

REPRESENTATIONS OF CLAIMS ARISING FROM A RISK PORTFOLIO

D. H. REID

London, U.K.

I. INTRODUCTION

This paper is concerned with certain problems which arise in the course of rate fixing for a portfolio comprising a large number of direct risks.

This is, of course, the traditional territory of the actuary and there is no lack of papers discussing specific aspects of the subject, such as the analysis of claim frequency, or the fitting of distributions to claim amount data. Equally there is a literature which discusses in general terms the philosophy which it is suggested should be followed by the non-life insurance ratemaker. It is when one tries in practice to follow a logical and consistent path through from portfolio records to rates that gaps in the literature become apparent and it is with covering these gaps, or at least some of them, that the present paper is concerned.

It is not the concern of the author in this paper to attempt to describe the entire ratemaking process. That there is no reference to the analysis of office expenses, commissions, or to the application of "judgment" or "experience" should not be taken to imply that these aspects are not susceptible to the kind of approach advocated here. Attention here is focused upon what might in a sense be regarded as the most basic aspect: the claim process or experience.

In many insurance operations the position which seems to be developing is that there is no lack of necessary computer hardware, and in principle, of data: the difficulties arising relate more to the form in which data should be retained and extracted, and the manner in which it should be used to assist in rate-making.

One point which might perhaps usefully be made here and from which confusion may have arisen in the past concerns the question of simplicity in analysis. Several authors have remarked upon the need for simplicity in any analysis of claims experience in order that

the resulting premium structure may be simple enough for practical use. There are really two separate points here: even though an analysis may be relatively complex its results may be capable of expression in simple terms and it is clear from what has been said above that claims experience is only one of several ingredients each of which may require definition in relatively complex terms. The requirement for simplicity, therefore, may be regarded as a need for simplicity in the final stages of the rate-making process. This, in turn, may be visualised as the deliberate fitting of a simple structure to the perhaps relatively complex structure of premiums determined up to that stage. Once again, the increasing impact of computers and associated hardware is having an impact upon the simplicity which was at one time a pre-requisite of clerical systems.

In this paper no account is taken of the time dependence of claim experience (the "tail" of the account). It is assumed that a sufficient period will have elapsed since the time of exposure to risk that claim distributions will for practical purposes be complete. Since the primary concern here is the analysis of the structure of claim experience rather than its absolute level, this is in practice not a severe limitation.

The subject under consideration, variously described as "factorial analysis" or "classification of risks", has been treated by many authors. Bailey & Simon (1) and Mehring (2) develop an iterative approach to the fitting of multiplicative models to claim frequency and Seal (3), more recently, describes another approach to the analysis of claim frequencies based upon the use of the "logit" transformation. Johnson & Hey (4) also briefly describe a method of approach to the analysis of claim frequencies. What the present paper attempts to do is to look at the whole problem of analysing claim experience and to set it into a context of established statistical methodology.

2. GENERAL APPROACH

In analysing claim experience the conventional approach is to separate claim frequency and measures of claim amount and to try to analyse each of these quantities. Experience has shown that the following difficulties arise:

- (a) Frequencies may be correlated with average amounts both over the portfolio at a given time and through secular time for a given risk class. Thus, risk classes having an additional share of nil claims but otherwise comparable costs to other risk classes have both distorted frequencies and average claim costs. Clearly this kind of correlation should be taken into account but there is a danger that whilst frequency may be quite readily examinable average claim cost may be much less so, with the result that misleading inferences may be drawn.
- (b) The separation of claim frequency and amount leads to difficulties in establishing significance and in view of the correlation noted above these difficulties can be so serious as to make the precise assessment of risk factors by this approach virtually impossible.
- (c) Considerable difficulties may arise in analysing claim costs. Typically no one theoretical distribution is particularly applicable and there are in any case great difficulties in finding probability models of claim amount which when applied to component risk groups aggregate into recognisable distributions. Even if such distributions could be found, there is not usually sufficient information to construct them for individual risk classes.
- (d) If in consequence of (c) the attempt is made to analyse average claim costs these may be found to have a very substantial variability for the skew claim distributions often encountered in practice. This can lead to difficulties in detecting the pattern of the claims experience.
- (e) It may be suspected that the incidence of large claims is unduly heavy in certain parts of the portfolio. Whilst it may be possible to examine such a hypothesis by tabulating experience of large claims, it is difficult, if not impossible, to integrate this experience in a consistent way with, for example, an analysis based upon average claim amounts.
- (f) All of these points are particularly relevant to collective risk statistics which collate the experience of several insurers who may have different claim recording practices, leading to differing claim frequencies from one company to another and associated differences in average claim amounts.

