Prediction of Hospital Bankruptcy Using a Hierarchical Generalized Linear Model

Maengseok Noh
Korea Health Industry Development Institute, Department of Health Industry Informatics
57-1 Nolyangjin-Dong Dongjac-Gu, Seoul, Korea
e-mail : msnoh@khidi.or.kr

Hyejung Chang
Korea Health Industry Development Institute, Department of Health Industry Informatics
57-1 Nolyangjin-Dong Dongjac-Gu, Seoul, Korea
e-mail : hjchang@khidi.or.kr

Youngjo Lee
Seoul National University, Department of Statistics
San 56-1, Shilim-dong, Kwanak-Gu, Seoul, Korea
e-mail : youngjo@plaza.snu.ac.kr

1. INTRODUCTION

Recently, the hospital bankruptcy rate is increasing in Korea. An important issue is to predict bankruptcy using existing hospital management information. The hospital bankruptcy is often measured in year intervals, called grouped duration data, not by the continuous time elapsed to the bankruptcy (Ryu, 1994). This study introduces a hierarchical generalized linear model (HGLM) for analyzing of hospital bankruptcy data. The hazard function for each hospital may be influenced by unobservable latent variables. These unobservables, termed as random effects or frailties, could describe correlations among repeated measures within the hospital and heterogeneities among hospitals.

2. MODEL

Let T_i be a continuous duration variable representing the time until hospital i goes bankrupt. The duration variable was measured in r intervals, $I_1 = [a_0 = 0, a_1)$, $I_2 = [a_1, a_2)$, …, $I_r = [a_{r-1}, a_r)$, and $I_{r+1} = [a_r, a_{r+1} = \infty)$. Let d_{ij} be 1 if hospital i survives in I_j conditional on $T_i \geq a_{j-1}$, and d_{ij} be 0 if hospital i goes bankrupt. If the observation is censored from I_{j+1}, the corresponding d_{ik} for $k \geq j+1$ simply does not appear in the data. Then the censoring scheme is automatically taken into account. For instance, if T_i is censored from I_{j+1}, then $d_{i1}=d_{i2}=\ldots=d_{ij} = 1$.

When $\boldsymbol{l}(t|x, u)$ is the hazard function at time t for a hospital with a latent variable u and covariates x, and $\boldsymbol{l}_0(t)$ is the baseline hazard function, the proportional hazard model (PHM) with an unobserved latent variable is expressed as (1).

$$\boldsymbol{l}(t|x,u)= \boldsymbol{l}_0(t)\exp(x^t\boldsymbol{b})u \qquad (1)$$

Considering the model (1), the sequential binary random variables d_{ij} can be modeled by a HGLM as

i) the conditional distribution of d_{ij} given u_i is Bernoulli with

$\alpha_{ij} = P(d_{ij} = 1 \mid u_i) = P(T_i >= a_j \mid T_i >= a_{j-1}, u_i)$, and

ii) the distribution of random variable u_i is arbitrary.

For the distribution of u_i, we consider gamma and log-normal distributions. Because $\alpha_{ij} = \exp(-\exp(x_{ij}^t\beta + \gamma_j)\, u_i)$ where γ_j is the log integral of $\lambda_0(t)$ from a_{j-1} to a_j, the complementary log-log link leads to $\log(-\log(\alpha_{ij})) = x_{ij}^t\beta + \gamma_j + \log(u_i)$. Suppose that hospital i goes bankrupt in the r_i interval or is censored from interval (r_i +1). Then the hierarchical likelihood (h-likelihood) for n hospitals is defined as (2) where σ_u^2 is the dispersion parameter of $v_i = \log(u_i)$.

$$h = \sum_{i=1}^{n} \sum_{j=1}^{r_i} \{ d_{ij} \log \frac{\alpha_{ij}}{1-\alpha_{ij}} + \log(1-\alpha_{ij}) \} + \sum_{i=1}^{n} l(\sigma_u^2; v_i) \qquad (2)$$

3. RESULTS

For estimating fixed and random effects we maximize h-likelihood estimators suggested by Lee and Nelder (1996). Dispersion components are estimated by using adjusted profile h-likelihood (APHL). The APHL is a first-order Laplace approximation to the marginal-likelihood (m-likelihood). Lee and Nelder (2001) developed the second-order Laplace approximation, which gives good properties for non-normal random effects. Because the gamma distribution for the random effect u gives an explicit form for the m-likelihood, we compare the first- and second-order Laplace estimators with the marginal-likelihood estimator by a simulation study.

4. EXAMPLE

Data from twenty bankrupt and sixty profitable hospitals were collected for the five years, and were fitted using various methods based upon h-likelihood and m-likelihood. The covariates are financial indicators observed one year before bankruptcy. The results were compared with those from a logit model (Lynch & Ozcan, 1994). While we can evaluate effects of covariates on the bankruptcy status only at a specific year with the logit analysis, the HGLM analysis allows us to evaluate them over consecutive years. Therefore, the new important covariates influencing hospital bankruptcy have been identified through the HGLM analysis.

REFERENCES

Lee, Y. and Nelder, J.A. (1996). Hierarchical Generalized Linear Models (with discussion). *Journal of the Royal Statistical Society B*, **58**, 619-678.

Lee, Y. and Nelder, J.A. (2001). Hierarchical Generalised Linear Models: a Synthesis of Generalised Linear models, Random-effect Models and Structured Dispersions. to appear in *Biometrika*.

Lynch, J.R, and Ozcan, Y.A. (1994), Hospital closure: an efficiency analysis. *Hospital Health Service Administration*, **39**, 205-220.

Ryu, K. (1994). Group Duration Analysis of the Proportional Hazard Model: Minimum Chi-square Estimators and Specification Tests. *Journal of the American Association*, **89**, 1386-1397.