IMPLICATIONS OF THE MANDATORY ELIMINATION OF A RATING VARIABLE

BY: FRANK J. KARLINSKI, III

BIOGRAPHY:

Mr. Karlinski is a principal and consulting actuary with William M. Mercer Meidinger Hansen, Inc. Prior to joining Mercer, he was executive vice president and chief operating officer of the Merchants Insurance Group. Prior to Merchants, Mr. Karlinski was employed by Prudential Property and Casualty Insurance Company, most recently as Vice President, Operations. He holds a bachelor of arts degree from Hamilton College. He received his CAS Fellowship in 1977.

ABSTRACT:

The amount of rate classification (rate dislocation) experienced by individual insureds when a rating variable is eliminated depends on the rate relativity associated with the factor and with its distribution. In may be possible to introduce a surrogate rating variable to replace the one eliminated which reduces the total rate dislocation in the system. A mathematical expression for rate dislocation can be used to determine rate relativities for the replacement variable that minimize rate dislocation. A very high correlation is necessary between the eliminated variable and its surrogate for the surrogate to be effective in reducing rate dislocation.

In private passenger automobile, mileage has been cited as a replacement variable for sex. When sex is eliminated alone or in conjunction with marital status and/or age, very little of the rate dislocation introduced is eliminated by the introduction of mileage.

Implications of the Mandatory Elimination of a Rating Variable

A body of literature exists on the attributes of an actuarially sound classification system and on the social questions inherent in the use of certain classification variables for certain lines of insurance. But relatively little has been written describing what happens when a rating variable is eliminated from a classification system. The purpose of this paper is to examine what happens to the rates of insureds under the mandatory elimination of a classification rating variable. We will introduce the concept of rate dislocation, define it mathematically, and use it to quantify the effects of eliminating a variable, both in the classes containing the variable to be eliminated and in those classes with which they are joined. We will examine the degree of rate dislocation that occurs under different scenarios and identify the parameters that affect the degree of dislocation. We will then introduce a surrogate variable and examine the effect of its introduction on rates and rate dislocation.

Having concluded the theoretical discussion, we will then apply the basic principles discussed to the actual case of private passenger automobile insurance. We will illustrate the effects on rates of eliminating one or more of age, sex and marital status. Finally, we will examine the impact of introducing mileage in place of the eliminated variable(s).

We begin with the concept of rate dislocation. Define rate dislocation conceptually as the degree by which the rates charged to individuals (or classes) differ after the elimination of a rating variable from those charged before the elimination. If rate dislocation can be expressed mathematically, the expression should, logically, have the following characteristics:

1. When two classes are joined by the elimination of a rating variable, the amount of rate dislocation caused by the joining should vary directly as the difference between the original rates of the two classes.

2. For classes that are being joined, the dislocation should be minimized by choosing the mean rate of the classes as the new combined rate.

We can simply define the rate dislocation of a single class caused by the elimination of a variable as the percentage difference between the rate charged after the elimination and the rate charged before it. Therefore, for a single class, the dislocation caused by eliminating variable A is:

$$\mathrm{Dis}(A) = \left|\frac{R'-R}{R}\right|$$

where

R is the rate charged before the variable is eliminated.

R' is the rate charged after the variable is eliminated.

Clearly the definition satisfies condition (1) above. To expand the definition over all classes, we must combine the dislocation of each class in a way that satisfies condition (2). We find that one such definition is one that is analogous to the statistical definition of coefficient of variation. Using that definition, the dislocation function can be expressed in equation (1) as:

$$\mathrm{Dis}(A) = \frac{\sqrt{\sum_{i=1}^{n} D_i (R'_i - R_i)^2}}{\sum_{i=1}^{n} D_i \; R_i} \qquad (1)$$

D_i = is the distribution of exposures in class i.

Thus Dis(A) is the square root of the weighted average of the squared differences between old and new rates divided by the average rate for all classes. That equation (1) satisfies condition 2 is shown in Appendix 1. Whether it is the only function that satisfies the condition is not known. (The author was unable to find others.)

