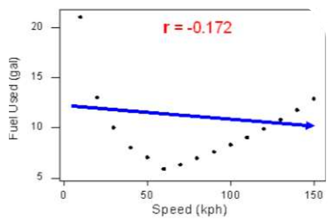
Applying Quantile Regression to Ratemaking: A Measured Approach

This study explores the use of telematics data to understand the relationship between driver behavior and insurance claims. It focuses on quantile regression techniques for modeling claim frequency, providing insights for insurance ratemaking.

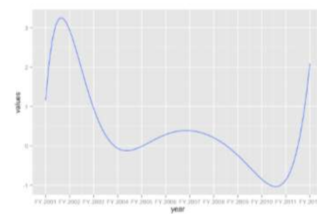
Market Modeling Approach: GLMs and GAMs

Widely adopted modeling techniques in the actuarial field, providing a powerful framework for pricing:



GLMs

•Extension of linear regression, allowing for non-normal response distributions and non-linear relationships through link functions.
•Handle diverse data structures by specifying appropriate distribution families (e.g., Gaussian, Poisson, Gamma, etc.).

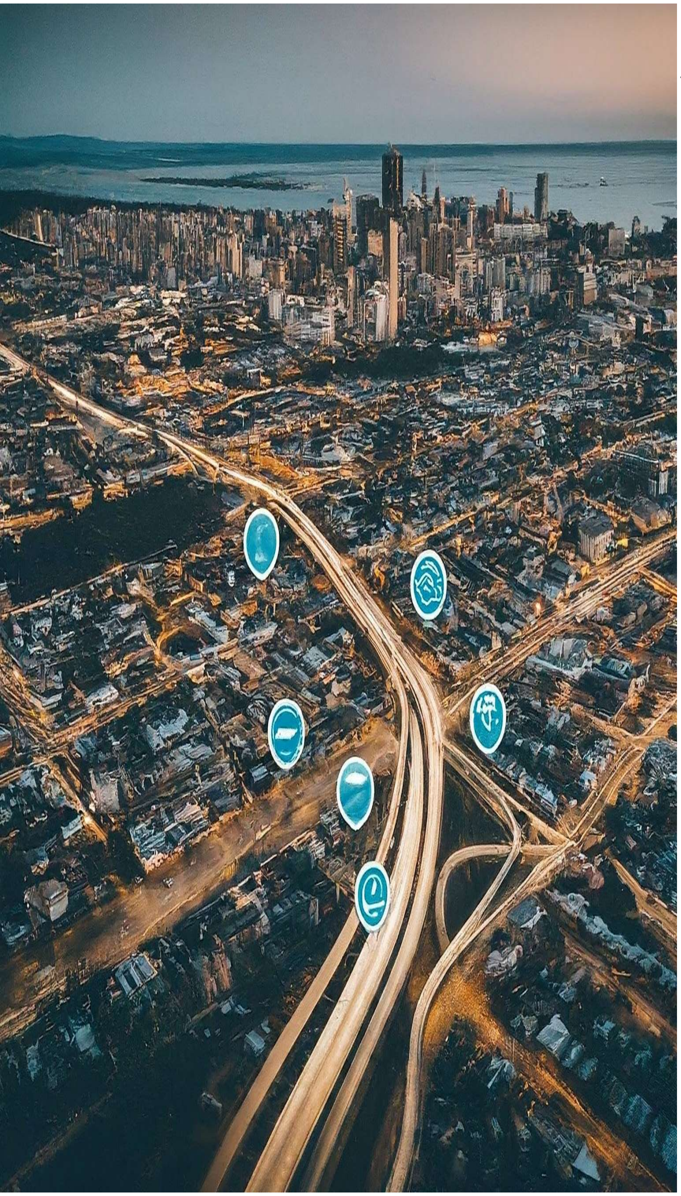


GAMs

•Further extension of GLMs, offering greater flexibility through the use of smooth functions (e.g., splines) to model complex, non-linear relationships.
•Non-parametric approach that allows for automatic identification of interactions and nonlinearities.

Limitations and Considerations:

- Distribution assumptions: The need for specifying a response distribution may not always accurately capture the underlying data. (Solution: testing several distributions)
- Sensitivity to outliers: GLMs and GAMs can be sensitive to extreme observations, leading to biased estimates. (Solution Penalized approaches)
- Interpretability: While GAMs provide increased flexibility, their complexity can make interpretation and communication of results more challenging.



Addressing Selection Challenges: Insights from Quantile Regression

Tailored pricing

Introduced by R.Koenker and G. Bassett Jr. in their 1978 paper, "Regression Quantiles”

Enhance Selection

Uncovers hidden relationship

Data Introduction

The data set used in the analysis contains 6,006 observations and describes the count of claims for automobile insurance

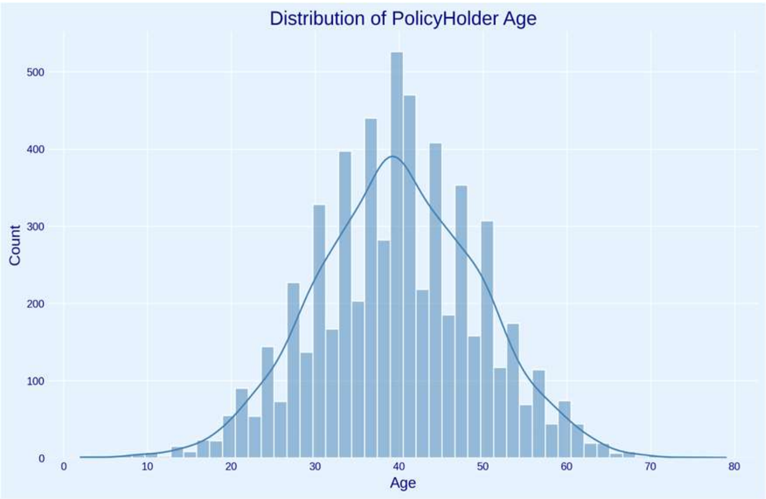
Standard Data

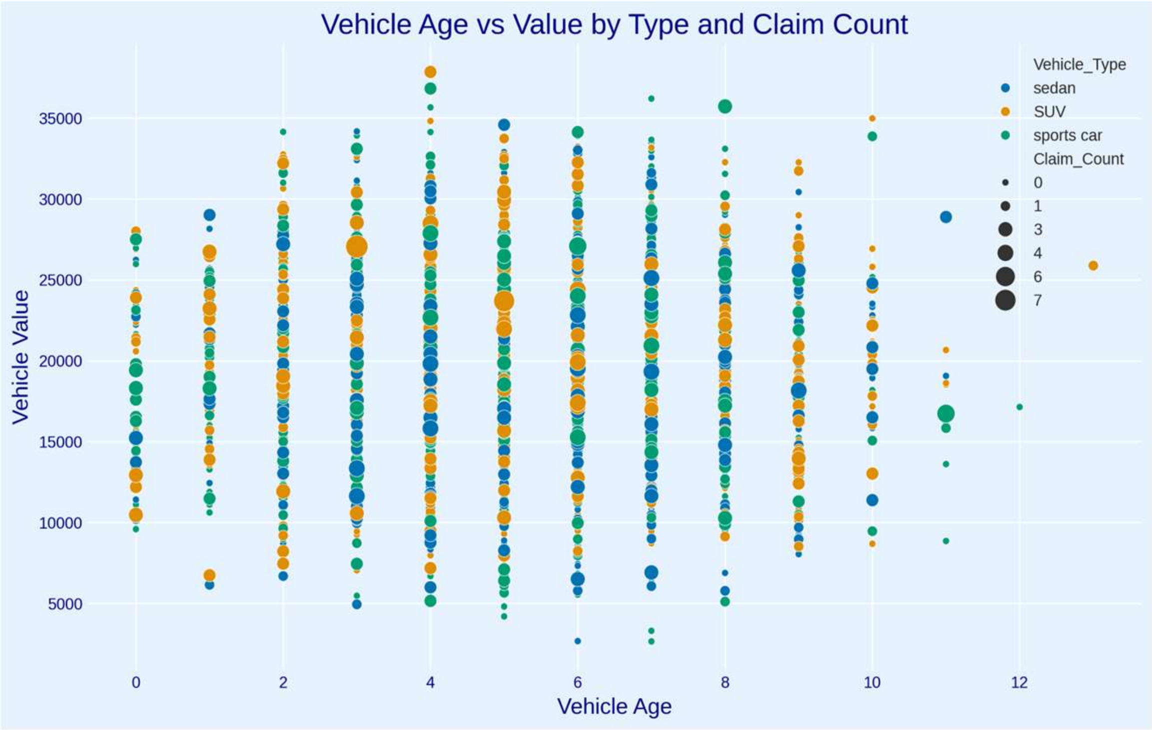
- Driver Age
- Location
- Vehicle_Age
- Vehicle_Type
- Vehicle_Value

