International Journal of Mathematical Analysis and Applications
2014; 1(4): 59-62
Published online September 30, 2014 (http://www.aascit.org/journal/ijmaa)

Keywords
Ridge Regression,
Generalized Linear Models,
Age-Period-Cohort Models

Received: August 29, 2014
Revised: September 26, 2014
Accepted: September 27, 2014

Ridge regression method for fitting mortality models

Oyebayo Olaniran, Moyosola Bamidele

Department of Statistics, University of Ilorin, PMB 1515, Ilorin, Nigeria

Email address
rid4stat@yahoo.com (O. Olaniran), bamidelemoyo@gmail.com (M. Bamidele)

Citation
Oyebayo Olaniran, Moyosola Bamidele. Ridge Regression Method for Fitting Mortality Models. *International Journal of Mathematical Analysis and Applications.* Vol. 1, No. 4, 2014, pp. 59-62.

Abstract
Some mortality models can be expressed in the form of generalized linear model framework (GLMs). The modelling approach of the GLMs centered on the assumptions of no correlation between the explanatory variables which may be age, cohort, year as the case may be for most of the mortality models. Many mortality models often had inherent trivial correlation between any of the earlier listed variables thus multicollinearity set in. In this paper, a new fitting methodology using Ridge Regression (RR) under the class of shrinkage methods is proposed to tackle the trivial correlation problem, which might be inherent in mortality models via model specification. By way of example, the Age-Period-Cohort model, which exhibits the trivial correlation problem, was used to demonstrate the fitting procedure using Monte-carlo simulation and France mortality real life data set. The results from the monte-carlo simulation and real life data set established the supremacy of the fitting methodology over the existing fitting procedure via mean square error of prediction, log-likelihood and Bayesian Information Criterion (BIC).

1. Introduction

Modelling mortality data has been the major focus of many authors in actuarial science. The major concern of most authors within the context of mortality is to establish a model that will adequately fits the historical data. For example, authors like Currie et al (2004), Currie (2006), Cairns et al (2006a), Renshaw and Haberman (2006) and Currie (2013) among others have focused on establishment of a model and fitting methodology that involves minimization of the deviance function. The current argument is that the existing fitting methodology does not consider the inherent correlation structured in the data and those introduced via model specification. In this paper we propose a new fitting methodology using Ridge Regression approach originally proposed by Hoerl and Kennard (1970). The proposed method is in turn applied to the Age-Period-Cohort (APC) model proposed by Currie (2006).

The paper is structured as follows; in section 2 existing literature is reviewed focusing on APC model and the existing fitting methodology as explained by Currie (2013), section 3 recalls the construction of Poisson Ridge regression Estimator introduced by Mansson and Shukur (2011). Section 4 & 5 provides the model assessment criteria and the simulation scheme respectively. Section 6 gives the results from the simulation study and the real life dataset and conclusions are given in section 7.

2. Literature Review

As noted by most authors, the following literature guides on stochastic models that are often used for modelling mortality force $\mu_{x,t}$ or mortality risks $q_{x,t}$; Lee and Carter

(1992), Renshaw and Haberman (2006), Cairns et al (2006a), Currie et al (2004) and Currie (2006).

2.1. Stochastic Mortality Models

Stochastic mortality models are often applied to mortality force $\mu_{x,t}$ or mortality risk $q_{x,t}$. The mortality rate or mortality force is often defines as;

$$\mu_{x,t}=D_{x,t}/E_{x,t} \quad (1)$$

Where $D_{x,t}$ is the number of death at age x during calendar year t and $E_{x,t}$ is the mid-year population of age x during calendar year t.

The initial mortality rate $q_{x,t}$ is the probability that a person aged x dies within the next year. The two mortality measures are linked by the following approximation:

$$q_{x,t} = 1 - \exp(-\mu_{x,t}) \quad (2)$$

2.2. Age-Period-Cohort Model (APC)

Currie (2006) introduces the Age-Period-Cohort model (APC)

$$\log(\mu_{x,t}) = \alpha_x + \kappa_t + \gamma_{t-x} \quad (3)$$

Currie (2006) uses P-splines to estimate α_x, κ_t *and* γ_{t-x} to ensure smoothness. For the purpose of this work the latest fitting methodology proposed by Currie (2013) which uses Poisson generalized linear models (GLMs) framework will be adapted to ensure a realistic comparison between the propose method and the existing one.

2.2.1. Fitting Age-Period-Cohort Models Using Poisson GLMs

Fitting mortality models using Poisson GLM follows the same procedure as in the usual Poisson GLMs explained by McCullagh and Nelder (1989), the additional task involved is the addition of the constraints to make the models identifiable. The main task is to specify the design matrix *X* which is the matrix of the predictors. Here *X* can be a combination of age effect, period effect and cohort effect as the case may be in any of the mortality model.

