Setting the optimal limit value of motor insurance coverage by stochastic optimization

Lucie HANELOVÁ a, Jiří VALECKÝ a*

a Department of Finance, Faculty of Economics, VŠB-Technical University of Ostrava, Sokolská třída 33, 702 00 Ostrava, Czech Republic.

Abstract

In this paper, we provide an alternative to a passive approach to the selection of insurance products or policy conditions. Specifically, we propose a method to make a decision about the optimal limit value for motor insurance coverage. Respecting the stochastic nature of individual loss, we formulate a problem of stochastic programming in which the total potential financial loss of the policyholder is minimized. Actually, we present a general optimization problem in which various relevant probability distributions of individual loss may be considered. In addition, we extend the work of Valecký (2017) and derive an insurance rate that describes better the dependence between the pure premium and the given limit value under the assumption that the individual potential loss follows a gamma distribution. Because of the absence of a closed-form solution, sample average approximation is applied to the objective function and the optimal solution to this approximated (SAA) problem is determined. Finally, the quality of the obtained solution is assessed by approximation to the optimality gap representing the difference between our candidate and the true solution.

Keywords

Decision making, limit value, motor insurance, optimal coverage, stochastic optimization.

JEL Classification: C44, C61, G22.

* jiri.valecky@vsb.cz (corresponding author)
This paper was supported within the Operational Programme Education for Competitiveness – Project No. CZ.1.07/2.3.00/20.0296 – and under the SGS project SP2017/32.
Setting the optimal limit value of motor insurance coverage by stochastic optimization

Lucie HANELOVÁ, Jiří VALECKÝ

1. Introduction

Decisions about the policy conditions of an insurance product, such as the limit value, coinsurance and so on, are generally made in a passive manner; that is, the policyholder chooses an insurance product from one of the available insurers or selects one of the available settings of a given insurance product. Clearly, the final choice represents a solution that is satisfactory rather than optimal. By contrast, this paper deals with an active approach to setting policy conditions, specifically the limit value, for motor insurance to maximize the coverage as well as to minimize the costs of insurance.

Thus, the decision making about the optimal insurance product is limited mainly to the coverage of potential insurance claims. However, higher coverage also yields a higher premium. In addition, when several criteria are considered, the decision-making process involves one of various multi-criteria methods that consist mainly of calculation weights representing the importance of each of the criteria for the policyholder. In the process, the weights need not be determined via the AHP process of Saaty (1980) (see for instance Borovcová, 2017) but may also be derived in a more understandable manner (see Tzeng and Huang, 2011; Zmeškal et al., 2013; or Rao, 2014 for an overview of the methods). However, this process is sensitive to the correct evaluation of the policyholder’s preferences, and the final choice may be influenced significantly by the importance weights.

If only financial criteria existed, it would be possible to evaluate the total costs related to each variant, that is, the claim severity, required premium, amount of coinsurance and deductibles and so on. In our case of two criteria, comparing the premium with the limit value appeared to be sufficient. However, insurers usually apply a multiplicative tariff, yielding a constant ratio of the premium and limit value. It follows that a given percentage increase in the limit value yields the same percentage increase in the premium.

Unfortunately, these approaches do not respect the stochastic nature of insurance claims and even the fact that the premium is supposed to be paid even if no insured accident occurs. As a result, the potential loss is totally covered but the policyholder, on the other hand, pays for expensive insurance that will never or rarely be used (with a very small probability). By contrast, an active approach to setting policy conditions provides the optimal condition when the insurance is not so expensive and the potential loss is not fully covered with very low probability. However, this alternative approach requires one of the stochastic programming techniques. For an introduction to stochastic optimization, we recommend some general books addressing this issue, for example those by Kall and Mayer (2011) and King and Wallace (2012).

Studies representing the active approach to setting insurance conditions, in particular the problem of the optimal limit value and deductibles, are available in the journal literature and are based mainly on the expected utility theory (e.g. Zhou et al., 2010; Lu and Meng, 2011; Liu et al., 2015). However, Wang and Huang (2016), for instance, extend this approach by incurring VaR and CVaR in optimization, and Pflug et al. (2017) even incorporate the model uncertainty as a decision variable. Further, an extension involving the prospect theory is available (e.g. Sung et al., 2011; Chen et al., 2015; Cheung et al., 2015).

