

Chapter 6

Stochastic Churn Modeling with Dynamic Attribution and Bayesian Estimation



Ping Chou and Howard Hao-Chun Chuang

Abstract Parametric probability models, such as Beta-Geometric, are workhorse models for contractual customer churn prediction. Due to their simplicity, robustness to missing and censored data, and managerially relevant statistics, those models are applied to different business sectors such as healthcare and finance. Nonetheless, the existent models tend to assume a stationary churn process or an identical distribution of latent churn rate. The Beta-Logistic model by Hubbard et al. (Survival prediction-algorithms, challenges and applications. PMLR, pp 22–39 [17]) allows for time-invariant covariates and captures non-identically distributed individual churn. To further accommodate time-varying determinants of churn rate, we apply a Grassia(II)-Geometric (G2G) model by Fader and Hardie (Incorporating time-varying covariates in a simple mixture model for discrete-time duration data [9]). Grounded on the flexible model structure, we propose Bayesian estimation and inference of G2G and empirically assess its prediction performance. Using a workforce dataset from an electronic manufacturing service company, we show that G2G with greater flexibilities outperform extant models in terms of model fitness and employee churn prediction. Additionally, we identify major determinants of churn processes in manufacturing plants and generate cohort-wised survival curves. With built-in interpretability and posterior inference, our Bayesian G2G modeling approach can be useful for churn prediction in marketing and operations management.

Keywords Stochastic modeling · Employee churn modeling · Bayesian estimation

P. Chou (✉)

Department of Management Information Systems, National Chengchi University, Taipei 11605, Taiwan

e-mail: 108356508@nccu.edu.tw

H. H.-C. Chuang

College of Commerce, National Chengchi University, Taipei 11605, Taiwan

e-mail: chuang@nccu.edu.tw

Introduction

Probability models for churn prediction in contractual settings, e.g., Beta-Geometric (BG) [8] and Beta-Weibull [13], have been used in customer retention across business sectors, e.g., online platforms [19], video/music streaming [17], and insurance [27]. The use of such models goes beyond churn prediction and extends to predicting time to purchase [14], patient revisit [20], conversion rate of online advertising [17], and customer-based valuation of business [5, 21]. The models are robust to incomplete information [7], able to leverage interval-censored and non-censored data [10], and provides managerially relevant statistics, e.g., residual lifetime, lifetime value, and their distributions [5, 6, 12]. Despite the proven efficacy, some criticize those models for being covariate-free and hence unable to fully capture group-level and individual-level heterogeneity [22, 23]. Hubbard et al. [17] introduce the Beta-Logistic (BL) model—an extended BG with time-invariant covariates—and empirically improve churn prediction performance in streaming subscribers. While relaxing the restrictive assumption that all individuals share an identical distribution of latent churn rate, BL assumes the heterogeneous distributions to be time-invariant (stationary). That is, BL does not allow for time-varying covariates such as seasonality and tactical marketing activities that are common in many business sectors and result in nonstationary distributions of latent attrition/churn rate [1, 21, 26].

To fill the gap, Fader and Hardie [9] propose a Grassia(II)-Geometric (G2G) model with time-varying covariates, which captures heterogeneous and non-stationary latent churn process. Notwithstanding the theoretical flexibility, G2G's efficacy has not been empirically examined, and it remains unclear whether G2G effectively improves the less generic BG (assuming a common churn rate distribution) and BL (assuming heterogeneous yet stationary churn rate distributions). We collaborate with a leading electronic manufacturing services (EMS) company and assess the G2G for predicting employee churn in its production plants. Due to high turnover rates, individual heterogeneities, and seasonal variations in employee churn risks (e.g., there are more competing offers in peak seasons), the focal EMS requires a flexible yet interpretable model for predicting churn proportion and remaining lifetime for labor planning and hiring in advance. Motivated by the conceptual similarity between customer and employee churn [24, 25], our analysis shows that stochastic churn modeling can be useful in not only consumer services but also workforce operations.

The contribution of this paper is multi-facet. First, we apply Bayesian estimation to BL and G2G, going beyond prior studies' focal interest in churn prediction and instill posterior inference in such models. Second, we empirically show that the more generic and flexible G2G significantly outperforms both BG and BL models. Third, we find that covariate-free BG precisely projects the global survival curves and job durations, whereas G2G effectively captures individual heterogeneity and satisfactorily projects local curves that BG cannot capture. Last, unlike some machine learning approaches that are short of rationales and difficult to obtain test statistics, the proposed Bayesian G2G is highly interpretable with readily available credible regions

of model estimates, such that managers can gauge the significance of predictors and their impact on the metrics of managerial relevance.