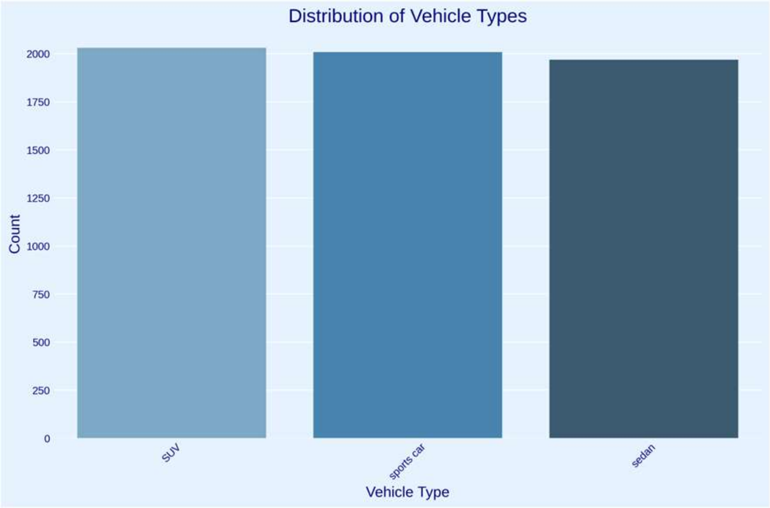
Telematics Data

- Harsh_Accel
- Harsh_Brakes
- Night_Driving
- Harsh_Lateral

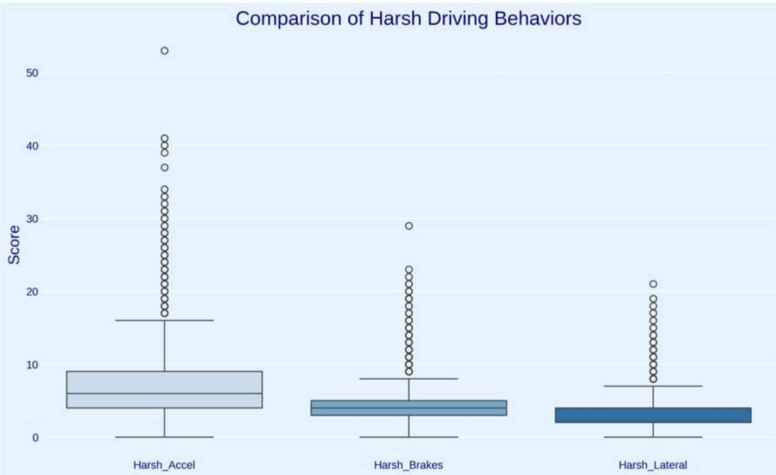
Classic Variables

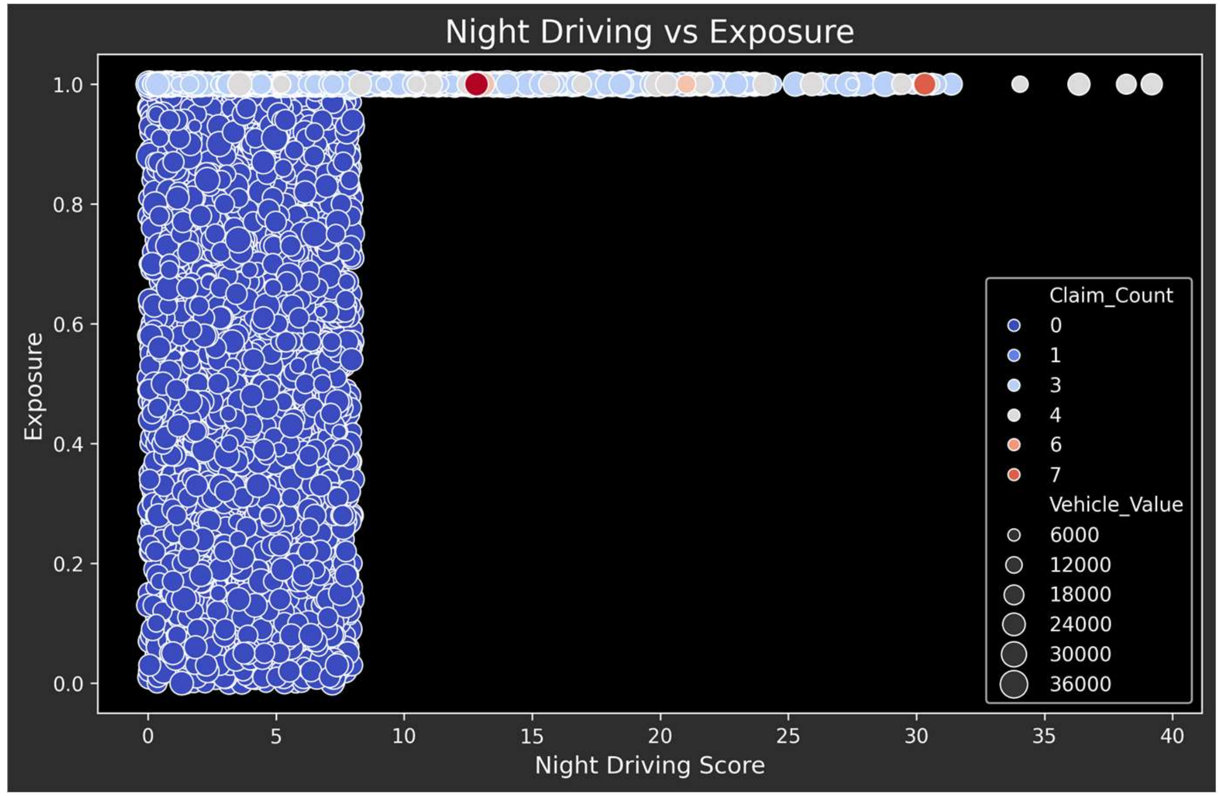




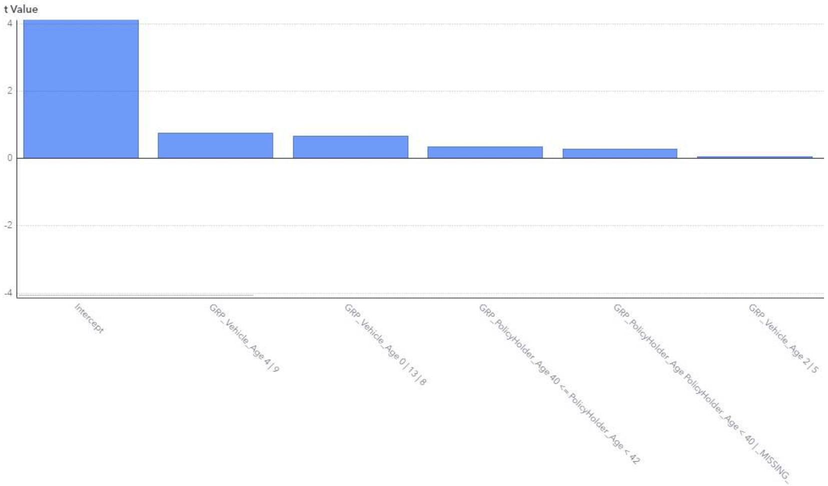


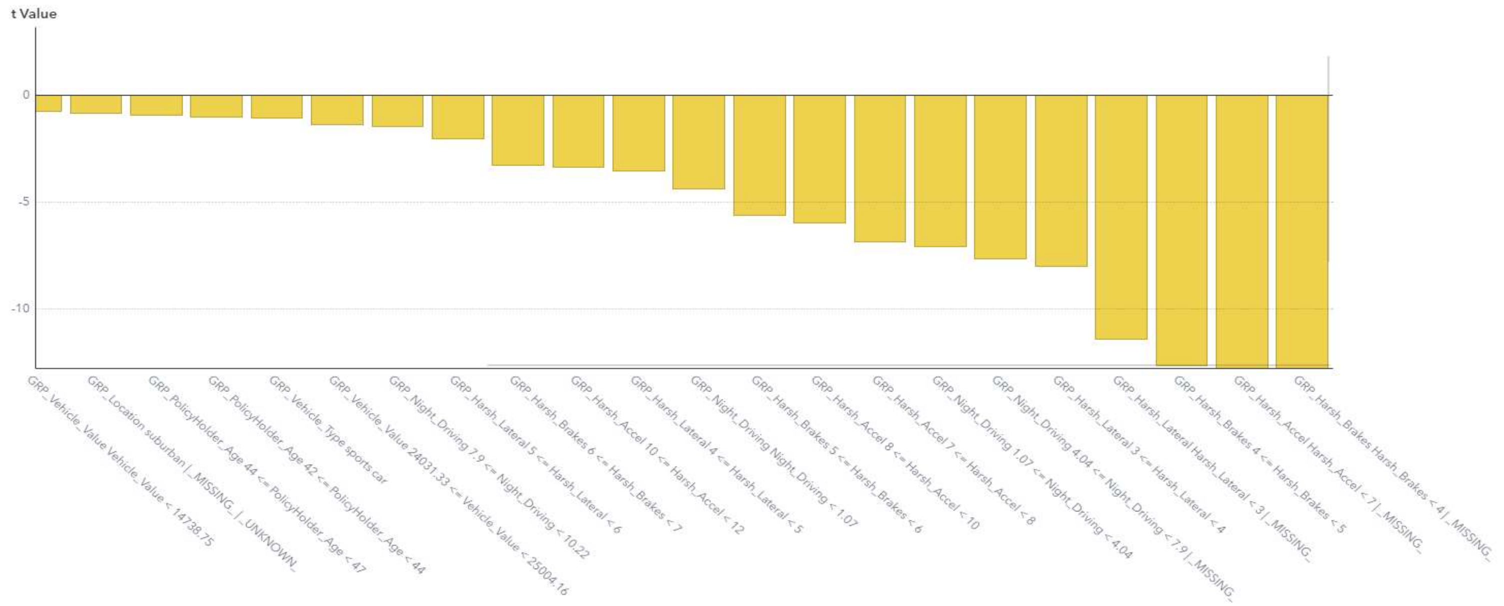
Telematics Variables



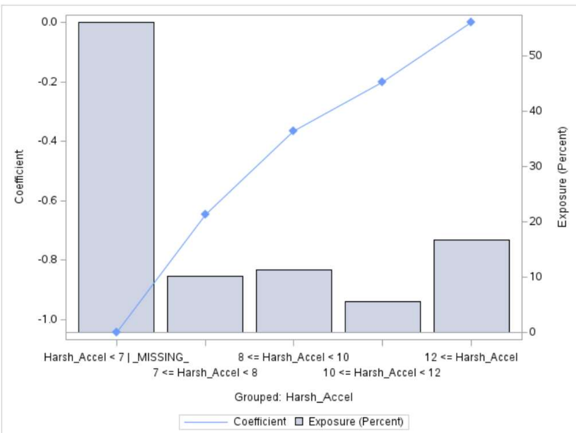


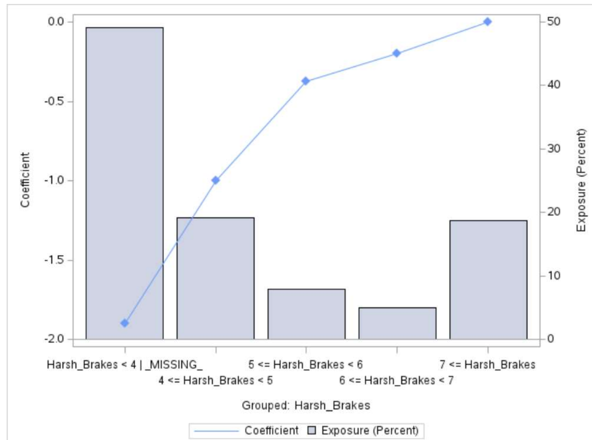
