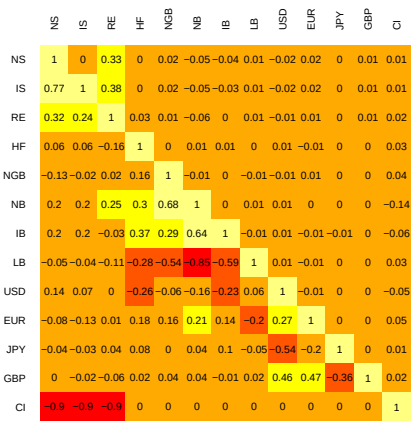
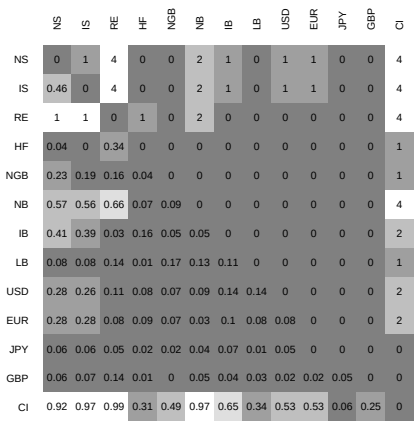
Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Rehabilitation of improper correlation matrices





Note no SAMBA/03/09

Authors Arnoldo Frigessi
Anders Løland
Antonio Pievatolo
Fabrizio Ruggeri

Date 29th January 2009

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Norwegian Computing Center

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modeling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Rehabilitation of improper correlation matrices**
Authors	**Arnoldo Frigessi** <arnoldo.frigessi@medisin.uio.no> **Anders Løland** <anders.loland@nr.no> **Antonio Pievatolo** <antonio.pievatolo@mi.imati.cnr.it> **Fabrizio Ruggeri** <fabrizio@mi.imati.cnr.it>
Date	29th January 2009
Publication number	SAMBA/03/09

Abstract

The simplest way to describe the dependence for a set of financial assets is their correlation matrix. This correlation matrix can be improper when it is specified element-wise. We describe a new method for obtaining a positive definite correlation matrix starting from an improper one. The expert's opinion and trust on each pairwise correlation is described by a beta distribution. Then, by combining these individual distributions, a joint distribution over the space of positive definite correlation matrices is obtained using the Cholesky factorisation, and its mode constitutes the new proper correlation matrix. The optimisation is complemented by a visual representations of the entries that were most affected by the legalisation procedure. We also sketch a Bayesian approach to the same problem.

Keywords	Bayesian inference, beta distribution, expert opinions, nearness problem, positive semidefinite matrix
Target group	
Availability	Open
Project	SFI02-Risk
Project number	220302
Research field	Finance, insurance and commodities
Number of pages	19

Contents

1 Introduction

The study of the dependence of financial assets is the key ingredient in a better understanding of the complex mechanisms underlying financial dynamics. Also, capturing the proper dependence between financial assets allows for diversification benefits. Indeed, it can be argued that the study of dependence is at the core of finance. There is a vast literature, recent and more classical, presenting methods to estimate dependence and to model dependent behaviours (Embrechts et al., 1999, 2002; Longin and Solnik, 2001). The simplest way to describe the dependence of a set of financial assets is their correlation matrix. This captures the linear component of pairwise dependence and is a basic element generally in multivariate statistics (Mardia et al., 1979), and more specifically in multivariate financial time series modelling, for classical methods like the Capital Asset Pricing Model (CAPM) and Markowitz portfolio theory, and more recent applications (Dimakos and Aas, 2004).

A correlation matrix is required to be a symmetric positive semidefinite matrix with unit elements on its diagonal. We refer to such matrices as proper correlation matrices. Estimation of a correlation matrix is based on observed time series of the n assets, $(Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t})$, for $t = 1, 2, ..., T$, under appropriate stationarity assumptions. It is easy to obtain empirical proper correlation matrices with the necessary non-negative eigenvalues.

There are at least two common situations were the construction of a proper correlation matrix is difficult. This is the case when not all assets are observed in the same time points (Higham, 2002). For example, the first two assets $(Z_{1,t}, Z_{2,t})$ are available every day t, while a third asset $(Z_{3,t})$ is sampled only weekly. Here the three pairwise correlations do not necessarily compose a proper correlation matrix. More generally, when sample correlation matrices are constructed from historical data, but data are not available for every time point for all variables of interest, the matrix with such pairwise correlations is easily improper. A second situation which leads to improper correlation matrices is when some assets are not observed at all, but an expert opinion is obtained on its correlation with the other assets. Relevant credit or operational loss data might for example be hard to obtain. In this case, expert opinions have to be called upon (Dimakos and Aas, 2004; Medova, 2000).

Once an improper correlation matrix is obtained, symmetric with unit elements on its diagonal, but not positive semidefinite matrix, it has to be transformed into a proper correlation matrix in order to proceed with the analysis. The objective is then to find a proper matrix, which is as close as possible to the illegal one. An important observation is that this legalised matrix can be very different from the original one, even if only few elements of the improper matrix originate from unmatched time series. In this paper we propose two new methods, which rectify the illegal matrix and deliver a proper (actually a positive-definite) correlation matrix, allowing to control the perturbation of each entry.

The methods proposed in the literature (Grubišić and Pietersz, 2007; Higham, 2002; Pietersz and Groenen, 2004) measure the distance between matrices by means of the Frobenius norm. Let us denote by $\boldsymbol{C}$ the improper correlation matrix. The problem is

to find the proper correlation matrix $\boldsymbol{M}$ which minimises the Frobenius norm of the difference

$$||\boldsymbol{M} - \boldsymbol{C}||_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |m_{ij} - c_{ij}|^2. \tag{1}$$

An extension is given by the weighted Frobenius norm (Pietersz and Groenen, 2004),

$$||\boldsymbol{M} - \boldsymbol{C}||_W^2 = ||\boldsymbol{W}^{1/2}(\boldsymbol{M} - \boldsymbol{C})\boldsymbol{W}^{1/2}||_F^2 \tag{2}$$

where $\boldsymbol{W}$ is a given, fixed, symmetric positive definite weight matrix. The optimal $\boldsymbol{M}$ matrix, according to either criteria, is found by numerical minimisation.

