Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2014, Article ID 645719, 5 pages
http://dx.doi.org/10.1155/2014/645719

Research Article

Parametric Regression Models Using Reversed Hazard Rates

Asokan Mulayath Variyath[1] and P. G. Sankaran[2]

[1] *Department of Mathematics and Statistics, Memorial University of Newfoundland, St. John's, NL, Canada A1C 5S7*
[2] *Department of Statistics, Cochin University of Science and Technology, Cochin, Kerala 682022, India*

Correspondence should be addressed to P. G. Sankaran; sankaranpg@yahoo.com

Received 19 June 2013; Accepted 5 October 2013; Published 6 January 2014

Academic Editor: Aera Thavaneswaran

Copyright © 2014 A. Mulayath Variyath and P. G. Sankaran. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Proportional hazard regression models are widely used in survival analysis to understand and exploit the relationship between survival time and covariates. For left censored survival times, reversed hazard rate functions are more appropriate. In this paper, we develop a parametric proportional hazard rates model using an inverted Weibull distribution. The estimation and construction of confidence intervals for the parameters are discussed. We assess the performance of the proposed procedure based on a large number of Monte Carlo simulations. We illustrate the proposed method using a real case example.

1. Introduction

In survival studies, covariates or explanatory variables are usually employed to represent heterogeneity in a population. The main objective in such situations is to understand and exploit the relationship between lifetime and covariates. Regres sion models are useful in such contexts to assess the effect of covariates on lifetime. These models can be formulated in many ways and several types are in common use. Parametric regression models for lifetime involve specification for the distribution of a lifetime T given a vector of covariates X. The most commonly used parametric model is the Weibull regression model, which satisfies the proportional relationship between hazard rate functions of the lifetimes of two subjects. The maximum likelihood technique is usually employed to find estimates of the parameters of the model. For more properties and applications of parametric regression models, one should refer to Lawless [1].

In survival studies, there are many occasions where lifetime data are left censored. For example, baboons in the Amboseli Reserve, Kenya, sleep in the trees and descend for ageing at certain times of the day. Observers often arrive later in the day after this descent has occurred and on such days they can only ascertain that the descent took place before a particular time, so that the descent times are left censored (see [2]). On such occasions, a reversed hazard rate is more appropriate than a hazard rate to analyze lifetime data due to the fact that estimators of hazard rates are unstable when data are left censored. The reversed hazard rate of T is defined as

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t - \Delta t \leq T \mid T \leq t)}{\Delta t}. \quad (1)$$

Introduced by Barlow et al. [3], the function $\lambda(t)$ has been used in various contexts such as the estimation of distribution function under left censoring [1], defining a new stochastic order [4], characterization of lifetime distributions [5–7], studying ageing behavior [8, 9], evolving new repair and maintenance strategies [10, 11], the mixed proportional hazards model [12], and stress hybrid hazards model [13].

Recently, Sengupta and Nanda [14] introduced the proportional reversed hazards model in a semiparametric setup. In the present work, we introduce a fully parametric regression model that satisfies the proportional reversed hazards property. The inverted Weibull distribution is employed as a lifetime model, which can be extended to any parametric model. A large number of simulation studies indicate that the proposed approach is performing well.

The rest of the paper is organized as follows. In Section 2, we introduce a parametric regression model using an inverted Weibull distribution. The proposed model has the property that the reversed hazard rate for the lifetime of pair of subjects is proportional. The estimation of the parameters of

the model is discussed in Section 3. Simulation studies are conducted in Section 4 to assess the finite sample behavior of the estimators. The proposed model is applied to real life data in Section 5 to illustrate its utility. Finally Section 6 provides the major conclusions of the study.

2. Statistical Model

Let T be a nonnegative random variable representing the lifetime of a subject with the distribution function $F(t)$. Assume that the probability density function of T, $f(t)$, exists. The reversed hazard rate of T given in (1) can be written as

$$\lambda(t) = \frac{f(t)}{F(t)}. \tag{2}$$

Let X be $p \times 1$ vector of auxiliary information which may be time dependent. The proportional reversed hazard (PRH) model is defined by

$$\lambda(t \mid X) = \lambda_0(t)\, g(\beta; X), \tag{3}$$

where $\lambda_0(t)$ is the baseline reversed hazard rate, $g(\cdot)$ is a nonnegative function of X and β, $p \times 1$ vector of regression parameters, and $\lambda(t \mid x)$ is the reversed hazard rate of T given the covariate X. The PRH model can be expressed in terms of the distribution function as

