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

Workshop on Customer-Base Analysis
Johann Wolfgang Goethe-Universität, Frankfurt
March 8–9, 2006

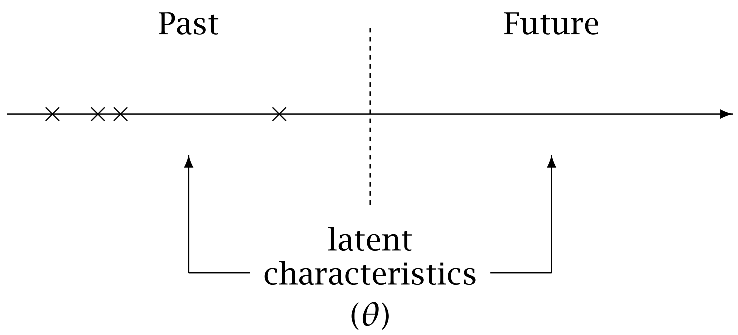
Customer-Base Analysis

- 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 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

Agenda

- The right way to think about computing CLV
- Models for noncontractual settings
  - The Pareto/NBD model
  - The BG/NBD model
  - The BG/BB model
- Models for contractual settings
- Beyond the basic models

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 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 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* CLV)

Models for Noncontractual Settings

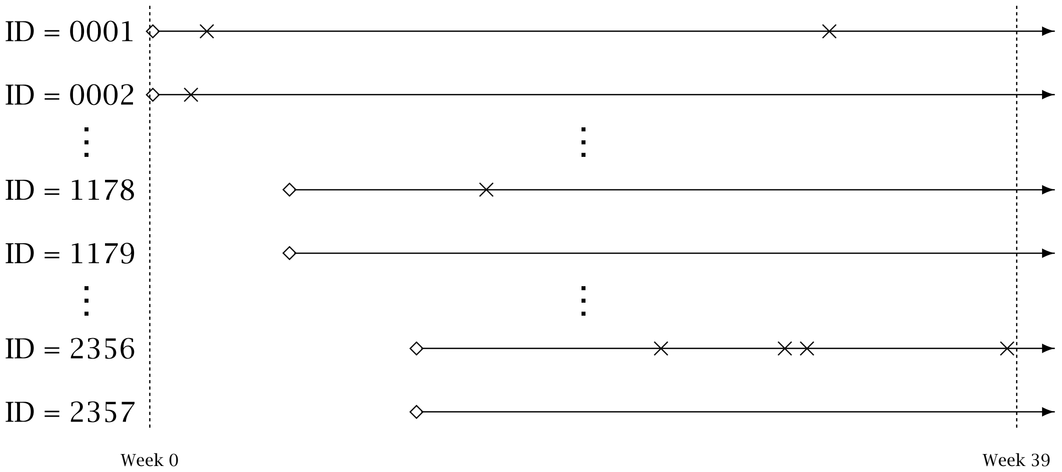
Classifying Customer Bases

Opportunities for Transactions		Type of Relationship With Customers: Noncontractual	Type of Relationship With Customers: 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

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

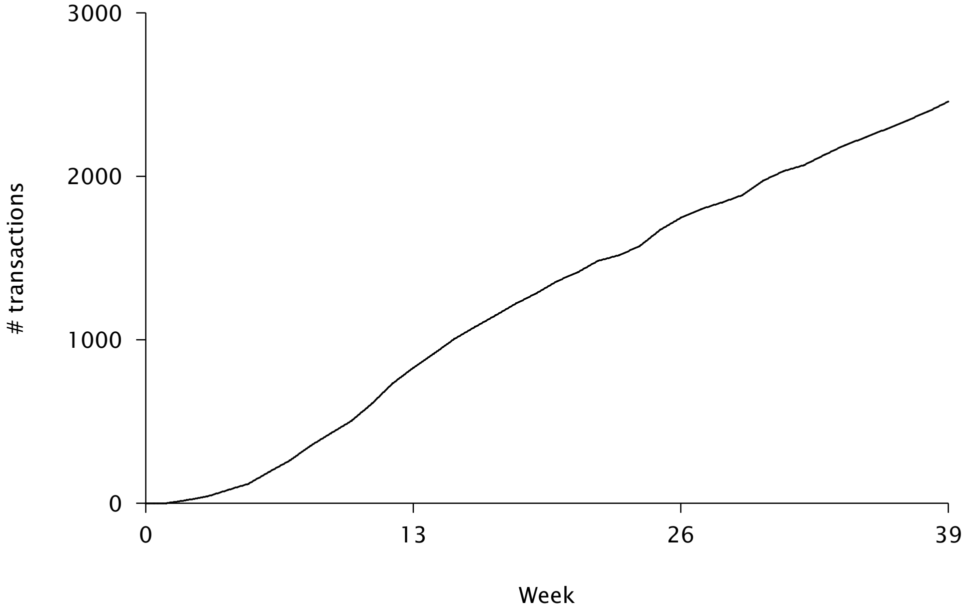
Purchase Histories



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:

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

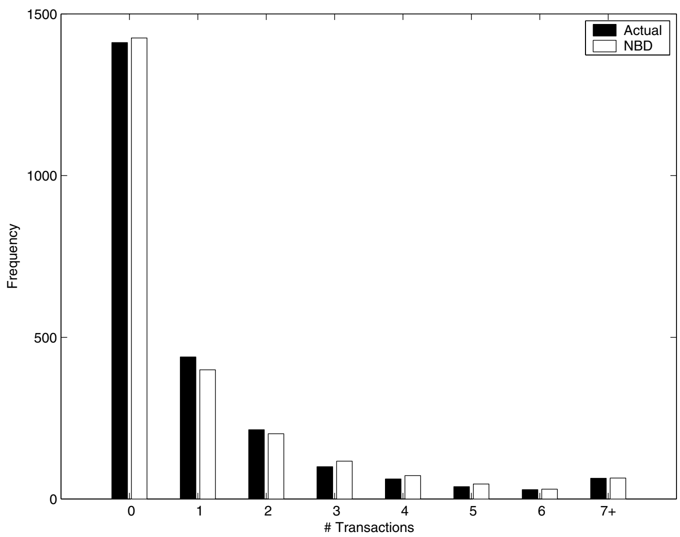
Modelling the Transaction Stream

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

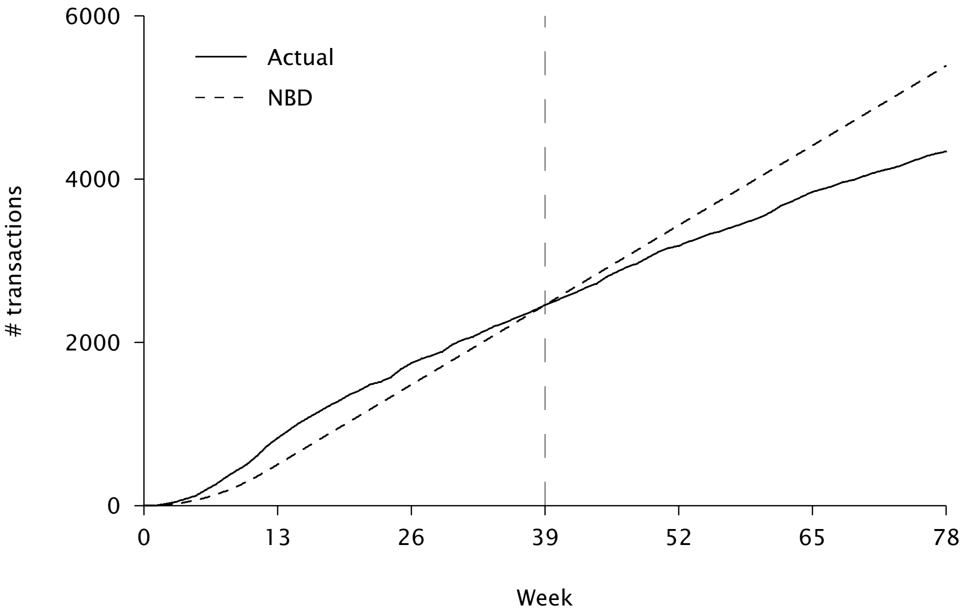
$$\begin{aligned} P(X(t) = x \mid r, \alpha) &= \int_0^\infty P(X(t) = x \mid \lambda)\, g(\lambda)\, d\lambda \\ &= \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x , \end{aligned}$$

which is the negative binomial distribution (NBD).