General Problem and Models

Contractual Churn and Probability Modeling

Per [11], the problem of churn modeling in general can be categorized into four quadrants by two dimensions: transaction opportunity (continuous/discrete) and relationship type (noncontractual/contractual). This paper focuses on contractual and discrete-time setting that matches the EMS operations, in which employees have an opportunity to decide on whether quitting from production plants at the end of each period. An employee is considered “survival” if he/she stays on-job in the beginning of the next period. Otherwise, he/she is considered “churn”.

Throughout the paper, we use index i to represent individuals, and $T_i \in Z^+$, $C_i \in \{0, 1\}$, and $X_i \in R^d$ refer to one’s duration of survival subject to right-censoring, censor indicator (“1” censored survival, “0” observed churn instance), and d -dimensional covariates. For notational brevity, we use $D = \{T_i, C_i, X_i\}_{i=1}^N$ to denote data with N individuals and $\{\cdot\}$ to represent all model parameters. Mathematically, the objective of stochastic churn modeling is to optimize $\{\cdot\}$, including the coefficients on covariate effects and distributional parameters, such that the overall likelihood function is maximized,

$$P(D|\cdot) = \left\{ \prod_{i:C_i=0} P(T_i|\cdot) \right\} \cdot \left\{ \prod_{i:C_i=1} S(T_i|\cdot) \right\} \quad (6.1)$$

where the first part is the probability mass function $P(T|\cdot)$ for employee churned ($C_i = 0$) and the second part is the survival function $S(T|\cdot)$ for employee survival ($C_i = 1$). Given the estimated model parameters, one can easily apply the survival function $S(T_i|\cdot)$ to infer an individual’s discounted expected residual lifetime (*DERL*) [5]:

$$DERL(d|\cdot, n) = \sum_{t=n}^{\infty} \frac{S(t|\cdot)}{S(n-1|\cdot)} \cdot \left(\frac{1}{1+d} \right)^{t-n} \quad (6.2)$$

where n stands for the number of periods an individual has survived and d denotes a discount rate. The DERL can be adapted to predict an unseen individual’s expected lifetime (*EL*) by setting $n = 1$. That is,

$$EL(d|\cdot) = \sum_{t=1}^{\infty} S(t|\cdot) \cdot \left(\frac{1}{1+d} \right)^{t-1} \quad (6.3)$$

Probability Churn Models from Marketing

Beta-Geometric (BG)

Beta-Geometric (BG) is built on a simple idea assuming that at each period, an individual determines whether to stay active by flipping a two-sided coin with a latent attrition parameter θ (i.e., “head” one ends the contract, “tail” one renews it). The number of trials before showing up of the first head follows a geometric distribution, whereas the probability of attrition θ is heterogeneous across individuals and static over time. Assuming heterogeneous θ follows a beta distribution, the geometric churn process results in the BG model, i.e.,

$$P(T_i|\theta) = \theta(1-\theta)^{T_i-1}, \quad S(T_i|\theta) = (1-\theta)^{T_i}, \quad f(\theta|\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \quad (6.4)$$

Integrating out the latent (θ), the probability mass and survival functions can be re-written as:

$$P(T_i|\alpha, \beta) = \frac{B(\alpha+1, \beta+T_i-1)}{B(\alpha, \beta)}, \quad S(T_i|\alpha, \beta) = \frac{B(\alpha, \beta+T_i)}{B(\alpha, \beta)} \quad (6.5)$$

where α and β are model parameters to be estimated such that the overall likelihood $P(D|\cdot)$ is maximized.

Beta-Logistic (BL)

Unlike BG that assumes a stationary and common beta distribution of latent churn rate θ , BL extends BG by leveraging covariate information and results in stationary yet heterogeneous beta distributions. More specifically, BL makes μ_i (mean) and σ_i^2 (variance) of the latent attrition rate θ functions of individual-specific and time-invariant covariates ($\mathbf{X}_i = [x_{i,1}, \dots, x_{i,d}]$), i.e.,

$$\begin{aligned} \mu_i &= \log \text{it}(\gamma_{\mu,0} + \gamma_{\mu,1}x_{i,1} + \dots + \gamma_{\mu,d}x_{i,d}), \\ \sigma_i^2 &= \exp(\gamma_{\sigma,0} + \gamma_{\sigma,1}x_{i,1} + \dots + \gamma_{\sigma,d}x_{i,d}) \end{aligned} \quad (6.6)$$

Our parameterization (μ, σ^2) differs from (α, β) in [17] for ease of interpretation. Technically, the formulas below allow one to transform (μ_i, σ_i^2) back into (α_i, β_i) , and compute the likelihood using $P(T_i|\alpha_i, \beta_i)$ and $S(T_i|\alpha_i, \beta_i)$.

