Probability Models for Customer Lifetime Value Analysis Day 2

Peter S. Fader
University of Pennsylvania

Bruce G.S. Hardie
London Business School

Stockholm School of Economics
June 3–4, 2024

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.

The logic 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. Individuals vary, but percentages remain constant.”
>
> Sir Arthur Conan Doyle, *The Sign of the Four*, 1890.

Building a probability model

i) Determine the marketing decision problem/information needed.

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

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

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

Building a probability model

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

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

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.

A probability model is an "as-if" story about the data-generating process.

Fundamental building blocks

There are three fundamental behavioral processes:

- Counting $\longrightarrow$ "how many"
- Timing $\longrightarrow$ "when / how long"
- "Choice" $\longrightarrow$ "whether / which"

There are four basic models:

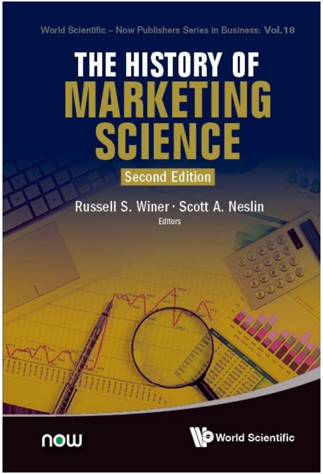
Phenomenon	Individual-level	Heterogeneity	Model
Counting	Poisson	gamma	NBD
Timing (continuous)	exponential	gamma	Pareto (II)
Timing (discrete) (or counting)	geometric	beta	BG
Choice	binomial	beta	BB

“Integrated” models

More complex phenomena can be captured by combining models from each of three basic behavioural processes.

		Stage 2		
		Counting	Timing	Choice
	Counting			
Stage 1	Timing			
	Choice			

Further reading



https://doi.org/10.1142/9789811272233_0007

Chapter 7

Stochastic Models of Buyer Behavior

Peter S. Fader[*,§], Bruce G. S. Hardie[†,¶], and Subrata Sen[‡,‖]

[*]*The Wharton School, University of Pennsylvania, Philadelphia PA, USA*
[†]*London Business School, London, UK*
[‡]*Yale School of Management, Yale University, New Haven CT, USA*
[§]*faderp@wharton.upenn.edu*
[¶]*bhardie@london.edu*
[‖]*subrata.sen@yale.edu*

7.1 Introduction

"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. Individuals vary, but percentages remain constant."

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

Buyer behavior is a complex phenomenon. We can safely assume that the actual data-generating process that lies behind any observed measure(s) of buyer behavior (e.g., the brands chosen across a series of purchase occasions, the number of times a product is purchased in a certain time period, and the time at which a new product is first purchased) 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. As such, any account of buyer

199

Customer lifetime value

Customer lifetime value is *the present value of the future profits 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 profit 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:

- According to Vodafone Group Plc’s Q3 2016 results, Vodafone UK had 12.3 million “pay monthly” customers at the end of that period.
- In his “Q4 2015 Earnings Conference Call” the CFO of Amazon made the comment that “worldwide active customer accounts were approximately 304 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 settings:

> 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 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, Morrison, and Colombo (1987).

PELOTON

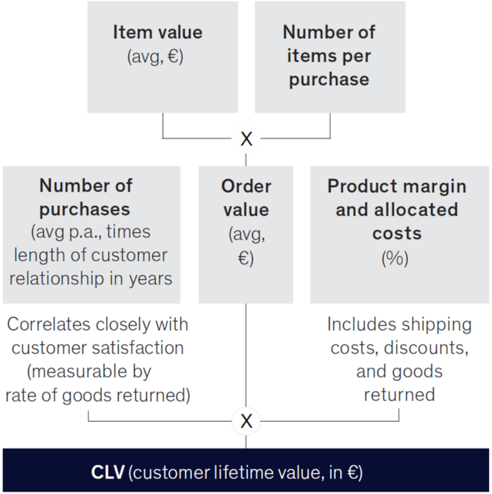
Connected Fitness Subscriber Lifetime Value

	FY 2019
Monthly subscription price	$39.00
×	
Subscriber LTV months	154
×	
(Subscription Contribution plus content costs for past use) divided by Subscription Revenue(1)	59.8%
LTV per Connected Fitness Subscriber	**$3,593**

Implied by 1 / **0.65%** Average Net Monthly Connected Fitness Churn

https://investor.onepeloton.com/static-files/73e7570a-6aca-47df-b080-a6a069139eed

https://www.tidio.com/blog/customer-lifetime-value/



https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/customer-lifetime-value-the-customer-compass

Motivating problem

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 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.

Motivating problem

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

Q1 Assuming our current prospect pool has the same characteristics as that from which these donors were acquired, what is the value of a new donor (ignoring donor acquisition costs)?

Q2 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)?

Expected value of a new donor

	1995	1996	1997		2001
$P(\text{donate})$	1.0	$\frac{5{,}652}{11{,}104}$	$\frac{4{,}674}{11{,}104}$	$\cdots$	$\frac{3{,}163}{11{,}104}$
Avg. donation	\$50	\$50	\$50	$\cdots$	\$50
Discount	1	$\frac{1}{1.1}$	$\frac{1}{(1.1)^2}$	$\cdots$	$\frac{1}{(1.1)^6}$

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

Motivating problem

We are ignoring any donations we could receive from the donor beyond Year 7.

What about Q2? ("Given their donation behavior to date, in how many of the subsequent five years can we expect a supporter to make a donation?")

> 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? …

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

A simple model for noncontractual settings

Notation

Let $Y(t)$ equal 1 if a transaction occurs at the tth transaction opportunity, 0 otherwise.

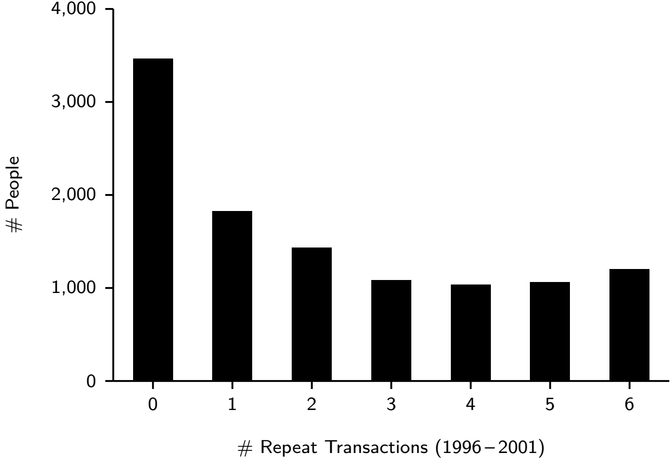
We denote the # transactions occurring in the interval $\{1, 2, \ldots, n\}$ by the random variable

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

We denote the # transactions occurring in the interval $\{n+1, n+2, \ldots, n+n^*\}$ by the random variable

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

Distribution of repeat transactions



Modelling the transaction stream

i) Each year an individual decides whether or not to support the charity by tossing a coin:

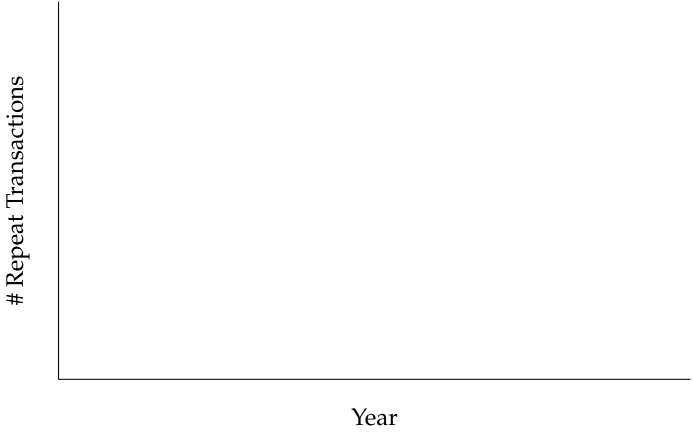
$$\begin{aligned} \mathbb{H} &\rightarrow \text{donate} \\ \mathbb{T} &\rightarrow \text{don't donate} \end{aligned}$$

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

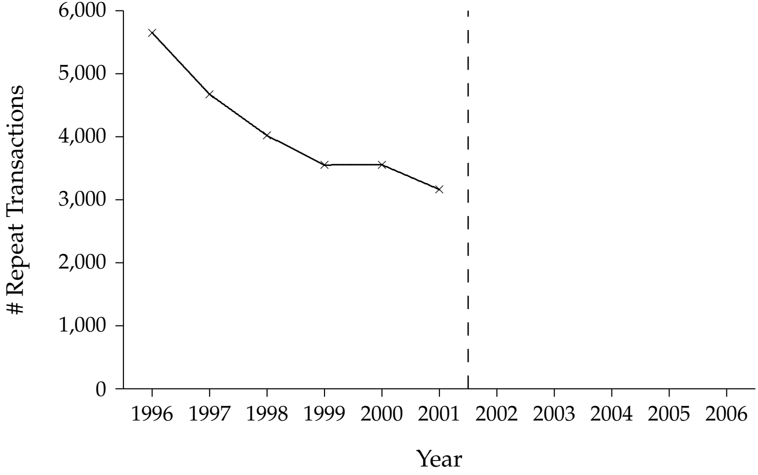
ii) An individual tosses the same coin each year.

iii) $P(\mathbb{H})$ varies across individuals.

