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

ESCP-EAP European School of Management (Paris campus)
March 7–8, 2007

Day 1
Building Blocks

Problem 1:
Projecting Customer Retention Rates

(Modelling Discrete-Time Duration Data)

Background

One of the most important problems facing marketing managers today is the issue of *customer retention*. It is vitally important for firms to be able to anticipate the number of customers who will remain active for $1, 2, \ldots, T$ periods (e.g., years or months) after they are first acquired by the firm.

The following dataset is taken from a popular book on data mining (Berry and Linoff, *Data Mining Techniques*, Wiley 2004). It documents the "survival" pattern over a seven-year period for a sample of customer who were all "acquired" in the same period.

Customers Surviving At Least 0–7 Years

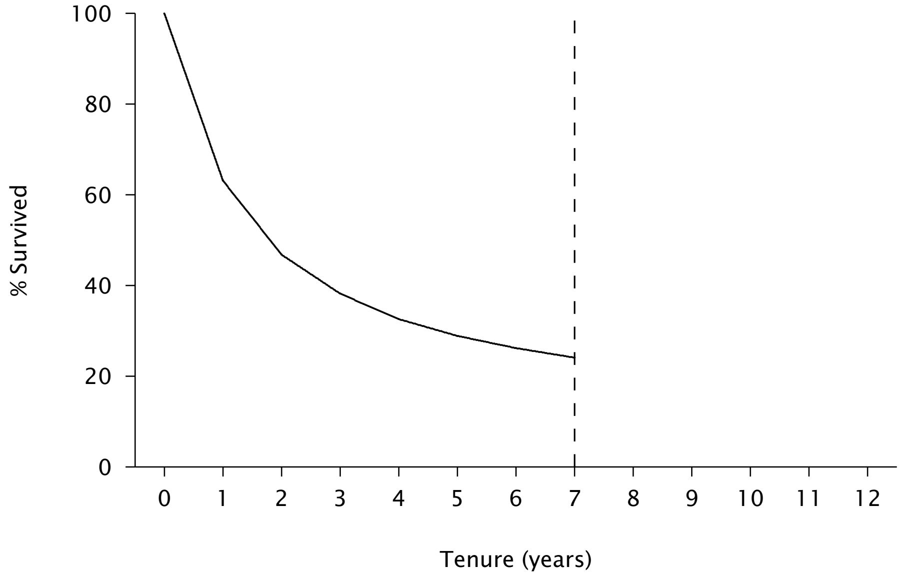
Year	# Customers	% Alive
0	1000	100%
1	631	63%
2	468	47%
3	382	38%
4	326	33%
5	289	29%
6	262	26%
7	241	24%

Of the 1000 initial customers, 631 renew their contracts at the end of the first year. At the end of the second year, 468 of these 631 customers renew their contracts.

Modelling Objective

Develop a model that enables us to project the survival curve (and therefore retention rates) over the next five years (i.e., out to $T = 12$).

Modelling Objective (I)

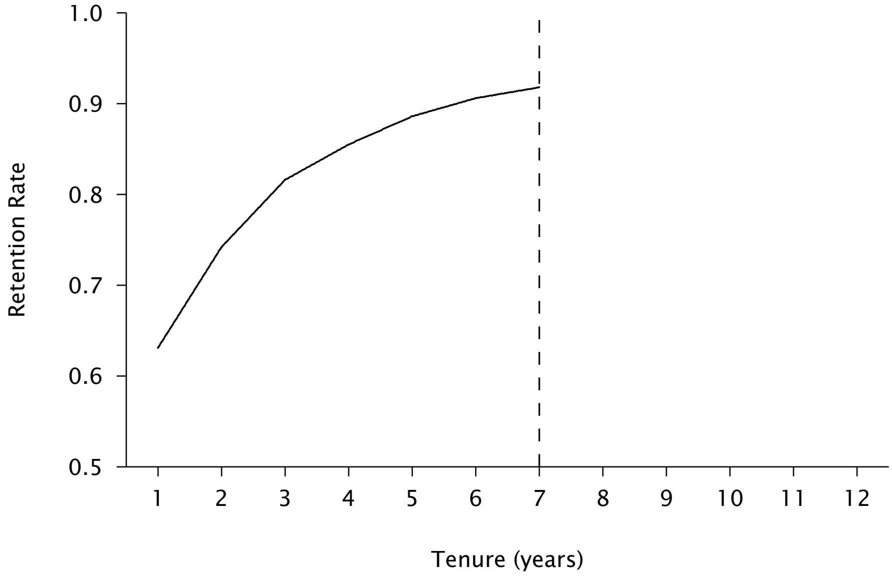


Implied Retention Rates

The retention rate for period t (r_t) is defined as the proportion of customers who had renewed their contract at the end of period $t-1$ who then renew their contract at the end of period t:

$$r_t = \frac{P(T > t)}{P(T > t - 1)}$$

Modelling Objective (II)



Natural Starting Point

Project survival using simple functions of time:

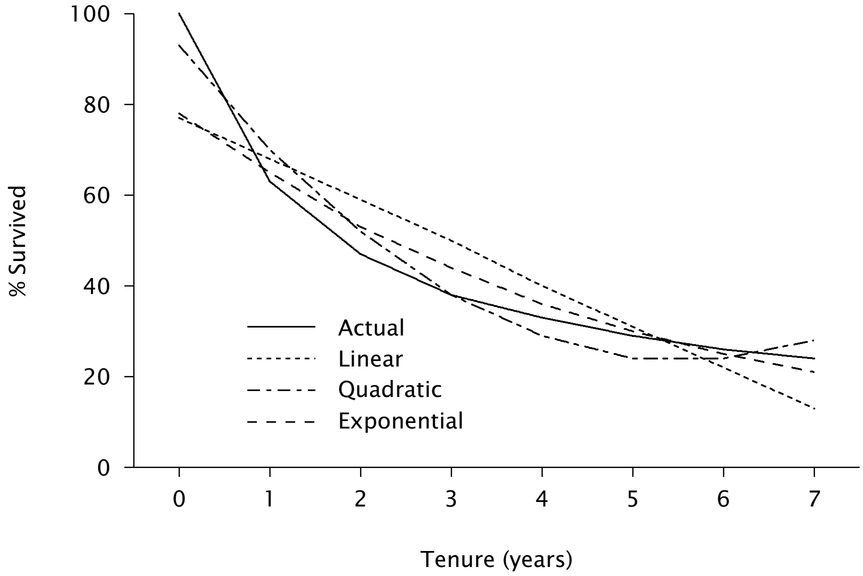
- Consider linear, quadratic, and exponential functions
- Let y = the proportion of customers surviving at least t years

$$y = 0.773 - 0.092t \qquad R^2 = 0.777$$

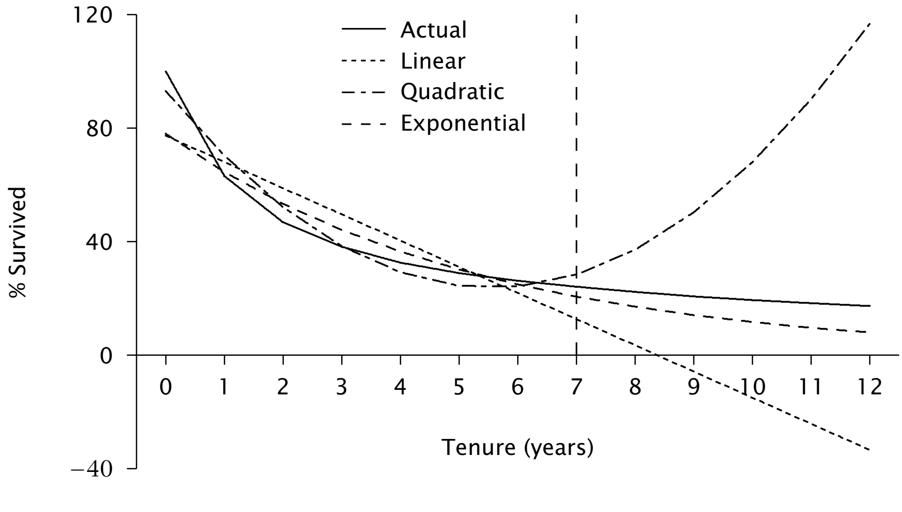
$$y = 0.930 - 0.249t + 0.022t^2 \qquad R^2 = 0.960$$

$$\ln(y) = -0.248 - 0.190t \qquad R^2 = 0.915$$

Model Fit



Survival Curve Projections



Developing a Better Model (I)

Consider the following story of customer behavior:

i. At the end of each period, an individual renews his contract with (constant and unobserved) probability $1 - \theta$.

ii. All customers have the same "churn probability" θ.

Developing a Better Model (I)

More formally:

- Let the random variable T denote the duration of the customer's relationship with the firm.
- We assume that the random variable T has a (shifted) geometric distribution with parameter θ:

$$P(T = t \mid \theta) = \theta(1 - \theta)^{t-1}, \quad t = 1, 2, 3, \ldots$$

$$P(T > t \mid \theta) = (1 - \theta)^{t}, \quad t = 1, 2, 3, \ldots$$

Developing a Better Model (I)

The probability of the observed pattern of contract renewals is:

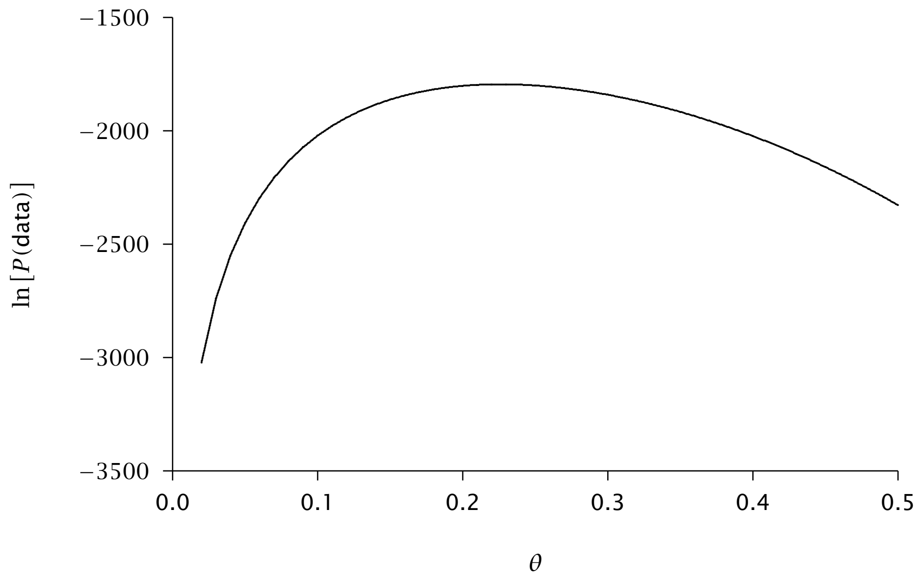
$$
\begin{aligned}
&[\theta]^{369}[\theta(1-\theta)^{1}]^{163}[\theta(1-\theta)^{2}]^{86} \\
&\quad \times [\theta(1-\theta)^{3}]^{56}[\theta(1-\theta)^{4}]^{37}[\theta(1-\theta)^{5}]^{27} \\
&\quad\quad \times [\theta(1-\theta)^{6}]^{21}[(1-\theta)^{7}]^{241}
\end{aligned}
$$

Estimating Model Parameters

- Let us assume that the observed data are the outcome of a process characterized the "coin-flipping" model of contract renewal.
- Which value of θ is more likely to have "generated" the data?

θ	$P(\text{data})$	$\ln[P(\text{data})]$
0.2	4.10×10^{-783}	-1801.5
0.5	1.31×10^{-1011}	-2327.6

Estimating Model Parameters



Estimating Model Parameters

We estimate the model parameters using the method of *maximum likelihood*:

- The likelihood function is defined as the probability of observing the sample data for a given set of the (unknown) model parameters
- This probability is computed using the model and is viewed as a function of the model parameters:

$$L(\text{parameters}|\text{data}) = p(\text{data}|\text{parameters})$$

- For a given dataset, the maximum likelihood estimates of the model parameters are those values that maximize $L(\cdot)$

Estimating Model Parameters

The log-likelihood function is defined as:

$$
\begin{aligned}
LL(\theta|\text{data}) = \; & 369 \times \ln[P(T=1)] + \\
& 163 \times \ln[P(T=2)] + \\
& \quad \ldots \qquad\qquad\quad + \\
& 21 \times \ln[P(T=7)] + \\
& 241 \times \ln[P(T>7)]
\end{aligned}
$$

The maximum value of the log-likelihood function is $LL = -1794.62$, which occurs at $\hat{\theta} = 0.226$.

Estimating Model Parameters

	A	B	C	D	E
1	theta	0.5000			
2	LL	-2327.59			
3					
4	Year	P(T=t)	# Cust.	# Lost	
5	0		1000		
6	1	0.5000	631	369	-255.77
7	2	0.2500	468	163	-225.97
8	3	0.1250	[illegible]	[illegible]	[illegible]8.83
9	4	0.0625	326	56	-155.26
10	5	0.0313	289	37	-128.23
11	6	0.0156	262	27	-112.29
12	7	0.0078	241	21	-101.89
13					-1169.3393
14					

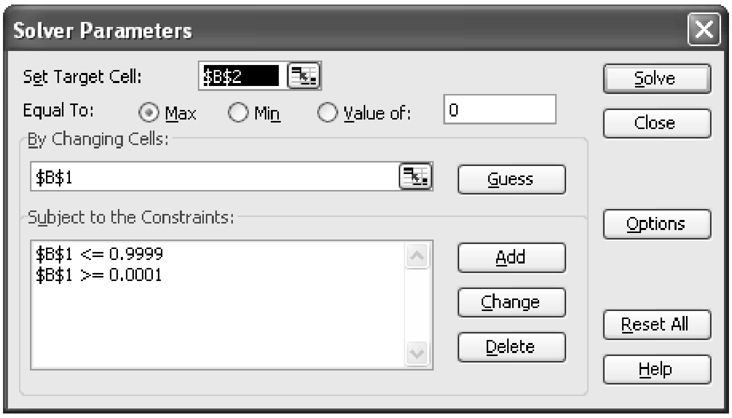
=SUM(E6:E13) → B2

=D6*LN(B6) → E6

=$B$1*(1-$B$1)^(A8-1) → B8

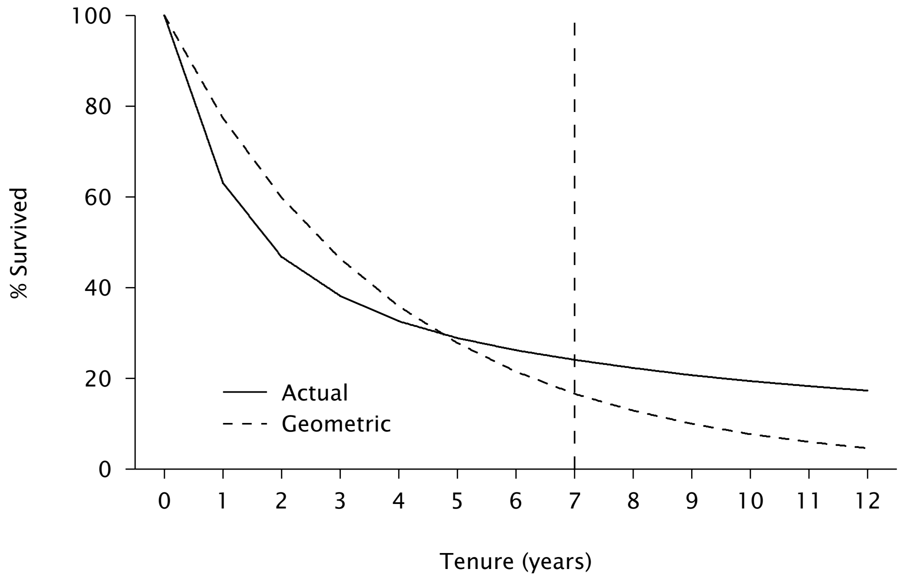
=C12*LN(1-SUM(B6:B12)) → E13

Estimating Model Parameters



21

Survival Curve Projection



What's wrong with this story of customer contract-renewal behavior?

Developing a Better Model (II)

Consider the following story of customer behavior:

i. At the end of each period, an individual renews his contract with (constant and unobserved) probability $1 - \theta$.

ii. "Churn probabilities" vary across customers.

Accounting for Heterogeneity (I)

Assume two segments of customers:

Segment	Size	Churn Prob.
1	π	θ_1
2	$1-\pi$	θ_2

Developing a Better Model (IIa)

- For a randomly-chosen individual,

$$\begin{aligned} P(T=t) &= P(T=t \mid \text{segment 1})P(\text{segment 1}) \\ &\quad + P(T=t \mid \text{segment 2})P(\text{segment 2}) \end{aligned}$$

- More formally,

$$\begin{aligned} &P(T=t \mid \theta_1, \theta_2, \pi) \\ &\qquad = \theta_1(1-\theta_1)^{t-1}\pi + \theta_2(1-\theta_2)^{t-1}(1-\pi) \end{aligned}$$

- We call this a “finite mixture” (of geometric distributions) model

Estimating Model Parameters

The log-likelihood function is defined as:

$$
\begin{aligned}
LL(\theta_1, \theta_2, \pi \,|\, \text{data}) = {} & 369 \times \ln[P(T = 1)] + \\
& 163 \times \ln[P(T = 2)] + \\
& \ldots \qquad\qquad + \\
& 21 \times \ln[P(T = 7)] + \\
& 241 \times \ln[P(T > 7)]
\end{aligned}
$$

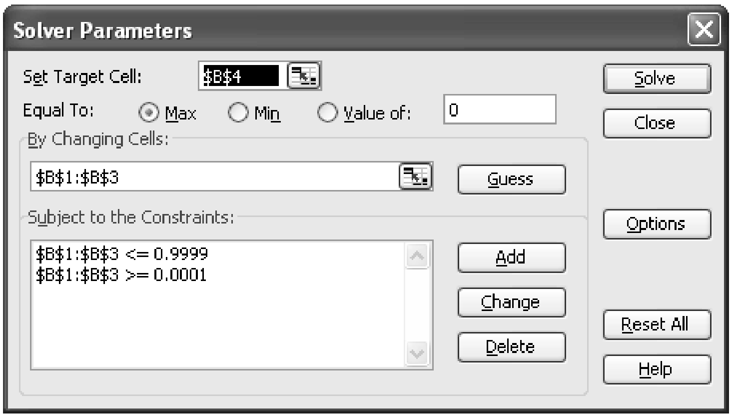
The maximum value of the log-likelihood function is $LL = -1680.05$, which occurs at $\hat{\theta}_1 = 0.083$, $\hat{\theta}_2 = 0.586$, and $\hat{\pi} = 0.439$.

Estimating Model Parameters

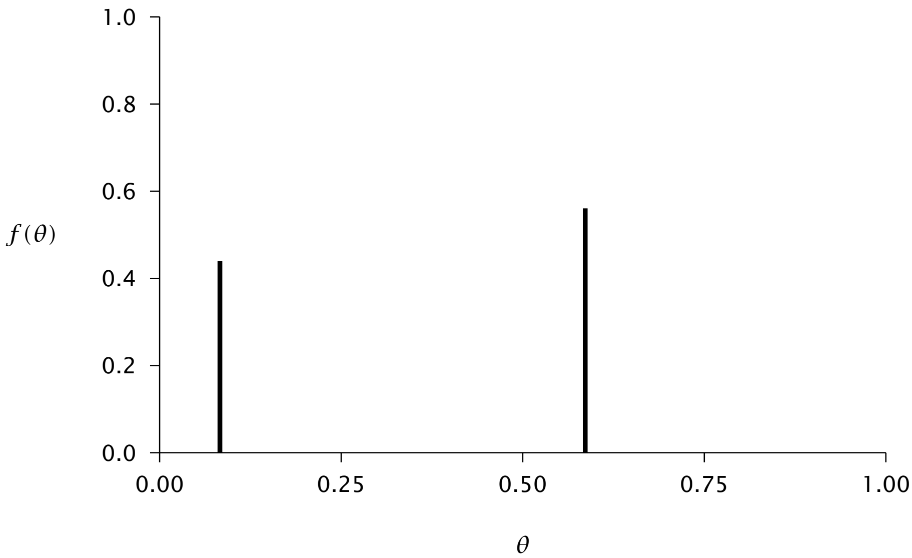
	A	B	C	D	E	F	G
1	theta_1	0.1000					
2	theta_2	0.5000					
3	pi	0.5000					
4	LL	-1694.35					
5							
6	Year	P(T=t|seg 1)	P(T=t|seg 2)	P(T=t)	# Cust.	# Lost	
7	0				1000		
8	1	0.1000	0.5000	0.3000	[illegible]	[illegible]	60
9	2	0.0900	[illegible]	[illegible]	468	163	-288.8290
10	[illegible]	[illegible]	0.1250	0.1030	382	86	-195.4803
11	4	0.0729	0.0625	0.0677	326	56	-150.7895
12	5	0.0656	0.0313	0.0484	289	37	-112.0225
13	6	0.0590	0.0156	0.0373	262	27	-88.7698
14	7	0.0531	0.0078	0.0305	241	21	-73.3055
15							-340.8870

- =$B$1*(1-$B$1)^(A8-1)
- =$B$2*(1-$B$2)^(A8-1)
- =B8*$B$3+C8*(1-$B$3)

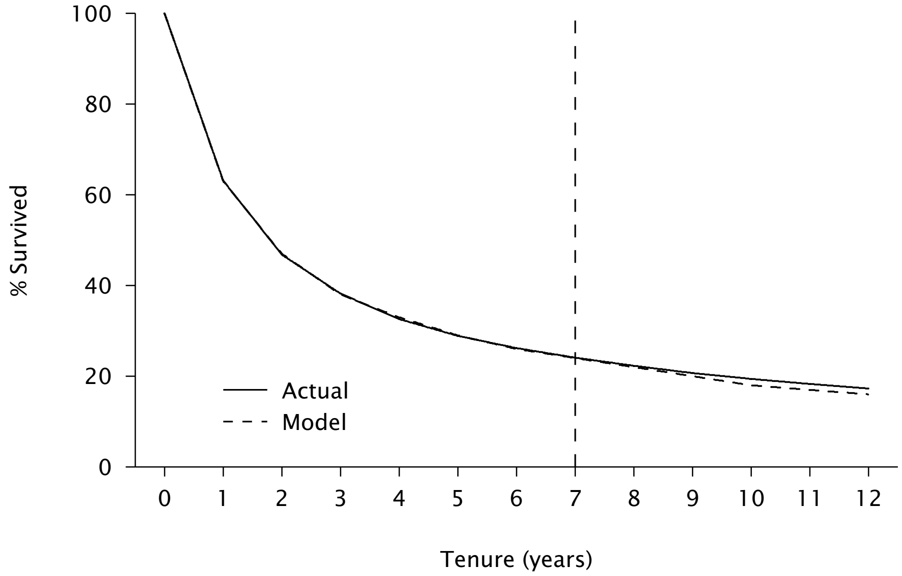
Estimating Model Parameters



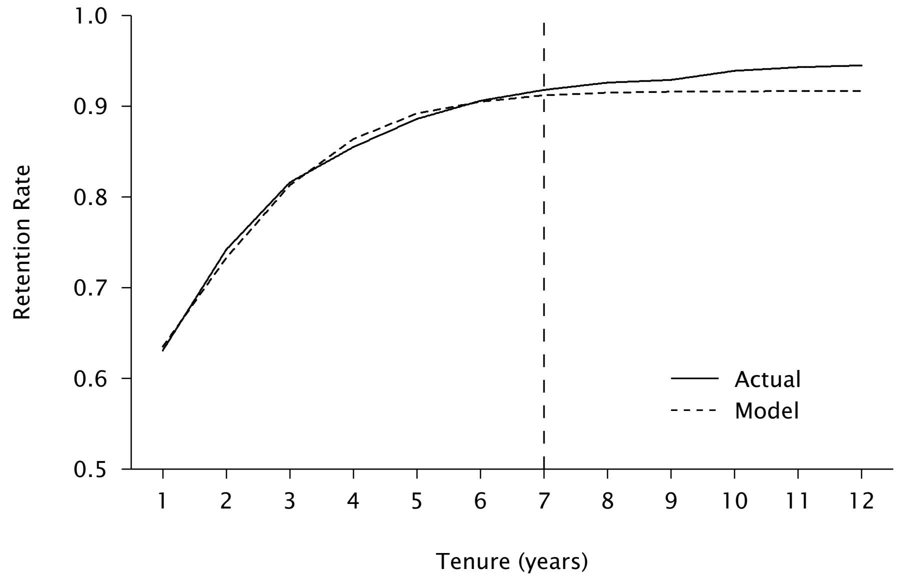
Estimated Distribution of Churn Probabilities



Survival Curve Projection



Projecting Retention Rates



Developing a Better Model (IIb)

Assume three segments of customers:

- For a randomly-chosen individual,

$$\begin{aligned} P(T = t) = {} & P(T = t \mid \text{segment } 1)P(\text{segment } 1) \\ & + P(T = t \mid \text{segment } 2)P(\text{segment } 2) \\ & + P(T = t \mid \text{segment } 3)P(\text{segment } 3) \end{aligned}$$

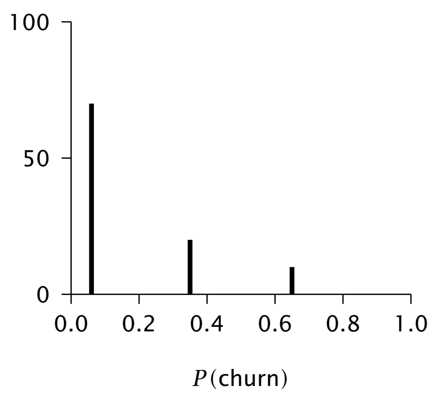
- More formally, we have a three-segment "finite mixture" model:

$$P(T = t \mid \theta_1, \theta_2, \theta_3, \pi_1, \pi_2) = \sum_{i=1}^{3} \theta_i (1 - \theta_i)^{t-1} \pi_i$$

  where $\pi_3 = 1 - (\pi_1 + \pi_2)$.

Vodafone Italia Churn Clusters

Cluster	$P(\text{churn})$	% CB
Low risk	0.06	70
Medium risk	0.35	20
High risk	0.65	10



Source: "Vodafone Achievement and Challenges in Italy" presentation (2003-09-12)

Accounting for Heterogeneity (II)

- We move from a finite number of segments to an infinite number of segments.
- Assume heterogeneity in θ is captured by a beta distribution with pdf

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

The Beta Function

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

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt,\ \alpha > 0, \beta > 0\,,$$

and can be expressed in terms of gamma functions:

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

- The gamma function $\Gamma(z)$ is defined by the integral

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt,\ z > 0\,,$$

and has the recursive property $\Gamma(z+1) = z\Gamma(z)$.

