

AIG Interview Presentation: Motor Third-Party Liability Claims Analysis and Prediction

Yi-Pei Chan

25 Jan. 2021

AIG Interview
Presentation:
Motor
Third-Party
Liability Claims
Analysis and
Prediction

Yi-Pei Chan

Project Concept

Data Exploration

The Dataset

Data Visualization

Model &
Prediction

Poisson GLM

Model &
Prediction

Poisson GLM

Poisson Lasso & Ridge

Gradient Boosting
Model

Final Validation

Q & A

Link to complete code and analysis :
<https://yipeichan.github.io/claims.html>

Project Concept

Data Exploration

The Dataset

Data Visualization

Model &
Prediction

Poisson GLM

Model &
Prediction

Poisson GLM

Poisson Lasso & Ridge

Gradient Boosting
Model

Final Validation

Q & A

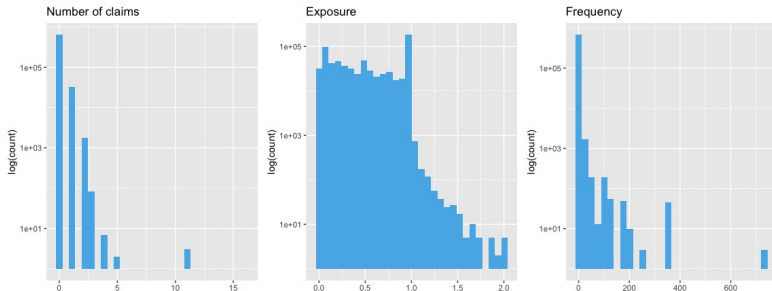
**AIG Interview
Presentation:
Motor
Third-Party
Liability Claims
Analysis and
Prediction**

Yi-Pei Chan

- Project Concept
- Data Exploration
 - The Dataset
 - Data Visualization
- Model & Prediction
 - Poisson GLM
- Model & Prediction
 - Poisson GLM
 - Poisson Lasso & Ridge
 - Gradient Boosting Model
- Final Validation
- Q & A

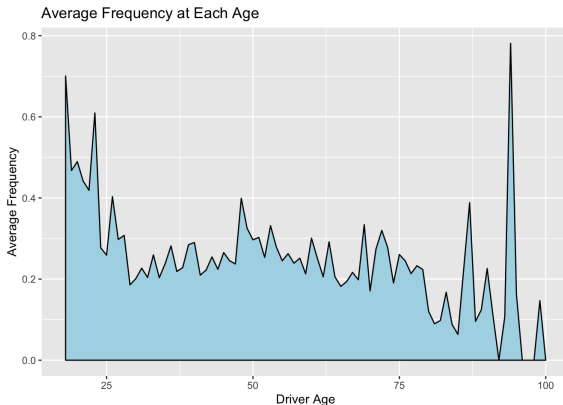
Data Exploration - Visualization

- ▶ Among the 678,013 policies, there were 34,060 filed claims, i.e. 5.02% notified claims.
- ▶ Potential Problems :
 1. Mean should equal to Variance in Poisson distribution
⇒ Use Negative binomial if Overdispersed
 2. More 0s than are expected in Poisson regression ?
⇒ Incorporate the logit model for predicting excess 0s
 3. Varied exposure periods (observations not comparable)
⇒ Add offset of Exposure term to the model



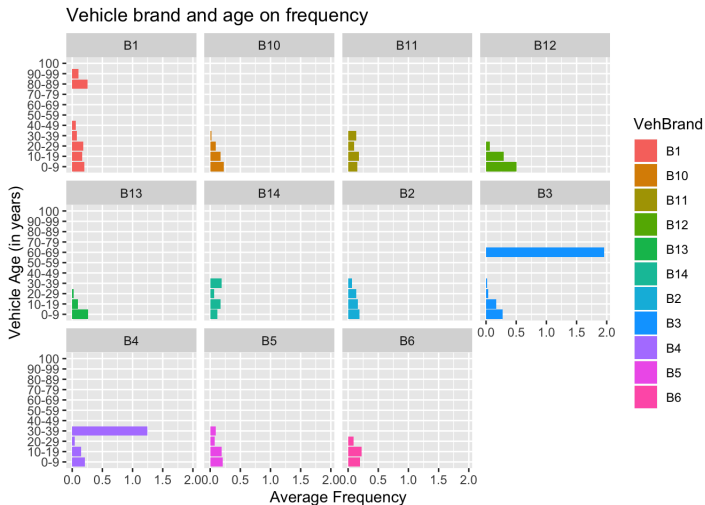
Data Exploration - Visualization

- ▶ Exposure : the duration of the insurance coverage
- ▶ Claim frequency : claim count per unit of exposure
- ▶ Did driver age influence frequency?
 1. The highest mean frequency happens at age 94
 2. Drivers between age 18 to 23 tends to have higher mean frequency



Data Exploration - Visualization

► Did vehicle brand and age influence frequency?



