HW1

Alan Wu
MKTG 7760: Probability Models in Marketing

January 29, 2026

1 Question 1

Look carefully at slide 18 (confusingly labeled as 19) from the Session 1 deck, which compares the survival curves for the actual data and the homogeneous shifted geometric model. Explain the major/systematic differences between the two curves. (Hint: think about why the sBG model tracks the actual data so well.)

The discrepancy between the actual survival curve and the homogeneous geometric model arises because the model incorrectly assumes that all customers share the exact same likelihood of leaving, regardless of how long they have been with the company. In the early periods ($t < 6$), the model's curve sits above the actual data, indicating that it underestimates how quickly people leave initially. However, in the later periods ($t > 7$), the model crosses and falls below the actual data, predicting a continued steep decline while the real curve flattens out. This happens because the geometric model applies a constant "average" churn rate to everyone forever. In reality, the people who stay for a long time are naturally those who are less likely to leave, meaning the group's overall tendency to drop out should decrease over time rather than staying constant.

2 Question 2

Run the sBG model and produce forecasts for both datasets ("High End" and "Regular") in two different ways: using the seven-period calibration period we used in class as well as a ten-period one. Briefly describe your results and any key conclusions that you draw from them

Below summarizes the results. The full excel sheet screenshots is in the appendix section. We will show the results of the forecasting and then discuss the results by comparing the regular cohort to the high end cohort, and also using 7 periods of calibration to 10 periods of calibration.

