A Novel Approach to Modeling Variable Annuity Policyholder Surrender Using Joint Modeling Techniques

Pstat 296-Graduate Research in Actuarial Science

Team Members:
John Randazzo
Terry Jiang
Fay Wu
Advisors:
Prof. Ian Duncan
Prof. Jiyoun Myung
Shannon Nicponski

Project Background

- Industry sponsor/data provider: [REDACTED]
- Event of interest: Policyholder **surrender** of Variable Annuity contract
  - **Surrender**: Policyholder withdrawals all money out of account and terminates the contract.
- Fit an Averaged Logistic Regression Model
- Fit a **Joint Model**
  - Fit a linear mixed effects model to a longitudinal, endogenous covariate
  - Fit a Cox Proportional Hazards Survival Model

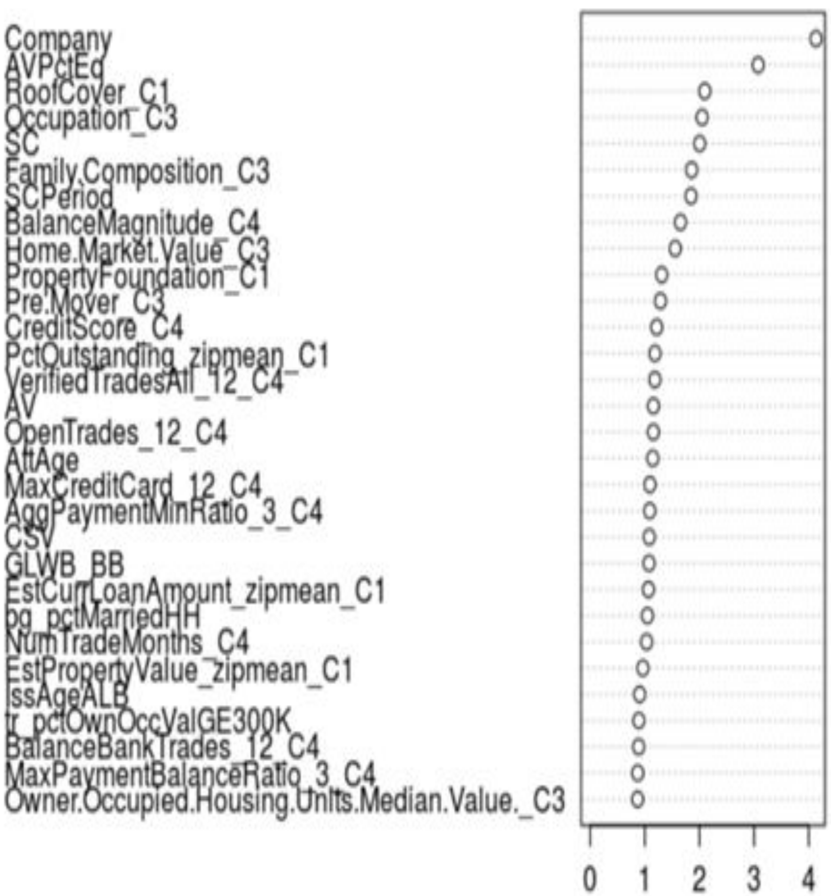
Data Description

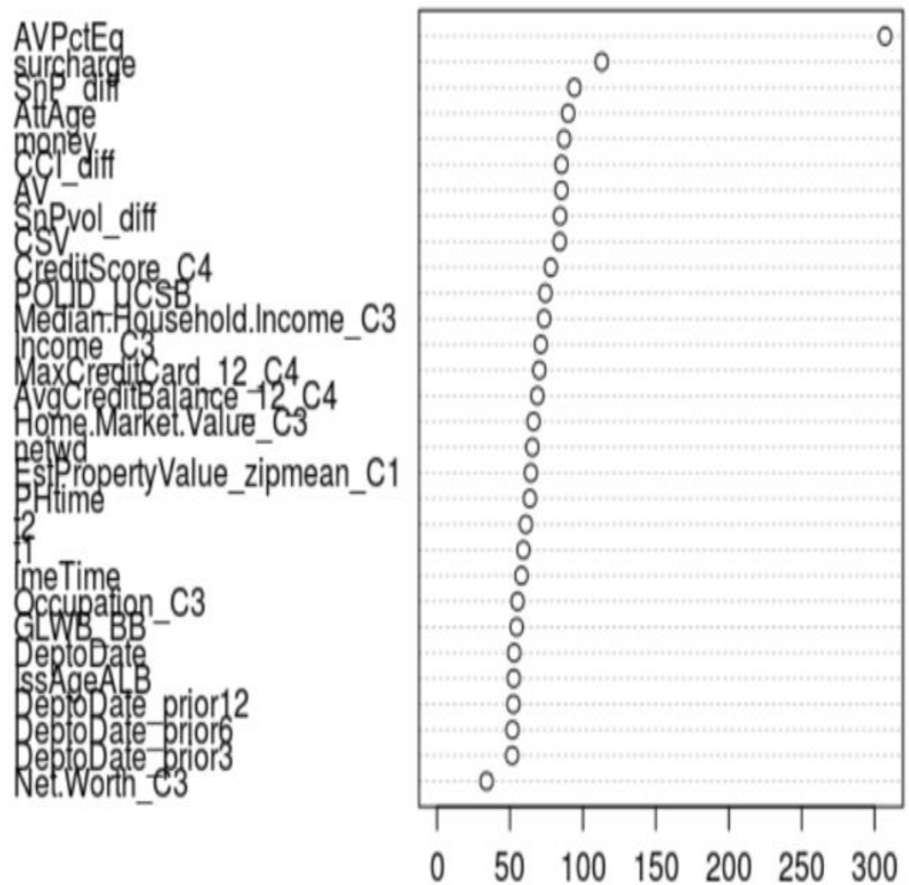
- Data obtained through the good graces of [REDACTED] consist of:
  - VA contract data from 2 companies
  - U.S Bureau census data
  - Consumer data (credit card, property, etc.)
- 4,732,698 rows consisting of quarterly observations on 518,120 policyholders
- 11 valuation quarters (June 2012 - December 2014)

