Weather-Informed Insurance Ratemaking and Product Management

Peng Shi[a]

Joint work with Zhengxiao Li[b] and Yifan Huang[b]

[a] University of Wisconsin-Madison
[b] University of International Business and Economics

Agenda

1. Classic experience rating
2. Incorporating weather information via a varying-coefficient model
    - Deep learning to encode weather information
    - Multi-task learning approach to allow forward-looking prediction
3. A case study in automobile insurance
4. Summary and concluding remarks

Classic Pricing and Experience Rating

Initial pricing:

- create homogenous risk classes based on rating variables
- premium is based on expected losses

Experience rating:

- adjust renewal premium based on claim history
- examples are credibility and bonus-malus

Classic Pricing and Experience Rating

- Weather-related hazards pose significant financial risks to individuals and businesses
- In automobile insurance, weather may affect claim likelihood through different channels
  - Adverse weather conditions can directly increase accident risk
  - weather may indirectly alter risk by influencing policyholder behavior
- In practice, however, weather information is rarely incorporated explicitly into insurance ratemaking

Project Overview

- We aim to develop a weather-informed framework for insurance ratemaking and product management
- We propose a deep varying-coefficient framework for longitudinal insurance claim counts that supports both initial underwriting and renewal pricing.
  - We use encoder–decoder architecture that maps multi-year historical weather sequences into weather-responsive pricing coefficients.
  - we introduce an auxiliary learning task that encourages the model to extract forward-looking weather patterns.

Basic Setting

- Consider a portfolio of n policyholders observed over T periods.
  - T_i denote the number of observed policy years for policyholder i, so that $T = \max\{T_1, \ldots, T_n\}$
  - y_{it} denote the number of claims by policyholder i in coverage period j,
  - $\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,q_x})'$ be a q_x–dimensional vector of rating variables

- We consider the classical Poisson–Gamma model for longitudinal count data

$$Y_{ij} \mid \Theta_i \sim \text{Poisson}(\lambda_{ij}\Theta_i), \text{ with } \Theta_i \sim \text{Gamma}(\alpha, \alpha)$$
$$\log \lambda_{ij} = f(\boldsymbol{x}_{ij}; \boldsymbol{\beta}) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}$$

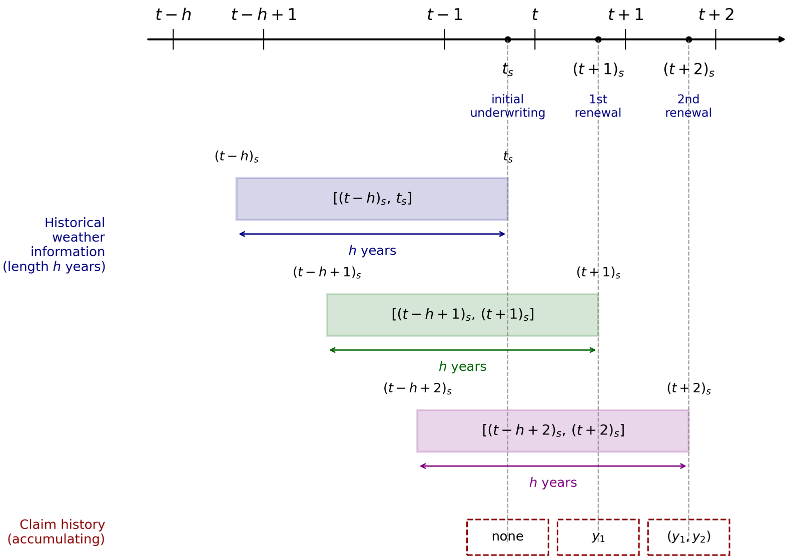
Basic Setting

- Consider a portfolio of n policyholders observed over T periods.
  - T_i denote the number of observed policy years for policyholder i, so that $T = \max\{T_1, \ldots, T_n\}$
  - y_{it} denote the number of claims by policyholder i in coverage period j,
  - $\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,q_x})'$ be a q_x–dimensional vector of rating variables

