

# **Applied Probability Models in Marketing Research: Extensions**

Bruce G. S. Hardie  
London Business School  
bhardie@london.edu

Peter S. Fader  
University of Pennsylvania  
faderp@wharton.upenn.edu

12th Annual Advanced Research Techniques Forum  
June 24-27, 2001

©2001 Bruce G. S. Hardie and Peter S. Fader

1

## **Probability Models 101**

2

## **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.

3

## **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

4

## 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*.

5

## 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:

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

6

## **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.

7

## **“Classes” of Models**

- The first tutorial introduced simple models for three behavioral processes:
  - Timing → “when”
  - Counting → “how many”
  - “Choice” → “whether/which”
- Each of these simple models has multiple applications.
- More complex behavioral phenomena can be captured by combining models from each of these processes.

8

## Outline

- Problem 4: Who is Visiting khakichinos.com?  
(Incorporating Covariates in Count Models)
- Problem 5: Understanding the Adoption of a Video-on-Demand Service  
(Introducing Additional Model Structures)
- Problem 6: Modeling Repeat Purchase Quantities at CDNOW  
(Building an “Integrated” Model)

### **Problem 4: 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.

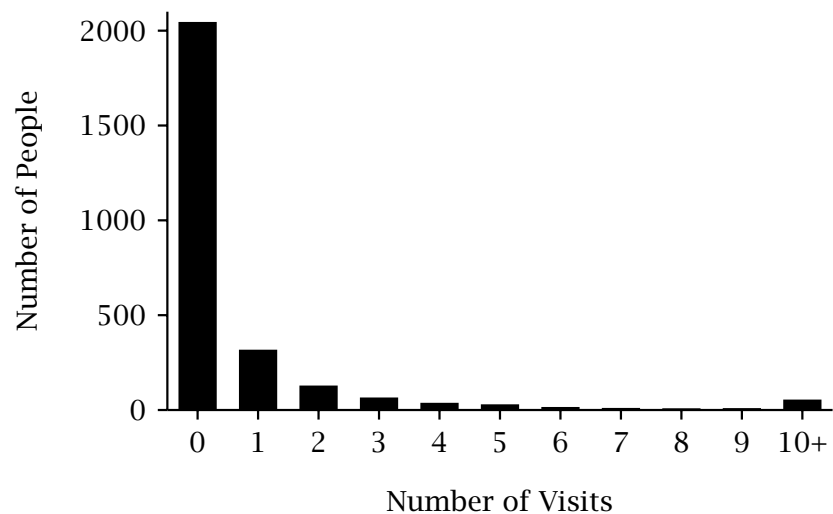
11

## Raw Data

ID	# Visits	ln(Income)	Sex	ln(Age)	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
...					

12

## Distribution of Visits



13

## Modeling Count Data

Recall the NBD:

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

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

14

## Observed vs. Unobserved Heterogeneity

Unobserved Heterogeneity:

- People differ in their mean (visiting) rate  $\lambda$
- To account for heterogeneity in  $\lambda$ , 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  $\lambda$  to her observable characteristics

15

## 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  $\lambda_i$ :

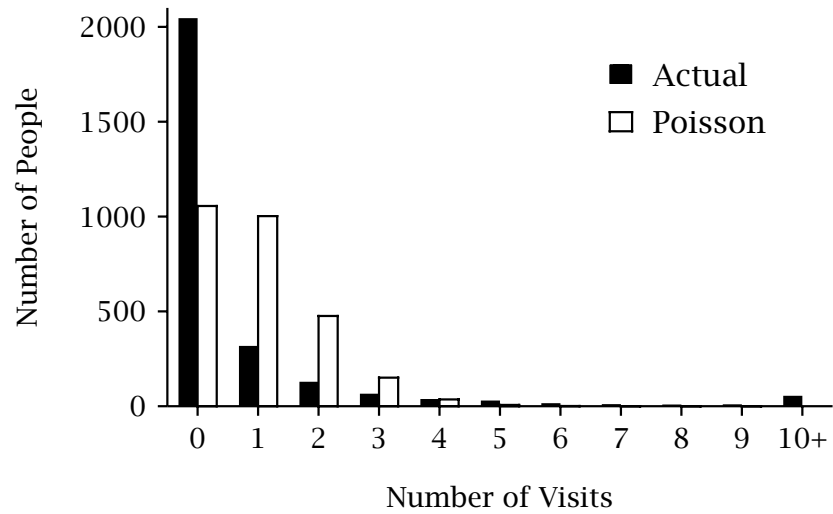
$$P(Y_i = y | \lambda_i) = \frac{\lambda_i^y e^{-\lambda_i}}{y!}$$

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

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

16

## Fit of Poisson



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

17

## Poisson Regression Results

Variable	Coefficient
$\lambda_0$	0.0439
Income	0.0938
Sex	0.0043
Age	0.5882
Size	-0.0359
<i>LL</i>	-6291.5
<i>LL</i> <sub>Poiss</sub>	-6378.6
LR (df = 4)	174.2

18

## Comparing Expected Visit Behavior

	Person A	Person B
Income	54,598	90,017
Sex	M	F
Age	55	33
Size	4	2

Who is less likely to have visited the web site?