$$F(t \mid X) = \left(F_0(t)\right)^{g(\beta;X)}, \tag{4}$$

where $F(t \mid X)$ is the distribution function of T given X and $F_0(t)$ is the baseline distribution function in the absence of covariates. It should be noted that, for two subjects, the ratio of reversed hazard rates is independent of the time t. Semiparametric analysis of the model (2) is recently discussed in Sengupta and Nanda [14]. Our objective here is to carry out the parametric analysis of an inverted Weibull distribution under left censoring. When the lifetime random variable follows the inverted Weibull distribution, the baseline distribution function is given by

$$F_0(t) = e^{-\gamma/t^{\alpha}}, \quad t > 0;\ \alpha, \gamma > 0. \tag{5}$$

The baseline reversed hazard rate of T is then obtained as

$$\lambda_0(t) = \frac{\gamma\alpha}{t^{\alpha+1}}. \tag{6}$$

Note that the baseline reversed hazard rate is decreasing as t increases. In the presence of the covariate X, we have

$$\lambda(t \mid X) = \frac{\gamma\alpha}{t^{\alpha+1}} g(\beta; X). \tag{7}$$

We assume that $g(\beta; X) = e^{\beta' X}$ so that

$$F(t \mid X) = e^{(-\gamma/t^{\alpha})e^{\beta' x}}, \tag{8}$$

with

$$f(t \mid X) = \frac{\gamma\alpha e^{\beta' x}}{t^{\alpha+1}} e^{-(\gamma/t^{\alpha})e^{\beta' x}}. \tag{9}$$

Suppose that the lifetime random variable T is randomly left censored by Z. In practice, one could observe the vectors (Y, δ, X), where $Y = \max(T, Z)$ and $\delta = I(T = Y)$ with $I(\cdot)$ being the indicator function. Let (y_i, δ_i, x_i), $i = 1, 2, \ldots, n$, be i.i.d. copies of (Y, δ, X). Then the likelihood function can be written as

$$L(\beta; y) = \prod_{i=1}^{n} f(y_i \mid x_i)^{\delta_i} F(y_i \mid x_i)^{1-\delta_i}. \tag{10}$$

Under the inverted Weibull distribution assumption, the likelihood function given in (10) is obtained as

$$L(\beta, \alpha, \gamma; y) \propto \prod_{i=1}^{n} \left(\frac{\gamma\alpha e^{\beta' x_i}}{y_i^{\alpha+1}} e^{-(\gamma/y_i^{\alpha})e^{\beta' x_i}} \right)^{\delta_i} \left(e^{(-\gamma/y_i^{\alpha})e^{\beta' x_i}} \right)^{1-\delta_i}, \tag{11}$$

so that the log likelihood function is

$$\begin{aligned} l(\beta, \alpha, \gamma; y) = {} & \log c + \sum_{i=1}^{n} \delta_i \beta' x_i \\ & + \sum_{i=1}^{n} \delta_i (\log\gamma + \log\alpha) - (\alpha + 1) \sum_{i=1}^{n} \delta_i \log y_i \\ & - \gamma \sum_{i=1}^{n} \frac{e^{\beta' x_i}}{y_i^{\alpha}}, \end{aligned} \tag{12}$$

where c is a real constant independent of β, γ, and α. We maximize (12) to estimate the parameters β, α, and γ by equating the partial derivatives with respect to each parameter to zero as

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_{i=1}^{n} \delta_i x_i - \gamma \sum_{i=1}^{n} \frac{e^{\beta' x_i} x_i}{y_i^{\alpha}} = 0, \\ \frac{\partial l}{\partial \alpha} &= \sum_{i=1}^{n} \frac{\delta_i}{\alpha} - \sum_{i=1}^{n} \delta_i \log y_i + \gamma \sum_{i=1}^{n} \frac{e^{\beta' x_i} \log y_i}{y_i^{\alpha}} = 0, \\ \frac{\partial l}{\partial \gamma} &= \sum_{i=1}^{n} \frac{\delta_i}{\gamma} - \sum_{i=1}^{n} \frac{e^{\beta' x_i}}{y_i^{\alpha}} = 0. \end{aligned} \tag{13}$$

