Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2009, Article ID 943926, 10 pages
doi:10.1155/2009/943926

Research Article

When Inflation Causes No Increase in Claim Amounts

Vytaras Brazauskas,[1] Bruce L. Jones,[2] and Ričardas Zitikis[2]

[1] *Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA*
[2] *Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON, Canada N6A 5B7*

Correspondence should be addressed to Bruce L. Jones, jones@stats.uwo.ca

Received 27 January 2009; Accepted 26 June 2009

Recommended by Tomasz J. Kozubowski

It is well known that when (re)insurance coverages involve a deductible, the impact of inflation of loss amounts is distorted, and the changes in claims paid by the (re)insurer cannot be assumed to reflect the rate of inflation. A particularly interesting phenomenon occurs when losses follow a Pareto distribution. In this case, the observed loss amounts (those that exceed the deductible) are identically distributed from year to year even in the presence of inflation. Nevertheless, in this paper we succeed in estimating the inflation rate from the observations. We develop appropriate statistical inferential methods to quantify the inflation rate and illustrate them using simulated data. Our solution hinges on the recognition that the distribution of the number of observed losses changes from year to year depending on the inflation rate.

1. Introduction

A number of challenges arise when an insurance policy covers only loss amounts that exceed a threshold known as the deductible. The insurer typically does not know about losses that are less than this amount, making appropriate characterization of the loss distribution impossible. This can even give rise to misleading and/or paradoxical observations about the distribution.

An interesting example of this has been observed in actuarial practice. A reinsurer desired to understand the impact of inflation on loss amounts. However, upon exploring the losses that were reported to the reinsurer, it was found that no inflation was present. The losses reported to the reinsurer were only those that exceeded a fixed deductible, which did not change over time as is typically the case. The losses reported in different years had near identical distributions. Specifically, the reinsurer found that the distribution of reported losses in each year could be accurately described by the same Pareto distribution. Moreover, attempts to model inflation by employing various macroeconomic indexes (e.g., consumer price index) also failed to yield satisfactory results as the reinsurance data was industry specific. The details of this problem were obtained through personal communications with reinsurance industry practitioners.

The Pareto distribution arises quite often in modelling insurance losses. This distribution uniquely possesses a property that gives rise to the reinsurer's observation regarding the inflation of loss amounts.

To examine this phenomenon statistically, we simulated losses corresponding to 10 successive years. The numbers of losses in these years is assumed to be independent Poisson random variables with mean 1000, and all loss amounts are independent. These are common assumptions in insurance loss modelling. The losses occurring during the jth year have a Pareto distribution with scale parameter $\theta = 1.05^{j-1}$ and shape parameter $\alpha = 2$. These parameter choices were arbitrary but reflect the phenomenon that has been observed. Throughout the paper, we will use the shorthand $Y \sim \text{Pareto}\ (\theta, \alpha)$ to indicate that a random variable Y has the Pareto distribution function

$$F_Y(y) = 1 - \left(\frac{\theta}{y}\right)^{\alpha}, \quad y > \theta,\ \theta, \alpha > 0 \tag{1.1}$$

with corresponding probability density function

$$f_Y(y) = \frac{\alpha}{y}\left(\frac{\theta}{y}\right)^{\alpha}, \quad y > \theta, \tag{1.2}$$

mean given by

$$\mathbf{E}[Y] = \frac{\alpha\theta}{\alpha - 1}, \quad \alpha > 1, \tag{1.3}$$

and median given by

$$\text{median}[Y] = 2^{1/\alpha}\theta. \tag{1.4}$$

So, losses during the jth year are distributed as Pareto $(1.05^{j-1}, 2)$. We assume that the insurer will pay only the amount of losses that exceed 5 and therefore will be unaware of any losses that are less than 5. The simulated data are summarized in Figure 1.

The left-hand graph shows box-and-whisker plots of loss amounts in each year. Each box extends from the first quartile to the third quartile, with the median indicated by the line inside the box. The whiskers extend to the most extreme observations that are not more than 1.5 times the interquartile range outside the box. We see very clearly from the left-hand graph the impact that inflation has on the loss distribution. The right-hand graph in Figure 1 summarizes the distribution of losses in each year that are greater than 5. These box-and-whisker plots do not show any signs of inflation of loss amounts.