19

## Is $\beta$ 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  $\beta = \mathbf{0}$ .

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

20

## 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  $LR > \chi^2_{.05,k}$

21

## Computing Standard Errors

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

22

## 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
Size	-0.035907406	0.01528397	-2.3493511

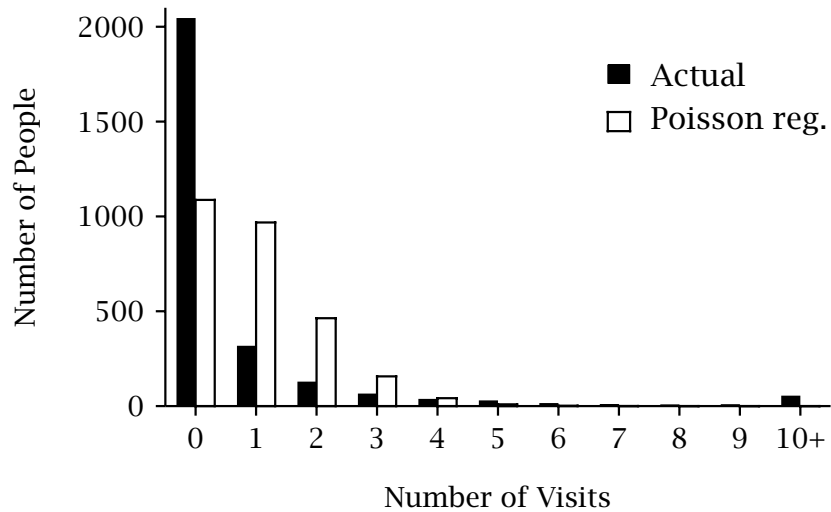
23

## 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
SIZE	-.3577795361E-01	.15287122E-01	-2.340

24

## Fit of Poisson Regression



25

## The ZIP Regression Model

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

- a proportion  $\pi$  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

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

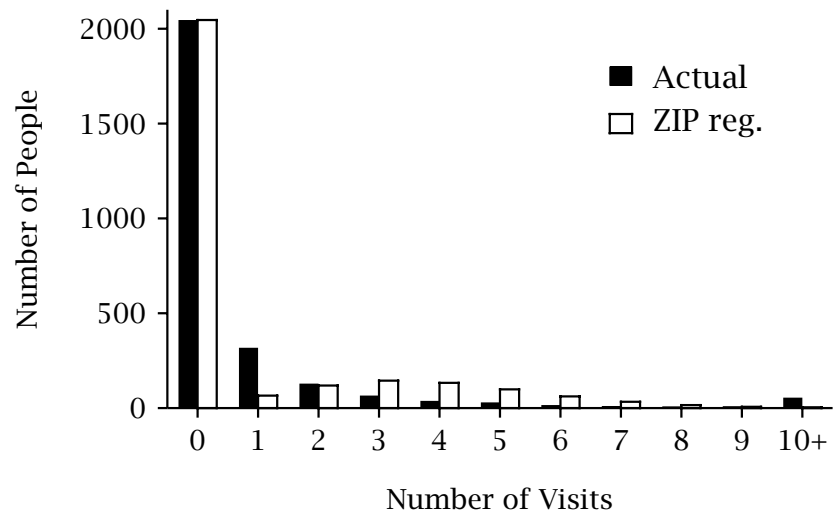
26

## ZIP Regression Results

Variable	Coefficient
$\lambda_0$	6.6231
Income	-0.0891
Sex	-0.1327
Age	0.1141
Size	0.0196
$\pi$	0.7433
$LL$	-4297.5
$LL_{\text{Poiss reg}}$	-6291.5
LR (df = 1)	3988.0

27

## Fit of ZIP Regression



28

## NBD Regression

The explanatory variables may not fully capture the differences among individuals

To capture the remaining (unobserved) component of differences among individuals, let  $\lambda_0$  vary across the population according to a gamma distribution with parameters  $r$  and  $\alpha$ :

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

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

29

## NBD Regression Results

Variable	Coefficient
$r$	0.1388
$\alpha$	8.1979
Income	0.0734
Sex	-0.0093
Age	0.9022
Size	-0.0243
$LL$	-2889.0

30

## S-Plus NBD Regression Results

Coefficients:

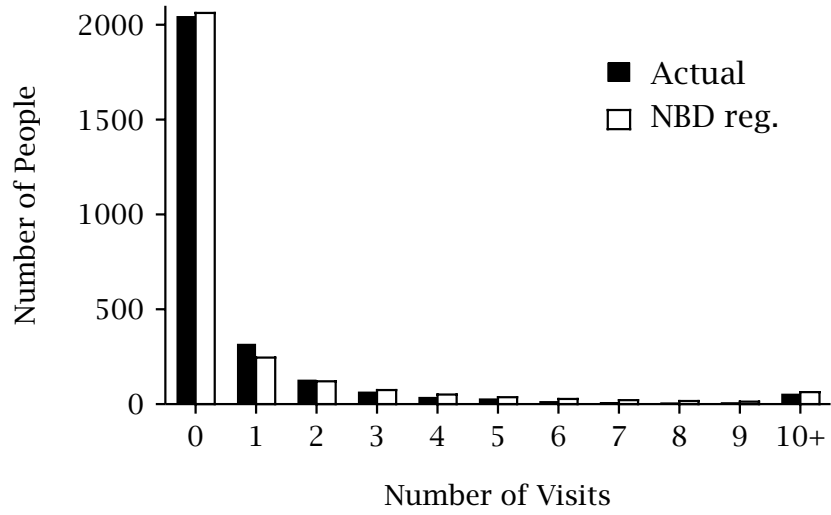
	Value	Std. Error	t value
(Intercept)	-4.047149702	1.10159557	-3.6738979
Income	0.074549233	0.09555222	0.7801936
Sex	-0.005240835	0.11592793	-0.0452077
Age	0.889862966	0.14072030	6.3236289
Size	-0.025094493	0.04187696	-0.5992435

Theta: 0.13878  
Std. Err.: 0.00726

## Limdep NBD Regression Results

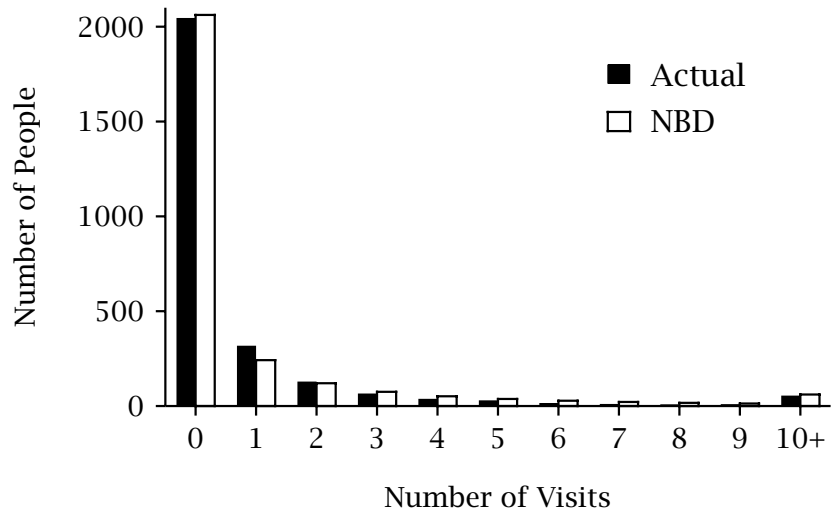
Variable	Coefficient	Standard Error	b/St.Er.
Constant	-4.077239653	1.0451741	-3.901
INCOME	.7237686001E-01	.76663437E-01	.944
SEX	-.9009160129E-02	.11425700	-.079
AGE	.9045111135	.17741724	5.098
SIZE	-.2406546843E-01	.38695426E-01	-.622
Overdispersion parameter			
Alpha	7.206708844	.33334006	21.620

## Fit of NBD Regression



33

## Fit of NBD



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

34

## Concepts and Tools Introduced

- Incorporating covariate effects in count models
- Poisson (and NBD) regression models
- The value of covariates is frequently over-emphasized

35

## 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 (1998), *Market Segmentation: Conceptual and Methodological Foundations*, Boston, MA: Kluwer Academic Publishers.

Winkelmann, Rainer (1997), *Econometric Analysis of Count Data*, 2nd, revised and enlarged edition, Berlin: Springer.

