How Not to Project Customer Retention

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie
www.brucehardie.com[†]

January 2007

Summary

Kumar and Reinartz (2006) suggest that future retention rates for a cohort of customers can be forecast as a function of time using the formula $r_t = r_\infty\big(1-\exp(-\lambda t)\big)$. Our analysis suggests this is not a good idea; we believe that any attempt to take a series of past retention numbers (for a given group of customers) and project them into the future should be based off a formal model of contract duration, such as the shifted-beta-geometric distribution proposed by Fader and Hardie (2007).

Analysis

1. The task of taking a series of past retention numbers (for a given group of customers) and projecting them into the future is an important component of any effort to make predictions about customer tenure, lifetime value, and so on in a contractual setting.

2. Berry and Linoff (2004) tackle the problem by fitting flexible functions of time to the survival data, which are then used to project the survivor function beyond the range of observations. The bad forecasts lead Berry and Linoff to conclude that "parametric approaches do not work" when seeking to project the survivor function beyond the range of observations.

3. Fader and Hardie (2007), hereafter FH, feel that such a conclusion is premature and propose that the shifted-beta-geometric (sBG) distribution be used as a model of customer contract duration in a discrete-time contractual setting. They demonstrate that it can provide accurate forecasts and other useful diagnostics about customer retention.

4. If our interest lies in projecting retention rates rather than the survivor function, why obtain the estimates of future retention rates indirectly from the projected survivor function? Why not model retention rates directly, fitting a flexible function of time to the observed retention rates and using the resulting formula to generate retention rate estimates for future periods?

5. Such an approach is proposed by Kumar and Reinartz (2006, p. 100), hereafter KR, who suggest we model retention rates using the function

$$r_t = r_\infty\big(1 - \exp(-\lambda t)\big)\,, \tag{1}$$

where r_t is the retention rate for period t, r_∞ is the retention rate ceiling, and λ determines how quickly retention rates converge over time to the retention ceiling.

6. While such an approach seems very plausible, does it work practice?

7. The retention rate numbers presented in Table 1 are for two segments of customers ("Regular" and "High End") for an unspecified subscription-type business. (These retention rates are computed from the survival data give in FH, Table 1.)

Year	Regular	High End
1	0.631	0.869
2	0.742	0.855
3	0.816	0.879
4	0.853	0.908
5	0.887	0.929
6	0.907	0.938
7	0.920	0.950
8	0.925	0.953
9	0.928	0.951
10	0.937	0.960
11	0.943	0.958
12	0.945	0.963

Table 1: Observed retention rates for years 1–12.

8. Suppose we only have the first seven years of data and wish to compute estimates of $r_8, r_9, \ldots$. We will fit (1) to the data and use the resulting equation to project the retention rates into the future.

9. Since (1) is a nonlinear function of time, we obtain estimates of the two model parameters via nonlinear least squares (NLS), which sees

us finding the values of r_∞ and λ that minimize the sum of squared errors

$$SSE = \sum_{t=1}^{7} (\text{Actual } r_t - \text{Model } r_t)^2 \,. \tag{2}$$

10. This is easy to do in Excel. Figure 1 shows we "code up" (1) and (2) for the "Regular" segment dataset.

	A	B	C	D
1	r_infinity	0.885		
2	lambda	1.092		
3	SSE	7.25E-03		
4				
5		Model	Actual	(M-A)^2
6	1	0.588	0.631	1.84E-03
7	2	0.785	0.742	1.88E-03
8	3	0.874	0.816	1.26E-03
9	4	0.874	0.853	4.29E-04
10	5	0.881	0.887	3.41E-05
11	6	0.884	0.907	5.45E-04
12	7	0.884	0.920	1.26E-03
13	8	0.885	0.925	
14	9	0.885	0.928	
15	10	0.885	0.937	
16	11	0.885	0.943	
17	12	0.885	0.945	

Figure 1: Estimating the "Regular" segment model parameters using NLS.

We find the NLS estimates of the two model parameters by using the Solver add-in to find the values of r_∞ and λ (cells `B1:B2`) that *minimize* SSE (cell `B3`).