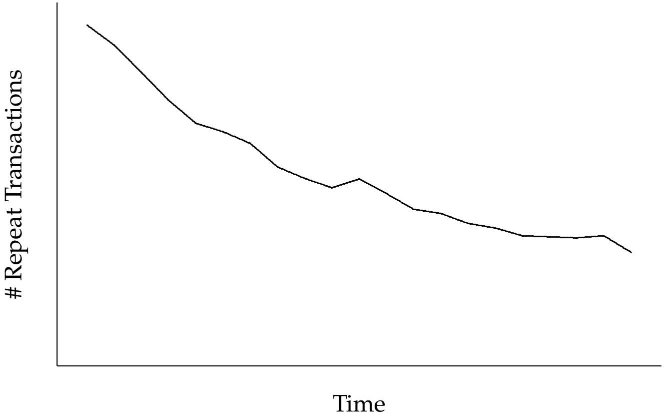
Modelling the transaction stream



Annual repeat transactions



Towards a more realistic model

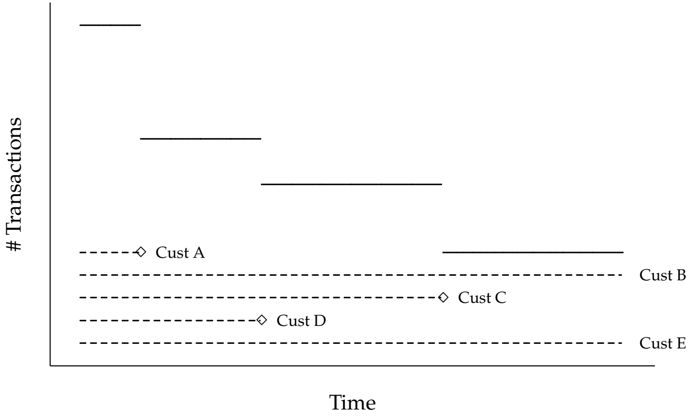


Towards a more realistic model

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 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 $\Longrightarrow$ 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

$\longrightarrow$

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),

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

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

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

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

Model development

Removing the conditioning on the latent traits p and θ,

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

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

Model implementation

	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)

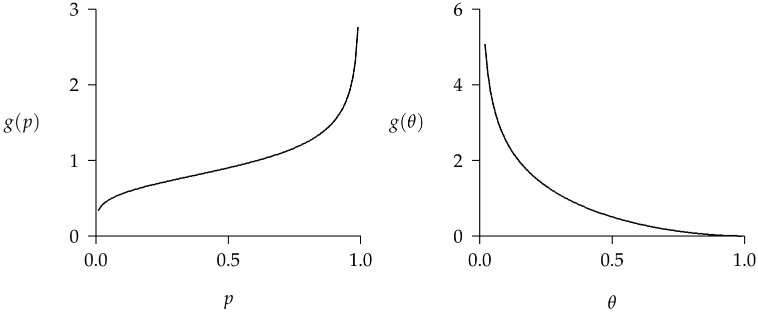
=IF(I$8<=$G9,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+I$8)-GAMMALN($B$1+$B$2+$B9+I$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+I$8)-GAMMALN($B$3+$B$4+$B9+I$8+1))/$E$3,0)

=C15-B15-1

=D19*LN(F19))

=SUM(H19:N19)

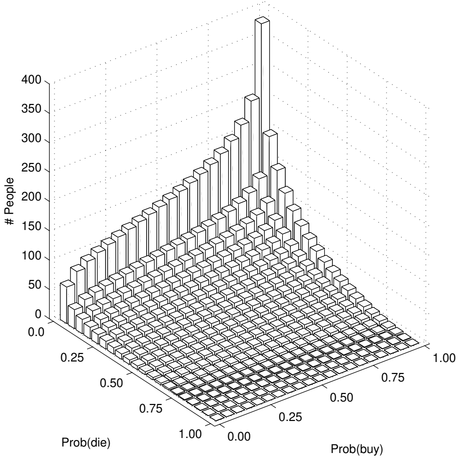
Estimated beta distributions



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

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

Joint distribution of “buy” and “die” coins



Key results: $P(Y(t) = 1)$

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

By conditioning, $P(Y(t) = 1) = P(Y(t) = 1 \mid \text{alive at } t) \times P(\text{alive at } t)$.

Recalling our model assumptions,

- $P(Y(t) = 1 \mid p, \text{alive at } t) = p$
- $P(\text{alive at } t \mid \theta) = S(t) = (1 - \theta)^t$

it follows that $P(Y(t) = 1 \mid p, \theta) = p(1 - \theta)^t$.

But p and θ are unobserved …

Key results: $P(Y(t) = 1)$

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 be computed using the recursion

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

Key results: $E[X(n)]$

We are interested in the expected # transactions occurring in $\{1, 2, \ldots, n\}$.

Conditional on p and θ,

	$t = 1$	$t = 2$	$\cdots$	$t = n$
$P(\text{buy} \mid \text{alive})$	p	p		p
$P(\text{alive})$	$(1-\theta)$	$(1-\theta)^2$		$(1-\theta)^n$

$$
\begin{aligned}
E[X(n) \mid p, \theta] &= \sum_{t=1}^{n} p(1-\theta)^t \\
&= \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n+1}}{\theta}\,.
\end{aligned}
$$

Key results: $E[X(n)]$

Removing the conditioning on p and θ gives us

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

Alternatively,

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

Key results: $P(X(n) = x)$

We are interested in the distribution of (repeat) transactions in the interval $\{1, 2, \ldots, n\}$.

Noting that in order to make x transactions, the customer must have been alive for at least x periods,

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

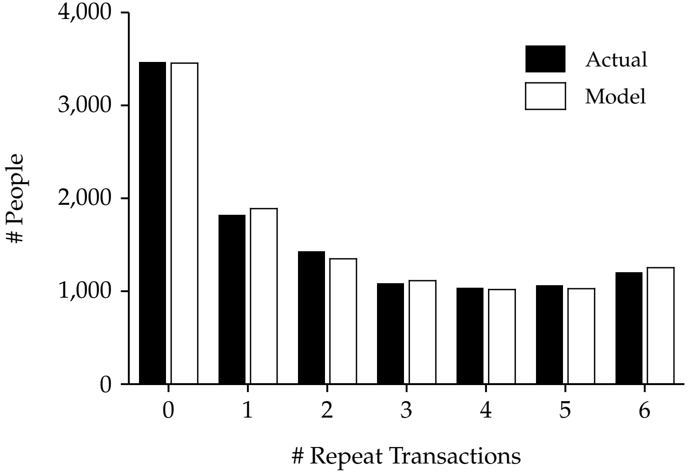
Key results: $P(X(n, n+n^*) = x^*)$

We are interested in the distribution of transactions in the post-calibration-period interval $\{n+1, \ldots, n+n^*\}$.

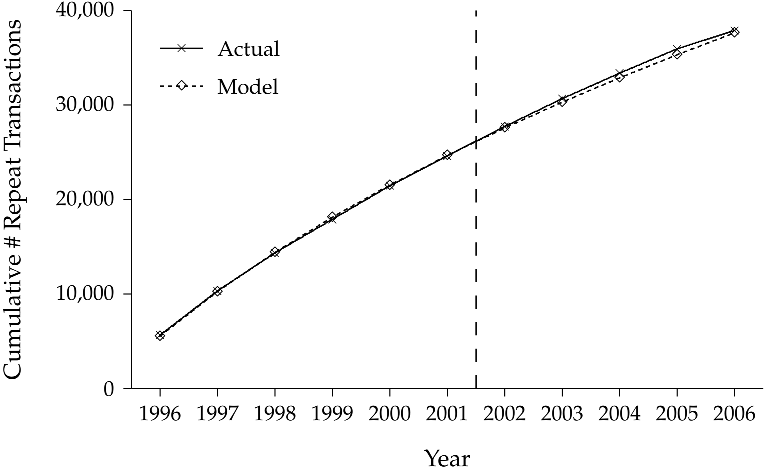
We note that one reason for $x^* = 0$ is that the customer died in the interval $\{1, 2, \ldots, n\}$.

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

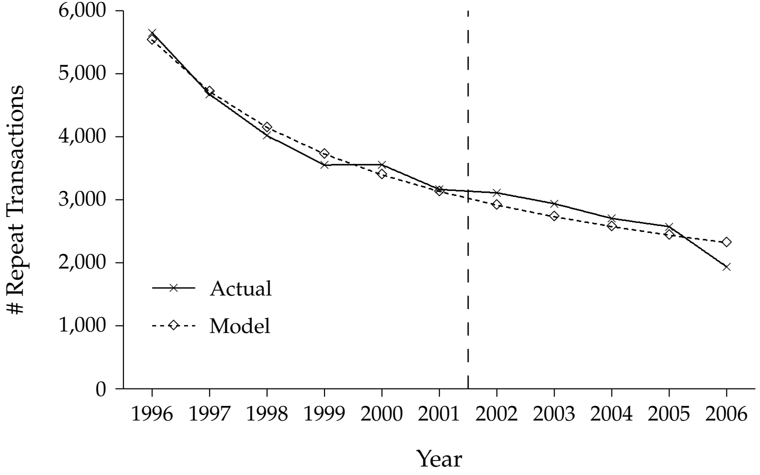
Fit of the BG/BB model



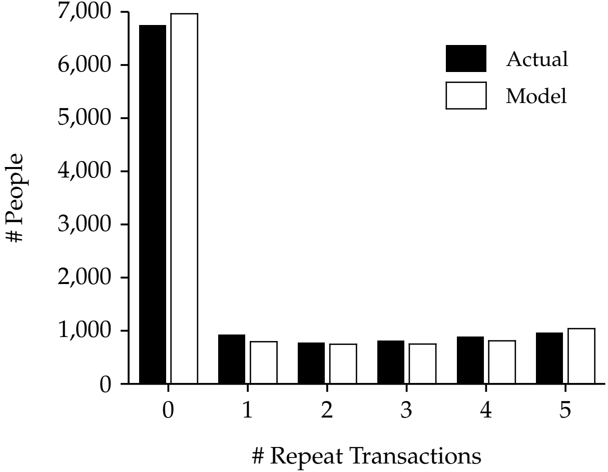
Tracking cumulative repeat transactions



Tracking annual repeat transactions



Repeat transactions in 2002 – 2006



Motivating problem (Q1)

	Year 1	Year 2	...	Year n
$P(\text{donate})$	1.0	$P(Y(1)=1)$		$P(Y(n-1)=1)$
Avg. donation	$v(0)$	$v(1)$		$v(n-1)$
Discount	1	$\frac{1}{(1+d)}$		$\frac{1}{(1+d)^{n-1}}$

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

which, assuming $v(t) = \bar{v}\ \forall t$,

$$= \bar{v}\left\{1 + \sum_{t=1}^{\infty} \frac{P(Y(t) = 1)}{(1+d)^t}\right\}$$

Motivating problem (Q1)

	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 (Q2)

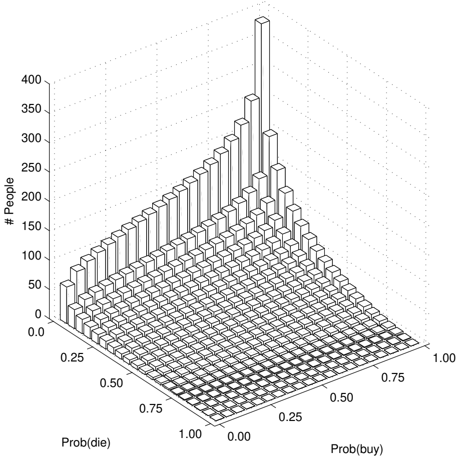
Given their donation behavior to date, in how many of the subsequent five years can we expect a supporter to make a donation?

