Deep Learning Model for Survival Analysis

Authors :

Sheeba Mohan

R. Shyaam Prasadh

Nov 25, 2021

GDI&A

Global Data Insight & Analytics

Agenda

- Ford Credit business
- Motivation
- Data
- Methodology
- Results
- Robustness tests
- Summary
- Future Work

Ford Credit Business

Example:

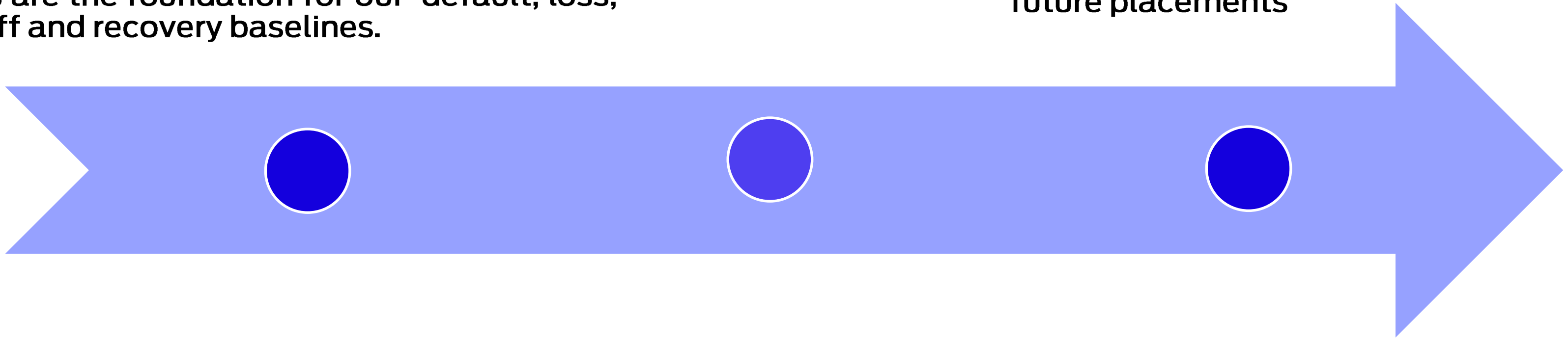
Retail contract, 60 months to maturity, consumer risk class tier 0, MSRP \$23 000, amount financed \$24 000, APR 0%, monthly payment of \$300, balloon payment at maturity end \$6 000.

Default after 20 months, amount outstanding at time of default \$18 000, residual value of the car after 20 months: \$15 000.

Loss	\$18 000 Amount outstanding at time of default
gross loss (after repossession)	\$3 000 (\$18 000 - \$15 000) Loss – sales proceed
net loss (after recoveries)	\$3 000 – recoveries

Contract level information from the past 15 years are the foundation for our default, loss, payoff and recovery baselines.

Loss Forecast depends on baselines, recent default/loss behavior, and future placements



FORD CREDIT

Ford Credit Business

What is a Baseline?

A baseline estimates how likely an account in the portfolio will default / payoff. For example, default baselines show the distribution of time to default.

– A baseline is the probability distribution of time to event data.

– If default/payoff/liquidated, when will it occur ➔ what is the likelihood that it will default/payoff/closed at t- months after origination?

Applications of Baselines

- Default baseline is used in projecting the lifetime performance of the portfolio, projected POP for scorecard development/ monitoring and IFRS9 modeling.
- Loss baseline is used for credit loss forecasting (CPE) and securitization.
- Payoff baseline is used for IFRS9 modeling and Effective term Calculation.

Research Objective :

Predict the survival probabilities using continuous time models

Motivation : Survival Function and Hazard Function

- **Survival analysis** - a series of observations and attempts to estimate the time it takes for an event of interest to occur.

- **Survival function :** probability that a subject of interest will survive beyond time t, or equivalently, the probability that the duration will be at least **'t'**

$$\mathbf{S(t)=Pr(T>t)}$$

- **Hazard function :** h(t) gives the probability of the default occurring at time t, given that the subject did not experience the default event until time t.

$$\mathbf{h(t)=\lim_{\delta t\to\infty}\left(\frac{Pr(t\leq T\leq t+\delta t\,|\,T>t)}{\delta t}\right)}$$

- **Objective of survival analysis -** Estimate and interpret survival and/or hazard functions from survival data.

