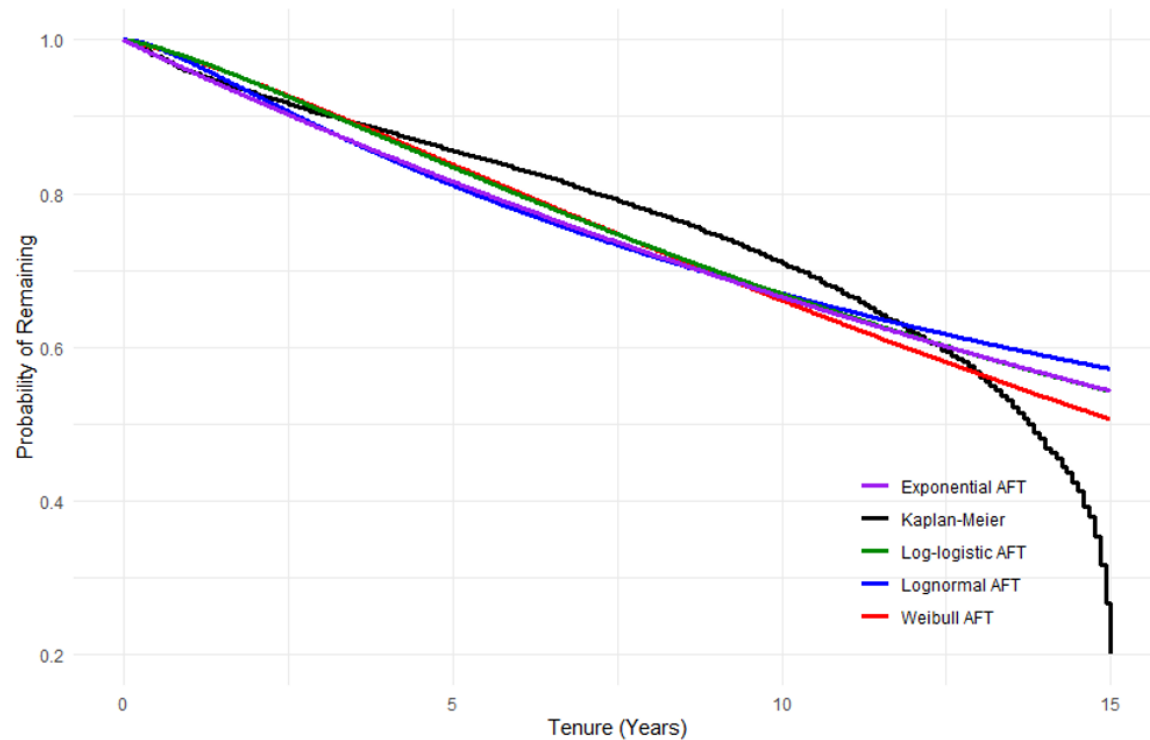




"Survival of the Fittest"

Survival Models: AFT and CPH Models

Sam Kloese, ACAS, CSPA
Manager, Model Review Team
October 27, 2026



NAIC Model Review Team

- State regulators can refer P&C rate models to us
 - State DOI must sign Rate Support Agreement
 - Guarantees confidentiality
 - Services are free to state insurance regulators
 - We assist with the review, but do not make the final decision
 - Reports made available to other states who also signed agreement
 - NAIC requires that filings contain comprehensive documentation

Predictive Model Training

- GLM Training Course available for Regulators
 - “Generalized Linear Model (GLM)” on Compass
- Prior Book Clubs:
 - October 2020: Gradient Boosting Machines
 - <https://www.youtube.com/watch?v=byLGB9ODOxY>
 - April 2021: From GLMs to GAMs
 - <https://youtu.be/vRbHqbNINx8>
 - December 2021: Review of Random Forest Models
 - https://youtu.be/Lm_hf8UHM
 - April 2024: Non-GLM Model Documentation
 - https://www.youtube.com/watch?v=SwB_E3wAB3Y

Introduction

- Today's focus
 - Introduce the topic of survival models
 - Focus on the technical aspects of reviewing survival models
- Survival models do not model insurance claim cost
- Appropriate uses of survival models might be a good future CASTF topic

Introduction

- Survival analysis is synonymous with “time-to-event” modeling. Examples:
 - Time until patients die from a disease
 - Time until unemployed individuals secure a new job
 - Time until batteries need to be replaced
 - Time until an insurance policyholder cancels their policy
- Time-to-event modeling requires different tools
 - Times are always positive
 - Data is “censored” (we don’t know the full-time period for all observations)
- Open Source Tools
 - R: survival
 - Python: lifelines

Data Censorship Issues

- Our typical modeling approaches don't work
 - Censorship skews typical statistics like minimum, average, and maximum
 - Regression models minimize the error terms
 - For censored data, we don't know the error terms
- Survival models are equipped to handle data with some censored records
 - Censorship rule of thumb: 10 events per variable (EPV)*
 - If the 10 EPV rule is violated
 - Lengthen the observed time period (collect more events)
 - Reduce the number of variables

*Peduzzi, P., Concato, J., Feinstein, A.R., & Holford, T.R. (1995). Importance of events per independent variable in proportional hazards regression analysis. *Journal of Clinical Epidemiology*, 48(12), 1503-1510.

