Explicit Description of the Input Data for the Program CRAC 2.0 Used in the Applications of the Credibility Theory

Virginia ATANASIU
Department of Mathematics, Academy of Economic Studies
virginia_atanasiu@yahoo.com

In this paper a brief overview of the structure and the possibilities of the program CRAC 2.0 is given. It will be shown how sectors can be determined in order to use the hierarchical model that is built into the software. Furthermore a general structure for defining insurance problems to be solved by CRAC 2.0 will be discussed.
Mathematics Subject Classification: 62P05.
***Keywords**: the software package CRAC 2.0, Jewell's hierarchical model, practical insurance problems, structure of CRAC – input.*

Introduction

In this paper we will discuss some practical insurance problems that can be solved by means of credibility theory. All numerical results in this paper were obtained using the software package CRAC 2.0, which uses Jewell's hierarchical model. Here we consider applications of credibility theory dealing with real life situations, and implemented on real insurance portfolios.

1. Structure of the software

Introduction and philosophy

The CRAC 2.0 input is designed in such a way that quite a number of practical problems can be solved. We describe the input format for the different variables. An example will illustrate the general guidelines.

The CRAC – software is based on a two – level semi – linear hierarchical credibility model. It is semi – linear in the sense that some (fixed) transformation of the claim figures is used. It is two – level hierarchical, as the individual contracts on the down – level are grouped together in sectors on an intermediate level.

So decisions can be made on three different levels:

-the *top – level* based on collective estimators for the whole portfolio;

-the *in – between sector level* based on the estimators for the different sectors defined by the user;

-the *down – level* based on the individual estimators for all contracts.

For every contract in the portfolio (a contract can also be an average of a group of contracts), data for T observation periods can be given. For each contract the system allows for missing data. For a successful application of the software, T should be at least 3.

1.2 Subdivision of the portfolio into sectors

The basic idea behind hierarchical credibility is described in Subsection 1.4. Instead of working with individual contracts at one side and the portfolio as a whole at the other side, an in – between or sectorial level is created. In this way we obtain three decision levels: global (for the portfolio as a whole) denoted by m, sectorial, denoted by N_p^a and individual, denoted by M_{pj}^a. The portfolio is split into a number of sectors and each of these sectors contains individual contracts. It is evident that this subdivision of the portfolio into smaller subgroups will be different from one application to another. The determination of the sectors could be performed based on classical multivariate statistical techniques such as discriminant analysis and cluster analysis. On the other hand it should be noted that for several reasons it is just not possible to construct an ideal subdivision of the portfolio. That should not prevent us from actually making the non – ideal subdivision and applying credibility theory.

The sectors can be determined by CRAC 2.0 according to either one two or three different

characteristics. Each individual record contains up to 3 different variables ranging from 1 to $S1$, $S2$, and $S3$ respectively. Then the following subdivisions are possible:
-on the basis of one of these variables resulting in $S1$, $S2$, or $S3$ sectors;
-on the basis of two combined criteria resulting in two subsequent subdivisions in cells according to the chosen pairs of S- variables. This gives for instance a total of $S1 \cdot S2$ cells if the first two criteria are used;
-on the basis of all three variables, producing $S1 \cdot S2 \cdot S3$ sectors.
Consider the **two** following **examples**.

Example 1: **Individual fire insurance**
A subdivision into sectors might be based on the following criteria:
• owned house or rented house;
• size of the house: small, medium or big house;
• geographical position of the house (district, state,…).
The choice of this subdivision may be brought on by statistical differences of the various groups, or simply imposed by the management.

Example 2: **Health insurance**
We consider the principle of this type of insurance (to be discussed in Section 2) to be that a firm insures all its employees as a group for medical costs. So the client of the insurance company is the firm and not the individual employee. A subdivision of a portfolio of such contracts might be based on the following criteria:
• kind of activities of the firm: services, chemical, agricultural,…;
• geographical position of the firm: e.g. Northern, Central and Southern part of the country;
• turnover of the firm: a number of classes could be defined.
The total portfolio in these two examples can then be split up according to one, two or all three of these criteria.

1.3 Definition of the problem; a standard procedure

It is obvious that not only the sectors but also the quantity to be calculated will be different from one application to another. In fire insurance we normally will be interested in calculating a permillage of the capital insured, whereas in the problem of health insurance we want to calculate a premium (in some currency) per individual insured.
For this reason a standard procedure is provided within the CRAC – software. This procedure allows the solution of quite a number of practical problems.
The standard form for defining problems is as follows:

$$X_{pjr} = \frac{\left(\left(numerator\right)_{pjr} - deductible\right)_+}{factor - \left(deno\min ator\right)_{pjr}} \quad (1),$$

where $q_+ = \max\{0, q\}$, p denotes the sector this policy belongs to, j denotes the contract and r denotes the observation period. For the numerator we usually take claim payments, while the denominator expresses some measure of exposure like premium volume or number of participants. We will illustrate this standard form for the two examples mentioned previously.