36

## 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 trait(s) a function of (time-invariant) covariates:

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

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

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

37

## Covariates in Timing Models

- If the covariates are time-invariant, we can make  $\lambda$  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

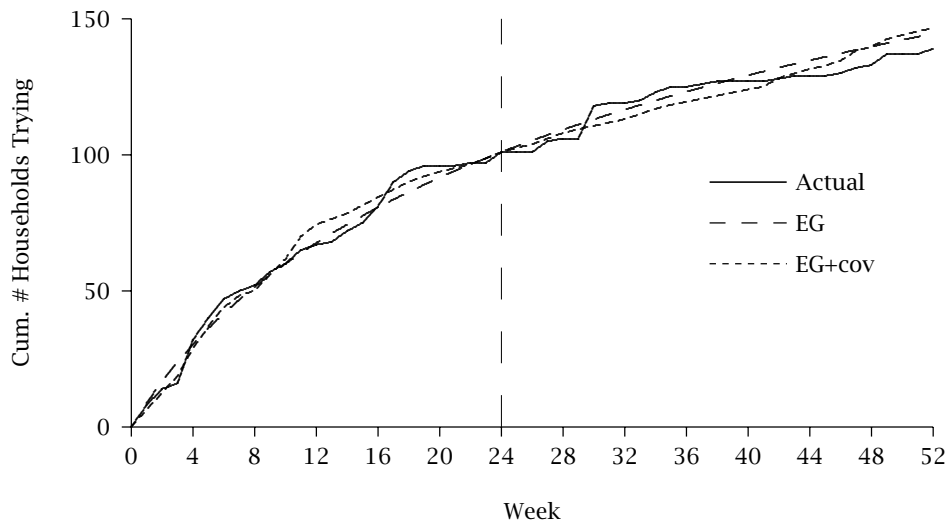
$$F(t) = 1 - e^{-\lambda_0 A(t)}$$

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

- Known as “proportional hazards regression”

38

## Comparing EG with EG+cov



39

## 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

40

**Problem 5:**  
**Understanding the Adoption of a**  
**Video-on-Demand Service**  
(Introducing Additional Model Structures)

41

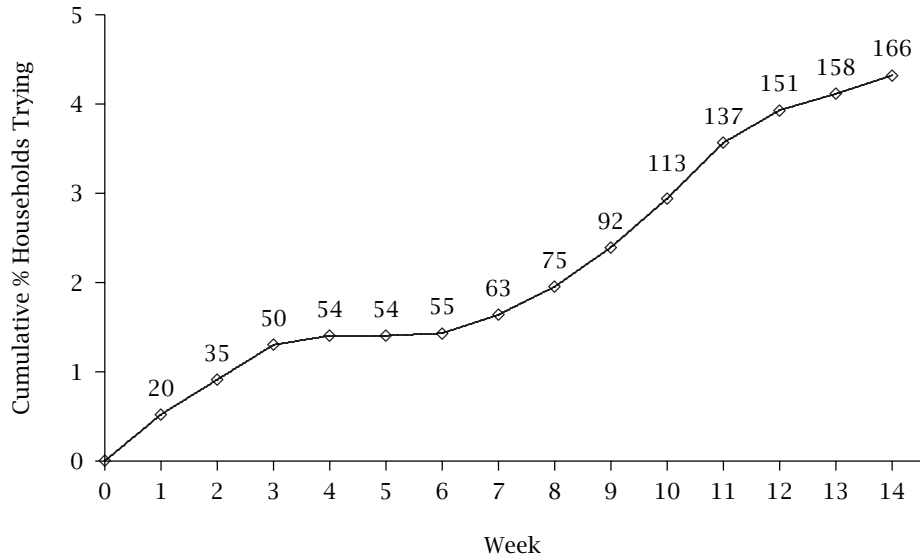
**Background**

A major telecommunications firm has just started a video-on-demand (VOD) service which covers 3841 households. After the first 14 weeks of operation, a total of 166 households have ordered videos through this service at least once.

The marketing manager wishes to understand the nature of the trial process, and would like an estimate of how many households will have tried the VOD service by the end of the first year of operation.

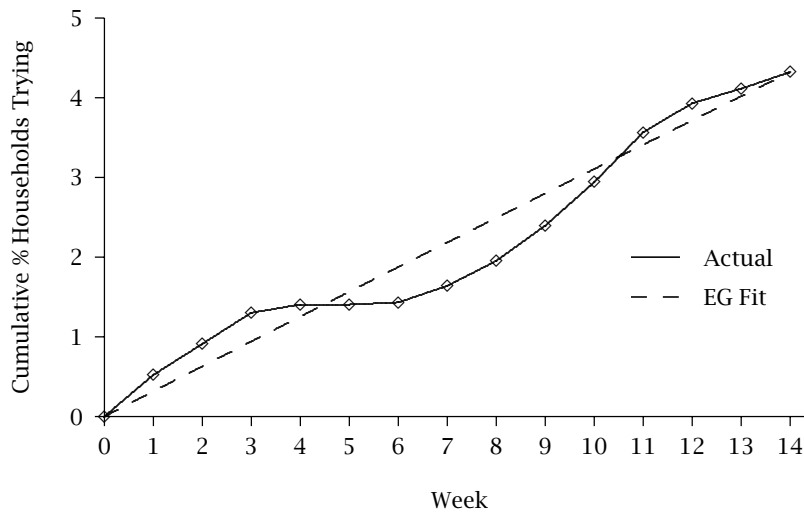
42

## Cumulative Trial of VOD Service



43

## Applying the EG Model



Predictions: 15% penetration at the end of one year  
almost 30% penetration after two years

44

## What is Wrong With the EG Model?

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

- i. at the individual-level, time-to-trial is not exponentially distributed
- ii. trial rates ( $\lambda$ ) are not gamma-distributed

45

## Relaxing the Gamma Assumption

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

$$F(t) = \sum_{s=1}^S p_s F(t|\lambda_s), \quad \sum_{s=1}^S p_s = 1$$

- Collapses to a single-segment exponential model with  $LL = -1122.02$  (no heterogeneity in  $\lambda$ )

46

## Reflecting on the Exponential Assumption

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

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

is also independent of  $t$

- How can we make  $P(t < T \leq t + \Delta t | T > t)$  depend on  $t$ ?

47

## The Hazard Rate Function

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

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

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

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

48

## The Hazard Rate Function

The hazard rate function uniquely defines the distribution of a nonnegative random variable:

$$F(t) = 1 - \exp\left(-\int_0^t h(u) du\right)$$

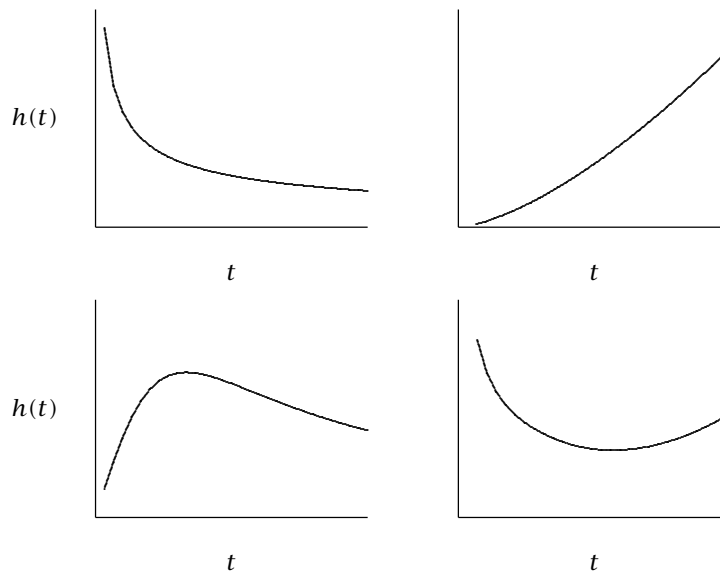
Example:

the exponential distribution has a *constant* hazard rate,  $\lambda$

$$\begin{aligned} F(t) &= 1 - \exp\left(-\int_0^t \lambda du\right) \\ &= 1 - e^{-\lambda t} \end{aligned}$$

