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

13th Annual Advanced Research Techniques Forum
June 2–5, 2002

Probability Models 101

The Logic of Probability Models

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

Uses of Probability Models

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

Building a Probability Model

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

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

- We denote this by x.

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

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

Building a Probability Model

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

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

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

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

Building a Probability Model

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

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

"Classes" of Models

- 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.

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.

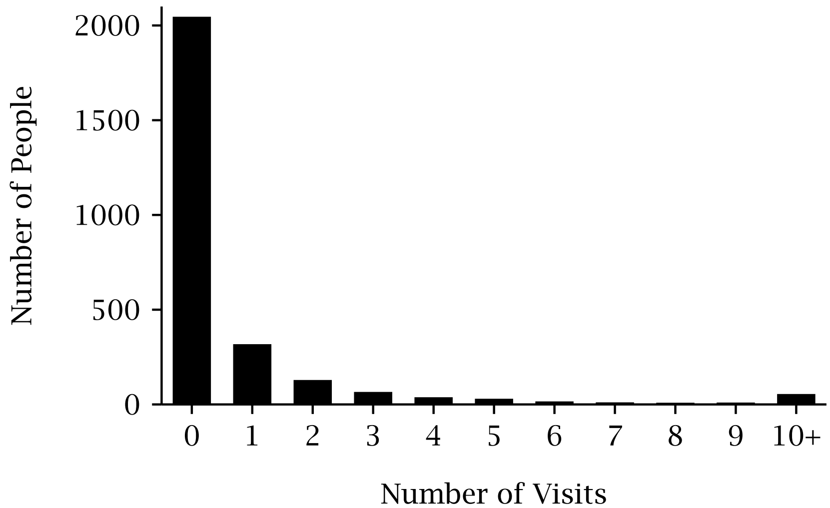
In order to gain an understanding of the aggregate visiting patterns, some Media Metrix panel data has been purchased. For a sample of 2728 people who visited an online apparel site at least once during the second-half of 2000, the dataset reports how many visits each person made to the khakichinos.com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

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

Raw Data

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

Distribution of Visits



13

Modeling Count Data

Recall the NBD:

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

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

Observed vs. Unobserved Heterogeneity

Unobserved Heterogeneity:

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

Observed Heterogeneity:

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

The Poisson Regression Model

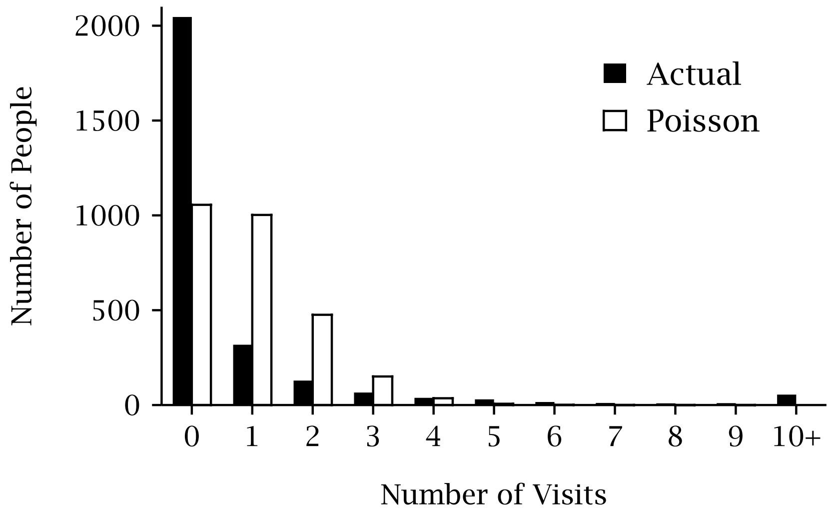
- Let the random variable Y_i denote the number of times individual i visits the site in a unit time period
- At the individual-level, Y_i is assumed to be distributed Poisson with mean λ_i:

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

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

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

Fit of Poisson



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

Poisson Regression Results

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

	A	B	C	D	E	F	G	H	I	J	K
1	\lambda_0	0.04387			LL =	=SUM(K9:K2736)					
2	B_inc	0.09385									
3	B_sex	0.00426									
4	B_age	0.58825									
5	B_size	-0.0359									
6				=TRANSPOSE(B2:B5)	=TRANSPOSE(B2:B5)	=TRANSPOSE(B2:B5)	=TRANSPOSE(B2:B5)				
7											
8	ID	Total		Income	Sex	Age	Size		lambda	P(Y=y)	ln[P(Y=y)]
9	1	0		11.3793940723457	1	3.87120101090789	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D9:G9))	=POISSON(B9,I9,FALSE)	=LN(J9)
10	2	5		9.76995615991161	1	4.04305126783455	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D10:G10))	=POISSON(B10,I10,FALSE)	=LN(J10)
11	3	0		11.0821425488778	0	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D11:G11))	=POISSON(B11,I11,FALSE)	=LN(J11)
12	4	0		10.9150884642146	1	3.95124371858143	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D12:G12))	=POISSON(B12,I12,FALSE)	=LN(J12)
13	5	0		10.9150884642146	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D13:G13))	=POISSON(B13,I13,FALSE)	=LN(J13)
14	6	0		10.9150884642146	0	2.94443897916644	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D14:G14))	=POISSON(B14,I14,FALSE)	=LN(J14)
15	7	0		11.1913418408428	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D15:G15))	=POISSON(B15,I15,FALSE)	=LN(J15)
16	8	1		11.7360690162844	0	4.07753744390572	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D16:G16))	=POISSON(B16,I16,FALSE)	=LN(J16)
17	9	0		10.0212705881925	0	4.24849524204936	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D17:G17))	=POISSON(B17,I17,FALSE)	=LN(J17)
18	10	0		10.9150884642146	0	3.85014760171006	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D18:G18))	=POISSON(B18,I18,FALSE)	=LN(J18)
19	11	1		10.7684849900227	0	3.93182563272433	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D19:G19))	=POISSON(B19,I19,FALSE)	=LN(J19)
20	12	0		10.9150884642146	0	3.98898404656427	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D20:G20))	=POISSON(B20,I20,FALSE)	=LN(J20)
21	13	3		10.5320962119585	0	3.63758615972639	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D21:G21))	=POISSON(B21,I21,FALSE)	=LN(J21)
22	14	0		10.9150884642146	0	3.61091791264422	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D22:G22))	=POISSON(B22,I22,FALSE)	=LN(J22)
23	15	0		10.2219412836547	1	3.58351893845611	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D23:G23))	=POISSON(B23,I23,FALSE)	=LN(J23)
24	16	1		10.7684849900227	1	3.25809653802148	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D24:G24))	=POISSON(B24,I24,FALSE)	=LN(J24)
25	17	2		12.2060726455302	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D25:G25))	=POISSON(B25,I25,FALSE)	=LN(J25)
26	18	0		10.7684849900227	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D26:G26))	=POISSON(B26,I26,FALSE)	=LN(J26)
27	19	6		11.1913418408428	1	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D27:G27))	=POISSON(B27,I27,FALSE)	=LN(J27)
28	20	0		10.3889953683178	1	3.58351893845611	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D28:G28))	=POISSON(B28,I28,FALSE)	=LN(J28)
29	21	2		10.7684849900227	1	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D29:G29))	=POISSON(B29,I29,FALSE)	=LN(J29)
30	22	0		11.1913418408428	1	3.46573590279973	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D30:G30))	=POISSON(B30,I30,FALSE)	=LN(J30)
31	23	0		11.1913418408428	1	3.43398720448515	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D31:G31))	=POISSON(B31,I31,FALSE)	=LN(J31)
32	24	2		11.7360690162844	1	3.80666248977032	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D32:G32))	=POISSON(B32,I32,FALSE)	=LN(J32)
33	25	0		11.3793940723457	0	4.27666611901606	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D33:G33))	=POISSON(B33,I33,FALSE)	=LN(J33)
34	26	0		10.3889953683178	0	4.21950770517611	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D34:G34))	=POISSON(B34,I34,FALSE)	=LN(J34)
35	27	0		10.6572593549125	1	3.49650756146648	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D35:G35))	=POISSON(B35,I35,FALSE)	=LN(J35)
36	28	0		12.0725412529057	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D36:G36))	=POISSON(B36,I36,FALSE)	=LN(J36)
37	29	0		10.9150884642146	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D37:G37))	=POISSON(B37,I37,FALSE)	=LN(J37)
38	30	0		10.9150884642146	0	3.52636052461616	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D38:G38))	=POISSON(B38,I38,FALSE)	=LN(J38)
39	31	0		11.1913418408428	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D39:G39))	=POISSON(B39,I39,FALSE)	=LN(J39)
40	32	0		10.2219412836547	1	3.13549421592915	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D40:G40))	=POISSON(B40,I40,FALSE)	=LN(J40)
41	33	0		11.3793940723457	0	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D41:G41))	=POISSON(B41,I41,FALSE)	=LN(J41)
42	34	0		9.07680897935166	1	3.40119738166216	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D42:G42))	=POISSON(B42,I42,FALSE)	=LN(J42)
43	35	0		10.0212705881925	1	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D43:G43))	=POISSON(B43,I43,FALSE)	=LN(J43)
44	36	0		11.0821425488778	0	4.06044301054642	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D44:G44))	=POISSON(B44,I44,FALSE)	=LN(J44)
45	37	2		10.2219412836547	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D45:G45))	=POISSON(B45,I45,FALSE)	=LN(J45)
46	38	2		12.0725412529057	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D46:G46))	=POISSON(B46,I46,FALSE)	=LN(J46)
47	39	1		11.0821425488778	0	4.17438726989564	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D47:G47))	=POISSON(B47,I47,FALSE)	=LN(J47)
48	40	0		9.52879410309472	1	2.70805020110221	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D48:G48))	=POISSON(B48,I48,FALSE)	=LN(J48)
49	41	0		11.0821425488778	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D49:G49))	=POISSON(B49,I49,FALSE)	=LN(J49)
50	42	0		11.3793940723457	1	4.12713438504509	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D50:G50))	=POISSON(B50,I50,FALSE)	=LN(J50)
51	43	0		11.3793940723457	0	4.17438726989564	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D51:G51))	=POISSON(B51,I51,FALSE)	=LN(J51)
52	44	0		10.5320962119585	1	3.55534806148941	6		=B$1*EXP(SUMPRODUCT(D$6:G$6,D52:G52))	=POISSON(B52,I52,FALSE)	=LN(J52)
53	45	0		10.7684849900227	0	3.2188758248682	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D53:G53))	=POISSON(B53,I53,FALSE)	=LN(J53)
54	46	0		11.3793940723457	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D54:G54))	=POISSON(B54,I54,FALSE)	=LN(J54)
55	47	0		11.7360690162844	0	3.04452243772342	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D55:G55))	=POISSON(B55,I55,FALSE)	=LN(J55)
56	48	0		10.7684849900227	0	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D56:G56))	=POISSON(B56,I56,FALSE)	=LN(J56)
57	49	0		10.3889953683178	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D57:G57))	=POISSON(B57,I57,FALSE)	=LN(J57)
58	50	0		10.3889953683178	1	2.63905732961526	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D58:G58))	=POISSON(B58,I58,FALSE)	=LN(J58)
59	51	0		11.0821425488778	0	3.73766961828337	5		=B$1*EXP(SUMPRODUCT(D$6:G$6,D59:G59))	=POISSON(B59,I59,FALSE)	=LN(J59)
60	52	6		9.76995615991161	0	3.29583686600433	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D60:G60))	=POISSON(B60,I60,FALSE)	=LN(J60)
61	53	16		9.76995615991161	0	3.13549421592915	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D61:G61))	=POISSON(B61,I61,FALSE)	=LN(J61)
62	54	1		10.2219412836547	1	3.46573590279973	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D62:G62))	=POISSON(B62,I62,FALSE)	=LN(J62)
63	55	0		11.3793940723457	1	3.55534806148941	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D63:G63))	=POISSON(B63,I63,FALSE)	=LN(J63)

Problem 4 -- Poisson reg

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	0.043874			LL =	-6291.497				
2	B_inc	0.093846								
3	B_sex	0.004262								
4	B_age	0.588246								
5	B_size	-0.035908								
6				0.093846	0.004262	0.588246	-0.035908			
7										
8	ID	Total		Income	Sex	Age	Size		lambda	P(Y=y)
9	1	0		11.38	1	3.87	2		1.163169	0.312494
10	2	5		9.77	1	4.04	1		1.146951	0.005253
11	3	0		11.08	0	3.33	2		0.820311	0.440295
12	4	0		10.92	1	3.95	3		1.126091	0.324298
13	5	0		10.92	1	2.83	3		0.583375	0.558012
14	6	0		10.92	0	2.94	3		0.620172	0.537852
15	7	0		11.19	0	3.66	2		1.007121	0.365269
16	8	1		11.74	0	4.08	2		1.352204	0.349774
17	9	0		10.02	0	4.25	1		1.319538	0.267259
18	10	0		10.92	0	3.85	3		1.056563	0.347649
19	11	1		10.77	0	3.93	2		1.133395	0.364883
20	12	0		10.92	0	3.99	2		1.188389	0.304712
21	13	3		10.53	0	3.64	2		0.932346	0.05317
22	14	0		10.92	0	3.61	1		0.986207	0.372989
23	15	0		10.22	1	3.58	3		0.849923	0.427448
24	16	1		10.77	1	3.26	3		0.738786	0.352913
25	17	2		12.21	0	3.66	2		1.107742	0.202657
26	18	0		10.77	0	3.95	2		1.146416	0.317774
27	19	6		11.19	1	3.33	2		0.8323	0.000201
28	20	0		10.39	1	3.58	2		0.894917	0.408641
29	21	2		10.77	1	3.33	4		0.744487	0.131631
30	22	0		11.19	1	3.47	2		0.900313	0.406442
31	23	0		11.19	1	3.43	2		0.883655	0.41327
32	24	2		11.74	1	3.81	2		1.157956	0.210601
33	25	0		11.38	0	4.28	2		1.470202	0.229879
34	26	0		10.39	0	4.22	2		1.295416	0.273784
35	27	0		10.66	1	3.50	4		0.811519	0.444183
36	28	0		12.07	0	3.95	2		1.29566	0.273717
37	29	0		10.92	1	3.81	3		1.034278	0.355483
38	30	0		10.92	0	3.53	3		0.873327	0.41756
39	31	0		11.19	1	3.37	2		0.849659	0.427561
40	32	0		10.22	1	3.14	4		0.629981	0.532602
41	33	0		11.38	0	3.33	4		0.785063	0.456091
42	34	0		9.08	1	3.40	1		0.736748	0.478668
43	35	0		10.02	1	3.53	1		0.866538	0.420404
44	36	0		11.08	0	4.06	4		1.171748	0.309825
45	37	2		10.22	1	3.69	2		0.937326	0.172058
46	38	2		12.07	1	3.69	2		1.115104	0.203853
47	39	1		11.08	0	4.17	1		1.395491	0.345679
48	40	0		9.53	1	2.71	3		0.475852	0.621356
49	41	0		11.08	1	3.81	3		1.05062	0.349721
50	42	0		11.38	1	4.13	3		1.304465	0.271318
51	43	0		11.38	0	4.17	3		1.335529	0.263019

Comparing Expected Visit Behavior

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

Who is less likely to have visited the web site?

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

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

Consider two models, A and B:

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

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

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

The Likelihood Ratio Test

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

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

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

Computing Standard Errors

- Excel
  - indirectly via a series of likelihood ratio tests
- 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

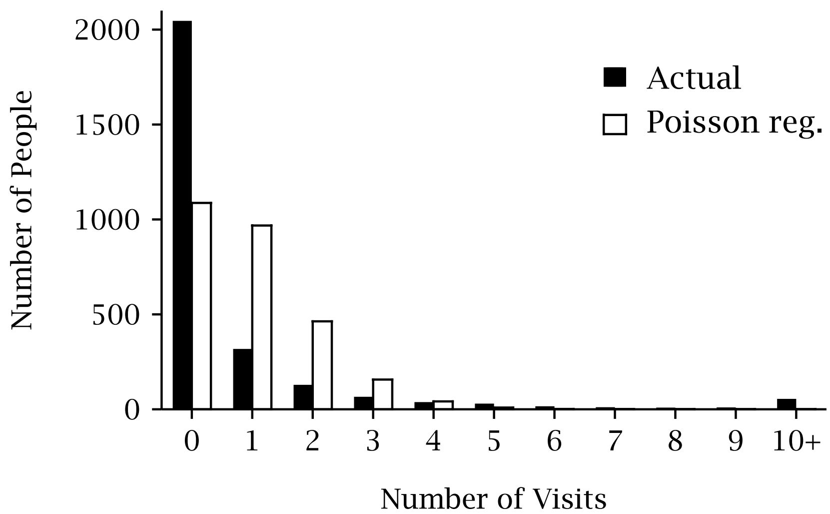
S-Plus Poisson Regression Results

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


Limdep Poisson Regression Results

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

Fit of Poisson Regression



The ZIP Regression Model

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

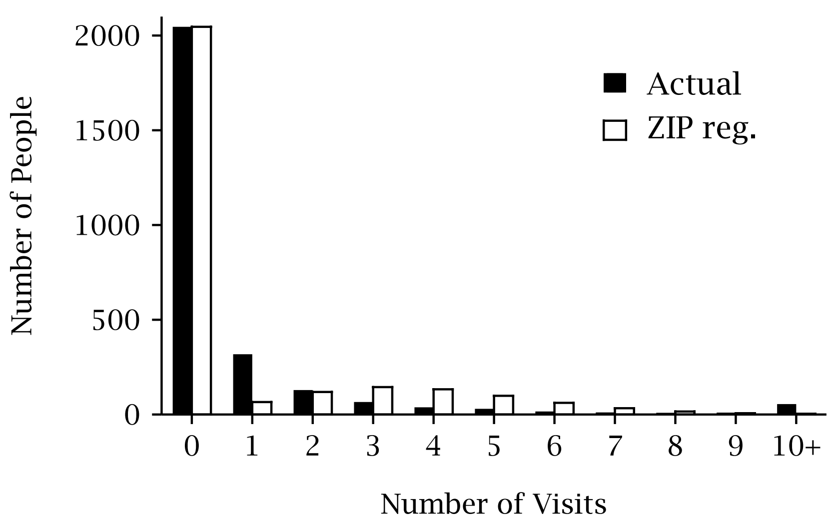
- a proportion π of those people who go to online apparel sites will never visit khakichinos.com
- the visiting behavior of the "ever visitors" can be characterized by the Poisson regression model