Since there is no closed form solution available for (13), we use numerical methods to estimate the parameters. The observed information matrix is given by

$$I(\beta, \alpha, \gamma) = \begin{bmatrix} \dfrac{-\partial^2 l}{\partial\beta\partial\beta} & \dfrac{-\partial^2 l}{\partial\beta\partial\alpha} & \dfrac{-\partial^2 l}{\partial\beta\partial\gamma} \\ \dfrac{-\partial^2 l}{\partial\alpha\partial\beta} & \dfrac{-\partial^2 l}{\partial\alpha^2} & \dfrac{-\partial^2 l}{\partial\alpha\partial\gamma} \\ \dfrac{-\partial^2 l}{\partial\gamma\partial\beta} & \dfrac{-\partial^2 l}{\partial\gamma\partial\alpha} & \dfrac{-\partial^2 l}{\partial\gamma^2} \end{bmatrix}. \tag{14}$$

Note that the matrix (14) is of order $(p + 2) \times (p + 2)$. Under the standard regularity conditions, the vector of estimates

$(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma})$ is asymptotically $(p+2)$-variate normal with mean vector (β, α, γ) and dispersion matrix I^{*-1}, where I^* is the Fisher information matrix obtained from I by taking the expected values of each entry.

There are different algorithms available to estimate the parameters by solving the score equations or directly optimizing the likelihood function. The Newton-Raphson method is the most common method used to estimate since it is easy to determine the derivatives of the score equations. In this numerical iterative method, the initial values play a vital role due to the logarithm function. In the simulation studies given in Section 4, we use the simplex method proposed by Neldar and Mead [15] to estimate the parameters. The simplex method is a simple method to use to estimate the parameters by maximizing the likelihood function where we do not need the derivatives of the function to be optimized.

3. Testing and Confidence Intervals for β

Tests and interval estimates of parameters can be derived by the likelihood ratio test procedure. We are mainly interested in the regression parameter β where the parameters $\theta = (\alpha, \gamma)$ are normally considered as nuisance parameters.

Let the p vector regression parameter be denoted as $\beta = (\beta_1, \beta_2)$, where β_1 and β_2 are of vectors of sizes k and $p-k$, respectively, and θ is the other parameter in the model. We are interested in testing

$$H_0 : \beta_1 = \beta_1^0 \text{ against } H_1 : \beta_1 \neq \beta_1^0, \tag{15}$$

where β_1^0 is the specified regression parameter value. To test H_0, we construct the likelihood ratio statistic

$$\Lambda^* = 2l\left(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\theta}\right) - 2l\left(\beta_1^0, \widehat{\beta}_2, \widehat{\theta}\right), \tag{16}$$

where $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\theta}$ are the maximum likelihood estimates under the full model. For a large value of n, Λ^* follows the χ_k^2 distribution under the null hypothesis.

Alternatively, we can use the test statistic

$$\Lambda_1 = \left(\widehat{\beta}_1 - \beta_1^0\right) V_{11}^{-1} \left(\widehat{\beta}_1 - \beta_1^0\right), \tag{17}$$

where V_{11} can be obtained from $V = I(\widehat{\beta}, \widehat{\theta})^{-1}$, which is partitioned as

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}. \tag{18}$$

Under the null hypothesis, Λ_1 follows χ_k^2 distribution.

Assuming asymptotic normality, we can construct the $100(1-\alpha)\%$ confidence interval for the individual regression parameter β_j as

$$\widehat{\beta}_j \pm z_{\alpha/2} \text{ s.e.}\left(\widehat{\beta}_j\right), \tag{19}$$

where s.e.$(\widehat{\beta}_j)$ can be obtained from V_{11}.

Another important problem is the selection of important covariates in the proportional reversed hazard models. Since we assume a parametric model, we can use variable selection methods such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). To test adequacy of the parametric model, Cox-Snell residuals can be used, which is explained in Section 5 with a case example.

4. Performance Analysis

To assess the performance of the proposed method, we carried out a large number of simulations. We generated samples of size 100 from an inverted Weibull distribution, with different values of parameters as $\alpha = (0.5, 1, 1.5, 2)$ and $\gamma = (0.5, 1, 1.5, 2)$. We considered a single covariate which is generated from Uniform (0,1) and a regression parameter assumed to be $\beta = (0.5, 1.1.5, 2)$. We developed the censoring mechanism using the random data generated from the exponential distribution with the parameter λ. We choose the value of λ such that the percent of censored data is between 10 and 20 percent. We used the simplex method proposed by Neldar and Mead [15] to estimate the parameters. We repeated the study for 10000 times and computed the mean and standard deviation of the parameter estimates. The entire study was repeated for a sample size of 250. The summary of the parameter estimates is given in Table 1.

