Introduction to Probability Models in Marketing

Oded Netzer*

Columbia Business School

* Slide thanks to Professor Eva Ascarza

The Data

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
100001	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
100004	1	0	1	0	1	1	1	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
100008	1	1	1	1	1	1	1	?	?	?	?	?
100009	1	1	1	1	1	1	0	?	?	?	?	?
100010	1	1	1	1	1	0	0	?	?	?	?	?
...												
111102	1	1	1	1	1	1	1	?	?	?	?	?
111103	1	0	1	1	0	1	1	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

Modelling the Transaction Stream

- Let random variable $X(n)$ denote the # transactions across n consecutive transaction opportunities.
- The customer buys at any given transaction opportunity with probability p:

$$P(X(n) = x \mid p) = \binom{n}{x} p^x (1-p)^{n-x} .$$

- Purchase probabilities (p) are distributed across the population according to a beta distribution:

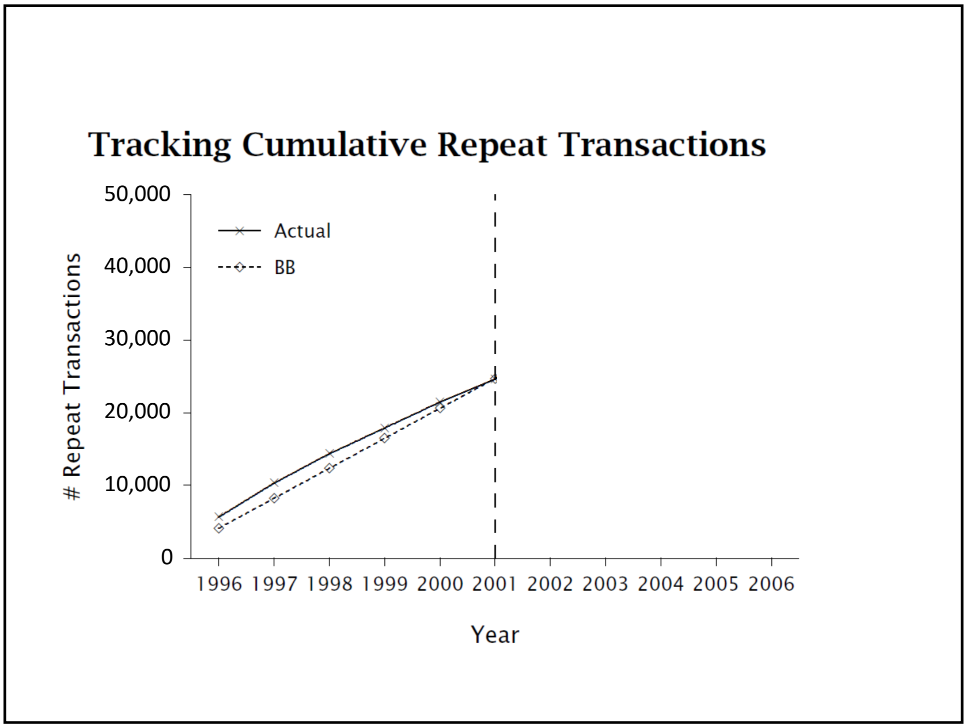
$$g(p \mid \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} .$$

Modelling the Transaction Stream

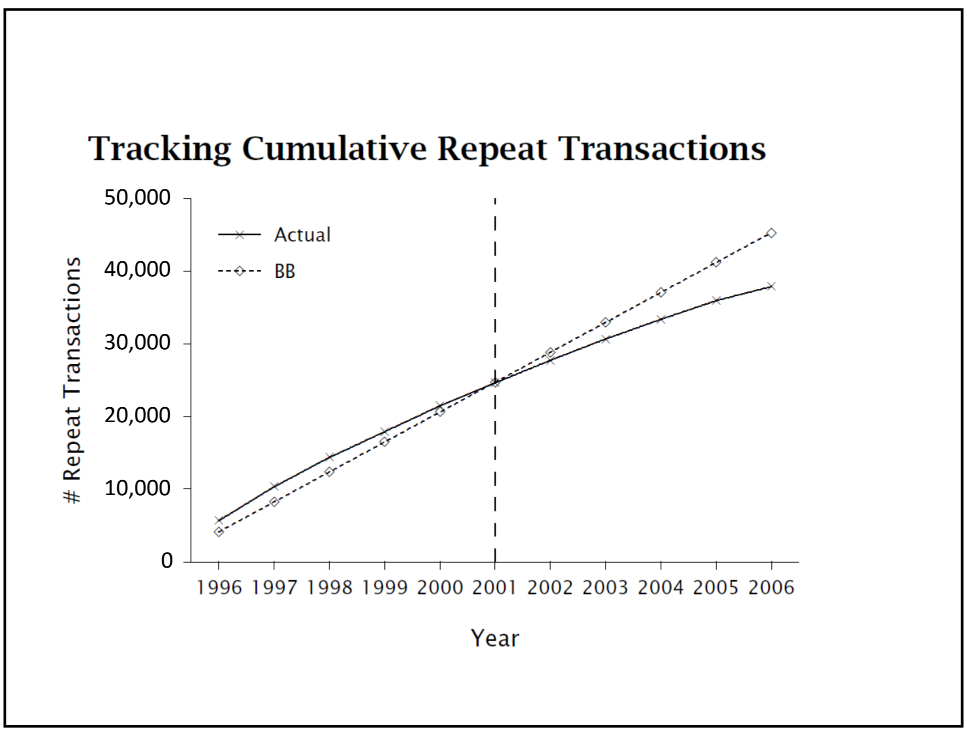
The distribution of transactions for a randomly-chosen individual is given by:

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

which is the beta-binomial (BB) distribution.

