Journal of the Korean Data & Information Science Society
2009, **20**(2), 445–452

한국데이터정보과학회지

Credibility estimation via kernel mixed effects model†

Jooyong Shim[1] · Tae Yoon Kim[2] · Sangyeol Lee[3] · Changha Hwang[4]

[1]Department of Applied Statistics, Catholic University of Daegu
[2]Department of Statistics, Keimyung University
[3]Department of Statistics, Seoul National University
[4]Department of Statistics, Dankook University

Received 2 January 2009, revised 3 February 2009, accepted 10 February 2009

Abstract

Credibility models are actuarial tools to distribute premiums fairly among a heterogeneous group of policyholders. Many existing credibility models can be expressed as special cases of linear mixed effects models. In this paper we propose a nonlinear credibility regression model by reforming the linear mixed effects model through kernel machine. The proposed model can be seen as prediction method applicable in any setting where repeated measures are made for subjects with different risk levels. Experimental results are then presented which indicate the performance of the proposed estimating procedure.

Keywords: Credibility model, kernel machine, longitudinal data, mixed effects model.

1. Introduction

The credibility ratemaking problem (Goovaerts and Hoogstad, 1987), which is a fundamental problem in actuarial sciences, is to predict the future claims amount using data from past claims. For details refer to Antonio and Beirlant (2008), Dannenburg *et al.* (1996). The linear mixed effects model (Hedeker and Gibbons, 2006) has been developed to describe the longitudinal data which are tracked over each subject's observing time and to predict the future observations. The linear mixed effects model can be applied to the credibility ratemaking problem (Frees *et al.*, 1999). For example, an insurance company's collection of risk classes can be viewed as generating longitudinal data in any policy period. An insurance company observes claims for those risk classes over several periods. Further, there may be input variables of the risk class which can be used to predict claims. Claims can be modeled across risk classes and time by the linear mixed effects model.

---

† This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund) (KRF-2007-314-C00045).

[1] Department of Applied Statistics, Catholic University of Daegu, Gyungbuk 712-702, Korea.
[2] Department of Statistics, Keimyung University, Daegu 704-701, Korea.
[3] Department of Statistics, Seoul National University, Seoul 151-742, Korea.
[4] Corresponding author: Department of Statistics, Dankook University, Gyeonggido 448-160, Korea. E-mail: chwang@dankook.ac.kr.

In this paper we propose a nonlinear credibility model within the framework of the credibility regression model of Hachemeister (1975). This is based on kernel trick of the support vector machine developed by Vapnik (1995). The rest of this paper is organized as follows. In Section 2 we give a review of the linear trend credibility regression model of Hachemeister (1975). In Section 3 and 4 we propose a nonlinear credibility model and estimation procedure of claims within the framework of the nonlinear credibility regression model. In Section 5 we perform the numerical studies through examples. In Section 6 we give the conclusions.

2. Linear credibility model

In this section we illustrate the linear trend credibility regression model of Hachemeister (1975) which is a special case of the linear mixed effects model. The Hachemeister model can be written as

$$y_{it} = \mathbf{x}_{it}\beta + \mathbf{x}_{it}\mathbf{b}_i + \epsilon_{it}. \tag{2.1}$$

The risk class, or subject, i is observed on T occasions, where $i = 1, \cdots, n$. y_{it} is the claims of the ith subject in period t, $\mathbf{x}_{it} = (1, t)$ is the known explanatory variable, β is the non-subject specific parameter vector, and $\mathbf{b}_i$ is the subject-specific random parameter which is assumed to follow a normal distribution $N(0, \sigma_B^2 I_2)$ with 2×2 identity matrix I_2 . A matrix form of this model is

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{X}_i\mathbf{b}_i + \epsilon_i, \tag{2.2}$$

where $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT})'$, $\mathbf{X}_i' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & T \end{pmatrix}$, $\beta = (\beta_1, \beta_2)'$, $\mathbf{b}_i = (b_{i1}, b_{i2})'$ and $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{iT})'$ for $i = 1, \cdots, n$.

