Method Article

Multi-factor association dependence modelling for project risk analysis

Byung-Cheol Kim*

*Corresponding author.

Project and Supply Chain Management, Black School of Business, Penn State Erie, the Behrend College, 5101 Jordan Road, Erie, PA, 16563-1400 USA

A B S T R A C T

This article presents an analytic (non-simulation) dependence model for quantitative project risk analysis. The multi-factor association model (MFAM) accounts for multiple association factors in typical project settings and provides a closed-form solution to a complete and mathematically consistent correlation matrix. Given standardized, ubiquitous project plans (e.g., work breakdown structure, resource allocation, or risk register), the MFAM establishes a hierarchical tree of association factors, which is subsequently encoded into an analytic model for quantitative risk analysis. In this article, we present the MFAM programmed in Microsoft Excel and demonstrate the computational efficiency of the MFAM using two alternative schedules with distinct resource utilizations. With the enhanced robustness in dealing with multiple risk factors in a project and the computational efficiency from the non-simulation second-moment approach, the MFAM offers additional flexibility for and scalability to large-scale project risk analysis problems. The modelling procedures and solutions presented in this article highlight three potentials of the MFAM as a robust quantitative risk analysis tool.

- The MFAM can be fully programmed in Microsoft Excel using only basic cell functions.
- Under mild assumptions, the MFAM provides reliable risk estimates comparable to Monte Carlo simulation.
- The MFAM is scalable to high-dimensional risk problems (i.e., with multi-thousands or more) with an affordable computational burden.

© 2021 The Author. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

A R T I C L E   I N F O

Method name: Multi-Factor Association Model (MFAM)

Keywords: Project risk analysis, Multi-factor dependence model, Structured association, Second-moment approach

Article history: Received 6 May 2021; Accepted 4 July 2021; Available online 5 July 2021

DOI of original article: 10.1016/j.ejor.2021.04.043

* Corresponding author.

E-mail address: buk70@psu.edu

https://doi.org/10.1016/j.mex.2021.101443

2215-0161/© 2021 The Author. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Specifications table

| Subject area | Engineering |
|--------------|-------------|
| More specific subject area | Project risk analysis |
| Method name | Multi-Factor Association Model (MFAM) |
| Name and reference of original method | EJOR-D-20-01833R1 |

"Multi-factor Dependence Modelling with Specified Marginals and Structured Association in Large-scale Project Risk Assessment", European Journal of Operational Research.

*Method details

Outline

This article presents the multi-factor association model (MFAM) for large-scale project risk analysis [3]. The MFAM is an analytic (non-simulation) model that generates a mathematically consistent correlation matrix given marginal distributions and auxiliary risk factors that induce inter-variable associations between the marginals.

In practice, the MFAM can be used to enhance the realism of dependence modelling and risk analysis in three notable ways.

- **Multi-factors.** The MFAM enhances the flexibility of dependence modelling by accounting for multiple risk factors commonly observed in real-life project settings (e.g., weather, work types, resources, and risk register).
- **Scalability.** The MFAM provides a parameterized, closed-form solution to the correlation matrix generation problem, which is scalable to large-scale projects according to the data available in individual projects.
- **Computational tractability.** The MFAM adopts an analytic approach (i.e., the second-moment approach) and provides a computationally tractable algorithm to generate a mathematically consistent correlation matrix.

This article highlights the modelling details and applications of the MFAM in typical project risk analysis problems. Extended applications of the KFAM in more general settings are straightforward.

Multi-factor association modelling

Dependence modelling for risk analysis can be carried out in two ways: the correlation-driven approach and the factor-driven approach [3]. The correlation-driven approach builds a correlation matrix from all pairwise correlation coefficients among the dependent variables in a system. In contrast, the factor-driven approaches introduce auxiliary variables as a proxy of the pairwise association among relevant random variables [1,5].

The primary advantages of the factor-driven approaches are twofold. First, it is often more intuitive to elicit inter-variable associations in terms of the common factors that may influence two or more random variables. Second, a correlation matrix generated by a factor-driven approach is guaranteed to be mathematically consistent [3].

Specifically, the MFAM can be implemented following the four steps in Fig. 1.