Data Preprocessing

- Remove extra observation after the policyholder surrendered
  - Left with 4,512,705 rows (95% of the original size)
- Add in new data from market
  - Historical closing prices for S&P 500 - we normalize the differences
  - S&P500 at Valuation Date – S&P 500 at Issue Date
- Created new variables for Joint Model
- Variable selection through random Forest and Variable Importance Plot.
- Missing values filled with Median or Mode

Variable Importance Plot

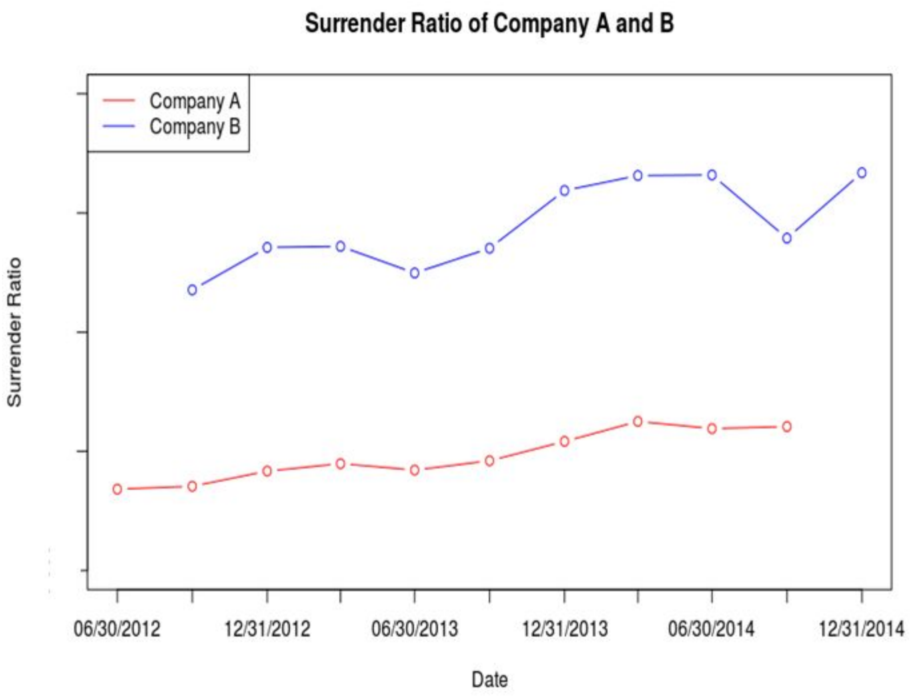




Important Terminology/Covariates

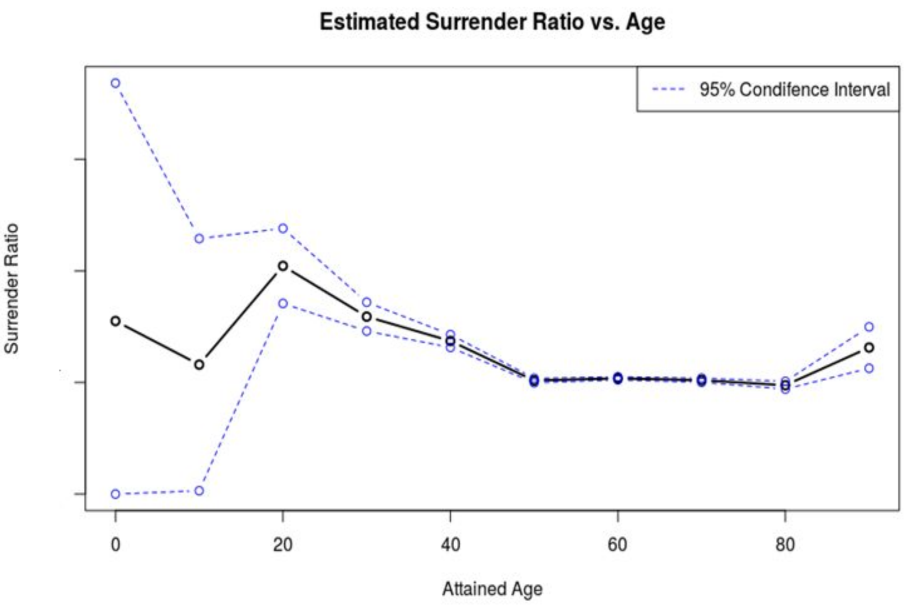
Variables	
AV	Account Value: dollar amount in the variable annuity account at observation time
GLWB_BB	Guaranteed Lifetime Withdrawal Benefit (Benefit Base), the amount of guaranteed return on the variable annuity product
Moneyness	GLWB_BB/AV Moneyness > 1 → "In the money" (less incentive to surrender) Moneyness < 1 → "Out of the money" (higher incentive to surrender) $0<\text{Moneyness}<2$
Phtime	The number of quarters a Policyholder has been in the contract at observation time
SCI	Surrender Charge Indicator In – The policyholder is in Surrender Charge Period Shock – The policyholder is at the end of Surrender Charge Period Out – The policyholder is out of the Surrender Charge Period