Claims between subjects are assumed independent and the errors in each subject are assumed to follow a normal distribution $N(0, \sigma_\epsilon^2 I_T)$. The variance-covariance matrix of each subject can be written as

$$Var(\mathbf{y}_i) = \sigma_B^2 \mathbf{X}_i\mathbf{X}_i' + \sigma_\epsilon^2 I_T = V_i. \tag{2.3}$$

The generalized least squares(GLS) estimator of β can be obtained as $\hat\beta_{GLS} = (\sum_{i=1}^n \mathbf{X}_i' V_i^{-1} \mathbf{X}_i)^{-}1(\sum_{i=1}^n \mathbf{X}_i' V_i^{-1}\mathbf{y}_i)$, the GLS estimator of $\beta + \mathbf{b}_i$ based only on data of the i th subject is $\hat{\mathbf{a}}_i = (\mathbf{X}_i' V_i^{-1}\mathbf{X}_i)^{-1}\mathbf{X}_i' V_i^{-1}\mathbf{y}_i$ and the linear Bayes estimator of $\mathbf{b}_i$ is $\hat{\mathbf{b}}_i^B = \sigma_B^2 \mathbf{X}_i' V_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\hat\beta_{GLS})$. Then the credibility estimator of $\beta + \mathbf{b}_i$ is obtained as

$$\begin{aligned} \hat\beta_{GLS} + \hat{\mathbf{b}}_i^{(B)} &= \hat\beta_{GLS} + \sigma_B^2 \mathbf{X}_i' V_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\hat\beta_{GLS}) \\ &= \hat\beta_{GLS} - \sigma_B^2 \mathbf{X}_i' V_i^{-1}\mathbf{X}_i'\hat\beta_{GLS} + \sigma_B^2 \mathbf{X}_i' V_i^{-1}\mathbf{X}_i\hat{\mathbf{a}}_i \\ &= (I - \xi_i)\hat\beta_{GLS} + \xi_i\hat{\mathbf{a}}_i, \end{aligned} \tag{2.4}$$

where $\xi_i = \sigma_B^2 \mathbf{X}_i' V_i^{-1}\mathbf{X}_i$ is the credibility factor. The GLS estimator of β, $\hat\beta_{GLS}$, can be written as the weighted average of $\hat{\mathbf{a}}_i$ with credibility factors, $\hat\beta_{GLS} = (\sum_{i=1}^n \xi_i)^{-1}\sum_{i=1}^n \xi_i\hat{\mathbf{a}}_i$.

For the i th subject, the predicted claims amount for $t = T + 1$, $\hat y_{i,T+1}^{(c)}$, can be obtained as

$$(1, T+1)((I - \xi_i)\hat\beta_{GLS} + \xi_i\hat{\mathbf{a}}_i)\text{for } i = 1, \cdots, n. \tag{2.5}$$

3. Nonlinear credibility model

We consider nonlinear form of the Hachemeister model (2.1) as follows,

$$y_{it} = \phi(\mathbf{x}_{it})\omega + \phi(\mathbf{x}_{it})\omega_i + \epsilon_{it}, \tag{3.1}$$

where $\phi(\mathbf{x}_{it})$ is the nonlinear feature mapping function $\phi(\cdot) : R \to R^{d_f}$ which maps the input space to the higher dimensional feature space where the dimension d_f is defined in an implicit way, and ω and ω_i are the corresponding weight parameter vectors.

