SYNCHRONOUS BOOTSTRAPPING OF SEEMINGLY UNRELATED REGRESSIONS

Greg Taylor

Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Professorial Associate,
Centre for Actuarial Studies
Faculty of Economics and Commerce
University of Melbourne
Parkville VIC 3052
Australia

Adjunct Professor in Actuarial Studies
Faculty of Commerce and Economics
University of New South Wales
Kensington NSW 2033
Australia

Phone: 61 2 9249 2901
Fax: 61 2 9249 2999
greg.taylor@taylorfry.com.au

Gráinne McGuire

Taylor Fry Consulting Actuaries
Level 8, 30 Clarence Street
Sydney NSW 2000
Australia

Phone: 61 2 9249 2906
Fax: 61 2 9249 2999
grainne.mcguire@taylorfry.com.au

August 2005

Summary

Consider the seemingly unrelated regression framework, in which regression models are applied to a number of data sets, with stochastic dependencies between them. The regression models are not restricted to general linear models (e.g. GLMs). Forecasts are required, with estimates of prediction errors that account for the dependencies between data sets.

Bootstrapping is used to estimate prediction errors. Specialised forms of bootstrapping that capture the dependencies are constructed.

Insurance and banking applications are mentioned. The former is investigated with numerical examples. The specific context is insurance loss reserving under the requirement that the entire distribution of loss reserve be estimated, where this reserve is aggregated across a number of stochastically dependent lines of business.

Keywords: bootstrap, correlation, loss reserving, seemingly unrelated regression, stochastic dependency.

1. Introduction

This paper is concerned with the simultaneous regression modelling of a number of data sets when observations within each data set are stochastically independent but there is a dependency structure between the data sets. This is the framework of **seemingly unrelated regressions** (SUR), introduced by Zellner (1962) and discussed by Srivastava and Giles (1987).

An estimation procedure follows from this framework, but difficulties will arise if the data under consideration do not conform with the framework. There may be additional difficulties in obtaining reliable estimates of the dependency structures when, *a priori*, these are totally unknown.

The SUR framework amounts to stacked general linear models, and so involves the restrictive conditions of normal error structure and linear relations between response variates and covariates (Section 2).

One would like to relax these restrictions by the use of generalised linear models (GLMs) instead of general linear models. However, as will appear in Section 3, estimation within the GLM framework becomes problematic as soon as any dependency structure is introduced between data sets.

This paper suggests forms of bootstrapping which capture such dependency structures, at least approximately, and even when the structures are unknown.

2. The seemingly unrelated regression framework

Consider *I* general linear models

$$Y_i = X_i\beta_i + \varepsilon_i, \quad i = 1, 2, ..., I \qquad\qquad (2.1)$$

$$\varepsilon_i \sim N\left(0, D_i\right) \qquad\qquad (2.2)$$

where

Y_i is a column n_i-vector of observations

X_i is an n_i x p_i design matrix

β_i is a column p_i-vector of parameters

ε_i is a column n-vector of stochastically independent error terms,

whence D_i is an n_i x n_i diagonal matrix.

While the components within each Y_i are stochastically independent, distinct Y_i are not necessarily independent, one from another.

Let Y denote the stacked n-vector:

$$Y = \left[Y_1^T, ..., Y_I^T\right]^T$$

where the upper T denotes matrix transposition and $n = n_1 + ... + n_I$.

Similarly, let β, ε denote the p-vector and n-vector respectively:

$$\beta = \left[\beta_1^T, ..., \beta_I^T\right]^T$$
$$\varepsilon = \left[\varepsilon_1^T, ..., \varepsilon_I^T\right]^T$$

where $p = p_1 + ... + p_I$ and X is the n x p block diagonal matrix

$$X = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_I \end{bmatrix}.$$

Then (2.1) yields

$$Y = X\beta + \varepsilon \qquad\qquad (2.3)$$

$$\varepsilon \sim N\left(0, \sigma^2 W\right) \qquad\qquad (2.4)$$

where $\sigma^2 > 0$ is an unknown scalar, W is an n x n matrix with diagonal blocks $D_1, ..., D_I$, but otherwise unspecified.

If W is known, then β is estimated by the standard general linear regression estimator:

$$\hat{\beta} = \left(X^T W^{-1} X \right)^{-1} X^T W^{-1} Y . \qquad (2.5)$$

This is a **maximum likelihood estimator** (MLE).

This regression yields a vector of residuals

$$\begin{aligned} R &= Y - X\hat{\beta} \\ &= \left[1 - X \left(X^T W^{-1} X \right)^{-1} X^T W^{-1} \right] Y . \end{aligned} \qquad (2.6)$$

It follows from this and (2.4) that

$$Var\left[R \right] = \sigma^2 W \left[1 - W^{-\frac{1}{2}} X \left(X^T W^{-1} X \right)^{-1} X^T W^{-\frac{1}{2}} \right]. \qquad (2.7)$$

In the case in which $W = 1$, the second number within the square bracket is of order $\left(\min n_i \right)^{-1}$. Provided that W does not differ too much from the identity, Var $[R]$ will approximate $\sigma^2 W$. This means that the n x n matrix RR^T can be taken as an estimator of $\sigma^2 W$.

Suppose that some structure is imposed on W, eg

$$W_{ij,kl} = \left(p_{ij} \right)^{\delta(k,l)}, 0 < p_{ij} < 1 \text{ for } i \neq j \qquad (2.8)$$

where $W_{ij,kl}$ denotes the (k,l) element of (i,j) block of W and $\delta(k,l)$ is some distance measure such that $\delta(k,k) = 0$ and $\delta(k,l) \uparrow$ as $| k - l | \uparrow$. Then the p_{ij} that define W may be estimated by reference to RR^T.

According to this description, W is estimated by reference to R, but R depends on W. Such estimation must be implemented iteratively. Let $W^{(m)}$ be the m-th iterative estimate of W. Then define the m-th residual vector $R^{(m)}$ by (2.6) with W replaced by $W^{(m)}$. The next iterate $W^{(m+1)}$ is then estimated by reference to $R^{(m)}$ as described above. The process may be initiated by $W^{(0)} = D$, the block diagonal matrix formed from the D_i.

3. Attempted extension of the SUR framework to include GLMs

The SUR framework represented by (2.1) and (2.2) assumes linear relations between the response variates and covariates, and also a normal error structure. Both of these restrictions may be relaxed if (2.1) and (2.2) are replaced by the following GLM:

$$Y_i \sim EDF\left(\mu_i, \phi_i, V_i \right) \qquad (3.1)$$

$$\mu_i = h_i^{-1}\left(X_i\beta_i\right) \tag{3.2}$$

where (3.1) means that each component of Y_i has a distribution from the **exponential dispersion family** (Nelder and Wedderburn, 1972) with $E[Y_i] = \mu_i$, **scale parameter** ϕ_i (a scalar), **variance function** $V_i(.)$, and, for given i, all components of Y_i are stochastically independent.

The notation in (3.2) adopts the convention that the inverse of the **link function** h_i, which maps a subset of the real numbers to the real numbers, operates on vectors component-wise.

Just as in the SUR framework, the models represented by (3.1) and (3.2) may be stacked to give

$$Y \sim EDF\left(\mu,\phi,V\right) \tag{3.3}$$

$$\mu = h^{-1}\left(X\beta\right) \tag{3.4}$$

where Y, X, β are stacked just as in Section 2, and so are μ, ϕ and V.

One might now attempt to introduce dependency between the data sets in a manner parallel to that used in the SUR framework of Section 2. Accordingly, it is supposed that Var [Y] is **not** diagonal.

Assume that this covariance matrix has same structure as W in (2.8). At this point, one wishes to extend (3.3) to the following:

$$Y \sim EDF\left(\mu,\phi,V,C\right) \tag{3.3a}$$

in which the covariance structure W has been represented by the correlation matrix

$$C = D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \tag{3.5}$$

with D the n x n diagonal matrix containing the diagonal of W.

Unfortunately, the precise likelihood to be associated with (3.3a) is unclear. In a general linear setting, correlations would be eliminated by orthogonal transformation of variates. The covariance W would be diagonalised:

$$W = P^T\Lambda P \tag{3.6}$$

with P orthogonal and Λ diagonal, so that, for $Z = PY$, $Var[Z] = \Lambda$.

In the case of (3.3a), however, this does **not** yield $Z \sim EDF\left(P\mu,.,.,1\right)$ because the EDF is not closed under linear transformation.

It follows that a model consisting of (3.3a) and (3.4) cannot in general be linearly transformed to a GLM with stochastically independent components of the response vector. Hence estimation in the extended framework (3.3a) becomes problematic.

4.   Synchronous bootstrapping

4.1   Preliminaries

In parallel with Sections 2 and 3, but more generally, consider I models

$$Y_i = g_i\left(X_i, \beta_i\right) + \varepsilon_i, \qquad i = 1, 2, ..., I \tag{4.1}$$

where

Y_i is a column n_i-vector of observations $Y_{ik}, k = 1, ..., n_i$

X_i is an $n_i \times p_i$ matrix whose rows consist of predictors of the Y_{ik}

β_i is a column p_i-vector of parameters

g_i is a vector-valued function specifying a model structure

ε_i is a column n_i-vector of stochastically independent, centred, equi-distributed error terms each with variance σ_i^2, referred to subsequently as **process error**.

For the present analysis, assume that $n_1 = ... = n_I = n$, say (not the same n as in Sections 2 and 3).

