Probability Models for Customer-Base Analysis

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

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

16th Annual Advanced Research Techniques Forum
June 12–15, 2005

Customer-Base Analysis

Faced with a customer transaction database, we may wish to determine

- which customers are most likely to be active in the future,
- the level of transactions we could expect in future periods from those on the customer list, both individually and collectively, and
- individual customer lifetime value (CLV).

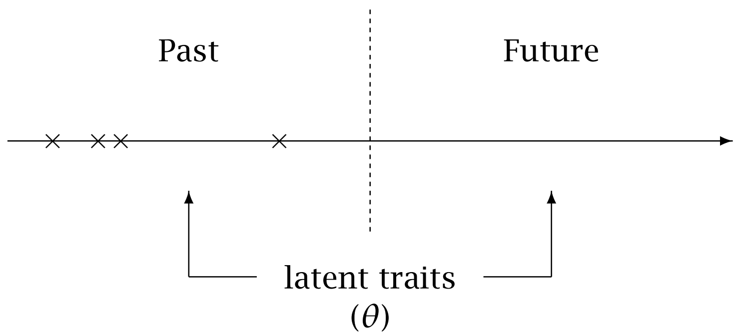
Classifying Customer Bases

- It is important to distinguish between contractual and noncontractual settings:
  - In a *contractual* setting, we observe the time at customers become inactive.
  - In a *noncontractual* setting, the time at which a customer becomes inactive is unobserved.
- The challenge of noncontractual markets:

  How do we differentiate between those customers who have ended their relationship with the firm versus those who are simply in the midst of a long hiatus between transactions?

Setting the Stage: Comparison of Modelling Approaches

Traditional approach
future = f(past)



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

Outline

Part 1: Review of Probability Models

Part 2: Models for Noncontractual Settings

Part 3: Model for Contractual Settings

Part 4: Conclusions

Part 1
Review of Probability Models

The Logic of Probability Models

- Many researchers attempt to describe/predict behavior using observed variables.
- However, they still use random components in recognition that not all factors are included in the model.
- We treat behavior as if it were "random" (probabilistic, stochastic).
- We propose a model of individual-level behavior which is "summed" across heterogeneous individuals to obtain a model of aggregate behavior.

Building a Probability Model

(i) Determine the marketing decision problem/ information needed.

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

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

Building a Probability Model

(iv) Specify a distribution to characterize the distribution of the latent trait variable(s) across the population.

- We denote this by $g(\theta)$.
- This is often called the *mixing distribution*.

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

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

Building a Probability Model

(vi) Estimate the parameters (of the mixing distribution) by fitting the aggregate distribution to the observed data.

(vii) Use the model to solve the marketing decision problem/provide the required information.

“Classes” of Models

- We focus on three fundamental behavioral processes:
  - Timing ⟶ “when”
  - Counting ⟶ “how many”
  - “Choice” ⟶ “whether/which”
- Our toolkit contains simple models for each behavioral process.
- More complex behavioral phenomena can be captured by combining models from each of these processes.

Individual-level Building Blocks

Count data arise from asking the question, “How many?”. As such, they are non-negative integers with no upper limit.

Let the random variable X be a count variable:

X is distributed Poisson with mean λ if

$$P(X = x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \; x = 0, 1, 2, \ldots$$

Individual-level Building Blocks

Timing (or duration) data are generated by answering "when" and "how long" questions, asked with regards to a specific event of interest.

> The models we develop for timing data are also used to model other non-negative continuous quantities (e.g., transaction value).

Let the random variable T be a timing variable:

> T is distributed exponential with rate parameter λ if
>
> $$F(t \mid \lambda) = P(T \le t \mid \lambda) = 1 - e^{-\lambda t}, \; t > 0 .$$

Individual-level Building Blocks

A Bernoulli trial is a probabilistic experiment in which there are two possible outcomes, 'success' (or '1') and 'failure' (or '0'), where p is the probability of success.

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

Individual-level Building Blocks

Let the random variable X be the number of independent and identically distributed Bernoulli trials required until the first success:

X is a (shifted) geometric random variable, where

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

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

Individual-level Building Blocks

Let the random variable X be the total number of successes occurring in n independent and identically distributed Bernoulli trials:

X is distributed binomial with parameter p, where

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

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

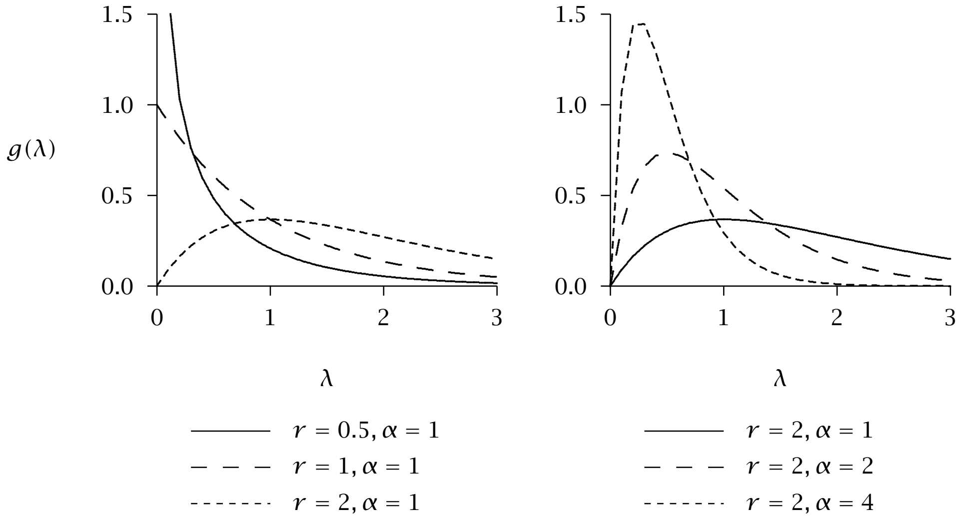
Capturing Heterogeneity in Latent Traits

The gamma distribution:

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

- $\Gamma(\cdot)$ is the gamma function
- r is the “shape” parameter and α is the “scale” parameter
- The gamma distribution is a flexible (unimodal) distribution … and is mathematically convenient.

Illustrative Gamma Density Functions



Capturing Heterogeneity in Latent Traits

The beta distribution:

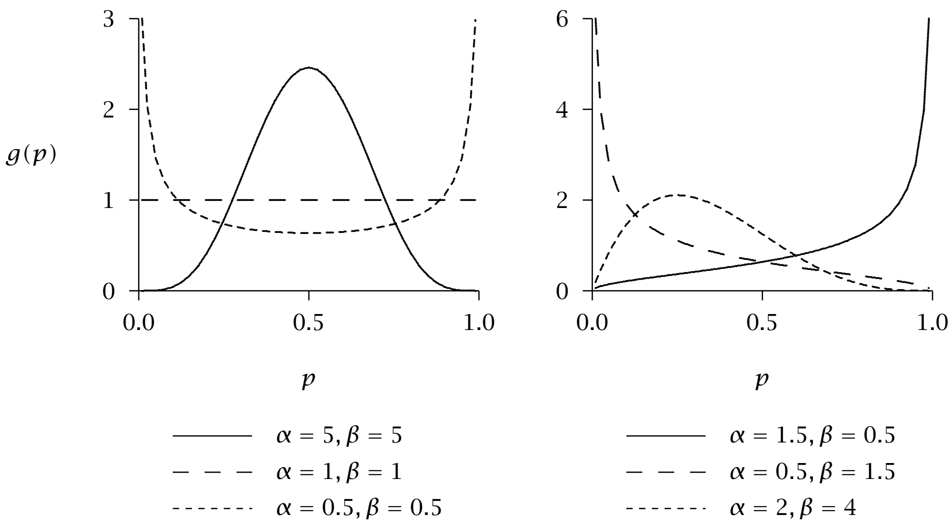
$$g(p \mid \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \; 0 < p < 1 .$$

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

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

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

Illustrative Beta Density Functions



The Negative Binomial Distribution (NBD)

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

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

The Exponential-Gamma Model (Pareto Distribution of the Second Kind)

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

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

The Beta-Geometric Model

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

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

The Beta-Binomial Distribution

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

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

Integrated Models

- Counting + Timing
  - catalog purchases (purchasing | "alive" & "death" process)
  - "stickiness" (# visits & duration/visit)
- Counting + Counting
  - purchase volume (# transactions & units/transaction)
  - page views/month (# visits & pages/visit)
- Counting + Choice
  - brand purchasing (category purchasing & brand choice)
  - "conversion" behavior (# visits & buy/not-buy)

A Template for Integrated Models

		Stage 2		
		Counting	Timing	Choice
	Counting			
Stage 1	Timing			
	Choice			

Integrated Models

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

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

- Solving the integral

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

often results in an expression that contains the Gaussian hypergeometric function, ${}_2F_1(a, b; c; z)$.

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

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

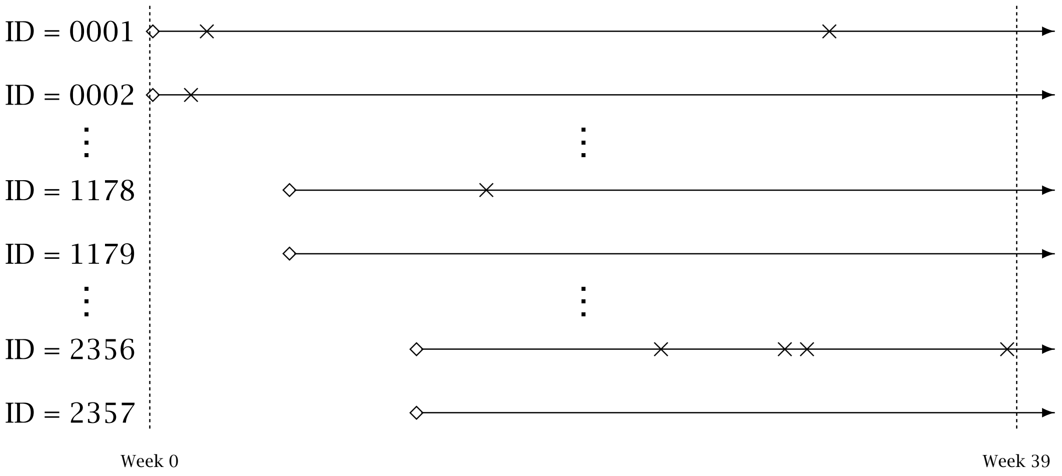
where $u_0 = 1$.