> 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)?

Motivating problem (Q2)

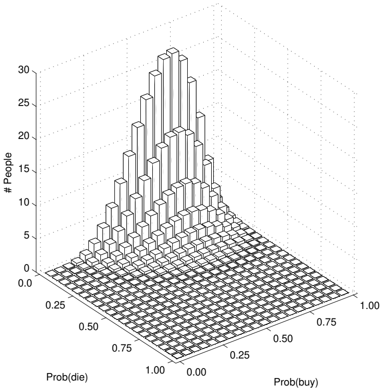
ID	1995	1996	1997	1998	1999	2000	2001	2002 – 2006
100001	1	0	0	0	0	0	0	?
100002	1	0	0	0	0	0	0	?
100003	1	0	0	0	0	0	0	?
100004	1	0	1	0	1	1	1	?
100005	1	0	1	1	1	0	1	?
100006	1	1	1	1	0	1	0	?
100007	1	1	0	1	0	1	0	?
100008	1	1	1	1	1	1	1	?
100009	1	1	1	1	1	1	0	?
100010	1	0	0	0	0	0	0	?
⋮			⋮			⋮		
111102	1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

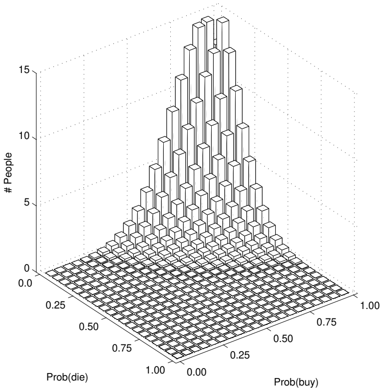


Joint distribution of “buy” and “die” coins

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

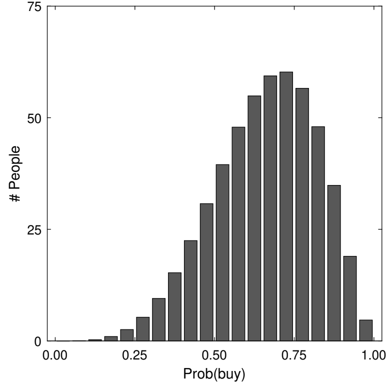


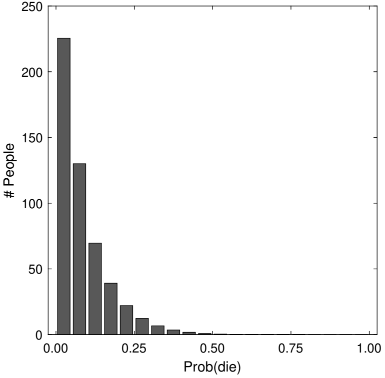
$x = 5, t_x = 5, n = 6$
335 people



Marginal distributions of “buy” and “die” coins

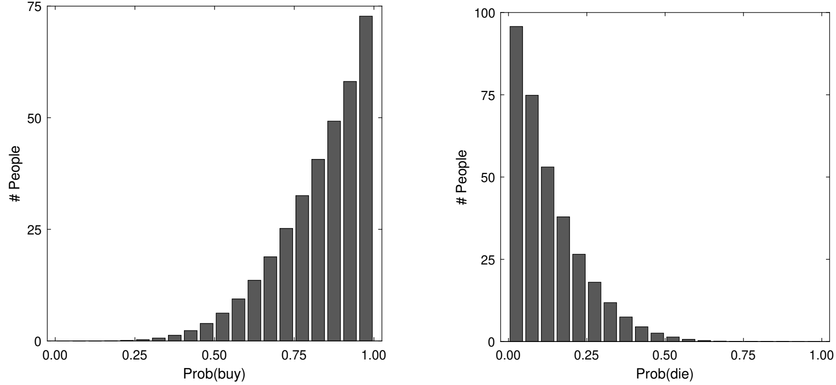
$x = 4, t_x = 6, n = 6$ (512 people)





Marginal distributions of “buy” and “die” coins

$\mathbf{x = 5, t_x = 5, n = 6}$ (335 people)



100004 vs. 100009

	100004	100009
$E(P)$	0.65	0.83
$E(\Theta)$	0.07	0.12
Expected # donations in $2002-2006 \mid$		
alive in 2001:	2.71	3.23
$P(\text{alive in } 2001)$:	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:

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

Conditional expectations

Therefore,

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

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)

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

- 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

$$\begin{aligned} \mathsf{A} &= P(x, t_x, n \mid \text{alive at } n)\, P(\text{alive at } n) \\ \mathsf{B} &= P(x, t_x, n \mid \text{dead at } n)\, P(\text{dead at } n) \end{aligned}$$

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

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

Conditional expectations

	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

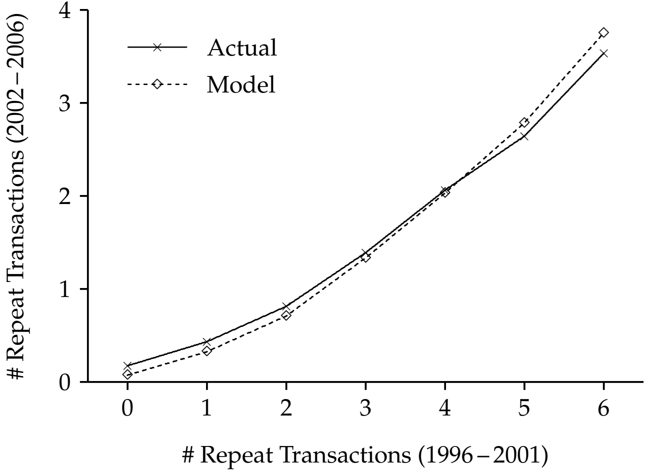
Expected # transactions in 2002 – 2006 as a function of recency and frequency

# 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

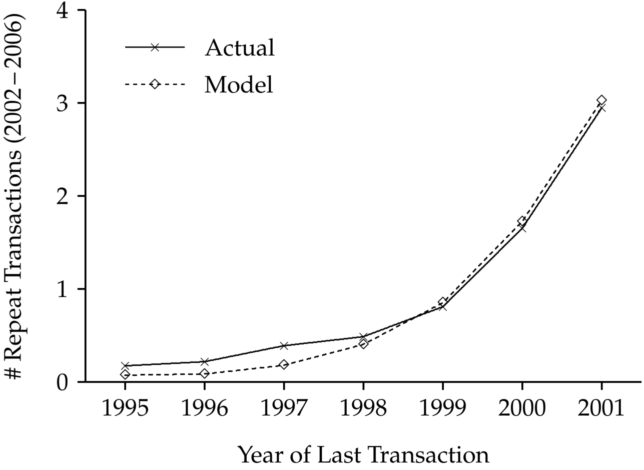
Actual # transactions in 2002 – 2006 as a function of recency and frequency

# 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



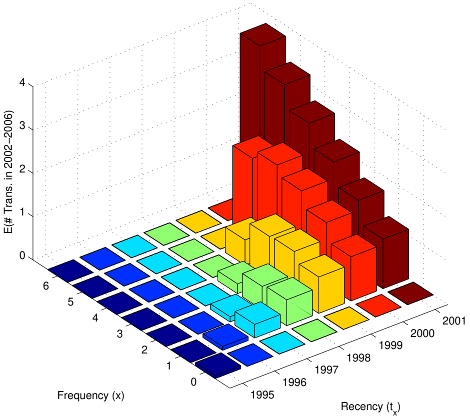
Conditional expectations by recency



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

Computing $E(CLV)$

By definition,

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

Assuming that an individual's contribution margin per transaction is constant, $v(t) = \text{CM / transaction} \times y(t)$.

The expected lifetime value of a "just-acquired" customer can be expressed as

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

Computing $E(DT)$

The quantity $E(DT)$ 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
$P(\text{buy} \mid \text{alive})$	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 $E(DT)$

Therefore,

$$E[DT(d) \mid p, \theta] = \sum_{s=1}^{\infty} p \left(\frac{1-\theta}{1+d} \right)^s$$
$$= \frac{p(1-\theta)}{d+\theta} \,.$$

Taking expectations over the distributions of p and θ,

$$E[DT(d) \mid \alpha, \beta, \gamma, \delta] = \int_0^1 \int_0^1 E[DT(d) \mid p, \theta] g(p \mid \alpha, \beta) g(\theta \mid \gamma, \delta) dp \, d\theta$$
$$= \left(\frac{\alpha}{\alpha+\beta} \right) \left(\frac{\delta}{\gamma+\delta} \right) \frac{{}_2F_1\big(1, \delta+1; \gamma+\delta+1; \frac{1}{1+d}\big)}{(1+d)}$$

Computing $E(RLV)$

Standing at time n,

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

The quantity $E(DRT)$ is the present value of the expected future transaction stream for a customer with a given transaction history.