Assume that there is a dependency between the data sets Y_i, described by correlation matrices

$$C_{ij} = Corr\left(Y_i, Y_j\right). \tag{4.2}$$

It follows from the above independence assumption that $C_{ii} = 1$ for each i.

The C_{ij} will be supposed unknown, in which case estimation of each β_i will be carried out by reference to just the data set Y_i. Suppose that there is some unbiased estimation procedure for mapping data Y_i to estimates $\hat{\beta}_i$, $\hat{\sigma}_i^2$ of β_i, σ_i^2. Let $\hat{Y}_i$ denote the vector of **fitted values**

$$\hat{Y}_i = g_i\left(X_i, \hat{\beta}_i\right). \tag{4.3}$$

Typically, Y_i and $\hat{Y}_i$ will be correlated, and furthermore the correlations (4.2) will induce correlations between $\hat{Y}_i$ and $\hat{Y}_j$. However, it will be assumed that these are of order n^{-1}.

Define the **standardised Pearson residual vector**

$$R_i = \left(Y_i - \hat{Y}_i\right)/\hat{\sigma}_i. \tag{4.4}$$

Suppose now that one wishes to estimate a vector $\zeta_i = E[Z_i]$ where

$$Z_i = \alpha_i\left(U_i, \beta_i\right) + \eta_i \tag{4.5}$$

which is of the same form as (4.1) but with $Y_i, g_i, X_i, \varepsilon_i$ replaced by $Z_i, \alpha_i,$ U_i, η_i. Dimensions may differ as between (4.1) and (4.5), other than for β_i. Suppose that Z_i is of dimension $n' \le n$.

An estimate of ζ_i is given by

$$\hat{\zeta}_i = \alpha_i\left(U_i, \hat{\beta}_i\right). \tag{4.6}$$

Define a **re-sampling matrix** as an $n' \times n \left(n' \le n\right)$ matrix whose rows are distinct natural coordinate vectors, ie have one unit component and the rest zero.

If $n' = n$, a re-sampling matrix is a **permutation matrix**. In general, if M is a re-sampling matrix with (k,s) element M_{ks}, then the product $M_{ks}M_{lt}$ has the following properties.

For $k = l, M_{ks}M_{lt}$ $= 1$ for $s = t$
                                        $= 0$ otherwise
For $k \ne l, M_{ks}M_{lt}$ $= 1$ for a single pair s, t with $s \ne t$
                                        $= 0$ otherwise.

4.2   Sources of correlation

There are two main possible sources of the correlations (4.2). It will be seen in the sub-sections below that these different sources of correlation dictate different forms of bootstrapping.

The first source is **correlated noise**, that is the $g_i\left(X_i, \beta_i\right)$ terms in (4.1) are deterministic and

$$Corr\left(\varepsilon_i, \varepsilon_j\right) = C_{ij}, \tag{4.7}$$

implying (4.2).

Alternatively, the source may be **correlated parameters**. This is an apparent form of correlation arising from **model mis-specification**.

Suppose that, while (4.1) is the assumed model form, the correct form is in fact

$$Y_i = g^+{}_i \left(X^+{}_i, \beta_i, \gamma_i \right) + \varepsilon^+{}_i \qquad\qquad (4.1a)$$

where γ_i is a vector of parameters unrecognised by the model, and $g^+{}_i$ and $X^+{}_i$ are modifications of g_i and X_i that do recognise these additional parameters.

By (4.1) and (4.1a),

$$\varepsilon_i = \varepsilon^+{}_i + b_i$$

where b_i denotes the **bias vector**

$$b_i = g^+{}_i \left(X^+{}_i, \beta_i, \gamma_i \right) - g_i \left(X_i, \beta_i \right)$$

Then

$$Cov\left(\varepsilon_i, \varepsilon_j \right) = Cov\left(\varepsilon^+_i, \varepsilon^+_j \right) + b_i b_j^T \qquad\qquad (4.8)$$

The left side of this relation is the covariance between ε_i and ε_j, relative to a model of the form (4.1). It consists of the first term on the right side, which is the correct covariance, plus an adjustment that is a measure of the mis-specification of the true model (4.1a).

4.3   Bootstrapping a single data set

The following is a description of the bootstrap (Efron, 1979) as it is conventionally applied.

Define **pseudo-data vectors** $Y^*_{i(r)}, r = 1, ..., R$ associated with Y_i as

$$Y^*_{i(r)} = \hat{Y}_i + \hat{\sigma}_i M_{i(r)} R_i \qquad\qquad (4.9)$$

where $M_{i(r)}$ is an n x n permutation matrix chosen randomly from the set of all such matrices and the $M_{i(r)}$ are independent.

This re-samples the data **without replacement**. The fact that $Y^*_{i(r)}$ has the same distribution as Y_i is an essential feature of the bootstrap.

Each pseudo-data vector $Y^*_{i(r)}$ yields **pseudo-estimates** $\hat{\beta}^*_{i(r)}, \hat{\sigma}^{*2}_{i(r)}$, and **pseudo-fitted values** $\hat{Y}^*_{i(r)}$. This also yields pseudo-estimates $\hat{\varsigma}^*_{i(r)}$ of ς_i where

$$\hat{\varsigma}^*_{i(r)} = \alpha_i \left(U_i, \hat{\beta}^*_{i(r)} \right), \qquad i = 1, ..., I. \qquad\qquad (4.10)$$

When $\hat{\varsigma}_i$ is in the nature of a **forecast** of Z_i, one may generate the **pseudo-forecasts**

$$Z^*_{i(r)} = \hat{\varsigma}^*_{i(r)} + \eta^*_{i(r)} \qquad\qquad (4.11)$$

where $\eta^*_{i(r)}$ is a component of noise, and is sampled just as in (4.9):

$$\eta^*_{i(r)} = \hat{\sigma}_i A_{i(r)} R_i \qquad\qquad (4.12)$$

where $A_{i(r)}$ is an $n' \times n$ re-sampling matrix chosen randomly from the set of all such matrices and the $A_{i(r)}$ are independent.

The collection of $\hat{\varsigma}^*_{i(r)}$, $r = 1,\dots,R$ induce an empirical distribution of $\hat{\varsigma}_i$, which will be called the **bootstrap distribution** of $\hat{\varsigma}_i$. Similarly the $Z^*_{i(r)}$, $r = 1,\dots,R$ induce a bootstrap distribution of Z_i.

4.4   Synchronous bootstrapping of multiple data sets

4.4.1   Motivation

The bootstrap described in Section 4.3 may be validly performed for each value of $i = 1,\dots,I$. Suppose, however, that one is interested in forecasting some quantity that combines data sets, eg

$$z = 1^T Z \qquad\qquad (4.13)$$

where $Z = \left(Z_1^T,\dots,Z_I^T \right)^T$ and 1 is the vector of the same dimension with all components unity.

Let $Z^*_{(r)} = \left(Z^{*T}_{1(r)}\dots,Z^{*T}_{I(r)} \right)^T$ but note that, because the I data sets have been independently bootstrapped, any correlation structure between the Z_i is likely to have been lost.

Hence the bootstrap distribution of Z will not be representative of the true distribution. Neither will the bootstrap distribution of z be representative of its true distribution.

4.4.2   Synchronising individual data points

Let Y_{ik} denote the k-th component of Y_i. Suppose that the correlation structure (4.2) is of the correlated noise form (4.7) and is given by

$$\begin{aligned} C_{ij,kl} &= \rho_{ij1} \ \text{ for } i \neq j, k = l \\ &= \rho_{ij0} \ \text{ for } i \neq j, k \neq l \end{aligned} \qquad\qquad (4.14)$$

where $C_{ij,kl}$ denotes the (k,l)-element of the matrix C_{ij}. Recall that $C_{ii} = 1$.

Here ρ_{ij0}, which may be zero, may be regarded as a "background correlation" between distinct data sets i and j, which ρ_{ij1} represents a spike in correlation that occurs for points in corresponding positions in those data sets.

The quantities $\{Z_i\}$ to be forecast are assumed subject to the same correlation structure (4.14) as the data $\{Y_i\}$ and $Corr\left[Y_{ik}, Z_{jl}\right] = \rho_{ij0}$.

Suppose the data to be modelled are of the form (4.1), and apply the bootstrap described in Section 4.3 with the exception that (4.9) and (4.12) are replaced by the following:

$$Y^*_{i(r)} = \hat{Y}_i + \hat{\sigma}_i M_{(r)} R_i \tag{4.9a}$$

$$\eta^*_{i(r)} = \hat{\sigma}_i A_{(r)} R_i \tag{4.12a}$$

where $M_{(r)}$ and $A_{(r)}$ are n x n and n' x n randomly chosen re-sampling matrices.

The difference between this form of bootstrap and the conventional form is that the present form applies the same matrices $M_{(r)}$ and $A_{(r)}$ to the bootstrapping of all I data sets. In this sense, the re-sampling of the residual vectors $R_1,..., R_I$ is **synchronised** for each r. This form of bootstrapping was suggested by Kirschner, Kerley and Isaacs (2002) in an insurance loss reserving context.