$$\alpha_i = \left(\frac{1 - \mu_i}{\sigma_i^2} - \frac{1}{\mu_i} \right) \mu_i^2, \quad \beta_i = \alpha_i \left(\frac{1}{\mu_i} - 1 \right) \quad (6.7)$$

Since the covariates in combination with the coefficients ($[\gamma_{\mu,0}, \dots, \gamma_{\mu,d}]$ and $[\gamma_{\sigma,0}, \dots, \gamma_{\sigma,d}]$) fully determine the parameters μ_i and σ_i^2 of latent churn rate distributions, the overall likelihood of BL is conditioned on the coefficients, which leaves them the only parameters to be optimized.

Grassia(II)-Geometric (G2G)

Previous studies have acknowledged that the parametric formulation of BG may fit data poorly sometimes owing to ignoring time-varying factors and duration dependence, e.g., long-living individuals tend to live longer [8, 13]. For such contexts, we need to consider a geometric process with dynamic attrition/churn rate that is generic and flexible. That is,

$$P(T_i|\theta_{i,1}, \dots, \theta_{i,T_i}) = \theta_{i,T_i} \prod_{t=1}^{T_i-1} (1 - \theta_{i,t}), \quad S(T_i|\theta_{i,1}, \dots, \theta_{i,T_i}) = \prod_{t=1}^{T_i} (1 - \theta_{i,t}) \quad (6.8)$$

However, it is impractical to apply separate prior for each $\theta_{i,t}$ and integrate over all the priors. Alternatively, Fader and Hardie [9] propose a Grassia(II)-Geometric (G2G) model that replaces the Beta prior in BG with a Grassia(II) distribution. Their formulations based on the clog-log link and gamma heterogeneity are¹:

$$\theta_{i,t} = 1 - \exp(-\eta\phi_{i,t}), \quad P(\eta) \sim \text{Gamma}(a, b) \quad (6.9)$$

where $\phi_{i,t}$ captures observable heterogeneity and subsumes effects of time-invariant and time-varying covariates over time. Let $\mathbf{X}_i^c = [x_{i,1}^c, \dots, x_{i,d}^c]$ be d -dimensional time-invariant covariates (with effect $\boldsymbol{\gamma}^c$) and $\mathbf{X}_i^{v,t} = [x_{i,1}^{v,t}, \dots, x_{i,d'}^{v,t}]$ be d' -dimensional time-varying covariate at timing t (with effect $\boldsymbol{\gamma}^v$), $\phi_{i,t}$ brings together the stationary component ($\boldsymbol{\gamma}^c \mathbf{x}^c$) and non-stationary component ($\boldsymbol{\gamma}^v \mathbf{x}^{v,t}$):

$$\phi_{i,t} = \exp(\boldsymbol{\gamma}^c \mathbf{X}_i^c + \boldsymbol{\gamma}^v \mathbf{X}_i^{v,t}) = \exp(\gamma_1^c x_{i,1}^c + \dots + \gamma_d^c x_{i,d}^c + \gamma_1^v x_{i,1}^{v,t} + \dots + \gamma_{d'}^v x_{i,d'}^{v,t}) \quad (6.10)$$

¹ a and b denote the shape and rate parameter of *Gamma*.

The above formulation breaks the limitation of stationarity and exclusively time-invariant covariates in BL. Integrating $P(T_i|\theta_{i,1}, \dots, \theta_{i,T_i})$ and $S(T_i|\theta_{i,1}, \dots, \theta_{i,T_i})$ with respect to the latent η , we can derive churn rate ($P(T|\cdot)$) and survival rate ($S(T|\cdot)$), which constitute the likelihood function and are directly conditioned on *Gamma* parameters (a, b) and coefficients (γ^c and γ^v),²

$$\begin{aligned} P(T_i|a, b, \gamma) &= \left\{ \frac{b}{b + \sum_{t=1}^{T_i-1} \phi_{i,t}} \right\}^a - \left\{ \frac{b}{b + \sum_{t=1}^{T_i} \phi_{i,t}} \right\}^a, \\ S(T_i|a, b, \gamma) &= \left\{ \frac{b}{b + \sum_{t=1}^{T_i} \phi_{i,t}} \right\}^a \end{aligned} \quad (6.11)$$

Empirical Methods and Data

EMS Workforce Data

Aimed at examining the model's efficacy and eliciting insights potentially valuable to workforce management practice, we co-work with an anonymous EMS company, Alpha, who provides us with a dataset on 20,000 employees over 145 weeks. In the dataset, in addition to T_i (duration of survival) and C_i (censor status), each employee is characterized by 28 time-invariant covariates, including gender, age, hometown, onboard month, category of recruit/contract, mean salary bonus, ratios working on rest day and working on daytime, and mean working hour. In addition, 16 time-varying covariates such as cumulative weeks of survival and season/month indicators for each period are included for capturing time dynamics in distributions of $\theta_{i,t}$. In Table 6.1, we summarize and explain our operationalization of time-invariant (x_1 to x_{28}) and time-varying covariates (x_{29} to x_{44}).

