Astin Bulletin 11 (1980) 77-90

TWO PRAGMATIC APPROACHES TO LOGLINEAR CLAIM COST ANALYSIS

PETER TER BERG

Parameter estimation in case of loglinear modelled claim cost distribution characteristics is mathematically tractable, especially with the Inverse Gaussian and Lognormal distribution.

1. INTRODUCTION

The collecting and presenting of statistical data is often in terms of aggregated quantities. In the field of non-life insurance, we can think of *total exposure, total number of claims* and *total claim costs.*

Such statistics can be produced for all kinds of more or less homogeneous *risk groups, insurance lines* and for different *time periods.*

In order to perform maximum likelihood estimation to such statistics, it is necessary to specify the probability distribution governing the stochastic process, which generates the data.

This paper will focus on positive claim cost analysis, conditionally on a known, positive, number of claims. Having the specification of the probability distribution for the cost of a single claim, there is still the task of deriving the n-fold convolution for this distribution in order to get the probability distribution for the aggregate quantity. For most distributions this will lead to intractable numerical procedures in relation with parameter estimation; the Gamma and Inverse Gaussian distribution being exceptions. As the specification of the probability distribution for the cost of a single claim is seldom justifiable on axiomatic grounds, it is advisable to use tractable distributions such as the Gamma and Inverse Gaussian distribution. This point of view was already stressed by HADWIGER (1942).

Furthermore, the form of the n-fold convolution will approach a Normal distribution, a property also shared by the Gamma, Inverse Gaussian and Lognormal distribution, being distributions which do not assign any probability mass to the negative axis and which are as such preferable on this ground.

The paper runs as follows. First tractable probability distributions will be specified. Then, a loglinear parametrization for the mean and shape parameter will be given. The next step is the application of maximum likelihood estimation, together with the study of the properties of the loglikelihood function as well as the derivation of the information matrix, whose inverse is the Rao-Cramér lower bound and which is a tool for analyzing the possible loss of information due to aggregation.

For the maximum likelihood analysis the Inverse Gaussian and Lognormal distribution are singled out, being the two pragmatic approaches.

The paper closes with final remarks, containing miscellaneous aspects, the most important one perhaps being the existence of a consistent estimator, which is easy to calculate and which might have good efficiency compared with fully efficient estimators.

2. THE GAMMA AND INVERSE GAUSSIAN DISTRIBUTION

Consider a sequence of n independent, identically distributed random variables, taking on positive, continuous values.

Which probability distributions have the property that the sum of these random variables will follow the same probability distribution with modified parameters depending on n? HADWIGER (1942) addressed himself to this question, making a plea for using such distributions for reasons of simplicity and lack of guidance to specify the distribution on axiomatic grounds.

The first distribution, he mentioned, was the Gamma distribution:

$$f(y|\mu, \phi) = \left(\frac{\phi}{\mu}\right)^{\phi} \frac{y^{\phi-1} \exp(-\phi\mu^{-1}y)}{\Gamma(\phi)} \tag{2.1}$$

and the second one is now known as the Inverse Gaussian distribution:

$$f(y|\mu, \phi) = \left(\frac{\mu\phi}{2\pi y^3}\right)^{1/2} \exp\left\{-\tfrac{1}{2}\phi\left[\frac{y}{\mu} + \frac{\mu}{y} - 2\right]\right\}. \tag{2.2}$$

Both distributions are parametrized in such a way that the mean and variance are given by μ and $\mu^2\phi^{-1}$.

The parameter ϕ characterizes the shape of the distribution and with increasing ϕ, both (2.1) and (2.2) approach the Normal distribution [1]. In case of the sum of n random variables, the parameters μ and ϕ are modified into $n\mu$ and $n\phi$. Taking the sample mean modifies ϕ into $n\phi$ and leaves μ unaltered. The use of (2.1) or (2.2) instead of the correct, in general unknown, distribution will imply a specification error for the n-fold convolution, which is mild in nature, at least if n or better $n\phi$ is large. Compared with the Gamma distribution, the Inverse Gaussian distribution is not well-known, whereas it is full of tractable properties.

For instance, maximum likelihood estimation results in closed form expressions whereas this is not possible for the shape parameter of the Gamma distribution.

[1] In the sense of $\sqrt{\varphi}\,(y-\mu)/\mu$ tending to a standard normally distributed random variable.

Furthermore, (2.2) is flexible in the possibility to take on very skew forms, approaching a member of the stable family:

$$f(y|\lambda) = \left(\frac{\lambda}{2\pi\, y^3}\right)^{1/2} \exp\{-\tfrac{1}{2}\lambda/y\} \tag{2.3}$$

This density results by defining $\lambda = \mu\phi$ and taking the limit for $\mu \to \infty$ in (2.2).

Some historical details and references for the Inverse Gaussian distribution are in order.

SCHRÖDINGER (1915) derived this distribution for the first hitting time in Brownian motion, WALD (1947) derived it in connection with sequential testing and TWEEDIE (1957) wrote on it from the viewpoint of mathematical statistics. FOLKS and CHHIKARA (1978) have written a review on the Inverse Gaussian distribution, clearly being unaware of the pioneering work by HADWIGER (1940a, b, 1942) and HADWIGER and RUCHTI (1941) who applied this distribution to age-specific fertility analysis.

For this reason (2.2.) is associated in demography with the name of *Hadwiger*.