**Four steps of the MFAM Analysis:**

1. **Specify marginal variables.** Primary random variables of an uncertain system are identified and specified in terms of probabilistic estimates. To this end, the uniform, normal, triangular, and PERT distributions are commonly used.
2. **Specify association factors.** Association factors (AFs), also called common causes, are identified and specified as independent random variables.

3. **Load AFs to marginals.** The relationships between the marginal variables and the risk factors are specified.

4. **Generate a correlation matrix.** Once a vector of factored random variables is established with a known mathematically consistent correlation matrix, the risk model can be analyzed as a fully-specified joint distribution. Then the results can be visualized in terms of the risk profiles (i.e., histograms and cumulative probability distributions).

**MFAM algorithm**

The MFAM adopts a multiplicative factor approach in which factor-driven association is generated by multiplying all relevant AFs to individual marginals.

\[
\{X_i\} = \{R_i b_i\}, \quad R_i = \prod_{k \in \theta_i} r_{ik}
\]

where \(i = 1, \ldots, d\); \(R_i\) is an aggregate factor defined as a product of the AFs \((r_{ik})\) linked to a variable \(b_i\), and; \(\theta_i\) is a set of AFs associated with \(b_i\). Since the marginals of \(X\) are constructed from \(b\) multiplied by the relevant AFs, the former is referred to as ‘factored’ variables, whereas the latter is referred to as ‘base’ variables hereinafter.

This section presents the generic algorithm for the MFAM programming. The key notations used in the derivation are summarized below.

- \(X = (X_1, \ldots, X_d)^T\): a \(d \times 1\) vector of factored variables.
- \(b = (b_1, \ldots, b_d)^T\): a \(d \times 1\) vector of base variables.
- \(r = (r_1, \ldots, r_K)^T\): a \(K \times 1\) vector of AFs.
- \(\theta_i\): a set of AFs associated with \(b_i\).
- \(\Psi = [d, K]\): a \(d \times K\) allocation matrix (AM) of the \(K\) association factors loaded onto \(d\) base variables.
- \(R_i = \prod_{k \in \theta_i} r_{ik}\): an aggregate factor as a product of the AFs associated with \(b_i\).
- \(E[\cdot]\): the expectation operator that returns the mean value of a variable.
- \(V[\cdot]\): an operator that returns the variance of a variable.
- \(Sd[\cdot]\): an operator that returns the standard deviation of a variable.

A generic algorithm for the MFA correlation computation is presented below.
**Fig. 2.** Resource allocation scenarios.

---

**MFAM algorithm [3]:**

Module 1: Read input parameters: \( b, r, \) and \( \Psi \).

- §1.1 Read the base variables, \( b = (b_1, ..., b_j)^T \): the means (\( \mathbf{E}_b \)) and variances (\( \mathbf{V}_b \)).
- §1.2 Read the association factors, \( r = (r_1, ..., r_j)^T \): the means (\( \mathbf{E}_r \)) and variances (\( \mathbf{V}_r \)).
- §1.3 Read the allocation matrix, \( \Psi (d \times K) \).

Module 2: Compute auxiliary parameter vectors \( \mathbf{E}_{b1} \) and \( \mathbf{E}_{b2} \).

- §2.1 \( \mathbf{E}_{b1} = [\mathbf{E}[b_i]] = \mathbf{V}_b + \mathbf{E}^{-2} \).
- §2.2 \( \mathbf{E}_{b2} = [\mathbf{E}[r_i]] = \mathbf{V}_r + \mathbf{E}^{-2} \).

Module 3: Compute the means and variances of the factored variables, \( \mathbf{X} = (X_1, ..., X_d)^T \).

