PREDICTIVE CUSTOMER ANALYTICS

Peter Fader

Frances and Pei-Yuan Chia Professor of Marketing
Co-Director, Wharton Customer Analytics Initiative
The Wharton School of the University of Pennsylvania
http://wcai.wharton.upenn.edu
Twitter: @faderp

Wharton UNIVERSITY of PENNSYLVANIA ONLINE

Organization: A public radio station supported primarily by contributions from its listeners

Challenge: Looking at listeners' histories of whether or not they gave each year, what can we predict about their future giving patterns?

Focal donors:

- Initial focus on 1995 cohort, ignoring donation amount
- 11,104 first-time supporters who made a total of 24,615 repeat donations over the next 6 years

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

HOW MUCH WILL DONORS GIVE IN THE FUTURE?
HOW DOES IT DEPEND ON THEIR PAST PATTERNS?

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

LET'S FIRST LOOK AT "BOB":

WHAT CAN WE PREDICT ABOUT HIS GIVING IN 2002-06?

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

WHAT CAN WE TELL ABOUT "SARAH"?

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
SARAH	**1**	**0**	**0**	**0**	**0**	**0**	**0**	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
100004	1	0	1	0	1	1	1	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
BOB	1	1	1	1	1	1	1	?	?	?	?	?
…												
111102	1	1	1	1	1	1	1	?	?	?	?	?
111103	1	0	1	1	0	1	1	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

HOW DO "MARY" AND "SHARMILA" COMPARE?

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
SARAH	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
MARY	**1**	**0**	**1**	**0**	**1**	**1**	**1**	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
BOB	1	1	1	1	1	1	1	?	?	?	?	?
SHARMILA	**1**	**1**	**1**	**1**	**1**	**1**	**0**	?	?	?	?	?
100010	1	0	0	0	0	0	0	?	?	?	?	?
...												

WHAT DONATION BEHAVIOR CHARACTERISTICS DO WE NEED TO TAKE INTO ACCOUNT?

GIVING BEHAVIORS

RECENCY: How recently did the donor give? When was the last time the donor gave?

FREQUENCY: How many times did the donor give in the past 6 years?

Most meaningful donation patterns can be described by these two metrics alone.

DONOR TYPES

ALIVE: Clearly an active donor: giving frequently and gave recently

DORMANT: Has not given recently, but is likely to give again with the right development prompts

LAPSED: Has not given recently and is not likely to give again

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
MARY	1	0	1	0	1	1	1	?	?	?	?	?
SHARMILA	1	1	1	1	1	1	0	?	?	?	?	?

Y	Y	N	N	**N**	**N**
Y	Y	N	N	**Y**	**Y**

What does it mean when there's one or more "no donation" at the end of a sequence?

a) The donor **lapsed** (i.e., left the donor pool)

b) The donor is **dormant** (i.e., decided not to give that year, didn't think of giving, etc.)

c) We don't know, but can build a model to come up with a "best guess"

Answer: c) We never know for sure whether the donor is lapsed or not; based on **recency** and **frequency** of his donation, we can make an educated guess about the probability of lapsing, so we can decide where to devote resources

Based on our best guesses about the probability of "death" and propensity to donate, we can calculate expected frequency of future donations for each donor

WHAT ABOUT "MARY" VERSUS "CHRIS"?

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
SARAH	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
MARY	**1**	**0**	**1**	**0**	**1**	**1**	**1**	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
BOB	1	1	1	1	1	1	1	?	?	?	?	?
...												
111102	1	1	1	1	1	1	1	?	?	?	?	?
CHRIS	**1**	**0**	**1**	**1**	**0**	**1**	**1**	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

You can “try this at home”

"BUY TILL YOU DIE" MODEL

We employ a "Buy Till You Die" model to predict future donation behaviors

The model only uses three inputs:

1. **Recency** (R)
2. **Frequency** (F)
3. **Number of people** for each combination of R/F

This requires a small amount of data and provides an easier structure to work with (i.e., data are aggregated from individual-level to R/F groups)

By assuming certain **probability distributions** for donors' **propensities**, we can construct a robust model that is **easy to implement on Excel**

This "BTYD" modeling approach has a **long track record of success** in a variety of different domains