Types of Data Censorship

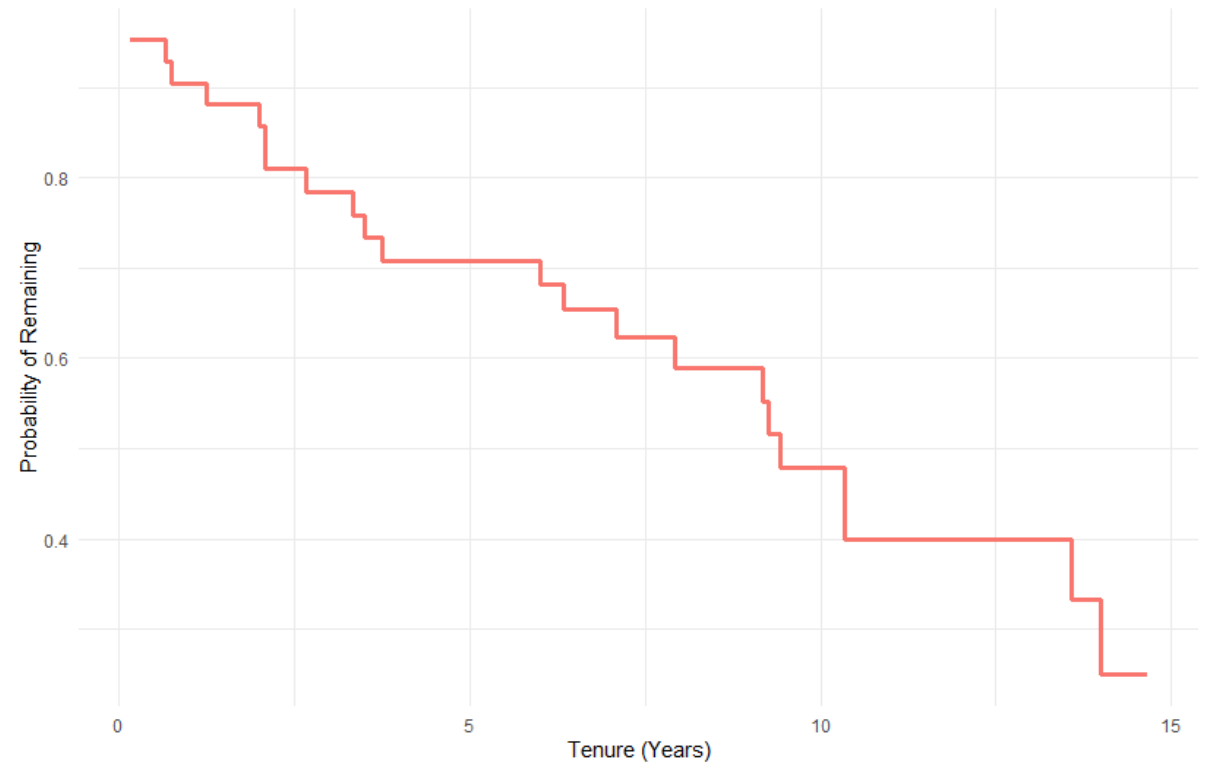
- Uncensored - The event occurred and we know the duration time
- Right-censored - The event did not occur, so the duration time must be greater than or equal to the observed time.
- Left-censored - The event occurred, but we don't have a defined start time.
- Interval-censored - The event occurred within a certain range, but not exactly known.

Right Censorship with Policy Duration

- Right censorship applies to insurance policy duration
- Example:
 - 1,000 policies written between 1/1/2021 and 12/31/2025
 - Some policies will cancel within the period, these durations are known
 - The remaining policies will still be active on 12/31/2025, these are “right-censored”
 - A policy written up to 12/30/2025 would be included in this dataset, and provides some signal
- Left and Interval censorship generally don’t apply to insurance policy duration
 - We generally have effective dates and cancellation dates (if applicable) for every policy

Survival Model Curves

- $S(t) = \Pr(T > t)$; The probability that the event occurs happens after time t
- $S(0) = 1$
- $S(t)$ gets smaller as t increases



Survival Models

Abbreviation	Model	Type
KM	Kaplan-Meier	Non-parametric (No parameters, based on observations)
AFT	Accelerated Failure Time	Parametric (Parameters for intercept and co-variates)
CPH	Cox Proportional Hazard	Semi-parametric (no parameter for base hazard)

Kaplan-Meier

Kaplan-Meier (KM)

- Also known as “the product-limit” estimator
- Does not assume any underlying distribution
- Easily calculated directly from the data
- Survival function $S(t)$
 - t_i : a duration time
 - d_i : number of events that happened at time t_i
 - n_i : number of individuals that survived up to time t_i
- Often KM is not the proposed model, but a reference for comparison.

$$S(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i} \right)$$

Kaplan-Meier (KM)

- Example: Start with a table of all policies effective 1/1/2021 - 12/31/2025

Policy Number	Effective Date	Years	Cancellation	Cancellation Date
001	1/1/2021	2	1	1/1/2023
002	1/1/2021	5	1	1/1/2026
003	1/1/2022	5	0	NA
004	1/1/2023	3	1	1/1/2026
005	1/1/2025	2	0	NA

- Arrange durations in increasing order

Policy Number	Effective Date	Years	Cancellation	Cancellation Date
001	1/1/2021	2	1	1/1/2023
005	1/1/2025	2	0	NA
004	1/1/2023	3	1	1/1/2026
003	1/1/2022	5	0	NA
002	1/1/2021	5	1	1/1/2026

Kaplan-Meier (KM)