Observation: Company

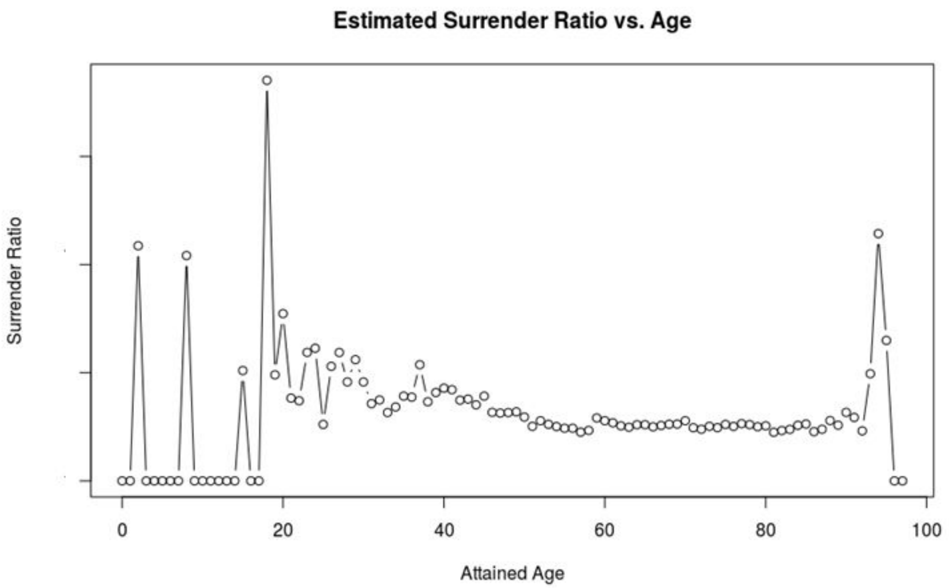


- Glm Model: Separate Glm for each company.
  - Different distributions in two companies.
  - Different observation time periods.
- Joint Model: Only modeling people from Company B.
  - Higher Surrender frequency.
  - Provides data on how much of each policyholder's account is in equity

Observation: Age



Time interval = 10 years



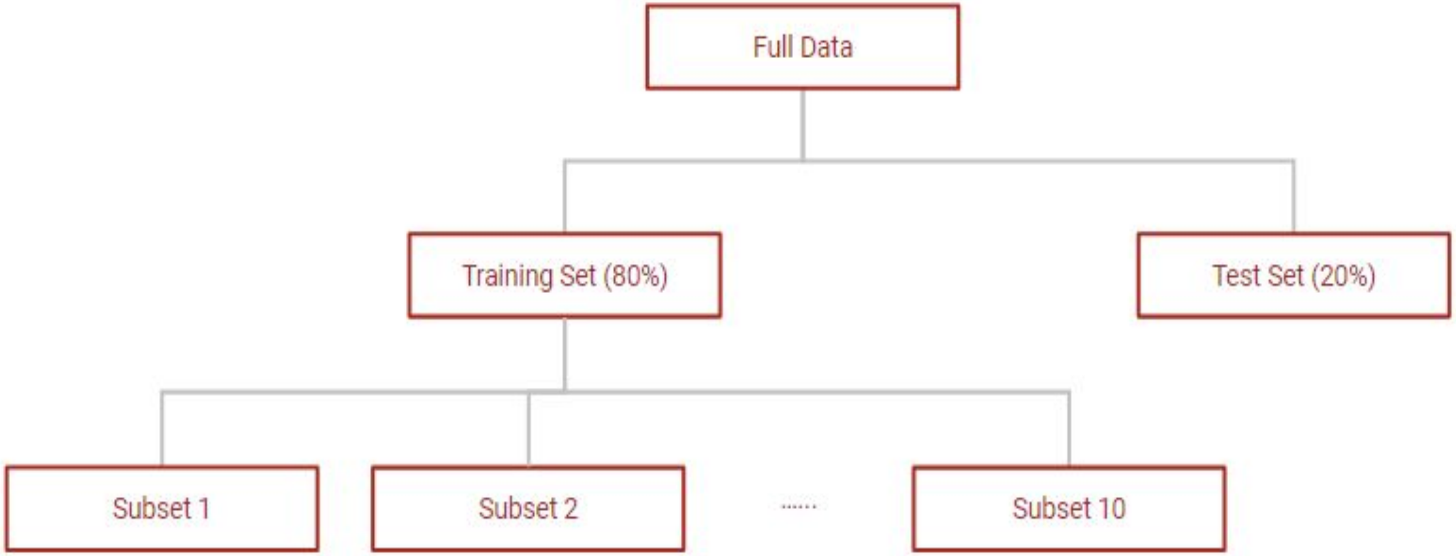
Time interval = 1 year

Observation: Age Cont.

- Distributions that occur before the IRA owner reaches the age of 59½ are subject to a 10% early-distribution penalty, in addition to any income tax owed.
- Required Minimum Distributions (RMDs) generally are minimum amounts that a retirement plan account owner must withdraw annually starting with the year that he or she reaches 70 ½ years of age or, if later, the year in which he or she retires.