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	alpha	1.204		B(alpha,beta)	1.146										
2	beta	0.750													
3	gamma	0.657		B(gamma,delta)	0.729										
4	delta	2.783													
5															
6	LL	-33225.6													
7												i			
8	x	t_x	n	# donors	L(.|X=x,t_x,n)			n - t_x - 1		0	1	2	3	4	5
9	6	6	6	1203	-2624.6	0.1129		-1	0.1129	0	0	0	0	0	0
10	5	6	6	728	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
11	4	6	6	512	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
12	3	6	6	357	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
13	2	6	6	234	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
14	1	6	6	129	-630.0	0.0076		-1	0.0076	0	0	0	0	0	0
15	5	5	6	335	-1245.1	[illegible]		0	0.0136	0.0107	0	0	0	0	0
16	4	5	6	284	-1447.1	0.0061		0	0.0046	0.0015	0	0	0	0	0
17	3	5	[illegible]	[illegible]	[illegible]3.5	0.0036		0	0.0030	0.0006	0	0	0	0	0
18	2	5	6	173	-952.6	0.0041		[illegible]	0.0035	0.0005	0	0	0	0	0
19	1	5	6	119	-567.3	0.0085		[illegible]	0.0076	0.0009	0	0	0	0	0
20	4	4	6	240	-923.6	0.0213		1	0.0046	0.0152	0.0015	0	0	0	0
21	3	4	6	181	-915.7	0.0063		1	0.0030	0.0027	0.0006	0	0	0	0
22	2	4	6	155	-805.3	0.0055		1	0.0035	0.0015	0.0005	0	0	0	0
23	1	4	6	78	-356.5	0.0104		1	0.0076	0.0018	0.0009	0	0	0	0
24	3	3	6	322	-1135.8	0.0294		2	0.0030	0.0230	0.0027	0.0006	0	0	0
25	2	3	6	255	-1151.6	0.0109		2	0.0035	0.0054	0.0015	0.0005	0	0	0
26	1	3	6	129	-545.0	0.0146		2	0.0076	0.0043	0.0018	0.0009	0	0	0
27	2	2	6	613	-1846.4	0.0492		3	0.0035	0.0383	0.0054	0.0015	0.0005	0	0
28	1	2	6	277	-993.9	0.0276		3	0.0076	0.0130	0.0043	0.0018	0.0009	0	0
29	1	1	6	1091	-2497.1	0.1014		4	0.0076	0.0737	0.0130	0.0043	0.0018	0.0009	0
30	0	0	6	3464	-4044.3	0.3111		5	0.0362	0.1909	0.0459	0.0189	0.0098	0.0058	0.0037

=EXP(GAMMALN(B1)+GAMMALN(B2)-GAMMALN(B1+B2))

=EXP(GAMMALN($B$1+A9)+GAMMALN($B$2+C9-A9)-GAMMALN($B$1+$B$2+C9))/$E$1*EXP(GAMMALN($B$3)+GAMMALN($B$4+C9)-GAMMALN($B$3+$B$4+C9))/$E$3

=SUM(E9:E19)

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

=C15-B15-1

=D19*LN(F19)

=SUM(I19:O19)

Detailed step-by-step instructions available at http://brucehardie.com/notes/010/

EXPECTED # OF DONATIONS IN 2002-2006

AS A FUNCTION OF RECENCY AND FREQUENCY

Name	1995	1996	1997	1998	1999	2000	2001	R	F
BOB	1	1	1	1	1	1	1	6	6
SARAH	1	0	0	0	0	0	0	1	1
MARY	1	0	1	0	1	1	1	6	4
SHARMILA	1	1	1	1	1	1	0	5	5
CHRIS	1	0	1	1	1	0	1	6	4

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

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

Mary and **Chris** have the same RF (6,4), so their expected number of donations going forward is the same

Even though **Mary** and **Chris** have lower F than **Sharmila**, their higher R suggests that they are Alive, thus they are 50% more valuable than **Sharmila**

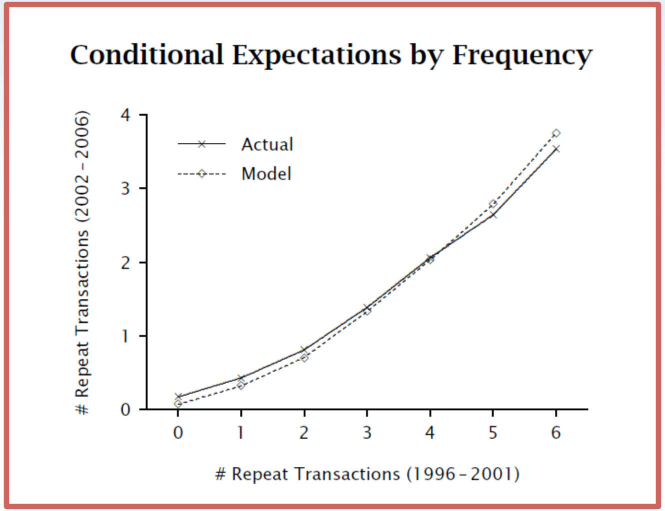
Sharmila (5,5), despite high donation rate, has likely lapsed