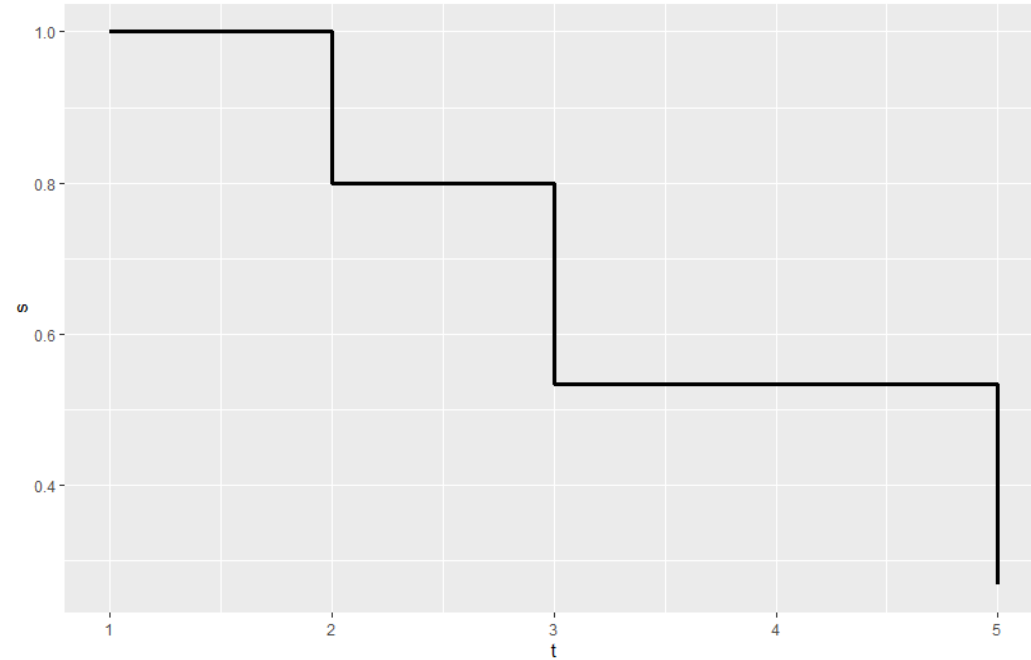
Policy Number	Effective Date	Years	Cancellation	Cancellation Date
001	1/1/2021	2	1	1/1/2023
005	1/1/2025	2	0	NA
004	1/1/2023	3	1	1/1/2026
003	1/1/2022	5	0	NA
002	1/1/2021	5	1	1/1/2026

$$S(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

- Calculate $S(t_i)$

t_i	d_i	n_i	$(1 - \frac{d_i}{n_i})$	$S(t_i)$
1	0	5	5/5	5/5 = 100%
2	1	5	4/5	5/5*4/5 = 80%
3	1	3	2/3	5/5*4/5*2/3 = 53%
4	0	2	2/2	5/5*4/5*2/3*2/2 = 53%
5	1	2	1/2	5/5*4/5*2/3*2/2*1/2 = 27%

Kaplan-Meier (KM)