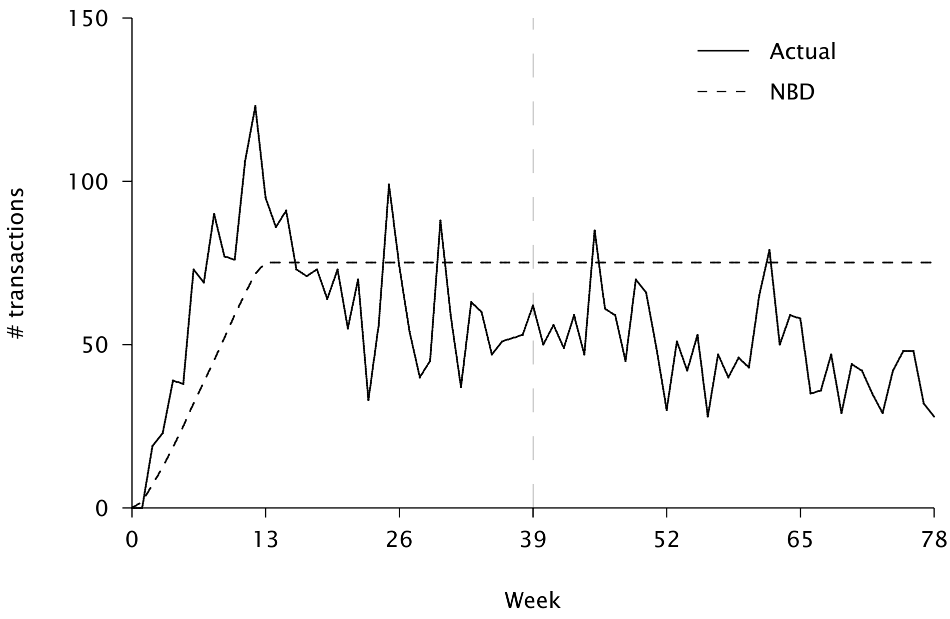
Frequency of Repeat Transactions



Tracking Cumulative Repeat Transactions



Tracking Weekly Repeat Transactions



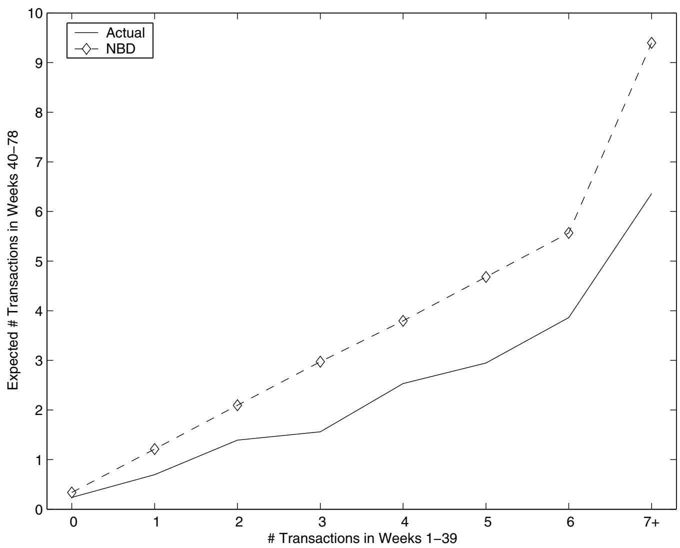
Conditional Expectations

- 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.

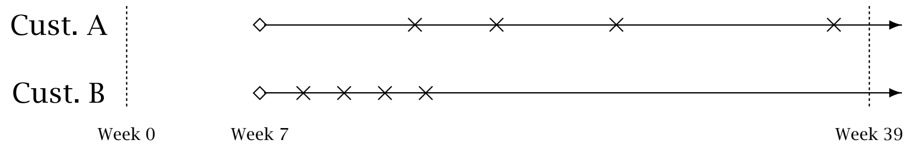
- For the NBD, a straight-forward application of Bayes' theorem gives us

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

Conditional Expectations



Conditional Expectations

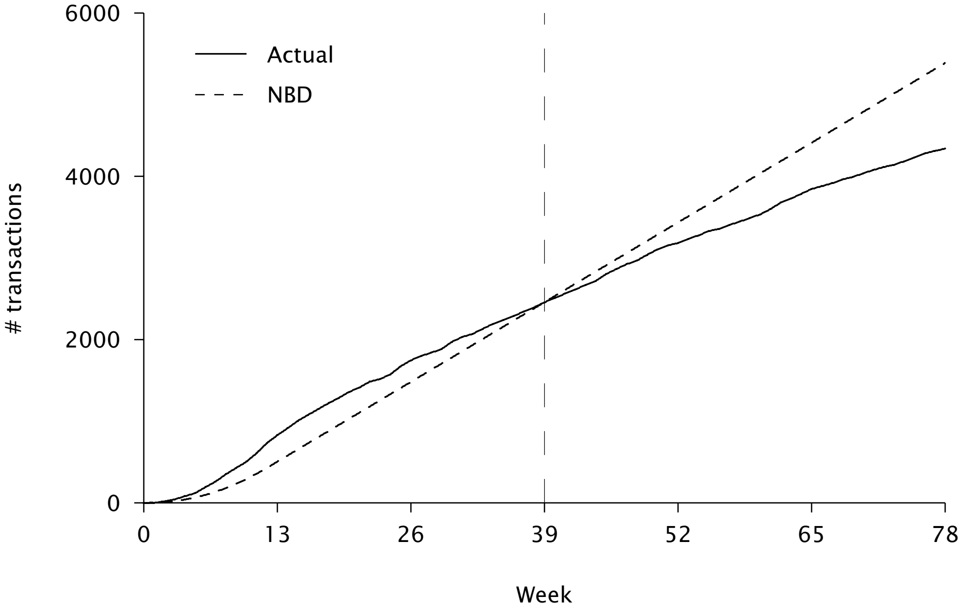


According to the NBD model:

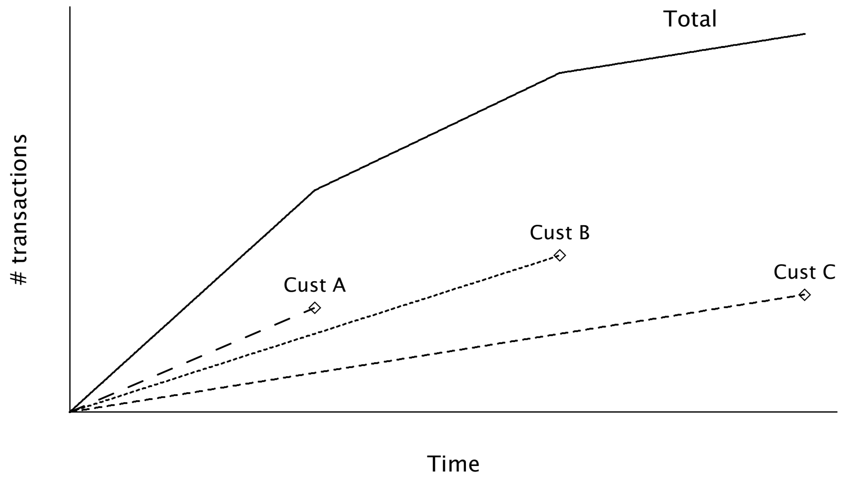
Cust. A: $E[Y(39) \mid x = 4, T = 32] = 3.88$

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,

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

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

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

Deriving the Model Likelihood Function

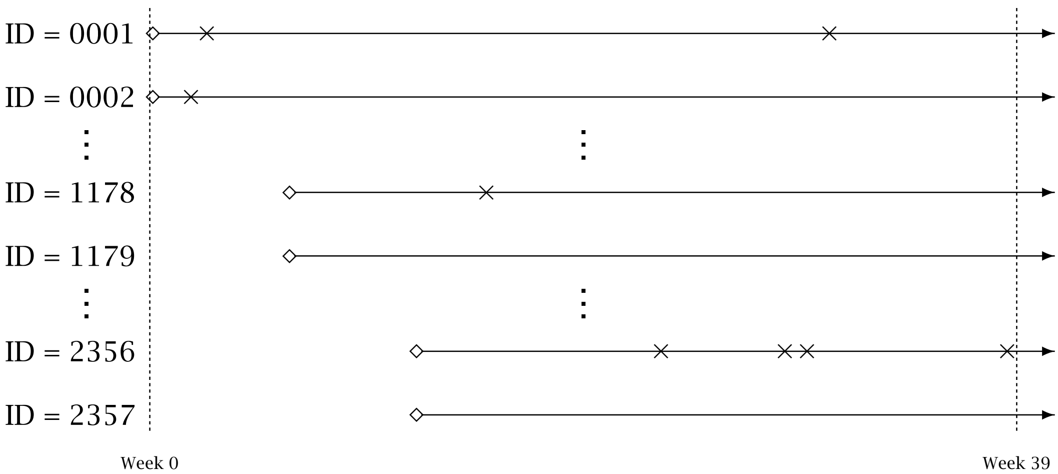
Removing the conditioning on τ,

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

Key Observation

- 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

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 $u_0 = 1$.

The Gaussian Hypergeometric Function

Euler's integral representation of the Gaussian hypergeometric function is

$$ {}_2F_1(a,b;c;z) = \frac{1}{B(b,c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a} dt $$

where and $c > b > 0$ and $B(\cdot,\cdot)$ is the beta function.