Denote the penalized negative log-likelihood estimator of ω from whole data by $\hat{\omega}^{(A)}$, the linear Bayes estimator of ω_i by $\hat{\omega}_i^{(B)}$ where ω_i is assumed to follow a normal distribution $N(\mathbf{0}, \sigma_B^2 I_{d_f})$. Then the credibility estimator of $\omega_i + \omega$ can be rewritten as

$$\begin{aligned} \hat{\omega}^{(A)} + \hat{\omega}_i^{(B)} &= \hat{\omega}^{(A)} + \sigma_B^2 \phi(\mathbf{X}_i)' V_i^{-1}(\mathbf{y}_i - \phi(\mathbf{X}_i)\hat{\omega}^{(A)}) \\ &= (I - \sigma_B^2 \phi(\mathbf{X}_i)' V_i^{-1} \phi(\mathbf{X}_i))\hat{\omega}^{(A)} + \sigma_B^2 \phi(\mathbf{X}_i)' V_i^{-1} \mathbf{y}_i \\ &= (I - \sigma_B^2 \phi(\mathbf{X}_i) V_i^{-1} \phi(\mathbf{X}_i))\hat{\omega}^{(A)} + \sigma_B^2 \phi(\mathbf{X}_i)' V_i^{-1} \phi(\mathbf{X}_i)\hat{\omega}_i^* \\ &= (I - \xi_i)\hat{\omega}^{(A)} + \xi_i \hat{\omega}_i^*, \end{aligned} \tag{3.2}$$

where $\hat{\omega}_i^* = (\phi(\mathbf{x}_i)' V_i^{-1} \phi(\mathbf{x}_i))^{-1} \phi(\mathbf{x}_i)' V_i^{-1} \mathbf{y}_i$ is the GLS estimator of $\omega_i + \omega$ based only on data of i th subject, $V_i = \sigma_B^2 \phi(\mathbf{X}_i)\phi(\mathbf{X}_i)' + \sigma_\epsilon^2 I_T$, and $\xi_i = \sigma_B^2 \phi(\mathbf{X}_i)' V_i^{-1} \phi(\mathbf{X}_i)$ is the credibility factor.

We know that $\phi(\mathbf{X}_i)\phi(\mathbf{X}_j)' = K(\mathbf{X}_i, \mathbf{X}_j)$ which are obtained from the application of Mercer's conditions (1909). The variance-covariance matrix of $\mathbf{y}_i$ can be written as $V_i = \sigma_B^2 K(\mathbf{X}_i, \mathbf{X}_i) + \sigma_\epsilon^2 I_T$. The penalized negative log-likelihood estimator is

$$\hat{\omega}^{(A)} = \phi(\mathbf{X})'\hat{\alpha} \quad \text{with } \hat{\alpha} = (K(\mathbf{X}, \mathbf{X}) + \tilde{V}/\lambda)^{-1}\mathbf{y}, \tag{3.3}$$

which is obtained by minimizing the penalized negative log-likelihood,

$$L = \frac{1}{2\lambda}\omega'\omega + \frac{1}{2}(\mathbf{y} - \phi(\mathbf{X})\omega)' \tilde{V}^{-1}(\mathbf{y} - \phi(\mathbf{X})\omega) \tag{3.4}$$

where $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^n$, $\tilde{V}$ is a block diagonal matrix of V_i's, and λ is a penalized parameter of the penalized negative log-likelihood. The penalized negative log-likelihood estimator in (3.3) can be seen as the results of the weighted least squares support vector machine (Valyon and Horvath, 2005) with bias included obtained from the equation (3.4).

The credibility estimate of $\mathbf{y}_i$ can be obtained as

$$\begin{aligned} \mathbf{y}_i^{(c)} &= \phi(\mathbf{X}_i)(\hat{\omega}^{(A)} + \hat{\omega}_i^{(B)}) \\ &= \phi(\mathbf{X}_i)\hat{\omega}^{(A)} + \sigma_B^2 \phi(\mathbf{X}_i)\phi(\mathbf{X}_i)' V_i^{-1}(\mathbf{y}_i - \phi(\mathbf{X}_i)\hat{\omega}^{(A)}) \\ &= K(\mathbf{X}_i, \mathbf{X}_i)\hat{\alpha} + \sigma_B^2 K(\mathbf{X}_i, \mathbf{X}_i) V_i^{-1}(\mathbf{y}_i - K(\mathbf{X}_i, \mathbf{X}_i)\hat{\alpha}) \\ &= (I - \sigma_B^2 K(\mathbf{X}_i, \mathbf{X}_i) V_i^{-1}) K(\mathbf{X}_i, \mathbf{X})\hat{\alpha} + \sigma_B^2 K(\mathbf{X}_i, \mathbf{X}_i) V_i^{-1} \mathbf{y}_i. \end{aligned} \tag{3.5}$$