Computing $E(DRT)$

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

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

Computing $E(DRT)$

	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)	E(DRT)	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			044.3	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)

$E(DRT)$ 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

Conditional expectations by 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

$E(DRT)$ versus conditional expectations

For any given analysis setting, the $E(DRT)$ numbers differ from the conditional expectations by a constant, independent of the customer's exact purchase history.

In this empirical setting, $E(DRT) = 1.575 \times \text{CE}$.

As a result, any ranking of customers on the basis of $E(DRT)$ 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 $E(DT)$ and $E(DRT)$ for noncontractual settings and their evaluation when the transaction stream is characterized by the BG/BB model.

Handling multiple cohorts

Approach 1:

Pool across cohorts (i.e., assume same parameters for each cohort).

Approach 2:

Estimate cohort-specific parameters that are a function of time of acquisition.

Approach 3:

Use a *changepoint model* to capture differences across "regimes of cohorts."

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/`

Gopalakrishnan, Arun, Eric T. Bradlow, and Peter S. Fader (2017), "A Cross-Cohort Changepoint Model for Customer- Base Analysis," *Marketing Science*, **36** (March–April), 195–213.

“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

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?

Modelling count data in continuous time

Let the random variable $X(t)$ denote the number of transactions in the interval $(0, t]$.

Assume that, conditional on λ (the mean number of transactions per unit of time), $X(t)$ is distributed Poisson:

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

Transaction means (λ) are distributed across the population according to a gamma distribution.

The gamma distribution

The gamma distribution is a flexible (and mathematically convenient) two-parameter distribution defined over the positive real line:

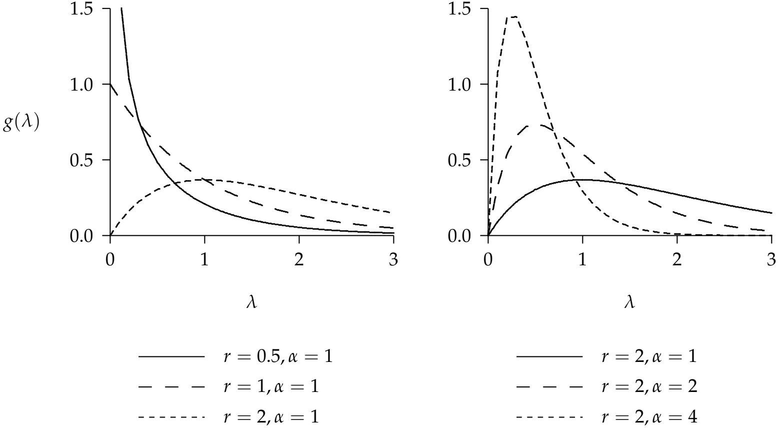
$$g(\lambda \,|\, r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)},$$

where r is the “shape” parameter and α is the “inverse scale” parameter.

The mean and variance are given by

$$E(\Lambda) = \frac{r}{\alpha} \;\text{ and }\; \text{var}(\Lambda) = \frac{r}{\alpha^2}.$$

Illustrative gamma density functions



Modelling count data in continuous time

For a randomly chosen individual,

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

This *gamma mixture of Poissons* is called the Negative Binomial Distribution (NBD).

The mean and variance are the NBD are

$$E[X(t) \mid r, \alpha] = \frac{rt}{\alpha} \text{ and } \operatorname{var}[X(t) \mid r, \alpha] = \frac{rt}{\alpha} + \frac{rt^2}{\alpha^2} .$$

Modelling timing data in continuous time

Let the random variable T denote the time at which the event of interest occurs.

Assume that, conditional on λ (the event rate), T is distributed exponential:

$$
\begin{aligned}
F(t \mid \lambda) &= P(T \leq t \mid \lambda) \\
&= 1 - e^{-\lambda t} \\
E(T \mid \lambda) &= \frac{1}{\lambda}
\end{aligned}
$$

Event rates (λ) are distributed across the population according to a gamma distribution.

Modelling timing data in continuous time

For a randomly chosen individual,

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

This *gamma mixture of exponentials* is known as the Pareto distribution of the second kind (Pareto Type II).

The mean and variance are the Pareto Type II are

$$E(T \mid r, \alpha) = \frac{\alpha}{r-1} \text{ and } \operatorname{var}(T \mid r, \alpha) = \frac{r\alpha^2}{(r-1)^2(r-2)} ,$$

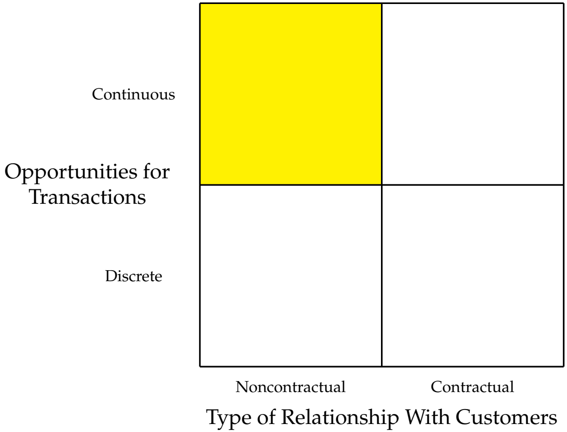
with median $\alpha(2^{1/r} - 1)$.

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

Classifying customer bases



Setting

New customers at CDNOW, January 1997 – March 1997

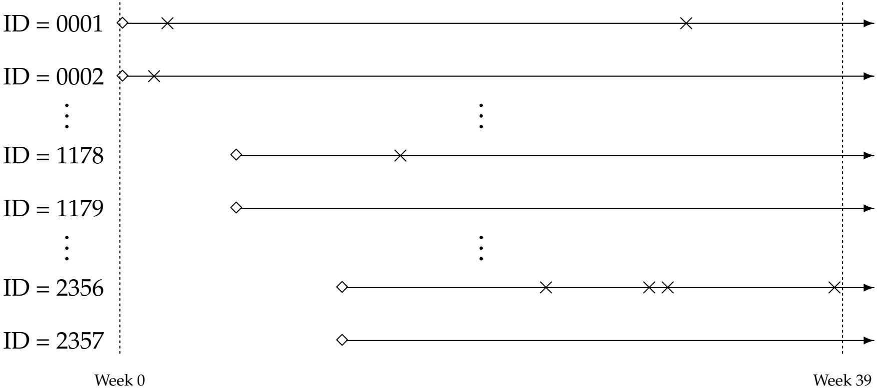
Systematic sample (1/10) drawn from panel of 23,570 new customers

39-week calibration period

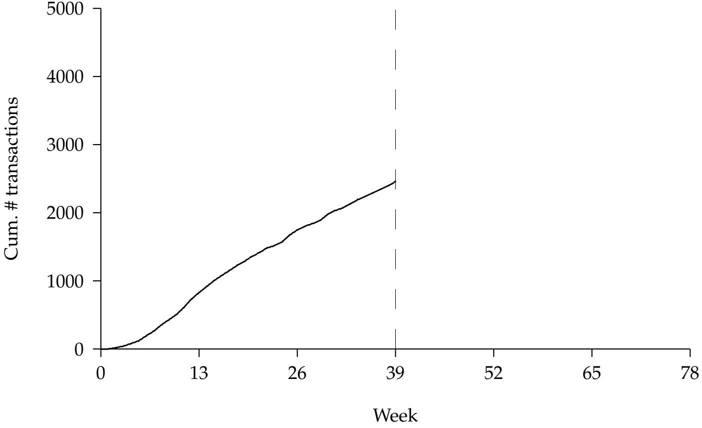
39-week forecasting (holdout) period

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).

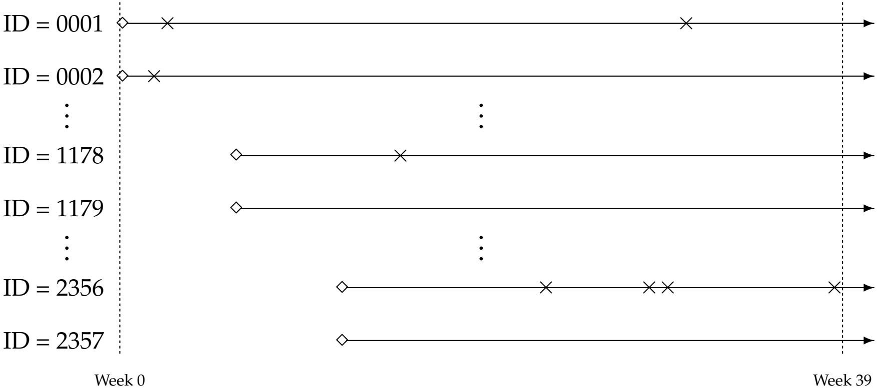
Summarizing purchase histories

Given the model assumptions, we do not require data on when each of the x transactions occurred.

