NONLINEAR NORMAL CORRELATED LOSS ARRAY

Integrated Financial Modeling of Portfolio and Runoff Risk

ASTIN Topic: *Dynamic Financial Analysis*

TER BERG, PETER

Posthuma Partners
Badhuisweg 11
2587 CA Den Haag
Netherlands

peter_ter_berg@posthuma-partners.nl

Abstract

The runoff triangle is viewed as an incompletely observed rectangular array which also may comprise future loss years. Together with multivariate normality, a parametric structure for the means and covariances completes the stochastic specification. The payment pattern is a core aspect. After maximum likelihood estimation and incorporation of estimation uncertainty, the predictive distribution of the unobserved entities is derived. This distribution is conditional on the observed history and marginal with respect to the estimated parameters. Besides predictive means, also percentiles can be derived. The latter become important when objectifying prudent provisions as well as solvency margins.

Keywords

incurred not settled, covered not incurred, autocorrelation, (negative) multinomial, distributed lag, selection matrix, maximum likelihood, estimation uncertainty, prudent provision, solvency margin.

NONLINEAR NORMAL CORRELATED LOSS ARRAY

Integrated Financial Modeling of Portfolio and Runoff Risk

By

PETER TER BERG

Posthuma Partners, The Hague, Netherlands

1. INTRODUCTION

Risk is part and parcel of the insurance business and presents itself to insurers in different ways. First of all there is the insurance risk they incur on running policies and claims still waiting to be settled. Both of these types of risk are related and have financial consequences. They reflect in particular on risk provision and the insurer's solvency position. Modeling them together in their entirety gives us the opportunity to assess both the adequacy of risk provisions and the solvency position generally.

From an actuarial viewpoint it is important to bring into focus the risk profile attached to the two risk components: the profile of insurance commitments undertaken by the insurer and the profile of claims still in the pipeline. This integrated view of past (incurred not settled) and future (covered not incurred) displays corresponding matters of prudential provision and solvency margin in a neat joint way. Indeed such matters as an actuarial provision for premium deficiency or the value of future underwriting business may be explicated.

Unlike other models which usually only make an isolated financial assessment of claims provision, the Integrated Financial Modeling (IFM) encompasses the risk inherent in the insurance proces as a whole and allows us to describe it. The analysis of loss years and the associated loss provisions usually is based on runoff triangles, which are aggregated representations of claims experience. However, we prefer to consider a rectangular loss array over a runoff triangle. Indeed, by viewing a rectangular array instead of a triangular one, we are able to encompass future loss years for which no entries have been observed as yet. This allows a symmetric treatment of aggregate loss by loss year, whether the loss years are lying in the past or ahead in the future. Schematically, the situation is as follows:

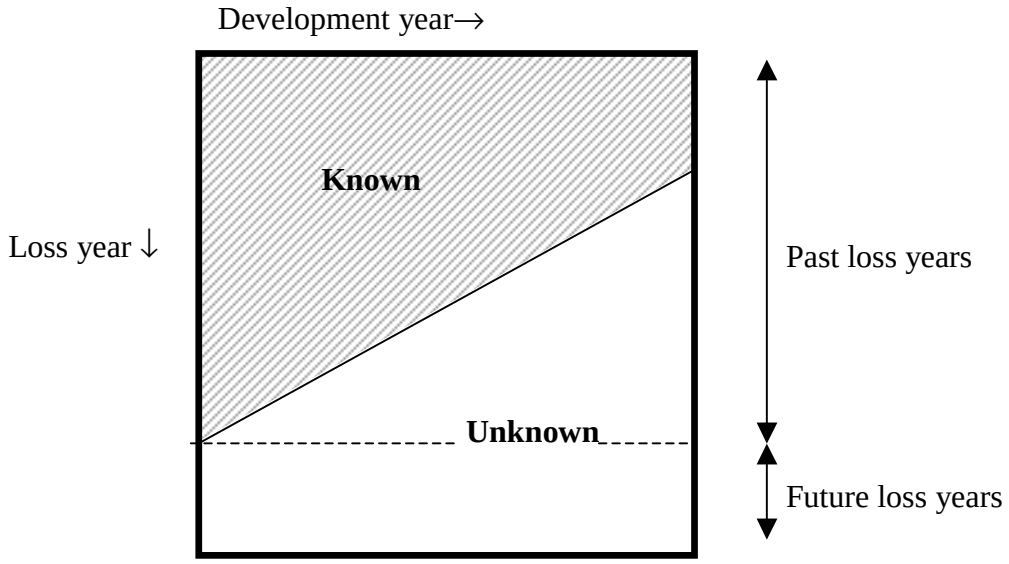

Figure 1 Rectangular loss array

The elements of this loss array are viewed as arising from a multivariate normal distribution. We do not belief this normality assumption to hold in the strict sense, but the impact of its violation need not be that serious. After all, the entries of the loss array are aggregates for which usually a central limit law holds. In exchange for this simplicity, we get convenient freedom to model a covariance structure for these entries.

By adding a measure of risk exposure, like gross earned premium, to the loss array the loss years are evaluated as a time series of loss ratios which fluctuate around a long term average. Besides that, development patterns are incorporated in the model and correlations between loss years and development years are taken into account. This means that if entries deviate from the value that was expected beforehand, this is taken into account for predicting neighbouring entries. These covariances are important. Not as much for efficient parameter estimation but the more so for prediction.

Formulating a parametric model for the expected values and covariances, the stochastic law with multivariate normality has been completed. An important aspect of the model is a smooth payment pattern which may require some multimodality. The remainder of the paper then follows the standard steps for maximum likelihood estimation and prediction in a multivariate normal environment. Parameter estimation uncertainty is explicitly accounted for.

The plan of the paper runs as follows. After recognizing a rectangular loss array as a matrix, it is vectorized such that selection matrices identify observed elements as well as elements to be predicted. Then inspired by time series analysis and the (negative) multinomial distribution the vector of means and the covariance matrix are derived. Then the payment pattern is viewed as a process in continuous time. This leads to a rather sophisticated discretized payment pattern. Its maximum likelihood estimation generates a delicate numerical optimization problem. Finally, the predictive probability distribution is set out for linear aggregates of the loss array. Here, estimation uncertainty is integrated out with an appeal to Bayesian statistical inference. So, this predictive distribution is conditional on the historical observations and marginal with respect to the unknown parameters.

2. MATRIX EMBEDDING OF THE INCOMPLETE LOSS ARRAY

We consider a rectangular data structure of aggregate nominal paid loss, where the implicit continuous time has been discretized. For ease we take these as years but any other interval may apply. So, we have a matrix **Y** with *m* rows for the loss years and *n* columns for the settlement durations. In the typical insurance application we have a triangular subset of **Y** as observations. In case of missing observations the triangle becomes a trapezium or some blurred variation. Furthermore we have a column vector **w** with volume measures for the *m* loss periods. For ease we may imagine **w** to contain gross earned premiums. So, we have:

$$\mathbf{Y} = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}$$

In a natural display the rows and columns of **Y** will be ordered chronologically, with equal length for the loss period and the settlement duration. In the settlement direction, the equal length assumption is not necessary, however. Indeed, if the final column comprises the distant upper tail of the settlement durations, this will aggregate an infinite length upper tail. In that case the *n*-th column of **Y** will have no observations. For prediction and provision purposes our later focus will be on a linear (discounted) combination of the unknown future elements of **Y** conditional on the known elements of **Y**.

We may arrange the *mn* elements of **Y** as a column vector by stacking its columns one beneath the other. This vector is denoted as vec**Y**. Next we may permute vec**Y** such that missing observations, known observations and future observations are put together. To this end we need an *mn*×*mn* permutation matrix **S** and arrive at:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_0 \\ \mathbf{S}_1 \\ \mathbf{S}_2 \end{bmatrix} \qquad \mathbf{S}\mathrm{vec}\mathbf{Y} = \begin{bmatrix} \mathbf{S}_0\mathrm{vec}\mathbf{Y} \\ \mathbf{S}_1\mathrm{vec}\mathbf{Y} \\ \mathbf{S}_2\mathrm{vec}\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \qquad \begin{matrix} r_0 \text{ missing elements} \\ r_1 \text{ known elements} \\ r_2 \text{ future elements} \end{matrix}$$

Adopting a normal distribution for vec**Y** and choosing a parametric model for its mean and covariance matrix, the next steps will be swift *in principle*. Parameter estimation through maximizing the likelihood function requires the normal density for $\mathbf{y}_1$ and prediction requires the normal density for $\mathbf{y}_2$ conditional on $\mathbf{y}_1$. The addition of estimation uncertainty is handled as a mixing operation.

