ABSTRACT

Mixture of normals is a more general and flexible distribution for modeling of daily changes in market variables with fat tails and skewness. An efficient analytical Monte Carlo method was proposed by Wang and Taaffe for generating daily changes using a multivariate mixture of normal distributions with arbitrary covariance matrix. However the usual Cholesky Decomposition will fail if the covariance matrix is not positive definite. In practice, the covariance matrix is unknown and has to be estimated. The estimated covariance may be not positive definite. We propose a modified Cholesky decomposition for semi-definite matrices and also suggest an optimal semi-definite approximation for indefinite matrices.

1 INTRODUCTION


We have proposed an efficient Monte Carlo method for generating daily changes in market variables using a multivariate mixture of normal distributions with an arbitrary covariance matrix (Wang and Taaffe 2001). The main idea is to transform a multivariate normal with an input covariance matrix into the desired multivariate mixture of normal distributions. This input covariance matrix can be derived analytically.

After we proposed our method, researchers, graduate students, and practitioners from both academic and financial institutions showed their great interests about the method. We have received many inquiries on implementation of our method and model fitting. The most common question from finance industry is how to implement our method when the input covariance matrix is not positive definite.

In theory, the covariance matrix of market variables is positive semi-definite if it exists. However it is usually unknown and has to be estimated from the existing market data. The estimated covariance matrix can be positive definite, or positive semi-definite, or indefinite due to numerical or estimation errors.

Modified Cholesky decomposition is widely used to handle positive semi-definite and indefinite matrices (Gill and Murray 1981; Higham 1988, 1990; Schlick 1993; Schnabel and Eskow 1990; and Cheng and Higham 1998). We propose an alternative modified Cholesky decomposition to deal with positive semi-definite matrices. It is simple, efficient, and easy to implement. An optimal positive semi-definite approximation in Frobenius norm is provided to deal with symmetric indefinite matrices.

This paper proceeds as follows. In Section 2, we discuss the covariance matrix and its estimation. In theory, the covariance matrix is positive semi-definite. The two most widely used sample covariance estimates have been proved to be positive semi-definite. A positive definite sample covariance matrix is constructed in order to apply the usual Cholesky decomposition directly. In Section 3, we briefly review the original Cholesky decomposition, which only works for positive definite matrices. The estimated covariance matrix may not be positive definite due to several reasons. In Section 4, we propose a modified Cholesky decomposition with diagonal pivoting to handle the positive semi-definite matrices. The algorithm is easy to implement and efficient. In Section 5, a best approximation of indefinite matrix is introduced. We just add a small diagonal matrix to the covariance matrix to form a positive definite or positive semi-definite matrix, so that the usual or modified Cholesky decomposition can be used. In Section 6, we review the general mixture of $k$ normal random variable. In Section 7, we propose a method for generating multivariate mixture of normals with arbitrary covariance matrix using the modified Cholesky decomposition. A detailed algorithm is provided to deal with the more general semi-definite matrices.
2 COVARIANCE AND SAMPLE COVARIANCE MATRICES

In this section, we derive that any covariance matrix is positive semi-definite by theory. In practice, the covariance matrix is unknown. We need to use the market data to estimate it. There is no guarantee that it would be positive semi-definite due to numerical and estimation errors. However we can prove that the sample covariance matrix is always positive semi-definite in theory.

Let \( X = (X_1, \ldots, X_n)^T \) be a multivariate random variable, we define its covariance as

\[
\Sigma = \text{Var}(X) = E((X - E(X))(X - E(X))^T).
\]

We have the following fundamental result:

**Theorem 2.1** If the covariance matrix \( \Sigma \) exists, then it must be positive semi-definite.

**Proof** For any constant vector \( c = (c_1, \ldots, c_n)^T \), we have

\[
\text{Var}(c^T X) \geq 0.
\]