While the un-weighted Frobenius norm does not allow control of how the entries of the improper matrix are perturbed, so does the weighted version, through the weight matrix $\boldsymbol{W}$, assigned by the user. If a weight w_{ij} is high for the correlation between the two assets i and j, then the correlation distance (1) for this correlation, if possible, is lower at the expense of other correlations with a lower weight. Because $\boldsymbol{W}$ must be a positive definite matrix, there is no complete liberty in choosing weights. This is indeed a significant problem in practice. In general, how confidence in the elements of $\boldsymbol{C}$ is expressed through the weight matrix $\boldsymbol{W}$ is not intuitive for the user. Furthermore, the norm (2) is symmetric around $\boldsymbol{C}$, while for more extreme correlations (close to -1 or 1), it is more natural to use a non-symmetric distribution.

Our approach assumes that confidence in every pairwise correlation of the improper matrix is described by a distribution with mean in the current value. We propose to use the beta distribution, scaled to $[-1, +1]$, which is sufficiently flexible, though easy to utilise in practice. In this way we express the problem by means of a probability density, where the unknown parameters, here the elements of the proper correlation matrix, each follow a beta distribution. We propose an algorithm that maximises the product of such beta densities within the set of proper (actually positive definite) correlation matrices. This is generally different from minimising (2), which can be seen as a special case.

It is important to reveal how the entries of the optimal proper matrix differ from the corresponding elements in the original improper matrix. We present new ways of visualising what we call transformation hotspots, that is positions where the correlation has been adjusted significantly. This allows a detailed monitoring of the effects of the procedure.

Our second approach makes use of the actual time series, from which the improper correlation matrix originate. We combine the historic data with additional expert opinions using a Bayesian approach. We consider the coefficients of the observed, invalid correlation matrix $\boldsymbol{C}$ as realisations of random variables centred at the coefficients of a proper correlation matrix Θ, on which we have prior opinions. Inference relies on Markov chain Monte Carlo in this case, and is therefore more time demanding. Hence we do not present results for this second approach, which should be regarded as a proof of concept, which can be fully implemented as sketched in this paper. For the first method, inference is much more rapid, as it is based on a numerical algorithm, and we show results for this situation only. Our approach provides a tool for analysts and practitioners. It is easy

to use, able to handle coherently both stringent and vague knowledge and expectations, and is ready to be inserted into the pipeline of risk analysis and decision making.

This paper proceeds as follows. In Section 2, we present the likelihood based approach, using beta densities. We describe the Bayesian approach, combining data and expert opinions, in Section 3. The algorithms are described in Section 4. We illustrate the use of the first method on an example for a Norwegian life insurance company in Section 5. Finally, we conclude in Section 6.

2 Beta correlation matrices

In this section we introduce a stochastic model for correlation matrices. The model is used as a tool for describing the willingness to relax from a current improper pairwise correlation matrix. It is based on the univariate beta distribution.

Let us denote by $\boldsymbol{C}$ the improper matrix containing the available, but incoherent, correlation values. This includes pairwise correlations estimated on data and possibly additional assets, whose correlation were elicited by experts. We build a probability distribution for correlation matrices which is centred in $\boldsymbol{C}$, with a spread around $\boldsymbol{C}$, given by the user to express his willingness to perturb individual entries in $\boldsymbol{C}$. This probability model is then used to determine a proper correlation matrix close to $\boldsymbol{C}$.

A naive probability distribution for a symmetric matrix with elements in $[-1, +1]$, as correlations are, is obtained by assuming that each upper triangular entry of the matrix is a realisation from a univariate distribution within $[-1, +1]$. One choice for such a distribution is the translated and scaled beta distribution. If we generate each upper triangular entry in this way, we obviously can also generate illegal correlation matrices. Hence we shall restrict the support of the distribution on symmetric and non-negative definite matrices with unit diagonal and elements in $[-1, +1]$.

Let Y_{ij} be a random variable with values in $[-1, 1]$, representing a random correlation between two variables, for $i = 1, \ldots, n$ and $j = i + 1, \ldots, n$. A symmetric matrix containing correlations is uniquely determined by these Y_{ij}'s. We call this matrix $\boldsymbol{Y}$. The probability distribution $f(\boldsymbol{Y})$ of $\boldsymbol{Y}$ is given as the product over all $i = 1, \ldots, n$ and $j = i+1, \ldots, n$ of the individual beta densities

$$f(y_{ij}; c_{ij}, a_{ij}, b_{ij}) = \frac{\Gamma(a_{ij} + b_{ij})}{\Gamma(a_{ij})\Gamma(b_{ij})} \frac{(1 + y_{ij})^{a_{ij}-1}(1 - y_{ij})^{b_{ij}-1}}{2^{a_{ij}+b_{ij}-1}}. \tag{3}$$

We shall centre $\boldsymbol{Y}$ around $\boldsymbol{C}$ by requiring that $\mathrm{E}(Y_{ij}) = c_{ij}$, which implies that the two parameters a_{ij} and b_{ij} are such that

$$\frac{a_{ij} - b_{ij}}{a_{ij} + b_{ij}} = c_{ij}.$$

The spread around this centre is given by

$$\mathrm{Var}(Y_{ij}) = \frac{4a_{ij}b_{ij}}{(a_{ij} + b_{ij})^2(a_{ij} + b_{ij} + 1)}.$$

This value describes the willingness to relax this specific entry to accommodate for positive definiteness, or in other words, the analyst's faith in the initial value c_{ij}.

Note that $\boldsymbol{Y}$ has $\boldsymbol{C}$ as its mean, as specified by the user, but $\boldsymbol{C}$ could also have been $\boldsymbol{Y}$'s mode. That would give a different solution to problem, but would probably not be easier for the user.