Pareto/NBD Likelihood Function

Removing the conditioning on λ and μ:

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

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

Key Results

$E[X(t)]$

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

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

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

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

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

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:

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

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

- 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

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

- Therefore,

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

Conditional Expectations

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

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

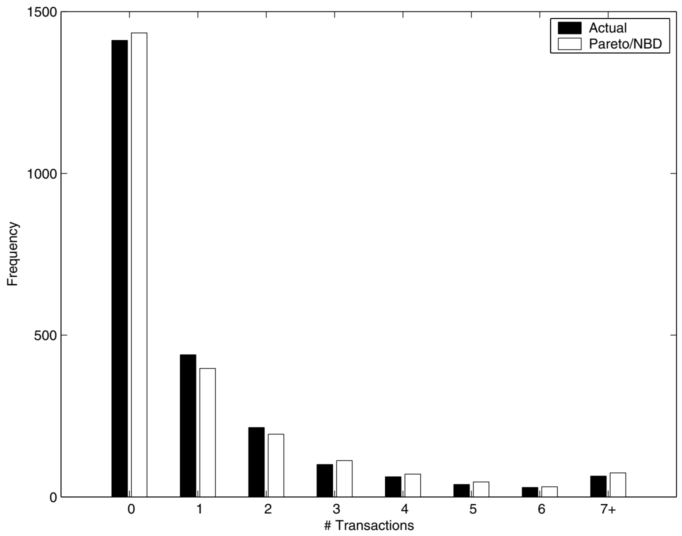
$$\begin{aligned} E[Y(t) \mid \lambda, \mu, x, t_x, T] &= E[Y(t) \mid \lambda, \mu, \text{alive at } T] \\ &\quad \times P(\tau > T \mid \lambda, \mu, x, t_x, T) \end{aligned}$$

Conditional Expectations

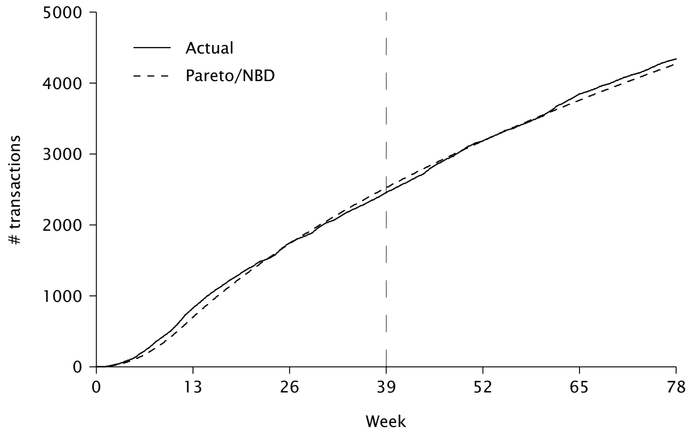
Taking the expectation over the joint posterior distribution of λ and μ yields:

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

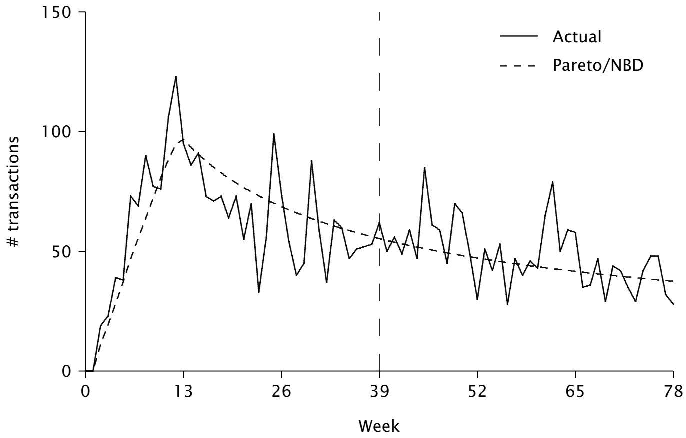
Frequency of Repeat Transactions



Tracking Cumulative Repeat Transactions

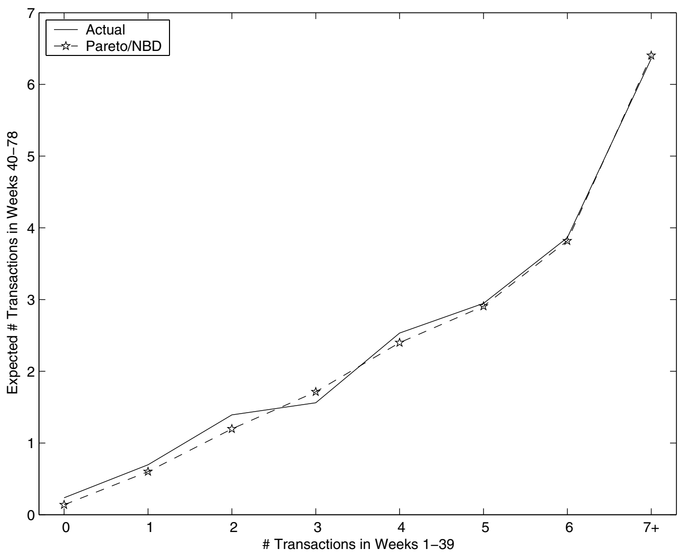


Tracking Weekly Repeat Transactions

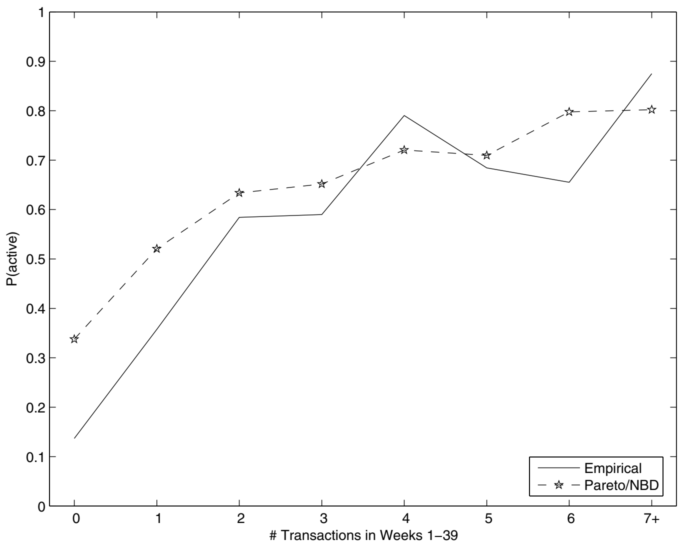


53

Conditional Expectations



Proportions of Active Customers



Computing 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

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

Computing CLV

- Standing at time T, we wish to compute the present value of the expected future transaction stream for a customer with purchase history (x, t_x, T).
- We call this quantity $DERT$, discounted expected residual transactions.

Computing DERT

- Standing at time T,

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

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

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

Computing DERT

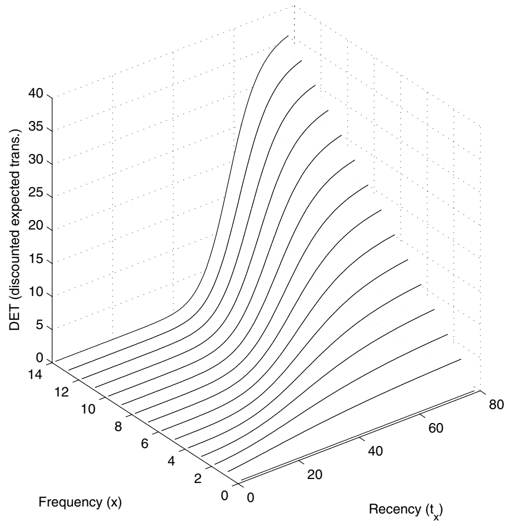
$$
\begin{aligned}
DERT(\delta \mid r, \alpha, s, \beta, x, t_x, T) \\
&= \int_0^\infty \int_0^\infty \Big\{ DERT(\delta \mid \lambda, \mu, \text{alive at } T) \\
&\qquad \times P(\text{alive at } T \mid \lambda, \mu, x, t_x, T) \\
&\qquad \times g(\lambda, \mu \mid r, \alpha, s, \beta, x, t_x, T) \Big\}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

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

Continuous Compounding

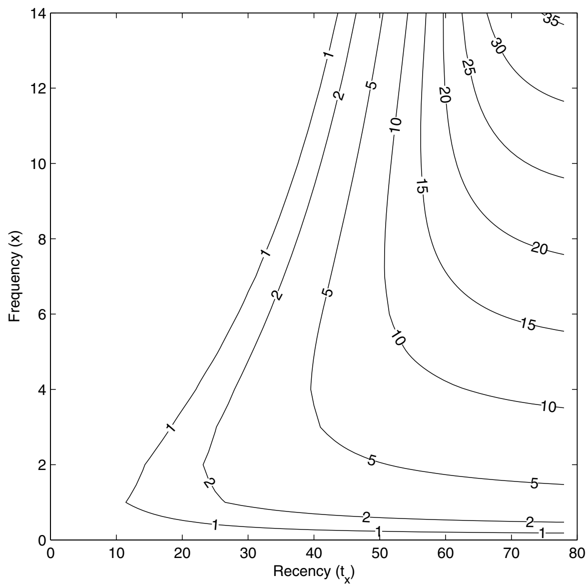
- An annual discount rate of $(100 \times d)\%$ is equivalent to a continuously compounded rate of $\delta = \ln(1 + d)$.
- If the data are recorded in time units such that there are k periods per year ($k = 52$ if the data are recorded in weekly units of time) then the relevant continuously compounded rate is $\delta = \ln(1 + d)/k$.