Since

\[
0 \leq \text{Var}(c^T X) = \text{Cov}(c^T X, c^T X) = E(c^T X - E(c^T X))^2 = E[(c^T X - E(c^T X))(c^T X - E(c^T X))^T] = E[c^T (X - E(X))(X - E(X))^T c] = c^T E(X - E(X))(X - E(X))^T c = c^T \Sigma c,
\]

therefore \( \Sigma \) is positive semi-definite. \( \square \)

If \( x_k = (x_{1k}, \ldots, x_{nk})^T \) is the \( k \)-th observation of \( X = (x_1, \ldots, x_n)^T \) for \( k = 1, \ldots, N \), then the sample mean of \( X \) is

\[
\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)^T = \left( \frac{1}{N} \sum_{k=1}^{N} x_{1k}, \ldots, \frac{1}{N} \sum_{k=1}^{N} x_{nk} \right)^T.
\]

The most widely used sample covariance matrices are

\[
\hat{\Sigma}_1 = \frac{1}{N} \sum_{k=1}^{N} (x_k - \bar{x})(x_k - \bar{x})^T
\]

and

\[
\hat{\Sigma}_2 = \frac{1}{N-1} \sum_{k=1}^{N} (x_k - \bar{x})(x_k - \bar{x})^T.
\]

There is no big difference between \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) if the sample size \( N \) is large. In statistics, sometimes \( \hat{\Sigma}_1 \) is called the maximum likelihood estimate and \( \hat{\Sigma}_2 \) is called an unbiased estimate. In addition, \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) have the following nice properties.

**Theorem 2.2** Both sample covariance matrices \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) are positive semi-definite.

**Proof** For any constant vector \( c = (c_1, \ldots, c_n)^T \),

\[
\begin{align*}
\hat{c}^T \hat{\Sigma}_1 \hat{c} &= \hat{c}^T \left( \frac{1}{N} \sum_{k=1}^{N} (x_k - \bar{x})(x_k - \bar{x})^T \right) c \\
&= \frac{1}{N} \sum_{k=1}^{N} \hat{c}^T (x_k - \bar{x})(x_k - \bar{x})^T c \\
&= \frac{1}{N} \sum_{k=1}^{N} \hat{c}^T (x_k - \bar{x})[\hat{c}^T (x_k - \bar{x})]^T \\
&= \frac{1}{N} \sum_{k=1}^{N} [\hat{c}^T (x_k - \bar{x})]^2 \geq 0.
\end{align*}
\]

Therefore \( \hat{\Sigma}_1 \) is positive semi-definite. Similarly we can prove that \( \hat{\Sigma}_2 \) is positive semi-definite too. \( \square \)

We can combine \( \hat{\Sigma}_1 \) and \( \hat{\Sigma}_2 \) to form a new positive definite sample covariance matrix. Define

\[
\hat{\sigma}_{ij} = \begin{cases} 
\frac{1}{N} \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)^2 & \text{for } i = j, \text{ and } i, j = 1, \ldots, n, \\
\frac{1}{N} \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) & \text{for } i \neq j, \text{ and } i, j = 1, \ldots, n.
\end{cases}
\]

We have the following result.

**Theorem 2.3** The combined sample covariance matrix

\[
\hat{\Sigma}_3 = (\hat{\sigma}_{ij})
\]

is positive definite when the \( X_i \)'s are not constants.

**Proof** We can view \( \hat{\Sigma}_3 \) as a sum of two matrices:

\[
\hat{\Sigma}_3 = \hat{\Sigma}_1 + E
\]

where \( E \) is a diagonal matrix with elements

\[
e_{ii} = \frac{1}{N(N-1)} \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)^2, \quad i = 1, \ldots, n.
\]

For any non-zero constant vector \( c = (c_1, \ldots, c_n)^T \),

\[
\hat{c}^T \hat{\Sigma}_3 \hat{c} = \hat{c}^T \hat{\Sigma}_1 \hat{c} + \hat{c}^T Ec \geq \hat{c}^T Ec = \frac{1}{N(N-1)} \sum_{k=1}^{N} (x_{ik} - \bar{x}_i)^2 c_i^2 > 0.
\]
We conclude that $\hat{\Sigma}_3$ is positive definite. □

Therefore the usual Cholesky decomposition can apply to $\hat{\Sigma}_3$ directly. All three estimates are good candidates to estimate the covariance matrix. They are positive semi-definite and converge to the true $\Sigma$ almost surely (as second order moment estimates).