In Fig. 6.1, we show the distribution of employee's on-job duration (with density rescaled). Employees onboarding in Year 1 (during weeks 1–52) are on the left panel, whereas employees onboarding in Year 2 (during weeks 53–104) are on the right panel. As can be seen, both distributions are right-skewed and have a majority of employees having duration less than 20 weeks. Moreover, both panels exhibit bimodal curves for duration between 0 and 10 weeks and multi-modals for duration longer than 20. The shape implies mixtures of heterogeneous employee cohorts, echoing the proposal of BL/G2G to bring in covariates to explicitly account for cross-sectional heterogeneity which covariate-free BG cannot capture.

² It should be reminded particularly that the coefficient vector does not involve an intercept.

Table 6.1 Summary of covariates

	Covariate	Definition		Covariate	Definition
x_1	Male	1 if male	x_{23}	C-Category (F)	1 if contract category = 'full-time'
x_2	Age (≤ 20)	1 if age ≤ 20	x_{24}	C-Category (D)	1 if contract category = 'dispatch'
x_3	Age (21–30)	1 if $20 < \text{age} \leq 30$	x_{25}	Salary bonus	Mean salary bonus
x_4	Age (31–40)	1 if $30 < \text{age} \leq 40$	x_{26}	Rest day work	Ratio working on rest day
x_5	Area (E)	1 if hometown = 'East'	x_{27}	Shift ratio	Ratio working on daytime
x_6	Area (N)	1 if hometown = 'North'	x_{28}	Work hour	Mean working hour
x_7	Area (NE)	1 if hometown = 'North-East'	x_{29}	Cum. Survival	Cumulative weeks of survival
x_8	Area (NW)	1 if hometown = 'North-West'	x_{30}	Cum. Survival ²	Squared term of Cum. Survival
x_9	Area (SC)	1 if hometown = 'South-Central'	x_{31}	Is Spring	1 if next period is in Spring
x_{10}	Onboard (1)	1 if onboard month = '1'	x_{32}	Is Summer	1 if next period is in Summer
x_{11}	Onboard (2)	1 if onboard month = '2'	x_{33}	Is Autumn	1 if next period is in Autumn
x_{12}	Onboard (3)	1 if onboard month = '3'	x_{34}	Is January	1 if next period is in January
x_{13}	Onboard (4)	1 if onboard month = '4'	x_{35}	Is February	1 if next period is in February
x_{14}	Onboard (5)	1 if onboard month = '5'	x_{36}	Is March	1 if next period is in March
x_{15}	Onboard (6)	1 if onboard month = '6'	x_{37}	Is April	1 if next period is in April
x_{16}	Onboard (7)	1 if onboard month = '7'	x_{38}	Is May	1 if next period is in May
x_{17}	Onboard (8)	1 if onboard month = '8'	x_{39}	Is June	1 if next period is in June
x_{18}	Onboard (9)	1 if onboard month = '9'	x_{40}	Is July	1 if next period is in July
x_{19}	Onboard (10)	1 if onboard month = '10'	x_{41}	Is August	1 if next period is in August
x_{20}	Onboard (11)	1 if onboard month = '11'	x_{42}	Is September	1 if next period is in September
x_{21}	R-Category (F)	1 if recruit category = 'full-time'	x_{43}	Is October	1 if next period is in October

(continued)

Table 6.1 (continued)

	Covariate	Definition		Covariate	Definition
x_{22}	R-Category (D)	1 if recruit category = 'dispatch'	x_{44}	Is November	1 if next period is in November