- We consider the classical Poisson–Gamma model for longitudinal count data

$$
\begin{aligned}
Y_{ij} \mid \Theta_i &\sim \mathrm{Poisson}(\lambda_{ij}\Theta_i), \text{ with } \Theta_i \sim \mathrm{Gamma}(\alpha, \alpha) \\
\log \lambda_{ij} &= f(\boldsymbol{x}_{ij}; \boldsymbol{\beta}) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}
\end{aligned}
$$

Weather Information



Varying Coefficient Model

We consider a varying-coefficient extension:

$$\log \lambda_{ij}(\boldsymbol{z}_{i[(t+j-1)_s,h]}, \boldsymbol{x}_{it}) = \log d_{ij} + \beta_0(\boldsymbol{z}_{i[(t+j-1)_s+,h]}) + \sum_{l=1}^{q_x} \beta_l(\boldsymbol{z}_{i[(t+j-1)_s,h]}) x_{it,l},$$

- Weather information is policyholder specific
- Regression coefficients remain interpretable

Multi-task Learning

A temporal mismatch: the risk-relevant weather and the historical information at pricing

We introduce a multi-task learning framework that jointly models

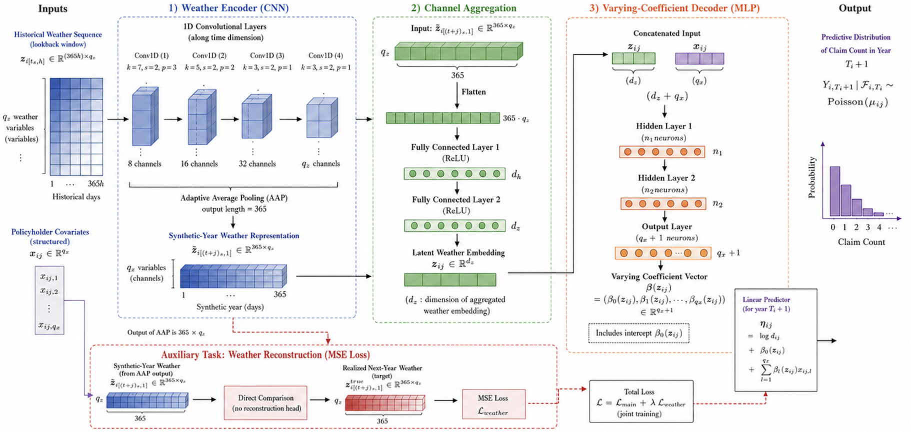
- claim frequency
- a forward-looking representation of policy-year weather

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{claim}} + \gamma \cdot \mathcal{L}_{\text{weather}}$$

where

$$\mathcal{L}_{\text{weather}} = \frac{1}{T_i} \sum_{j=1}^{T_i} \frac{1}{365\, q_z} \left\| \tilde{\boldsymbol{z}}_{i[(t+j)_s,1]} - \boldsymbol{z}^{\text{true}}_{i[(t+j)_s,1]} \right\|_F^2 .$$

Multi-task Learning Architecture



Data

Data are collected via two sources:

- A longitudinal automobile insurance dataset covering a portfolio of policyholders, including their claim outcomes and associated rating variables
- A multidimensional time series of daily weather measurements recorded at meteorological stations

Data: Insurance

Table: Empirical distribution of claim counts by year and overall.

Count	Percentage by Year					
	2010	2011	2012	2013	2014	Overall
0	73.76%	75.71%	76.30%	77.53%	82.28%	77.62%
1	16.86%	16.04%	15.75%	15.38%	13.82%	15.38%
2	6.24%	5.74%	5.79%	5.53%	3.19%	5.18%
3	2.13%	1.84%	1.63%	1.29%	0.58%	1.38%
4+	1.01%	0.66%	0.53%	0.27%	0.13%	0.44%
obs.	14,065	23,598	31,315	41,581	30,942	141,501

Data: Insurance

Table: Description and summary statistics of rating variables.

