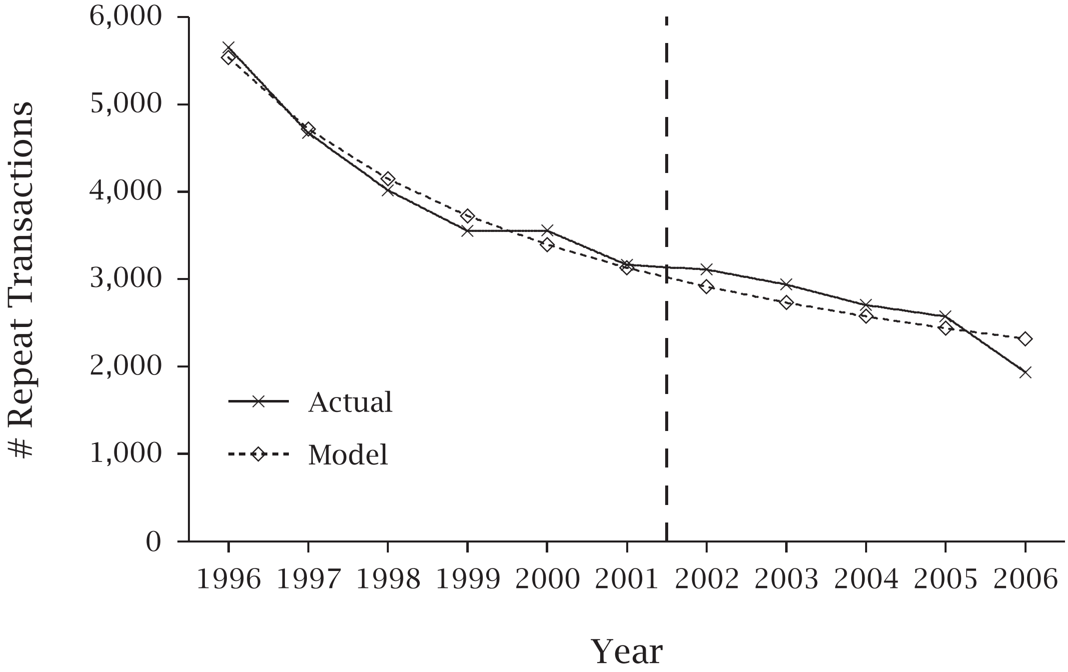
Distribution of transactions in $\{n + 1, \ldots, n + n^*\}$:

$$
\begin{aligned}
P(X(n, n + n^*) &= x^* \mid \alpha, \beta, \gamma, \delta) \\
&= \delta_{x^*=0} \left\{ 1 - \frac{B(\gamma, \delta + n)}{B(\gamma, \delta)} \right\} \\
&\quad + \binom{n^*}{x^*} \frac{B(\alpha + x^*, \beta + n^* - x^*)}{B(\alpha, \beta)} \frac{B(\gamma, \delta + n + n^*)}{B(\gamma, \delta)} \\
&\quad + \sum_{i=x^*}^{n^*-1} \binom{i}{x^*} \frac{B(\alpha + x^*, \beta + i - x^*)}{B(\alpha, \beta)} \frac{B(\gamma + 1, \delta + n + i)}{B(\gamma, \delta)} .
\end{aligned}
$$

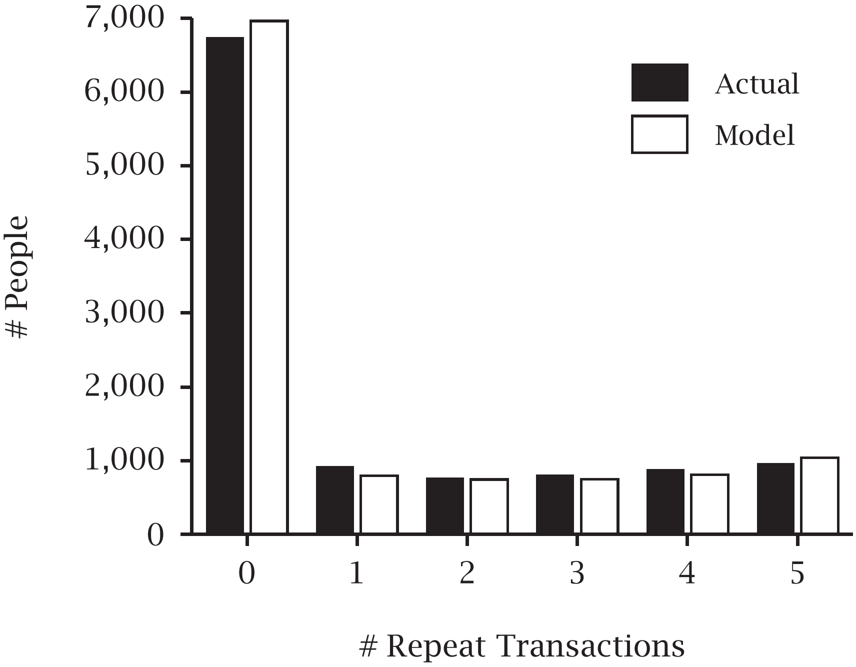

Fit of the BG/BB Model
4,000
3,000
2,000
1,000
0
People
Actual
Model
0
1
2
3
4
5
6
Repeat Transactions