DERT by Recency and Frequency



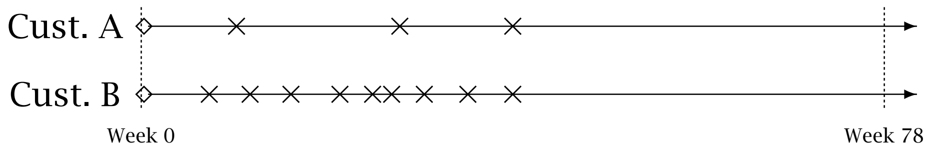
61

Iso-Value Representation of DERT



62

The “Increasing Frequency” Paradox



	DERT
Cust. A	4.6
Cust. B	1.9

Modeling 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.

Modeling the Spend Process

- For a customer with x transactions, let $z_1, z_2, \dots, z_x$ 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

Modeling 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

Modeling the Spend Process

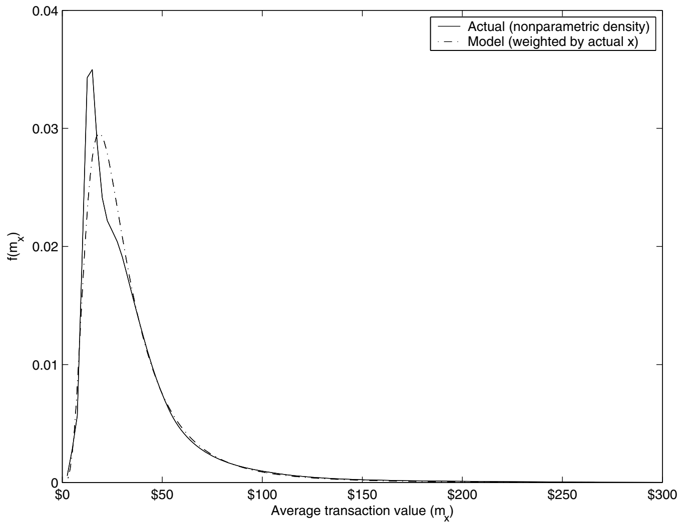
Marginal distribution for m_x:

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

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

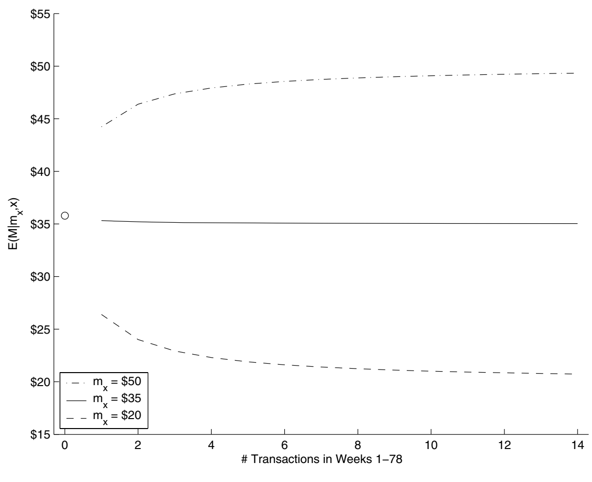
$$E(M|p,q,\gamma,m_x,x) = \left(\frac{q-1}{px+q-1}\right)\frac{\gamma p}{q-1} + \left(\frac{px}{px+q-1}\right) m_x$$

Distribution of Average Transaction Value



69

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

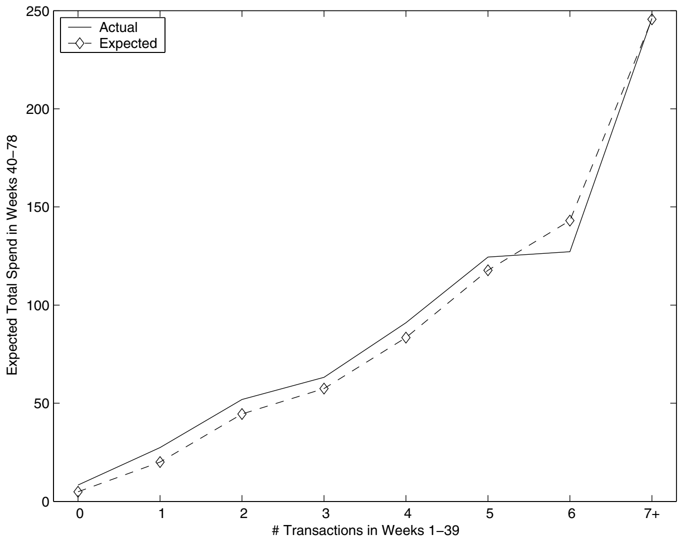


Conditional Expectations of Total Value

We are interested in predicting an individual's expected total spend in the future period $(T, T+t]$ conditional on his RFM

$$= \text{conditional expectation of transactions}$$
$$\times \text{ conditional expectation of revenue/transaction}$$
$$= E(Y(t) \mid r, \alpha, s, \beta, x, t_x, T) \times E(M|p, q, \gamma, m_x, x)$$

Conditional Expectations of Total Value



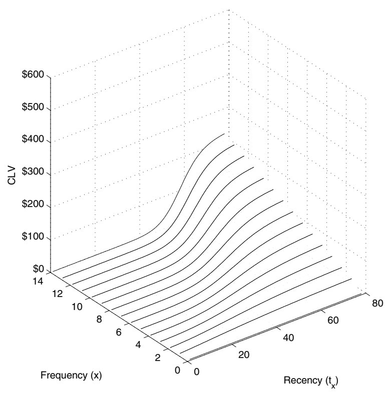
Computing Expected Residual CLV

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

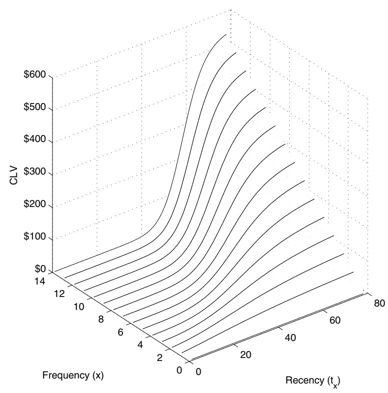
$$\begin{aligned} E(RCLV) &= \text{margin} \times \text{revenue/transaction} \times DERT \\ &= \text{margin} \times E(M|p,q,\gamma,m_x,x) \\ &\quad\ \times DERT(\delta \mid r,\alpha,s,\beta,x,t_x,T) \end{aligned}$$

Estimates of Expected Residual CLV

M = $20

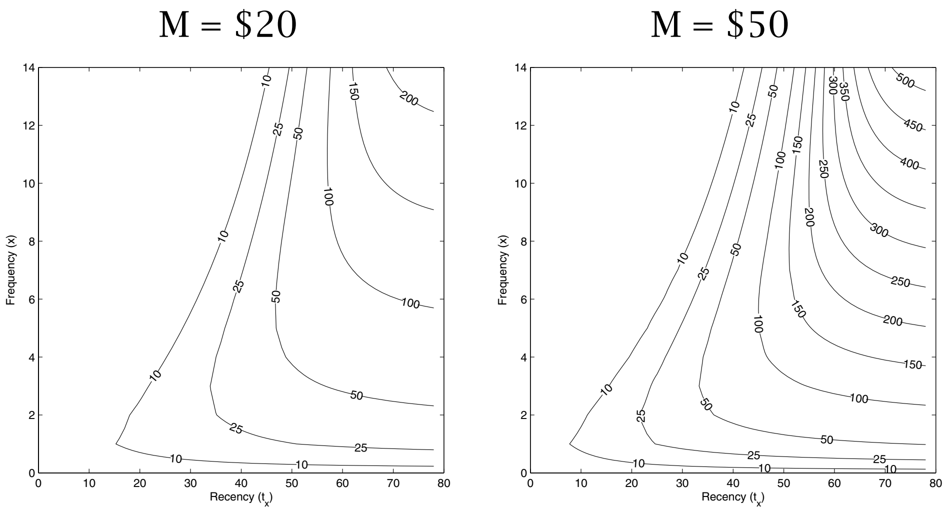


M = $50



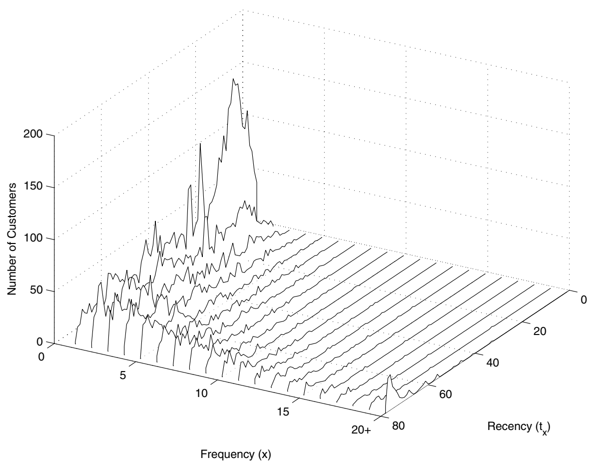
(Margin = 30%, 15% discount rate)