We now illustrate the use of the function with an example. Assume two rating variables, A and B, with the occurrence of A independent of the occurrence of B. Assume the occurrence of the variables and the rate relativities to be as follows:

Variable	Rate Relativity	% of Risks With Variable
A	1.2	10%
B	1.3	25%
A and B	2.0	-

Assuming a base rate of $100, variables A and B define four classes, with rates and distribution as follows:

Class	Distribution	Rate
Base	.675	$100
A	.075	120
B	.225	130
AB	.025	200

Now if rating variable A is eliminated from the classification system, the classes will be reduced to two, Base (containing the old Base class and the old class A) and B (containing old B and old AB).

Applying the dislocation function, Dis(A) can be calculated as follows:

Class	Distribution	Old rate(R)	New rate(R')	$(D)(R' - R)^2$
Base	.675	$100	$102	2.7
A	.075	120	102	24.3
B	.225	130	137	11.0
AB	.025	200	137	99.2
			Total	137.2
			Square Root	11.7
			Dis(A) =	10.6%

The elimination of A causes a rate dislocation of 10.6%. The same exercise performed for the elimination of variable B (that is, without eliminating A) gives the following result:

Class	Distribution	Old rate(R)	New rate(R')	$(D)(R' - R)^2$
Base	.675	$100	$107.5	38.0
A	.075	120	140	30.0
B	.225	130	107.5	113.9
AB	.025	200	140	90.0
			Total	271.9
			Square Root	16.5
			Dis(B) =	14.9%

The rate dislocation is 14.9% for the elimination of variable B, much higher than A. Why is Dis(B) greater than Dis(A)? One reason (from characteristic 1 above) must be that the rate relativity for B is greater than for A. In addition, we note that the distribution of B is greater than A. So we ask: What effect does the distribution of the variable being eliminated have on the dislocation? It can be shown that the expansion of equation (1) for n=2 is given by:

$$\text{Dis} = \frac{\sqrt{(1-p)p}(r-1)}{1+p(r-1)} \qquad (2)$$

Where p is the percentage of the population in the class being eliminated, and r is the rate relativity for that class. Equation (2) is derived in Appendix 2. Now let us see how the dislocation function behaves as p varies by differentiating the function with respect to p. If for notational ease, we say Dis=f, then:

$$\ln(f) = 1/2\ \ln(p) + 1/2\ \ln(1-p) + \ln(r-1) - \ln(1+p(r-1))$$

and

$$\frac{d\ \ln(f)}{dp} = \frac{1}{2p} - \frac{1}{2(1-p)} - \frac{(r-1)}{1+p(r-1)}$$

and finding the common denominator, we get

$$= \frac{1 - p(r+1)}{2p(1-p)(1+p(r-1))} = \frac{f'}{f}$$

Therefore,

$$\frac{df}{dp} = f(1-p(r+1))/2p(1-p)(1+p(r-1))$$

$$= \frac{(1-p(r+1))(r-1)((1-p)p)^{1/2}}{2p(1-p)[1+p(r-1)]^2}$$

And by inspection of the numerator we find that df/dp = 0 when p = 1/(r+1). Therefore, the dislocation is maximized as a function of the distribution of the variable at 1/(r+1). And dislocation increases monotonically for 0<p<1/(r+1) and decreases monotonically for $p>1/(r+1)$.

This is a very curious and unexpected result. We know by definition that, for rating variables with equal distributions, the elimination of the variable with the greatest rate variance causes the greatest rate dislocation. Now we find that the same cannot be said of distribution. That is, for a set of variables with equal rate variance, the elimination of the one with the greatest distribution does not necessarily cause the greatest rate dislocation.