Tracking Cumulative Repeat Transactions
40,000
30,000
20,000
10,000
0
Cumulative # Repeat Transactions
Actual
Model
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
Year

Tracking Annual Repeat Transactions



Repeat Transactions in 2002–2006



Motivating Problem 2 (Q2a)

	Year 1	Year 2		Year n
$P(\text{donate})$	1.0	$P(Y(1) = 1)$		$P(Y(n-1) = 1)$
Net CF	$v(0)$	$v(1)$		$v(n-1)$
discount	1	$\dfrac{1}{(1+d)}$		$\dfrac{1}{(1+d)^{n-1}}$

$$
\begin{aligned}
E(CLV) &= v(0) + \sum_{t=1}^{\infty} \frac{v(t)\, P(Y(t) = 1)}{(1+d)^t} \\
&= \bar{v} \left\{ 1 + \sum_{t=1}^{\infty} \frac{P(Y(t) = 1)}{(1+d)^t} \right\}
\end{aligned}
$$

Motivating Problem 2 (Q2a)

	A	B	C	D	E	F	G
1	alpha	1.204		**E(CLV)**	$185		
2	beta	0.750					
3	gamma	0.657					
4	delta	2.783					
5	d	0.100					
6	v bar	$50					
7							
8	Year	t	P(Y(t)=1)	disc.			
9	2	1	0.4985	0.9091			
10	3	2	0.4248	0.8264			
11	4	3	0.3735	0.7513			
12							
13	6	5	0.3058	0.6209			
14	7	6	0.2820	0.5645			
15	8	7	0.2624	0.5132			
16	9	8	0.2459	0.4665			
17	10	9	0.2318	0.4241			
106	99	98	0.0561	0.0001			
107	100	99	0.0558	0.0001			

=B6*(1+SUMPRODUCT(C9:C107,D9:D107))

=B1/(B1+B2)*B4/(B3+B4)

=C9*($B$4+B10-1)/($B$3+$B$4+B10-1)

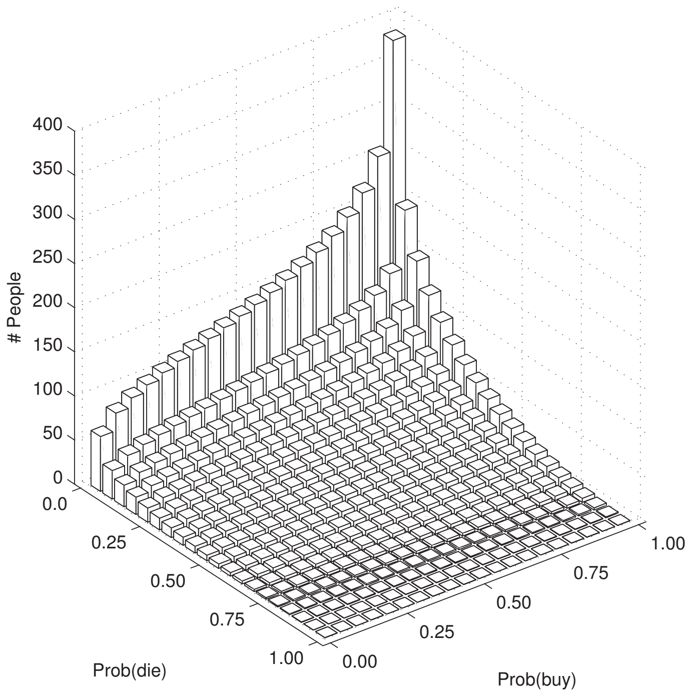
Motivating Problem 2 (Q2b)

Given their donation behavior to date, in how many of the subsequent five years can we expect a supporter to make a donation?

> What about 100004 (who made repeat donations in four years with the last occurring in 2001) versus 100009 (who made repeat donations in five years with the last occurring in 2000)?