Part 2
Models for Noncontractual Settings

Setting

- New customers at CDNOW, 1/97–3/97
- Systematic sample (1/10) drawn from panel of 23,570 new customers
- 39-week calibration period
- 39-week forecasting (holdout) period

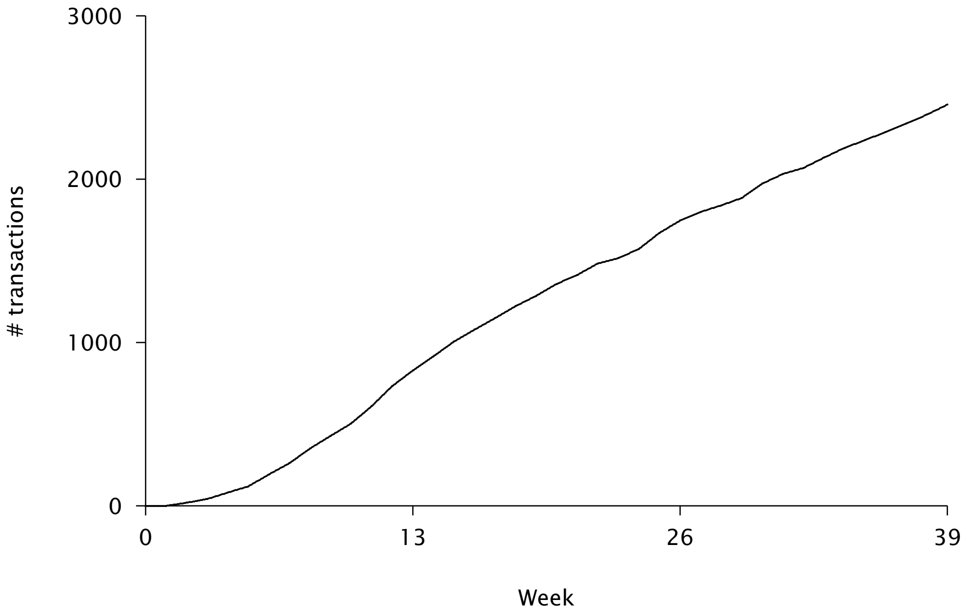
Purchase Histories



Raw Data

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

Cumulative Repeat Transactions



Modelling Objective

Given this customer database, we wish to determine the level of transactions that should be expected in next period (e.g., 39 weeks) by those on the customer list, both individually and collectively.

Modelling the Transaction Stream

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

Modelling the Transaction Stream

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

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

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

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

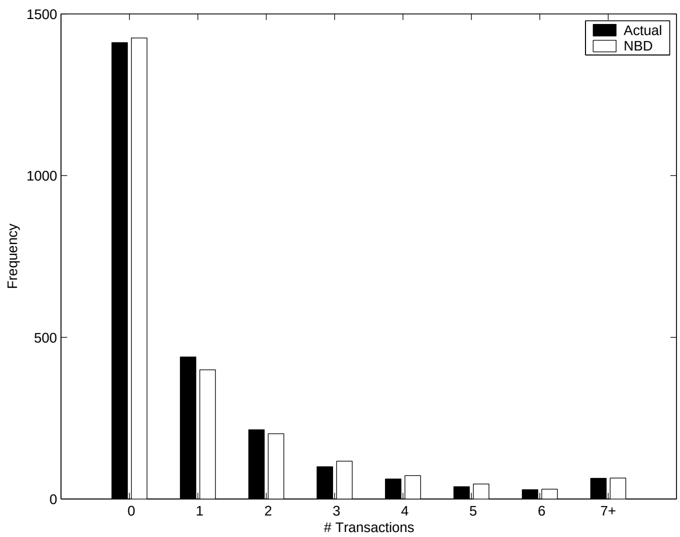
Modelling the Transaction Stream

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

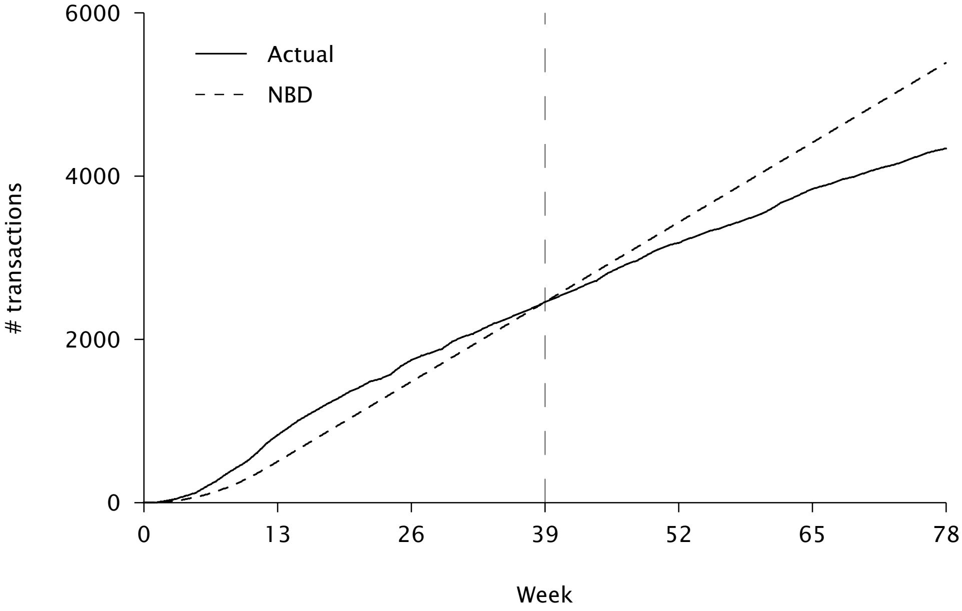
$$P(X(t) = x) = \int_0^\infty P(X(t) = x|\lambda)\, g(\lambda)\, d\lambda$$
$$= \frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x,$$

which is the negative binomial distribution (NBD).

Frequency of Repeat Transactions

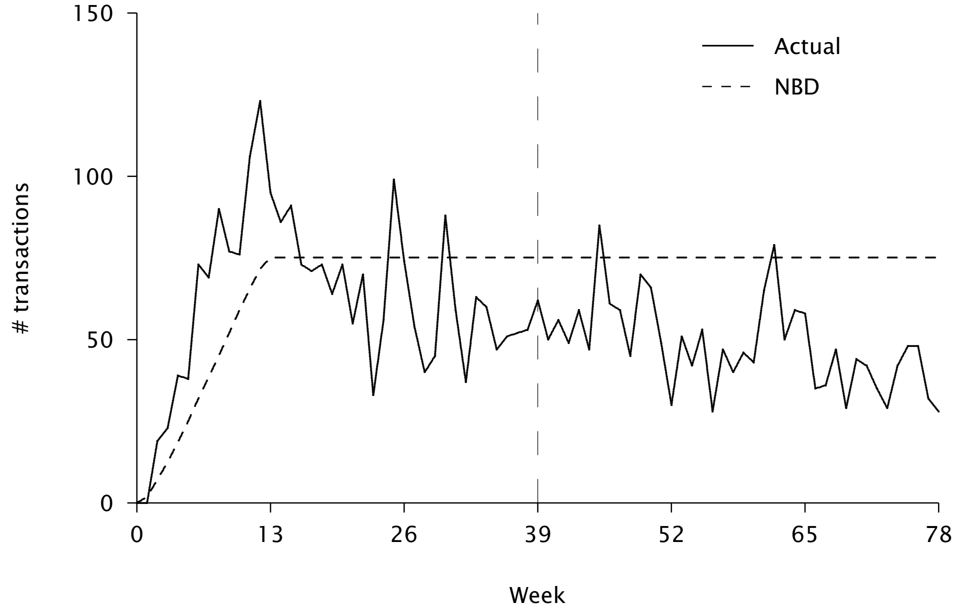


Tracking Cumulative Repeat Transactions



39

Tracking Weekly Repeat Transactions

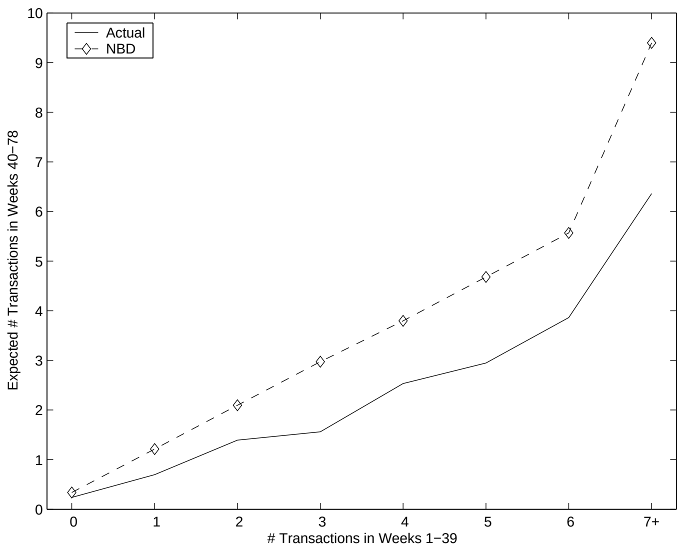


Conditional Expectations

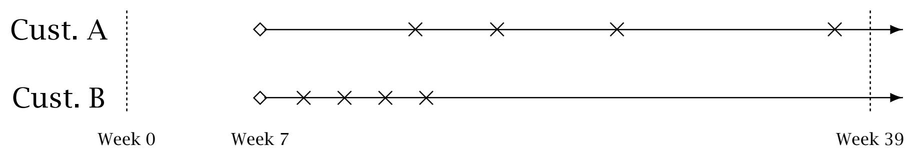
- We are interested in computing $E(Y(t)|\text{data})$, the expected number of transactions in an adjacent period $(T, T+t]$, conditional on the observed purchase history.
- For the NBD, a straight-forward application of Bayes' theorem gives us