Estimates of Expected Residual CLV



(Margin = 30%, 15% discount rate)

Distribution of Repeat Customers



(12,054 customers make no repeat purchases)

Total CLV by RFM Group

		Recency			
	Frequency	0	1	2	3
M=0	0	$53,029			
M=1	1		$7,654	$9,893	$1,794
	2		$2,790	$15,260	$17,410
	3		$259	$12,477	$52,912
M=2	1		$5,863	$7,629	$2,341
	2		$3,551	$26,527	$25,754
	3		$450	$37,212	$203,004
M=3	1		$11,257	$19,738	$3,738
	2		$7,289	$45,911	$47,906
	3		$1,038	$62,698	$414,938

Substantive Observations

- There is a highly nonlinear relationship between recency/frequency and future transactions
  $\longrightarrow$ the "increasing frequency" paradox
- The underlying process for monetary value appears to be stationary and independent of recency and frequency
- A thorough analysis of the customer base requires careful consideration of the "zero class"
- Iso-value curves can be used to identify customers with different purchase histories but similar CLVs

Summary

- We are able to generate 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 estimates of expected (residual) CLV as a function of RFM in a noncontractual setting:

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

An Alternative to the Pareto/NBD Model

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

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

Purchase Process:

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

Dropout Process:

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

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:
  1. The customer is still alive at the end of the observation period (i.e., $\tau > T$)
  2. The customer became inactive immediately after the xth transaction (i.e., $\tau = t_x$)

Deriving the Model Likelihood Function

Conditional on λ,

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

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

Deriving the Model Likelihood Function

Removing the conditioning on τ,

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

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

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

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]
14	0007	1	5.00	38.86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
15	0008	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
16	0009	2	35.71	38.86	-9.5367	-0.8390	-0.4910	-8.4489	-9.7432
17	0010	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
2362	2355	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000
2363	2356	4	26.57	27.00	-14.1284	1.1450	-0.7922	-14.6252	-16.4902
2364	2357	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000

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

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

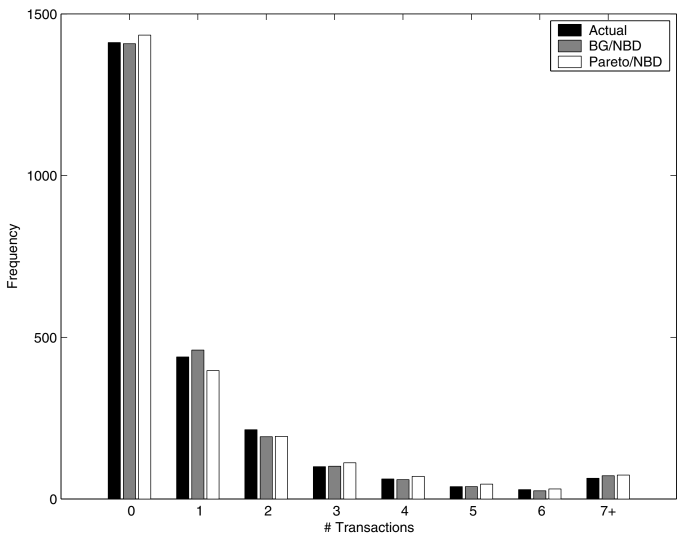
=-(B$1+B8)*LN(B$2+D8)

=SUM(E8:E2364)

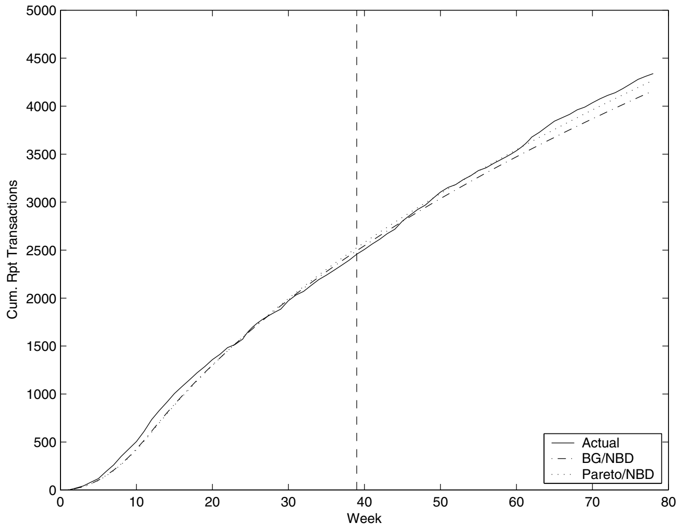
=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))

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

Frequency of Repeat Transactions

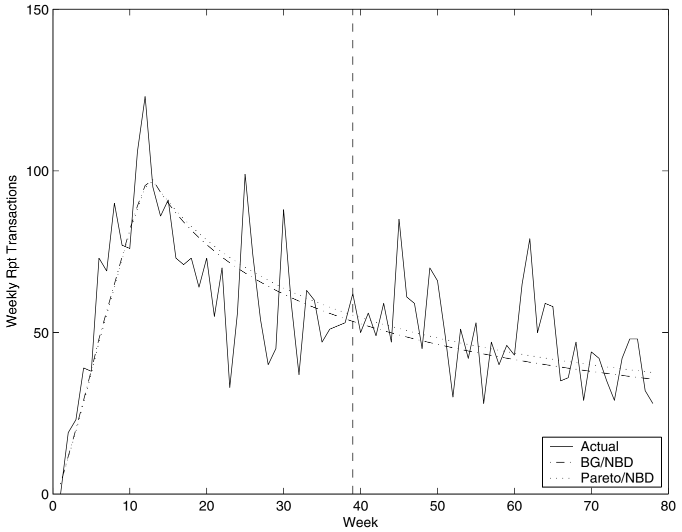


Tracking Cumulative Repeat Transactions



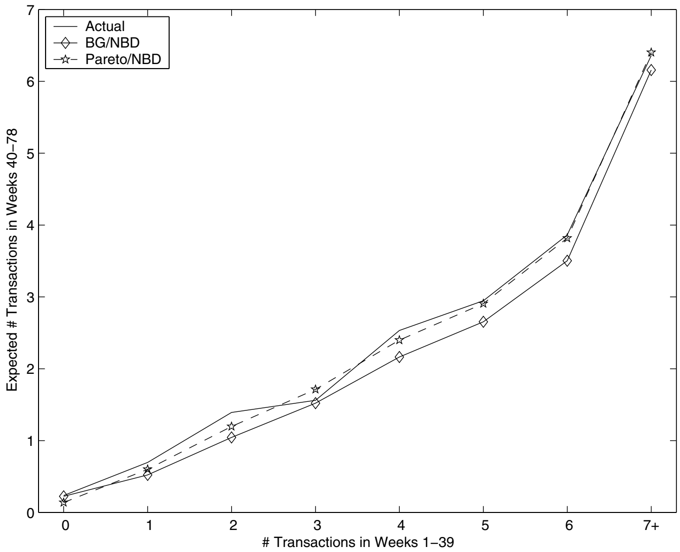
89

Tracking Weekly Repeat Transactions



90

Conditional Expectations



Further Reading

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

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

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

Further Reading

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

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

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

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

Modelling the Transaction Stream

How valid is the assumption of Poisson purchasing?

⟶ can transactions occur at any point in time?

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

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	[illegible]	[illegible]		[illegible]	[illegible]	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]	[illegible]	0.0087	0.0060	0.0043	0
19	0	0	4	4482	-1373.9	0.7360		[illegible]	[illegible]	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)

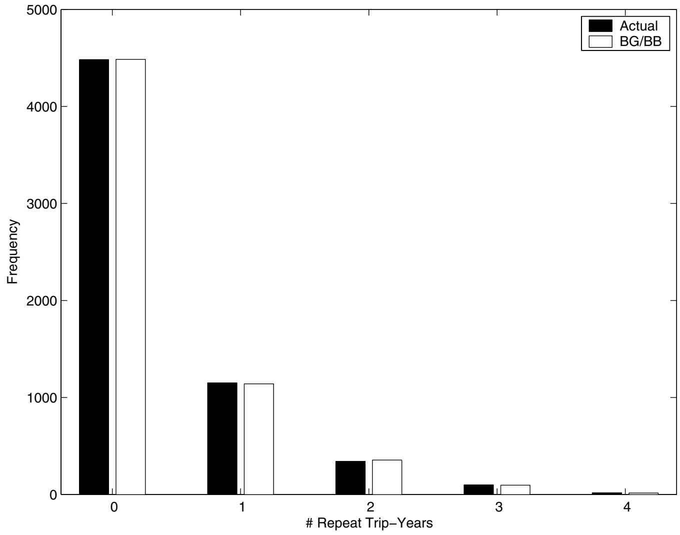
=IF($H9<J$8,0,EXP(GAMMALN($B$1+$A9)+GAMMALN($B$2+$B9-$A9+J$8)-GAMMALN($B$1+$B$2+$B9+J$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+J$8)-GAMMALN($B$3+$B$4+$B9+J$8+1))/$E$3)

=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,

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

Conditional Expectations

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

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

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

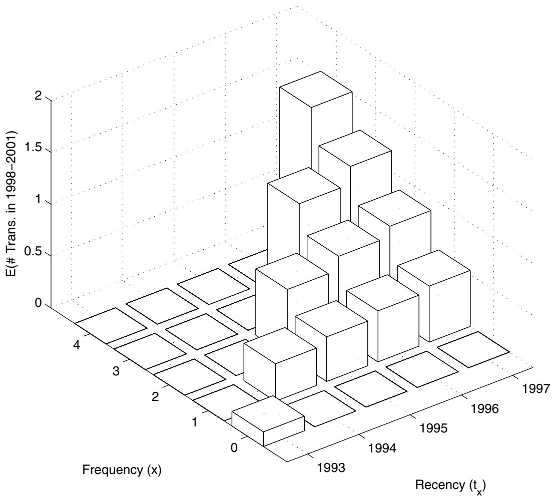
Posterior Mean of p as a Function of Recency and Frequency

	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



Computing DERT

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

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

- However, p and θ are unobserved.