- §3.1 \( \mathbf{E}_i = \{\{\mathbf{E}^{-\theta_i}\}(\mathbf{E}_b)\}, \quad i = 1, ..., d \)
- §3.2 \( \mathbf{V}_i = \{\{\mathbf{E}^{-\theta_i}\}(\mathbf{E}_b)\} - \{\{\mathbf{E}^{-2\theta_i}\}(\mathbf{E}_b)^2\}, \quad i = 1, ..., d \)

Module 4: Generate the correlation coefficient matrix (\( i = 1, ..., d; \quad j = 1, ..., d \))

- §4.1 \( \Sigma_{i,j}(b, r, \Psi) = \begin{cases} \frac{\mathbf{E}_{i,j}}{\mathbf{V}_{i,j} \mathbf{V}_{j,j}} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \)

where \( \mathbf{V}_{i,j} = \{\mathbf{E}^{\theta_i \theta_j}\}(\mathbf{E}_b) \)

\( \mathbf{E}_{i,j} = \{\mathbf{E}^{\theta_i \theta_j}\}(\mathbf{E}_b) \)

\( \mathbf{E}[\mathbf{R}_{j,j}] = \{\mathbf{E}^{\theta_i \theta_i}\}(\mathbf{E}_b) \)

Note: \( \theta_{i,j} \) is the intersection of \( \theta_i \) and \( \theta_j \); whereas \( \partial \theta_{i,j} \) (or \( \partial \theta_{j,i} \)) is the set difference of \( \theta_i \) and \( \theta_j \) (or \( \theta_j \) and \( \theta_i \)). \( \Psi \) is a vector of the \( i \)-th (\( j \)-th) row of \( \Psi \); and, \( \odot \) is the Hadamard operator for entry-wise product or power computation [2].

**Test projects**

The MFAM algorithm provides a robust solution to the problem of project risk modelling and analysis. In this paper, we consider a subproject of six work packages (WPs): \( wp_1, wp_2, ..., wp_6 \). Our goal is to assess the cost uncertainty of the subproject under two resource scenarios: Plan 1 and Plan 2.

- **Plan 1 (One-crew scenario).** All WPs are carried out by one crew (Crew A) along with one crane (and crane operator) for \( wp_2 \) and \( wp_6 \). As a result, the six WPs form a single path of work sequence. See Fig. 2.

- **Plan 2 (Two-crew scenario).** Two crews (Crew A and Crew B) are allocated concurrently, and each crew carries out three WPs. See Fig. 2 for the details. As a result, Plan 2 offers a time-saving schedule. Yet, Plan 2 increases the workflow complexity with two parallel paths.

The two plans reflect two common challenges in project resource planning: resource leveling and schedule acceleration. That is, whenever possible the maximum resource demands in a project should be minimized simply to make things easier to manage. Yet, there are situations where hiring more people or other renewable resources (e.g., equipment) accelerates the progress and ultimately the project completion.

Now we are going to carry out a quantitative analysis of the three plans and provide the project manager with some actionable information. To start with, we observe that the cost performance of individual WPs is influenced by the production rates of the resources allocated to the WP. This notion of qualitative association induced by common work types or common resources shared by multiple
performance units is defined as the structured association (SA) in Kim [3]. Using the SA approach, the two plans in Fig. 2 can be schematically represented as the association factor trees in Fig. 3.

**Fig. 3**. Association factor trees. $r_1$ (Crew A), $r_2$ (Crane), and $r_3$ (Crew B)

**Fig. 4**. Excel MFAM for Plan 2. (Source file: XL-MFAM [3]_Plan-2.xlsx)

**MFAM programming in Microsoft Excel®**

The MFAM algorithm in Section 3 is programmed in Microsoft Excel, following the four-steps in Fig. 1. The resulting Excel MFAM for Plan 2 is shown in Fig. 4.
Step 1. Specify marginal variables:

The WP costs are estimated in terms of the three-point technique. We consider a situation where all WPs have the same three-point estimates: Optimistic ($o$) = 0.7, Most Likely ($m$) = 1.0, and Pessimistic ($p$) = 1.6. Note that the three-point estimates of the WP costs are stored in Cells “C8:E13” of the Excel MFAM in Fig. 4.

