Event Variables in Client Analytics

The Co-operators, assurance et services financiers

A. Alinas, T. Duchesne, E. Lavoie-Charland, M. Malekiha, J. McVittie, I. Saïdani, A. Sen, J. Wang, M. Zhao

Seventh Montreal Industrial Problem Seminar

Outline

1. Summary of Problem
2. Proposed Models
3. Model Validation
4. Further Considerations

Main questions

1. Client Retention:
   - When will an existing client leave the company?

     e.g. At a given point in time, which clients are most at risk of leaving the company?
2. Life Cross-Sale:
   - When will an existing client add a new product?

     e.g. Given a client's current product(s), what is the probability that they will leave before adding life insurance, or vice-versa?

The objectives

1. Determine the relevant response variable for each question.
2. Develop a statistical approach to determine a reasonable model.
3. Incorporate the roles of potential covariates (fixed and time-dependent).
4. Perform validation, measure performance, and assess prediction.

The objectives

1. Determine the relevant response variable for each question.
2. Develop a statistical approach to determine a reasonable model.
3. Incorporate the roles of potential covariates (fixed and time-dependent).
4. Perform validation, measure performance, and assess prediction.

The objectives

1. Determine the relevant response variable for each question.
2. Develop a statistical approach to determine a reasonable model.
3. Incorporate the roles of potential covariates (fixed and time-dependent).
4. Perform validation, measure performance, and assess prediction.

The objectives

1. Determine the relevant response variable for each question.
2. Develop a statistical approach to determine a reasonable model.
3. Incorporate the roles of potential covariates (fixed and time-dependent).
4. Perform validation, measure performance, and assess prediction.

Available data

- Data:
  1. TRAINING SET
  2. TESTING SET
- Time intervals for every client, start and end times defined by the events of interest:
  1. Addition of one or more products
  2. Removal of one or more products
  3. Earliest termination of all products
- Approx. 25 covariates to potentially be used in the model

Available data

- Data:
  1. TRAINING SET
  2. TESTING SET
- Time intervals for every client, start and end times defined by the events of interest:
  1. Addition of one or more products
  2. Removal of one or more products
  3. Earliest termination of all products
- Approx. 25 covariates to potentially be used in the model

Available data

- Data:
  1. TRAINING SET
  2. TESTING SET
- Time intervals for every client, start and end times defined by the events of interest:
  1. Addition of one or more products
  2. Removal of one or more products
  3. Earliest termination of all products
- Approx. 25 covariates to potentially be used in the model

Client retention problem: difficulties with covariates

- The Cox Proportional Hazards (PH) model assumes this hazard function for the time until an event ("earliest" termination of all products):

$$\lambda(t) = \lambda_0(t) e^{\boldsymbol{\beta}^\top \boldsymbol{z}}$$

$\lambda_0(\cdot)$ is an unspecified baseline hazard, $\boldsymbol{z}$ are the covariates known at time t, and $\boldsymbol{\beta}$ are the parameters to be estimated.

- R packages easily fit Cox PH and predict quantities such as the cumulative hazard, when $\boldsymbol{z}$ is time-*independent*

$$\Lambda(t) = \int_0^t \lambda(u) du = e^{\boldsymbol{\beta}^\top \boldsymbol{z}} \int_0^t \lambda_0(u) du$$

Client retention problem: difficulties with covariates

- The Cox Proportional Hazards (PH) model assumes this hazard function for the time until an event ("earliest" termination of all products):

$$\lambda(t) = \lambda_0(t) e^{\boldsymbol{\beta}^\top \mathbf{z}}$$

  $\lambda_0(\cdot)$ is an unspecified baseline hazard, $\mathbf{z}$ are the covariates known at time t, and $\boldsymbol{\beta}$ are the parameters to be estimated.