2.1 Overall results of each run + graph

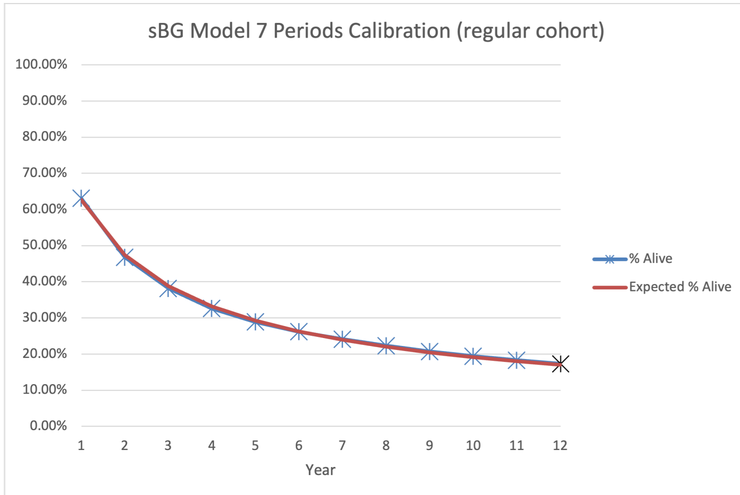


Figure 1: sBG Model on regular cohort with 7 periods of calibration

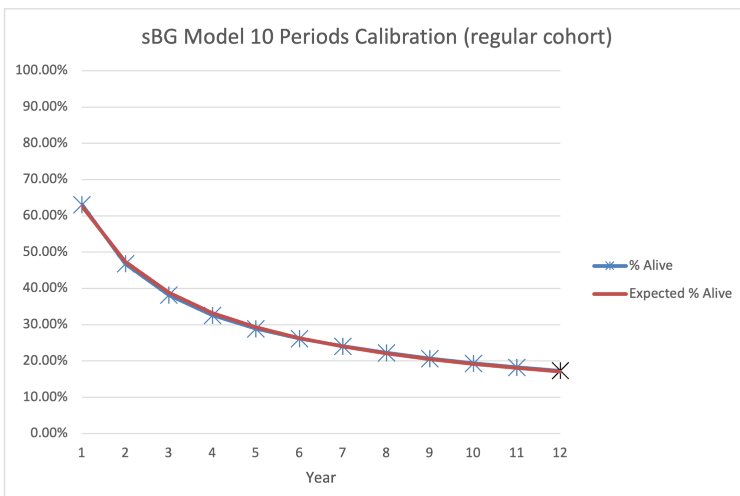


Figure 2: sBG Model on regular cohort with 10 periods of calibration

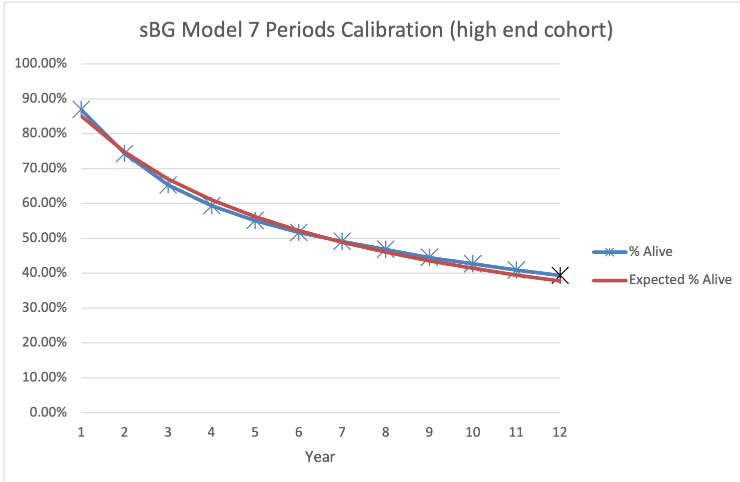


Figure 3: sBG Model on high end cohort with 7 periods of calibration

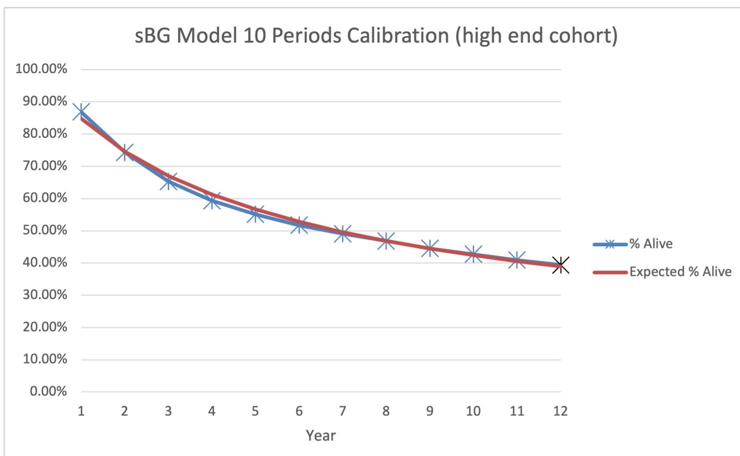


Figure 4: sBG Model on high end cohort with 10 periods of calibration

Table 1: Comparison of MAPE Across Cohorts and Calibration Lengths

		MAPE (%)	
Cohort	**Calibration Length**	**Calibration**	**Forecasting**
Regular	7 Periods	1.07	1.49
	10 Periods	1.07	1.16
High End	7 Periods	1.64	2.83
	10 Periods	1.53	0.94

Some basic conclusions that we can come to are that regardless of the calibration length and the cohort, it seems that the beta-geometric model has a very good fit for the calibration. With low MAPE for all 4 experiements, the probability model with heterogeneity

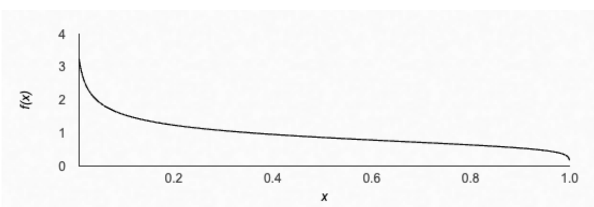
Table 2: Solver Parameter Estimates (α and β) by Cohort and Calibration Length

Cohort	**Calibration Length**	α	β
Regular	7 Periods	0.704	1.182
	10 Periods	0.697	1.172
High End	7 Periods	0.668	3.806
	10 Periods	0.605	3.389

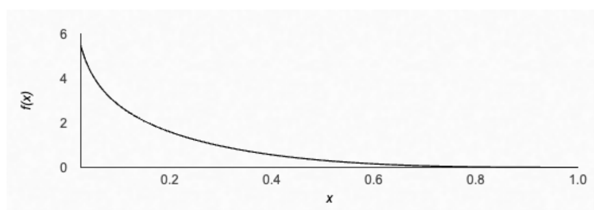
assumptions by beta distribution performs much better than the homogeneous or regression attempts at the churn problem.