3 CHOLESKY DECOMPOSITION AND SINGULAR CASES

Cholesky decomposition is the most commonly used numerical algorithm to decompose a symmetric and positive definite matrix into a lower and upper triangular matrix.

$$A = LL^T$$

or

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & \ldots & 0 \\ l_{21} & l_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \ldots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & \ldots & l_{n1} \\ 0 & l_{22} & \ldots & l_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & l_{nn} \end{bmatrix}.$$ 

The procedure of decomposition is as follows.

$$l_{ii} = \sqrt{(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2)}, \quad i = 1, \ldots, n \quad (1)$$

and

$$l_{ji} = (a_{ji} - \sum_{k=1}^{i-1} l_{jk}l_{ik})/l_{ii}, \quad j = i + 1, \ldots, n. \quad (2)$$

When matrix $A$ is symmetric and positive definite, the expression under the square root is positive and therefore, all elements in $L$ are real. Because of this, the Cholesky decomposition is also called the “square root” decomposition.

Cholesky decomposition is one of the most numerically stable of all matrix algorithms (Wilkinson 1968). Without any pivoting, the decomposition process is stable and the propagation round-off error can be controlled.

In order for the above decomposition to proceed, the matrix $A$ must be positive definite. Mathematically, this requires all eigenvalues of the matrix be positive. Covariance matrix estimated from historic data can be proved to be at least positive semi-definite, i.e., all its eigenvalues are greater than or equal to 0.

However, in certain situations, the eigenvalues of a covariance matrix can be 0, and hence, cause it to be not positive definite. This can happen in the following three cases:

Case 1 If one random variable $X_i$ is indeed a constant, then the entire $i$-th row (and column) of the covariance matrix is 0.

Case 2 If one variable has a perfect linear dependency on one or more other variables, then the matrix is singular and has at least a zero eigenvalue.

Case 3 When the sample size is small, the covariance matrix may be singular due to mere sampling fluctuation. Because the sample covariance matrix does converge to its population covariance matrix, this will not be a problem when the sample size gets large.

When the covariance matrix is not positive definite, the Cholesky decomposition process cannot proceed. Many statistics programs simply send an error message and halt. In order to simulate such situations properly, we need to find a way to process covariance matrices that are not positive definite.

4 CHOLESKY DECOMPOSITION WITH DIAGONAL PIVOTING

Lemma 4.1 The maximum value of a symmetric and positive semi-definite matrix can be achieved on its diagonal.

This lemma guarantees that the pivots of Gaussian Elimination with complete pivoting can always be chosen from the diagonal, and therefore, maintain the symmetry in the decomposition process.

Algorithm 4.1 (Cholesky Decomposition with Diagonal Pivoting) Input: integer $n$, positive semi-definite $n \times n$ matrix $A$.

For $i = 1, \ldots, n$, repeat the following steps.

1. Find the largest diagonal elements on or below the $i$-th row and column. Let it be $a_{ii, ri}$.
2. If $a_{ii, ri} = 0$, stop. The decomposition is complete.
3. If $r_i \neq i$, interchange the $i$-th row and the $r_i$-th row, and the $i$-th column and the $r_i$-th column.
4. Calculate

$$l_{ii} = \sqrt{(a_{ii} - \sum_{k=1}^{i-1} l_{ik}^2)} \quad (3)$$

and

$$l_{ji} = (a_{ji} - \sum_{k=1}^{i-1} l_{jk}l_{ik})/l_{ii}, \quad j = i + 1, \ldots, n. \quad (4)$$
Each iteration in the above algorithm reduces one column of the original matrix \( A \) towards an upper triangular matrix through row/column interchanges \( I_{r,i} \) and \( I_{T,r_i}^T \) and Gaussian elimination \( L_i^{-1} \):

\[
A_{i+1} = L_{i}^{-1}I_{r,i}A_{i}I_{T,r_i}^T
\]

where \( A_1 = A \),

\[
L_i = I + [0, \ldots, 0, I_{i+1,j}, \ldots, I_{n,j}]^T e_i^T
\]
and \( e_i \) is the \( i \)-th unit coordinate vector.