- R packages easily fit Cox PH and predict quantities such as the cumulative hazard, when $\mathbf{z}$ is time-*independent*

$$\Lambda(t) = \int_0^t \lambda(u) du = e^{\boldsymbol{\beta}^\top \mathbf{z}} \int_0^t \lambda_0(u) du$$

Typical approaches and difficulties with covariates

- However, with time-*varying* covariates, to obtain an estimate of the hazard $\lambda(t)$, we need an estimate of $\Lambda_0(t)$ for which the slope can be obtained
- Either we:
  - smooth $\Lambda_0(t)$ or
  - assume a parametric form for the baseline hazard

Typical approaches and difficulties with covariates

- However, with time-*varying* covariates, to obtain an estimate of the hazard $\lambda(t)$, we need an estimate of $\Lambda_0(t)$ for which the slope can be obtained
- Either we:
  - smooth $\Lambda_0(t)$ or
  - assume a parametric form for the baseline hazard

Typical approaches and difficulties with covariates

- However, with time-*varying* covariates, to obtain an estimate of the hazard $\lambda(t)$, we need an estimate of $\Lambda_0(t)$ for which the slope can be obtained
- Either we:
  - smooth $\Lambda_0(t)$ or
  - assume a parametric form for the baseline hazard

Typical approaches and difficulties with covariates

- However, with time-*varying* covariates, to obtain an estimate of the hazard $\lambda(t)$, we need an estimate of $\Lambda_0(t)$ for which the slope can be obtained
- Either we:
  - smooth $\Lambda_0(t)$ or
  - assume a parametric form for the baseline hazard

Our immediate options

1. Assume baseline hazard takes a parametric form so that the hazard can be computed directly:

$$\lambda(t) = \lambda_0(t; \theta) e^{\boldsymbol{\beta}^\top \boldsymbol{z}}$$

2. Assume baseline hazard is linear piecewise

Our immediate options

1. Assume baseline hazard takes a parametric form so that the hazard can be computed directly:

$$\lambda(t) = \lambda_0(t; \theta) e^{\boldsymbol{\beta}^\top \boldsymbol{z}}$$

2. Assume baseline hazard is linear piecewise

More elaborate approaches

These models can be used for either the client retention or the life cross-sale problems:

1. Multi-state models
2. Self-exciting process models

More elaborate approaches

These models can be used for either the client retention or the life cross-sale problems:

1. Multi-state models
2. Self-exciting process models

Fitting a parametric baseline hazard

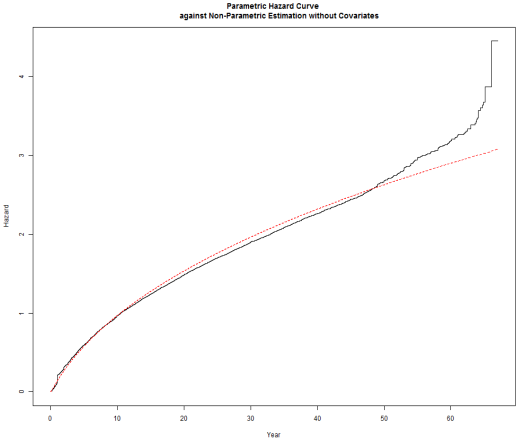
- Implementation of the Cox PH model with a parametric baseline hazard is relatively straightforward with the `R` library `flexsurv`
- Choices for $\lambda_0(\cdot)$ include the 4-parameter generalized gamma (includes gamma, weibull, etc.), generalized F, and Gompertz

Fitting a parametric baseline hazard

- Implementation of the Cox PH model with a parametric baseline hazard is relatively straightforward with the `R` library `flexsurv`
- Choices for $\lambda_0(\cdot)$ include the 4-parameter generalized gamma (includes gamma, weibull, etc.), generalized F, and Gompertz