11. This is replicated for the "High End" segment data in Table 2.

	A	B	C	D
1	r_infinity	0.911		
2	lambda	2.940		
3	SSE	6.49E-03		
4				
5		Model	Actual	(M-A)^2
6	1	0.863	0.869	3.54E-05
7	2	0.909	0.855	2.88E-03
8	3	0.911	0.879	1.03E-03
9	4	0.911	0.908	1.05E-05
10	5	0.911	0.929	3.15E-04
11	6	0.911	0.938	7.16E-04
12	7	0.911	0.950	1.50E-03
13	8	0.911	0.953	
14	9	0.911	0.951	
15	10	0.911	0.960	
16	11	0.911	0.958	
17	12	0.911	0.963	

Figure 2: Estimating the "High End" segment model parameters using NLS.

12. The model-based retention rate numbers are plotted in Figure 3, along with the corresponding actual retention rates. Clearly the model is not working! For both datasets, it fails to track *and* predict the retention rates. We note that its forecasts for r_t level-off too early, to the extent that it is under-forecasting by the end of the model calibration period.

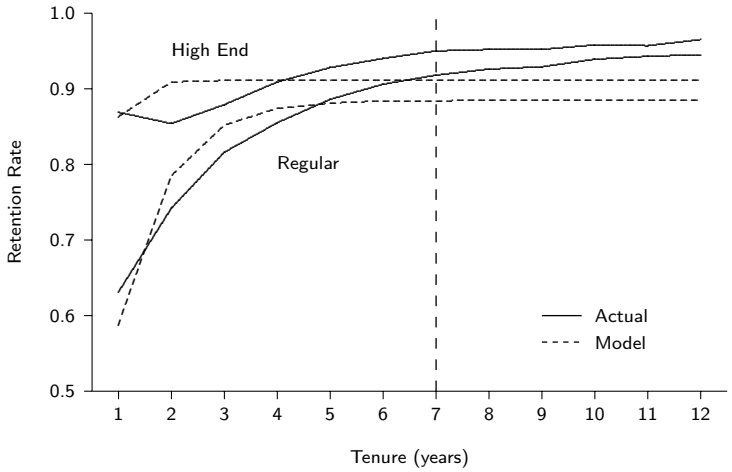

Figure 3: Actual versus model estimates of retention rates by tenure for the High End and Regular segments.

13. In contrast, consider the projections associated with the sBG model (Figure 4). No comment is required!

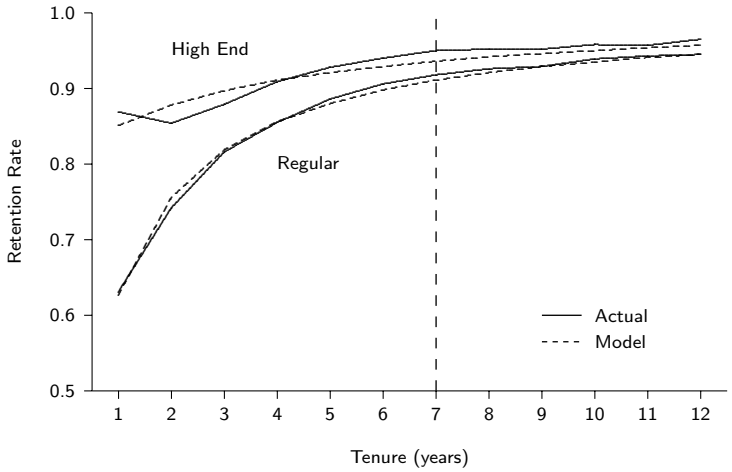

Figure 4: Actual versus sBG estimates of retention rates by tenure for the High End and Regular segments.

14. Central to KR's examination of (1) is their Exhibit 5-3, a copy of which is presented in Figure 5.

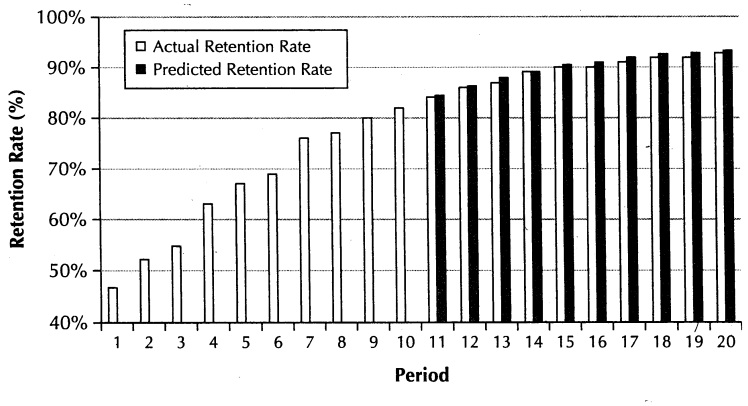

Figure 5: Exhibit 5-3 (KR, p. 100)

This shows the actual retention rates numbers (white bars) for a credit card company over a period of twenty quarters. The black bars show the estimates of the retention rates for periods 11–20 for $r_\infty = 0.95$ and $\lambda = 0.20$. (Both sets of retention rate numbers are expressed as percentages.) KR conclude that "[i]t can be seen that the method *to approximate* the actual retention rates was very close."