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

ZIP Regression Results

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

Fit of ZIP Regression



	A	B	C	D	E	F	G	H	I	J	K
1	\lambda_0	6.6231			LL =	=SUM(K10:K2737)					
2	pi	0.7433									
3	B_inc	-0.0891									
4	B_sex	-0.1327									
5	B_age	0.1141									
6	B_size	0.0196									
7				=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)				
8											
9	ID	Total		Income	Sex	Age	Size		lambda	P(Y=y)	ln[P(Y=y)]
10	1	0		11.3793940723457	1	3.87120101090789	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D10:G10))	=IF(B10=0,B$2,0)+(1-B$2)*POISSON(B10,I10,FALSE)	=LN(J10)
11	2	5		9.76995615991161	1	4.04305126783455	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D11:G11))	=IF(B11=0,B$2,0)+(1-B$2)*POISSON(B11,I11,FALSE)	=LN(J11)
12	3	0		11.0821425488778	0	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D12:G12))	=IF(B12=0,B$2,0)+(1-B$2)*POISSON(B12,I12,FALSE)	=LN(J12)
13	4	0		10.9150884642146	1	3.95124371858143	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D13:G13))	=IF(B13=0,B$2,0)+(1-B$2)*POISSON(B13,I13,FALSE)	=LN(J13)
14	5	0		10.9150884642146	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D14:G14))	=IF(B14=0,B$2,0)+(1-B$2)*POISSON(B14,I14,FALSE)	=LN(J14)
15	6	0		10.9150884642146	0	2.94443897916644	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D15:G15))	=IF(B15=0,B$2,0)+(1-B$2)*POISSON(B15,I15,FALSE)	=LN(J15)
16	7	0		11.1913418408428	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D16:G16))	=IF(B16=0,B$2,0)+(1-B$2)*POISSON(B16,I16,FALSE)	=LN(J16)
17	8	1		11.7360690162844	0	4.07753744390572	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D17:G17))	=IF(B17=0,B$2,0)+(1-B$2)*POISSON(B17,I17,FALSE)	=LN(J17)
18	9	0		10.0212705881925	0	4.24849524204936	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D18:G18))	=IF(B18=0,B$2,0)+(1-B$2)*POISSON(B18,I18,FALSE)	=LN(J18)
19	10	0		10.9150884642146	0	3.85014760171006	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D19:G19))	=IF(B19=0,B$2,0)+(1-B$2)*POISSON(B19,I19,FALSE)	=LN(J19)
20	11	1		10.7684849900227	0	3.93182563272433	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D20:G20))	=IF(B20=0,B$2,0)+(1-B$2)*POISSON(B20,I20,FALSE)	=LN(J20)
21	12	0		10.9150884642146	0	3.98898404656427	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D21:G21))	=IF(B21=0,B$2,0)+(1-B$2)*POISSON(B21,I21,FALSE)	=LN(J21)
22	13	3		10.5320962119585	0	3.63758615972639	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D22:G22))	=IF(B22=0,B$2,0)+(1-B$2)*POISSON(B22,I22,FALSE)	=LN(J22)
23	14	0		10.9150884642146	0	3.61091791264422	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D23:G23))	=IF(B23=0,B$2,0)+(1-B$2)*POISSON(B23,I23,FALSE)	=LN(J23)
24	15	0		10.2219412836547	1	3.58351893845611	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D24:G24))	=IF(B24=0,B$2,0)+(1-B$2)*POISSON(B24,I24,FALSE)	=LN(J24)
25	16	1		10.7684849900227	1	3.25809653802148	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D25:G25))	=IF(B25=0,B$2,0)+(1-B$2)*POISSON(B25,I25,FALSE)	=LN(J25)
26	17	2		12.2060726455302	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D26:G26))	=IF(B26=0,B$2,0)+(1-B$2)*POISSON(B26,I26,FALSE)	=LN(J26)
27	18	0		10.7684849900227	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D27:G27))	=IF(B27=0,B$2,0)+(1-B$2)*POISSON(B27,I27,FALSE)	=LN(J27)
28	19	6		11.1913418408428	1	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D28:G28))	=IF(B28=0,B$2,0)+(1-B$2)*POISSON(B28,I28,FALSE)	=LN(J28)
29	20	0		10.3889953683178	1	3.58351893845611	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D29:G29))	=IF(B29=0,B$2,0)+(1-B$2)*POISSON(B29,I29,FALSE)	=LN(J29)
30	21	2		10.7684849900227	1	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D30:G30))	=IF(B30=0,B$2,0)+(1-B$2)*POISSON(B30,I30,FALSE)	=LN(J30)
31	22	0		11.1913418408428	1	3.46573590279973	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D31:G31))	=IF(B31=0,B$2,0)+(1-B$2)*POISSON(B31,I31,FALSE)	=LN(J31)
32	23	0		11.1913418408428	1	3.43398720448515	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D32:G32))	=IF(B32=0,B$2,0)+(1-B$2)*POISSON(B32,I32,FALSE)	=LN(J32)
33	24	2		11.7360690162844	1	3.80666248977032	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D33:G33))	=IF(B33=0,B$2,0)+(1-B$2)*POISSON(B33,I33,FALSE)	=LN(J33)
34	25	0		11.3793940723457	0	4.27666611901606	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D34:G34))	=IF(B34=0,B$2,0)+(1-B$2)*POISSON(B34,I34,FALSE)	=LN(J34)
35	26	0		10.3889953683178	0	4.21950770517611	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D35:G35))	=IF(B35=0,B$2,0)+(1-B$2)*POISSON(B35,I35,FALSE)	=LN(J35)
36	27	0		10.6572593549125	1	3.49650756146648	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D36:G36))	=IF(B36=0,B$2,0)+(1-B$2)*POISSON(B36,I36,FALSE)	=LN(J36)
37	28	0		12.0725412529057	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D37:G37))	=IF(B37=0,B$2,0)+(1-B$2)*POISSON(B37,I37,FALSE)	=LN(J37)
38	29	0		10.9150884642146	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D38:G38))	=IF(B38=0,B$2,0)+(1-B$2)*POISSON(B38,I38,FALSE)	=LN(J38)
39	30	0		10.9150884642146	0	3.52636052461616	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D39:G39))	=IF(B39=0,B$2,0)+(1-B$2)*POISSON(B39,I39,FALSE)	=LN(J39)
40	31	0		11.1913418408428	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D40:G40))	=IF(B40=0,B$2,0)+(1-B$2)*POISSON(B40,I40,FALSE)	=LN(J40)
41	32	0		10.2219412836547	1	3.13549421592915	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D41:G41))	=IF(B41=0,B$2,0)+(1-B$2)*POISSON(B41,I41,FALSE)	=LN(J41)
42	33	0		11.3793940723457	0	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D42:G42))	=IF(B42=0,B$2,0)+(1-B$2)*POISSON(B42,I42,FALSE)	=LN(J42)
43	34	0		9.07680897935166	1	3.40119738166216	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D43:G43))	=IF(B43=0,B$2,0)+(1-B$2)*POISSON(B43,I43,FALSE)	=LN(J43)
44	35	0		10.0212705881925	1	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D44:G44))	=IF(B44=0,B$2,0)+(1-B$2)*POISSON(B44,I44,FALSE)	=LN(J44)
45	36	0		11.0821425488778	0	4.06044301054642	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D45:G45))	=IF(B45=0,B$2,0)+(1-B$2)*POISSON(B45,I45,FALSE)	=LN(J45)
46	37	2		10.2219412836547	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D46:G46))	=IF(B46=0,B$2,0)+(1-B$2)*POISSON(B46,I46,FALSE)	=LN(J46)
47	38	2		12.0725412529057	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D47:G47))	=IF(B47=0,B$2,0)+(1-B$2)*POISSON(B47,I47,FALSE)	=LN(J47)
48	39	1		11.0821425488778	0	4.17438726989564	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D48:G48))	=IF(B48=0,B$2,0)+(1-B$2)*POISSON(B48,I48,FALSE)	=LN(J48)
49	40	0		9.52879410309472	1	2.70805020110221	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D49:G49))	=IF(B49=0,B$2,0)+(1-B$2)*POISSON(B49,I49,FALSE)	=LN(J49)
50	41	0		11.0821425488778	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D50:G50))	=IF(B50=0,B$2,0)+(1-B$2)*POISSON(B50,I50,FALSE)	=LN(J50)
51	42	0		11.3793940723457	1	4.12713438504509	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D51:G51))	=IF(B51=0,B$2,0)+(1-B$2)*POISSON(B51,I51,FALSE)	=LN(J51)
52	43	0		11.3793940723457	0	4.17438726989564	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D52:G52))	=IF(B52=0,B$2,0)+(1-B$2)*POISSON(B52,I52,FALSE)	=LN(J52)
53	44	0		10.5320962119585	1	3.55534806148941	6		=B$1*EXP(SUMPRODUCT(D$7:G$7,D53:G53))	=IF(B53=0,B$2,0)+(1-B$2)*POISSON(B53,I53,FALSE)	=LN(J53)
54	45	0		10.7684849900227	0	3.2188758248682	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D54:G54))	=IF(B54=0,B$2,0)+(1-B$2)*POISSON(B54,I54,FALSE)	=LN(J54)
55	46	0		11.3793940723457	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D55:G55))	=IF(B55=0,B$2,0)+(1-B$2)*POISSON(B55,I55,FALSE)	=LN(J55)
56	47	0		11.7360690162844	0	3.04452243772342	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D56:G56))	=IF(B56=0,B$2,0)+(1-B$2)*POISSON(B56,I56,FALSE)	=LN(J56)
57	48	0		10.7684849900227	0	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D57:G57))	=IF(B57=0,B$2,0)+(1-B$2)*POISSON(B57,I57,FALSE)	=LN(J57)
58	49	0		10.3889953683178	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D58:G58))	=IF(B58=0,B$2,0)+(1-B$2)*POISSON(B58,I58,FALSE)	=LN(J58)
59	50	0		10.3889953683178	1	2.63905732961526	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D59:G59))	=IF(B59=0,B$2,0)+(1-B$2)*POISSON(B59,I59,FALSE)	=LN(J59)
60	51	0		11.0821425488778	0	3.73766961828337	5		=B$1*EXP(SUMPRODUCT(D$7:G$7,D60:G60))	=IF(B60=0,B$2,0)+(1-B$2)*POISSON(B60,I60,FALSE)	=LN(J60)
61	52	6		9.76995615991161	0	3.29583686600433	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D61:G61))	=IF(B61=0,B$2,0)+(1-B$2)*POISSON(B61,I61,FALSE)	=LN(J61)
62	53	16		9.76995615991161	0	3.13549421592915	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D62:G62))	=IF(B62=0,B$2,0)+(1-B$2)*POISSON(B62,I62,FALSE)	=LN(J62)
63	54	1		10.2219412836547	1	3.46573590279973	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D63:G63))	=IF(B63=0,B$2,0)+(1-B$2)*POISSON(B63,I63,FALSE)	=LN(J63)
64	55	0		11.3793940723457	1	3.55534806148941	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D64:G64))	=IF(B64=0,B$2,0)+(1-B$2)*POISSON(B64,I64,FALSE)	=LN(J64)
65	56	0		11.7360690162844	0	3.93182563272433	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D65:G65))	=IF(B65=0,B$2,0)+(1-B$2)*POISSON(B65,I65,FALSE)	=LN(J65)
66	57	0		10.3889953683178	1	3.73766961828337	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D66:G66))	=IF(B66=0,B$2,0)+(1-B$2)*POISSON(B66,I66,FALSE)	=LN(J66)
67	58	1		11.1913418408428	1	3.49650756146648	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D67:G67))	=IF(B67=0,B$2,0)+(1-B$2)*POISSON(B67,I67,FALSE)	=LN(J67)
68	59	0		10.2219412836547	0	3.61091791264422	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D68:G68))	=IF(B68=0,B$2,0)+(1-B$2)*POISSON(B68,I68,FALSE)	=LN(J68)
69	60	0		10.9150884642146	0	3.89182029811063	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D69:G69))	=IF(B69=0,B$2,0)+(1-B$2)*POISSON(B69,I69,FALSE)	=LN(J69)

Problem 4 -- ZIP reg

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	6.623121			LL =	-4297.472				
2	pi	0.743284								
3	B_inc	-0.089147								
4	B_sex	-0.132673								
5	B_age	0.114116								
6	B_size	0.019557								
7				-0.089147	-0.132673	0.114116	0.019557			
8										
9	ID	Total		Income	Sex	Age	Size		lambda	P(Y=y)
10	1	0		11.38	1	3.87	2		3.401926	0.751835
11	2	5		9.77	1	4.04	1		3.926982	0.039364
12	3	0		11.08	0	3.33	2		3.750935	0.749316
13	4	0		10.92	1	3.95	3		3.648895	0.749964
14	5	0		10.92	1	2.83	3		3.211824	0.753626
15	6	0		10.92	0	2.94	3		3.714352	0.749541
16	7	0		11.19	0	3.66	2		3.857748	0.748705
17	8	1		11.74	0	4.08	2		3.852663	0.020991
18	9	0		10.02	0	4.25	1		4.488797	0.746168
19	10	0		10.92	0	3.85	3		4.118794	0.74746
20	11	1		10.77	0	3.93	2		4.130478	0.017045
21	12	0		10.92	0	3.99	2		4.103526	0.747524
22	13	3		10.53	0	3.64	2		4.079151	0.049143
23	14	0		10.92	0	3.61	1		3.854132	0.748725
24	15	0		10.22	1	3.58	3		3.721968	0.749493
25	16	1		10.77	1	3.26	3		3.41574	0.028807
26	17	2		12.21	0	3.66	2		3.524096	0.046992
27	18	0		10.77	0	3.95	2		4.139641	0.747374
28	19	6		11.19	1	3.33	2		3.253065	0.016334
29	20	0		10.39	1	3.58	2		3.595932	0.750327
30	21	2		10.77	1	3.33	4		3.512782	0.047222
31	22	0		11.19	1	3.47	2		3.303015	0.752724
32	23	0		11.19	1	3.43	2		3.29107	0.752838
33	24	2		11.74	1	3.81	2		3.271277	0.052139
34	25	0		11.38	0	4.28	2		4.068542	0.747675
35	26	0		10.39	0	4.22	2		4.415197	0.746389
36	27	0		10.66	1	3.50	4		3.614933	0.750195
37	28	0		12.07	0	3.95	2		3.685319	0.749725
38	29	0		10.92	1	3.81	3		3.589186	0.750375
39	30	0		10.92	0	3.53	3		3.969385	0.748133
40	31	0		11.19	1	3.37	2		3.266118	0.753079
41	32	0		10.22	1	3.14	4		3.606303	0.750255
42	33	0		11.38	0	3.33	4		3.798555	0.749036
43	34	0		9.08	1	3.40	1		3.882258	0.748574
44	35	0		10.02	1	3.53	1		3.620111	0.750159
45	36	0		11.08	0	4.06	4		4.23856	0.746988
46	37	2		10.22	1	3.69	2		3.694033	0.043564
47	38	2		12.07	1	3.69	2		3.132227	0.054931
48	39	1		11.08	0	4.17	1		4.049343	0.018123
49	40	0		9.53	1	2.71	3		3.58278	0.750421
50	41	0		11.08	1	3.81	3		3.53613	0.750761
51	42	0		11.38	1	4.13	3		3.571927	0.750499

NBD Regression

The explanatory variables may not fully capture the differences among individuals