Table 1 provides some additional information about the simulated loss data. The table shows that while the average loss amount increases with inflation, the average observed loss amount does not appear to increase. We also see that the number of observed losses

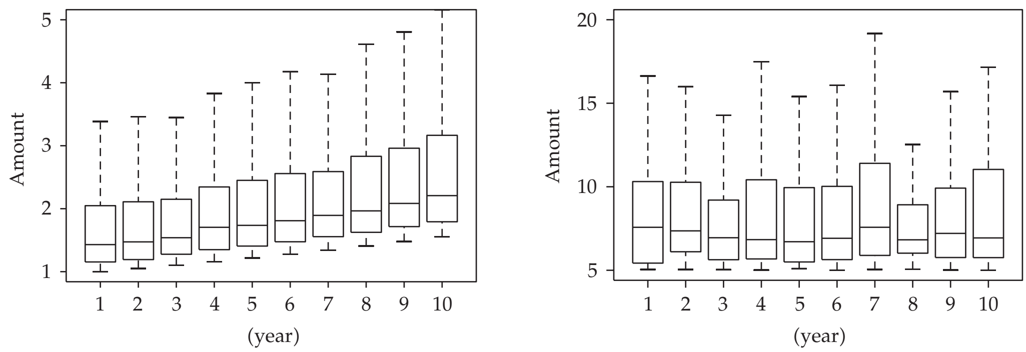

Figure 1: Box-and-whisker plots of all loss amounts (a) and observed loss amounts (b).

Table 1: Summary of simulated loss data.

Year	Number of losses	Average loss	Number of observed losses	Average of observed losses	Sum of observed losses
1	1004	1.9813	37	9.8732	365.3071
2	971	2.1358	43	10.6640	458.5501
3	1029	2.1206	44	9.4408	415.3972
4	1063	2.3359	56	9.8994	554.3648
5	1026	2.3554	62	8.2097	509.0030
6	1030	2.5579	78	9.2125	718.5715
7	1003	2.7498	75	10.7216	804.1190
8	955	2.7866	71	10.0545	713.8679
9	982	3.1771	89	12.1130	1078.0582
10	1029	3.0533	92	9.8543	906.5962

tends to increase over time, and this is how the information about inflation is captured. The sum of observed losses also increases over time. However, the increases reflect the so-called leveraging effect of the deductible (see [1, page 189]) and do not properly represent the increases due to inflation. This is because, if the deductible is kept unchanged, then total observed losses will not increase by the inflation rate because losses that were previously below the deductible may, with inflation, exceed the deductible.

The rest of the paper is organized as follows. In Section 2, we provide some background information and derive two methods for estimation of the inflation rates. In Section 3, numerical illustrations based on our simulated data are presented.

2. Estimating Inflation Rates

If we were observing every loss, then we would have a realization of the following array of random variables:

$$\left\{Y_{j,1},\ldots,Y_{j,N_j}\right\} \sim \text{Pareto}\left(\theta_j,\alpha\right),\quad j=1,\ldots,J, \tag{2.1}$$

where J represents the total number of years for which losses are observed, and N_j is the number of losses that occur in the year j. All random variables in array (2.1) are assumed independent and, row-wise, have Pareto distributions with the specified parameters, which are unknown and thus need to be estimated from available data. The data consist of only those losses whose amounts $Y_{j,k}$ exceed a specified threshold d, as the insurer is not informed of the losses which are less than this deductible. Hence, our data set is a realization of the following array:

$$\left\{X_{j,1}, \dots, X_{j,M_j}\right\}, \quad j = 1, \dots, J, \tag{2.2}$$

which is a subarray of (2.1). Obviously, the observed M_j do not exceed the unobserved N_j for every $1 \leq j \leq J$. All random variables in array (2.2) are independent and every $X_{j,k} \sim$ Pareto (d, α). The latter fact can be seen by noting that the $X_{j,k}$'s are copies of a random variable X_j, and the $Y_{j,k}$'s are copies of a random variable Y_j. Now $X_j = Y_j \mid Y_j > d$. Therefore, for all $x \geq d$,

$$\begin{aligned} \mathbf{P}\left[X_j > x\right] &= \mathbf{P}\left[Y_j > x \mid Y_j > d\right] \\ &= \frac{\mathbf{P}\left[Y_j > x\right]}{\mathbf{P}\left[Y_j > d\right]} \\ &= \left(\frac{d}{x}\right)^{\alpha}. \end{aligned} \tag{2.3}$$

The fact that this distribution does not depend on j is unique to the Pareto loss distribution and is reflected in the title of this paper. The property identified in the above equations raises the question of how to estimate the rate of inflation given the observed losses $X_{j,k}$. We note in passing that this property has been noted and utilized in a number of contexts including econometrics and engineering sciences (see [2, 3]).