GLM Factor Importance

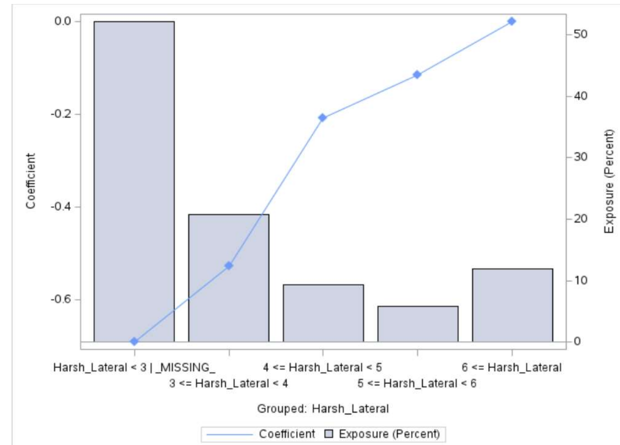


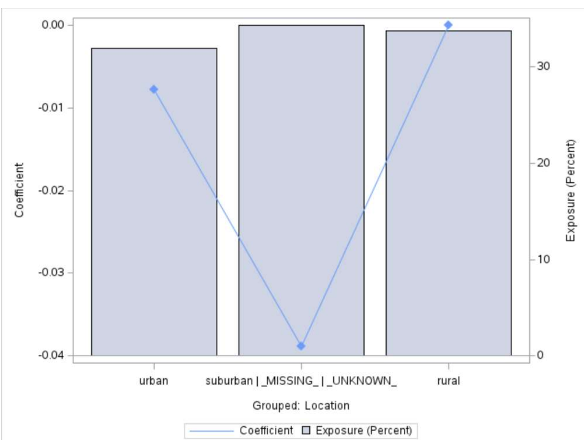


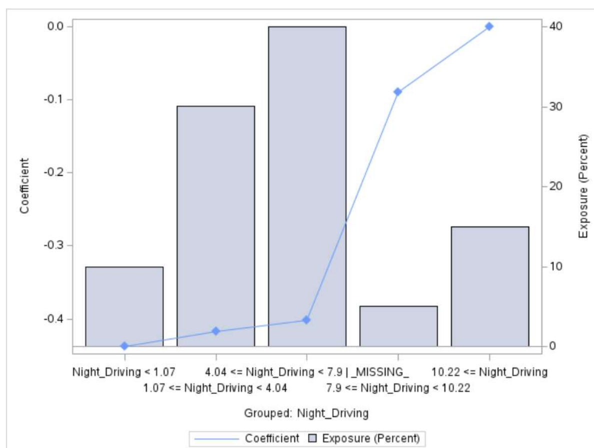
GLM Coefficients

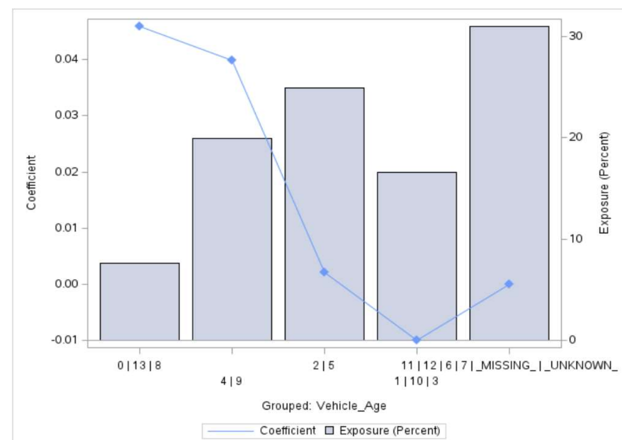






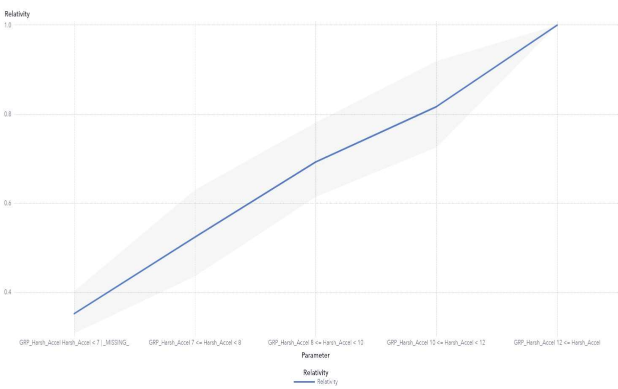




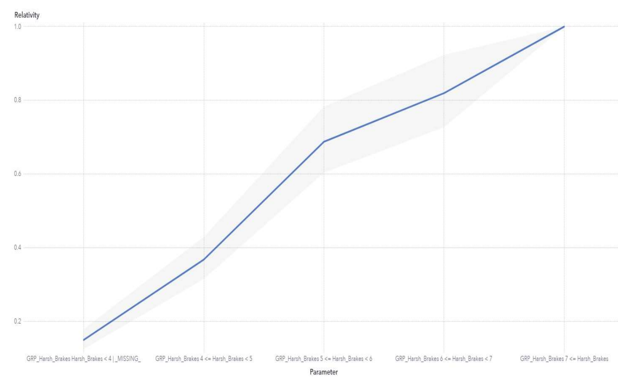


GLM Relativity Plots

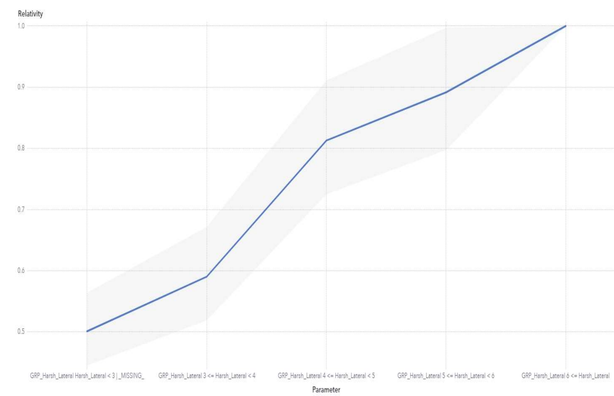
Acceleration



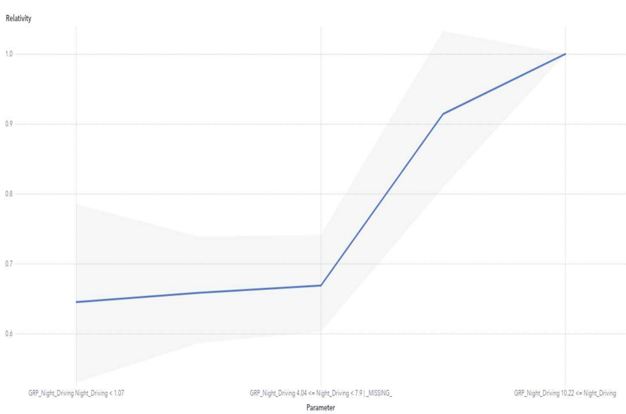