From Table 1 we can see that the mean of the parameter estimates based on 10000 simulation is very close to the true parameter values and the standard deviation is also small. When the sample size increases the standard error of the estimates decreases and bias reduces. It should be noted that there is a slight positive bias in all cases, even though it is negligibly small. Since there are no comparable models based on reversed hazard rates, we did not perform any comparison studies.

5. An Example

We consider an extract of left censored data from an Australian twin study given in Duffy et al. [16]. The data consist of information on the age of appendectomy of monozygotic and dizygotic twins. There are observations with missing age at onset and therefore the data are left censored. The individuals having age at onset of less than 11 are left censored. The covariate, namely, Zygocity, has values from 1 to 6. This data set consists of 54 observations of which 15 are left censored. We use this data to illustrate the utility of the parametric reverse hazard rate model. Probability plotting and statistical test confirmed the distribution of data as inverted Weibull distribution. We use the simplex method to estimate the parameters. Since the parameter values are unknown and to avoid the effect of inappropriate initial values, we consider different initial values and choose the estimates which have maximum likelihood. Estimates of the parameters are $\widehat{\alpha} = 2.3940$, $\widehat{\beta} = -0.0142$, and $\widehat{\gamma} = 444.0586$. The 95% confidence interval for β indicates that the regression coefficient corresponding to Zygocity is not significantly different than zero; that is, the effect of Zygocity is negligible. This conclusion is also verified through the likelihood ratio test statistic value 0.0074 having a P value of 0.93.

Table 1: Averages and standard deviation (in brackets) of parameter estimates.

True parameter values			Estimate of parameters ($n = 100$)			Estimate of parameters ($n = 250$)		
α	β	γ	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$	$\widehat{\alpha}$	$\widehat{\beta}$	$\widehat{\gamma}$
1	1	1	1.0228	1.0189	1.0351	1.0080	1.0127	1.0096
			(0.0924)	(0.4040)	(0.2486)	(0.0556)	(0.2434)	(0.1444)
1	0.5	1	1.0199	0.5104	1.0287	1.0075	0.5046	1.0115
			(0.0906)	(0.3984)	(0.2389)	(0.0564)	(0.2489)	(0.1476)
1	1.5	1	1.0196	1.5321	1.0302	1.0083	1.5113	1.0121
			(0.0877)	(0.4033)	(0.2413)	(0.0547)	(0.2457)	(0.1430)
1	2	1	1.0200	2.0348	1.0313	1.0090	2.0178	1.0113
			(0.0862)	(0.4123)	(0.2384)	(0.0535)	(0.2530)	(0.1423)
1	1	0.5	1.0234	1.0260	0.5107	1.0085	1.0090	0.5049
			(0.0977)	(0.4338)	(0.1347)	(0.0607)	(0.2708)	(0.0824)
1	1	1.5	1.0199	1.0191	1.5590	1.0081	1.0064	1.5225
			(0.0860)	(0.3848)	(0.3553)	(0.0539)	(0.2381)	(0.2128)
1	1	2	1.0200	1.0223	2.0800	1.0081	1.0104	2.0301
			(0.0843)	(0.3759)	(0.4708)	(0.0534)	(0.2361)	(0.2823)
0.5	1	1	0.5099	1.0277	1.0252	0.5040	1.0088	1.0109
			(0.0459)	(0.3986)	(0.2400)	(0.0285)	(0.2452)	(0.1459)
1.5	1	1	1.5308	1.0156	1.0372	1.5142	1.0081	1.0121
			(0.1330)	(0.4072)	(0.2480)	(0.0814)	(0.2474)	(0.1449)
2	1	1	2.0460	1.0229	1.0329	2.0160	1.0129	1.0100
			(0.1794)	(0.4050)	(0.2444)	(0.1111)	(0.2512)	(0.1484)

We use a Cox-Snell residual plot to assess the goodness of fit. The Cox-Snell residual is defined by

$$r_i = -\log F(y_i) = \frac{\gamma}{y_i^{\alpha}} e^{\beta x_i}, \quad i = 1, \ldots, n. \tag{20}$$

If the model fits the data, then the residuals should have a standard exponential distribution, so that a hazard plot of residuals versus the Nelson-Aalen estimator of the cumulative hazard of the residuals will be a straight line with slope one. A plot of Cox-Snell residuals against the Nelson-Aalen estimates of the cumulative hazard rate of residuals is given in Figure 1, which shows that the fit is reasonably good.