This paper proposes an alternative approach to the analysis of claims experience which is a natural one if the artificial device of splitting claim frequency and claim amount by the introduction of the question "has a claim occurred?" be dropped. If the methods of the theory of random variables are considered it will be recalled that for continuous variables the approach is based upon the dissection of the real axis into intervals which in the limit may be regarded as vanishingly small. There is no reason why the same approach should not be adopted in looking at claims experience in relation to the exposure to risk for unit time of a unit of risk. Omitting the final step of progression to the limit yields the method developed below.

Extending this approach slightly we may consider analysing claims by type as well as by size for those portfolios where different types of claim may be recorded. The results of this process may be visualised in the following form for a particular period of exposure to risk. Associated with each individual risk will be a vector of observed numbers of claims, analysed both by type and by size interval. As we are considering in the following development that risks are grouped into classes defined by combinations of risk factors, such a vector could describe aggregate claim experience for the risks in each factor combination.

In order to reduce space it is convenient in this paper without loss of generality to regard claims as being of only one type so that the components of the vector are distinguished only by the ranges of claim size to which they relate.

3. METHOD OF ANALYSIS

3.1 *The Problems*

It is convenient here to repeat in more detail the problems with which we are concerned. We are given a risk portfolio comprising risk units each classified by a number of risk factors. The structure of the risk factors is that of a complete cross-classification—each level of each factor may be associated with each level of the others: the portfolio may, of course, have no risks falling into a particular factor combination.

The questions with which we are concerned are the following:

- (a) The estimation of expected claim cost for each risk group and the attaching of some kind of estimated accuracy to this.

- (b) The examination of the risk factor classification in order to see whether this might be simplified with no loss of ability to discriminate differences in claims experience, or, if, on the other hand, it should be elaborated.

3.2 Notation

The data resulting from this approach may be described with the following notation. Risk factors r in number are denoted by I, J, K, \dots (thus, for example, I might represent age of policyholder), with associated numbers of factor levels n_i, n_j, \dots . A particular combination is denoted by i, j, k, \dots and a particular risk within that factor combination by superscript (t) .

It is assumed to be possible to define a meaningful "exposure to risk" which might be regarded as measured in terms of unit risk exposed for unit time. Exposure to risk is denoted by e . Thus, ${}^{(1)}e_{12} \dots$ denotes the exposure to risk of the first individual risk of the factor combination I at level one, J at level two, etc.

Claims are classified into mutually exclusive and exhaustive categories by size which are indexed by (k) : the number of claims arising is denoted by y . Thus, for example, ${}^{(t)}y_{12}^{(k)}$ denotes the number of claims falling into the k th size range arising from the t th risk in factor combination I at level one, J at level two \dots

3.3 The Model

In order to proceed it is necessary to set up some kind of model of the data, and a natural model to choose might be one which represents the numbers of claims arising on an individual risk in unit time as the sum of a number of elements. One element would represent the supposed effect of each risk factor and a further element would represent an error term. Numbers of claims here, of course, refers to the numbers of claims analysed by size range.