Brakes



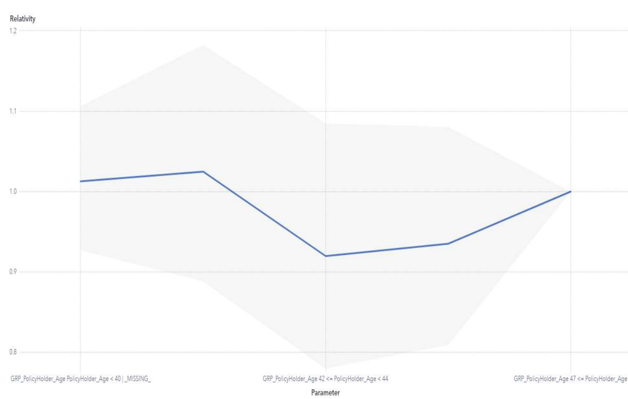
Lateral



Night Driving

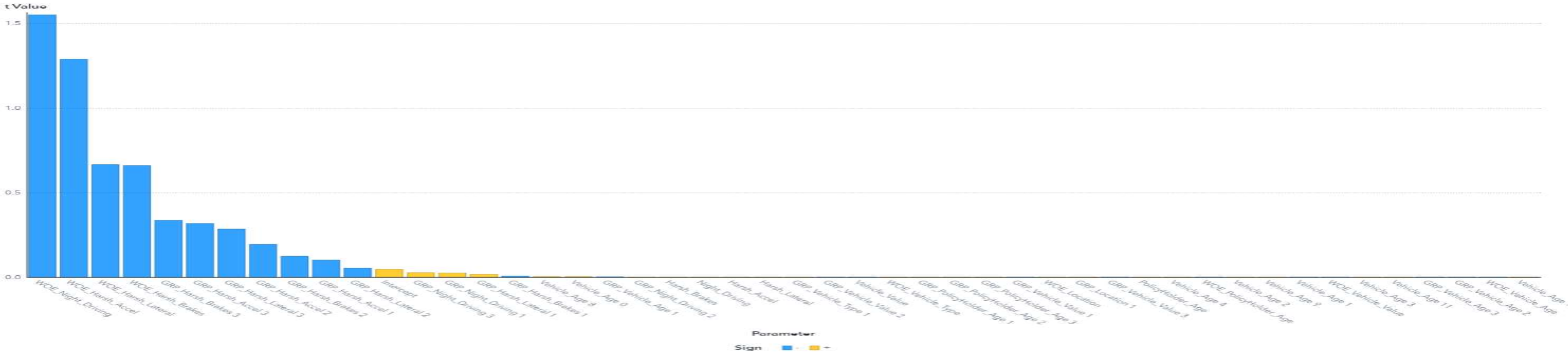


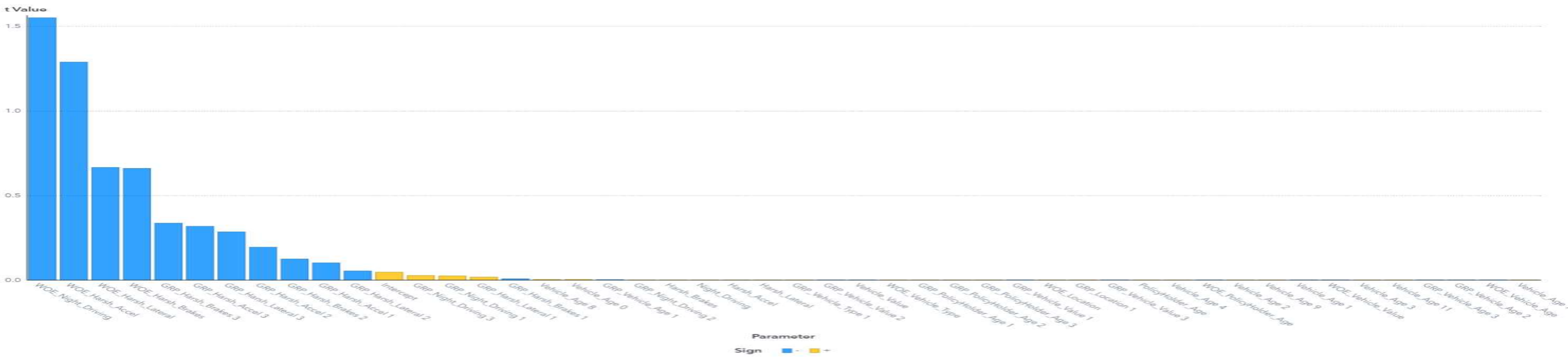
Policy Holder Age



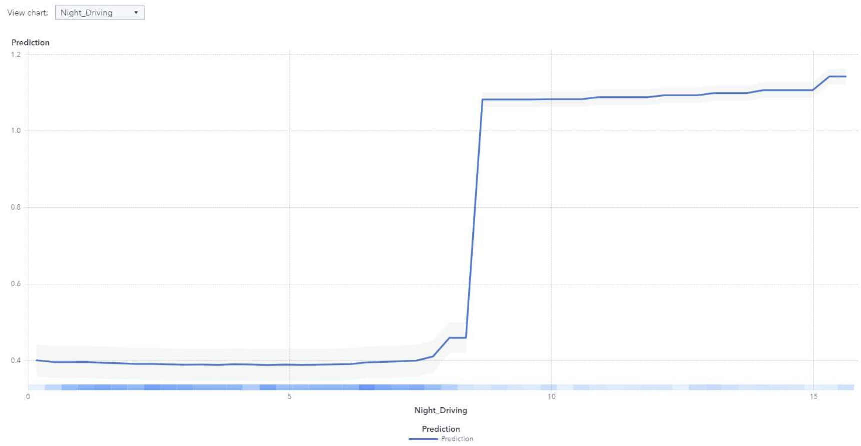
Vehicle Age

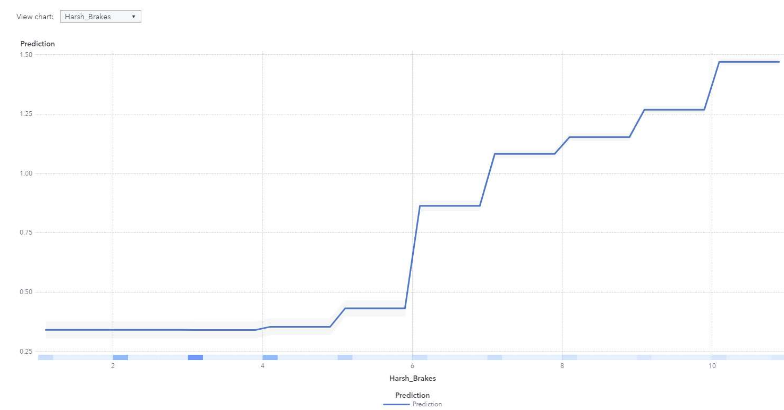
QR factor Importance

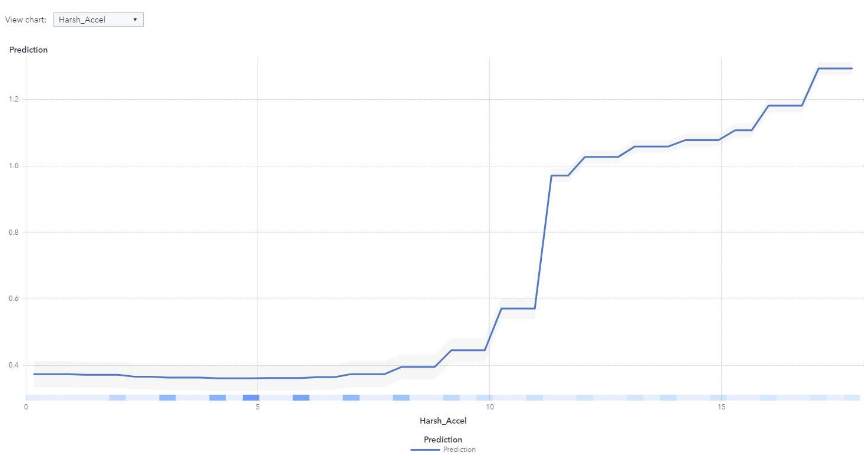


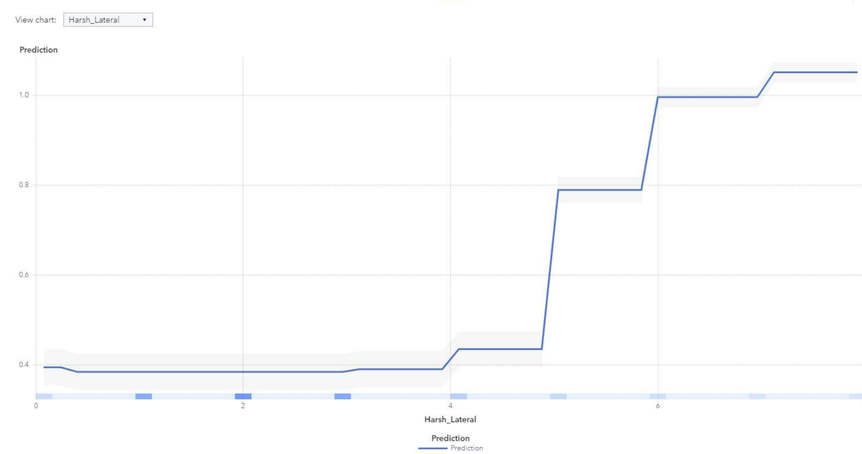