Notwithstanding the plea by HADWIGER (1942) to use (2.2), it has virtually remained unnoticed in insurance mathematics; SEAL (1969, 1978) is an exception.

3. A NORMALIZING TRANSFORMATION

Consider again a sequence of n independent, identically distributed random variables, taking on positive continuous values with mean and variance equal to μ and $\mu^2\phi^{-1}$.

The arithmetic mean of these random variables has a probability distribution with mean μ and variance $\mu^2/n\phi$. The density function of this sample mean may still be skew and a logarithmic transformation will induce a more symmetric picture and a density close to the Normal one, at least for $n\phi$ sufficiently large.

The logarithm of the sample mean has asymptotically mean and variance equal to $\log \mu$ and $(n\ \phi)^{-1}$.

However, for $n\phi$ being small, it may be good to refine the expectation of the logarithm of the sample mean.

Denoting the sample mean by $\bar{y}$, we have:

$$\begin{aligned} \log \bar{y} &= \log \mu + \log (\bar{y}/\mu) \\ &= \log \mu + \log (1 + [\bar{y} - \mu]/\mu) \\ &= \log \mu + \left(\frac{\bar{y} - \mu}{\mu}\right) - \tfrac{1}{2}\left(\frac{\bar{y} - \mu}{\mu}\right)^2 + \text{remainder}. \end{aligned} \tag{3.1}$$

Taking expectations gives:

$$E(\log \bar{y}) \doteq \log \mu - (2n\phi)^{-1} \tag{3.2}$$

The approximation (3.2) is—as it should—compatible with the asymptotic expansion for the mean of the logarithm of a Gamma distributed random variable as well as an Inverse Gaussian distributed random variable, as investigated by WHITMORE and YALOVSKY (1978), who state that for $n\phi \geqq 10$ this approximation will be satisfactory.

4. LOGLINEAR PARAMETRIZATION AND MATRIX NOTATION

The following loglinear parametrizations are adopted:

$$\left.\begin{aligned} \log \mu_r &= \mathbf{x}_r' \boldsymbol{\theta} \qquad (4.1) \\ \log \phi_r &= \mathbf{z}_r' \boldsymbol{\eta} \qquad (4.2) \end{aligned}\right\} r = 1, \ldots\ldots\ldots, \text{R.}$$

The column vectors $\mathbf{x}_r$ and $\mathbf{z}_r$ characterize a risk group, an insurance line or a time period, where r is an identifying index. The use of $\mathbf{x}_r$ and $\mathbf{z}_r$ allows for the possibility of different variables explaining the mean and shape parameter.

The column vectors $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ contain the parameters which are subject to estimation. The dimension of $\mathbf{x}_r$ and $\boldsymbol{\theta}$ is denoted by K and the dimension of $\mathbf{z}_r$ and $\boldsymbol{\eta}$ equals L. In most cases, the elements of $\mathbf{x}_r$ and $\mathbf{z}_r$ will be dummy variables taking on the values 0 and 1. But this need not be the case and the mathematics do not depend on it.

So, $\mathbf{x}_r$ and $\mathbf{z}_r$ can also contain variables of economic and demographic nature, being measured on a continuous scale.

It is clear that (4.2) is not encouraging in connection with the Gamma distribution (2.1).

In case that $\mathbf{z}_r$ boils down to a scalar, the analysis for the Gamma distribution is simple however, as shown in TER BERG (1980).

So, maximum likelihood analysis with (4.1) and (4.2) will be performed for the Inverse Gaussian distribution and the Normal, after logarithmic transformation, distribution.

These will be the two pragmatic approaches.

Pragmatic in the sense that they do not depend on a specification for the probability distribution of the claim cost for a single claim and that they use the smoothing effect of aggregating.

The use of parametric forms such as (4.1) and (4.2) is in the spirit of the generalized linear model methodology of NELDER and WEDDERBURN (1972).

Other specifications are possible, such as linear ones:

$$\mu_r = \mathbf{x}_r' \boldsymbol{\theta}; \quad \phi_r = \mathbf{z}_r' \boldsymbol{\eta}. \tag{4.3}$$

or power transformations of (4.3). However, (4.1) and (4.2) have the logical property that μ_r and ϕ_r are always positive, an advantage compared with (4.3).

Furthermore, (4.1) is natural conjugate for the normalizing transformation (3.1), which forms a reason of simplicity. Finally combining a loglinear Poisson distribution, with a Gamma or Inverse Gaussian distribution with loglinear mean, gives a Compound Poisson distribution, also with loglinear mean. This forms a reason of mathematical beauty.

The following matrix-notation will be used:

$$\begin{aligned} \mathbf{N} &= \text{diag } (n_1, \ldots\ldots, n_R) \;, && \text{R} \times \text{R} \\ \mathbf{\Phi} &= \text{diag } (\phi_1, \ldots\ldots, \phi_R) \;, && \text{R} \times \text{R} \\ \mathbf{X} &= [\mathbf{x}_1, \ldots\ldots, \mathbf{x}_R]' \;, && \text{R} \times \text{K} \\ \mathbf{Z} &= [\mathbf{z}_1, \ldots\ldots, \mathbf{z}_R]' \;, && \text{R} \times \text{L} \end{aligned} \tag{4.4}$$

where n_r denotes the number of claims indexed r.

The first column of $\mathbf{X}$ and $\mathbf{Z}$ are equal to $\boldsymbol{\iota}$, a vector with all elements equal to 1; this is the so-called constant term.