Fig 1 : Survival Function Curve*

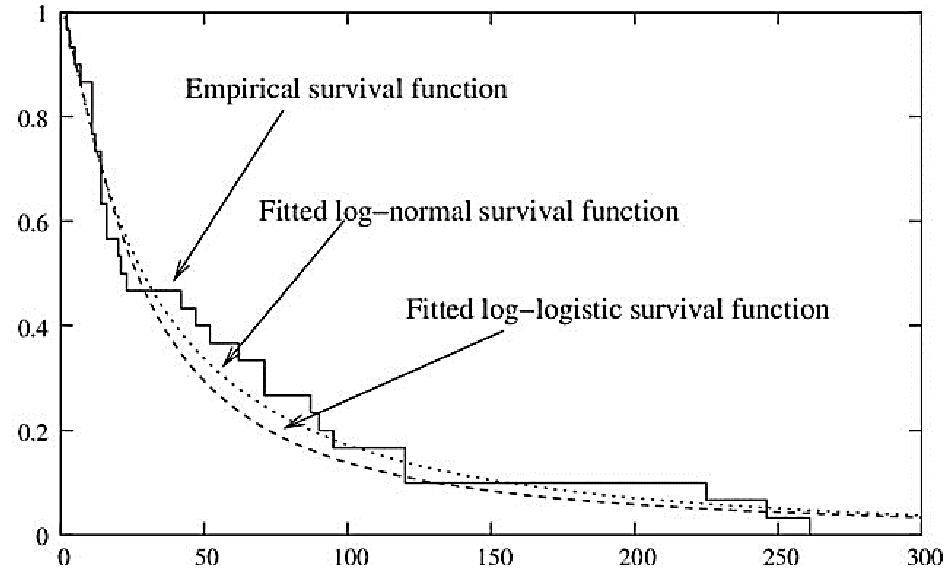
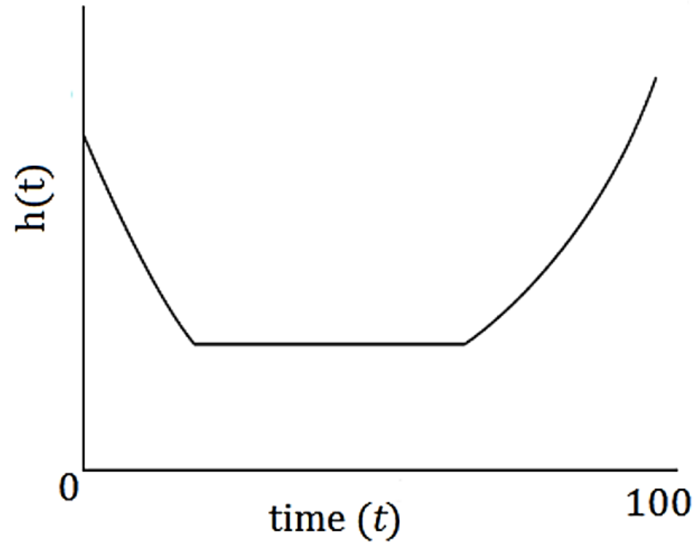


Fig 2 : Hazard Function Curve*



Motivation : Standard methods in Survival Analysis

Non-parametric, Semi-parametric, and Parametric

1. **Kaplan-Meir**
2. **Cox Proportional Hazard Regression Model**

1. **Kaplan - Meier estimator:** break the estimation of the survival function S(t) into smaller steps depending on the observed event times.

   For each interval, the probability of surviving until the end of this interval is calculated.

$$\hat{S}(t) = \prod_{j:t_j \le t} \left[1 - \frac{d_j}{n_j} \right]$$

Assumptions:

- Both censored and defaulted - are used in estimation.
- No cohort effect on survival.
- The survival probability is equal for all subjects.