The matrix $\boldsymbol{Y}$ is symmetric and positive definite if and only if there exists a unique lower triangular matrix $\boldsymbol{X}$, with positive diagonal entries, such that $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{X}'$. This defines an invertible transform $g(\boldsymbol{Y}) = \boldsymbol{X}$. We can then transform the beta based density on $\boldsymbol{Y}$ to obtain the density for $\boldsymbol{X}$, applying the density transformation rules. We get

$$f(\boldsymbol{X}) \propto \prod_{i=1}^{n} \prod_{j=i+1}^{n} f(\boldsymbol{x}_i \boldsymbol{x}_j'; c_{ij}, a_{ij}, b_{ij}) \, |J(\boldsymbol{Y} \to \boldsymbol{X})|$$

where $\boldsymbol{x}_i$ denotes the i-th row of $\boldsymbol{X}$, $J(\boldsymbol{Y} \to \boldsymbol{X})$ is the Jacobian of the transformation and the Euclidean norm $||\boldsymbol{x}_i||_2$ is 1 for all i's. This last requirement ensures that the diagonal elements of $\boldsymbol{Y}$ are 1 and also, by Schwarz inequality, that the off diagonal elements have modulus less than one. Observe that we only obtain the density up to a proportionality constant, because the transformation restricts the space to positive definite matrices, so that the normalising constant is different. The Jacobian can be computed as

$$J(\boldsymbol{Y} \to \boldsymbol{X}) = 2^n \prod_{i=1}^{n} x_{ii}^{n-i+1},$$

see (Gupta and Nagar, 1991, pp. 14).

The equations for the values of the parameters (a_{ij}, b_{ij}) take a simple form. Omitting the indexes ij for the moment, if μ and σ^2 are the mean and the variance of a standard beta random variable V, with parameters a and b, then simple algebra leads to

$$\begin{aligned} a =& \mu \left[\frac{\mu(1-\mu)}{\sigma^2} - 1 \right] \\ b =& (1-\mu) \left[\frac{\mu(1-\mu)}{\sigma^2} - 1 \right], \end{aligned}$$

with the constraint $\sigma^2 < \mu(1-\mu)$ to assure existence. The random correlation Y is obtained as $Y = 2V - 1$. Requiring it to have mean c and variance σ_Y^2, we use the above equations with $\mu = (c+1)/2$ and $\sigma^2 = \sigma_Y^2/4$ to find (a, b). The constraint in the transformed space is now $\sigma_Y^2 < (c+1)(1-c)$.

In practice, for ease of interpretation, we will restrict both a and b to be larger than one, for example at least $1+\varepsilon$ with some small $\varepsilon > 0$, so that the transformed beta distributions are always bell-shaped with their mode in the interval $(-1, 1)$. We will allow the user to specify an interval $c \pm \Delta$ where he thinks c belongs to with high probability, say such as 99.73% as in the Gaussian case ($Z \sim \mathcal{N}(0, \sigma^2) \Rightarrow P(-3\sigma < Z < 3\sigma) \approx 0.9973$). We fix σ^2 through the equation $3\sigma_Y = \Delta$, that is, $\sigma^2 = (\Delta/6)^2$, in such a way that the constraints $a > 1 + \varepsilon$ and $b > 1 + \varepsilon$ are met. This gives

$$\sigma^2 = \min \left\{ \left(\frac{\Delta}{6} \right)^2, \frac{\mu^2(1-\mu)}{(1+\varepsilon+\mu)}, \frac{\mu(1-\mu)^2}{(2+\varepsilon-\mu)} \right\}.$$

It is easy to check that the above value of σ^2 is always admissible. Since we do this for every individual element of $\boldsymbol{Y}$, we obtain a set of variances $\sigma^2_{Y,ij}$, by which we can express the willingness to change the value of the correlation indexed by ij from its current value c_{ij}, for each pair i, j.

The mode of $f(\boldsymbol{X})$, now specified up to a constant of proportionality, is a good candidate (when it exists) for a proper approximation of $\boldsymbol{C}$. Passing to the logarithm of $f(\boldsymbol{X})$, the task is to maximise

$$\begin{aligned}\log f(\boldsymbol{X}) = &\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left[(b_{ij}-1)\log(1-\boldsymbol{x}_i\boldsymbol{x}_j')+(a_{ij}-1)\log(1+\boldsymbol{x}_i\boldsymbol{x}_j')\right]\\ &+\sum_{i=1}^{n}(n-i+1)\log(x_{ii})\end{aligned} \tag{4}$$

subject to $||\boldsymbol{x}_i||_2 = 1$ and $x_{ii} > 0$ for all i's. There is no simple solution to this maximisation problem, and in Section 4 we describe a method for optimisation based on available computing libraries.

Let $\tilde{\boldsymbol{X}}$ be a local maximum of f, giving the proper and non-singular correlation matrix $\tilde{\boldsymbol{C}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}'$. Clearly, $\tilde{\boldsymbol{C}}$ is a proper matrix that resembles the original improper $\boldsymbol{C}$. How strong this similarity is, and which of the entries have relaxed the most, depends on the user's choice of the set of variances σ^2_{ij}. Indeed a very stringent choice can make the solution impossible, or very hard. In general, there are many ways to accommodate an improper matrix, even if the elements causing problems are just a few: $\tilde{\boldsymbol{X}}$ can have many elements which have changed.

An important question is how much $\tilde{\boldsymbol{C}}$ differs from $\boldsymbol{C}$ and where. While previous approaches focused on Euclidean-type metrics such as the Frobenius norm (Higham, 2002; Pietersz and Groenen, 2004; Rebonato and Jäckel, 1999), our probabilistic formulation induces a metric which is quite different. Two matrices $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ can have the same Euclidean distance but very different probabilistic distances, as measured by $f(\boldsymbol{X}_1) - f(\boldsymbol{X}_2)$, depending on the position within the support of the distribution. Our proposal is to identify *hotspots* in $\tilde{\boldsymbol{X}}$, that is positions ij in the correlation matrix where $\tilde{c}_{ij}$ differs most from c_{ij}: these would represent the correlations mostly affected by the rehabilitation procedure. To decide whether a given location ij is a hotspot, we calculate the following quantity

$$h_{ij} = \begin{cases} \frac{P\left(Y_{ij}\in(\tilde{c}_{ij},c_{ij})\right)}{P(Y_{ij}\leq c_{ij})} & \text{if} \quad \tilde{c}_{ij} \leq c_{ij} \\ \frac{P\left(Y_{ij}\in(c_{ij},\tilde{c}_{ij})\right)}{P(Y_{ij}> c_{ij})} & \text{if} \quad \tilde{c}_{ij} > c_{ij} \end{cases} \tag{5}$$

where $Y_{ij} = 2V_{ij} - 1$, with $V_{ij} \sim \text{beta}(a_{ij}, b_{ij})$. We can interpret h_{ij} as a (standardised) measure of how far $\tilde{c}_{ij}$ is out in the tail of the distribution of Y_{ij}.