Example 1: **Individual fire insurance**
When we are interested in calculating a risk premium as a permillage of the capital insured for each sector as defined earlier, the numerator is the amount paid by the insurance company, the denominator is the capital insured. By choosing the scaling factor as 0,001 the quantity X_{pjr} represents the risk permillage we are interested in.

Example 2: **Health insurance**
In this example we calculate a risk premium per individual, so we have to take the numerator as the amount paid by the insurer; the denominator represents the number of insured employees in the firm.

1.4 Structure of CRAC – input

An input – file to be used in CRAC 2.0 must contain the following data for each record:
1.Policy number or identification number.
2.Specification of the sector.

Three fields containing the values according to which the subdivisions in sectors is organized. These variables range from 1 to $S1$, $S2$ or $S3$.

3.Number of observations.

This value must be greater than zero and at most equal to T.

4.For each observation period i ($i = -1, -2, ..., -T$) the following fields:

- An indicator which is set to zero in case no observation for the period i is available; otherwise it is 1;
- A numerator for the observation period i;
- A denominator for the observation period i;

This field can only contain a value of zero when field 4a indicates a non – available period;

- A weight factor for the observation period i;

This field reflects the importance in the portfolio of each individual contract.

1.5 A practical example

In this subsection we will combine all elements from the previous subsections into one example. The goal of this is to illustrate how practical problems can be translated into problems that can be handled by the CRAC – software. In the following sections we will concentrate on the results obtained rather than on the practical problems arising from working with this software package.

To determine the risk premium for individual fire insurance, assume statistical information sis available for the last four years. The risk premium must be expressed as a permillage of the capital insured. The user judges a subdivision into the following sections necessary:

1) houses in the North (N) and houses in the South (S);

2) houses with an insured capital of less than \$100.000 (SH), from \$100.000 up to \$250.000 (MH), and more than \$250.000 (BH);

3) houses in the city (IC) and houses outside the city (OC).

It is not the intention to set up an individual tariff. The only aim of the user is to set up a sectorial investigation. The user should make the following selections:

1.$T = 4$;

2.no individual calculations;

3.$S1 = 2$ (N/S), $S2 = 3$ (SH/MH/BH) and $S3 = 2$ (IC/OC);

4.the variables defining the subdivision in sections; when the three criteria are combined, the following sectors appear:

-sector 1:	N	SH	IC	-sector 7:	S	SH	IC
-sector 2:	N	SH	OC	-sector 8:	S	SH	OC
-sector 3:	N	MH	IC	-sector 9:	S	MH	IC
-sector 4:	N	MH	OC	-sector 10:	S	MH	OC
-sector 5:	N	BH	IC	-sector 11:	S	BH	IC
-sector 6:	N	BH	OC	-sector 12:	S	BH	OC

5.names of the sectors must be given.

In this situation a typical record for an individual contract would have the following layout:

-policy number: not required in this case

-three numerical zones, the first one containing the value 1 (N) or 2 (S); the second one containing the values 1 (SH), 2 (MH) or 3 (BH); the third one containing the values s1 (IC) or 2 (OC).

-the number of observation periods available ($\geq T$),

-for observation period i (for $i = -1, -2, \ldots, -T$)

- 1 (observation available) or 0 (observation not available);
- numerator, the claim size for period i;
- denominator, the capital insured for time period i;
- weight, e.g. the premium paid in period i.

Because the estimate for the risk premium has to be expressed in terms of a permillage of the capital insured SF has to be put equal to 0.001. A total of 75 bytes are necessary in this case for each record. An example of a record looks as follows:

0103022000000000000000001000100300000060010000002800000560000000000000000000

From this record we see that this contract belongs to the North (N), it is a big house (BH) outside the city (OC). There are data for two years:

- year -2: claim size of 000100, capital insured equal to 300.000 and premium of 0600;
- year -3: claim size of 000000, capital insured equal to 280.000 and premium of 0560.

2. An example in health insurance

Description of health insurance

In the example of health-insurance considered here, a firm insures all its employees as a group for medical costs. The clients of the insurance company are firms, not individual employees.