Fig 3 : Code Snippet Kaplan Meir estimator

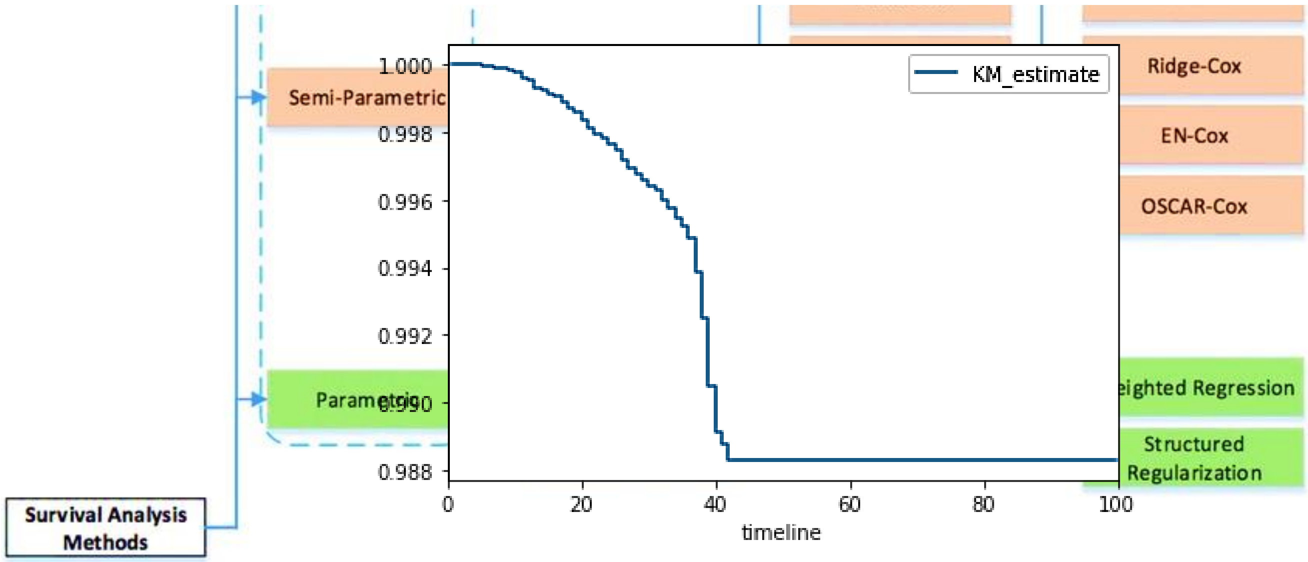
Fig 4 : Survival Analysis Methods*

```
from lifelines import KaplanMeierFitter
kmf = KaplanMeierFitter()

T = data_cox["exposure1"]
E = data_cox["odefind"]

kmf.fit(T, event_observed=E)
```

```
<lifelines.KaplanMeierFitter:"KM_estimate", fitted with 100325 total observations, 99517 right-censored observations>
```



Motivation : Standard methods in Survival Analysis

Disadvantages: KM cannot estimate survival probabilities considering all covariates in the data (univariate approach)

2. Cox Proportional Hazard : not only time and censorship features but also additional data as covariates. The dependent variable is expressed by the hazard function (or default intensity) as follows

$$h(t;x) = h_0(t)\exp\{\beta_1 x_1 + \ldots + \beta_k x_k\}$$

Cox PH's proportionality assumption

Proportionality property does not hold due to a violation of some covariates.

3. Time-varying Cox regression

Time-varying covariates can be included in the survival models.

Predicted hazard depends on later values of the covariate than the value of the covariate at the baseline.

Fig 5 : Code Snippet Time varying cox proportion

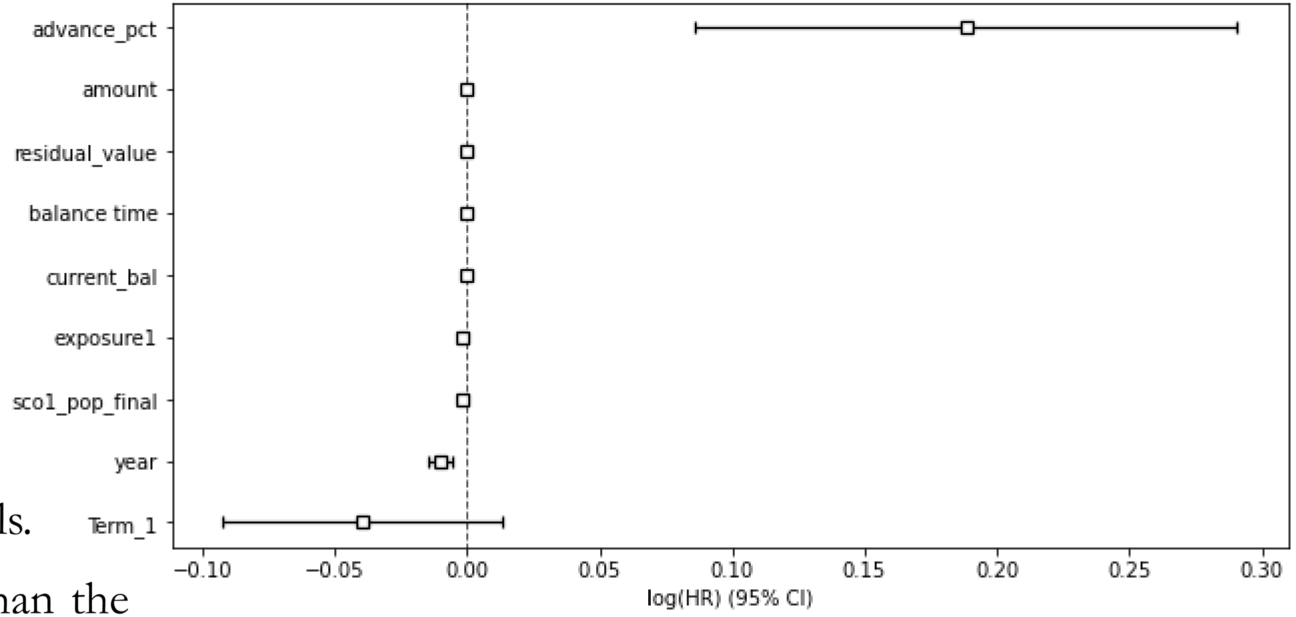
```
from lifelines import CoxTimeVaryingFitter

cox_tv = CoxTimeVaryingFitter(penalizer=0.1)
cox_tv.fit(data_cox, id_col='id', event_col='odefind',start_col='start', stop_col='stop', show_progress=True, step_size=1)

#Iteration 30: norm_delta = 0.00000, step_size = 1.00000, ll = -179884.48600,
#newton_decrement = 0.00000, seconds_since_start = 207.0
#Convergence completed after 30 iterations.

Iteration 3: norm_delta = 0.00000, step_size = 1.00000, ll = -5888.79485, newton_decrement = 0.00000, seconds_since_start = 0.2
Convergence completed after 3 iterations.

<lifelines.CoxTimeVaryingFitter: fitted with 100325 periods, 100325 subjects, 808 events>
```



