Commun.Fac.Sci.Univ.Ank.Series A1
Volume 55, Number 2, Pages 9–16 (2006)
ISSN 1303–5991

A STUDY TO EXAMINE BÜHLMANN-STRAUB CREDIBILITY MODEL IN GENERALIZED LINEAR MODELS

MERAL EBEGİL

ABSTRACT. Credibility theory is interested calculating weighted estimation which is used to make balanced partition between past data and recent data. In recent years, some well-known statistical methods are applied to credibility theory. Nelder and Verrall (1997) set the relationship between credibility theories and generalized linear models by using Bühlmann model among credibility models and hierarchical generalized linear model in likelihood basis among statistical models [7]. In this study, the aim is to bring forward a statistical view to the credibility theory, which is used to determine the premium in the non-life branches of insurance. For this aim, it is tried to extend the study of Nelder and Verrall (1997) by setting the relationship between Bühlmann-Straub credibility model and hierarchical generalized linear models in generalized linear mixed models. For this purpose, credibility formula of Bühlmann-Straub model is formed by the help of hierarchical generalized linear models. Such studies will provide both of the insurance companies and insured people with fairer, more sensible and more reliable determinations for the premium by making use of different types of data in insurance sector. Moreover, establishment of a statistical view to the credibility theory will save this theory from being a tool used merely for the determination of the amount of premium in insurance sector and will enable its applicability as a method that can be used in the science of statistics.

1. INTRODUCTION

Over the last decade generalized linear models (GLMs) has become a widely-used statistical tool to model actuarial data. In this matter, start was given by McCullagh and Nelder (1989) who offered actuarial illustrations in their study [6]. Another study in actuarial sense was conducted by Haberman and Renshaw (1996) [5]. One of the advantages of GLMs is that regression is no longer restricted to normal data, but extended to distributions from the exponential family. This property enables appropriate modeling of e.g. frequency counting, skewed data

Received by the editors Jan 3, 2006; Rev. July 21, 2006 Accepted: Aug. 31, 2006.
2000 *Mathematics Subject Classification.* Primary 62J12.
Key words and phrases. Credibility Theory; Generalized Linear Models; Hierarchical Generalized Linear Models; Bühlmann-Straub Credibility Models.

and binary data. Furthermore, a GLM models the additive effect of explanatory variables on a transformation of the mean, instead of the mean itself.

For the data of an insurance portfolio, a model including both fixed effects and random effects is necessary and random variables used in standard GLMs must be independent. However the assumption that variables are independent can not be provided in most of the actuarial and general statistics problems. To model this type of data, generalized linear mixed models (GLMMs), which is a version of GLMs, can be used. GLMMs extend GLMs by including random effects in the linear predictor. The random effects not only determine the correlation structure between observations on the same subject, they also take account of heterogeneity among subjects, due to unobserved characteristics.

In actuarial context Frees et al. (1999, 2001) provide an introduction to linear mixed models (LMMs) and illustrate how they can be used ratemaking [2, 3]. Also, Frees et al. (2004) has done a study about normal hierarchical linear models in actuarial context [4]. The study of Russell and Greig (1999) offered the definition and application of random effects models, as a better alternatives to empirical credibility [8]. In these studies, in general, examples were given; however the relation of the credibility models to the linear models were not theoretically shown. Using likelihood-based hierarchical generalized linear models (HGLMs) Nelder and Verrall (1997) describe a possible method to interpret traditional credibility theory in the framework of GLMs. In this method, derivation for exponential families of credibility formula for Bühlmann model by making the use of HGLM is shown. In other words, connection of Bühlmann credibility model with GLMs is formed [7].

In this study, the study of Nelder and Verrall (1997) is based and it is tried to form credibility formula theoretically for Bühlmann-Straub model which is the generalized form of Bühlmann model. Thus, connection of Bühlmann-Straub model with HGLMs is formed. HGLMs are GLMMs with random effects that are not necessarily normally distributed; an assumption that is traditionally made. Consequently, the study of Nelder and Verrall (1997) is extended to the Bühlmann-Straub model which is the generalized form of Bühlmann model. Thereby, a statistical approach was put forward to the credibility theory that is used for determination of the premium to be paid by the insured in non-life insurance branches. In the credibility theory, credibility estimations are done from the distribution free, without making an assumption for distribution. However provided that the relationship between credibility models and statistical models is formed, under the assumptions relating to the statistical model in use, it is possible to form hypothesis tests and confidence intervals for the credibility models too.

The second section of the study dwells upon GLMs. In third section also, forming Bühlmann-Straub credibility model by the help of HGLMs and defining Bühlmann-Straub credibility model in the GLMs framework are tried.