$$E\big(Y(t)|X(T)=x\big) = \left(\frac{r+x}{\alpha+T}\right)t$$

Conditional Expectations



Conditional Expectations

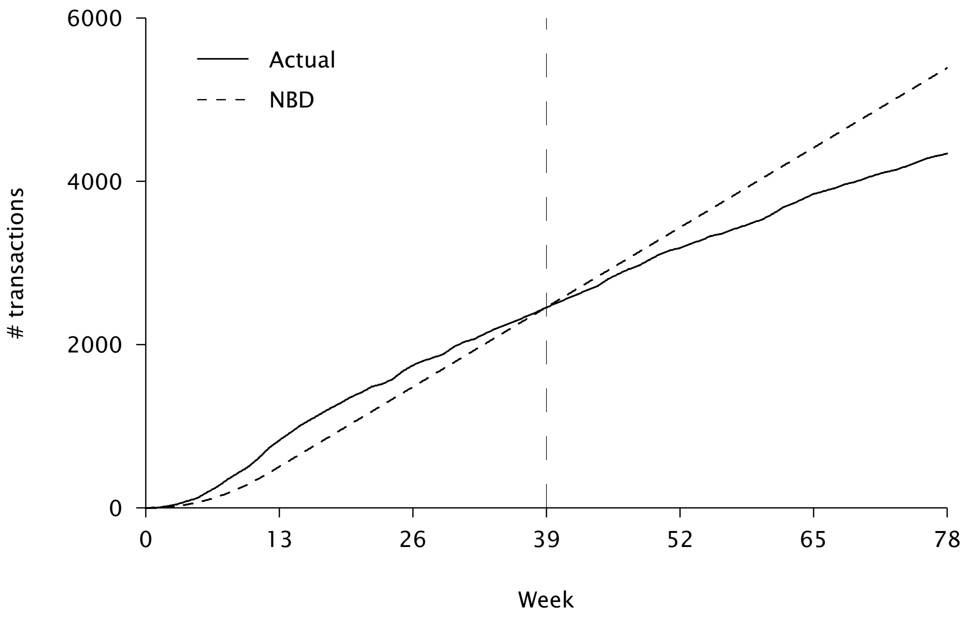


According to the NBD model:

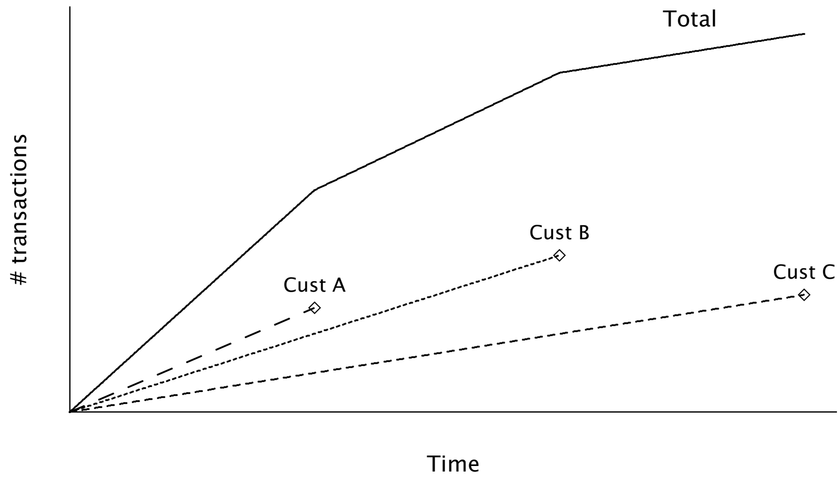
Cust. A: $E[Y(39)|X(32) = 4] = 3.88$

Cust. B: $E[Y(39)|X(32) = 4] = ?$

Tracking Cumulative Repeat Transactions



Towards a More Realistic Model



Modelling the Transaction Stream

Transaction Process:

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

Dropout Process:

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

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

Transaction Process:

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

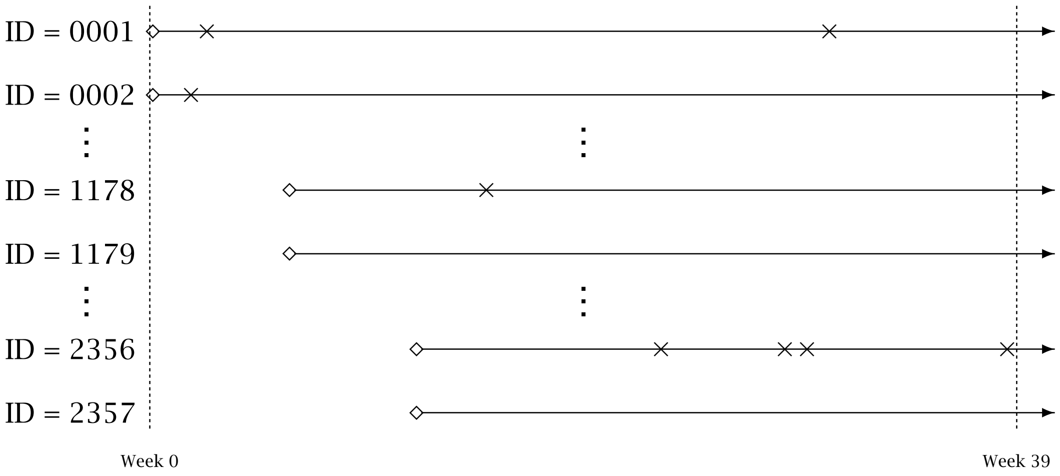
Dropout Process:

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

Summarizing Purchase Histories

- Given the model assumptions, we do not require information on when each of the x transactions occurred.
- The only customer-level information required by this model is *recency* and *frequency*.
- The notation used to represent this information is $(X = x, t_x, T)$, where x is the number of transactions observed in the time interval $(0, T]$ and t_x $(0 < t_x \leq T)$ is the time of the last transaction.

Purchase Histories



Raw Data

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

Pareto/NBD Likelihood Function

For a randomly-chosen individual with purchase history $(X = x, t_x, T)$,

$$L(r, \alpha, s, \beta \mid X = x, t_x, T) = \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)} \left\{ \frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + \left(\frac{s}{r+s+x}\right) \mathsf{A}_0 \right\}$$

where

$$\alpha \geq \beta: \quad \mathsf{A}_0 = \frac{1}{(\alpha+t_x)^{r+s+x}}\, {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+t_x}\big) - \frac{1}{(\alpha+T)^{r+s+x}}\, {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big)$$

$$\alpha \leq \beta: \quad \mathsf{A}_0 = \frac{1}{(\beta+t_x)^{r+s+x}}\, {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+t_x}\big) - \frac{1}{(\beta+T)^{r+s+x}}\, {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big).$$

Key Results

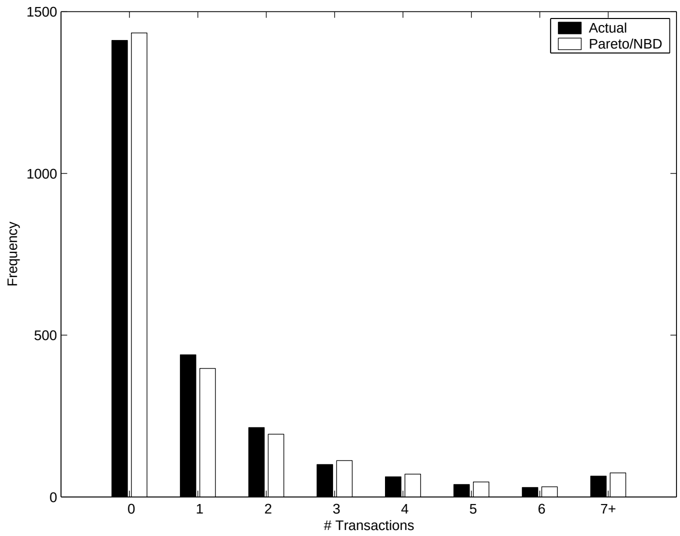
$P(\text{“active”} \mid X = x, t_x, T)$

The probability that an individual with observed behavior $(X = x, t_x, T)$ is still “active” at time T.

$E(Y(t) \mid X = x, t_x, T)$

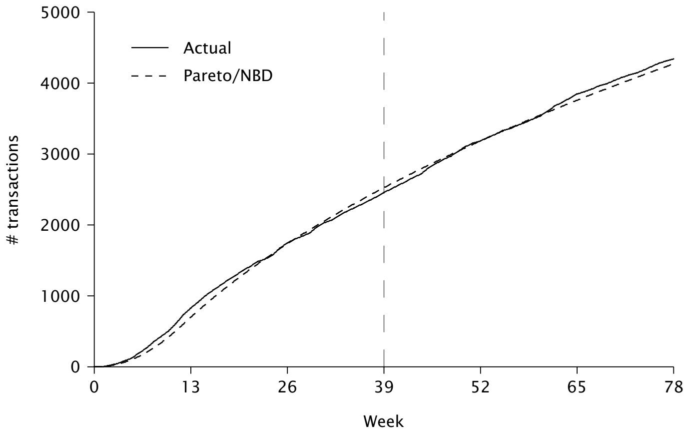
The expected number of transactions in the future period $(T, T+t]$ for an individual with observed behavior $(X = x, t_x, T)$.