This result can be illustrated with a change in the values of our prior example. Give A and B each a rate relativity of 2.0. Set the distribution of A equal to 30% and the distribution of B equal to 50%. Then the rates and dislocations are given in the following table:

		Original	Eliminate A		Eliminate B	
Class	Distribution	Rate	Rate	$D(R'-R)^2$	Rate	$D(R'-R)^2$
Base	.35	$100	$130	315	$150	875
A	.15	200	130	735	300	1500
B	.35	200	260	1260	150	875
A B	.15	400	260	2940	300	1500
		$195		5250		4750
		Dislocation		37.2%		35.3%

And we see that eliminating variable A causes more dislocation than eliminating B, even though their relativities are equal and B commands more distribution. The elimination of A causes greater dislocation because its distribution is closer to the maximum dislocating distribution of 1/(r+1), which, given the rate relativity of two is equal to one-third.

Now let us see what happens when a rating variable is eliminated and another variable, correlated with the first, is introduced. Assume variable A, currently in use, defines classes A and Base A, with rates and distribution as follows:

Class	Distribution	Rates
Base A	.5	$100
A	.5	160

Suppose A is eliminated and a surrogate variable B, not currently in use, is introduced which identifies, to some extent, the same risks as A. Let B define classes B and Base B and be distributed, relative to A, as follows:

		Current	Variable B Distribution	
	Current Rates	Distribution	Base B	B
Base A	$100	.5	.5	.0
A	$160	.5	.1	.4
	$130	1.0	.6	.4

As the table shows, A and B are highly correlated, in that B identifies 80% of the risks identified by A, and does not include in class B any risks previously rated in Base A. Now suppose data is available which indicates a rate relativity for B (relative to Base B) of 1.2. Using that experience, the rate charged for Base B would be x, given by:

$$.6x + .4(1.2x) = 130$$

$$x = \$120$$

and the class B rate would be 1.2x, or $144.

The various rate dislocations are as follows:

Class Old	Class New	Dist.	Orig. Rate	Eliminate A w/o using B Rate	Eliminate A w/o using B $D(R'-R)^2$	Eliminate A Substitute B Rate	Eliminate A Substitute B $D(R'-R)^2$
Base A	Base B	.5	$100	$130	450	$120	200
Base A	B	.0	100	130	0	144	0
A	Base B	.1	160	130	90	120	160
A	B	.4	160	130	360	144	102
			$130	$130	900	$130	462
				Square root	30.0		21.5
				Dislocation	23.1%		16.5%

The 23.1% rate dislocation caused by eliminating A without introducing a substitute is reduced to 16.5% by introducing B. This reduction of less than 30% seems small considering that, of the risks that fell into class A, variable B identified 80% of them.

We now ask ourselves whether the use of a factor other than the 1.2 experience relativity for class B will give rates which produce a smaller rate dislocation relative to the original rates for classes Base A and A. The answer is that there are rates which produce less rate dislocation. And of course the rates which produce the least dislocation can be derived by differentiating the dislocation function, setting it equal to zero, and solving for B. (Recognizing that the average rate must remain at $130 allows the elimination from the equation of the rate for the Base B class.) The result is a rate of $160 for class B and $110 for class Base B. The rate dislocation then drops to 13.3%. But this could have been expected, since the $160 and $110 rates are just the weighted averages of the component original rates which, by definition, must produce the minimum dislocation. In fact, it can be shown that minimizing the dislocation function is mathematically equivalent to taking the weighted average of the original classes.

In considering the substitution of rating variable B for A, note that, at best B can eliminate about 42% (1-(13.3%/23.1%)) of the dislocation. This may seem surprising since B identifies 80% of the risks formerly in class A and does not include in class B any of the formerly class Base A risks. If, for this special case in which class B does not contain any risks formerly in class Base A, we define the correlation between A and B as the percentage of formerly class A risks that full into class B, we can examine dislocation as a function of correlation. A graph of the minimum dislocation as a function of the correlation of A and B is shown for this example as Exhibit 1. The shape of the curve reveals that the rate dislocation is not reduced very much by the introduction of B until B reaches a high degree of correlation with A.

