Kangaroo Auto Insurance Company Modeling Problem

Group 1: Riki Saito, Ved Piyush, Zhuoran Wang, and Zach Gunderson

Final Model (Two-Part Model)

Frequency: numclaims ~ offset(log(exposure)) + factor(agecat) + area + veh_value + veh_age + veh_value:veh_age + area:veh_value, (Family = Poisson, Link = Log)

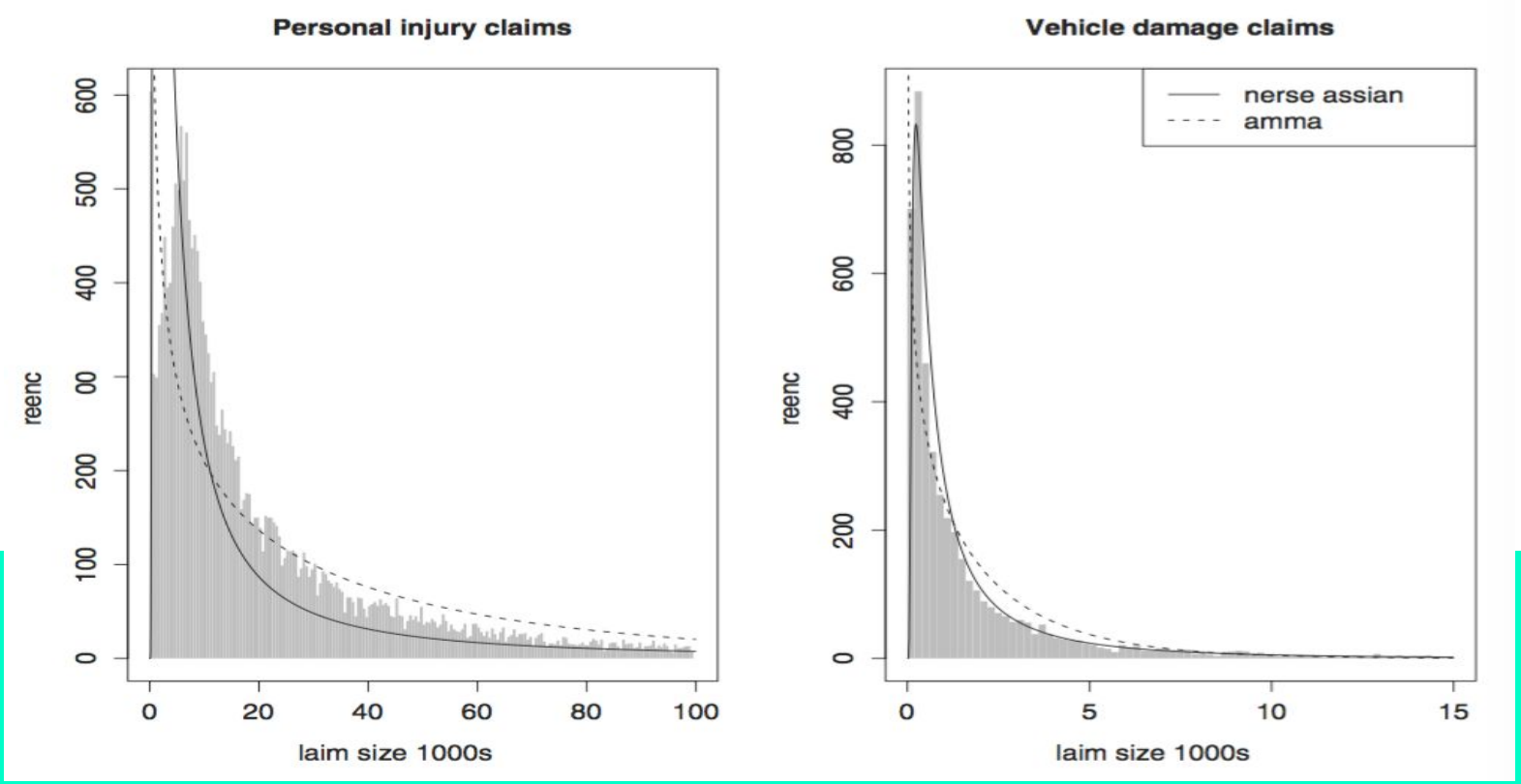
Severity: (claimcst0/numclaims) ~ gender + veh_age + agecat, (Family = Inverse Gaussian, Link = Log)

Loss (Claimcst0) = Frequency(Numclaims)*Severity(claimcst0/numclaims)

Methods considered

- Tweedie GLM
- Two-part Model
  - Frequency: Poisson (Weight = Exposure) vs. Poisson (Offset = Log (Exposure))
  - severity: Gamma GLM vs Inverse Gaussian

Gamma GLM vs Inverse Gaussian



Jong, Piet de and Heller, Gillain Z. Generalized Linear Models for Insurance Data. Cambridge. 2008.

Our Process

- Model Selection - Stepwise Selection, Trial-and-error, and Intuition
- Evaluation and Detecting overfitting - Cross-Validation
- Estimating Unbiased Model Coefficients - Bootstrapping

Concerns

- Few predictors in the severity model can lead to underfitting
- Cases where Veh_Value=0. Are these missing Values?
- Didn't consider Veh_Body because there are too many categories

Other variables that might be useful

- Exact age
- Driver history or record
- Credit score
- Income, # Children under 18 (Possible correlation with other variables)

Thank you for listening!