2.2 Comparison of high end to regular

It is evident that the high end and the regular cohorts are two different segments. This is evident by analyzing the resulting α and β values. If we simply compare the 7 period calibration length experiments across high end and regular cohorts, we notice that they have the following distributions:



(a) (Regular) Beta Distribution for $(\alpha, \beta) = (0.704, 1.182)$



(b) (High End) Beta Distribution for $(\alpha, \beta) = (0.668, 3.806)$

The figure illustrates the heterogeneity in churn probabilities (θ) estimated by the model for both cohorts. The shape of the Beta distribution, governed by parameters α and β, reveals a distinct difference in customer retention.

For the **High End** cohort, the estimated parameters ($\alpha = 0.668, \beta = 3.806$) produce a distribution that is heavily skewed toward zero. This indicates a highly loyal customer base with a low expected churn probability of approximately 14.9% ($E[\theta] = \frac{\alpha}{\alpha+\beta}$). The large β value suggests that the vast majority of these customers have a negligible risk of dropping out in any given period.

In contrast, the **Regular** cohort ($\alpha = 0.704, \beta = 1.182$) exhibits a much flatter distribution with a significantly heavier tail. The lower β value pushes the expected churn probability up to 37.3%, implying that this segment is far more volatile. The model perceives the Regular cohort as containing a substantial subset of "high-risk" customers who are likely to churn early, unlike the uniformly sticky High End group.

2.3 Comparison of seven-period to ten-period

The extension of the calibration window from 7 to 10 periods yields a marked improvement in predictive accuracy, particularly in the out-of-sample forecasting phase. While the in-sample calibration fit remains relatively stable, the forecasting performance benefits significantly from the additional data, with the High End cohort seeing its error drop from 2.83% to 0.94%.

This improvement is likely driven by the model's ability to observe the evolution of the customer base. In the initial periods, the cohort contains a mix of high-risk and low-risk customers, resulting in higher aggregate churn. By extending the window to 10 periods, the model captures the stabilization of the retention curve as high-risk customers exit. This allows the solver to better identify the underlying loyalty of the remaining customers, preventing the overestimation of future churn that occurs with the shorter, 7-period window.

3 Question 3

For an sBG timing process with parameters (α, β) = (1, 2), what is the expected "half life" of a cohort of newly acquired customers (i.e., how many renewal cycles should we expect to wait until 50% of the customers have churned)?

For an sBG timing process parameterized by $(\alpha, \beta) = (1, 2)$ and for the half-life of this cohort, we would like to be able to find the t, where $t \in 1, 2, .., n$ where n is the total number of periods we tracked, in which $P(T > t|\alpha, \beta) = 0.5$. This is because in order to survive up until the t-th cycle, then we are equivalently saying we would like to find $P(T \leq t|\alpha, \beta)$. However, since the probability we are looking for is 0.5, or half, then our equation $P(T > t|\alpha, \beta) = 0.5$ is equivalent and easier to deal with since we've already derived the survival function.

Therefore, to solve the equation $P(T > t|\alpha, \beta) = 0.5$, we have

$$\begin{aligned} 0.5 &= P(T > t|\alpha, \beta) \\ &= \frac{\beta(\alpha, \beta + t)}{\beta(\alpha, \beta)} \\ &= \frac{\frac{\Gamma(1)\Gamma(2+t)}{\Gamma(3+t)}}{\frac{\Gamma(1)\Gamma(2)}{\Gamma(3)}} \\ &= \frac{\Gamma(2+t)}{\Gamma(3+t)} \cdot \frac{\Gamma(3)}{\Gamma(2)} \\ &= \frac{2}{2+t} \end{aligned}$$

Solving

$$0.5 = \frac{2}{2+t}$$

We get $t = 2$. And thus, the expected half life of a cohort of newly acquired customers is $\mathbf{t = 2}$ periods.