Let $d = vec(d_{x,t})$ *and* $e = vec(e_{x,t})$ be the vectors of observed deaths and central exposures; here, the vec operator stacks the columns of a matrix in column order on top of each other. It is pertinent to note that with this definition the age suffix varies faster than the year suffix in *d and e*. With the above definition, it's easier to describe the Poisson GLM fitting methodology for the APC model. The fitting procedure as explained by Currie (2013) is described below:

The APC model defined in equation (3) above, is specified by its model matrix, $X = [Xa : Xy : Xc]$ where Xa, $n \times n_a$, defines the age effects, Xy, $n \times n_y$, the period effects, and Xc, $n \times n_c$, the cohort effects; thus X is $n \times (n_a + n_y + n_c)$. Recalling that the age suffix in the vector of deaths varies faster than the year suffix, we see that Xa consists of n_y copies of In_a stacked on top of each other where *Is* denotes the identity matrix of size s. The Kronecker product is a good way of writing matrices with a row-column structure and here $Xa = 1n_y \otimes In_a$; Searle (1982) provides a good discussion of Kronecker products. In a similar way we have $Xy = In_y \otimes 1n_a$. There is no simple way of writing Xc but it is easily obtained by noting that row i of Xc consists of 0's except for a single 1 which occurs in column c if the data point corresponding to row i belongs to cohort c, $c = 1, \ldots, nc$. The model matrix for the APC model can be written as:

$$X = [1n_y \otimes In_a : In_y \otimes 1n_a : Xc]$$

It is now a simple matter to check that the rank of $\boldsymbol{X}$ is less than its number of columns. Thus, the model is not identifiable.

From the APC model structure defined earlier, we can denote the coefficients of *Xa*, *Xy* and *Xc* by $\boldsymbol{\alpha}$, $\boldsymbol{\kappa}$, $\boldsymbol{\gamma}$. If we let $\boldsymbol{\theta}=(\alpha', \boldsymbol{\kappa}', \boldsymbol{\gamma}')'$, to solve the maximum likelihood equations we need three constraints on $\boldsymbol{\theta}$. There is no unique way of choosing these constraints but whatever choice we make, we always obtain the same table of fitted forces of mortality rate $\mu_{x,t}$. One possible set of constraints is;

$$\sum_{t=1}^{n_y} \kappa_t = 0 \quad (4)$$

$$\sum_{c=1}^{n_c} \gamma_c = 0 \quad (5)$$

$$\sum_{c=1}^{n_c} c\gamma_c = 0 \quad (6)$$

where c runs from 1 (youngest cohort) to n_c (oldest cohort).

We express these constraints in the form of $H\boldsymbol{\theta}=0$ where H is defined as follows:

$$h_1=(0'n_a, 1'n_y, 0'n_c) \quad (7)$$

$$h_2=(0'n_a+n_y, 1'n_c) \quad (8)$$

$$h_3=(0'n_a+n_y, n'_c) \quad (9)$$

$$H=(h_1, h_2, h_3) \quad (10)$$

Here 0s is a vector of 0's of the indicated length and $n_c=(1, \ldots, n_c)'$. We refer to H, $3\times(n_a+n_y+n_c)$ as the constraints matrix. The matrix

$$X_{aug} = \begin{pmatrix} X \\ H \end{pmatrix} \quad (11)$$

is known as the augmented matrix, the maximum likelihood equations have a unique solution, $\boldsymbol{\theta}$, subject to the constraint $H\boldsymbol{\theta}=0$, if X_{aug} is of full column rank. The glm function in R uses a different method of obtaining a solution. In effects, R sets $\kappa_1 = \gamma_c = \gamma_{nc} = 0$. These can also be similarly illustrated as a constraints matrix HR defines as follows:

$$h^R_1=(0'n_a, 1', 0'n_y+n_c\text{-}1) \quad (12)$$

$$h^R_2=(0'n_a+n_y, 1,1'n_c\text{-}1) \quad (13)$$

$$h^R_3=(0'n_a+n_y+n_c,1) \quad (14)$$

$$H_R=(h_1, h_2, h_3) \quad (15)$$

Suppose and $\boldsymbol{\theta}_R$, are the MLE of θ under the preferred constraints H and R constraint HR, $\boldsymbol{\theta}$ can be obtained from θ_R.