For the i th subject, the predicted claims amount for $t = T + 1$, $\hat{y}_{i,T+1}^{(c)}$, can be obtained as

$$\begin{aligned} &K((1, T+1), \mathbf{X})\hat{\alpha} - \sigma_B^2 K((1, T+1), \mathbf{X}_i) V_i^{-1} K((1, T+1), \mathbf{X})\hat{\alpha} \\ &+\sigma_B^2 K((1, T+1), \mathbf{X}_i) V_i^{-1} \mathbf{y}_i. \end{aligned} \tag{3.6}$$

4. Credibility estimation

In this section we propose the estimation procedure of $(\alpha, \sigma_B^2, \sigma_\epsilon^2)$ with tuning hyperparameters in a one procedure. With these estimates we obtain the credibility estimates and predicted values of claims using (3.5) and (3.6).

Given the values of σ_B^2 and σ_ϵ^2, the functional structures of credibility estimates are characterized by a set of hyperparameters θ (the kernel parameter and the penalized parameter). We consider the cross validation(CV) function (Craven and Wahba, 1979) similar to what follows:

$$CV(\theta) = \frac{1}{nT}\mathbf{y}'(I - S(\theta))'\tilde{V}(\theta)^{-1}(I - S(\theta))\mathbf{y}, \tag{4.1}$$

where $S(\theta) = A(\theta) + B(\theta)$, $S(\theta)\mathbf{y}$ is the estimate of $\mathbf{y}$ from whole data, which is equivalent to $K(\mathbf{X}, \mathbf{X})\hat{\alpha}$ in (3.3), $A(\theta)$ is a $nT \times nT$ block diagonal matrix of $\sigma_B^2 K_\theta(\mathbf{X}_i, \mathbf{X}_i)V(\theta)_i^{-1}$, and $B(\theta)$ is given as follows

$$B(\theta) = \{(\mathrm{I} - \sigma_B^2 K_\theta(\mathbf{X}_i, \mathbf{X}_i)V_i(\theta)^{-1})K_\theta(\mathbf{X}_i, \mathbf{X})\} \times (K_\theta(\mathbf{X}, \mathbf{X}) + \tilde{V}(\theta)/\lambda)^{-1}\}_{i=1}^n.$$

Replacing the k th diagonal element of $S(\theta)$ by their average $tr(S(\theta))/(nT)$, the generalized cross validation(GCV) function can be obtained as

$$GCV(\theta) = \frac{nT\mathbf{y}'(\mathrm{I} - S(\theta))'\tilde{V}(\theta)^{-1}(\mathrm{I} - S(\theta))\mathbf{y}}{(nT - tr(S(\theta)))^2} \tag{4.2}$$

The estimation procedure is based on the model:

$$\mathbf{y} = \phi(\mathbf{X})\omega + \epsilon = \tilde{K}\alpha + \epsilon.$$

Here $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^n$, $\epsilon = \mathbf{y} - \phi(\mathbf{X})\omega$ is assumed to follow $N(\mathbf{0}, \tilde{V})$, and $\tilde{V}$ is a block diagonal matrix of V_i's.

We assume α follows $N(\mathbf{0}, \frac{1}{\lambda}\tilde{K}^{-1})$, where $\tilde{K}$ is a kernel matrix constructed from $\mathbf{X}$. Then the estimates of $(\alpha, \sigma_B^2, \sigma_\epsilon^2)$ can be obtained from

$$\begin{aligned} minL(\alpha, \sigma_D^2, \sigma_e^2) &= -\log P(\mathbf{y}|\alpha, \sigma_D^2, \sigma_\epsilon^2) - \log P(\alpha) \\ &= \frac{1}{2}\log\det(\tilde{V}(\sigma_B^2, \sigma_\epsilon^2)) + \frac{1}{2}(\mathbf{y} - \tilde{K}\alpha)'\tilde{V}(\sigma_B^2, \sigma_\epsilon^2)^{-1}(\mathbf{y} - \tilde{K}\alpha) + \frac{1}{2}\lambda\alpha'\tilde{K}\alpha. \end{aligned}$$