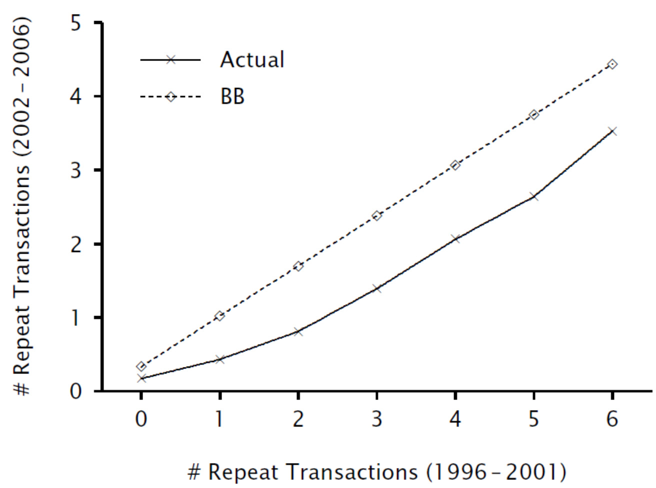


Tracking Cumulative Repeat Transactions
50,000
40,000
30,000
20,000
10,000
0
Repeat Transactions
Actual
BB
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006
Year



Tracking Cumulative Repeat Transactions
50,000
40,000
30,000
20,000
10,000
0
Repeat Transactions
Actual
BB
1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006
Year

Conditional Expectations



The Data

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Sara	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
Mary	1	0	1	0	1	1	1	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
Bob	1	1	1	1	1	1	1	?	?	?	?	?
100009	1	1	1	1	1	1	0	?	?	?	?	?
Pete	1	1	1	1	1	0	0	?	?	?	?	?
...												
111102	1	1	1	1	1	1	1	?	?	?	?	?
Chris	1	0	1	1	0	1	1	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

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 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 with probability θ

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

Model Development

Consider the following transaction pattern:

1996	1997	1998	1999	2000	2001
1	0	0	1	0	0

- The customer must have been alive in 1999 (and therefore in 1996–1998)
- Three scenarios give rise to no purchasing in 2000 and 2001