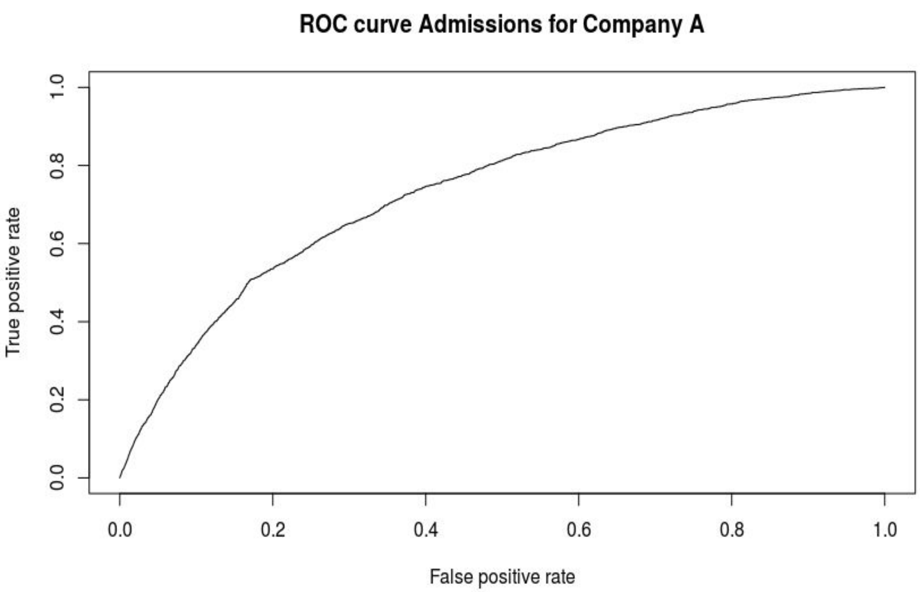
	Z Value
Age 59 - IRA	5.39915
Age 59 - No IRA	2.92530
Age 70 - IRA	-3.70759
Age 70 - No IRA	-1.16711

Averaged Logistic Regression Model Steps

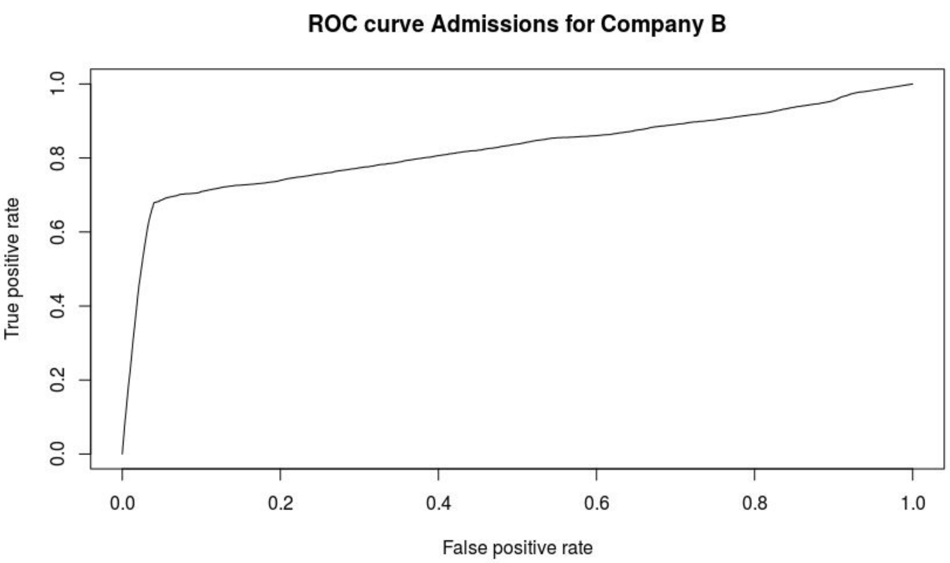


Use the model with averaged coefficients in glm's from 10 subsets to predict the response in the test set

Model Performance on the Test Set



AUC = 0.735



AUC = 0.8221

Averaged Logistic Regression Model Results

Variable Name	Relationship to Surrender
Maximum Credit Card Owned	Positive
Credit Score	Negative
Surrender Charge Indicator - Out	Positive
Snp 500 Normalized Difference	Negative
Moneyness	Negative

A Crash Course on Joint Modeling
Cox Proportional Hazard Model

- The most popular approach to modeling a survival process (such as time-to-surrender) is to use a Cox Proportional Hazards model:

$$\lambda(t|X_i) = \lambda_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_p X_{ip}) = \lambda_0(t)\exp(X_i \cdot \beta).$$

- The crux of the matter is that the model relies on the Proportional Hazards assumption, that the effects of any covariates included in the model are constant over time (time-independent), ie:

$$\frac{\lambda(t|X_i)}{\lambda(t|X_j)} = c \in \mathbb{R}, \forall t$$

Why Not a Simple Cox Model?