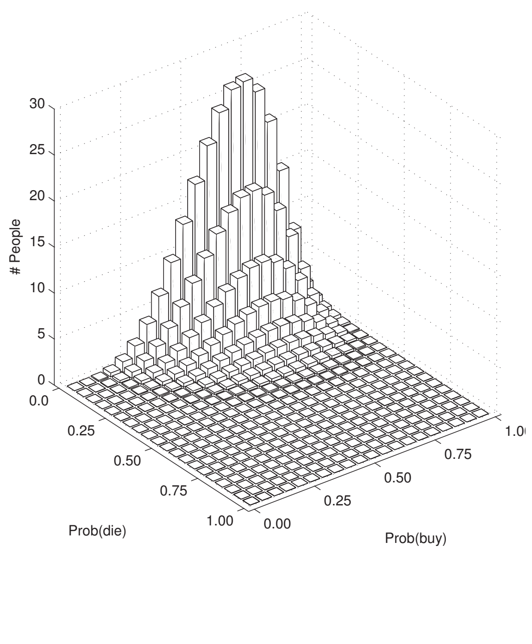
ID	1995	1996	1997	1998	1999	2000	2001	2002 - 2006
100001	1	0	0	0	0	0	0	?
100002	1	0	0	0	0	0	0	?
100003	1	0	0	0	0	0	0	?
100004	1	0	1	0	1	1	1	?
100005	1	0	1	1	1	0	1	?
100006	1	1	1	1	0	1	0	?
100007	1	1	0	1	0	1	0	?
100008	1	1	1	1	1	1	1	?
100009	1	1	1	1	1	1	0	?
100010	1	0	0	0	0	0	0	?
⋮			⋮			⋮		
111102	1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

Joint Distribution of “Buy” and “Die” Coins

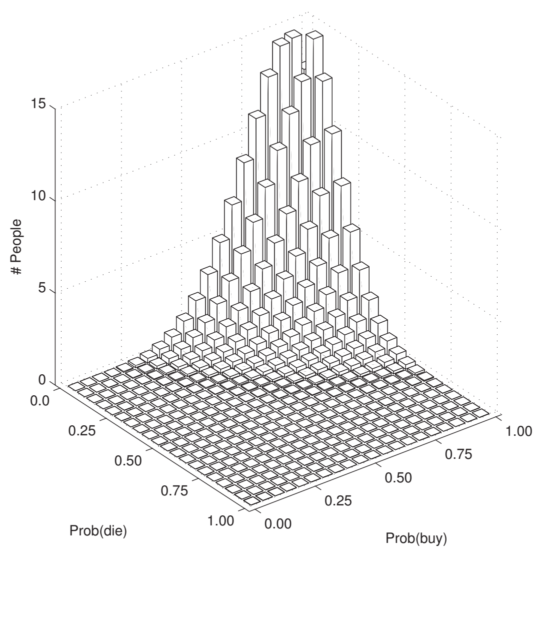


Joint Distribution of “Buy” and “Die” Coins

$x = 4, t_x = 6, n = 6$
512 people



$x = 5, t_x = 5, n = 6$
335 people



100004 vs. 100009

	100004	100009
$E(P)$	0.65	0.83
$E(\Theta)$	0.07	0.12
Expected # donations in 2002 – 2006 | alive in 2001:	2.71	3.23
P(alive in 2001):	1.00	0.56
Expected # donations in 2002 – 2006:	2.71	1.81

Conditional Expectations

- We wish to derive an expression for the expected number of transactions across the next n^* transaction opportunities for an individual with observed behavior (x, t_x, n).
- Suppose we know that the individual is alive at n, and we know their p and θ:

	$n+1$	$n+2$	$\cdots$	$n+n^*$
P(buy)	p	p		p
P(alive)	$(1-\theta)$	$(1-\theta)^2$		$(1-\theta)^{n^*}$

Conditional Expectations

- Therefore,

$$E[X(n, n+n^*) \mid p, \theta, \text{alive at } n] = \sum_{s=1}^{n^*} p(1-\theta)^s = \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n^*+1}}{\theta}.$$

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

Conditional Expectations

What is the probability that an individual with observed behavior (x, t_x, n) is alive at n?

- Recall that for a customer with transaction history (x, t_x, n)

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

- This was formulated by assuming the customer is alive or dead, and then removing the conditioning.

Conditional Expectations

- According to Bayes' theorem

$$P(\text{alive at } n \mid x, t_x, n) = \frac{\mathsf{A}}{\mathsf{A} + \mathsf{B}}$$

where

$$\mathsf{A} = P(x, t_x, n \mid \text{alive at } n)\, P(\text{alive at } n)$$

$$\mathsf{B} = P(x, t_x, n \mid \text{dead at } n)\, P(\text{dead at } n)$$

- It follows that

$$P(\text{alive at } n \mid p, \theta; x, t_x, n) = \frac{p^x (1-p)^{n-x} (1-\theta)^n}{L(p, \theta \mid x, t_x, n)}\,.$$

Conditional Expectations

We can now remove the conditioning on the customer being alive at n, but p and θ are still unobserved.