Another related hotspot classification method is that of considering, for every entry in the correlation matrix, a set of intervals $(q_{(\alpha/2)}, q_{(1-\alpha/2)})$, based on the quantiles of the relevant beta distribution, as α takes different values, such as 0.10, 0.25, 0.50, 0.75. Then a location ij would be coded 0 if $\tilde{c}_{ij}$ belongs to the innermost interval, and correspondingly, from 1 up to 4. Code 4 is given if it does not belong to any interval, hence out in the tail, and the location is classified as a hotspot.

3 Sketching a Bayesian approach

In the previous section we assumed that an improper correlation matrix was available along with expert opinions on the reliability of each entry, expressed with the help of beta densities. In this section, we consider a different situation: we assume that each entry of the illegal correlation matrix comes along with an inherent measure of precision, obtained, for example, directly from the same data that gave origin to the point estimate of the correlation. In addition, we assume that expert opinions are elicited on their strength of belief in each entry, and we assume that expert opinions are not based on the same data. We combine the two sources of information following a Bayesian approach, considering the coefficients of the observed improper correlation matrix $\boldsymbol{C}$ as realisations of random variables centred at the coefficients of an unknown, "true", symmetric positive-definite matrix Θ. On this latent matrix Θ, we have obtained prior opinions. Symmetry and positive-definiteness are ensured by considering the Cholesky decomposition $\Theta = \boldsymbol{L}\boldsymbol{L}'$; the posterior distribution of the coefficients of $\boldsymbol{L}$ is then used to obtain Bayes estimators of the coefficient of a proper correlation matrix $\tilde{\boldsymbol{C}}$, which acts as an estimate for Θ.

We assume that an empirical correlation c_{ij}, $i \neq j$, is observed along with a measure of its spread given by upper and lower bounds, b_{ij} and a_{ij}, representing for example the 95% confidence interval of the ij correlation obtained from data analysis. We use triangular distributions $c_{ij} \sim \mathcal{T}r(\theta_{ij}, a_{ij}, b_{ij})$, defined on $[a_{ij}, b_{ij}]$ and with mode at θ_{ij}, so that the density is given by

$$f(c_{ij}|\theta_{ij}, a_{ij}, b_{ij}) = \begin{cases} \frac{2}{b_{ij}-a_{ij}} \frac{c_{ij}-a_{ij}}{\theta_{ij}-a_{ij}} & c_{ij} \leq \theta_{ij} \\ \frac{2}{b_{ij}-a_{ij}} \frac{c_{ij}-b_{ij}}{\theta_{ij}-b_{ij}} & c_{ij} > \theta_{ij}. \end{cases} \tag{6}$$

Other likelihood models for c_{ij} are possible, e.g. transformed beta distributions, but we found triangular distributions flexible and mathematically easy to treat.

Next, we choose a prior model for the unknown parameters θ_{ij}, $i \neq j$; note that $\theta_{ii} = 1, i = 1, \ldots, n$. We will explore various possible choices of priors. We assume first that the θ_{ij}'s are independently beta distributed on the interval $[-1, 1]$, with densities given as in (3), but now for θ_{ij}:

$$\pi(\theta_{ij}|\alpha_{ij}, \beta_{ij}) = \frac{\Gamma(\alpha_{ij} + \beta_{ij})}{\Gamma(\alpha_{ij})\Gamma(\beta_{ij})} \frac{(1+\theta_{ij})^{\alpha_{ij}-1}(1-\theta_{ij})^{\beta_{ij}-1}}{2^{\alpha_{ij}+\beta_{ij}-1}}. \tag{7}$$

Denoting by $\boldsymbol{\theta}$ the vector with components $\theta_{ij}, 1 \leq j \leq i \leq n$ (and analogously for other matrix elements, e.g. $\boldsymbol{c}$), we combine (7) for all θ_{ij} to obtain the joint prior density

$$\pi(\boldsymbol{\theta}|\boldsymbol{a}, \boldsymbol{b}) \propto \prod_{1 \leq j \leq i \leq n} \left\{ (1+\theta_{ij})^{\alpha_{ij}-1}(1-\theta_{ij})^{\beta_{ij}-1} \right\}. \tag{8}$$

The matrix $\Theta = \{\theta_{ij}\}_{1 \leq i,j \leq n}$, with off-diagonal entries drawn from (8), is symmetric since we take $\theta_{ji} = \theta_{ij}$ for all pairs (i, j), but its values, even when drawn from the corresponding posterior distribution, are not leading, in general, to a positive definite matrix. Therefore we should restrict ourselves to θ_{ij}'s such that $\Theta \in \mathcal{S}_\Theta$, the space of

all symmetric positive-definite matrices, with coefficients within $[-1, 1]$. In this way, we would give up the independence property of all θ_{ij} and, even worse, we would impose very complex bounds on θ_{ij} and simulation from the posterior distribution would be a very burdensome task.

To restrict the prior measure to just symmetric positive-definite matrices, we recall that all positive definite matrices can be expressed by the Cholesky decomposition as $\Theta = \boldsymbol{L}\boldsymbol{L}'$, with $\boldsymbol{L}$ being a lower triangular matrix with positive diagonal entries. Denote its elements by l_{ij}. The relation between the coefficients θ_{ij}, $i \geq j$, of Θ and l_{ij} is given by

$$\theta_{ij} = \sum_{k=1}^{j} l_{ik} l_{jk}.$$

We consider now the prior distribution on $\boldsymbol{l}$. The expression of the Jacobian is

$$J(\Theta \to \boldsymbol{L}) = 2^n \prod_{i=1}^{n} l_{ii}^{n-i+1} \tag{9}$$

(Gupta and Nagar, 1991, pp. 14).