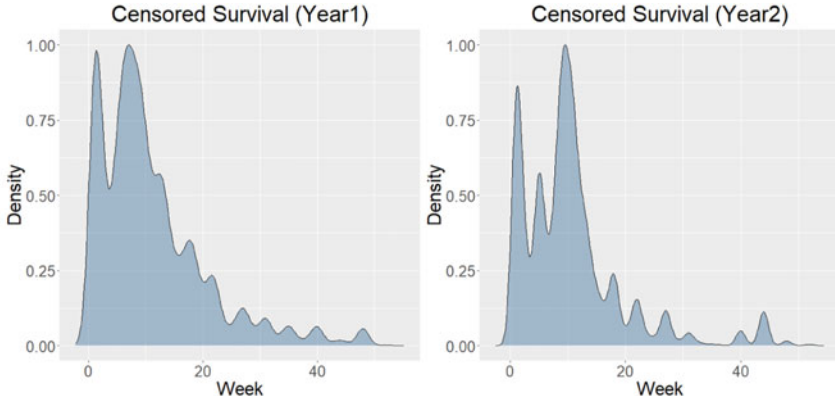


Fig. 6.1 Density plots of survival (on-job) duration

Bayesian Estimation and Inference

For BL and G2G with covariates, to construct the posterior distributions of covariate effects γ from Bayesian inference, we adopt Gibbs sampling—a Markov Chain Monte Carlo (MCMC) method for model calibration (see [4] for an extensive review). Based on a preliminary test on convergence and distribution of covariate effects, we adopt independent normal priors [15]³ with identical parameters. Specifically, we set $N(0, 0.2)$ for G2G and $N(0, 0.5)$ for BL. Based on the simulated posterior distributions, we estimate posterior modes (i.e., *maximum a posterior*, MAP) and the credible regions of covariate effects. Let $N(\cdot)$ be the normal density function, and define collection $\{\cdot\} = \{\gamma_j, \gamma_{-j}\}$, where γ_j denotes the j th coefficient being updated and γ_{-j} be all the other coefficients, the full conditional distribution $P(\gamma_j | \gamma_{-j}, D)$ can be simply written as:

$$P(\gamma_j | \gamma_{-j}, D) \propto P(D | \cdot) \times N(\gamma_j) \tag{6.12}$$

We calibrate the models on 2000 randomly sampled employees who start their jobs in Year 1, whose survival (employment) durations are censored at week 52. For the models, we run 2000 MCMC iterations and discard the first 1000 draws

³ In the preliminary analysis, we found that improper prior such as uniform may lead to unstableness in optimization.

as burn-in samples. On the other hand, we select employees who join in Year 2 as test samples for the sake of assessing out-of-sample prediction performance. Survival durations for those in test set who survive after week 104 are censored. We apply log-likelihood (LL), Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) to assess the models' fitness, and use C-statistics [16] to evaluate the models' performance on projecting hold-out employees' job duration:

$$C = \frac{\sum_{i \neq j} \{I(T_i > T_j) \cdot I(EL_i > EL_j)\} \cdot I(C_j = 0)}{\sum_{i \neq j} I(T_i > T_j) \cdot I(C_j = 0)} \quad (6.13)$$

of which EL denotes the expected lifetime assuming no discount (i.e., $d = 0$).

Data Analysis

We first assess the computational efficiency that is critical for model adoption and application. In Table 6.2, we report the CPU time it takes to calibrate the foregoing models. We implement the code in R, and execute the programs on a platform with a Win10 operating system, an Intel i7-1165G7 processor, and a 16 GB RAM. Compared with the 2-parameter BG with gradient-based optimization, which takes 35.67 s for calibration, MCMC-based BL (with 30 parameters) and G2G (with 44 parameters) take 701.31 and 10,009.55 s (equivalently, 19.66 and 280.62 times of BG's execution time). Not surprisingly, the most sophisticated G2G—with time-varying and individual-specific latent churn rate—comes at higher computing cost. Nonetheless, given its unique flexibility in modelling heterogeneous and nonstationary churn distributions, we posit the estimation cost induced by generic formulations is acceptable and affordable in modern computing. We then formally assess the performance gains of G2G.

Table 6.3 reports the fitness and performance of the foregoing models (i.e., LL, AIC, BIC based on training samples as well as LL and C-index on hold-out samples). Taking BG as our benchmark, we find that modeling time-invariant covariates makes BL rather flexible and able to capture cohort-level stationary heterogeneity, leading to better fitness on in- and hold-out samples. G2G with time-invariant and time-varying covariates enhances performance further and outperforms BL. The improvement of G2G over BL is more salient than that of BL over BG, evidencing the non-trivial

Table 6.2 Computational cost

	# of parameters	Method	CPU time (s)	Normalized time
BG	2	Gradient	35.67	1
BL	30	MCMC	701.31	19.66
G2G	44	MCMC	10,009.55	280.62

Table 6.3 Performance evaluation

	Year 1: in-sample (N = 2000)			Year 2: out-of-sample (N = 11,673)	
	LL	AIC	BIC	LL	C-Statistic
BG	-6360.33	12,724.66	12,735.86	-34,875.05	0.500
BL	-5741.50	11,542.99	11,711.02	-33,179.62	0.826
G2G	-5031.14	10,150.28	10,396.72	-31,235.94	0.832

value of accounting for time-varying covariates and nonstationary heterogeneity distributions of latent churn.