3. MEAN AND COVARIANCE OF THE LOSS ARRAY

We will infer simple forms for the mean and covariance by viewing obvious candidates for the column vector of row totals and the row vector of column totals of **Y**. In the following the column vector ι has all its elements equal to 1, its order being clear from the context.

3.1 *Row Totals*

Let us form the vector **Y**ι which is, once observed, a time series on ultimate aggregate loss for the *m* loss years. Well-known approaches for the analysis of such a series can be borrowed from econometrics and time series analysis. For the mean a linear model suggests itself and the covariance matrix may be inspired by autoregressive and/or moving average models. The volumes **w** enter as a diagonal matrix **W** defined as:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_m \end{bmatrix} = \mathbf{w}_\Delta$$

This also introduces the suffix Δ notation due to THEIL (1983) to indicate the diagonal matrix reshaping of a column vector, which will be used later. We arrive at the following specification for the mean and the covariance:

$$E(\mathbf{Y}\iota) = \mathbf{W}\mathbf{X}\beta$$

$$V(\mathbf{Y}\iota) = \sigma^2\mathbf{W}\Omega\mathbf{W}$$

Here $\mathbf{X}\beta$ shapes the linear model through an $m \times k$ matrix $\mathbf{X}$ and k parameters in β. The $m \times m$ matrix Ω is generic for the ARMA form. Confining ourselves to AR forms, the case of first order autocorrelation corresponds with:

$$\Omega = \begin{bmatrix} 1 & \rho & \cdots & \rho^{m-1} \\ \rho & 1 & \ddots & \rho^{m-2} \\ \vdots & \ddots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \cdots & 1 \end{bmatrix}$$

Higher order autocorrelations are easily handled through the inverse of Ω, using results originating with SIDDIQUI (1958) as well as ANDERSON and MENTZ (1980). For the vector of ultimate loss ratios, the foregoing implies

$$E(\mathbf{W}^{-1}\mathbf{Y}\iota) = \mathbf{X}\beta \qquad V(\mathbf{W}^{-1}\mathbf{Y}\iota) = \sigma^2\Omega$$

When β boiles down to a scalar, $\mathbf{X}$ becomes the column vector ι and the the time series models a stationary loss ratio proces.

3.2 *Column Totals*

The row vector $\iota'\mathbf{Y}$ divided by its total gives the empirical form of the settlement duration in the form of a payment pattern. This payment pattern can be viewed as a probability vector $\pi \in \Re^n$. This leads to the mean:

$$E(\mathbf{Y}'\iota) \propto \pi \qquad \iota'\pi = 1$$

For the covariance matrix we borrow from the (negative) multinomial distribution. This leads to a general form:

$$V(\mathbf{Y}'\iota) \propto (\pi_\Delta + \tau\pi\pi') \qquad \tau \geq -1$$

The sign of the correlations coincides with the sign of τ. So when $\tau > 0$ large losses are indicative for other large losses, etc. This particular specification of the covariance matrix is insensitive for odd choices of the settlement duration intervals. Indeed let $\mathbf{G}$ be a grouping matrix . Then we have $\tilde{\pi} = \mathbf{G}\pi$ and

$$\mathbf{G}(\pi_\Delta + \tau\pi\pi')\mathbf{G}' = (\mathbf{G}\pi)_\Delta + \tau\mathbf{G}\pi(\mathbf{G}\pi)' = \tilde{\pi}_\Delta + \tau\tilde{\pi}\tilde{\pi}'$$

This property is important in order to handle a shifting upper tail.