Combining (8) and (9), we get the prior distribution

$$\pi(\boldsymbol{l}|\boldsymbol{a}, \boldsymbol{b}) \propto \prod_{i=1}^{n} l_{ii}^{n-i+1} \prod_{1 \leq j \leq i \leq n} \left\{ (1 + \sum_{k=1}^{j} l_{ik} l_{jk})^{\alpha_{ij}-1} (1 - \sum_{k=1}^{j} l_{ik} l_{jk})^{\beta_{ij}-1} \right\}. \tag{10}$$

Constraints on the elements of $\boldsymbol{L}$ are needed to ensure that the positive definite matrix is actually a correlation matrix. We need to impose that l_{ij}'s are in $\mathcal{L}$ such that

$$\mathcal{L} = \left\{ l_{ij} : \sum_{j=1}^{i} l_{ij}^2 = 1, i = 1, \ldots, n \right\}.$$

Because of Schwarz inequality, it turns out that l_{ij}'s satisfy also $-1 \leq \sum_{k=1}^{j} l_{ik} l_{jk} \leq 1,\ 1 \leq j < i \leq n$.

We combine (6) and (10) by Bayes theorem to obtain the posterior distribution of $\boldsymbol{l}$

$$\begin{aligned} \pi(\boldsymbol{l}|\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{b}) \propto \prod_{i=1}^{n} l_{ii}^{n-i+1} \prod_{1 \leq j \leq i \leq n} \Bigg\{ & \frac{(1 + \sum_{k=1}^{j} l_{ik} l_{jk})^{\alpha_{ij}-1}}{(\sum_{k=1}^{j} l_{ik} l_{jk} - a_{ij})^{I\{c_{ij} \leq \sum_{k=1}^{j} l_{ik} l_{jk}\}}} \\ & \cdot \frac{(1 - \sum_{k=1}^{j} l_{ik} l_{jk})^{\beta_{ij}-1}}{(b_{ij} - \sum_{k=1}^{j} l_{ik} l_{jk})^{I\{c_{ij} > \sum_{k=1}^{j} l_{ik} l_{jk}\}}} \Bigg\} I\{\mathcal{L}\}. \end{aligned} \tag{11}$$

where $I\{\cdot\}$ is the indicator function of a statement.

A point estimate $\hat{\boldsymbol{L}}$ can be obtained as the posterior mean of the posterior density (11). The uncertainty around $\hat{\boldsymbol{L}}$ is well described by an appropriate credibility interval. It is necessary to run an MCMC (Markov chain Monte Carlo) algorithm to obtain these estimates, as described in Section 4.

We could have expressed the distributions of c_{ij} directly with the parameters l_{ij}; in this case the mode θ_{ij} of the triangular density (6) would be replaced by

$$\frac{\sum_{k=1}^{j} l_{ik} l_{jk}}{\sqrt{\sum_{k=1}^{i} l_{ik}^2} \sqrt{\sum_{k=1}^{j} l_{jk}^2}}.$$

However, while expert opinions on θ_{ij} can be easily obtained, it is not realistic that experts are confident to elicit priors on the l_{ij}'s.

4 Algorithms

We describe the algorithms to obtain an estimated proper correlation matrix, for both our approaches, as described in Section 2 and Section 3.

4.1 Beta correlation matrices

The constrained maximisation of $\log f(\boldsymbol{X})$ in equation (4) should be done over $n(n+1)/2-1$ variables, that is, the lower triangle of $\boldsymbol{X}$ except for the entry x_{11}, which takes unit value. We can transform the problem in an equivalent one with simpler constraints by dividing each row of $\boldsymbol{X}$ by its norm:

Find the maximum of

$$\begin{aligned}\log \tilde{f}(\boldsymbol{X}) = &\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left[(b_{ij}-1)\log\left(1-\frac{\boldsymbol{x}_i}{||\boldsymbol{x}_i||}\frac{\boldsymbol{x}'_j}{||\boldsymbol{x}_j||}\right)+(a_{ij}-1)\log\left(1+\frac{\boldsymbol{x}_i}{||\boldsymbol{x}_i||}\frac{\boldsymbol{x}'_j}{||\boldsymbol{x}_j||}\right)\right] \\ &+\sum_{i=1}^{n}(n-i+1)\log\left(\frac{x_{ii}}{||\boldsymbol{x}_i||}\right)\end{aligned} \tag{12}$$

subject to $x_{11}=1$ and $x_{ii}>0$, $i=2,\ldots,n$.

Thus we operate on unnormalised row vectors and the new function $\tilde{f}(\boldsymbol{X})$ does not change its value as long as the row vectors of $\boldsymbol{X}$ do not change direction. If $\boldsymbol{X}^*$ is a local maximum of $\tilde{f}$, then an admissible solution $\tilde{\boldsymbol{X}}$ is obtained by normalising its rows and it is a local maximum of f as well. Although it is quite involved, it is possible to write down the gradient of $\log \tilde{f}$ explicitly and to use the quasi-Newton optimisation method with box constraints as in Byrd et al. (1995). The optimal legal and non singular correlation matrix is then obtained as $\tilde{\boldsymbol{C}}=\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}'$.

With regards to the initial value for the optimisation routine, we cannot use the output of Rebonato and Jäckel (1999) or of Pietersz and Groenen (2004), because these may produce numerically singular matrices and the unique Cholesky factorisation then does not apply. Therefore we apply the following procedure:

1. calculate the spectral decomposition of $\boldsymbol{C}=\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}'$ with the eigenvalues in the diagonal matrix $\boldsymbol{\Lambda}$ sorted into descending order ($\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_n$);

2. if i denotes the index of the last positive eigenvalue, then modify the eigenvalues with index $j=i+1,\ldots,n$ letting $\lambda_j=\lambda_i/2^{(j-i)}$ and denote by $\boldsymbol{\Lambda}_0$ the new matrix with the modified eigenvalues;

3. set $\boldsymbol{C}_0=\boldsymbol{S}\boldsymbol{\Lambda}_0\boldsymbol{S}'$, which is now a non-singular and positive definite matrix;

4. apply the Cholesky factorisation to $\boldsymbol{C}_0=\boldsymbol{X}_0\boldsymbol{X}'_0$ and use $\boldsymbol{X}_0$ to initialise the algorithm.