It can now be shown that the joint distribution of $Y^*_{1(r)},..., Y^*_{I(r)}$ is approximately the same as for $Y_1,..., Y_I$. To see this, note that, from (4.4) and (4.9a),

$$Y^*_{i(r)} = \left[1 - M_{(r)}\right]\hat{Y}_i + M_{(r)} Y_i \tag{4.15}$$

Write μ_i to denote $g_i\left(X_i, \beta_i\right)$ so that (4.1) takes the form

$$Y_i = \mu_i + \varepsilon_i \tag{4.1a}$$

and substitute this result into (4.15), to obtain

$$Y^*_{i(r)} - \mu_i = \left[1 - M_{(r)}\right]\left(\hat{Y}_i - \mu_i\right) + M_{(r)}\varepsilon_i \tag{4.16}$$

From (4.2) and the assumption made immediately after (4.3),

$$Corr\left[Y^*_{i(r)}, Y^*_{j(r)}\right] = M_{(r)} C_{ij} M^T_{(r)} + O\left(n^{-1}\right) \tag{4.17}$$

The (k,l) element of this correlation matrix is

$$\sum_{s,t} M_{(r)ks} M_{(r)lt} C_{ij,st} \qquad\qquad (4.18)$$

where $M_{(r)ks}$ denotes the (k,s) element of $M_{(r)}$.

The property of re-sampling matrices stated just after its definition in Section 4.1, applied to (4.18), reduces it to ρ_{ij1} for $k = l$ and ρ_{ij0} for $k \neq l$, in which case (4.17) is just

$$Corr\left[Y^*_{i(r)}, Y^*_{j(r)} \right] = C_{ij} + O\left(n^{-1}\right) \qquad\qquad (4.19)$$

Thus, to within $O\left(n^{-1}\right), \left\{Y^*_{1(r)},...,Y^*_{I(r)}\right\}$ has the same marginals and correlation structure as $Y_1,...,Y_I$, and therefore has approximately the same joint distribution. The basic requirement for application of the bootstrap, stated just after (4.9) is thus satisfied.

By the same reasoning as above, it follows that the $Z^*_{i(r)}$ have the same on-diagonal/off-diagonal structure as (4.19). It also follows that the correlation between any component of $Y^*_{i(r)}$ and any component of $Z^*_{j(r)}$ is ρ_{ij0}, as required.

Then the $Z^*_{(r)}$ induce a bootstrap distribution that is representative of Z, including its dependencies.

4.4.3 Synchronising data subsets

Let $N_1, N_2,..., N_Q$ be disjoint subsets of $\{1,...,n\}$ whose union is equal to this set.

Suppose the correlation structure (4.2) is of the correlated parameter form described in Section 4.2 and is given by

$$\begin{aligned} C_{ij,kl} &= \rho_{ij1} \text{ for } i \neq j, \quad k,l \in \text{ some } N_q \\ &= \rho_{ij0} \text{ for } i \neq j, \quad k,l \text{ from distinct } N_q \end{aligned} \qquad\qquad (4.20)$$

Assume that there is no correlated noise, ie the ε_i, $i = 1,...,I$ are stochastically independent.

As in Section 4.4.2, ρ_{ij0} may be viewed as a background level of correlation between distinct data sets, while ρ_{ij1} represents a different level of correlation that occurs for points in corresponding subsets of those data sets.

Now suppose that the Z_i extend the correlation structure of the Y_i in the following sense. Let $N^*_1,..., N^*_Q$ be disjoint subsets of $\{1,...,n'\}$ whose union is equal to this set. Let n_q and n'_q be the orders of N_q and N^*_q respectively. Suppose that $n'_q \leq n_q$ for each q.

Let $Y_i^{(q)}$ denote the sub-vector of Y_i consisting of just the components $Y_{ik}, k \in N_q$. Similarly, let $Z_i^{(q)}$ denote the sub-vector of Z_i consisting of just the components $Z_{il}, l \in N_q^*$. Define

$$W_i^{(q)} = \begin{bmatrix} Y_i^{(q)} \\ Z_i^{(q)} \end{bmatrix} \qquad\qquad (4.21)$$

Let $W_{ik}^{(q)}$ denote the k-th component of $W_i^{(q)}$ and let $D_{ij,kl}^{(q,s)}$ denote $Corr\left(W_{ik}^{(q)}, W_{jl}^{(s)}\right)$.

Suppose that

$$\begin{aligned}
D_{ij,kl}^{(q,s)} &= 1 \text{ for } i = j, q = s, k = l \\
&= 0 \text{ for } i = j, q = s, k \neq l \\
&= 0 \text{ for } i = j, q \neq s \\
&= \rho_{ij1} \text{ for } i \neq j, q = s \\
&= \rho_{ij0} \text{ for } i \neq j, q \neq s
\end{aligned} \qquad\qquad (4.22)$$

which is consistent with (4.20).

Suppose once again that the data to be modelled are of the form (4.1), and apply the bootstrap described in Section 4.3 with the exception that (4.9) and (4.12) are replaced by the following:

$$Y_{i(r)}^{(q)*} = \hat{Y}_i^{(q)} + \hat{\sigma}_i M_{i(r)}^{(q)} R_i^{(q)} \qquad\qquad (4.9b)$$

$$\eta_{i(r)}^{(q)*} = \hat{\sigma}_i A_{i(r)}^{(q)} R_i^{(q)} \qquad\qquad (4.12b)$$

and

$$Z_{i(r)}^{(q)*} = \hat{\varsigma}_{i(r)}^{(q)*} + \eta_{i(r)}^{(q)*} \qquad\qquad (4.23)$$

where $R_i^{(q)}$ is the sub-vector of R_i corresponding to $Y_i^{(q)}$ as a sub-vector of Y_i, and $M_{i(r)}^{(q)*}$ and $A_{i(r)}^{(q)*}$ are $n_q \times n_q$ and $n_q' \times n_q$ randomly chosen re-sampling matrices.

Note that **different matrices** $M_{i(r)}^{(q)}$ and $A_{i(r)}^{(q)}$ apply to **different subsets within the data sets**. This differs from the situation described by (4.9a) and (4.12a) where common re-sampling matrices were applied to whole data sets. The re-sampling of residual vectors is now synchronised within each subset N_q and its companion N_q^*.

Corresponding to (4.21), define

$$W_{i(r)}^{(q)*} = \begin{bmatrix} Y_{i(r)}^{(q)*} \\ Z_{i(r)}^{(q)*} \end{bmatrix} \qquad . \qquad W_{i(r)}^{*} = \begin{bmatrix} Y_{i(r)}^{*} \\ Z_{i(r)}^{*} \end{bmatrix} \qquad (4.24)$$

The argument given in Section 4.4.2 may be adapted to the present situation to demonstrate that $W_{1(r)}^{(q)*}, ..., W_{I(r)}^{(q)*}$ have approximately the same joint distribution as $W_{1(r)}^{(q)}, ..., W_{I(r)}^{(q)}$; and that $W_{1(r)}^{*}, ..., W_{I(r)}^{*}$ have approximately the same joint distribution as $W_{1(r)}, ..., W_{I(r)}$ for $r = 1, ..., R$.

Define Z and $Z_{(r)}^{*}$ as in Section 4.4.1. Then the $Z_{(r)}^{*}$ induce a bootstrap distribution that is representative of Z, including its dependencies.

Remark 4.1. Synchronised re-sampling within subsets N_q is **not** the natural extension of the synchronised point re-sampling in Section 4.4.2. In that earlier case, if $R_{ik} \rightarrow R_{il}$ in the re-sampling of data set i, then $R_{jk} \rightarrow R_{jl}$ in the re-sampling of data set j. The natural extension of this to the present sub-section would be for:

(a)   the re-sampling $R_{ik} \rightarrow R_{il}$ to involve the mapping of each subset N_q to a single other subset N_s; **and**

(b)   the same mapping of subsets (though not of members of those subsets) in $R_{jk} \rightarrow R_{jl}$; ie for all $k \in N_q$, there is a single N_s such that

$R_{ik} \rightarrow R_{il}$ with $l \in N_s$

$R_{jk} \rightarrow R_{jl}$ with $j \in N_s$.

This, however, creates some difficulty when $n_q < n_s$, and the alternative procedure described above will often be more convenient.

Remark 4.2. The re-sampling scheme represented by the matrices $M_{i(r)}^{(q)}$ and $A_{i(r)}^{(q)}$ will encounter difficulties if N_q is small. There are $N_q!$ choices of the permutation matrix $M_{i(r)}^{(q)}$, and so, for given r, there are $n_1! n_2! ... n_Q!$ possible re-samplings in (4.9b). This needs to be at least as large as R, and preferably a good deal larger. If not, it may be necessary to aggregate some of the smaller N_q into larger ones.

Remark 4.3. The correlation structure (4.20) was assumed to involve no correlated noise. Such noise can be added with structure (just for the noise) (4.14), changing the total correlation structure to the following:

$$C_{ij,kl} = \rho_{ij2} \text{ for } i \neq j, \, k = l$$
$$= \rho_{ij1} \text{ for } i \neq j, \, k,l \in \text{ some } N_q, k \neq l \qquad (4.25)$$
$$= \rho_{ij0} \text{ for } i \neq j, \, k,l \text{ from distinct } N_q$$

This correlation structure will be reproduced by a different form of bootstrap in which (4.9b) and (4.12b) are replaced by the following:

$$Y_{i(r)}^{(q)*} = \hat{Y}_i^{(q)} + \hat{\sigma}_i M_{(r)}^{(q)} R_i^{(q)} \qquad (4.9c)$$

$$\eta_{i(r)}^{(q)*} = \hat{\sigma}_i A_{(r)}^{(q)} R_i^{(q)} \qquad (4.12c)$$

That is, the same re-sampling is applied (at the subset level) to each data set. The modification of (4.9b) and (4.12b) to (4.9c) and (4.12c) is the same as from (4.9) and (4.12) to (4.9a) and 4.12a).