Computing DERT

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

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

Computing DERT

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

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

DET

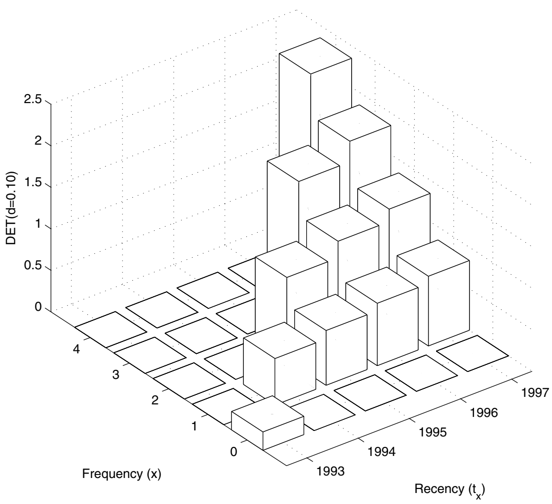
=EXP(GAMMALN($B$1+A11+1)+GAMMALN($B$2+C11-A11)-GAMMALN($B$1+$B$2+C11+1))*EXP(GAMMALN($B$3)+GAMMALN($B$4+C11+1)-GAMMALN($B$3+$B$4+C11+1))*$K$24/(E$1*E$3*(1+$B$6))/F11

=SUM(M24:M174)

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

	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		

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

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

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

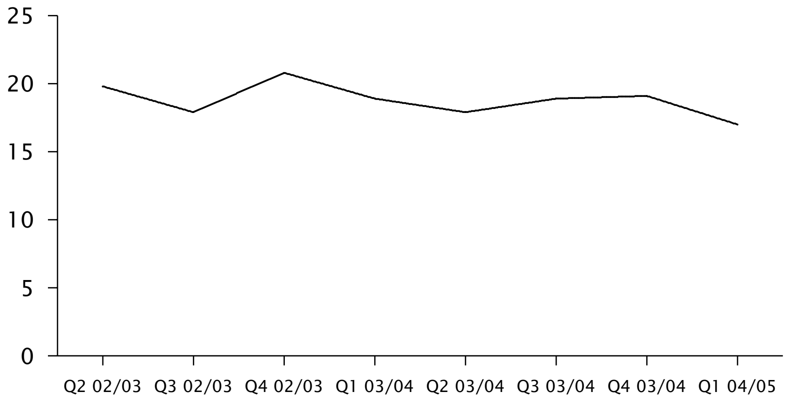
Type of Relationship With Customers

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

Vodafone Germany
Quarterly Annualized Churn Rate (%)



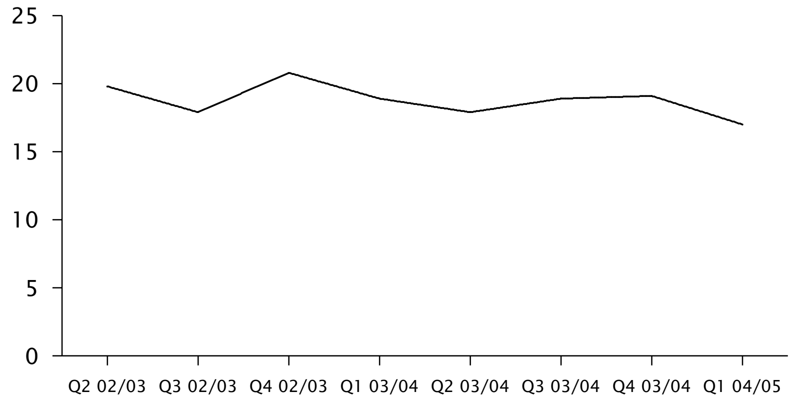
Source: Vodafone Germany "Vodafone Analyst & Investor Day" presentation (2004-09-27)

Valuing the Customer Base

Assuming a 10% discount rate, a MBA student's back-of-the-envelope calculation would value Vodafone Germany's existing customer base as

$$\begin{aligned} E(\text{Residual Value}) &= \#\text{ customers} \times \text{average margin per user} \\ &\quad \times \sum_{t=1}^{\infty} \frac{(1-0.18)^t}{(1+0.1)^t} \\ &= 2.93 \times \#\text{ customers} \times \text{AMPU} \end{aligned}$$

Vodafone Germany
Quarterly Annualized Churn Rate (%)



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

A Look Behind the Aggregate Retention Rates

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

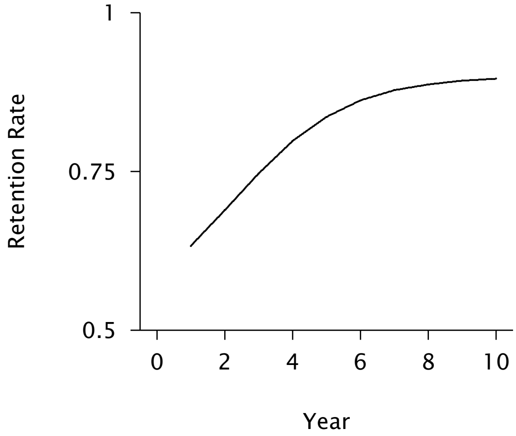
2000	2001	2002	2003	2004
13,000	8,233	5,677	4,242	3,385
	15,000	9,500	6,550	4,895
		17,500	11,083	7,642
			19,000	12,033
				20,500
13,000	23,233	32,677	40,875	48,455
	0.63	0.65	0.67	0.68

A Look Behind the Aggregate Retention Rates

Annual retention rates by cohort:

2000	2001	2002	2003	2004
--	0.63	0.69	0.75	0.80
	--	0.63	0.69	0.75
		--	0.63	0.69
			--	0.63
				--
--	0.63	0.65	0.67	0.68

A Real-World Consideration



In practice, we (almost) always observe increasing retention rates.

A Real-World Example

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.

Calculating CLV

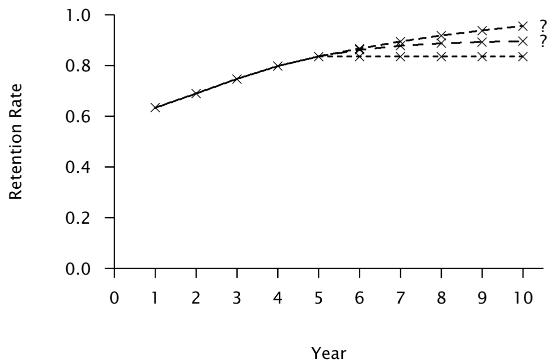
Modified classroom formula:

$$E(\text{CLV}) = \sum_{t=0}^{\infty} m\Big\{\prod_{i=0}^{t} r_i\Big\}\Big/\Big(\frac{1}{1+d}\Big)^t$$

where the retention rate for period i (r_i) is defined as the proportion of customers active at the end of period $i-1$ who are still active at the end of period i.

Implementation Questions

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



- What is the expected (remaining) CLV of a customer who has been with us for n periods?