- It makes no sense to use $g(p \mid \alpha, \beta)$ and $g(\theta \mid \gamma, \delta)$ as we know something about the customer's behavior, (x, t_x, n).
- By Bayes' theorem, the joint posterior distribution of P and Θ is given by

$$g(p, \theta \mid \alpha, \beta, \gamma, \delta; x, t_x, n) = \frac{L(p, \theta \mid x, t_x, n) g(p \mid \alpha, \beta) g(\theta \mid \gamma, \delta)}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}\,.$$

Conditional Expectations

Combining the various elements give us

$$
\begin{aligned}
E[X(n, n+n^*) \mid \alpha, \beta, \gamma, \delta; x, t_x, n]
&= \int_0^1 \int_0^1 \Big\{ E[X(n, n+n^*) \mid p, \theta, \text{alive at } n] \\
&\qquad \times P(\text{alive at } n \mid p, \theta; x, t_x, n) \\
&\qquad \times g(p, \theta \mid \alpha, \beta, \gamma, \delta; x, t_x, n) \Big\} dp\, d\theta \\
&= \frac{1}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)} \frac{B(\alpha + x + 1, \beta + n - x)}{B(\alpha, \beta)} \\
&\qquad \times \left(\frac{\delta}{\gamma - 1}\right) \frac{\Gamma(\gamma + \delta)}{\Gamma(1 + \delta)} \left\{ \frac{\Gamma(1 + \delta + n)}{\Gamma(\gamma + \delta + n)} - \frac{\Gamma(1 + \delta + n + n^*)}{\Gamma(\gamma + \delta + n + n^*)} \right\}.
\end{aligned}
$$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	alpha	1.204		B(alpha,beta)	1.146		n*	5								
2	beta	0.750														
3	gamma	0.657		B(gamma,delta)	0.729		"constant"	1.896								
4	delta	2.783														
5																
6	LL	-33225.6														
7																
8	p1x	tx	n	# donors		L(.|x,t_x,n)	CE		n -t_x - 1		0	1	2	3	4	5
9	6	6	6	1203	-2624.5	0.1129	3.753	0.2233	-1	0.1129	0	0	0	0	0	0
10	5	6	6	728	-3126.7	0.0136	3.232	0.0232	-1	0.0136	0	0	0	0	0	0
11	4	6	6	512	-2757	[illegible]	2.711	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0
12	3	6	6	357	-2073.9	0.0030	2.190	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0
13	2	6	6	234	-1322.5	0.0035	1.669	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0
14	1	6	6	129	-630.0	0.0076	1.148	[illegible]	[illegible]	[illegible]	0	0	0	0	0	0
15	5	5	6	335	-1245.1	0.0243	1.813	0.0232	0	0.0136	0.0107	0	0	0	0	0
16	4	5	6	284	-1447.1	0.0061	2.030	0.0066	0	0.0046	0.0015	0	0	0	0	0
17	3	5	6	225	-1263.5	0.0036	1.805	0.0035	0	0.0030	0.0006	0	0	0	0	0
18	2	5	6	173	-952.6	0.0041	1.443	0.0031	0	0.0035	0.0005	0	0	0	0	0
19	1	5	6	119	-567.3	0.0085	1.022	0.0046	0	0.0076	0.0009	0	0	0	0	0
20	4	4	6	240	-923.6	0.0213	0.583	0.0066	1	0.0046	0.0152	0.0015	0	0	0	0
21	3	4	6	181	-915.7	0.0063	1.035	0.0035	1	0.0030	0.0027	0.0006	0	0	0	0
22	2	4	6	155	-805.3	0.0055	1.058	0.0031	1	0.0035	0.0015	0.0005	0	0	0	0
23	1	4	6	78	-356.5	0.0104	0.839	0.0046	1	0.0076	0.0018	0.0009	0	0	0	0
24	3	3	6	322	-1135.8	0.0294	0.224	0.0035	2	0.0030	0.0230	0.0027	0.0006	0	0	0
25	2	3	6	255	-1151.6	0.0109	0.536	0.0031	2	0.0035	0.0054	0.0015	0.0005	0	0	0
26	1	3	6	129	-545.0	0.0146	0.594	0.0046	2	0.0076	0.0043	0.0018	0.0009	0	0	0
27	2	2	6	613	-1846.4	0.0492	0.119	0.0031	3	0.0035	0.0383	0.0054	0.0015	0.0005	0	0
28	1	2	6	277	-993.9	0.0276	0.314	0.0046	3	0.0076	0.0130	0.0043	0.0018	0.0009	0	0
29	1	1	6	1091	-2497.1	0.1014	0.086	0.0046	4	0.0076	0.0737	0.0130	0.0043	0.0018	0.0009	0
30	0	0	6	3464	-4044.3	0.3111	0.073	0.0120	5	0.0362	0.1909	0.0459	0.0189	0.0098	0.0058	0.0037