The difficulty with this approach is that in many practical situations the numbers recorded would almost certainly take values 0 or 1 and so have a standard deviation almost, if not precisely, functionally related to the mean. Thus, the assumption of homoscedasticity usually made in fitting such linear models is not applicable and conventional methods of analysis lose much of their attraction.

In order to proceed use may be made of a variance-stabilising transformation. The natural choice here appears to be the inverse-sine transformation which possesses the considerable advantage that the variance of the transformed variable depends only on the exposure to risk being inversely proportional to it. See e.g. Cox (4). This transformation will, therefore, be adopted in the following development although the logit transformation could be used in a similar way. The model now becomes

$$\sin^{-1} \sqrt{\frac{\sum_{(i)} y_{ij}^{(k)}}{\sum_{(i)} e_{ij} \dots}} = z_{ij}^{(k)} \dots, \text{ say}$$

$$= \mu^{(k)} + \alpha_i^{(k)} + \beta_j^{(k)} + \dots + \varepsilon_{ij}^{(k)} \dots,$$

where now $\{\varepsilon_{ij}^{(k)} \dots\}$ are assumed to be uncorrelated with zero means and with variances proportional to $\{\sum_{(i)} e_{ij} \dots\}^{-1}$, and $\mu^{(k)}, \alpha_i^{(k)}, \beta_j^{(k)} \dots$ are constants.

3.4 Estimation of Parameters

The conventional least squares approach proceeds by consideration of the quantities $(z^{(k)} - X\theta^{(k)})' V^{-1} (z^{(k)} - X\theta^{(k)})$, where $\theta^{(k)'}$ is the row vector $(\mu^{(k)}, \alpha_1^{(k)} \dots \alpha_{n_1-1}^{(k)}, \beta_1^{(k)}, \dots \dots)$, $X = (x_{ij})$ has elements of zero except where

$$x_{i1} = 1, \text{ or } x_{ij} = 1 \quad (j \neq 1),$$

the latter case occurring when the factor level corresponding to position j of the vector $\theta^{(k)}$ of factor effects is present in the factor combination corresponding to the i th position of the vector $z^{(k)}$. $z^{(k)}$ represents the observations $z_{ij}^{(k)} \dots$ taken in a convenient order.

It will be observed that X is not dependent upon (k) . V^{-1} is the diagonal matrix $\text{diag} (\{\sum_{(i)} e_{ij} \dots\})$ taken in the order corresponding to vector $z^{(k)}$. Then we have the least squares estimates of $\theta^{(k)}$ as $\hat{\theta}^{(k)} = (X' V^{-1} X)^{-1} X' V^{-1} z^{(k)}$.

Note that the number of parameters chosen to represent the levels of each factor is at least one less than the corresponding number of factor levels in order to secure full rank of the matrix X .

3.5 Estimates of Cost

If C_k represents the mean cost of claims in size range (k) the estimated claim cost is $\bar{C} \cdot (\sin^2 \underline{x} \cdot \hat{\theta}^{(k)})$ for the factor combination represented by row \underline{x} of X .

If $S_{(k)}^2$ is defined as $\frac{1}{N - n_0} \{ \underline{z}^{(k)'} V^{-1} \underline{z}^{(k)} - \hat{\underline{\theta}}' X' V^{-1} X \hat{\underline{\theta}} \}$ where N is the number of elements of $\underline{z}^{(k)}$ and n_0 the number of elements of $\hat{\underline{\theta}}^{(k)}$, then the variance—co-variance matrix of $\hat{\underline{\theta}}^{(k)}$ may be estimated by $S_{(k)}^2 (X' V^{-1} X)^{-1}$ so that the sampling variance of the above estimated cost is approximately $\underline{x}' (X' V^{-1} X)^{-1} \underline{x} \sum_k S_{(k)}^2 C_k^2 \{ \sin 2x \cdot \hat{\underline{\theta}}^{(k)} \}^2$, assuming that the elements C_k are assumed known. This can be modified to incorporate estimated variances of the terms C_k (assuming also zero correlation of the numbers of claims falling in different size ranges for a given factor combination: this point is discussed further in paragraph 4)

3.6 Tests on Risk Factors

Question (b) of paragraph 3.1 related to whether the inclusion of a particular risk factor was justified by its power of discrimination, and this question may be re-phrased in the context of the simplified model under consideration here as one of testing the significance of the groups of parameters corresponding to each risk factor. At the same time the question of the overall significance of the model at each level of claim, which more properly should have been discussed earlier, can be dealt with.