The MFAM adopts the second-moment approach. Thus, a three-point estimate needs to be converted to the corresponding mean and variance. We consider a case where the WP costs have the triangular distribution. Then the corresponding second-moment estimates (SME) are determined as:

$$E[b_i] = \frac{o_i + m_i + p_i}{3}, \quad i = 1, ..., 6$$

$$Sd[b_i] = \sqrt{\frac{o_i^2 + m_i^2 + p_i^2 - o_i m_i - m_i p_i - p_i o_i}{6}}, \quad i = 1, ..., 6$$

Note that the variances are replaced by standard deviations (Sd) for computational simplicity. In the Excel MFAM, the SMEs are calculated in Cells “F8:G13”. For example, $E[b_1]$ and $Sd[b_1]$ in Cell “F8:G8” are

“F8”: “=(C8+D8+E8)/3” and

“G8”: “=SQRT((SUMSQ(C8:E8)-C8^2-D8^2-E8^2-C8)/18)”.

In addition, an auxiliary variable, the mean of a squared base estimate ($E_{b0}$) in MFAM Algorithm §2.1, is introduced in Cells “H8:H13”.

$$E[b_i^2] = V[b_i] + E[b_i]^2, \quad i = 1, ..., 6$$

For example, $E[b_1^2]$ in Cell “H8” is:

“H8”: “=G8^2+F8^2”

Step 2. Specify association factors:

The AFs identified in the association factor trees (Fig. 3) are estimated in terms of their means and standard deviations and presented in Table 1. Note that Crew B parameters are chosen in a way that represents a situation where a secondary crew (Crew B) is less efficient than the primary crew (Crew A) available in the project. Specifically, the mean = 1.2 (> 1.0) reflects that Crew B is less productive, on average, than Crew A. Moreover, the standard deviation = 0.3 also indicates that the Crew B performance is less reliable than Crew A (with standard deviation = 0.1).

In the Excel MFAM, the risk factor means and standard deviations are stored in Cells “C17:E18”. Note the means of the squared risk factors ($E_{r_k}$) in MFAM Algorithm §2.2 are also introduced in Cells “C19:E19”.

$$E[r_k^2] = V[r_k] + E[r_k]^2, \quad k = 1, 2, 3$$

For example, $E[r_2^2]$ in Cell “C19” is:

“C19”: “=C18^2+C17^2”
Fig. 5. Excel MFAM coding - Modules 1 and 2.

Module 1. Read input parameters: b, r, and Ψ.

§1.1 Means of the base variables, Es = Cells "F8:F13". Copy (F1.1) to the range.
Standard deviations of the base variables, Sds = Cells "G8:G13". Copy (F1.2) to the range.
§1.2 Means of AFs, Er = Cells "C17:E17". Manual input.
Standard deviations of the AFs, Sdr = Cells "C18:E18". Manual input.
§1.3 The allocation matrix, AM(Ψ) = Cells "I8:K13". Manual input.

Module 2. Auxiliary parameter vectors Ebs and Ers.

§2.1 Ebs = Cells "H8:H13". Copy (F1.3) to the range.
§2.2 Ers = Cells "C20:E20". Copy (F2) to the range.

Fig. 6. Excel MFAM coding - Module 3.

Module 3. Factored variables, X.

§3.1 The means of X, Ex = Cells "L8:L13". Copy (F3.1) to the range.
§3.2 The standard deviations of X, Sdx = Cells "M8:M13". Copy (F3.2) to the range.

Step 3. Load AFs to marginals.

Given a set of base (factor-free) estimates and a set of AFs, the confounding relationships between them are represented using an allocation matrix (AM, Ψ). This process is straightforward and self-evident. For example, the Plan 2 factor tree in Fig. 3 is coded in Cell "I8:K13" of the Excel MFAM (Fig. 4).

It is worth noting that the three steps discussed above correspond to the first two modules of the MFAM algorithm. The corresponding cells of the modules in the Plan 2 Excel MFAM are summarized in Fig. 5 below.