Now let us turn to an actual class plan for private passenger automobile. For this exercise we will limit ourselves to the variables of age, sex and marital status and not consider the senior citizen classes. Using data on Exhibit 2 columns 1 through 5, we can calculate the rate dislocation caused by elimination of one or more of the variables. If we assume that the rate for any class after the elimination of a variable will be set at the weighted average of the rates of the merging classes, the revised rates are those found on Exhibit 2, columns 6 through 9. Rate dislocations associated with those rates are as follows:

Variable(s) Eliminated	Rate Dislocation
Sex	16.2%
Sex and Marital Status	18.3
Age	30.0
Age, Sex and Marital Status	35.2

Three general statements can be made upon inspection of these results:

1. The most important of the three variables, by far, is age.

2. Once sex is eliminated, also eliminating marital status adds very little to the rate dislocation.
3. Once age is eliminated, also eliminating sex and marital status adds relatively little additional rate dislocation.

Annual mileage is the variable most often cited as a possible surrogate for age, sex and marital status, and it is most often linked with the variable of sex. Several studies have shown that males drive, on average, significantly more than females. The data from one such study[1] shows the following:

Annual Mileage	Distribution Males	Distribution Females	Rate Relativity
0 - 6,000	24%	54%	1.00
6,000 +	76	46	1.48
Total	55%	45%	

If we divide each class into two classes based on mileage, with short mileage defined as use less than 6,000 miles annually, the data shows that 76% of males and 46% of females would be classified as long mileage. The data also gives a long/short accident frequency relativity of 1.48. From the prior analysis, we know

1 From a Study by the U.S. Department of Transportation entitled "1977 Nationwide Personal Transportation Study, Characteristics of 1977 Licensed Drivers and Their Travel" as quoted in an article in the Journal of Insurance Regulation entitled "Sex-Divided Mileage, Accident, and Insurance Cost Data Show That Auto Insurers Overcharge Most Women", June, 1988, pages 399-401.

that it takes a high degree of correlation between the eliminated variable and the surrogate to reduce significantly the rate dislocation. We also know that use of the long/short relativity of 1.48 will not necessarily produce the greatest possible reduction in dislocation. The results of introducing long and short mileage classes at a relativity of 1.48 are:

Variables(s) Eliminated	Rate Dislocation Without Mileage	Rate Dislocation With Mileage
Sex[2]	16.2%	17.5%
Age, Sex & Mar. Status	35.2%	38.6%

Introduction of mileage actually produces additional rate dislocation; a most surprising result. By the methods previously described, the long/short relativity that produces the minimum dislocation can be calculated. It turns out to be between 1.13 and 1.20 depending on class and produces a rate dislocation of 15.4%, a minimal (4%) improvement over the dislocation caused by rates with no mileage variable at all.

Over the entire distribution, the average rating factor for men is 1.14, for women it is 1.15. These numbers are so close that, for

[2] In the analysis of sex eliminated rates, we did not split the married classes into long and short mileage classes. Since they were not merged with other classes when sex was eliminated, splitting them into long and short classes would create an artificial rate dislocation which would distort the comparisons.

the age, sex and marital status eliminated rates, the minimum dislocation long/short relativity is approximately 1.01. Use of this factor produces dislocation which is, of course, very close to that with no mileage variable at all, 35.1%.

<u>Summary</u>

We have examined, from a theoretical perspective, the implications for insureds of eliminating a classification rating variable. We have introduced the concept of rate dislocation and defined it mathematically. We have used it to quantify the impact on rates of the elimination of a rating variable and of the introduction of a surrogate variable. From our theoretical analysis, we reached the following conclusions:

1. Rate dislocation is a function of the rate relativity and the distribution of the eliminated variable.

2. Rate dislocation increases monotonically as the relativity of the eliminated variable increases.

3. Rate dislocation increases as the distribution of the eliminated variable increases within the range zero to 1/(1+r), where r is the rate relativity, and decreases from 1/(1+r) to 1.

4. The amount of rate dislocation eliminated by the introduction of a surrogate variable is related to the correlation between the surrogate and the original (i.e., the efficiency of the surrogate in identifying the same risks identified by the original).

5. A very high degree of correlation is necessary between the original and the surrogate variables to reduce significantly the rate dislocation caused by the elimination of the original variable.

6. The rate dislocation function (or equivalently, distributional data) can be used to find a value for the rate relativity of the surrogate variable which minimizes rate dislocation, without reference to experience data of the surrogate.