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

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

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

NBD Regression Results

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

	A	B	C	D	E	F	G	H	I	J	K
1	r	0.1388			LL =	=SUM(K10:K2737)					
2	alpha	8.1979									
3	B_inc	0.0734									
4	B_sex	-0.0093									
5	B_age	0.9022									
6	B_size	-0.0243									
7				=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)				
8											
9	ID	Total		Income	Sex	Age	Size		exp(BX)	P(Y=y)	ln[P(Y=y)]
10	1	0		11.3793940723457	1	3.87120101090789	2		=EXP(SUMPRODUCT(D$7:G$7,D10:G10))	=EXP(GAMMALN(B$1+B10)-GAMMALN(B$1))/FACT(B10)*(B$2/(B$2+I10))^B$1*(I10/(B$2+I10))^B10	=LN(J10)
11	2	5		9.76995615991161	1	4.04305126783455	1		=EXP(SUMPRODUCT(D$7:G$7,D11:G11))	=EXP(GAMMALN(B$1+B11)-GAMMALN(B$1))/FACT(B11)*(B$2/(B$2+I11))^B$1*(I11/(B$2+I11))^B11	=LN(J11)
12	3	0		11.0821425488778	0	3.3322045101752	2		=EXP(SUMPRODUCT(D$7:G$7,D12:G12))	=EXP(GAMMALN(B$1+B12)-GAMMALN(B$1))/FACT(B12)*(B$2/(B$2+I12))^B$1*(I12/(B$2+I12))^B12	=LN(J12)
13	4	0		10.9150884642146	1	3.95124371858143	3		=EXP(SUMPRODUCT(D$7:G$7,D13:G13))	=EXP(GAMMALN(B$1+B13)-GAMMALN(B$1))/FACT(B13)*(B$2/(B$2+I13))^B$1*(I13/(B$2+I13))^B13	=LN(J13)
14	5	0		10.9150884642146	1	2.83321334405622	3		=EXP(SUMPRODUCT(D$7:G$7,D14:G14))	=EXP(GAMMALN(B$1+B14)-GAMMALN(B$1))/FACT(B14)*(B$2/(B$2+I14))^B$1*(I14/(B$2+I14))^B14	=LN(J14)
15	6	0		10.9150884642146	0	2.94443897916644	3		=EXP(SUMPRODUCT(D$7:G$7,D15:G15))	=EXP(GAMMALN(B$1+B15)-GAMMALN(B$1))/FACT(B15)*(B$2/(B$2+I15))^B$1*(I15/(B$2+I15))^B15	=LN(J15)
16	7	0		11.1913418408428	0	3.66356164612965	2		=EXP(SUMPRODUCT(D$7:G$7,D16:G16))	=EXP(GAMMALN(B$1+B16)-GAMMALN(B$1))/FACT(B16)*(B$2/(B$2+I16))^B$1*(I16/(B$2+I16))^B16	=LN(J16)
17	8	1		11.7360690162844	0	4.07753744390572	2		=EXP(SUMPRODUCT(D$7:G$7,D17:G17))	=EXP(GAMMALN(B$1+B17)-GAMMALN(B$1))/FACT(B17)*(B$2/(B$2+I17))^B$1*(I17/(B$2+I17))^B17	=LN(J17)
18	9	0		10.0212705881925	0	4.24849524204936	1		=EXP(SUMPRODUCT(D$7:G$7,D18:G18))	=EXP(GAMMALN(B$1+B18)-GAMMALN(B$1))/FACT(B18)*(B$2/(B$2+I18))^B$1*(I18/(B$2+I18))^B18	=LN(J18)
19	10	0		10.9150884642146	0	3.85014760171006	3		=EXP(SUMPRODUCT(D$7:G$7,D19:G19))	=EXP(GAMMALN(B$1+B19)-GAMMALN(B$1))/FACT(B19)*(B$2/(B$2+I19))^B$1*(I19/(B$2+I19))^B19	=LN(J19)
20	11	1		10.7684849900227	0	3.93182563272433	2		=EXP(SUMPRODUCT(D$7:G$7,D20:G20))	=EXP(GAMMALN(B$1+B20)-GAMMALN(B$1))/FACT(B20)*(B$2/(B$2+I20))^B$1*(I20/(B$2+I20))^B20	=LN(J20)
21	12	0		10.9150884642146	0	3.98898404656427	2		=EXP(SUMPRODUCT(D$7:G$7,D21:G21))	=EXP(GAMMALN(B$1+B21)-GAMMALN(B$1))/FACT(B21)*(B$2/(B$2+I21))^B$1*(I21/(B$2+I21))^B21	=LN(J21)
22	13	3		10.5320962119585	0	3.63758615972639	2		=EXP(SUMPRODUCT(D$7:G$7,D22:G22))	=EXP(GAMMALN(B$1+B22)-GAMMALN(B$1))/FACT(B22)*(B$2/(B$2+I22))^B$1*(I22/(B$2+I22))^B22	=LN(J22)
23	14	0		10.9150884642146	0	3.61091791264422	1		=EXP(SUMPRODUCT(D$7:G$7,D23:G23))	=EXP(GAMMALN(B$1+B23)-GAMMALN(B$1))/FACT(B23)*(B$2/(B$2+I23))^B$1*(I23/(B$2+I23))^B23	=LN(J23)
24	15	0		10.2219412836547	1	3.58351893845611	3		=EXP(SUMPRODUCT(D$7:G$7,D24:G24))	=EXP(GAMMALN(B$1+B24)-GAMMALN(B$1))/FACT(B24)*(B$2/(B$2+I24))^B$1*(I24/(B$2+I24))^B24	=LN(J24)
25	16	1		10.7684849900227	1	3.25809653802148	3		=EXP(SUMPRODUCT(D$7:G$7,D25:G25))	=EXP(GAMMALN(B$1+B25)-GAMMALN(B$1))/FACT(B25)*(B$2/(B$2+I25))^B$1*(I25/(B$2+I25))^B25	=LN(J25)
26	17	2		12.2060726455302	0	3.66356164612965	2		=EXP(SUMPRODUCT(D$7:G$7,D26:G26))	=EXP(GAMMALN(B$1+B26)-GAMMALN(B$1))/FACT(B26)*(B$2/(B$2+I26))^B$1*(I26/(B$2+I26))^B26	=LN(J26)
27	18	0		10.7684849900227	0	3.95124371858143	2		=EXP(SUMPRODUCT(D$7:G$7,D27:G27))	=EXP(GAMMALN(B$1+B27)-GAMMALN(B$1))/FACT(B27)*(B$2/(B$2+I27))^B$1*(I27/(B$2+I27))^B27	=LN(J27)
28	19	6		11.1913418408428	1	3.3322045101752	2		=EXP(SUMPRODUCT(D$7:G$7,D28:G28))	=EXP(GAMMALN(B$1+B28)-GAMMALN(B$1))/FACT(B28)*(B$2/(B$2+I28))^B$1*(I28/(B$2+I28))^B28	=LN(J28)
29	20	0		10.3889953683178	1	3.58351893845611	2		=EXP(SUMPRODUCT(D$7:G$7,D29:G29))	=EXP(GAMMALN(B$1+B29)-GAMMALN(B$1))/FACT(B29)*(B$2/(B$2+I29))^B$1*(I29/(B$2+I29))^B29	=LN(J29)
30	21	2		10.7684849900227	1	3.3322045101752	4		=EXP(SUMPRODUCT(D$7:G$7,D30:G30))	=EXP(GAMMALN(B$1+B30)-GAMMALN(B$1))/FACT(B30)*(B$2/(B$2+I30))^B$1*(I30/(B$2+I30))^B30	=LN(J30)
31	22	0		11.1913418408428	1	3.46573590279973	2		=EXP(SUMPRODUCT(D$7:G$7,D31:G31))	=EXP(GAMMALN(B$1+B31)-GAMMALN(B$1))/FACT(B31)*(B$2/(B$2+I31))^B$1*(I31/(B$2+I31))^B31	=LN(J31)
32	23	0		11.1913418408428	1	3.43398720448515	2		=EXP(SUMPRODUCT(D$7:G$7,D32:G32))	=EXP(GAMMALN(B$1+B32)-GAMMALN(B$1))/FACT(B32)*(B$2/(B$2+I32))^B$1*(I32/(B$2+I32))^B32	=LN(J32)
33	24	2		11.7360690162844	1	3.80666248977032	2		=EXP(SUMPRODUCT(D$7:G$7,D33:G33))	=EXP(GAMMALN(B$1+B33)-GAMMALN(B$1))/FACT(B33)*(B$2/(B$2+I33))^B$1*(I33/(B$2+I33))^B33	=LN(J33)
34	25	0		11.3793940723457	0	4.27666611901606	2		=EXP(SUMPRODUCT(D$7:G$7,D34:G34))	=EXP(GAMMALN(B$1+B34)-GAMMALN(B$1))/FACT(B34)*(B$2/(B$2+I34))^B$1*(I34/(B$2+I34))^B34	=LN(J34)
35	26	0		10.3889953683178	0	4.21950770517611	2		=EXP(SUMPRODUCT(D$7:G$7,D35:G35))	=EXP(GAMMALN(B$1+B35)-GAMMALN(B$1))/FACT(B35)*(B$2/(B$2+I35))^B$1*(I35/(B$2+I35))^B35	=LN(J35)
36	27	0		10.6572593549125	1	3.49650756146648	4		=EXP(SUMPRODUCT(D$7:G$7,D36:G36))	=EXP(GAMMALN(B$1+B36)-GAMMALN(B$1))/FACT(B36)*(B$2/(B$2+I36))^B$1*(I36/(B$2+I36))^B36	=LN(J36)
37	28	0		12.0725412529057	0	3.95124371858143	2		=EXP(SUMPRODUCT(D$7:G$7,D37:G37))	=EXP(GAMMALN(B$1+B37)-GAMMALN(B$1))/FACT(B37)*(B$2/(B$2+I37))^B$1*(I37/(B$2+I37))^B37	=LN(J37)
38	29	0		10.9150884642146	1	3.80666248977032	3		=EXP(SUMPRODUCT(D$7:G$7,D38:G38))	=EXP(GAMMALN(B$1+B38)-GAMMALN(B$1))/FACT(B38)*(B$2/(B$2+I38))^B$1*(I38/(B$2+I38))^B38	=LN(J38)
39	30	0		10.9150884642146	0	3.52636052461616	3		=EXP(SUMPRODUCT(D$7:G$7,D39:G39))	=EXP(GAMMALN(B$1+B39)-GAMMALN(B$1))/FACT(B39)*(B$2/(B$2+I39))^B$1*(I39/(B$2+I39))^B39	=LN(J39)
40	31	0		11.1913418408428	1	3.36729582998647	2		=EXP(SUMPRODUCT(D$7:G$7,D40:G40))	=EXP(GAMMALN(B$1+B40)-GAMMALN(B$1))/FACT(B40)*(B$2/(B$2+I40))^B$1*(I40/(B$2+I40))^B40	=LN(J40)
41	32	0		10.2219412836547	1	3.13549421592915	4		=EXP(SUMPRODUCT(D$7:G$7,D41:G41))	=EXP(GAMMALN(B$1+B41)-GAMMALN(B$1))/FACT(B41)*(B$2/(B$2+I41))^B$1*(I41/(B$2+I41))^B41	=LN(J41)
42	33	0		11.3793940723457	0	3.3322045101752	4		=EXP(SUMPRODUCT(D$7:G$7,D42:G42))	=EXP(GAMMALN(B$1+B42)-GAMMALN(B$1))/FACT(B42)*(B$2/(B$2+I42))^B$1*(I42/(B$2+I42))^B42	=LN(J42)
43	34	0		9.07680897935166	1	3.40119738166216	1		=EXP(SUMPRODUCT(D$7:G$7,D43:G43))	=EXP(GAMMALN(B$1+B43)-GAMMALN(B$1))/FACT(B43)*(B$2/(B$2+I43))^B$1*(I43/(B$2+I43))^B43	=LN(J43)
44	35	0		10.0212705881925	1	3.52636052461616	1		=EXP(SUMPRODUCT(D$7:G$7,D44:G44))	=EXP(GAMMALN(B$1+B44)-GAMMALN(B$1))/FACT(B44)*(B$2/(B$2+I44))^B$1*(I44/(B$2+I44))^B44	=LN(J44)
45	36	0		11.0821425488778	0	4.06044301054642	4		=EXP(SUMPRODUCT(D$7:G$7,D45:G45))	=EXP(GAMMALN(B$1+B45)-GAMMALN(B$1))/FACT(B45)*(B$2/(B$2+I45))^B$1*(I45/(B$2+I45))^B45	=LN(J45)
46	37	2		10.2219412836547	1	3.68887945411394	2		=EXP(SUMPRODUCT(D$7:G$7,D46:G46))	=EXP(GAMMALN(B$1+B46)-GAMMALN(B$1))/FACT(B46)*(B$2/(B$2+I46))^B$1*(I46/(B$2+I46))^B46	=LN(J46)
47	38	2		12.0725412529057	1	3.68887945411394	2		=EXP(SUMPRODUCT(D$7:G$7,D47:G47))	=EXP(GAMMALN(B$1+B47)-GAMMALN(B$1))/FACT(B47)*(B$2/(B$2+I47))^B$1*(I47/(B$2+I47))^B47	=LN(J47)
48	39	1		11.0821425488778	0	4.17438726989564	1		=EXP(SUMPRODUCT(D$7:G$7,D48:G48))	=EXP(GAMMALN(B$1+B48)-GAMMALN(B$1))/FACT(B48)*(B$2/(B$2+I48))^B$1*(I48/(B$2+I48))^B48	=LN(J48)
49	40	0		9.52879410309472	1	2.70805020110221	3		=EXP(SUMPRODUCT(D$7:G$7,D49:G49))	=EXP(GAMMALN(B$1+B49)-GAMMALN(B$1))/FACT(B49)*(B$2/(B$2+I49))^B$1*(I49/(B$2+I49))^B49	=LN(J49)
50	41	0		11.0821425488778	1	3.80666248977032	3		=EXP(SUMPRODUCT(D$7:G$7,D50:G50))	=EXP(GAMMALN(B$1+B50)-GAMMALN(B$1))/FACT(B50)*(B$2/(B$2+I50))^B$1*(I50/(B$2+I50))^B50	=LN(J50)
51	42	0		11.3793940723457	1	4.12713438504509	3		=EXP(SUMPRODUCT(D$7:G$7,D51:G51))	=EXP(GAMMALN(B$1+B51)-GAMMALN(B$1))/FACT(B51)*(B$2/(B$2+I51))^B$1*(I51/(B$2+I51))^B51	=LN(J51)
52	43	0		11.3793940723457	0	4.17438726989564	3		=EXP(SUMPRODUCT(D$7:G$7,D52:G52))	=EXP(GAMMALN(B$1+B52)-GAMMALN(B$1))/FACT(B52)*(B$2/(B$2+I52))^B$1*(I52/(B$2+I52))^B52	=LN(J52)
53	44	0		10.5320962119585	1	3.55534806148941	6		=EXP(SUMPRODUCT(D$7:G$7,D53:G53))	=EXP(GAMMALN(B$1+B53)-GAMMALN(B$1))/FACT(B53)*(B$2/(B$2+I53))^B$1*(I53/(B$2+I53))^B53	=LN(J53)
54	45	0		10.7684849900227	0	3.2188758248682	1		=EXP(SUMPRODUCT(D$7:G$7,D54:G54))	=EXP(GAMMALN(B$1+B54)-GAMMALN(B$1))/FACT(B54)*(B$2/(B$2+I54))^B$1*(I54/(B$2+I54))^B54	=LN(J54)
55	46	0		11.3793940723457	1	3.36729582998647	2		=EXP(SUMPRODUCT(D$7:G$7,D55:G55))	=EXP(GAMMALN(B$1+B55)-GAMMALN(B$1))/FACT(B55)*(B$2/(B$2+I55))^B$1*(I55/(B$2+I55))^B55	=LN(J55)
56	47	0		11.7360690162844	0	3.04452243772342	4		=EXP(SUMPRODUCT(D$7:G$7,D56:G56))	=EXP(GAMMALN(B$1+B56)-GAMMALN(B$1))/FACT(B56)*(B$2/(B$2+I56))^B$1*(I56/(B$2+I56))^B56	=LN(J56)
57	48	0		10.7684849900227	0	3.52636052461616	1		=EXP(SUMPRODUCT(D$7:G$7,D57:G57))	=EXP(GAMMALN(B$1+B57)-GAMMALN(B$1))/FACT(B57)*(B$2/(B$2+I57))^B$1*(I57/(B$2+I57))^B57	=LN(J57)
58	49	0		10.3889953683178	1	2.83321334405622	3		=EXP(SUMPRODUCT(D$7:G$7,D58:G58))	=EXP(GAMMALN(B$1+B58)-GAMMALN(B$1))/FACT(B58)*(B$2/(B$2+I58))^B$1*(I58/(B$2+I58))^B58	=LN(J58)
59	50	0		10.3889953683178	1	2.63905732961526	3		=EXP(SUMPRODUCT(D$7:G$7,D59:G59))	=EXP(GAMMALN(B$1+B59)-GAMMALN(B$1))/FACT(B59)*(B$2/(B$2+I59))^B$1*(I59/(B$2+I59))^B59	=LN(J59)
60	51	0		11.0821425488778	0	3.73766961828337	5		=EXP(SUMPRODUCT(D$7:G$7,D60:G60))	=EXP(GAMMALN(B$1+B60)-GAMMALN(B$1))/FACT(B60)*(B$2/(B$2+I60))^B$1*(I60/(B$2+I60))^B60	=LN(J60)
61	52	6		9.76995615991161	0	3.29583686600433	1		=EXP(SUMPRODUCT(D$7:G$7,D61:G61))	=EXP(GAMMALN(B$1+B61)-GAMMALN(B$1))/FACT(B61)*(B$2/(B$2+I61))^B$1*(I61/(B$2+I61))^B61	=LN(J61)
62	53	16		9.76995615991161	0	3.13549421592915	1		=EXP(SUMPRODUCT(D$7:G$7,D62:G62))	=EXP(GAMMALN(B$1+B62)-GAMMALN(B$1))/FACT(B62)*(B$2/(B$2+I62))^B$1*(I62/(B$2+I62))^B62	=LN(J62)
63	54	1		10.2219412836547	1	3.46573590279973	1		=EXP(SUMPRODUCT(D$7:G$7,D63:G63))	=EXP(GAMMALN(B$1+B63)-GAMMALN(B$1))/FACT(B63)*(B$2/(B$2+I63))^B$1*(I63/(B$2+I63))^B63	=LN(J63)
64	55	0		11.3793940723457	1	3.55534806148941	2		=EXP(SUMPRODUCT(D$7:G$7,D64:G64))	=EXP(GAMMALN(B$1+B64)-GAMMALN(B$1))/FACT(B64)*(B$2/(B$2+I64))^B$1*(I64/(B$2+I64))^B64	=LN(J64)
65	56	0		11.7360690162844	0	3.93182563272433	4		=EXP(SUMPRODUCT(D$7:G$7,D65:G65))	=EXP(GAMMALN(B$1+B65)-GAMMALN(B$1))/FACT(B65)*(B$2/(B$2+I65))^B$1*(I65/(B$2+I65))^B65	=LN(J65)
66	57	0		10.3889953683178	1	3.73766961828337	2		=EXP(SUMPRODUCT(D$7:G$7,D66:G66))	=EXP(GAMMALN(B$1+B66)-GAMMALN(B$1))/FACT(B66)*(B$2/(B$2+I66))^B$1*(I66/(B$2+I66))^B66	=LN(J66)
67	58	1		11.1913418408428	1	3.49650756146648	2		=EXP(SUMPRODUCT(D$7:G$7,D67:G67))	=EXP(GAMMALN(B$1+B67)-GAMMALN(B$1))/FACT(B67)*(B$2/(B$2+I67))^B$1*(I67/(B$2+I67))^B67	=LN(J67)
68	59	0		10.2219412836547	0	3.61091791264422	4		=EXP(SUMPRODUCT(D$7:G$7,D68:G68))	=EXP(GAMMALN(B$1+B68)-GAMMALN(B$1))/FACT(B68)*(B$2/(B$2+I68))^B$1*(I68/(B$2+I68))^B68	=LN(J68)
69	60	0		10.9150884642146	0	3.89182029811063	1		=EXP(SUMPRODUCT(D$7:G$7,D69:G69))	=EXP(GAMMALN(B$1+B69)-GAMMALN(B$1))/FACT(B69)*(B$2/(B$2+I69))^B$1*(I69/(B$2+I69))^B69	=LN(J69)
70	61	0		10.9150884642146	1	3.46573590279973	3		=EXP(SUMPRODUCT(D$7:G$7,D70:G70))	=EXP(GAMMALN(B$1+B70)-GAMMALN(B$1))/FACT(B70)*(B$2/(B$2+I70))^B$1*(I70/(B$2+I70))^B70	=LN(J70)
71	62	0		10.3889953683178	1	3.36729582998647	2		=EXP(SUMPRODUCT(D$7:G$7,D71:G71))	=EXP(GAMMALN(B$1+B71)-GAMMALN(B$1))/FACT(B71)*(B$2/(B$2+I71))^B$1*(I71/(B$2+I71))^B71	=LN(J71)
72	63	0		10.3889953683178	0	3.46573590279973	2		=EXP(SUMPRODUCT(D$7:G$7,D72:G72))	=EXP(GAMMALN(B$1+B72)-GAMMALN(B$1))/FACT(B72)*(B$2/(B$2+I72))^B$1*(I72/(B$2+I72))^B72	=LN(J72)
73	64	0		10.7684849900227	1	3.78418963391826	2		=EXP(SUMPRODUCT(D$7:G$7,D73:G73))	=EXP(GAMMALN(B$1+B73)-GAMMALN(B$1))/FACT(B73)*(B$2/(B$2+I73))^B$1*(I73/(B$2+I73))^B73	=LN(J73)
74	65	0		9.76995615991161	1	3.71357206670431	2		=EXP(SUMPRODUCT(D$7:G$7,D74:G74))	=EXP(GAMMALN(B$1+B74)-GAMMALN(B$1))/FACT(B74)*(B$2/(B$2+I74))^B$1*(I74/(B$2+I74))^B74	=LN(J74)
75	66	0		10.0212705881925	1	3.17805383034795	2		=EXP(SUMPRODUCT(D$7:G$7,D75:G75))	=EXP(GAMMALN(B$1+B75)-GAMMALN(B$1))/FACT(B75)*(B$2/(B$2+I75))^B$1*(I75/(B$2+I75))^B75	=LN(J75)
76	67	0		11.3793940723457	0	4.00733318523247	2		=EXP(SUMPRODUCT(D$7:G$7,D76:G76))	=EXP(GAMMALN(B$1+B76)-GAMMALN(B$1))/FACT(B76)*(B$2/(B$2+I76))^B$1*(I76/(B$2+I76))^B76	=LN(J76)

Problem 4 -- NBD reg

	A	B	C	D	E	F	G	H	I	J
1	r	0.139			LL =	-2888.966				
2	alpha	8.198								
3	B_inc	0.073406								
4	B_sex	-0.009272								
5	B_age	0.902164								
6	B_size	-0.024319								
7				0.073406	-0.009272	0.902164	-0.024319			
8										
9	ID	Total		Income	Sex	Age	Size		exp(BX)	P(Y=y)
10	1	0		11.38	1	3.87	2		71.51161	0.729357
11	2	5		9.77	1	4.04	1		76.02589	0.015868
12	3	0		11.08	0	3.33	2		43.42559	0.774672
13	4	0		10.92	1	3.95	3		72.50603	0.728104
14	5	0		10.92	1	2.83	3		26.44384	0.818759
15	6	0		10.92	0	2.94	3		29.50734	0.809189
16	7	0		11.19	0	3.66	2		59.02749	0.746801
17	8	1		11.74	0	4.08	2		89.25195	0.090138
18	9	0		10.02	0	4.25	1		94.07931	0.70456
19	10	0		10.92	0	3.85	3		66.80224	0.735547
20	11	1		10.77	0	3.93	2		72.89216	0.090752
21	12	0		10.92	0	3.99	2		77.57994	0.72197
22	13	3		10.53	0	3.64	2		54.93643	0.027954
23	14	0		10.92	0	3.61	1		56.51751	0.750754
24	15	0		10.22	1	3.58	3		49.45389	0.762891
25	16	1		10.77	1	3.26	3		38.38151	0.089842
26	17	2		12.21	0	3.66	2		63.59217	0.045873
27	18	0		10.77	0	3.95	2		74.18036	0.726032
28	19	6		11.19	1	3.33	2		43.37107	0.008591
29	20	0		10.39	1	3.58	2		51.2965	0.759568
30	21	2		10.77	1	3.33	4		40.04943	0.042567
31	22	0		11.19	1	3.47	2		48.9236	0.76387
32	23	0		11.19	1	3.43	2		47.54218	0.766469
33	24	2		11.74	1	3.81	2		69.25654	0.046252
34	25	0		11.38	0	4.28	2		104.0559	0.69552
35	26	0		10.39	0	4.22	2		91.8963	0.706672
36	27	0		10.66	1	3.50	4		46.07074	0.769319
37	28	0		12.07	0	3.95	2		81.63227	0.717361
38	29	0		10.92	1	3.81	3		63.63946	0.739957
39	30	0		10.92	0	3.53	3		49.88038	0.762111
40	31	0		11.19	1	3.37	2		44.76608	0.771921
41	32	0		10.22	1	3.14	4		32.21815	0.801431
42	33	0		11.38	0	3.33	4		42.27648	0.777095
43	34	0		9.08	1	3.40	1		40.49327	0.780983
44	35	0		10.02	1	3.53	1		48.58828	0.764494
45	36	0		11.08	0	4.06	4		79.79024	0.719426
46	37	2		10.22	1	3.69	2		55.72407	0.045151
47	38	2		12.07	1	3.69	2		63.83215	0.045892
48	39	1		11.08	0	4.17	1		95.12148	0.089876
49	40	0		9.53	1	2.71	3		21.33489	0.837088
50	41	0		11.08	1	3.81	3		64.42466	0.738842
51	42	0		11.38	1	4.13	3		87.92064	0.710658

