Insurance: Mathematics and Economics 112 (2023) 97–109

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A note on portfolios of averages of lognormal variables

Phelim Boyle [a], Ruihong Jiang [b,*]

[a] Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, ON, N2L 3C5, Canada
[b] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

ARTICLE INFO

Article history:
Received September 2022
Received in revised form April 2023
Accepted 17 June 2023
Available online 24 June 2023

This paper is dedicated to the memory of Ken Seng Tan who passed away on January 1, 2023

JEL classification:
C46

Keywords:
Lognormal distribution
Sum of lognormals
Block covariance matrix
Limiting distribution
Moment formulae

ABSTRACT

This paper establishes conditions under which a portfolio consisting of the averages of K blocks of lognormal variables converges to a K-dimensional lognormal variable as the number of variables in each block increases. The associated block covariance matrix has to have a special structure where the correlations and variances within the block submatrices are equal. We show why the variance homogeneity assumption plays a key role in the derivation.

1. Introduction

The lognormal distribution has extensive applications in the social and physical sciences. It is widely used in finance to model stock prices and in actuarial science to model insurance liabilities. Crow and Shimizu (1987) document applications in biology, ecology, atmospheric sciences and geology. The lognormal distribution also has applications in engineering and in particular in the field of wireless communications. The distribution of the sum of lognormals has an intuitive interpretation and plays an important role in many applications.

The distribution of the sum of random variables is of interest in mathematical statistics and applied probability. Asmussen and Rojas-Nandayapa (2008) study the tail behaviour of the sum of correlated lognormals. Gulisashvili and Tankov (2016) propose an efficient importance sampling estimator for the left tail of the distribution function of the sum of lognormal variables. Botev et al. (2019) develop a Monte Carlo approach to estimate the distribution of the sum of dependent lognormal random variables.

There are several references in the actuarial and finance literature to the distribution of the sum of lognormals. Dufresne (2008) considers the distribution of the sum of two lognormals with applications to stock prices. Asmussen (2018) uses conditional Monte Carlo to investigate the tails of sums of random variables and uses his approach to compute the Value at Risk and Conditional Tail Expectation. Furman et al. (2020) develop an approximation for the sums of independent lognormal variables and illustrate the method with applications to risk measures and economic capital.

It is straightforward to derive closed-form expressions for the moments of a sum of correlated lognormals. However, the distribution itself does not have a closed form. Many approximations of this distribution have been developed over the years. For a review of these approaches, see Asmussen et al. (2011). It is often more convenient to work with the distribution of the average rather than the sum and we will follow this practice. It turns out that under certain conditions, the distribution of the average of a group of lognormals

* Corresponding author.
E-mail addresses: pboyle@wlu.ca (P. Boyle), rhjiang@uwaterloo.ca (R. Jiang).

https://doi.org/10.1016/j.insmatheco.2023.06.001
0167-6687/

converges to a univariate lognormal, as the number of variables tends to infinity. This result first appeared in the engineering literature. See Szyszkowicz and Yanikomeroglu (2009) and Beaulieu (2011). These authors showed that the distribution of the average of n correlated lognormal variables converges to a univariate lognormal, as n tends to infinity.

The current paper extends the Szyszkowicz-Yanikomeroglu-Beaulieu (SYB) result to the case where there are homogeneous groups of correlated lognormals with a specified correlation structure. We assume that the underlying variables can be divided into a finite number of K groups where there is strong homogeneity within groups but there can be considerable heterogeneity across the different groups. These properties are captured through a covariance matrix with a specific block structure. In our case, the correlations and variances are constant within each block but the correlations among the different blocks are unequal. This type of block structure is mathematically convenient and it also has practical applications.

We present two situations which have been modelled using this type of covariance structure. First, consider a stock portfolio consisting of different sub-portfolios. The sub-portfolios could correspond, for instance, to industry groupings. Engle and Kelly (2012) and Archakov and Hansen (2020) discuss portfolios with this structure. The last named authors develop procedures for estimating covariance matrices in these circumstances. Archakov and Hansen (2020) illustrate their approach using a sample of daily stock returns. The model discussed later in our paper corresponds to the case where the stocks in each sub-portfolio are equally weighted. DeMiguel et al. (2009) have shown that equally weighted stock portfolios possess desirable risk-return characteristics.

The second example relates to a portfolio of loans where each loan is subject to credit risk. These risks can be modelled using a block correlation matrix. The application is discussed by Huang and Yang (2010) in connection with loan portfolio models used by rating agencies. They observe that the default risk correlation between loans is specified by analysts who often classify the loans into groups based on sectors, industries or regions. The correlation between two different blocks depends only on the blocks they belong to. Huang and Yang (2010) demonstrate the efficiency of their approach in a Monte Carlo context using two large loan portfolios.

These examples provide the motivation for us to study models with this type of block correlation structure. Essentially we show that the within-group averages converge to a K-dimensional lognormal distribution as the numbers in each group become large. The significance of this result is that it reduces the dimension of the problem. The derivation depends on certain conditions. A critical assumption is that the within-block covariances are equal which means that the corresponding variances are equal since we assume constant correlations within each block. We highlight the importance of this assumption by showing that even when the differences among the within-block variances are miniscule, the block averages do not converge to a lognormal distribution. The between-group correlations are constant for each pair of groups and also have to satisfy additional conditions to ensure the covariance matrix is positive definite.

Associated with every multivariate lognormal distribution is a corresponding multivariate normal distribution. It is often more convenient to work with the covariance matrix associated with this normal distribution than the covariance matrix of the lognormal distribution. If there are m variables in each block and there are K blocks the overall covariance matrix is of size $Km \times Km$. The variables within each block are assumed to have the same variance. The pairwise correlations between these variables are all equal. The large matrix is assumed to be positive definite and this implies that the $K \times K$ matrix composed of the averages of the block submatrices is also positive definite.

The rest of the paper is organized as follows. Section 2 introduces our notations and assumptions. We start with the properties of the univariate lognormal random variable. Then we derive expressions for the first four moments of the average of a number of lognormal variables, under fairly general assumptions. We will rely on these results in section 4. Section 3 contains our convergence result. Given our assumptions, the distribution of the averages of the K blocks of lognormal variables converges to a K-dimensional lognormal distribution. The derivation relies on Kolmogorov's strong law of large numbers (Serfling (2009)) and is an extension of the approach used by Szyszkowicz and Yanikomeroglu (2009) and Beaulieu (2011). In Section 4 we consider an example where the heterogeneity in the within-block variances is controlled by an arbitrarily small parameter ϵ. We show that even for a miniscule degree of heterogeneity, the third and fourth moments of the limiting distribution lack the lognormal properties so that the convergence no longer holds. The final section summarises the paper and outlines several future research topics.