Frequency of Repeat Transactions

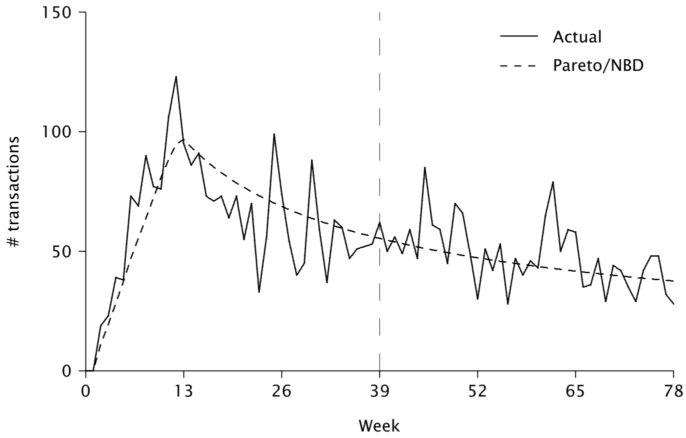


53

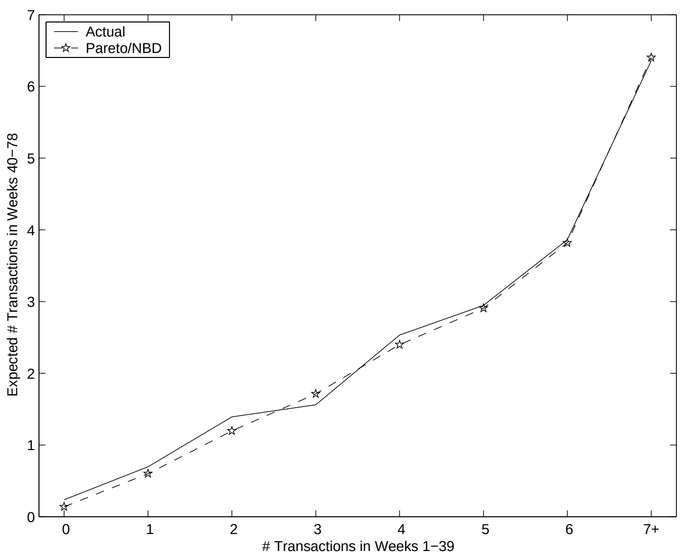
Tracking Cumulative Repeat Transactions



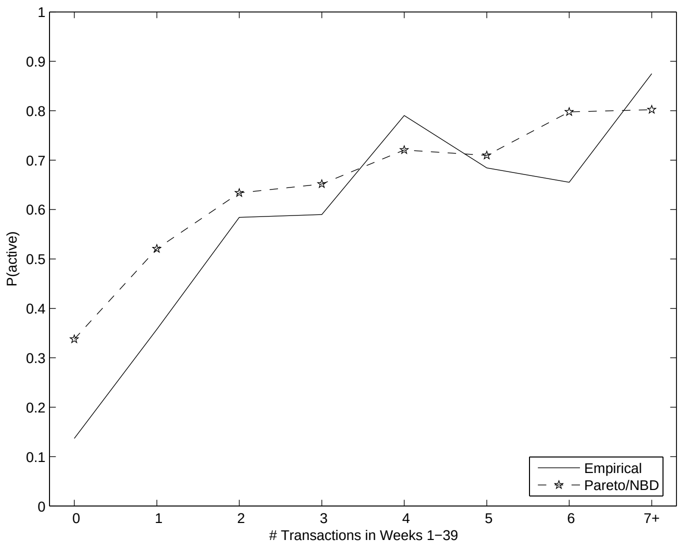
Tracking Weekly Repeat Transactions



Conditional Expectations



Proportions of Active Customers



Computing CLV

CLV is the present value of the future cashflows associated with the customer

The general formula for computing CLV is

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

where $v(t)$ = customer's net cashflow at t

$S(t)$ = probability the customer is alive at t

$d(t)$ = discount factor at t

Computing CLV

If we assume that an individual's spend per transaction is constant,

$$v(t) = \text{net cashflow} / \text{transaction} \times t(t)$$

where $t(t)$ is the transaction rate at t.

$$\Longrightarrow E(CLV) = E(\text{net cashflow} / \text{transaction}) \times \underbrace{\int_0^\infty t(t) S(t) d(t) dt}_{DET}.$$

DET is the present value of the expected number of future transactions (discounted expected transactions).

Computing DET

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

$$DET(\delta \mid \lambda, \mu) = \int_0^\infty \lambda e^{-\mu t} e^{-\delta t} dt = \frac{\lambda}{\mu + \delta}$$

  ... but λ and μ are unobserved.

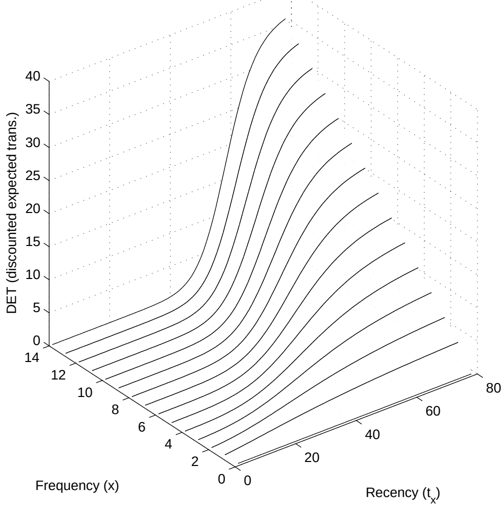
- Standing at time T, we wish to compute the present value of the expected future transaction stream for a customer with purchase history $(X = x, t_x, T)$, with continuous compounding at rate of interest δ.

Computing DET

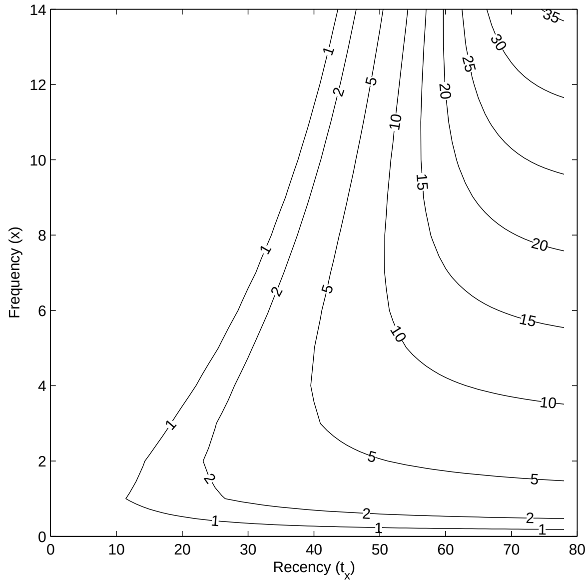
$$
\begin{aligned}
&DET(\delta \mid r, \alpha, s, \beta, X = x, t_x, T) \\
&\quad = \int_0^\infty \int_0^\infty \Big\{ DET(\delta \mid \lambda, \mu) \\
&\qquad\qquad \times P(\text{“active” at } T \mid \lambda, \mu, X = x, t_x, T) \\
&\qquad\qquad \times g(\lambda, \mu \mid r, \alpha, s, \beta, X = x,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

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

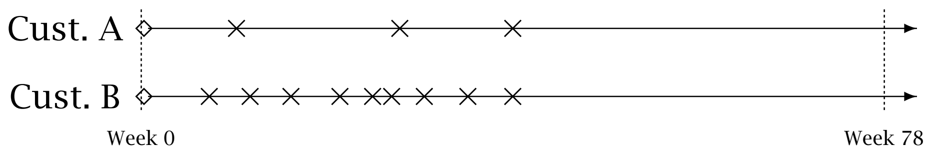
DET as a Function of Recency and Frequency



Iso-Value Representation of DET



The "Increasing Frequency" Paradox



	DET
Cust. A	4.6
Cust. B	1.9

Key Contribution

- We are able to generate forward-looking estimates of DET as a function of recency and frequency in a noncontractual setting:

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

- Adding a sub-model for spend per transaction enables us to generate forward-looking estimates of CLV as a function of RFM in a noncontractual setting:

$$CLV = f(R, F, M) = DET \times g(F, M)$$

An Alternative to the Pareto/NBD Model

- Estimation of model parameters can be a barrier to Pareto/NBD model implementation
- Recall the dropout process story:
  - Each customer has an unobserved "lifetime"
  - Dropout rates vary across customers
- Let us consider an alternative story:
  - After any transaction, a customer tosses a coin
    heads → become inactive
    tails → remain active
  - P(heads) varies across customers

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

Purchase Process:

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

Dropout Process:

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

BG/NBD Likelihood Function

For a randomly-chosen individual with purchase history ($X = x, t_x, T$),

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

where

$$\mathsf{A}_1 = \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)}$$

$$\mathsf{A}_2 = \frac{\Gamma(a+b)\Gamma(b+x)}{\Gamma(b)\Gamma(a+b+x)}$$

$$\mathsf{A}_3 = \left(\frac{1}{\alpha+T}\right)^{r+x}$$

$$\mathsf{A}_4 = \left(\frac{a}{b+x-1}\right)\left(\frac{1}{\alpha+t_x}\right)^{r+x}$$

BGNBD Estimation

	A	B	C	D	E	F	G	H	I
1	r	0.243							
2	alpha	4.414							
3	a	0.793							
4	b	2.426							
5	LL	-9582.4							
6									
7	ID	x	t_x	T	ln(.)	ln(A_1)	ln(A_2)	ln(A_3)	ln(A_4)
8	0001	2	30.43	38.86	-9.4596	-0.8390	-0.4910	-8.4489	-9.4265
9	0002	1	1.71	38.86	-4.4711	-1.0562	-0.2828	-4.6814	-3.3709
10	[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]	0.0000
13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]596
14	0007	1	5.00	38.86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]043
15	0008	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
16	0009	2	35.71	38.86	-9.5367	-0.8390	-0.4910	-8.4489	-9.7432
17	0010	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
2362	2355	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000
2363	2356	4	26.57	27.00	-14.1284	1.1450	-0.7922	-14.6252	-16.4902
2364	2357	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000

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

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

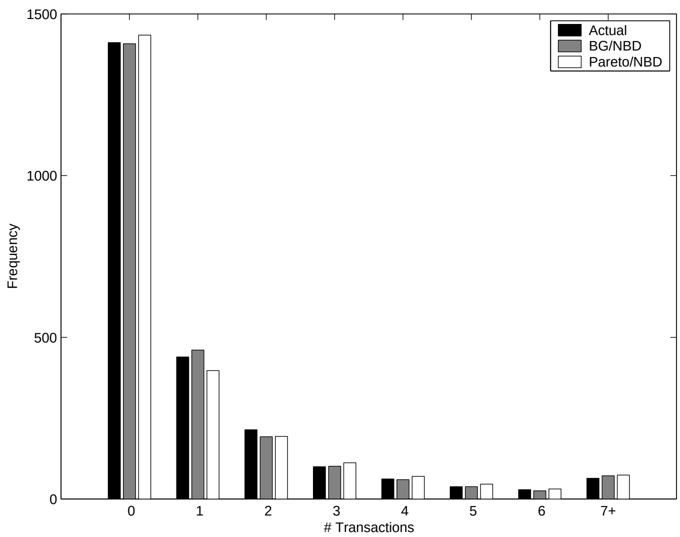
=-(B$1+B8)*LN(B$2+D8)