15. How do we reconcile the conflicting conclusions regarding model performance as indicated by Figures 3 and 5?

16. Looking closely at Figure 5, we see that there are no estimates (black bars) for the first ten periods. Figure 6 replicates this figure, adding in the estimates of $r_1, r_2, \ldots, r_{10}$. (We obtain the actual numbers by using a ruler to measure them off a photo-enlarged copy of the graph.)

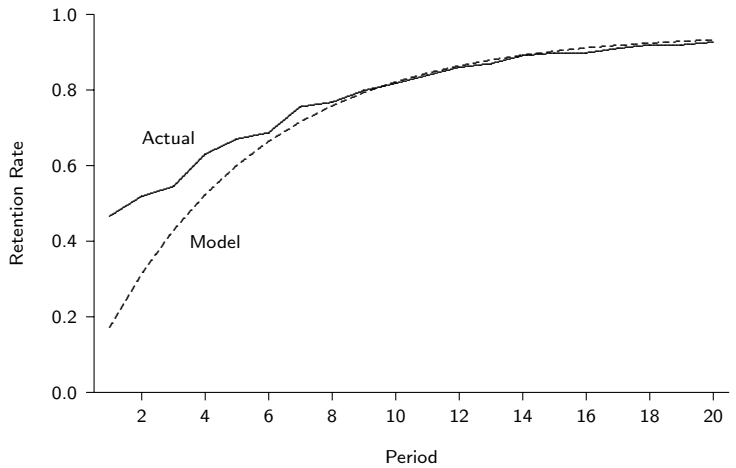

Figure 6: Exhibit 5-3 with retention rate estimates for the first ten periods.

17. Clearly the parameter values $r_\infty = 0.95$ and $\lambda = 0.20$ were not estimated from, say, the first ten periods of data. (Let's just say that Figure 5 is potentially misleading.)

18. Using NLS, we fit (1) to the first ten periods of data (Figure 7) and examine the how well the model tracks the period-by-period retention numbers (Figure 8).

	A	B	C	D
1	r_infinity	0.757		
2	lambda	0.577		
3	SSE	0.0375		
4				
5		Model	Actual	(M-A)^2
6	1	0.3317	0.4664	0.0181
7	2	0.5181	0.5186	0.0000
8	3	0.6228	0.5447	0.0061
9	4	0.6816	0.6300	0.0027
10	5	0.7146	0.6704	0.0020
11	6	0.7332	0.6870	0.0021
12	7	0.7436	0.7557	0.0001
13	8	0.7495	0.7676	0.0003
14	9	0.7528	0.7984	0.0021
15	10	0.7546	0.8174	0.0039
16	11	0.7556	0.8387	
17	12	0.7562	0.8601	
18	13	0.7566	0.8696	
19	14	0.7567	0.8909	
20	15	0.7568	0.8980	
21	16	0.7569	0.8980	
22	17	0.7569	0.9099	
23	18	0.7570	0.9194	
24	19	0.7570	0.9194	
25	20	0.7570	0.9265	

Figure 7: Fitting (1) to the KR data using NLS.

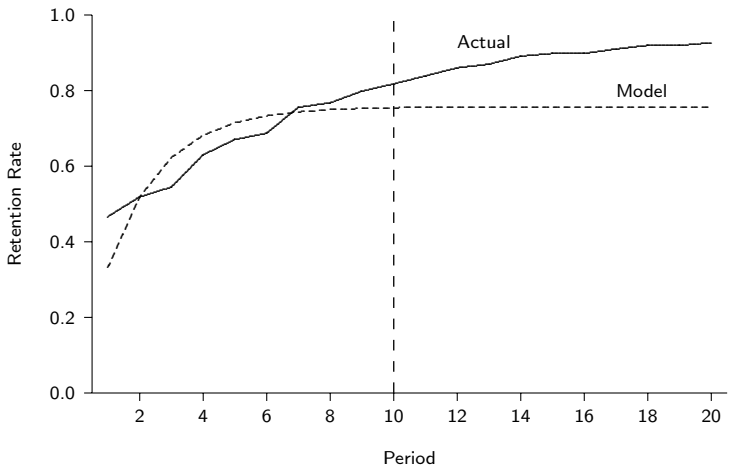

Figure 8: Actual versus model estimates of retention rates for the KR data.