The Beta Distribution

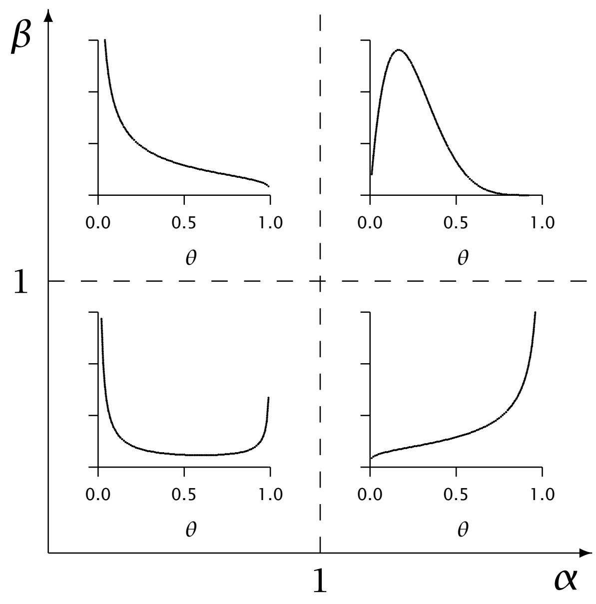
$$f(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \; 0 < \theta < 1 .$$

- The mean of the beta distribution is

$$E(\theta) = \frac{\alpha}{\alpha + \beta}$$

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

General Shapes of the Beta Distribution



Developing a Better Model (IIc)

For a randomly-chosen individual,

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

We call this "continuous mixture" model the shifted-beta-geometric (sBG) distribution

Computing sBG Probabilities

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

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

Estimating Model Parameters

The log-likelihood function is defined as:

$$\begin{aligned} LL(\alpha, \beta|\text{data}) = \; & 369 \times \ln[P(T = 1)] + \\ & 163 \times \ln[P(T = 2)] + \\ & \quad \ldots \qquad\qquad + \\ & 21 \times \ln[P(T = 7)] + \\ & 241 \times \ln[P(T > 7)] \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -1680.27$, which occurs at $\hat{\alpha} = 0.704$ and $\hat{\beta} = 1.182$.

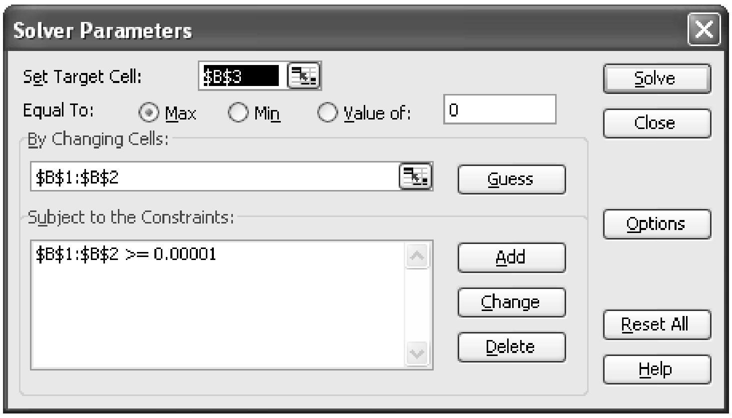
Estimating Model Parameters

	A	B	C	D	E
1	alpha	1.000			
2	beta	1.000			
3	LL	-1741.73			
4					
5	Year	P(T=t)	# Cust.	# Lost	
6	0		1000		
7	1	0.5000		369	-255.7713
8	2	0.1667	468	163	-292.0568
9	[illegible]	[illegible]	[illegible]	86	-213.7020
10	[illegible]	[illegible]	[illegible]	56	-167.7610
11	5	0.0333	289	37	-125.8443
12	6	0.0238	262	27	-100.9171
13	7	0.0179	241	21	-84.5324
14					-501.1454

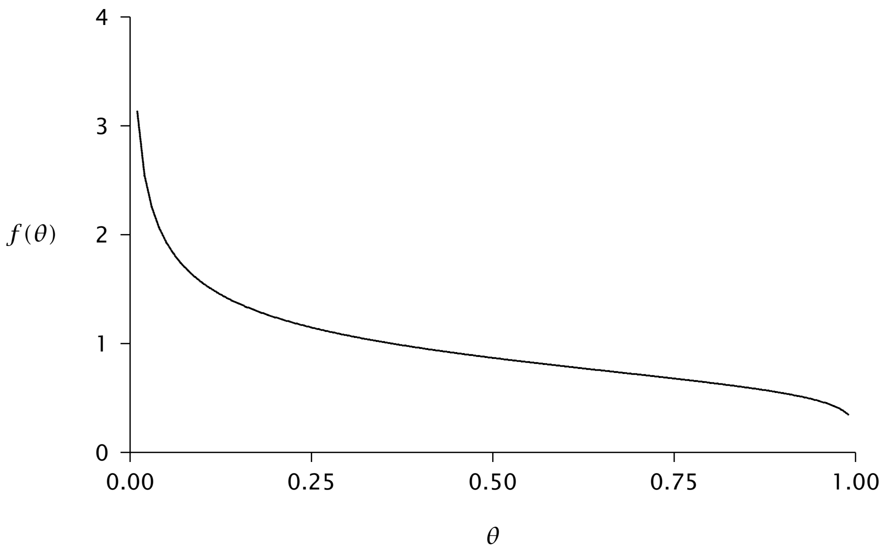
=B1/(B1+B2)

=B7*($B$2+A8-2)/($B$1+$B$2+A8-1)

Estimating Model Parameters

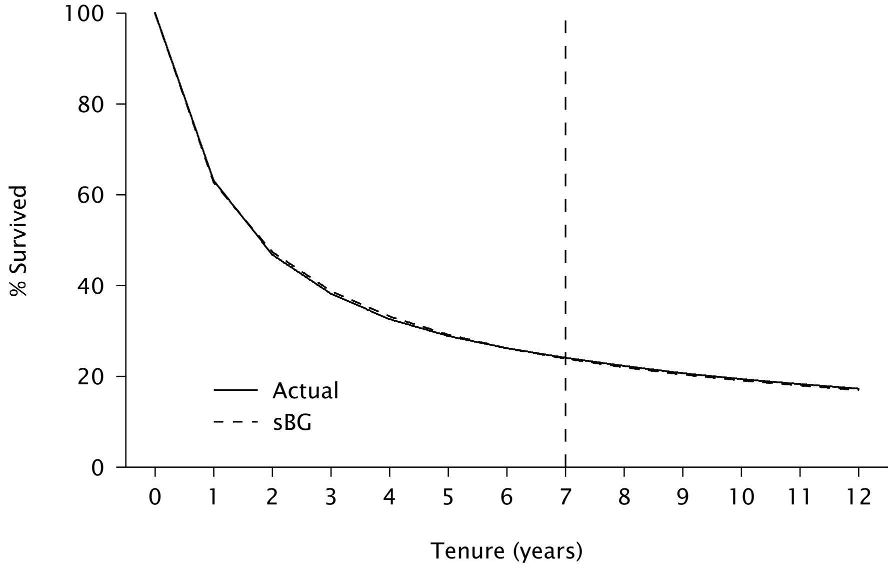


Estimated Distribution of Churn Probabilities

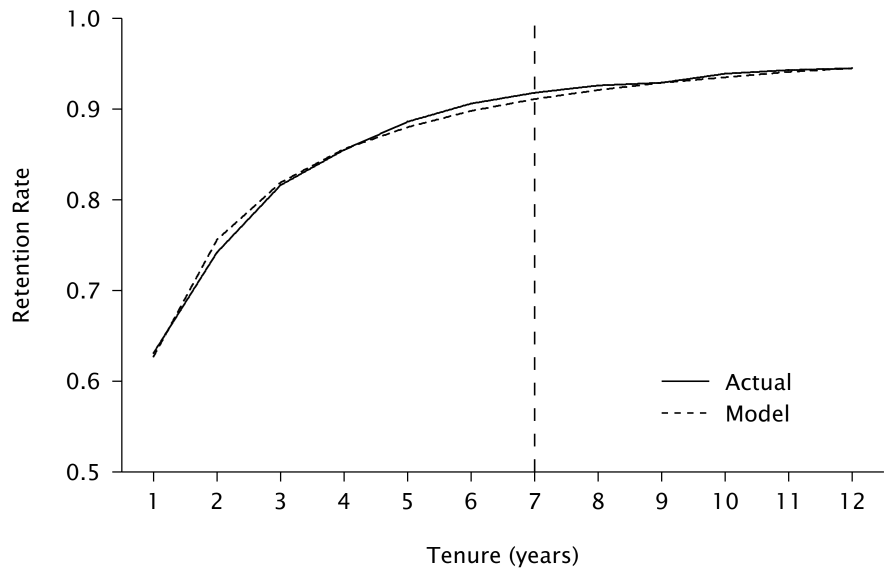


$$\hat{\alpha} = 0.704, \hat{\beta} = 1.182, \widehat{E(\theta)} = 0.373$$

Survival Curve Projection



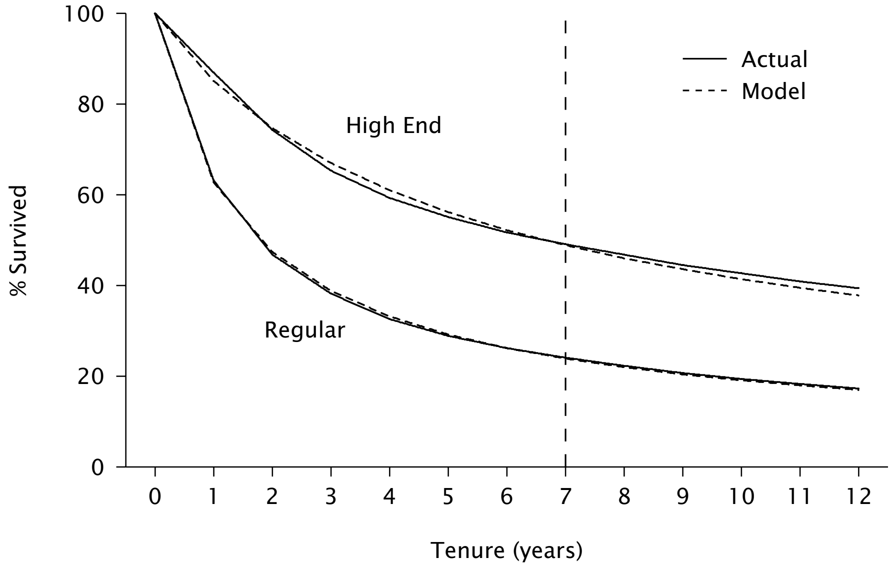
Projecting Retention Rates



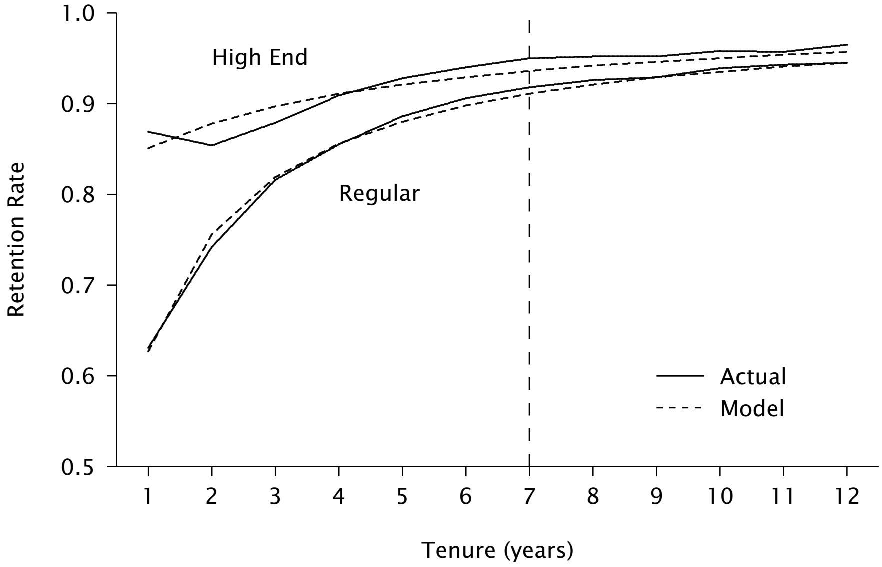
A Further Test of the sBG Model

- The dataset we have been analyzing is for a "regular" segment of customers.
- We also have a dataset for a "high end" customer segment.
- Fitting the sBG model to the data on contract renewals for this segment yields $\hat{\alpha} = 0.668$ and $\hat{\beta} = 3.806$ ($\Longrightarrow \widehat{E(\theta)} = 0.149$).

Survival Curve Projections



Projecting Retention Rates



Concepts and Tools Introduced

- Probability models
- Discrete and continuous mixture models
- Maximum-likelihood estimation of model parameters
- Modelling discrete-time (single-event) duration data
- Models of contract renewal behavior

Further Reading

Buchanan, Bruce and Donald G. Morrison (1988), "A Stochastic Model of List Falloff with Implications for Repeat Mailings," *Journal of Direct Marketing*, **2** (Summer), 7–15.

Fader, Peter S. and Bruce G. S. Hardie (2007), "How to Project Customer Retention," *Journal of Interactive Marketing*, **21** (Winter), 76–90.

Fader, Peter S. and Bruce G. S. Hardie (2007), "How Not to Project Customer Retention."
`<http://brucehardie.com/notes/016/>`

Weinberg, Clarice Ring and Beth C. Gladen (1986), "The Beta-Geometric Distribution Applied to Comparative Fecundability Studies," *Biometrics*, **42** (September), 547–560.

Introduction to 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 individuals (taking individual differences into account) to obtain a model of aggregate behavior.

Uses of Probability Models

- Understanding market-level behavior patterns
- Prediction
  - To settings (e.g., time periods) beyond the observation period
  - Conditional on past behavior
- Profiling behavioral propensities of individuals
- Benchmarks/norms

Building a Probability Model

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

(ii) Identify the *observable* individual-level behavior of interest.

- We denote this by x.

(iii) Select a probability distribution that characterizes this individual-level behavior.

- This is denoted by $f(x|\theta)$.
- We view the parameters of this distribution as individual-level *latent characteristics*.

Building a Probability Model

(iv) Specify a distribution to characterize the distribution of the latent characteristic 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.

Outline

- Problem 1: Projecting Customer Retention Rates
  (Modelling Discrete-Time Duration Data)
- Problem 2: Predicting New Product Trial
  (Modelling Continuous-Time Duration Data)
- Problem 3: Estimating Billboard Exposures
  (Modelling Count Data)
- Problem 4: Test/Roll Decisions in Segmentation- based Direct Marketing
  (Modelling "Choice" Data)
- Problem 5: Characterizing the Purchasing of Hard-Candy
  (Introduction to Finite Mixture Models)
- Problem 6: Who is Visiting khakichinos.com?
  (Incorporating Covariates in Count Models)

Problem 2:
Predicting New Product Trial

(Modelling Continuous-Time Duration Data)

Background

Ace Snackfoods, Inc. has developed a new shelf-stable juice product called Kiwi Bubbles. Before deciding whether or not to "go national" with the new product, the marketing manager for Kiwi Bubbles has decided to commission a year-long test market using IRI's BehaviorScan service, with a view to getting a clearer picture of the product's potential.

The product has now been under test for 24 weeks. On hand is a dataset documenting the number of households that have made a trial purchase by the end of each week. (The total size of the panel is 1499 households.)

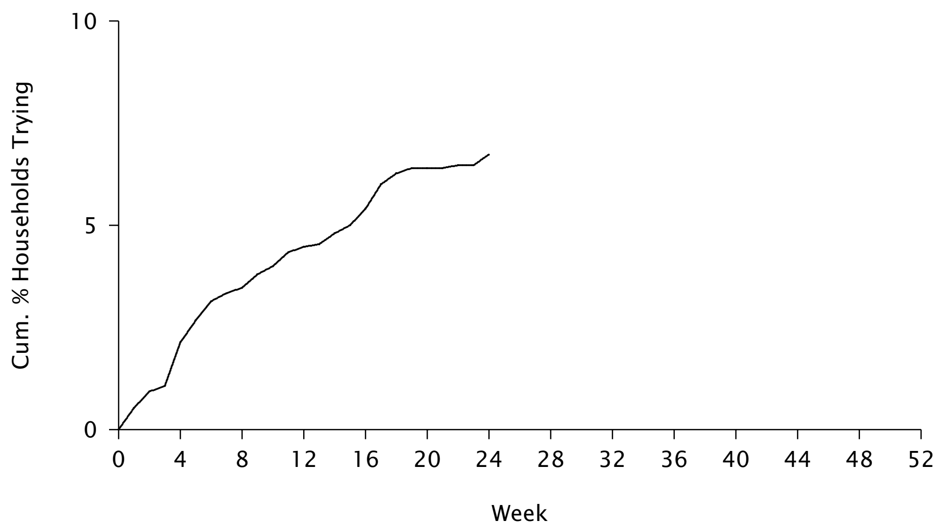
The marketing manager for Kiwi Bubbles would like a forecast of the product's year-end performance in the test market. First, she wants a forecast of the percentage of households that will have made a trial purchase by week 52.

Kiwi Bubbles Cumulative Trial

Week	# Households	Week	# Households
1	8	13	68
2	14	14	72
3	16	15	75
4	32	16	81
5	40	17	90
6	47	18	94
7	50	19	96
8	52	20	96
9	57	21	96
10	60	22	97
11	65	23	97
12	67	24	101

61

Kiwi Bubbles Cumulative Trial



Developing a Model of Trial Purchasing

- Start at the individual-level then aggregate.

  Q: What is the individual-level behavior of interest?

  A: Time (since new product launch) of trial purchase.

- We don't know exactly what is driving the behavior $\Rightarrow$ treat it as a random variable.

The Individual-Level Model

- Let T denote the random variable of interest, and t denote a particular realization.

- Assume time-to-trial is characterized by the exponential distribution with parameter λ (which represents an individual's trial rate).

- The probability that an individual has tried by time t is given by:

$$F(t \mid \lambda) = P(T \leq t \mid \lambda) = 1 - e^{-\lambda t} .$$

Distribution of Trial Rates

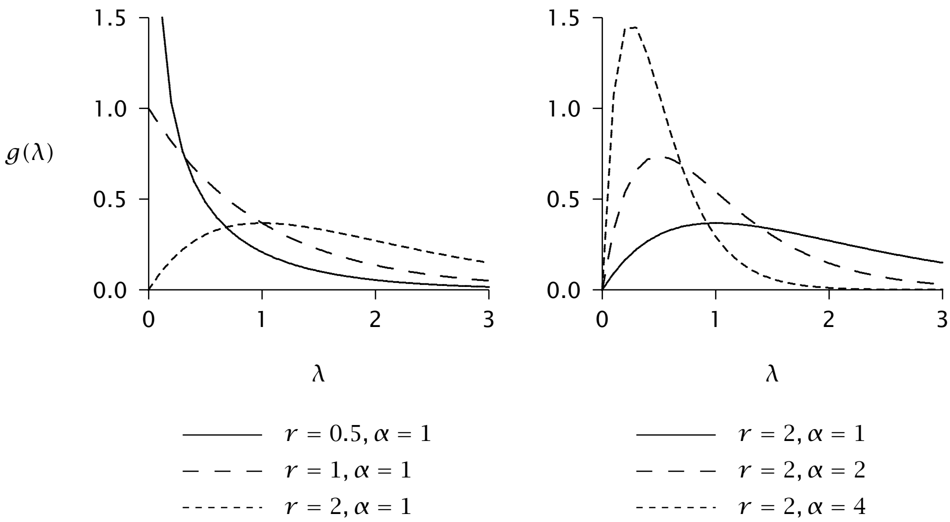
- Assume trial rates are distributed across the population according to a gamma distribution:

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

  where 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



Market-Level Model

The cumulative distribution of time-to-trial at the market-level is given by:

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

We call this the "exponential-gamma" model.

Estimating Model Parameters

The log-likelihood function is defined as:

$$
\begin{aligned}
LL(r, \alpha|\text{data}) = {} & 8 \times \ln[P(0 < T \le 1)] \quad + \\
& 6 \times \ln[P(1 < T \le 2)] \quad + \\
& \ldots \quad + \\
& 4 \times \ln[P(23 < T \le 24)] + \\
& (1499 - 101) \times \ln[P(T > 24)]
\end{aligned}
$$

The maximum value of the log-likelihood function is $LL = -681.4$, which occurs at $\hat{r} = 0.050$ and $\hat{\alpha} = 7.973$.

Estimating Model Parameters

	A	B	C	D	E	F
1	Product:	Kiwi Bubbles			r	1.000
2	Panelists:	1499			alpha	1.000
3					LL	-4909.5
4		Cum_Trl				
5	Week	# HHs	Incr_Trl	P(T <= t)	P(try week t)	
6	[illegible]	[illegible]	8	0.50000	0.50000	-5.545
7	2	14	6	0.66667	0.16667	-10.751
8	3	16	2	[illegible]	0.08333	-4.970
9	4	32	16	[illegible]	0.05000	-47.932
10	5	40	8	0.83333	[illegible]	-27.210
11	6	47	7	0.85714	0.02381	-26.164
12	7	50	3	0.87500	0.01786	-12.076
13	8	52	2	0.88889	0.01389	-8.553
14	9	57	5	0.90000	0.01111	-22.499
15	10	60	3	0.90909	0.00909	-14.101
29	24	101	4	0.96000	0.00167	-25.588
30						-4499.988

=SUM(F6:F30)

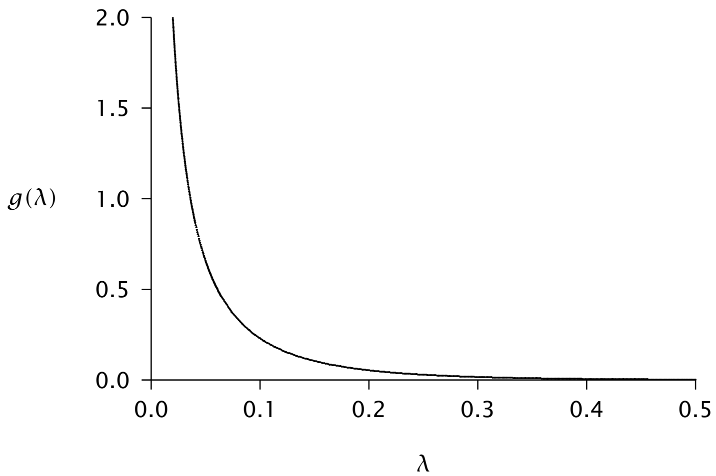
=1-(F$2/(F$2+A6))^F$1

=D7-D6

=C8*LN(E8)

=(B2-B29)*LN(1-D29)

Estimated Distribution of λ



$\hat{r} = 0.050, \hat{\alpha} = 7.973$

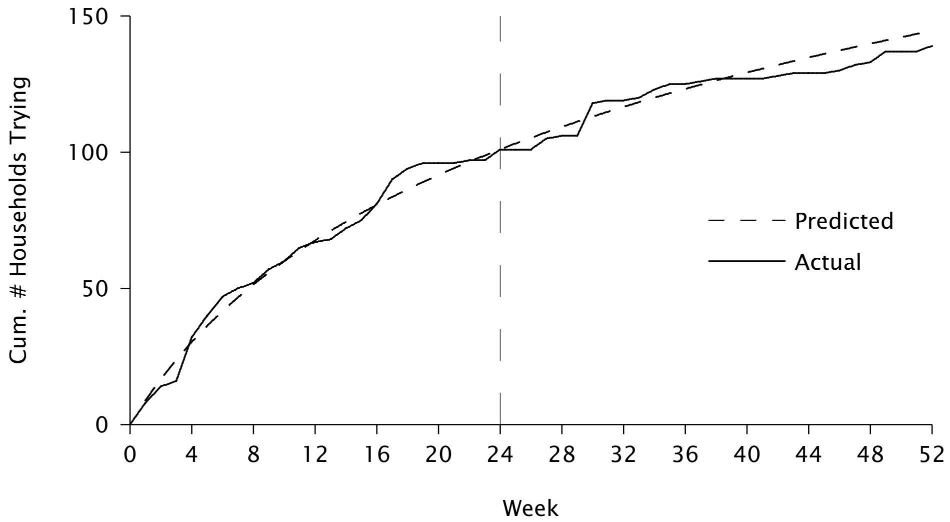
Forecasting Trial

- $F(t)$ represents the probability that a randomly chosen household has made a trial purchase by time t, where $t = 0$ corresponds to the launch of the new product.
- Let $T(t)$ = cumulative # households that have made a trial purchase by time t:

$$\begin{aligned} E[T(t)] &= N \times \hat{F}(t) \\ &= N \left\{ 1 - \left(\frac{\hat{\alpha}}{\hat{\alpha} + t} \right)^{\hat{r}} \right\} . \end{aligned}$$

  where N is the panel size.
- Use projection factors for market-level estimates.

Cumulative Trial Forecast



Further Model Extensions

- Add a “never triers” parameter.
- Incorporate the effects of marketing covariates.
- Model repeat sales using a “depth of repeat” formulation, where transitions from one repeat class to the next are modeled using an “exponential-gamma”-type model.

Concepts and Tools Introduced

- Modelling continuous-time (single-event) duration data
- Models of new product trial

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Robert Zeithammer (2003), "Forecasting New Product Trial in a Controlled Test Market Environment," *Journal of Forecasting*, **22** (August), 391–410.

Hardie, Bruce G. S., Peter S. Fader, and Michael Wisniewski (1998), "An Empirical Comparison of New Product Trial Forecasting Models," *Journal of Forecasting*, **17** (June–July), 209–229.

Kalbfleisch, John D. and Ross L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, 2nd edn., New York: Wiley.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: Wiley.

Problem 3: Estimating Billboard Exposures

(Modelling Count Data)

Background

One advertising medium at the marketer's disposal is the outdoor billboard. The unit of purchase for this medium is usually a "monthly showing," which comprises a specific set of billboards carrying the advertiser's message in a given market.

The effectiveness of a monthly showing is evaluated in terms of three measures: reach, (average) frequency, and gross rating points (GRPs). These measures are determined using data collected from a sample of people in the market.

Respondents record their daily travel on maps. From each respondent's travel map, the total frequency of exposure to the showing over the survey period is counted. An "exposure" is deemed to occur each time the respondent travels by a billboard in the showing, on the street or road closest to that billboard, going towards the billboard's face.

Background

The standard approach to data collection requires each respondent to fill out daily travel maps for *an entire month*. The problem with this is that it is difficult and expensive to get a high proportion of respondents to do this accurately.

B&P Research is interested in developing a means by which it can generate effectiveness measures for a monthly showing from a survey in which respondents fill out travel maps for *only one week*.

Data have been collected from a sample of 250 residents who completed daily travel maps for one week. The sampling process is such that approximately one quarter of the respondents fill out travel maps during each of the four weeks in the target month.

Effectiveness Measures

The effectiveness of a monthly showing is evaluated in terms of three measures:

- Reach: the proportion of the population exposed to the billboard message at least once in the month.
- Average Frequency: the average number of exposures (per month) among those people reached.
- Gross Rating Points (GRPs): the mean number of exposures per 100 people.

Distribution of Billboard Exposures (1 week)

# Exposures	# People	# Exposures	# People
0	48	12	5
1	37	13	3
2	30	14	3
3	24	15	2
4	20	16	2
5	16	17	2
6	13	18	1
7	11	19	1
8	9	20	2
9	7	21	1
10	6	22	1
11	5	23	1

Average # Exposures = 4.456

Modelling Objective

Develop a model that enables us to estimate a billboard showing's reach, average frequency, and GRPs for the month using the one-week data.

Modelling Issues

- Modelling the exposures to showing in a week.
- Estimating summary statistics of the exposure distribution for a longer period of time (i.e., one month).

Model Development (I)

- Let the random variable X denote the number of exposures to the showing in a week.
- At the individual-level, X is assumed to be Poisson distributed with (exposure) rate parameter λ:

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

- All individuals are assumed to have the same exposure rate.

Estimating Model Parameters

The log-likelihood function is defined as:

$$\begin{aligned} LL(\lambda \mid \text{data}) = 48 \times \ln[P(X = 0)] &+ \\ 37 \times \ln[P(X = 1)] &+ \\ 30 \times \ln[P(X = 2)] &+ \\ \ldots &+ \\ 1 \times \ln[P(X = 23)] & \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -929.0$, which occurs at $\hat{\lambda} = 4.456$.