QR factor Interpretability





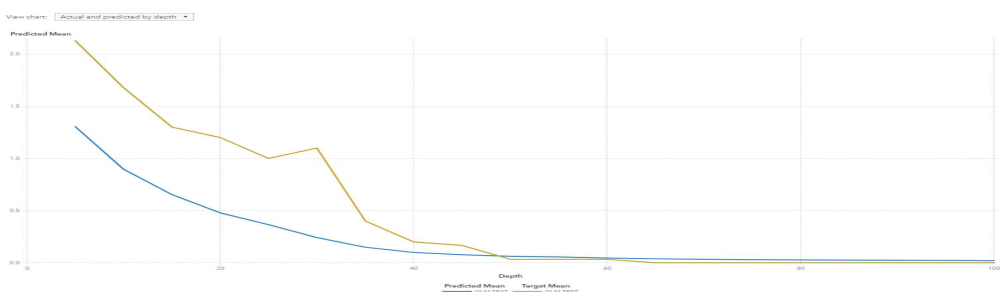


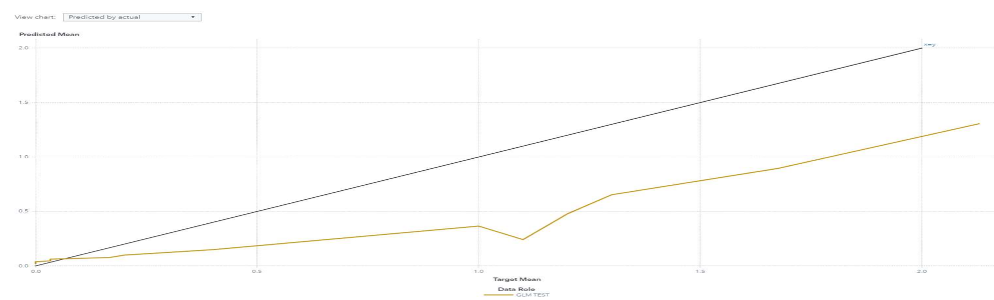


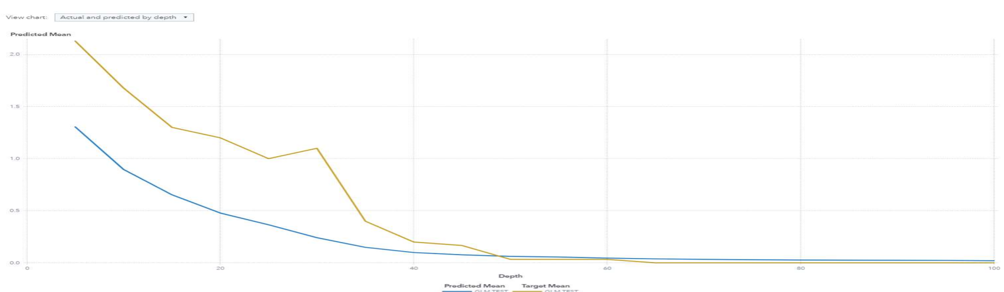
GLM VS QR

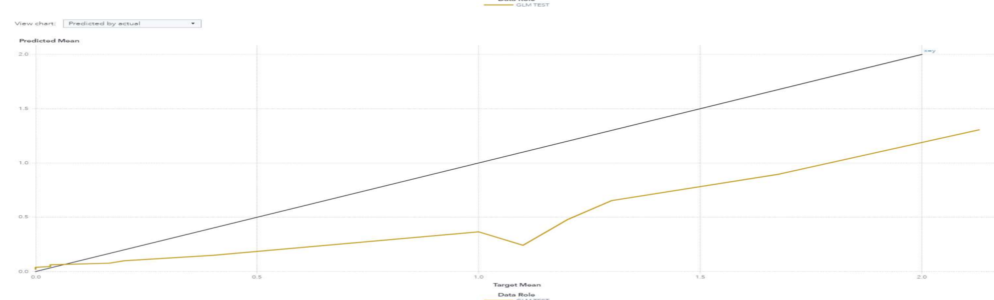
Model Comparison

Champion	Name	Algorithm Name	Average Squared Error	Root Average Squared Error	Root Mean Absolute Error	Root Mean Squared Logarithmic Error
★	QR 0.5	Quantile Regression	0.1534	0.3917	0.4133	0.1729