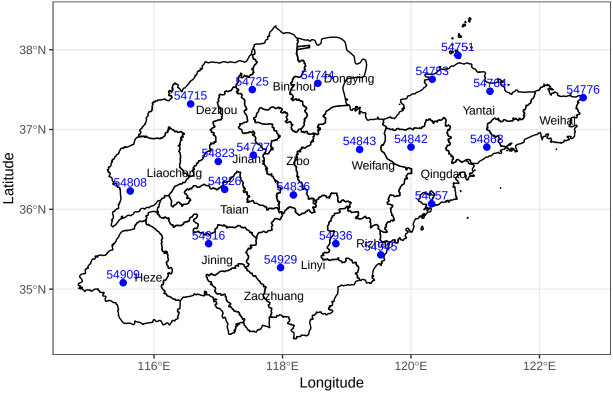
Variables	Type	Description	Mean	SD
exposure	numeric	policy exposure (in years)	0.98	0.07
age	numeric	driver's age (in years)	39.40	9.08
carage	numeric	age of the vehicle (in years)	4.01	2.94
carvalue	numeric	vehicle purchase price (in RMB)	97,228	88,558
gender	categorical	driver's gender		
		male	68.32%	
		female	24.95%	
		unknown	6.73%	
cartype	categorical	vehicle type		
		sedans	70.82%%	
		wagons	19.99%	
		SUVs	7.02%	
		microcars	2.17%	
carorigin	categorical	origin of the vehicle		
		joint-venture	53.98%	
		domestic	43.73%	
		imported	2.29%	

Data: Insurance

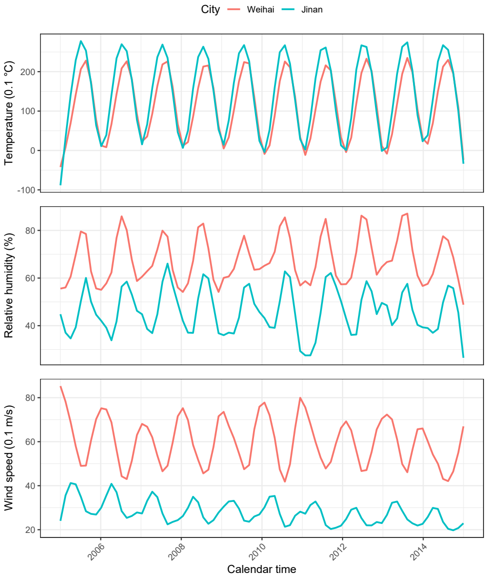
Table: Description and summary statistics of rating variables.

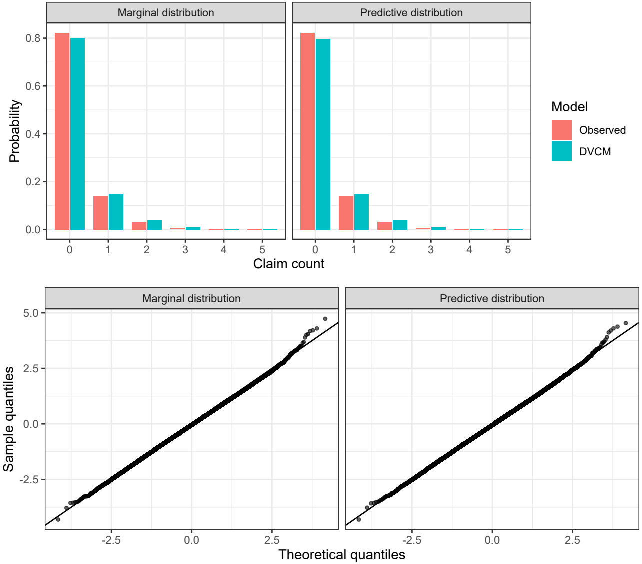
Variables	Type	Description	Mean	SD
`svolume`	categorical	engine displacement (swept volume)		
		$\leq$1.0L	16.48%	
		1.1–1.6L	55.54%	
		1.7–2.4L	23.93%	
		$\geq$2.5L	4.05%	
`carseats`	categorical	number of seats in vehicle		
		< 6 `seats`	82.41%	
		≥ 6 `seats`	17.59%	
`coverage`	categorical	type of coverage		
		`mandatory only`	34.05%	
		`mandatory and voluntory`	65.95%	

We use daily measurements of eight meteorological variables from stations in Shandong Province, corresponding to the region covered by the insurance data.