The only customer-level data required by this model are *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



Purchase histories

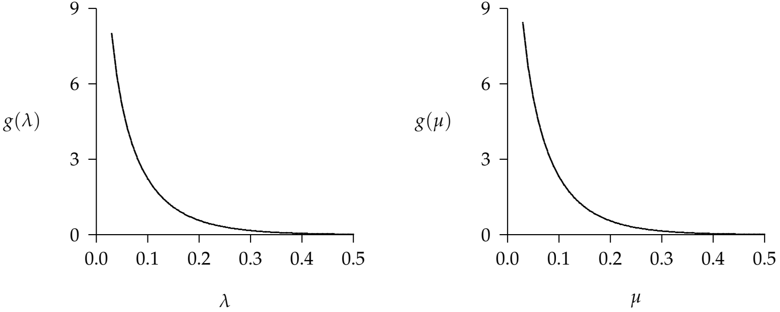
	A	B	C	D
1	ID	x	t_x	T
2	0001	2	30.43	38.86
3	0002	1	1.71	38.86
4	0003	0	0.00	38.86
5	0004	0	0.00	38.86
6	0005	0	0.00	38.86
7	0006	7	29.43	38.86
8	0007	1	5.00	38.86
9	0008	0	0.00	38.86
10	0009	2	35.71	38.86
11	0010	0	0.00	38.86
12	0011	5	24.43	38.86
13	0012	0	0.00	38.86
14	0013	0	0.00	38.86
15	0014	0	0.00	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

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

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

Estimated gamma distributions



$r = 0.553$, $\alpha = 10.578$
$E(\Lambda) = 0.0522$

$s = 0.606$, $\beta = 11.669$
$E(M) = 0.0519$

Key results

Distribution of transactions in the interval $(t, t+t^*]$:

$$
\begin{aligned}
&P(X(t, t+t^*) = x \,|\, r, \alpha, s, \beta) \\
&\quad = \delta_{x=0}\left[1 - \left(\frac{\beta}{\beta+t}\right)^s\right] + \frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha+t^*}\right)^r\left(\frac{t^*}{\alpha+t^*}\right)^x\left(\frac{\beta}{\beta+t+t^*}\right)^s \\
&\qquad + \alpha^r\beta^s\frac{B(r+x, s+1)}{B(r,s)}\left\{\mathsf{B}_1 - \sum_{i=0}^{x}\frac{\Gamma(r+s+i)}{\Gamma(r+s)i!}{t^*}^i\mathsf{B}_2\right\}
\end{aligned}
$$

where

$$
\mathsf{B}_1 = \begin{cases} {}_2F_1\left(r+s, s+1; r+s+x+1; \frac{\alpha-(\beta+t)}{\alpha}\right)/\alpha^{r+s} & \text{if } \alpha \geq \beta+t \\ {}_2F_1\left(r+s, r+x; r+s+x+1; \frac{\beta+t-\alpha}{\beta+t}\right)/\beta^{r+s} & \text{if } \alpha \leq \beta+t \end{cases}
$$

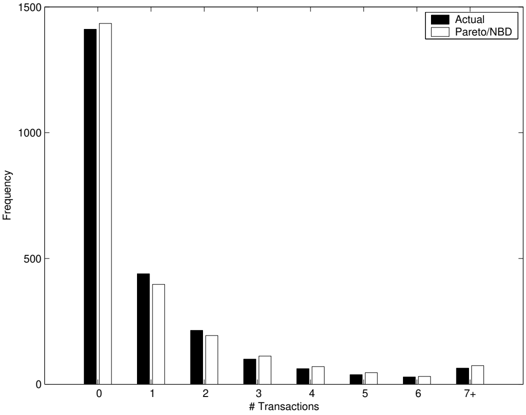
$$
\mathsf{B}_2 = \begin{cases} {}_2F_1\left(r+s+i, s+1; r+s+x+1; \frac{\alpha-(\beta+t)}{\alpha+t^*}\right)/(\alpha+t^*)^{r+s+i} & \text{if } \alpha \geq \beta+t \\ {}_2F_1\left(r+s+i, r+x; r+s+x+1; \frac{\beta+t-\alpha}{\beta+t+t^*}\right)/(\beta+t+t^*)^{r+s+i} & \text{if } \alpha \leq \beta+t \end{cases}
$$

Key results

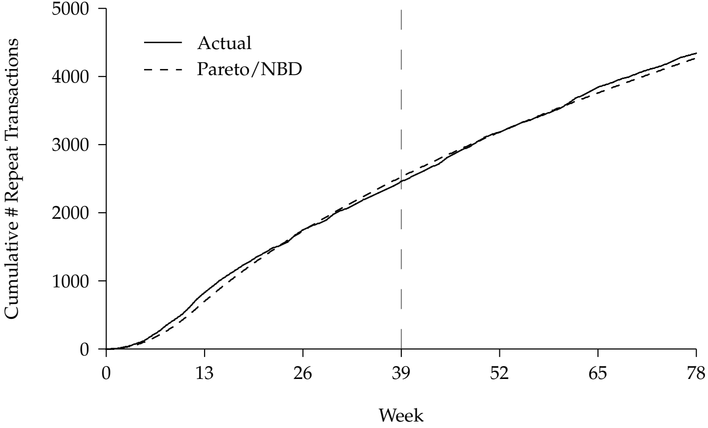
Expected number of transactions in the interval $(0, t]$:

$$E[X(t) \mid r, \alpha, s, \beta] = \frac{r\beta}{\alpha(s-1)} \left[1 - \left(\frac{\beta}{\beta + t} \right)^{s-1} \right].$$

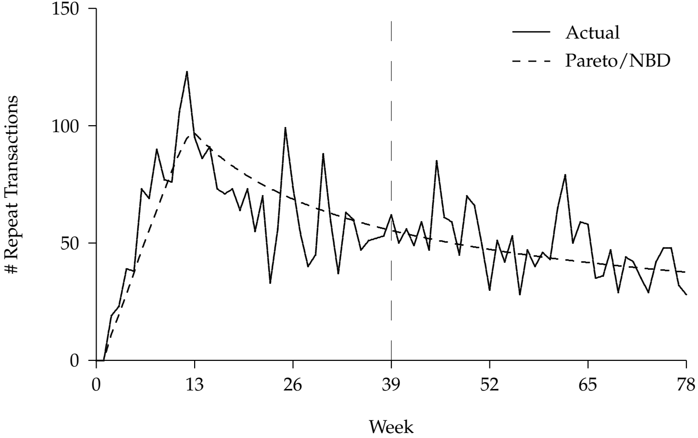
Frequency of repeat transactions



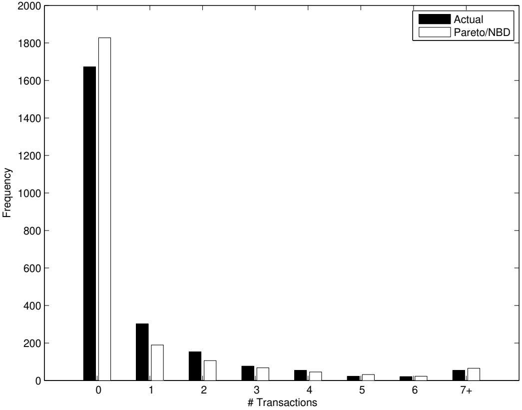
Tracking cumulative repeat transactions



Tracking weekly repeat transactions



Repeat transactions in weeks 40–78



Conditional expectations

We wish to derive an expression for the expected number of transactions in the interval $(T, T + t]$ for an individual with observed behavior (x, t_x, T).

Suppose we know that the individual is alive at T and we know their λ and μ:

$$\begin{aligned}
E[X(T, T+t) \,|\, \lambda, \mu, \omega > T] \\
&= \lambda t \, P(\Omega > T + t \,|\, \mu, \omega > T) \\
&\qquad + \int_T^{T+t} \lambda \omega f(\omega \,|\, \mu, \omega > T) \, d\omega \\
&= \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t} \,.
\end{aligned}$$

Conditional expectations

However,

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

We derive an expression for the probability that the individual is alive at T, $P(\Omega > T \mid \lambda, \mu; x, t_x, T)$.

It makes no sense to use $g(\lambda \mid r, \alpha)$ and $g(\mu \mid s, \beta)$ as we know something about the individual's behavior.

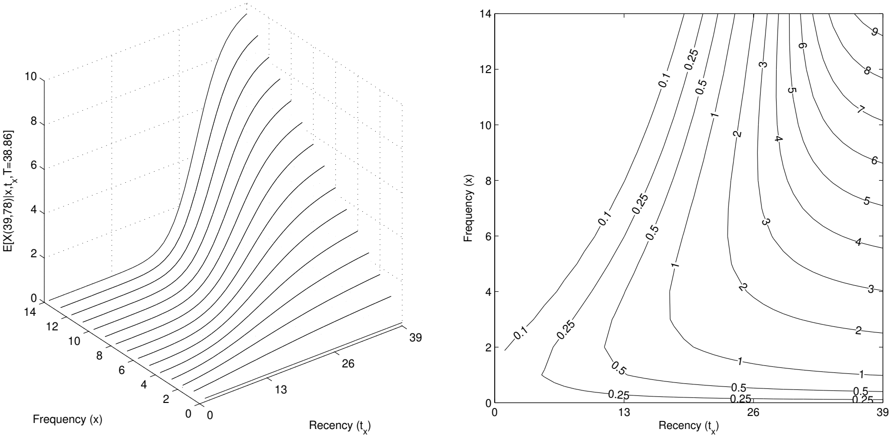
We use Bayes' theorem to derive the joint posterior distribution of Λ and M, $g(\lambda, \mu \mid r, \alpha, s, \beta; x, t_x, T)$.