Suppose that the annual inflation rates for the observation period are represented by $r_2, \dots, r_J$, where these rates are related to the Pareto-scale parameters by the equation

$$\frac{\theta_j}{\theta_{j-1}} = 1 + r_j. \tag{2.4}$$

Equation (2.4) arises from the very reasonable requirement that if r_j is the rate of loss inflation as one goes from year $j-1$ to year j, then $Y_j =_d (1+r_j)Y_{j-1}$. Note that if the Pareto distributions have finite first moments (i.e., $\alpha > 1$), then the ratio θ_j/θ_{j-1} in (2.4) can be replaced by $\mathbf{E}[Y_j]/\mathbf{E}[Y_{j-1}]$. However, we do not require the finiteness of first moments in this paper.

We first present a simple and intuitively appealing approach to estimating the inflation rate when we assume that it is the same in each year. We can also view this as a method of estimating the average inflation rate during the observation period. That is, the inflation rate r is such that

$$\theta_j = \theta(1 + r)^{j-1}, \tag{2.5}$$

with $\theta = \theta_1$. This method allows us to estimate α and r recognizing that most of the information about α is provided by the $X_{j,k}$'s, and given α, most of the information about r is provided by the M_j's.

We assume that $N_1, \ldots, N_J$ are independent Poisson random variables, and for each j, N_j has mean λ_j such that $\lambda_j = \lambda e_j$, where e_j represents the known number of exposure units in year j and λ is a parameter representing the claim rate per exposure unit. In other words, the e_j values indicate the amount of insurance in force in year j, and it is appropriate that the claim rate is proportional to e_j. The assumption that the number of losses has a Poisson distribution is common in actuarial science, though our first method generalizes easily to mixed Poisson distributions.

Now since the number of losses N_j has a Poisson distribution with mean λe_j, the number of observed losses M_j has a Poisson distribution with mean $\lambda e_j (\theta_j / d)^\alpha$. Thus,

$$\begin{aligned} \mathbf{E}[M_j] &= \lambda e_j \left(\frac{\theta_j}{d} \right)^\alpha, \\ \log \mathbf{E}[M_j] &= \log \lambda + \log e_j - \alpha \log d + \alpha \log \theta_j \\ &= \log \lambda + \log e_j - \alpha \log d + \alpha \log \theta + \{\alpha \log(1 + r)\}(j - 1). \end{aligned} \tag{2.6}$$

Therefore,

$$\log \left(\frac{\mathbf{E}[M_j]}{e_j} \right) = \log \lambda - \alpha \log d + \alpha \log \theta + \{\alpha \log(1 + r)\}(j - 1). \tag{2.7}$$

Notice that the right-hand side of (2.7) is a linear function of j with the slope $\alpha \log(1 + r)$. We could therefore estimate r by first estimating α by maximum likelihood using the conditional likelihood of the $X_{j,k}$'s, and then fit a linear function to the points $(j, \log(m_j / e_j))$, $j = 1, \ldots, J$, by ordinary least squares and estimate r using the estimate of the slope along with the MLE of α. This gives

$$\widehat{\alpha} = \frac{\sum_{j=1}^{J} m_j}{\sum_{j=1}^{J} \sum_{k=1}^{m_j} \log(x_{j,k} / d)}, \tag{2.8}$$

$$\widehat{r} = \exp \left\{ \frac{12 \sum_{j=1}^{J} j \log(m_j / e_j) - 6(J + 1) \sum_{j=1}^{J} \log(m_j / e_j)}{\widehat{\alpha} J(J^2 - 1)} \right\} - 1, \tag{2.9}$$

where $x_{j,k}$ is the realized value of $X_{j,k}$, and m_j is the realized value of M_j.

This approach allows us to estimate r without estimating the parameters λ and θ, which we consider nuisance parameters in our problem.