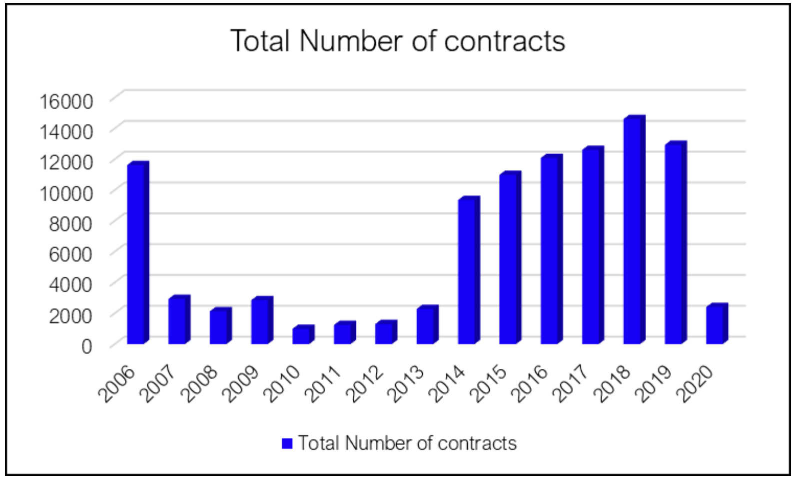
Data

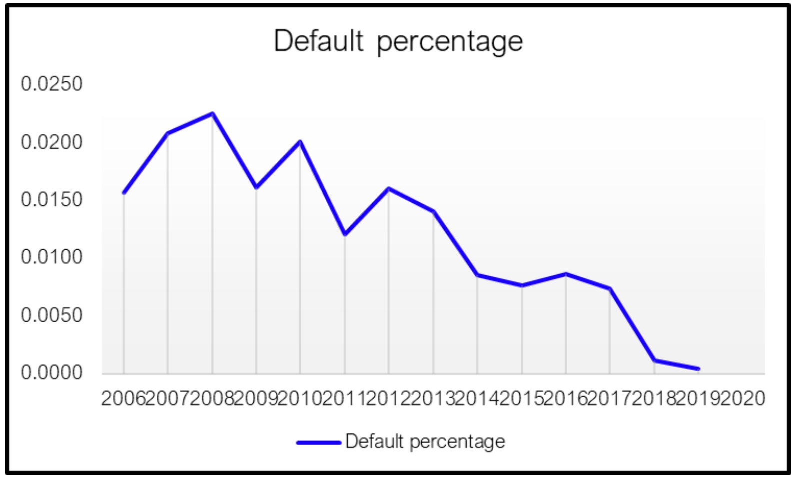
- Source : FCE's SCOPE system – Germany Ford contracts ; Sample Period : 2006-2020