- The Cox PH model can be adjusted to accommodate time-varying effects, but we run into another problem that invalidates this approach
- The Cox approach assumes that all covariates are **exogenous** (external), meaning that they are deterministic and tell the “full story”
  - Examples: age, sex, occupation, number of beers drunk
- However, in most cases, variables of interest can be classified as **endogenous** (internal), meaning that they are prone to measurement error
  - Examples: T cell count, blood alcohol levels, moneyness
- One can model an endogenous marker’s evolution over time using a linear mixed effects model, which fits a model with random and fixed effects

A Crash Course on Joint Modeling
Linear Mixed Effect Model

- The mixed model approach remedies our problem of measurement error by postulating that the observed longitudinal outcome $y_i(t)$ equals the "true" level plus a random term

$$
\begin{aligned}
y_i(t) &= m_i(t) + \epsilon_i(t) \\
m_i(t) &= x_i^T \cdot \beta_i + z_i^T \cdot b_i \\
b_i &\sim \mathcal{N}(0, D), \epsilon_i \sim \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

- $\mathbf{X_i}$ = fixed effects vector
- $\mathbf{Z_i}$ = random effects vector
- $\mathbf{B_i}$ = fixed effects coefficients
- $\mathbf{b_i}$ = random effects coefficients
- $\mathbf{y_i(t)}$ is the observed level of the longitudinal outcome at time t for subject i
- $\mathbf{m_i(t)}$ is the fitted value of the longitudinal outcome at time t for subject i

Arriving at the Joint Model

- The mixed model approach remedies our problem of measurement error by postulating that the observed longitudinal outcome $y_i(t)$ equals the "true" level m_i (t) plus a random term representing measurement error

- We then use our estimates for the "true" value of $m_i(t)$ and impute it into the Cox model, yielding a joint model:

$$\lambda_i(t) = \lambda_0(t) exp(x_i^T \cdot \beta + \alpha m_i(t))$$

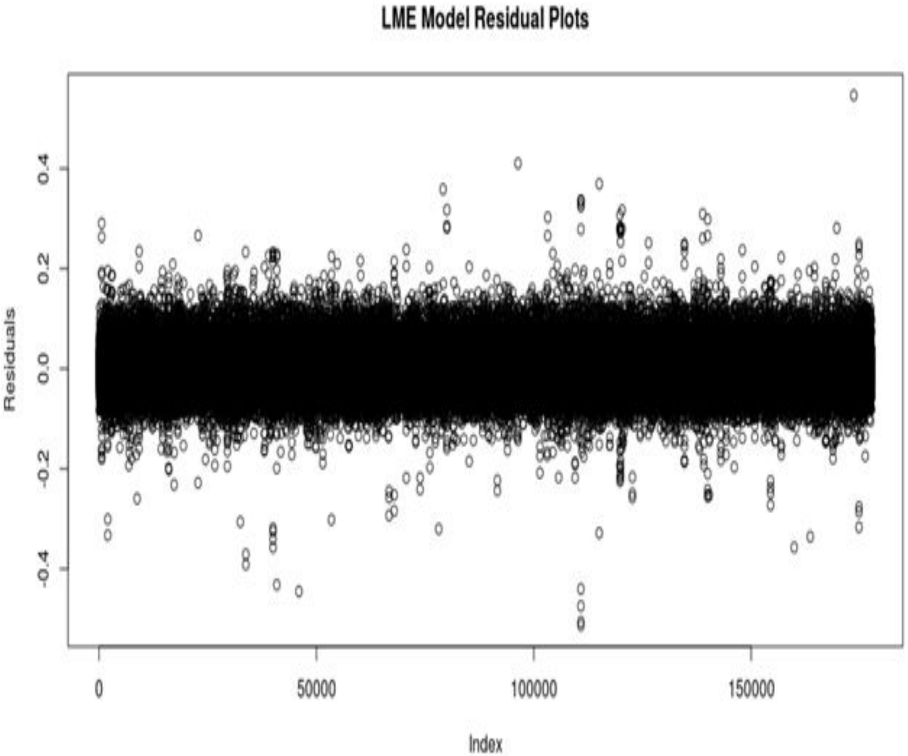
Important Terminology/Covariates

Variables	
AV	Account Value: dollar amount in the variable annuity account at observation time
GLWB_BB	Guaranteed Lifetime Withdrawal Benefit (Benefit Base), the amount of guaranteed return on the variable annuity product
Moneyness	GLWB_BB/AV Moneyness > 1 → "In the money" (less incentive to surrender) Moneyness < 1 → "Out of the money" (higher incentive to surrender) 0<Moneyness<2
Phtime	The number of quarters a Policyholder has been in the contract at observation time
SCI	Surrender Charge Indicator In – The policyholder is in Surrender Charge Period Shock – The policyholder Surrender Charge Period just ended Out – The policyholder is out of the Surrender Charge Period

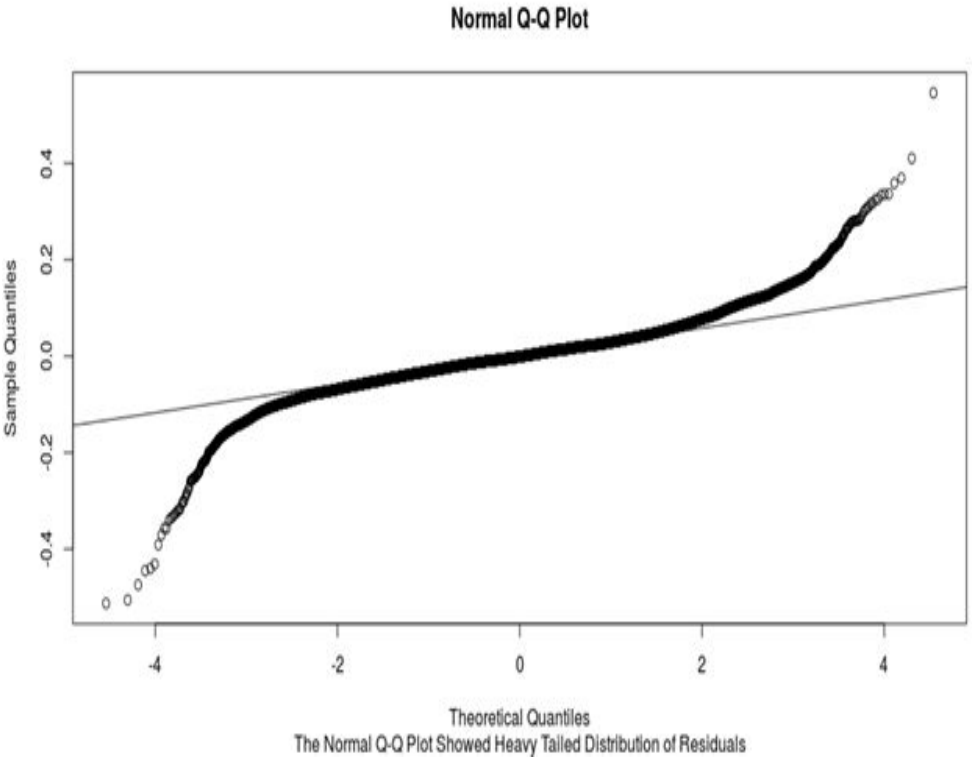
Our Linear Mixed Effects Model Moneyness as response variable