We assume that the K × K matrix $\mathbf{X}'\mathbf{N}\mathbf{X}$ and L × L matrix $\mathbf{Z}'\mathbf{N}\mathbf{Z}$ are non-singular and for asymptotic analysis we need that the elements of $\mathbf{N}$, being the design of the sample, grow in such a way that:

$$\lim \mathbf{N}/(tr\mathbf{N}) = \bar{\mathbf{N}} \tag{4.5}$$

exists, leaving $\mathbf{X}'\bar{\mathbf{N}}\mathbf{X}$ and $\mathbf{Z}'\bar{\mathbf{N}}\mathbf{Z}$ non-singular.

As tr denotes the trace of a matrix, being the sum of the diagonal elements and $\mathbf{N}$ is diagonal, we have:

$$\boldsymbol{\iota}'\bar{\mathbf{N}}\boldsymbol{\iota} = tr\bar{\mathbf{N}} = 1. \tag{4.6}$$

The norming induced by (4.5) is not essential, however.

5. LOGLINEAR MODELLING WITH THE INVERSE GAUSSIAN DISTRIBUTION

The probability density function for total claim costs y_r with n_r claims reads as follows:

$$f(y_r \mid n_r\mu_r, n_r\phi_r) = n_r \left(\frac{\mu_r\phi_r}{2\pi y_r^3}\right)^{1/2} \exp\left\{-\tfrac{1}{2}n_r\phi_r\left[\frac{y_r}{n_r\mu_r} + \frac{n_r\mu_r}{y_r} - 2\right]\right\}. \tag{5.1}$$

Substituting the loglinear parametrizations given in (4.1) and (4.2) and forming the logarithm of the likelihood function gives[2]:

$$\begin{aligned} \log L = \text{const} \; &+ \tfrac{1}{2}\boldsymbol{\iota}'\mathbf{X}\boldsymbol{\theta} + \tfrac{1}{2}\boldsymbol{\iota}'\mathbf{Z}\boldsymbol{\eta} - \tfrac{1}{2}\Sigma y_r \exp(\mathbf{z}_r'\boldsymbol{\eta} - \mathbf{x}_r'\boldsymbol{\theta}) + \\ &- \tfrac{1}{2}\Sigma n_r^2 y_r^{-1} \exp(\mathbf{x}_r'\boldsymbol{\theta} + \mathbf{z}_r'\boldsymbol{\eta}) + \Sigma n_r \exp(\mathbf{z}_r'\boldsymbol{\eta}). \end{aligned} \tag{5.2}$$

[2] All summations run from $r = 1$ to $r = R$.

Differentiation with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ gives:

$$\frac{\partial \log L}{\partial \boldsymbol{\theta}} = \tfrac{1}{2}\mathbf{X}'\boldsymbol{\iota} + \tfrac{1}{2}\Sigma y_r \exp(\mathbf{z}_r'\boldsymbol{\eta} - \mathbf{x}_r'\boldsymbol{\theta})\mathbf{x}_r + - \Sigma n_r^2 y_r^{-1} \exp(\mathbf{x}_r'\boldsymbol{\theta} + \mathbf{z}_r'\boldsymbol{\eta})\mathbf{x}_r \tag{5.3}$$

$$\frac{\partial \log L}{\partial \boldsymbol{\eta}} = \tfrac{1}{2}\mathbf{Z}'\boldsymbol{\iota} - \tfrac{1}{2}\Sigma y_r \exp(\mathbf{z}_r'\boldsymbol{\eta} - \mathbf{x}_r'\boldsymbol{\theta})\mathbf{z}_r + - \tfrac{1}{2}\Sigma n_r^2 y_r^{-1} \exp(\mathbf{x}_r'\boldsymbol{\theta} + \mathbf{z}_r'\boldsymbol{\eta})\mathbf{z}_r + \Sigma n_r \exp(\mathbf{z}_r'\boldsymbol{\eta})\mathbf{z}_r. \tag{5.4}$$

Equating the elements of (5.3) and 5.4) to o defines the maximum likelihood equations for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

The next step is the derivation of the Hessian of the loglikelihood function.

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = -\tfrac{1}{2}\Sigma \phi_r \left(\frac{y_r}{\mu_r} + \frac{n_r^2 \mu_r}{y_r}\right) \mathbf{x}_r \mathbf{x}_r' \tag{5.5}$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\eta}'} = \tfrac{1}{2}\Sigma \phi_r \left(\frac{y_r}{\mu_r} - \frac{n_r^2 \mu_r}{y_r}\right) \mathbf{x}_r \mathbf{z}_r' \tag{5.6}$$

$$\frac{\partial^2 \log L}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} = -\tfrac{1}{2}\Sigma \phi_r \left(\frac{y_r}{\mu_r} + \frac{n_r^2 \mu_r}{y_r} - 2\right) \mathbf{z}_r \mathbf{z}_r'. \tag{5.7}$$

Although (5.5.) and (5.7) are negative definite matrices, the Hessian of the loglikelihood function is not negative definite for all values of $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. This implies that the loglikelihood function is not concave in the whole parameter space.

But it is concave in $\boldsymbol{\theta}$ conditionally on $\boldsymbol{\eta}$ and concave in $\boldsymbol{\eta}$ conditionally on $\boldsymbol{\theta}$.