4.4.4 Synchronising geometrically related subsets

Section 4.4.3 was concerned with correlations between subsets whose positions within their respective data sets were fixed in **absolute** terms. The present subsection will be concerned with correlation between data sets that depend on the **relative** positions of members of those sets.

This may arise because the data sets, abstractly represented as vectors hitherto, actually have some geometric structure. This is perhaps best explained by way of an example.

Suppose that the data vector Y_i comprises observations taken over some 2-dimensional lattice, so that any component is naturally identified by a **co-ordinate pair** thus: $Y_{ik} = Y_{ik_1k_2}, k_1 = 1,...,K_1; \, k_2 = 1,...,K_2$.

Now suppose that the correlations between Y_i and Y_j each take the form

$$Corr\left(Y_{ik_1k_2}, Y_{jl_1l_2}\right) = \text{ function } \left(|k_1 - l_1|, |k_2 - l_2|\right) \qquad (4.26)$$

The arguments $|k_1 - l_1|, |k_2 - l_2|$ measure the relative difference between the spatial positioning of the observations Y_{ik} and Y_{jl}.

It would be possible to express (4.26) in terms of vector components Y_{ik}, Y_{jl}, as before, but to do so would seem contrived, and the geometric form appearing in (4.26) will be retained.

Consider the following as an example of (4.26), in which the left side is denoted $C_{ij,k_1k_2l_1l_2}$:

$$C_{ij,k_1k_2l_1l_2} = \rho_{ij2} \text{ if } i \neq j, (k_1, k_2) = (l_1, l_2)$$
$$= \rho_{ij1} \text{ if } i \neq j, (k_1, k_2) \neq (l_1, l_2), |k_1 - k_2|, |l_1 - l_2| \leq 1 \qquad (4.27)$$
$$= \rho_{ij0} \text{ if } i \neq j, \text{ at least one of } |k_1 - k_2|, |l_1 - l_2| > 1$$

Assume that there is no correlated noise. Geometrically, the situation is as depicted in Figure 4.1.

Figure 4.1
Geometric representation of (4.27)
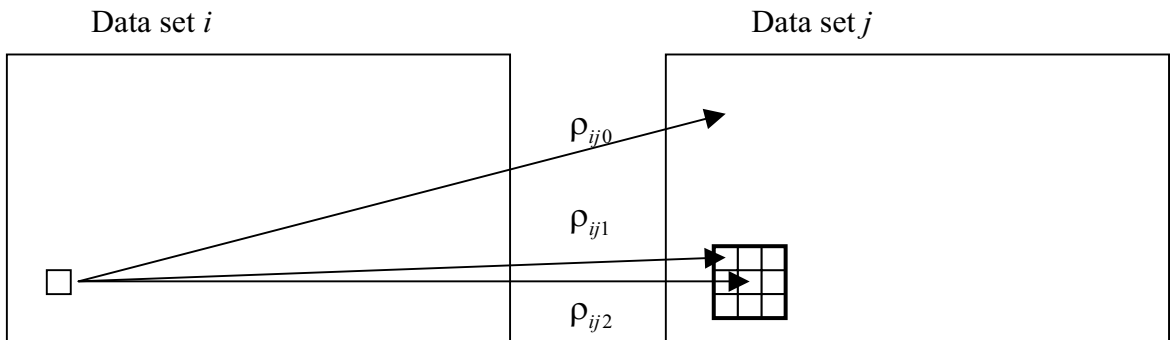
Data set i Data set j

Now suppose that the Z_i extend the lattice on which the i-th data set is defined and that the above correlation structure extends in the natural way.

Define

$$W_i = \begin{bmatrix} Y_i \\ Z_i \end{bmatrix}.$$

Suppose that the data to be modelled are of the form (4.1), and apply the bootstrap described in Section 4.3 with the exception that (4.9) and (4.12) are replaced as follows.

Just as in (4.9), consider all permutations $R_i \rightarrow M_{i(r)} R_i$. However, for each such permutation, allow whole 3 x 3 blocks of cells to be permuted, as illustrated in Figure 4.2. The permutation acts on the centre cell of the block shown in the figure, but induces the mapping shown on the neighbouring cells.

Figure 4.2
Effect of a single bootstrap permutation
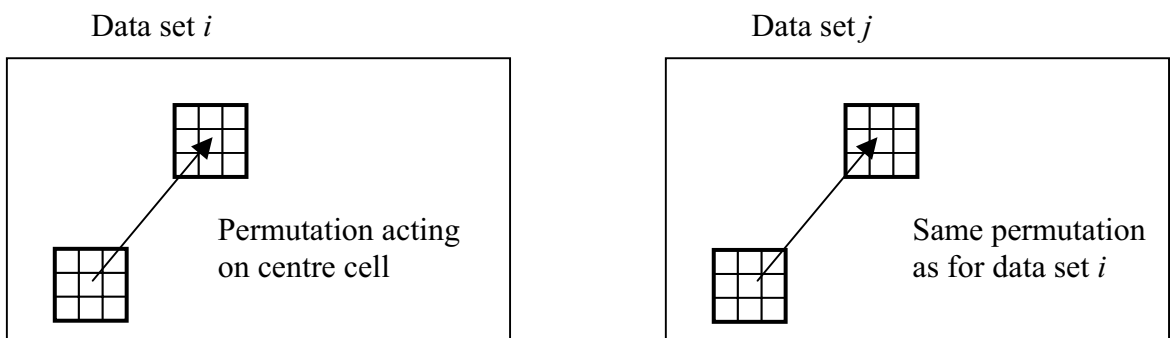
Data set i Data set j

The permutation of whole blocks is reminiscent of the **moving block bootstrap** (Carlstein, 1986, Kunsch, 1989), though that is used for preservation of correlation **within** a data set. Here there is assumed to be no correlation within a data set, and the synchronisation of whole blocks is used to preserve correlation **between** data sets. If, however, within-set correlation were present, that would be preserved too.

It is also reminiscent of the **historical simulation** procedure used to estimate VaR for a banking portfolio subject to correlation between risk factors (Duffie and Pan, 1997).

An exception will arise when the source or target centre cell of a permutation lies on the boundary of the data set. In this case, permuted neighbouring cells that do not make sense are simply ignored, as illustrated in Figure 4.3.

Figure 4.3
Examples of exceptional permutations

Data set *i*   Data set *j*
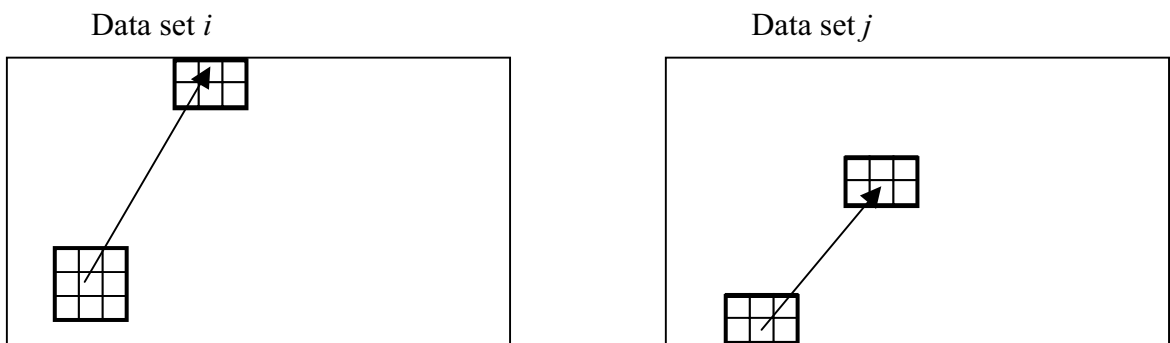
At this stage it might seem natural to construct pseudo-data by the following replacement of (4.9):

$$Y_{i(r)}^* = \hat{Y}_i + \hat{\sigma}_i R_{i(r)}^Y \qquad\qquad (4.9d)$$

where $R_{i(r)}^Y$ is the array of residuals $R_{i(r)k_1k_2}^Y$, with each of these elements calculated as the average, over all permutations, of the residuals mapped to the (k_1, k_2) position in Figure 4.2.

However, as will be demonstrated below, this form of averaging causes the correlations between some of the $R_{i(r)k_1k_2}^Y$ to be less than specified by (4.27). Hence, the correct form of the pseudo-data vector is

$$Y_{i(r)}^* = \hat{Y}_i + \hat{\sigma}_i R_{i(r)}^{Y*} \qquad\qquad (4.9e)$$

where $R_{i(r)}^{Y*}$ is an adjusted form of $R_{i(r)}^Y$ yielding the correct correlation structure.

The correlation structure illustrated in Figure 4.1 is (partially) preserved by synchronising (4.9d) across $i = 1,\ldots,I$ in the sense of matched permutations, ie if a permutation maps $(k_1, k_2) \rightarrow (l_1, l_2)$ in data set i, then the same permutation acts on all other data sets j, as illustrated in Figure 4.2. The fact that the preservation of correlation is only partial is discussed further below.

In the present geometric context, pseudo-forecasts (4.11) represent a geometric extension Z_i of re-sampled residuals from Y_i. This is illustrated in Figure 4.4, where the synchronisation of data sets is also illustrated.

The same boundary issues arise as were illustrated in Figure 4.3, and are dealt with in the same way.