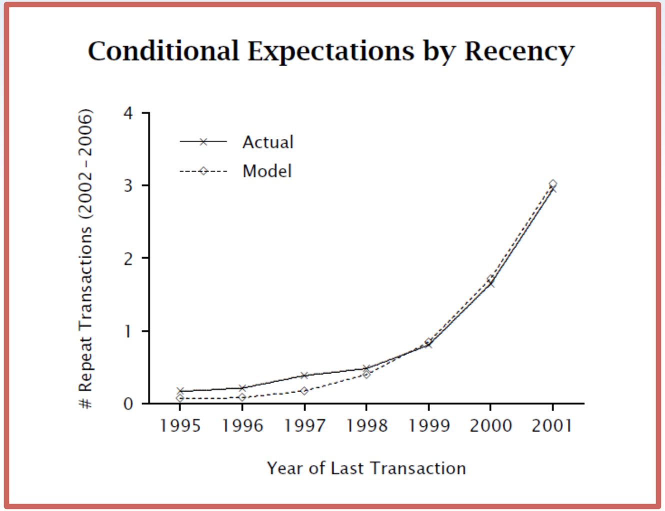
Sarah, with very low R and F, is lapsed and/or a very light donor (hard to tell)

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

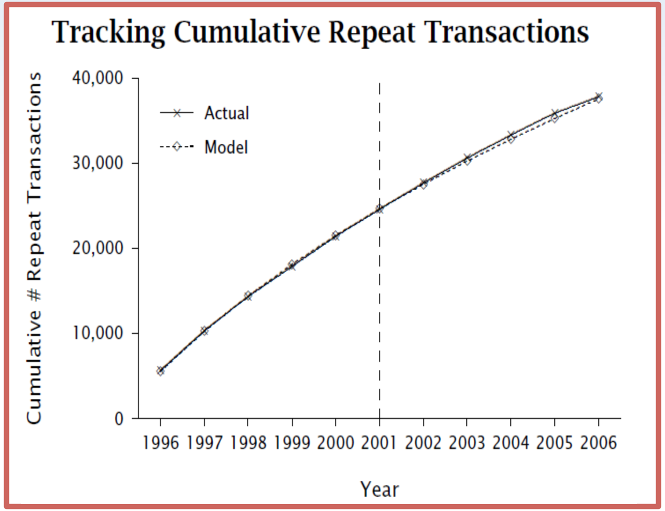
HOW DO WE KNOW THE MODEL WORKS?

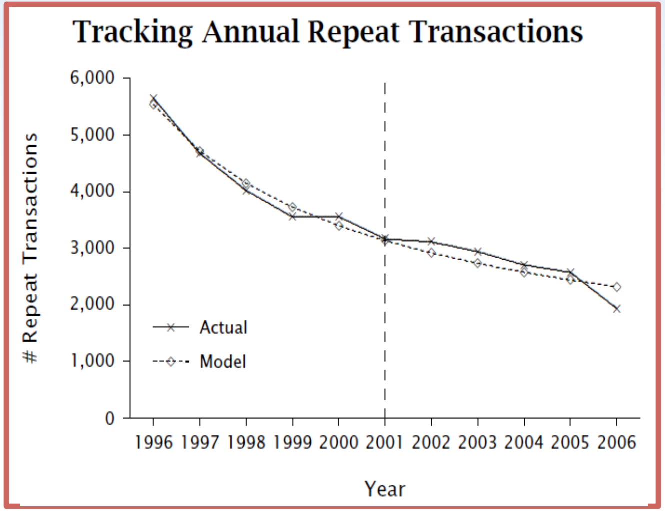
The model does a solid job of **making predictions** about future donation behaviors, as **conditional expectations** correspond very well with actual holdout period data (2002 – 2006)





Overall, the model is exceptional good at **fitting the historical data** as well as the **holdout period data** (i.e., model predictions vs. actual number of donations between 2002 and 2006)





A SECOND ILLUSTRATION

Using a larger dataset from a different non-profit firm, we create a “heat map” that shows which combinations of RF will likely yield the most valuable donors