So, (5.2) is a well-behaved function in $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, which can be maximized by a zig-zag iterative procedure such as set out in OBERHOFER and KMENTA (1974).

That is: maximize (5.2) with respect to $\boldsymbol{\theta}$ conditionally on an initial value of $\boldsymbol{\eta}$, then maximize with respect to $\boldsymbol{\eta}$ conditionally on the maximizing value of $\boldsymbol{\theta}$, then maximize with respect to $\boldsymbol{\theta}$ conditionally on the maximizing value of $\boldsymbol{\eta}$, etc., until convergence.

Minus the expectation of the Hessian gives the information matrix, whose inverse forms the Rao-Cramér lower bound for the covariance matrix of the maximum likelihood estimators for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$.

This information matrix can be written as:

$$-E\left(\frac{\partial^2 \log L}{\partial \begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\eta}\end{bmatrix} \partial \begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\eta}\end{bmatrix}'}\right) = \left[\begin{array}{c|c} \mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X} + \tfrac{1}{2}\mathbf{X}'\mathbf{X} & \tfrac{1}{2}\mathbf{X}'\mathbf{Z} \\ \hline \tfrac{1}{2}\mathbf{Z}'\mathbf{X} & \tfrac{1}{2}\mathbf{Z}'\mathbf{Z} \end{array}\right] \tag{5.8}$$

Applying a well-known inversion formula for partitioned matrices—see for instance THEIL (1971, p. 18) or RAO (1973, p. 33)—we get for the inverse of (5.8.):

$$(5.9)\quad V\begin{pmatrix}\hat{\boldsymbol{\theta}}\\ \hat{\boldsymbol{\eta}}\end{pmatrix}=\left[\begin{array}{c|c} V(\hat{\boldsymbol{\theta}}) & -V(\hat{\boldsymbol{\theta}})\,\mathbf{X}'\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1} \\ \hline -(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{X}\,V(\hat{\boldsymbol{\theta}}) & 2\,(\mathbf{Z}'\mathbf{Z})^{-1}+(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\,V(\hat{\boldsymbol{\theta}})\mathbf{X}'\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\end{array}\right]$$

where $V(\hat{\boldsymbol{\theta}})$ is given by:

$$(5.10)\qquad V(\hat{\boldsymbol{\theta}})=(\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X}+\tfrac{1}{2}\mathbf{X}'[\mathbf{I}-\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}']\,\mathbf{X})^{-1}.$$

It is interesting and important to study the loss of information due to aggregation in the estimation of $\boldsymbol{\theta}$; the elements of $\boldsymbol{\eta}$ being nuisance parameters. Forming the likelihood function for the original data shows that we only need one additional aggregate: the sum of the reciprocals of the claim costs.

Without aggregation the information matrix results in [3]:

$$(5.11)\qquad -E\left(\frac{\partial^2 \log L}{\partial\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\eta}\end{bmatrix}\partial\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\eta}\end{bmatrix}'}\right)=\left[\begin{array}{c|c} \mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X}+\tfrac{1}{2}\mathbf{X}'\mathbf{N}\mathbf{X} & \tfrac{1}{2}\mathbf{X}'\mathbf{N}\mathbf{Z} \\ \hline \tfrac{1}{2}\mathbf{Z}'\mathbf{N}\mathbf{X} & \tfrac{1}{2}\mathbf{Z}'\mathbf{N}\mathbf{Z}\end{array}\right]$$

The covariance matrix for $\hat{\boldsymbol{\theta}}$ becomes now:

$$(5.12)\qquad V(\hat{\boldsymbol{\theta}})=(\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X}+\tfrac{1}{2}\mathbf{X}'[\mathbf{N}-\mathbf{N}\mathbf{Z}\,(\mathbf{Z}'\mathbf{N}\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{N}]\mathbf{X})^{-1}.$$

So, we should compare (5.10) and (5.12).

First of all, we see that in case $\mathbf{X}=\mathbf{Z}$ or even more general, in case that the column(s) of $\mathbf{X}$ can be written as a linear combination of the columns of $\mathbf{Z}$:

$$(5.13)\qquad \mathbf{X}=\mathbf{Z}\mathbf{C},$$

both (5.10) and (5.12) boil down to:

$$(5.14)\qquad V(\hat{\boldsymbol{\theta}})=(\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X})^{-1}$$

implying no loss of efficiency in the estimation of $\boldsymbol{\theta}$.

But if (5.13) is not valid, there will be a loss of efficiency, even asymptotically. To see this, consider the asymptotic forms of (5.10) and (5.12):

$$(5.15)\qquad \lim\ (tr\mathbf{N})\ (\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X}+\tfrac{1}{2}\mathbf{X}'\,[\mathbf{I}-\mathbf{Z}\,(\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}']\mathbf{X})^{-1}=(\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}\mathbf{X})^{-1}$$

$$(5.16)\qquad \lim\ (tr\mathbf{N})\ (\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X}+\tfrac{1}{2}\mathbf{X}'\,[\mathbf{N}-\mathbf{N}\mathbf{Z}\,(\mathbf{Z}'\mathbf{N}\mathbf{Z})^{-1}\,\mathbf{Z}'\mathbf{N}]\,\mathbf{X})^{-1}=$$
$$(\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}\mathbf{X}+\tfrac{1}{2}\mathbf{X}'\,[\bar{\mathbf{N}}-\bar{\mathbf{N}}\mathbf{Z}(\mathbf{Z}'\bar{\mathbf{N}}\mathbf{Z})^{-1}\,\mathbf{Z}'\bar{\mathbf{N}}]\mathbf{X})^{-1}.$$

[3] This is most easily seen by putting $\mathbf{N}=\mathbf{I}$ in (5.8), changing the interpretation of $\mathbf{X}$ and $\mathbf{Z}$ accordingly, taking similar rows of $\mathbf{X}$ and $\mathbf{Z}$ together and introducing $\mathbf{N}$ again, resulting in (5.11).