Similarly, we can simulate the values using the excel chart by inputting $(\alpha, \beta) = (1, 2)$ and realize the expected percentage alive becomes 50% after two periods.

4 Question 4

Derive the "unshifted" beta-geometric model and its key characteristics: $P(t|\alpha, \beta)$, $S(t|\alpha, \beta)$, and the formulas needed to compute $P(t|\alpha, \beta)$ via forward recursion

We derive the unshifted beta-geometric model by returning to the unshifted geometric probability mass function. We know that the shifted geometric distribution starts at t = 1,2,.. . However, the unshifted distribution starts at t = 0,1,... .

Going back to the homogeneous geometric model, parameterized by θ we now find that $P(T = t|\theta) = \theta(1-\theta)^t$, for $t = 0, 1, \ldots$. Furthermore, for the overall survival probability, we see that this shifts to $P(T > t|\theta) = (1-\theta)^{t+1}$, for $t = 0, 1, \ldots$. This is because that when we shift to include 0, we assume that there can be some churn probability. Thus, the survival probability at time t = 0 should be $1 - \theta$, which is confirmed in our survival probability function.

Now, we will derive the probability of churn at any time t, $P(T = t|\alpha, \beta)$ and the probability of survival $P(T > t|\alpha, \beta)$ for the unshifted beta-geometric model, now parameterized by α, β.

Before deriving, we should note that the beta distribution does not change from the sBG model, and we maintain that the pdf of the Beta distribution remains the same, as $f(\theta|\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\beta(\alpha,\beta)}$. And to derive the formulas, we will be performing the integration by cleaning up constants, consolidating like terms, and then matching the resultant integral to a beta distribution with modified parameters.

- $\mathbf{P(T = t|\alpha, \beta)}$. From our previous derivations, we used $P(T = t|\alpha, \beta) = \int_{\theta=0}^{1} P(T = t|\theta) f(\theta|\alpha, \beta) d\theta$. We will do the same here

$$
\begin{aligned}
P(T=t|\alpha,\beta) &= \int_{\theta=0}^{1} P(T=t|\theta) f(\theta|\alpha,\beta) d\theta \\
&= \int_{\theta=0}^{1} \theta(1-\theta)^t \cdot \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\beta(\alpha,\beta)} d\theta \\
&= \frac{1}{\beta(\alpha,\beta)} \int_{\theta=0}^{1} \theta^{\alpha}(1-\theta)^{\beta+t-1} d\theta \\
&= \frac{\beta(\alpha+1,\beta+t)}{\beta(\alpha,\beta)} \int_{\theta=0}^{1} \frac{\theta^{\alpha}(1-\theta)^{\beta+t-1}}{\beta(\alpha+1,\beta+t-1)} d\theta \\
&= \frac{\beta(\alpha+1,\beta+t)}{\beta(\alpha,\beta)}
\end{aligned}
$$

- $\mathbf{S}(\mathbf{T}=\mathbf{t}|\alpha,\beta)$. From our previous derivation for the survival function, we used that $S(T=t|\alpha,\beta) = P(T>t|\alpha,\beta) = \int_{\theta=0}^{1} P(T>t|\theta) f(\theta|\alpha,\beta) d\theta$ The equation is below:

$$
\begin{aligned}
P(T>t|\alpha,\beta) &= \int_{\theta=0}^{1} P(T>t|\theta) f(\theta|\alpha,\beta) d\theta \\
&= \int_{\theta=0}^{1} (1-\theta)^{t+1} \cdot \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\beta(\alpha,\beta)} d\theta \\
&= \frac{1}{\beta(\alpha,\beta)} \int_{\theta=0}^{1} \theta^{\alpha-1}(1-\theta)^{\beta+t} d\theta \\
&= \frac{\beta(\alpha,\beta+t+1)}{\beta(\alpha,\beta)} \int_{\theta=0}^{1} \frac{\theta^{\alpha}(1-\theta)^{\beta+t}}{\beta(\alpha+1,\beta+t+1)} d\theta \\
&= \frac{\beta(\alpha,\beta+t+1)}{\beta(\alpha,\beta)}
\end{aligned}
$$

Using these two formulas, we can find the forward recursive step $P(T=t|\alpha,\beta)$ for each $t = 0, 1, \ldots.$ We derive the forward recursion the same way we do with the original shifted beta geometric. We have two cases: at t = 0, and at t ¿ 0.