Figure 4.4
Synchronised pseudo-forecasts

Data set i Data set j
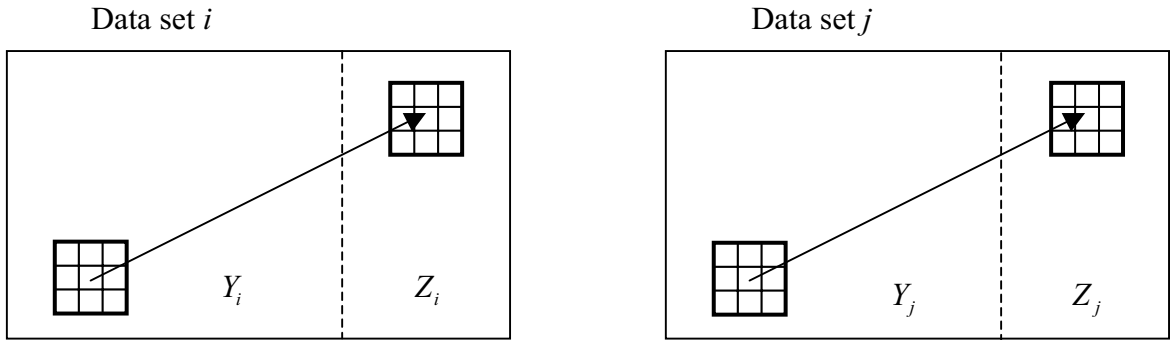

In sympathy with (4.9e), (4.12) is replaced by

$$\eta^{*}_{i(r)} = \hat{\sigma}_i R^{Z*}_{i(r)} \qquad\qquad (4.12e)$$

where $R^{Z*}_{i(r)}$ is a correlation-corrected form of $R^{Z}_{i(r)}$, the array of residuals $R^{Z}_{i(r)l_1 l_2}$ with each of these elements calculated as the average, over all re-samplings (within the bootstrap replication r), of the residuals mapped to the (l_1, l_2) position of Z_i in Figure 4.4.

The fact that (4.9d) only partially preserves the correlation between Y_i and Y_j is illustrated in Figure 4.5. Suppose the two target blocks there cover co-ordinates (k,l) with $k_1 \le k \le k_1 + 5$, $l_1 - 1 \le l \le l_1 + 1$. Then the particular permutation illustrated will have preserved correlation between Y_{i,k_1+1,l_1} and Y_{j,k_1+2,l_1} (for example) but not between Y_{i,k_1+2,l_1} and Y_{i,k_1+3,l_1}.

In fact, the above re-sampling scheme will degrade the excess of the covariance between any pair of neighbouring cells over the background covariance (corresponding to ρ_{ij0}) by a factor of 1-p where p is the proportion of re-samplings in which those cells occur within a single 3 x 3 block.

Figure 4.5
Source of partial correlations
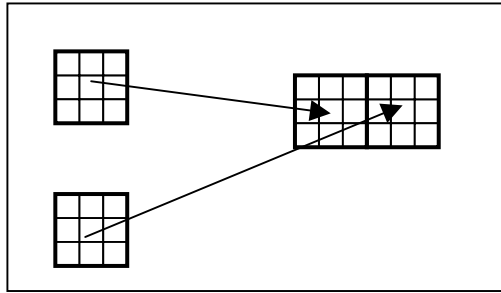
Let C_{ij}^Y denote $Corr\left(R_{i(r)}^Y, R_{j(r)}^Y\right)$, noting that this quantity is independent of r. Let C_{ij} denote the correlation matrix defined by (4.26). Note that C_{ij}^Y will be a distorted version of C_{ij} in which ρ_{ij1} is replaced by $\lambda_{ij,k_1k_2l_1l_2}\rho_{ij1} + \left(1 - \lambda_{ij,k_1k_2l_1l_2}\right)\rho_{ij0}$, for some known factor $\lambda_{ij,k_1k_2l_1l_2}$ between 0 and 1, when it occurs in the $\left(k_1, k_2, l_1, l_2\right)$ position in C_{ij}.

Let $\hat{C}_{ij}^Y$ be the estimate of C_{ij}^Y obtained by estimating ρ_{ij0}, ρ_{ij2} and $\rho_{ij0} + \lambda_{ij,k_1k_2l_1l_2}\left(\rho_{ij1} - \rho_{ij0}\right)$ by means of averages of bootstrapped covariances. For example, the third of these quantities would be estimated by averaging the bootstrapped covariances corresponding to $C_{ij,k_1k_2l_1l_2}$ over the set $S = \left\{i, j, k_1, k_2, l_1, l_2 : i \neq j, \left(k_1, k_2\right) \neq \left(l_1, l_2\right), \left|k_1 - k_2\right| \leq 1, \left|l_1 - l_2\right| \leq 1\right\}$ (as well as averaging over r).

As explained above, the factor $\lambda_{ij,k_1k_2l_1l_2}$ is just the proportion of bootstrap permutations that place cells *(k₁,k₂)* and *(l₁,l₂)* within a single permuted 3x3 block. These quantities depend only on the permutations, as distinct from the data. They may therefore be calculated ahead of the application of the bootstrap, possibly by a prior application of the bootstrap, in which just **positions** (and not the values) of permuted cells are used.

Let $\rho'_{ij1,k_1k_2l_1l_2}$ denote the quantity $\rho_{ij0} + \lambda_{ij,k_1k_2l_1l_2}\left(\rho_{ij1} - \rho_{ij0}\right)$. Then $\left(\hat{\rho}'_{ij1,k_1k_2l_1l_2} - \hat{\rho}_{ij0}\right)$ estimates $\lambda_{ij,k_1k_2l_1l_2}\left(\rho_{ij1} - \rho_{ij0}\right)$, where $\hat{\rho}_{ij0}$ and $\hat{\rho}'_{ij1,k_1k_2l_1l_2}$ are the bootstrap estimates described above. A single estimate of $\left(\rho_{ij1} - \rho_{ij0}\right)$ is then given by

$$\sum\left(\hat{\rho}'_{ij1,k_1k_2l_1l_2} - \hat{\rho}_{ij0}\right) / \sum \lambda_{ij,k_1k_2l_1l_2}$$

where the summations run over all subscripts in the set *S*, and also over *r*.

Combination of this estimate of $\left(\rho_{ij1} - \rho_{ij0}\right)$ with $\hat{\rho}_{ij0}$ then gives a single estimate $\hat{\rho}_{ij1}$ of ρ_{ij1}.

Let C be the n x n correlation matrix consisting of n_i x n_j blocks C_{ij}, and define $\hat{C}, C^Y$ and $\hat{C}^Y$ similarly.   Let $\Gamma, \hat{\Gamma}, \Gamma^Y, \hat{\Gamma}^Y$ be the covariance matrices corresponding to $C, \hat{C}, C^Y \hat{C}^Y$. Consider the diagonalised forms of $\hat{\Gamma}, \hat{\Gamma}^Y$:

$$\hat{\Gamma} = PDP^T, \qquad \hat{\Gamma}^Y = P^Y D^Y \left(P^Y\right)^T \tag{4.28}$$

where D, D^Y are diagonal and P, P^Y orthogonal.
Then

$$\hat{\Gamma} = L^Y \hat{\Gamma}^Y \left(L^Y\right)^T \tag{4.29}$$

where

$$L^Y = PD^{\frac{1}{2}} \left(D^Y\right)^{-\frac{1}{2}} \left(P^Y\right)^T \tag{4.30}$$

Now $\hat{\Gamma}^Y$ is an estimate of $\Gamma^Y = Var\left[R_{(r)}^Y\right]$, where $R_{(r)}^Y$ is the n-vector obtained by stacking the vectors $R_{i(r)}^Y, i = 1,...,I$. Therefore, $\hat{\Gamma}$ is an estimate of $Var\left[L^Y R_{(r)}^Y\right]$.

But $\hat{\Gamma}$ is also an estimate of the covariance matrix Γ corresponding to C, the correlation matrix summarising (4.27). This means that, under the definition

$$R_{(r)}^{Y*} = L^Y R_{(r)}^Y \tag{4.31}$$

the correlation matrix of $R_{(r)}^{Y*}$ will approximate C, and so this definition will yield the correct $R_{(r)}^{Y*}$ for substitution in (4.9e).

By similar reasoning, the use of $R_{i(r)}^Z$ in (4.12e) would only partially preserve the correlation between Z_i and Z_j. A process precisely parallel to the above yields a matrix L^Z such that the residual vector $R_{i(r)}^{Z*}$ required by (4.12e) is given by

$$R_{(r)}^{Z*} = L^Z R_{(r)}^Z \tag{4.32}$$

Remark 4.4. It would be necessary to check $\hat{C}^Y$ and $\hat{C}$ for positive definiteness. In the event that it is not found, a degenerate choice of (4.30) might be indicated, eg $\rho_{ij1} = \rho_{ij0}$ (synchronised individual data points (see (4.14)) or $\rho_{ij2} = \rho_{ij1}$.

5.   Practical application

Sections 4.4.2 to 4.4.4 consider specific correlation structures and produce forms of synchronous bootstrap that generate pseudo-data with the same correlations. In these situations, the bootstrap will be said to be **adapted** to the given correlation structure.

A poorly adapted bootstrap will tend to destroy correlation between data points. An example of this is given in Section 4.4.1, where poor adaptation consists of bootstrapping correlated data sets independently.

For this reason, it is to be expected that measured correlations between bootstrap forecasts $Z_{i(r)}^*$ and $Z_{j(r)}^*$ will be maximised when the form of bootstrap is close to adapted to the true correlation structure.