Then, the factored variable calculation in Module 3 of the MFAM algorithm can be coded as follows. First, the mean values of the factored variables are computed from MFAM §3.1:

\[
E_X = \left\{ \left( \prod E_r^{\psi_i} \right) E_b \right\}_i, \quad i = 1, \ldots, 6
\]

For example, \(E_X(i = 1)\) in Cell "L8" is

"L8": "={F8*PRODUCT($C$17:$E$17^I8:K8)}". (F3.1)

Likewise, the standard deviation of a factor variable is computed from MFAM §3.2:

\[
Sd_X = \sqrt{\left( \prod E_r^{\psi_i} \right) E_b - \left( \prod E_r^{2\psi_i} \right) E_b^2}, \quad i = 1, \ldots, 6
\]

For example, \(Sd_X(i = 1)\) in Cell "M8" is

"M8": "={SQRT(H8*PRODUCT($C$19:$E$19^I8:K8)-F8^2*PRODUCT($C$17:$E$17^2*I8:K8))}". (F3.2)

Then Module 3 of the Excel MFAM is completed as in Fig. 6.

Step 4. Construct the correlation matrix.

Given the three inputs (i.e., b, r, and Ψ), the MFAM offers a closed-form solution to the problem of generating a mathematically consistent correlation matrix. In the Excel MFAM, the correlation formula
Excel MFAM.

Module.4 The correlation coefficient matrix \( (i = 1, \ldots, d; \ j = 1, \ldots, d) \)

\[ \Sigma_r(i, j | b, r, \Psi) = \text{Cells "C24:H29"} \]

Copy (F4) to the off-diagonal cells in the range.

Input 1’s to diagonal cells in the range.

For ease of understanding, we decomposed the correlation formula into five components as follows.

\[ \Sigma_r(i, j | b, r, \Psi)_{i \neq j} = \frac{V[R_{i,j}]}{(S_d X)_i(S_d X)_j} [E[R_{i,j}]](E_r)_i(E_r)_j \]

For example, the cell formula for \( \Sigma_r(1, 2) \) in Cell “D24” consists of five components:

\[ "D24": \{(=F4.a)\cdot(F4.b)^\ast(F4.c)^\ast(F4.d)/\ast(F4.e))\} \]

\[ \begin{array}{l}
V[R_{1:2}] = \{(\text{PRODUCT}(E.r[INDEX(AM,D$23,0)]))\cdot(INDEX(AM,$B24,0))\cdot(INDEX(AM,$D23,0)))\} \quad \text{(F4.a)}
\end{array} \]

\[ \begin{array}{l}
E[R_{1:2}] = \{(\text{PRODUCT}(E.r[INDEX(AM,$B24,0)]))\cdot(INDEX(AM,$B24,0))\cdot(INDEX(AM,$D23,0)))\} \quad \text{(F4.b)}
\end{array} \]

\[ \begin{array}{l}
E[R_{1:2}] = \{(\text{PRODUCT}(E.r[INDEX(AM,$B24,0)]))\cdot(INDEX(AM,$B24,0))\cdot(INDEX(AM,$D23,0)))\} \quad \text{(F4.c)}
\end{array} \]

\[ \begin{array}{l}
(E_b)_1(\ E_b)_2 = \{(\text{INDEX}(E.b,$B24))\cdot(INDEX(E.b,$D23))\} \quad \text{(F4.d)}
\end{array} \]

\[ \begin{array}{l}
(S_d X)_1(\ S_d X)_2 = \{(\text{INDEX}(S.x,$B24))\cdot(INDEX(S.x,$D23))\} \quad \text{(F4.e)}
\end{array} \]

Note: \( E.b = "F8:F13" \); \( S.x = "M8:M13" \); \( AM = "I8:K13" \); \( E.r = "C17:E17" \); \( E.rs = "C19:E19" \).

Once the correlation formula is entered for \( \Sigma_r(1, 2) \), it can be copied to all other off-diagonal elements of the correlation matrix. Finally, the diagonal elements of the correlation matrix are entered as 1’s, which completes the MFAM modelling.

Method validation

The Excel MFAM in Section 5 is generic and can be used for both Plan 1 and 2 in Fig. 3. The inputs and resulting correlation matrix for Plan 1 are presented in Fig. 8.