S-Plus NBD Regression Results

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

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

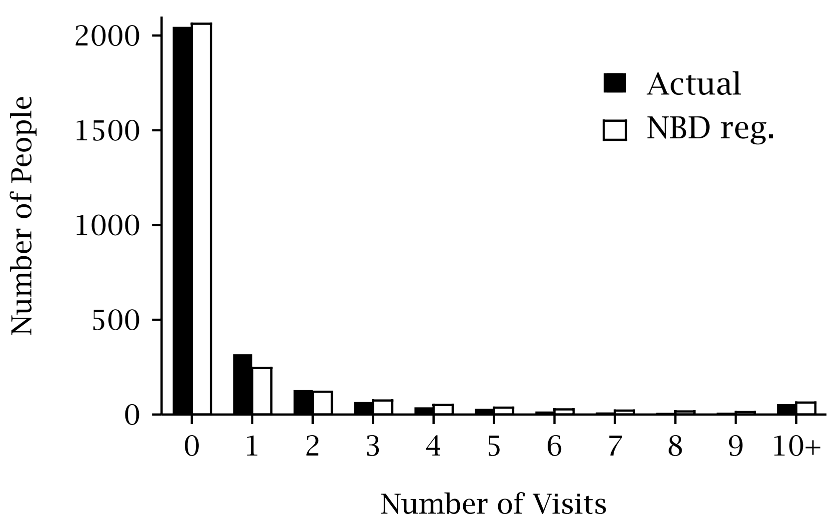

Limdep NBD Regression Results

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

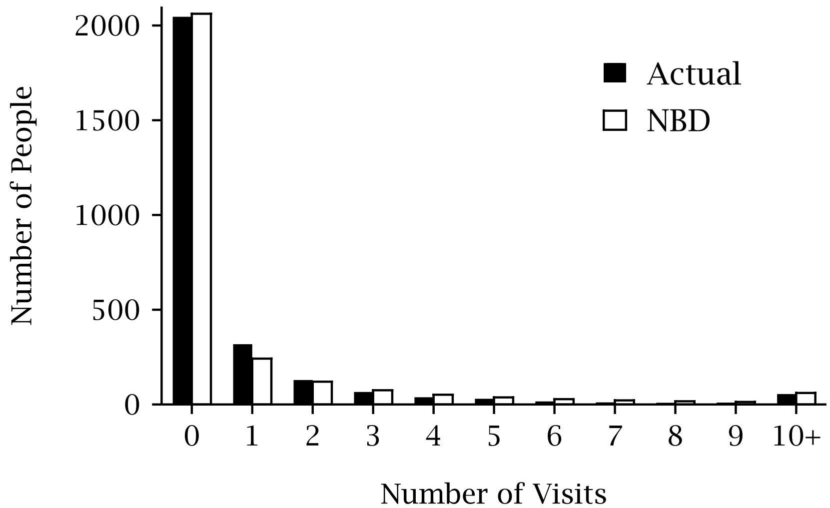
Summary of Regression Results

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

Fit of NBD Regression



Fit of NBD



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

Concepts and Tools Introduced

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

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.

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 θ_i not "explained" by $\mathbf{x}_i$
- Derive the corresponding aggregate distribution

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

Covariates in Timing Models

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

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

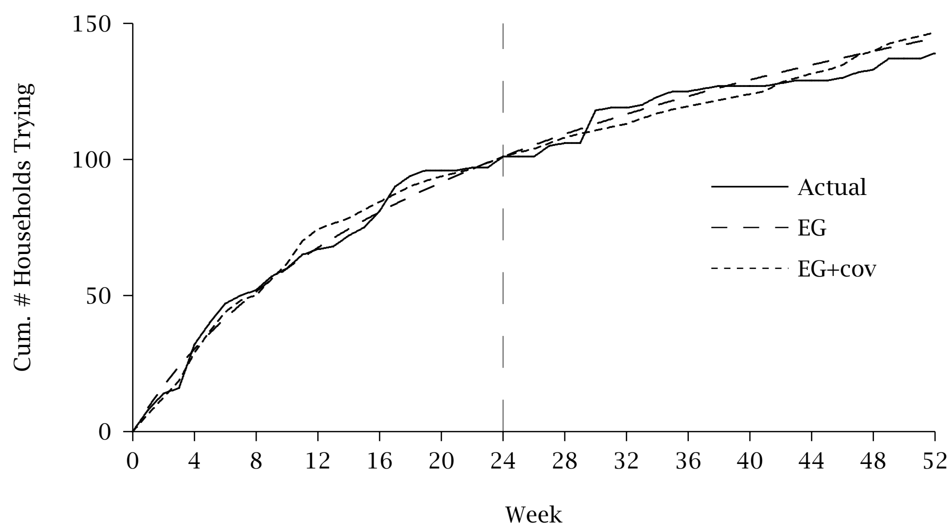
- If the covariates are time-varying (i.e., $\mathbf{x}_{it}$), we incorporate their effects via the hazard rate function

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

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

- Known as "proportional hazards regression"

Comparing EG with EG+cov



Covariates in "Choice" Models

Two options for binary choice:

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

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

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

Problem 5: Understanding the Adoption of a Video-on-Demand Service

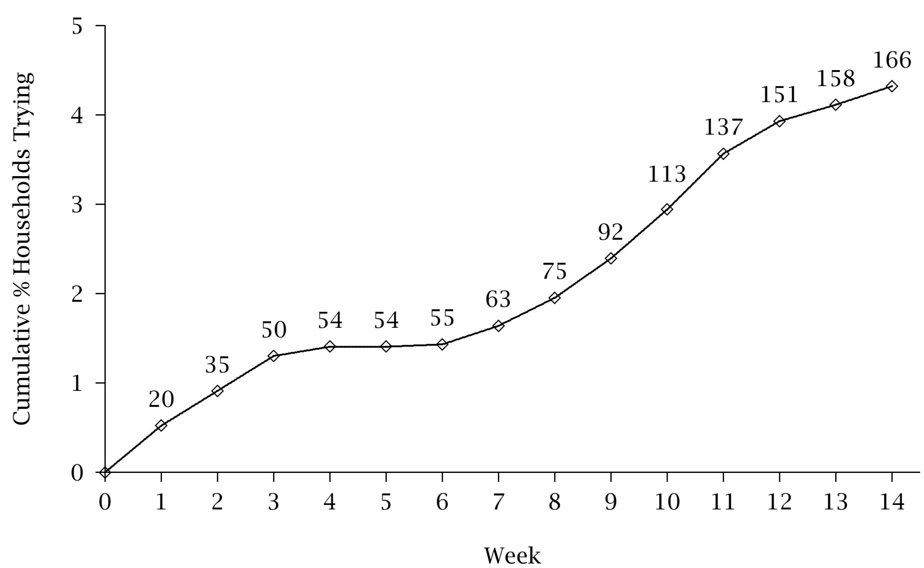
(Introducing Additional Model Structures)

Background

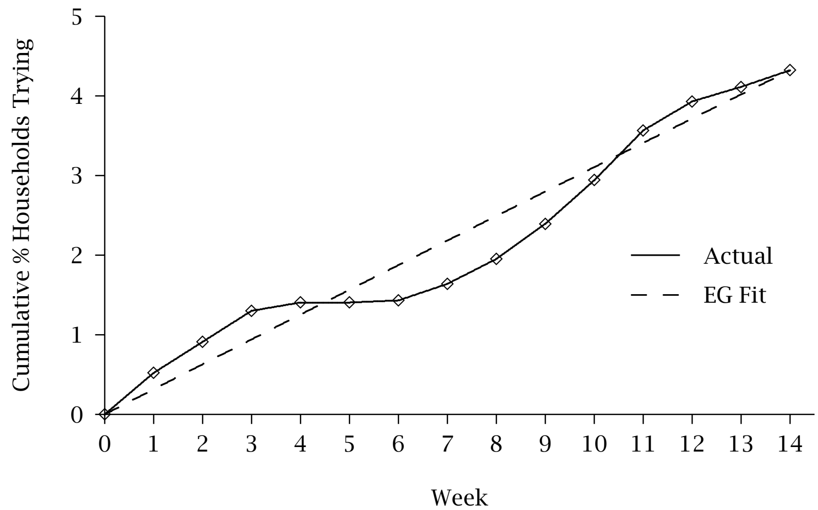
A major telecommunications firm has just started a video-on-demand (VOD) service which covers 3841 households. After the first 14 weeks of operation, a total of 166 households have ordered videos through this service at least once.

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.

Cumulative Trial of VOD Service



Applying the EG Model



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

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 (λ) are not gamma-distributed

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 λ)

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?

Problem 5 -- 2 segment Exponential

	A	B	C	D	E	F	G	H	I	J
1	Video-on-Demand Service				lambda	0.003154	0.003154			
2	# Connected Households		3841		pi	0.452512				
3					LL =	=SUM(I8:I22)				
4										
5										
6		Cum_Trl					P(T<=t)			
7	Week	# HHs	Incr_Trl			Seg 1	Seg 2	Overall		E[T(t)]
8	1	20	=B8			=1-EXP(-F$1*$A8)	=1-EXP(-G$1*$A8)	=F$2*F8+(1-F$2)*G8	=C8*LN(H8)	=C$2*H8
9	2	35	=B9-B8			=1-EXP(-F$1*$A9)	=1-EXP(-G$1*$A9)	=F$2*F9+(1-F$2)*G9	=C9*LN(H9-H8)	=C$2*H9
10	3	50	=B10-B9			=1-EXP(-F$1*$A10)	=1-EXP(-G$1*$A10)	=F$2*F10+(1-F$2)*G10	=C10*LN(H10-H9)	=C$2*H10
11	4	54	=B11-B10			=1-EXP(-F$1*$A11)	=1-EXP(-G$1*$A11)	=F$2*F11+(1-F$2)*G11	=C11*LN(H11-H10)	=C$2*H11
12	5	54	=B12-B11			=1-EXP(-F$1*$A12)	=1-EXP(-G$1*$A12)	=F$2*F12+(1-F$2)*G12	=C12*LN(H12-H11)	=C$2*H12
13	6	55	=B13-B12			=1-EXP(-F$1*$A13)	=1-EXP(-G$1*$A13)	=F$2*F13+(1-F$2)*G13	=C13*LN(H13-H12)	=C$2*H13
14	7	63	=B14-B13			=1-EXP(-F$1*$A14)	=1-EXP(-G$1*$A14)	=F$2*F14+(1-F$2)*G14	=C14*LN(H14-H13)	=C$2*H14
15	8	75	=B15-B14			=1-EXP(-F$1*$A15)	=1-EXP(-G$1*$A15)	=F$2*F15+(1-F$2)*G15	=C15*LN(H15-H14)	=C$2*H15
16	9	92	=B16-B15			=1-EXP(-F$1*$A16)	=1-EXP(-G$1*$A16)	=F$2*F16+(1-F$2)*G16	=C16*LN(H16-H15)	=C$2*H16
17	10	113	=B17-B16			=1-EXP(-F$1*$A17)	=1-EXP(-G$1*$A17)	=F$2*F17+(1-F$2)*G17	=C17*LN(H17-H16)	=C$2*H17
18	11	137	=B18-B17			=1-EXP(-F$1*$A18)	=1-EXP(-G$1*$A18)	=F$2*F18+(1-F$2)*G18	=C18*LN(H18-H17)	=C$2*H18
19	12	151	=B19-B18			=1-EXP(-F$1*$A19)	=1-EXP(-G$1*$A19)	=F$2*F19+(1-F$2)*G19	=C19*LN(H19-H18)	=C$2*H19
20	13	158	=B20-B19			=1-EXP(-F$1*$A20)	=1-EXP(-G$1*$A20)	=F$2*F20+(1-F$2)*G20	=C20*LN(H20-H19)	=C$2*H20
21	14	166	=B21-B20			=1-EXP(-F$1*$A21)	=1-EXP(-G$1*$A21)	=F$2*F21+(1-F$2)*G21	=C21*LN(H21-H20)	=C$2*H21
22	15					=1-EXP(-F$1*$A22)	=1-EXP(-G$1*$A22)	=F$2*F22+(1-F$2)*G22	=(C2-B21)*LN(1-H21)	=C$2*H22
23	16					=1-EXP(-F$1*$A23)	=1-EXP(-G$1*$A23)	=F$2*F23+(1-F$2)*G23		=C$2*H23
24	17					=1-EXP(-F$1*$A24)	=1-EXP(-G$1*$A24)	=F$2*F24+(1-F$2)*G24		=C$2*H24
25	18					=1-EXP(-F$1*$A25)	=1-EXP(-G$1*$A25)	=F$2*F25+(1-F$2)*G25		=C$2*H25
26	19					=1-EXP(-F$1*$A26)	=1-EXP(-G$1*$A26)	=F$2*F26+(1-F$2)*G26		=C$2*H26
27	20					=1-EXP(-F$1*$A27)	=1-EXP(-G$1*$A27)	=F$2*F27+(1-F$2)*G27		=C$2*H27
28	21					=1-EXP(-F$1*$A28)	=1-EXP(-G$1*$A28)	=F$2*F28+(1-F$2)*G28		=C$2*H28
29	22					=1-EXP(-F$1*$A29)	=1-EXP(-G$1*$A29)	=F$2*F29+(1-F$2)*G29		=C$2*H29
30	23					=1-EXP(-F$1*$A30)	=1-EXP(-G$1*$A30)	=F$2*F30+(1-F$2)*G30		=C$2*H30
31	24					=1-EXP(-F$1*$A31)	=1-EXP(-G$1*$A31)	=F$2*F31+(1-F$2)*G31		=C$2*H31
32	25					=1-EXP(-F$1*$A32)	=1-EXP(-G$1*$A32)	=F$2*F32+(1-F$2)*G32		=C$2*H32
33	26					=1-EXP(-F$1*$A33)	=1-EXP(-G$1*$A33)	=F$2*F33+(1-F$2)*G33		=C$2*H33
34	27					=1-EXP(-F$1*$A34)	=1-EXP(-G$1*$A34)	=F$2*F34+(1-F$2)*G34		=C$2*H34
35	28					=1-EXP(-F$1*$A35)	=1-EXP(-G$1*$A35)	=F$2*F35+(1-F$2)*G35		=C$2*H35
36	29					=1-EXP(-F$1*$A36)	=1-EXP(-G$1*$A36)	=F$2*F36+(1-F$2)*G36		=C$2*H36
37	30					=1-EXP(-F$1*$A37)	=1-EXP(-G$1*$A37)	=F$2*F37+(1-F$2)*G37		=C$2*H37
38	31					=1-EXP(-F$1*$A38)	=1-EXP(-G$1*$A38)	=F$2*F38+(1-F$2)*G38		=C$2*H38
39	32					=1-EXP(-F$1*$A39)	=1-EXP(-G$1*$A39)	=F$2*F39+(1-F$2)*G39		=C$2*H39
40	33					=1-EXP(-F$1*$A40)	=1-EXP(-G$1*$A40)	=F$2*F40+(1-F$2)*G40		=C$2*H40
41	34					=1-EXP(-F$1*$A41)	=1-EXP(-G$1*$A41)	=F$2*F41+(1-F$2)*G41		=C$2*H41
42	35					=1-EXP(-F$1*$A42)	=1-EXP(-G$1*$A42)	=F$2*F42+(1-F$2)*G42		=C$2*H42
43	36					=1-EXP(-F$1*$A43)	=1-EXP(-G$1*$A43)	=F$2*F43+(1-F$2)*G43		=C$2*H43
44	37					=1-EXP(-F$1*$A44)	=1-EXP(-G$1*$A44)	=F$2*F44+(1-F$2)*G44		=C$2*H44
45	38					=1-EXP(-F$1*$A45)	=1-EXP(-G$1*$A45)	=F$2*F45+(1-F$2)*G45		=C$2*H45
46	39					=1-EXP(-F$1*$A46)	=1-EXP(-G$1*$A46)	=F$2*F46+(1-F$2)*G46		=C$2*H46
47	40					=1-EXP(-F$1*$A47)	=1-EXP(-G$1*$A47)	=F$2*F47+(1-F$2)*G47		=C$2*H47
48	41					=1-EXP(-F$1*$A48)	=1-EXP(-G$1*$A48)	=F$2*F48+(1-F$2)*G48		=C$2*H48
49	42					=1-EXP(-F$1*$A49)	=1-EXP(-G$1*$A49)	=F$2*F49+(1-F$2)*G49		=C$2*H49
50	43					=1-EXP(-F$1*$A50)	=1-EXP(-G$1*$A50)	=F$2*F50+(1-F$2)*G50		=C$2*H50
51	44					=1-EXP(-F$1*$A51)	=1-EXP(-G$1*$A51)	=F$2*F51+(1-F$2)*G51		=C$2*H51
52	45					=1-EXP(-F$1*$A52)	=1-EXP(-G$1*$A52)	=F$2*F52+(1-F$2)*G52		=C$2*H52
53	46					=1-EXP(-F$1*$A53)	=1-EXP(-G$1*$A53)	=F$2*F53+(1-F$2)*G53		=C$2*H53
54	47					=1-EXP(-F$1*$A54)	=1-EXP(-G$1*$A54)	=F$2*F54+(1-F$2)*G54		=C$2*H54
55	48					=1-EXP(-F$1*$A55)	=1-EXP(-G$1*$A55)	=F$2*F55+(1-F$2)*G55		=C$2*H55
56	49					=1-EXP(-F$1*$A56)	=1-EXP(-G$1*$A56)	=F$2*F56+(1-F$2)*G56		=C$2*H56
57	50					=1-EXP(-F$1*$A57)	=1-EXP(-G$1*$A57)	=F$2*F57+(1-F$2)*G57		=C$2*H57
58	51					=1-EXP(-F$1*$A58)	=1-EXP(-G$1*$A58)	=F$2*F58+(1-F$2)*G58		=C$2*H58
59	52					=1-EXP(-F$1*$A59)	=1-EXP(-G$1*$A59)	=F$2*F59+(1-F$2)*G59		=C$2*H59