In many practical situations, one may have a range of possible correlation structures in mind as candidates, but little idea of the likelihood of each. Examples are given in Section 6.

It is desirable in these situations to allow the data to select the adapted form of bootstrap.   This is done by implementing a separate form of the bootstrap adapted to each candidate correlation structure, and selecting that which yields the greatest correlations.

One needs to decide on a criterion for the "greatest correlations" when considering estimates of a whole matrix of correlations $C_{ij,kl}$. It may be possible to do this without considering correlations explicitly.   For example, if the objective is to estimate z defined by (4.13), then one might choose the bootstrap which maximises the absolute difference between the estimate of Var $[z]$ and the estimate of the same quantity obtained by independent bootstrapping of distinct data sets.

6.   Extensions

6.1   Other geometrically related subsets

As stated at the beginning of Section 4.4.4, the synchronisation of geometrically related subsets was described there only in terms of an example. The example involved a specific geometric relationship.   Clearly, other relationships are possible.

For example, one might wish to increase the size of the block illustrated in Figure 4.1 from 3 x 3 to 5 x 5 in order to extend the correlation from "nearest neighbour" to "nearest neighbour but one".   The bootstrap adapted to this structure can be developed along exactly the same lines as Section 4.4.4.

While the correlation structure may be made increasingly complex, and the bootstrap adapted accordingly, this does not come without cost. The greater the complexity, the greater the effort in co-ordinating the bootstrap's permutations of whole subsets, the more numerous the boundary problems associated with those permutations, and the more extensive the required correction of correlation biases such as described by the factors λ in Section 4.4.4.

6.2   Sampling with replacement not permitted

All bootstrapping in the foregoing sections has been defined in terms of permutations, or sampling without replacement. Sampling with replacement does **not** preserve correlation structures.

Sampling with replacement would be effected by re-defining the re-sampling matrix of Section 4.1, no longer requiring that rows be distinct. Such a re-sampling matrix M has the following multiplication properties:

$$\text{For } k = l, M_{ks}M_{lt} \quad = 1 \text{ for } s = t$$
$$= 0 \text{ otherwise}$$
$$\text{For } k \neq l, M_{ks}M_{lt} \quad = 1 \text{ for a single pair } s, t \left(\text{no longer } s \neq t \text{ necessarily}\right)$$
$$= 0 \text{ otherwise.}$$

The allowance here that $M_{ks}M_{lt}$ may be 1 for $k \neq l, s = t$, means that (4.18) can take the value $C_{ij,ss}$ for $k \neq l$, in which case the re-sampling matrix M has failed to preserve correlation structure by transferring an on-diagonal correlation to off-diagonal within C_{ij}.

This can be easily understood by consideration of an example of synchronisation of individual data points in which all rows of M are the same. In this case, (4.9a) becomes

$$Y_{i(r)}^* = \hat{Y}_i + \hat{\sigma}_i R_{is} 1 \qquad\qquad (6.1)$$

where R_{is} is the s-th component of R_i and 1 is a vector with all components unity. In other words, the single residual R_{is} has been re-sampled for all components of $Y_{i(r)}^*$ and so, by synchronisation, R_{is} is sampled for all components of $Y_{i(r)}^*, i = 2, ..., I$.

Then $Corr\left[Y_{i(r)}^*, Y_{j(r)}^*\right]$ will (apart from a correction of $0\left(n^{-1}\right)$) have all elements the same, and equal to $Corr\left(R_{is}, R_{js}\right) = C_{ij,ss}$, as found algebraically above.

It follows that, strictly, only re-sampling **without** replacement should be used within a synchronous bootstrap. In practice, however, the error introduced by a sampling with replacement will often be relatively small. Provided that the dimensions n_i of the vectors Y_i are not too small, the relative incidence of repetitions in sampling with replacement will be low, and the distortion of correlation structure caused by them correspondingly low.

6.3   Subsets of different dimensions

The observations vectors Y_i were defined in (4.1) as of dimension n_i, $i = 1,...,I$. However, for subsequent analysis, it was assumed that $n_1 = ... = n_I = n$. Consider now the implications of allowing the n_i to differ one from another.

Consider, for example, the synchronisation of individual data points (Section 4.4.2) within this extended framework. The correlation structure (4.14) is still meaningful provided that the "correlation matrix" C_{ij} is understood to be of dimension $n_i \times n_j$.

The required bootstrap design is as follows. Suppose, without loss of generality, that $I = 3$ and $n_1 \le n_2 \le n_3$. Then vector Y_2 needs to be decomposed into two sub-vectors: one of dimension n_2 containing the components that have correlation ρ_{211} with their counterparts in Y_1, and the other of dimension $n_2 - n_1$ containing no such components.

Likewise, Y_3 needs to be decomposed into three sub-vectors of respective dimensions $n_1, n_2 - n_1$ and $n_3 - n_2$. Bootstrapping is then synchronised on sub-vectors of dimension n_1, and separately on sub-vectors of dimension $n_2 - n_1$. Other sub-vectors are not synchronised. The synchronisation is illustrated schematically in Figure 6.1.

Figure 6.1
Bootstrapping data sets of different dimensions
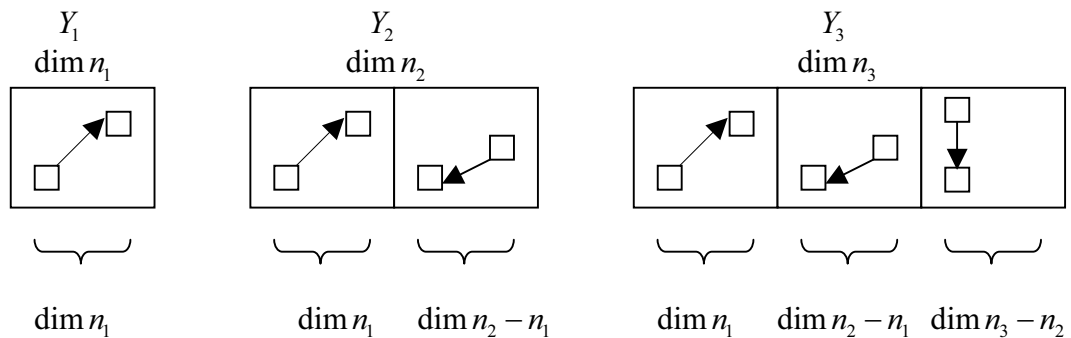
7.   Numerical examples

7.1   Insurance data

The bootstrap designs discussed in Sections 4.4.2 and 4.4.3 were tested numerically on hypothetical insurance data. The data take the form of **run-off triangles** for different lines of business (LoBs) of an insurer.

Thus each Y_i denotes a triangular array of data rather than a vector and the notation Y_{ik} introduced at the start of Section 4.4.2 is conveniently extended to Y_{ikl}, denoting the (k,l) element of triangle $Y_i, k = 1,...,n_i; l = 1,...,n_i - k + 1$.

Here *k* denotes **accident year** and *l* **development year** in insurance parlance (Taylor, 2000). Years in which claims experience occurs are represented by diagonals in the triangle. The latest year of observation corresponds to the n_i-th diagonal.

Note that the use of n_i in this dimensioning is different from that in previous sections. In fact, the dimension of Y_i in the present framework is $\frac{1}{2} n_i (n_i + 1)$.

The vector Z_i to be forecast consists of the future diagonals, ie the $(n_i + 1) - \text{th},...,(2n_i - 1) - \text{th}$, as far as they relate to accident years $1,...,n_i$. That is, Z_i consists of $\{Y_{ikl}, k = 1,...,n_i; l = n_i - k + 2,...,n_i\}$. If the Y_{ikl} denote claim payments, then the ½$n_i(n_i$-1) values subject to forecast represent all future claim costs in respect of the n_i accident years. The total of these is called the **loss reserve**.

It may be noted that the number of potentially distinct correlations in this situation is large. For the case n_i=n, it is ½*I(I-1)*x½*n(n-1)*. This can be a large number. For example, in the case *I*=10, *n*=20, it is 8,550. this will be too many to estimate from the data, and allowance must be made for them either by introducing some structure to the correlations, such as a SUR framework, or by incorporating them implicitly, such as in a synchronous bootstrap.

The actuarial literature contains a few examples of the inclusion of such correlations in loss reserving procedures, but sometimes in a limited form.

For example, Brehm (2002) estimates correlations between LoB-specific loss reserves on the assumption that those correlations arise from dependencies between the rates of superimposed inflation (diagonal effects) influencing the LoBs.

Braun (2004) considers the chain ladder model with (in the terminology of earlier sections) point-wise correlations between age-to-age factors, and relies on empirical estimation of them.

As noted earlier, Kirschner, Kerley and Isaacs (2002) consider the case of a point-wise bootstrap.

In each example below, *I* =3, n_i =20 and Y_1, Y_2, Y_3 have identical marginal distributions, but have correlations superimposed. The marginal distributions take the form:

$Y_{ikl} \sim$ Gamma   *(7.1)*

$E[Y_{ikl}] = exp[\beta_{k0} + \beta_{k1} (l+1) + \beta_{k2} \, log(l+1)]$ *(7.2)*

$Var[Y_{ikl}] = \varphi \, E[Y_{ikl}]$ *(7.3)*

Note that the distribution of Y_{ikl} is independent of i. The expected value (7.2), as a function of l, is a Hoerl curve, i.e. is gamma shaped. By (7.2) and (7.3), the coefficient of variation (CoV) of Y_{ikl} is $\{\varphi/E[Y_{ikl}]\}^{1/2}$.