=SUM(E8:E2364)

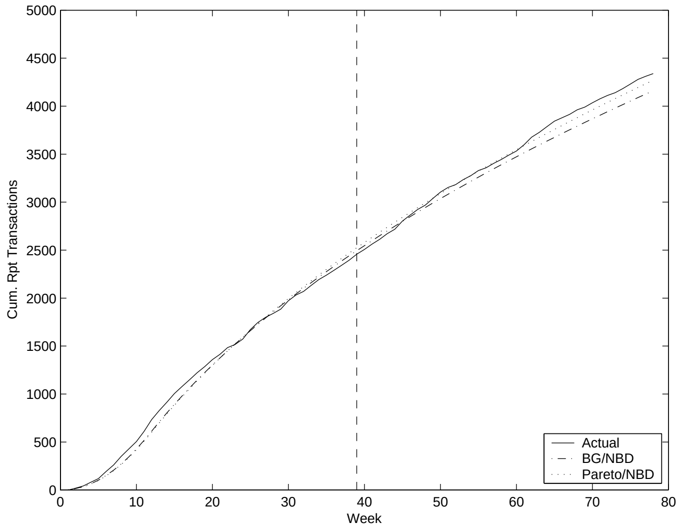
=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))

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

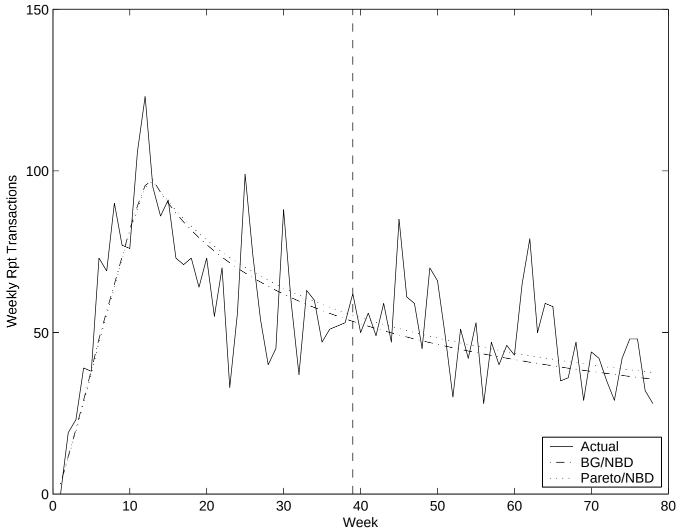
Frequency of Repeat Transactions



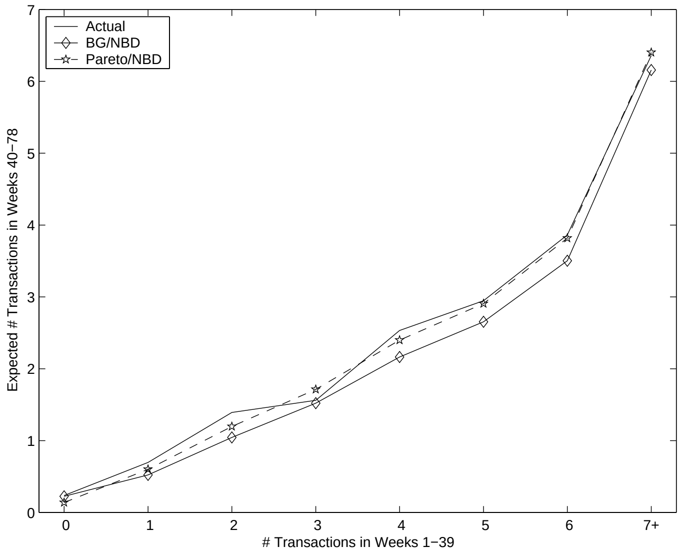
Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions



Conditional Expectations



Further Reading

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

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

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

Further Reading

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

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

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

Fader, Peter S. 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/`>

Modelling the Transaction Stream

How valid is the assumption of Poisson purchasing?

→ can transactions occur at any point in time?

“Discrete-Time” Transaction Opportunities

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

“Discrete-Time” Transaction Data

A *transaction opportunity* is

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

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

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

Repeat Purchasing for Luxury Cruises

(Berger, Weinberg, and Hanna 2003)

1993	1994	1995	1996	1997	1994	1995	1996	1997	# Customers
Y	Y	Y	Y	Y	1	1	1	1	18
				N	1	1	1	0	34
			N	Y	1	1	0	1	36
				N	1	1	0	0	64
		N	Y	Y	1	0	1	1	14
				N	1	0	1	0	62
			N	Y	1	0	0	1	18
				N	1	0	0	0	302
	N	Y	Y	Y	0	1	1	1	16
				N	0	1	1	0	118
			N	Y	0	1	0	1	36
				N	0	1	0	0	342
		N	Y	Y	0	0	1	1	44
				N	0	0	1	0	292
			N	Y	0	0	0	1	216
				N	0	0	0	0	4482

Research Objectives

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

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

Model Development

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

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

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

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

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

Model Development

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

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

Alive?

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

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

Model Development

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

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

Model Development

- Bernoulii purchasing while alive ⇒ the order of a given number of transactions (prior to the last observed transaction) doesn't matter
- For example, $f(10100 \mid p, \theta) = f(01100 \mid p, \theta)$
- *Recency* (time of last transaction, m) and *frequency* (number of transactions, $x = \sum_{t=1}^{n} y_t$) are sufficient summary statistics

  $\Longrightarrow$ we do not need the complete binary string representation of a customer's transaction history

Repeat Purchasing for Luxury Cruises

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

→

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

Model Development

For a customer with purchase history $(X = x, m, n)$,

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

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,

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

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

BGBB Estimation

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

BGBB Estimation

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

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

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

=SUM(E9:E19)

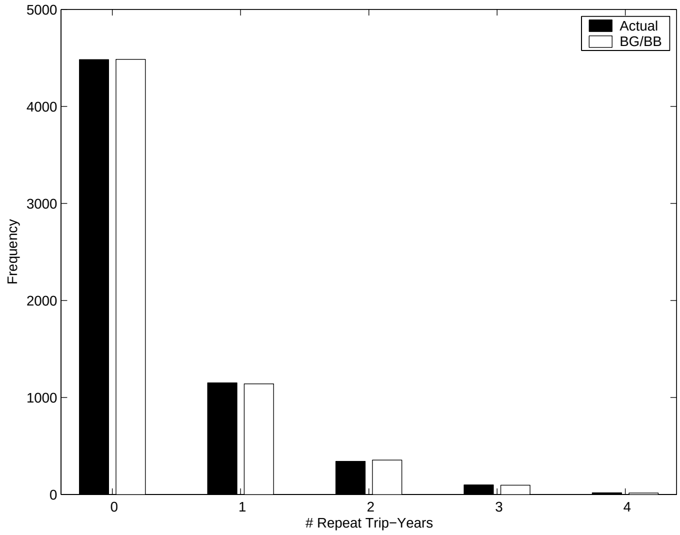
=IF($H9<J$8,0,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+J$8)-GAMMALN($B$1+$B$2+$B9+J$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+J$8)-GAMMALN($B$3+$B$4+$B9+J$8+1))/$E$3)

=C15-B15-1

=D19*LN(F19))

=SUM(I19:M19)

Model Fit



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

Computing P(alive in period $n+1$)

- According to Bayes' theorem,

$$P(\text{alive in } n \mid \text{data}) = \frac{P(\text{data} \mid \text{alive in } n) P(\text{alive in } n)}{P(\text{data})}$$

- Recalling the individual-level likelihood function,

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

it follows that

$$P(\text{alive in period } n \mid X=x,m,n,p,\theta) = p^x(1-p)^{n-x}(1-\theta)^n \Big/ L(p,\theta \mid X=x,m,n)$$

Computing P(alive in period $n+1$)

For a customer with purchase history $(X=x,m,n)$,

$$\begin{aligned} &P(\text{alive in period } n+1 \mid X=x,m,n,\alpha,\beta,\gamma,\delta) \\ &= \int_0^1\int_0^1 \Big\{P(\text{alive in period } n \mid X=x,m,n,p,\theta)(1-\theta) \\ &\qquad \times g(p,\theta \mid X=x,m,n,\alpha,\beta,\gamma,\delta)\Big\}\, dp\, d\theta \\ &= \frac{B(\alpha+x,\beta+n-x)B(\gamma,\delta+n+1)}{B(\alpha,\beta)B(\gamma,\delta)} \Big/ L(\alpha,\beta,\gamma,\delta \mid x,n,m) \end{aligned}$$

P(alive)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	alpha	0.66		B(alpha,beta)	0.4751										
2	beta	5.19													
3	gamma	173.76		B(gamma,delta)	4E-260										
4	delta	1882.93													
5															
6	LL	-7130.7													
7													i		
8	x	m	n	# cust.		L(.|X=x,m,n)		P(alive in 1998)		n-m-1		0	1	2	3
9	4	4	4	18	-106.7	0.0027		0.92		-1	0.0027	0	0	0	0
10	3	4	4	66	-368.0	0.0038		0.92		-1	0.0038	0	0	0	0
11	2	4	4	98	-463.5	0.0088		0.92		-1	0.0088	0	0	0	0
12	1	4	4	216	-704.4	0.0384		0.92		-1	0.0384	0	0	0	0
13	3	3	4	34	-184.6	0.0044		0.79		0	0.0038	0.0006	0	0	0
14	2	3	4	180	-829.0	0.0100		0.81		0	0.0088	0.0012	0	0	0
15	1	3	4	292	-920.8	0.0427		0.82		0	0.0384	0.0043	0	0	0
16	2	2	4	64	-283.5	0.0119		0.68		1	0.0088	0.0019	0.0012	0	0
17	1	2	4	342	-1033.4	0.0487		0.72		1	0.0384	0.0060	0.0043	0	0
18	1	1	4	302	-863.0	0.0574		0.61		2	0.0384	0.0087	0.0060	0.0043	0
19	0	0	4	4482	-1373.9	0.7360		0.60		3	0.4785	0.0845	0.0686	0.0568	0.0476

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

Computing CLV

CLV is the present value of the future cashflows associated with the customer

The general formula for computing CLV is

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

where $v(t)$ = customer's net cashflow at t

$S(t)$ = probability the customer is alive at t

$d(t)$ = discount factor at t

Computing CLV

If we assume that an individual's spend per transaction is constant,

$$v(t) = \text{net cashflow / transaction} \times t(t)$$

where $t(t)$ is the transaction rate at t.

$$\Longrightarrow E(CLV) = E(\text{net cashflow / transaction}) \times \underbrace{\int_0^\infty t(t)S(t)d(t)dt}_{DET}.$$

DET is the present value of the expected number of future transactions (discounted expected transactions).