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

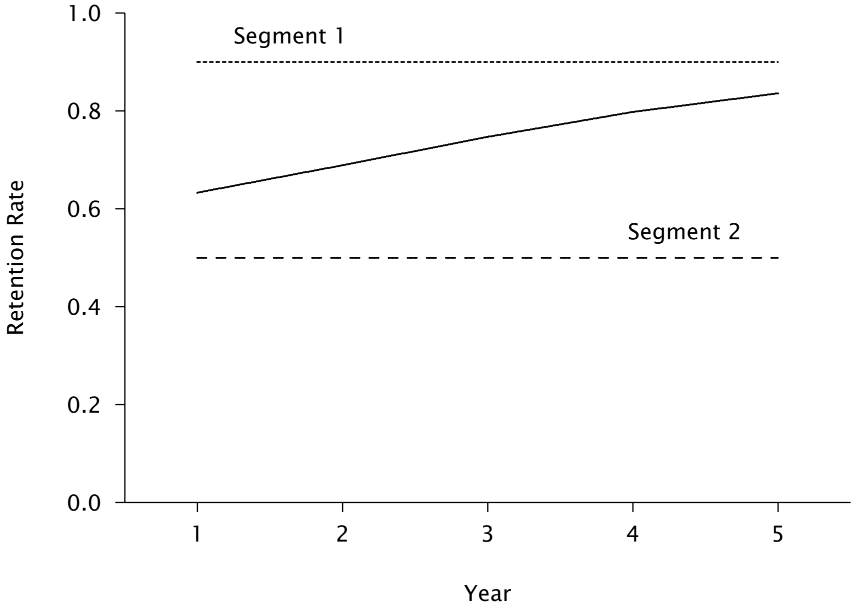
Suppose we track a cohort of 15,000 customers, comprising two underlying segments:

- Segment 1 comprises 5,000 customers, each with a time-invariant annual retention probability of 0.9.
- Segment 2 comprises 10,000 customers, each with a time-invariant annual retention probability of 0.5.

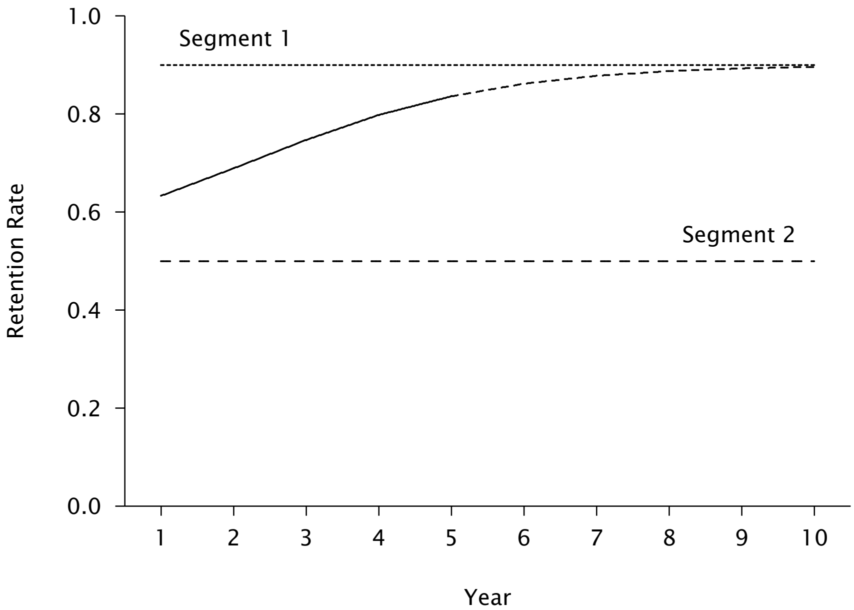
The Role of Heterogeneity

	# Active Customers			r_t		
Year	Seg 1	Seg 2	Total	Seg 1	Seg 2	Total
0	5,000	10,000	15,000			
1	4,500	5,000	9,500	0.900	0.500	0.633
2	4,050	2,500	6,550	0.900	0.500	0.689
3	3,645	1,250	4,895	0.900	0.500	0.747
4	3,281	625	3,906	0.900	0.500	0.798
5	2,953	313	3,266	0.900	0.500	0.836

The Role of Heterogeneity

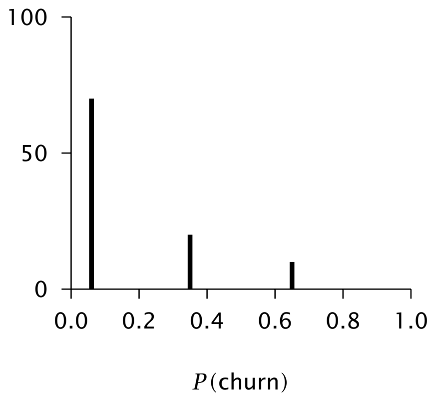


Projecting Retention Rates Given "Story"



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)

What is the expected residual CLV of a customer with a tenure of n periods?

Fleshing Out the Story

- Assume we're dealing with an insurance company.
- The annual gross contribution (after accounting for retention marketing activities) equals $100.
- Contracts renewed on 1 January.
- The finance department has suggested that a 10% discount rate for marketing activities.

$E[RCLV(d = 10\% \mid \text{survived two years})]$

- If this person belongs to segment 1:

$$E[RCLV(d = 10\%)] = \sum_{t=1}^{\infty} 100 \times \frac{0.9^t}{(1+0.1)^{t-1}}$$
$$= \$495.0$$

- If this person belongs to segment 2:

$$E[RCLV(d = 10\%)] = \sum_{t=1}^{\infty} 100 \times \frac{0.5^t}{(1+0.1)^{t-1}}$$
$$= \$91.7$$

But to which segment does this person belong?

The Role of Heterogeneity

	# Active Customers			r_t		
Year	Seg 1	Seg 2	Total	Seg 1	Seg 2	Total
0	5,000	10,000	15,000			
1	4,500	5,000	9,500	0.900	0.500	0.633
2	4,050	2,500	6,550	0.900	0.500	0.689
3	3,645	1,250	4,895	0.900	0.500	0.747
4	3,281	625	3,906	0.900	0.500	0.798
5	2,953	313	3,266	0.900	0.500	0.836

$E[RCLV(d = 10\% \mid \text{survived two years})]$

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

$$\frac{P(\text{survived two years} \mid \text{segment 1}) \times P(\text{segment 1})}{P(\text{survived two years})}$$

$$= \frac{0.9^2 \times 0.333}{0.9^2 \times 0.333 + 0.5^2 \times 0.667}$$

$$= 0.618$$

$E[RCLV(d = 10\% \mid \text{survived two years})]$

It follows that the CLV for an individual with a tenure of two years is

$$
\begin{aligned}
E[RCLV(d &= 10\% \mid \text{survived two years})] \\
&= CLV(d = 10\% \mid \text{seg. } 1)P(\text{seg. } 1 \mid \text{survived two yrs}) \\
&\quad + CLV(d = 10\% \mid \text{seg. } 2)P(\text{seg. } 2 \mid \text{survived two yrs}) \\
&= 495.0 \times 0.618 + 91.7 \times (1 - 0.618) \\
&= 341.1
\end{aligned}
$$

Implications

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

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

Our two-segment model is but one such story ...

How much is our existing customer base worth?

Valuing an Existing Customer Base

As we move from a single cohort of customers (defined by year-of-acquisition), to a customer base composed of a number of different cohorts, we must condition any calculations on time-of-entry into the firm's customer base.

Valuing an Existing Customer Base

Number of active customers by cohort:

2000	2001	2002	2003	2004
13,000	8,233	5,677	4,242	3,385
	15,000	9,500	6,550	4,895
		17,500	11,083	7,642
			19,000	12,033
				20,500
13,000	23,233	32,677	40,875	48,455

Agg. 03–04 retention rate = 27,955/ 40,875 = 0.684

Valuing an Existing Customer Base

Standing at December 31, 2004, how do we compute the total CLV for the existing customer base?

Two approaches:

- Naïve (aggregate) retention rate (0.684)
- Segment-based (i.e., model-based)

Approach 1

Using the aggregate 03–04 retention rate:

$$\begin{aligned}\text{Total CLV} &= 48,455 \times \sum_{t=1}^{\infty} 100 \times \frac{0.684^t}{(1+0.1)^{t-1}} \\ &= \$8,761,000\end{aligned}$$

Approach 2

Recognizing the underlying segments:

Cohort	# Active in 2004	P(seg. 1)
2000	3,385	0.840
2001	4,895	0.745
2002	7,642	0.618
2003	12,033	0.474
2004	20,500	0.333

Total CLV = \$14,020,000

Valuing an Existing Customer Base

Cohort	Total CLV	“Loss”
Naïve	$8,761,000	38%
Segment (model)	$14,020,000	

Exploring the Magnitude of the “Loss”

- Systematically vary heterogeneity in retention rates
- First need to specify 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 CLV

- Recall:

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

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

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

Computing DERL

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

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

- But θ is unobserved ....