49

## Shapes of the Hazard Rate Function



50

## The Weibull Distribution

- A generalization of the exponential distribution that can represent decreasing or increasing hazard rates

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

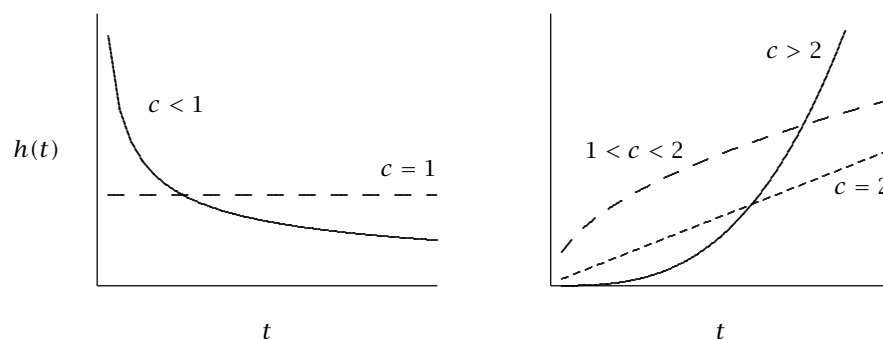
$$h(t) = c\lambda t^{c-1}$$

where  $c$  is the “shape” parameter and  $\lambda$  is the “scale” parameter

- Collapses to the exponential when  $c = 1$
- $F(t)$  is S-shaped for  $c > 1$

51

## The Weibull Hazard Rate Function



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

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

52

## The Weibull-Gamma Model

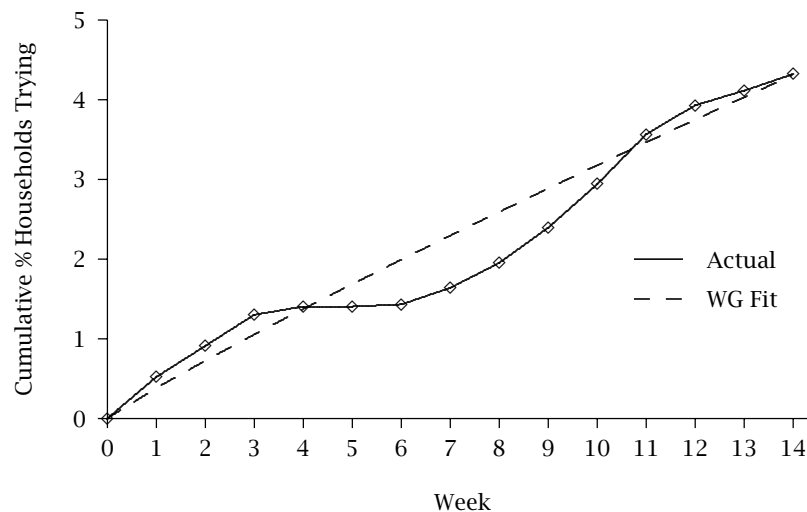
- Assuming  $\lambda$  is distributed across the population according to a gamma distribution, we have

$$\begin{aligned}
 P(T \leq t) &= \int_0^{\infty} (1 - e^{-\lambda t^c}) \frac{\alpha^r \lambda^{r-1} e^{-\alpha \lambda}}{\Gamma(r)} d\lambda \\
 &= 1 - \left( \frac{\alpha}{\alpha + t^c} \right)^r
 \end{aligned}$$

- This collapses to the exponential-gamma model when  $c = 1$
- Also known as the Burr Type XII distribution

53

## Applying the WG Model



Predictions mirror those of the EG model  
Is  $c$  different from 1?

54

## The Latent-Class Weibull Model

To allow for heterogeneity in both  $\lambda$  and  $c$ , we postulate the existence of discrete segments of households, each with their own values of  $\lambda$  and  $c$

- For two segments, we have

$$F(t) = \pi_1 \left(1 - e^{-\lambda_1 t^{c_1}}\right) + (1 - \pi_1) \left(1 - e^{-\lambda_2 t^{c_2}}\right)$$

- For three segments, we have

$$F(t) = \pi_1 \left(1 - e^{-\lambda_1 t^{c_1}}\right) + \pi_2 \left(1 - e^{-\lambda_2 t^{c_2}}\right) + (1 - \pi_1 - \pi_2) \left(1 - e^{-\lambda_3 t^{c_3}}\right)$$

55

## Parameter Estimates

- For the two segment model, the maximum value of the log-likelihood function is  $LL = -1100.9$

The associated parameter estimates are:

	Seg 1	Seg 2
$\lambda$	0.0063	1.51E-06
$c$	0.4846	5.6915
$\pi$	0.9789	

- Adding a third segment does not lead to any improvement in the value of the log-likelihood function

