Journal of the Korean Statistical Society 40 (2011) 1–9

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Kernel Poisson regression machine for stochastic claims reserving

Jooyong Shim [a], Changha Hwang [b,*]

[a] *Department of Applied Statistics, Catholic University of Daegu, Gyeongbuk 702-701, South Korea*

[b] *Department of Statistics, Dankook University, Gyeonggido 448-160, South Korea*

ARTICLE INFO

Article history:
Received 7 July 2009
Accepted 28 January 2010
Available online 26 February 2010

AMS 2000 subject classifications:
primary 62P05
secondary 62J12

Keywords:
Canonical parameter
Chain ladder
Claims reserving
Cross-validation function
Kernel machine
Penalized likelihood
Poisson regression

ABSTRACT

This paper shows how the kernel Poisson regression machine (KPRM) can be applied in the context of claims reserving. The paper concentrates on the chain-ladder technique, within the framework of the chain-ladder linear model. It is shown that the KPRM can provide stable reserve estimates. The ordinary cross-validation (OCV) and the generalized cross-validation (GCV) functions are introduced to determine hyperparameters which affect the performance of the KPRM. Experimental results are then presented which indicate the performance of the KPRM.

1. Introduction

Claims reserving is central to the insurance industry. The reserve represents money that is set aside to meet claims that arise in the future on written policies. The correct estimation of the reserve represents an important task for the insurance company. The chain-ladder model is probably the most popular one for estimating claims reserves. The main reason for this is its simplicity and the fact that it seems to work with almost no assumptions except for a consistent delay pattern in the payment of claims. This method is deterministic and provides only a point estimate. However, it is necessary to be able to estimate the variability of claims reserves. Stochastic claims reserving models extend traditional models to allow this additional measure to be estimated. Stochastic claims reserving models produce estimates not only of the expected value of the future payments, but also of the variation about that expected value.

Many stochastic claims reserving models have been proposed over the last two decades. The two main models are the Poisson and lognormal models, with gamma and negative binomial models both receiving occasional reference. The Poisson model first appeared in stochastic claims reserving literature in the early 1990s (Mack, 1991). It was later formalized into the generalized linear model (GLM) framework by Renshaw and Verrall (1994), which provides the link to contingency tables (England & Verrall, 1999). This model has less rigid restrictions on negative incremental claims than the lognormal model. Indeed, it assumes that the sum of the incremental claims in every column (development year) and row (accident year) of the loss data triangle has to be non-negative (England & Verrall, 1999, 2002). The Poisson model assumes that the variance is equal to the mean, which may not always be met. An overdispersed Poisson model is usual in stochastic claims

* Corresponding author.
E-mail addresses: ds1631@hanmail.net (J. Shim), chwang@dankook.ac.kr (C. Hwang).

1226-3192/$ – see front matter
doi:10.1016/j.jkss.2010.01.004

reserving. This overdispersed model is usually estimated based on a quasi-likelihood approach, since the log-likelihood function is no longer proper (Montgomery, Peck, & Vining, 2001). Verrall (1996) first considered the use of a generalized additive model (GAM) in the context of claims reserving. This preserves the underlying structure of GLM, but replaces the linear predictor by a nonparametric smoothing procedure. GAM (Hastie & Tibshirani, 1990) allows greater flexibility in the fitted model, and removes the requirement that the parametric model is valid over the whole range of values to be smoothed.

We will propose the kernel Poisson regression machine (KPRM) which can be applied in the context of claims reserving within the framework of the chain-ladder model. We replace the linear predictor in GLM by the additive predictor including both parametric and nonparametric components. This machine is derived based on the kernel trick, which was introduced in Aizerman, Braverman, and Rozonoer (1964) and has been popularized since Vapnik (1995). Cawley, Janacek, and Talbot (2007) first considered the use of the kernel trick in GLM. The KPRM makes it possible to derive the generalized cross-validation (GCV) method for choosing the hyperparameters which affect the performance of the machine. The rest of this paper is organized as follows. In Section 2 we review stochastic claims reserving models that are related to the chain-ladder technique. In Section 3 we propose the KPRM which is based on the penalized negative log-likelihood, and propose GCV function for the model selection. Section 4 illustrates how to evaluate the precision of reserves estimates. Section 5 presents some numerical studies to illustrate our method. In Section 6 we give the conclusion.

2. Stochastic models for chain-ladder technique

All the main stochastic claims reserving models, including Poisson, lognormal, gamma and negative binomial, as well as the mixture model, can be written in terms of GLMs. In this section we briefly illustrate the link between the chain-ladder technique and stochastic claims reserving models including Poisson model, which provides a fundamental basis for the KPRM to be proposed in Section 3. For details see England and Verrall (2001, 2002), Renshaw and Verrall (1994, 1998) and Verrall (1996).

2.1. Chain-ladder technique

The standard chain-ladder technique uses cumulative claims data. It is well known that it is irrelevant whether cumulative or incremental data are used when considering claims reserving in a stochastic context, and it is easier for the explanations here to use incremental. Thus, without loss of generality, we assume that the data consist of a triangle of incremental claims:

$$\{y_{ij} : i = 1, \dots, n; j = 1, \dots, n - i + 1\}. \tag{1}$$

Here y_{ij} denotes the incremental claims amount arising from accident year i paid in development year j. Accident year refers to year in which the accident giving rise to a claim occurs and development year refers to number of years elapsed since the accident. The cumulative claims are defined by

$$C_{ij} = \sum_{k=1}^{j} y_{ik}$$

and the development factors of the chain-ladder technique are denoted by $\{\lambda_j : j = 2, \dots, n\}$. The estimates of the development factor from the standard chain-ladder technique are

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}}.$$

The chain-ladder technique obtains the estimate of C_{ij} $(j > n-i+1)$, from the observed value of $C_{i,n-i+1}$ in a recursive way:

$$\hat{C}_{i,n-i+2} = C_{i,n-i+1}\hat{\lambda}_{n-i+2}$$
$$\hat{C}_{i,k} = \hat{C}_{i,k-1}\hat{\lambda}_k, \quad k = n-i+3, n-i+4, \dots, n.$$

2.2. Stochastic claims reserving and Poisson regression

We now briefly review the connections between the chain-ladder technique and some stochastic claims reserving models. We assume that all column sums of the triangle (1) are non-negative, i.e.

$$\sum_{i=1}^{n-j+1} y_{ij} \geq 0 \quad \text{for all } j. \tag{2}$$