Project Management Development – Practice and Perspectives
Fifth International Scientific Conference on Project Management in the Baltic Countries
April 14-15, 2016, Riga, University of Latvia

THE IMPACT OF CORRELATIONS IN THE CRITICAL FIELD

Wolfgang Tysiak, University of Applied Sciences, Dortmund, Germany[49];
Tobias Tietz, Heinrich Heine University, Düsseldorf, Germany

Abstract

In almost every project there will be uncertainties in the duration of the individual tasks. Therefore an elaborated risk management is needed and the best known tool to handle this in practice is the PERT approach. It has been shown in former publications (Tysiak (2015a), (2015b)) that the PERT approach comprises a lot of inaccuracies, contortions, and systematic miscalculations. Some of these disadvantages can be avoided or reduced by substituting PERT with Monte Carlo simulation. Unfortunately one unrealistic assumption still remains even in the Monte Carlo approach: The supposition that the durations of the individual tasks are independent from each other. In the present contribution it will be shown how to introduce correlations into the Monte Carlo approach and which impact they might have.

Key words: *project management, risk management, PERT, critical field, Iman-Conover approach, correlated distributions*
JEL code: G32

Introduction

Since a project is said to be “a temporary endeavor undertaken to create a unique product, service, or result” (PMI (2010)), there will always be the need of implementing some kind of risk management in project management (c.f. PMI (2010), Schelle/Ottmann/Pfeiffer (2006)). Risks in projects can occur in different dimensions, such as time, cost, quality etc. In this contribution we will only consider uncertainties related to time, but as everybody knows, who has experience in project management, a prolongation of a project normally will also affect its costs. A commonly used approach to deal with this situation is PERT (project evaluation and review technique, c.f. Kerzner (2009), Taylor (2010)), which has been developed by the United States Navy together with the OR department of Booz, Allen and Hamilton in the 1950s. Purpose of this development was to support the deployment of the Polaris-Submarine weapon system (c.f. Fazar (1959)). Unfortunately there are still some weaknesses, disadvantages, errors, and inaccuracies in using this method and therefore there still is the strong need for further improvement.

The PERT approach

PERT is based on the Critical Path Method (CPM) that was invented by DuPont in the late 50s of the last century (c.f. Kelley/Walker (1959)). CPM assumes deterministic durations of the different activities of a project and by calculating the earliest starting and finishing dates of the individual activities, it achieves the earliest finishing date of the whole project. Calculating the whole project backwards, one also gets the latest finishing and latest starting dates of the

[49] Corresponding author – e-mail address: Wolfgang.Tysiak@FH-Dortmund.de

activities. Those activities that have no buffers (the difference between the earliest starting and latest starting date) then constitute the critical path.

In PERT we do not assume deterministic durations. The durations are usually estimated by so-called three-point-estimates (optimistic, most probable, and pessimistic durations). PERT then assumes beta distributions for the given three-point-estimates, calculates the expected durations and the variances for each activity, then performs a CPM approach with the expected durations, and by this identifies a critical path. Afterwards the non-deterministic approach is introduced by calculating the convolution of the distributions along the critical path. This convolution implies that the distributions of the durations of the activities are assumed to be independent. Furthermore the resulting distribution for the duration of the whole project is assumed to be normal, which is quite reasonable due to the Central Limit Theorem.

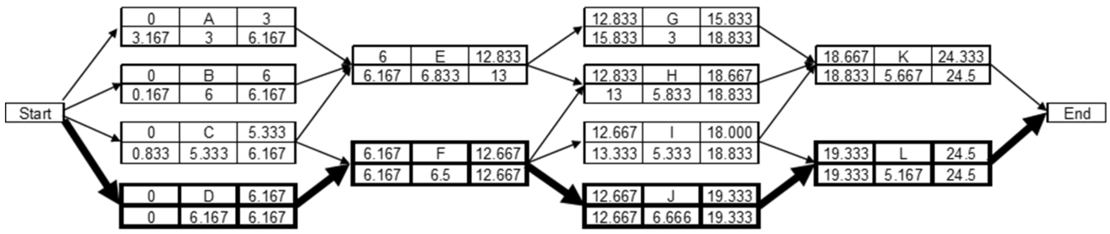
Fig. 1 shows a fictitious project plan with predecessors, optimistic durations, most probable durations, pessimistic durations, expected durations, and their variances (with assumed beta distributions), whereas fig. 2 gives us the resulting critical path. The bold line in fig. 3 shows the estimated distribution of the duration of the whole project as the independent convolution along the critical path.