Problem 5 -- 2 segment Exponential

	A	B	C	D	E	F	G	H	I	J
1	Video-on-Demand Service				lambda	0.0032	0.0032			
2	# Connected Households		3841		pi	0.4525				
3					LL =	-1122.02				
4										
5										
6		Cum_Trl				P(T<=t)				
7	Week	# HHs	Incr_Trl			Seg 1	Seg 2	Overall		E[T(t)]
8	1	20	20			0.00315	0.00315	0.00315	-115.21	12.09
9	2	35	15			0.00629	0.00629	0.00629	-86.46	24.15
10	3	50	15			0.00942	0.00942	0.00942	-86.51	36.17
11	4	54	4			0.01254	0.01254	0.01254	-23.08	48.15
12	5	54	0			0.01565	0.01565	0.01565	0.00	60.09
13	6	55	1			0.01875	0.01874	0.01875	-5.78	72.00
14	7	63	8			0.02184	0.02183	0.02184	-46.24	83.87
15	8	75	12			0.02492	0.02491	0.02492	-69.39	95.70
16	9	92	17			0.02799	0.02799	0.02799	-98.36	107.49
17	10	113	21			0.03105	0.03105	0.03105	-121.57	119.25
18	11	137	24			0.03410	0.03410	0.03410	-139.01	130.97
19	12	151	14			0.03714	0.03714	0.03714	-81.14	142.65
20	13	158	7			0.04017	0.04017	0.04017	-40.59	154.30
21	14	166	8			0.04319	0.04319	0.04319	-46.41	165.91
22	15					0.04621	0.04621	0.04621	-162.27	177.48
23	16					0.04921	0.04921	0.04921		189.01
24	17					0.05220	0.05220	0.05220		200.51
25	18					0.05519	0.05519	0.05519		211.98
26	19					0.05816	0.05816	0.05816		223.40
27	20					0.06113	0.06113	0.06113		234.80
28	21					0.06409	0.06408	0.06409		246.15
29	22					0.06703	0.06703	0.06703		257.47
30	23					0.06997	0.06997	0.06997		268.76
31	24					0.07290	0.07290	0.07290		280.00
32	25					0.07582	0.07582	0.07582		291.22
33	26					0.07873	0.07873	0.07873		302.40
34	27					0.08163	0.08163	0.08163		313.54
35	28					0.08452	0.08452	0.08452		324.65
36	29					0.08741	0.08740	0.08740		335.72
37	30					0.09028	0.09028	0.09028		346.76
38	31					0.09314	0.09314	0.09314		357.76
39	32					0.09600	0.09600	0.09600		368.73
40	33					0.09885	0.09884	0.09884		379.66
41	34					0.10168	0.10168	0.10168		390.56
42	35					0.10451	0.10451	0.10451		401.43
43	36					0.10733	0.10733	0.10733		412.26
44	37					0.11014	0.11014	0.11014		423.05
45	38					0.11295	0.11294	0.11294		433.82
46	39					0.11574	0.11573	0.11574		444.54
47	40					0.11852	0.11852	0.11852		455.24
48	41					0.12130	0.12129	0.12130		465.90
49	42					0.12407	0.12406	0.12406		476.53
50	43					0.12682	0.12682	0.12682		487.12
51	44					0.12957	0.12957	0.12957		497.68
52	45					0.13232	0.13231	0.13231		508.21
53	46					0.13505	0.13504	0.13504		518.71
54	47					0.13777	0.13777	0.13777		529.17
55	48					0.14049	0.14048	0.14048		539.60
56	49					0.14319	0.14319	0.14319		549.99
57	50					0.14589	0.14589	0.14589		560.36
58	51					0.14858	0.14857	0.14858		570.69
59	52					0.15126	0.15126	0.15126		580.98

The Hazard Rate Function

The hazard rate 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 Rate Function

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

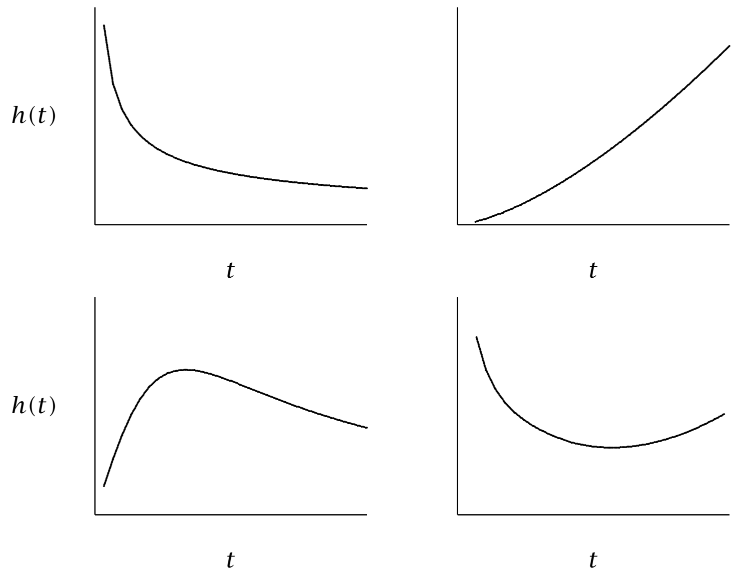
$$F(t) = 1 - \exp\left(-\int_0^t h(u)\,du\right)$$

Example:

the exponential distribution has a *constant* hazard rate, λ

$$F(t) = 1 - \exp\left(-\int_0^t \lambda\,du\right)$$
$$= 1 - e^{-\lambda t}$$

Shapes of the Hazard Rate Function



The Weibull Distribution

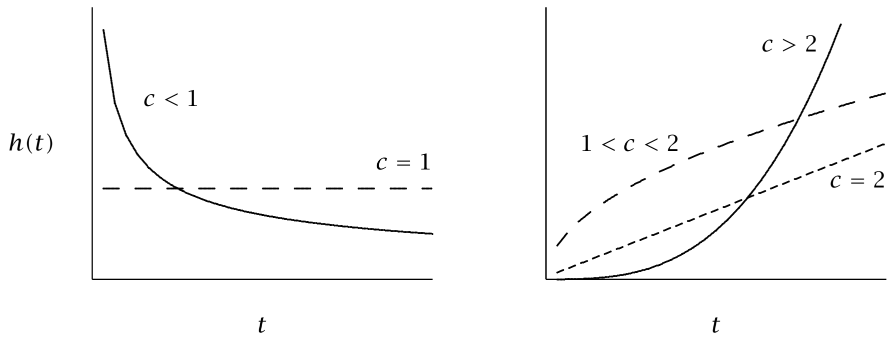
- A generalization of the exponential distribution that can represent decreasing or increasing hazard rates

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

  where c is the "shape" parameter and λ is the "scale" parameter

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

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 $c > 1$

The Weibull-Gamma Model

- Assuming λ is distributed across the population according to a gamma distribution, we have

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

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

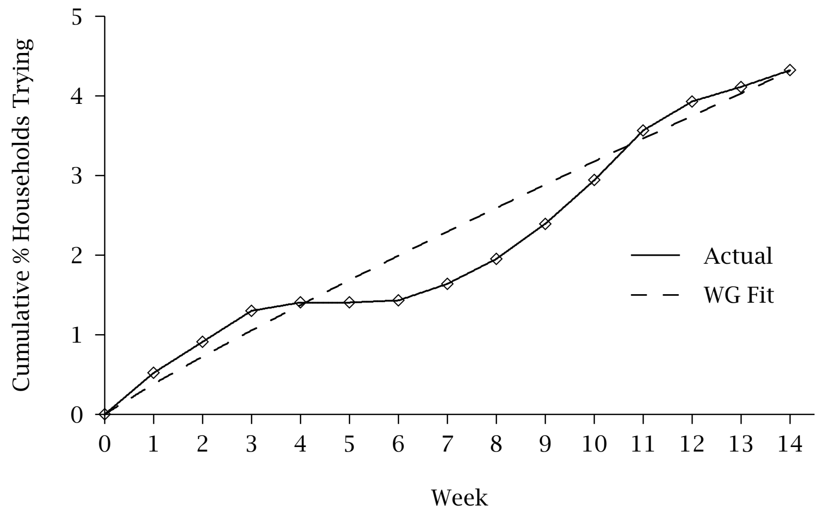
Problem 5 -- WG model

	A	B	C	D	E	F	G
1	Video-on-Demand Service				r	1254.68	
2	# Connected Households		3841		alpha	324084.72	
3					c	0.922	
4					LL =	=SUM(F7:F21)	
5		Cum_Trl					
6	Week	# HHs	Incr_Trl		P(T <= t)		E[T(t)]
7	1	20	=B7		=1-(F$2/(F$2+A7^F$3))^F$1	=C7*LN(E7)	=C$2*E7
8	2	35	=B8-B7		=1-(F$2/(F$2+A8^F$3))^F$1	=C8*LN(E8-E7)	=C$2*E8
9	3	50	=B9-B8		=1-(F$2/(F$2+A9^F$3))^F$1	=C9*LN(E9-E8)	=C$2*E9
10	4	54	=B10-B9		=1-(F$2/(F$2+A10^F$3))^F$1	=C10*LN(E10-E9)	=C$2*E10
11	5	54	=B11-B10		=1-(F$2/(F$2+A11^F$3))^F$1	=C11*LN(E11-E10)	=C$2*E11
12	6	55	=B12-B11		=1-(F$2/(F$2+A12^F$3))^F$1	=C12*LN(E12-E11)	=C$2*E12
13	7	63	=B13-B12		=1-(F$2/(F$2+A13^F$3))^F$1	=C13*LN(E13-E12)	=C$2*E13
14	8	75	=B14-B13		=1-(F$2/(F$2+A14^F$3))^F$1	=C14*LN(E14-E13)	=C$2*E14
15	9	92	=B15-B14		=1-(F$2/(F$2+A15^F$3))^F$1	=C15*LN(E15-E14)	=C$2*E15
16	10	113	=B16-B15		=1-(F$2/(F$2+A16^F$3))^F$1	=C16*LN(E16-E15)	=C$2*E16
17	11	137	=B17-B16		=1-(F$2/(F$2+A17^F$3))^F$1	=C17*LN(E17-E16)	=C$2*E17
18	12	151	=B18-B17		=1-(F$2/(F$2+A18^F$3))^F$1	=C18*LN(E18-E17)	=C$2*E18
19	13	158	=B19-B18		=1-(F$2/(F$2+A19^F$3))^F$1	=C19*LN(E19-E18)	=C$2*E19
20	14	166	=B20-B19		=1-(F$2/(F$2+A20^F$3))^F$1	=C20*LN(E20-E19)	=C$2*E20
21	15				=1-(F$2/(F$2+A21^F$3))^F$1	=(C2-B20)*LN(1-E20)	=C$2*E21
22	16				=1-(F$2/(F$2+A22^F$3))^F$1		=C$2*E22
23	17				=1-(F$2/(F$2+A23^F$3))^F$1		=C$2*E23
24	18				=1-(F$2/(F$2+A24^F$3))^F$1		=C$2*E24
25	19				=1-(F$2/(F$2+A25^F$3))^F$1		=C$2*E25
26	20				=1-(F$2/(F$2+A26^F$3))^F$1		=C$2*E26
27	21				=1-(F$2/(F$2+A27^F$3))^F$1		=C$2*E27
28	22				=1-(F$2/(F$2+A28^F$3))^F$1		=C$2*E28
29	23				=1-(F$2/(F$2+A29^F$3))^F$1		=C$2*E29
30	24				=1-(F$2/(F$2+A30^F$3))^F$1		=C$2*E30
31	25				=1-(F$2/(F$2+A31^F$3))^F$1		=C$2*E31
32	26				=1-(F$2/(F$2+A32^F$3))^F$1		=C$2*E32
33	27				=1-(F$2/(F$2+A33^F$3))^F$1		=C$2*E33
34	28				=1-(F$2/(F$2+A34^F$3))^F$1		=C$2*E34
35	29				=1-(F$2/(F$2+A35^F$3))^F$1		=C$2*E35
36	30				=1-(F$2/(F$2+A36^F$3))^F$1		=C$2*E36
37	31				=1-(F$2/(F$2+A37^F$3))^F$1		=C$2*E37
38	32				=1-(F$2/(F$2+A38^F$3))^F$1		=C$2*E38
39	33				=1-(F$2/(F$2+A39^F$3))^F$1		=C$2*E39
40	34				=1-(F$2/(F$2+A40^F$3))^F$1		=C$2*E40
41	35				=1-(F$2/(F$2+A41^F$3))^F$1		=C$2*E41
42	36				=1-(F$2/(F$2+A42^F$3))^F$1		=C$2*E42
43	37				=1-(F$2/(F$2+A43^F$3))^F$1		=C$2*E43
44	38				=1-(F$2/(F$2+A44^F$3))^F$1		=C$2*E44
45	39				=1-(F$2/(F$2+A45^F$3))^F$1		=C$2*E45
46	40				=1-(F$2/(F$2+A46^F$3))^F$1		=C$2*E46
47	41				=1-(F$2/(F$2+A47^F$3))^F$1		=C$2*E47
48	42				=1-(F$2/(F$2+A48^F$3))^F$1		=C$2*E48
49	43				=1-(F$2/(F$2+A49^F$3))^F$1		=C$2*E49
50	44				=1-(F$2/(F$2+A50^F$3))^F$1		=C$2*E50
51	45				=1-(F$2/(F$2+A51^F$3))^F$1		=C$2*E51
52	46				=1-(F$2/(F$2+A52^F$3))^F$1		=C$2*E52
53	47				=1-(F$2/(F$2+A53^F$3))^F$1		=C$2*E53
54	48				=1-(F$2/(F$2+A54^F$3))^F$1		=C$2*E54
55	49				=1-(F$2/(F$2+A55^F$3))^F$1		=C$2*E55
56	50				=1-(F$2/(F$2+A56^F$3))^F$1		=C$2*E56
57	51				=1-(F$2/(F$2+A57^F$3))^F$1		=C$2*E57
58	52				=1-(F$2/(F$2+A58^F$3))^F$1		=C$2*E58

Problem 5 -- WG model

	A	B	C	D	E	F	G
1	Video-on-Demand Service				r	1254.68	
2	# Connected Households		3841		alpha	324084.72	
3					c	0.9220	
4					LL =	-1121.52	
5		Cum_Trl					
6	Week	# HHs	Incr_Trl		P(T <= t)		E[T(t)]
7	1	20	20		0.00386	-111.12	14.84
8	2	35	15		0.00731	-85.06	28.07
9	3	50	15		0.01060	-85.73	40.73
10	4	54	4		0.01380	-22.98	53.02
11	5	54	0		0.01693	0.00	65.02
12	6	55	1		0.02000	-5.79	76.81
13	7	63	8		0.02302	-46.42	88.40
14	8	75	12		0.02599	-69.81	99.83
15	9	92	17		0.02893	-99.11	111.12
16	10	113	21		0.03183	-122.68	122.27
17	11	137	24		0.03471	-140.46	133.31
18	12	151	14		0.03755	-82.08	144.23
19	13	158	7		0.04037	-41.10	155.05
20	14	166	8		0.04316	-47.05	165.77
21	15				0.04593	-162.13	176.41
22	16				0.04867		186.96
23	17				0.05140		197.42
24	18				0.05410		207.81
25	19				0.05679		218.13
26	20				0.05946		228.37
27	21				0.06210		238.55
28	22				0.06474		248.65
29	23				0.06735		258.70
30	24				0.06995		268.68
31	25				0.07253		278.60
32	26				0.07510		288.47
33	27				0.07766		298.27
34	28				0.08019		308.03
35	29				0.08272		317.72
36	30				0.08523		327.37
37	31				0.08773		336.96
38	32				0.09021		346.51
39	33				0.09268		356.00
40	34				0.09514		365.45
41	35				0.09759		374.85
42	36				0.10003		384.20
43	37				0.10245		393.51
44	38				0.10486		402.77
45	39				0.10726		411.99
46	40				0.10965		421.17
47	41				0.11203		430.30
48	42				0.11440		439.39
49	43				0.11675		448.44
50	44				0.11910		457.45
51	45				0.12143		466.42
52	46				0.12376		475.35
53	47				0.12607		484.25
54	48				0.12838		493.10
55	49				0.13067		501.92
56	50				0.13296		510.70
57	51				0.13524		519.44
58	52				0.13750		528.15

Applying the WG Model



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

The Latent-Class Weibull Model

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

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

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
λ	0.0063	1.51E-06
c	0.4846	5.6915
π	0.9789	