3.3 *Joint Specification*

The foregoing specifications are compatible with the following form for the mean

$$E(\mathbf{Y}) = (\mathbf{W}\mathbf{X}\beta)\pi'$$

which emphasises the rank-1 structure of this expectation. Following chapter 2 in MAGNUS and NEUDECKER (1999) this can be rewritten, using the Kronecker product of a matrix, in vectorized form as:

$$E(\text{vec}\mathbf{Y}) = (\pi \otimes \mathbf{W}\mathbf{X})\beta$$

which emphasises the linearity in β. For the covariance matrix we get:

$$V(\text{vec}\mathbf{Y}) = \sigma^2(\pi_\Delta + \tau\pi\pi') \otimes \mathbf{W}\Omega\mathbf{W}$$

Observe the different treatment of the direction of loss time and settlement duration time, as compared with the chain-ladder method.

3.4 *Final Joint Specification*

Using the selection matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ we get for the mean of $\mathbf{y}_1$ and $\mathbf{y}_2$:

$$E\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{S}_1\\ \mathbf{S}_2\end{bmatrix} E(\text{vec}\mathbf{Y}) = \begin{bmatrix}\mathbf{S}_1(\pi \otimes \mathbf{W}\mathbf{X})\\ \mathbf{S}_2(\pi \otimes \mathbf{W}\mathbf{X})\end{bmatrix}\beta = \begin{bmatrix}\mathbf{Z}_1\\ \mathbf{Z}_2\end{bmatrix}\beta = \begin{bmatrix}\mathbf{Z}_1\beta\\ \mathbf{Z}_2\beta\end{bmatrix}$$

Observe that the auxiliary matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ depend on the parameters in π. Likewise we get for the covariance matrix:

$$V\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{S}_1\\ \mathbf{S}_2\end{bmatrix} V(\text{vec}\mathbf{Y}) \begin{bmatrix}\mathbf{S}_1\\ \mathbf{S}_2\end{bmatrix}' = \sigma^2\begin{bmatrix}\mathbf{C}_{11} & \mathbf{C}_{12}\\ \mathbf{C}_{21} & \mathbf{C}_{22}\end{bmatrix}$$

$$\mathbf{C}_{ij} = \mathbf{S}_i\left[(\pi_\Delta + \tau\pi\pi') \otimes \mathbf{W}\Omega\mathbf{W}\right]\mathbf{S}_j' \qquad i, j = 1,2$$

The auxiliary matrices $\mathbf{C}_{ij}$ depend, in addition to the parameters in π, also on τ and the parameters in Ω. We remark that, for prediction purposes, the selection matrix $\mathbf{S}_2$ may be collapsed to a grouping matrix or more general any matrix defining appropriate (discounted) aggregates. Likewise the selection matrix $\mathbf{S}_1$ may also have aspects of a grouping matrix when certain elements of $\mathbf{Y}$ are observed in grouped form. What matters in the foregoing are the linear transformation aspects.

3.5 *Conditional Specification*

Let us denote all unknown parameters to be estimated as ζ. The mean and covariance of $\mathbf{y}_2$ conditional on $\mathbf{y}_1$ and ζ are given as:

$$E(\mathbf{y}_2 \mid \mathbf{y}_1, \zeta) = \mathbf{Z}_2\beta + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{y}_1 - \mathbf{Z}_1\beta) = \mu(\zeta)$$

$$V(\mathbf{y}_2 \mid \mathbf{y}_1, \zeta) = \sigma^2\left(\mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\right) = \Psi(\zeta)$$

These are classical formulae. See for instance page 522 in RAO (1973). The auxiliary vector μ and matrix Ψ with elements as function of the parameters in ζ are for later use.

4. PAYMENT PATTERN IN CONTINUOUS TIME

The payment pattern as modelled by the probability vector π could be subject to parameter estimation. This amounts to $n-1$ free parameters for a discrete time payment

pattern with finite horizon and as a result no uppertail. This model has its merits. However, when n becomes large the estimated pattern of π may become erratic whereas a smooth pattern is likely to be the actuarial a priori view. Furthermore, an upper tail may be more realistic. To this end we imagine the payment pattern to coincide with settlement time duration. It has a density and distribution depending on a few parameters η:

$$g(t\,|\,\eta) \qquad\qquad G(t\,|\,\eta) = \int_0^t g(z\,|\,\eta)\mathrm{d}z$$

Examples of potential useful densities are Weibull, lognormal, gamma, inverse Gaussian, etc. De discretized probabilities π result as the solution of the linear system:

$$\pi_1 + \cdots + \pi_i = \int_0^1 G(i - z\,|\,\eta)\mathrm{d}z \qquad\qquad i = 1,2,\cdots,n-1$$
$$\pi_1 + \cdots + \pi_{n-1} + \pi_n = 1$$