Data: Weather




Marginal distribution
Predictive distribution
Probability
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
Claim count
Model
Observed
DVCM
Sample quantiles
5.0
2.5
0.0
-2.5
Theoretical quantiles

Prediction Comparison

We consider the following two modeling frameworks:

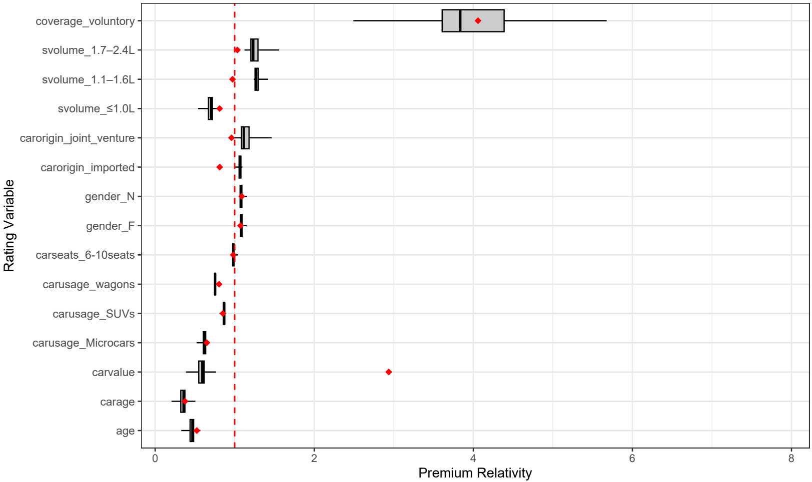
- Baseline model: A classical Poisson regression model for longitudinal count data with gamma-distributed random effects.
- Deep varying-coefficient model (DVCM): The proposed varying-coefficient Poisson regression model.

Table: Ordered Gini indices for champion–challenger comparisons†.

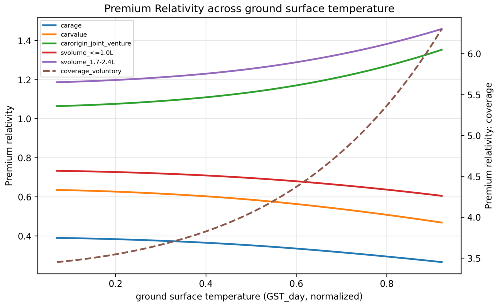
Base Premium	Marginal Distribution	Predictive Distribution
Baseline model	22.88 (0.80)	20.15 (0.96)
DVCM	-30.88 (0.77)	-21.76 (0.96)

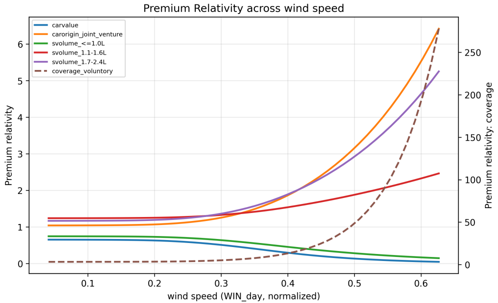
† Standard errors are reported in parentheses.

Implications: Relativity

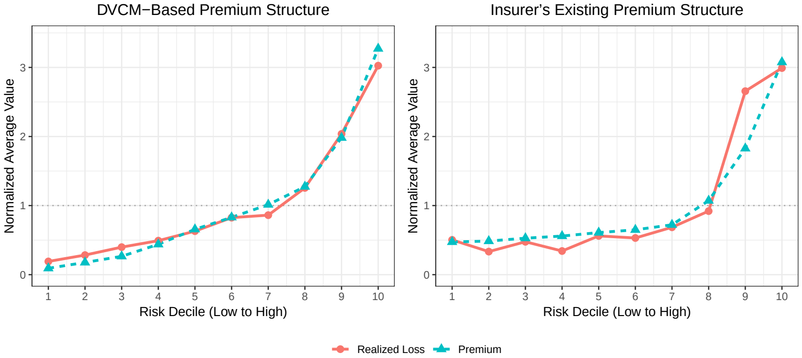


Implications: Relativity





Comparison with Current Premium



Conclusion

- Developed a deep varying-coefficient framework that incorporates weather information into automobile insurance pricing while preserving actuarial interpretability.
- Weather information improved risk differentiation and generated substantial heterogeneity in premium relativities and Bonus–Malus adjustments.
- Results suggest weather affects insurance risk through both baseline claim frequency and interactions with traditional rating factors.
- Future insurance pricing systems may become increasingly adaptive to changing environmental conditions.