Computing DERL

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

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

- It follows that

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

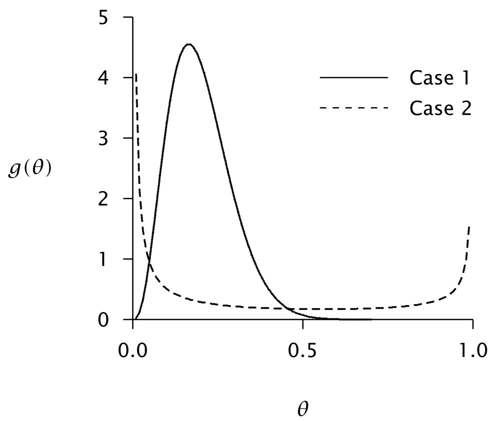
Impact of Heterogeneity on Error

- Assume the following arrival of new customers:

2000	2001	2002	2003	2004
10000	15000	17500	19000	20500

- Assume a 10% discount rate.
- For given values of α and β, determine the "loss" associated with computing the 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	$E(\theta)$
1	3.80	15.20	0.20
2	0.067	0.267	0.20

Number of Active Customers: Case 1

2000	2001	2002	2003	2004	n	CLV
10,000	8,000	6,480	5,307	4,391	4	3.84
	15,000	12,000	9,720	7,961	3	3.72
		17,500	14,000	11,340	2	3.59
			19,000	15,200	1	3.45
				20,500	0	3.31
10,000	23,000	35,980	48,027	59,392		

Aggregate 03–04 retention rate = 38,892/48,027 = 0.81

Number of Active Customers: Case 2

2000	2001	2002	2003	2004	n	CLV
10,000	8,000	7,600	7,383	7,235	4	10.19
	15000	12000	11,400	11,074	3	10.06
		17500	14000	13,300	2	9.86
			19000	15200	1	9.46
				20500	0	7.68
10,000	23,000	37,100	51,783	67,309		

Aggregate 03–04 retention rate = 46,809/51,783 = 0.90

Impact of Heterogeneity on Error: Case 1

$$\begin{aligned}\text{Naïve valuation} &= 59{,}392 \times \sum_{t=1}^{\infty} \frac{0.81^t}{(1+0.1)^{t-1}} \\ &= 182{,}292\end{aligned}$$

$$\begin{aligned}\text{Correct valuation} &= 4{,}391 \times 3.84 + 7{,}961 \times 3.72 \\ &\quad + 11{,}340 \times 3.59 + 15{,}200 \times 3.45 \\ &\quad + 20{,}500 \times 3.31 \\ &= 207{,}438\end{aligned}$$

Naïve underestimates correct by 12%.

Impact of Heterogeneity on Error: Case 2

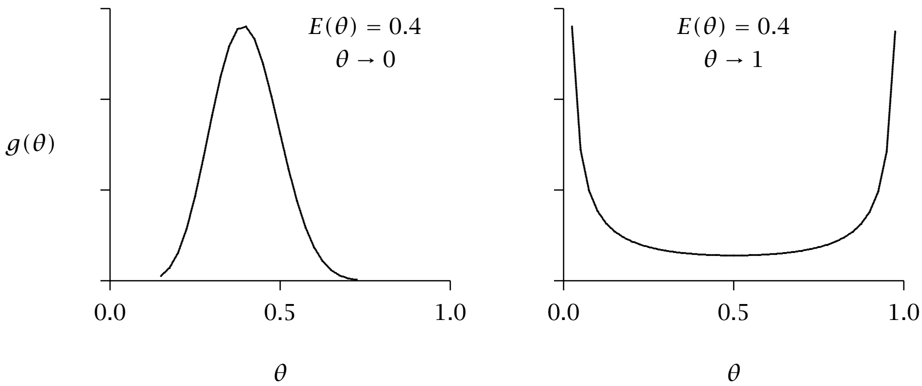
$$\begin{aligned}\text{Naïve valuation} &= 67{,}309 \times \sum_{t=1}^{\infty} \frac{0.9^t}{(1+0.1)^{t-1}} \\ &= 341{,}402\end{aligned}$$

$$\begin{aligned}\text{Correct valuation} &= 7{,}235 \times 10.19 + 11{,}074 \times 10.06 \\ &\quad + 13{,}300 \times 9.86 + 15{,}200 \times 9.46 \\ &\quad + 20{,}500 \times 7.68 \\ &= 617{,}536\end{aligned}$$

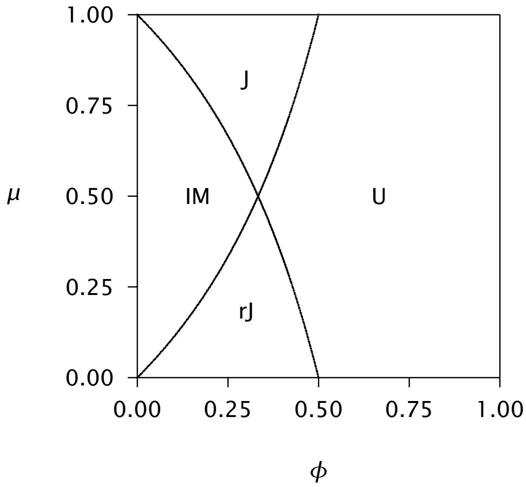
Naïve underestimates correct by 45%.

Interpreting the Beta Distribution Parameters

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

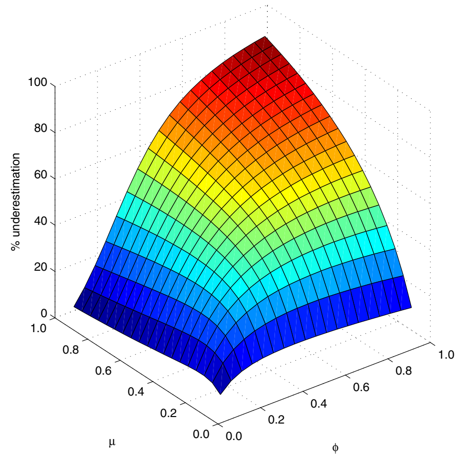


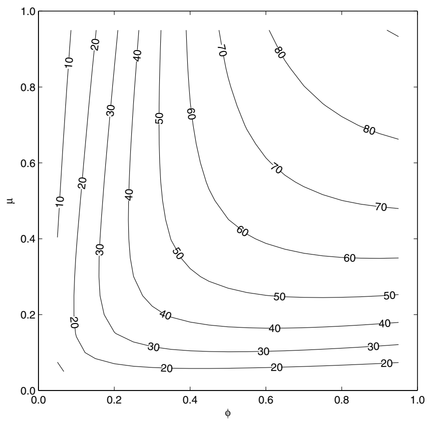
Shape of the Beta Distribution



Error as a Function of μ and ϕ

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





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

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

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

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

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

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

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

Beyond the Basic Models

Customer-Base Analysis

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

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

when faced with a customer transaction database.

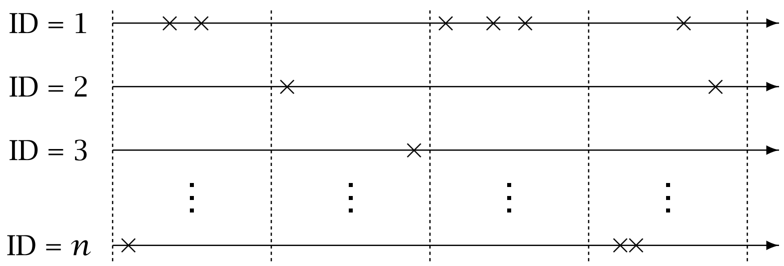
Philosophy of Model Building

- 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.

Implementation Issues

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

Implications of Data Recording Processes



The model likelihood function must match the data structure:

- Interval-censored individual-level data
- Period-by-period (cross-sectional) histograms

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

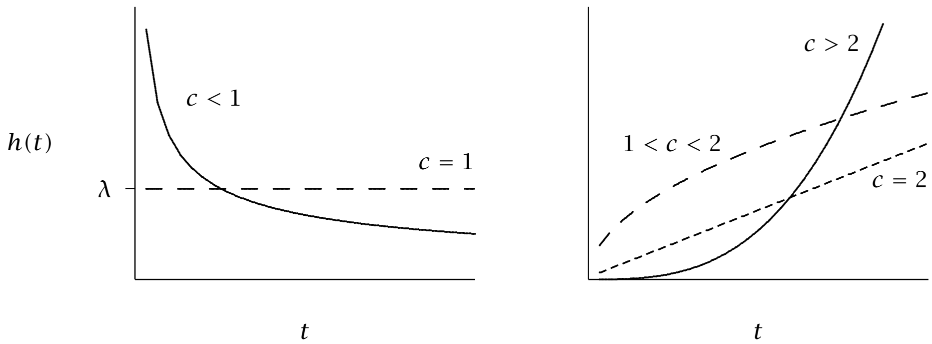
- A generalization of the exponential distribution that can have an increasing and decreasing hazard function:

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

  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.

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)
  - used to create segments (and apply no-covariate models)
- Need to be wary of endogeneity bias and sample selection effects

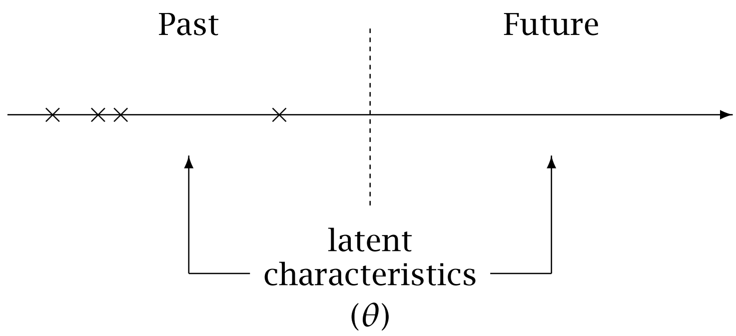
The Cost of Model Extensions

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

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