Estimating Model Parameters

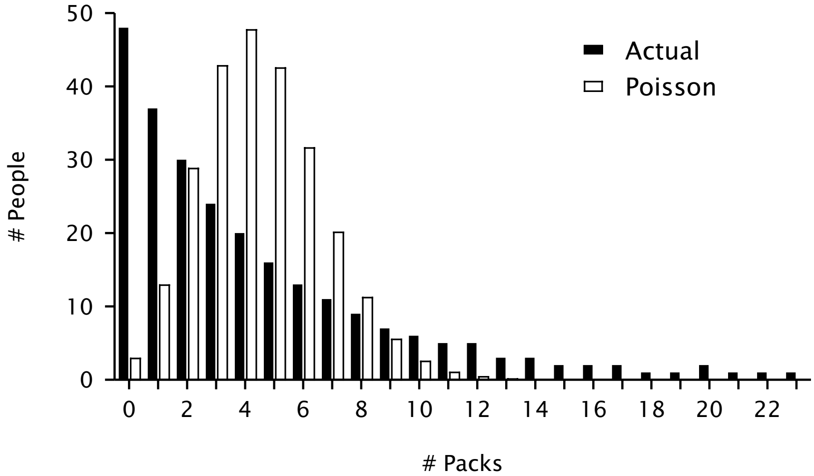
	A	B	C	D
1	lambda	3.000		
2	LL	-1005.8		
3				
4	x	f_x	P(X=x)	
5	0	48	0.04979	-144.00
6	[illegible]	[illegible]	[illegible]	-70.35
7	[illegible]	[illegible]	[illegible]	-44.88
8	3	24	0.22404	-35.90
9	4	[illegible]	[illegible]	-35.67
10	5	16	0.10082	-36.71
11	6	13	0.05041	-38.84
12	7	11	0.02160	-42.18
13	8	9	0.00810	-43.34
14	9	7	0.00270	-41.40
27	22	1	0.00000	-27.30
28	23	1	0.00000	-29.34

=SUM(D5:D28)

=POISSON(A5,$B$1,FALSE)

=B9*LN(C9)

Fit of the Poisson Model



Model Development (II)

- Let the random variable X denote the number of exposures to the showing in a week.
- At the individual-level, X is assumed to be Poisson distributed with (exposure) rate parameter λ:

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

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

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

Model Development (II)

The distribution of exposures at the population- level is given by:

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

This is called the Negative Binomial Distribution, or NBD model.

Mean of the NBD

We can derive an expression for the mean of the NBD *by conditioning*:

$$\begin{aligned} E(X) &= E\big[E(X \mid \lambda)\big] \\ &= \int_0^\infty E(X \mid \lambda) g(\lambda \mid r, \alpha)\, d\lambda \\ &= \frac{r}{\alpha}\,. \end{aligned}$$

Computing NBD Probabilities

- Note that

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{r + x - 1}{x(\alpha + 1)}$$

- We can therefore compute NBD probabilities using the following *forward recursion* formula:

$$P(X = x) = \begin{cases} \left(\dfrac{\alpha}{\alpha + 1}\right)^r & x = 0 \\[2ex] \dfrac{r + x - 1}{x(\alpha + 1)} \times P(X = x - 1) & x \geq 1 \end{cases}$$

Estimating Model Parameters

The log-likelihood function is defined as:

$$
\begin{aligned}
LL(r, \alpha|\text{data}) = 48 \times \ln[P(X=0)] &+ \\
37 \times \ln[P(X=1)] &+ \\
30 \times \ln[P(X=2)] &+ \\
\ldots &+ \\
1 \times \ln[P(X=23)]&
\end{aligned}
$$

The maximum value of the log-likelihood function is $LL = -649.7$, which occurs at $\hat{r} = 0.969$ and $\hat{\alpha} = 0.218$.

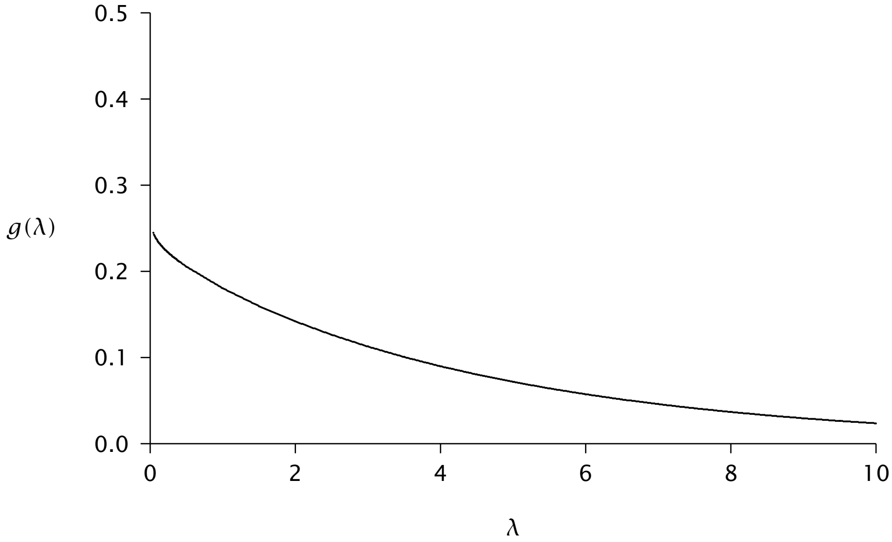
Estimating Model Parameters

	A	B	C	D
1	r	1.000		
2	alpha	1.000		
3	LL	-945.5		
4				
5	x	f_x	P(X=x)	
6	0	48	0.50000	-33.27
7	1	37	0.25000	-51.29
8	2	30	0.12500	-62.38
9	[illegible]	[illegible]	[illegible]	-66.54
10	4	20	0.03125	-69.31
11	5	16	0.01563	-66.54
12	6	13	0.00781	-63.08
13	7	11	0.00391	-61.00
14	8	9	0.00195	-56.14
15	9	7	0.00098	-48.52
28	22	1	0.00000	-15.94
29	23	1	0.00000	-16.64

=(B2/(B2+1))^B1

=C6*($B$1+A7-1)/(A7*($B$2+1))

Estimated Distribution of λ



$\hat{r} = 0.969, \hat{\alpha} = 0.218$

NBD for a Non-Unit Time Period

- Let $X(t)$ be the number of exposures occuring in an observation period of length t time units.
- If, for a unit time period, the distribution of exposures *at the individual-level* is distributed Poisson with rate parameter λ, then $X(t)$ has a Poisson distribution with rate parameter λt:

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

NBD for a Non-Unit Time Period

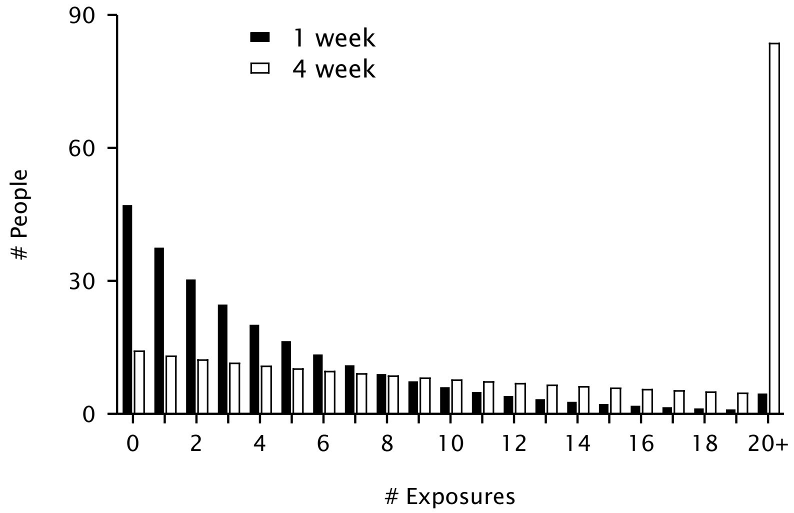
- The distribution of exposures at the population-level is given by:

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

- The mean of this distribution is given by

$$E[X(t)] = \frac{rt}{\alpha}$$

Exposure Distributions: 1 week vs. 4 week



Effectiveness of Monthly Showing

- For $t = 4$, we have:
  - $P(X(t) = 0) = 0.056$, and
  - $E[X(t)] = 17.82$
- It follows that:
  - Reach $= 1 - P(X(t) = 0)$
    $= 94.4\%$
  - Frequency $= E[X(t)]/(1 - P(X(t) = 0))$
    $= 18.9$
  - GRPs $= 100 \times E[X(t)]$
    $= 1782$

Concepts and Tools Introduced

- Counting processes
- The NBD model
- Extrapolating an observed histogram over time
- Using models to estimate "exposure distributions" for media vehicles

Further Reading

Ehrenberg, A. S. C. (1988), *Repeat-Buying*, 2nd edn., London: Charles Griffin & Company, Ltd. (Available online at <http://www.empgens.com/A/rb/rb.html>.)

Greene, Jerome D. (1982), *Consumer Behavior Models for Non-Statisticians*, New York: Praeger.

Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, **6** (April), 145–159.

Problem 4:
Test/Roll Decisions in Segmentation-based Direct Marketing

(Modelling "Choice" Data)

The "Segmentation" Approach

i. Divide the customer list into a set of (homogeneous) segments.

ii. Test customer response by mailing to a random sample of each segment.

iii. Rollout to segments with a response rate (RR) above some cut-off point,

$$\text{e.g., RR} > \frac{\text{cost of each mailing}}{\text{unit margin}}$$

Ben's Knick Knacks, Inc.

- A consumer durable product (unit margin = $161.50, mailing cost per 10,000 = $3343)
- 126 segments formed from customer database on the basis of past purchase history information
- Test mailing to 3.24% of database

Ben's Knick Knacks, Inc.

Standard approach:

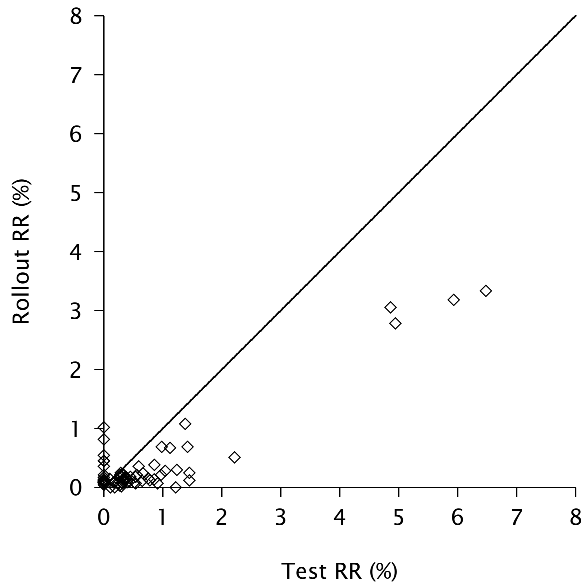
- Rollout to all segments with

$$\text{Test RR} > \frac{3{,}343/10{,}000}{161.50} = 0.00207$$

- 51 segments pass this hurdle

103

Test vs. Actual Response Rate



Modelling Objective

Develop a model that leverages the whole data set to make better informed decisions.

Model Development

i. Assuming all members of segment s have the same (unknown) response probability p_s, X_s has a binomial distribution:

$$P(X_s = x_s | m_s, p_s) = \binom{m_s}{x_s} p_s^{x_s} (1 - p_s)^{m_s - x_s} \,,$$

with $E(X_s | m_s, p_s) = m_s p_s$.

ii. Heterogeneity in p_s is captured using a beta distribution:

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

The Beta Binomial Model

The aggregate distribution of responses to a mailing of size m_s is given by

$$
\begin{aligned}
P(X_s = x_s | m_s\ \alpha, \beta) \\
&= \int_0^1 P(X_s = x_s | m_s, p_s)\, g(p_s \mid \alpha, \beta)\, dp_s \\
&= \binom{m_s}{x_s} \frac{B(\alpha + x_s, \beta + m_s - x_s)}{B(\alpha, \beta)}\,.
\end{aligned}
$$

Estimating Model Parameters

The log-likelihood function is defined as:

$$
\begin{aligned}
LL(\alpha, \beta | \text{data}) &= \sum_{s=1}^{126} \ln[P(X_s = x_s | m_s, \alpha, \beta)] \\
&= \sum_{s=1}^{126} \ln\Bigg[\frac{m_s!}{(m_s - x_s)!\, x_s!} \underbrace{\frac{\Gamma(\alpha + x_s)\Gamma(\beta + m_s - x_s)}{\Gamma(\alpha + \beta + m_s)}}_{B(\alpha + x_s, \beta + m_s - x_s)} \underbrace{\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{1/B(\alpha, \beta)} \Bigg]
\end{aligned}
$$

The maximum value of the log-likelihood function is $LL = -200.5$, which occurs at $\hat{\alpha} = 0.439$ and $\hat{\beta} = 95.411$.

Estimating Model Parameters

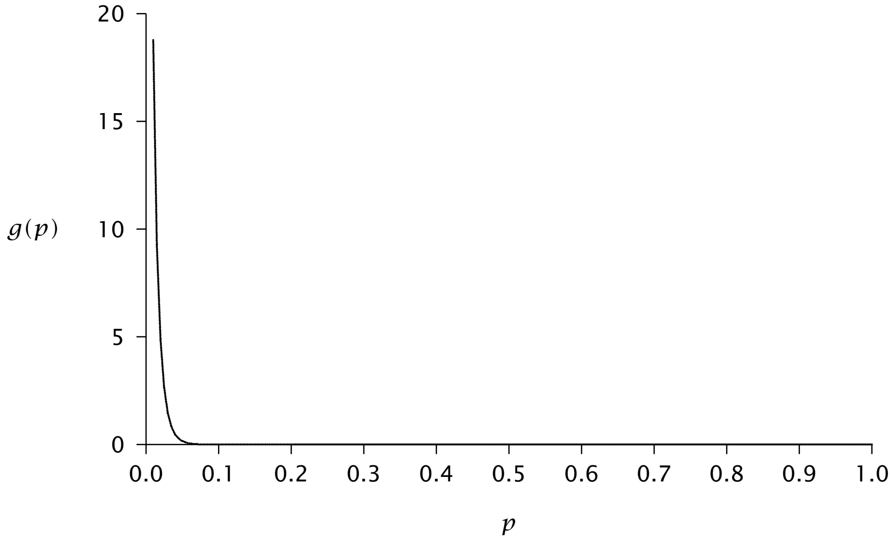
	A	B	C	D	E
1	alpha	1.000		B(alpha,beta)	1.000
2	beta	1.000			
3	LL	-718.9			
4					
5	Segment	m_s	x_s	P(X=x|m)	
6	1	34	0	0.02857	-3.555
7	[illegible]	[illegible]	[illegible]	[illegible]	-4.635
8	[illegible]	[illegible]	[illegible]	[illegible]	-3.989
9	[illegible]	[illegible]	[illegible]	[illegible]	-4.984
10	[illegible]	[illegible]	[illegible]	[illegible]	-7.135
11	6	144	7	0.00690	-4.977
12	7	1235	80	[illegible]	-7.120
13	8	573	34	[illegible]	-6.353
14	9	1083	24	0.00092	-6.988
130	125	383	0	0.00260	-5.951
131	126	404	0	0.00247	-6.004

=SUM(E6:E131)

=COMBIN(B6,C6)*EXP(GAMMALN(B$1+C6)+GAMMALN(B$2+B6-C6)-GAMMALN(B$1+B$2+B6))/E$1

=LN(D11)

Estimated Distribution of *p*



$$\hat{\alpha} = 0.439, \hat{\beta} = 95.411, \bar{p} = 0.0046$$

Applying the Model

What is our best guess of p_s given a response of x_s to a test mailing of size m_s?

Intuitively, we would expect

$$E(p_s|x_s, m_s) \approx \omega\frac{\alpha}{\alpha+\beta} + (1-\omega)\frac{x_s}{m_s}$$

Bayes' Theorem

- The *prior distribution* $g(p)$ captures the possible values p can take on, prior to collecting any information about the specific individual.
- The *posterior distribution* $g(p|x)$ is the conditional distribution of p, given the observed data x. It represents our updated opinion about the possible values p can take on, now that we have some information x about the specific individual.
- According to Bayes' theorem:

$$g(p|x) = \frac{f(x|p)g(p)}{\int f(x|p)g(p)\,dp}$$

Bayes' Theorem

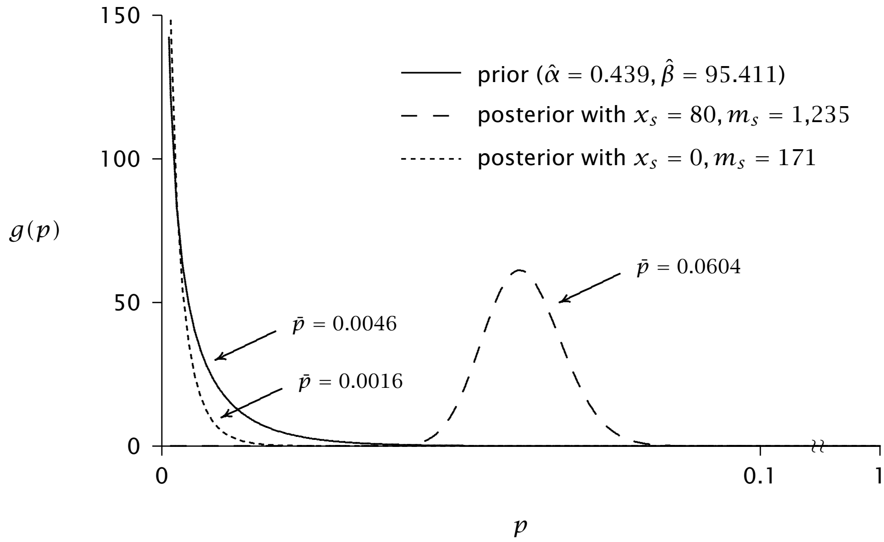
For the beta-binomial model, we have:

$$g(p_s|X_s = x_s, m_s) = \frac{\overbrace{P(X_s = x_s|m_s, p_s)}^{\text{binomial}}\,\overbrace{g(p_s)}^{\text{beta}}}{\underbrace{\int_0^1 P(X_s = x_s|m_s, p_s)\, g(p_s)\, dp_s}_{\text{beta-binomial}}}$$

$$= \frac{1}{B(\alpha + x_s, \beta + m_s - x_s)} p_s^{\alpha + x_s - 1}(1 - p_s)^{\beta + m_s - x_s - 1}$$

which is a beta distribution with parameters $\alpha + x_s$ and $\beta + m_s - x_s$.

Distribution of p



Applying the Model

Recall that the mean of the beta distribution is $\alpha/(\alpha+\beta)$. Therefore

$$E(p_s|X_s = x_s, m_s) = \frac{\alpha + x_s}{\alpha + \beta + m_s}$$

which can be written as

$$\left(\frac{\alpha+\beta}{\alpha+\beta+m_s}\right)\frac{\alpha}{\alpha+\beta} + \left(\frac{m_s}{\alpha+\beta+m_s}\right)\frac{x_s}{m_s}$$

- a weighted average of the test RR (x_s/m_s) and the population mean ($\alpha/(\alpha+\beta)$).
- "Regressing the test RR to the mean"

Model-Based Decision Rule

- Rollout to segments with:

$$E(p_s|X_s = x_s, m_s) > \frac{3{,}343/10{,}000}{161.5} = 0.00207$$

- 66 segments pass this hurdle
- To test this model, we compare model predictions with managers' actions. (We also examine the performance of the "standard" approach.)

Results

	Standard	Manager	Model
# Segments (Rule)	51		66
# Segments (Act.)	46	71	53
Contacts	682,392	858,728	732,675
Responses	4,463	4,804	4,582
Profit	$492,651	$488,773	$495,060

Use of model results in a profit increase of $6,287; 126,053 fewer contacts, saved for another offering.

Concepts and Tools Introduced

- “Choice” processes
- The Beta Binomial model
- “Regression-to-the-mean” and the use of models to capture such an effect
- Bayes’ theorem (and “empirical Bayes” methods)
- Using “empirical Bayes” methods in the development of targeted marketing campaigns

Further Reading

Colombo, Richard and Donald G. Morrison (1988), "Blacklisting Social Science Departments with Poor Ph.D. Submission Rates," *Management Science*, **34** (June), 696–706.

Morrison, Donald G. and Manohar U. Kalwani (1993), "The Best NFL Field Goal Kickers: Are They Lucky or Good?" *Chance*, **6** (August), 30–37.

Morwitz, Vicki G. and David C. Schmittlein (1998), "Testing New Direct Marketing Offerings: The Interplay of Management Judgment and Statistical Models," *Management Science*, **44** (May), 610–628.

Bayes' Theorem and the NBD Model

Recall:

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

The mean of the NBD is $E(X) = r/\alpha$.

Bayes' Theorem and the NBD Model

Applying Bayes' theorem:

$$g(\lambda \mid r, \alpha, X = x) = \frac{\overbrace{\frac{\lambda^x e^{-\lambda}}{x!}}^{P(X=x|\lambda)} \overbrace{\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}}^{g(\lambda|r,\alpha)}}{\underbrace{\frac{\Gamma(r+x)}{\Gamma(r)\,x!}\left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^x}_{P(X=x\,|\,r,\alpha)}}$$

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

Expected behavior in a non-overlapping period:

$$E(Y \mid X = x) = \left(\frac{\alpha}{\alpha+1}\right)\frac{r}{\alpha} + \left(\frac{1}{\alpha+1}\right)x$$

Bayes' Theorem and the EG Model

Recall:

$$P(T \le t \mid r, \alpha) = \int_0^\infty P(T \le t|\lambda)\, g(\lambda \mid r, \alpha)\, d\lambda$$

$$= 1 - \left(\frac{\alpha}{\alpha+t}\right)^r .$$

Applying Bayes' theorem:

$$g(\lambda \mid r, \alpha, T > s) = e^{-\lambda s}\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} \Big/ \left(\frac{\alpha}{\alpha+s}\right)^r$$

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

Bayes' Theorem and the EG Model

The predictive distribution for the EG model is:

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

Problem 5:
Characterizing the Purchasing of Hard-Candy

(Introduction to Finite Mixture Models)

Distribution of Hard-Candy Purchases

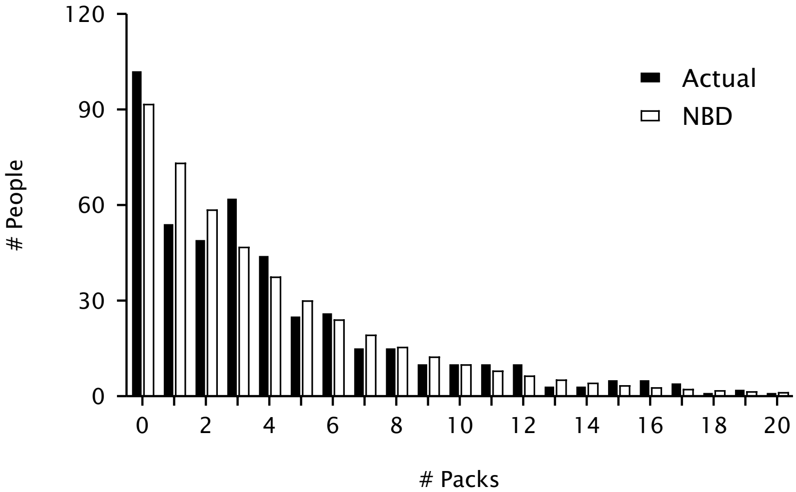
# Packs	# People	# Packs	# People
0	102	11	10
1	54	12	10
2	49	13	3
3	62	14	3
4	44	15	5
5	25	16	5
6	26	17	4
7	15	18	1
8	15	19	2
9	10	20	1
10	10		

Source: Dillon and Kumar (1994)

Fitting the NBD

	A	B	C	D	E	F	G	H	I	J
1	r	0.998								
2	alpha	0.250								
3	LL	-1140.02								
4										
5	# Packs	Observed	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
6	0	102	0.20073	-163.79	91.5		0	102	91.5	1.20
7	1	54	0.16021	-98.89	73.1		1	54	73.1	4.97
8	2	49	0.12802	-100.72	58.4	=B$27*C6	2	49	=(H6-I6)^2/I6	1.51
9	3	62	0.10234	-141.32	46.7		3	62	46.7	5.04
10	4	44	0.08183	-110.14	37.3		4	44	37.3	1.20
11	5	25	0.06543	-68.17	29.8		5	25	29.8	0.78
12	6	26	0.05233	-76.71	23.9		6	26	23.9	0.19
13	7	15	0.04185	-47.60	19.1		7	15	19.1	0.87
14	8	15	0.03347	-50.96	15.3		8	15	15.3	0.00
15	9	10	0.02677	-36.20	12.2		9	10	12.2	0.40
16	10	10	0.02141	-38.44	9.8		10	10	9.8	0.01
17	11	10	0.01713	-40.67	7.8		11	10	7.8	0.61
18	12	10	0.01370	-42.90	6.2		12	10	6.2	2.25
19	13	3	0.01096	-13.54	5.0		13	3	5.0	0.80
20	14	3	0.00876	-14.21	4.0		14	3	4.0	0.25
21	15	5	0.00701	-24.80	3.2		15+	18	11.8	3.27
22	16	5	0.00561	-25.92	2.6					23.35
23	17	4	0.00449	-21.63	2.0					
24	18	1	0.00359	-5.63	1.6				# params	2
25	19	2	0.00287	-11.71	1.3			=CHIDIST(J22,J25)	df	13
26	20	1	0.00230	-6.08	1.0					
27		456							p-value	0.038

Fit of the NBD



The Zero-Inflated NBD Model

Because of the "excessive" number of zeros, let us consider the zero-inflated NBD (ZNBD) model:

- a proportion π of the population never buy hard-candy
- the visiting behavior of the "ever buyers" can be characterized by the NBD model

$$P(X = x) = \delta_{x=0}\pi + (1 - \pi) \times \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x$$

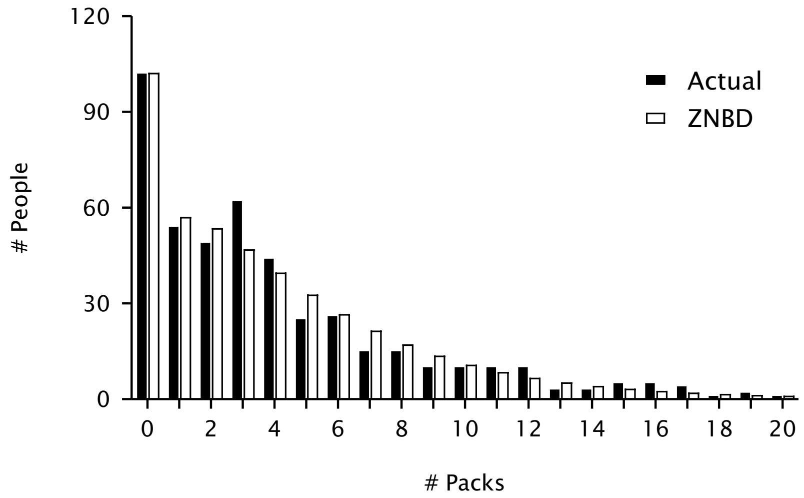
This is sometimes called the "NBD with hard-core non-buyers" model.