As for projecting employees' survival, BL and G2G achieve C-statistics of 0.826 and 0.832, respectively. These are high values considering the high uncertainty of employee churn data and simplicity of the models' parametric formulations [28]. Because the expected lifetimes that BG produces are not differentiable between individuals, leading to a C-statistic of 0.500, we apply the same protocol to train a survival forest [18]. The survival forest achieves a C-statistic of 0.835. This finding implies that simple parametric yet interpretable formulations do not necessarily undermine prediction accuracy.

Given the significantly better model fitness and prediction performance of G2G, in Table 6.4, we report the MAP effect estimates and 90% credible regions of covariates [2]. For brevity, we report only the top-10 covariates based on their absolute effect size in descending order.

As expected, *Salary Bonus* as a motivating factor outranks other covariates and reduces employee churn. The ratio working on daytime (*Shift Ratio*) tells if one often works on normal shift hours, thus reducing employee churn. Full-time (*R-Category (F)*) and dispatch job categories (*R-Category (D)*), compared to part-time job (the baseline category), are positions seeking regular and long-term workers. Therefore, employees who have full-time and dispatch jobs are less likely to churn. Negative

Table 6.4 Effect estimates and credible regions of G2G estimates

	Covariate	Estimates (MAP)	Credible region
1	Salary bonus	-3.995	(-4.24, -3.52)
2	Work hour	-2.526	(-2.64, -2.35)
3	Rest day work	-1.775	(-2.30, -1.41)
4	Shift ratio	-1.620	(-1.84, -1.37)
5	R-category (F)	-1.063	(-1.31, -0.70)
6	R-category (D)	-1.048	(-1.23, -0.83)
7	Age (≤ 20)	-0.806	(-0.95, -0.55)
8	Age (21-30)	-0.727	(-0.87, -0.55)
9	Age (31-40)	-0.607	(-0.77, -0.37)
10	Is Summer	-0.518	(-0.69, -0.36)

relationship between age and churn rate (*Age* (≤ 20), *Age* (21–30), *Age* (31–40)) makes sense in that physically fit younger workers are generally better suited for labor-intensive tasks.

Some of the estimated effects are counterintuitive: (a) Why are employees having longer mean working hour (*Work Hour*) and having higher ratio working on rest day (*Rest Day Work*) less likely to churn? (b) Additionally, why is the latent churn rate of an employee lower, when the upcoming period is in Summer (*Is Summer*)? An unpublished interview by the authors' colleague reveals that the majority of Alpha's employees are job-hoppers that seek over-time working for extra bonus, thus making (a) reasonable. For (b), it reveals that salary level for the industry is lower during May to June, explaining why employees are unlikely to churn in Summer.

The above discussion focuses specifically on the model fitness and the covariate effects. Below we examine if the estimated effects consistently reflect survival curve at aggregate-level. In Fig. 6.2, we project survival curves using the calibrated BG and G2G models. In addition to the global curve (on the left panel), we take *Salary Bonus* with the strongest effect as an illustrative example to divide the sample into two sub-groups. We project two local curves, i.e., one with bonus below-median (on the middle) and another with bonus above-median (on the right).

For the full sample, except for a few shocks uncaptured, overall both BG and G2G fit the global curve fairly well (with mean discrepancy less than 4%). The result is surprising given the simplicity of BG as we introduced in Sect. 2.2.1. As for the two groups, despite the larger discrepancy between data and model projections, G2G does a decent job at characterizing the two local curves. Arguably, BG will be useful and robust if a decision-maker is only interested in inferring the curve reflecting global survival rate. Nonetheless, the proposed Bayesian G2G will be a more effective and appropriate technique for one to predict the survival rate/remaining job duration of an individual or a specific cohort over time.

To offer managers practical insights in workforce planning, we project survival curves for each individual employee, and by pairwise CORT dissimilarity [3] and hierarchical clustering we cluster the curves. In Fig. 6.3, we show the aggregated