4.2 Bayesian approach

In order to obtain the posterior mean of (11), we need to implement a Metropolis-Hastings algorithm, which samples from the posterior distribution of $\boldsymbol{l}$. For example, the Gibbs Sampler would be feasible, since all the conditional posterior densities are given up to a proportionality constant. Since we have supposed that the l_{ij}'s are in $\mathcal{L}$, i.e. such that $\sum_{j=1}^{i} l_{ij}^2 = 1$, $i = 1, \ldots, n$, we preserve this condition by updating the blocks $\boldsymbol{l}_{(i)} = \{l_{ij}, j = 1, \ldots, i\}, i = 1, \ldots, n$. We need to specify the full conditional distributions $\pi(\boldsymbol{l}_{(i)}|\boldsymbol{c}, \boldsymbol{l}_{(-i)})$, where $\boldsymbol{l}_{(-i)} = \boldsymbol{l} \setminus \{\boldsymbol{l}_{(i)}\}$.

For a given $i, i = 1, \ldots, n$, the elements of the i-th row appear in both the priors and in the likelihood in correspondence with either θ_{it} or θ_{si}, with $1 \leq t \leq i$ and $i < s \leq n$. We define the set of such indices as $\mathcal{A}_i$.

Therefore, we get

$$\pi(\boldsymbol{l}_{(i)}|\boldsymbol{c}, \boldsymbol{l}_{(-i)}) \propto l_{ii}^{n-i+1} \prod_{(s,t)\in\mathcal{A}_i} \left\{ \frac{(1+\sum_{k=1}^{t} l_{sk}l_{tk})^{\alpha_{st}-1}}{(\sum_{k=1}^{t} l_{sk}l_{tk} - a_{st})^{I\{c_{st}\leq\sum_{k=1}^{t} l_{sk}l_{tk}\}}} \cdot \right.$$
$$\left. \cdot \frac{(1-\sum_{k=1}^{t} l_{sk}l_{tk})^{\beta_{st}-1}}{(b_{st} - \sum_{k=1}^{t} l_{sk}l_{tk})^{I\{c_{st}>\sum_{k=1}^{t} l_{sk}l_{tk}\}}} \right\} I\{\mathcal{L}\}.$$

Once m samples $\{\boldsymbol{l}^{(r)}\}_{r=1}^{m}$ from the posterior distribution of $\boldsymbol{l}$ are obtained, after the necessary burn-in which assures that approximate convergence has been reached, posterior point estimators of the entries of Θ can be computed, thus forming $\tilde{\boldsymbol{C}}$, which can be used instead of the improper matrix $\boldsymbol{C}$. We choose the estimator $\tilde{c}_{ij}$ that minimises the posterior expected squared loss

$$E\left(\sum_{k=1}^{j} l_{ik}l_{jk} - \tilde{c}_{ij}\right)^2, \tag{13}$$

under the posterior distribution of $\boldsymbol{l}$. It is well known that the posterior mean

$$E\sum_{k=1}^{j} l_{ik}l_{jk}$$

minimises (13) and, therefore, it is the Bayes optimal estimate of θ_{ij} under the squared loss function. For the (i, j) entry, this estimate can be approximated as

$$\frac{1}{m}\sum_{r=1}^{m}\sum_{k=1}^{j} l_{ik}^{(r)}l_{jk}^{(r)},$$

by ergodicity of the MCMC. Each element

$$\sum_{k=1}^{j} l_{ik}^{(r)}l_{jk}^{(r)}$$

in this sum is the (i, j) entry of a correlation matrix $\tilde{\boldsymbol{C}}^{(r)}$. Then, $\tilde{\boldsymbol{C}} = \frac{1}{m}\sum_{r=1}^{m} \tilde{\boldsymbol{C}}^{(r)}$ is still a correlation matrix.

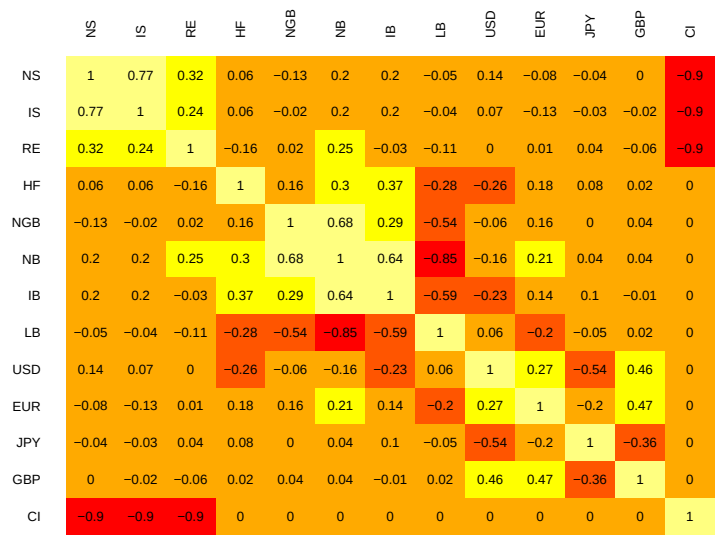


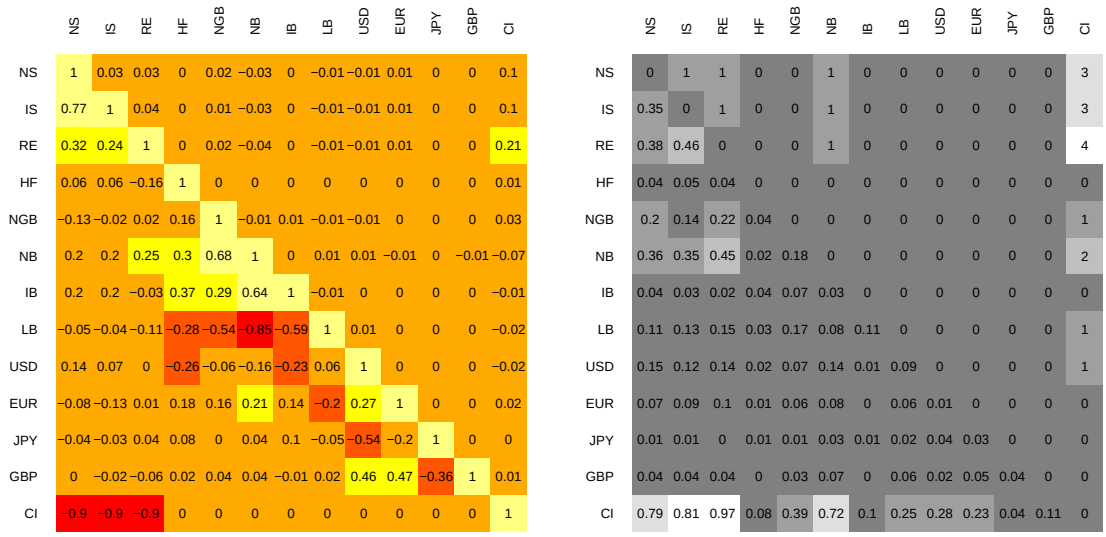
Figure 1. Illegal correlation matrix C among major risk factors. The shades of red correspond to a grid with step 0.4 over the interval $[-1, 1]$. See Table 1 for abbreviations.