Fitting the ZNBD

	A	B	C	D	E	F	G	H	I	J	K
1	r	1.504									
2	alpha	0.334									
3	pi	0.113									
4	LL	-1136.17									
5											
6			P(X=x)								
7	# Packs	Observed	NBD	ZNBD	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
8	0	102	0.12468	0.22368	-152.75	102.0		0	102	102.0	0.00
9	1	54	0.14054	0.12465	-112.44	56.8		1	54	56.8	0.14
10	2	49	0.13188	0.11[illegible]	[illegible]	[illegible]		2	49	53.3	0.35
11	3	62	0.11545	0.10239	-141.29	46.7		3	62	46.7	5.02
12	4	44	0.09743	0.08641	-107.74	39.4		4	44	39.4	0.54
13	5	25	0.08039	0.07130	-66.02	32.5		5	25	32.5	1.74
14	6	26	0.06531	0.05793	-74.06	26.4		6	26	26.4	0.01
15	7	15	0.05248	0.04654	-46.01	21.2		7	15	21.2	1.82
16	8	15	0.04181	0.03708	-49.42	16.9		8	15	16.9	0.22
17	9	10	0.03309	0.02935	-35.28	13.4		9	10	13.4	0.86
18	10	10	0.02605	0.02311	-37.68	10.5		10	10	10.5	0.03
19	11	10	0.02042	0.01811	-40.11	8.3		11	10	8.3	0.37
20	12	10	0.01595	0.01415	-42.58	6.5		12	10	6.5	1.95
21	13	3	0.01242	0.01101	-13.53	5.0		13	3	5.0	0.81
22	14	3	0.00964	0.00855	-14.28	3.9		14	3	3.9	0.21
23	15	5	0.00747	0.00663	-25.08	3.0		15+	18	10.4	5.48
24	16	5	0.00578	0.00512	-26.37	2.3					19.54
25	17	4	0.00446	0.00395	-22.13	1.8					
26	18	1	0.00343	0.00305	-5.79	1.4				# params	3
27	19	2	0.00264	0.00234	-12.11	1.1				df	12
28	20	1	0.00203	0.00180	-6.32	0.8					
29		456								p-value	0.076

=(A8=0)*B$3+(1-B$3)*C8

Fit of the ZNBD



What is Wrong With the NBD Model?

The assumptions underlying the model could be wrong on two accounts:

i. at the individual-level, the number of purchases is not Poisson distributed

ii. purchase rates (λ) are not gamma-distributed

Relaxing the Gamma Assumption

- Replace the continuous distribution with a discrete distribution by allowing for multiple (discrete) segments each with a different (latent) buying rate:

$$P(X = x) = \sum_{s=1}^{S} \pi_s P(X = x|\lambda_s), \quad \sum_{s=1}^{S} \pi_s = 1$$

- This is called a finite mixture model.

Fitting the One-Segment Model

	A	B	C	D
1	lambda	3.991		
2	LL	-1545.00		
3				
4	# Packs	Observed	P(X=x)	LL
5	0	102	0.01848	-407.11
6	1	54	0.07375	-140.78
7	2	49	0.14717	-93.89
8	3	62	0.19579	-101.10
9	4	44	0.19536	-71.85
10	5	25	0.15595	-46.46
25	20	1	0.00000	-18.64
26		456		

Fitting the Two-Segment Model

	A	B	C	D	E	F
1	lambda_1	1.802				
2	lambda_2	9.121				
3	pi	0.701				
4	LL	-1188.83				
5						
6	# Packs	Observed	Seg1	Seg2	P(X=x)	LL
7	0	102	0.16494	0.00011	0.11564	-220.04
8	[illegible]	[illegible]	[illegible]725	0.00100	0.20864	-84.63
9	[illegible]	49	[illegible]	0.00455	0.189[illegible]	[illegible]
10	[illegible]	[illegible]	[illegible]	0.01383	0.11691	-133.07
11	4	44	0.07249	0.0[illegible]	[illegible]	[illegible]23.61
12	5	25	0.02613	0.05753	0.03552	-83.44
27	20	1	0.00000	0.00071	0.00021	-8.45
28		456				

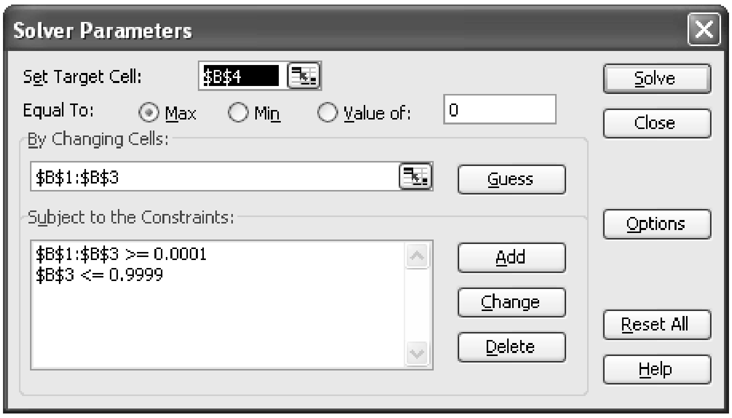
=POISSON(A7,B$1,FALSE)

=POISSON(A7,B$2,FALSE)

=B$3*C7+(1-B$3)*D7

=B7*LN(E7)

Fitting the Two-Segment Model



Estimating the Mixing Proportions

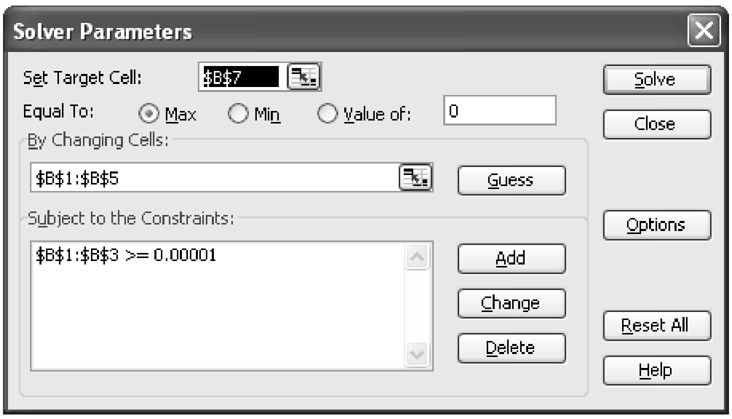
- For more than two segments, satisfying the constraints that $0 < \pi_s < 1$ while ensuring that $\sum_{s=1}^{S} \pi_s = 1$ can be computationally difficult.
- We therefore reparameterize the mixing proportions:

$$\pi_s = \frac{\exp(\theta_s)}{\sum_{s'=1}^{S} \exp(\theta_{s'})}, \quad \theta_S = 0.$$

Fitting the Three-Segment Model

	A	B	C	D	E	F	G
1	lambda_1	3.483					
2	lambda_2	11.216					
3	lambda_3	0.291					
4	theta_1	0.674	1.963	←	=EXP(B4)		
5	theta_2	-0.430	0.650				
6	theta_3	0	1.000				
7	LL	-1132.04		=C4/SUM(C4:C6)			
8							
9			0.543	0.180	0.277		
10	# Packs	Observed	Seg1	Seg2	Seg3	P(X=x)	LL
11	0	102	0.03071	0.00001	0.74786	0.22367	-152.76
12	1	54	0.10696	0.00015	0.21728	0.11827	-115.28
13	2	49	=SUMPRODUCT(C$9:E$9,C11:E11)			1009	-108.12
14	3	62	0.21629	0.00317	0.00306	0.11892	-132.02
15	4	44	0.18835	0.00887	0.00022	0.10399	-99.59
16	5	25	0.13122	0.01991	0.00001	0.07487	-64.80
31	20	1	0.00000	0.00549	0.00000	0.00099	-6.92
32		456					

Fitting the Three-Segment Model



Fitting the Four-Segment Model

	A	B	C	D	E	F	G	H
1	lambda_1	3.002						
2	lambda_2	0.205						
3	lambda_3	7.418						
4	lambda_4	12.873						
5	theta_1	1.598	4.943					
6	theta_2	0.876	2.401					
7	theta_3	0.398	1.489					
8	theta_4	0	1.000					
9	LL	-1130.07						
10								
11			0.503	0.244	0.151	0.102		
12	# Packs	Observed	Seg1	Seg2	Seg3	Seg4	P(X=x)	LL
13	0	102	0.04969	0.81487	0.00060	0.00000	0.22406	-152.58
14	1	54	0.14917	0.16683	0.00445	0.00003	0.11641	-116.14
15	2	49	0.22390	0.01708	0.01652	0.00021	0.11925	-104.20
16	3	62	0.22404	0.00117	0.04084	0.00091	0.11919	-131.88
17	4	44	0.16814	0.00006	0.07574	0.00294	0.09631	-102.97
18	5	25	0.10095	0.00000	0.11237	0.00756	0.06853	-67.01
33	20	1	0.00000	0.00000	0.00006	0.01647	0.00168	-6.39
34		456						

Parameter Estimates

	Seg 1	Seg 2	Seg 3	Seg 4	*LL*
λ	3.991				−1545.00
λ_s	1.802	9.121			−1188.83
π_s	0.701	0.299			
λ_s	0.291	3.483	11.216		−1132.04
π_s	0.277	0.543	0.180		
λ_s	0.205	3.002	7.418	12.873	−1130.07
π_s	0.244	0.503	0.151	0.102	

How Many Segments?

- Controlling for the extra parameters, is an $S + 1$ segment model better than an S segment model?
- We can't use the likelihood ratio test because its properties are violated
- It is standard practice to use "information-theoretic" model selection criteria
- A common measure is the Bayesian information criterion:

$$\text{BIC} = -2LL + p\ln(N)$$

where p is the number of parameters and N is the sample size
- Rule: choose S to minimize BIC

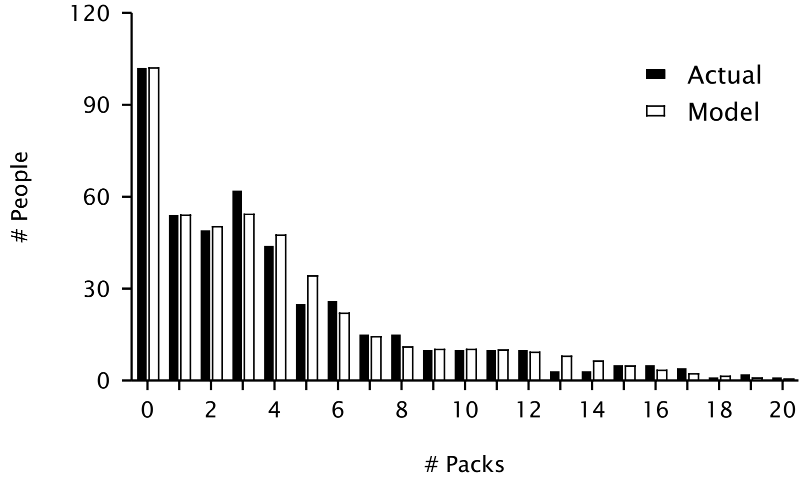
Summary of Model Fit

Model	LL	# params	BIC	χ^2 p-value
NBD	−1140.02	2	2292.29	0.04
ZNBD	−1136.17	3	2290.70	0.08
Poisson	−1545.00	1	3096.12	0.00
2 seg Poisson	−1188.83	3	2396.03	0.00
3 seg Poisson	−1132.04	5	2294.70	0.22
4 seg Poisson	−1130.07	7	2303.00	0.33

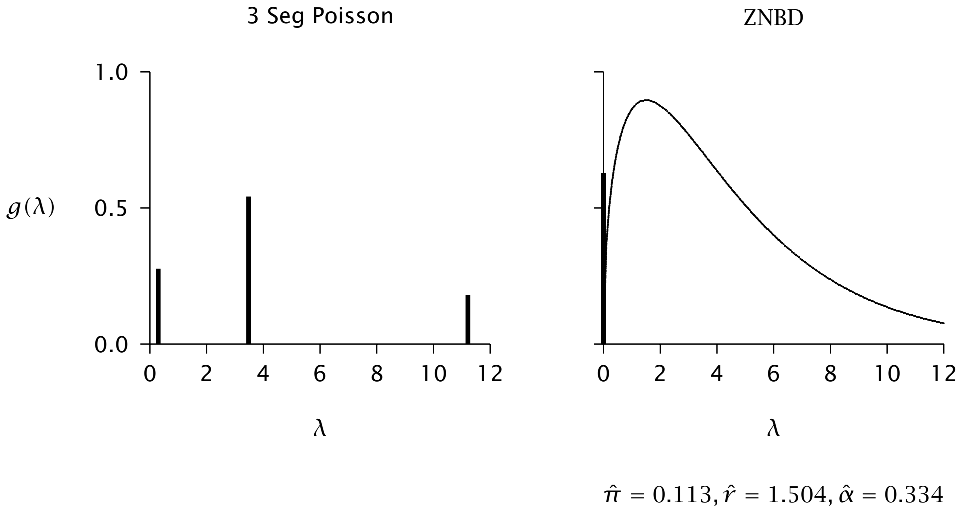
LatentGOLD Results

	Seg 1	Seg 2	Seg 3	Seg 4	*LL*
mean	3.991				−1545.00
class size	1.000				
mean	1.801	9.115			−1188.83
class size	0.700	0.300			
mean	3.483	0.291	11.210		−1132.04
class size	0.542	0.277	0.181		
mean	2.976	0.202	7.247	12.787	−1130.07
class size	0.500	0.243	0.156	0.106	

Fit of the Three-Segment Poisson Model



Implied Heterogeneity Distribution



Classification Using Bayes Theorem

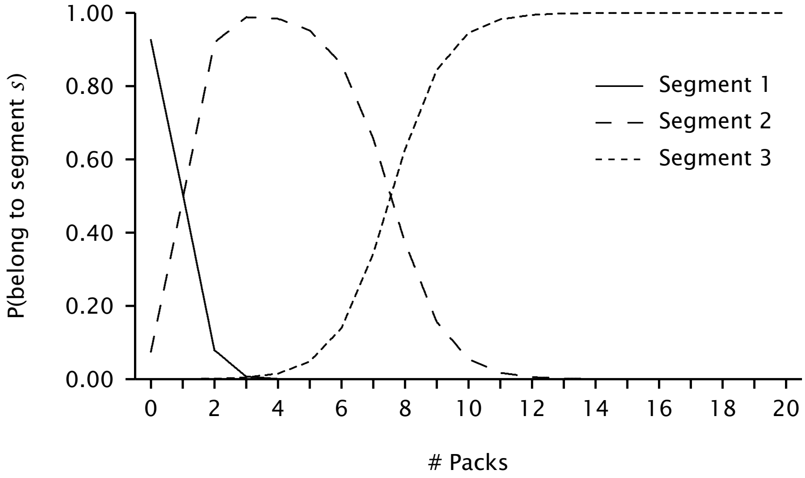
To which "segment" of the mixing distribution does each observation x belong?

- π_s can be interpreted as the prior probability of λ_s
- By Bayes theorem,

$$P(s \mid X = x) = \frac{P(X = x \mid \lambda_s)\pi_s}{\sum_{s'=1}^{S} P(X = x \mid \lambda_{s'})\pi_{s'}},$$

which can be interpreted as the posterior probability of λ_s

Posterior Probabilities



Conditional Expectations

What is the expected purchase quantity over the next month for a customer who purchased seven packs last week?

$$
\begin{aligned}
E[X(4)] &= E[X(4)|\text{seg } 1]\,P(\text{seg } 1|X = 7) \\
&\quad + E[X(4)|\text{seg } 2]\,P(\text{seg } 2|X = 7) \\
&\quad + E[X(4)|\text{seg } 3]\,P(\text{seg } 3|X = 7) \\
&= (4 \times 0.291) \times 0.0000 \\
&\quad + (4 \times 3.483) \times 0.6575 \\
&\quad + (4 \times 11.216) \times 0.3425 \\
&= 24.5
\end{aligned}
$$

… or 13.9 with “hard assignment” to segment 2.

Concepts and Tools Introduced

- Finite mixture models
- Discrete vs. continuous mixing distributions
- Probability models for classification

Further Reading

Dillon, Wiliam R. and Ajith Kumar (1994), "Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview," in Richard P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, Oxford: Blackwell.

McLachlan, Geoffrey and David Peel (2000), *Finite Mixture Models*, New York: John Wiley & Sons.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Problem 6:
Who is Visiting khakichinos.com?

(Incorporating Covariates in Count Models)

Background

Khaki Chinos, Inc. is an established clothing catalog company with an online presence at khakichinos.com. While the company is able to track the online *purchasing* behavior of its customers, it has no real idea as to the pattern of *visiting* behaviors by the broader Internet population.

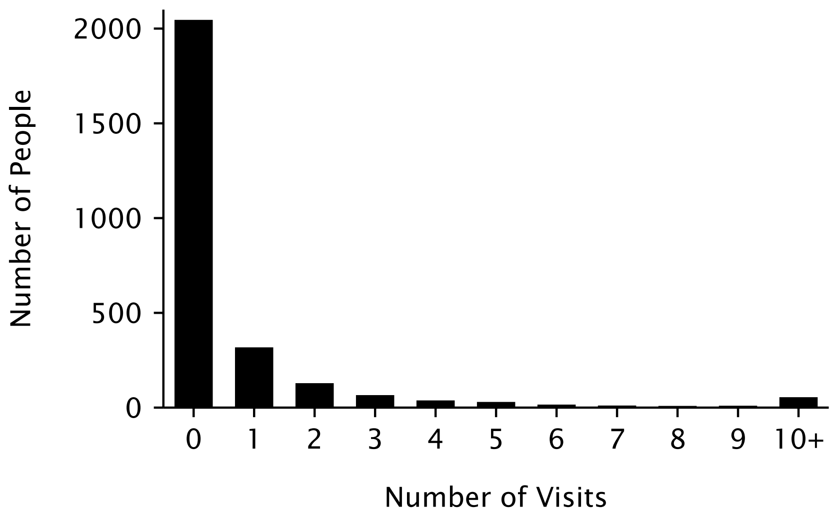
In order to gain an understanding of the aggregate visiting patterns, some Media Metrix panel data has been purchased. For a sample of 2728 people who visited an online apparel site at least once during the second-half of 2000, the dataset reports how many visits each person made to the khakichinos.com web site, along with some demographic information.

Management would like to know whether frequency of visiting the web site is related to demographic characteristics.

Raw Data

ID	# Visits	ln(Income)	Sex	ln(Age)	HH Size
1	0	11.38	1	3.87	2
2	5	9.77	1	4.04	1
3	0	11.08	0	3.33	2
4	0	10.92	1	3.95	3
5	0	10.92	1	2.83	3
6	0	10.92	0	2.94	3
7	0	11.19	0	3.66	2
8	1	11.74	0	4.08	2
9	0	10.02	0	4.25	1
	...				

Distribution of Visits



Modelling Count Data

Recall the NBD:

- At the individual-level, $Y \sim \text{Poisson}(\lambda)$
- λ is distributed across the population according to a gamma distribution with parameters r and α

$$P(Y = y) = \frac{\Gamma(r+y)}{\Gamma(r)y!}\left(\frac{\alpha}{\alpha+1}\right)^r\left(\frac{1}{\alpha+1}\right)^y$$

Observed vs. Unobserved Heterogeneity

Unobserved Heterogeneity:

- People differ in their mean (visiting) rate λ
- To account for heterogeneity in λ, we assume it is distributed across the population according to some (parametric) distribution
- But there is no attempt to *explain* how people differ in their mean rates

Observed Heterogeneity:

- We observe how people differ on a set of observable independent (explanatory) variables
- We explicitly link an individual's λ to her observable characteristics

The Poisson Regression Model

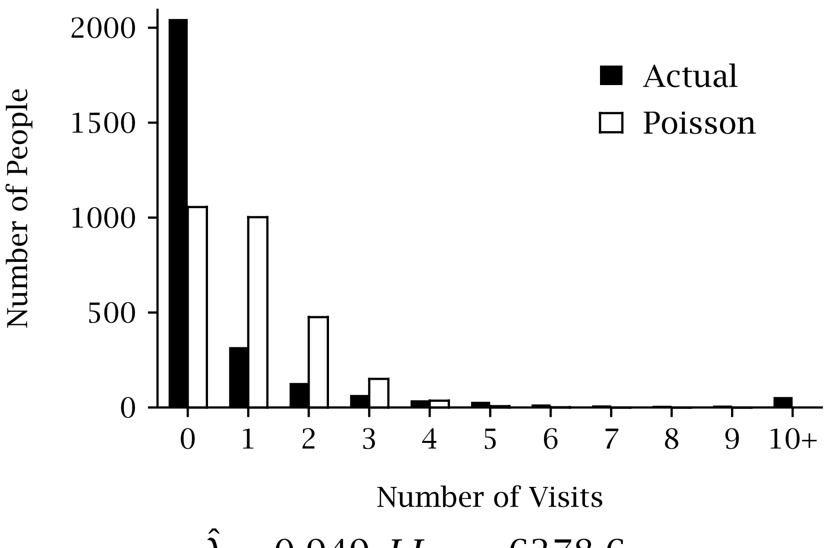
- Let the random variable Y_i denote the number of times individual i visits the site in a unit time period
- At the individual-level, Y_i is assumed to be distributed Poisson with mean λ_i:

$$P(Y_i = y|\lambda_i) = \frac{\lambda_i^y e^{-\lambda_i}}{y!}$$

- An individual's mean is related to her observable characteristics through the function

$$\lambda_i = \lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

Fit of the Poisson Model



$$\hat{\lambda} = 0.949, LL = -6378.6$$

Fitting the Poisson Regression Model

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	0.0439			LL	-6291.497				
2	B_inc	0.0938								
3	B_sex	0.0043								
4	B_age	0.5882								
5	B_size	-0.0359								
6				0.0938	0.0043	0.5882	-0.0359			
7										
8	ID	Total		Income	Sex	Age	HH Size	lambda	P(Y=y)	ln[P(Y=y)]
9	1	0		11.38	1	3.87	2	1.16317	0.31249	-1.163
10	2	5		9.77	1	4.04	1	1.14695	0.00525	-5.249
11	3	0		[illegible]	0	[illegible]	2	0.82031	0.44030	[illegible]20
12	4	0		10.92	1	[illegible]	[illegible]	[illegible]	0.32430	-1.126
13	5	0		10.92	1	2.83	3	0.58338	0.55801	-0.583
14	6	0		10.92	0	2.94	3	0.62017	0.53785	-0.620
15	7	0		11.19	0	3.66	2	1.00712	0.36527	-1.007
16	8	1		11.74	0	4.08	2	1.35220	0.34977	-1.050
17	9	0		10.02	0	4.25	1	1.31954	0.26726	-1.320
18	10	0		10.92	0	3.85	3	1.05656	0.34765	-1.057
2735	2727	0		10.53	1	2.89	4	0.56150	0.57035	-0.561
2736	2728	0		11.74	1	2.83	3	0.63010	0.53254	-0.630

{=TRANSPOSE(B2:B5)}

=B$1*EXP(SUMPRODUCT(D$6:G$6,D9:G9))

=H9^B9*EXP(-H9)/FACT(B9)

=LN(I9)

Poisson Regression Results

Variable	Coefficient
λ_0	0.0439
Income	0.0938
Sex	0.0043
Age	0.5882
HH Size	−0.0359
LL	−6291.5
LL_{Poiss}	−6378.6
LR (df = 4)	174.2

Comparing Expected Visit Behavior

	Person A	Person B
Income	59,874	98,716
Sex	M	F
Age	55	33
HH Size	4	2

Who is less likely to have visited the web site?

$$\begin{aligned}\lambda_A &= 0.0439 \times \exp\big(0.0938 \times \ln(59,874) + 0.0043 \times 0 \\ &\quad + 0.5882 \times \ln(55) - 0.0359 \times 4\big) \\ &= 1.127 \\ \lambda_B &= 0.0439 \times \exp\big(0.0938 \times \ln(98,716) + 0.0043 \times 1 \\ &\quad + 0.5882 \times \ln(33) - 0.0359 \times 2\big) \\ &= 0.944\end{aligned}$$

Is $\boldsymbol{\beta}$ Different from 0?

Consider two models, A and B:

> If we can arrive at model B by placing k constraints on the parameters of model A, we say that model B is *nested* within model A.

The Poisson model is nested within the Poisson regression model by imposing the constraint $\boldsymbol{\beta} = \mathbf{0}$.

We use the *likelihood ratio test* to determine whether model A, which has more parameters, fits the data better than model B.