A more general approach involves estimating the parameters α and r_j, $j = 1, \ldots, J$, by maximum likelihood estimation using the full likelihood function. That is,

$$\begin{aligned} L &= \prod_{j=1}^{J} \frac{\left[\lambda e_j (\theta_j/d)^{\alpha}\right]^{m_j} \exp\{-\lambda e_j (\theta_j/d)^{\alpha}\}}{m_j!} \left(\prod_{k=1}^{m_j} f_{X_j}(x_{j,k})\right) \\ &= \prod_{j=1}^{J} \frac{\left[\lambda e_j (\theta_j/d)^{\alpha}\right]^{m_j} \exp\{-\lambda e_j (\theta_j/d)^{\alpha}\}}{m_j!} \left(\prod_{k=1}^{m_j} \frac{\alpha}{x_{j,k}} \left(\frac{d}{x_{j,k}}\right)^{\alpha}\right). \end{aligned} \tag{2.10}$$

Note that we have an identifiability problem because λ could be replaced by $\lambda' = c\lambda$ and θ_j by $\theta_j' = \theta_j / c^{1/\alpha}$, and the likelihood is unchanged. So, while we can determine estimates of λ and $\theta_1, \ldots, \theta_J$ that maximize the likelihood, these estimates are not unique. However, this is not a concern because we are not interested in λ, and we are interested in $\theta_1, \ldots, \theta_J$ only to the extent that they tell us the year-to-year inflation rates. We proceed with this in mind.

By cancelling multiplicative constants in the likelihood function and taking logs, we have

$$\begin{aligned} \ell = \sum_{j=1}^{J} \Bigg[& m_j \log \lambda + \alpha m_j \log \theta_j - \alpha m_j \log d - \lambda e_j \left(\frac{\theta_j}{d}\right)^{\alpha} \\ & + \sum_{k=1}^{m_j} (\log \alpha + \alpha \log d - \alpha \log x_{j,k}) \Bigg]. \end{aligned} \tag{2.11}$$

Differentiating with respect to θ_j, we have

$$\frac{\partial \ell}{\partial \theta_j} = \frac{\alpha m_j}{\theta_j} - \frac{\alpha \lambda e_j \theta_j^{\alpha - 1}}{d^{\alpha}}. \tag{2.12}$$

Therefore,

$$m_j - \hat{\lambda} e_j \left(\frac{\hat{\theta}_j}{d}\right)^{\hat{\alpha}} = 0, \tag{2.13}$$

$$\hat{\theta}_j = d \left(\frac{m_j}{\hat{\lambda} e_j}\right)^{1/\hat{\alpha}}. \tag{2.14}$$

This allows us to obtain the MLE of the inflation rate in year j, $r_j = \theta_j / \theta_{j-1} - 1$, $j = 2, \ldots, J$. That is,

$$\hat{r}_j = \left(\frac{m_j e_{j-1}}{m_{j-1} e_j}\right)^{1/\hat{\alpha}} - 1. \tag{2.15}$$

Differentiating the log-likelihood with respect to α, we have

$$\frac{\partial \ell}{\partial \alpha} = \sum_{j=1}^{J} \left[m_j \log\left(\frac{\theta_j}{d}\right) - \lambda e_j \log\left(\frac{\theta_j}{d}\right)\left(\frac{\theta_j}{d}\right)^{\alpha} + \sum_{k=1}^{m_j}\left(\frac{1}{\alpha} + \log d - \log x_{j,k}\right)\right]. \tag{2.16}$$

Replacing the parameters in (2.16) by their MLE's and using (2.14), we have

$$\sum_{j=1}^{J}\sum_{k=1}^{m_j}\left(\frac{1}{\widehat{\alpha}} + \log d - \log x_{j,k}\right) = 0, \tag{2.17}$$

which leads to (2.8), the same estimate we obtained using the first method.