The difference between the inverses of (5.16) and (5.15) is equal to:

$$\tfrac{1}{2}\mathbf{X}'[\bar{\mathbf{N}} - \bar{\mathbf{N}}\mathbf{Z}(\mathbf{Z}'\bar{\mathbf{N}}\mathbf{Z})^{-1}\,\mathbf{Z}'\bar{\mathbf{N}}]\mathbf{X}$$

a positive semidefinite matrix. This implies the difference between (5.15) and (5.16) also to be positive semidefinite, implying the loss of efficiency.

A very popular choice for $\mathbf{Z}$ will be $\mathbf{Z} = \boldsymbol{\iota}$. For this case, the loss of efficiency is most easily studied by calculating the ratio of the generalized variances of (5.15) and (5.16). The generalized variance is equal to the determinant of the covariance matrix and this is equal to the product of its latent roots. Substituting $\mathbf{Z} = \boldsymbol{\iota}$, $\boldsymbol{\Phi} = \phi\mathbf{I}$ and using $\boldsymbol{\iota}'\bar{\mathbf{N}}\boldsymbol{\iota} = 1$ and the fact that the first column of $\mathbf{X}$ equals $\boldsymbol{\iota}$, this ratio results in:

$$\begin{aligned}(5.17)\qquad & |\phi\mathbf{X}'\bar{\mathbf{N}}\mathbf{X}| \;|\; (\phi + \tfrac{1}{2})\mathbf{X}'\bar{\mathbf{N}}\mathbf{X} - \tfrac{1}{2}\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\iota}\,\boldsymbol{\iota}'\bar{\mathbf{N}}\mathbf{X}|^{-1} = \\ & \phi^{K}\,(\phi + \tfrac{1}{2})^{-K} \,|\, \mathbf{I} - (2\phi + 1)^{-1}\,(\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1}\,\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\iota}\,\boldsymbol{\iota}'\bar{\mathbf{N}}\mathbf{X}\,|^{-1} = \\ & \phi^{K}\,(\phi + \tfrac{1}{2})^{-K}\,\{1 - (2\phi + 1)^{-1}\,\boldsymbol{\iota}'\bar{\mathbf{N}}\mathbf{X}\,(\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1}\,\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\iota}\}^{-1} = \left(\frac{\phi}{\phi + \frac{1}{2}}\right)^{K-1}\end{aligned}$$

This efficiency ratio depends on the dimensionality K.

Calculating the ratio of the traces of 5.15) and (5. 16) does not depend on this dimensionality. The trace of a matrix is equal to the sum of its latent roots. So, for efficiency purposes, we can consider the arithmetic mean of these latent roots. Applying this line of reasoning to (5.17) implies the geometric mean of the latent roots and the efficiency ratio is transformed into:

$$(5.18)\qquad \left(\frac{\phi}{\phi + \frac{1}{2}}\right)^{1-K^{-1}} \approx \frac{\phi}{\phi + \frac{1}{2}} \quad \text{for K large.}$$

The larger the coefficient of variation, given by $\phi^{-\frac{1}{2}}$, the smaller (5.18) will be, implying a large loss of efficiency due to aggregation.

6. LINEAR MODELS WITH LOGLINEAR HETEROSCEDASTICITY

Now we will consider the logarithmic transformation of section 3 in relation with the parametrizations (4.1) and (4.2).

Denoting $w_r = \log(y_r/n_r)$, we have approximately:

$$(6.1)\qquad E(w_r) = \mathbf{x}_r'\boldsymbol{\theta} - (2n_r\phi_r)^{-1}$$

$$(6.2)\qquad V(w_r) = (n_r\phi_r)^{-1}.$$

If the second term in (6.1) is deleted [4], we have the model as analyzed by HARVEY (1976).

[4] And both sides are multiplied by $\sqrt{n_r}$.

Introducing a dummy variable d, taking on the values 0 and 1, both models can be represented by replacing (6.1):

$$E(w_r) = \mathbf{x}_r'\boldsymbol{\theta} - d(2n_r\phi_r)^{-1}. \tag{6.3}$$

This will enable us to see more clearly the effect of adopting the model induced by $d = 0$ against $d = 1$.