Thus, following convention, if it were desired to investigate whether a particular factor whose effects were represented in the vectors $\underline{\theta}^{(k)}$ by $\underline{\Psi}^{(k)}$ (a subset of $\underline{\theta}^{(k)}$) were worth including in the analysis, the approach would be as follows, assuming that the usual normality assumptions were met:

Taking the matrix $(X' V^{-1} X)^{-1}$ the sub-matrix $A\psi$ of elements corresponding to the position of $\underline{\psi}$ in $\underline{\theta}$ is selected. From this is formed the quantity $\underline{\psi}^{(k)'} (A\psi)^{-1} \underline{\psi}^{(k)}$, which is then divided by its number of degrees of freedom and compared with the residual mean square from the earlier analysis ($S_{(k)}^2$) against an appropriate F distribution. This process would be performed consecutively for each value of k , i.e. each level of claim cost, significance at any one level indicating the probable justification of that factor's inclusion in the rating process.

The other part of question (b) of paragraph 3.1 concerns the assessment of the sub-divisions within a particular risk factor. Once

again, within the model proposed here, this assessment would be made separately at each claim level by setting up an appropriate comparison. Normally this would be done in the course of the analysis and after the opportunity was available for inspecting results: in this case care is required in assessment of parameter differences and the appropriate method is one of assessing "multiple comparisons".

Scheffé [6] supplies the details and it is merely noted here that if $\hat{\xi}$ represents the estimated value of the comparison under consideration and if $\hat{\sigma}_{\hat{\xi}}$ represents its estimated standard deviation then according to Scheffé the confidence interval is given by

$$\hat{\xi} - s\hat{\sigma}_{\hat{\xi}} \leq \xi \leq \hat{\xi} + s\hat{\sigma}_{\hat{\xi}},$$

where

$$s = (qF_{\alpha; q, n-k})^{1/2},$$

$F_r : v_1, v_2$ is the upper α point of an F distribution with v_1, v_2 degrees of freedom, q is the dimensionality of the space of estimable functions of which ξ is a member, and $n - r$ represents the number of degrees of freedom of the residual variance.

4. MORE PRECISE SPECIFICATION

The formulation and methods proposed in section 3 may appear somewhat unnatural when the problem is viewed ab initio. Essentially we are concerned with a set of vector variables which we are trying to model by a linear combination of parameters $\theta^{(k)}, k = 1, 2 \dots p$. The problem might be regarded simply as one of multivariate multiple regression or rather multivariate analysis of variance and might be reformulated as follows:

$$V^{-1/2} Z = V^{-1/2} X\Theta + E,$$

where Z is the matrix with columns $\underline{z}^{(k)},$
 $\Theta \dots \dots \dots \underline{\theta}^{(k)},$

and E is an error matrix, assumed to represent a sample from the multivariate normal distribution with mean $\underline{0}$ and var-covariance matrix Σ .

It will be observed that the formulation is now more general in that specific account is taken of potential correlations between the

different levels of claim amount. As in the univariate case, the matrix Σ is assumed constant and unrelated to Θ .