1996	1997	1998	1999	2000	2001
A	A	A	A	D	D
A	A	A	A	A	D
A	A	A	A	A	A

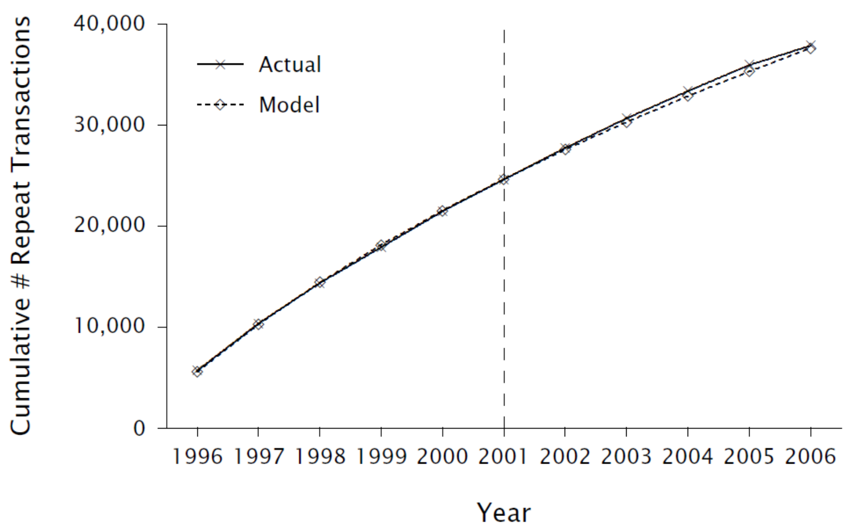
Model Development

Removing the conditioning on the latent traits p and θ,

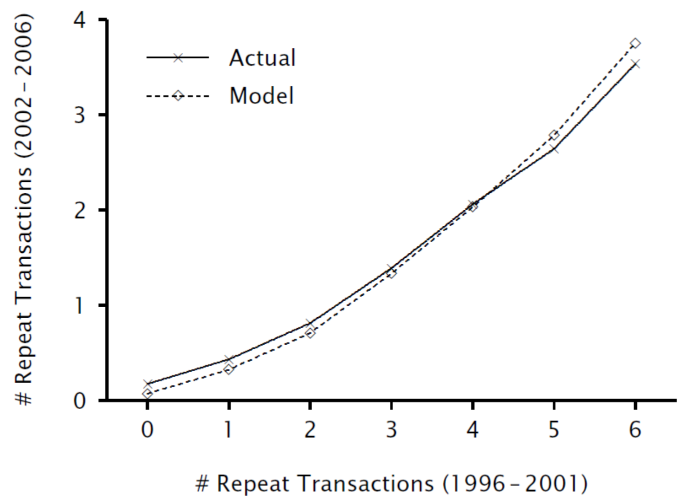
$$\begin{aligned}
L(\alpha,\beta,\gamma,\delta \mid x, t_x, n) &= \int_0^1 \int_0^1 L(p,\theta \mid x, t_x, n) g(p \mid \alpha,\beta) g(\theta \mid \gamma,\delta)\, dp\, d\theta \\
&= \frac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)} \frac{B(\gamma,\delta+n)}{B(\gamma,\delta)} \\
&\quad + \sum_{i=0}^{n-t_x-1}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

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

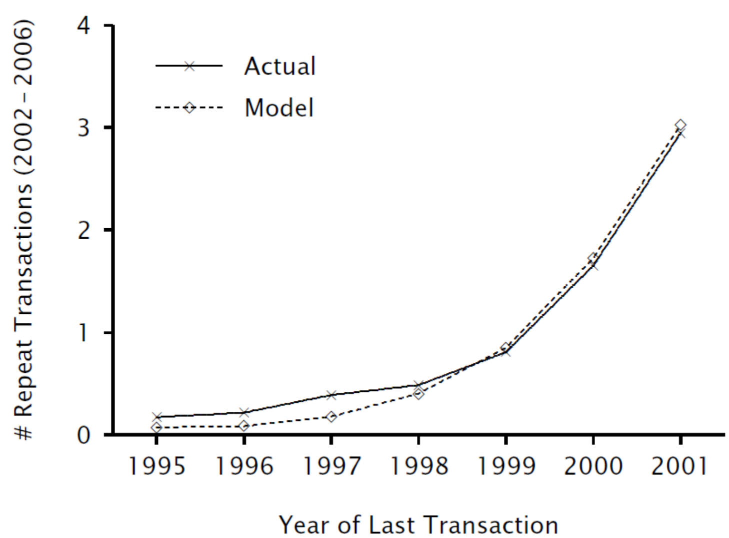
Tracking Cumulative Repeat Donations



Conditional Expectations by Frequency



Conditional Expectations by Recency



Expected # Transactions in 2002 – 2006

Bob (R:6, F:6) is expected to donate 3.75 times out of 5 opportunities between 2002 and 2006, surprisingly low given his 100% donation rate

Mary (R:6, F:4) and **Pete** (R:4, F:4) made the same number of transactions, but their recency is different. **Mary** is expected to donate almost 5 times more than **Pete**.

Sarah, with very low R and F, is lapsed and/or a very light donor

# Rpt Trans (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.07						
1		0.09	0.31	0.59	0.84	1.02	1.15
2			0.12	0.54	1.06	1.44	1.67
3				0.22	1.03	1.80	2.19
4					0.58	2.03	2.71
5						1.81	3.23
6							3.75

Probability of being alive in 2002

Bob (R:6, F:6) is expected to donate 3.75 times out of 5 opportunities between 2002 and 2006, surprisingly low given his 100% donation rate

Mary (R:6, F:4) and **Pete** (R:4, F:4) made the same number of transactions, but their recency is different. **Mary** is expected to donate almost 5 times more than **Pete**.

Sarah, with very low R and F, is lapsed and/or a very light donor (hard to tell)

# Rpt Trans (1996-2001)	Year of Last Transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.11						
1		0.07	0.25	0.48	0.68	0.83	0.93
2			0.07	0.30	0.59	0.80	0.93
3				0.10	0.44	0.77	0.93
4					0.20	0.52	0.93
5						1.81	0.93
6							0.93

Summary of BGBB model

- Noncontractual setting
  - Transactions: Binary process
  - "Death" process: Discrete time
- Data Generating Process
- Conditional expectations
- P(alive)

Fader, Peter S., Bruce G.S. Hardie, and Jen Shang (2010), "Customer-Base Analysis in a Discrete-Time Noncontractual Setting," *Marketing Science*, 29 (6), 1086-1108.

Different Business Settings

Noncontractual setting with "infinite" time periods

- Transactions
  - Binomial → Poisson
  - BB → NBD

- "Death" process
  - Geometric → Exponential
  - BG → EG (Pareto of second kind)

Different Business Settings

What if "death" was observed?

What if individual transaction behavior changes over time?

Remember…

- Behavior predicts behavior (Duh!)
- Data generating process – likelihood
- Do you observe "death"?
- Validate your model!
- You know how to:
  - Build BB → BG/BB → BG/NBD
- How would you…
  - Add other behaviors?
  - Incorporate dynamics?