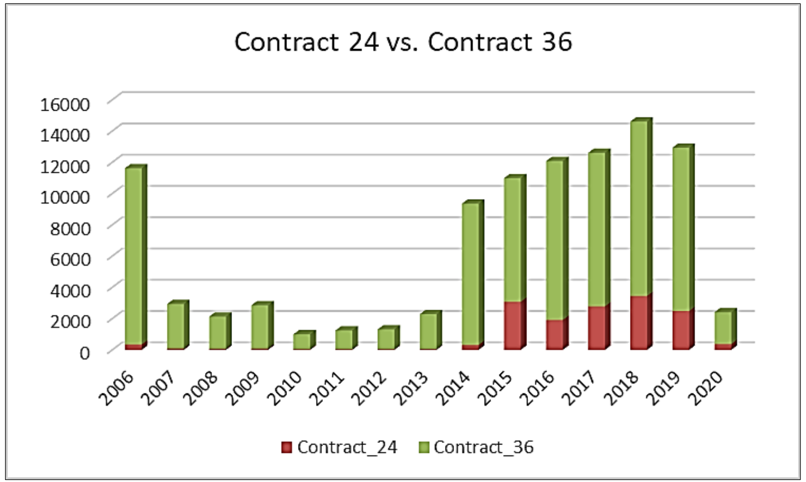
Variable	Description
Scope id	Borrower id
odefind	Default / payoff (goodzbad_indic = 7 or 8 / date_into_lru)
Amount	Asset value
Year	Origination year of the contract
Residual value	the estimated amount that an asset's owner would earn by disposing of the asset, less any disposal cost.
POP	Probability of Payment
Current balance	Outstanding balance at observation time 't'
Term	Contract term - 24, 36 months
Product / Business	FAF / Consumer
Exposure	Purchase date – Date of observation
Survival Probabilities	the proportion of units that survive beyond a specified time
Balance time	Date of observation – Closure date
Advance_pct	Prepaid amount to value ratio at observation time (%)

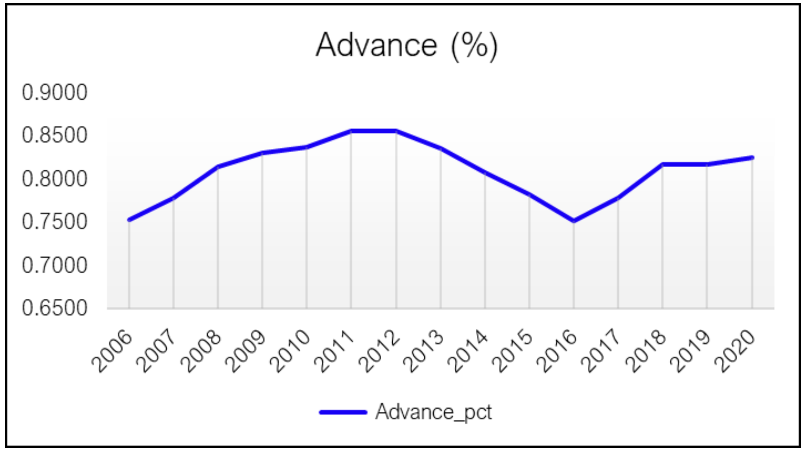
Data

- No. of Observations: 1,00,325 contracts









Methodology : DeepSurv

- Generalization of the Cox proportional hazards model with only a single hidden layer (Farragi and Simon, 1995).
- Learn relationships between primary covariates and the corresponding hazard risk function.
- Farragi and Simon's network extended non-linearity quality.

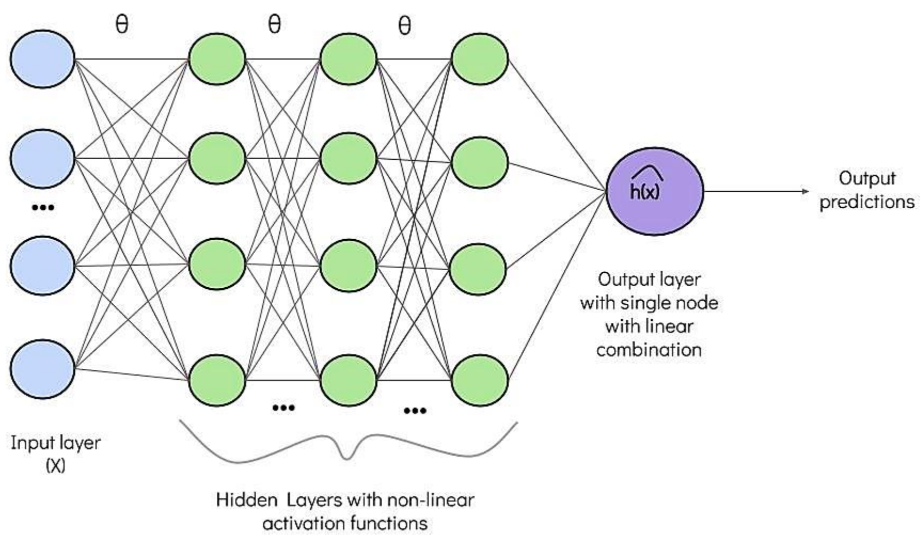
4. DeepSurv (Katzman et al.,2018)

-Handle the main proportional hazards constraint.