	GLM	Count GLM	0.3312	0.5755	0.5501	0.2686





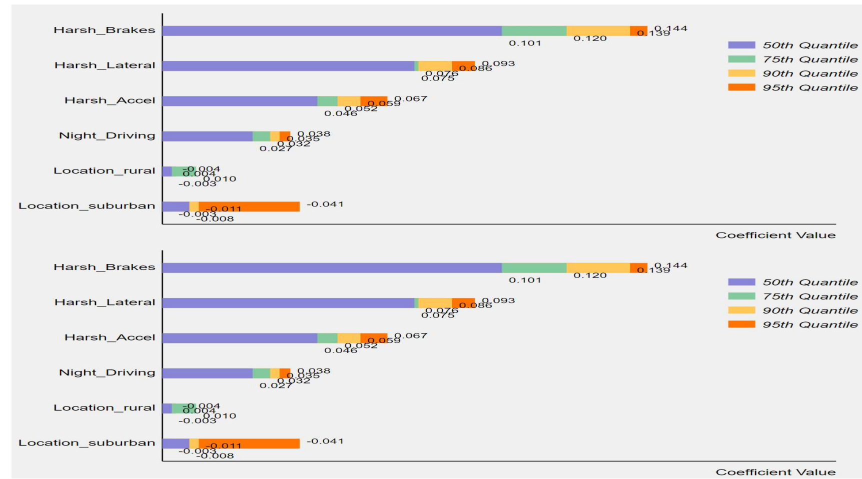




Quantiles Study and Observations

Coefficient Magnitudes Across Quantiles

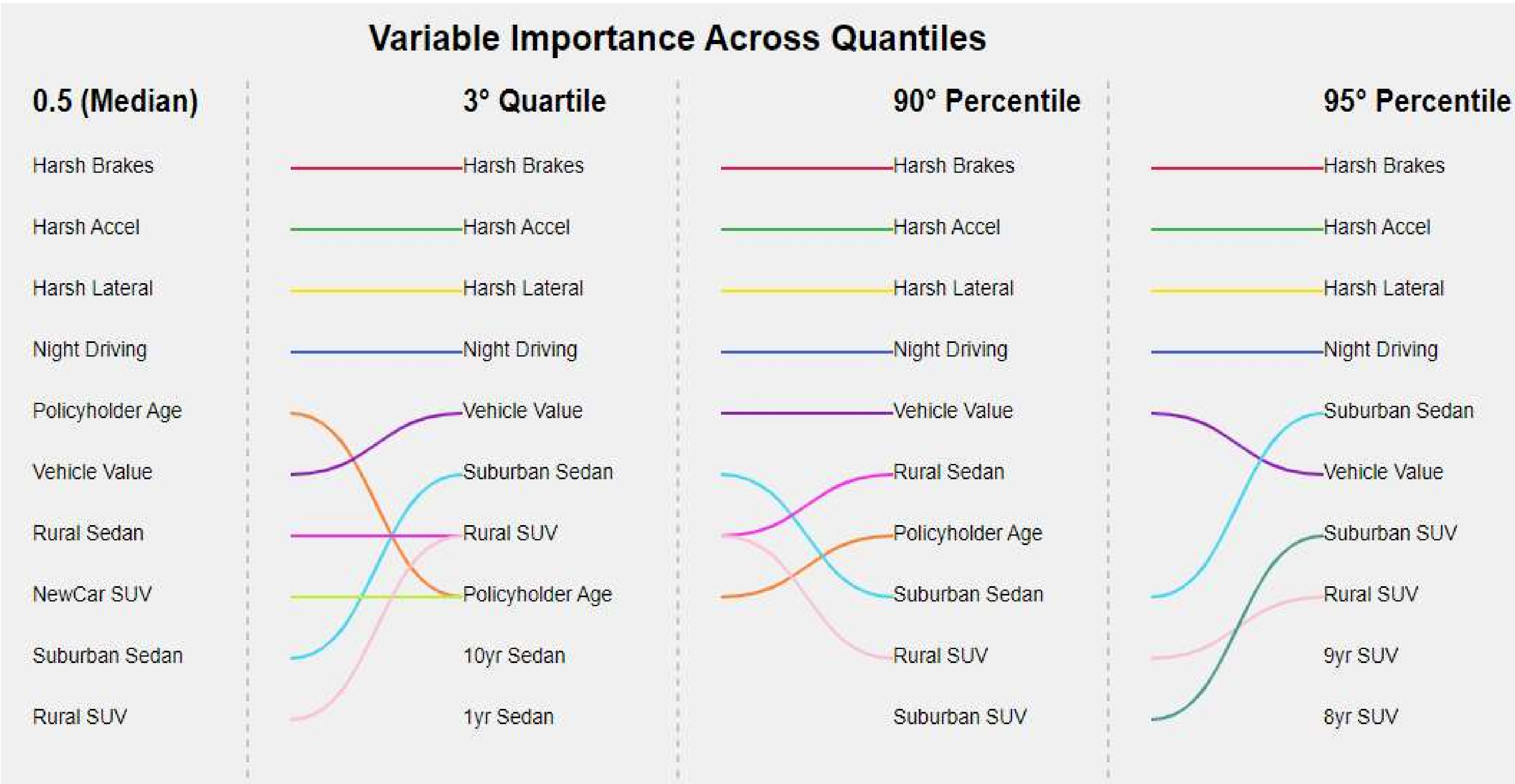
- Harsh braking has the largest impact across all quantiles.
- All driving behaviors show increasing impact in higher quantiles.
- Night driving has the smallest, but still significant, effect.