Variable Name	Effect	P-value
Normalized S&P Difference	Positive	0
Policyholder Time	Positive	0
Number of Withdrawals in past 3 months	Positive	0
Average Percentage of Equity	Positive	0
Interaction between S&P_Diff and Policyholder Time	Positive	0

LME Diagnostic



Residual Plot



Normal Q-Q Plot

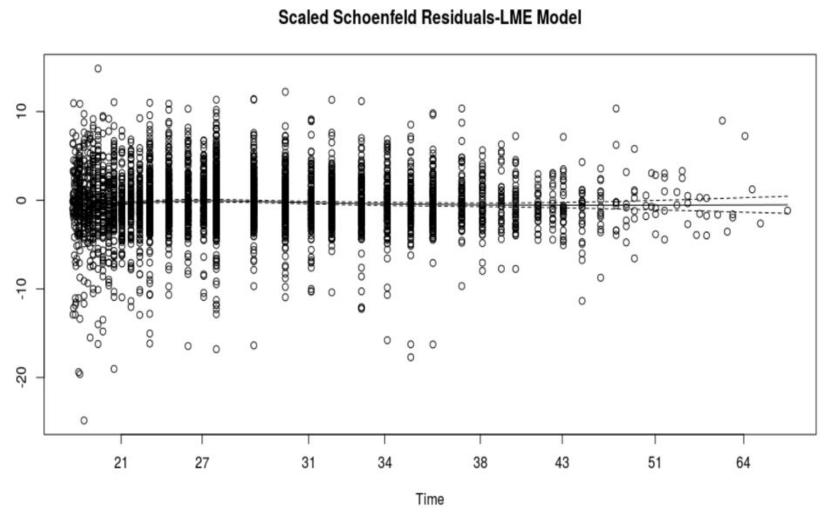
The Q-Q plot demonstrates heavy-tailed distribution of residuals due to left skewed distribution an capping Moneyness

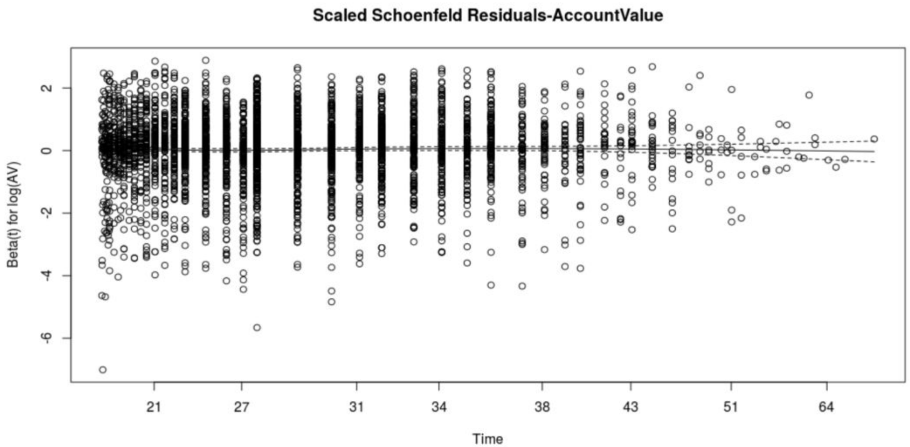
Our Cox PH Model

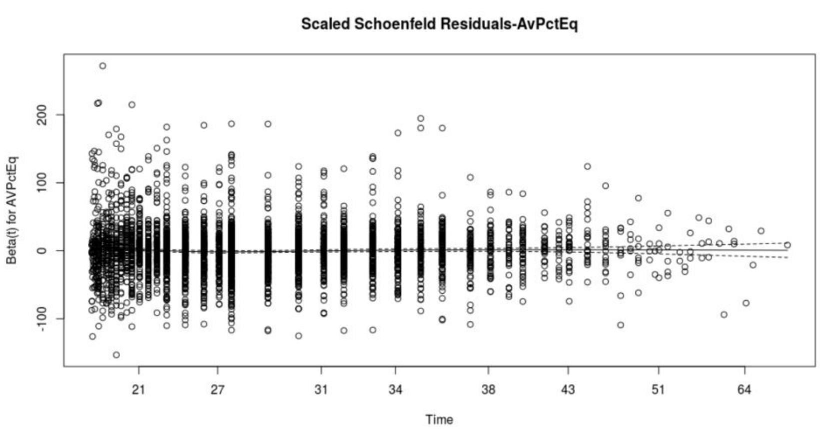
Variable	Effect	P-value
S&P Normalized Difference	Positive	0
Log(Account Value)	Positive	0.0.12
Average Percentage in Equity	Positive	0.35
Surrender Charge Indicator-Out	Negative	0
Surrender Charge Indicator - Shock	Negative	0
Credit Score	Negative	0
Interaction: Log(Account Value) and Average Percentage in Equity	Negative	0