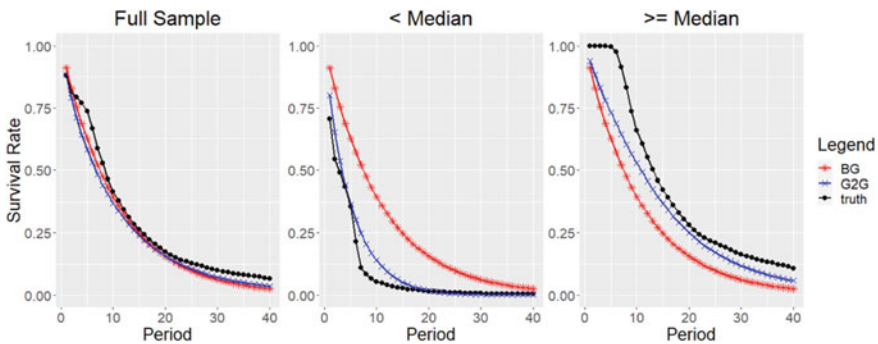


Fig. 6.2 The survival curves

Fig. 6.3 Survival curves clustering

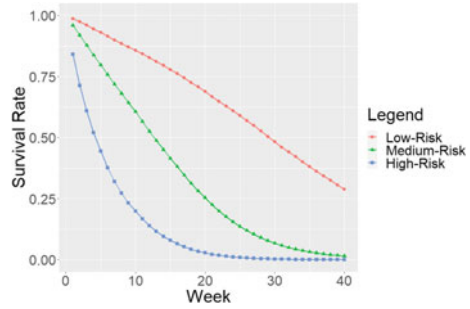


Table 6.5 Averages of features in all clusters

	High-risk	Medium-risk	Low-risk
Salary bonus	0.043	0.244	0.591
Work hour	0.768	0.798	0.807
Rest day work	0.440	0.474	0.469
Shift ratio	0.289	0.526	0.512
R-category (F)	0.110	0.185	0.282
R-category (N) ⁵	0.890	0.815	0.718
Age (≤ 20)	0.100	0.129	0.148
Age (21–30)	0.725	0.723	0.657
Age (> 30)	0.175	0.148	0.195

survival curves,⁴ and summarize their mean characteristics in Table 6.5. We identify three clusters: *High-Risk*, *Medium-Risk*, and *Low-Risk*. *High-Risk* has more of 20s and dispatch workers, who seldom work overtime and in irregular hour for extra bonus. *Medium-Risk* is younger, while having more full-time workers and extra working hours and bonus. The *Low-Risk*, while being the oldest and having fewer shifts on daytime and rest day, has the highest proportion of full-time workers and the most working hours and bonus. Clearly, the clusters are high in some drivers of survival tendencies but low in the others. For churn anticipation and retention, we suggest one to focus on the factors strictly increasing or decreasing survival risk (e.g., *Salary Bonus*), or to focus on the combination of drivers (e.g., high *Salary Bonus* and low *Shift Ratio*) for the sake of developing countermeasures against employee churn.

⁴ For each of the clusters, we produce the aggregated survival curve by taking average over all member curves in the cluster.

⁵ For ease of interpretation, we denote dispatch and part-time jobs by *R-Category (N)*.

Conclusion

Marketing modelers have invented a lasting stream of contractual churn models and exerted their influences over applications outside the realm of marketing in the past two decades. However, numerous models ignore time-varying covariates by assuming stationary and identical distributions of latent attrition. Such models are thus restrictive and not flexible enough for tenure duration predictions in many business sectors and problem settings. Fader and Hardie [9] propose G2G that offers a generic structure for incorporating time-varying covariates. Instilling Bayesian inference and estimation into G2G, this paper empirically assesses the model's efficacy and the value of time-varying covariates in predicting employee churn, which shares a great similarity with customer churn.

Our analysis of workforce data from manufacturing plants shows that stochastic churn modeling from marketing can be useful for operations management. We find that G2G allowing for nonstationary heterogeneity distributions and time-varying covariates has the best performance in model fitness and survival rate projection. Furthermore, we identify the drivers of employee churn, including factors that are immediately comprehensible (e.g., upcoming seasonality, recruitment category, and age) and counterintuitive ones (mean working hour and working on rest day). When project survival curves globally and locally, the consistently decent performance of G2G suggests that managers can apply G2G to predict churn proportion and remaining lifetime for labor planning and hiring in practice.

Notwithstanding the promising result, our modeling effort is not without limitations and leaves several directions for continual explorations. First, we encourage subsequent studies to triangulate our findings using datasets from different firms and industries. Such efforts will enhance the validity of our Bayesian G2G modeling approach. Second, the fine-tuning of Gamma prior and the MCMC calibration of covariate effects is computationally intensive. Hence, computationally efficient optimization method of G2G deserves more investigations. Third, in this paper, we assume homogeneous employee value and calibrate the models accordingly. Models aimed at delivering maximum workforce performance (e.g., working output and return-on-investment ratio) should take heterogeneous employee value into account. In sum, the probability models provide unique value on prediction and inference front, and G2G having more realistic assumptions is a promising framework in theory and practice of stochastic churn modeling.