We know that $X\boldsymbol{\theta}=X\boldsymbol{\theta}_R$ also $H\boldsymbol{\theta}=H\boldsymbol{\theta}_R=0$. Hence

$$Xaug\boldsymbol{\theta}=XH\boldsymbol{\theta}=X\boldsymbol{\theta}H\boldsymbol{\theta}=X\boldsymbol{\theta}_RH\boldsymbol{\theta}_R=XH\boldsymbol{\theta}_R \quad (16)$$

Now $Xaug$ has full column rank $X'augXaug=X'X+H'H$ is positive definite and hence non-singular. Thus multiplying both sides of (16) by $X'aug=[X' :H']$ gives

$$\widehat{\boldsymbol{\theta}} = (X'X + H'H)^{-1}(X'X + H'H_R)\widehat{\boldsymbol{\theta}_R} \quad (17)$$

Using the above formulation we can define the Maximum Likelihood Estimator $\boldsymbol{\theta}_{ML}$ since the result from R statistical package software assumes MLE framework. Thus

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{ML}} = (X'X + H'H)^{-1}(X'X + H'H_R)\widehat{\boldsymbol{\theta}_R} \quad (18)$$

3. Poisson Ridge Regression Estimator

As an alternative to the maximum likelihood estimator, to remedy the problem caused by multicollinearity Mansson and Shukur (2011) proposed the Poisson Ridge Regression method (PRR) applied to count data. The PRR estimator using the mortality parameterization earlier defined is;

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{PR}} = (X'WX + kI)^{-1}(X'WX\widehat{\boldsymbol{\theta}}_{\boldsymbol{ML}}) \quad (19)$$

Where W is the diag($\mu_{x,ti}$) and k is a small non-negative constant value i.e. k>0 and I is the identity matrix.

There is need to establish the linkage between (18) and (17) since (18) does not considered the constraints impose on each parameter of the model.

Note that if k = 0 the ridge estimator becomes the Maximum Likelihood estimate. The values of k will be selected by the analyst. The corresponding values of the ridge parameter produces different regression coefficient. As the value of k increases from zero the smaller the variance but the greater the bias introduced. It is always difficult to select the optimal value of k that produces the stable regression coefficients. Some various methods are used in selecting the appropriate k.

3.2. Methods for Selecting the Value of Parameter k in Ridge Regression Modelling

The main step in the ridge regression analysis is to select a value of k and to obtain the corresponding estimates of the regression coefficients. There are several methods of selecting the appropriate value of k. One of the procedures and the one used can be summarized as follows:

i. Ridge trace: Hoerl and Kennard (1970) suggested a graphical method called ridge trace to select the value of the ridge parameter k. This is a plot of the values of the regression coefficient plotted against the range of the values of k.

ii. Cross validation: the cross validation method does not make use of a particular statistic, like all of the above; rather it plot several models, using ridge parameter between 0 and 1. For this work, an interval of 0.0001 between successive ridge parameter was used. The one, which gives the minimum Root Mean Square Error of Prediction RMSEP, is selected as the best model.

Having explained the Poisson Ridge Regression estimator, the next thing is to establish the linkage between the mortality model and the PRR estimator. Thus, the PRR model for the Age-Period-Cohort model is;

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{PR}} = ((X'X + H'H) + kI)^{-1}(X'X + H'H_R)\widehat{\boldsymbol{\theta}_R} \quad (21)$$

The above estimator in (21) tends to improve the solution of (18) by adding a small value k.

4. Criteria for Assessing the Performances of the Estimators

The assessments of the estimators considered in this paper were based on the following criteria.

i. Root mean square error Prediction
ii. Log-Likelihood
iii. Bayesian Information criterion (BIC)

4.1. Root Mean Squared Error

The root mean squared error, denoted by RMSEP, and of an estimator is the square root of the mean squared error of prediction for the estimators.

i.e. $RMSEP = \sqrt{E\left(\widehat{D}_{x,t} - D_{x,t}\right)^2}$

ii. Log-likelihood

$$Log\left(L(\theta)\right) = \sum_{xt}\{D_{xt}\log[e_{xt}\hat{\theta}] - e_{xt}\hat{\theta} - \log(D_{xt}!)\}$$

iii. Bayesian Information Criterion (BIC)

$$BIC = Log\left(L(\hat{\theta})\right) - 1/2\, Klog(P)$$

For the likelihood based criteria explained above, model with higher criterion estimate is better while for the MSEP and RMSEP criteria smaller estimate is better.

5. Simulation Study

Assuming a starting population size N which is regarded as initial exposure to risk size, let q_{xti} denote the mortality rate for a fixed number of years n_y and varying ages $i=1, 2, \ldots, n_a$. Also, it's very easy to define $1\text{-}q_{xti}$ as the probability of surviving from age x to $x+i$. Assuming baseline survival probability of 0.5, q_{xti} is generated as follows:

1. $1 - q_{xt}{}^{i} = 0.5, \ldots, 0.001$

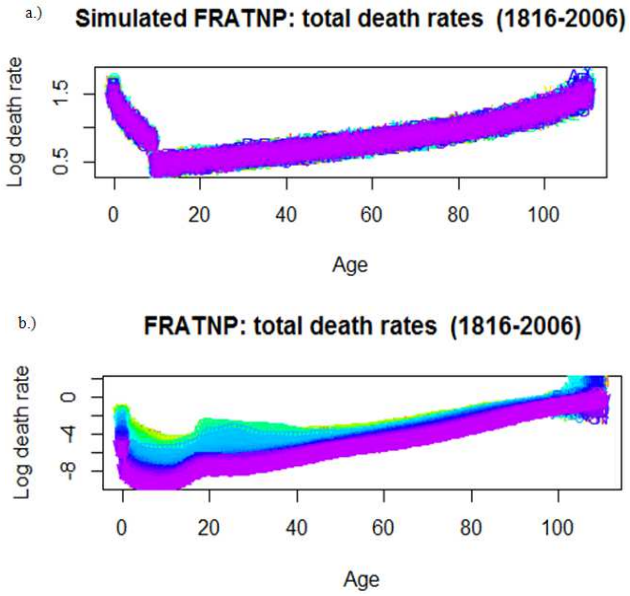
2. $y_i \sim bernoulli(1 - q_{xt}{}^i)$
3. $\widehat{q_{xt}}^{\iota}$=1-$\bar{y}$

$\mu_{x,t}$ follows from the link $q_{x,t} \approx 1 - exp(-\mu_{x,t})$, hence
$\mu_{x,t} = -\log(1 - q_{x,t})$
$d_{xt} = N * q_{x,t}$; $e_{xt} = N - d_{xt}$

The sample matrices generated were replicated 1000 times to ensure stability.

6. Analysis and Result

In this section, we present the result from the monte-carlo simulation and the real life data set; French mortality rates and populations (1816-2006) for ages 0-110. The dataset was extracted from R package demography within R statistical software environment (www.cran.org). The data consists of numbers of deaths Dxt and the corresponding exposures Ext. The data simulation and analysis were implemented using the same software.



a) Simulated French Mortality rates, b.) Actual French Mortality rates.

Figure 1. *Illustration of log of death rates.*

6.2. Results of Estimator Assessments Criteria

Table 6.2. *Model assessment criteria results for the Simulated Data*

Estimator	MSEP	RMSEP	Log-Likelihood	BIC
APC-RIDGE	1,590,857	1,261.292	-1,987,801	-1,990,804
APC-ML	4,690,508	2,165.758	-7,148,649	-7,151,653

Table 6.3. *Model assessment criteria results for the France Mortality Data*

Estimator	MSEP	RMSEP	Log-Likelihood	BIC
APC-RIDGE	3,632,685	1,905.960	-4,650,784	-4,653,787
APC-ML	5,951,300	2,439.529	-8,735,145	-8,738,148

7. Discussion of Results and Conclusion

Figure 1 shows the simulated and the actual French log of death rates, the simulated data tends to approximate the actual log of death rates, which indicates that the simulation strategy is reasonable. The result from the estimator assessments criteria in table 6.2 & 6.3 reveals that APC model estimated using ridge approach is better in terms of prediction and fitness than the one estimated using maximum likelihood approach. As a concluding remark, we therefore recommend the proposed method since it provides a better fit to the data than the existing method. Furthermore, as noted by Currie (2013) that it is better to model mortality rate using binomial link than modellling mortality force using Poisson link. Therefore, we intend to apply the Ridge logistic regression approach to model mortality rate in our future research.

References

[1] Currie, I.D., Durban, M. and Eilers, P.H.C. (2004) Smoothing and forecasting mortality rates. *Statistical Modelling*, 4, 279-298.

[2] Currie, I.D., Durban, M. and Eilers, P.H.C. (2006) Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society*, Series B, 68, 259-280.

[3] Cairns, A.J.G., Blake, D., and Dowd, K. (2006a) "Pricing death: Frameworks for the valuation and securitization of mortality risk", *ASTIN Bulletin*, 36: 79-120.

[4] Currie, I. D. (2013). Fitting models of mortality with generalized linear and non-linear models. Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK.

[5] Hoerl, A. E. and Kennard R. W. (1970): Ridge regression biased estimation for non-orthogonal problems: *Technometrics* 12 (1): 55-67.

[6] Lee, R.D., and Carter, L.R. (1992) "Modeling and forecasting U.S. mortality", *Journal of the American Statistical Association*, 87: 659-675.

[7] Mansson K. and Shukur B.M.G. (2011). A Poisson Ridge Regression Estimator. *Communication in Statistics Theory and Methods*.

[8] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. London: Chapman and Hall.

[9] R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (http://www.R-project.org).

[10] Renshaw, A.E., and Haberman, S. (2006) "A cohort-based extension to the Lee-Carter model for mortality reduction factors", *Insurance: Mathematics and Economics*, 38: 556-570.

[11] Searle, S.R. (1982). *Matrix Algebra Useful for Statistics*. New York: Wiley.