Computing DET

- For a customer with purchase history $(X = x, m, n)$,

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

- However, p and θ are unobserved.

Computing DET

For a just-acquired customer ($x = m = n = 0$),

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

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

Computing DET

For a customer with purchase history $(X = x, m, n)$, we have to integrate over the *posterior* distribution of p and θ:

$$\begin{aligned}
&DET(d \mid X = x, m, n, \alpha, \beta, \gamma, \delta) \\
&\quad = \int_0^1 \int_0^1 \frac{p(1-\theta)}{d+\theta} g(p, \theta \mid X = x, m, n, \alpha, \beta, \gamma, \delta)\, dp\, d\theta \\
&\quad = \frac{B(\alpha+x+1, \beta+n-x) B(\gamma, \delta+n+1)}{B(\alpha, \beta) B(\gamma, \delta)(1+d)} \\
&\qquad\qquad \times \frac{{}_2F_1(1, \delta+n+1; \gamma+\delta+n+1; \frac{1}{1+d})}{L(\alpha, \beta, \gamma, \delta \mid X = x, m, n)}
\end{aligned}$$

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 $u_0 = 1$.

DET

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

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

`=SUM(M24:M174)`

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

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

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

Posterior Mean of p as a Function of Recency and Frequency

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

Expected # Remaining Transactions (DET for $d = 0$)

	Year of Last Cruise				
# Cruise-years	1997	1996	1995	1994	1993
4	5.16				
3	4.05	3.50			
2	2.95	2.60	2.18		
1	1.84	1.65	1.45	1.23	
0					0.47

Discounted Expected Transactions (DET) at a 10% Annual Discount Rate ($d = 0.10$)

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

- See <`http://brucehardie.com/papers/020/`> for a copy of the paper that develops the BG/BB model.
- See <`http://brucehardie.com/notes/010/`> for a note on how to implement the BG/BB model in Excel, along with a copy of the associated spreadsheet.

Part 3
Models for Contractual Settings

Contractual Settings

- Examples of contractual settings:

  gym membership, cable TV, airport lounge, cellular phone, theatre subscription, utilities, credit card …

- In contrast to noncontractual markets, we *do* know when a customer ends their relationship with the firm.
- Focus on retention (= 1 − churn).

Typical Contractual Setting

An analysis of our customer records yields the following retention rates:

Years since acquisition	1	2	3	4	5
Retention rate	0.633	0.689	0.747	0.798	0.836

The retention rate for period t (r_t) is defined as the proportion of customers active at the end of period $t-1$ who are still active at the end of period t.

Motivating Questions

- How much should I spend on customer acquisition?
- How much is my existing customer base worth?

  What is the expected (remaining) tenure of a customer who has been with us for n years?

Computing CLV

Recall the general formula for computing CLV:

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

Assuming an individual's value at t is independent of their tenure, we can factor it out of the calculation:

$$E(CLV) = E(\text{net cashflow rate}) \times \underbrace{\int_0^\infty S(t)d(t)dt}_{DEL}\,.$$

DEL is the present value of the expected lifetime of the customer (discounted expected lifetime).

Computing DEL

Switching to a discrete-time formulation, we have

$$DEL_0 = \sum_{t=0}^{\infty} \frac{S(t)}{(1+d)^t} \text{ or } DEL_1 = \sum_{t=1}^{\infty} \frac{S(t)}{(1+d)^t}$$

- DEL_0 feeds into the calculation of CLV for an *as-yet-to-be-acquired* customer (excluding cost of acquisition).
- DEL_1 feeds into the calculation of CLV for a *just-acquired customer*; we ignore the first purchase that identifies them as a new customer.
- Clearly $DEL_0 = DEL_1 + 1$; we will focus on DEL_1.

Computing DEL

Given the relationship between retention rates and the survivor function,

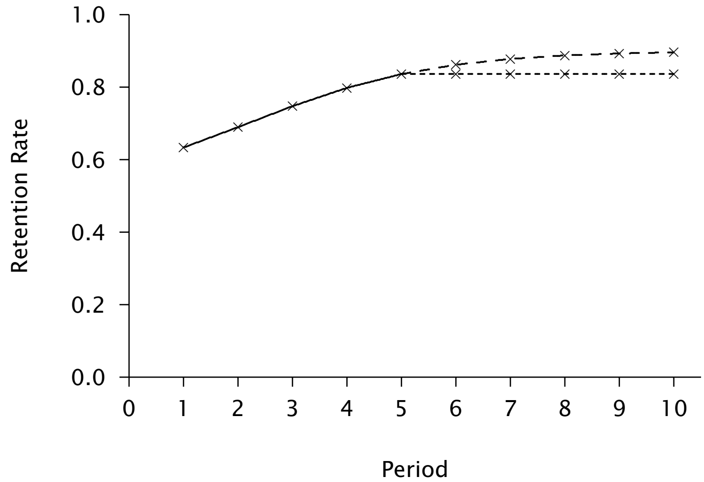
$$r_t = \frac{S(t)}{S(t-1)} \Longleftrightarrow S(t) = \prod_{i=1}^{t} r_i\,,$$

it follows that

$$DEL(d) = \sum_{t=1}^{\infty} \left\{ \prod_{i=1}^{t} r_i \right\} \left(\tfrac{1}{1+d}\right)^t$$

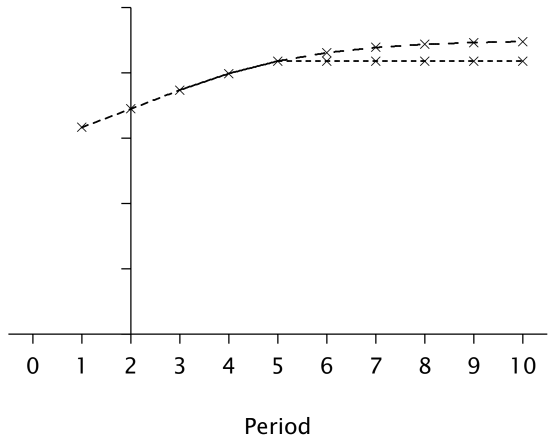
Implementation Questions

How do we project r_t beyond the set of observed retention rates?



Implementation Questions

How do we compute DEL for customers with an observed tenure of 2 periods, $DEL(d \mid n = 2)$?



$$r_1' = r_3 = 0.747,\ r_2' = r_4 = 0.798, \text{ etc.}$$

$DEL(d = 0.1 \mid$ survived two years$)$

What is DEL for a customer with a two-year tenure (assuming a 10% discount rate)?

Using the “shifted” observed retention rates ($r'_1 = 0.747$, $r'_2 = 0.798$, $r'_{3+} = 0.836$) gives us

$$DEL(d = 0.1) = \sum_{t=1}^{\infty} \left\{ \prod_{i=1}^{t} r'_i \right\} \left(\frac{1}{1+d} \right)^t$$
$$= 2.73$$

	A	B	C	D	E
1	DEL	Year	r_t	S(t)	1/(1+d)^t
2	2.73	1	0.747	0.747	0.909
3		2	0.798	0.596	0.826
4		3	0.836	0.49[illegible]	[illegible]
5		4	0.836	0.41[illegible]	[illegible]83
6		5	0.836	0.348	0.621
7		6	0.83[illegible]	[illegible]	0.564
8		7	0.836	0.243	0.513
9		[illegible]	[illegible]	[illegible]04	0.467
10		[illegible]	[illegible]	[illegible]170	0.424
100		99	0.836	1.7E-08	8.0E-05
101		100	0.836	1.4E-08	7.3E-05

=D2*C3

=1/(1.1)^B7

=SUMPRODUCT(D2:D101,E2:E101)

Why Do Retention Rates Increase Over Time?

Individual-level time dynamics (e.g., increasing loyalty as the customer gains more experience with the firm).

vs.

An artifact of cross-sectional heterogeneity.

"The Ruse of Heterogeneity"

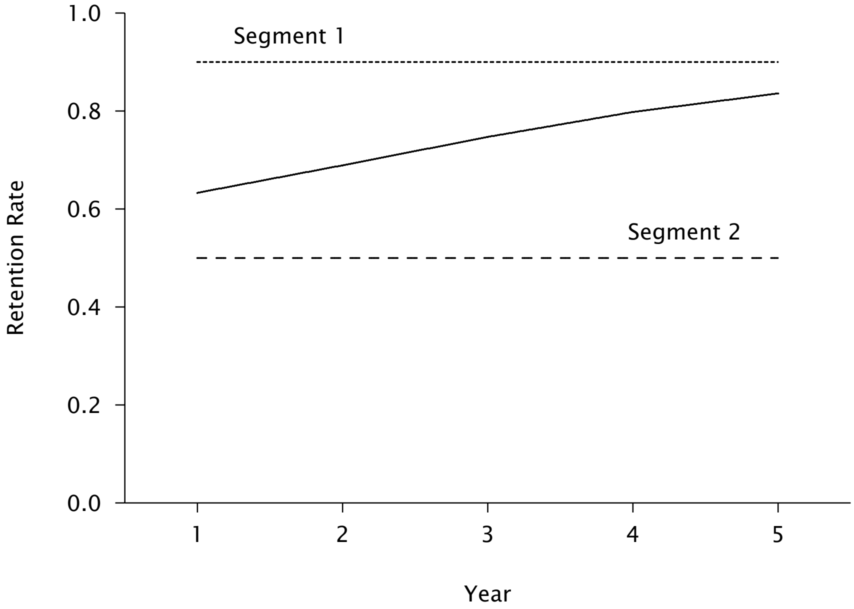
Suppose we track a cohort of 150,000 customers, comprising two unobserved segments:

- Segment 1 comprises 50,000 customers, each with a time-invariant annual retention probability of 0.9.
- Segment 2 comprises 100,000 customers, each with a time-invariant annual retention probability of 0.5.