2. The lognormal distribution

We start with a brief review of the univariate lognormal distribution. Next, we consider the multivariate lognormal distribution and derive the formulae for the first four moments of linear combinations of lognormal variables. Then we outline the block structure used in our convergence result and discuss the positive definite issue.

2.1. Univariate and multivariate lognormal distributions

If a normal random variable X has variance σ^2 and mean $\mu - \frac{\sigma^2}{2}$, then

$$Y = e^X$$

is said to have a lognormal distribution. We have

$$E[Y] = e^{\mu}$$

and for all positive integers n,

$$E[Y^n] = \exp\left(\left[\mu - \frac{\sigma^2}{2}\right]n + \frac{n^2\sigma^2}{2}\right). \tag{1}$$

We now introduce the n-dimensional multivariate lognormal distribution. Let

$$W = [W_1, W_2, \cdots, W_n]^T$$

be a multivariate normal distribution with the mean vector

$$\left[\mu_1 - \frac{\sigma_1^2}{2}, \cdots, \mu_n - \frac{\sigma_n^2}{2}\right]^T \tag{2}$$

and the covariance matrix

$$C = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \cdots & \cdots & \sigma_n^2 \end{bmatrix}. \tag{3}$$

Then

$$X = [X_1, X_2, \cdots, X_n]^T$$

where

$$X_i = e^{W_i}$$

has a multivariate lognormal distribution.

We now list some of the cross moments. These moments will be useful when we derive the moments of linear combinations of the X's. For $i \neq j$, we have

$$E[X_i X_j] = e^{\mu_i + \mu_j + C_{ij}}$$
$$E[X_i X_j^2] = e^{\mu_i + 2\mu_j + C_{ij} + 2C_{jj}}.$$

For $i \neq j$, $i \neq k$ and $k \neq j$,

$$E[X_i X_j X_k] = e^{\mu_i + \mu_j + \mu_k + C_{ij} + C_{jk} + C_{ik}}.$$

Next we consider the case when there are several variables in the cross moments. For $i \neq j$,

$$E[X_i^3 X_j] = e^{3\mu_i + \mu_j + 3C_{ii} + 3C_{ij}}$$
$$E[X_i^2 X_j^2] = e^{2\mu_i + 2\mu_j + C_{ii} + C_{jj} + 4C_{ij}}.$$

For $i \neq j$, $i \neq k$ and $k \neq j$, we have

$$E[X_i^2 X_j X_k] = e^{2\mu_i + \mu_j + \mu_k + C_{ii} + 2C_{ij} + 2C_{jk} + C_{jk}}.$$

The last cross moment corresponds to the case when no pair of the four indices i, j, k and l are equal.

$$E[X_i X_j X_k X_l] = e^{\mu_i + \mu_j + \mu_k + \mu_l + C_{ij} + C_{ik} + C_{il} + C_{jk} + C_{jl} + C_{kl}}$$

Later in the paper, we will need expressions for the moments of the average of a group of lognormal variables. To this end, we state the formulae for the first four moments of any linear combination of lognormal variables with the mean and covariance structure given by equations (2) and (3). The proofs of these expressions are straightforward but sometimes lengthy and are omitted. We assume that the weights w_i are all positive where

$$\sum_{i=1}^{n} w_i = 1.$$

We define the linear combination A_n by

$$A_n = \sum_{i=1}^{n} w_i X_i. \tag{4}$$

The first moment of A_n is

$$E[A_n] = \sum_{i=1}^{n} w_i e^{\mu_i}. \tag{5}$$

The second moment of A_n is

$$E[A_n^2] = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j e^{\mu_i + \mu_j + C_{ij}}. \tag{6}$$

The third moment is

$$E[A_n^3] = \sum_{i=1}^{n} w_i^3 e^{3\mu_i} e^{3C_{ii}} + 3\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n} w_i w_j^2 e^{(\mu_i+2\mu_j)} e^{(C_{jj}+2C_{ij})}$$
$$+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i,k\neq j}^{n} w_i w_j w_k e^{(\mu_i+\mu_j+\mu_k)} e^{(C_{ij}+C_{jk}+C_{ik})}.$$

For the fourth moment, we have

$$E[A_n^4] = \sum_{i=1}^{n} w_i^4 e^{4\mu_i} e^{6C_{ii}} + 4\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n} w_i^3 w_j e^{(3\mu_i+\mu_j)} e^{3(C_{ii}+C_{ij})}$$
$$+3\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n} w_i^2 w_j^2 e^{2(\mu_i+\mu_j)} e^{(C_{ii}+C_{jj}+4C_{ij})}$$
$$+6\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}\sum_{k=1,k\neq i,k\neq j}^{n} w_i^2 w_j w_k e^{(2\mu_i+\mu_j+\mu_k)} e^{(C_{ii}+2C_{ij}+2C_{ik}+C_{jk})}$$
$$+\sum_{i=1}^{n}\sum_{j=1,i\neq j}^{n}\sum_{k=1,k\neq j,k\neq i}^{n}\sum_{l=1,l\neq k,l\neq j,l\neq i}^{n} w_i w_j w_k w_l e^{(\mu_i+\mu_j+\mu_k+\mu_l)} e^{(C_{ij}+C_{ik}+C_{il}+C_{jk}+C_{jl}+C_{kl})}.$$

These formulae for the moments of A_n are quite general. We can simplify the expressions for the moments by considering a special case. Assume that

$$w_i = \frac{1}{n}, \quad 1 \le i \le n,$$
$$\mu_i = \mu, \quad 1 \le i \le n,$$
$$\sigma_i = \sigma, \quad 1 \le i \le n,$$
$$\rho_{ij} = \rho, \quad 1 \le i < j \le n.$$

Let us denote these four assumptions as Assumptions A. In this case, we show that as n tends to infinity the limits simplify. We have

$$\lim_{n\to\infty} E[A_n^2] = e^{2\mu} \lim_{n\to\infty} \frac{\left(\sum_{i=1}^{n}\ \sum_{j=1,j\neq i}^{n} e^{\rho\sigma^2} + \sum_{i=1}^{n} e^{\sigma^2}\right)}{n^2}$$
$$= e^{2\mu} \lim_{n\to\infty} \frac{\left(n(n-1)e^{\rho\sigma^2} + n\, e^{\sigma^2}\right)}{n^2}$$
$$= e^{2\mu+\rho\sigma^2}.$$

Szyszkowicz and Yanikomeroglu (2009) have shown[1] that under Assumptions A, the distribution of A_n tends to a univariate lognormal distribution. We see from the last equation that the variance of the associated (normal) limiting distribution is given by $\rho\sigma^2$. This is the average variance of the covariance matrix C under Assumptions A. Using the same approach, we can derive expressions for the limits of the higher moments under Assumptions A. For the third moment, under Assumptions A, we can show that

$$\lim_{n\to\infty} E[A_n^3] = e^{3\mu\, +\, 3\rho\sigma^2}. \tag{7}$$

For the fourth moment, under Assumptions A, we can show that

$$\lim_{n\to\infty} E[A_n^4] = e^{4\mu\, +\, 6\rho\sigma^2}. \tag{8}$$

These last two moments correspond to the moments of the limiting univariate lognormal distribution with $\rho\sigma^2$ playing the role of the variance.

