Title: Model misspecification and data-driven model ranking approach for insurance loss and claims data

Abstract:

Statistical models play a crucial role in analyzing insurance loss and claims data, offering insights into various risk elements. The prevailing statistical notion that "all models are wrong" can wield significant influence, particularly when multiple competing statistical models are considered. This becomes particularly pertinent when all models portray similar characteristics within certain subsets of the support of the random variable under scrutiny. Since the true model is unknown in practical scenarios, the challenge of model selection becomes daunting, complicating the study of associated characteristics of the true data generation process. To address these challenges, the concept of model averaging is embraced. Often, averaging over multiple models helps alleviate the risk of model misspecification, as different models may capture distinct aspects of the data or modeling assumptions. This enhances the robustness of the model selection process and yields a more acceptable and reasonable estimate compared to relying solely on a single model. This paper introduces two novel data-based model selection methods—one based on the likelihood function and the other on the density power divergence measure. The focus is estimating the Value-at-Risk (VaR) of non-life insurance claim size data, providing comprehensive insights into potential insurer losses. The performance of the proposed procedures is compared through Monte Carlo simulations, both in uncontaminated data scenarios and in the presence of contamination. Additionally, the applicability of the methods is demonstrated using two real non-life insurance datasets, estimating VaRs at different confidence levels.