Given a set of the factored costs and the corresponding correlation matrix, calculation of the first two moments of the total cost is straightforward using basic statistics. That is, the mean of the sum of random variables is the sum of their means; and the variance of the sum of dependent random variables is the sum of the elements in the variance-covariance matrix.

\[ E[TC] = \sum_{i=1}^{6} X_i \]

\[ V[TC] = S_d X \Sigma S_d X \]

For example, given the Excel MFAM (Figs. 4 and 8), the mean and variance of the total cost (TC) are calculated as:

\[ \text{mean(TC)} = \text{SUM(L8:L13)} \]

\[ \text{variance(TC)} = \text{MMULT(MMULT(TRANSPOSE(M8:M13),C24:H29), M8:M13)} \]

The MFAM results of the two plans are summarized in Table 2. The table also shows the results from Monte Carlo simulation. The simulations were carried out using @RISK from Palisade (https://www.palisade.com) with 50,000 iterations.
Table 2

| Cases    | Plan 1          | Plan 2          |
|----------|-----------------|-----------------|
|          | Mean | St. dev. | P-90 | Mean | St. dev. | P-90 |
| Independent | 6.600 | 0.458    | 7.187 | 6.600 | 0.458    | 7.187 |
| MFAM      | 6.600 | 0.835    | 7.670 | 7.260 | 1.190    | 8.785 |
| Simulation| 6.600 | 0.834    | 7.694 | 7.260 | 1.191    | 8.819 |

Note: P-90 represents the total cost at 90% confidence level.

For ease of comparison, the resulting risk profiles of the total costs are displayed in Fig. 9. The results also show the effects of the factor-induced correlations against the independence assumption. Note that the MFAM distributions in Fig. 9 are fitted by the lognormal distribution to account for possible skewness commonly observed in project costs and times. Overall, the results show a good match between the MFAM and the simulation. For more detailed comparison, please refer to the supplementary Excel MFAM files.

In practice, quantitative risk assessments provide vital information for risk-adjusted project planning and control. Specifically, the results from MFAM can be used to set the amount of contingency reserve at a chosen confidence level. For this purpose, the total cost risks in Fig. 9 can be displayed as cumulative risk profiles as in Fig. 10. From the graph, the decision maker intuitively grasps the effects of dependence on the total cost risk under the two resource plans. More specifically,
Fig. 9. Cost risk distributions. ("IND" represents the total cost under the independence assumption. "Sim" represents the results from simulation)

Fig. 10. Effects of factor-induced dependence on risk-adjusted budgets.

$\Delta_1$ and $\Delta_2$ in Fig. 10 represent the additional budgets for the contingency at P-90 under Plan 1 and Plan 2, respectively.

Conclusions

This article presents the underlying ideas, generic algorithm, and programming codes of the MFAM for project risk analysis with correlation. The MFAM was programmed in Microsoft Excel using only the cell functions commonly used in practice.

The MFAM provides an analytic (non-simulation) and closed-form solution to the problem of correlation matrix specification and estimation. As a result, the MFAM offers distinct computational advantages over simulation. The MFAM can be further calibrated and optimized according to the data available in individual projects, while preserving the scalability to high-dimensional project risks.
The computational efficiency of the MFAM can be further improved by programming the MFAM algorithm into a VBA (Visual Basic for Applications) code in Excel, which is reserved for the readers as an exciting learning-by-doing exercise.

Uncited references

[4]
Supplementary material and/or Additional information:

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author sincerely appreciates the insightful comments and encouraging suggestions from the Editor and the three reviewers during the review of the European Journal of Operational Research.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2021.101443.

References

[1] S. Cho, A linear Bayesian stochastic approximation to update project duration estimates, Eu. J. Operation. Res. 196 (2) (2009) 585–593.
[2] S.M. Fallat, C.R Johnson, Hadamard powers and totally positive matrices, Linear Algebra Appl. 423 (2-3) (2007) 420–427.
[3] B.-C. Kim, Dependence modeling for large-scale project cost and time risk assessment: additive risk factor approaches, IEEE Trans. Eng. Manage. (2021) Early Access: https://ieeexplore.ieee.org/abstract/document/9332228.
[4] B.-C. Kim, Multi-factor dependence modelling with specified marginals and structured association in large-scale project risk assessment, Eu. J. Operation. Res. (2021) Under Review.
[5] J.R. van Dorp, A dependent project evaluation and review technique: A Bayesian network approach, Eu. J. Operation. Res. 280 (2) (2020) 689–706.