The Ruse of Heterogeneity

	# Active Customers			r_t		
Year	Seg 1	Seg 2	Total	Seg 1	Seg 2	Total
0	50,000	100,000	150,000			
1	45,000	50,000	95,000	0.900	0.500	0.633
2	40,500	25,000	65,500	0.900	0.500	0.689
3	36,450	12,500	48,950	0.900	0.500	0.747
4	32,805	6,250	39,055	0.900	0.500	0.798
5	29,525	3,125	32,650	0.900	0.500	0.836

The Ruse of Heterogeneity



$DEL(d = 0.1 \mid$ survived two years)

- If this person belongs to segment 1:

$$\begin{aligned} DEL(d = 0.1) &= \sum_{t=1}^{\infty} \left\{ \prod_{i=1}^{t} 0.9 \right\} \left(\tfrac{1}{1+0.1}\right)^t \\ &= 4.50 \end{aligned}$$

- If this person belongs to segment 2:

$$\begin{aligned} DEL(d = 0.1) &= \sum_{t=1}^{\infty} \left\{ \prod_{i=1}^{t} 0.5 \right\} \left(\tfrac{1}{1+0.1}\right)^t \\ &= 0.833 \end{aligned}$$

- But to which segment does this person belong?

$DEL(d = 0.1 \mid$ survived two years)

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

$$\begin{aligned} &\frac{P(\text{survived two years} \mid \text{segment 1}) \times P(\text{segment 1})}{P(\text{survived two years})} \\ &= \frac{0.9^2 \times 0.333}{0.9^2 \times 0.333 + 0.5^2 \times 0.667} \\ &= 0.618 \end{aligned}$$

$DEL(d = 0.1 \mid \text{survived two years})$

It follows that the discounted expected lifetime for an individual with a tenure of two years is

$$\begin{aligned}
&DEL(d = 0.1 \mid \text{survived two years}) \\
&\quad = DEL(d = 0.1 \mid \text{seg. } 1)P(\text{seg. } 1 \mid \text{survived two yrs}) \\
&\qquad + DEL(d = 0.1 \mid \text{seg. } 2)P(\text{seg. } 2 \mid \text{survived two yrs}) \\
&\quad = 4.5 \times 0.618 + 0.833 \times (1 - 0.618) \\
&\quad = 3.10 \quad (\text{cf. } 2.73 \text{ for naïve approach})
\end{aligned}$$

Implications

To the extent that the observed retention dynamics may be largely driven by heterogeneity, they are not indicative of true individual-level behavior.

⟶ any estimates of CLV for existing customers must be based off a correct story of individual-level buyer behavior that explicitly accounts for heterogeneity.

A Discrete-Time Model for Contract Duration

1. 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 (shifted) geometric distribution:

$$S(t \mid \theta) = (1-\theta)^t\,, \quad t = 1, 2, 3, \ldots$$

2. Heterogeneity in θ is captured by a beta distribution with pdf

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

A Discrete-Time Model for Contract Duration

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

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

- The aggregate survivor function is

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

A Discrete-Time Model for Contract Duration

- The (aggregate) retention rate is given by

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

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

Application

- Consider a cohort of 10,000 customers from which 4583 leave by the end of their first year, and 1752, 916, 560, 376, 269 in years 2–6 respectively.

- Fitting the sBG model to these data yields parameter estimates of $\hat{\alpha} = 1.10$ and $\hat{\beta} = 1.30$.

	A	B	C	D	E
1	alpha	1.10		B(alpha,beta)	0.689
2	beta	1.30			
3					
4	LL	-15521.8			
5					
6	Year	# Cancelled	P(cancel)		
7	1	4583	0.458	-3575.7	
8	2	1752	0.175	-3051.7	
9	3	916	0.092	-2189.7	
10	4	560	0.056	-1614.4	
11	5	376	0.038	-1233.5	
12	6	269	0.027	-972.2	
13			0.154	-2884.6	

Simplified Parameter Estimation

For a single cohort with observed retention rates r_1 and r_2, the maximum likelihood estimators of α and β are:

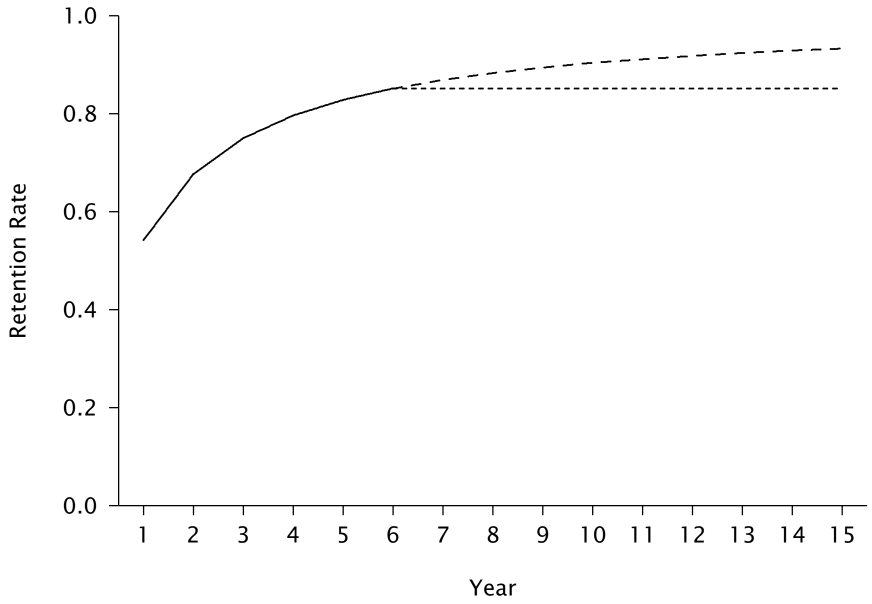
$$\hat{\alpha} = \frac{(1 - r_1)(1 - r_2)}{(r_2 - r_1)}$$

$$\hat{\beta} = \frac{r_1(1 - r_2)}{(r_2 - r_1)}$$

Retention Rates

Year	# Inactive	# Active	Retention Rate Empirical	Retention Rate Model
0		10000		
1	4583	5417	0.542	0.542
2	1752	3665	0.677	0.677
3	916	2749	0.750	0.750
4	560	2189	0.796	0.796
5	376	1813	0.828	0.828
6	269	1544	0.852	0.851
7				0.869
8				0.883
9				0.894
10				0.904
11				0.911
12				0.918
13				0.924
14				0.929
15				0.933

Retention Rates



Computing DEL

- For a just-acquired customer,

$$
\begin{aligned}
DEL(d \mid \theta) &= \sum_{t=1}^{\infty} \frac{S(t \mid \theta)}{(1+d)^t} = \sum_{t=1}^{\infty} \left(\frac{1-\theta}{1+d} \right)^t \\
&= \frac{1-\theta}{d+\theta}
\end{aligned}
$$

- Removing the conditioning on θ,

$$
\begin{aligned}
DEL(d \mid \alpha, \beta) &= \int_0^1 \frac{1-\theta}{d+\theta} f(\theta \mid \alpha, \beta)\, d\theta \\
&= \frac{\beta}{(\alpha+\beta)(1+d)} \\
&\qquad \times {}_2F_1\big(1, \beta+1; \alpha+\beta+1; \tfrac{1}{1+d}\big)\,.
\end{aligned}
$$

Computing DEL

- Model-based DEL:

$$DEL(d = 0.1 \mid \alpha = 1.1, \beta = 1.3) = 1.77$$

- Empirical DEL (where, for $i > 6$, $r_i = r_6$):

$$
\begin{aligned}
DEL(d = 0.1) &= \sum_{t=1}^{\infty} \Big\{ \prod_{i=1}^{t} r_i \Big\} \big(\tfrac{1}{1+0.1}\big)^t \\
&= 1.65
\end{aligned}
$$

- Under-estimates true DEL by 7%.

DEL Given Tenure of *n* Periods

What is DEL for a customer with a two-year tenure (assuming a 10% discount rate)?

Using the "shifted" observed retention rates ($r'_1 = r_3 = 0.750$, $r'_2 = r_4 = 0.796$, ..., $r'_{4+} = r_{6+} = 0.852$) gives us

$$\begin{aligned} DEL(d = 0.1) &= \sum_{t=1}^{\infty} \left\{ \prod_{i=1}^{t} r'_i \right\} \left(\tfrac{1}{1+0.1}\right)^t \\ &= 2.82 \end{aligned}$$

DEL Given Tenure of *n* Periods

- For a customer with a tenure of n periods,

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

- But θ is unobserved ....