It has been decided that the subdivision of the portfolio of these contracts will be based on the following criteria (the values between brackets give the corresponding sector number in the CRAC software).

- kind of activities of the firm: metal (1), chemical (2), agricultural (3), kind 4 (4), kind 5 (5), kind 6 (6) and food (7);
- geographical positions of the firm: North (1), Central (2), South (3);
- number of employees of the firm: we defined 7 different groups from small (1) to immense (7).

Notice that the portfolio could be split up according to three criteria: the first and the third have values ranging from 1 to 7, the second has values ranging from 1 to 3. By combining all criteria we obtain a subdivision into 147 = 7·3·7 sectors.

Example of the input data

In the table below, an example is given of how the input data might look. The first record of this file contains information on policy Z 1024. The three following fields, each of length 2, have the values 030302, indicating that this policy covers an agricultural firm (03) in the Southern part (03) of the country; on the basis of the number of employees, the firm is classified as „size 2” (02).

Next, we see that there is only one observation period available, namely period -3. In this period a amount of 00092357 has been paid; the number of employees insured equals 000045 and the weight has been chosen equal to this number of employees. The interpretation of the other records is straightforward.

Z1024	03	03	02	1	0	00000000	000000	000000	0	00000000
Z1028	06	01	03	3	1	00194514	000547	000547	1	00095224
Z1185	06	01	05	3	1	00039453	001036	001036	1	00141023
Z1188	06	03	05	3	1	00030216	000621	000621	1	00258241

000000	000000	1	00092357	000045	000045
000547	000547	1	00109245	000547	000547
001036	001036	1	00363791	001036	001036
000621	000621	1	00453345	000621	000621

The input data has to be checked first for various errors, such as denominators that are zero, sectors that do not exist, and so on. In our demonstration run 3 out of 703 policies had to be rejected in advance.

Construction of sectors

We could make a subdivision of this portfolio in seven different ways:

1.based on kind of activities: 7 sectors;
2.based on geographical position: 3 sectors;
3.based on a number of employees: 7 sectors;
4.based on a combination of (1.) + (2.):7 × 3 = 21 sectors;
5.based on a combination of (1.) + (3.):7 × 7 = 49 sectors;
6.based on a combination of (2.) + (3.):3 × 7 = 21 sectors;
7.based on a combination of (1.) + (2.) + (3.):7 × 3 × 7 = 147 sectors.

In this application we will use subdivision 3.. In the following section we will comment on the credibility results generated by the software for this particular case and subdivision.

Conclusions

We give a rather explicit description of the input data for the program CRAC 2.0 used, only to show that in practical situations there will always be enough data to apply credibility theory to a real insurance portfolio. The point we want to emphasize is that practical application of credibility theory is feasible nowadays using appropriate software. In this paper we consider applications of credibility theory dealing with real life situations, and implemented on real insurance portfolios. Though more examples could be given, we limit ourselves to the introduction of a problem of health insurance, and (briefly) one in fire insurance. In these examples we try to demonstrate what kind of data is needed to apply credibility theory. The examples show that credibility theory is really a useful tool - perhaps the only existing tool - for such insurance applications. The fact that it is based on complicated mathematics, involving conditional expectations, needs not bother the user more than it does when he applies statistical tools like SAS, GLIM, discriminant analysis, and scoring models. These techniques can be applied by anybody on his own field of endeavor, be it economics, medicine, or insurance.

References

[1]. *De Vylder, F.* & *Goovaerts, M.J.*, **Semilinear credibility with several approximating functions**, Insurance: Mathematics and Economics, 4, 155- 162. (Zbl. No. 0167-6687), 1985.

[2].*Gerber, H.U.* & *Jones, D.A.,* **Credibility formulas of the updating type**, in “Credibility theory and application”, Proceedings of the Berkeley Actuarial Research Conference on credibility, Academic Press, New York, 89-109, 1975.

[3]. *Gerber, H.U.*, **Credibility for Esscher premiums**, Mitteilungen der VSVM, 80, 307-312, 1980.

[4]. *Goovaerts M. J.*, *Kaas R., Van Heerwaarden A., E.*, *Bauwelincks T.*, **Insurance Series**, **Volume 3, Effective Actuarial Methods**, Elsevier Science Publishers B.V., 187-211, 1990.

[5]. *Pentikäinen T.*, Daykin C. D., Pesonen M., **Practical Risk Theory for Actuaries**, Chapman & Hall, 1993.

[6]. Sundt B., **An Introduction to Non-Life Insurance Mathematics**, volume of the “Mannheim Series”, 22-54, 1984.