On estimation of insurance claim numbers by combining local regression and distribution fitting ideas

Meelis Käärik, Raul Kangro and Liina Muru

University of Tartu, Estonia, meelis.kaarik@ut.ee

In this paper we address the well-known problem of premium estimation in non-life insurance. More precisely, we are modelling the claim numbers, which can be seen as one component of the premium. We are looking for certain dynamic regression type model to avoid the "price shock" issue of static classification. We also take into account that it is hard to specify the form of suitable regression functions, and simple choices of such functions usually have undesirable effects by implicitly implying that risk behaviour of clients corresponding to one region of values of regression variables contain information about the risk behaviour of clients corresponding to a very different region of the same variables. Thus we are proposing certain local regression model, where for each new client we first fix a neighborhood of similar clients (depending on the values of certain argument variables) and then apply the local regression model on this neighborhood. Local maximum likelihood estimation is used to determine the parameters of the model and cross-validation techniques are used to determine the optimal size of the neighborhood. As a result, we propose certain semiparametric model for estimating the risk parameters for each new client. A case study with real vehicle casco insurance dataset is included, the results obtained by proposed method are compared by the ones obtained by global regression and the classification and regression trees (C&RT) approach.