If the algorithm stops at step \( i \), then \( I_{r,i} \) is the unit matrix \( I \). \( L_i \) should be equal to the diagonal matrix with first \( i - 1 \) diagonal elements being 1 and all other elements being 0. \( I_{r,i} \) and \( L_i \) for \( i = 1, \ldots, n \) can be regarded as unit matrix \( I \).

When the algorithm terminates, matrix \( A \) is decomposed into

\[
A = L_n L_n^T
\]
where \( L_n = I_{1,r_1}L_1I_{2,r_2}L_2 \cdots I_{n,r_n}L_n. \)

Notice that \( L_n \) is not a lower triangular matrix. But if we define

\[
P = I_{1,r_1}I_{2,r_2} \cdots I_{n-1,r_{n-1}}
\]
then the matrix

\[
L = P^T L_n
\]
is a lower triangular matrix. Obviously, \( P \) is a permutation matrix: \( P^T P = I \).

From the above constructive algorithm, we can have the following theorem.

**Theorem 4.1** A symmetric and positive semi-definite matrix can always be decomposed to \( A = (PL)(PL)^T \) form, where \( P \) is a permutation matrix, and \( L \) is a lower triangular matrix.

Because of the diagonal pivoting, the absolute value of all elements in the Gaussian elimination matrix \( L_i \) is less than or equal to 1. Therefore, the decomposition process is numerically stable.

## 5 BEST APPROXIMATION OF INDEFINITE MATRICES

In real practice, because of missing data or numerical errors, the estimated covariance matrix \( \Sigma \) may be indefinite (i.e., it contains one or more negative eigenvalue). In such abnormal cases, we propose to add a small diagonal matrix \( E \) to \( \Sigma \) and use the new matrix \( \hat{\Sigma} = \Sigma + E \) as the covariance. If the smallest element in \( E \) is greater than or equal to the absolute value of the negative eigenvalues of \( \Sigma \), then \( \hat{\Sigma} \) will be positive definite or semi-definite. We can use the usual Cholesky decomposition or our modified Cholesky algorithm.

Recall the definition of Frobenius norm:

\[
\|A\|_F = \sqrt{\sum_{i,j} |a_{i,j}|^2}.
\]

Under the Frobenius norm,

\[
\|\Sigma - \Sigma\|_F = \|E\|_F
\]
can reach a minimum for a special construction of \( E \). For example, see Higham (1988). We consider a spectral decomposition of \( \Sigma \). Let

\[
\Sigma = Q \Lambda \Lambda^T Q^T
\]
with \( Q \) orthogonal and \( \Lambda \) diagonal. We define \( D \) and \( \hat{\Sigma} \) by

\[
D = \text{diag}(\max(0, \lambda_{11}), \ldots, \max(0, \lambda_{nn}))
\]
and

\[
\hat{\Sigma} = Q \Lambda Q^T.
\]

Here \( E \) can be picked as

\[
E = \hat{\Sigma} - \Sigma.
\]

\( \hat{\Sigma} \) is the unique best semi-positive approximation of \( \Sigma \) with respect to the Frobenius norm.

Is it possible that some of the main diagonal elements of \( \hat{\Sigma} \) generated according to above procedure are not positive? If it is true, its corresponding correlation matrix of \( \hat{\Sigma} \) will not have ones on the diagonal. Certainly having ones on the diagonal of the correlation matrix are important and necessary. We use the following result as an answer.