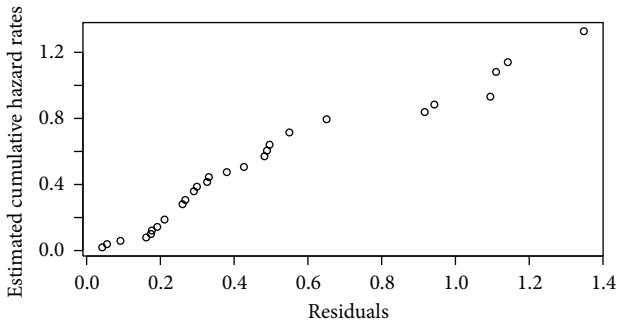


Figure 1: Plot of cumulative hazard rates of Cox-Snell residuals versus residuals.

6. Conclusions

Proportional reversed hazard rate models are more suitable for modeling the left censored lifetime data. In this paper, we proposed a parametric PRH model assuming that the lifetime data follows an inverted Weibull distribution. The estimation and hypothesis testing of the parameters of the model have been discussed in detail. The performance of the proposed model is assessed based on a large number of Monte Carlo simulations. Our simulations results clearly indicated that the proposed model is performing well. We applied the proposed model to a real life example to illustrate the utility of the method. Recently, Bayesian methodology was extensively employed in the analysis of lifetime data. The inference procedures of the proposed model by selecting an appropriate prior distribution are topics of research to be explored. The present work can be easily extended to any location-scale families of distributions.

Conflict of Interests

The authors declare that there is no conflict of interest for this paper.

Acknowledgments

The authors would like to thank the editor and an anonymous referee for their valuable comments and suggestions that

substantially improved the overall quality of an earlier version of this paper. The research of Dr. Variyath is supported by a grant from the Natural Science and Engineering Council of Canada.

References

[1] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, NY, USA, 2003.

[2] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding, *Statistical Models Based on Counting Process*, Springer, New York, NY, USA, 1993.

[3] R. E. Barlow, A. W. Marshall, and P. Proschan, "Properties of probability distributions with monotone hazard rate," *Annals of Mathematical Statistics*, vol. 34, pp. 375–389, 1963.

[4] J. Keilson and U. Sumitha, "Uniform stochastic ordering and related inequalities," *Canadian Journal of Statistics*, vol. 15, pp. 63–69, 1982.

[5] H. W. Block, T. H. Savits, and H. Singh, "The reversed hazard rate function," *Probability in the Engineering and Informational Sciences*, vol. 12, no. 1, pp. 69–90, 1998.

[6] M. S. Finkelstein, "On the reversed hazard rate," *Reliability Engineering and System Safety*, vol. 78, pp. 71–75, 2002.

[7] N. U. Nair, P. G. Sankaran, and G. Asha, "Characterizations of distributions using reliability concepts," *Journal of Applied Statistical Science*, vol. 14, no. 3-4, pp. 237–241, 2005.

[8] R. C. Gupta, P. L. Gupta, and R. D. Gupta, "Modeling failure time data by Lehman alternatives," *Communications in Statistics*, vol. 27, no. 4, pp. 887–904, 1998.

[9] C.-D. Lai and M. Xie, *Stochastic Ageing and Dependence for Reliability*, Springer, New York, NY, USA, 2006.

[10] A. W. Marshall and I. Olkin, *Life Distributions*, Springer, New York, NY, USA, 2007.

[11] X. Li and M. Xu, "Reversed hazard rate order of equilibrium distributions and a related aging notion," *Statistical Papers*, vol. 49, no. 4, pp. 749–767, 2008.

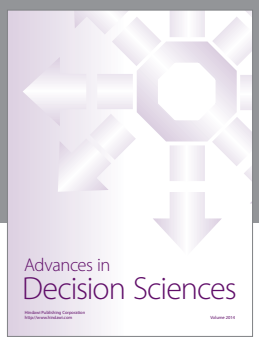
[12] G. Horny, "Inference in mixed proportional hazard models with K random effects," *Statistical Papers*, vol. 50, no. 3, pp. 481–499, 2009.

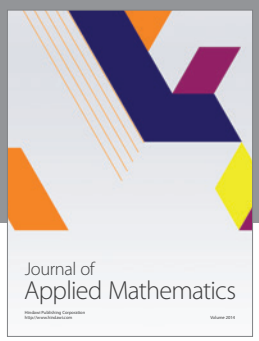
[13] C. A. V. Tojeiro and F. Louzada, "A general threshold stress hybrid hazard model for lifetime data," *Statistical Papers*, vol. 53, no. 4, pp. 833–848, 2012.