2.2. Groups of lognormal variables

This section describes the structure of the groups of lognormal variables that we will be dealing with. The within-group correlations and variances are equal, and the correlation between any two variables from a given pair of groups is constant. Essentially there is a high degree of homogeneity within groups and scope for some heterogeneity across groups. The structure of the relationships among the variables is captured through a well-defined block structure. The properties of block covariance and block correlation matrices are discussed by several authors including Cadima et al. (2010), Huang and Yang (2010), Roustant and Deville (2017) and Archakov and Hansen (2020).

[1] Assumptions A also correspond to a special case of the assumptions used by Beaulieu (2011). Beaulieu assumes equal covariances instead of equal variances and equal correlations.

We assume there are K groups of lognormal variables where each group consists of m variables denoted by

$$\{X_i^1\}_{i=1}^m,\ \{X_i^2\}_{i=1}^m,\ \cdots,\ \{X_i^K\}_{i=1}^m.$$

Let $n = Km$. This n-dimensional lognormal distribution has an associated n-dimensional normal distribution. The m lognormal variables in the k^{th} block are denoted by

$$[X_1^k,\ X_2^k, \cdots, X_m^k]^T.$$

Let

$$W_i^k = log X_i^k, \quad 1 \le i \le m$$

so that

$$\{W_i^1\}_{i=1}^m,\ \{W_i^2\}_{i=1}^m,\ \cdots,\ \{W_i^K\}_{i=1}^m$$

follow a multivariate normal distribution.

Denote by C_n the covariance matrix of this multivariate normal distribution. The matrix C_n is assumed to have a special block structure where the within-block variances and correlations are constant. We have

$$C_n = \begin{bmatrix} B_{[1,1]} & \cdots & B_{[1,K]} \\ \vdots & \vdots & \vdots \\ B_{[K,1]} & \cdots & B_{[K,K]} \end{bmatrix} \tag{9}$$

where each $B_{[k,k]}$ is of dimension $m \times m$.

The diagonal blocks $B_{[k,k]}$ have the following structure

$$B_{[k,k]} \ = \ \sigma_k^2 \begin{bmatrix} 1 & \rho_k & \cdots & \rho_k \\ \rho_k & 1 & \cdots & \rho_k \\ \vdots & \vdots & \vdots & \vdots \\ \rho_k & \rho_k & \cdots & 1 \end{bmatrix} \tag{10}$$

where $0 < \rho_k < 1$. The off-diagonal blocks are defined as follows

$$B_{[k,l]} \ = \ \rho_{kl}\sigma_k\sigma_l \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \tag{11}$$

where $0 < \rho_{kl} < 1$. We will assume that C_n is positive definite and we will see that this property imposes some additional restrictions on the ρ_k, ρ_l and ρ_{kl}.

We now describe the moments of these lognormal variables. We have for $1 \le k \le K$ and $1 \le i \le m$,

$$E[X_i^k] = e^{\mu_i^k}.$$

The within-block second moments are, for $1 \le k\ \le K$,

$$E[(X_i^k)^2] = e^{2\mu_i^k + \sigma_k^2}$$
$$E[X_i^k X_j^k] = e^{\mu_i^k + \mu_j^k + \rho_k\sigma_k^2}.$$

The cross block product moments are for $1 \le k \le K,\ 1 \le l \le K,\ k \ne l$,

$$E[X_i^k X_j^l] = e^{\mu_i^k + \mu_j^l + \rho_{kl}\sigma_k\sigma_l}.$$

We now turn to the positive definite issue. When matrix C_n is positive definite, this has implications for the relations among the correlation parameters. It is known that the matrix C_n is positive definite if and only if a certain smaller $K \times K$ matrix is positive definite (Cadima et al. (2010) and Archakov and Hansen (2020)). The elements of this smaller matrix, herein denoted by $C_K(m)$, correspond to the averages of the elements of each of the block matrices.

$$C_K(m) = \begin{bmatrix} \sigma_1^2\left(\frac{1}{m} + (1-\frac{1}{m})\rho_1\right) & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1K}\sigma_1\sigma_K \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2\left(\frac{1}{m} + (1-\frac{1}{m})\rho_2\right) & \cdots & \rho_{2K}\sigma_2\sigma_K \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1K}\sigma_1\sigma_K & \rho_{2K}\sigma_2\sigma_K & \cdots & \sigma_K^2\left(\frac{1}{m} + (1-\frac{1}{m})\rho_K\right) \end{bmatrix}.$$

We therefore assume that $C_K(m)$ is a positive definite matrix for all m and in particular that it is positive definite as m tends to infinity. This means the matrix $\bar{C}_K$ is positive definite, where

$$\bar{C}_K = \lim_{m\to\infty} C_K(m) = \begin{bmatrix} \sigma_1^2\rho_1 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1K}\sigma_1\sigma_K \\ \rho_{12}\sigma_1\sigma_2 & \rho_2\sigma_2^2 & \cdots & \rho_{2K}\sigma_2\sigma_K \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{1K}\sigma_1\sigma_K & \rho_{2K}\sigma_2\sigma_K & \cdots & \rho_K\sigma_K^2 \end{bmatrix}.$$

We know that the determinants of all the principal submatrices of $\bar{C}_K$ have to be positive because $\bar{C}_K$ is positive definite. Hence we have

$$\rho_1\rho_2 \geq \rho_{12}^2.$$

There is a similar result for all two by two matrices that straddle the diagonal so that, for $1 \leq i < j \leq K$, we have

$$\rho_i\rho_j \geq \rho_{ij}^2. \tag{12}$$

It will be convenient to define r_{kl} as

$$r_{kl} = \frac{\rho_{kl}}{\sqrt{\rho_k\rho_l}} \tag{13}$$

where $0 \leq r_{kl} < 1$.