**Theorem 5.1** All diagonal elements of \( \hat{\Sigma} \) are greater than or equal to their corresponding diagonal elements in \( \Sigma \) and therefore, are positive:

\[
\tilde{\sigma}_{ii} \geq \sigma_{ii} > 0, \quad i = 1, \ldots, n.
\]

**Proof** From the spectral decomposition of \( \Sigma \), we have

\[
\sigma_{ii} = \sum_{j=1}^{n} \lambda_{jj} q_{ij}^2, \quad i = 1, \ldots, n
\]
and

\[
\tilde{\sigma}_{ii} = \sum_{j=1}^{n} d_{jj} q_{ij}^2, \quad i = 1, \ldots, n.
\]
By the definition of \( D \), we have 
\[
d_{ii} = \max(0, \lambda_{ii}), \quad i = 1, \ldots, n
\]
and
\[
d_{ii} \geq \lambda_{ii}, \quad i = 1, \ldots, n.
\]
Therefore
\[
\tilde{\sigma}_{ii} \geq \sigma_{ii} > 0, \quad i = 1, \ldots, n.
\]
\[\square\]

6 Mixture of Normal Distributions

In this section, we review the univariate mixture of \( k \) normal distributions.

In general, the cumulative distribution function (cdf) of a mixture of \( k \) normal random variable \( X \) is defined by
\[
F(x) = \sum_{j=1}^{k} p_j \Phi \left( \frac{x - \mu_j}{\sigma_j} \right),
\]
where \( \Phi \) is the cdf of \( N(0,1) \). Therefore its probability density function (pdf) is
\[
f(x) = \sum_{j=1}^{k} p_j \phi_j(x; \mu_j, \sigma_j^2),
\]
where, for \( j = 1, \ldots, k, \)
\[
\phi_j(x; \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi} \sigma_j} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}},
\]
\[
0 \leq p_j \leq 1, \quad \sum_{j=1}^{k} p_j = 1,
\]
with mean
\[
\mu = \sum_{j=1}^{k} p_j \mu_j
\]
and variance
\[
\sigma^2 = \sum_{j=1}^{k} p_j (\sigma_j^2 + \mu_j^2) - \mu^2.
\]

7 Generating Multivariate Mixtures of Normal Variates

In this section, we propose a modified Cholesky decomposition in generating a multivariate mixture of normal distributions with arbitrary covariance matrix.

We assume that \( X = (X_1, \ldots, X_n)^T \) is a random vector of daily changes in market variables. The marginal distribution of each component \( X_i \) is a univariate mixture of \( k_i \) normals with pdf:
\[
f_{X_i}(x) = \sum_{h=1}^{k_i} p_{ih} \frac{1}{\sqrt{2\pi} \sigma_{ih}} e^{-\frac{(x-\mu_{ih})^2}{2\sigma_{ih}^2}},
\]
where
\[
0 \leq p_{ih} \leq 1, \quad h = 1, \ldots, k_i, \quad \sum_{h=1}^{k_i} p_{ih} = 1, \quad i = 1, \ldots, n.
\]
The covariance matrix of \( X \) is
\[
\Sigma_X = [\sigma_{ij}(X)],
\]
where \( \sigma_{ij}(X) = \text{Cov}(X_i, X_j) \) and \( i, j = 1, \ldots, n \).

Based on our results of Propositions 3.2.2 and 3.2.3 of Wang and Taffee (2001), generating a multivariate mixture of normals with the marginal pdfs of (10) and covariance matrix of \( \Sigma_X = [\sigma_{ij}(X)] \) can be accomplished as follows:

Algorithm 7.1 Inputs: integer \( n \), positive semi-definite covariance matrix \( \Sigma_X \), mixture of normal parameters \( p_{ij}, \mu_{ij}, \sigma_{ij}, \quad l = 1, \ldots, k_i, \quad i = 1, \ldots, n \).