=B4/(B3-1)*EXP(GAMMALN(B3+B4)-GAMMALN(B4+1))*(EXP(GAMMALN(1+B4+C9)-GAMMALN(B3+B4+C9))-EXP(GAMMALN(1+B4+C9+H1)-GAMMALN(B3+B4+C9+H1)))

=H9*$H$3/F9

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

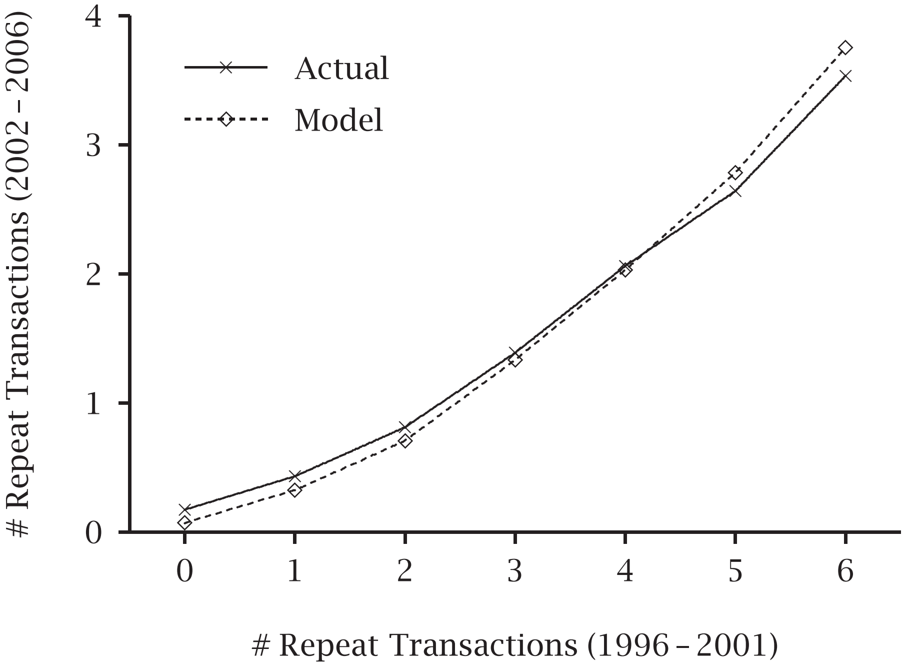
Expected # Transactions in 2002 – 2006 as a Function of Recency and Frequency

# 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

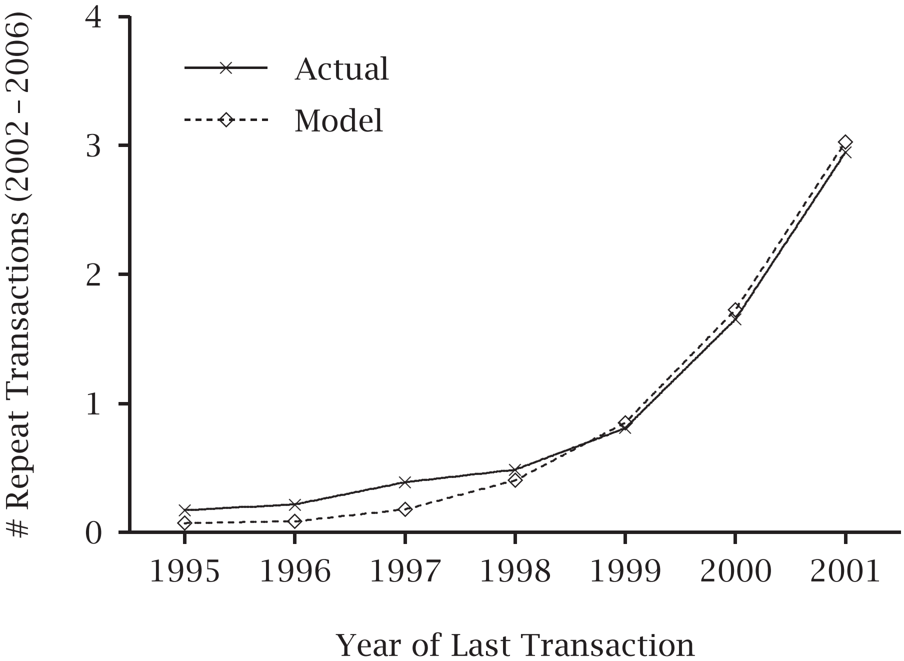
# Rpt Trans.	Year of Last Transaction						
(1996 – 2001)	1995	1996	1997	1998	1999	2000	2001
0	0.17						
1		0.22	0.37	0.60	0.56	1.14	1.47
2			0.40	0.46	0.74	1.41	1.89
3				0.46	0.94	1.66	2.29
4					0.84	1.91	2.72
5						1.74	3.06
6							3.53