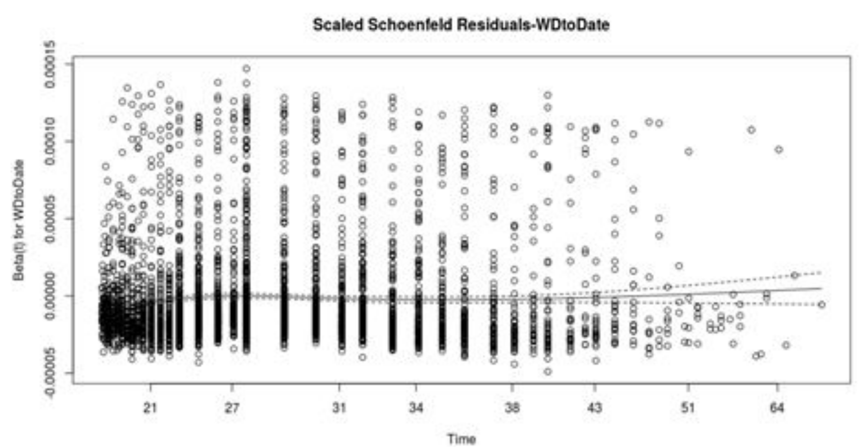
John

Cox PH Model Diagnostic









At Long Last… Our Joint Model

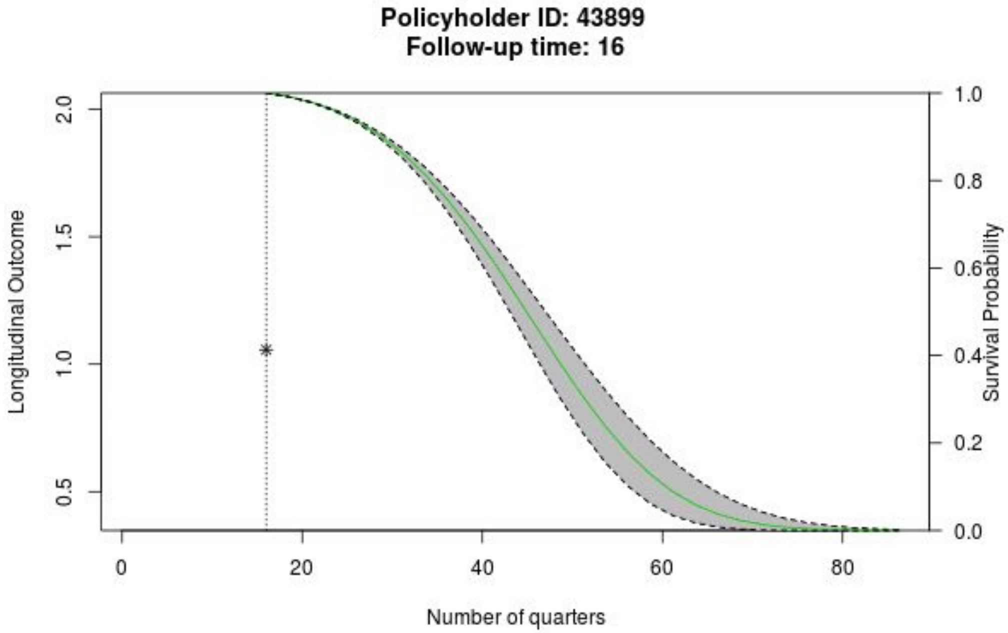
Negative Association	P-value = 0

Variable Name-Longitudinal	Effe ct	P
Normalized S&P Difference	+	0
Policyholder Time	+	0
Number of Withdrawals in past 3 months	+	0
Average Percentage of Equity	+	0
Interaction between S&P_Diff and Policyholder Time	+	0

Variable	Effect	P
S&P Normalized Difference	+	0
Log(Account Value)	-	0.3697
Average Percentage in Equity	-	0
Surrender Charge Indicator-Out	-	0
Surrender Charge Indicator - Shock	-	0
Credit Score	-	0
Interaction: Log(Account Value) and Average Percentage in Equity	+	0.3667

Model Prediction

- The R package “JM” has a really cool function “runDynPred()” which allows us to visualize dynamic survival predictions and how the longitudinal response’s fluctuation affects them over time