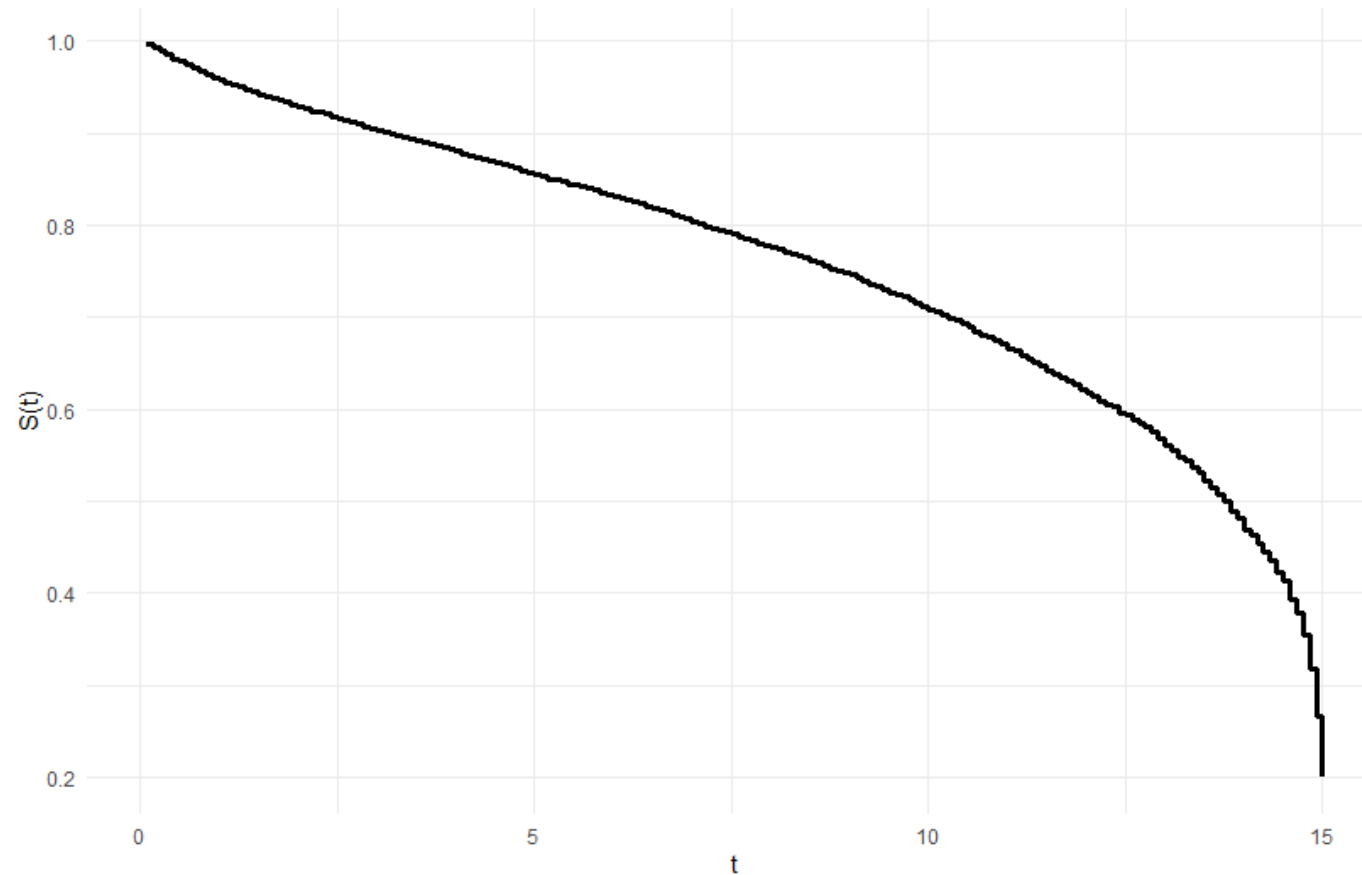
- Calculate $S(t_i)$

t_i	d_i	n_i	$(1 - \frac{d_i}{n_i})$	$S(t_i)$
1	0	5	5/5	5/5 = 100%
2	1	5	4/5	5/5 * 4/5 = 80%
3	1	3	2/3	5/5 * 4/5 * 2/3 = 53%
4	0	2	2/2	5/5 * 4/5 * 2/3 * 2/2 = 53%
5	1	2	1/2	5/5 * 4/5 * 2/3 * 2/2 * 1/2 = 27%

Data for future examples

- Synthetic insurance churn dataset:
<https://www.kaggle.com/datasets/andrewdiachkov/realistic-synthetic-insurance-churn-dataset-for-ml?resource=download>
- **Caveats**
 - Data for illustration purposes only
 - Data is synthetic, unsure if data is realistic
 - Using data in a way that the contributor didn't intend
 - Data was built for sample binary classification models (calculate probability on cancellation)
 - I am taking "customer tenure", an intended predictor, and using it in the target variable
- **Modifications**
 - Filter for "Auto"
 - Further grouped age bands into under 35, 35-54, and 55+

Data for future examples



Accelerated Failure Time

Accelerated Failure Time

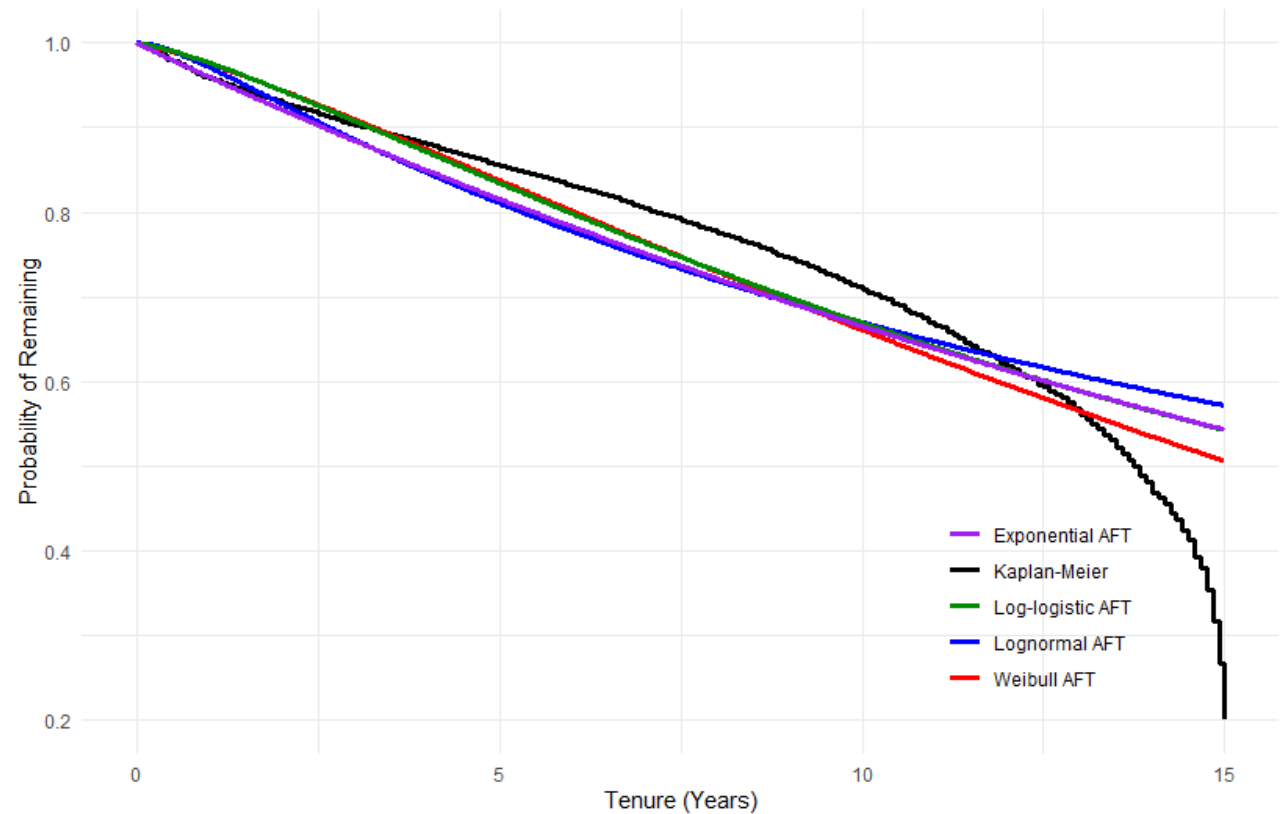
- Different risk classes are speeding up or slowing down in their rate to approach t
- This acceleration/deceleration can be explained by model covariates
- Acceleration example: Every human year is worth 7 dog years
 - $S_{dog}(t) = S_{human}(t * 7)$
- Produces smooth survival curves, not step functions
- Allows for multiple assumed distributions
 - Weibull
 - Log Normal
 - Log Logistic
 - Exponential