Activity	Predecessors	OD	MD	PD	ED	VAR
A	-	2	3	4	3.000	0.111
B	-	3	6	9	6.000	1.000
C	-	2	5	10	5.333	1.778
D	-	4	6	9	6.167	0.694
E	A, B, C	3	7	10	6.833	1.361
F	C, D	2	7	9	6.500	1.361
G	E	2	3	4	3.000	0.111
H	E, F	3	6	8	5.833	0.694
I	F	3	5	9	5.333	1.000
J	F	2	7	10	6.667	1.778
K	G, H, I	2	6	8	5.667	1.000
L	I, J	3	5	8	5.167	0.694

source: Tysiak (2015a)

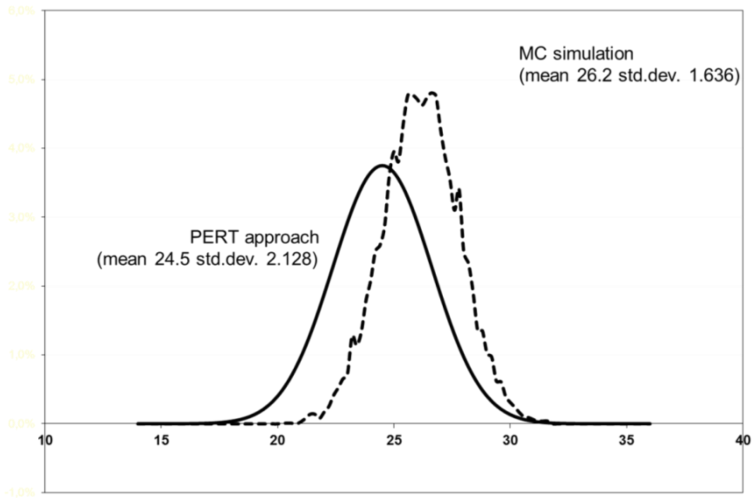
Fig. 1. **A fictitious project plan**

One of the main problems of PERT is the fact, that unfortunately the assumption of a unique critical path is not realistic. Due to the varying durations of the individual activities, it is not the case that the activities can be divided into critical and non-critical: Each activity possesses a probability between 0 and 1 to become critical. This was already mentioned by Van Slyke (1963), who called this property “criticality”. Van Slyke performed a lot of Monte Carlo simulations, which at his time of course were only possible with the deployment of large mainframes. In the end we have to admit that in the case of uncertainties, there is no unique critical path, but only a “critical field”.

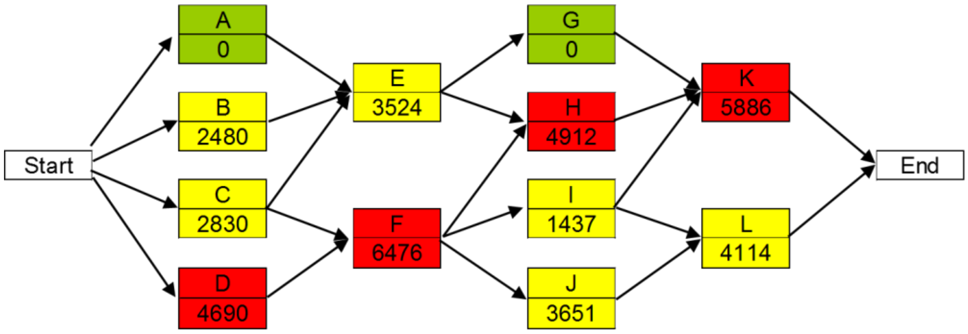


source: Tysiak (2015a)
Fig. 2. **The critical path due to PERT**

To illustrate this critical field idea, we perform Monte Carlo simulations with the given example. To make the results comparable to the PERT approach, we use exactly the same beta distributions for the durations of the individual activities (see Tysiak (2015b) for computational background). Fig. 4 shows the "critical field" as a result of 10,000 simulations, whereas the dotted line in fig. 3 compares the distribution of the total duration of the project with the PERT result. It is obvious in fig. 3 that the mean increased from 24.5 in the PERT approach to 26.2 in the Monte Carlo simulation, whereas the standard deviation decreased from 2.128 to 1.636. More meaningful in the context of risk management is to observe quantiles. If we look at the maximum duration that will be achieved with a probability of 95% (here denoted as VaR95 (value-at-risk)), we find that the VaR95 increased from 28.0 in the PERT approach to 28.8 in the Monte Carlo simulation. By this, it becomes obvious, that PERT underestimates the real risk. A more detailed explanation of this fact can be found in Tysiak (2015b). It can also be shown that this underestimation is systematical and perceivable in quite almost every project plan.



source: authors' construction
Fig. 3. **Results of PERT and Monte Carlo simulation with assumed independence**



source: authors' construction

Fig. 4. **The critical field (number of times that a node is critical)**

Correlations of the durations

By using Monte Carlo simulation instead of PERT we might get a more realistic model of the real distribution of the duration of a project. But there is still one rather unrealistic assumption: The independence of the distributions of the durations of the individual activities. If we look for example at a construction project. Then one severe risk that may occur and affect the duration of tasks might be the weather. But if we have really bad weather, this will usually affect several tasks. Or if we fear in some other project that some tasks may last a little longer because of the qualifications and talents of the staff, then this will be the case for all the activities where this staff is involved. If you think of projects, you will easily identify reasons for dependencies/correlations between the durations of individual activities and quite seldom you can really believe that all the activities are totally independent from each other. Therefore in this contribution we will introduce correlations into the Monte Carlo simulation approach and additionally we will show their impact to the resulting duration of the whole project.