Conditional Expectations by Frequency



111

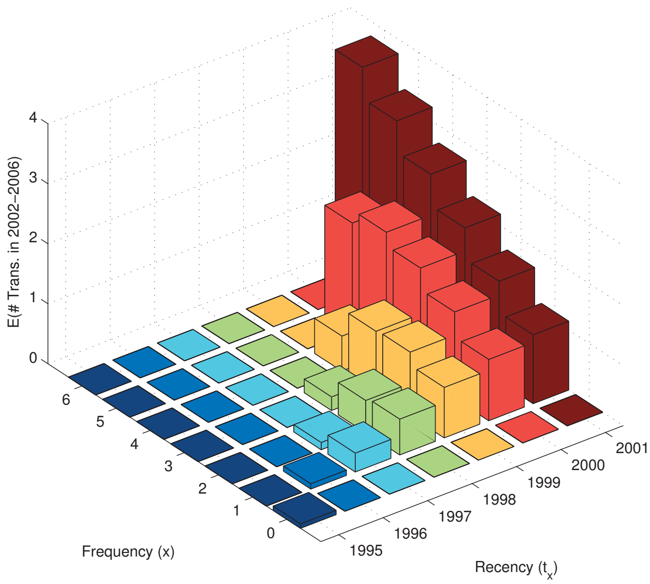
Conditional Expectations by Recency



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

Computing $E(CLV)$

- Recall:

$$E(CLV) = \sum_{t=1}^{\infty} \frac{v(t)\, S(t)}{(1+d)^t}\,.$$

- Assuming that an individual's spend per transaction is constant, $v(t) = \text{net cashflow / transaction} \times y(t)$.
- The expected lifetime value of a "just-acquired" customer can be expressed as

$$E(CLV) = E(\text{net cashflow / transaction}) \times \underbrace{\sum_{t=1}^{\infty} \frac{E[Y(t) \mid \text{alive at } t]\, S(t)}{(1+d)^t}}_{\text{discounted expected transactions}}\,.$$

Computing DET

- The quantity *DET*, discounted expected transactions, is the present value of the expected transaction stream for a customer "just acquired" in Period "0".
- Suppose we know their p and θ:

	1	2	$\cdots$	t
$P(\text{buy})$	p	p		p
$P(\text{alive})$	$(1-\theta)$	$(1-\theta)^2$		$(1-\theta)^t$
discount	$\frac{1}{(1+d)}$	$\frac{1}{(1+d)^2}$		$\frac{1}{(1+d)^t}$

Computing DET

- Therefore, $$DET(p,\theta,d) = \sum_{s=1}^{\infty} p\left(\frac{1-\theta}{1+d}\right)^s = \frac{p(1-\theta)}{d+\theta}.$$
- Taking expectations over the distributions of p and θ,

$$\begin{aligned} &DET(\alpha,\beta,\gamma,\delta,d) \\ &\quad = \int_0^1\int_0^1 DET(p,\theta,d)g(p\mid\alpha,\beta)g(\theta\mid\gamma,\delta)dp\,d\theta \\ &\quad = \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\gamma}{\gamma+\delta}\right)\frac{{}_2F_1(1,\delta+1;\gamma+\delta+1;\frac{1}{1+d})}{(1+d)}, \end{aligned}$$

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 $u_0 = 1$.

Computing $E(RLV)$

- Standing at time n,

$$E(RLV) = E(\text{net cashflow / transaction}) \times \underbrace{\sum_{t=n+1}^{\infty} \frac{E[Y(t) \mid \text{alive at } t]\, S(t \mid t > n)}{(1+d)^{t-n}}}_{\text{discounted expected residual transactions}}.$$

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

Computing DERT

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

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	alpha	1.204		B(alpha,beta)	1.146		d	0.100							
2	beta	0.750													
3	gamma	0.657		B(gamma,delta)	0.729										
4	delta	2.783													
5															
6	LL	-33225.6													
7															
8	p1x	tx	n	# donors		L(.|x,t_x,n)	DERT	n -t_x - 1		0	1	2	3	4	5
9	6	6	6	1203	-2624.5	0.1129	5.910	-1	0.1129	0	0	0	0	0	0
10	5	6	6	728	-3126.7	0.0136	5.089	-1	0.0136	0	0	0	0	0	0
11	4	6	6	512	-2757.0	0.0046	4.269						0	0	0
29	1	1	6	1091	-2497.1	0.1014	0.135						.0018	0.0009	0
30	0	0			44.3	0.3111	0.115						.0098	0.0058	0.0037
31															
32					j										
33			2F1	7.714	0	1									
34			a	1	1	0.8519									
35			b	9.78	2	0.7300									
36			c	10.44	3	0.6286									
37			z	0.91	4	0.5435									
38					5	0.4717									
182					149	1.079E-07									
183					150	9.768E-08									