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

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

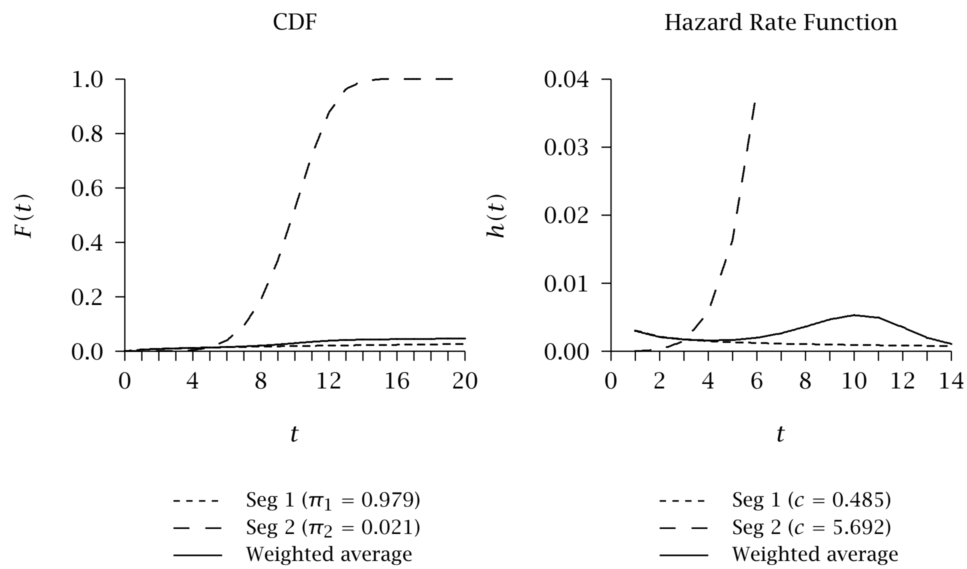
Problem 5 -- 2 segment Weibull

	A	B	C	D	E	F	G	H	I	J
1	Video-on-Demand Service				lambda	0.0063	1.51E-06			
2	# Connected Households		3841		c	0.4846	5.6915			
3					pi	0.9789				
4					LL =	=SUM(I8:I22)				
5										
6		Cum_Trl					P(T<=t)			
7	Week	# HHs	Incr_Trl			Seg 1	Seg 2	Overall		E[T(t)]
8	1	20	=B8			=1-EXP(-F$1*($A8^F$2))	=1-EXP(-G$1*($A8^G$2))	=F$3*F8+(1-F$3)*G8	=C8*LN(H8)	=C$2*H8
9	2	35	=B9-B8			=1-EXP(-F$1*($A9^F$2))	=1-EXP(-G$1*($A9^G$2))	=F$3*F9+(1-F$3)*G9	=C9*LN(H9-H8)	=C$2*H9
10	3	50	=B10-B9			=1-EXP(-F$1*($A10^F$2))	=1-EXP(-G$1*($A10^G$2))	=F$3*F10+(1-F$3)*G10	=C10*LN(H10-H9)	=C$2*H10
11	4	54	=B11-B10			=1-EXP(-F$1*($A11^F$2))	=1-EXP(-G$1*($A11^G$2))	=F$3*F11+(1-F$3)*G11	=C11*LN(H11-H10)	=C$2*H11
12	5	54	=B12-B11			=1-EXP(-F$1*($A12^F$2))	=1-EXP(-G$1*($A12^G$2))	=F$3*F12+(1-F$3)*G12	=C12*LN(H12-H11)	=C$2*H12
13	6	55	=B13-B12			=1-EXP(-F$1*($A13^F$2))	=1-EXP(-G$1*($A13^G$2))	=F$3*F13+(1-F$3)*G13	=C13*LN(H13-H12)	=C$2*H13
14	7	63	=B14-B13			=1-EXP(-F$1*($A14^F$2))	=1-EXP(-G$1*($A14^G$2))	=F$3*F14+(1-F$3)*G14	=C14*LN(H14-H13)	=C$2*H14
15	8	75	=B15-B14			=1-EXP(-F$1*($A15^F$2))	=1-EXP(-G$1*($A15^G$2))	=F$3*F15+(1-F$3)*G15	=C15*LN(H15-H14)	=C$2*H15
16	9	92	=B16-B15			=1-EXP(-F$1*($A16^F$2))	=1-EXP(-G$1*($A16^G$2))	=F$3*F16+(1-F$3)*G16	=C16*LN(H16-H15)	=C$2*H16
17	10	113	=B17-B16			=1-EXP(-F$1*($A17^F$2))	=1-EXP(-G$1*($A17^G$2))	=F$3*F17+(1-F$3)*G17	=C17*LN(H17-H16)	=C$2*H17
18	11	137	=B18-B17			=1-EXP(-F$1*($A18^F$2))	=1-EXP(-G$1*($A18^G$2))	=F$3*F18+(1-F$3)*G18	=C18*LN(H18-H17)	=C$2*H18
19	12	151	=B19-B18			=1-EXP(-F$1*($A19^F$2))	=1-EXP(-G$1*($A19^G$2))	=F$3*F19+(1-F$3)*G19	=C19*LN(H19-H18)	=C$2*H19
20	13	158	=B20-B19			=1-EXP(-F$1*($A20^F$2))	=1-EXP(-G$1*($A20^G$2))	=F$3*F20+(1-F$3)*G20	=C20*LN(H20-H19)	=C$2*H20
21	14	166	=B21-B20			=1-EXP(-F$1*($A21^F$2))	=1-EXP(-G$1*($A21^G$2))	=F$3*F21+(1-F$3)*G21	=C21*LN(H21-H20)	=C$2*H21
22	15					=1-EXP(-F$1*($A22^F$2))	=1-EXP(-G$1*($A22^G$2))	=F$3*F22+(1-F$3)*G22	=(C2-B21)*LN(1-H21)	=C$2*H22
23	16					=1-EXP(-F$1*($A23^F$2))	=1-EXP(-G$1*($A23^G$2))	=F$3*F23+(1-F$3)*G23		=C$2*H23
24	17					=1-EXP(-F$1*($A24^F$2))	=1-EXP(-G$1*($A24^G$2))	=F$3*F24+(1-F$3)*G24		=C$2*H24
25	18					=1-EXP(-F$1*($A25^F$2))	=1-EXP(-G$1*($A25^G$2))	=F$3*F25+(1-F$3)*G25		=C$2*H25
26	19					=1-EXP(-F$1*($A26^F$2))	=1-EXP(-G$1*($A26^G$2))	=F$3*F26+(1-F$3)*G26		=C$2*H26
27	20					=1-EXP(-F$1*($A27^F$2))	=1-EXP(-G$1*($A27^G$2))	=F$3*F27+(1-F$3)*G27		=C$2*H27
28	21					=1-EXP(-F$1*($A28^F$2))	=1-EXP(-G$1*($A28^G$2))	=F$3*F28+(1-F$3)*G28		=C$2*H28
29	22					=1-EXP(-F$1*($A29^F$2))	=1-EXP(-G$1*($A29^G$2))	=F$3*F29+(1-F$3)*G29		=C$2*H29
30	23					=1-EXP(-F$1*($A30^F$2))	=1-EXP(-G$1*($A30^G$2))	=F$3*F30+(1-F$3)*G30		=C$2*H30
31	24					=1-EXP(-F$1*($A31^F$2))	=1-EXP(-G$1*($A31^G$2))	=F$3*F31+(1-F$3)*G31		=C$2*H31
32	25					=1-EXP(-F$1*($A32^F$2))	=1-EXP(-G$1*($A32^G$2))	=F$3*F32+(1-F$3)*G32		=C$2*H32
33	26					=1-EXP(-F$1*($A33^F$2))	=1-EXP(-G$1*($A33^G$2))	=F$3*F33+(1-F$3)*G33		=C$2*H33
34	27					=1-EXP(-F$1*($A34^F$2))	=1-EXP(-G$1*($A34^G$2))	=F$3*F34+(1-F$3)*G34		=C$2*H34
35	28					=1-EXP(-F$1*($A35^F$2))	=1-EXP(-G$1*($A35^G$2))	=F$3*F35+(1-F$3)*G35		=C$2*H35
36	29					=1-EXP(-F$1*($A36^F$2))	=1-EXP(-G$1*($A36^G$2))	=F$3*F36+(1-F$3)*G36		=C$2*H36
37	30					=1-EXP(-F$1*($A37^F$2))	=1-EXP(-G$1*($A37^G$2))	=F$3*F37+(1-F$3)*G37		=C$2*H37
38	31					=1-EXP(-F$1*($A38^F$2))	=1-EXP(-G$1*($A38^G$2))	=F$3*F38+(1-F$3)*G38		=C$2*H38
39	32					=1-EXP(-F$1*($A39^F$2))	=1-EXP(-G$1*($A39^G$2))	=F$3*F39+(1-F$3)*G39		=C$2*H39
40	33					=1-EXP(-F$1*($A40^F$2))	=1-EXP(-G$1*($A40^G$2))	=F$3*F40+(1-F$3)*G40		=C$2*H40
41	34					=1-EXP(-F$1*($A41^F$2))	=1-EXP(-G$1*($A41^G$2))	=F$3*F41+(1-F$3)*G41		=C$2*H41
42	35					=1-EXP(-F$1*($A42^F$2))	=1-EXP(-G$1*($A42^G$2))	=F$3*F42+(1-F$3)*G42		=C$2*H42
43	36					=1-EXP(-F$1*($A43^F$2))	=1-EXP(-G$1*($A43^G$2))	=F$3*F43+(1-F$3)*G43		=C$2*H43
44	37					=1-EXP(-F$1*($A44^F$2))	=1-EXP(-G$1*($A44^G$2))	=F$3*F44+(1-F$3)*G44		=C$2*H44
45	38					=1-EXP(-F$1*($A45^F$2))	=1-EXP(-G$1*($A45^G$2))	=F$3*F45+(1-F$3)*G45		=C$2*H45
46	39					=1-EXP(-F$1*($A46^F$2))	=1-EXP(-G$1*($A46^G$2))	=F$3*F46+(1-F$3)*G46		=C$2*H46
47	40					=1-EXP(-F$1*($A47^F$2))	=1-EXP(-G$1*($A47^G$2))	=F$3*F47+(1-F$3)*G47		=C$2*H47
48	41					=1-EXP(-F$1*($A48^F$2))	=1-EXP(-G$1*($A48^G$2))	=F$3*F48+(1-F$3)*G48		=C$2*H48
49	42					=1-EXP(-F$1*($A49^F$2))	=1-EXP(-G$1*($A49^G$2))	=F$3*F49+(1-F$3)*G49		=C$2*H49
50	43					=1-EXP(-F$1*($A50^F$2))	=1-EXP(-G$1*($A50^G$2))	=F$3*F50+(1-F$3)*G50		=C$2*H50
51	44					=1-EXP(-F$1*($A51^F$2))	=1-EXP(-G$1*($A51^G$2))	=F$3*F51+(1-F$3)*G51		=C$2*H51
52	45					=1-EXP(-F$1*($A52^F$2))	=1-EXP(-G$1*($A52^G$2))	=F$3*F52+(1-F$3)*G52		=C$2*H52
53	46					=1-EXP(-F$1*($A53^F$2))	=1-EXP(-G$1*($A53^G$2))	=F$3*F53+(1-F$3)*G53		=C$2*H53
54	47					=1-EXP(-F$1*($A54^F$2))	=1-EXP(-G$1*($A54^G$2))	=F$3*F54+(1-F$3)*G54		=C$2*H54
55	48					=1-EXP(-F$1*($A55^F$2))	=1-EXP(-G$1*($A55^G$2))	=F$3*F55+(1-F$3)*G55		=C$2*H55
56	49					=1-EXP(-F$1*($A56^F$2))	=1-EXP(-G$1*($A56^G$2))	=F$3*F56+(1-F$3)*G56		=C$2*H56
57	50					=1-EXP(-F$1*($A57^F$2))	=1-EXP(-G$1*($A57^G$2))	=F$3*F57+(1-F$3)*G57		=C$2*H57
58	51					=1-EXP(-F$1*($A58^F$2))	=1-EXP(-G$1*($A58^G$2))	=F$3*F58+(1-F$3)*G58		=C$2*H58
59	52					=1-EXP(-F$1*($A59^F$2))	=1-EXP(-G$1*($A59^G$2))	=F$3*F59+(1-F$3)*G59		=C$2*H59

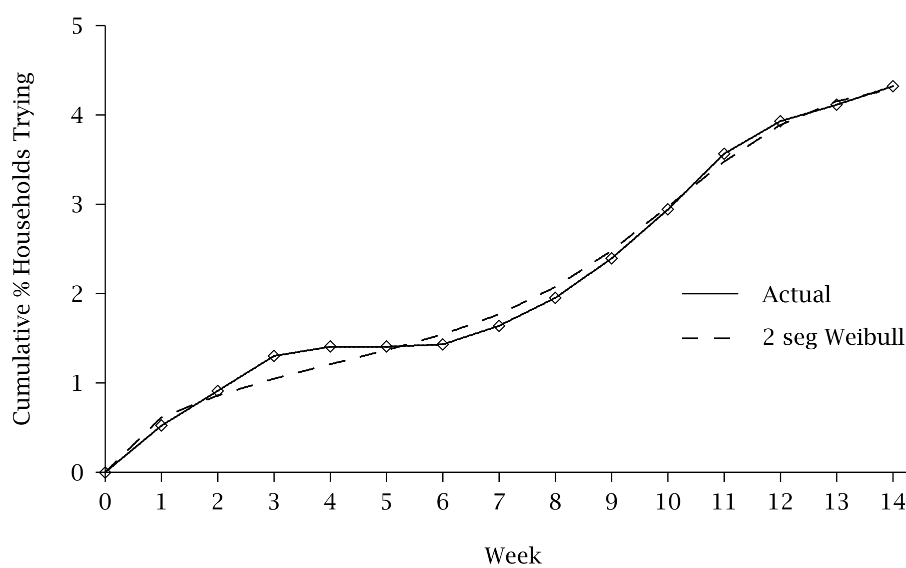
Problem 5 -- 2 segment Weibull

	A	B	C	D	E	F	G	H	I	J
1	Video-on-Demand Service				lambda	0.0063	1.51E-06			
2	# Connected Households		3841		c	0.4846	5.6915			
3					pi	0.9789				
4					LL =	-1100.91				
5										
6		Cum_Trl					P(T<=t)			
7	Week	# HHs	Incr_Trl			Seg 1	Seg 2	Overall		E[T(t)]
8	1	20	20			0.00628	0.00000	0.00615	-101.82	23.63
9	2	35	15			0.00878	0.00008	0.00860	-90.20	33.02
10	3	50	15			0.01068	0.00078	0.01047	-94.22	40.21
11	4	54	4			0.01226	0.00402	0.01209	-25.70	46.44
12	5	54	0			0.01366	0.01423	0.01367	0.00	52.50
13	6	55	1			0.01491	0.03966	0.01543	-6.34	59.27
14	7	63	8			0.01605	0.09272	0.01768	-48.79	67.89
15	8	75	12			0.01712	0.18784	0.02073	-69.50	79.62
16	9	92	17			0.01812	0.33419	0.02480	-93.57	95.25
17	10	113	21			0.01906	0.52332	0.02972	-111.61	114.15
18	11	137	24			0.01995	0.72044	0.03476	-126.96	133.51
19	12	151	14			0.02080	0.87648	0.03889	-76.85	149.38
20	13	158	7			0.02161	0.96305	0.04152	-41.59	159.47
21	14	166	8			0.02239	0.99345	0.04293	-52.53	164.88
22	15					0.02314	0.99942	0.04379	-161.24	168.19
23	16					0.02387	0.99998	0.04451		170.97
24	17					0.02457	1.00000	0.04520		173.62
25	18					0.02525	1.00000	0.04587		176.18
26	19					0.02592	1.00000	0.04652		178.67
27	20					0.02656	1.00000	0.04715		181.08
28	21					0.02719	1.00000	0.04776		183.44
29	22					0.02780	1.00000	0.04836		185.74
30	23					0.02839	1.00000	0.04894		187.98
31	24					0.02898	1.00000	0.04951		190.17
32	25					0.02955	1.00000	0.05007		192.32
33	26					0.03011	1.00000	0.05062		194.42
34	27					0.03065	1.00000	0.05115		196.47
35	28					0.03119	1.00000	0.05168		198.49
36	29					0.03172	1.00000	0.05219		200.47
37	30					0.03223	1.00000	0.05270		202.41
38	31					0.03274	1.00000	0.05319		204.32
39	32					0.03324	1.00000	0.05368		206.20
40	33					0.03373	1.00000	0.05416		208.04
41	34					0.03421	1.00000	0.05464		209.86
42	35					0.03469	1.00000	0.05510		211.65
43	36					0.03516	1.00000	0.05556		213.41
44	37					0.03562	1.00000	0.05601		215.14
45	38					0.03607	1.00000	0.05646		216.85
46	39					0.03652	1.00000	0.05690		218.54
47	40					0.03696	1.00000	0.05733		220.20
48	41					0.03740	1.00000	0.05776		221.84
49	42					0.03783	1.00000	0.05818		223.46
50	43					0.03826	1.00000	0.05860		225.06
51	44					0.03868	1.00000	0.05901		226.64
52	45					0.03909	1.00000	0.05941		228.20
53	46					0.03950	1.00000	0.05981		229.75
54	47					0.03991	1.00000	0.06021		231.27
55	48					0.04031	1.00000	0.06060		232.78
56	49					0.04071	1.00000	0.06099		234.27
57	50					0.04110	1.00000	0.06138		235.74
58	51					0.04149	1.00000	0.06176		237.20
59	52					0.04187	1.00000	0.06213		238.65

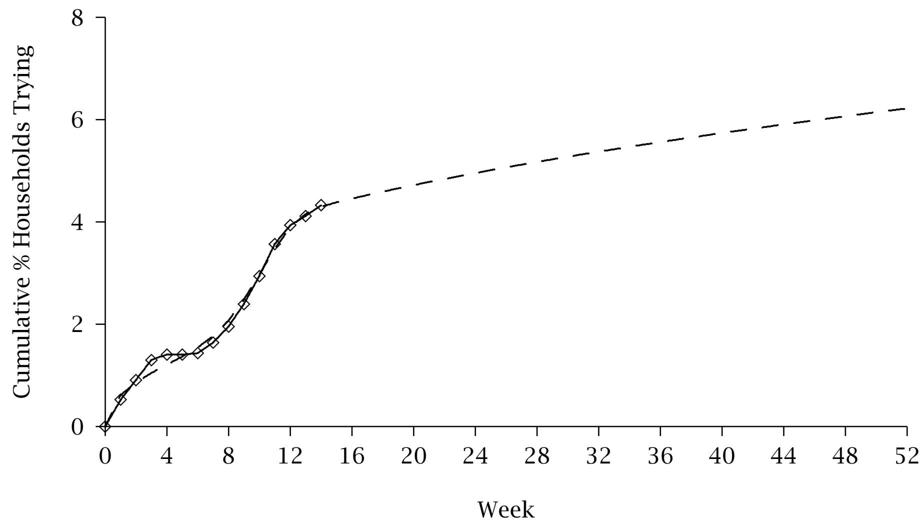
CDF and Hazard Rate Function Plots



Model Fit



Cumulative Trial Forecast



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 λ_s and c_s
- λ_s is distributed across the segment members according to a gamma distribution with parameters r_s and α_s

$\longrightarrow$ a finite mixture of Weibull-gamma models:

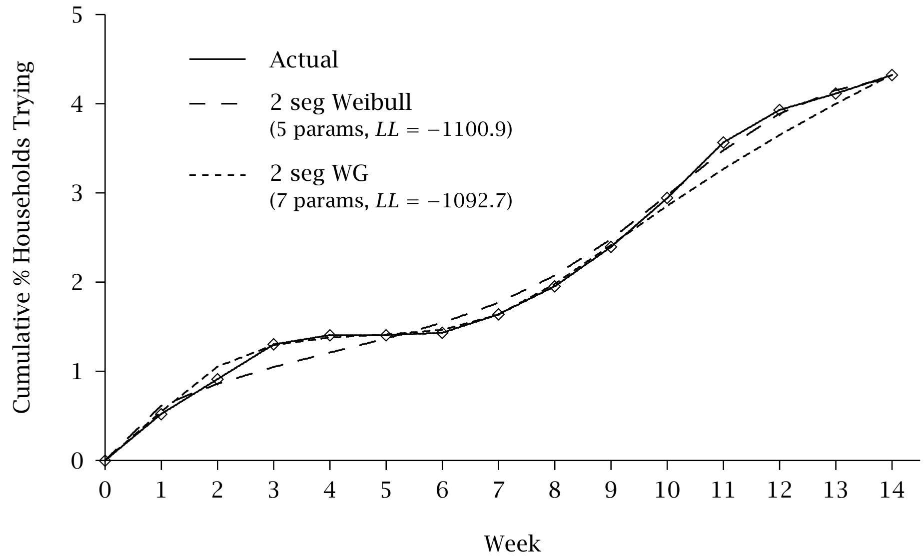
$$F(t) = \pi_1\left[1 - \left(\frac{\alpha_1}{\alpha_1 + t^{c_1}}\right)^{r_1}\right] + (1 - \pi_1)\left[1 - \left(\frac{\alpha_2}{\alpha_2 + t^{c_2}}\right)^{r_2}\right]$$