56

## 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:

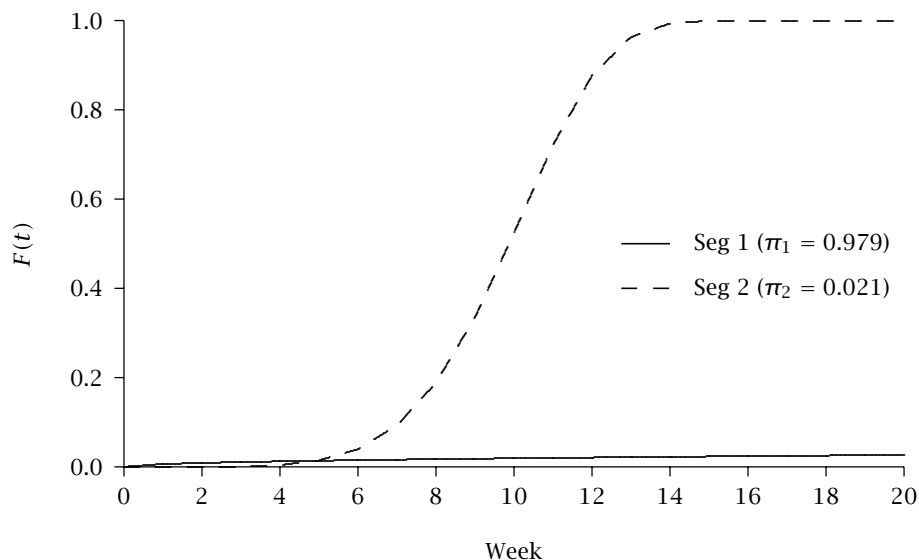
$$\text{BIC} = -2LL + p \ln(N)$$

where  $p$  is the number of parameters and  $N$  is the sample size

- Rule: choose  $S$  to minimize BIC

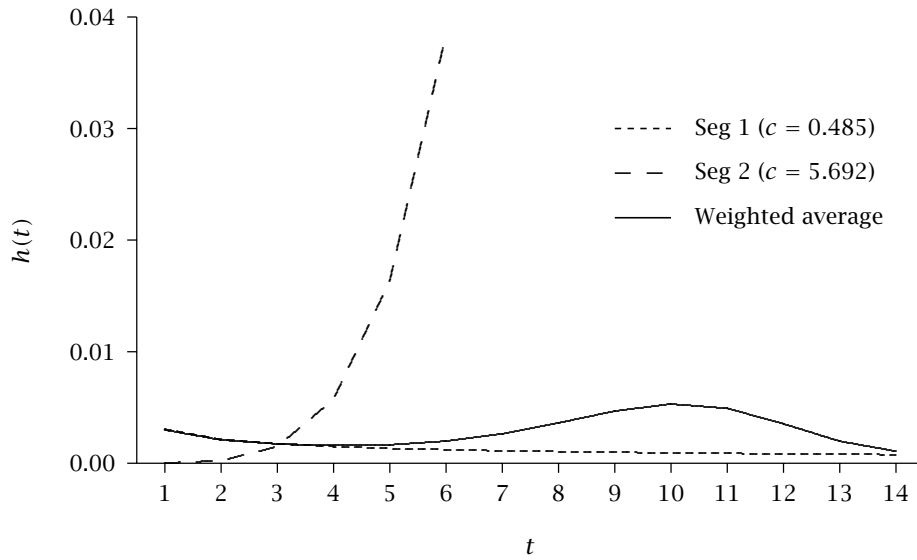
57

## Segment-level CDFs



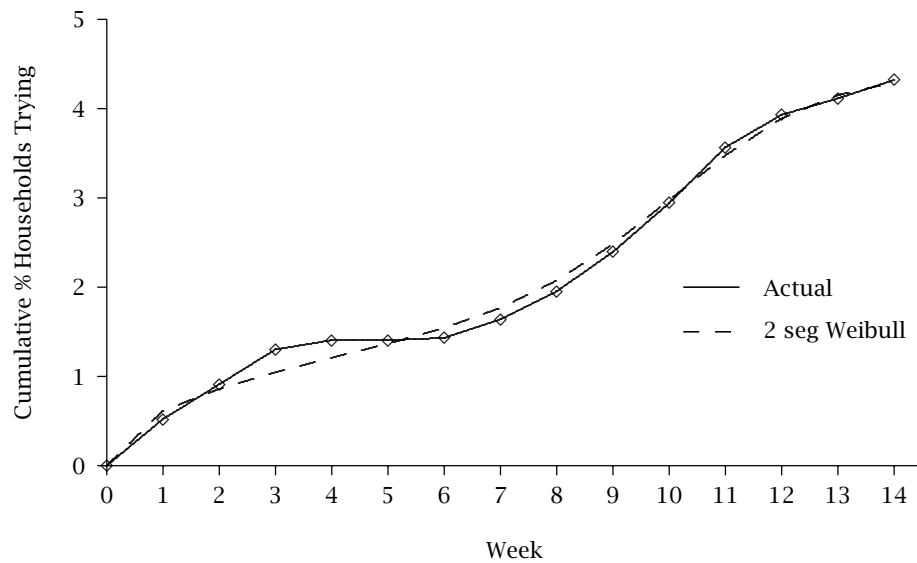
58

## Hazard Rate Function Plot



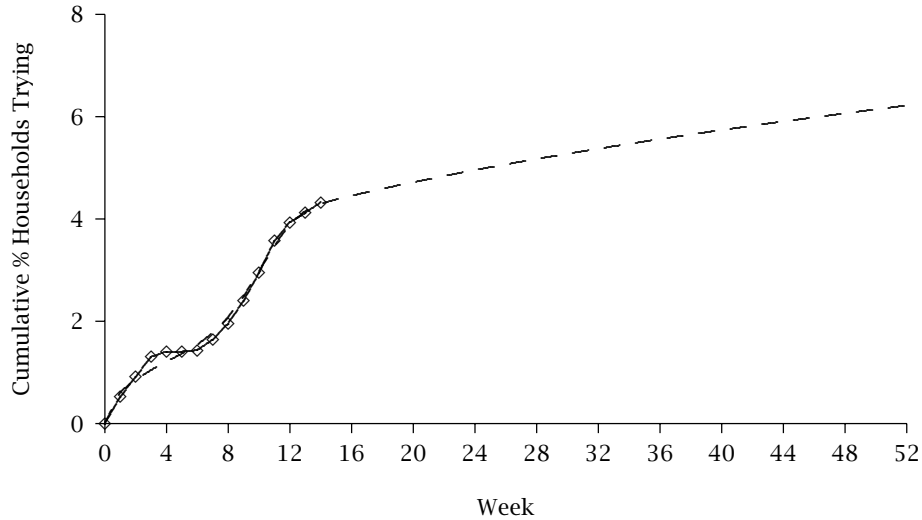
59

## Model Fit



60

## Cumulative Trial Forecast



61

## Variation on a Theme

We postulate the existence of discrete segments of households:

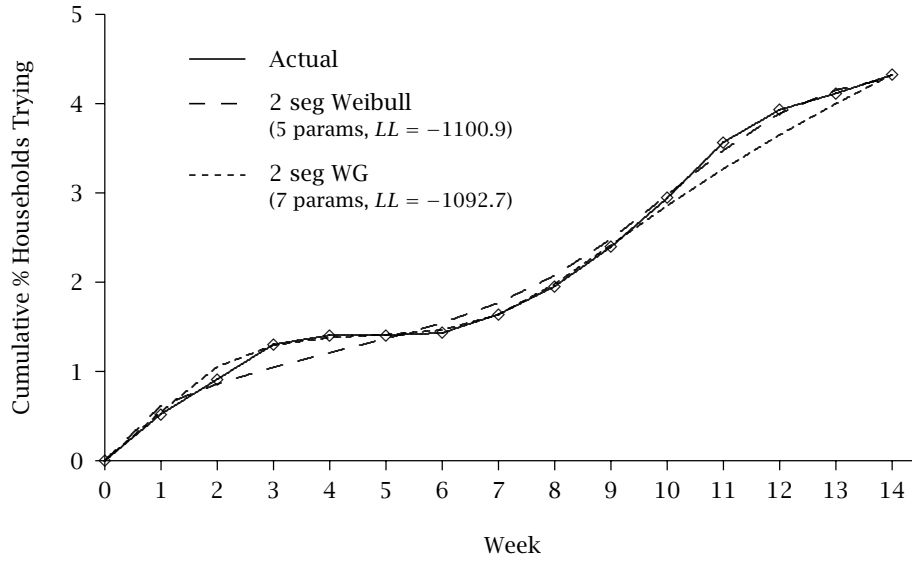
- within each segment, individual-level time-to-trial is Weibull-distributed with parameters  $\lambda_s$  and  $c_s$
- $\lambda_s$  is distributed across the segment members according to a gamma distribution with parameters  $r_s$  and  $\alpha_s$