The Likelihood Ratio Test

- The null hypothesis is that model A is not different from model B
- Compute the test statistic

$$LR = -2(LL_B - LL_A)$$

- Reject null hypothesis if $LR > \chi^2_{.05,k}$

Computing Standard Errors

- Excel
  - indirectly via a series of likelihood ratio tests
  - easily computed from the Hessian matrix (computed using difference approximations)
- General modelling environments (e.g., MATLAB, Gauss)
  - easily computed from the Hessian matrix (as a by-product of optimization or computed using difference approximations)
- Advanced statistics packages (e.g., Limdep, R, S-Plus)
  - they come for free

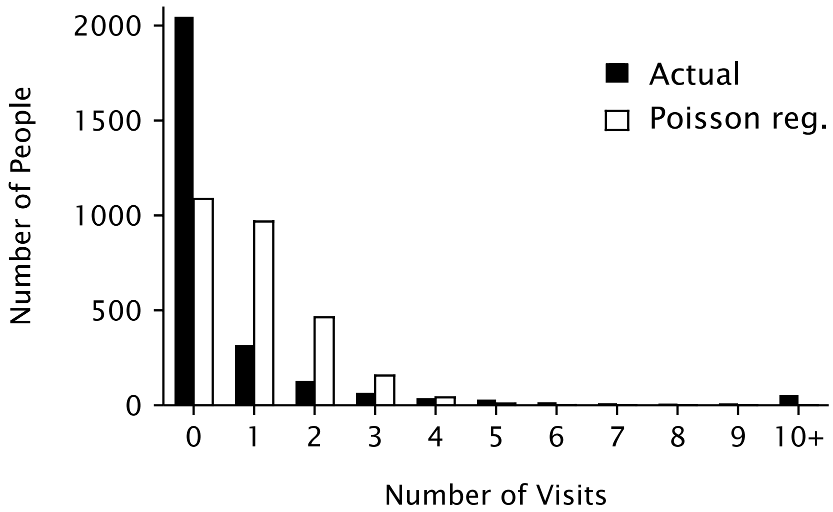
S-Plus Poisson Regression Results

```
Coefficients:
                   Value Std. Error    t value
(Intercept) -3.126238804 0.40578080  -7.7042552
     Income  0.093828021 0.03436347   2.7304580
        Sex  0.004259338 0.04089411   0.1041553
        Age  0.588249213 0.05472896  10.7484079
    HH Size -0.035907406 0.01528397  -2.3493511
```


Limdep Poisson Regression Results

```
+---------+--------------+----------------+--------+
|Variable | Coefficient  | Standard Error |b/St.Er.|
+---------+--------------+----------------+--------+
 Constant -3.122103284      .40565119       -7.697
 INCOME     .9305546493E-01  .34332533E-01    2.710
 SEX        .4312514407E-02  .40904265E-01     .105
 AGE        .5893014445      .54790230E-01   10.756
 HH SIZE  -.3577795361E-01  .15287122E-01   -2.340
```

Fit of the Poisson Regression



The ZIP Regression Model

Because of the "excessive" number of zeros, let us consider the zero-inflated Poisson (ZIP) regression model:

- a proportion π of those people who go to online apparel sites will never visit khakichinos.com
- the visiting behavior of the "ever visitors" can be characterized by the Poisson regression model

$$P(Y_i = y) = \delta_{y=0}\pi + (1-\pi) \times \frac{[\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)]^y e^{-\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)}}{y!}$$

Fitting the ZIP Regression Model

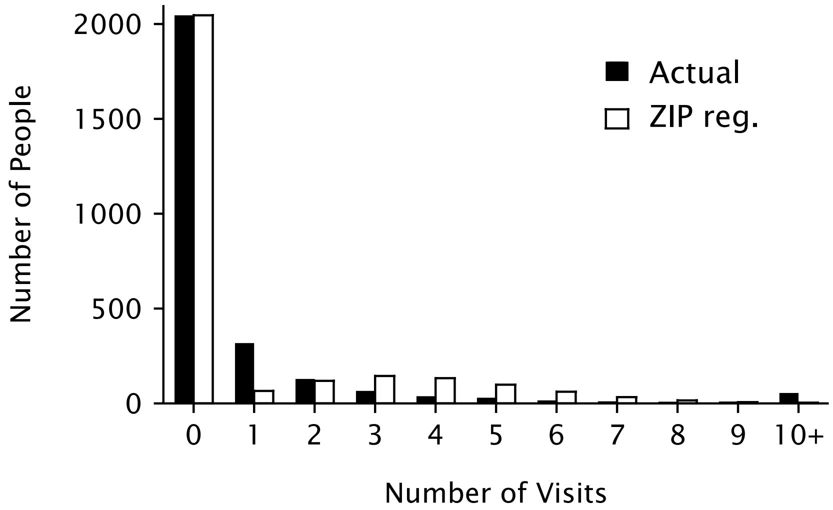
	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	6.6231			LL	-4297.472				
2	pi	0.7433								
3	B_inc	-0.0891								
4	B_sex	-0.1327								
5	B_age	0.1141								
6	B_size	0.0196								
7				-0.0891	-0.1327	0.1141	0.0196			
8										
9	ID	Total		Income	Sex	Age	HH Size	lambda	P(Y=y)	ln[P(Y=y)]
10	1	0		11.38	1	3.87	2	3.40193	0.75184	-0.285
11	2	5		9.77	1	4.04	1	3.92698	0.03936	-3.235
12	3	0		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]2	-0.289
13	4	0		10.92	1	3.95	3	3.64889	0.74996	-0.288
14	5	0		10.92	1	2.83	3	3.21182	0.75363	-0.283
15	6	0		10.92	0	2.94	3	3.71435	0.74954	-0.288
16	7	0		11.19	0	3.66	2	3.85775	0.74871	-0.289
17	8	1		11.74	0	4.08	2	3.85266	0.02099	-3.864
18	9	0		10.02	0	4.25	1	4.48880	0.74617	-0.293
19	10	0		10.92	0	3.85	3	4.11879	0.74746	-0.291
2736	2727	0		10.53	1	2.89	4	3.41119	0.75176	-0.285
2737	2728	0		11.74	1	2.83	3	2.98515	0.75626	-0.279

=IF(B10=0,B$2,0)+(1-B$2)*H10^B10*EXP(-H10)/FACT(B10)

ZIP Regression Results

Variable	Coefficient
λ_0	6.6231
Income	−0.0891
Sex	−0.1327
Age	0.1141
HH Size	0.0196
π	0.7433
LL	−4297.5
$LL_{\text{Poiss reg}}$	−6291.5
LR (df = 1)	3988.0

Fit of the ZIP Regression



NBD Regression

The explanatory variables may not fully capture the differences among individuals

To capture the remaining (unobserved) component of differences among individuals, let λ_0 vary across the population according to a gamma distribution with parameters r and α:

$$P(Y_i = y) = \frac{\Gamma(r + y)}{\Gamma(r)y!}\left(\frac{\alpha}{\alpha + \exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right)^r\left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\alpha + \exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right)^y$$

- Known as the “Negbin II” model in most textbooks
- Collapses to the NBD when $\boldsymbol{\beta} = \mathbf{0}$

Fitting the NBD Regression Model

	A	B	C	D	E	F	G	H	I	J
1	r	0.1388			LL	-2888.966				
2	alpha	8.1979								
3	B_inc	0.0734								
4	B_sex	-0.0093								
5	B_age	0.9022								
6	B_size	-0.0243								
7				0.0734	-0.0093	0.9022	-0.0243			
8										
9	ID	Total		Income	Sex	Age	HH Size	exp(BX)	P(Y=y)	ln[P(Y=y)]
10	1	0		11.38	1	3.87	2	71.51161	0.72936	-0.316
11	2	5		9.77	1	4.04	1	76.02589	0.01587	-4.143
12	3	0		[illegible]	[illegible]	[illegible]	[illegible]	43.42559	0.77467	-0.255
13	4	0		[illegible]	[illegible]	[illegible]	[illegible]	72.50603	0.72810	-0.317
14	5	0		10.92	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	6	0		10.92	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
16	7	0		11.19	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
17	8	1		11.74	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
18	9	0		10.02	0	4.25	1	94.07931	0.70456	-0.350
19	10	0		10.92	0	3.85	3	66.80224	0.73555	-0.307
2736	2727	0		10.53	1	2.89	4	26.42093	0.81883	-0.200
2737	2728	0		11.74	1	2.83	3	28.08647	0.81351	-0.206

=EXP(SUMPRODUCT(D$7:G$7,D10:G10))

=EXP(GAMMALN(B$1+B10)-GAMMALN(B$1))/FACT(B10)*(B$2/(B$2+H10))^B$1*(H10/(B$2+H10))^B10

NBD Regression Results

Variable	Coefficient
r	0.1388
α	8.1979
Income	0.0734
Sex	−0.0093
Age	0.9022
HH Size	−0.0243
LL	−2889.0

S-Plus NBD Regression Results

```
Coefficients:
                   Value Std. Error    t value
(Intercept) -4.047149702 1.10159557 -3.6738979
     Income  0.074549233 0.09555222  0.7801936
        Sex -0.005240835 0.11592793 -0.0452077
        Age  0.889862966 0.14072030  6.3236289
    HH Size -0.025094493 0.04187696 -0.5992435

              Theta:  0.13878
          Std. Err.:  0.00726
```

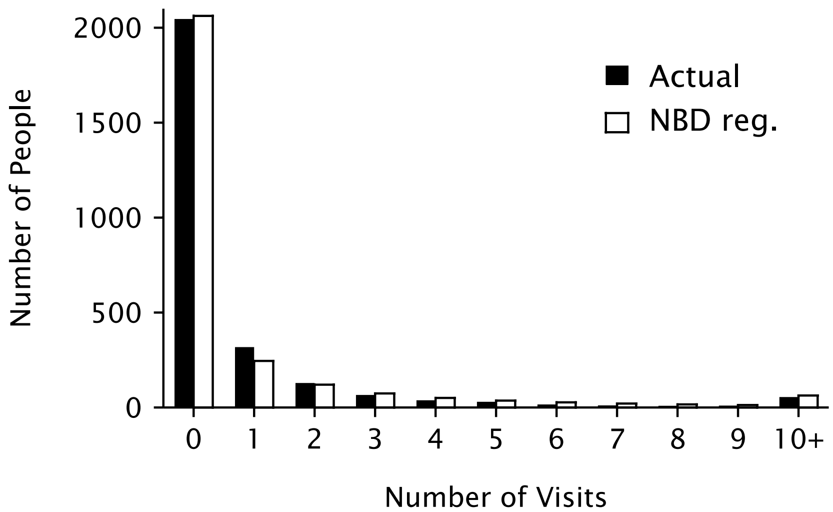

Limdep NBD Regression Results

```
+---------+-------------+---------------+--------+
|Variable | Coefficient  | Standard Error |b/St.Er.|
+---------+-------------+---------------+--------+
 Constant -4.077239653      1.0451741       -3.901
 INCOME    .7237686001E-01  .76663437E-01     .944
 SEX      -.9009160129E-02  .11425700        -.079
 AGE       .9045111135      .17741724        5.098
 HH SIZE  -.2406546843E-01  .38695426E-01    -.622
          Overdispersion parameter
 Alpha     7.206708844      .33334006       21.620
```

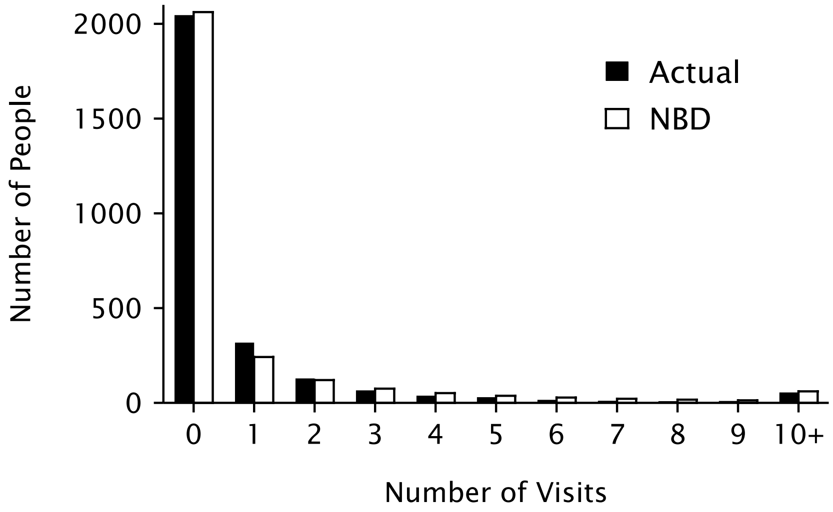
Summary of Regression Results

Variable	Poisson	ZIP	NBD
λ_0	0.0439	6.6231	
r			0.1388
α			8.1979
Income	0.0938	−0.0891	0.0734
Sex	0.0043	−0.1327	−0.0093
Age	0.5882	0.1141	0.9022
HH Size	−0.0359	0.0196	−0.0243
π		0.7433	
LL	−6291.5	−4297.5	−2889.0

Fit of the NBD Regression



Fit of the NBD



$\hat{r} = 0.134, \hat{\alpha} = 0.141, LL = -2905.6$

Concepts and Tools Introduced

- Incorporating covariate effects in count models
- Poisson (and NBD) regression models
- The possible over-emphasis of the value of covariates

Further Reading

Cameron, A. Colin and Pravin K. Trivedi (1998), *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Winkelmann, Rainer (2003), *Econometric Analysis of Count Data*, 4th edn., Berlin: Springer.

Introducing Covariates: The General Case

- Select a probability distribution that characterizes the individual-level behavior of interest:

$$f(y|\theta_i)$$

- Make the individual-level latent characteristic(s) a function of (time-invariant) covariates:

$$\theta_i = s(\theta_0, \mathbf{x}_i)$$

- Specify a mixing distribution to capture the heterogeneity in θ_i not "explained" by $\mathbf{x}_i$
- Derive the corresponding aggregate distribution

$$f(y|\mathbf{x}_i) = \int f(y|\theta_0, \mathbf{x}_i) g(\theta_0)\, d\theta_0$$

Covariates in Timing Models

- If the covariates are time-invariant, we can make λ a direct function of covariates:

$$F(t) = 1 - e^{-\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)t}$$

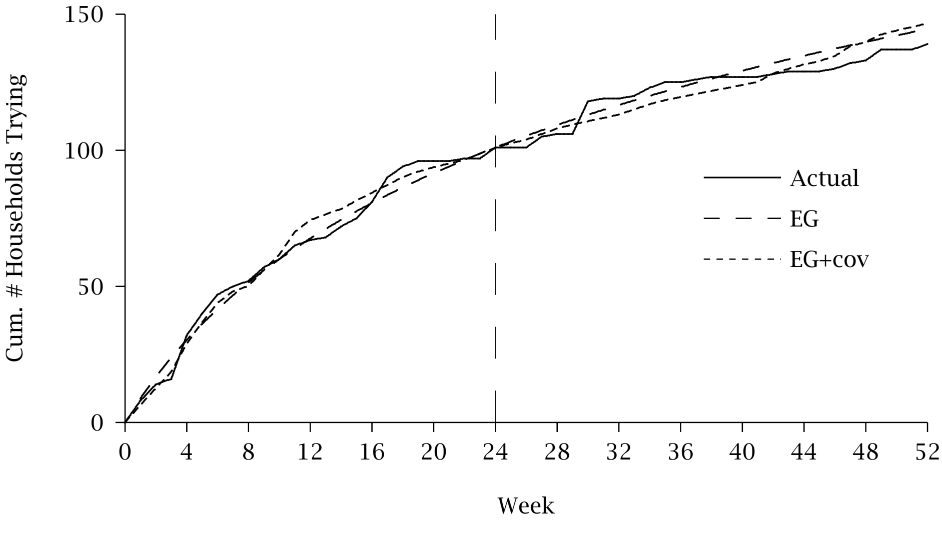
- If the covariates are time-varying (i.e., $\mathbf{x}_{it}$), we incorporate their effects via the hazard rate function

$$F(t) = 1 - e^{-\lambda_0 A(t)}$$

  where $A(t) = \sum_{j=1}^{t} \exp(\boldsymbol{\beta}'\mathbf{x}_{ij})$

- Known as "proportional hazards regression"

Comparing EG with EG+cov



Covariates in “Choice” Models

Two options for binary choice:

- The beta-logistic model
  - a generalization of the beta-binomial model in which the mean is made a function of (time-invariant) covariates
  - covariate effects not introduced at the level of the individual
- Finite mixture of binary logits:

$$P(Y = 1) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\exp(\boldsymbol{\beta}'\mathbf{x}_i) + 1}$$

with some elements of $\boldsymbol{\beta}$ varying across segments

Discussion

Recap

- The preceding five problems introduce simple models for three behavioral processes:
  - Timing ⟶ "when"
  - Counting ⟶ "how many"
  - "Choice" ⟶ "whether/which"
- Each of these simple models has multiple applications.

Further Applications: Timing Models

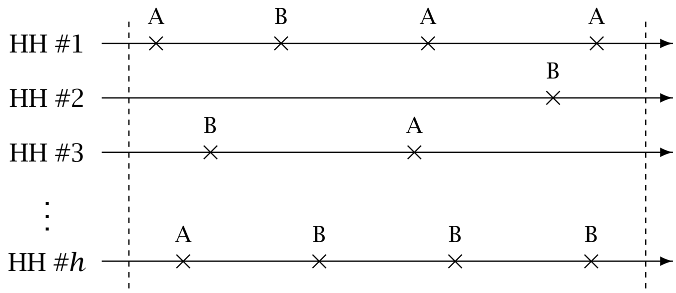
- Repeat purchasing of new products
- Response times:
  - Coupon redemptions
  - Survey response
  - Direct mail (response, returns, repeat sales)
- Other durations:
  - Salesforce job tenure
  - Length of web site browsing session

Further Applications: Count Models

- Repeat purchasing
- Customer concentration ("80/20" rules)
- Salesforce productivity/allocation
- Number of page views during a web site browsing session

Further Applications: "Choice" Models

- Brand choice



- Media exposure
- Multibrand choice (BB → Dirichlet Multinomial)
- Taste tests (discrimination tests)
- "Click-through" behavior

The Excel spreadsheets associated with this tutorial, along with electronic copies of the tutorial materials, can be found at:

`http://brucehardie.com/talks.html`

Day 2
Models for Customer-Base Analysis

Agenda

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

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?”

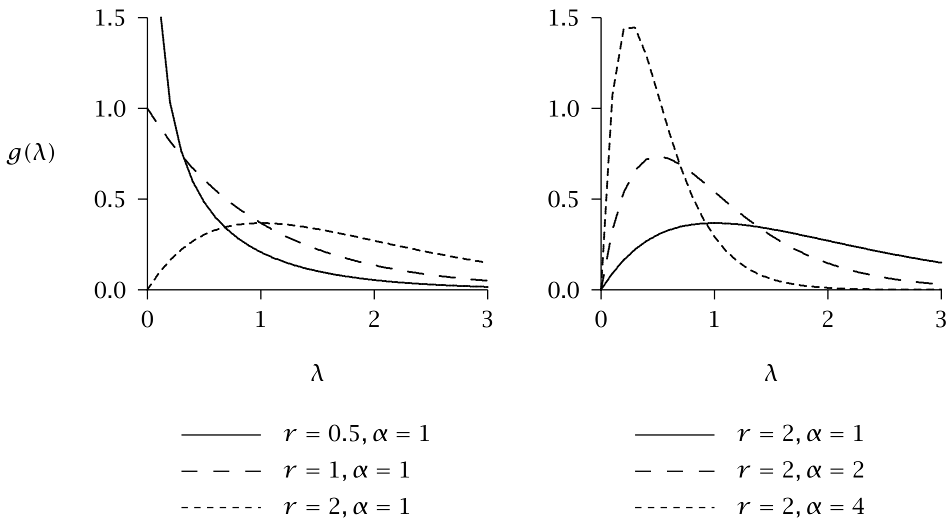
Capturing Heterogeneity in Latent Traits

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:

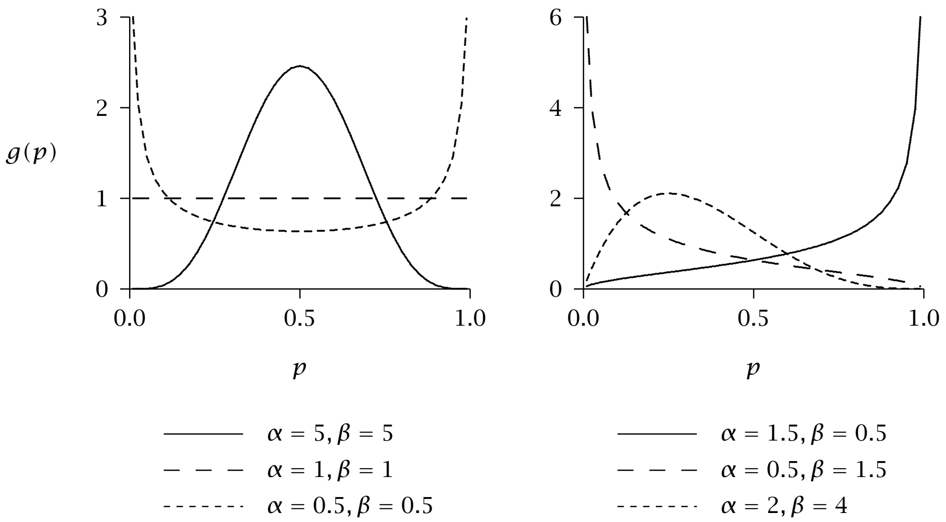
$$g(p \mid \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,,\ 0 < p < 1\,.$$

- $B(\alpha,\beta)$ 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 intermediate result of the form

$$= \text{constant} \times \int_0^1 t^{\alpha}(1-t)^{\beta}(u+vt)^{-\gamma} dt$$

The "Trick" for Integrated Models

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

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

where ${}_2F_1(a, b; c; z)$ is the Gaussian hypergeometric function.

The Gaussian Hypergeometric Function

$$ {}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!} $$

Easy to compute, albeit tedious, in Excel as

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

using the recursion

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

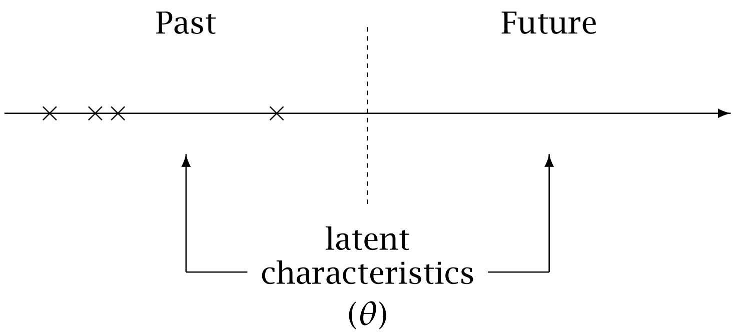
where $u_0 = 1$.

Customer-Base Analysis

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

Comparison of Modelling Approaches

Traditional approach
future = f(past)



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

Classifying Analysis Settings

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

- Based on numbers presented in a January 2006 press release that reported Vodafone Group Plc’s third quarter key performance indicators, we see that Vodafone UK has 6.3 million “pay monthly” customers.
- In his “Q3 2005 Financial Results Conference Call”, the CFO of Amazon made the comment that “[a]ctive customer accounts, representing customers who ordered in the past year, surpassed 52 million, up 19%”.

Classifying Analysis Settings

- 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?

Classifying Analysis Settings

Consider the following four specific business settings:

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

Classifying Customer Bases

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

Type of Relationship With Customers

The Right Way to Think About Computing Customer Lifetime Value

Calculating CLV

Customer lifetime value is *the present value of the future cash flows associated with the customer*.

- A forward-looking concept
- Not to be confused with (historic) customer profitability

Calculating CLV

Standard classroom formula:

$$CLV = \sum_{t=0}^{T} m \frac{r^t}{(1+d)^t}$$

where	m	=	net cash flow per period (if active)
	r	=	retention rate
	d	=	discount rate
	T	=	horizon for calculation

Calculating $E(CLV)$

A more correct starting point:

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

where $E[v(t)]$ = expected value (or net cashflow) of the customer at time t (if active)

$S(t)$ = the probability that the customer has remained active to at least time t

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

Calculating $E(CLV)$

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

Models for Contractual Settings

Classifying Customer Bases

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

Type of Relationship With Customers

SUNIL GUPTA, DONALD R. LEHMANN, and JENNIFER AMES STUART*

It is increasingly apparent that the financial value of a firm depends on off-balance-sheet intangible assets. In this article, the authors focus on the most critical aspect of a firm: its customers. Specifically, they demonstrate how valuing customers makes it feasible to value firms, including high-growth firms with negative earnings. The authors define the value of a customer as the expected sum of discounted future earnings. They demonstrate their valuation method by using publicly available data for five firms. They find that a 1% improvement in retention, margin, or acquisition cost improves firm value by 5%, 1%, and .1%, respectively. They also find that a 1% improvement in retention has almost five times greater impact on firm value than a 1% change in discount rate or cost of capital. The results show that the linking of marketing concepts to shareholder value is both possible and insightful.

Valuing Customers

THE CUSTOMER LIFETIME VALUE CONCEPT AND ITS CONTRIBUTION TO CORPORATE VALUATION

*by Hans H. Bauer, Maik Hammerschmidt and Matthias Braehler**

ABSTRACT

The shareholder value and the customer lifetime value approach are conceptually and methodically analogous. Both concepts calculate the value of a particular decision unit by discounting the forecasted net cash flows by the risk-adjusted cost of capital. However, virtually no scholarly attention has been devoted to the question if any of the components of the shareholder value could be determined in a more market-oriented way using individual customer lifetime values. Therefore, the main objective of this paper is to systematically explore the contribution of both concepts to the field of corporate valuation.

At first we present a comprehensive calculation method for estimating both the individual lifetime value of a customer and the customer equity. After a critical examination of the shareholder value concept, a synthesis of both value approaches allowing for a disaggregated and more realistic corporate valuation will be presented.

Hypothetical Contractual Setting

Number of active customers each year by year-of-acquisition cohort:

2001	2002	2003	2004	2005
10,000	6,334	4,367	3,264	2,604
	10,000	6,334	4,367	3,264
		10,000	6,334	4,367
			10,000	6,334
				10,000
10,000	16,334	20,701	23,965	26,569

Hypothetical Contractual Setting

Assume

- Each contract is annual, starting on January 1 and expiring at 11:59pm on December 31.
- An average net cashflow of $100/year.
- A 10% discount rate

What is the expected residual value of the customer base at December 31, 2005?

Hypothetical Contractual Setting

Aggregate retention rate:

$$\frac{2{,}604 + 3{,}264 + 4{,}367 + 6{,}334}{3{,}264 + 4{,}367 + 6{,}334 + 10{,}000} = 0.691$$

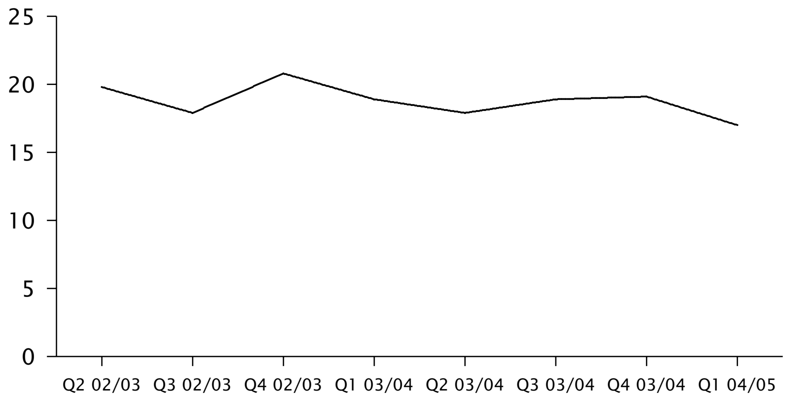
Expected residual value of the customer base at December 31, 2005:

$$26{,}569 \times \sum_{t=1}^{\infty} \$100 \times \frac{0.691^t}{(1+0.1)^{t-1}} = \$4{,}945{,}049$$

Annual Retention Rates by Cohort

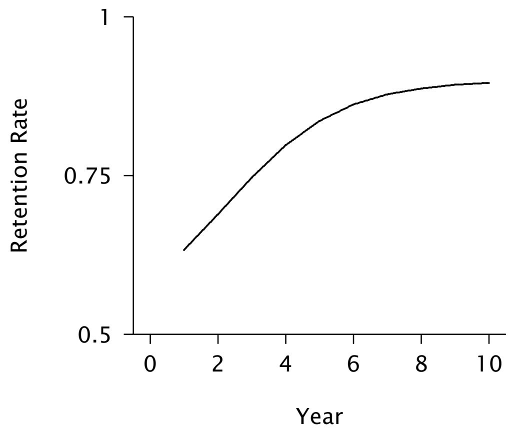
2001	2002	2003	2004	2005
--	0.633	0.689	0.747	0.798
	--	0.633	0.689	0.747
		--	0.633	0.689
			--	0.633
				--
--	0.633	0.655	0.675	0.691

Vodafone Germany
Quarterly Annualized Churn Rate (%)



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

A Real-World Consideration



At the cohort level, we (almost) always observe increasing retention rates.

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

Fielding, Michael (2005), "Get Circulation Going: DM Redesign Increases Renewal Rates for Magazines," *Marketing News*, September 1, 9–10.

Key Considerations

- Need to recognize inter-cohort differences (at any point in time) when valuing a customer base.
- Need to project retention beyond the set of observed retention rates.
- 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.

  A sorting effect in a heterogeneous population.

The Role of Heterogeneity

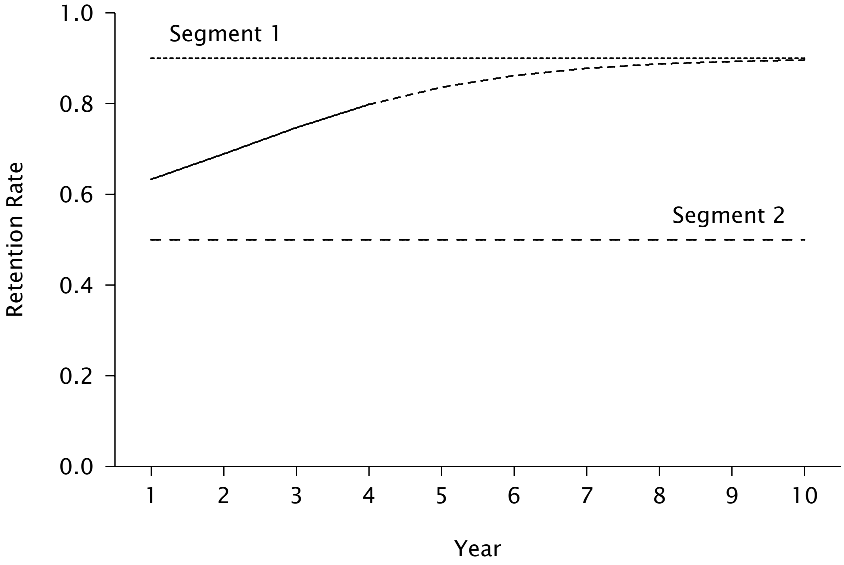
Suppose we track a cohort of 10,000 customers, comprising two underlying segments:

- Segment 1 comprises one-third of the customers, each with a time-invariant annual retention probability of 0.9.
- Segment 2 comprises two-thirds of the customers, each with a time-invariant annual retention probability of 0.5.