Assuming w_r to be (approximately) normally distributed, the loglikelihood function results in:

$$\log L = \text{const} + \tfrac{1}{2}\,\boldsymbol{\iota}'\mathbf{Z}\boldsymbol{\eta} - \tfrac{1}{2}\Sigma\, n_r \exp(\mathbf{z}_r'\boldsymbol{\eta})\,[w_r - \mathbf{x}_r'\boldsymbol{\theta}]^2 + \tfrac{1}{2}d\boldsymbol{\iota}'\mathbf{X}\boldsymbol{\theta} - \tfrac{1}{8}\, d\Sigma\, n_r^{-1} \exp(-\mathbf{z}_r'\boldsymbol{\eta}) \tag{6.4}$$

Differentiating with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$ gives:

$$\frac{\partial \log L}{\partial\, \boldsymbol{\theta}} = \Sigma n_r \exp(\mathbf{z}_r'\boldsymbol{\eta})\,[w_r - \mathbf{x}_r'\boldsymbol{\theta}]\,\mathbf{x}_r + \tfrac{1}{2}\, d\, \mathbf{X}'\boldsymbol{\iota} \tag{6.5}$$

$$\frac{\partial \log L}{\partial\, \boldsymbol{\eta}} = \tfrac{1}{2}\mathbf{Z}'\boldsymbol{\iota} - \tfrac{1}{2}\Sigma\, n_r \exp(\mathbf{z}_r'\boldsymbol{\eta})\,[w_r - \mathbf{x}_r'\boldsymbol{\theta}]^2\,\mathbf{z}_r + \tfrac{1}{8}\, d\, \Sigma\, n_r^{-1} \exp(-\mathbf{z}_r'\,\boldsymbol{\eta})\,\mathbf{z}_r. \tag{6.6}$$

Equating the elements of (6.5) and (6.6) equal to 0 defines the maximum likelihood equations. Conditionally on $\boldsymbol{\eta}$, the solution for $\boldsymbol{\theta}$ exists in closed form:

$$\boldsymbol{\theta} = [\Sigma n_r \exp(\mathbf{z}_r'\,\boldsymbol{\eta})\,\mathbf{x}_r\mathbf{x}_r']^{-1}\,[\Sigma n_r \exp(\mathbf{z}_r'\,\boldsymbol{\eta})\,w_r\mathbf{x}_r + \tfrac{1}{2}d\,\mathbf{X}'\boldsymbol{\iota}] = (\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X})^{-1}\,[\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{w} + \tfrac{1}{2}d\mathbf{X}'\,\boldsymbol{\iota}] \tag{6.7}$$

where $\mathbf{w} = (w_1, \ldots\ldots, w_r)'$.

The submatrices of the Hessian are as follows:

$$\frac{\partial^2 \log L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -\,\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X} \tag{6.8}$$

$$\frac{\partial^2 \log L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\eta}'} = \Sigma n_r\phi_r[w_r - \mathbf{x}_r'\boldsymbol{\theta}]\,\mathbf{x}_r\mathbf{z}_r' \tag{6.9}$$

$$\frac{\partial^2 \log L}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}'} = -\,\tfrac{1}{2}\Sigma n_r\phi_r[w_r - \mathbf{x}_r'\boldsymbol{\theta}]^2\,\mathbf{z}_r\mathbf{z}_r' - \tfrac{1}{8}\, d\, \mathbf{Z}'(\mathbf{N}\boldsymbol{\Phi})^{-1}\mathbf{Z}. \tag{6.10}$$

The matrices (6.8) and 6.10) are negative definite, whereas the Hessian is not. So, the very same situation applies as in case of the Inverse Gaussian distribution in section 5.

Due to the fact of a closed form expression for $\boldsymbol{\theta}$, given by (6.7), the zig-zag iterative maximizing procedure is even more simple in this case, however.

The information matrix takes the following form:

$$(6.11) \qquad -E\left(\frac{\partial^2 \log L}{\partial\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\eta}\end{bmatrix}\partial\begin{bmatrix}\boldsymbol{\theta}\\ \boldsymbol{\eta}\end{bmatrix}'}\right)=\left[\begin{array}{c|c}\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X} & \tfrac{1}{2}d\mathbf{X}'\mathbf{Z}\\ \hline \tfrac{1}{2}d\mathbf{Z}'\mathbf{X} & \tfrac{1}{2}\mathbf{Z}'\mathbf{Z}+\tfrac{1}{4}d\mathbf{Z}'(\mathbf{N}\boldsymbol{\Phi})^{-1}\mathbf{Z}\end{array}\right]$$

Applying partitioned matrix invertation gives:

$$(6.12) \qquad V(\hat{\boldsymbol{\theta}})=(\mathbf{X}'\mathbf{N}\boldsymbol{\Phi}\mathbf{X}-d\mathbf{X}'\mathbf{Z}[2\mathbf{Z}'\mathbf{Z}+d\mathbf{Z}'\ (\mathbf{N}\boldsymbol{\Phi})^{-1}\mathbf{Z}]^{-1}\ \mathbf{Z}'\mathbf{X})^{-1}$$

and the asymptotic form of (6.12) results in:

$$(6.13) \qquad \lim\ (tr\mathbf{N})\ V(\hat{\boldsymbol{\theta}})=(\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}\mathbf{X})^{-1}.$$

Whether $d=0$ or $d=1$ does not matter for large $\mathbf{N}$, which does not surprise us, as both models converge to each other for growing $\mathbf{N}$.

For the model induced by $d=0$, HARVEY (1976) contains an analysis of a two-step and three-step estimator for $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$, which are asymptotically efficient and which result in an economy of computational effort.

7. FINAL REMARKS

7.1. *A Simple Estimator*

Consider again (6.3) with $d=0$, multiplied by $n_T^{1/2}$ and rewritten in the following form:

$$(7.1) \qquad \mathbf{N}^{1/2}\mathbf{w}=\mathbf{N}^{1/2}\mathbf{X}\boldsymbol{\theta}+\boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon}$ is a disturbance term, approximately normally distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}^{-1}$.