Relative Impact Increase (vs. 50th Quantile)

- Harsh acceleration shows the highest relative increase in the 95th quantile.
- Harsh lateral movements have the least relative change across quantiles.
- Night driving impact grows steadily across higher quantiles.

Variable Importance across Quantile

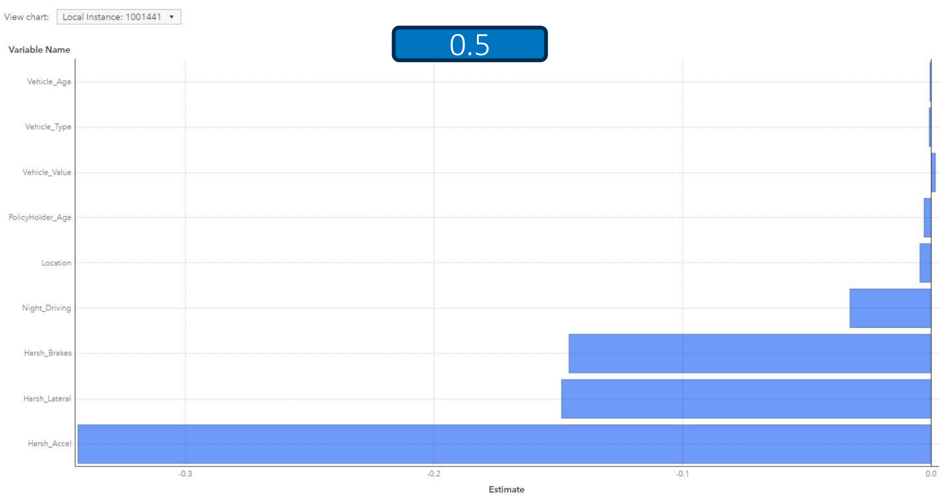


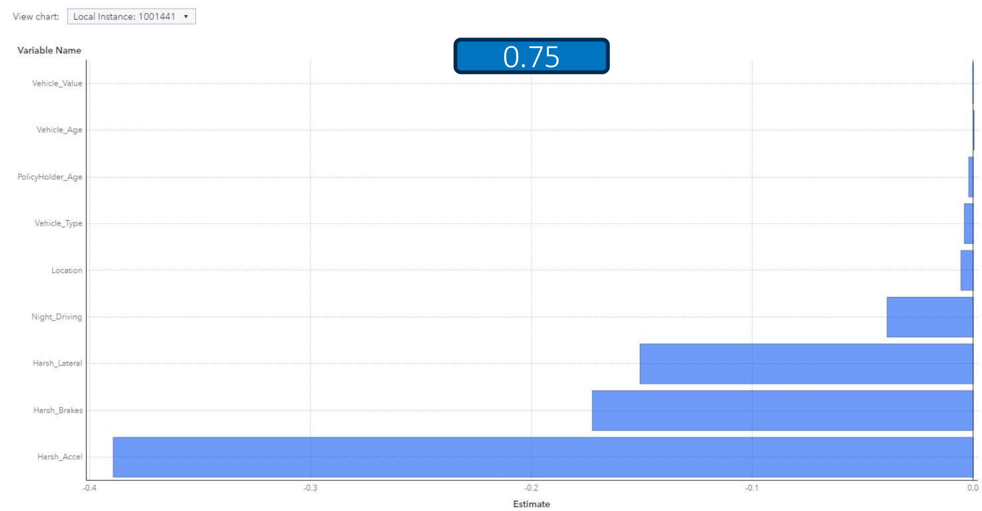
Quantile Assessment

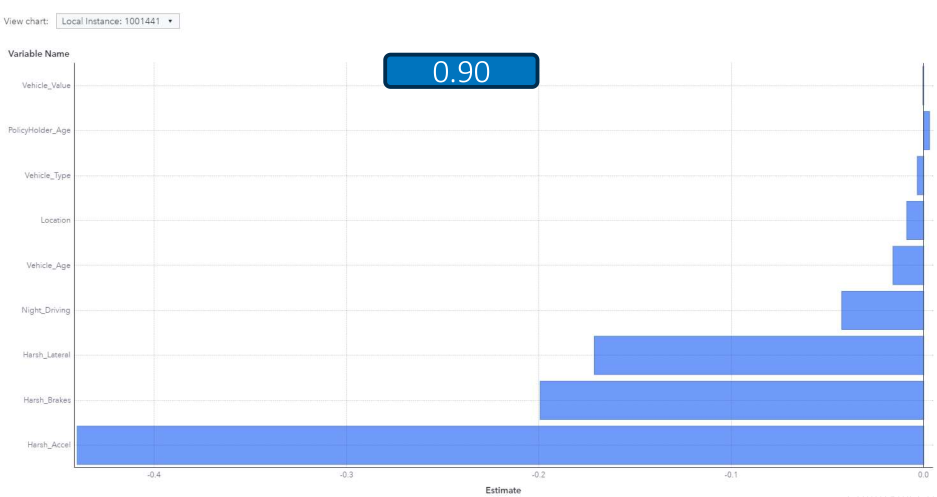
Policy N.	Median Predicted: Claim_Count	3° Quartile Predicted: Claim_Count	90° Percentile Predicted: Claim_Count	95° Percentile Predicted: Claim_Count
1000001	0.004747044	0.1425601658	0.2891028698	0.3820723342
1000002	0.0746615004	0.2022707809	0.375220882	0.4851351452
1000003	0.0036593336	0.1173456127	0.3047124235	0.4346234754
1000004	0.036009962	0.1174893193	0.2611172364	0.3841585473
1000005	0.0937414437	0.215132897	0.3765309345	0.4516364939
1000006	0.065947667	0.0497370919	0.1802781134	0.2968332382
1000007	0.0720394021	0.2004008823	0.3498224578	0.6462116376
1000008	0.0863966894	0.2529513444	0.4027705525	0.6438219887
1000009	0.2446003253	0.4177903825	0.6090034987	0.7484135563
1000010	0.0865132254	0.2516162677	0.4377096144	0.5819673069