To summarize, matrix C_n is positive definite if and only if the smaller matrix $C_K(m)$ is positive definite. We assume that $C_K(m)$ is positive definite for all values of m. This implies that equation (12) is true.

We now look at the multivariate lognormal distribution. The $Km \times Km$ covariance matrix of the lognormal distribution is denoted by D_n. It inherits the block structure from C_n and it is also positive definite. Similar to C_n, we know that, for D_n, a smaller $K \times K$ matrix formed by replacing each block with its average also has the positive definite property. We denote this matrix by $D_K(m)$. The elements of this matrix are lengthy so we will just exhibit some representative ones. The diagonal $(k,k)^{th}$ element is

$$\sum_{i=1}^{m} \frac{e^{2\mu_i^k+\rho_k\sigma_k^2}}{m^2} + \sum_{i=1}^{m}\sum_{j=1, j\neq i}^{m} \frac{e^{\mu_i^k+\mu_j^k+\rho_k\sigma_k^2}}{m^2} - \sum_{i=1}^{m}\frac{e^{\mu_i^k}}{m}\sum_{j=1}^{m}\frac{e^{\mu_j^k}}{m}.$$

The off-diagonal $(k,l)^{th}$ element of $D_K(m)$ is

$$\sum_{i=1}^{m}\sum_{j=1,}^{m} \frac{e^{\mu_i^k+\mu_j^l+\rho_{kl}\sigma_k\sigma_l}}{m^2} - \sum_{i=1}^{m}\frac{e^{\mu_i^k}}{m}\sum_{j=1}^{m}\frac{e^{\mu_j^l}}{m}.$$

We can now use these moments to obtain the moments of the limiting distribution of the K averages as m tends to infinity. Accordingly, we define the K-dimensional random variable

$$Y_m^K = \left[\frac{\sum_{i=1}^m X_i^1}{m}, \cdots, \frac{\sum_{i=1}^m X_i^K}{m}\right]^T. \tag{14}$$

As m tends to infinity, the expected value of Y_m^K is

$$\begin{aligned}\lim_{m\to\infty} E\left[Y_m^K\right] &= \left[\lim_{m\to\infty}\sum_{i=1}^{m}\frac{E[X_i^1]}{m}, \cdots, \lim_{m\to\infty}\sum_{i=1}^{m}\frac{E[X_i^K]}{m}\right]^T \\ &= \left[e^{\mu^{(1)}}, \cdots, e^{\mu^{(K)}}\right]^T\end{aligned}$$

where, for $1 \leq k \leq K$,

$$\lim_{m\to\infty}\sum_{i=1}^{m}\frac{E[X_i^k]}{m} = \lim_{m\to\infty}\sum_{i=1}^{m}\frac{e^{\mu_i^k}}{m} = e^{\mu^{(k)}}.$$

As m tends to infinity the covariance matrix $D_K(m)$ tends to $\bar{D}_K$ where

$$\bar{D}_K = \begin{bmatrix} e^{2\mu^{(1)}}(e^{\rho_1\sigma_1^2}-1) & e^{\mu^{(1)}+\mu^{(2)}}(e^{\rho_{12}\sigma_1\sigma_2}-1) & \cdots & e^{\mu^{(1)}+\mu^{(K)}}(e^{\rho_{1K}\sigma_1\sigma_K}-1) \\ e^{\mu^{(1)}+\mu^{(2)}}(e^{\rho_{12}\sigma_1\sigma_2}-1) & e^{2\mu^{(2)}}(e^{\rho_2\sigma_2^2}-1) & \cdots & e^{\mu^{(2)}+\mu^{(K)}}(e^{\rho_{2K}\sigma_2\sigma_K}-1) \\ \vdots & \vdots & \vdots & \vdots \\ e^{\mu^{(1)}+\mu^{(K)}}(e^{\rho_{1K}\sigma_1\sigma_K}-1) & e^{\mu^{(2)}+\mu^{(K)}}(e^{\rho_{2K}\sigma_2\sigma_K}-1) & \cdots & e^{2\mu^{(K)}}(e^{\rho_K\sigma_K^2}-1) \end{bmatrix} \tag{15}$$

While we have not yet proved that the vector Y_m^K tends to a K-dimensional lognorml distribution there are several encouraging signs we are on the right track. Note that the individual average, on the right-hand side of equation (14), each tends to a univariate lognormal from the results of Szyszkowicz and Yanikomeroglu (2009) and Beaulieu (2011). Thus as m tends to infinity, the limit of

$$\sum_{i=1}^{m} \frac{X_i^k}{m}$$

tends to a lognormal distribution with mean $e^{\mu^{(k)}}$ and variance $e^{2\mu^{(k)}}(e^{\rho_k \sigma_k^2} - 1)$. Hence the marginal distribution has the correct convergence. In addition the $K \times K$ matrix $\bar{D}_K$ has the right structure to be the covariance matrix of a K dimensional lognormal distribution with these marginals.

3. The convergence result

We are now ready to derive our convergence result. In this section, we show that, as m tends to infinity, the distribution of a portfolio made up of the averages of the K blocks of lognormal variables converges to a K-dimensional lognormal variable. Our approach is based on those of Szyszkowicz and Yanikomeroglu (2009) and Beaulieu (2011) who prove the result when $K = 1$. The proof relies on an application of Kolmogorov's strong law of large numbers (Serfling (2009)).

In the previous section we generated the multivariate lognormal distribution in an obvious way by associating each lognormal variable with a corresponding single normal variable. To prove our convergence result it will be convenient to generate each lognormal variate from a linear combination of normal variates. Here is a simple example that illustrates the intuition. Suppose Z is a normal variable with mean zero and variance σ^2. Let Z_1 and Z_2 be independent normal variates where Z_1 has mean μ and variance $\frac{\sigma^2}{2}$ and Z_2 has mean $-\mu$ and variance $\frac{\sigma^2}{2}$. Their sum $(Z_1 + Z_2)$, will be a normal variable with mean zero and variance σ^2. This normal variable has the same mean and variance as Z.

This means that we can generate essentially the same lognormal distribution using different linear combinations of normal variables. The basic idea in the proof is to generate an n-dimensional multivariate lognormal distribution X from an underlying normal distribution which is structured as a combination of correlated and independent normal variables. The structure is designed so that, when m tends to infinity, we can apply the strong law of large numbers to the independent components, while maintaining the desired correlation properties.