Conditional expectations

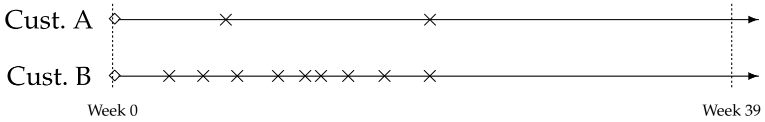
Combining the various elements gives us

$$
\begin{aligned}
E[X(T, T+t) \mid r, \alpha, s, \beta; x, t_x, T] & \\
= \int_0^\infty \int_0^\infty \Big\{ & E[X(T, T+t) \mid \lambda, \mu, \omega > T] \\
& \times P(\Omega > T \mid \lambda, \mu; x, t_x, T) \\
& \times g(\lambda, \mu \mid r, \alpha, s, \beta; x, t_x, T) \Big\} d\lambda \, d\mu \\
= \left\{ \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \middle/ L(r, \alpha, s, \beta \mid x, t_x, T) \right\} & \\
\times \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)} \left[1 - \left(\frac{\beta+T}{\beta+T+t} \right)^{s-1} \right] . &
\end{aligned}
$$

Conditional expectations by recency and frequency

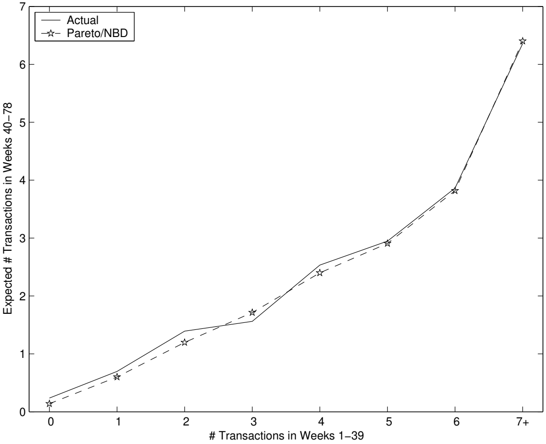


The “increasing frequency” paradox

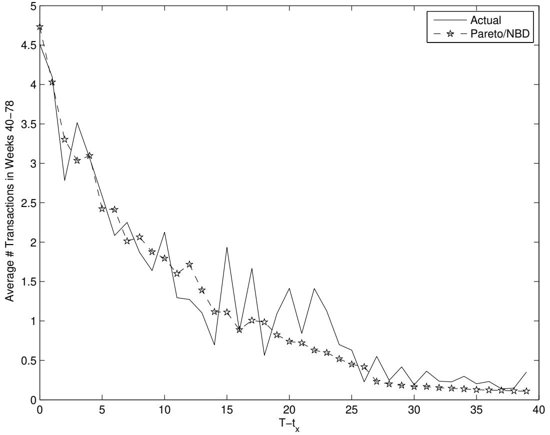


	x	t_x	T	$E[X(39,78) \mid x, t_x, T]$
Cust. A	2	20	38.86	1.00
Cust. B	9	20	38.86	0.72

Conditional expectations by frequency



Conditional expectations by recency



Computing $E(DRT)$

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

$$
\begin{aligned}
&E[DRT(\delta) \mid r, \alpha, s, \beta; x, t_x, T] \\
&\quad = \int_0^\infty \int_0^\infty \Big\{ E[DRT(\delta) \mid \lambda, \mu, \omega > T] \\
&\qquad\qquad \times P(\Omega > T \mid \lambda, \mu; x, t_x, T) \\
&\qquad\qquad\quad \times g(\lambda, \mu \mid r, \alpha, s, \beta; x, t_x, T) \Big\} d\lambda \, d\mu \\
&\quad = \frac{\alpha^r \beta^s \delta^{s-1} \Gamma(r + x + 1) \Psi\big(s, s; \delta(\beta + T)\big)}{\Gamma(r)(\alpha + T)^{r+x+1} L(r, \alpha, s, \beta \mid x, t_x, T)}
\end{aligned}
$$

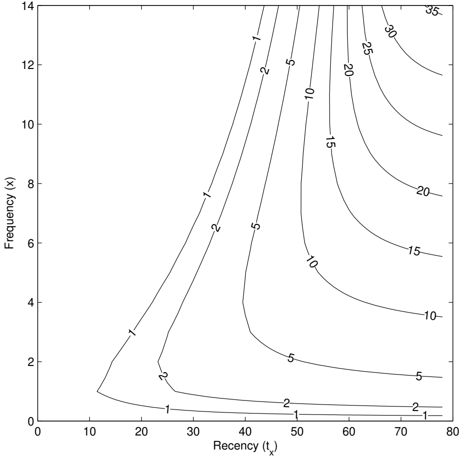
where $\Psi(\cdot)$ is the confluent hypergeometric function of the second kind.

Continuous compounding

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

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

Iso-value representation of $E(DRT)$



Key contribution

We are able to generate forward-looking estimates of $E(DRT)$ as a function of recency and frequency in a noncontractual setting:

$$E(DRT) = f(R, F)$$

Adding a sub-model for spend per transaction enables us to generate estimates of expected residual lifetime value as a function of RFM in a noncontractual setting:

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

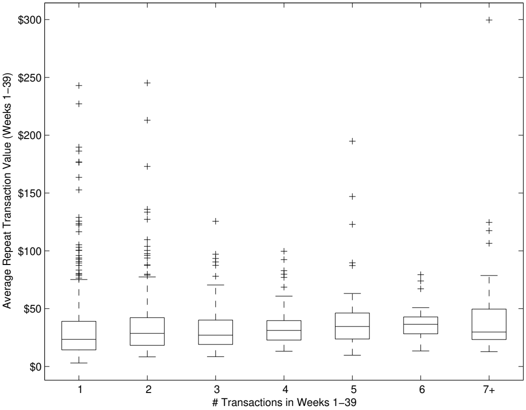
Modelling the spend process

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) made 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

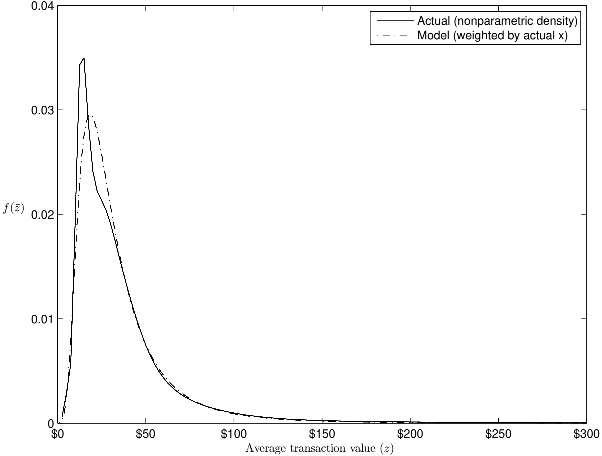
i) The dollar value of a customer's given transaction is distributed gamma with shape parameter p and scale parameter ν (which implies $\zeta = p/\nu$)

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



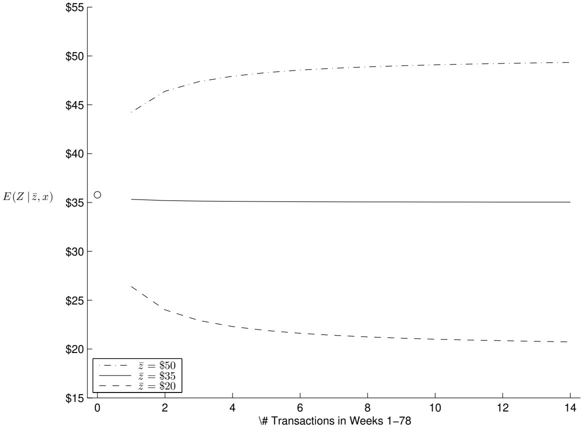
Modelling the spend process

Applying Bayes' theorem, the expected average transaction value for a customer with an average spend of $\bar{z}$ across x transactions is

$$E(Z \mid p, q, \gamma; \bar{z}, x) = \left(\frac{q-1}{px+q-1}\right)\frac{\gamma p}{q-1} + \left(\frac{px}{px+q-1}\right)\bar{z}$$

As we observe more transactions, our estimate of the individual's true mean spend tends to their observed average spend.

$E(Z)$ as a function of M and F

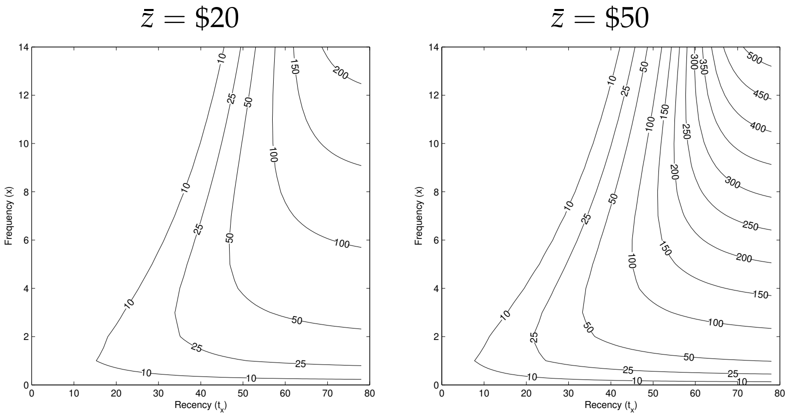


Computing expected residual lifetime value

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

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

Estimates of $E(RLV)$



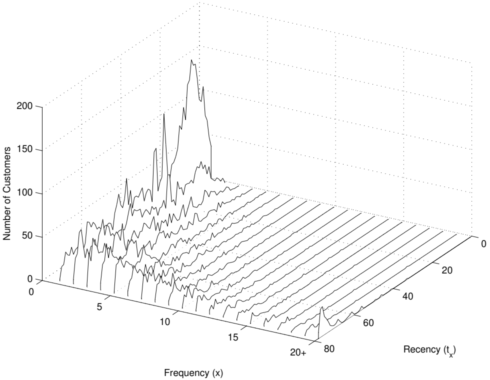
(Margin = 30%, 15% discount rate)

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

Further reading

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.

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), "RFM and CLV: Using Iso-value Curves for Customer Base Analysis," *Journal of Marketing Research*, **42** (November), 415–430.