- Estimate the log-risk function h(X) with the CoxPH model and the linear combination of static features from given data X and t the baseline hazards.

- DeepSurv is **a deep feed-forward neural network** which estimates each individual contracts' effect on their hazard rates with respect to parametrized weights of the network θ.

Fig 6 : DeepSurv - deep feed-forward neural network



Likelihood with parametrized weights β

$$Lc(\beta) = \prod i : ei = 1 \frac{exp(\hat{h\beta}(xi))}{\sum j \in R(ti) exp(\hat{h\beta}(x_j))},$$

Loss function

$$l(\theta) = -\frac{1}{Ne = 1} \sum i : ei = 1(\hat{h\beta}(xi) - log \sum j \in R(ti)(e^{\hat{h\beta}(x_j)})) + \lambda * ||\theta||_2^2,$$

Implementation

- Split survival dataset into train, test, validation subsets (64%,16%,20%), then standardize the given data (only the continuous variables) since our output layer is a linear Cox regression activation.
- Building the Vanilla MLP with four hidden layers,
- Batch normalization (for stabilization and reducing data noise),
- Dropout 40% between the hidden layers,
- ReLU were chosen as an optimal activation layer (alternatively, Scaled Exponential Linear Units (SELU) was implemented),
- Adam optimizer was used for model training, without setting initial learning rate value.
- the learning rate was too high and, hence, we put a 0.0001 value, in order to improve the performance:

Fig 7 : Learning rate vs. Batch Loss

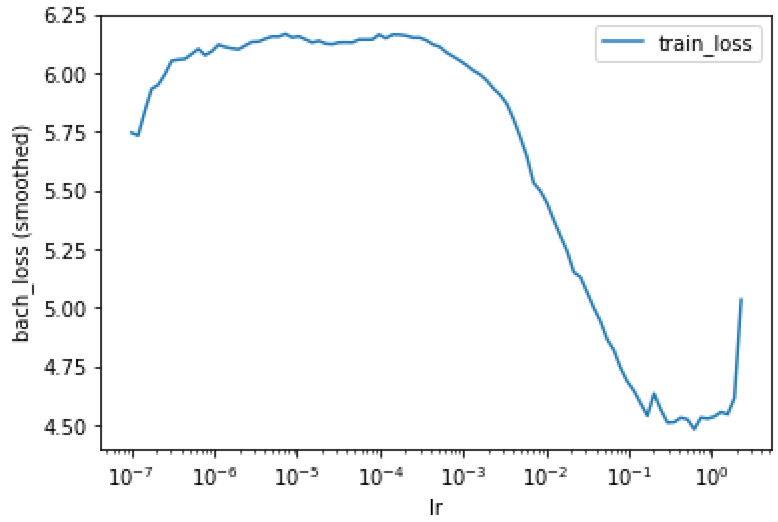
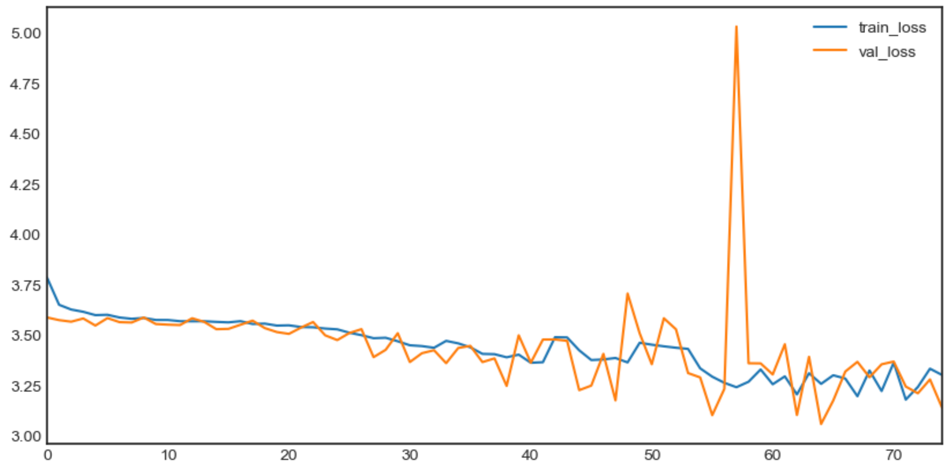


Fig 8 : Training Loss Vs. Validation Loss



Results

- Survival curve represents a statistical graphical interpretation of the survival behavior of subjects.
- **Fig.9** Estimated survival lifetimes for 100 contracts from the test dataset using the output of the DeepSurv model.
- Contracts with term 24/36, the predicted survival times decrease within the first two/ three years.
- Flatten part at t≈40 months. Maturity time of the contracts.
- **Concordance index**

  - model's ability to correctly provide a reliable ranking of the survival times based on the individual risk scores

  - concordance is that a subject that dies at time t should have a higher risk at time t than a subject who survives beyond time t.