AIG Interview
Presentation:
Motor
Third-Party
Liability Claims
Analysis and
Prediction

Yi-Pei Chan

Project Concept

Data Exploration

The Dataset

Data Visualization

Model &
Prediction

Poisson GLM

Model &
Prediction

Poisson GLM

Poisson Lasso & Ridge

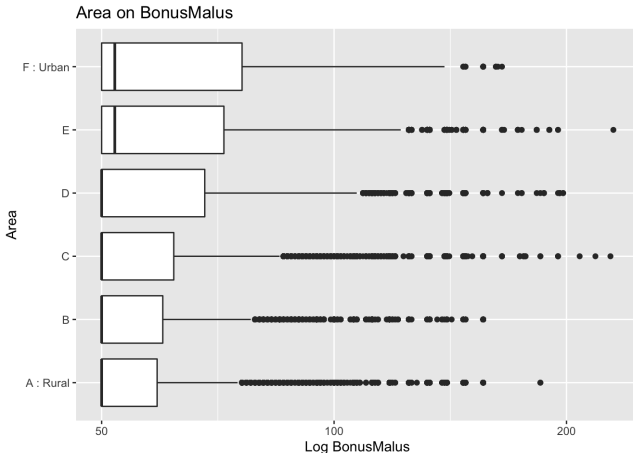
Gradient Boosting
Model

Final Validation

Q & A

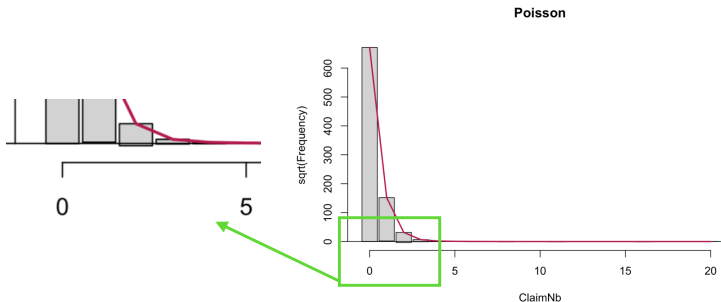
Data Exploration - Visualization

- What is the relationship between area and bonus-malus?



Model and Prediction - Poisson GLM

- ▶ Hanging rootogram :
Only 2 count is a little under predicted



Model & Prediction - Poisson Lasso & Ridge Regression

```
glm.ridge$lambda.min; coef(glm.ridge, s = "lambda.min")
```

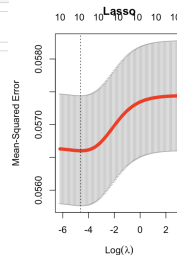
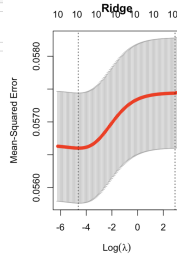
```
## [1] 0.009804138
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"  
##              1  
## (Intercept) -2.9270950727  
## Exposure    -1.0400993812  
## VehPower     0.0061349023  
## VehAge      -0.0263678397  
## DrivAge      0.0060848768  
## BonusMalus   0.0169722817  
## VehBrand    -0.0010265539  
## VehGas       0.0502432492  
## Area         0.0169615264  
## Density      0.0194397589  
## Region      -0.0009442549
```

```
glm.lasso$lambda.min; coef(glm.lasso, s = "lambda.min")
```

```
## [1] 0.001635429
```

```
## 11 x 1 sparse Matrix of class "dgCMatrix"  
##              1  
## (Intercept) -2.696642397  
## Exposure    -1.193913018  
## VehPower     .  
## VehAge      -0.024586132  
## DrivAge      0.006071144  
## BonusMalus   0.017390359  
## VehBrand     .  
## VehGas       0.004379296  
## Area         .  
## Density      0.016603390  
## Region       .
```



AIG Interview
Presentation:
Motor
Third-Party
Liability Claims
Analysis and
Prediction