Estimating $\boldsymbol{\theta}$ by ordinary least squares, results in the following estimator for $\boldsymbol{\theta}$:

$$(7.2) \qquad \tilde{\boldsymbol{\theta}}=(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\ \mathbf{X}'\mathbf{N}\mathbf{w}.$$

This estimator is approximately normally distributed with mean and covariance matrix given by:

$$(7.3) \qquad \begin{aligned} E(\tilde{\boldsymbol{\theta}}) &\doteq \boldsymbol{\theta}\\ V(\tilde{\boldsymbol{\theta}}) &\doteq (\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\ \mathbf{X}'\mathbf{N}\boldsymbol{\Phi}^{-1}\ \mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}. \end{aligned}$$

If the elements of $\mathbf{N}$ grow large, the approximate nature disappears. The asymptotic covariance matrix for $\tilde{\boldsymbol{\theta}}$ becomes $(\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1}\ \mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}^{-1}\mathbf{X}(\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1}$, which we should compare with $(\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}\mathbf{X})^{-1}$, given by (6.13).

The difference of these asymptotic matrices can be written as:

$$(7.4) \qquad (\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1}\,\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}^{-1}\,\mathbf{X}(\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1} - (\mathbf{X}'\bar{\mathbf{N}}\boldsymbol{\Phi}\mathbf{X})^{-1} = (\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1}\,\mathbf{A}'\boldsymbol{\Phi}^{-1}\,[\mathbf{I} - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\,\mathbf{A}']\,\boldsymbol{\Phi}^{-1}\mathbf{A}(\mathbf{X}'\bar{\mathbf{N}}\mathbf{X})^{-1},$$

where $\mathbf{A} = (\bar{\mathbf{N}}\boldsymbol{\Phi})^{1/2}\mathbf{X}$. The matrix between brackets is idempotent with rank (R-K). So, (7.4) is positive semidefinite, making $\tilde{\boldsymbol{\theta}}$ an inefficient estimator.

WATSON (1967, pp. 1684-1687) compares such matrices by deriving a lower bound for the ratio of the generalized variances.

Making use of the inequality of Hadamard as well as applying the inequality of Kantorovich, he finds a lower bound for this efficiency ratio, which can be written as:

$$(7.5) \qquad \{4\rho(1+\rho)^{-2}\}^{K}$$

where $\rho \geqq 1$ denotes the ratio of the largest to the smallest latent root of $\boldsymbol{\Phi}$. As $\boldsymbol{\Phi}$ is diagonal, the latent roots are given by the diagonal elements.

Adopting the geometric mean latent root criterion instead of the generalized variance, the lowerbound is obtained by deleting the exponent K in (7.5).

This gives a more scalar interpretable idea about the possible loss of efficiency for various values of ρ. Besides the fact that (7.2) can be a good estimator in practice, it will also serve as a convenient starting value for the more complex approaches, given by maximizing (5.2) and (6.4).

7.2. *A Simple Test*

It is also possible to test whether $\boldsymbol{\Phi} = \phi\mathbf{I}$ or not. For the loglinear specification of ϕ_r, GODFREY (1978) derived a Lagrange multiplier test, which only depends on the least squares residuals and which is asymptotically equivalent to the likelihood ratio test.

This test runs as follows. Returning to (7.1) and (7.2), $\boldsymbol{\epsilon}$ can be estimated by

$$(7.6) \qquad \tilde{\boldsymbol{\epsilon}} = \mathbf{N}^{1/2}\,\mathbf{w} - \mathbf{N}^{1/2}\,\mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{N}\mathbf{w}.$$

The variance of the elements of $\boldsymbol{\epsilon}$, which is equal to ϕ^{-1}, can be estimated by

$$(7.7) \qquad \frac{\tilde{\boldsymbol{\epsilon}}'\tilde{\boldsymbol{\epsilon}}}{\mathrm{R}} = \frac{\mathbf{w}'[\mathbf{N} - \mathbf{N}\mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^{-1}\mathbf{X}'\mathbf{N}]\mathbf{w}}{\mathrm{R}}.$$

The next step is to form the vector $\mathbf{q}$ with typical element given by the square of the typical element of (7.6), divided by (7.7) and subtracting 1:

$$(7.8) \qquad q_r = \left(\frac{\mathrm{R}\,\tilde{\epsilon}_r^2}{\tilde{\boldsymbol{\epsilon}}'\tilde{\boldsymbol{\epsilon}}} - 1\right).$$

The test statistic can now be written as:

$$T = \tfrac{1}{2}\mathbf{q}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{q}. \tag{7.9}$$

and has a Chi-square distribution with (L-1) degrees of freedom as limiting distribution under the hypothesis that $\mathbf{\Phi}$ is a scalar matrix, making (7.2) an efficient estimator.

7.3. *Outliers.*

What about the practice of giving observations, taking on large values, a special treatment by putting them apart? I think that we should not do this. Outlying observations should be identified with the help of the model.

Now, by forming sample means, we construct aggregate observations, perhaps outlying, but less outlying than the constituting parts. Adopting the Inverse Gaussian distribution, which can be very skew, or applying the logarithmic transformation, there is room for the data to be "outlying" in the sense that they are not distributed symmetrically around the central tendency. In order to identify outliers, we should apply maximum likelihood estimation and form *standardized residuals* and study these.