However, some studies do not use the utility theory and establish the expected costs of the insured directly; for example, Gaffney and Ben-Israel (2016) derive the optimal deductibles as well as the limit value under the insurance budget, and Valecký (2017) determines the optimal limit value under the assumption that the potential loss follows an exponential probability distribution. Although the premium is set as an increasing function of the limit value, the insurance rate is determined ad hoc and does not respect the true relationship between the pure premium and the limit value.

The goal of the paper is to determine an optimal limit value of motor insurance coverage and to respect the trade-off between the pure premium and the given limit value as well as the stochastic nature of the potential individual loss that is assumed to follow a gamma distribution. However, we consider the pure premium only because the safety, as well as the expense loadings, are not known.

The remainder of the paper is organized as follows. The general approach to setting the pure premium is described in Section 2. In Section 3, we formulate the optimization problem of stochastic programming and describe the approximation to the closed-form solution as well as showing how to evaluate the quality of the solution obtained from this approximation. Finally, we
present an illustrative example in Section 4, in which
the optimal limit value is determined for a specific pol-
cyholder using sample average approximation and the
quality of the solution is evaluated. Section 5 provides
a discussion and concluding remarks.

2. Pricing motor hull insurance with a limit value
In this section, the fundamental principle for setting the
pure premium is described. For more details on how the
final premium is determined, we refer the reader to

2.1 Insurance claim and benefit
A policyholder may suffer an individual potential loss
X that is split as follows:

\[ X = X^{[\ell]} + X^{[p]}, \]

where \( X^{[p]} \) is the part retained by the policyholder and
\( X^{[\ell]} \) represents the loss reimbursement claimed
by the policyholder and paid by the insurer, that is, the
insurance benefit.

The method for determining the benefit \( X^{[\ell]} \) de-
PENDs on the policy conditions; let us say that it is a
function of the individual potential loss, thus

\[ X^{[\ell]} = f(X). \]

Assuming an insurance policy with a limit value, say \( L \), both parts of the potential loss are defined as

\[ X^{[\ell]} = \begin{cases} X & \text{if } X < L, \\ 0 & \text{if } X \geq L, \end{cases} \]

\[ X^{[p]} = \begin{cases} L & \text{if } X < L, \\ X - L & \text{if } X \geq L, \end{cases} \]

(2)

where \( X \) is the random potential loss.

2.2 Setting the pure premium
However, the policyholder has the possibility of suffer-
ing several individual losses with varying severity.
Thus, the total potential loss \( S \) is a result of the random
sum of individual random losses, that is

\[ S = X_1 + X_2 + \ldots + X_n, \]

(3)

where \( n \) is the random number of insured accidents.

Recalling the rule for splitting the potential loss (2)
applied to each individual loss, the total insurance ben-
EFIT is a random variable defined as follows:

\[ S^{[\ell]} = X_1^{[\ell]} + X_2^{[\ell]} + \ldots + X_n^{[\ell]}. \]

(4)

2.3 General formula for the pure premium
In addition to the insurer’s costs and required profit as
well as the market conditions in the insurance market,
the final premium is determined primarily by the
amount of the total potential benefit \( S^{[\ell]} \). The part
of the premium that is supposed to cover these potential
claims is referred to as the pure premium and is yielded
by the equivalence principle as a present value of the
expected total potential benefit, thus

\[ P = PV \left( E \left( S^{[\ell]} \right) \right) \]

(5)

However, in motor insurance or non-life insurance
generally, the discounting of future losses is omitted be-
cause of short maturity (less than 1 year). The formula
for a pure premium may be rewritten as

\[ P = E \left( S^{[\ell]} \right) = E \left[ N \right] E \left[ X^{[\ell]} \right], \]

(6)

where \( E \left[ N \right] \) and \( E \left[ X^{[\ell]} \right] \) are the expected claim fre-
QUENCY and the expected claim severity, respectively.