- t = 0. We have:

$$
\begin{aligned}
P(T=0|\alpha,\beta) &= \frac{\beta(\alpha+1,\beta)}{\beta(\alpha,\beta)} \\
&= \frac{\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}
$$

- $t > 0$. To define the recursion rule, we solve $\frac{P(T=t|\alpha,\beta)}{P(T=t-1|\alpha,\beta)}$

$$
\begin{aligned}
\frac{P(T=t|\alpha,\beta)}{P(T=t-1|\alpha,\beta)} &= \frac{\frac{\beta(\alpha+1,\beta+t)}{\beta(\alpha,\beta)}}{\frac{\beta(\alpha+1,\beta+t-1)}{\beta(\alpha,\beta)}} \\
&= \frac{\beta(\alpha+1,\beta+t)}{\beta(\alpha+1,\beta+1-1)} \\
&= \frac{\frac{\Gamma(\alpha+1)\Gamma(\beta+t)}{\Gamma(\alpha+\beta+1+t)}}{\frac{\Gamma(\alpha+1)\Gamma(\beta+t-1)}{\Gamma(\alpha+\beta+t)}} \\
&= \frac{\Gamma(\beta+t)}{\Gamma(\beta+t-1)} \cdot \frac{\Gamma(\alpha+\beta+t)}{\Gamma(\alpha+\beta+t+1)} \\
&= \frac{\beta+t-1}{\alpha+\beta+t}
\end{aligned}
$$

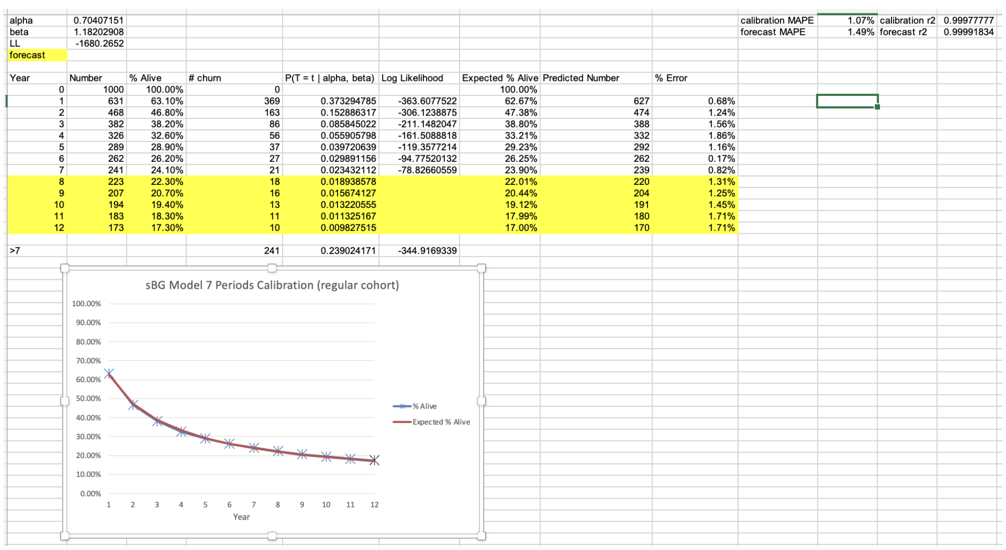
And thus, $P(T=t|\alpha,\beta) = P(T=t-1|\alpha,\beta) \cdot \frac{\beta+t-1}{\alpha+\beta+t}$ for $t = 1, 2, \ldots$