Yi-Pei Chan

Project Concept

Data Exploration

The Dataset
Data Visualization

Model &
Prediction

Poisson GLM

Model &
Prediction

Poisson GLM

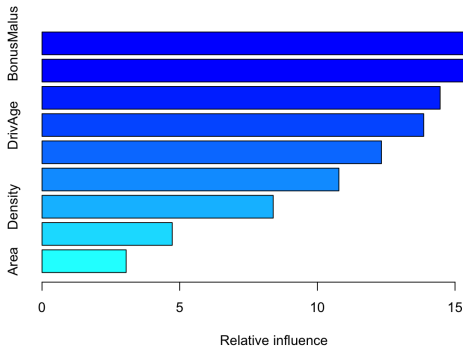
Poisson Lasso & Ridge

Gradient Boosting
Model

Final Validation

Q & A

Model & Prediction - Gradient Boosting Model



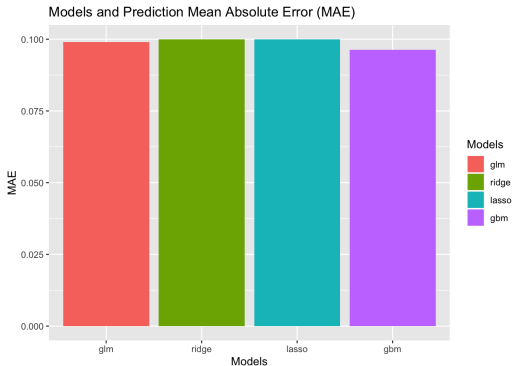
	var	rel.inf
BonusMalus	17.014808	
Region	15.372979	
VehAge	14.459134	
DrivAge	13.862481	
VehBrand	12.328304	
VehPower	10.782009	
Density	8.396521	
VehGas	4.728894	
Area	3.054871	

Final Validation

Use the test set to find the best fitting model

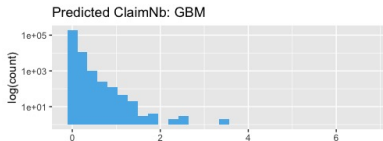
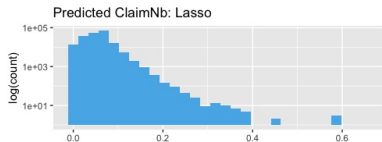
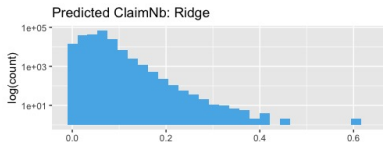
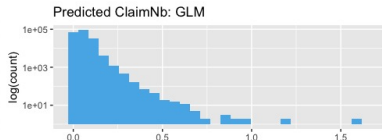
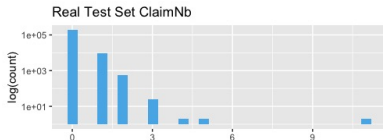
► The claim number prediction MAE for test set with

1. Poisson GLM : 0.09905573
2. Poisson Ridge GLM : 0.09988506
3. Poisson Lasso GLM : 0.09996999
4. Gradient Boosting Model : 0.09630762



Final Validation

Evaluation of the Predicted Number of Claims in the Test Set



AIG Interview
Presentation:
Motor
Third-Party
Liability Claims
Analysis and
Prediction

Yi-Pei Chan

Project Concept

Data Exploration

The Dataset

Data Visualization

Model &
Prediction

Poisson GLM

Model &
Prediction

Poisson GLM

Poisson Lasso & Ridge

Gradient Boosting
Model

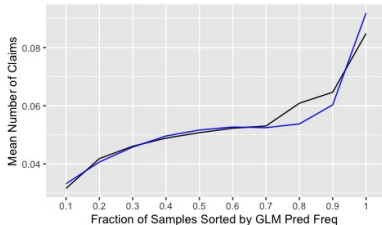
Final Validation

Q & A

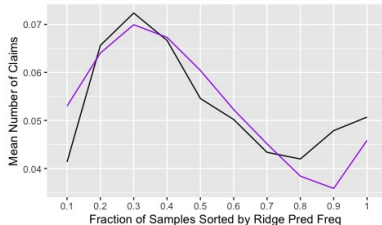
Final Validation

Evaluation of the Predicted Number of Claims in the Test Set

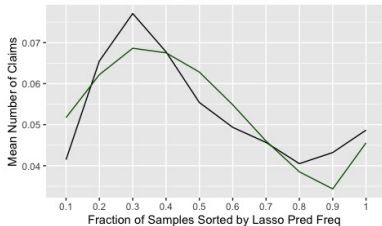
Real v.s. GLM Pred ClaimNb (blue)



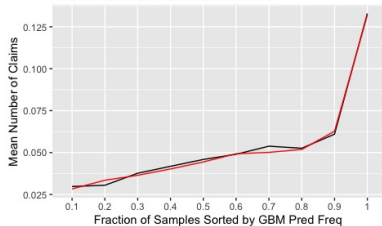
Real v.s. Ridge Pred ClaimNb (purple)



Real v.s. Lasso Pred ClaimNb (green)



Real v.s. GBM Pred ClaimNb (red)



Q & A

Link to complete code :

<https://yipeichan.github.io/claims.html>

Project Concept

Data Exploration

The Dataset

Data Visualization

Model &
Prediction

Poisson GLM

Model &
Prediction

Poisson GLM

Poisson Lasso & Ridge

Gradient Boosting
Model

Final Validation

Q & A