We then applied the principles of the theoretical discussion to actual data for private passenger automobile. Using actual class plan factors and distributions, we first eliminated various combinations of age, sex and marital status and then substituted mileage. We determined the following:

1. By far the most important of the three variables is age.

2. Once sex is eliminated, eliminating marital status adds little additional rate dislocation.

3. Once age is eliminated, also eliminating sex and marital status adds little additional rate dislocation.

4. There are circumstances under which the introduction of mileage to replace age, sex and/or marital status actually increases the dislocation from the original rates.

5. Despite the use of a long/short relativity that minimizes rate dislocation, introduction of mileage has very little effect on the rate dislocation caused by the elimination of sex and almost no effect on that caused by elimination of age, sex and marital status.

It should not be inferred from the above that the use of mileage as a rating variable is not valid and desirable, either in conjunction with age, sex and marital status or in place of them. The data clearly shows that mileage differentiates drivers with significantly different accident rates. What mileage cannot do to any significant degree, no matter what relativity is chosen, is to reduce dislocation caused by the elimination of one or all of those variables.

Dislocation as a Function of

The Correlation Between Two Variables

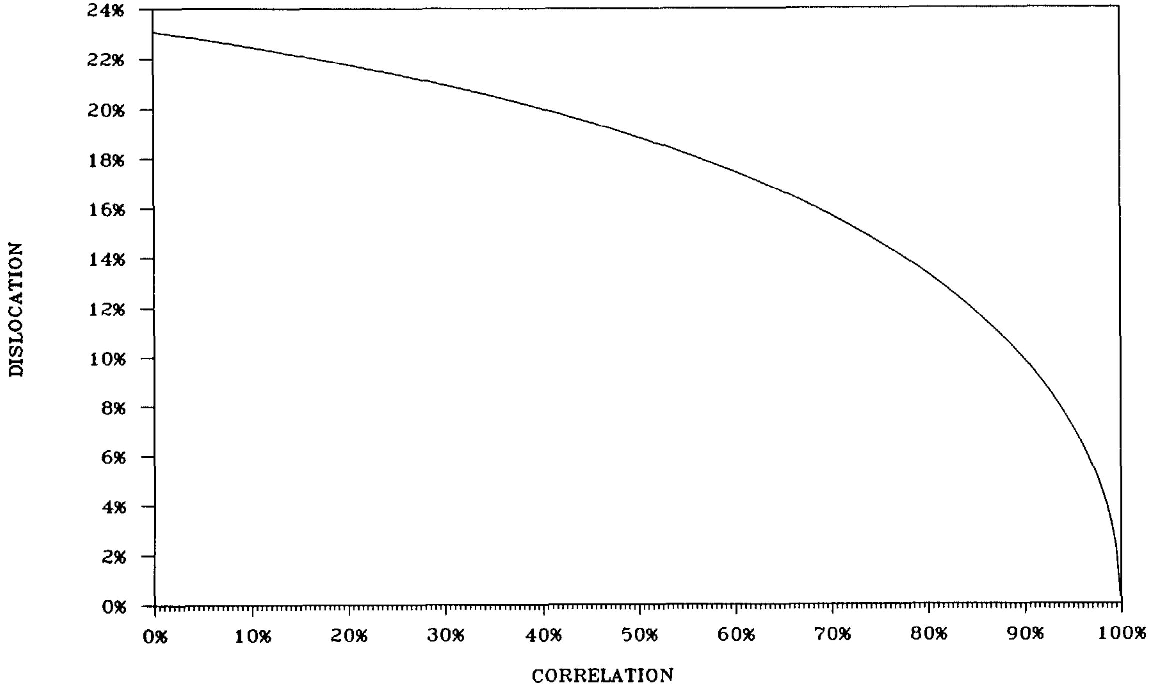

EXHIBIT 1

PRIVATE PASSENGER AUTOMOBILE CLASSIFICATION FACTORS AND DISTRIBUTIONS

Rating Variables			Distribution	Original Rating Factor	*Revised Rating Factors Variable(s) Eliminated*			
Sex	Marital Status	Age			Sex	Age	Sex & M/S	Age/Sex & M/S
Female	Single	17 to 20	2%	1.6	2.44	1.15	2.33	1.14
Female	Single	21 to 24	2%	1.3	1.78	1.15	1.64	1.14
Male	Married	17 to 20	1%	1.8	1.80	1.02	2.33	1.14
Male	Married	21 to 24	2%	1.3	1.30	1.02	1.64	1.14