=EXP(GAMMALN($B$1+A9+1)+GAMMALN($B$2+C9-A9)-GAMMALN($B$1+$B$2+C9+1))*EXP(GAMMALN($B$3)+GAMMALN($B$4+C9+1)-GAMMALN($B$3+$B$4+C9+1))/($E$1*$E$3*(1+$H$1))*$D$33/F9

=SUM(F33:F183)

=F33*($D$34+E34-1)*($D$35+E34-1)*$D$37/(($D$36+E34-1)*E34)

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 = $1.575 \times$ 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.

Concepts and Tools Introduced

- Computing CLV in noncontractual settings (where “death” is unobserved).
- The notion of latent attrition (“buy till you die”) models.
- The BG/BB model for discrete-time noncontractual settings.
- Recency and frequency as sufficient statistics.
- The notion of DET and DERT for noncontractual settings and their evaluation when the transaction stream is characterized by the BG/BB model.

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?

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.

From Discrete to Continuous Time

The BG/BB model integrates two processes: timing and counting.

Timing: The BG component captures the time until death.

Counting: The BB component captures the counting of transactions while alive.

What are the equivalent distributions in a continuous-time setting?

From Discrete to Continuous Time

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

binomial	$\longrightarrow$	Poisson
beta-binomial	$\longrightarrow$	NBD
geometric	$\longrightarrow$	exponential
beta-geometric	$\longrightarrow$	Pareto Type II
BG/BB	$\longrightarrow$	Pareto/NBD

MANAGEMENT SCIENCE
Vol. 33, No. 1, January 1987
Printed in U.S.A.

COUNTING YOUR CUSTOMERS: WHO ARE THEY AND WHAT WILL THEY DO NEXT?

DAVID C. SCHMITTLEIN, DONALD G. MORRISON AND RICHARD COLOMBO
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
Graduate School of Business, Columbia University, New York, New York 10027
Department of Marketing, Tisch Hall, New York University, New York, New York 10002

This article is concerned with counting and identifying those customers who are still active. The issue is important in at least three settings: monitoring the size and growth rate of a firm's ongoing customer base, evaluating a new product's success based on the pattern of trial and repeat purchases, and targeting a subgroup of customers for advertising and promotions. We develop a model based on the number and timing of the customers' previous transactions. This approach allows computation of the probability that any particular customer is still active. Several numerical examples are used to illustrate applications of the model.
(MARKETING; CONSUMER BEHAVIOR; POISSON PROCESS; PROBABILITY MIXTURE MODELS; NEW PRODUCT INTRODUCTIONS; MARKET SEGMENTATION; BROKERAGE FIRMS)

PETER S. FADER, BRUCE G.S. HARDIE, and KA LOK LEE*

The authors present a new model that links the well-known RFM (recency, frequency, and monetary value) paradigm with customer lifetime value (CLV). Although previous researchers have made a conceptual link, none has presented a formal model with a well-grounded behavioral "story." Key to this analysis is the notion of "iso-value" curves, which enable the grouping of individual customers who have different purchasing histories but similar future valuations. Iso-value curves make it easy to visualize the interactions and trade-offs among the RFM measures and CLV. The stochastic model is based on the Pareto/NBD framework to capture the flow of transactions over time and a gamma-gamma submodel for spend per transaction. The authors conduct several holdout tests to demonstrate the validity of the model's underlying components and then use it to estimate the total CLV for a cohort of new customers of the online music site CDNOW. Finally, the authors discuss broader issues and opportunities in the application of this model in actual practice.

RFM and CLV: Using Iso-Value Curves for Customer Base Analysis

Journal of Marketing Research
Vol. XLII (November 2005), 415–430

Further Reading

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 (2005), "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>

Fader, Peter S. and Bruce G. S. Hardie (2013), "The Gamma-Gamma Model of Monetary Value."
<http://brucehardie.com/notes/025/>

Pareto/NBD Likelihood Function

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

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

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

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

MARKETING SCIENCE
Vol. 24, No. 2, Spring 2005, pp. 275–284
ISSN 0732-2399 | EISSN 1526-548X | 05 | 2402 | 0275

DOI 10.1287/mksc.1040.0098

"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model

Peter S. Fader
The Wharton School, University of Pennsylvania, 749 Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, faderp@wharton.upenn.edu

Bruce G. S. Hardie
London Business School, Regent's Park, London NW1 4SA, United Kingdom, bhardie@london.edu

Ka Lok Lee
Catalina Health Resource, Blue Bell, Pennsylvania 19422, kaloklee@alumni.upenn.edu

Today's managers are very interested in predicting the future purchasing patterns of their customers, which can then serve as an input into "lifetime value" calculations. Among the models that provide such capabilities, the Pareto/NBD "counting your customers" framework proposed by Schmittlein et al. (1987) is highly regarded. However, despite the respect it has earned, it has proven to be a difficult model to implement, particularly because of computational challenges associated with parameter estimation.

