Enhancing Business Insurance Loss Models through InsurTech Innovations

Zhiyu Quan* Changyue Hu* Panyi Dong* Emiliano A. Valdez†

1 Abstract

Recent transformative and disruptive advancements in the insurance industry have embraced various InsurTech innovations. In particular, with the rapid progress in data science and computational capabilities, InsurTech is able to integrate a multitude of emerging data sources, shedding light on opportunities to enhance risk classification and claims management. This paper presents a groundbreaking effort as we integrate real-life proprietary business insurance claims information with InsurTech data to enhance the loss model, a fundamental component of insurance companies' risk management. This integrated information further allows us to utilize various machine learning techniques to quantify the predictive improvement of the InsurTech-enhanced loss model over that of the insurance in-house. The quantification process provides a deeper understanding of the value of InsurTech innovation, as it uncovers potential risks previously overlooked in conventional insurance loss modeling. This work represents a successful undertaking of an academic-industry collaboration, suggesting an inspiring path for future partnerships between industry and academic institutions.

Keywords: InsurTech, business insurance, loss models, academic-industry collaboration.

2 Background on InsurTech

InsurTech combines the words "insurance" and "technology" to describe the applications of emerging technology to modernize the entire insurance value chain by improving efficiency, enriching customer service, enhancing underwriting and actuarial processes, and further uncovering new opportunities, to name a few. InsurTech emerged around 2010 and was initially categorized as a subsector of FinTech (Kelley and Wang, 2021). InsurTech companies are conventionally funded as start-up companies by venture capitalists; InsurTech reached more than 2,000 start-up deals at the end of 2022. The amount of financing has been on the rise globally over the past decade; See Figure 1.

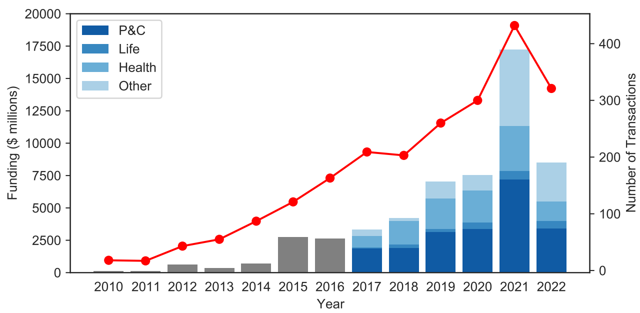


Figure 1: InsurTech financing for 2010-2022.

Based on datasets obtained from our industry partners, we focus on improving BOP loss models by leveraging InsurTech innovations. P&C insurers adopted predictive modeling in the early 2000s by integrating analytics for various purposes like actuarial ratemaking, which assesses insurance product risks using historical loss data. As ratemaking hinges on data quality and quantity, we sought to use external data to enhance our loss models. Therefore, in this paper, our goal is two-fold: mining predictive risk characteristics from InsurTech data that improve insurance in-house loss model, and explaining these

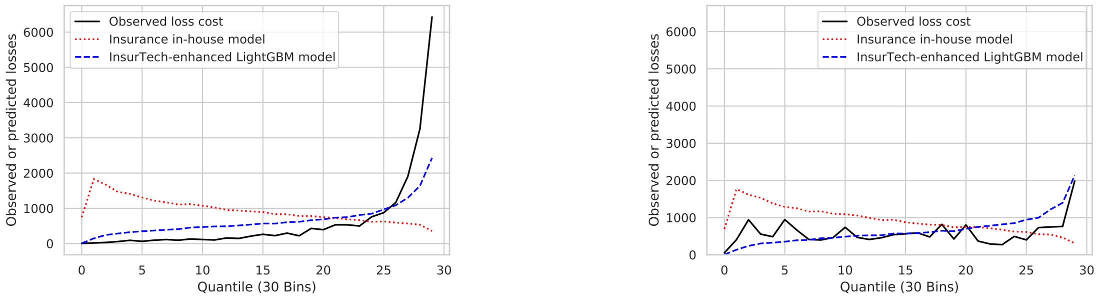
*Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL, 61801, USA

†Presenter: emiliano.valdez@uconn.edu; Department of Mathematics, University of Connecticut, 341 Mansfield Road, Storrs, CT, 06269-1009, USA.

risk characteristics using interpretable machine learning techniques. We also intend to further propose potential rating factors useful for business insurance and also in compliance with insurance regulation.

3 Comparison of model performance

Here we compare the performance of two models: insurance in-house models versus InsurTech-enhanced models. We model each coverage group separately using LightGBM and Tweedie GLM after elastic net feature selection, and thus we have six InsurTech-enhanced models in total. Here we give results for BG coverage only. For LightGBM model calibration, we use Bayesian optimization with the Optuna framework, and the optimization objective loss function is the mean absolute error (MAE). For the Tweedie GLM after elastic net feature selection model calibration, we use grid search to find the optimal hyperparameters. We apply a 10-fold cross-validation on the training set to find the best models for both methods. We use double lift charts, commonly used to compare predictive power of two distinct models, to visually compare the predictive performance of InsurTech-enhanced models and insurance in-house models.



(a) Double lift chart for BG train

(b) Double lift chart for BG test

Figure 2: Double lift charts for model comparison

Coverage	Dataset	Model	Gini	PE	RMSE	MAE
BG	train	Insurance in-house model	0.29	-0.40	5761.94	1526.47
		Tweedie GLM + elastic net	0.44	-0.04	5660.01	1286.31
		LightGBM	0.84	0.00	5364.05	1198.07
	test	Insurance in-house model	0.32	-0.54	5328.02	1461.92
		Tweedie GLM + elastic net	0.32	-0.16	5284.90	1238.94
		LightGBM	**0.37**	**-0.08**	**5198.57**	**1181.47**

Table 1: Model performance based on validation measures

Table 1 compares the predictive performance of InsurTech-enhanced models, Tweedie GLM with elastic net feature selection and LightGBM, and the insurance in-house model based on the training and test datasets of each coverage group. We highlight the best-performing model with the value of each validation measure in bold. In general, both InsurTech-enhanced models consistently outperform the insurance in-house model, which suggests the improvement stems from the additional information provided by InsurTech, irrespective of the chosen loss model. We note that InsurTech-enhanced models significantly reduce the absolute value of percentage error (PE) of observed loss cost predictions compared to insurance in-house models. This indicates that InsurTech-enhanced models have superior predictive performance at the portfolio level, usually a key concern for insurers particularly from a financial statement perspective.

References

Kelley, C. and Wang, K. (2021). Insurtech: A guide for the actuarial community. Technical report, Willis Tower Watson. Published by the Society of Actuaries.