References

1. Bachmann, P., Meierer, M., & Näf, J. (2021). The role of time-varying contextual factors in latent attrition models for customer base analysis. *Marketing Science*, 40(4), 783–809.
2. Besag, J., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, 3–41.

3. Chouakria, A. D., & Nagabhushan, P. N. (2007). Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis and Classification*, 1(1), 5–21.
4. Dunson, D. B., & Johndrow, J. E. (2020). The Hastings algorithm at fifty. *Biometrika*, 107(1), 1–23.
5. Fader, P. S., & Hardie, B. G. (2010). Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science*, 29(1), 85–93.
6. Fader, P. S., & Hardie, B. G. (2017). Exploring the distribution of customer lifetime value (in contractual settings). http://www.brucehardie.com/notes/035/distribution_of_CLV_contractual.pdf
7. Fader, P. S., & Hardie, B. G. (2009). Fitting the sBG model to multi-cohort data. http://www.brucehardie.com/notes/017/sBG_estimation.pdf
8. Fader, P. S., & Hardie, B. G. (2007). How to project customer retention. *Journal of Interactive Marketing*, 21(1), 76–90.
9. Fader, P. S., & Hardie, B. G.: Incorporating time-varying covariates in a simple mixture model for discrete-time duration data. http://www.brucehardie.com/notes/037/time-varying_covariates_in_BG.pdf
10. Fader, P. S., & Hardie, B. G. (2020). Musings on fitting the P(II) distribution to single-event timing data. http://www.brucehardie.com/notes/038/musings_on_pareto-ii_parameter_estimation.pdf
11. Fader, P. S., & Hardie, B. G. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing*, 23(1), 61–69.
12. Fader, P. S., & Hardie, B. G. (2018). The mean and variance of customer lifetime value in contractual settings. http://www.brucehardie.com/notes/036/mean_and_var_of_CLV_in_contractual_settings.pdf
13. Fader, P. S., Hardie, B. G., Liu, Y., Davin, J., & Steenburgh, T. (2018). “How to project customer retention” revisited: the role of duration dependence. *Journal of Interactive Marketing*, 43, 1–16.
14. Fader, P. S., Hardie, B. G., McCarthy, D., & Vaidyanathan, R. (2019). Exploring the equivalence of two common mixture models for duration data. *The American Statistician*, 73(3), 288–295.
15. Ghosh, J., Li, Y., & Mitra, R. (2018). On the use of Cauchy prior distributions for Bayesian logistic regression. *Bayesian Analysis*, 13(2), 359–383.
16. Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA*, 247(18), 2543–2546.
17. Hubbard, D., Rostykus, B., Raimond, Y., & Jebara, T. (2021). Beta survival models. In *Survival prediction-algorithms, challenges and applications* (pp. 22–39). PMLR.
18. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, 2(3), 841–860.
19. Kumar, S. (2018). Long Wix.com: Sohn investment idea contest entry (NASDAQ:Wix). <https://seekingalpha.com/article/3976965-long-wix-com-sohn-investment-idea-contest-entry>
20. Lee, K. L., Fader, P., & Hardie, B. (2007). How to project patient persistency. *Foresight: The International Journal of Applied Forecasting*, 8, 31–35.
21. McCarthy, D. M., Fader, P. S., & Hardie, B. G. (2017). Valuing subscription-based businesses using publicly disclosed customer data. *Journal of Marketing*, 81(1), 17–35.
22. Park, C. H. (2012). Essays on shopping dynamics in customer base analysis. <https://ecommons.cornell.edu/bitstream/handle/1813/30999/cp247.pdf>
23. Pfeifer, P. E. (2011). On estimating current-customer equity using company summary data. *Journal of Interactive Marketing*, 25(1), 1–14.
24. Ribes, E., Touahri, K., & Perthame, B. (2017). Employee turnover prediction and retention policies design: A case study. arXiv preprint [arXiv:1707.01377](https://arxiv.org/abs/1707.01377)
25. Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, 38(3), 1999–2006.
26. Schweidel, D. A., Fader, P. S., & Bradlow, E. T. (2008). Understanding service retention within and across cohorts using limited information. *Journal of Marketing*, 72(1), 82–94.

27. Soleymanian, M., Weinberg, C., & Zhu, T. (2021). 'Privacy concerns, economic benefits, and consumer decisions: a multi-period panel study of consumer choices in the automobile insurance industry'. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3905034
28. Wang, L., Li, Y., & Chignell, M. (2021). Combining ranking and point-wise losses for training deep survival analysis models. In *2021 IEEE international conference on data mining (ICDM)*, pp. 689–698. IEEE.