Accelerated Failure Time

- Selecting a preliminary distribution
 - Plot KM versus smoothed AFT curves
 - Cox-snell plots
 - Compare AIC

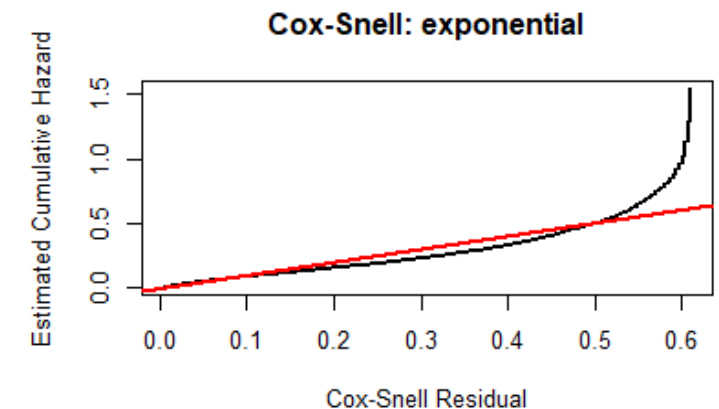
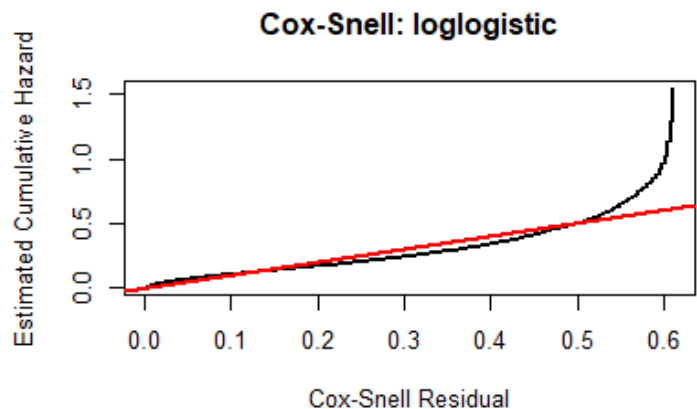
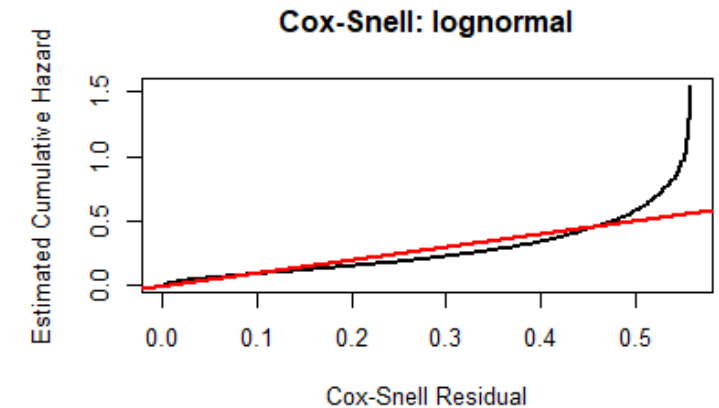
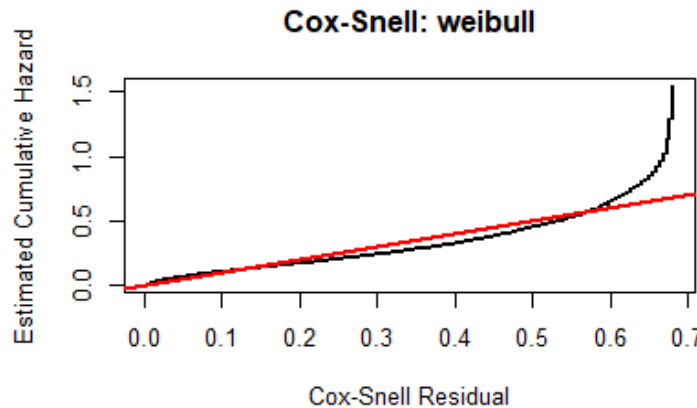
KM vs. AFT Curves

- The AFT models have a similar shape
- Weibull seems to be the “least incorrect” for $t > 12$



Cox-snell plots

- Distribution is better when the curve follows the diagonal reference line
- No clear winner based on these plots alone



Compare AFT AIC values

Model	AIC
Weibull	41,029
Exponential	41,279
Log-logistic	41,443
Lognormal	42,097

- Lower AIC is better (similar to GLMs)
 - Weibull is best based on this analysis
- *AIC should be checked again after finalizing predictor variables

