OPPONENT ASSESSMENT OF DIPLOMA THESIS

Title: Hurdle models in non-life insurance
Name: Cheng Tian

Thesis summary

The thesis deals with the hurdle regression models which are suitable for modelling of semi-continuous responses. In particular, it elaborates in details the log-normal, Gamma, Inverse-Gaussian and Weibull hurdle models. The logistic regression is used to model the probability that the response crosses a hurdle. Chapter 2 summarizes the procedures and statistical tests suitable for model selection including residual diagnostics, information criteria and tests based on likelihood functions. Chapter 3 provides an extensive numerical study using nonlife insurance data. Several hurdle models are built and compared.

Overall thesis evaluation

Thesis topic. I believe that the diploma assignment has been fulfilled. In my opinion, the topic was rather difficult.

Own contribution. The main contribution of the thesis relies in deriving the likelihood functions, their derivatives and information matrices for several semi-continuous hurdle models. In some cases, it was even possible to obtain a closed form expressions for the estimates. These new models were then implemented in R software and applied successfully to nonlife (car) insurance data.

Mathematical level. The thesis contains rigorous mathematical formulations, especially in the first chapter. The mathematical quality is adequate to the study branch *Financial and insurance mathematics*. However, I have several comments and questions to this point, see below.

References. The references are properly cited in the text and included in the list which appears at the end of the thesis. I do not see any copied parts.

Formal arrangement. I found many type errors, but their frequency is appropriate to the length of the thesis. Sometimes the articles are used in a confusing way.

Remarks and questions

1. Page 4: I am afraid that the Pareto distribution is not suitable for claim counts modelling.
2. Pages 5 and 9: You claim that the hurdle models can handle overdispersion. However, I cannot see it. Could you please convince me?
3. Maybe I missed that but I would expect at least a short discussion about the convexity of the likelihood functions with respect to the estimated parameters.

Moreover, the positive definiteness of the Hessian matrix is also desirable at least for the Newton-Raphson algorithm, see page 8.

4. Chapter 2: At several places, you are referring to some symbols which are usually used in the definition of Generalized Linear Models (GLM). However, GLM are nowhere defined in the thesis.
5. Chapter 3, sequential model selection: If I understood well your approach, you applied “backward” model selection in all cases. Why you did so? I would consider the “forward” selection as a safer one, especially with the continuous regression component of the hurdle models.
6. I think that the tables 3.4, 3.6 containing the observed information matrices are not necessary. It can be sufficient including them into an appendix.
7. Page 39: The rewritten model (3.4) can be skipped.
8. Page 39, Model diagnostics: I do not fully understand to the conclusions based on the deviance test. Would you mind formulating the hypothesis and explaining the test in more details?
9. Page 43: I am afraid that nowadays it is forbidden to increase the premium to male drivers.

Conclusion

I suggest accepting the work as a diploma thesis.

RNDr. Martin Branda, Ph.D.
KPMS MFF UK
16. 1. 2018