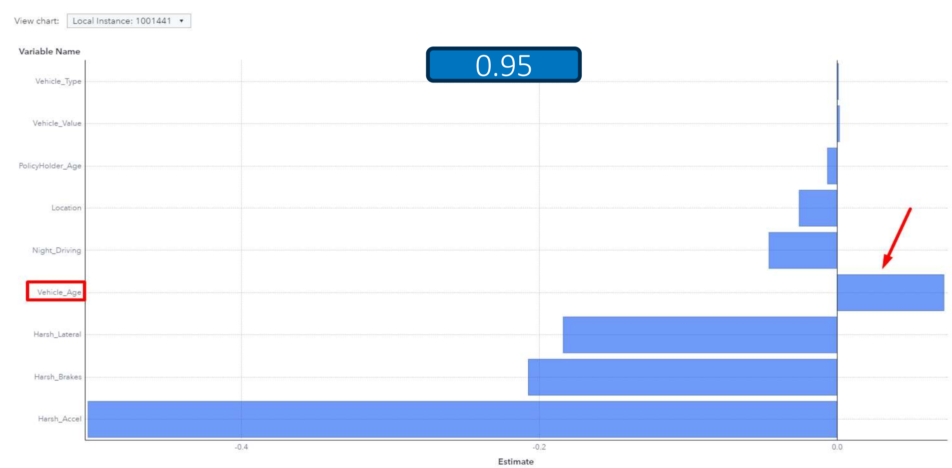
QR offers a detailed view of potential outcomes across different risk levels enabling more tailored and effective risk management and pricing strategies. This can be done at single risk or portfolio level.

Micro Analyisis





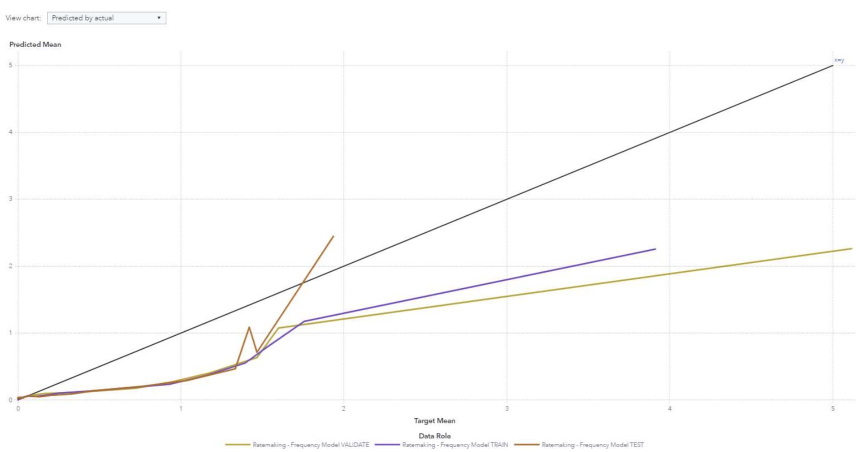


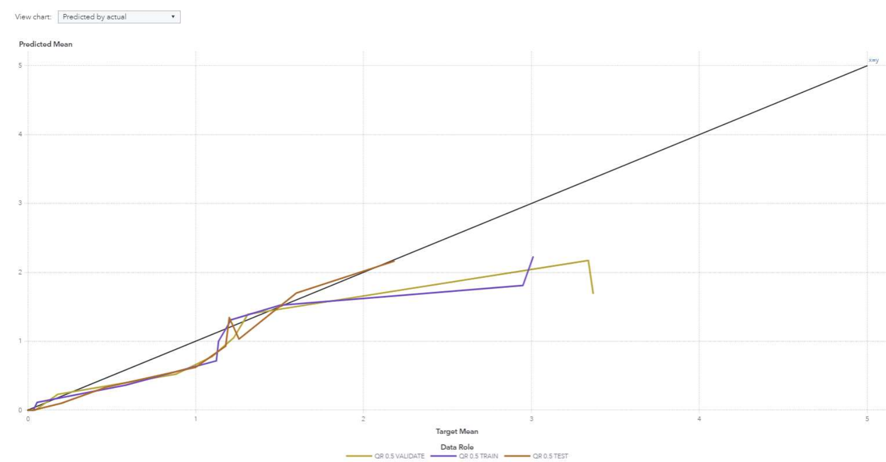


GLM VS QR with Outlier

Model Comparison

Champion	Name	Algorithm Name	Average Squared Error	Root Average Squared Error	Root Mean Absolute Error	Root Mean Squared Logarithmi...
★	QR 0.5	Quantile Regression	0.1283	0.3582	0.4106	0.1722
	Ratemaking - Frequency Model	Count GLM	0.3558	0.5965	0.5622	0.2888





Challenges and Limitations

Lack of Data

When focusing on extreme events there's often a lack of data in the upper quantiles.

Crossing

Fitting multiple quantile regression models at different percentiles, the resulting quantile curves can cross each other.

Inconsistent Inference

The significance of predictors can change dramatically from one quantile to another.

Separated Estimation

Many of these issues arise because quantile functions are typically estimated separately for each quantile, without considering the relationships between them.

SAS Viya

Advantages of Quantile Regression

Handling Heterogeneity

Quantile regression (QR) directly models the relationship between predictors and different quantiles of the response variable, not just the conditional mean. This allows QR to capture heteroscedasticity and provide information about the entire distribution of the response, unlike generalized linear models which focus on modeling the mean response.

Robustness to Outliers

Quantile regression is more robust to the presence of outliers compared to generalized linear modeling. QR minimizes the sum of asymmetrically weighted absolute residuals, reducing the impact of extreme values and providing a more accurate representation of the underlying risk factors.

Flexibility of Assumptions

Quantile regression does not rely heavily on the precise nature of the dependence structure among observations, unlike traditional regression methods that assume strict independence. This flexibility allows QR to capture the relationship between predictors and different parts of the response distribution, even when there are violations of the independence assumption.

sas

Conclusion

Our study focused on the advantages of quantile regression over generalized linear modeling in the ratemaking process. We used the SAS actuarial tool SAS Dynamic Actuarial Modeling software to perform all the necessary analysis in the study. We began by introducing our telematics data on automobile drivers, followed by a brief overview of quantile regression and generalized linear modeling theory, and then we highlighted the key benefits of quantile regression.

In a practical example, we compared quantile regression to generalized linear modeling and demonstrated that the model created by quantile regression was more accurate than the one created by generalized linear modeling when applied to our data. Quantile regression is distribution-free, meaning that there is no need to transform any of the variables or to determine the correct distribution for the target variable. We also showed the robustness of quantile regression to outliers, which can lead to more accurate predictions and premium calculations, improving underwriting performance and profitability.

References

- Buchinsky, M. (1998). **The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach**. *Journal of Applied Econometrics, 13*, 1-30. https://doi.org/10.1002/(SICI)1099-1255(199801/02)13:1%3c1::AID-JAE474%3e3.0.CO;2-A
- Fahrmeir, L., and Tutz, G. (2001). **Models for Multicategorical Responses: Multivariate Extensions of Generalized Linear Models**. In *Multivariate Statistical Modelling Based on Generalized Linear Models* (pp. 69-137). Springer, New York. https://link.springer.com/chapter/10.1007/978-1-4757-3454-6_3
- Huber, P. J. (1981). **Robust Statistics**. Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/0471725250
- Karst, O. J. (1958). **Linear Curve Fitting Using Least Deviations**. *Journal of the American Statistical Association, 53*, 118-132. https://www.tandfonline.com/doi/abs/10.1080/01621459.1958.10501430
- Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). **Loss Models: From Data to Decisions** (4th ed.). Hoboken, NJ: Wiley.