- C-index : > 95.6 %

Fig 9 : Survival Probabilities

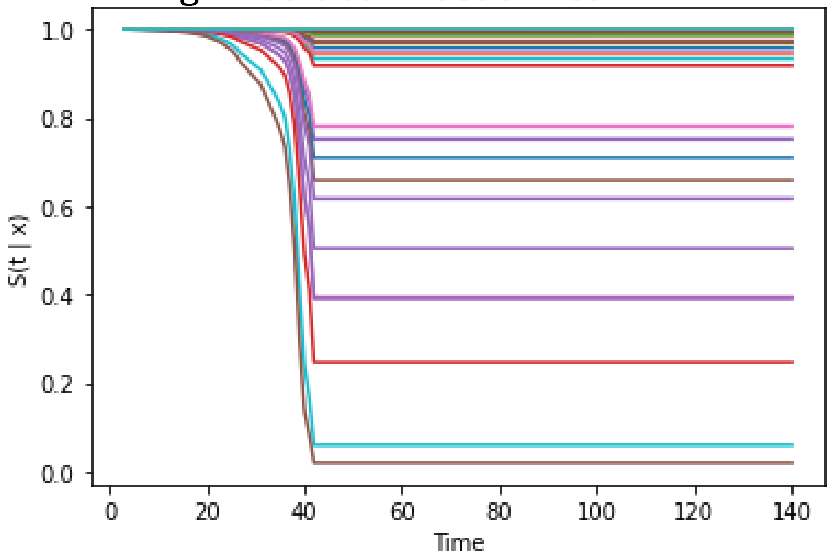


Fig 10 : Hyperparameter Tuning

Parameters	Value
Learning rate	0.0001
Batch size	{64,128,256}
Nodes per layer	{32,64,128,256}
Optimizer	Adam, SGD
Max (C-index) ≈97.06%	

Robustness Tests:

5. Cox-CC

- Proportional version of the Cox-Time model
- Case-control approximation
- C-index ≈79.8%

Fig 11 : Training Loss Vs. Validation Loss

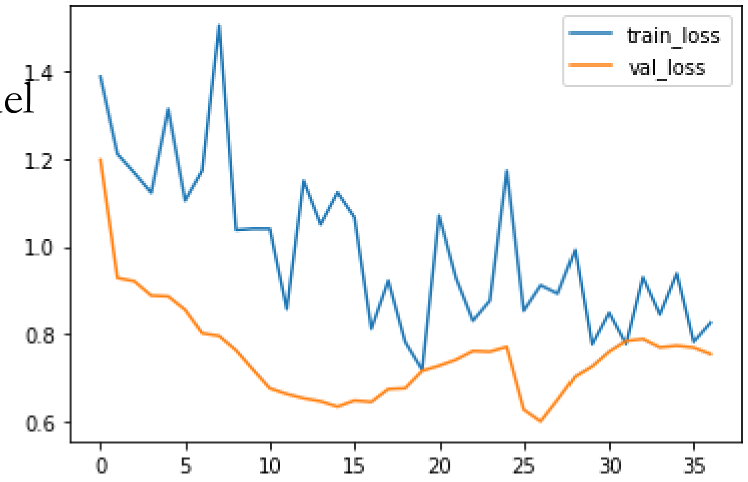
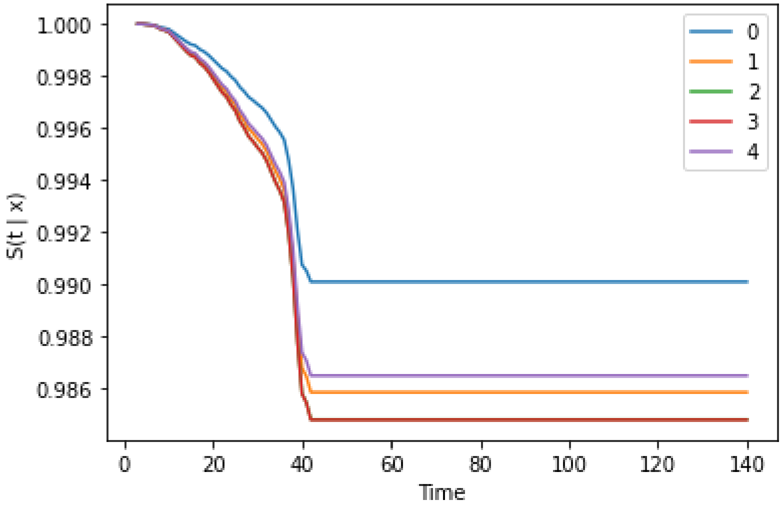