Models of this kind are extensively investigated by Anderson [7] who shows that the estimates of the parameters are identical to those derived by the approach of Section 3. The advantage of this approach however, is that it enables estimates which are in principle more precise to be made of the variance of claim cost. Indeed, Anderson shows that the row vector $(\theta^{(1)'}, \theta^{(2)'}, \dots, \theta^{(p)'})$ has covariance matrix

$$\begin{pmatrix} \sigma_{11}(X'V^{-1}X)^{-1} & \sigma_{12}(X'V^{-1}X)^{-1} & \dots & \dots \\ \sigma_{21}(X'V^{-1}X)^{-1} & & \dots & \dots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \dots & \dots & \dots & \sigma_{pp}(X'V^{-1}X)^{-1} \end{pmatrix}$$

where $(\sigma_{ij}) = \Sigma$,

estimated by $\hat{\Sigma} = \Sigma \sum_{\alpha} \zeta'_{\alpha} \xi_{\alpha} - \hat{\Theta}'X'V^{-1}X\hat{\Theta}$,

where ξ_{α} represents the α th row of $V^{-1/2}Z$.

The estimated claim cost is $\underline{C} \cdot (\sin^2 \underline{x} \cdot \hat{\theta}^{(k)})$,

and if we write $v_k = \underline{x} \cdot \hat{\theta}^{(k)}$ we have

$$\begin{aligned} \text{cov}(v_k, v_j) &= E \{ \underline{x} \cdot (\hat{\theta}^{(k)} - E[\hat{\theta}^{(k)}]) (\hat{\theta}^{(j)} - E[\hat{\theta}^{(j)}])' \underline{x}' \}, \\ &= \sigma_{jk} \underline{x} (X'V^{-1}X)^{-1} \underline{x}' \end{aligned}$$

and since $\text{cov}(\sin^2 v_k, \sin^2 v_j) \sim \sin 2v_k \sin 2v_j \text{cov}(v_k, v_j)$, (see e.g. Rao [8]), the estimated variance of the estimated claim cost may readily be determined. This approach may be extended along similar lines to that of paragraph 3.

5. AN EXAMPLE

The following example is based upon a survey of part of a U.K. private car portfolio in which the total units of exposure amounted to 289,853. The rating factors selected are:

geographical area	(4 levels)
type of vehicle	(3 levels)
date of birth of policyholder	(3 levels)
no-claims-bonus % at last renewal	(3 levels)

The levels of claim size (k) selected are:

(k)	1	2	3
range (£)	0—10	10—75	75+
C_k (£)	1.55	31.04	251.15

A sample of the input data is shown in Appendix I, exposure and claim vectors associated with factor levels being displayed. The derived matrix $X' V^{-1} X$ is shown in Appendix II and its inverse in Appendix III. The values of certain derived quantities are as follows:

(k)	1	2	3
$\underline{z}^{(k)'} V^{-1} \underline{z}^{(k)}$	12420	12610	13230
$S^2_{(k)}$.2611	.2615	.3381

and the values of $\hat{\theta}^{(k)}$ are:

(k)	1	2	3
$\mu^{(k)}$.217	.218	.218
area	— .018 — .017 — .007	— .018 — .017 — .007	— .018 — .017 — .007
vehicle	— .034 — .022	— .035 — .022	— .035 — .022
age	+ .035 + .018	+ .002 + .018	+ .092 + .018
N.C.B.	+ .059 + .030	+ .115 + .029	+ .062 + .030

Tests

The overall significance of the factors for $k = 1$ is tested by referring $\frac{213.75}{9} / .2611$ i.e. 90.96

to an F distribution with 9 and 98 degrees of freedom. The result is significant at the 0.1% level and similar results are obtained for the other values of k .

The results of tests carried out on the factors singly, doubly, and triply are shown in Appendix IV for $k = 1$. The extreme right-most column headed T represents the quantity to be compared with an F distribution with N and 98 degrees of freedom: in each case the results are highly significant. Similar results apply for other values of k .

Estimation of Costs

For each of a selection of incidence vectors x the estimated costs with associated approximate variances (as paragraph 3) may be derived as follows:

<i>Incidence Vector</i>	<i>Cost Estimate (£)</i>	<i>Variance</i>
1100000101	16.94	.33
1010100000	7.71	.05
1001001010	33.61	1.04

Corresponding results for all factor combinations are readily calculable.