This average of the cumulative distribution is due to the difference between loss occurrences at the beginning of the loss year and towards the end of the loss year. An implicit assumption here is that loss occurrences are uniformly distributed within the year. We illustrate the foregoing with exponentially distributed settlement duration. In that case we have $g(t\,|\,\eta) = \eta\exp(-t\eta)$ and π results as:

$$\pi_1 = 1 - ce^{-\eta} \qquad\qquad c = \eta^{-1}(e^{\eta} - 1)$$
$$\pi_i = c^2\eta e^{-i\eta} \qquad\qquad i = 2,3,\cdots,n-1$$
$$\pi_n = ce^{(1-n)\eta}$$

This is a geometric distribution with a deflated first probability. The appearance of such an averaged cumulative distribution is not new. In the econometric topic of distributed lag analysis THEIL and STERN (1963) used a simple Erlang density given by $g(t\,|\,\eta) = \eta^2 t\exp(-t\eta)$ where similar averaged integrals occur. For a survey on distributed lags, see GRILICHES (1967).

The empirical modelling of the payment pattern may need a multimodal density. In that case a mixture of basic densities is an option. Observe that for most densities the average of the cumulative distribution requires numerical integration.

5. ASPECTS OF PARAMETER ESTIMATION

The data available for parameter estimation are stored in the r_1 elements of $\mathbf{y}_1$. We remember the mean and covariance as:

$$E(\mathbf{y}_1) = \mathbf{Z}_1\beta \qquad\qquad V(\mathbf{y}_1) = \sigma^2\mathbf{C}_{11}$$

Minus the logarithm of the likelihood function can be written as:

$$-\ln L = r\ln\sigma + \tfrac{1}{2}\ln\det\mathbf{C} + \tfrac{1}{2}\sigma^{-2}(\mathbf{y} - \mathbf{Z}\beta)'\mathbf{C}^{-1}(\mathbf{y} - \mathbf{Z}\beta)$$

where, for notational simplicity, all indices 1 have been suppressed. Optimization with respect to β and σ conditional on the other parameters gives rise to the following closed form expressions:

$$\hat{\beta} = \left(\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{C}^{-1}\mathbf{y} \qquad\qquad \hat{\sigma} = \sqrt{Q/r}$$

where Q is a quadratic form given by:

$$Q = (\mathbf{y} - \mathbf{Z}\hat{\beta})'\mathbf{C}^{-1}(\mathbf{y} - \mathbf{Z}\hat{\beta}) = \mathbf{y}'\mathbf{C}^{-1}\mathbf{y} - \mathbf{y}'\mathbf{C}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{C}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{C}^{-1}\mathbf{y} = \begin{bmatrix}\mathbf{y}\\ \mathbf{0}\end{bmatrix}'\begin{bmatrix}\mathbf{C} & \mathbf{Z}\\ \mathbf{Z}' & \mathbf{O}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{y}\\ \mathbf{0}\end{bmatrix}$$

The estimation criterion after concentrating out β and σ boiles down to:

$$-\ln L^* = \frac{\ln\det\mathbf{C} + r\ln Q}{2}$$

Minimization of this estimation criterion with respect to the remaining unknown parameters, needs numerical optimization. In generaly this is a straightforward matter. When the modelling of the payment pattern π requires mixtures, it becomes more delicate. Local optima are a possibility and these need care. See page 742-743 in QUANDT (1983). At the minimal stationary point, the implied estimates for β and σ can be derived. The Hessian matrix of the original minus loglikelihood function at this optimum may be derived by numerical differentation. The inverse of this Hessian gives an approximation for the covariance matrix of the parameter estimator. We denote this covariance matrix as Φ. In a Bayesian way we may state that the posterior density for the parameters ζ is approximately normal with mean and covariance matrix:

$$E(\zeta) \approx \hat{\zeta} \qquad\qquad V(\zeta) \approx \Phi$$

These inferential results not only play a role at the modelling stage but also when accounting for the impact of parameter estimation uncertainty on the predictive distribution of $\mathbf{y}_2$.