In the last section we explained how to construct a multivariate lognormal distribution with a given block structure based on Km individual normal variables. The next lemma shows how to construct essentially the same Km-dimensional lognormal distribution from a different set of $K + Km$ underlying normal variables. These underlying normal variables have been selected to facilitate the convergence proof.

Lemma 3.1. *Assume we are given the mean and covariance of the Km-dimensional lognormal distribution with the block structure described in the previous section. We can generate an equivalent Km-dimensional lognormal distribution with the same block structure from a specified multivariate normal distribution of dimension $K + Km$.*

The first K of these normal variables are

$$Z_{01}, Z_0, \cdots, Z_{0K}.$$

These variables all have mean zero and variance one:

$$E[Z_{0k}] = 0 \quad and \quad Var[Z_{0k}] = 1, \quad k = 1, 2, \cdots, K.$$

These K random variables are correlated with one another. The pairwise correlations among them are

$$E[Z_{0k} Z_{0l}] = r_{kl}, \quad 1 \le k \ne l \le K,$$

where the r_{kl} were defined in equation (13).

The remaining Km normal variables can be divided into K groups with m variables in each group

$$\{Z_i^1\}_{i=1}^m, \ \{Z_i^2\}_{i=1}^m, \cdots, \{Z_i^K\}_{i=1}^m.$$

These variables are all pairwise independent. They are also all independent of each of the first K normal variables. For $1 \le i \le m$ and $1 \le k \le K$, we have

$$E[Z_i^k] = \mu_i^k - \frac{\sigma_k^2}{2} < \infty$$
$$Var[Z_i^k] = \sigma_k^2(1 - \rho_k) < \infty.$$

The Km random variables[2] defined as

$$\{X_i^k\}_{i=1}^m := \{e^{\sqrt{\rho_k}\,\sigma_k Z_{0k} + Z_i^k}\}_{i=1}^m, \quad 1 \le k \le K,$$

represent the desired multivariate lognormal distribution.

[2] We have used the same notation as before $\{X_i^k\}_{i=1}^m$ to economize on symbols.

Proof. We note that the exponent $(\sqrt{\rho_k}\ \sigma_k Z_{0k} + Z_i^k)$ in the last equation is normal with mean $\mu_i^k - \frac{\sigma_k^2}{2}$ and variance σ_k^2.

Also for $1 \le k, l \le K$ and $1 \le i, j \le m$, we can show

$$\begin{aligned}
E[X_i^k] &= e^{\mu_i^k} \\
E[(X_i^k)^2] &= e^{2\mu_i^k + \sigma_k^2} \\
E[X_i^k X_j^k] &= e^{\mu_i^k + \mu_j^k + \rho_k \sigma_k^2}, \quad i \neq j \\
E[X_i^k X_j^l] &= e^{\mu_i^k + \mu_j^l + \rho_{kl}\sigma_k \sigma_l},\ k \neq l.
\end{aligned}$$

Combining these results we see that the $\{X_i^k\}_{i=1}^m$ generated in this way form a multivariate lognormal distribution with the desired properties. □

We are now ready to state and prove our convergence result.

Theorem 3.2. *Suppose we have the Km-dimensional lognormal distribution described in Lemma (3.1)*

$$\{X_i^1\}_{i=1}^m,\ \{X_i^2\}_{i=1}^m,\ \cdots,\ \{X_i^K\}_{i=1}^m.$$

These n variables have the block covariance matrix denoted by D_n in the previous section. The K-dimensional random variable Y_m^K defined in equation (14) denotes the averages of the K blocks of lognormal variables.

Under our assumptions, as m tends to infinity, Y_m^K converges in probability to a K-dimensional lognormal distribution with expected value

$$\left[e^{\mu^{(1)}}, \cdots, e^{\mu^{(K)}}\right]^T$$

and covariance matrix $\bar{D}_K$.

Proof. Our proof is an extension of that in Szyszkowicz and Yanikomeroglu (2009) and Beaulieu (2011) who consider the simpler case when $K = 1$. The general idea of the proof is that we express the random variable of interest Y_m^K as a product of two random variables. One follows a lognormal distribution. The other one can be proven to converge to a constant, by the law of large numbers. Then, invoking Slutsky's Theorem (Rohatgi and Saleh (2015)), we prove the convergence of Y_m^K to a lognormal random variable. The detailed proof is as follows.

We first consider the average associated with the k^{th} block. From Lemma (3.1), for $1 \le k \le K$,

$$\frac{1}{m}\sum_{i=1}^m X_i^k = \frac{1}{m}\sum_{i=1}^m e^{\sqrt{\rho_k}\ \sigma_k Z_{0k} + Z_i^k} = e^{\sqrt{\rho_k}\ \sigma_k Z_{0k}}\left(\frac{1}{m}\sum_{i=1}^m e^{Z_i^k}\right).$$

Now we take the limit as m tends to infinity to get, for $1 \le k \le K$,

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m X_i^k = e^{\sqrt{\rho_k}\ \sigma_k Z_{0k}}\left(\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m e^{Z_i^k}\right).$$

We can invoke Kolmogorov's strong law of large numbers, with the relaxation on the identically distributed condition, according to page 27 of Serfling (2009). We require that $e^{Z_i^k}$, $1 \le i \le m$, are independent and that the series $\sum_{i=1}^m Var\left(e^{Z_i^k}\right)/i^2$ converges as m tends to infinity. The first condition is satisfied since the Z_i^k, $1 \le i \le m$, are independent. The second condition is satisfied as long as σ_k^2, ρ_k and μ_i^k are all less than infinity, which is assumed in the model. This can be easily verified as in Beaulieu (2011).

Hence, from Kolmogorov's strong law of large numbers, for $1 \le k \le K$,

$$\begin{aligned}
\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m e^{Z_i^k} &= \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m E[e^{Z_i^k}] \\
&= \lim_{m\to\infty}\left(\frac{1}{m}\sum_{i=1}^m e^{\mu_i^k}\right)e^{-\frac{\rho_k\sigma_k^2}{2}} \\
&= e^{\mu^{(k)} - \frac{\rho_k\sigma_k^2}{2}}.
\end{aligned}$$

The convergence is in probability.

We define

$$Y_k = e^{\sqrt{\rho_k}\ \sigma_k Z_{0k} + \mu^{(k)} - \frac{\rho_k\sigma_k^2}{2}}. \tag{16}$$

Hence,

$$\lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m X_i^k = Y_k$$

with probability 1, according to Slutsky's Theorem (Rohatgi and Saleh (2015)).