The Role of Heterogeneity

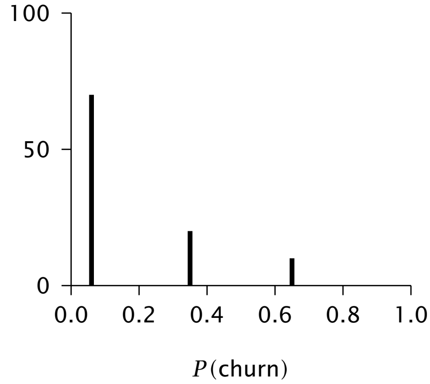
	# Active Customers			r_t		
Year	Seg 1	Seg 2	Total	Seg 1	Seg 2	Total
1	3,333	6,667	10,000			
2	3,000	3,334	6,334	0.900	0.500	0.633
3	2,700	1,667	4,367	0.900	0.500	0.689
4	2,430	834	3,264	0.900	0.500	0.747
5	2,187	417	2,604	0.900	0.500	0.798

The Role of Heterogeneity



Vodafone Italia Churn Clusters

Cluster	P(churn)	%CB
Low risk	0.06	70
Medium risk	0.35	20
High risk	0.65	10



Source: "Vodafone Achievement and Challenges in Italy" presentation (2003-09-12)

$E(RLV)$ of an Active 2001 Cohort Member

- If this person belongs to segment 1:

$$
\begin{aligned}
E(RLV) &= \sum_{t=1}^{\infty} 100 \times \frac{0.9^t}{(1+0.1)^{t-1}} \\
&= \$495
\end{aligned}
$$

- If this person belongs to segment 2:

$$
\begin{aligned}
E(RLV) &= \sum_{t=1}^{\infty} 100 \times \frac{0.5^t}{(1+0.1)^{t-1}} \\
&= \$92
\end{aligned}
$$

$E(RLV)$ of an Active 2001 Cohort Member

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

$$
\frac{P(\text{renewed contract four times} \mid \text{segment 1}) \times P(\text{segment 1})}{P(\text{renewed contract four times})}
$$

$$
\begin{aligned}
&= \frac{0.9^4 \times 0.333}{0.9^4 \times 0.333 + 0.5^4 \times 0.667} \\
&= 0.84
\end{aligned}
$$

$$
\Longrightarrow E(RLV) = 0.84 \times \$495 + (1 - 0.84) \times \$92 = \$430
$$

Valuing the Existing Customer Base

Recognizing the underlying segments:

Cohort	# Active in 2005	$P(\text{seg. }1)$	$E(RLV)$
2001	2,604	0.840	$430
2002	3,264	0.745	$392
2003	4,367	0.618	$341
2004	6,334	0.474	$283
2005	10,000	0.333	$226

Total expected residual value = $7,940,992

Valuing the Existing Customer Base

Cohort	Total RV	Underestimation
Naïve	$4,945,049	38%
Segment (model)	$7,940,992	

Exploring the Magnitude of the Error

- Systematically vary heterogeneity in retention rates
- First need to specify (and validate) a flexible model of contract duration

A Discrete-Time Model for Contract Duration

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

$\longrightarrow$ the duration of the customer's relationship with the firm is characterized by the (shifted) geometric distribution:

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

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

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

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

$$\begin{aligned} r_t &= \frac{S(t)}{S(t-1)} \\ &= \frac{\beta + t - 1}{\alpha + \beta + t - 1}. \end{aligned}$$

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

Computing $E(CLV)$

- Recall:

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

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

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

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

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

Computing DERL

- Standing at the end of period n, just prior to the point in time at which the customer makes her contract renewal decision,

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

- But θ is unobserved ....

Computing DERL

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

$$f(\theta \mid \alpha, \beta, n-1 \text{ renewals}) = \frac{S(n-1 \mid \theta) f(\theta \mid \alpha, \beta)}{S(n \mid \alpha, \beta)} = \frac{\theta^{\alpha-1}(1-\theta)^{\beta+n-2}}{B(\alpha, \beta+n-1)}$$

- It follows that

$$DERL(d \mid \alpha, \beta, n-1 \text{ renewals}) = \left(\frac{\beta+n-1}{\alpha+\beta+n-1}\right) {}_2F_1(1, \beta+n; \alpha+\beta+n; \tfrac{1}{1+d})$$

Computing DERL

Alternative derivation:

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

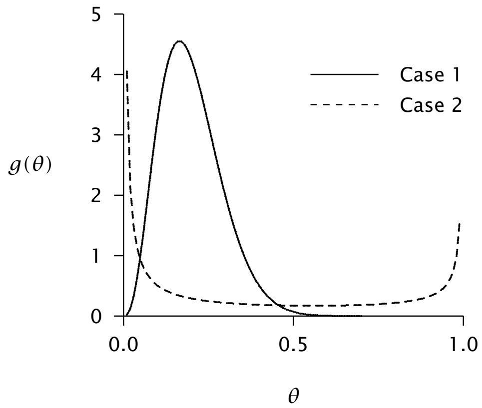
Impact of Heterogeneity on Error

- Assume the following arrival of new customers:

2001	2002	2003	2004	2005
10,000	10,000	10,000	10,000	10,000

- Assume $\bar{v} = \$1$ and a 10% discount rate.
- For given values of α and β, determine the error associated with computing the residual value of the existing customer base using the naïve approach (a constant aggregate retention rate) compared with the "correct" model-based approach.

Two Scenarios



Case	α	β	$E(\theta)$	$S(1)$	$S(2)$	$S(3)$	$S(4)$
1	3.80	15.20	0.20	0.800	0.684	0.531	0.439
2	0.067	0.267	0.20	0.800	0.760	0.738	0.724

Computing DERL Using Excel

Recall our alternative derivation:

$$DERL(d \mid \alpha, \beta, n-1 \text{ renewals}) = \sum_{t=n}^{\infty} \frac{S(t \mid \alpha, \beta)}{S(n-1 \mid \alpha, \beta)} \Big(\frac{1}{1+d}\Big)^{t-n}$$

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

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

Calculating DERL

	A	B	C	D	E	F
1	alpha	3.8	**DERL**	3.59		
2	beta	15.2				
3				2 renewals (n=3)		
4	t	S(t)		S(t|t>n-1)	disc.	
5	0	1.0000				
6	1	0.8000				
7	2	0.6480				
8	3	0.5307		0.8190	1.0000	
9	[illegible]	[illegible]		0.6776	0.9091	
10	[illegible]	[illegible]		0.5656	0.8264	
11	6	0.3085		[illegible]	[illegible]	
12	7	0.2616		[illegible]	[illegible]	
13	8	0.2234		0.3447	0.6209	
14	9	0.1919		0.2962	0.5645	
15	10	0.1659		0.2560	0.5132	
204	199	5.82E-05		8.98722E-05	7.71E-09	
205	200	5.72E-05		8.83056E-05	7.01E-09	

=SUMPRODUCT(D6:D205,E6:E205)

=B8/$B$7

=(B$2+A6-1)/(B$1+B$2+A6-1)*B5

=1/1.1^($A9-3)

Number of Active Customers: Case 1

2001	2002	2003	2004	2005	n	$E(RLV)$
10,000	8,000	6,480	5,307	4,391	5	$3.84
	10,000	8,000	6,480	5,307	4	$3.72
		10,000	8,000	6,480	3	$3.59
			10,000	8,000	2	$3.45
				10,000	1	$3.31
10,000	18,000	24,480	29,787	34,178		

Aggregate 04–05 retention rate = 24,178/29,787 = 0.81

Impact of Heterogeneity on Error: Case 1

$$\begin{aligned}\text{Naïve valuation} &= 34{,}178 \times \sum_{t=1}^{\infty} \frac{0.81^t}{(1+0.1)^{t-1}} \\ &= \$105{,}845\end{aligned}$$

$$\begin{aligned}\text{Correct valuation} = &\ 4{,}391 \times \$3.84 + 5{,}307 \times \$3.72 \\ &+ 6{,}480 \times \$3.59 + 8{,}000 \times \$3.45 \\ &+ 10{,}000 \times \$3.31 \\ = &\ \$120{,}543\end{aligned}$$

Naïve underestimates correct by 12%.

Number of Active Customers: Case 2

2001	2002	2003	2004	2005	n	$E(RLV)$
10,000	8,000	7,600	7,383	7,235	5	\$10.19
	10,000	8,000	7,600	7,383	4	\$10.06
		10,000	8,000	7,600	3	\$9.86
			10,000	8,000	2	\$9.46
				10,000	1	\$7.68
10,000	18,000	25,600	32,983	40,218		

Aggregate 04–05 retention rate = 30,218/32,983 = 0.92

Impact of Heterogeneity on Error: Case 2

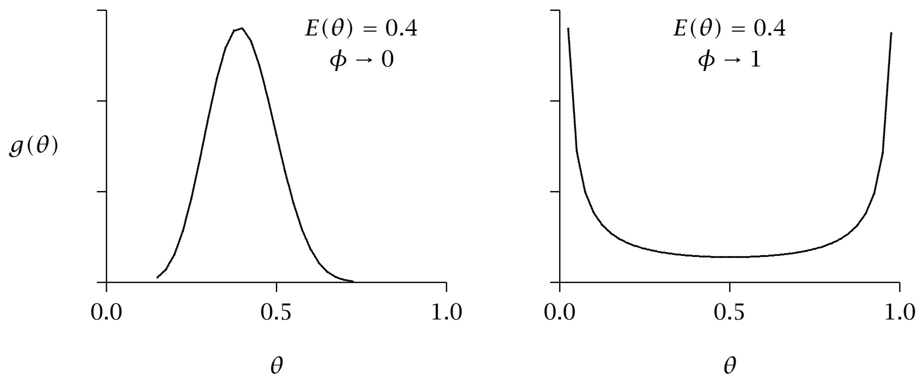
$$\begin{aligned}\text{Naïve valuation} &= 40{,}218 \times \sum_{t=1}^{\infty} \frac{0.92^t}{(1+0.1)^{t-1}} \\ &= \$220{,}488\end{aligned}$$

$$\begin{aligned}\text{Correct valuation} &= 7{,}235 \times \$10.19 + 7{,}383 \times \$10.06 \\ &\quad + 7{,}600 \times \$9.86 + 8{,}000 \times \$9.46 \\ &\quad + 10{,}000 \times \$7.68 \\ &= \$375{,}437\end{aligned}$$

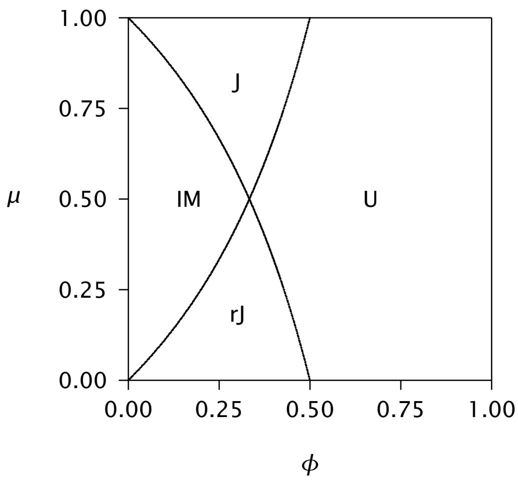
Naïve underestimates correct by 41%.

Interpreting the Beta Distribution Parameters

$$\text{mean } \mu = \frac{\alpha}{\alpha+\beta} \text{ and polarization index } \phi = \frac{1}{\alpha+\beta+1}$$

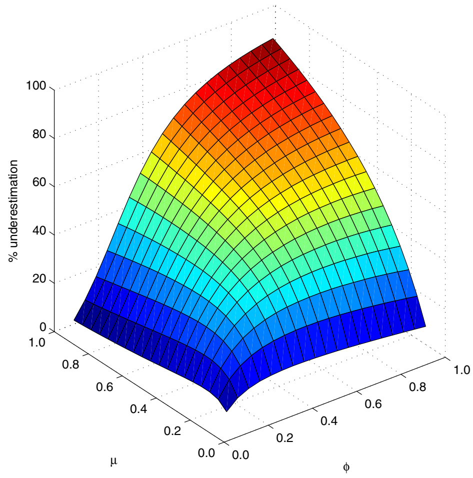


Shape of the Beta Distribution

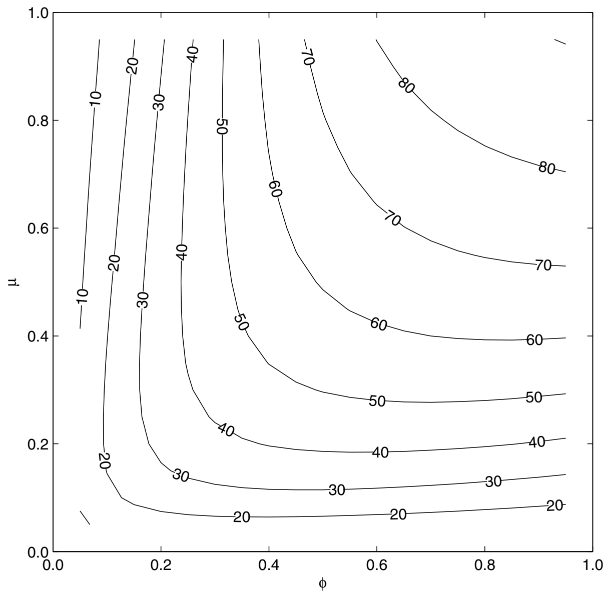


Error as a Function of μ and ϕ

For a fine grid of points in the (μ, ϕ) space, we determine the corresponding values of (α, β) and compute % underestimation:



Error as a Function of μ and ϕ



SUNIL GUPTA, DONALD R. LEHMANN, and JENNIFER AMES STUART*

It is increasingly apparent that the financial value of a firm depends on off-balance-sheet intangible assets. In this article, the authors focus on the most critical aspect of a firm: its customers. Specifically, they demonstrate how valuing customers makes it feasible to value firms, including high-growth firms with negative earnings. The authors define the value of a customer as the expected sum of discounted future earnings. They demonstrate their valuation method by using publicly available data for five firms. They find that a 1% improvement in retention, margin, or acquisition cost improves firm value by 5%, 1%, and .1%, respectively. They also find that a 1% improvement in retention has almost five times greater impact on firm value than a 1% change in discount rate or cost of capital. The results show that the linking of marketing concepts to shareholder value is both possible and insightful.

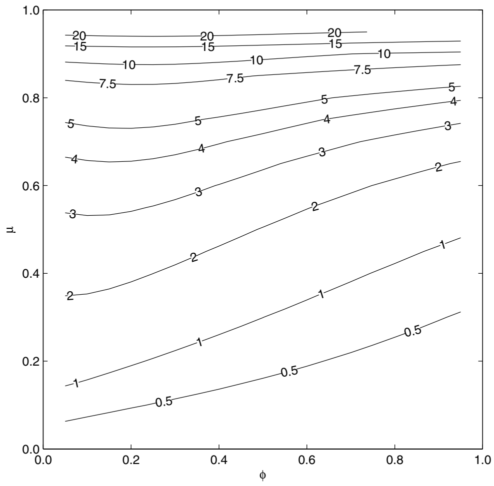
Valuing Customers

Retention Elasticities

- Widely-held belief that improvement in customer retention can have a major impact on customer (and therefore firm) value (Reichheld 1996).
- Gupta et al. (2004) report an average retention elasticity of 5.
- What happens when we recognize heterogeneity-induced dynamics in retention rates?

Retention Elasticities as a Function of μ and ϕ

We determine the retention elasticity for the values of α and β associated with each point on the (μ, ϕ) unit square:



Re-analysis Using (r_1, r_2)

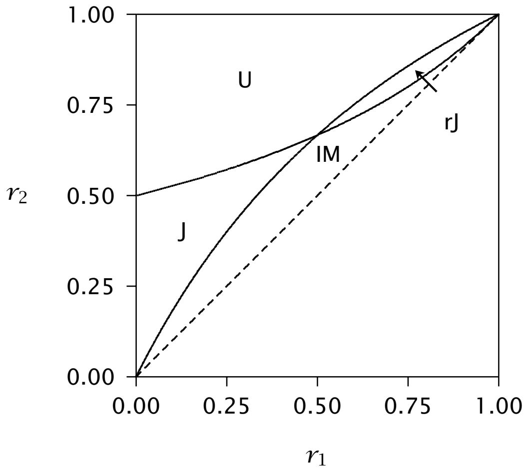
- μ and ϕ are not quantities that most managers or analysts think about; retention rates are easier to comprehend.
- Since the period 1 and 2 retention rates are, respectively,

$$r_1 = \frac{\beta}{\alpha + \beta} \text{ and } r_2 = \frac{\beta + 1}{\alpha + \beta + 1},$$

it follows that

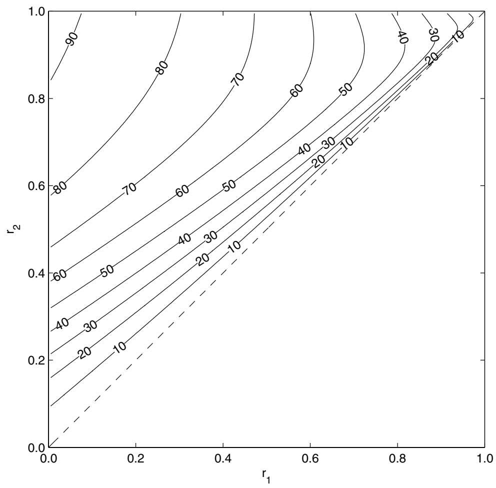
$$\alpha = \frac{(1 - r_1)(1 - r_2)}{r_2 - r_1} \text{ and } \beta = \frac{r_1(1 - r_2)}{r_2 - r_1}.$$

Shape of the Beta Distribution (r_1, r_2)



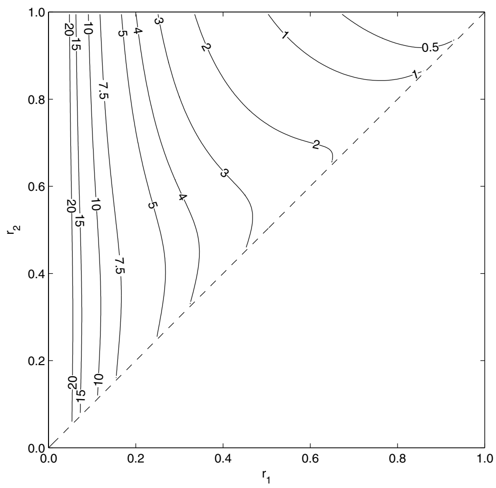
Error as a Function of (r_1, r_2)

For a fine grid of points in the (r_1, r_2) space, we determine the corresponding values of (α, β) and compute % underestimation:



Retention Elasticities as a Function of (r_1, r_2)

We determine the retention elasticity for the values of α and β associated with each point on the (r_1, r_2) space:



Further Reading

Fader, Peter S. and Bruce G. S. Hardie (2007), "How to Project Customer Retention," *Journal of Interactive Marketing*, **21** (Winter), 76–90.

Fader, Peter S. and Bruce G. S. Hardie (2006), "Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity." <http://brucehardie.com/papers/022/>

Fader, Peter S. and Bruce G. S. Hardie (2007), "Computing DERL for the sBG Model Using Excel." <http://brucehardie.com/notes/018/>

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

Classifying Customer Bases

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

Type of Relationship With Customers

Contract Duration in Continuous-Time

i. The duration of an individual customer's relationship with the firm is characterized by the exponential distribution with pdf and survivor function,

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

ii. Heterogeneity in λ follows a gamma distribution with pdf

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

Contract Duration in Continuous-Time

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

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

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

The Hazard Function

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

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

and represents the instantaneous rate of “failure” at time t conditional upon “survival” to t.

The probability of “failing” in the next small interval of time, given “survival” to time t, is

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

The Hazard Function

- For the exponential distribution,

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

- For the EG model,

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

- In applying the EG model, we are assuming that the increasing retention rates observed in the aggregate data are simply due to heterogeneity and not because of underlying time dynamics at the level of the individual customer.

Computing DERL

- Standing at time s,

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

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

$$\begin{aligned} DERL(\delta \mid \lambda, \text{tenure of at least } s) &= \int_0^\infty \lambda e^{-\lambda t} e^{-\delta t} dt \\ &= \frac{1}{\lambda + \delta} \end{aligned}$$

- But λ is unobserved ....

Computing DERL

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

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

Computing DERL

It follows that

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

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

Models for Noncontractual Settings

Classifying Customer Bases

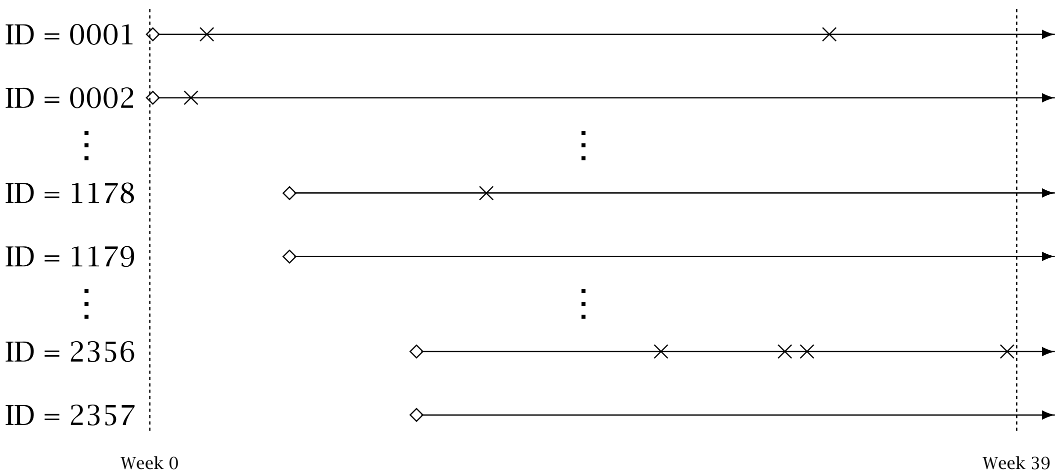
Opportunities for Transactions	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student mealplan Mobile phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subs Insurance policy Health club m'ship

Type of Relationship With Customers

Setting

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

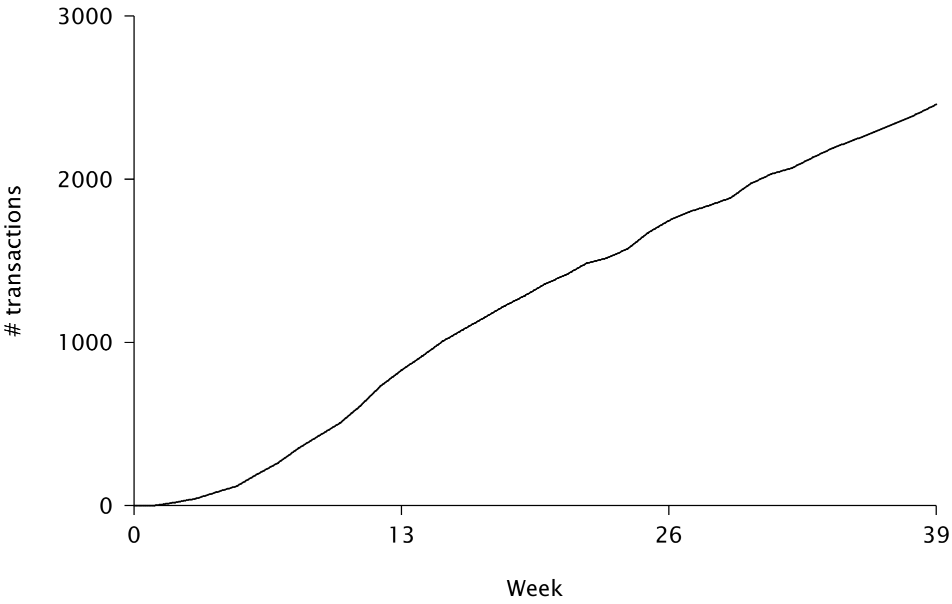
Purchase Histories



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:

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

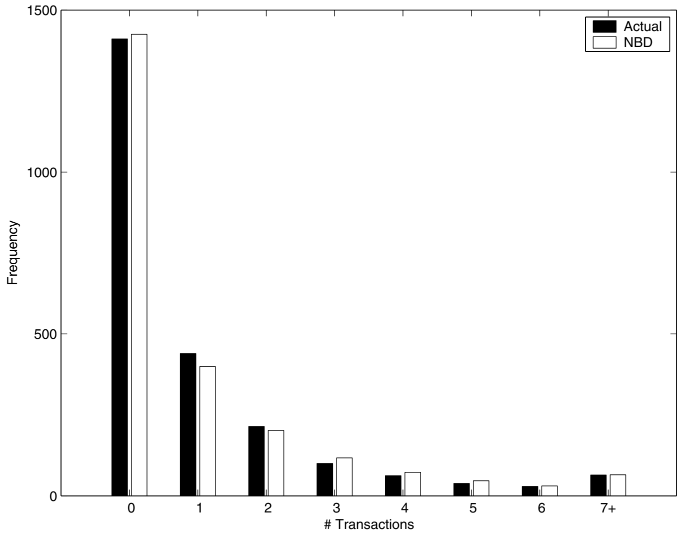
Modelling the Transaction Stream

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

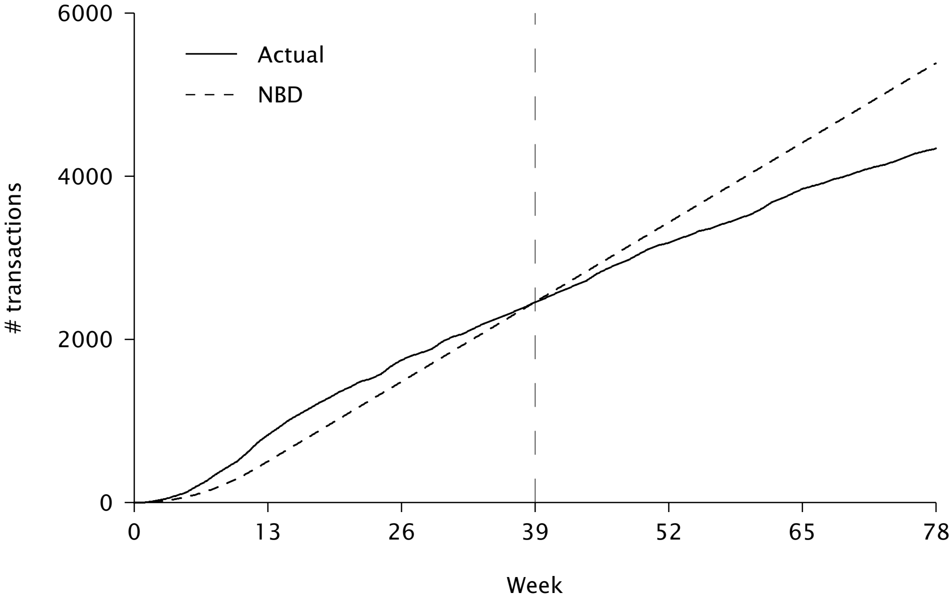
$$\begin{aligned} P(X(t) = x|r,\alpha) &= \int_0^\infty P(X(t) = x|\lambda)\, g(\lambda)\, d\lambda \\ &= \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x , \end{aligned}$$

which is the negative binomial distribution (NBD).