The latter approach does not assume any structure between $r_2, \ldots, r_J$. However, as we did earlier, it might be reasonable to assume that $r_2 = \cdots = r_J$, in which case we denote the inflation rate by r. Hence, as before, $\theta_j = \theta(1+r)^{j-1}$, with $\theta = \theta_1$. In this case, we have only four unknown parameters, λ, α, θ, and r, and the log-likelihood function is

$$\begin{aligned}\ell = \sum_{j=1}^{J}\Bigg[& m_j \log\lambda + \alpha m_j \log\left\{\theta(1+r)^{j-1}\right\} - \alpha m_j \log d - \lambda e_j \left(\frac{\theta(1+r)^{j-1}}{d}\right)^{\alpha} \\ & + \sum_{k=1}^{m_j}\left(\log\alpha + \alpha\log d - \alpha \log x_{j,k}\right)\Bigg].\end{aligned} \tag{2.18}$$

Our identifiability problem remains. However, we can eliminate the problem by letting $\phi = \lambda(\theta/d)^{\alpha}$. Then

$$\ell = \sum_{j=1}^{J}\left[m_j \log\phi + m_j \alpha (j-1)\log(1+r) - e_j \phi (1+r)^{\alpha(j-1)} + \sum_{k=1}^{m_j}\left(\log\alpha + \alpha\log d - \alpha\log x_{j,k}\right)\right], \tag{2.19}$$

and we can determine the unique MLE's of α, ϕ, and r. Differentiating with respect to ϕ we have

$$\frac{\partial \ell}{\partial \phi} = \sum_{j=1}^{J}\left[\frac{m_j}{\phi} - e_j(1+r)^{\alpha(j-1)}\right], \tag{2.20}$$

and hence,

$$\widehat{\phi} = \frac{\sum_{j=1}^{J} m_j}{\sum_{j=1}^{J} e_j (1+\widehat{r})^{\widehat{\alpha}(j-1)}}. \tag{2.21}$$

Next we differentiate with respect to r and obtain

$$\frac{\partial \ell}{\partial r} = \sum_{j=1}^{J} \left[\frac{\alpha(j-1)m_j}{1+r} - \phi\alpha(j-1)e_j(1+r)^{\alpha(j-1)-1} \right], \tag{2.22}$$

which leads to

$$\sum_{j=1}^{J} \left[(j-1)m_j - \widehat{\phi}(j-1)e_j(1+\widehat{r})^{\widehat{\alpha}(j-1)} \right] = 0. \tag{2.23}$$

Finally, differentiating with respect to α, we have

$$\frac{\partial \ell}{\partial \alpha} = \sum_{j=1}^{J} \left[(j-1)m_j \log(1+r) - \phi(j-1)e_j(1+r)^{\alpha(j-1)} \log(1+r) + \sum_{k=1}^{m_j} \left(\frac{1}{\alpha} + \log d - \log x_{j,k} \right) \right]. \tag{2.24}$$

Replacing the parameters by their MLE's, setting the right-hand side of (2.24) equal to 0, and using (2.23), we obtain (2.8), as before. Substituting (2.21) into (2.23) and dividing by the numerator of (2.21), we have

$$\frac{\sum_{j=1}^{J}(j-1)m_j}{\sum_{j=1}^{J} m_j} - \frac{\sum_{j=1}^{J}(j-1)e_j(1+\widehat{r})^{\widehat{\alpha}(j-1)}}{\sum_{j=1}^{J} e_j(1+\widehat{r})^{\widehat{\alpha}(j-1)}} = 0. \tag{2.25}$$

Since (2.8) provides an explicit expression for $\widehat{\alpha}$, we can obtain $\widehat{r}$ by solving (2.25).

In practice, rather than simply assuming that all r_j's are equal, we should perform a hypothesis test with the null hypothesis $H_0 : r_2 = \cdots = r_J$. This can be accomplished by employing the well-known likelihood ratio test (LRT) whose test statistic is given by

$$-2\{[\text{maximum of } \ell \text{ under } H_0] - [\text{maximum of } \ell \text{ over full parameter space}]\}. \tag{2.26}$$

As follows, for example, from Casella and Berger [4, Section 10.3], the asymptotic distribution of the statistic given by (2.26) is chi-squared with $(J+1)-3$ degrees of freedom.

3. Numerical Illustrations

In this section we provide numerical illustrations of the methods presented in Section 2. We use the simulated data discussed in Section 1. However, assume we do not know the number of losses and average loss amounts shown in the second and third columns of Table 1. We do know the number of observed losses given in the fourth column as well as the amount of each observed loss that occurred in each year. Also, it is reasonable for us to assume that we know that the exposure is the same each year. The same Poisson parameter was used to generate the number of losses in each year. Therefore, suppose that $e_j = 1$ for $j = 1, \dots, 10$.