Note that since $[Z_{01}, Z_{02}, \cdots, Z_{0K}]^T$ form a multivariate normal distribution, then $[Y_1, Y_2, \cdots, Y_K]^T$ form a multivariate lognormal distribution.

Hence we have shown that

$$\lim_{m\to\infty} Y_m^K = \left[\lim_{m\to\infty} \sum_{i=1}^{m} \frac{X_i^1}{m}, \cdots, \lim_{m\to\infty} \sum_{i=1}^{m} \frac{X_i^K}{m} \right]^T = [Y_1, \cdots, Y_K]^T \tag{17}$$

with probability 1.

From equation (16) it is easy to check that

$$E[Y_k] = e^{\mu^{(k)}}$$

and

$$Var(Y_k) = e^{2\mu^{(k)}} \left(e^{\rho_k \sigma_k^2} - 1 \right).$$

Furthermore, for $1 \le k \le K,\ 1 \le l \le K,\ k \ne l$, we have

$$\begin{aligned} Cov(Y_k, Y_l) &= e^{\mu^{(k)}+\mu^{(l)}} \left(e^{r_{kl}\sqrt{\rho_k}\sigma_k\sqrt{\rho_l}\sigma_l} - 1 \right) \\ &= e^{\mu^{(k)}+\mu^{(l)}} \left(e^{\rho_{kl}\sigma_k\sigma_l} - 1 \right). \end{aligned}$$

Hence the covariance matrix of the K-dimensional lognormal distribution $[Y_1,\ Y_2\ \cdots, Y_K]^T$ is $\bar{D}_K$ as anticipated. □

Note that Theorem 3.2 works on Y_m^K whose elements are within-block sums with equal weights. We can also extend to the general case with unequal weights. In the general case, we define a new K-dimensional random variable

$$Y_m^K(w) = \left[\frac{\sum_{i=1}^m w_i^1 X_i^1}{m}, \cdots, \frac{\sum_{i=1}^m w_i^K X_i^K}{m} \right]^T \tag{18}$$

where $w = \{w_i^k : 1 \le i \le m, 1 \le k \le K\}$ are weights satisfying[3]

$$\sum_{i=1}^{m} \frac{w_i^k}{m} = 1, \quad 1 \le k \le K.$$

The expected value of $Y_m^K(w)$ is

$$\begin{aligned} \lim_{m\to\infty} E\left[Y_m^K(w)\right] &= \left[\lim_{m\to\infty} \sum_{i=1}^{m} \frac{w_i^1 E[X_i^1]}{m}, \cdots, \lim_{m\to\infty} \sum_{i=1}^{m} \frac{w_i^K E[X_i^K]}{m} \right]^T \\ &= \left[e^{\mu^{(1)}(w)}, \cdots, e^{\mu^{(K)}(w)} \right]^T \end{aligned}$$

where, for $1 \le k \le K$,

$$\lim_{m\to\infty} \sum_{i=1}^{m} \frac{w_i^k E[X_i^k]}{m} = e^{\mu^{(k)}(w)}.$$

The covariance matrix of the K-dimensional limiting lognormal distribution is $\bar{D}_K(w)$ which is the same as $\bar{D}_K$ with $\mu^{(k)}(w)$ replacing $\mu^{(k)}$.

Proposition 3.3. *Suppose for every $1 \le k \le K$, we have*

$$\lim_{m\to\infty} \sum_{i=1}^{m} \frac{(w_i^k)^2}{i^2} < \infty.$$

Under our block covariance structure, as m tends to infinity, the K-dimensional random variable, $Y_m^K(w)$ converges in probability to a K-dimensional lognormal distribution with mean

[3] Here is an example of such weights that converge naturally to a reasonable limit. Assume that for $1 \le i \le m$, v_i are numbers on $(0, 2)$. We could define $w_i = \lambda v_i$ where

$$\lambda = \frac{m}{\sum_{i=1}^m v_i}$$

so that $\sum_{i=1}^m \frac{w_i}{m} = 1$.

$$\left[e^{\mu^{(1)}(w)}, \cdots, e^{\mu^{K}(w)}\right]^T$$

and with covariance matrix equal to $\bar{D}_K(w)$.

The proof is similar to Theorem 3.2 and is omitted. Theorem 3.2 is a special case of Proposition 3.3 when $w_i^k = 1$ for all i and k.

4. Example with heterogeneous variances

This section highlights the importance of the (within-block) homogeneous variance assumption. Under our assumptions, the within-block homogeneous variance leads to homogeneous covariance. This assumption guarantees the convergence results. Note that our block covariance structure is the same as in Szyszkowicz and Yanikomeroglu (2009), but different from that in Beaulieu (2011). The latter paper assumes heterogeneity within-block variance but imposes a homogeneous covariance assumption in equation (T3), which essentially leads to the convergence result.

To show the importance of our assumption, it will be enough to consider the case $K = 1$. We provide an example where the variances differ by very miniscule amounts and show that the average of a large group of lognormal variables does not converge to a univariate lognormal. Under our assumptions, we are able to find closed-form expressions for the moments of the average. We show that the relations among these moments do not correspond to the relations among the moments of a univariate lognormal. Hence, we conclude that the limiting distribution is not a univariate lognormal.

We assume there are n lognormal variables where the standard deviations (of the associated normal distributions) lie in the interval $[\sigma_{low}, \sigma_{high}]$. It will be convenient to use the following notation

$$\begin{aligned}
\sigma_{low} &= \sigma, \\
\sigma_{high} &= \sigma(1+\epsilon), \quad \epsilon > 0, \\
\delta &= \frac{\epsilon}{2n}.
\end{aligned}$$

We assume the corresponding standard deviations are given by the following vector

$$\sigma\left[1+\delta,\ 1+3\delta, \cdots, 1+(2n-1)\delta\right].$$

Hence the standard deviations are evenly distributed in the interval $[\sigma, \sigma(1+\epsilon)]$. The corresponding variances are contained in the interval $(\sigma^2,\ \ \sigma^2(1+\epsilon)^2)$. This interval can be made arbitrarily small by reducing the value of ϵ.

For convenience, we assume that all the μ_i are equal to zero and that all the pairwise correlations are $\rho > 0$. It is convenient to let $\alpha = 1+\epsilon$.

Under these assumptions, the limiting distribution of the second moment (about zero) of the average is, recalling equation (6),

$$\begin{aligned}
\lim_{n\to\infty} E\left[A_n^2\right] &= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{n^2} e^{\rho\sigma^2(1+(2i-1)\delta)^2} \\
&\quad + \lim_{n\to\infty} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{n^2} e^{\rho\sigma^2(1+(2i-1)\delta)(1+(2j-1)\delta)} \\
&= \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2 xy} dx dy.
\end{aligned} \tag{19}$$

The first term on the right-hand side tends to zero. The double summation on the second line tends to the double integral from the Riemann integrability of the continuous bivariate exponential function in the rectangle $[1, \alpha] \times [1, \alpha]$.