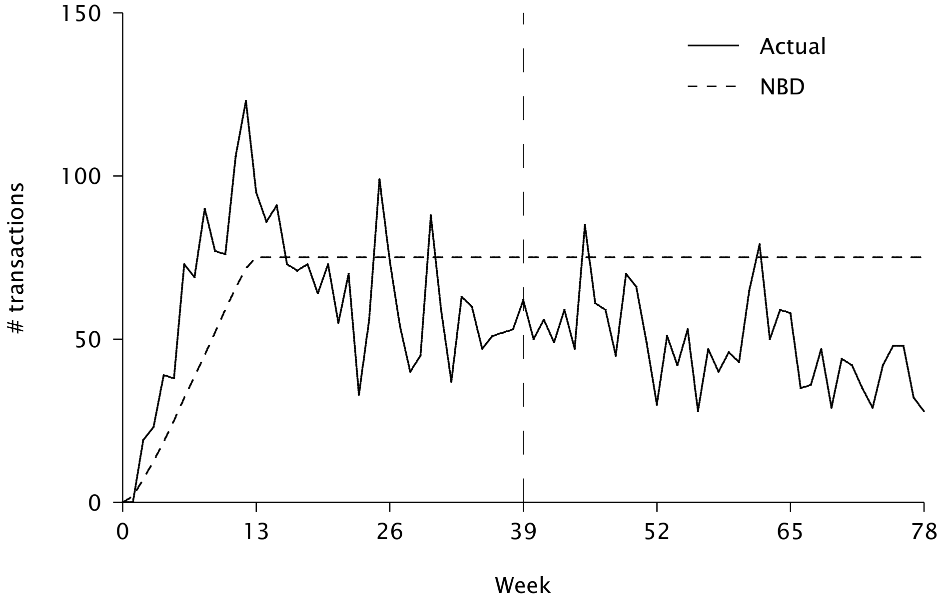
Frequency of Repeat Transactions



Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions

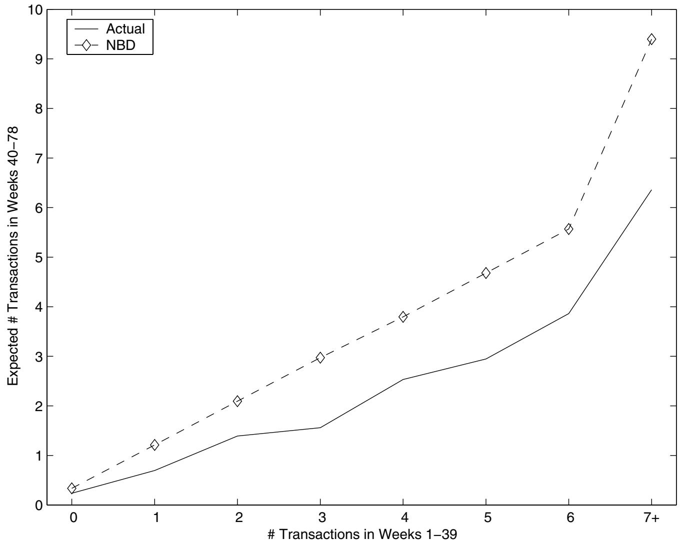


Conditional Expectations

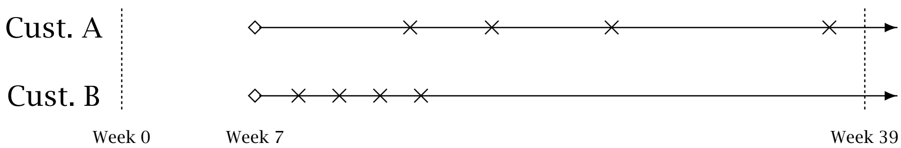
- We are interested in computing $E(Y(t)|\text{data})$, the expected number of transactions in an adjacent period $(T, T+t]$, conditional on the observed purchase history.
- For the NBD, a straight-forward application of Bayes' theorem gives us

$$E[Y(t)|r, \alpha, x, T] = \left(\frac{r+x}{\alpha+T}\right)t$$

Conditional Expectations



Conditional Expectations

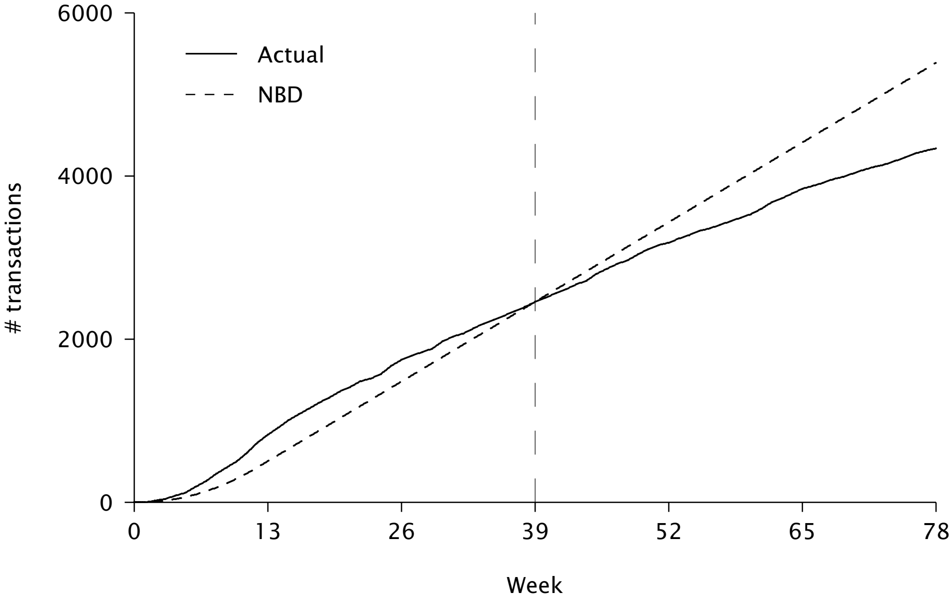


According to the NBD model:

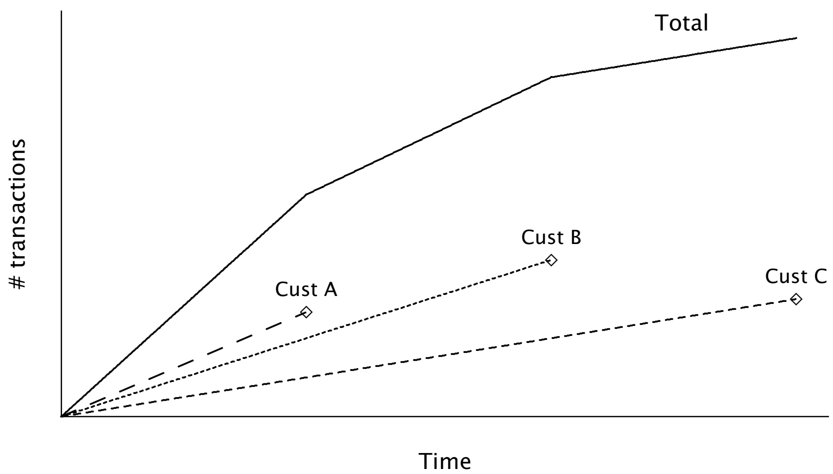
Cust. A: $E[Y(39) \mid x = 4, T = 32] = 3.88$

Cust. B: $E[Y(39) \mid x = 4, T = 32] = ?$

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).

Deriving the Model Likelihood Function

- Let us assume we know when each of a customer's x transactions occurred during the period $(0, T]$ (denoted by $t_1, t_2, \ldots, t_x$)
- There are two possible ways this pattern of transactions could arise:
  i. The customer is still alive at the end of the observation period (i.e., $\tau > T$)
  ii. The customer became inactive at some time τ in the interval $(t_x, T]$

Deriving the Model Likelihood Function

Conditional on λ,

$$\begin{aligned} L(\lambda \mid t_1, \ldots, t_x, T, \tau > T) \\ &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_x - t_{x-1})} e^{-\lambda(T - t_x)} \\ &= \lambda^x e^{-\lambda T} \end{aligned}$$

$$\begin{aligned} L(\lambda \mid t_1, \ldots, t_x, T, \text{inactive at } \tau \in (t_x, T]) \\ &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_x - t_{x-1})} e^{-\lambda(\tau - t_x)} \\ &= \lambda^x e^{-\lambda \tau} \end{aligned}$$

(Note: we do not need $t_1, \ldots, t_x$; x and t_x are sufficient.)

Deriving the Model Likelihood Function

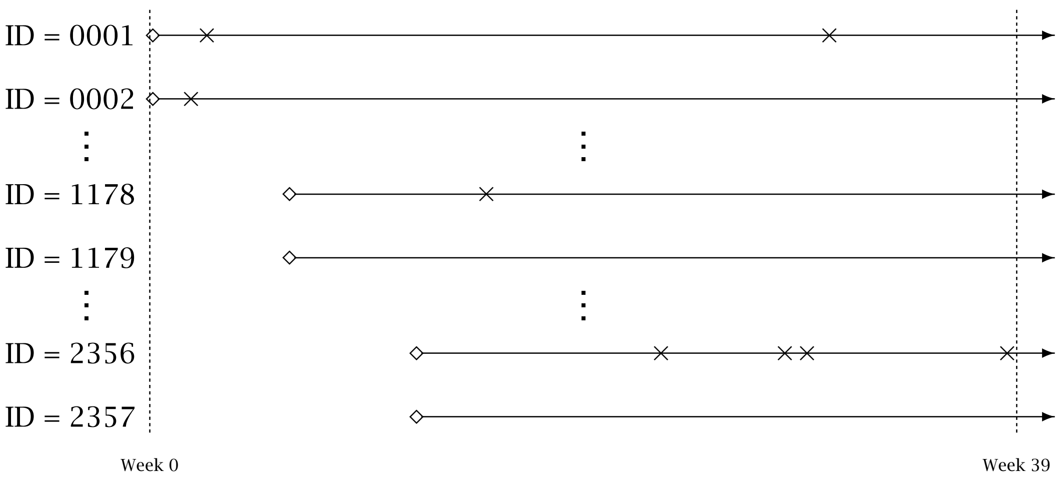
Removing the conditioning on τ,

$$
\begin{aligned}
L(\lambda,\mu \mid x, t_x, T) \\
&= L(\lambda \mid x, T, \tau > T) P(\tau > T \mid \mu) \\
&\quad + \int_{t_x}^{T} L(\lambda \mid x, T, \text{inactive at } \tau \in (t_x, T]) f(\tau \mid \mu)\, d\tau \\
&= \frac{\lambda^x \mu}{\lambda + \mu} e^{-(\lambda+\mu)t_x} + \frac{\lambda^{x+1}}{\lambda + \mu} e^{-(\lambda+\mu)T}
\end{aligned}
$$

Summarizing Purchase Histories

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

Removing the conditioning on λ and μ:

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

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

Key Results

$E[X(t)]$

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

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

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

$E(Y(t)\,|\,x,t_x,T)$

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

Expected Number of Transactions

Given that the number of transactions follows a Poisson process while the customer is alive,

i. if $\tau > t$,the expected number of transactions is simply λt.

ii. if $\tau \leq t$, the expected number of transactions in the interval $(0, \tau]$ is $\lambda\tau$.

Expected Number of Transactions

Removing the conditioning on τ:

$$\begin{aligned} E[X(t) \mid \lambda, \mu] &= \lambda t P(\tau > t \mid \mu) + \int_0^t \lambda\tau f(\tau \mid \mu)\, d\tau \\ &= \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t} \end{aligned}$$

Taking the expectation over the distributions of λ and μ:

$$\begin{aligned} &E[X(t) \mid r, \alpha, s, \beta] \\ &\qquad = \int_0^\infty \int_0^\infty E[X(t) \mid \lambda, \mu] g(\lambda \mid r, \alpha) g(\mu \mid s, \beta)\, d\lambda d\mu \\ &\qquad = \frac{r\beta}{\alpha(s-1)} \left[1 - \left(\frac{\beta}{\beta + t} \right)^{s-1} \right]. \end{aligned}$$

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

- The probability that a customer with purchase history (x, t_x, T) is "alive" at time T is the probability that the (unobserved) time at which he becomes inactive (τ) occurs after T, $P(\tau > T)$.
- By Bayes' theorem:

$$\begin{aligned} P(\tau > T \mid \lambda, \mu, x, t_x, T) &= \frac{L(\lambda \mid x, T, \tau > T)P(\tau > T \mid \mu)}{L(\lambda, \mu \mid x, t_x, T)} \\ &= \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(\lambda, \mu \mid x, t_x, T)} . \end{aligned}$$

- But λ and μ are unobserved.

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

We take the expectation of $P(\tau > T \mid \lambda, \mu, x, t_x, T)$ over the distribution of λ and μ, updated to take account of the information (x, t_x, T):

$$\begin{aligned} &P(\text{alive} \mid r, \alpha, s, \beta, x, t_x, T) \\ &\quad = \int_0^\infty \int_0^\infty \Big\{ P(\tau > T \mid \lambda, \mu, x, t_x, T) \\ &\qquad\qquad \times g(\lambda, \mu \mid r, \alpha, s, \beta, x, t_x, T) \Big\} \, d\lambda d\mu \end{aligned}$$

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

- By Bayes' theorem, the joint posterior distribution of λ and μ is

$$g(\lambda, \mu \mid r, \alpha, s, \beta, x, t_x, T) = \frac{L(\lambda, \mu \mid x, t_x, T) g(\lambda \mid r, \alpha) g(\mu \mid s, \beta)}{L(r, \alpha, s, \beta \mid x, t_x, T)} .$$

- Therefore,

$$P(\text{alive} \mid r, \alpha, s, \beta, x, t_x, T) = \frac{\Gamma(r + x)\alpha^r \beta^s}{\Gamma(r)(\alpha + T)^{r+x}(\beta + T)^s} \Big/ L(r, \alpha, s, \beta \mid x, t_x, T) .$$

Conditional Expectations

Let $Y(t)$ = the number of purchases made in the period $(T, T + t]$.

$$\begin{aligned} E[Y(t) \mid \lambda, \mu, \text{alive at } T] &= \lambda t P(\tau > T + t \mid \mu, \tau > T) \\ &\quad + \int_T^{T+t} \lambda \tau f(\tau \mid \mu, \tau > T)\, d\tau \\ &= \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t} . \end{aligned}$$

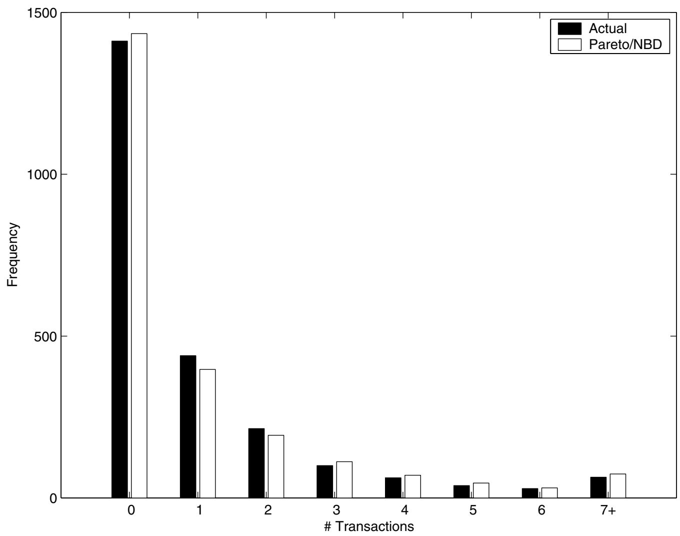
$$\begin{aligned} E[Y(t) \mid \lambda, \mu, x, t_x, T] &= E[Y(t) \mid \lambda, \mu, \text{alive at } T] \\ &\quad \times P(\tau > T \mid \lambda, \mu, x, t_x, T) \end{aligned}$$

Conditional Expectations

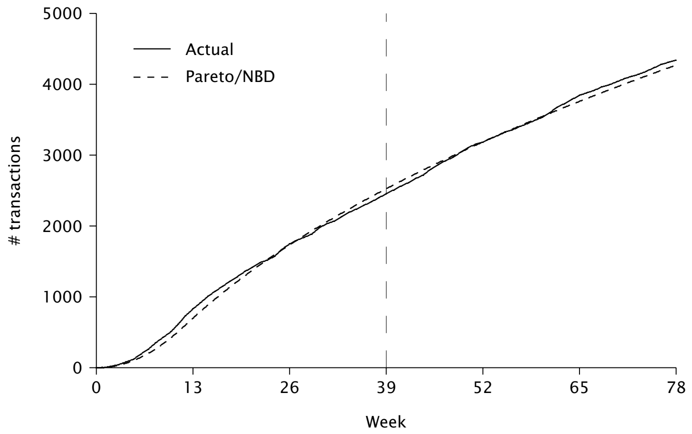
Taking the expectation over the joint posterior distribution of λ and μ yields:

$$E[Y(t) \mid r, \alpha, s, \beta, x, t_x, T]$$
$$= \left\{ \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \Big/ L(r, \alpha, s, \beta \mid x, t_x, T) \right\}$$
$$\times \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)} \left[1 - \left(\frac{\beta+T}{\beta+T+t} \right)^{s-1} \right].$$

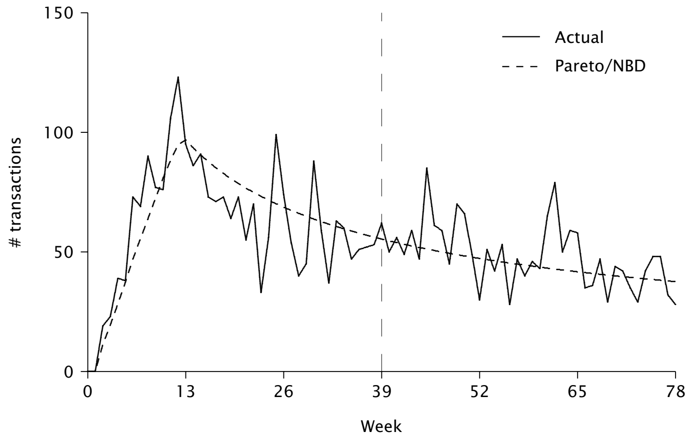
Frequency of Repeat Transactions



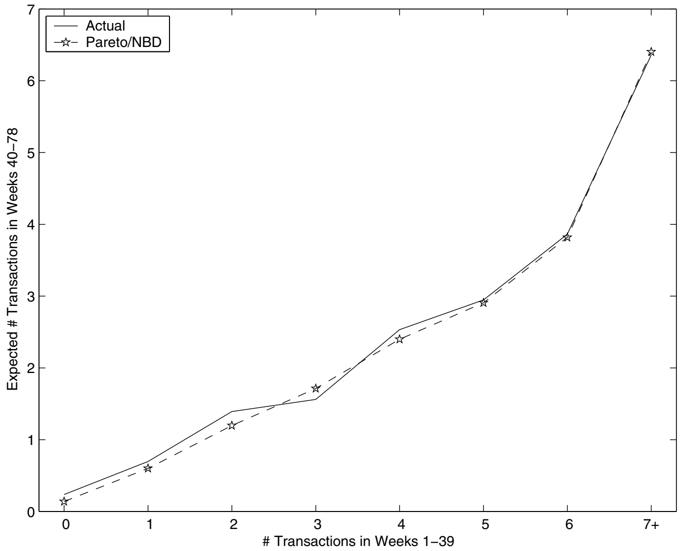
Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions



Conditional Expectations



Computing $E(CLV)$

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

If we assume that an individual's spend per transaction is constant, $v(t)$ = net cashflow / transaction $\times\, t(t)$ (where $t(t)$ is the transaction rate at t) and

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

Computing $E(RLV)$

- Standing at time T,

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

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

Computing DERT

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

$$\begin{aligned} DERT(\delta \mid \lambda, \mu, \text{alive at } T) &= \int_T^\infty \lambda \left(\frac{e^{-\mu t}}{e^{-\mu T}} \right) e^{-\delta(t-T)} dt \\ &= \int_0^\infty \lambda e^{-\mu s} e^{-\delta s} ds \\ &= \frac{\lambda}{\mu + \delta} \end{aligned}$$

Computing DERT

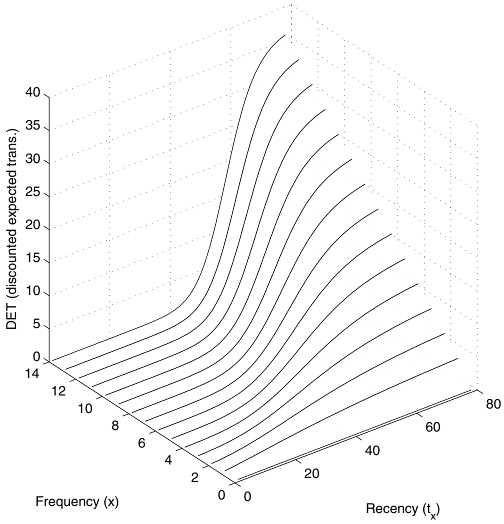
$$
\begin{aligned}
&DERT(\delta \mid r, \alpha, s, \beta, x, t_x, T) \\
&\quad = \int_0^\infty \int_0^\infty \Big\{ DERT(\delta \mid \lambda, \mu, \text{alive at } T) \\
&\qquad\qquad \times P(\text{alive at } T \mid \lambda, \mu, x, t_x, T) \\
&\qquad\qquad\quad \times g(\lambda, \mu \mid r, \alpha,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

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

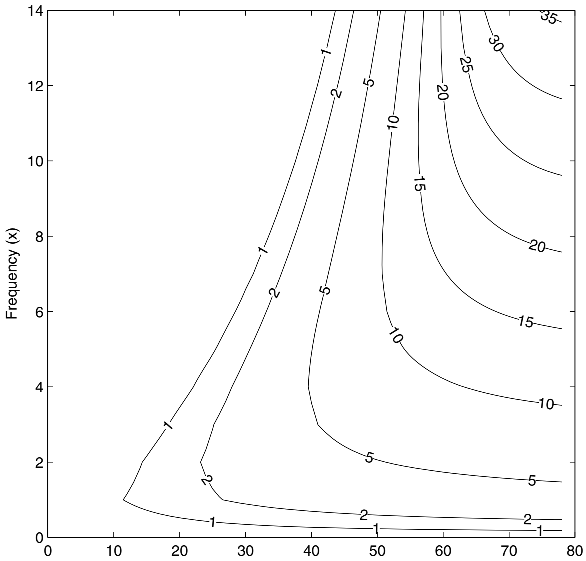
Continuous Compounding

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

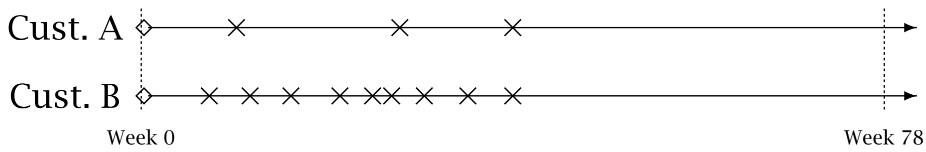
DERT by Recency and Frequency



Iso-Value Representation of DERT



The “Increasing Frequency” Paradox



	DERT
Cust. A	4.6
Cust. B	1.9

Key Contribution

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

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

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

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

Modelling the Spend Process

- The dollar value of a customer's given transaction varies randomly around his average transaction value
- Average transaction values vary across customers but do not vary over time for any given individual
- The distribution of average transaction values across customers is independent of the transaction process.

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 average observed transaction value

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

  is an imperfect estimate of his (unobserved) mean transaction value $E(M)$
- Our goal is to make inferences about $E(M)$ given m_x, which we denote as $E(M|m_x, x)$

Summary of Average Transaction Value

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

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

Modelling the Spend Process

- The dollar value of a customer's given transaction is distributed gamma with shape parameter p and scale parameter ν
- Heterogeneity in ν across customers follows a gamma distribution with shape parameter q and scale parameter γ

Modelling the Spend Process

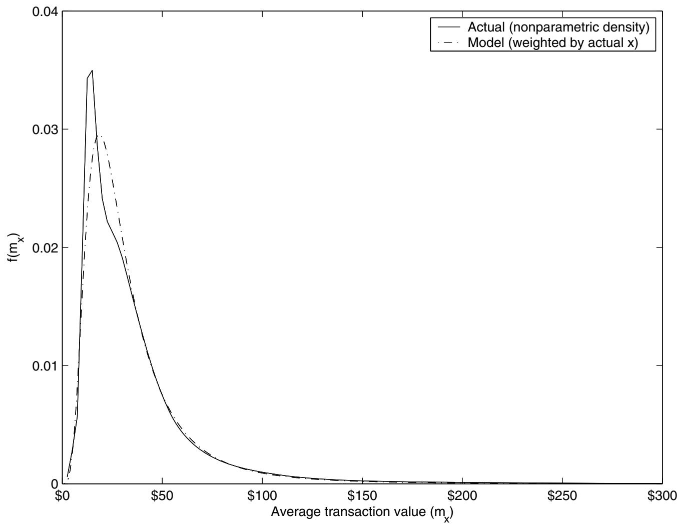
Marginal distribution for m_x:

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

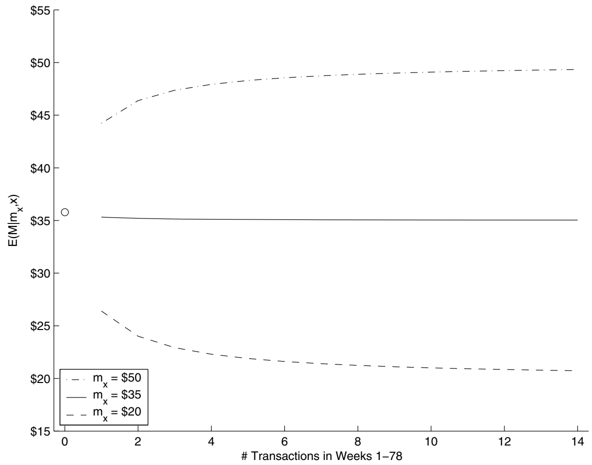
Expected average transaction value for a customer with an average spend of m_x across x transactions:

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

Distribution of Average Transaction Value



E(Monetary Value) as a Function of M and F



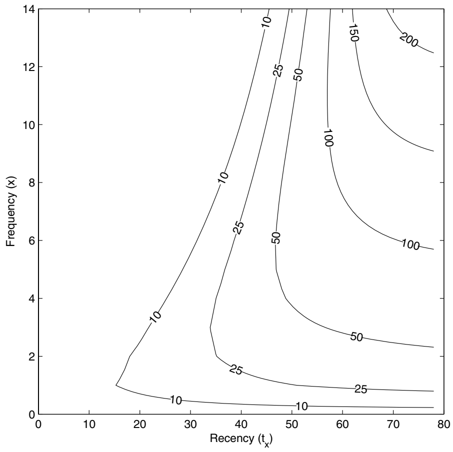
Computing Expected Residual Lifetime Value

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

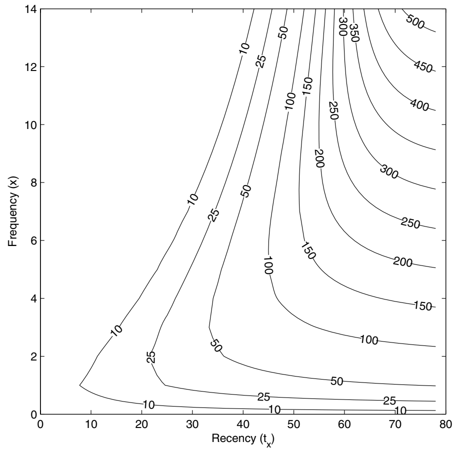
$$\begin{aligned} E(RLV) &= \text{margin} \times \text{revenue/transaction} \times DERT \\ &= \text{margin} \times E(M|p,q,\gamma,m_x,x) \\ &\qquad \times DERT(\delta \mid r,\alpha,s,\beta,x,t_x,T) \end{aligned}$$

Estimates of $E(RLV)$

M = $20



M = $50



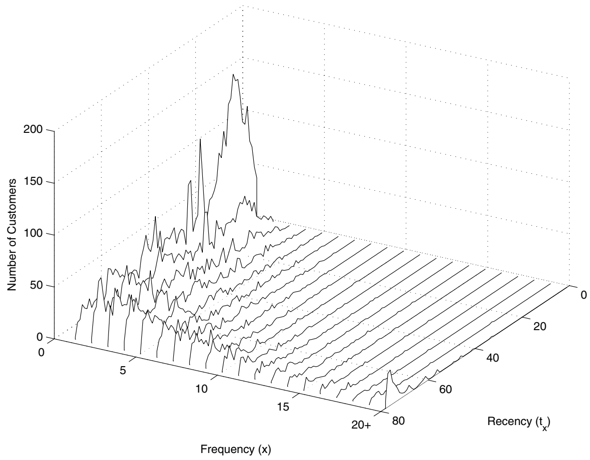
(Margin = 30%, 15% discount rate)

Summary: 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).
- Computed 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

	Frequency	Recency 0	1	2	3
M=0	0	$53,000			
M=1	1		$7,700	$9,900	$1,800
	2		$2,800	$15,300	$17,400
	3		$300	$12,500	$52,900
M=2	1		$5,900	$7,600	$2,300
	2		$3,600	$26,500	$25,800
	3		$500	$37,200	$203,000
M=3	1		$11,300	$19,700	$3,700
	2		$7,300	$45,900	$47,900
	3		$1,000	$62,700	$414,900

An Alternative to the Pareto/NBD Model

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

Deriving the Model Likelihood Function

- Let us assume we know when each of a customer's x transactions occurred during the period $(0, T]$ (denoted by $t_1, t_2, \ldots, t_x$)
- There are two possible ways this pattern of transactions could arise:
  i. The customer is still alive at the end of the observation period (i.e., $\tau > T$)
  ii. The customer became inactive immediately after the xth transaction (i.e., $\tau = t_x$)

Deriving the Model Likelihood Function

Conditional on λ,

$$L(\lambda \mid t_1, \ldots, t_x, T, \tau > T)$$
$$= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_x - t_{x-1})} e^{-\lambda(T - t_x)}$$
$$= \lambda^x e^{-\lambda T}$$

$$L(\lambda \mid t_1, \ldots, t_x, T, \text{inactive at } \tau = t_x)$$
$$= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_x - t_{x-1})}$$
$$= \lambda^x e^{-\lambda t_x}$$

Deriving the Model Likelihood Function

Removing the conditioning on τ,

$$L(\lambda, p \mid x, t_x, T)$$
$$= L(\lambda \mid x, T, \tau > T) P(\tau > T \mid p)$$
$$+ L(\lambda \mid x, T, \text{inactive at } \tau = t_x) P(\tau = t_x \mid p)$$
$$= (1 - p)^x \lambda^x e^{-\lambda T} + \delta_{x>0}\, p (1 - p)^{x-1} \lambda^x e^{-\lambda t_x}$$

where $\delta_{x>0} = 1$ if $x > 0$, 0 otherwise.