Estimates of $(\alpha, \sigma_B^2, \sigma_\epsilon^2)$ and optimal hyperparameters (λ, σ^2) can be obtained by the following iteration procedure:

[Step 1] With $\sigma_\epsilon^2 = \hat{\sigma}_\epsilon^2$

$\hat{\sigma}_B^2$ is obtained from

$$minL(\sigma_B^2) = \frac{1}{2}\log\det(\tilde{V}(\sigma_B^2, \hat{\sigma}_\epsilon^2)) + \frac{1}{2}(\mathbf{y} - \tilde{K}\hat{\alpha})'\tilde{V}(\sigma_B^2, \hat{\sigma}_\epsilon^2)^{-1}(\mathbf{y} - \tilde{K}\hat{\alpha}),$$

where $\hat{\alpha}$ is obtained from GCV function (4.2) and (3.3).

[Step 2] With $\hat{\sigma}_B^2$

$\hat{\sigma}_\epsilon^2$ is obtained from

$$minL(\sigma_\epsilon^2) = \frac{1}{2}\log\det(\tilde{V}(\hat{\sigma}_B^2, \sigma_\epsilon^2))) + \frac{1}{2}(\mathbf{y} - \tilde{K}\hat{\alpha})'\tilde{V}(\hat{\sigma}_B^2, \sigma_\epsilon^2))^{-1}(\mathbf{y} - \tilde{K}\hat{\alpha}),$$

where $\hat{\alpha}$ is obtained from GCV function (4.2) and (3.3).

[Step 3] Iterate Steps 1 and 2 until convergence.

5. Numerical studies

We illustrate the performance of the nonlinear credibility estimation using kernel machine through the simulated examples on the linear and nonlinear models and data from Hachemeister (1975). For the linear model, we consider the model as follows,

$$y_{it} = (1,t)\beta + (1,t)\mathbf{b}_i + \epsilon_{it} \text{ for } i = 1, \cdots, 4 \text{ and } t = 1, \cdots, 10,$$

where y_{it} is the response variable observed at time t of the i th subject with input vector $\mathbf{x}_{it} = (1,t)$, $\beta = \begin{pmatrix} 15 \\ 2.5 \end{pmatrix}$, $\mathbf{b}_i$'s are generated from $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}\right)$, and ϵ_{it}'s are generated independently from $N(0, 1.5^2)$. We apply the linear model of Hachemeister (1975) using SAS procedure(Proc Mixed) and obtain the estimates as follows,

$$\hat{\beta} = \begin{pmatrix} 12.8333 \\ 3.1121 \end{pmatrix}, \hat{C}OV(\mathbf{b}_i) = \begin{pmatrix} 2.4271 & 0 \\ 0 & 1.1855 \end{pmatrix}, \hat{\sigma}_\epsilon = 1.856,$$
$$\hat{\mathbf{b}} = \begin{pmatrix} 1.6319 & -1.5423 & 0.3682 & -0.4578 \\ -0.183 & 1.3473 & -1.2812 & 0.1169 \end{pmatrix}.$$

With a linear kernel function, we obtain the estimates as follows,

$$\lambda = 100, \hat{\beta} = \begin{pmatrix} 12.7745 \\ 3.1083 \end{pmatrix}, \hat{C}OV(\mathbf{b}_i) = \begin{pmatrix} 0.984761 & 0 \\ 0 & 0.98476 \end{pmatrix}, \hat{\sigma}_e = 1.9066$$
$$\hat{\mathbf{b}} = \begin{pmatrix} 1.1968 & -1.0549 & 0.27854 & -0.29462 \\ -0.10889 & 1.2889 & -1.2551 & 0.10569 \end{pmatrix}.$$

Figure 5.1 shows the true responses(dot and solid line) in each subject, credibility estimates of responses by Hachemeister model(dotted), credibility estimates of responses by proposed procedure(solid). From Figure 5.1 we can see that both estimates provide similar results for linear model case. The root mean squared errors(RMSE) of the proposed procedure and Hachemeister modelare obtained as 1.271 and 1.234, respectively, which lead to hard distinction of both estimates in the figure.