Likewise, the limiting distribution of the third moment (about zero) of the average is

$$\lim_{n\to\infty} E\left[A_n^3\right] = \int_1^\alpha \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2(xy+xz+yz)} dx dy dz. \tag{20}$$

Finally, the limiting distribution of the fourth moment (about zero) of the average is

$$\lim_{n\to\infty} E\left[A_n^4\right] = \int_1^\alpha \int_1^\alpha \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2(xy+xz+xw+yz+yw+zw)} dx dy dz dw. \tag{21}$$

Now consider the case of a true univariate lognormal distribution, X, with mean zero ($\mu = 0$) and variance σ_0^2. The first four moments of X about zero are

$$\begin{aligned}
E[X] &= 1, \\
E[X^2] &= e^{\sigma_0^2}, \\
E[X^3] &= e^{3\sigma_0^2}, \\
E[X^4] &= e^{6\sigma_0^2}.
\end{aligned}$$

Hence, in the case of a true univariate lognormal distribution X, we have

$$E[X^3] = \left(E[X^2]\right)^3$$

and

$$E[X^4] = \left(E[X^2]\right)^6.$$

If the distribution of A_m, as m tends to infinity, converges to a univariate lognormal, we would have

$$\int_1^\alpha \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2(xy+xz+yz)} dxdydz = \left(\int_1^\alpha \int_1^\alpha e^{\rho\sigma^2 xy} dxdy\right)^3 \tag{22}$$

and

$$\int_1^\alpha \int_1^\alpha \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2(xy+xz+xw+yz+yw+zw)} dxdydzdw = \left(\int_1^\alpha \int_1^\alpha e^{\rho\sigma^2 xy} dxdy\right)^6. \tag{23}$$

However the following two propositions[4] show that both equation (22) and equation (23) are false. This means that the limiting distribution of A_m, under our assumption of heterogeneous variances, does **not** converge to a lognormal distribution. Note that this result is true for arbitrarily small ϵ.

Proposition 4.1. *For a monotone function $f : (0, \infty) \to (0, \infty)$ and iid (independent and identically distributed) positive random variables X, Y, Z, we have*

$$E[f(XY)f(XZ)f(YZ)] \geq E[f(XY)]^3. \tag{24}$$

Moreover, if f is strictly monotone and X, Y, Z are non-degenerate, then the inequality (24) is strict.

Proof. Take X', Y', Z' such that X, Y, Z, X', Y', Z' are iid. Let x, y, z, x', y', z' be arbitrary positive real numbers. We first note that

$$\begin{aligned} E[f(Xy)f(Xz)] &= E[f(Xy)]E[f(Xz)] + Cov(f(Xy), f(Xz)) \\ &\geq E[f(Xy)]E[f(Xz)] = E[f(Xy)f(X'z)], \end{aligned}$$

where we use $Cov(f(Xy), f(Xz)) \geq 0$, which is due to the Fréchet-Hoeffding (or Hardy-Littlewood) inequality (Lieb and Loss (2001)) by noting that $f(Xy)$ and $f(Xz)$ are both increasing (or decreasing) functions of X. Therefore, we have

$$E[f(Xy)f(Xz)f(yz)] \geq E[f(Xy)f(X'z)f(yz)],$$

and by integrating over (y, z), we get

$$E[f(XY)f(XZ)f(YZ)] \geq E[f(XY)f(X'Z)f(YZ)]. \tag{25}$$

By the same argument,

$$E[f(x'Y)f(Yz)] \geq E[f(x'Y)f(Y'z)],$$

which gives

$$E[f(xY)f(x'z)f(Yz)] \geq E[f(xY)f(x'z)f(Y'z)],$$

and by integrating over (x, x', z), we get

$$E[f(XY)f(X'Z)f(YZ)] \geq E[f(XY)f(X'Z)f(Y'Z)]. \tag{26}$$

Finally,

$$E[f(xZ)f(y'Z)] \geq E[f(xZ)f(y'Z')],$$

which gives

$$E[f(xy)f(x'Z)f(y'Z)] \geq E[f(xy)f(x'Z)f(y'Z')],$$

and by integrating over (x, x', y, y'), we get

[4] We thank Ruodu Wang of the University of Waterloo for his proof of the first proposition and for inspiring the proof of the second proposition.

$$E[f(XY)f(X'Z)f(Y'Z)] \geq E[f(XY)f(X'Z)f(Y'Z')]. \tag{27}$$

Therefore, we obtain

$$\begin{aligned} E[f(XY)f(XZ)f(YZ)] &\geq E[f(XY)f(X'Z)f(YZ)] && \text{(by (25))} \\ &\geq E[f(XY)f(X'Z)f(Y'Z)] && \text{(by (26))} \\ &\geq E[f(XY)f(X'Z)f(Y'Z')] && \text{(by (27))} \\ &\geq E[f(XY)]^3. && \text{(by the iid assumption)} \end{aligned}$$

This shows (24). The last statement follows by noting that $Cov(f(Xy), f(Xz)) > 0$ as soon as $f(Xy)$ and $f(Xz)$ are non-degenerate. □

In our case, choosing X, Y, Z as uniform distributed on $[1, \alpha]$ and $f : x \mapsto e^{\rho\sigma^2 x}$ with $\rho\sigma^2 \neq 0$, we get the inequality

$$\int_1^\alpha \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2(xy+xz+yz)} dxdydz > \left(\int_1^\alpha \int_1^\alpha e^{\rho\sigma^2 xy} dxdy \right)^3.$$

Therefore, equation (22) does not hold.

Proposition 4.2. *For a monotone function $f : (0, \infty) \to (0, \infty)$ and iid positive random variables X, Y, Z, W, we have*

$$E[f(XY)f(XZ)f(XW)f(YZ)f(YW)f(ZW)] \geq E[f(XY)]^6. \tag{28}$$

Moreover, if f is strictly monotone and X, Y, Z, W are non-degenerate, then the inequality (28) is strict.