Deriving the Model Likelihood Function

Removing the conditioning on λ and p,

$$
\begin{aligned}
&L(r,\alpha,a,b \mid x,t_x,T)\\
&\quad= \int_0^1 \int_0^\infty L(\lambda,p \mid x,t_x,T) f(\lambda \mid r,\alpha) f(p \mid a,b)\, d\lambda dp\\
&\quad= \int_0^1 \int_0^\infty (1-p)^x \lambda^x e^{-\lambda T} \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} d\lambda dp\\
&\qquad + \delta_{x>0} \int_0^1 \int_0^\infty \Big\{ p(1-p)^{x-1} \lambda^x e^{-\lambda t_x}\\
&\qquad\qquad \times \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} \Big\} d\lambda dp
\end{aligned}
$$

BG/NBD Likelihood Function

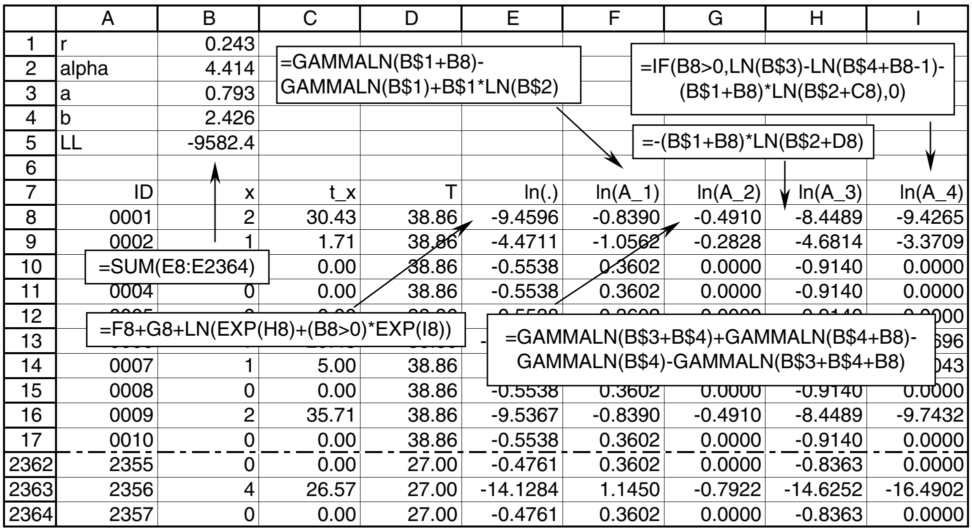
We can express the model likelihood function as:

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

where

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

BGNBD Estimation

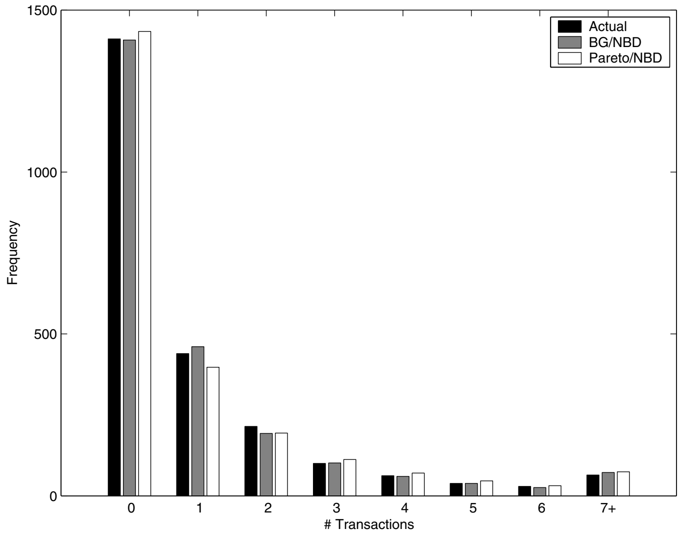


	A	B	C	D	E	F	G	H	I
1	r	0.243							
2	alpha	4.414							
3	a	0.793							
4	b	2.426							
5	LL	-9582.4							
6									
7	ID	x	t_x	T	ln(.)	ln(A_1)	ln(A_2)	ln(A_3)	ln(A_4)
8	0001	2	30.43	38.86	-9.4596	-0.8390	-0.4910	-8.4489	-9.4265
9	0002	1	1.71	38.86	-4.4711	-1.0562	-0.2828	-4.6814	-3.3709
10	[illegible]	[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]	[illegible]
13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]696
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

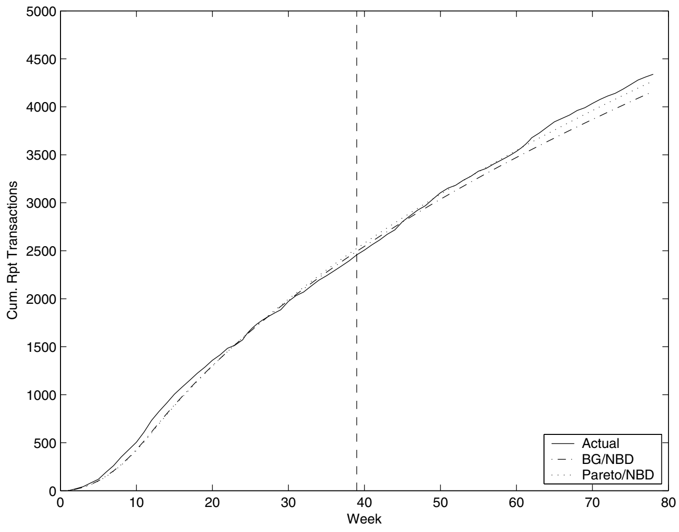
Model Estimation Results

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

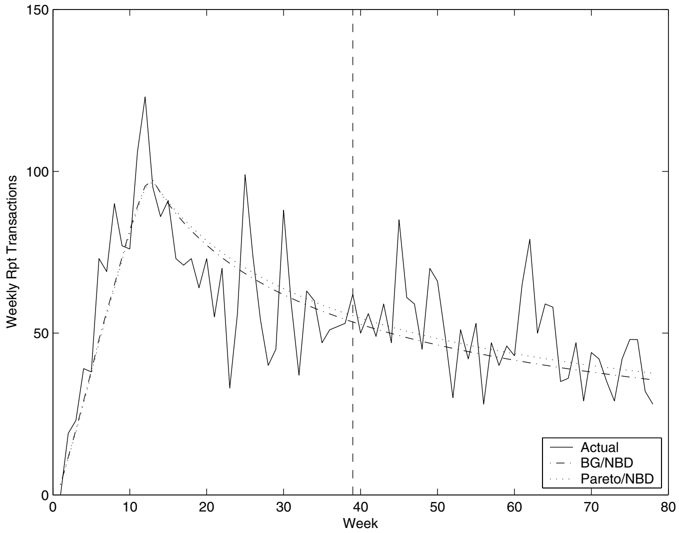
Frequency of Repeat Transactions



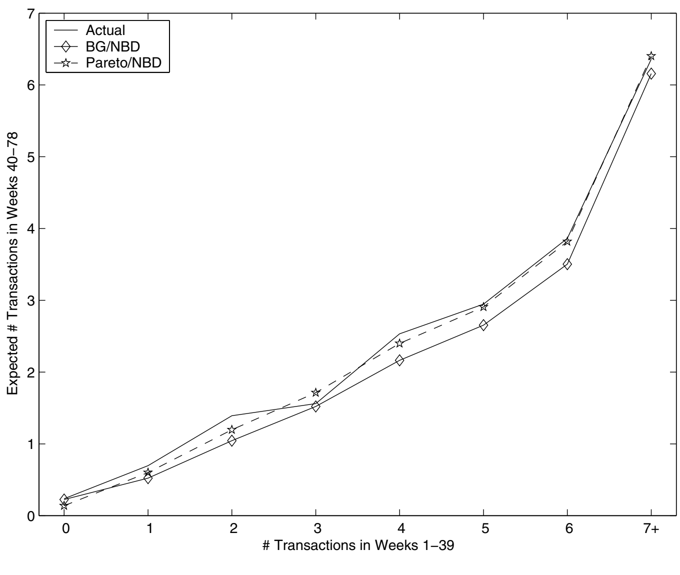
Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions



Conditional Expectations



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/`>

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005b), “RFM and CLV: Using Iso-value Curves for Customer Base Analysis,” *Journal of Marketing Research*, **42** (November), 415–430.

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005c), “"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model,” *Marketing Science*, **24** (Spring), 275–284.

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

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2007), “Creating a Fit Histogram for the BG/NBD Model .” <`http://brucehardie.com/notes/014/`>

Fader, Peter S. and Bruce G. S. Hardie (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?

Classifying Customer Bases

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

Type of Relationship With Customers

"Discrete-Time" Transaction Opportunities

"necessarily discrete"	attendance at sports events 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

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 ($DERT \longrightarrow E(RLV)$ 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

- Bernoulli 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, 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

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	t_x	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, t_x, n),

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

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

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

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

Model Development

Removing the conditioning on p and θ,

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

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

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	t_x	n	# cust.	L(.|X=x,t_x,n)			n-t_x-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	t_x	n	# cust.	L(.|X=x,t_x,n)			n-t_x-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)

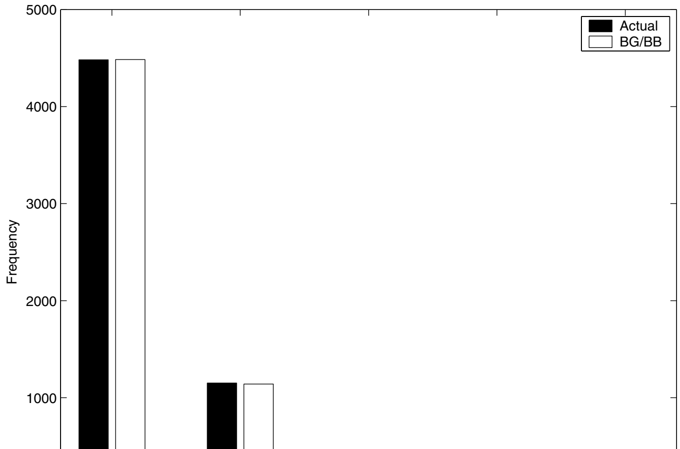
=IF($H9<J$8,0,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+J$8)-GAMMALN($B$1+$B$2+$B9+J$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+J$8)-GAMMALN($B$3+$B$4+$B9+J$8+1))/$E$3)

=C15-B15-1

=D19*LN(F19))

=SUM(I19:M19)

Model Fit



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

Key Results

$P(\text{alive in period } n+1 \mid x, t_x, n)$

The probability that an individual with observed behavior (x, t_x, n) will be "active" in the next period.

$E(X^* \mid n^*, x, t_x, n)$

The expected number of transactions across the next n^* transaction opportunities for an individual with observed behavior (x, t_x, n).

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

The discounted expected residual transactions for an individual with observed behavior (x, t_x, n).

P(alive in period $n + 1 \mid x, t_x, n$)

- 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,

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

it follows that

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

P(alive in period $n+1 \mid x, t_x, n$)

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

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

Conditional Expectations

Let X^* denote the number of purchases over the next n^* periods (i.e., in the interval $(n, n+n^*]$).

Assuming the customer is alive in period n,

$$
\begin{aligned}
&E(X^* \mid n^*, p, \theta, \text{alive in period } n) \\
&\qquad = \sum_{t=n+1}^{n^*} P(Y_t = 1 \mid p, \text{alive at } t) P(\text{alive at } t \mid t > n, \theta) \\
&\qquad = \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n^*+1}}{\theta}\,.
\end{aligned}
$$

Conditional Expectations

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

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

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	t_x	n	# cust.	L(.|X=x,t_x,n)			P(alive in 1998)		n-t_x-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
```

Conditional Expectations

	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)	4.3E-260										
4	delta	1882.93													
5	n*	4													
6															
7	LL	-7130.7													
8														i	
9	x	t_x	n	# cust.	L(.|X=x,t_x,n)			E(X*|n*)		n-t_x-1		0	1	2	3
10	4	4	4	18	-106.67	0.00267		1.52		-1	0.00267	0	0	0	0
11	3	4	4	66	-367.99	0.00379		1.20		-1	0.00379	0	0	0	0
12	2	4	4	98	-463.46	0.00883		0.87		-1	0.00883	0	0	0	0
13	1	4	4	216	-704.36	0.03835		0.54		-1	0.03835	0	0	0	0
14	3	3	4	34	-184.61	0.00438		1.03		0	0.00379	0.000595	0	0	0
15	2	3	4	180	-828.99	0.01000		0.77		0	0.00883	0.001163	0	0	0
16	1	3	4	292	-920.83	0.04270		0.49		0	0.03835	0.004348	0	0	0
17	2	2	4	64	-283.50	0.01192		0.64		1	0.00883	0.001921	0.001163	0	0
18	1	2	4	342	-1033.39	0.04872		0.43		1	0.03835	0.006022	0.004348	0	0
19	1	1	4	302	-863.02	0.05740		0.36		2	0.03835	0.008679	0.006022	0.004348	0
20	0	0	4	4482	-1373.92	0.73599		0.14		3	0.47847	0.084486	0.068631	0.056783	0.047618

```
=(EXP(GAMMALN($B$3-1)+GAMMALN($B$4+C10+1)-GAMMALN($B$3+$B$4+C10))-
EXP(GAMMALN($B$3-1)+GAMMALN($B$4+C10+$B$5+1)-
GAMMALN($B$3+$B$4+C10+$B$5)))*EXP(GAMMALN($B$1+A10+1)+
GAMMALN($B$2+C10-A10)-GAMMALN($B$1+$B$2+C10+1))/(E$1*E$3)/F10
```


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

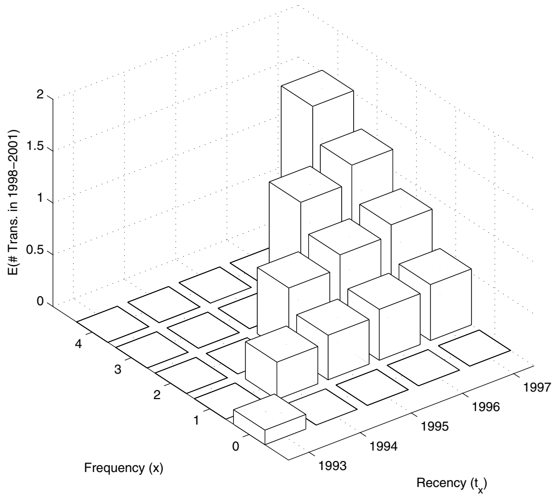
Posterior Mean of p as a Function of Recency and Frequency

	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	0.17	0.18	0.19	
0					0.08

Expected # Transactions in 1998–2001 as a Function of Recency and Frequency

	Year of Last Cruise				
# Cruise-years	1997	1996	1995	1994	1993
4	1.52				
3	1.20	1.03			
2	0.87	0.77	0.64		
1	0.54	0.49	0.43	0.36	
0					0.14

Expected # Transactions in 1998–2001 as a Function of Recency and Frequency



387

Computing DERT

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

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

- However, p and θ are unobserved.

Computing DERT

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

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

Computing DERT

For a customer with purchase history (x, t_x, n), we multiply $DERT(d \mid p, \theta, \text{alive at } n)$ by the probability that he is alive at transaction opportunity n and integrate over the *posterior* distribution of p and θ, giving us

$$\begin{aligned} &DERT(d \mid \alpha, \beta, \gamma, \delta, x, t_x, n) \\ &\qquad = \frac{B(\alpha+x+1, \beta+n-x)B(\gamma, \delta+n+1)}{B(\alpha, \beta)B(\gamma, \delta)(1+d)} \\ &\qquad\qquad \times \frac{{}_2F_1(1, \delta+n+1; \gamma+\delta+n+1; \frac{1}{1+d})}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)} \end{aligned}$$

DET

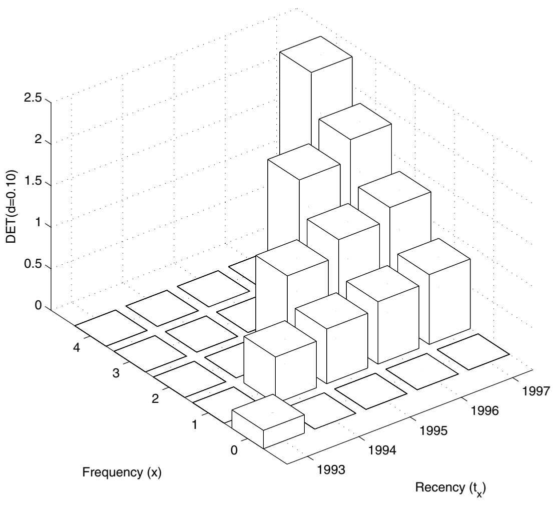
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	alpha	0.66	B(alpha,beta)		0.4751										
2	beta	5.19													
3	gamma	173.76	B(gamma,delta)		4E-260										
4	delta	1882.93													
5															
6	d	0.1	annual discount rate												
7															
8	LL	-7130.7													
9													i		
10	x	t_x	n	# cust.	L(.|X=x,t_x,n)			DET		n-t_x-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)

DERT as a Function of R & F ($d = 0.10$)



DERT as a Function of R & F ($d = 0.10$)

	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

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Paul D. Berger (2004), "Customer-Base Analysis with Discrete-Time Transaction Data." <http://brucehardie.com/papers/020/>

Fader, Peter S., Bruce G. S. Hardie, and Paul D. Berger (2005), "Implementing the BG/BB Model for Customer-Base Analysis in Excel." <http://brucehardie.com/notes/010/>

Beyond the Basic Models

Implementation Issues

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

Implications of Data Recording Processes (Contractual Settings)

Cohort	Calendar Time →				
1	n_{11}	n_{12}	n_{13}	$\dots$	n_{1I}
2		n_{22}	n_{23}	$\dots$	n_{2I}
3			n_{33}	$\dots$	n_{3I}
$\vdots$				$\ddots$	$\vdots$
I					n_{II}
	$n_{.1}$	$n_{.2}$	$n_{.3}$	$\dots$	$n_{.I}$

Implications of Data Recording Processes (Contractual Settings)

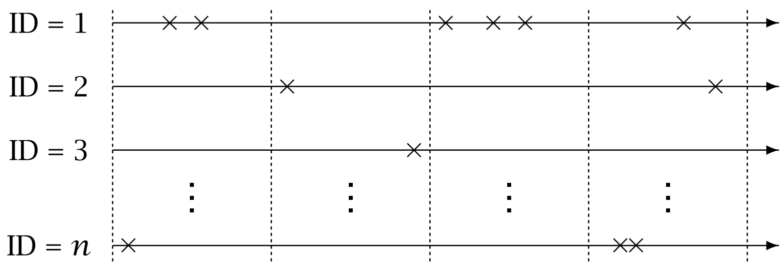
Cohort	Calendar Time →				
1	n_{11}				n_{1I}
2		n_{22}			n_{2I}
$\vdots$					$\vdots$
I−1				$n_{I-1,I-1}$	$n_{I-1,I}$
I					n_{II}

Cohort	Calendar Time →				
1	n_{11}				
2		n_{22}			
$\vdots$					
I−1				$n_{I-1,I-1}$	
I					n_{II}
	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.I-1}$	$n_{.I}$

Cohort	Calendar Time →				
1					n_{1I}
2					n_{2I}
$\vdots$					$\vdots$
I−1					$n_{I-1,I}$
I					n_{II}
	$n_{.1}$	$n_{.2}$	$\dots$	$n_{.I-1}$	$n_{.I}$

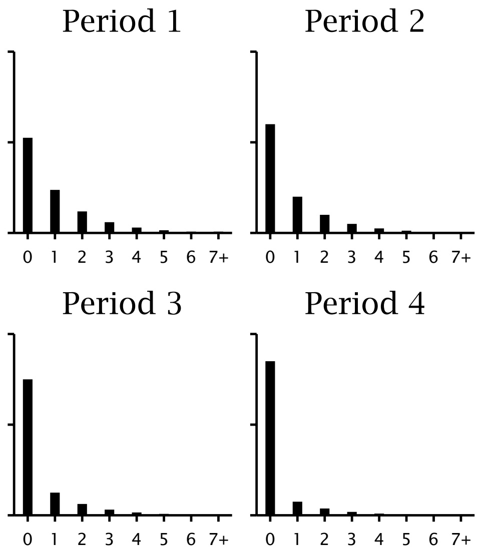
Cohort	Calendar Time →				
1				n_{1I-1}	n_{1I}
2				n_{2I-1}	n_{2I}
$\vdots$				$\vdots$	$\vdots$
I−1				$n_{I-1,I-1}$	$n_{I-1,I}$
I					n_{II}

Implications of Data Recording Processes (Noncontractual Settings)



Implications of Data Recording Processes (Noncontractual Settings)

	Period			
ID	1	2	3	4
1	2	0	3	1
2	0	1	0	1
3	0	1	0	0
⋮				
n	1	0	0	2



Implications of Data Recording Processes (Noncontractual Settings)

The model likelihood function must match the data structure:

- Interval-censored individual-level data

  Fader, Peter S. and Bruce G. S. Hardie (2005), “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 (2006), “Estimating CLV Using Aggregated Data: The *Tuscan Lifestyles* Case Revisited .”
  <http://brucehardie.com/papers/024/>

Model Extensions

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

Individual-Level Duration Dependence

- The exponential distribution is often characterized as being "memoryless".
- This means the probability that the event of interest occurs in the interval $(t, t + \Delta t]$ given that it has not occurred by t is independent of t:

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

- This is equivalent to a constant hazard function.

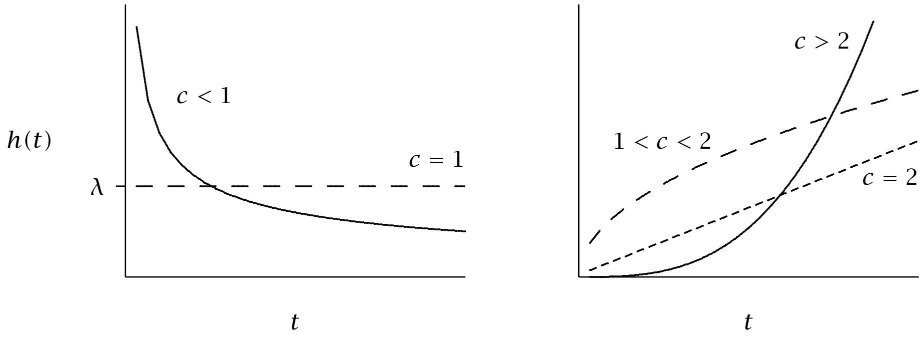
The Weibull Distribution

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

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

  where c is the "shape" parameter and λ is the "scale" parameter.
- Collapses to the exponential when $c = 1$.
- F(t) is S-shaped for $c > 1$.

The Weibull Hazard Function



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

- Decreasing hazard function (negative duration dependence) when $c < 1$.
- Increasing hazard function (positive duration dependence) when $c > 1$.

Individual-Level Duration Dependence

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

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

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

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

  using standard numerical integration techniques.

Individual-Level Duration Dependence

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

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

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

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

Nonstationarity

- "Buy then die" ⇔ latent characteristics governing purchasing are constant then become 0.

- Perhaps more realistic to assume that these latent characteristics can change over time.

- Nonstationarity can be handled using a hidden Markov model

  Netzer, Oded, James Lattin, and V. Srinivasan (2005), "A Hidden Markov Model of Customer Relationship Dynamics," working paper, Columbia Business School.

  or a (dynamic) changepoint model

  Fader, Peter S., Bruce G. S. Hardie, and Chun-Yao Huang (2004), "A Dynamic Changepoint Model for New Product Sales Forecasting," *Marketing Science*, **23** (Winter), 50–65.

Covariates

- Types of covariates:
  - customer characteristics
  - customer attitudes and behavior
  - marketing activities
- Handling covariate effects:
  - explicit integration (via latent characteristics and hazard functions)

    Schweidel, David A., Peter S. Fader, and Eric Bradlow (2006), "Modeling Service Retention Within and Across Cohorts under Limited Information." <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=742884>
  - used to create segments (and apply no-covariate models)
- Need to be wary of endogeneity bias and sample selection effects

The Cost of Model Extensions

- No closed-form likelihood functions; need to resort to simulation methods.
- Need full datasets; summaries (e.g., RFM) no longer sufficient.

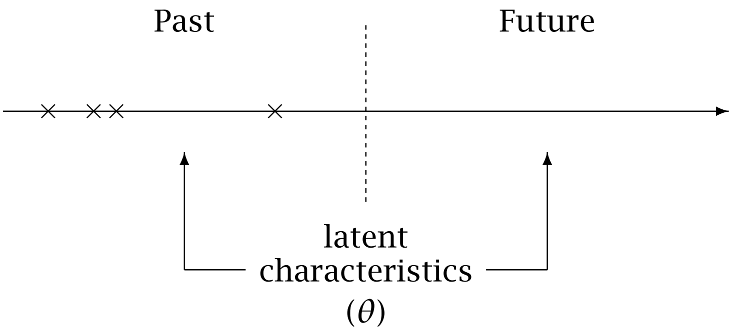
Philosophy of Model Building

- 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

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

Classifying Customer Bases

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

Type of Relationship With Customers