→ a finite mixture of Weibull-gamma models:

$$F(t) = \pi_1 \left[ 1 - \left( \frac{\alpha_1}{\alpha_1 + t^{c_1}} \right)^{r_1} \right] + (1 - \pi_1) \left[ 1 - \left( \frac{\alpha_2}{\alpha_2 + t^{c_2}} \right)^{r_2} \right]$$

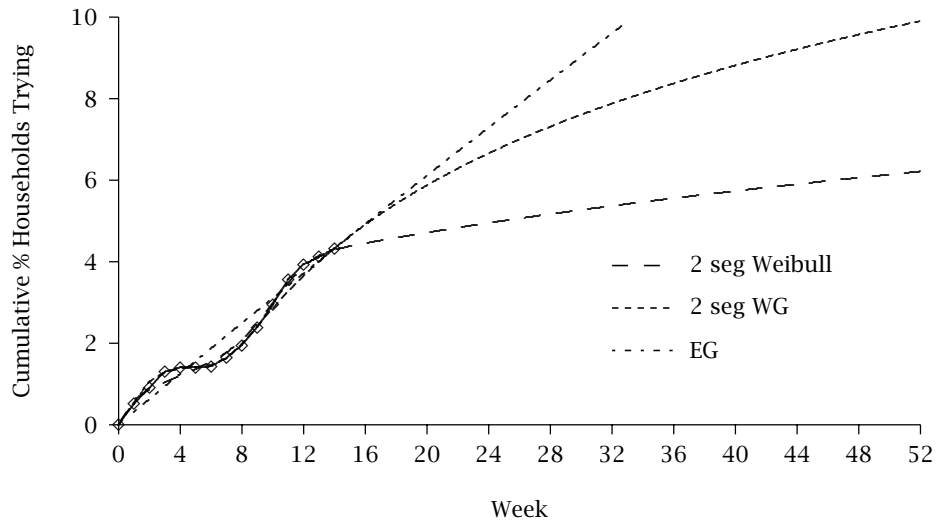
62

# Model Fit



63

# Cumulative Trial Forecast



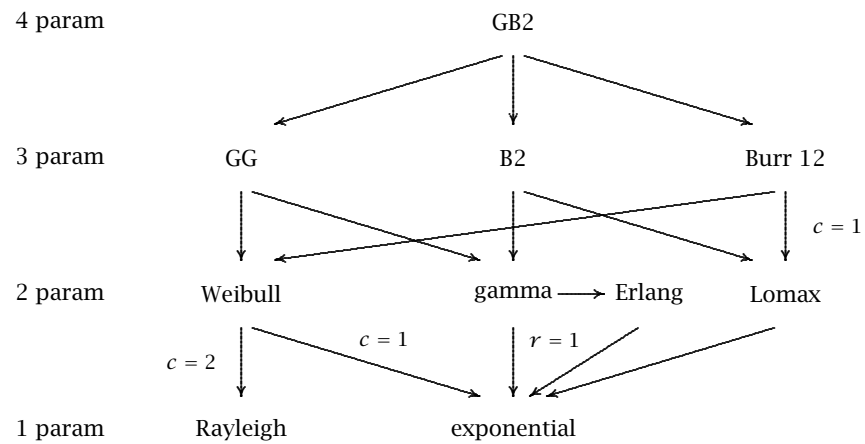
64

## Alternative Explanations

- Covariate effects
- “Time shifting”
- More complex behavioral stories

65

## A Family Tree of Distributions



66

## Concepts and Tools Introduced

- Hazard rate functions
- Alternative individual-level timing models (e.g., the Weibull)
- Finite mixture models
- The art/science of model building

67

## Further Reading

Bury, Karl (1999), *Statistical Distributions in Engineering*, Cambridge, UK: Cambridge University Press.

Evans, Merran, Nicholas Hastings, and Brian Peacock (2000), *Statistical Distributions*, 3rd edition, New York: John Wiley & Sons.

McDonald, James B. and Dale O. Richards (1987), "Hazard Rates and Generalized Beta Distributions," *IEEE Transactions on Reliability*, **R-36** (October), 463–466.

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

Meeker, William Q. and Luis A. Escobar (1998), *Statistical Methods for Reliability Data*, New York: John Wiley & Sons.

68

## **Problem 6: Modeling Repeat Purchase Quantities at CDNOW**

(Building an “Integrated” Model)

69

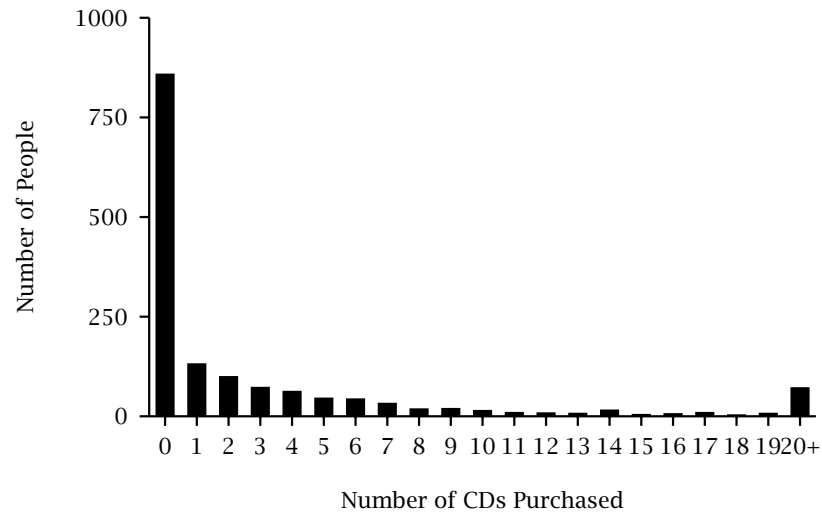
### **Problem Setting**

During the first week of 1997, 1574 individuals made their first-ever purchase of CDs at the CDNOW web site. Over the subsequent 51 weeks, these individuals purchased a total of 6357 CDs.

Our objective is to develop a simple model that describes the distribution of repeat purchase quantity (# CDs) for this group of 1574 individuals.