2. Generalized Linear Models (GLMs)

An important feature of GLMs is that they extend the framework of general (normal) linear models to the class of distributions from the exponential family. The basis of GLMs is the assumption that the data are sampled from a one-parameter exponential distribution family of distribution. Here, some principal properties about GLMs will be clarified. First, take a consideration of an observation y. A one-parameter exponential family of distribution has a log-likelihood of the form below:

$$\frac{y\phi - \Psi(\phi)}{\varphi} + c(y, \varphi) \tag{2.1}$$

As $\Psi(.)$ and $c(.)$ are known functions; ϕ is the canonical or natural parameter of one-parameter exponential family and φ is the scale or dispersion parameter which is assumed known. Below canonical form is used to show of densities from exponential family.

$$f(y) = \exp\left(\frac{y\phi - \Psi(\phi)}{\varphi} + c(y, \varphi)\right) \tag{2.2}$$

Let $Y_1, Y_2, ..., Y_n$ be independent random variables from a member of the exponential family. The following two well-known equations can be written for distributions from this family.

$$\mu_i = E(Y_i) = \frac{\partial \Psi(\phi_i)}{\partial \phi_i} \qquad i = 1, 2, ..., n \tag{2.3}$$

and

$$Var(Y_i) = \varphi \frac{\partial^2 \Psi(\phi_i)}{\partial \phi_i^2} = \varphi V(\mu_i) \qquad i = 1, 2, ..., n \tag{2.4}$$

$V(.)$ is called as the variance function. $Var(Y_i)$ is the product of two quantities and depends on the canonical parameter (and hence on the mean).

Link function defines the relation between linear predictor and mean. A GLM can be defined by the help of a distribution determined by a link function, like equation (2.4), and a linear predictor.

Another important property of GLMs is that they provide around the transformation of data. Instead of a transformed data vector, a transformation of the mean is modeled as a linear function of explanatory variables. In this context, a linear predictor can be defined as follows

$$g(\mu_i) = \eta_i = (X\beta)_i \tag{2.5}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_p)'$ is the parameter vector including model parameters. $X; (n \times p)$ is the design matrix, g is the link function and η_i is the i^{th} element of the so-called linear predictor. $g(\mu) = \phi$ is a special case called as canonical link function.

3. HGLMs and Bühlmann-Straub Credibility Model

In the study carried out by Nelder and Verrall (1997), how credibility theory takes part in HGLMs is shown by the help of Bühlmann credibility theory [7]. Moving here, in this part of study credibility formula of Bühlmann-Straub model will be obtained for exponential distribution family under the same assumptions stated by Bühlmann-Straub (1970) [1]. When positive weights stemming from difference in model assumptions are taken into consideration, a credibility formula like that in Bühlmann model can be formed by using HGLM for Bühlmann-Straub model.

Let data for $i = 1, 2, ..., n$; let $t = 1, 2, ..., T_i$ be described as y_{it} and assume that $T_i = s$ for every i. Thus i index the risks within the collective. In credibility theory, it is assumed that each risk has a risk parameter and this parameter for risk i is defined as θ_i. The assumptions of the model of Bühlmann-Straub (1970) are below [1].

i. The risks and hence θ_is are independently but not identically distributed.

ii. $y_{it} \mid \theta_i$ are independently but not identically distributed.

Let it be assumed that $y_{it} \mid \theta_i$ is distributed according to an exponential family. Here, define $\mu(\theta_i) = E[y_{it} \mid \theta_i]$. Note that; under the assumptions of the model, $E[y_{it} \mid \theta_i]$ does not depend on t. In this case, the canonical parameter for observation y_{it} will not depend on t. So, let it be assumed that the statement can be written as follows

$$\phi_i' = \phi(\mu(\theta_i)) = \phi(\alpha_i) \tag{3.1}$$

Here, ϕ is the canonical link function and α_i is a random effect for ith group or insured. Thus, it is $\mu(\theta_i) = \alpha_i$ for standard credibility model. If it is defined that $\nu_i = \phi(\alpha_i)$, the equation below is obtained:

$$\phi_i' = \nu_i \tag{3.2}$$

The expression in equation (3.2) does not depend on t, too. This also makes that $Var[y_{it} \mid \theta_i]$ does not depend on t [7].