We develop a new model, the beta-geometric/NBD (BG/NBD), which represents a slight variation in the behavioral "story" associated with the Pareto/NBD but is vastly easier to implement. We show, for instance, how its parameters can be obtained quite easily in Microsoft Excel. The two models yield very similar results in a wide variety of purchasing environments, leading us to suggest that the BG/NBD could be viewed as an attractive alternative to the Pareto/NBD in most applications.

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "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), "Overcoming the BG/NBD Model's #NUM! Error Problem." <http://brucehardie.com/notes/027/>

Approaches to Customer-Base Analysis

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
  - CLV/RLV
- Forward-looking/predictive versus descriptive.

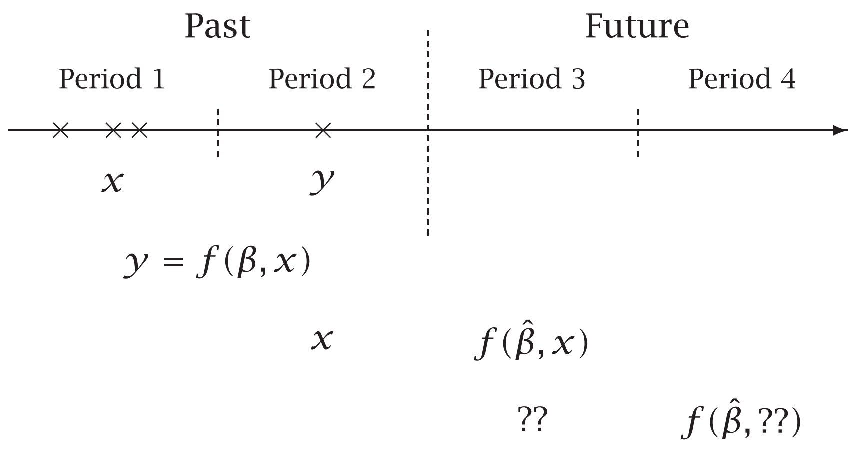
Traditional Modelling Approach

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).

Traditional Modelling Approach

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



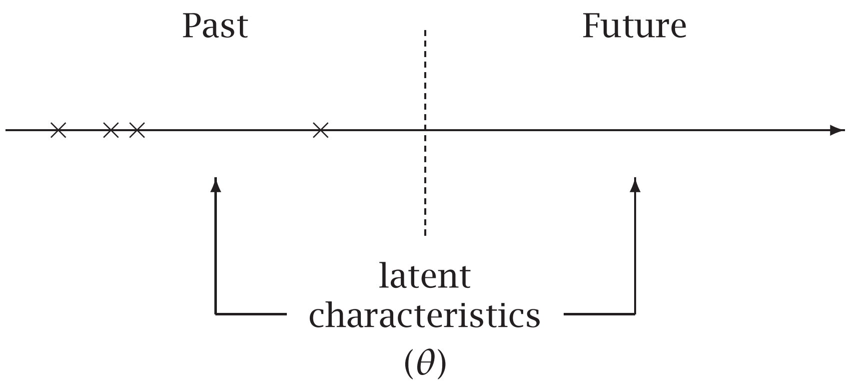
Traditional Modelling Approach

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)



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

Further Considerations

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

Covariates

Contractual:

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.

Noncontractual:

Fader, Peter S. and Bruce G. S. Hardie (2007), “Incorporating Time-Invariant Covariates into the Pareto/NBD and BG/NBD Models.” <`http://brucehardie.com/notes/019/`>

Schweidel, David A. and George Knox (2013), “Incorporating Direct Marketing Activity into Latent Attrition Models,” *Marketing Science*, **32** (May–June), 471–487.

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/`>

  Fader, Peter S. and Bruce G. S. Hardie (2015), "A Correlated Pareto/NBD Model." <`http://brucehardie.com/notes/034/`>

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

Discussion

Tell Us Your Stories

Pete: `faderp@wharton.upenn.edu`
`www.petefader.com`
`@faderp`

Bruce: `bhardie@london.edu`
`www.brucehardie.com`