5 Case study

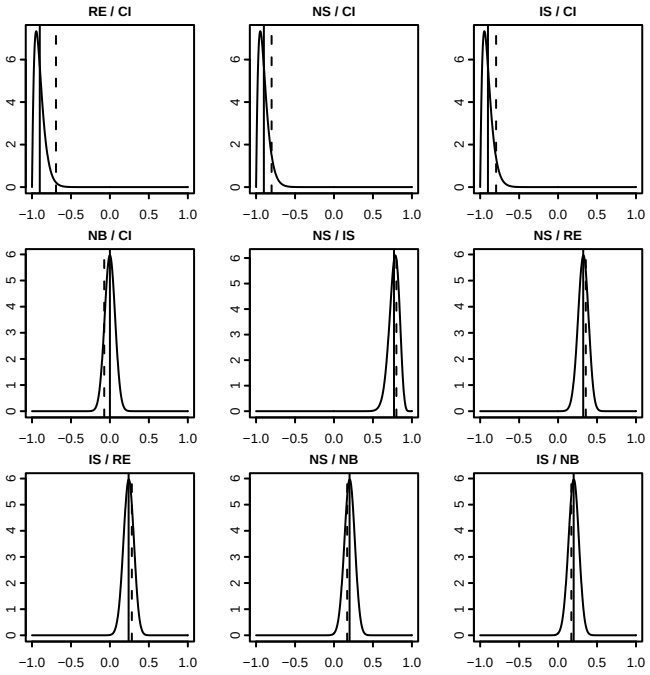
Abbreviation	Risk factor
NS	Norwegian stocks
IS	International stocks
RE	Real estate
HF	Hedge fund
NGB	Norwegian government bonds
NB	Norwegian bonds
IB	International bonds
LB	Long term bonds
USD	USD/NOK
EUR	EUR/NOK
JPY	JPY/NOK
GBP	GBP/NOK
CI	Credit indicator

Table 1. Abbreviations for case study correlation matrix.

To illustrate the beta correlation method (Section 2), we provide an example based on real data. The correlation matrix (Figure 1, Table 1 for abbreviations) contains major risk factors for the assets of a Norwegian life insurance company; market risk (stocks, bonds, hedge funds and real estate) and credit risk. Their stock and bond portfolio is exposed to currency risk, and here the four most important currencies are included. Finally, the company is exposed to credit risk (as recent events during 2008 have so clearly demon-

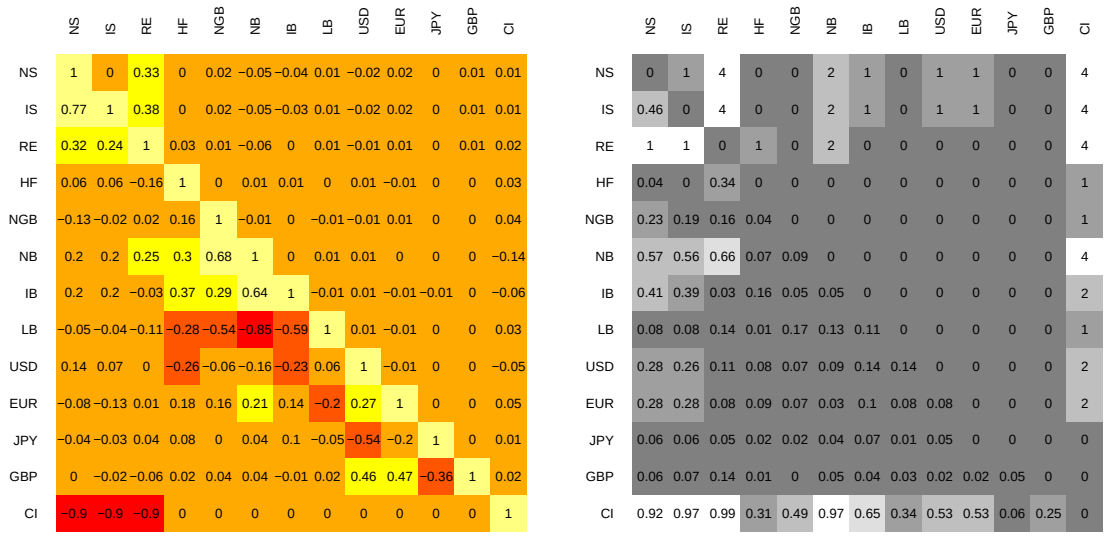


(a) Left panel: $\boldsymbol{C}$ in lower triangle and diagonal; $\tilde{\boldsymbol{C}} - \boldsymbol{C}$ in upper triangle. Right panel: hotspots with normalised tail probabilities (lower triangle) and hotspots with quantile intervals (upper triangle). The five greylevels are relative to a grid over $[0, 1]$ with step 0.2 for the former and they increase from dark grey to white as the optimised values move out of intervals for the latter.