For the nonlinear model, we consider the model as follows,

$$y_{it} = b_{0i} + 5\exp(\sin(1/\pi + 2t/\pi)) + 5\exp(\cos((1,t)\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix})) + \epsilon_{it},$$

where $b_{01} = 10, b_{02} = 15, b_{03} = 5, b_{04} = 20$, $\mathbf{b}_i$'s are generated from $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}\right)$, and ϵ_{it}'s are generated independently from $N(0, 2^2)$ for $t = 1, \cdots, 10, i = 1, \cdots, 4$. We apply

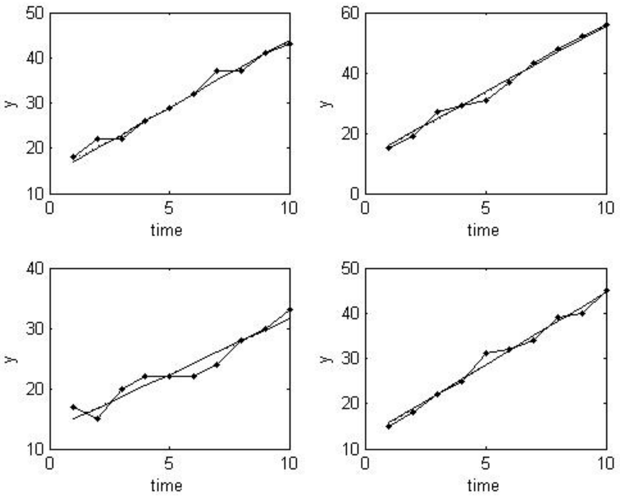

Figure 5.1 Results of the linear model. Observation (dot-solid line), credibility estimates by Hachemeister model(dotted line) and credibility estimates by proposed procedure(solid line).

the linear model of Hachemeister (1975) using SAS procedure(Proc Mixed). The proposed procedure is applied with the Gaussian kernel,

$$K(\mathbf{u}, \mathbf{v}) = \exp(-\frac{1}{\sigma^2}||\mathbf{u} - \mathbf{v}||^2),$$

and we obtain the following results: $\lambda = 1600, \sigma^2 = 70$, $\hat{\sigma}_B^2 = 17.305$, $\hat{\sigma}_\epsilon = 19.44$. Figure 5.2 shows the true responses(dot and solid line) in each subject, credibility estimates of responses by Hachemeister model(dotted), credibility estimates of responses by proposed procedure(solid). From Figure 5.2 we can see that the proposed procedure is suitable for the nonlinear model.

We now consider analyzing data from Hachemeister (1975) using the proposed model. Hachemeister (1975) analyzed the claim data of private passengers auto insurance of bodily injury coverage by applying the linear trend model,

$$y_{it} = (1, t)\beta + (1, t)\mathbf{b}_i + \epsilon_{it} \text{ for } i = 1, \cdots, 5 \text{ states in US and } t = 1, \cdots, 12\text{quarters}.$$

Using SAS procedure(proc Mixed) we obtained the estimates as follows,

$$\hat{\beta} = \begin{pmatrix} 1460.32 \\ 32.4147 \end{pmatrix}, \hat{C}OV \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} = \begin{pmatrix} 2.4271 & 0 \\ 0 & 1.1855 \end{pmatrix}, \hat{\sigma}_e = 1.856,$$
$$\hat{\mathbf{b}} = \begin{pmatrix} 1685.4 & 1380.9 & 1545.3 & 1220.5 & 1469.5 \\ 55.876 & 20.765 & 41.653 & 24.023 & 19.756 \end{pmatrix}.$$