The expressions in equation (3.1) and equation (3.2) have defined the distribution of random variable in each risk under the condition of risk parameter θ_i. At the same time, it is also necessary to define the structure of the collective the distribution of $\{\theta_i \quad i = 1, 2, ..., n\}$. For this, a "hierarchic likelihood", h, is described and this h is maximized. HGLM is defined by defining the kernel of the log-likelihood for $\phi(\alpha_i)$ as [7].

$$a_1\phi_i' - a_2\Psi(\phi_i') \tag{3.3}$$

In the actuarial literature, the distribution of the random effects is known as the structure of the collective. Besides, log-likelihood of θ_i is described on the statement of $\phi(\mu(\theta_i))$. The condition on θ_i is originated by $\mu(\theta_i) = \alpha_i$. This statement described in equation (3.3) is the latter that we wish to estimate. From equation (3.3) and the distribution of $y_{it} \mid \theta_i$, we may define a hierarchical log-likelihood as [7].

$$h = \sum_{i,t} l\left(\phi_i' y_{it} \mid \nu_i\right) + \sum_i l\left(\nu_i\right) \tag{3.4}$$

When equation (3.4) is re-arranged according to assumptions of Bühlmann-Straub model, it is obtained as follows

$$h = \sum_{i,t} \left(\frac{y_{it}\phi_i - \Psi(\phi_i)}{\varphi/m_t} \right) + c(y_{it}, \varphi/m_t) + a_1\phi_i' - a_2\Psi(\phi_i') \tag{3.5}$$

where m_t is a known constant representing the amount of exposure during the t^{th} policy period. In this case, an estimate of $\mu(\theta_i) = \alpha_i$ is required. The mean random effects $\{\alpha_i : \ i = 1, 2, ..., n\}$ are estimated by maximizing the hierarchical likelihood, equation (3.4), as follows. During this process, with a difference from Bühlmann model, resulting weights from difference in assumptions of Bühlmann-Straub model will be taken into consideration. These weights are shown as m_t in equation (3.5).

Using equation (2.3) is concluded as follows,

$$\frac{\partial \Psi(\phi(\alpha_i))}{\partial \nu_i} = \alpha_i \qquad i = 1, 2, ..., n$$

Thus, $\dfrac{\partial h}{\partial \nu_i} = \sum_{t=1}^{s} \left(\dfrac{y_{it} - \alpha_i}{\varphi/m_t} \right) + a_1 - a_2\alpha_i$ is stated. Equating $\dfrac{\partial h}{\partial \nu_i}$ to zero gives

$$\sum_{t=1}^{s} m_t \left(\frac{y_{it} - \widehat{\alpha}_i}{\varphi} \right) + a_1 - a_2\widehat{\alpha}_i = 0 \qquad i = 1, 2, ..., n$$

$$\sum_{t=1}^{s} m_t y_{it} - m\widehat{\alpha}_i + \varphi a_1 - \varphi a_2 \widehat{\alpha}_i = 0 \tag{3.6}$$

where, $m = \sum_{t=1}^{s} m_t$ After necessary computations are done by using equation (3.6), it is obtained as follows

$$\widehat{\alpha}_i = \frac{\sum_{t=1}^{s} m_t y_{it} + \varphi a_1}{m + \varphi a_2} \qquad i = 1, 2, ..., n$$

$$= \frac{m}{m + \varphi a_2} \frac{\sum_{t=1}^{s} m_t y_{it}}{m} + \frac{\varphi a_2}{m + \varphi a_2} \frac{a_1}{a_2}$$

$$= Z\overline{y}_i + (1 - Z)\mu$$

where, $\overline{y}_i = \frac{1}{m}\sum_{t=1}^{s} m_t y_{it}$, $Z = \frac{m}{m + \varphi a_2}$ and $\mu = \frac{a_1}{a_2}$. Z is called as Bühlmann-Straub credibility factor and m is the total exposure for all policy periods.

Thus, estimation of α_i, by choosing distribution described in equation (3.3) for random effects and by using canonical link function, is in the form of a credibility estimation providing $E\left(\mu\left(\theta_i\right)\right) = \frac{a_1}{a_2}$. This obtained form corresponds to Bühlmann-Straub model from known credibility models. Thus, credibility Formula of Bühlmann-Straub model is derived by the using of GLMs.

4. Conclusion

Linear models are statistical tools which are broadly used in modeling many types of data. For distributions from the exponential family GLMMs extend GLMs by including random or subject-specific effects next to the fixed effects in the structure for the linear predictor. HGLMs are GLMMs with random effects that are not necessarily normally distributed. The data in an insurance portfolio include both fixed effects and random effects. For this reason, Actuarial data can be modeled by GLMs.