The outlying nature of these standardized residuals depends also on **N**, **X** and **Z** and not only on the value of the observations itself.

For the framework of the standard linear model induced by (6.3) with $d = 0$ and $\mathbf{\Phi}$ a scalar matrix, the analysis of residuals is set out in THEIL (1971, ch. 5).

The probability distribution for the *standardized* residuals was derived by ELLENBERG (1973) and shown to be the Inverted-Student distribution. This supplies us with sound methods to analyze outliers at least for the standard linear model.

7.4. *Bayesian Estimation and Inequality Restrictions.*

As regards estimation, this paper was written around the maximum likelihood method. Bayesian estimation is possible too, however. Bayesian inference is easy if integration can be performed analytically. If this is not possible, numerical procedures are called for, which are cumbersome, especially for large parameter dimensions.

Monte Carlo integration appears to be feasable, however, as set out in a clear way by KLOEK and VAN DIJK (1978) in the context of Bayesian estimation.

The use of numerical integration creates the freedom to specify prior distributions without forcing these to take on forms which facilitate analytical integration.

Consider now homogeneous risk groups in a certain line of insurance, for instance automobile insurance. In this field it may be possible to specify a

lower bound for the population mean of the claim cost distribution for the best risk group as well as an upper bound in case of the worst risk group. Furthermore, it may be possible to rank certain elements of $\boldsymbol{\theta}$ in size [5]. This *a priori* knowledge can be formalized as a set of linear inequality restrictions on $\boldsymbol{\theta}$, forming a convex subset of the natural parameter space. Adopting a uniform prior distribution on this convex subset completes the specification of an informative prior distribution, which easily can be used in a Monte Carlo integration. The incorporation of such *a priori* knowledge may guide parameter estimation and shrink the highest posterior density intervals.

This will especially apply to multicollinear situations, making matrices such as (5.8) and (6.11) ill-conditioned, resulting in likelihood functions which are relatively flat in certain directions.

REFERENCES

BERG, P. TER (1980). On the Loglinear Poisson and Gamma Model, *Astin Bulletin* **11**, 35-40.

ELLENBERG, J. H. (1973). The Joint Distribution of the Standardized Least Squares Residuals from a General linear Regression, *Journal of the American Statistical Association* **68**, 941-943.

FOLKS, J. L., and R. S. CHHIKARA (1978). The Inverse Gaussian Distribution and its Statistical Application—A Review (with Discussion), *Journal of the Royal Statistical Society B* **40**, 263-289.

GODFREY, L. G. (1978). Testing for Multiplicative Heteroscedasticity, *Journal of Econometrics* **8**, 227-236.

HADWIGER, H. (1940a). Natürliche Ausscheidefunktionen für Gesamtheiten und die Lösung der Erneuerungsgleichung, *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker* **40**, 31-39.

HADWIGER, H. (1940b). Eine analytische Reproduktionsfunktion für biologische Gesamtheiten, *Skandinavisk Aktuarietidskrift* **23**, 101-113.

HADWIGER, H. (1942). Wahl einer Näherungsfunktion für Verteilungen auf Grund einer Funktionalgleichung, *Blätter für Versicherungsmathematik* **5**, 345-352.

HADWIGER, H., and W. RUCHTI (1941). Darstellung der Fruchtbarkeit durch eine biologische Reproduktionsformel, *Archiv für mathematische Wirtschafts- und Sozialforschung* **7**, 30-34.

HARVEY, A. C. (1976). Estimating Regression Models with Multiplicative Heteroscedasticity, *Econometrica* **44**, 461-465.

KLOEK, T., and H. K. VAN DIJK (1978). Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo, *Econometrica* **46**, 1-19. Reprinted in A. ZELLNER, ed. (1980). *Bayesian Analysis in Econometrics and Statistics*. Amsterdam: North-Holland Publishing Company.

NELDER, J. A., and R. W. M. WEDDERBURN (1972). Generalized Linear Models, *Journal of the Royal Statistical Society A* **135**, 370-384.

OBERHOFER, W., and J. KMENTA (1974). A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models, *Econometrica* **42**, 579-590.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: John Wiley & Sons, Inc.

SCHRÖDINGER, E. (1915). Zur Theorie der Fall- und Steigversuche an Teilchen mit Brownscher Bewegung, *Physikalische Zeitschrift* **16**, 289-295.

[5] Similar restrictions may apply to claim frequencies in relation to the loglinear Poisson model.

SEAL, H. L. (1969). *Stochastic Theory of a Risk Business*. New York: John Wiley & Sons, Inc.

SEAL, H. L. (1978). From Aggregate Claims Distribution to Probability of Ruin, *Astin Bulletin* **10**, 47-53.

THEIL, H. (1971). *Principles of Econometrics*. New York: John Wiley & Sons, Inc.

TWEEDIE, M. C. K. (1957). Statistical Properties of Inverse Gaussian Distributions, *Annals of Mathematical Statistics* **28**, 362-377 and 696-705.

WALD, A. (1947). *Sequential Analysis*. New York: John Wiley & Sons, Inc.

WATSON, G. S. (1967). Linear Least Squares Regression, *Annals of Mathematical Statistics* **38**, 1679-1699.

WHITMORE, G. A., and M. YALOVSKI (1978). A Normalizing Logarithmic Transformation for Inverse Gaussian Random Variables, *Technometrics* **20**, 207-208.