	A	B	C	D	E	F	G	H	I	J
1	Video-on-Demand Service				r	17514623	0.0044			
2	# Connected Households		3841		alpha	35549899	7.99E+08			
3					c	1.4783	10.3366			
4					pi	0.0141				
5					LL =	=SUM(I9:I23)				
6										
7		Cum_Trl				P(T<=t)				
8	Week	# HHs	Incr_Trl			Seg 1	Seg 2	Overall		E[T(t)]
9	1	20	=B9			=1-(F$2/(F$2+$A9^F$3))^F$1	=1-(G$2/(G$2+$A9^G$3))^G$1	=F$4*F9+(1-F$4)*G9	=C9*LN(H9)	=C$2*H9
10	2	35	=B10-B9			=1-(F$2/(F$2+$A10^F$3))^F$1	=1-(G$2/(G$2+$A10^G$3))^G$1	=F$4*F10+(1-F$4)*G10	=C10*LN(H10-H9)	=C$2*H10
11	3	50	=B11-B10			=1-(F$2/(F$2+$A11^F$3))^F$1	=1-(G$2/(G$2+$A11^G$3))^G$1	=F$4*F11+(1-F$4)*G11	=C11*LN(H11-H10)	=C$2*H11
12	4	54	=B12-B11			=1-(F$2/(F$2+$A12^F$3))^F$1	=1-(G$2/(G$2+$A12^G$3))^G$1	=F$4*F12+(1-F$4)*G12	=C12*LN(H12-H11)	=C$2*H12
13	5	54	=B13-B12			=1-(F$2/(F$2+$A13^F$3))^F$1	=1-(G$2/(G$2+$A13^G$3))^G$1	=F$4*F13+(1-F$4)*G13	=C13*LN(H13-H12)	=C$2*H13
14	6	55	=B14-B13			=1-(F$2/(F$2+$A14^F$3))^F$1	=1-(G$2/(G$2+$A14^G$3))^G$1	=F$4*F14+(1-F$4)*G14	=C14*LN(H14-H13)	=C$2*H14
15	7	63	=B15-B14			=1-(F$2/(F$2+$A15^F$3))^F$1	=1-(G$2/(G$2+$A15^G$3))^G$1	=F$4*F15+(1-F$4)*G15	=C15*LN(H15-H14)	=C$2*H15
16	8	75	=B16-B15			=1-(F$2/(F$2+$A16^F$3))^F$1	=1-(G$2/(G$2+$A16^G$3))^G$1	=F$4*F16+(1-F$4)*G16	=C16*LN(H16-H15)	=C$2*H16
17	9	92	=B17-B16			=1-(F$2/(F$2+$A17^F$3))^F$1	=1-(G$2/(G$2+$A17^G$3))^G$1	=F$4*F17+(1-F$4)*G17	=C17*LN(H17-H16)	=C$2*H17
18	10	113	=B18-B17			=1-(F$2/(F$2+$A18^F$3))^F$1	=1-(G$2/(G$2+$A18^G$3))^G$1	=F$4*F18+(1-F$4)*G18	=C18*LN(H18-H17)	=C$2*H18
19	11	137	=B19-B18			=1-(F$2/(F$2+$A19^F$3))^F$1	=1-(G$2/(G$2+$A19^G$3))^G$1	=F$4*F19+(1-F$4)*G19	=C19*LN(H19-H18)	=C$2*H19
20	12	151	=B20-B19			=1-(F$2/(F$2+$A20^F$3))^F$1	=1-(G$2/(G$2+$A20^G$3))^G$1	=F$4*F20+(1-F$4)*G20	=C20*LN(H20-H19)	=C$2*H20
21	13	158	=B21-B20			=1-(F$2/(F$2+$A21^F$3))^F$1	=1-(G$2/(G$2+$A21^G$3))^G$1	=F$4*F21+(1-F$4)*G21	=C21*LN(H21-H20)	=C$2*H21
22	14	166	=B22-B21			=1-(F$2/(F$2+$A22^F$3))^F$1	=1-(G$2/(G$2+$A22^G$3))^G$1	=F$4*F22+(1-F$4)*G22	=C22*LN(H22-H21)	=C$2*H22
23	15					=1-(F$2/(F$2+$A23^F$3))^F$1	=1-(G$2/(G$2+$A23^G$3))^G$1	=F$4*F23+(1-F$4)*G23	=(C2-B22)*LN(1-H22)	=C$2*H23
24	16					=1-(F$2/(F$2+$A24^F$3))^F$1	=1-(G$2/(G$2+$A24^G$3))^G$1	=F$4*F24+(1-F$4)*G24		=C$2*H24
25	17					=1-(F$2/(F$2+$A25^F$3))^F$1	=1-(G$2/(G$2+$A25^G$3))^G$1	=F$4*F25+(1-F$4)*G25		=C$2*H25
26	18					=1-(F$2/(F$2+$A26^F$3))^F$1	=1-(G$2/(G$2+$A26^G$3))^G$1	=F$4*F26+(1-F$4)*G26		=C$2*H26
27	19					=1-(F$2/(F$2+$A27^F$3))^F$1	=1-(G$2/(G$2+$A27^G$3))^G$1	=F$4*F27+(1-F$4)*G27		=C$2*H27
28	20					=1-(F$2/(F$2+$A28^F$3))^F$1	=1-(G$2/(G$2+$A28^G$3))^G$1	=F$4*F28+(1-F$4)*G28		=C$2*H28
29	21					=1-(F$2/(F$2+$A29^F$3))^F$1	=1-(G$2/(G$2+$A29^G$3))^G$1	=F$4*F29+(1-F$4)*G29		=C$2*H29
30	22					=1-(F$2/(F$2+$A30^F$3))^F$1	=1-(G$2/(G$2+$A30^G$3))^G$1	=F$4*F30+(1-F$4)*G30		=C$2*H30
31	23					=1-(F$2/(F$2+$A31^F$3))^F$1	=1-(G$2/(G$2+$A31^G$3))^G$1	=F$4*F31+(1-F$4)*G31		=C$2*H31
32	24					=1-(F$2/(F$2+$A32^F$3))^F$1	=1-(G$2/(G$2+$A32^G$3))^G$1	=F$4*F32+(1-F$4)*G32		=C$2*H32
33	25					=1-(F$2/(F$2+$A33^F$3))^F$1	=1-(G$2/(G$2+$A33^G$3))^G$1	=F$4*F33+(1-F$4)*G33		=C$2*H33
34	26					=1-(F$2/(F$2+$A34^F$3))^F$1	=1-(G$2/(G$2+$A34^G$3))^G$1	=F$4*F34+(1-F$4)*G34		=C$2*H34
35	27					=1-(F$2/(F$2+$A35^F$3))^F$1	=1-(G$2/(G$2+$A35^G$3))^G$1	=F$4*F35+(1-F$4)*G35		=C$2*H35
36	28					=1-(F$2/(F$2+$A36^F$3))^F$1	=1-(G$2/(G$2+$A36^G$3))^G$1	=F$4*F36+(1-F$4)*G36		=C$2*H36
37	29					=1-(F$2/(F$2+$A37^F$3))^F$1	=1-(G$2/(G$2+$A37^G$3))^G$1	=F$4*F37+(1-F$4)*G37		=C$2*H37
38	30					=1-(F$2/(F$2+$A38^F$3))^F$1	=1-(G$2/(G$2+$A38^G$3))^G$1	=F$4*F38+(1-F$4)*G38		=C$2*H38
39	31					=1-(F$2/(F$2+$A39^F$3))^F$1	=1-(G$2/(G$2+$A39^G$3))^G$1	=F$4*F39+(1-F$4)*G39		=C$2*H39
40	32					=1-(F$2/(F$2+$A40^F$3))^F$1	=1-(G$2/(G$2+$A40^G$3))^G$1	=F$4*F40+(1-F$4)*G40		=C$2*H40
41	33					=1-(F$2/(F$2+$A41^F$3))^F$1	=1-(G$2/(G$2+$A41^G$3))^G$1	=F$4*F41+(1-F$4)*G41		=C$2*H41
42	34					=1-(F$2/(F$2+$A42^F$3))^F$1	=1-(G$2/(G$2+$A42^G$3))^G$1	=F$4*F42+(1-F$4)*G42		=C$2*H42
43	35					=1-(F$2/(F$2+$A43^F$3))^F$1	=1-(G$2/(G$2+$A43^G$3))^G$1	=F$4*F43+(1-F$4)*G43		=C$2*H43
44	36					=1-(F$2/(F$2+$A44^F$3))^F$1	=1-(G$2/(G$2+$A44^G$3))^G$1	=F$4*F44+(1-F$4)*G44		=C$2*H44
45	37					=1-(F$2/(F$2+$A45^F$3))^F$1	=1-(G$2/(G$2+$A45^G$3))^G$1	=F$4*F45+(1-F$4)*G45		=C$2*H45
46	38					=1-(F$2/(F$2+$A46^F$3))^F$1	=1-(G$2/(G$2+$A46^G$3))^G$1	=F$4*F46+(1-F$4)*G46		=C$2*H46
47	39					=1-(F$2/(F$2+$A47^F$3))^F$1	=1-(G$2/(G$2+$A47^G$3))^G$1	=F$4*F47+(1-F$4)*G47		=C$2*H47
48	40					=1-(F$2/(F$2+$A48^F$3))^F$1	=1-(G$2/(G$2+$A48^G$3))^G$1	=F$4*F48+(1-F$4)*G48		=C$2*H48
49	41					=1-(F$2/(F$2+$A49^F$3))^F$1	=1-(G$2/(G$2+$A49^G$3))^G$1	=F$4*F49+(1-F$4)*G49		=C$2*H49
50	42					=1-(F$2/(F$2+$A50^F$3))^F$1	=1-(G$2/(G$2+$A50^G$3))^G$1	=F$4*F50+(1-F$4)*G50		=C$2*H50
51	43					=1-(F$2/(F$2+$A51^F$3))^F$1	=1-(G$2/(G$2+$A51^G$3))^G$1	=F$4*F51+(1-F$4)*G51		=C$2*H51
52	44					=1-(F$2/(F$2+$A52^F$3))^F$1	=1-(G$2/(G$2+$A52^G$3))^G$1	=F$4*F52+(1-F$4)*G52		=C$2*H52
53	45					=1-(F$2/(F$2+$A53^F$3))^F$1	=1-(G$2/(G$2+$A53^G$3))^G$1	=F$4*F53+(1-F$4)*G53		=C$2*H53
54	46					=1-(F$2/(F$2+$A54^F$3))^F$1	=1-(G$2/(G$2+$A54^G$3))^G$1	=F$4*F54+(1-F$4)*G54		=C$2*H54
55	47					=1-(F$2/(F$2+$A55^F$3))^F$1	=1-(G$2/(G$2+$A55^G$3))^G$1	=F$4*F55+(1-F$4)*G55		=C$2*H55
56	48					=1-(F$2/(F$2+$A56^F$3))^F$1	=1-(G$2/(G$2+$A56^G$3))^G$1	=F$4*F56+(1-F$4)*G56		=C$2*H56
57	49					=1-(F$2/(F$2+$A57^F$3))^F$1	=1-(G$2/(G$2+$A57^G$3))^G$1	=F$4*F57+(1-F$4)*G57		=C$2*H57
58	50					=1-(F$2/(F$2+$A58^F$3))^F$1	=1-(G$2/(G$2+$A58^G$3))^G$1	=F$4*F58+(1-F$4)*G58		=C$2*H58
59	51					=1-(F$2/(F$2+$A59^F$3))^F$1	=1-(G$2/(G$2+$A59^G$3))^G$1	=F$4*F59+(1-F$4)*G59		=C$2*H59
60	52					=1-(F$2/(F$2+$A60^F$3))^F$1	=1-(G$2/(G$2+$A60^G$3))^G$1	=F$4*F60+(1-F$4)*G60		=C$2*H60

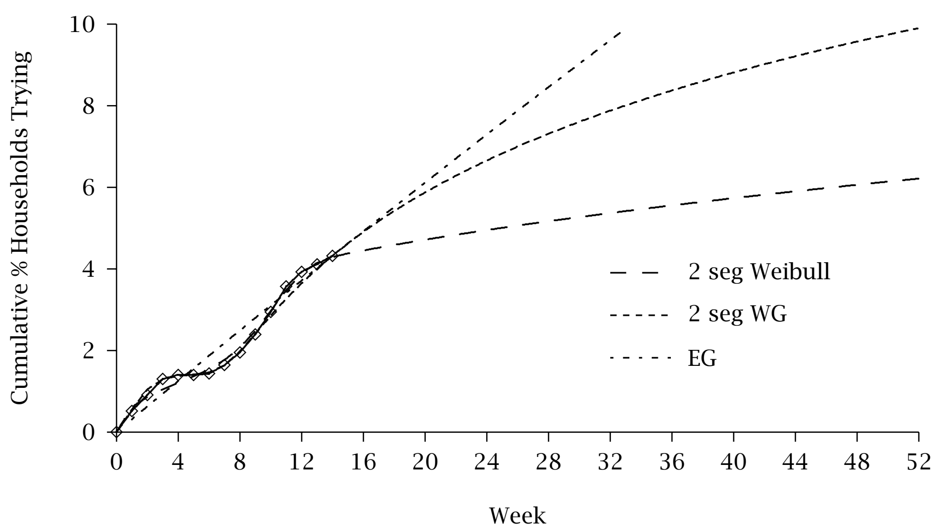
Problem 5 -- 2 segment WG

	A	B	C	D	E	F	G	H	I	J
1	Video-on-Demand Service				r	17514623	0.0044			
2	# Connected Households		3841		alpha	35549899	7.99E+08			
3					c	1.4783	10.3366			
4					pi	0.0141				
5					LL =	-1092.66				
6										
7		Cum_Trl					P(T<=t)			
8	Week	# HHs	Incr_Trl			Seg 1	Seg 2	Overall		E[T(t)]
9	1	20	20			0.38901	0.00000	0.00548	-104.13	21.05
10	2	35	15			0.74658	0.00000	0.01052	-79.36	40.40
11	3	50	15			0.91789	0.00000	0.01293	-90.40	49.68
12	4	54	4			0.97817	0.00001	0.01379	-28.24	52.97
13	5	54	0			0.99510	0.00009	0.01411	0.00	54.20
14	6	55	1			0.99906	0.00057	0.01464	-7.54	56.24
15	7	63	8			0.99984	0.00229	0.01635	-50.98	62.80
16	8	75	12			0.99998	0.00578	0.01979	-68.07	76.01
17	9	92	17			1.00000	0.01020	0.02415	-92.40	92.75
18	10	113	21			1.00000	0.01467	0.02855	-113.94	109.66
19	11	137	24			1.00000	0.01886	0.03268	-131.74	125.53
20	12	151	14			1.00000	0.02272	0.03649	-77.98	140.16
21	13	158	7			1.00000	0.02628	0.04000	-39.57	153.64
22	14	166	8			1.00000	0.02957	0.04325	-45.85	166.10
23	15					1.00000	0.03263	0.04626	-162.46	177.68
24	16					1.00000	0.03548	0.04907		188.48
25	17					1.00000	0.03815	0.05170		198.59
26	18					1.00000	0.04066	0.05418		208.11
27	19					1.00000	0.04303	0.05652		217.08
28	20					1.00000	0.04528	0.05873		225.58
29	21					1.00000	0.04741	0.06083		233.64
30	22					1.00000	0.04943	0.06283		241.31
31	23					1.00000	0.05136	0.06473		248.63
32	24					1.00000	0.05321	0.06655		255.62
33	25					1.00000	0.05498	0.06829		262.31
34	26					1.00000	0.05667	0.06996		268.72
35	27					1.00000	0.05830	0.07157		274.89
36	28					1.00000	0.05986	0.07311		280.82
37	29					1.00000	0.06137	0.07460		286.53
38	30					1.00000	0.06283	0.07603		292.04
39	31					1.00000	0.06423	0.07742		297.36
40	32					1.00000	0.06559	0.07876		302.51
41	33					1.00000	0.06691	0.08005		307.48
42	34					1.00000	0.06818	0.08131		312.31
43	35					1.00000	0.06942	0.08253		316.99
44	36					1.00000	0.07061	0.08371		321.53
45	37					1.00000	0.07178	0.08486		325.94
46	38					1.00000	0.07291	0.08597		330.22
47	39					1.00000	0.07401	0.08706		334.39
48	40					1.00000	0.07508	0.08812		338.45
49	41					1.00000	0.07613	0.08915		342.41
50	42					1.00000	0.07715	0.09015		346.26
51	43					1.00000	0.07814	0.09113		350.03
52	44					1.00000	0.07911	0.09208		353.70
53	45					1.00000	0.08006	0.09302		357.28
54	46					1.00000	0.08098	0.09393		360.78
55	47					1.00000	0.08188	0.09482		364.21
56	48					1.00000	0.08277	0.09569		367.55
57	49					1.00000	0.08363	0.09654		370.83
58	50					1.00000	0.08448	0.09738		374.04
59	51					1.00000	0.08531	0.09820		377.18
60	52					1.00000	0.08612	0.09900		380.25

Model Fit



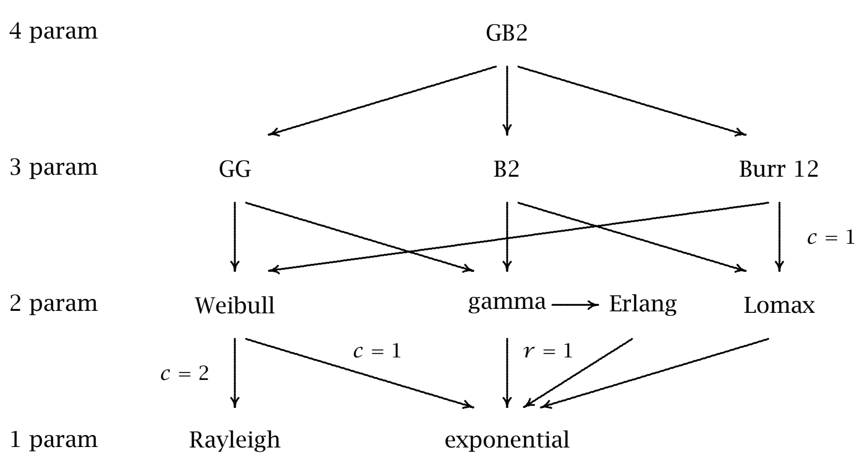
Cumulative Trial Forecast



Alternative Explanations

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

A Family Tree of Distributions



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

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.

Problem 6:
Modeling Repeat Purchase Quantities at CDNOW

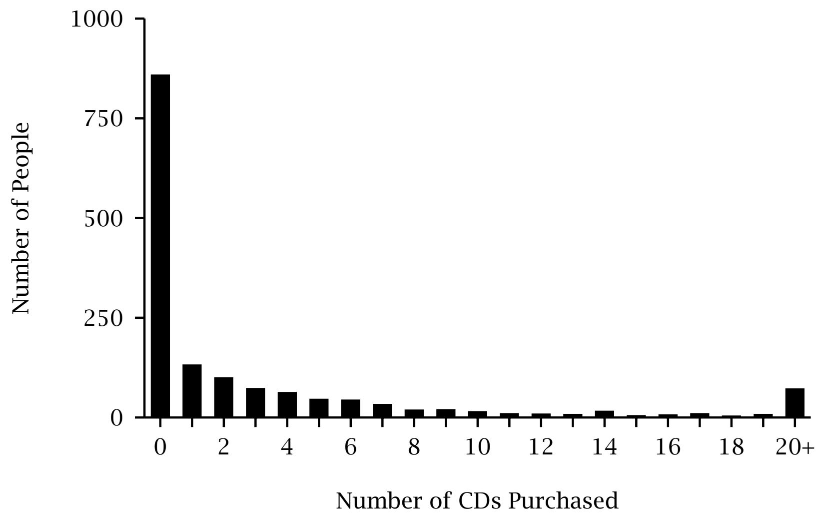
(Building an "Integrated" Model)

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.

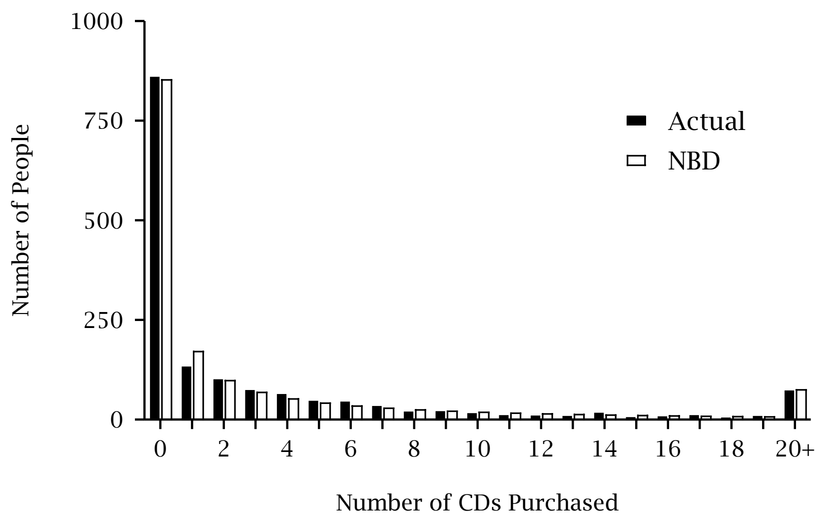
Our objective is to develop a simple model that describes the distribution of repeat purchase quantity (# CDs) for this group of 1574 individuals.