In this study, by using HGLMs which is a special type of GLMMs included in GLMs, it is shown that how credibility formula of Bühlmann-Straub credibility model are formed theoretically. Thus, the link between GLMMs which are appropriate to model actuarial data and Bühlmann-Straub credibility model is established. In conclusion, the models in credibility theory can be used more extensive range by GLMs

The advantages of putting the credibility models into such a statistical model framework are (i) to enrich actuaries' choices of the credibility models (ii) to provide actuaries with a disciplined process for choosing the models and (iii) to provide actuaries with the chance of obtaining more reliable and accurate credibility premium estimates by blending available information in actuarial science. So, the actuarial data can be modeled in a broad field. That is, the framework of HGLMs allows a more extensive range of models to be used than straightforward credibility theory. By linking Bühlmann-Straub credibility model to HGLMs, usage of Bühlmann-Straub model which is one of the credibility models is extended by making it be used in modeling many different types of data such as skewed data and binary data. In addition, such studies will contribute to the establishment of a scientific basis for the insurance sector. Credibility models, like linear models, form a subset of the applications of statistical models to the actuarial science.

In conclusion, by establishing the relation of the credibility models to the statistical models, it was possible to put forward a statistical view to the credibility theory ,which is a method used for the assessment of the amount of insurance premium. Establishment of a statistical view to the credibility theory redeems the credibility theory from being merely an instrument for the determination of premium payment in insurance sector and makes this theory a useful method, which can also find application in the science of statistics. In short, by linking the credibility models to statistical models, it becomes possible to apply the statistical methods the credibility models as well. Thus, it becomes possible to follow a disciplined process with scientific basis for the credibility theory.

ÖZET. Kredibilite teorisi, geçmişe ait veriler ile son döneme ilişkin veriler arasında dengeli bir paylaştırma yapmak için kullanılan ağırlıklı tahmini hesaplama yöntemidir. Son yıllarda, iyi bilinen bazı istatistiksel metotlar kredibilite teorisine uygulanmaktadır. Nelder ve Verrall (1997)'deki çalışmalarında, kredibilite modellerinden Bühlmann modeli ile istatistiksel modellerden likelihood'a dayalı hiyerarşik genelleştirilmiş doğrusal modelleri kullanarak, kredibilite teorisi ile genelleştirilmiş doğrusal modeller arasındaki ilişkiyi kurmuşlardır [7]. Bu çalışmada, sigortacılıkta hayat dışı sigorta branşlarında prim belirlemekte kullanılan kredibilite teorisine istatistiksel bir bakış açısı getirilmesi amaçlanmıştır. Bu amaçla, Nelder ve Verrall (1997)'in çalışmaları Bühlmann-Straub kredibilite modelinin genelleştirilmiş doğrusal karma modeller içinde yer alan hiyerarşik genelleştirilmiş doğrusal modellerle bağlantısı kurularak genişletilmeye çalışılmıştır. Bu nedenle, Bühlmann-Straub modelinin hiyerarşik genelleştirilmiş doğrusal modeller yardımıyla kredibilite formülü oluşturulmuştur. Böylece bu tür çalışmalar, sigortacılık sektöründe farklı veri tipleri kullanılarak, hem sigorta şirketi hem de sigortalılar açısından daha adil, duyarlı ve sağlıklı prim miktarları belirlenmesini sağlayacaktır. Ayrıca kredibilite teorisine istatistiksel bakış açısı kazandırmak, kredibilite teorisini sadece sigortacılık sektöründe prim miktarını belirlemede kullanılan bir araç olmaktan kurtarıp istatistik biliminde de kullanılabilecek bir yöntem haline getirecektir.

REFERENCES

[1] Bühlmann, H. and Straub, E., Glaubwürdigkeit für Schadensätze, Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker, (1970) 111-133.

[2] Frees, E.W.,Young, V.R. and Luo, Y., A Longitudinal Data Analysis Interpretation of Credibility Models, Insurance: Mathematics and Economics, 24(3) (1999), 229-247.

[3] Frees, E.W.,Young, V.R. and Luo, Y., Case Studies using Panel Data Models, North American Actuarial Journal, 5(4) (2001), 24-42.

[4] Frees, E.W.,Young, V.R. and Luo, Y., Credibility Rate Making using Collateral Information, Scand. Actuarial Journal, 6 (2004), 448-461.

[5] Haberman,S and Renshaw, A.E., Generalized Linear Models and Actuarial Science, The Statistician, 45 (1996), 407-436.

[6] McCullach, P. and Nelder, J.A., 1989, Generalized Linear Models, Monographs on Statistics and Applied Probability, Chapman and Hall, New York.

[7] Nelder, J.A. and Verrall,R.J., Credibility Theory and Generalized Linear Models, ASTIN Bulletin, 27(1) (1997), 71-82.

[8] Russell H., Greig Jr., Random Effects Linear Statistical Models and Bühlmann–Straub Credibility, Casualty Actuarial Society Vol: Winter,: Arlington, Virginia (1999), 387-404.

Current address: Department of Statistics, Faculty of Science and Arts,University of Gazi, Beşevler, 06500, Ankara, Turkey

E-mail address: `mdemirel@gazi.edu.tr`

URL: `http://math.science.ankara.edu.tr`