Recall that the loss covered by the insurer depends on
rule (2) and that the potential individual loss \( X \) is
random. Then, the expected severity depends on the
probability that \( X \) exceeds the limit value \( L \), thus

\[ E \left[ X^{[\ell]} \right] = Pr(X < L)E[X | X < L] + \\
+ Pr(X \geq L) \cdot L, \]

(7)

where \( Pr(X < L) \) is the cumulative distribution func-
tion of \( X \), that is, \( F_X(L) \), and

\[ Pr(X \geq L) = 1 - F_X(L). \]

Note that the expected frequency, as well as the
probability distribution of \( X \), does not depend on the
limit value \( L \). Then, the equivalence principle involv-
ing the limit value gives the following general formula
for the pure premium:

\[ P = E \left[ N \right] \left( F_X(L)E[X | X < L] + (1 - F_X(L)) \cdot L \right). \]

Pure premium with gamma-distributed severity
Let us assume that the distribution of claim frequency
follows a discrete probability distribution, for example
a negative binomial with the rate parameter \( \lambda \) and
the overdispersion parameter \( \kappa \), while the claim severity
is gamma distributed, \( Ga(\alpha, \theta) \), where \( \alpha \) is the shape
and \( \theta \) is the scale parameter.

The probability density function of the gamma dis-
tribution is defined as

\[ f_X(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}} \]

(8)

and the cumulative distribution function (cdf) is given by

\[ F_X(x) = \frac{\gamma(\alpha, x/\theta)}{\Gamma(\alpha)}, \]

(9)

where \( \gamma(\alpha, x/\theta) \) is the lower incomplete gamma func-
tion.
Let us assume a policyholder with an expected claim frequency with the value of 0.0841 and a gamma-distributed severity that has shape and scale parameter values of 1.1193 and 29,362.736, respectively. The next figure shows the pure premium of this policyholder for various limit values.

![Figure 1](image1)

**Figure 1** Relationship between the pure premium and the limit value for the given policyholder

Clearly, the pure premium is increasing, and it appears to be constant for the limit values higher than 250,000. In fact, it is still increasing as the limit value moves towards infinity. However, the probability that the individual loss will exceed the given limit value is so small that it yields a very small increase in the pure premium.

2.3 Rate making

The pure premium actually represents the statistical premium that covers the expected benefit. However, the premium is rarely set in this way; rather, insurers apply rates. Then, the pure premium, as well as the final premium, is determined as a multiple of the limit value.

To derive the rate for our purpose, we regress the pure premium on the limit values in the interval \([0.92000]\), setting the constant to zero. The final rate for a given policyholder corresponds to the slope parameter, which is estimated to have a value of 0.0334.

Then, the pure premium is given by

\[ P = r \cdot L, \]

(10)

where \( r \) is the insurance rate.

3. Stochastic optimization problem

The goal of the optimization problem is to set the optimal limit value to minimize the potential loss that is retained by the policyholder, that is, \( X^{[p]} \), who is also supposed to pay the premium \( P \).

Thus, the total financial costs of the policyholder can be defined as

\[ X^{[p]} + P = X - L + P, \]

(11)

which may be rewritten as

\[ X - L + r \cdot L = X - (1 - r) L. \]

(12)

However, even if the loss is smaller than the limit value, the policyholder only pays the premium. Both cases may be rewritten as the objective function in the form of

\[ G(L, X) = \begin{cases} X - (1 - r) L & \text{if } L < X, \\ r \cdot L & \text{if } L \geq X, \end{cases} \]

(13)

or

\[ G(L, X) = r \cdot L + \max(X - L, 0). \]

(14)

Having a loss of value of 450,000, the next figure shows the premium as well as the part of the loss retained by the policyholder for various limit values.

![Figure 2](image2)

**Figure 2** Loss retained by the policyholder (dotted line) and the pure premium (dashed line) for various limit values when the loss is assumed to be 450,000

By contrast, as the limit value moves towards infinity, the premium increases while the policyholder’s loss decreases and equals zero for limit values higher than the loss at 450,000. Clearly, the minimum of the objective function is 450,000; thus, the limit value is identical to the loss incurred, as shown in the next figure.
Clearly, the solution generally depends on the probability distribution of $X$ and solving the integral of the relevant cdf.

### 3.2 Sample average approximation

The closed-form solution to (16) is not always available, especially if the loss is gamma distributed. In these cases, it is possible to approximate the objective function using the Monte Carlo technique.