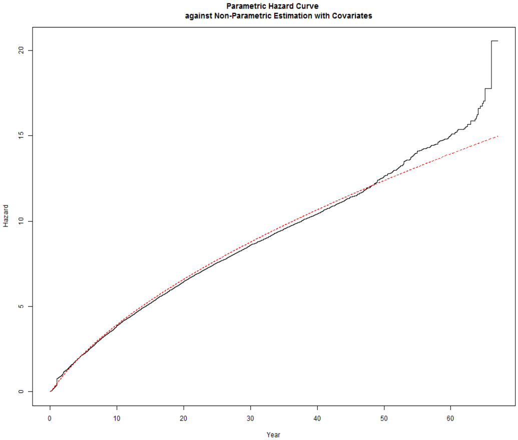
Preliminary Results

Generalized gamma baseline cumulative hazard without covariates

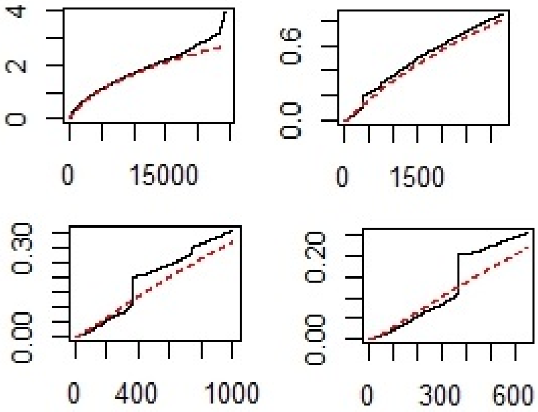


Preliminary Results

Generalized gamma baseline cumulative hazard with covariates



Zooming in tells a different story...



Considerations

- Note: Running `flexsurv` on a large data set takes much longer than running `coxph`
- Our initial plan:
  - Find a suitable set of covariates using the semi-parametric Cox PH model
  - Use these covariates in a parametric Cox PH model to do prediction

Considerations

- Note: Running `flexsurv` on a large data set takes much longer than running `coxph`
- Our initial plan:
    - Find a suitable set of covariates using the semi-parametric Cox PH model
    - Use these covariates in a parametric Cox PH model to do prediction

Considerations

- Note: Running `flexsurv` on a large data set takes much longer than running `coxph`
- Our initial plan:
  - Find a suitable set of covariates using the semi-parametric Cox PH model
  - Use these covariates in a parametric Cox PH model to do prediction

Selection of "optimal" covariates

What we tried:

- Given all covariates, cycle through all possible additive models and compare AIC (or BIC)
- Forward and backward selection based on likelihood ratio tests
- Other algorithms (eg. PCA)

Selection of “optimal” covariates

What we tried:

- Given all covariates, cycle through all possible additive models and compare AIC (or BIC)
- Forward and backward selection based on likelihood ratio tests

- Other algorithms (eg. PCA)

Selection of "optimal" covariates

What we tried:

- Given all covariates, cycle through all possible additive models and compare AIC (or BIC)
- Forward and backward selection based on likelihood ratio tests

- Other algorithms (eg. PCA)

Measurement of performance

How would we compare candidate models?

- The proportional hazards assumption can be checked for a fitted model by examining plots of the β coefficients over time
- Whether or not this assumption is met, it is more useful to the company to have a practical *prediction*-based measure
- Prediction with time-varying covariates is a unique challenge for survival models

Measurement of performance

How would we compare candidate models?

- The proportional hazards assumption can be checked for a fitted model by examining plots of the β coefficients over time
- Whether or not this assumption is met, it is more useful to the company to have a practical *prediction*-based measure
- Prediction with time-varying covariates is a unique challenge for survival models

Measurement of performance

How would we compare candidate models?

- The proportional hazards assumption can be checked for a fitted model by examining plots of the β coefficients over time
- Whether or not this assumption is met, it is more useful to the company to have a practical *prediction*-based measure
- Prediction with time-varying covariates is a unique challenge for survival models

Measurement of performance

Which clients are "most likely" to leave as of a given date, say July 1, 2012?

- The model can give us a ranking of the instantaneous probabilities of leaving the company, $\lambda_i(t)$; $i = 1, ..., m$, for the m clients in the test set
- Compare this with the *actual* ranking of clients in order of when they left the company after July 1, 2012
- Various measures: gain curves, lift curves, ROC, etc.