Male	Single	17 to 20	3%	3.0	2.44	1.57	2.33	1.14
Male	Single	21 to 24	3%	2.1	1.78	1.57	1.64	1.14
Male	Single	25 to 29	3%	1.5	1.30	1.57	1.14	1.14
Male	Married	25 to 29	6%	1.0	1.00	1.02	1.14	1.14
Female	Single	25 to 29	2%	1.0	1.30	1.15	1.14	1.14
Female	Single	30+	6%	1.0	1.00	1.15	1.00	1.14
Male	Single	30+	10%	1.0	1.00	1.57	1.00	1.14
Male	Married	30+	60%	1.0	1.00	1.02	1.00	1.14

APPENDIX 1

Proof that the dislocation function satisfies condition 2.

$$\mathrm{Dis}(A) = \frac{\sqrt{\sum_{i=1}^{n} D_i (R'_i - R_i)^2}}{\sum_{i=1}^{n} D_i \; R_i}$$

Eliminate classifications and charge entire population one rate r, so that $R_i = r$ for every class i;

$$\text{then } \mathrm{Dis}(A) = \sqrt{\sum_{i=1}^{n} D_i (r-R_i)^2} \Big/ \sum_{i=1}^{n} D_i R_i$$

$$\ln \mathrm{Dis}(A) = 1/2 \ln \left(\sum D_i (r-R_i)^2 \right) - \ln \left(\sum D_i R_i \right)$$

$$\frac{d \ln \mathrm{Dis}(A)}{dr} = (1/2) \; \frac{d[D_i (r-R_i)^2]/dr}{\sum D_i (r-R_i)^2}$$

$$= \frac{\sum D_i (r-R_i)}{\sum D_i (r-R_i)^2}$$

So $\dfrac{d \ln \mathrm{Dis}(A)}{dr} = \dfrac{1}{\mathrm{Dis}(A)} \; \dfrac{d[\mathrm{Dis}(A)]}{dr}$ implies that

$$\frac{d[\mathrm{Dis}(A)]}{dr} = 0 \text{ if and only if } \frac{d[\ln(\mathrm{Dis}(A))]}{dr} = 0$$

$$\text{or } \sum D_i (r-R_i) = 0$$

$$= \sum D_i r - \sum D_i R_i$$

$$= r - \sum D_i R_i$$

since $\sum D_i = 1$

Thus $r = \sum D_i R_i$ minimizes Dis(A).

APPENDIX 2

Expansion of the dislocation function for n=2

$$Dis(A) = \frac{\sqrt{\sum_{i=1}^{n} D_i(R'_i-R_i)^2}}{\sum_{i=1}^{n} D_i \; R_i}$$

Let n=2 and let i=1 be the base class. If A_1=1.0 and A_2=r, than r is the rate relativity for class 2. Now since $D_2=1-D_1$, if we set $D_2=p$, the summation in the numerator of (1) becomes:

$$= D_1(A_1-D_1A_1-D_2A_2)^2+D_2(A_2-D_1A_1-D_2A_2)^2$$

and substituting 1 for A_1, r for A_2 and p for D_2, we get

$$= (1-p)[1-(1-p)-pr]^2+p[r-(1-p)-pr]^2$$

$$= (1-p)p^2(r-1)^2+p(1-p)^2(r-1)^2$$

$$= (1-p)p(r-1)^2[p+(1-p)]$$

$$= (1-p)p(r-1)^2$$

and Dis then becomes

$$Dis = \frac{\sqrt{(1-p)p}(r-1)}{1+p(r-1)}$$