Therefore, for the forward recursion of the unshifted beta-geometric model, we have:

$$
P(T=t|\alpha,\beta) = \begin{cases} \frac{\alpha}{\alpha+\beta} & \text{for } t = 0 \\ \frac{\beta+t-1}{\alpha+\beta+t} \cdot P(T=t-1|\alpha,\beta) & \text{for } t = 1, 2, \ldots \end{cases}
$$

5 Appendix

5.1 Total Excel Tables



alpha	0.70407151	calibration MAPE	1.07%
beta	1.18202908	forecast MAPE	1.49%
LL	-1680.2652	calibration r2	0.99977777
forecast		forecast r2	0.99991834

Year	Number	% Alive	# churn	P(T = t | alpha, beta)	Log Likelihood	Expected % Alive	Predicted Number	% Error
0	1000	100.00%	0			100.00%		
1	631	63.10%	369	0.373294785	-363.6077522	62.67%	627	0.68%
2	468	46.80%	163	0.152886317	-306.1238875	47.38%	474	1.24%
3	382	38.20%	86	0.085845022	-211.1482047	38.80%	388	1.56%
4	326	32.60%	56	0.055905798	-161.5088818	33.21%	332	1.86%
5	289	28.90%	37	0.039720639	-119.3577214	29.23%	292	1.16%
6	262	26.20%	27	0.029891156	-94.77520132	26.25%	262	0.17%
7	241	24.10%	21	0.023432112	-78.82660559	23.90%	239	0.82%
8	223	22.30%	18	0.018938578		22.01%	220	1.31%
9	207	20.70%	16	0.015674127		20.44%	204	1.25%
10	194	19.40%	13	0.013220555		19.12%	191	1.45%
11	183	18.30%	11	0.011325167		17.99%	180	1.71%
12	173	17.30%	10	0.009827515		17.00%	170	1.71%
>7			241	0.239024171	-344.9169339			

Figure 6: sBG Model on high end cohort with 7 periods of calibration

alpha	0.69867434								calibration MAPE	1.07%	calibration r2	0.999736843
beta	1.171508255								forecast MAPE	1.16%	forecast r2	1
LL	-1850.441163											
validation												
Year	Number	% Alive	# churn	P(T = t | alpha, beta)	Log Likelihood	Expected % Alive	Predicted Num	% Error				
0	1000	100.00%	0			1						
1	631	63.10%	369	0.373586163	-363.319839	62.64%	626	0.73%				
2	468	46.80%	163	0.152484819	-306.5525081	47.39%	474	1.27%				
3	382	38.20%	86	0.085557215	-211.4370155	38.84%	388	1.67%				
4	326	32.60%	56	0.055715655	-161.6996703	33.27%	333	2.04%				
5	289	28.90%	37	0.03959303	-119.476781	29.31%	293	1.41%				
6	262	26.20%	27	0.029803528	-94.85446942	26.33%	263	0.48%				
7	241	24.10%	21	0.023370833	-78.88159656	23.99%	240	0.46%				
8	223	22.30%	18	0.018895228	-71.43922612	22.10%	221	0.90%				
9	207	20.70%	16	0.015643328	-66.52337197	20.54%	205	0.80%				
10	194	19.40%	13	0.013198759	-56.25922247	19.22%	192	0.95%				
11	183	18.30%	11	0.011309959		18.08%	181	1.18%				
12	173	17.30%	10	0.009817211		17.10%	171	1.14%				
>10			194	0.192151442	-319.9974623							