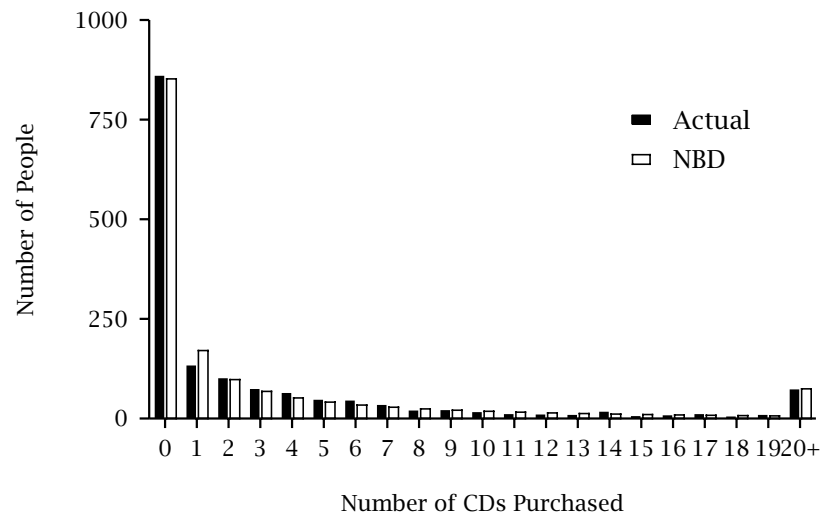
70

## Distribution of Purchase Quantity



71

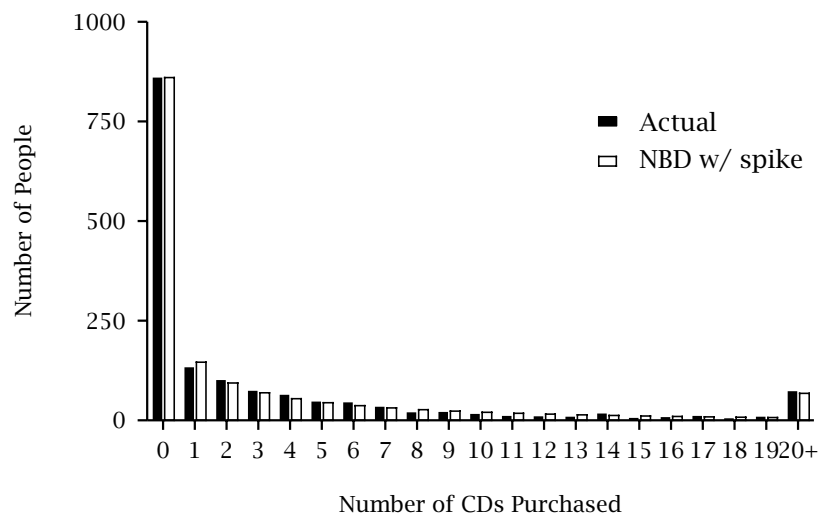
## Fit of NBD



$$\hat{r} = 0.212, \hat{\alpha} = 0.0584$$

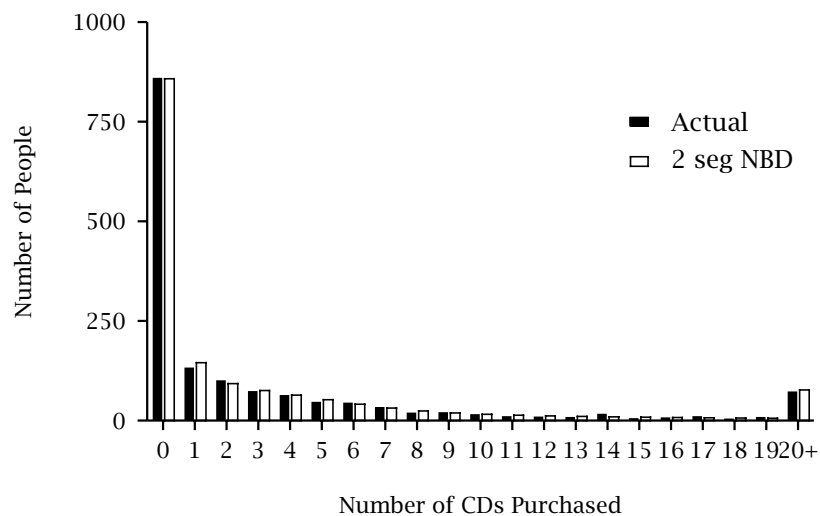
72

## Fit of NBD with Spike at 0



73

## Fit of Two-Segment Latent-Class NBD



74

## Reflection

The observed variables (# CDs purchased in the 51 week period) is the outcome of two separate processes:

- the number of purchase occasions (in the 51 week period), and
- the number of CDs purchased on each purchase occasion ( $\geq 1$ ).

75

## How Can I Buy Three CDs?

- One purchase occasion on which I buy three CDs
- Two purchase occasions, with me buying one CD on one of the purchase occasions and two CDs on the other.
- Three purchase occasions, with me buying one CD on each purchase occasion.

76

## Notation

- Let  $X_j$  = # CDs purchased on the  $j$ th purchase occasion
- $Y_n$  = # CDs purchases over  $n$  purchase occasions  
 $\Rightarrow Y_n = X_1 + X_2 + \cdots + X_n$
- $Y$  = # CDs purchases in the unit time interval
- $N$  = # purchase occasions in the unit time interval

77

## At the Level of the Individual

Our goal is identify the distribution of  $Y$

1. We specify the distribution of  $X_j$
2. This implies the distribution of  $Y_n$
3. We specify the distribution of  $N$

The distribution of  $Y$  follows naturally:

$$P(Y = y) = \sum_{n=1}^y P(Y_n = y)P(N = n)$$

78

## Deriving the Distribution of $Y$

- (i) We assume  $X_j$  is distributed according to the shifted-geometric distribution with “quantity” parameter  $p$ :

$$P(X_j = x|p) = p(1 - p)^{x-1}, \quad x = 1, 2, 3, \dots$$

- (ii) If  $X_j$  are iid shifted-geometric with parameter  $p$ ,  $Y_n = X_1 + X_2 + \dots + X_n$  is distributed according to the *Pascal* distribution:

$$P(Y_n = y|p) = \binom{y-1}{n-1} p^n (1-p)^{y-n}, \quad y \geq n$$

79

## Deriving the Distribution of $Y$

- (iii) We assume  $N$  is distributed according to the Poisson distribution with rate parameter  $\lambda$

The unconditional distribution of  $Y$  is given by

$$P(Y = y) = \begin{cases} e^{-\lambda} & x = 0 \\ \sum_{n=1}^y \binom{y-1}{n-1} p^n (1-p)^{y-n} \frac{\lambda^n e^{-\lambda}}{n!} & x = 1, 2, \dots \end{cases}$$

This is known as the Pólya-Aeppli distribution

80

## Introducing Heterogeneity

- The rate parameter  $\lambda$  is distributed across the population according to a gamma distribution:

$$g_1(\lambda) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

- The “quantity” parameter  $p$  is distributed across the population according to a beta distribution:

$$g_2(p) = \frac{1}{B(a, b)} p^{a-1} (1-p)^{b-1}$$

- $\lambda$  and  $p$  are independent

81

## The Aggregate Distribution of $Y$

For  $Y = 0$ :

$$P(Y = 0) = P_{\text{NBD}}(N = 0) = \left(\frac{\alpha}{\alpha + 1}\right)^r$$

For  $Y > 0$ :

$$\begin{aligned} P(Y = y) &= \int_0^1 \int_0^\infty \sum_{n=1}^y P(Y_n = y|p) P(N = n|\lambda) g_1(\lambda) g_2(p) d\lambda dp \\ &= \sum_{n=1}^y \left\{ \int_0^1 P(Y_n = y|p) g_2(p) dp \right\} \left\{ \int_0^\infty P(N = n|\lambda) g_1(\lambda) d\lambda \right\} \\ &= \sum_{n=1}^y \underbrace{\binom{y-1}{n-1} \frac{B(a+n, b+y-n)}{B(a, b)}}_{\text{beta-Pascal}} \underbrace{\frac{\Gamma(r+n)}{\Gamma(r)n!} \left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^n}_{\text{NBD}} \end{aligned}$$