DEL Given Tenure of n Periods

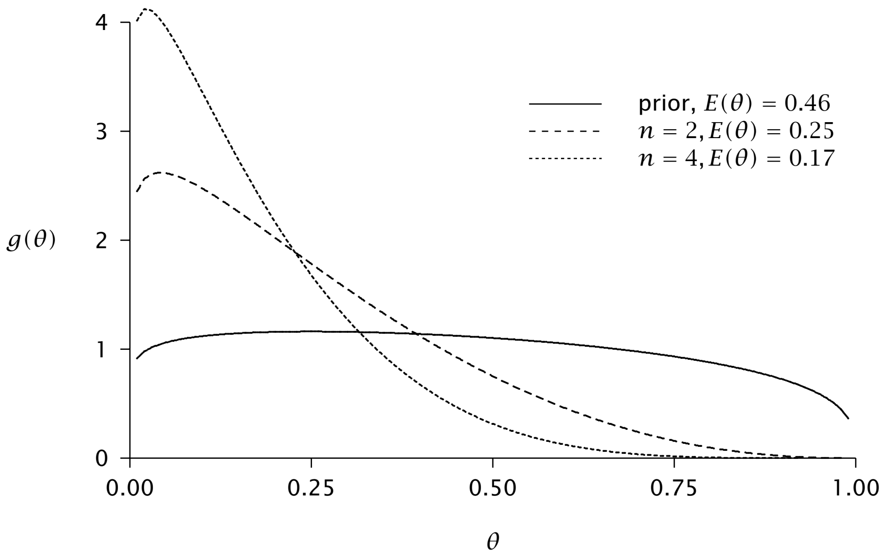
- By Bayes' theorem, the posterior distribution of θ is

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

- It follows that

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

Posterior Distribution of θ



$\hat{\alpha} = 1.1, \hat{\beta} = 1.3$

DEL Given Tenure of n Periods

Tenure	Model	Naïve	% Error
0	1.77	1.65	7%
1	2.59	2.35	9%
2	3.21	2.82	12%
3	3.70	3.14	15%
4	4.11	3.33	19%
5	4.46	3.43	23%
6	4.76	3.43	28%

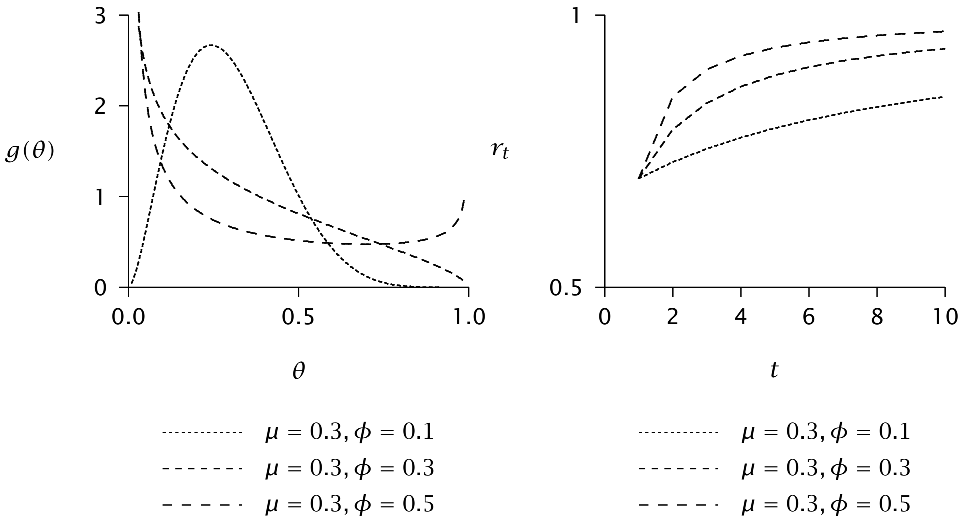
Impact of Heterogeneity on DEL Error

Consider three scenarios:

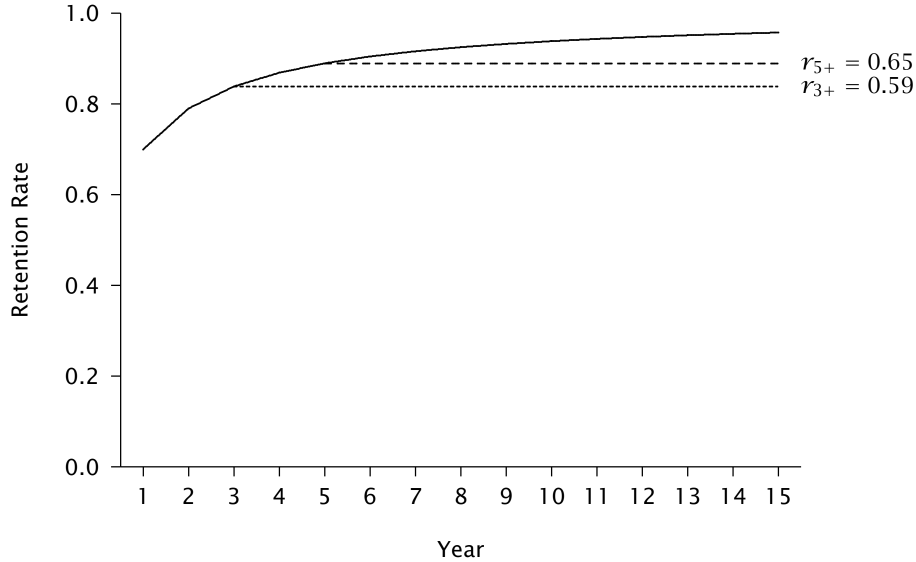
Scenario	μ	ϕ	α	β
1	0.3	0.1	2.7	6.3
2	0.3	0.3	0.7	1.6
3	0.3	0.5	0.3	0.7

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \phi = \frac{1}{\alpha + \beta + 1}$$

Impact of Heterogeneity on DEL Error



Varying the Retention Rate Censoring Point ($\mu = 0.3, \phi = 0.3$)



Impact of Heterogeneity on DEL Error

DEL for a just-acquired customer, varying the retention rate censoring point ($d = 0.1$):

				Censored at	
Scenario	μ	ϕ	Model	$t = 3$	$t = 5$
1	0.3	0.1	2.17	1.98	2.08
				9%	4%
2	0.3	0.3	3.24	2.59	2.88
				21%	11%
3	0.3	0.5	4.36	3.34	3.86
				23%	12%

143

Impact of Heterogeneity on DEL Error

DEL given tenure of n periods, varying the empirical retention rate censoring point ($d = 0.1$):

				$n = 0$			$n = 2$	
Scenario	μ	ϕ	Model	$t = 3$	$t = 5$	Model	$t = 3$	$t = 5$
1	0.3	0.1	2.17	1.98	2.08	2.64	2.18	2.41
				9%	4%		17%	9%
2	0.3	0.3	3.24	2.56	2.88	4.70	3.21	3.91
				21%	11%		32%	17%
3	0.3	0.5	4.36	3.34	3.86	6.57	4.50	5.55
				23%	12%		31%	16%

Valuing an Existing Customer Base

As we move from a single cohort of customers (defined by year-of-acquisition),

Year	0	1	2	3	4	5	6
# Active	10,000	5417	3665	2749	2189	1813	1544

to a customer base composed of a number of different cohorts, we must condition any calculation on time-of-entry into the firm's customer base.

Valuing an Existing Customer Base

- Number of active customers by cohort:

2001	2002	2003	2004
5000	2708	1832	1374
	10000	5417	3665
		15000	8125
			16000
5000	12708	22249	29163

- Fitting the sBG model, $\hat{\alpha} = 1.1, \hat{\beta} = 1.3$
- Agg. 03–04 retention rate = 13163/ 22249 = 0.592

Valuing an Existing Customer Base

		$DEL(d = 0.1)$	
Cohort	Active in 2004	Model	Naïve
2001	1374	3.70	2.14
2002	3664	3.20	2.14
2003	8125	2.59	1.93
2004	16000	1.77	1.44
	Total DEL	66080	49602

"Textbook" approach:

$$\begin{aligned}\text{Total DEL} &= 29163 \times \frac{0.592}{0.1 + (1 - 0.592)} \\ &= 33938\end{aligned}$$

Modelling Contract Duration in Continuous Time

- The duration of an individual's relationship with the firm is characterized by the exponential distribution:

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

- Heterogeneity in λ follows a gamma distribution with shape and scale parameters r and α, respectively.

$\longrightarrow$ the EG model: $S(t \mid r, \alpha) = \left(\dfrac{\alpha}{\alpha + t}\right)^{r}$

Modelling Contract Duration in Continuous Time

With continuous compounding at rate of interest δ, we compute DEL for a customer with tenure s as:

$$\begin{aligned} DEL(\delta \mid \text{survival to } s, r, \alpha) &= \int_s^\infty S(t \mid t > s, r, \alpha) e^{-\delta(t-s)} dt \\ &= \int_s^\infty \left(\frac{\alpha + s}{\alpha + t}\right)^r e^{-\delta(t-s)} dt \\ &= (\alpha + s)^r \delta^{r-1} \Psi(r, r; (\alpha + s)\delta) \end{aligned}$$

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

Incorporating Individual-Level Dynamics

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

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

- DEL for a customer with tenure s is computed by solving

$$\int_s^\infty \left(\frac{\alpha + s^c}{\alpha + t^c}\right)^r e^{-\delta(t-s)} dt$$

  using standard numerical integration techniques.

See <http://brucehardie.com/papers/021/> for a paper (and supporting spreadsheet) that explores these issues associated with calculating CLV in a contractual setting.

Part 4
Conclusions

Customer-Base Analysis

We have proposed a set of models that enable us to answer questions such as

- which customers are most likely to be active in the future,
- the level of transactions we could expect in future periods from those on the customer list, both individually and collectively, and
- individual customer lifetime value (CLV)

when faced with a customer transaction database.

Extensions

- Bring in submodel for value of each transaction

  $\Longrightarrow$ DET $\longrightarrow$ CLV

- Introduce covariates (customer descriptors and marketing activities).

  need to be wary of endogeneity bias and sample selection effects

Classifying Customer Bases

(Adapted from Schmittlein et al. 1987)

Opportunities for Transactions	Unobserved	Observed
Continuous	CPG purchases Visits to doctor Music downloading	Brokerage account “Warehouse club” Health club usage
Discrete	Church attendance Prescription refills (Charitable giving)	Cable TV Subscriptions Health club m’ship

Time At Which Customers Become Inactive

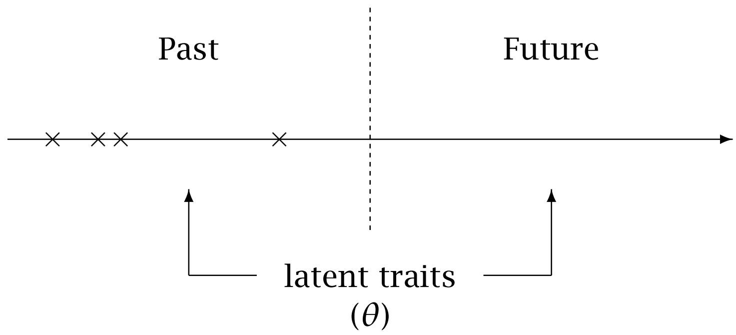
Philosophy of Model Building

- Keep it as simple as possible
- Minimize cost of implementation
  - Use of readily available software (e.g., Excel)
  - Use of data summaries
- Purposively ignore the effects of covariates (customer descriptors and marketing activities) so as to highlight the important underlying components of buyer behavior.

Central Tenet

Traditional approach

future = f(past)



Probability modelling approach

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