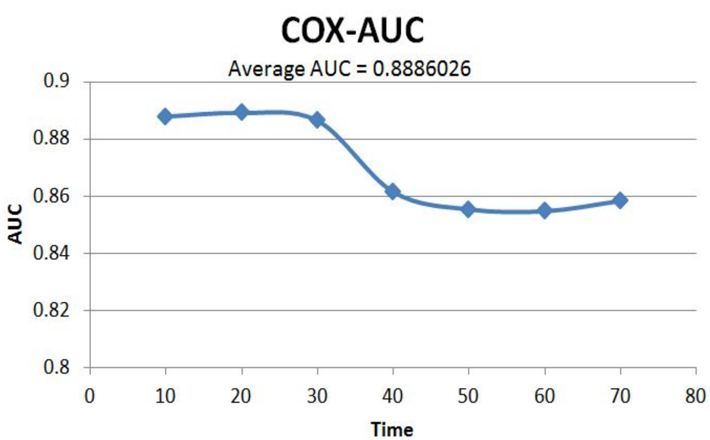


Model "Comparison" by AUC

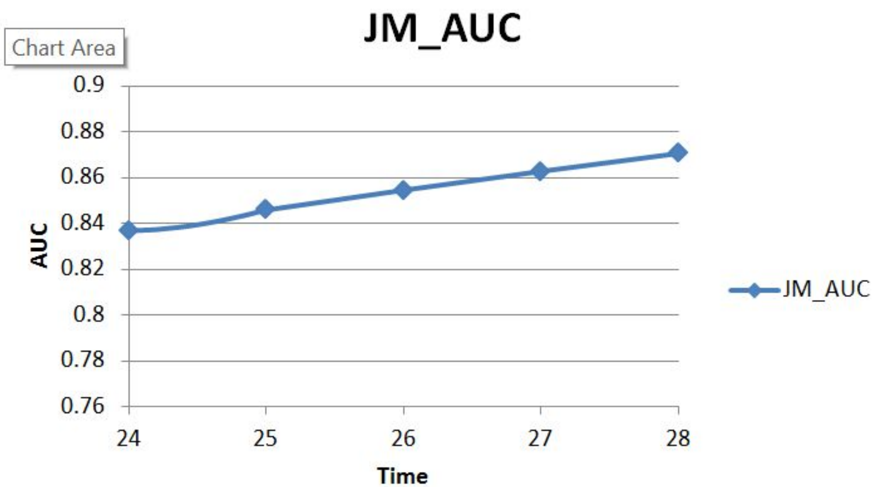
- GLM Model:

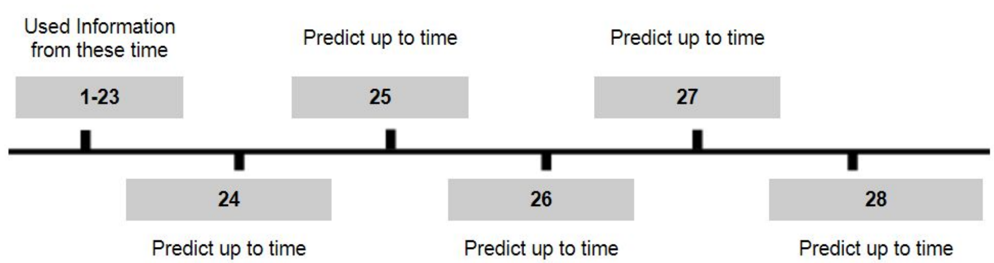
Com	AUC
Company A	0.735
Company B	0.8221

- Cox Model



Joint Model:





Challenges and Insights

- Joint models are very complex from a computational perspective
- Data need to be well-organized before a joint model can be fitted without issue
- We couldn't get some of the diagnostic functions in library(JM) to work if our model had categorical variables
- Models are imperfect - there are other endogenous covariates related to surrender that can be modeled through fixed and random effects
- library(JM) supports multivariate approaches… not for the squeamish. There's also a Bayesian approach.

Challenges and Insights

- Jo… pe…
- D… can be fi…

…M)

…s

```
> data$money <- (data$GLWB_BB - data$AV) / data$GLWB_BB
> coxph(Surv(data$t1,data$t2,data$Surr) ~ data$money)
Error in fitter(X, Y, strats, offset, init, control, weights = weights,  :
  routine failed due to numeric overflow.This should never happen.  Please contact the author.
```

```
Error in jointModel
it se
chec
96, 97, 101, 102, 105, 112, 114, 122, 1
231, 240, 242, 248, 261, 270, 271, 272, 274, 275, 292).
```

```
Error in `contrasts<-`(`*tmp*`, value = contr.funs[1 + isOF[nn]]) :
  contrasts can be applied only to factors with 2 or more levels
```

```
Error in lme.formula((money) ~ SnP_diff_
  nlminb problem, convergence error code
  message = false convergence (8)
```

```
Warning message:
In jointModel(mod33, mod22, timeVar = "PHtime") :
  infinite or missing values in Hessian at convergence.
```

```
Estimation: Monte Carlo ( samples)
Error in if (x$diffType == "absolute") { : argument is of length zero
```

Conclusion

- Joint Modeling is the correct approach for the phenomenon at hand.
- Moneyness has a direct inverse relation with surrenders, although many explanatory variables exists to predict surrenders.
- A multivariate approach is needed to arrive at a industry-quality model.
- This problem deserves more attention than we were able to give it

Closing Remarks

- In particular, we think that the process of applying JM to this context and building a model from the ground up (rigorously and correctly) deserves to be investigated in a Master's thesis or PhD dissertation
- We maintain that the joint modeling approach is a viable one for the phenomena of interest in this study, and urge you to consider it in your approaches for studying time-to-event and longitudinal data

Acknowledgments

Major props to:

- Dimitris Rizopolous (author of R package "JM")
- Terry Therneau (author of R package "survival")
- Jose Pinhiero (author of R package "nlme)
- Mr. and Mrs. Duncan
- Prof. Jiyoun Myung
- Shannon Nicponski
- Nhan Huynh

Thank You For Listening!

Questions/Observations?