The proposed procedure is applied with the Gaussian kernel, and we obtain the following results: $\lambda = 450000, \sigma^2 = 780$, $\hat{\sigma}_B^2 = 59350$, $\hat{\sigma}_\epsilon = 31149$. Figure 5.3 shows the observed

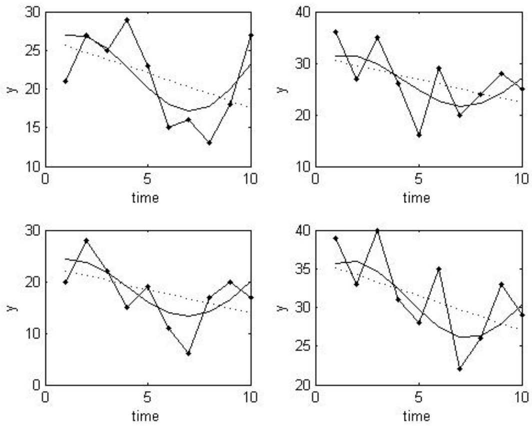

Figure 5.2 Results of the nonlinear model. Observation (dot-solid line), credibility estimates by Hachemeister model(dotted line) and credibility estimates by proposed procedure(solid line).

claims(dot and solid line) in each state, credibility estimates of responses by Hachemeister model(dotted), nonlinear credibility estimates of claims by proposed procedure(solid). RMSEs of the proposed procedure and Hachemeister modelare obtained as 166.16 and 161.954, respectively. From Figure 5.3 we can see the both procedures can be applied in this example.

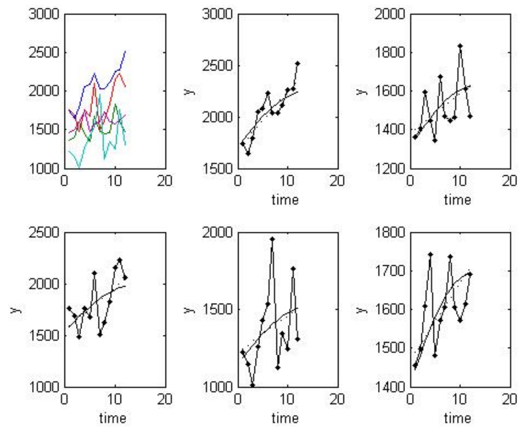

Figure 5.3 Results of data from Hachemeister(1975). Data observed in each state(upper-left). Observation (dot-solid line), credibility estimates by Hachemeister model(dotted line) and credibility estimates by proposed procedure(solid line).

6. Conclusions

In this paper, we dealt with a nonlinear "kernelized" variant of credibility model by reforming the linear mixed effects model through kernel machine. The proposed procedure can be applied for both the linear and the nonlinear models. Through the examples we showed that the proposed procedure derives the satisfying results.

References

Antonio, K. and Beirlant, J. (2008). Issues in claims reserving and credibility: a semiparametric approach with mixed models. *Journal of Risk & Insurance*, **75**, 643-676.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics*, **31**, 377-403.

Dannenburg, D. R., Kaas, R. and Goovaerts, M. J. (1996). *Practical actuarial credibility models*, Institute of Actuarial Science and Econometrics, University of Amsterdam, Amsterdam, The Netherlands.

Frees, E. W., Young, V. R. and Luo, Y. (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, **14**, 229-247.

Goovaerts, M. J. and Hoogstad, W. (1987). Credibility theory, surveys of actuarial studies. *Journal of the American Statistical Associations*, **57**, 369-375.

Hachemeister, C. A. (1975). Credibility for regression models with applications to trend. In *Credibility: Theory and applications*, Ed. P. M. Kahn, 129-163, Academic Press, New York.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal data analysis*, John Wiley & Sons, New York.

Mercer, J. (1909). Functions of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society*, A. **209**, 415-446.

Valyon, J. and Horvath, G. (2005). A robust LS-SVM regression. *Proceedings of World Academy of Science, Engineering and Technology*, **7**, 148-153, 2005.

Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.