We call this the BP/NBD distribution

82

## Parameter Estimation

*Case 1:* We observe both the number of purchases and the number of purchase occasions ( $y_i$  and  $n_i$ ) for each individual ( $i = 1, \dots, I$ )

- We estimate  $r$  and  $\alpha$  by fitting the NBD to the data on the number of the purchase occasions:

$$LL(r, \alpha) = \sum_{i=1}^I \ln[P_{\text{NBD}}(N = n_i)]$$

- We estimate  $a$  and  $b$  by fitting the beta-Pascal distribution to the data on the number of purchases, given the number of the purchase occasions:

$$LL(a, b) = \sum_{i:n_i>0} \ln[P_{\text{BP}}(Y = y_i | n_i)]$$

83

## Parameter Estimation

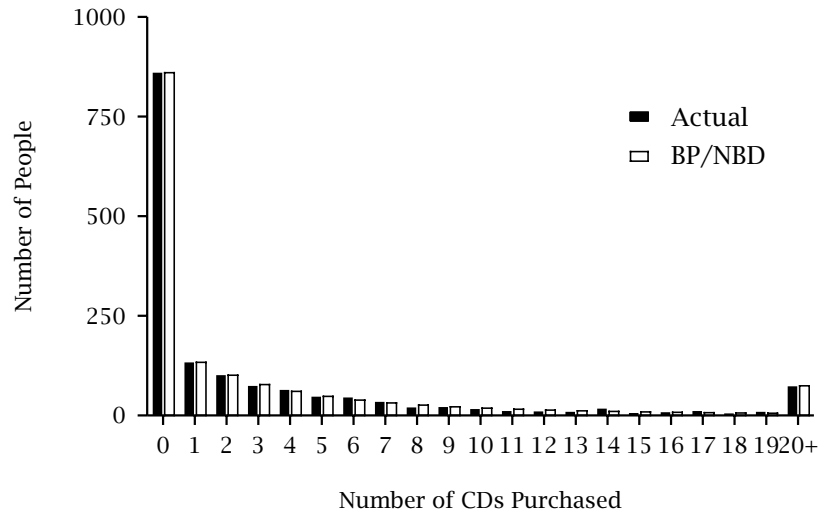
*Case 2:* The number of purchase occasions ( $n_i$ ) is not observed; we only observe the number of purchases ( $y_i$ ) for each individual ( $i = 1, \dots, I$ )

- We estimate the four model parameters ( $r, \alpha, a, b$ ) by fitting the BP/NBD to the data on the number of the purchases:

$$LL(r, \alpha, a, b) = \sum_{i=1}^I \ln[P_{\text{BP/NBD}}(Y = y_i)]$$

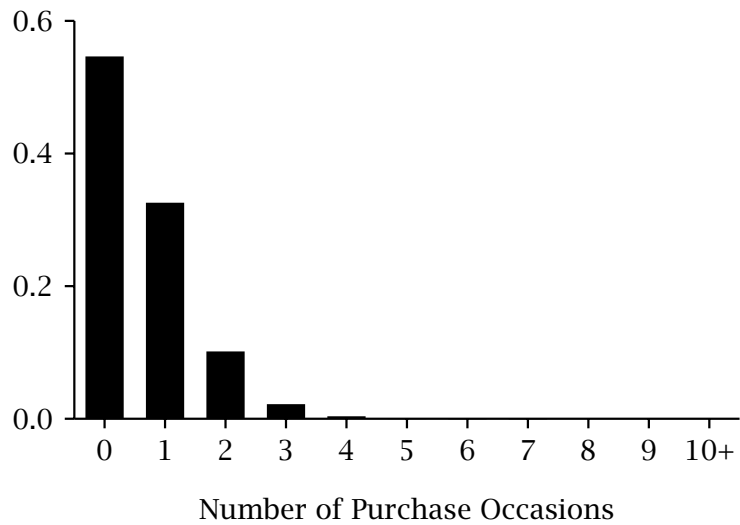
84

## Fit of the BP/NBD ("Case 2")



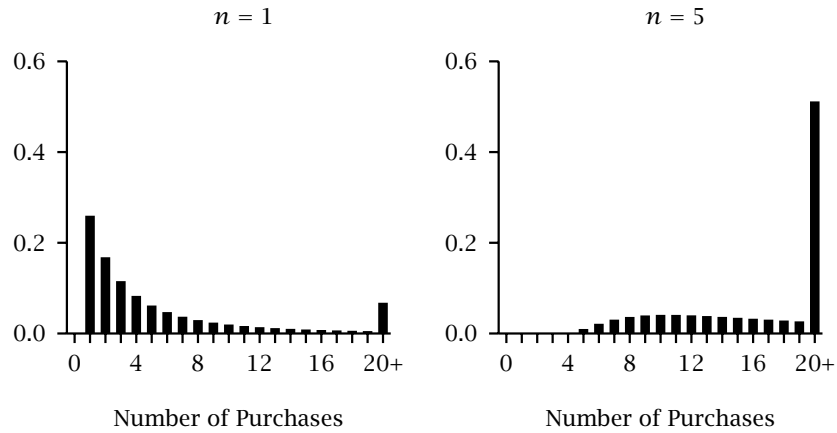
85

## Implied NBD for Purchase Occasions



86

## Implied Beta-Pascal Distributions



87

## Making Predictions

*Case 1:* We observe both the number of purchases and the number of purchase occasions ( $y$  and  $n$ ) for each individual

$$\begin{aligned}
 E(Y_2 | Y_1 = y, N_1 = n) &= \underbrace{E(N_2 | N_1 = n)}_{\text{NBD cond. exp.}} \underbrace{E(X_j | Y_1 = y, N_1 = n)}_{\text{shifted beta-geometric cond. exp.}} \\
 &= \left( \frac{r + n}{\alpha + 1} \right) \left( \frac{\alpha + \beta + y - 1}{\alpha + n - 1} \right)
 \end{aligned}$$

*Case 2:* We only observe the number of purchases ( $y$ ) for each individual

$$E(Y_2 | Y_1 = y) = \text{a horrible mess}$$

88

## A Template for Integrated Models

		Stage 2		
		Counting	Timing	Choice
Stage 1	Counting	CDNOW		
	Timing			
	Choice			

89

### Recap

- Modeling timing data
  - Modeling count data
  - Modeling “choice” data
- 
- Incorporating covariates in count models
  - Introducing additional model structures
  - Building an “integrated” model

90

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

<http://brucehardie.com/talks.html>