	1990	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
x / tx	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0.01																			
1		0.00	0.01	0.01	0.02	0.03	0.04	0.06	0.07	0.09	0.11	0.13	0.15	0.17	0.19	0.22	0.24	0.26	0.29	0.31
2			0.00	0.00	0.01	0.02	0.03	0.05	0.07	0.09	0.13	0.16	0.20	0.25	0.29	0.34	0.39	0.43	0.47	0.51
3				0.00	0.00	0.01	0.01	0.03	0.05	0.07	0.11	0.16	0.22	0.29	0.37	0.44	0.52	0.59	0.66	0.71
4					0.00	0.00	0.00	0.01	0.02	0.05	0.08	0.13	0.21	0.30	0.41	0.52	0.64	0.74	0.84	0.91
5						0.00	0.00	0.00	0.01	0.02	0.05	0.09	0.17	0.28	0.41	0.57	0.74	0.89	1.01	1.12
6							0.00	0.00	0.00	0.01	0.02	0.06	0.12	0.23	0.39	0.59	0.81	1.02	1.19	1.32
7								0.00	0.00	0.00	0.01	0.03	0.08	0.17	0.33	0.58	0.86	1.14	1.36	1.52
8									0.00	0.00	0.00	0.01	0.04	0.11	0.26	0.52	0.88	1.23	1.52	1.72
9										0.00	0.00	0.00	0.02	0.06	0.19	0.44	0.85	1.31	1.68	1.92
10											0.00	0.00	0.01	0.03	0.12	0.34	0.79	1.36	1.83	2.12
11												0.00	0.00	0.01	0.06	0.24	0.68	1.37	1.97	2.32
12													0.00	0.00	0.03	0.15	0.54	1.33	2.10	2.53
13														0.00	0.01	0.07	0.39	1.23	2.20	2.73
14															0.00	0.03	0.23	1.06	2.28	2.93
15																0.01	0.11	0.81	2.30	3.13
16																	0.03	0.51	2.25	3.33
17																		0.20	2.04	3.53
18																			1.48	3.73
19																				3.94

PROBABILITY OF BEING "ALIVE"

x/tx	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
0	0.06																			
1		0.01	0.03	0.05	0.07	0.10	0.13	0.17	0.22	0.27	0.32	0.38	0.45	0.51	0.58	0.65	0.73	0.80	0.87	0.93
2			0.00	0.01	0.02	0.03	0.06	0.09	0.12	0.17	0.23	0.30	0.37	0.45	0.54	0.62	0.70	0.79	0.86	0.93
3				0.00	0.00	0.01	0.02	0.03	0.06	0.10	0.15	0.21	0.29	0.38	0.48	0.58	0.68	0.77	0.86	0.93
4					0.00	0.00	0.00	0.01	0.02	0.05	0.08	0.14	0.21	0.31	0.42	0.53	0.65	0.76	0.85	0.93
5						0.00	0.00	0.00	0.01	0.02	0.04	0.08	0.14	0.23	0.35	0.48	0.62	0.74	0.85	0.93
6							0.00	0.00	0.00	0.01	0.02	0.04	0.09	0.16	0.28	0.42	0.58	0.72	0.84	0.93
7								0.00	0.00	0.00	0.01	0.02	0.05	0.10	0.21	0.35	0.53	0.70	0.84	0.93
8									0.00	0.00	0.00	0.01	0.02	0.06	0.14	0.28	0.48	0.67	0.83	0.93
9										0.00	0.00	0.00	0.01	0.03	0.09	0.22	0.41	0.64	0.82	0.93
10											0.00	0.00	0.00	0.01	0.05	0.15	0.35	0.60	0.81	0.93
11												0.00	0.00	0.00	0.03	0.10	0.27	0.55	0.79	0.93
12													0.00	0.00	0.01	0.05	0.20	0.49	0.78	0.93
13														0.00	0.00	0.03	0.13	0.42	0.75	0.93
14															0.00	0.01	0.07	0.34	0.73	0.93
15																0.00	0.03	0.24	0.69	0.93
16																	0.01	0.14	0.63	0.93
17																		0.05	0.54	0.93
18																			0.37	0.93
19																				0.93

SUMMARY: HOW IS THIS METHOD DIFFERENT?

There are many models that predict future donation behaviors; we believe our method is different / superior because:

1. The model requires a **very small amount of data** (Recency and Frequency), compared to other models that require a large dataset (typically detailed individual-level characteristics, e.g., demographics)

2. The model has demonstrated robust **out-of-sample validation**

3. The model can be **generalized to other types of behaviors**; it is not excessively customized to the donation domain

4. The model can easily be **implemented on Excel**; it does not require any proprietary or specialized software

WANT MORE?

For large-scale databases, use our open-source "BTYD" R Library:

http://cran.r-project.org/web/packages/BTYD/BTYD.pdf

While this model offers accurate predictions and useful insights about how to understand donation propensities, it stops short of offering any specific advice about **which donors to target** and **when / how to do so**

Building on this model, Schweidel and Knox explore the **impact of fundraising efforts** on donation activity in: "Incorporating Direct Marketing Activity into Latent Attrition Models" (http://ssrn.com/abstract=1670060)

Professor Peter Fader

www.petefader.com

Twitter: @faderp

Wharton Customer Analytics Initiative

www.wharton.upenn.edu/wcai/