1. Calculate \( \sum_{ij} \), where \( \sigma_{ij}(Y) = \)
   \[
   \left\{ \begin{array}{ll}
   \sigma_{ij}(X) = \sum_{h=1}^{k_i} \sum_{l=1}^{k_l} p_{ih} p_{lj} (\mu_{ih} - \mu_i) (\mu_{lj} - \mu_j), \\
   \sum_{i=1}^{k_i} p_{ih} 
   \end{array} \right. 
   \]
   for \( i \neq j \), and \( i, j = 1, \ldots, n \)
   \[1, \quad \text{for} \ i = j, \text{and} \ i, j = 1, \ldots, n.\]

2. Decompose the \( \Sigma_X \) using Cholesky Decomposition with Diagonal Pivoting.
3. Generate \( Z = (Z_1, \ldots, Z_n)^T \), where the \( Z_i \)s are from \( N(0,1) \).
4. Generate \( U = (U_1, \ldots, U_n)^T \), where the \( U_i \)s are from \( U(0,1) \).
5. Calculate \( Y = (Y_1, \ldots, Y_n)^T \) from \( Y = L.Z = I_{1:p_1} L_1 I_{2:p_2} L_2 \cdots I_{L:p_L} L_L.Z \).
6. Return \( \tilde{X} = (X_1, \ldots, X_n)^T \), where \( X_i = \sum_{h=1}^{k_i} (\sigma_{ih} Y_i + \mu_{ih}) f \left( \int_{-\infty}^{(x_{hi} - \mu_{hi})/\sigma_{hi}} \right), \quad \sum_{i=1}^{n} p_{ih} \leq U_i < \sum_{i=1}^{n} p_{ih} \}
   \) and \( \sum_{i=1}^{n} p_{ih} = 0. \)

Theorem 7.1 The random vector \( X \) generated from the previous algorithm is a multivariate mixture of normals.
with the marginal pdfs of (10) and covariance matrix of
\[ \Sigma_X = \begin{bmatrix} \sigma_{ij}(X) \end{bmatrix}. \]

8 CONCLUSIONS

Mixture of normals is a more general and flexible distribution for fitting the market data of daily changes. How to handle the covariance matrix is very difficult sometimes in generating random vectors. The classical three-step generating method is not efficient. Instead, generating market variables using the mixture of normal distributions is very efficient and more accurate. The input covariance matrix can be derived analytically without solving any numerical equations. Feedback from financial industry shows that a modified Cholesky decomposition is needed to handle the semi-definite situation. Sometimes the estimated covariance is indefinite. A best approximation of indefinite matrices in the Frobenius norm is provided here to form a semi-definite matrix. Thus the modified Cholesky decomposition can still be used. In theory, the covariance matrix is positive semi-definite. Quality market data and reasonable estimation should produce a positive semi-definite covariance estimate. Here, sample covariance is a good candidate to estimate the covariance. The purpose of this paper is to answer some questions from finance industry while implementing our algorithm.

ACKNOWLEDGMENTS

We wish to express our gratitude to Colm O’Cinneide of Deutsche Bank for providing us this topic. Thanks are also due to Charles Kicey of Valdosta State University, Michael Taaffe of Virginia Tech, and two anonymous referees for their valuable comments and suggestions.

Jin Wang’s research was supported in part by NSF grant DMI-0122173 to Valdosta State University.

REFERENCES


AUTHOR BIOGRAPHIES

JIN WANG is a Professor in the Department of Mathematics and Computer Science at Valdosta State University. He received his Ph.D. degree in Operations Research from Purdue University. His research interests include mathematical finance, Monte Carlo methods, portfolio optimization, risk management, stochastic modeling and optimization, supply chain management, and applied probability and statistics. His e-mail address is <jwang@valdosta.edu>.

CHUNLEI LIU is an Assistant Professor in the Department of Mathematics and Computer Science at Valdosta State University. He received his M.S. in Computational Mathematics from Wuhan University in China and his Ph.D. in Computer and Information Science from The Ohio State University. His research is in the areas of numerical analysis and computer networks. His email address is <cliu@valdosta.edu>.