Measurement of performance

Which clients are “most likely” to leave as of a given date, say July 1, 2012?

- The model can give us a ranking of the instantaneous probabilities of leaving the company, $\lambda_i(t)$; $i = 1, ..., m$, for the m clients in the test set
- Compare this with the *actual* ranking of clients in order of when they left the company after July 1, 2012
- Various measures: gain curves, lift curves, ROC, etc.

Measurement of performance

Which clients are "most likely" to leave as of a given date, say July 1, 2012?

- The model can give us a ranking of the instantaneous probabilities of leaving the company, $\lambda_i(t)$; $i = 1, \ldots, m$, for the m clients in the test set
- Compare this with the *actual* ranking of clients in order of when they left the company after July 1, 2012
- Various measures: gain curves, lift curves, ROC, etc.

Modeling the cross-sale

- Multi-state models are extremely flexible (can be implementated with packages such as `mstate`)
- The states and allowable transitions must be carefully defined
- Quantities of the form

$$\Pr\{\text{ in state } j \text{ at time } t \mid \text{ in state } i \text{ at time } s\}$$

could then be accessible

Modeling the cross-sale

- Multi-state models are extremely flexible (can be implementated with packages such as `mstate`)
- The states and allowable transitions must be carefully defined
- Quantities of the form

$$\Pr\{\text{ in state } j \text{ at time } t \mid \text{ in state } i \text{ at time } s\}$$

could then be accessible

Modeling the cross-sale

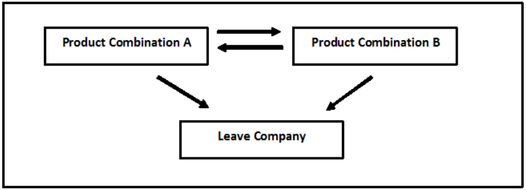
- Multi-state models are extremely flexible (can be implementated with packages such as `mstate`)
- The states and allowable transitions must be carefully defined
- Quantities of the form

$$\Pr\{\text{ in state } j \text{ at time } t \,\big|\, \text{in state } i \text{ at time } s\}$$

could then be accessible

Addressing the cross-sale problem

- We considered a simple 3-state model



Relevant and numerically accessible:

$$\Pr\{\text{will have combo } B \textbf{ before} \text{ leaving the company} \mid \text{currently have combo } A\}$$

Modeling the cross-sale

- Other difficulties that require more time:
  - Formatting the data set such that transitions are clearly defined
  - Expressions of conditional transition probabilities are usually not analytically tractable

Modeling the cross-sale

- Other difficulties that require more time:
  - Formatting the data set such that transitions are clearly defined
  - Expressions of conditional transition probabilities are usually not analytically tractable

Modeling the cross-sale

- Other difficulties that require more time:
  - Formatting the data set such that transitions are clearly defined
  - Expressions of conditional transition probabilities are usually not analytically tractable

Other Models and Future Work

- Self-exciting models and other counting models
- Survival trees and ensemble methods

Other Models and Future Work

- Self-exciting models and other counting models
- Survival trees and ensemble methods

References

- T. Petersen. *Fitting Parametric Survival Models with Time-Dependent Covariates*, Journal of Royal Statistical Society, 1986.
- C. Jackson. *flexsurv: A Platform for Parametric Survival Modelling in R*, Journal of Statistical Software, 2016.
- F. Gao, A. Manatunga, S. Chen. *Non-parametric estimation for baseline hazards function and covariate effects with time-dependent covariates*, Statistics in Medicine, 2007.
- F. Larocque, H. Ben-Ameur. *A review of survival trees*, Statistics Surveys, 2011.
- J. Yan, J. Huang. *Model Selection for Cox Models with Time-Varying Coefficients*, Biometrics, 2012.
- C. Jackson. *A Playform for Parametric Survival Modeling in R*, Journal of Statistical Software, 2016.