6. CONCLUSION

This paper is intended to be provocative rather than definitive. It is clear that the statistical analysis of claim experience provides an area in which there is much scope for research and the application of ingenuity. Nevertheless it seems to the author that an approach along the lines described in paragraph 3 may be expanded to deal with most of the problems arising in practice. The other thesis presented here is that an undue emphasis may have been placed on the representation of claim experience by claim frequency and distributions of claim size. It appears to the author that there is much to be gained by looking at experience from a viewpoint other than this, and it is this conviction which has provided the impetus for the development described above.

7. REFERENCES

- [1] BAILEY, R. A. and SIMON, L. J., Two studies in automobile insurance ratemaking. A.S.T.I.N. Bulletin, 1960, I, 192.
- [2] MEHRING, J., Über den Einfluss des Anfängerrisikos auf das Tarifmerkmal der Schadenfreiheit in der Kraftfahrt-Haftpflichtversicherung. Trans. of the 18th International Congress of Actuaries, 1968, Vol. II 551.
- [3] SEAL, H., Stochastic Theory of a Risk Business, Wiley, 1969.
- [4] JOHNSON, P. D. and HEY, G. B., Statistical Studies in Motor Insurance. Journal of the Institute of Actuaries, 1971, 97, 199.
- [5] COX, D. R., Analysis of Binary Data. Methuen 1970.
- [6] SCHEFFÉ, H., The Analysis of Variance, Wiley, 1959.
- [7] ANDERSON, T. W., An Introduction to Multivariate Statistical Analysis, Wiley, 1958.
- [8] RAO, C. R., Linear Statistical Inference and its Applications, Wiley 1973

APPENDIX I

area	vehicle	age	N.C.B.	exposure	Nos. of Claims		
					$k = 1$	$k = 2$	$k = 3$
I	I	I	I	323	26	29	39
I	I	I	2	738	38	29	57
I	I	I	3	250	11	8	17
I	I	2	I	220	11	17	11
I	I	2	2	630	27	27	27
I	I	2	3	960	35	35	35
I	I	3	I	351	19	29	19
I	I	3	2	1319	47	47	47
I	I	3	3	4887	126	126	126
I	2	I	I	203	20	23	30
I	2	I	2	623	32	24	48
I	2	I	3	238	10	8	15
I	2	2	I	210	10	15	10
I	2	2	2	727	35	35	35
I	2	2	3	1552	52	52	52
I	2	3	I	325	18	27	18
I	2	3	2	1488	56	56	56
I	2	3	3	6869	223	223	223
I	3	I	I	53	6	7	9
I	3	I	2	181	16	12	24
I	3	I	3	103	5	4	8
I	3	2	I	129	9	14	9
I	3	2	2	480	30	30	30
I	3	2	3	1007	54	54	54
I	3	3	I	158	11	17	11
I	3	3	2	802	39	39	39
I	3	3	3	3591	141	141	141

APPENDIX 2

E+04	1	2.899E+05	2.842E+04	1.005E+05	7.176E+04	9.486E+04	1.315E+05	2.528E+04	6.061E+04	1.779E+04	7.468
E+03	2	2.842E+04	2.842E+04	0.000E+00	0.000E+00	9.678E+03	1.224E+04	2.712E+03	5.915E+03	1.972E+03	6.988
E+04	3	1.005E+05	0.000E+00	1.005E+05	0.000E+00	3.367E+04	4.547E+04	9.118E+03	2.095E+04	5.937E+03	2.527
E+04	4	7.176E+04	0.000E+00	0.000E+00	7.176E+04	2.303E+04	3.221E+04	6.668E+03	1.598E+04	4.312E+03	1.887
E+04	5	9.486E+04	9.678E+03	3.367E+04	2.303E+04	9.486E+04	0.000E+00	1.254E+04	1.889E+04	7.277E+03	2.824
E+04	6	1.315E+05	1.224E+04	4.547E+04	3.221E+04	0.000E+00	1.315E+05	1.016E+04	2.599E+04	7.412E+03	3.173
E+04	7	2.528E+04	2.712E+03	9.118E+03	6.668E+03	1.254E+04	1.016E+04	2.528E+04	0.000E+00	4.886E+03	1.474
E+04	8	6.061E+04	5.915E+03	2.095E+04	1.598E+04	1.889E+04	2.599E+04	0.000E+00	6.061E+04	5.186E+03	2.041
E+00	9	1.779E+04	1.972E+03	5.937E+03	4.312E+03	7.277E+03	7.412E+03	4.886E+03	5.186E+03	1.779E+04	0.000
E+04	10	7.468E+04	6.988E+03	2.527E+04	1.887E+04	2.824E+04	3.173E+04	1.474E+04	2.041E+04	0.000E+00	7.468