AFT Model Output

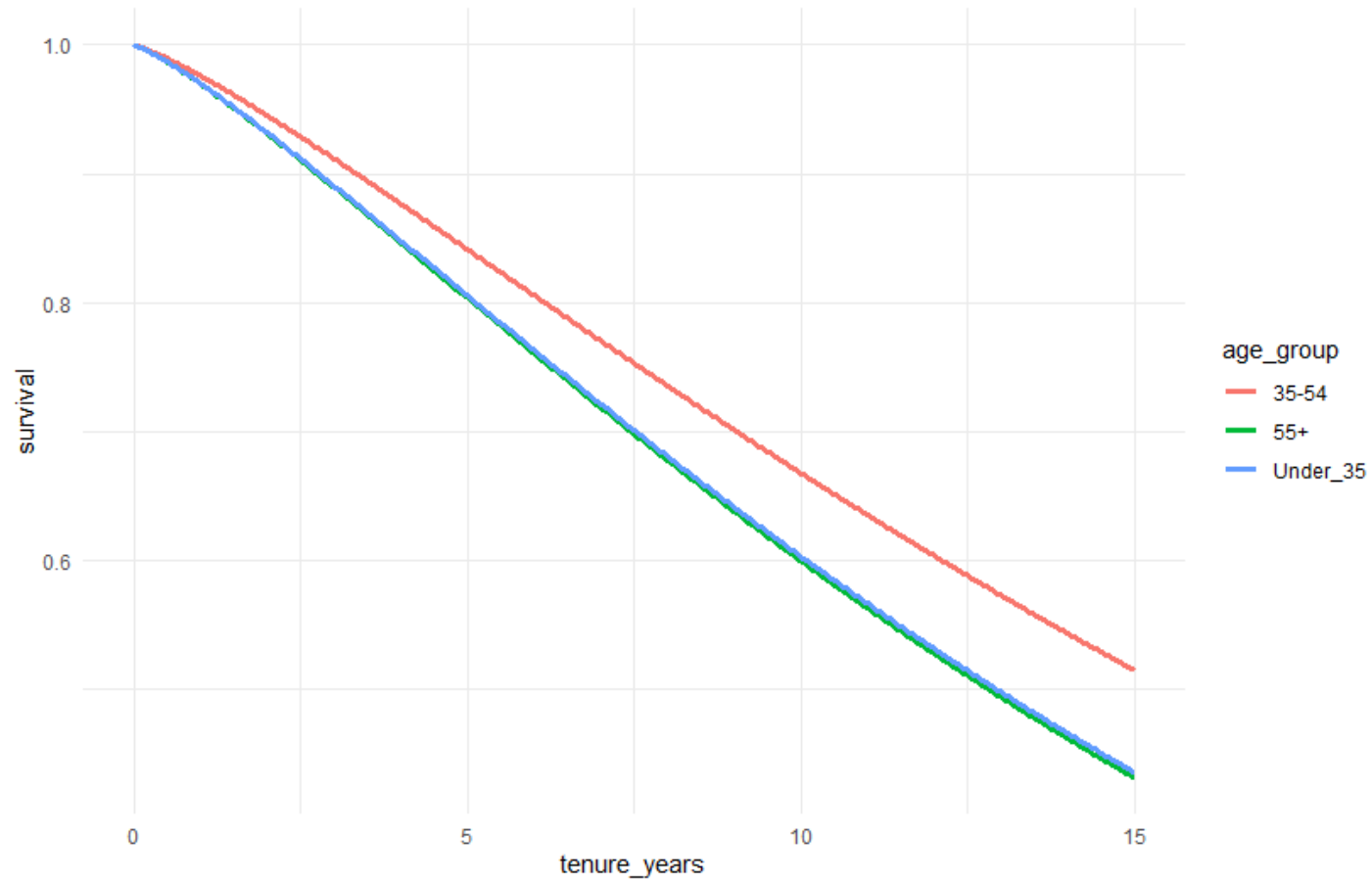
- The T in “AFT” stands for “Time”. Positive coefficients mean longer durations.
- P-values have a similar interpretation to GLMs
- AFT is fully parametric, it includes an intercept term

Variable	Value	P-value
(Intercept)	3.041	<0.001
Age Group: 55+	-0.193	<0.001
Age Group: Under 35	-0.183	<0.001
Marital Status: Single	-.016	0.498
Multi Policy Flag	0.670	<0.001
Complaint Flag	-0.549	<0.001