First, a sufficient number of scenarios $j = 1, \ldots, N$ is drawn from the assumed probability distribution. Second, the value of (16) is evaluated for each of the scenarios and then the expected value of the objective function is approximated by averaging, thus

\[ g(L) \approx \bar{g}_S(L) = \frac{1}{N} \sum_{i=1}^{N} G(L, X^{(i)}) , \quad (17) \]

where $N$ is the number of scenarios and $X^{(i)}$ is the loss of the $i$th scenario.

### 3.3 Evaluating candidate solutions

The solution to (17) is actually not the solution to the general objective function (16), and the optimal value obtained from the SAA problem may be far from the true optimal limit value.

To evaluate the quality of the SAA solution as an approximation of the true problem, the so-called optimality gap is given by

\[ d(L_{\text{SAA}}) = \bar{g}_S(L_{\text{SAA}}) - g(L_{\text{opt}}) , \quad (18) \]

where $\bar{g}_S(L_{\text{SAA}})$ and $g(L_{\text{opt}})$ are the objective function value for the optimal limit value from the SAA and the true problem, respectively.

We do not know the value of $g(L_{\text{opt}})$, but we can estimate it by solving the SAA problems $M$ times and averaging the values of (17), thus

\[ g(L_{\text{opt}}) \approx \bar{g}_{S,M}(L) = \frac{1}{M} \sum_{i=1}^{M} \bar{g}_S^{(i)}(L) , \]

where $\bar{g}_S^{(i)}(L)$ is the value of (17) for the $i$th SAA problem each of size $N$, thus

\[ \bar{g}_S^{(i)}(L) = \frac{1}{N} \sum_{j=1}^{N} G(L, X^{(i,j)}) . \]

It follows that the optimality gap can only be approximated. To evaluate the quality of $L_{\text{SAA}}$, it is necessary to obtain the statistically valid bound on the true optimality gap (18).

First, the sample variance of $\bar{g}_S(L_{\text{SAA}})$ is calculated as

\[ \bar{g}_S(L_{\text{SAA}}) = \frac{1}{N} \sum_{i=1}^{N} G(L, X^{(i)}) . \]
\[ \sigma^2_N = \frac{1}{N(N-1)} \sum_{j=1}^{N} \left[ G(L_{\text{SAA}}, X^{(j)}) - \bar{g}_N(L_{\text{SAA}}) \right]^2, \]

which gives the approximate 100(1-\(\alpha\))% confidence upper bound for \(\bar{g}_N(L_{\text{SAA}})\), that is,

\[ U_{\alpha}(L_{\text{SAA}}) = \bar{g}_N(L_{\text{SAA}}) + z_{\alpha,\sigma} \sigma_N(L_{\text{SAA}}), \]

where \(z_{\alpha,\sigma}\) is the critical value of the standard normal distribution.

Second, the variance of \(\bar{g}_{N,M}(L)\), defined as

\[ \sigma^2_{N,M} = \frac{1}{M(M-1)} \sum_{i=1}^{M} \left[ \bar{g}_N^{(i)}(L) - \bar{g}_{N,M}(L) \right]^2, \]

gives the 100(1-\(\alpha\))% confidence lower bound for \(\bar{g}_{N,M}(L)\), thus

\[ L_{\alpha,M} = \bar{g}_{N,M}(L) - t_{\alpha,\nu} \sigma_{N,M}. \]

where \(t_{\alpha,\nu}\) is the critical value of the t-distribution with \(\nu = M - 1\) degrees of freedom.

Then, the statistically valid bound on the true optimality gap with confidence of at least \(1 - 2\alpha\) is determined as follows:

\[ \hat{d}(L_{\text{SAA}}) = U_{\alpha}(L_{\text{SAA}}) - L_{\alpha,M}. \tag{19} \]

4. Setting the optimal limit value for a given policyholder

The goal is to set the optimal limit value for a given policyholder with the expected frequency at the value of 0.0841. Recall that the insurer sets the rate at 0.0334 per unit of limit value and that the individual loss follows a gamma distribution with shape and scale parameters of 1.1193 and 29,362.736, respectively. Further, we assume a car value of 700,000.

Because the closed-form solution is not available, we solve the problem of stochastic optimization in Matlab using the SAA technique. Thus, we minimize the following objective function:

\[ \min \frac{1}{N} \sum_{j=1}^{N} G(L, X^{(j)}) \tag{20} \]

subject to

\[ L \geq 0, \]

where \( G(L, X^{(j)}) \) is function (14) of the \(j\)th scenario.