7.2   Point-wise bootstrapping

In the model (7.1)-(7.3), the parameter vector β_k was made independent of k, i.e. the same Hoerl curve applied to each accident year as well as to each LoB. The parameter values were as set out in Table 7.1.

Table 7.1
Parameter values

Parameter	Value
β_{k0}	+5.022
β_{k1}	-0.4
β_{k2}	+2.4
Φ	100

Data triangles for the three LoBs were sampled according to this model, with correlated noise, as defined in Section 4.2, introduced by means of the following steps applied to each pair (k,l):

- Sample x from a trivariate standard random normal.
- Transform x to $u=Ax$ where A is the matrix

$$\begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{bmatrix}$$

and the resulting vector u consists of identically distributed components with all pairwise correlations equal to approximately 80%.

- Transform u to $y = \Gamma_{kl}^{-1} [\Phi(u^{(s)})]$ where $u^{(s)}$ is the standardised version of u and Γ_{kl} is the gamma d.f. associated with Y_{ikl}.
- Let y be a realisation of the vector $[Y_{1kl}, Y_{2kl}, Y_{3kl}]^T$.

Each of the resulting Y_{ikl} has d.f. Γ_{kl} and, $Corr[Y_{ikl}, Y_{jkl}]$ is independent of i,j for $i \neq j$ and is equal to about 80%.

This procedure is carried out independently for each pair (k,l), i.e. the values of x are sampled independently for distinct (k,l). As a result,

$$Corr[Y_{ikl}, Y_{jst}] = 0 \text{ for } (k,l) \neq (s,t) \tag{7.4}$$

This creates a correlation structure of the form (4.13), to which the synchronised bootstrap of individual points described in Section 4.4.2 is adapted. This will be referred to as a **point-wise bootstrap**. It is applied to the data triangles $\{Y_{ikl}\}$ generated as above, with the results of 10,000 replications set out in Table 7.2.

Here, and in the subsequent numerical examples, sampling with replacement is considerably easier to implement that sampling without replacement, and has been used even though theoretically less accurate. This is justified by the final remark of Section 6.2.

The true values in the table are obtained by simulation. Values of $E[Y_{ikl}]$ are generated according to (7.2), applying equally to the LoBs i=1,2,3. For each (k,l), correlated gamma noise terms, generated as described above, are added. This is done independently for different (k,l).

The true loss reserve for LoB i is calculated as the sum of the simulated values of Y_{ikl} over future cells (k,l). The correlated noise of about 80% translates into a similar correlation between loss reserves of pairs of LoBs.

Table 7.2
Results of point-wise bootstrap

Basis of estimation	Pair-wise correlation of LoB loss reserves	CoV of aggregate loss reserve across 3 LoBs
		%
True (simulated)	0.81	5.4
Independent bootstrap	-0.00	3.0
Synchronous point-wise bootstrap	0.79	5.0

Naturally, conventional independent bootstrapping produces no correlation between the loss reserves of different LoBs. The CoV of loss reserve for a single LoB is 5.4%, and is reduced by a factor of √3 to 3.0% when the loss reserve is aggregated over all three LoBs.

However, the correlation between LoBs prevents this degree of reduction, and the true CoV of the aggregate is 5.4%. The point-wise synchronous bootstrap captures the majority of the correlation between LoB-specific loss reserves, and therefore produces a substantially more accurate estimate of CoV of the aggregate.

7.3   Row-wise bootstrapping

Data

Model (7.1)-(7.3) was applied once again, with the following variation. As in Section 7.2, the parameters β_{k1} and β_{k2} were made independent of k, i.e. the same Hoerl curve applied to each LoB, and also, up to a multiplier, to each accident year also. These two parameters, and φ, were assigned the same values as in Table 7.1.

However, β_{k0} followed a random walk over k, with the same initial value β_{10} as in Table 7.1, and single-step variance of 0.4.

The model specification was therefore:

$$Y_{ikl} \sim \text{Gamma} \tag{7.1}$$

$$E[Y_{ikl}] = exp[\beta_{k0} + \beta_1 (l+1) + \beta_2 \, log(l+1)] \tag{7.2a}$$

$$Var[Y_{ikl}] = \varphi \, E[Y_{ikl}] \tag{7.3}$$

Data triangles for the three LoBs were sampled according to this model, but this time with independent process error for the different LoBs. The model of each data set is now deliberately mis-specified to omit the accident year multiplier. That is, instead of the correct specification (7.2a), the following model was fitted:

$$E[Y_{ikl}] = exp[\beta'_0 + \beta'_1 (l+1) + \beta'_2 \, log(l+1)] \tag{7.2b}$$

where the parameters in the mis-specified model are distinguished by a prime, and the notation indicates that the same set of parameters is estimated for each of the three data triangles.

This creates a bias vector b_i, as in Section 4.2, whose *(k,l)* component is of the form:

$$b_i = exp[\beta_{k0} + \beta_1 (l+1) + \beta_2 \, log(l+1)] - exp[\beta'_0 + \beta'_1 (l+1) + \beta'_2 \, log(l+1)] \tag{7.5}$$

and which may be observed to be independent of i.

The mis-specification with respect to rows of the triangles creates parameter correlation of the form (4.20), where the subsets N_q are the rows. A bootstrap with synchronised rows, as described in Section 4.4.3, is therefore applied. This will be referred to subsequently as a **row-wise bootstrap**.

Performance measurement

As in Section 4, let Y_i, Z_i denote the past and future data respectively in LoB i. The measurement of bootstrap performance will be concerned with Z_i.

Let

$$\zeta_i = E[Z_i] \tag{7.6}$$

which is the **true** mean, given by an equation like (7.2a), and let ζ'_i denote the corresponding mean in the mis-specified model (7.2b).

Let $\hat{\zeta}'_i$ denote the estimate of ζ'_i derived from the data set Y_i and let $Z^*_{i(r)}$ denote the row-wise bootstrap forecast of Z_i.

The **prediction error** in the bootstrap forecast is

$$Z_i - Z^*_{i(r)} = (Z_i - \zeta_i) \qquad \text{[process error]}$$
$$+ (\zeta_i - \zeta'_i) \qquad \text{[specification error]}$$
$$+ (\zeta'_i - \hat{\zeta}'_i) \text{[parameter error]}$$
$$+ (\hat{\zeta}'_i - Z^*_{i(r)}) \qquad \text{[internal bootstrap error]} \qquad (7.7)$$

It will be convenient to write this in the notation

$$Z_i - Z^*_{i(r)} = \varepsilon_i + \xi_i + \chi_{i(r)} \qquad (7.8)$$

where ε_i, $\chi_{i(r)}$ denote the process and internal bootstrap errors respectively and ξ_i denotes the sum of the specification and parameter errors.

Note that:

- ξ_i is a deterministic quantity;
- ε_i is stochastically independent of ε_j for jγi;
- ε_i is stochastically independent of ξ_j for any i,j.

and typically

$$E[\varepsilon_i] = 0 \qquad\qquad (7.9)$$
$$\xi_i \neq 0 \qquad\qquad (7.10)$$
$$E[\chi_{i(r)}] \neq 0 \qquad\qquad (7.11)$$

Relation (7.10) relates mainly to the existence of specification error. Relation (7.11) reflects the fact that forecasts from a synchronous bootstrap will contain a bias that tends to correct for specification error.

Now consider the **predictive covariance**

$$\begin{aligned} E[(Z_i - Z^*_{i(r)})(Z_j - Z^*_{j(r)})] &= E[(\varepsilon_i + \xi_i + \chi_{i(r)})(\varepsilon_j + \xi_j + \chi_{j(r)})] \\ &= \delta_{ij} \, Var[\varepsilon_i] + Cov[\chi_{i(r)}, \chi_{j(r)}] + E[\xi_i + \chi_{i(r)}] \, E[\xi_j + \chi_{j(r)}] \\ &= \delta_{ij} \, Var[\varepsilon_i] + Cov[\chi_{i(r)}, \chi_{j(r)}] + \{\xi_i + E[\chi_{i(r)}]\} \{\xi_j + E[\chi_{j(r)}]\} \end{aligned}$$
$$(7.12)$$

It will now be convenient to write

$$\xi_i = s_i + p_i \qquad\qquad (7.13)$$

recognising ξ_i as the sum of specification and parameter errors respectively. Substitution of this in (7.12) gives

$$\begin{aligned} E[(Z_i - Z^*_{i(r)})(Z_j - Z^*_{j(r)})] &= \delta_{ij} \, Var[\varepsilon_i] + Cov[\chi_{i(r)}, \chi_{j(r)}] + \{p_i + E[\chi_{i(r)}]\} \{p_j + E[\chi_{j(r)}]\} \\ &+ s_i s_j + s_i \{p_j + E[\chi_{j(r)}]\} + s_j \{p_i + E[\chi_{i(r)}]\} \end{aligned} \qquad (7.14)$$

Express this relation in the notation

$$(PC)_{ij} = (PC)^{ns}_{ij} + (PC)^{s}_{ij} \qquad\qquad (7.15)$$

where $(PC)_{ij}$ denotes predictive covariance, $(PC)^{s}_{ij}$ denotes the last three members of (7.14), involving specification error and $(PC)^{ns}_{ij}$ denotes the remainder of the right side of (7.14), not involving specification error.