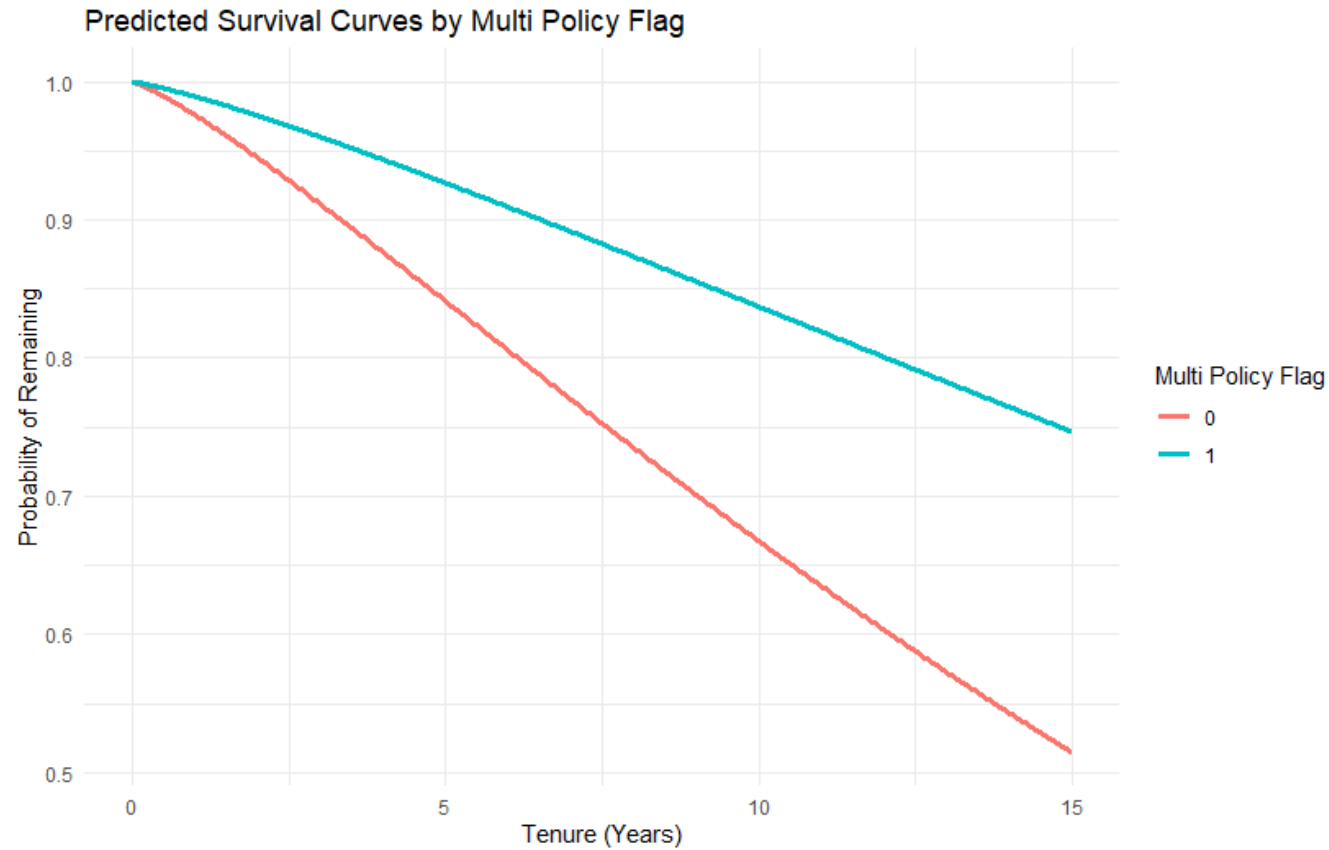
Survival Model Validations

- Survival Curves by Variable (each level plotted separately)
- Survival Curves for a high-retaining class and low-retaining class
- C-Index value

Survival Curves by Variable



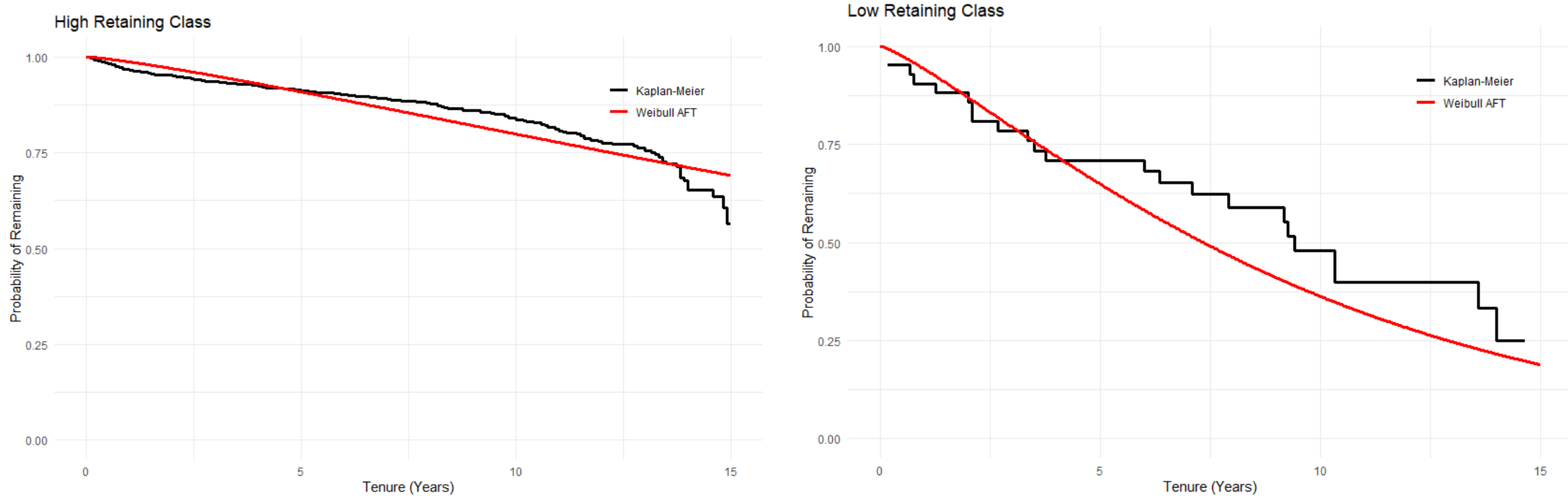
Survival Curves by Variable



Survival Curves for High/Low Retaining Classes

- Identify a high retaining class and a low retaining class
 - Easily determined by looking for classes with higher/lower coefficients
 - Pick classes with sufficient data to calculate a reasonable KM estimator
- Shows that the model is producing materially different survival curves
 - Make sure the y-axis is consistent (0 to 100%) for easy comparison
- Plot survival curve against the KM estimator for each group
 - Calculate a different KM estimator for each class by filtering for the specific class
 - The AFT curve should look similar to a smoothed version of the KM stepped line
 - Shows how well the smoothed model reflects actual experience
- Provide these plots based on holdout data

Survival Curves for High/Low Retaining Classes



- High Retaining Class: 55+, married, multi policy, no complaints
- Low Retaining Class: Under 35, single, single policy, complaint(s) made

C-Index

- Analogous to the AUC in binary classification
- Higher values represent better segmentation power
- "If I randomly pick two policyholders, what percentage of the time does the model correctly rank who cancels first?"
- Rules of thumb*:

C-Index Value	Interpretation
0.50	No better than chance
0.50-0.70	Poor to moderate discrimination
0.70-0.80	Reasonable threshold for survival models
>0.80	Good to excellent discrimination
1.00	Perfect discrimination

- My Weibull AFT C-Index: 0.588

Statistical Tests by Variable

- P-values from model output
- AIC with the full model and excluding one variable at a time
- C-Index with the full model and excluding one variable at a time (on holdout data)

Model	C-Index	AIC
Full Model	0.588	40,314
Model excluding age group	0.576	40,362
Model excluding marital status	0.588	40,312
Model excluding multi policy	0.538	40,856
Model excluding complaint flag	0.579	40,441

Cox Proportional Hazard

Cox Proportional Hazard

- Hazard function $h(t)$: describes the probability that event happens at some time, given survival up to that time.

$$h(t) = -\frac{d}{dt} \log S(t)$$

- The hazard function $h(t)$ can be derived from the survival function $S(t)$, and vice versa.
- Assumption: all individuals' hazards are proportional to one another.
 - A risk's hazards is a function of a baseline hazard and scaling factors
 - The relative survival impact between risks does not change over time
- Semi-parametric
 - Parameters for the covariates, but not the base hazard
 - Less strict distributional assumptions

Cox Proportional Hazard

- Based on the proportional hazards assumption:

$$h(t|x) = b_0(t) \exp\left(\sum_{i=1}^n b_i(x_i - \overline{x_i})\right)$$

- $b_0(t)$: baseline hazard function that changes with time
 - Baseline is NOT just the case where all covariates = 0
 - Baseline is based on covariate averages (or median) depending on the software
- $\exp(\dots)$: the linear relationship between covariates and the log of hazard

Cox Proportional Hazard

- R can provide a test of the proportional hazards assumption with `cox.zph()`
- Performs a Chi-square test on each term, and a global test
- In this test, higher p-values are good
 - p-value > 0.05: no evidence that the PH assumption is violated
 - p-value < 0.05: evidence that the effect changes over time
- If the global test fails (low p-value), the CPH model assumptions are violated

Test	Chi-square	d.f.	P-value
Age Group	1.863	2	0.394
Marital Status	0.181	1	0.671
Multi Policy Flag	22.547	1	<0.001
Complaint Flag	6.321	1	0.012
GLOBAL	31.823	5	<0.001

CPH Model Output

- The H in “CPH” stands for “Hazard”. Positive coefficients mean shorter durations.
- P-values have a similar interpretation to GLMs
- CPH is semi-parametric, there is no parameter for the baseline hazard (no intercept)

Variable	Value	P-value
(Intercept)	NA	NA
Age Group: 55+	0.224	<0.001
Age Group: Under 35	0.231	<0.001
Marital Status: Single	0.015	0.593
Multi Policy Flag	-0.825	<0.001
Complaint Flag	0.698	<0.001

Model Output Comparison

- Weibull AFT

Variable	Value	P-value
(Intercept)	3.041	<0.001
Age Group: 55+	-0.193	<0.001
Age Group: Under 35	-0.183	<0.001
Marital Status: Single	-.016	0.498
Multi Policy Flag	0.670	<0.001
Complaint Flag	-0.549	<0.001

- CPH

Variable	Value	P-value
(Intercept)	NA	NA
Age Group: 55+	0.224	<0.001
Age Group: Under 35	0.231	<0.001
Marital Status: Single	0.015	0.593
Multi Policy Flag	-0.825	<0.001
Complaint Flag	0.698	<0.001

CPH AIC Value

- AIC can be calculated for CPH too
- The AIC from CPH is not comparable to AIC from AFT
 - AFT AIC is based on log-likelihood, while CPH AIC is based on partial log-likelihood
- CPH AIC can be used to compare across multiple CPH models

CPH Model	AIC
Full model (with marital status)	86,233
Subset model (without marital status)	86,231

CPH C-Index

- C-Index is applicable to all survival models
- Useful for comparison to AFT models
- Equal in our example

Model	C-Index
AFT Weibull	0.588
CPH	0.588

Useful Model Documentation

Useful Model Documentation

- For Both:
 - Provide a table summarizing the data volume by originally effective year with policy count, cancellation count, and active (censored) count
 - Provide the KM estimator and a survival curve for the specified model type using no covariates
 - Provide the survival curves by variable with separate lines for each variable level or by decile for continuous variables
 - Plot the KM estimator and the survival curve for a high retaining group
 - Plot the KM estimator and the survival curve for a low retaining group
 - Provide coefficient values and P-values for each variable
 - Provide AIC for the full model and each subset model excluding one variable
 - Provide the C-Index

Useful Model Documentation

- For AFT:
 - Describe how the error distribution was chosen
- For CPH:
 - Provide the results of the statistical test for the proportional hazard assumption

Survival Model Predictions

Predictions from Survival Models

- Predictions
 - $S(t)$ for any risk class for any selected t
 - Predicted mean or median policy duration by risk class
- Predictions can be used to inform insurer business decisions

Thank you!

Contact: skloese@naic.org

References

References

- DataCamp Courses:
 - Survival Analysis in R
 - Survival Analysis in Python
- AFT for insurance paper:
 - Majeed, Accelerated Failure Time Models: An Application in Insurance Attrition, The Journal of Risk Management and Insurance, Vol. 24 No. 2
- Synthetic Kaggle Dataset:
 - <https://www.kaggle.com/datasets/andrewdiachkov/realistic-synthetic-insurance-churn-dataset-for-ml?resource=download>