Proof. The proof is parallel to that of Proposition 4.1, hence is omitted. □

In our case, choosing X, Y, Z, W as uniform distributed on $[1, \alpha]$ and $f : x \mapsto e^{\rho\sigma^2 x}$ with $\rho\sigma^2 \neq 0$, we get the inequality

$$\int_1^\alpha \int_1^\alpha \int_1^\alpha \int_1^\alpha e^{\rho\sigma^2(xy+xz+xw+yz+yw+zw)} dxdydzdw > \left(\int_1^\alpha \int_1^\alpha e^{\rho\sigma^2 xy} dxdy \right)^6.$$

Therefore, equation (23) does not hold.

5. Summary and future research

The distribution of the average of lognormal variables has widespread applications in insurance and finance. This paper analyzed the limiting distribution of the means of K groups of correlated lognormal variables. We showed that under certain assumptions on the block covariance structure, the distribution converges to a K-dimensional lognormal distribution. The within-block means can vary but the within-block variances and the within-block correlations are constant. We have shown that the homogeneity of the within-block variances plays an essential role in the derivation.

To round out the paper it may be helpful to discuss some future research topics. These relate to the effectiveness of the approximation both in terms of convergence and its application in practice. As a starting point, it would be of interest to analyze how sensitive the convergence result is to the underlying assumptions. For example, if the underlying variables are lognormal and all the assumptions are satisfied, how quickly does the portfolio of averages converge to the K-dimensional lognormal distribution? The convergence will clearly depend on the number of variables in each block. In a risk management context, tail behaviour is important so that the tail convergence would furnish a useful measuring rod.

Another project would be to examine the robustness of the convergence result to variations in the underlying assumptions. For example, while the homogeneity of within-block variances is essential for the portfolio of block averages to converge to a K-dimensional lognormal it would be interesting to explore what divergence is introduced when we use heterogeneous variances. For example, when $K = 1$ we could use a benchmark model with each variance equal to the average variance (of the heterogeneous model). In the benchmark model, we assume all pairwise correlations are equal to the average correlation. The benchmark model satisfies our conditions and the average under the benchmark model converges to a univariate lognormal. We could compare this distribution with the actual distribution of the average in the case of heterogeneous variances. Propositions 4.1 and 4.2 can provide numerical values for the moments so we can compare the first four moments of the two distributions. We could use metrics like the Kullback-Leibler divergence to compare the two distributions.

Perhaps the most interesting future task would be to assess how well the model works in a real-world setting. This would involve calibrating the model to one or more empirical data sets and estimating its parameters. The performance of the model could be compared to that of the empirical distribution using various criteria. The K-dimensional lognormal distribution would be much simpler to handle in terms of portfolio optimization and Monte Carlo simulation. However, the key issue will be how reliable the results obtained are as compared to those based on the parent empirical distribution.

In this connection, we would like to mention one caveat that is related to risk measurement calculations. These are typically concerned with tail risk estimation: e.g. based on Value at Risk or Conditional Tail Expectation. In the limit as m tends to infinity and we converge to the K-dimensional lognormal model, all of the unsystematic risks have been diversified away. This implies that tail risk quantities based on this model would underestimate their actual values. This type of bias would be anathema to any prudent risk manager. However, it should be possible to offset this bias by increasing either the input variances or the input correlations or both.

Declaration of competing interest

The authors report no conflicts of interest.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors are grateful to two referees and the editor for helpful comments on an earlier version of this paper. They thank the Natural Sciences and Engineering Research Council of Canada (Grant number: RGPIN-04687-2014) for support. They are also grateful to Christian Genest, Raymond Kwan and Qihe Tang for their insightful suggestions. They are profoundly grateful to Ruodu Wang for his essential contribution to this paper.

References

Archakov, Ilya, Hansen, Peter Reinhard, 2020. A canonical representation of block matrices with applications to covariance and correlation matrices. arXiv preprint. arXiv:2012.02698.

Asmussen, Søren, 2018. Conditional Monte Carlo for sums, with applications to insurance and finance. Annals of Actuarial Science 12 (2), 455–478.

Asmussen, Søren, Rojas-Nandayapa, Leonardo, 2008. Asymptotics of sums of lognormal random variables with Gaussian copula. Statistics & Probability Letters 78 (16), 2709–2714.

Asmussen, Søren, Ledet Jensen, Jens, Rojas-Nandayapa, Leonardo, 2011. A literature review on log-normal sums. Preprint. University of Queensland.

Beaulieu, Norman C., 2011. An extended limit theorem for correlated lognormal sums. IEEE Transactions on Communications 60 (1), 23–26.

Botev, Zdravko I., Salomone, Robert, Mackinlay, Daniel, 2019. Fast and accurate computation of the distribution of sums of dependent log-normals. Annals of Operations Research 280, 19–46.

Cadima, Jorge, Lage Calheiros, Francisco, Preto, Isabel P., 2010. The eigenstructure of block-structured correlation matrices and its implications for principal component analysis. Journal of Applied Statistics 37 (4), 577–589.

Crow, Edwin L., Shimizu, Kunio, 1987. Lognormal Distributions. Marcel Dekker, New York.

DeMiguel, Victor, Garlappi, Lorenzo, Uppal, Raman, 2009. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? The Review of Financial Studies 22 (5), 1915–1953.

Dufresne, Daniel, 2008. Sums of lognormals. Actuarial Research Conference Proceedings. Regina, Canada, pp. 14–16.

Engle, Robert, Kelly, Bryan, 2012. Dynamic equicorrelation. Journal of Business & Economic Statistics 30 (2), 212–228.

Furman, Edward, Hackmann, Daniel, Kuznetsov, Alexey, 2020. On log-normal convolutions: An analytical-numerical method with applications to economic capital determination. Insurance: Mathematics and Economics 90, 120–134.

Gulisashvili, Archil, Tankov, Peter, 2016. Tail behavior of sums and differences of log-normal random variables. Bernoulli, 444–493.

Huang, Jinggang, Yang, Liming, 2010. Correlation matrix with block structure and efficient sampling methods. Journal of Computational Finance 14 (1), 81.

Lieb, Elliott H., Loss, Michael, 2001. Graduate studies in mathematics. Analysis 14.

Rohatgi, Vijay K., Saleh, A.K. Md, 2015. An Introduction to Probability and Statistics. John Wiley & Sons.

Roustant, Olivier, Deville, Yves, 2017. On the validity of parametric block correlation matrices with constant within and between group correlations. arXiv preprint. arXiv:1705.09793.

Serfling, Robert J., 2009. Approximation Theorems of Mathematical Statistics. John Wiley & Sons.

Szyszkowicz, Sebastian S., Yanikomeroglu, Halim, 2009. Limit theorem on the sum of identically distributed equally and positively correlated joint lognormals. IEEE Transactions on Communications 57 (12), 3538–3542.