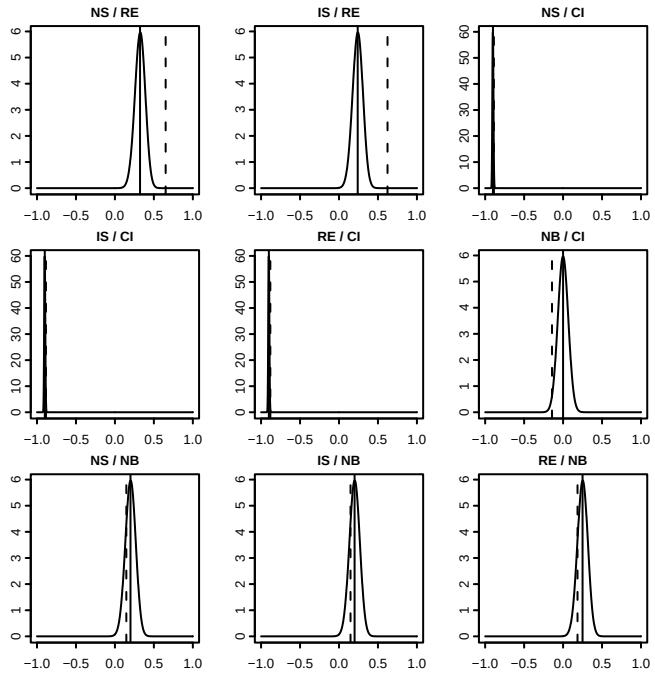


(b) Beta distributions with mean (solid vertical line) and optimal correlation (dashed line) for the 9 hottest hotspots.

Figure 2. Example 1: optimisation with $\Delta = 0.2$ for all correlations.



(a)



(b)

Figure 3. Same as Figure 2, but with $\Delta = 0.02$ for the correlations between CI and NS, IS, RE (-0.9) and between IS and NS (0.77).

strated), against both reinsurers and bond issuers. The correlation matrix was originally estimated from historical data. Then some of the correlations were manually altered by experts. Furthermore, the correlations of all assets with a new credit indicator were added by experts, expressing a priori guesses, since relevant credit indicator data were absent. The non-zero correlations of -0.9 of the credit indicator may seem extreme. In some cases they are chosen to be extra high, to perform sensitivity testing.

The correlation matrix is illegal, with one negative eigenvalue (≈ -0.3). We adjust the matrix with two sets of expert opinions on the correlations. In both sets the beta distributions have the original correlation values as means; the variances are all equal with $\Delta = 0.2$ in the first set, whereas $\Delta = 0.02$ for the correlations between CI and NS, IS, RE (-0.9) and between IS and NS (0.77) in the second set. This second specification has the purpose of checking what is the effect, on the whole matrix, of a strong unwillingness to modify certain particular entries. The results of the optimisation with the two different sets of variance parameters are shown in Figures 2 and 3, respectively. Clearly, a very concentrated distribution for some correlations may have an undesirable effect on other correlations, like the correlations between IS and RE and between NS and RE in Figure 3. In the same figure, the correlations between CI and NS, IS, RE have not changed much in absolute terms (0.01, 0.01 and 0.02, respectively), but in terms of the hotspots, the change is substantial, since the normalised tail probabilities for the three pairs are all above 0.9. This is not surprising, because the beta distributions assign 99.73% probability to the interval -0.9 ± 0.02 for these three correlations. In the same example, the correlations between IS and RE and between NS and RE are also extreme hotspots, with a normalised tail probability of 1, but here the change in absolute terms is also large. This can also be seen from the two upper left windows of Figure 3(b), which shows how the solution deviates from the a priori guess.

The estimated matrices shown are obtained from an initial starting value as specified in the previous Section. We also tried very different starting values, obtained from randomly generated correlation matrices, and the resulting local maxima are very close to the maximum reached from the fixed starting point.

6 Conclusions

We consider two situations, which are often encountered in practical financial analysis. In one case we assume that an improper correlation matrix is given, possibly in part obtained as a collection of estimated pairwise correlations from time series data, that are not fully aligned in time. In addition, experts can express their belief in these point estimates, by assigning beta densities to each entry which are more peaked around the current correlation if this is considered reliable, while it is flatter for correlations which are less safe.

In the second case, for each entry of the improper correlation matrix, we assume that a measure of precision is available along with the point estimate of the pair correlation, for example its empirical variance. Note that in this case, we cannot assume that the

improper correlation matrix can be enlarged with a new non-observed asset, since in this case the same expert opinion would be used twice. Unfortunately, the Bayesian approach relies on MCMC for inference. Such an algorithm is less robust in use compared with the simple numerical optimisation necessary to obtain a proper correlation matrix in the first case. More research is needed to devise a fast alternative to MCMC, before it can enter financial practice. One alternative approach is given by Rue and Martino (2009), who are using an integrated nested Laplace approximation to compute very accurate approximations of the posterior marginals.

Our approach leads to a very practical operational tool. A user friendly version can be based on a screen with the improper correlation matrix, the possibility to assign beta densities on each off-diagonal entry as a measure of confidence in the given correlation and the willingness to accept changes. The user may then elicit beta densities via their mean and standard deviation. When a beta distribution is given, a small figure appears, with the plot of the actual beta density, illustrating the choice made. The optimal proper matrix has actually followed all wishes as expressed by beta densities. However, for some entries, it might be necessary to move out to the tails of the corresponding beta densities. Such cases are shown in our hotspot matrix, that represent where major stress has been introduced to legalise the full matrix. Alternatives to beta densities are of course possible.

Acknowledgements

This work was done within (sfi)^2 Statistics for Innovation.

References

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.

Dimakos, X. K. and Aas, K. (2004). Integrated risk modelling. *Statistical modelling*, 4(4):265–277.

Embrechts, P., McNeil, A. J., and Straumann, D. (1999). Correlation: Pitfalls and alternatives. *Risk*, 12:69–71.

Embrechts, P., McNeil, A. J., and Straumann, D. (2002). *Risk Management: Value at Risk and Beyond*, chapter Correlation and dependence in risk management: properties and pitfalls, pages 176–223. Cambridge University Press.

Grubišić, I. and Pietersz, R. (2007). Efficient rank reduction of correlation matrices. *Linear Algebra and its Applications*, 422(2-3):629–653.

Gupta, A. K. and Nagar (1991). *Matrix variate distributions*. Chapmann & Hall.

Higham, N. J. (2002). Computing the nearest correlation matrix – a problem from finance. *IMA Journal of Numerical Analysis*, 22:329–343.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.

Medova, E. A. (2000). Extreme value theory: extreme values and the measurement of operational risk. *Operational Risk*, 1(7):13–17.

Pietersz, R. and Groenen, P. J. F. (2004). Rank reduction of correlation matrices by majorization. *Quantitative Finance*, 4:649–662.

Rebonato, R. and Jäckel, P. (1999). The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. Technical report, Quantitative Research Centre of the NatWest Group.

Rue, H. and Martino, S. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B*, 71:1–35.