Fig 12 : Survival Probabilities



6. Piecewise Constant Hazard (PC-Hazard) model

- The hazard function value comes between defined pieces or interval time in months of all contracts
- Significant factors- estimates obtained by means and hazard function at various time points.

$$h(t) = h_0 \sum_{l=0}^{m} g_l I_l(t) \quad \text{with} \quad I_l(t) = \begin{cases} 1 & \text{if} \quad \tau_l \leq t < \tau_{l+1} \\ 0 & \text{if} \quad \text{elsewhere} \end{cases} .$$

Fig 13 : Training Loss Vs. Validation Loss

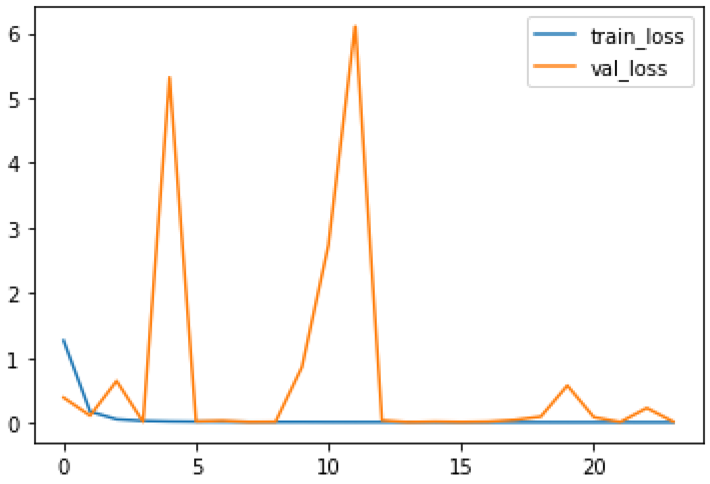
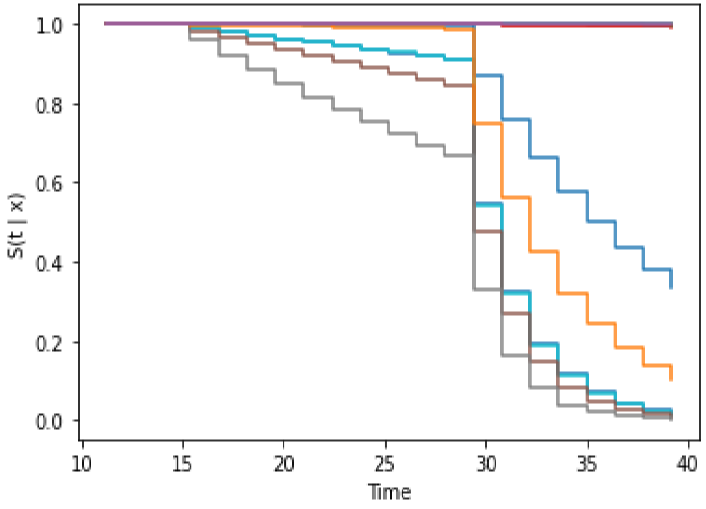


Fig 14 : Survival Probabilities



GDI&A
Global Data Insight & Analytics

Summary

Continuous time models	Description
CoxTime	Cox-Time is a relative risk model that extends Cox regression beyond the proportional hazards
CoxCC	Cox-CC is a proportional version of the Cox-Time model
CoxPH (DeepSurv)	CoxPH is a Cox proportional hazards model also referred to as DeepSurv
PC-Hazard	The Piecewise Constant Hazard (PC-Hazard) model assumes that the continuous-time hazard function is constant in predefined intervals.

Model	C-index
Deep Surv	**97.06 %**
Cox-CC	**79.08%**
(PC-Hazard) model	**99.08%**

Future work

1. Modern recommendation for regularization is to use early stopping with dropout and a weight constraint.
   Overfitting in regression models for time-to-event data- Extend the dataset
2. **Discrete time models**
   Deep hit model- deep neural network that learns the distribution of survival times. Single risk and multiple competing risks
3. Moving to SAS Viya – Experiment different Deep Learning architectures to predict survival probabilities

References

1. Kvamme, H., Borgan., & Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. *arXiv preprint arXiv:1907.00825*.
2. Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology,* 18(1), 2018.
3. Laura Antolini, Patrizia Boracchi, and Elia Biganzoli. A time-dependent discrimination index for survival data. *Statistics in Medicine*, 24(24):3927–3944, 2005.
4. Changhee Lee, William R Zame, Jinsung Yoon, and Mihaela van der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. *AAAI Conference on Artificial Intelligence*, 2018.

Thank You