First, we draw 100,000 scenarios from the gamma distribution that represent the realizations of random individual loss. The histogram is shown in the next figure.

![Figure 4 Histogram of the generated individual loss from a gamma distribution with shape and scale parameters with the values of 1.1193 and 29,362.736, respectively](image)

Clearly, all the generated realizations of the individual loss are significantly smaller than the car value. It follows that there is a low probability of maximum damage to the car.

The minimum of the objective function (20) is found in 16 seconds at the limit value 106,472, giving a premium of 3,552 and total expected financial costs with the value of 4,553. In addition, the probability that the individual loss will exceed the optimal limit value is 0.0338. The SAA approximation for the true problem is shown in the next figure.

![Figure 5 SAA approximation for \(E[G(L, X)]\)](image)

Remember that the optimal solution to (20) is only a candidate for the solution to the true problem. To evaluate how close it is to the closed-form solution, we solve the SAA problem 100 times, each with size
100,000. With the significance level of 5%, the approximated optimality gap has the value of 35.98, indicating that the obtained solution gives the minimum of an approximated objective function that is not farther from the true minimum than 35.98 with 90% confidence.

Note that the rate per limit value used does not correspond to the real rate of commercial insurers, because we mainly consider the pure premium rather than the final price of insurance, which also incorporates a safety loading as well as a loading for the insurer’s expenses and profit margin and which is affected by the competition in the insurance market. The real rates are actually unknown to the policyholder, but the premium corresponding to the various limit values can be obtained from a web calculator of the commercial insurer, and the rates can be determined when these premiums are regressed on the limit value.

Further, the parameters of the loss probability distribution are unknown to the policyholder. However, they may be estimated from publicly available statistics or just simply assumed. The parameters of our gamma distribution are actually adopted from a gamma regression model that respects the various characteristics of the policyholder, including the car value as one of the explanatory variables. However, one may point out the imperfections of this distribution. Thus, the distribution is unbounded above but the possible maximal loss is actually constrained to the car value rather than unlimited. In addition, although the car value entered the gamma regression model at 700,000, providing the parameters of our gamma distribution, there was little probability of such damage, yielding a very small increase in the pure premium for a limit value above 200,000.

Finally, in contrast to the pure premium calculation derived from the part of expected loss covered by the insurer, the other part, retained by the policyholder, does not respect the probability of insured accidents, that is, the claim frequency. The goal of our problem is to optimize the coverage in the case of an insured accident, which concerns the expected potential loss (the probability occurrence is neglected) rather than the expected loss itself, which is crucial for setting the pure premium. In addition, respecting the probability of an insured accident would induce the situation that any limit value is optimal. The pure premium is equivalent to the expected loss, and it is not important which part is retained by the policyholder and which part is transferred to the insurer for the pure premium, because their sum is invariant to the limit value. However, in real settings, when the specific loadings are added to the pure premium, the expected loss is always smaller than the final premium; therefore, solving the problem would require the use of a utility function.

5. Conclusion

The paper presented an illustrative example showing how to set the optimal limit value when the random individual loss is considered. The presented optimization problem is actually general, and various relevant probability distributions may be considered, not only gamma distribution. In addition, the problem can be extended by the chance constraint, which represents the need for the probability of individual loss exceeding the limit value to be at least a predefined value. Furthermore, adding fixed as well as variable deductibles (co-insurance) might be an interesting extension.

The example does not involve real rates, but they can be obtained from commercial web calculators and can be approximated by regression of the real premium on various limit values, as shown in this paper. We also pointed out that the potential loss is unlimited because of the gamma distribution considered in the optimization problem. In addition, it yields a relatively small optimal limit value, because there is only a small probability of losses exceeding this level. Therefore, the gamma distribution might be replaced with another distribution that sets the maximum of the potential loss and corresponds better to the loss occurrence.

However, the objective function of the optimization problem was presented in a general form and can accommodate any probability assumptions. Therefore, all imperfections related to the gamma distribution are a matter of this specific distribution rather than our optimization problem, which can incorporate more convenient probability assumptions as well as additional constraints. In addition, the problem allowed the exclusion of the expected utility theory, because the objective function minimized the potential rather than the expected loss.

References