19. Once again, we see that model fails to track and predict the retention rates. We note that its forecasts for r_t level-off too early, to the extent that it is under-forecasting by the end of the model calibration period.

20. Based on the results presented in Figures 3 and 8, we conclude that (1) should **not** be used as a model of retention rates.

21. But is the sBG model any better for this dataset?

22. FH derive the following expression for the period t retention rate as implied by the sBG model:

$$r_t = \frac{\beta + t - 1}{\alpha + \beta + t - 1}\,. \tag{3}$$

23. Using NLS, we fit (3) to the first ten periods of data (Figure 9) and examine the how well the model tracks the period-by-period retention numbers (Figure 10).

	A	B	C	D
1	alpha	3.010		
2	beta	2.346		
3	SSE	0.0055		
4				
5		Model	Actual	(M-A)^2
6	1	0.4381	0.4664	0.0008
7	2	0.5265	0.5186	0.0001
8	=(B2+A6-1)/(B1+B2+A6-1)			0.0021
9	4	0.6398	0.6300	0.0001
10	5	0.6783	0.6704	0.0001
11	6	0.7094	0.6870	0.0005
12	7	0.7350	0.7557	0.0004
13	8	0.7564	0.7676	0.0001
14	9	0.7746	0.7984	0.0006
15	10	0.7903	0.8174	0.0007
16	11	0.8040	0.8387	
17	12	0.8160	0.8601	
18	13	0.8266	0.8696	
19	14	0.8360	0.8909	
20	15	0.8445	0.8980	
21	16	0.8521	0.8980	
22	17	0.8591	0.9099	
23	18	0.8654	0.9194	
24	19	0.8711	0.9194	
25	20	0.8764	0.9265	

Figure 9: Fitting (3) to the KR data using NLS.

While the tracking/predictive performance of the sBG model on this dataset is not up to the standard of that for the first two datasets (Figure 4), it is so much better than that of (1), under-predicting the period 20 retention rate by 5.4% (versus 18.3%).

24. This analysis, coupled with that presented in FH, suggests that any attempt to take a series of past retention numbers (for a given group

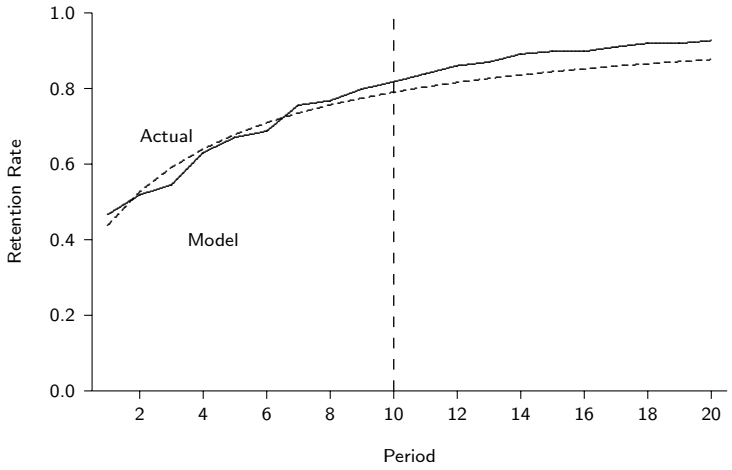

Figure 10: Actual versus sBG model estimates of retention rates for the KR data.

of customers) and project them into the future should not tackle the problem by fitting flexible functions of time to either the survival or retention rate data.

25. Such an analysis task should use a formal model of contract duration; the sBG distribution is one such model.

References

Berry, Michael J. A. and Gordon S. Linoff (2004), *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd edition, Indianapolis, IN: Wiley Publishing, Inc. Reference

Fader, Peter S. and Bruce G. S. Hardie (2007), "How to Project Customer Retention," *Journal of Interactive Marketing*, **21** (Winter), 76–90.

Kumar, V. and Werner J. Reinartz (2006), *Customer Relationship Management: A Databased Approach*, Hoboken, NJ: John Wiley & Sons, Inc.