Define z as in (4.12). Then the **mean square error of prediction** of z is

$$MSEP[z] = E[(z - z^{*}_{(r)})^2] = \Sigma_{ij} \, 1^{T}(PC)_{ij}1$$
$$= MSEP^{ns}[z] + MSEP^{s}[z] \qquad\qquad (7.16)$$
where

$$MSEP^{ns}[z] = \Sigma_{ij} \, 1^{T}(PC)^{ns}_{ij}1, \quad MSEP^{s}[z] = \Sigma_{ij} \, 1^{T}(PC)^{s}_{ij}1$$

Also define

$$MSEP_{ind}[z] = \Sigma_{i} \, 1^{T}(PC)_{ii}1 \qquad\qquad (7.17)$$

which is the MSEP in the case of independence between all LoBs, and define $MSEP^{s}_{ind}[z]$ and $MSEP^{ns}_{ind}[z]$ similarly. Then $MSEP[z] - MSEP_{ind}[z]$ is the inter-LoB covariance contribution to $MSEP[z]$.

The true MSEP is $MSEP[z]$. However, in practice, one attempts to eliminate specification error from a model, and so assumes that $MSEP^{s}[z]=0$, so that the observable MSEP is $MSEP^{ns}[z]$. Therefore, a measure of the efficiency of the synchronous bootstrap in capturing the inter-LoB covariances is

$$\text{Efficiency} = \frac{MSEP^{ns}[z] - MSEP^{ns}_{ind}[z]}{MSEP[z] - MSEP_{ind}[z]} \qquad\qquad (7.18)$$

Numerical results

The properties of the row-wise bootstrap depend on the degree of mis-specification that it needs to correct. If there is little variation in the row parameters β_{k0}, there will be correspondingly little parameter correlation, and little need for the bootstrap to estimate it efficiently.

Five different examples were considered, each consisting of a triple of data sets for the three LoBs. The random walk of parameters β_{k0} differed between the examples. Figure 7.1 plots (on a logarithmic scale) the sampled values of $\exp(\beta_{k0} - \beta_{10})$, which is the multiplier that scales the expected values of accident year k relative to year 1.

In Example 1 claims experience drifts to very high levels in the later accident years. Example 5 provides more or less the mirror image, with claims experience drifting to low levels. Examples 2, 3 and 4 are more moderate.

The effects of these parameters are illustrated in Table 7.3, which displays the mean and prediction error of the loss reserve for a single LoB, as estimated by a simple (unsynchronised) bootstrap with 1,000 replications. The prediction error takes the form of the root mean square error of prediction (RMSEP), including specification error, and estimated by simulation. The effect of the parameter drift on prediction error is evident in Examples 1 and 5.

Figure 7.1
Accident year multipliers
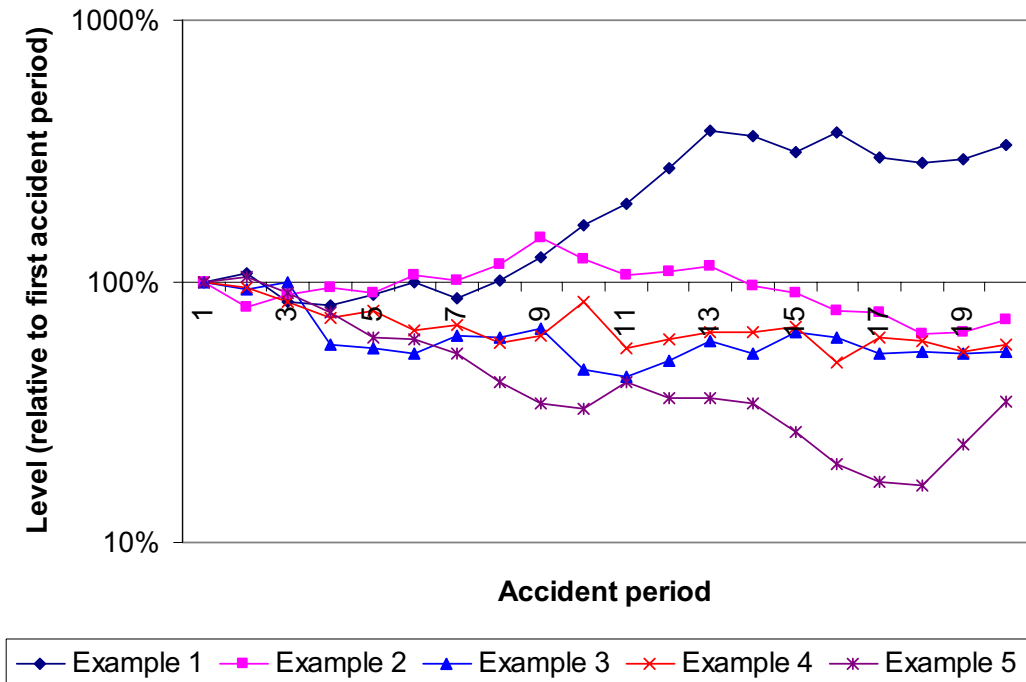
Table 7.3
Mean and prediction error of loss reserve for a single LoB

Example	Mean loss reserve	RMSEP
	$	% of mean
1	211,325	60
2	47,572	22
3	39,610	32
4	42,831	33
5	27,840	160

Table 7.4 displays values of the efficiency statistic (7.18) for the row-wise bootstrap applied to the five examples, with 10,000 replications in each case. The table also displays values of $\{1 - \text{MSEP}_{ind}[z] / \text{MSEP}[z]\}$, representing the relative contribution of the inter-LoB correlations to $\text{MSEP}[z]$. It has a maximum value of $\frac{2}{3}$.

The efficiency of the row-wise bootstrap is displayed in the column headed "row-wise". The column headed "point-wise" displays the corresponding efficiency measure for the point-wise bootstrap.

Just the presence of noise in the data will prevent any forecast procedure from capturing full parameter correlation. However, the row-wise bootstrap is seen to perform well in cases of high correlation and moderately well in cases of lower correlation.

Table 7.4
Efficiency of row-wise and point-wise bootstrap estimation of prediction error

Example	Contribution of inter-LoB correlations to prediction error	Efficiency	
		Row-wise bootstrap	Point-wise bootstrap
	%	%	%
1	66.4	73	32
2	59.7	60	9
3	63.9	24	39
4	63.1	26	25
5	65.0	103	51

Somewhat surprisingly, the point-wise bootstrap also performs moderately well, though it does not match the row-wise version in cases of high correlation.

It should be recognised that, despite the semi-effectiveness of the synchronous bootstrap in capturing the contribution of parameter correlation to RMSEP, the existence of specification error nevertheless causes a substantial part of RMSEP to remain unrecognised. Table 7.5 illustrates.

Table 7.5
Prediction error unrecognised due to specification error

Example	RMSEP (as % of mean)		Excluding specification error according to row-wise synchronised bootstrap
	Including specification error		
	According to unsynchronised bootstrap	According to row-wise synchronised bootstrap	
	%	%	%
1	60	50	10
2	22	15	10
3	32	24	9
4	33	24	10
5	160	62	80

The first column of results in Table 7.5 is simply reproduced from Table 7.3. The synchronised bootstrap RMSEP including specification error is based on $(PC)_{ij}$ from (7.15); excluding specification error based on $(PC)^{ns}_{ij}$.

The table illustrates how the RMSEP to be estimated (i.e. including specification error) is reduced by the introduction of the synchronous bootstrap. This occurs because this form of bootstrap assigns part of what is otherwise interpreted as noise to row effects.

The final column of the table contains the estimate of RMSEP obtained in practice where specification cannot be measured and is assumed zero. It is seen that 80% of the RMSEP is lost in one of the cases of high parameter correlation, and of the order of half in the cases of milder correlation.

Such losses are, of course, serious. However, two points should be made:

- The examples, particularly those involving high parameter correlation, are rather exaggerated. In practice, specification errors of the magnitude contained in these examples would hardly go unnoticed.
- The failure to correct for specification error is not a failure of the synchronous bootstrap as such. Indeed, Table 7.5 demonstrates that this form of bootstrap makes a significant contribution to correcting for such errors. Specification error will bedevil any form of forecast. This accepted, the synchronous bootstrap performs reasonably well in recognising inter-LoB dependencies.

8.   Acknowledgements

References

Braun, C. (2004). The prediction error of the chain ladder method applied to correlated run-off triangles. **Astin Bulletin**, 34, 399-423.

Brehm, P.J. (2002). Correlation and the aggregation of unpaid loss distributions. **Casualty Actuarial Society Fall Forum on Reserving Call Papers**, 1-21.

Carltstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. **The annals of statistics**, 14, 1171-1179.

Duffie, D. and Pan, J. (1997). An overview of value at risk. **Journal of Derivatives**, 4(3),7-49.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. **Annals of Statistics**, 7, 1-26.

Kirschner, G.S., Kerley, C. and Isaacs, B. (2002). Two approaches to calculating correlated reserve indications across multiple lines of business. **Casualty Actuarial Society Fall Forum on Reserving Call Papers**, 211-245.

Kunsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. **The annals of statistics**, 17, 1217-1241.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalised linear models. **Journal of the Royal Statistical Society, Series A**, 135, 370-384.

Srivastava, V.K. and Giles, D.E. (1987). **Seemingly unrelated regression equation models: estimation and inference**. Marcel Dekker, New York.

Taylor, G.C. (2000). **Loss reserving: an actuarial perspective**. Kluwer Academic Publishers. London, New York, Dordrecht.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregate bias. **Journal of the American Statistical Association**, 57,346-368.