APPENDIX 3

E-06	1	2.593E-05	-1.146E-05	-1.120E-05	-1.137E-05	-1.556E-05	-1.608E-05	-3.149E-07	-4.469E-06	-3.719E-06	-4.19
E-07	2	-1.146E-05	4.645E-05	1.124E-05	1.125E-05	5.382E-08	5.215E-07	-1.096E-06	-3.464E-07	-1.939E-09	5.373
E-07	3	-1.120E-05	1.124E-05	2.119E-05	1.124E-05	-1.765E-07	8.204E-08	-9.324E-07	-3.651E-07	5.652E-07	4.591
E-07	4	-1.137E-05	1.125E-05	1.124E-05	2.519E-05	2.710E-07	3.882E-07	-1.077E-06	-6.361E-07	4.802E-07	2.727
E-07	5	-1.556E-05	5.382E-08	-1.765E-07	2.710E-07	2.675E-05	1.594E-05	-3.346E-06	7.916E-07	-1.345E-06	-8.982
E-08	6	-1.608E-05	5.215E-07	8.204E-08	3.882E-07	1.594E-05	2.347E-05	-1.344E-06	9.205E-07	-2.956E-07	-8.070
E-06	7	-3.149E-07	-1.096E-06	-9.342E-07	-1.077E-06	-3.346E-06	-1.344E-06	5.055E-05	6.778E-06	-1.292E-05	-8.985
E-06	8	-4.469E-06	-3.464E-07	-3.651E-07	-6.361E-07	7.916E-07	9.205E-07	6.778E-06	2.212E-05	-4.233E-06	-3.288
E-06	9	-3.719E-06	-1.939E-09	5.652E-07	4.802E-07	-1.345E-06	-2.956E-07	-1.292E-05	-4.233E-06	6.507E-05	7.747
E-05	10	-4.199E-06	5.373E-07	4.591E-07	2.727E-07	-8.982E-07	-8.070E-08	-8.985E-06	-3.288E-06	7.747E-06	2.036

APPENDIX 4

Generation and analysis of subsets

SR = 2.559E+01

Degrees of freedom = 98

Factor	Subscripts	Factor	Subscripts	Factor	Subscripts	S	N	T
1						+1.66785E+01	3	+2.12906E+01
2						+4.40110E+01	2	+8.42719E+01
3						+3.24313E+01	2	+6.20992E+01
4						+8.00488E+01	2	+1.53276E+02
1		2				+6.11027E+01	5	+4.67995E+01
1		3				+4.78715E+01	5	+3.66655E+01
1		4				+9.83488E+01	5	+7.53269E+01
2		3				+7.31233E+01	4	+7.00079E+01
2		4				+1.19203E+02	4	+1.14124E+02
3		4				+1.65321E+02	4	+1.58278E+02
1		2		3		+8.90585E+01	7	+4.87223E+01
1		2		4		+1.37814E+02	7	+7.53961E+01
1		3		4		+1.81981E+02	7	+9.95591E+01
2		3		4		+1.96748E+02	6	+1.25576E+02