Further reading

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/`

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

Fader, Peter S. and Bruce G. S. Hardie (2008), "Creating an RFM Summary Using Excel." `http://brucehardie.com/notes/022/`

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

An alternative to the Pareto/NBD model

Estimation of model parameters can be a barrier to Pareto/NBD model implementation

Recall the dropout process story:

- Each customer has an unobserved "lifetime"
- Death rates vary across customers

Let us consider an alternative story:

- After any transaction, a customer tosses a coin

  $\mathsf{H} \rightarrow$ remain alive

  $\mathsf{T} \rightarrow$ die

- $P(\mathsf{T})$ varies across customers

The BG/NBD model

(Fader, Hardie and Lee 2005)

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

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

Model implementation

	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	[illegible]	[illegible]	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	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	0007	1	5.00	38.86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
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

Key results

Distribution of transactions in the interval $(0, t]$:

$$
\begin{aligned}
P(X(t) = x \mid r, \alpha, a, b) \\
&= \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \\
&\quad + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \\
&\quad\quad \times \left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r \left\{ \sum_{j=0}^{x-1} \frac{(\Gamma(r+j)}{\Gamma(r)j!} \left(\frac{t}{\alpha+t}\right)^j \right\}\right]
\end{aligned}
$$

Key results

Distribution of transactions in the interval $(t, t+t^*]$:

$$P(X(t,t+t^*) = x \mid r, \alpha, a, b) = \mathsf{A}_1 + \delta_{x=0}\,\mathsf{A}_2 + \delta_{x>0}\,\mathsf{A}_3 - \delta_{x>0}\sum_{j=0}^{x-1}\mathsf{A}_{4j}$$

where

$$\begin{aligned}
\mathsf{A}_1 &= \frac{\Gamma(r+x)}{\Gamma(r)x!}\,\frac{B(a,b+x)}{B(a,b)}\Big(\frac{\alpha}{\alpha+t+t^*}\Big)^r\Big(\frac{t^*}{\alpha+t+t^*}\Big)^x {}_2F_1\big(r+x, b+x; a+b+x; \tfrac{t}{\alpha+t+t^*}\big)\\
\mathsf{A}_2 &= 1 - \Big(\frac{\alpha}{\alpha+t}\Big)^r {}_2F_1\big(r, b; a+b; \tfrac{t}{\alpha+t}\big)\\
\mathsf{A}_3 &= \frac{B(a+1,b+x-1)}{B(a,b)}\Big(\frac{\alpha}{\alpha+t}\Big)^r {}_2F_1\big(r, b+x-1; a+b+x; \tfrac{t}{\alpha+t}\big)\\
\mathsf{A}_{4j} &= \frac{\Gamma(r+j)}{\Gamma(r)j!}\,\frac{B(a+1,b+x-1)}{B(a,b)}\Big(\frac{\alpha}{\alpha+t+t^*}\Big)^r\Big(\frac{t^*}{\alpha+t+t^*}\Big)^j {}_2F_1\big(r+j, b+x-1; a+b+x; \tfrac{t}{\alpha+t+t^*}\big)\,.
\end{aligned}$$

Key results

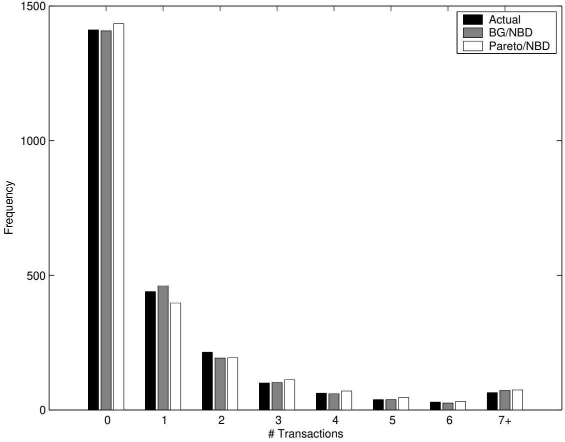
Expected number of transactions in the interval $(0, t]$:

$$E[X(t) \mid r, \alpha, a, b] = \frac{a+b-1}{a-1}\left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\left(r, b; a+b-1; \tfrac{t}{\alpha+t}\right)\right].$$

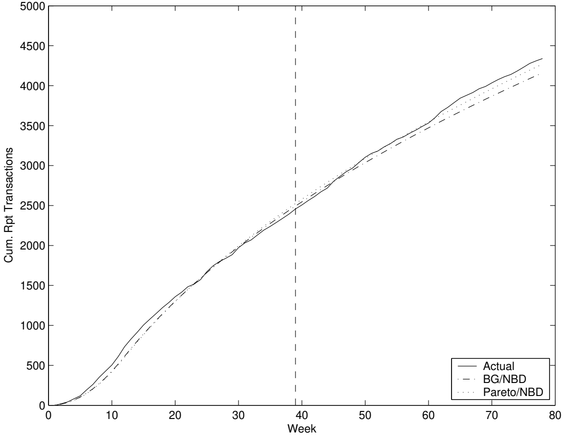
Expected number of transactions in the interval $(T, T+t]$ for an individual with observed behavior (x, t_x, T):

$$\begin{aligned}
E[X(T, T+t) \mid r, \alpha, a, b; x, t_x, T) &= \left(\frac{a+b+x-1}{a-1}\right) \\
&\times \left[1 - \left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left(r+x, b+x; a+b+x-1; \tfrac{t}{\alpha+T+t}\right)\right] \\
&\times 1 \Bigg/ \left(1 + \delta_{x>0}\frac{a}{b+x-1}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}\right).
\end{aligned}$$

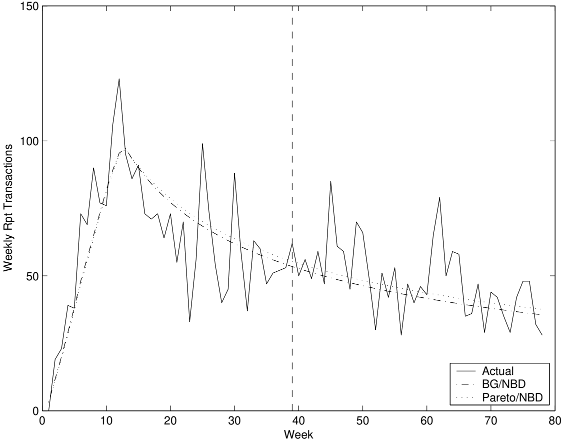
Frequency of repeat transactions



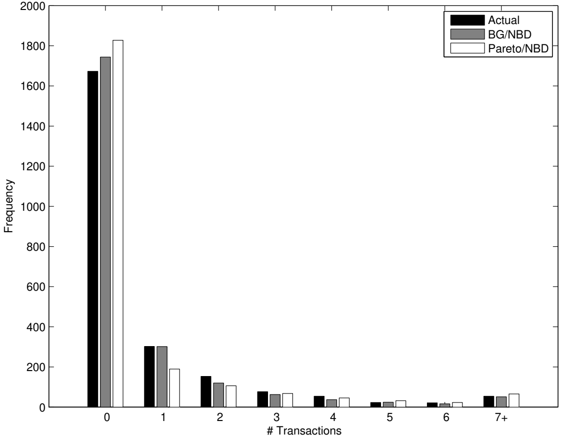
Tracking cumulative repeat transactions



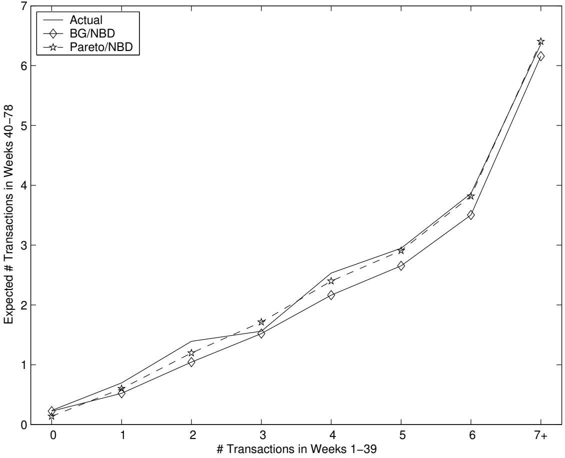
Tracking weekly repeat transactions



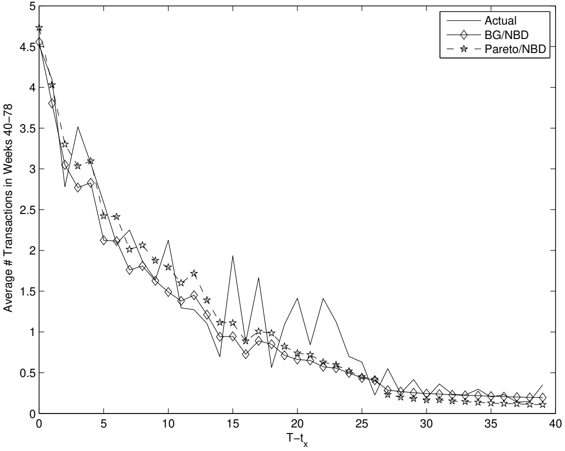
Repeat transactions in weeks 40–78



Conditional expectations by frequency



Conditional expectations by recency



Computing CLV using the BG/NBD model

The closed-form expressions for $E(DT)$ and $E(DRT)$ are very messy when the flow of transactions is characterized by the BG/NBD.

It is easier to compute these quantities in the following manner:

$$E(DT) = \sum_{t=1}^{\infty} \left(\frac{1}{1+d}\right)^{t-0.5} \Big\{ E\big[X(t)\big] - E\big[X(t-1)\big] \Big\}$$

$$\begin{aligned} E(DRT) = \sum_{t=1}^{\infty} \left(\frac{1}{1+d}\right)^{t-0.5} \Big\{ & E\big[X(T, T+t) \,|\, x, t_x, T\big] \\ & - E\big[X(T, T+t-1) \,|\, x, t_x, T\big] \Big\} \end{aligned}$$

Further reading

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), “"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 (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 (2008), “Computing P(alive) Using the BG/NBD Model.” `http://brucehardie.com/notes/021/`

Fader, Peter S. and Bruce G. S. Hardie (2013), “Overcoming the BG/NBD Model’s `#NUM!` Error Problem.” `http://brucehardie.com/notes/027/`

Further reading

Batislam, Emine Persentili, Meltem Denizel, and Alpay Filiztekin (2007), "Empirical Validation and Comparison of Models for Customer Base Analysis," *International Journal of Research in Marketing*, **24** (September), 201–209.