Table 2: Maximum likelihood estimates of r_j for $j = 2, \ldots, 10$.

j	2	3	4	5	6	7	8	9	10
$\hat{r}_j$	0.0786	0.0116	0.1291	0.0526	0.1226	−0.0196	−0.0272	0.1205	0.0168

Table 3: Point estimates and approximate 95% confidence intervals of r and α using the full likelihood and using the first approach. Note: the true parameter values are $r = 0.05$, $\alpha = 2$.

	Full likelihood approach		First approach	
Parameter	Estimate	Asymptotic CI	Estimate	Bootstrap CI
r	0.0503	(0.0353; 0.0654)	0.0526	(0.0375; 0.0702)
α	1.9858	(1.8328; 2.1389)	1.9858	(1.8246; 2.1495)

Applying the first method, we can estimate α and then r using (2.8) and (2.9). We obtain the estimates 1.9858 and 0.0526, respectively. Recall that the "true" parameter values are $\alpha = 2$ and $r = 0.05$.

In practice, we do not know that the loss inflation rate is the same each year, and our full maximum likelihood approach allows us to estimate the individual inflation rates. The estimates reported in Table 2 were obtained using (2.15), with $\hat{\alpha}$ obtained from (2.8).

If we then impose the restriction that the inflation rate is the same each year, we can obtain the maximum likelihood estimate of r by solving (2.25). Alternatively, rather than solving (2.25), the estimates can be obtained by numerically maximizing the log-likelihood function using, for example, the `optim` function in `R` (see [5]). This approach has the advantage of allowing one to obtain the Hessian matrix as a by-product of the maximization. Since the Hessian matrix equals (minus) the observed information matrix evaluated at the maximum likelihood estimates, an estimated variance-covariance matrix for the parameter estimators can be found by matrix inversion. This approach was used to obtain the point estimates and approximate 95% confidence intervals presented in Table 3. The estimates obtained using the first approach are also provided for comparison. In this case, the approximate confidence intervals were constructed by producing 1000 parametric bootstrap samples.

Having maximized the log-likelihood with and without the restriction that the inflation rate in each year is the same, we can perform a likelihood ratio test of the hypothesis that the inflation rates are the same. Using the LRT statistic in (2.26), we find that its value is 4.5741. Based on a chi-squared distribution with 8 degrees of freedom we find that the P-value is .8020 and conclude that the r_j's are statistically equal.

Acknowledgments

The authors sincerely thank Editor Tomasz J. Kozubowski and two anonymous referees for queries and suggestions that have guided them in revising the paper. The first author gratefully acknowledges the stimulating scientific atmosphere at the 38th ASTIN Colloquium in Manchester, United Kingdom, and Michael Fackler in particular for posing the problem whose solution makes the contents of the present paper. The second and third authors are grateful to the University of Wisconsin-Milwaukee for the most productive and pleasant stay during which results of the present paper had evolved to fruition.

References

[1] R. V. Hogg and S. A. Klugman, *Loss Distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1984.

[2] B. C. Arnold, *Pareto Distributions*, vol. 5 of *Statistical Distributions in Scientific Work*, International Co-operative Publishing House, Burtonsville, Md, USA, 1983.

[3] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions. Vol. 1*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 1994.

[4] G. Casella and R. L. Berger, *Statistical Inference*, Duxbury, Pacific Grove, Calif, USA, 2nd edition, 2001.

[5] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, http://www.r-project.org.


Advances in
Operations Research
Advances in
Decision Sciences
Journal of
Applied Mathematics
Algebra
Journal of
Probability and Statistics
The Scientific
World Journal
International Journal of
Differential Equations
Hindawi
Submit your manuscripts at
http://www.hindawi.com
International Journal of
Combinatorics
Advances in
Mathematical Physics
Journal of
Complex Analysis
Journal of
Mathematics
Mathematical Problems
in Engineering
Abstract and
Applied Analysis
Discrete Dynamics in
Nature and Society
International
Journal of
Mathematics and
Mathematical
Sciences
Journal of
Discrete Mathematics
Journal of
Function Spaces
International Journal of
Stochastic Analysis
Journal of
Optimization