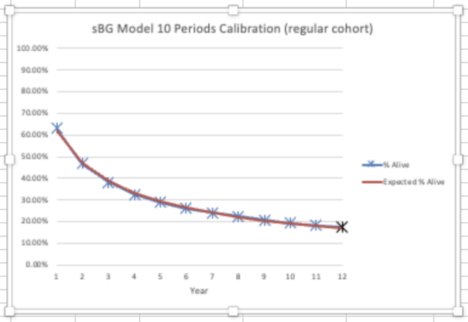


Figure 7: sBG Model on regular cohort with 10 periods of calibration

beta	3.806004788								forecast MAPE	2.83%	forecast r2	0.999702591
LL	-1611.158147											
validation												
Year	Number	% Alive	# churn	P(T = t | alpha, beta)	Log Likelihood	Expected % Alive	Predicted Number	% Error				
0	1000	100.00%	0			1						
1	869	86.90%	131	0.149321382	-249.1167222	85.07%	851	2.11%				
2	743	74.30%	126	0.103819788	-285.4024355	74.69%	747	0.52%				
3	653	65.30%	90	0.077070154	-230.6735264	66.98%	670	2.57%				
4	593	59.30%	60	0.059869528	-168.9352572	60.99%	610	2.85%				
5	551	55.10%	42	0.048084542	-127.4613697	56.18%	562	1.97%				
6	517	51.70%	34	0.039618426	-109.7676725	52.22%	522	1.01%				
7	491	49.10%	26	0.033308895	-88.45020099	48.89%	489	0.43%				
8	468	46.80%	23	0.028466523		46.04%	460	1.62%				
9	445	44.50%	23	0.024659884		43.58%	436	2.07%				
10	427	42.70%	18	0.021607018		41.42%	414	3.00%				
11	409	40.90%	18	0.019116902		39.51%	395	3.41%				
12	394	39.40%	15	0.017056137		37.80%	378	4.06%				
>7			491	0.488907285	-351.3509629							

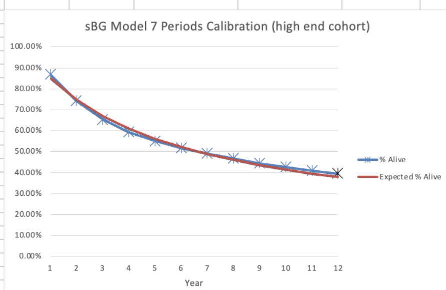


Figure 8: sBG Model on high end cohort with 7 periods of calibration

alpha	0.60517809
beta	3.38870392
LL	-1871.8718
validation	

calibration MAPE	1.53%	calibration r2	0.9965825
forecast MAPE	0.94%	forecast r2	1

Year	Number	% Alive	# churn	P(T = t | alpha, beta)	Log Likelihood	Expected % Alive	Predicted Nu	% Error
0	1000	100.00%	0			1		
1	869	86.90%	131	0.151526282	-247.1965012	84.85%	848	2.36%
2	743	74.30%	126	0.102821353	-286.6200414	74.57%	746	0.36%
3	653	65.30%	90	0.075285512	-232.7820812	67.04%	670	2.66%
4	593	59.30%	60	0.058006602	-170.8319065	61.24%	612	3.26%
5	551	55.10%	42	0.046358829	-128.996428	56.60%	566	2.72%
6	517	51.70%	34	0.038084963	-111.1098157	52.79%	528	2.11%
7	491	49.10%	26	0.031967905	-89.51859368	49.59%	496	1.01%
8	468	46.80%	23	0.027300384	-82.81965385	46.86%	469	0.14%
9	445	44.50%	23	0.02364669	-86.12423941	44.50%	445	0.00%
10	427	42.70%	18	0.020725534	-69.77499841	42.43%	424	0.64%
11	409	40.90%	18	0.018348197		40.59%	406	0.75%
12	394	39.40%	15	0.016383921		38.95%	390	1.13%
>10			427	0.424275946	-366.0975103			

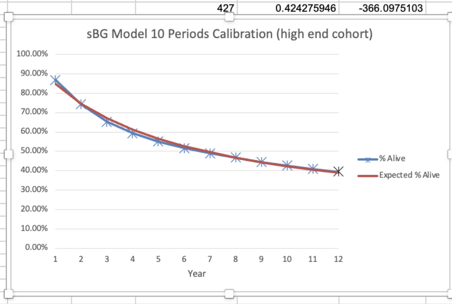


Figure 9: sBG Model on high end cohort with 10 periods of calibration