Hoppe, Daniel and Udo Wagner (2007), "Customer Base Analysis: The Case for a Central Variant of the Betageometric/NBD Model," *Marketing — Journal of Research and Management*, **3** (2), 75–90.

Wagner, Udo and Daniel Hoppe (2008), "Erratum on the MBG/NBD Model," *International Journal of Research in Marketing* **25** (September), 225–226.

Batislam, E. P., M. Denizel, and A. Filiztekin (2008), "Formal response to "Erratum on the MBG/NBD Model"," *International Journal of Research in Marketing* **25** (September), 227.

R and Python packages

[R] BTYD: Implementing Buy 'Til You Die Models.
`http://cran.r-project.org/package=BTYD`

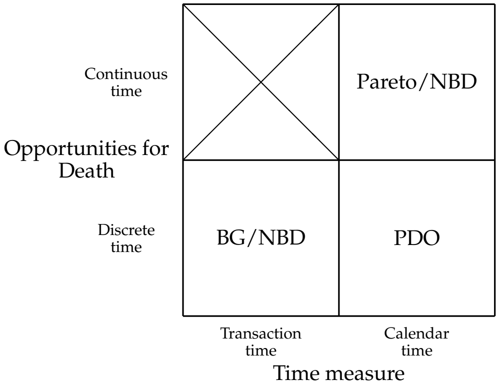
[R] BTYDplus: Probabilistic Models for Assessing and Predicting your Customer Base. `https://CRAN.R-project.org/package=BTYDplus`

[R] CLVTools: Tools for Customer Lifetime Value Estimation.
`https://CRAN.R-project.org/package=CLVTools`

[Python] lifetimes `https://github.com/CamDavidsonPilon/lifetimes`

[Python] PyMC Marketing: Open Source Marketing Analytics Solution
`https://www.pymc-marketing.io/en/stable/`

Classifying differences in assumed death processes



Jerath, Kinshuk, Peter S. Fader, and Bruce G. S. Hardie (2011), "New Perspectives on Customer 'Death' Using a Generalization of the Pareto/NBD Model," *Marketing Science*, **30** (September–October), 866–880.

Extending the basic models

Individual-level behaviour:

- Relax the assumption of Poisson/Bernoulli purchasing
- Relax the assumption of exponential/geometric lifetimes
- Relax the standard latent attrition assumption

Cross-sectional heterogeneity:

- Change the assumption of gamma/beta-distributed transaction rates
- Change the assumption of gamma/beta-distributed death rates
- Relax the assumption of independence

Covariates:

- Allow for time-invariant covariates
- Allow for time-varying covariates

Latent attrition

The standard "alive then dead" structure is but one way of capturing the decline in cohort-level sales.

> 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.
>
> Moe, Wendy W. and Peter S. Fader (2004), "Capturing Evolving Visit Behavior in Clickstream Data," *Journal of Interactive Marketing*, **18** (1), 5–19.

This standard "alive then dead" structure can be viewed as a constrained HMM. The structure of the (latent) state transition matrix can be relaxed and generalized.

Correlation

We assume independent latent traits for "mathematical convenience."

Not as "problematic" as some may think; independent priors $\not\Rightarrow$ independent posteriors

The easiest way to accommodate correlated priors is to use transformations of multivariate normals.

> 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/`

Experience to date suggests limited value … if any.

Covariates

Types of covariates:

- customer characteristics (e.g., demographics, attitudes)
- seasonality
- 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

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/

Fader, Peter S., Bruce G. S. Hardie, and Chun-Yao Huang (2020), "Incorporating Time-Varying Covariates in the BG/NBD Model." http://brucehardie.com/notes/040/

Bachmann, Patrick, Markus Meierer, and Jeffrey Näf (2021), "The Role of Time-Varying Contextual Factors in Latent Attrition Models for Customer Base Analysis," *Marketing Science*, **40** (4), 783-809.

Wünderlich, Robin, Nancy V. Wünderlich, and Florian v. Wangenheim (2022), "A Seasonal Model with Dropout to Improve Forecasts of Purchase Levels," *Journal of Interactive Marketing*, **57** (2), 212–236.

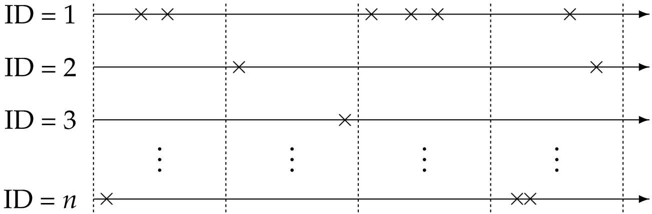
Fader, Peter S. and Bruce G. S. Hardie (2024), "Incorporating Covariates in the Gamma-Gamma Spend Model." http://brucehardie.com/notes/045/

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.

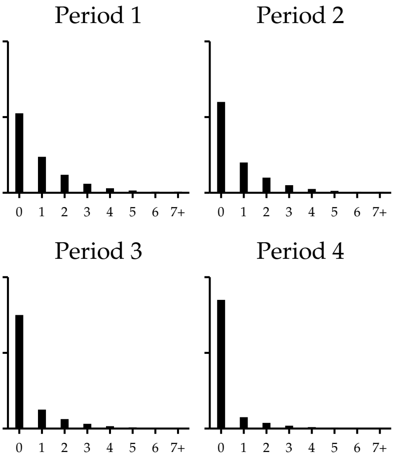
Reflections on data structures



Individual-level (x, t_x, T) are sufficient statistics for standard models

Reflections on data structures

	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



Reflections on data structures

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 (2016), "Customer-Base Analysis Using Repeated Cross-Sectional Summary (RCSS) Data," *European Journal of Operational Research*, **249** (1), 340–350.

Reflections on data structures

Overstock.com (Publicly Disclosed Data)

Period	QREV	QADD	QTO	QAU
Q1 1999	NA	NA	NA	NA
Q2 1999	NA	NA	NA	NA
Q3 1999	NA	NA	NA	NA
Q4 1999	NA	NA	NA	NA
Q1 2000	NA	NA	NA	NA
Q2 2000	NA	NA	NA	NA
Q3 2000	NA	NA	NA	NA
Q4 2000	NA	NA	NA	NA
Q1 2001	9578	99	151	NA
Q2 2001	7407	68	112	NA
Q3 2001	8744	81	131	NA
Q4 2001	14274	137	223	NA
Q1 2002	12067	105	177	NA
Q2 2002	14380	118	212	NA
Q3 2002	23808	164	291	NA
Q4 2002	41529	348	579	NA
Q1 2003	29164	264	491	NA
Q2 2003	28833	283	522	NA
Q3 2003	57788	342	643	NA
Q4 2003	123160	744	1376	NA
Q1 2004	82078	425	1126	NA
Q2 2004	87792	414	1023	NA
Q3 2004	103444	514	1178	864
Q4 2004	221321	1110	2440	1682
⋮	⋮	⋮	⋮	⋮
Q4 2015	480270	1163	3463	2423
Q1 2016	413677	857	2675	1894
Q2 2016	418540	2602	1816	NA
Q3 2016	441564	2703	1886	NA
Q4 2016	526182	3465	2372	NA
Q1 2017	432435	2606	1811	NA

Reflections on data structures

Article

Journal of Marketing Research
2018, Vol. 55(5) 617-635

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0022243718802843
journals.sagepub.com/home/jmx

Customer-Based Corporate Valuation for Publicly Traded Noncontractual Firms

Daniel M. McCarthy and Peter S. Fader

Abstract

There is growing interest in "customer-based corporate valuation"—that is, explicitly tying the value of a firm's customer base to its overall financial valuation using publicly disclosed data. While much progress has been made in building a well-validated customer-based valuation model for contractual (or subscription-based) firms, there has been little progress for noncontractual firms. Noncontractual businesses have more complex transactional patterns because customer churn is not observed, and customer purchase timing and spend amounts are more irregular. Furthermore, publicly disclosed data are aggregated over time and across customers, are often censored, and may vary from firm to firm, making it harder to estimate models for customer acquisition, ordering, and spend per order. The authors develop a general customer-based valuation methodology for noncontractual firms that accounts for these issues. They apply this methodology to publicly disclosed data from e-commerce retailers Overstock.com and Wayfair, provide valuation point estimates and valuation intervals for the firms, and compare the unit economics of newly acquired customers.

Keywords

customer equity, customer lifetime value, unit economics, valuation

Online supplement: https://doi.org/10.1177/0022243718802843

Summary

Customer lifetime value

Customer lifetime value is *the present value of the future profits 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 cash flow per period while “alive”?

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, Morrison, and Colombo (1987).

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
- Purposely ignore the effects of covariates and other "complexities" at the outset.

Make everything as simple as possible, but not simpler.

Albert Einstein

Further reading

Ascarza, Eva, Peter S. Fader, and Bruce G. S. Hardie (2017), "Marketing Models for the Customer-Centric Firm," in *Handbook of Marketing Decision Models* (Second Edition), Berend Wierenga and Ralf van der Lans (eds.), Cham, Switzerland: Springer International, 297–329.

Fader, Peter S. and Bruce G. S. Hardie (2009), "Probability Models for Customer-Base Analysis," *Journal of Interactive Marketing*, **23** (January), 61–69.

Fader, Peter S. and Bruce G. S. Hardie (2015), "Simple Probability Models for Computing CLV and CE," in *The Handbook of Customer Equity*, V. Kumar and Denish Shah (eds.), Cheltenham, UK: Edward Elgar Publishers, 77–100.

Discussion

Keep in contact / tell us your stories

Pete:	`faderp@wharton.upenn.edu`
	`www.petefader.com`
	`@faderp`
Bruce:	`bhardie@london.edu`
	`www.brucehardie.com`