Distribution of Purchase Quantity



71

Fit of NBD



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

Problem 6 -- NBD w spike Estimation

	A	B	C	D	E	F
1	r	0.38416				
2	alpha	0.07772			LL	=SUM(F6:F26)
3	pi	0.28658				
4						
5	Y	P(Y=y)	w/ spike	f_y		
6	0	=(B2/(B2+1))^B1	=B3+(1-B3)*B6	860		=LN(C6)*D6
7	1	=(B$1+A7-1)/(A7*(B$2+1))*B6	=(1-B$3)*B7	133		=LN(C7)*D7
8	2	=(B$1+A8-1)/(A8*(B$2+1))*B7	=(1-B$3)*B8	101		=LN(C8)*D8
9	3	=(B$1+A9-1)/(A9*(B$2+1))*B8	=(1-B$3)*B9	74		=LN(C9)*D9
10	4	=(B$1+A10-1)/(A10*(B$2+1))*B9	=(1-B$3)*B10	64		=LN(C10)*D10
11	5	=(B$1+A11-1)/(A11*(B$2+1))*B10	=(1-B$3)*B11	47		=LN(C11)*D11
12	6	=(B$1+A12-1)/(A12*(B$2+1))*B11	=(1-B$3)*B12	45		=LN(C12)*D12
13	7	=(B$1+A13-1)/(A13*(B$2+1))*B12	=(1-B$3)*B13	34		=LN(C13)*D13
14	8	=(B$1+A14-1)/(A14*(B$2+1))*B13	=(1-B$3)*B14	20		=LN(C14)*D14
15	9	=(B$1+A15-1)/(A15*(B$2+1))*B14	=(1-B$3)*B15	21		=LN(C15)*D15
16	10	=(B$1+A16-1)/(A16*(B$2+1))*B15	=(1-B$3)*B16	16		=LN(C16)*D16
17	11	=(B$1+A17-1)/(A17*(B$2+1))*B16	=(1-B$3)*B17	11		=LN(C17)*D17
18	12	=(B$1+A18-1)/(A18*(B$2+1))*B17	=(1-B$3)*B18	10		=LN(C18)*D18
19	13	=(B$1+A19-1)/(A19*(B$2+1))*B18	=(1-B$3)*B19	9		=LN(C19)*D19
20	14	=(B$1+A20-1)/(A20*(B$2+1))*B19	=(1-B$3)*B20	17		=LN(C20)*D20
21	15	=(B$1+A21-1)/(A21*(B$2+1))*B20	=(1-B$3)*B21	6		=LN(C21)*D21
22	16	=(B$1+A22-1)/(A22*(B$2+1))*B21	=(1-B$3)*B22	8		=LN(C22)*D22
23	17	=(B$1+A23-1)/(A23*(B$2+1))*B22	=(1-B$3)*B23	11		=LN(C23)*D23
24	18	=(B$1+A24-1)/(A24*(B$2+1))*B23	=(1-B$3)*B24	5		=LN(C24)*D24
25	19	=(B$1+A25-1)/(A25*(B$2+1))*B24	=(1-B$3)*B25	9		=LN(C25)*D25
26	20+	=1-SUM(B6:B25)	=1-SUM(C6:C25)	73		=LN(C26)*D26

Problem 6 -- NBD w spike Estimation

	A	B	C	D	E	F
1	r	0.38416				
2	alpha	0.07772			LL	-2933.57
3	pi	0.28658				
4						
5	Y	P(Y=y)	w/ spike	f_y		
6	0	0.36416	0.54638	860		-519.82
7	1	0.12981	0.09261	133		-316.46
8	2	0.08336	0.05947	101		-285.05
9	3	0.06147	0.04385	74		-231.39
10	4	0.04826	0.03443	64		-215.61
11	5	0.03926	0.02801	47		-168.03
12	6	0.03269	0.02332	45		-169.13
13	7	0.02766	0.01974	34		-133.46
14	8	0.02369	0.01690	20		-81.60
15	9	0.02048	0.01461	21		-88.75
16	10	0.01783	0.01272	16		-69.83
17	11	0.01562	0.01114	11		-49.47
18	12	0.01375	0.00981	10		-46.24
19	13	0.01215	0.00867	9		-42.73
20	14	0.01078	0.00769	17		-82.75
21	15	0.00959	0.00684	6		-29.91
22	16	0.00856	0.00611	8		-40.79
23	17	0.00765	0.00546	11		-57.31
24	18	0.00686	0.00489	5		-26.60
25	19	0.00616	0.00439	9		-48.85
26	20+	0.06019	0.04294	73		-229.80

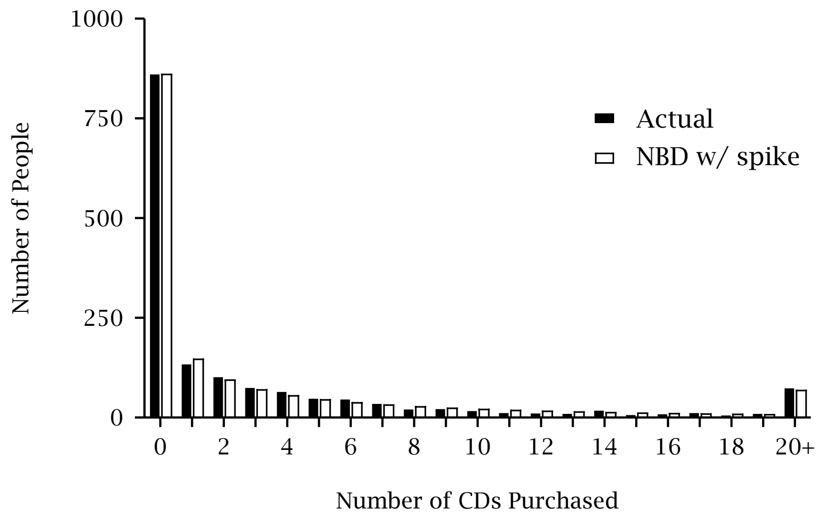
Problem 6 -- 2 seg NBD Estimation

	A	B	C	D	E	F	G
1	Seg 1	r	0.168				
2		alpha	0.045				
3	Seg 2	r	138.136				
4		alpha	32.182				
5		pi_1	0.922			LL	=SUM(G9:G29)
6							
7			P(Y=y)				
8	Y	Seg 1	Seg 2	Total		f_y	
9	0	=(C2/(C2+1))^C1	=(C4/(C4+1))^C3	=B9*C$5+(1-C$5)*C9		860	=LN(D9)*F9
10	1	=(C$1+$A10-1)/($A10*(C$2+1))*B9	=(C$3+$A10-1)/($A10*(C$4+1))*C9	=B10*C$5+(1-C$5)*C10		133	=LN(D10)*F10
11	2	=(C$1+$A11-1)/($A11*(C$2+1))*B10	=(C$3+$A11-1)/($A11*(C$4+1))*C10	=B11*C$5+(1-C$5)*C11		101	=LN(D11)*F11
12	3	=(C$1+$A12-1)/($A12*(C$2+1))*B11	=(C$3+$A12-1)/($A12*(C$4+1))*C11	=B12*C$5+(1-C$5)*C12		74	=LN(D12)*F12
13	4	=(C$1+$A13-1)/($A13*(C$2+1))*B12	=(C$3+$A13-1)/($A13*(C$4+1))*C12	=B13*C$5+(1-C$5)*C13		64	=LN(D13)*F13
14	5	=(C$1+$A14-1)/($A14*(C$2+1))*B13	=(C$3+$A14-1)/($A14*(C$4+1))*C13	=B14*C$5+(1-C$5)*C14		47	=LN(D14)*F14
15	6	=(C$1+$A15-1)/($A15*(C$2+1))*B14	=(C$3+$A15-1)/($A15*(C$4+1))*C14	=B15*C$5+(1-C$5)*C15		45	=LN(D15)*F15
16	7	=(C$1+$A16-1)/($A16*(C$2+1))*B15	=(C$3+$A16-1)/($A16*(C$4+1))*C15	=B16*C$5+(1-C$5)*C16		34	=LN(D16)*F16
17	8	=(C$1+$A17-1)/($A17*(C$2+1))*B16	=(C$3+$A17-1)/($A17*(C$4+1))*C16	=B17*C$5+(1-C$5)*C17		20	=LN(D17)*F17
18	9	=(C$1+$A18-1)/($A18*(C$2+1))*B17	=(C$3+$A18-1)/($A18*(C$4+1))*C17	=B18*C$5+(1-C$5)*C18		21	=LN(D18)*F18
19	10	=(C$1+$A19-1)/($A19*(C$2+1))*B18	=(C$3+$A19-1)/($A19*(C$4+1))*C18	=B19*C$5+(1-C$5)*C19		16	=LN(D19)*F19
20	11	=(C$1+$A20-1)/($A20*(C$2+1))*B19	=(C$3+$A20-1)/($A20*(C$4+1))*C19	=B20*C$5+(1-C$5)*C20		11	=LN(D20)*F20
21	12	=(C$1+$A21-1)/($A21*(C$2+1))*B20	=(C$3+$A21-1)/($A21*(C$4+1))*C20	=B21*C$5+(1-C$5)*C21		10	=LN(D21)*F21
22	13	=(C$1+$A22-1)/($A22*(C$2+1))*B21	=(C$3+$A22-1)/($A22*(C$4+1))*C21	=B22*C$5+(1-C$5)*C22		9	=LN(D22)*F22
23	14	=(C$1+$A23-1)/($A23*(C$2+1))*B22	=(C$3+$A23-1)/($A23*(C$4+1))*C22	=B23*C$5+(1-C$5)*C23		17	=LN(D23)*F23
24	15	=(C$1+$A24-1)/($A24*(C$2+1))*B23	=(C$3+$A24-1)/($A24*(C$4+1))*C23	=B24*C$5+(1-C$5)*C24		6	=LN(D24)*F24
25	16	=(C$1+$A25-1)/($A25*(C$2+1))*B24	=(C$3+$A25-1)/($A25*(C$4+1))*C24	=B25*C$5+(1-C$5)*C25		8	=LN(D25)*F25
26	17	=(C$1+$A26-1)/($A26*(C$2+1))*B25	=(C$3+$A26-1)/($A26*(C$4+1))*C25	=B26*C$5+(1-C$5)*C26		11	=LN(D26)*F26
27	18	=(C$1+$A27-1)/($A27*(C$2+1))*B26	=(C$3+$A27-1)/($A27*(C$4+1))*C26	=B27*C$5+(1-C$5)*C27		5	=LN(D27)*F27
28	19	=(C$1+$A28-1)/($A28*(C$2+1))*B27	=(C$3+$A28-1)/($A28*(C$4+1))*C27	=B28*C$5+(1-C$5)*C28		9	=LN(D28)*F28
29	20+	=1-SUM(B9:B28)	=1-SUM(C9:C28)	=B29*C$5+(1-C$5)*C29		73	=LN(D29)*F29

Problem 6 -- 2 seg NBD Estimation

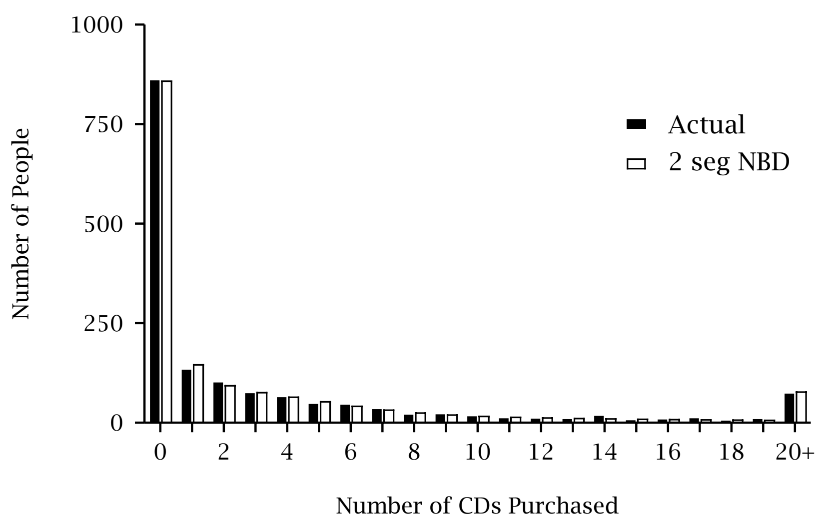
	A	B	C	D	E	F	G
1	Seg 1	r	0.168				
2		alpha	0.045				
3	Seg 2	r	138.136				
4		alpha	32.182				
5		pi_1	0.922			LL	-2929.533
6							
7			P(Y=y)				
8	Y	Seg 1	Seg 2	Total		f_y	
9	0	0.5900	0.0146	0.5449		860	-522.15
10	1	0.0949	0.0608	0.0922		133	-317.01
11	2	0.0530	0.1274	0.0589		101	-286.11
12	3	0.0367	0.1793	0.0478		74	-224.95
13	4	0.0278	0.1907	0.0405		64	-205.14
14	5	0.0222	0.1634	0.0332		47	-160.02
15	6	0.0183	0.1175	0.0260		45	-164.19
16	7	0.0154	0.0729	0.0199		34	-133.19
17	8	0.0132	0.0399	0.0153		20	-83.63
18	9	0.0115	0.0195	0.0121		21	-92.74
19	10	0.0100	0.0086	0.0099		16	-73.79
20	11	0.0089	0.0035	0.0085		11	-52.50
21	12	0.0079	0.0013	0.0074		10	-49.08
22	13	0.0071	0.0005	0.0066		9	-45.24
23	14	0.0064	0.0001	0.0059		17	-87.31
24	15	0.0058	0.0000	0.0053		6	-31.43
25	16	0.0052	0.0000	0.0048		8	-42.69
26	17	0.0047	0.0000	0.0044		11	-59.74
27	18	0.0043	0.0000	0.0040		5	-27.61
28	19	0.0040	0.0000	0.0037		9	-50.51
29	20+	0.0529	0.0000	0.0488		73	-220.51

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 (≥ 1).

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.

Notation

Let X_j = # CDs purchased on the jth 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

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:

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

Deriving the Distribution of Y

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

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

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

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

Deriving the Distribution of Y

(iii) We assume N is distributed according to the Poisson distribution with rate parameter λ

The unconditional distribution of Y is given by

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

This is known as the Pólya-Aeppli distribution

Introducing Heterogeneity

- The rate parameter λ 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:

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

- λ and p are independent

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

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 ($i = 1, \ldots, I$)

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

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

- 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:

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

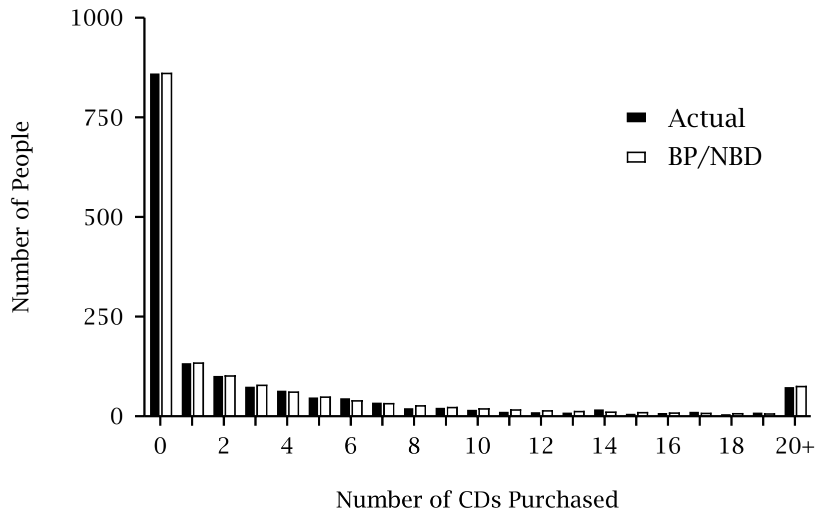
Parameter Estimation

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

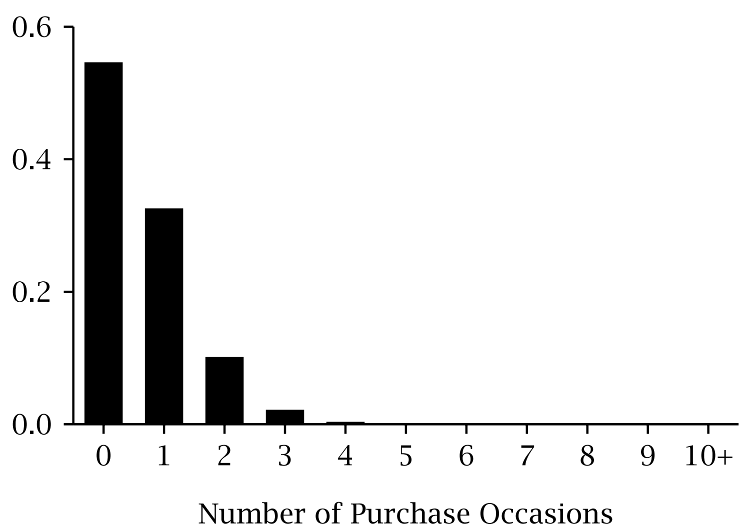
- We estimate the four model parameters (r, α, a, b) by fitting the BP/NBD to the data on the number of the purchases:

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

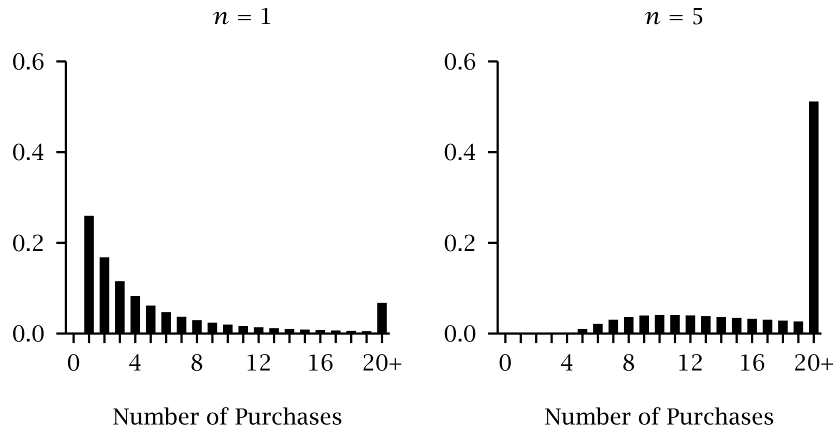
Fit of the BP/NBD ("Case 2")



Implied NBD for Purchase Occasions



Implied Beta-Pascal Distributions



Making Predictions

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

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

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

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

A Template for Integrated Models

		Stage 2		
		Counting	Timing	Choice
	Counting	CDNOW		
Stage 1	Timing			
	Choice			

Recap

- Modeling timing data
- Modeling count data
- Modeling "choice" data

— — — — — — — — — — — — — — —

- Incorporating covariates in count models
- Introducing additional model structures
- Building an "integrated" model

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

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