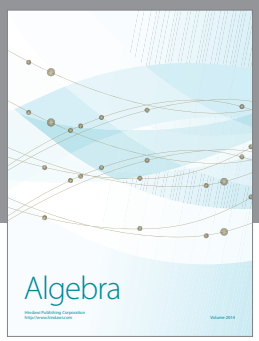
[14] D. Sengupta and A. K. Nanda, "The proportional reversed hazards regression model," *Journal of Statistical Theory and Applications*, vol. 18, no. 4, pp. 461–476, 2011.

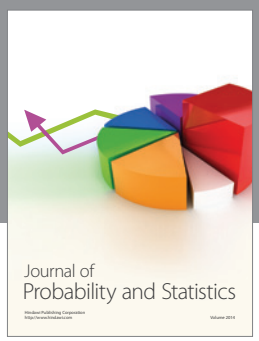
[15] J. A. Neldar and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–331, 1965.

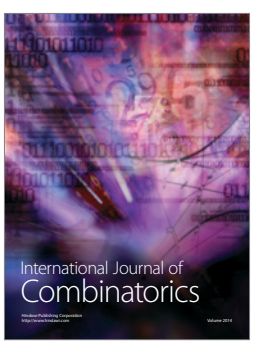
[16] D. Duffy, N. G. Martin, and J. D. Mathews, "Appendectomy in Australian twins," *American Journal of Human Genetics*, vol. 47, no. 3, pp. 590–592, 1990.

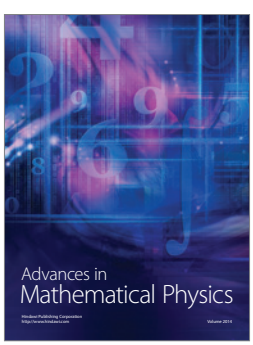


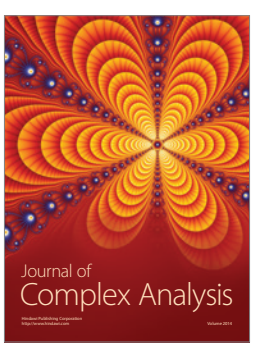


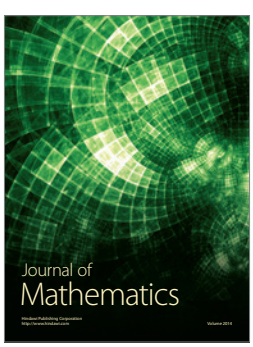


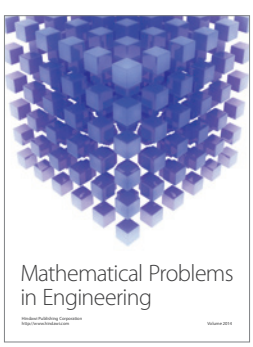


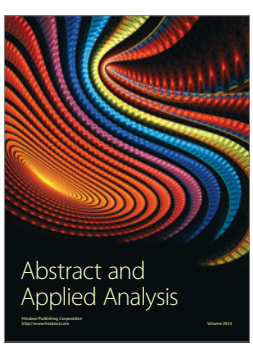


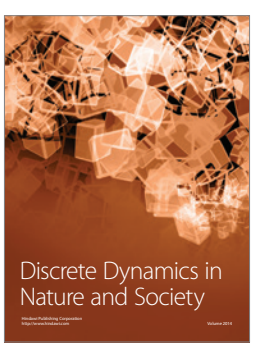


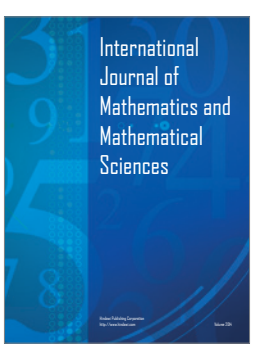


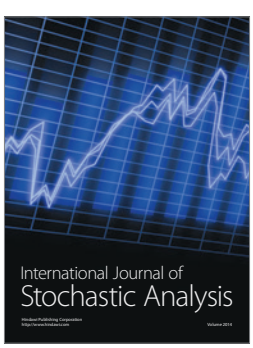












Hindawi

Submit your manuscripts at
http://www.hindawi.com