6. PREDICTIVE DISTRIBUTION

Remembering the mean and covariance of $\mathbf{y}_2$ conditional on $\mathbf{y}_1$ and ζ:

$$E(\mathbf{y}_2 \mid \mathbf{y}_1, \zeta) = \mu(\zeta) \qquad\qquad V(\mathbf{y}_2 \mid \mathbf{y}_1, \zeta) = \Psi(\zeta)$$

we write the matrix of cross-moments as:

$$E(\mathbf{y}_2\mathbf{y}_2' \mid \mathbf{y}_1, \zeta) = \Psi(\zeta) + \mu(\zeta)\mu'(\zeta)$$

We linearize $\mu(\zeta)$ about the maximum likelihood estimate

$$\mu(\zeta) \approx \mu(\hat{\zeta}) + \mathbf{J}(\zeta - \hat{\zeta})$$

where **J** is the Jacobian matrix:

$$\mathbf{J} = \left.\frac{\partial\mu(\zeta)}{\partial\zeta'}\right|_{\zeta=\hat{\zeta}}$$

The columns of **J** corresponding with the elements of β and σ allow simple closed form expressions:

$$\frac{\partial\mu(\zeta)}{\partial\beta'} = \mathbf{Z}_2 - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{Z}_1 \qquad\qquad \frac{\partial\mu(\zeta)}{\partial\sigma} = \mathbf{0}$$

The other columns of **J** are most easily derived by numerical differentiation. Remembering the mean and covariance of $\boldsymbol{\zeta}$ we marginalize the vector of means and matrix of cross-moments of $\mathbf{y}_2$:

$$E(\mathbf{y}_2 \mid \mathbf{y}_1) \approx \mu(\hat{\zeta})$$

$$E(\mathbf{y}_2\mathbf{y}_2' \mid \mathbf{y}_1) \approx \Psi(\hat{\zeta}) + \mathbf{J}\Phi\mathbf{J}' + \mu(\hat{\zeta})\mu'(\hat{\zeta})$$

from which the covariance matrix results as

$$V(\mathbf{y}_2 \mid \mathbf{y}_1) \approx \Psi(\hat{\zeta}) + \mathbf{J}\Phi\mathbf{J}'$$

This covariance matrix is additive in stochastic process uncertainty and parameter estimation uncertainty. When $\mathbf{S}_2$ is a matrix consisting of a single row, $\mathbf{y}_2$ will be scalar and the share of estimation uncertainty in total uncertainty results as:

$$\frac{\mathbf{J}\Phi\mathbf{J}'}{\Psi(\hat{\zeta}) + \mathbf{J}\Phi\mathbf{J}'}$$

The predictive probability distribution of $\mathbf{y}_2$ is approximately multivariate Student with the foregoing mean and covariance matrix. The degrees of freedom being informally derived as the number of observations r_1 minus the number of estimated parameters in the expected value specification, that is β and the free parameters in π.

Once we have avalaible this predictive distribution, its mean generates an unbiased statistical forecast whereas an upper percentile will correspond with a prudent actuarial provision. Ruin-like percentiles bring solvency margins into perspective.

REFERENCES

ANDERSON, T.W. and R.P. MENTZ (1980). On the structure of the likelihood function of autoregressive and moving average models. *Journal of Time Series Analysis* **1** 83-94.

GRILICHES, Z. (1967). Distributed lags: a survey. *Econometrica* **35** 16-49.

GRILICHES, Z. and M.D. INTRILIGATOR (1983). *Handbook of Econometrics, Volume* 1. Amsterdam: North-Holland Publishing Company.

MAGNUS, J.R. and H. NEUDECKER (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Chichester: John Wiley & Sons.

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

SIDDIQUI, M.M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *The Annals of Mathematical Statistics* **29** 585-588.

THEIL, H. (1983). Linear algebra and matrix methods in econometrics. Chapter 1 in GRILICHES and INTRILIGATOR (1983).

THEIL, H. and R.M. STERN (1960). A simple unimodal lag distribution. *Metroeconomica* **12** 111-119.

QUANDT, R.E. (1983). Computational Problems and Methods. Chapter 12 in GRILICHES and INTRILIGATOR (1983).