In recent years, the interest in the generation of correlated random numbers (so-called "copulas"), that follow given distributions, rapidly increased. One main driver in this progress was certainly the need of such numbers in the vast field of simulation in finance (c.f. Mai/Scherer (2014), Huynh/Lai/Soumare (2008), Brandimarte (2014)).

In the current publication we follow the technique that was proposed by Iman/Conover (1982). This approach is based on the well-known method to generate multivariate normal distributed random numbers that follow a given correlation matrix C. This can easily be achieved by calculating the Cholesky decomposition $C = L \times L^t$ of the given correlation matrix into a lower triangular matrix L and its transpose L^t. By multiplying the matrix of the independently generated normal distributed random numbers with the matrix L, we get the correlated normal distributed random numbers. Iman/Conover realized that correlations can easily be achieved just by reordering the existing data. Therefore they proposed for arbitrary distributions to create independent random numbers and afterwards reorder them by using the ranks according to a multivariate normal distribution with the desired correlation matrix. Especially in the case that these arbitrary distributions are quite "good-natured" and not "too different" from the normal distribution these approach works very well.

Since beta distributions fulfil these conditions quite well, these results could be verified in our analysis. To show this, we present the following example: If we take the 12 beta distributions defined for the tasks A to L with the given correlation matrix C_1 (fig. 5) and we perform the whole process of creating the independent beta distributed random numbers, generate the Cholesky decomposition of the given correlation matrix, create the normal distributed random numbers, and then reorder the beta distributed random numbers due to the ranks given by the normal distributed random numbers, we will get the empirical correlation matrix $\hat{C}_1$ (fig. 5). The congruence between C_1 and $\hat{C}_1$ is quite obvious.

The matrix C_1 has been chosen with a lot of zero entries. The main reason for this is the fact that it is quite difficult to create correlation matrices with several negative values. It is well known, that apart from the symmetry, a correlation matrix has to be positive definite. This is equivalent to the property to possess only positive eigenvalues. In practice this complicates the construction of a valid correlation matrix very much. It is much easier to create correlation matrices with positive values than with negative values. In some sense this might be seen as a validation of "Murphy's law", because in risk management, positive correlations lead to an increase of risk, whereas negative correlations reduce risk.

This is also the reason why in the following examples the matrix C_α is chosen with a lot of positive entries, whereas the matrix C_β, which additionally contains negative values, is very sparse and only contains a few non-zero entries.

$$C_1 = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0.8 & 0 & 0 & 0 & -0.2 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -0.8 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 0.2 \\
0 & 0 & 0.6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -0.6 & 0 \\
0.8 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & -0.8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -0.4 \\
-0.2 & 0 & 0 & 0 & 0 & -0.6 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0 & -0.4 & 0 & 1
\end{pmatrix}$$

$$\hat{C}_1 = \begin{pmatrix}
1 & 0.012 & -0.007 & 0.010 & -0.003 & -0.004 & 0.797 & 0.016 & 0.003 & 0.003 & -0.195 & 0.014 \\
0.012 & 1 & 0.140 & -0.011 & 0.001 & 0.000 & 0.008 & -0.003 & -0.8 & 0.013 & -0.008 & 0.005 \\
-0.007 & 0.140 & 1 & 0.005 & 0.601 & 0.018 & -0.006 & 0.002 & -0.015 & -0.002 & -0.011 & 0.002 \\
0.010 & -0.011 & 0.005 & 1 & -0.006 & -0.007 & -0.004 & 0.409 & 0.013 & -0.006 & 0.014 & 0.212 \\
-0.003 & 0.001 & 0.601 & -0.006 & 1 & 0.015 & 0.000 & 0.006 & -0.006 & -0.003 & -0.011 & -0.002 \\
-0.004 & 0.000 & 0.018 & -0.007 & 0.015 & 1 & 0.002 & 0.008 & -0.002 & 0.007 & -0.596 & -0.001 \\
0.797 & 0.008 & -0.006 & -0.004 & 0.000 & 0.002 & 1 & 0.015 & 0.007 & 0.000 & 0.001 & -0.006 \\
0.016 & -0.003 & 0.002 & 0.409 & 0.006 & 0.008 & 0.015 & 1 & 0.006 & 0.003 & -0.002 & 0.009 \\
0.003 & -0.800 & -0.015 & 0.013 & -0.006 & -0.002 & 0.007 & 0.006 & 1 & -0.013 & 0.008 & -0.003 \\
0.003 & 0.013 & -0.002 & -0.006 & -0.003 & 0.007 & 0.000 & 0.003 & -0.013 & 1 & -0.003 & -0.406 \\
-0.195 & -0.008 & -0.011 & 0.014 & -0.011 & -0.596 & 0.001 & -0.002 & 0.008 & -0.003 & 1 & -0.008 \\
0.014 & 0.005 & 0.002 & 0.212 & -0.002 & -0.001 & -0.006 & 0.009 & -0.003 & -0.406 & -0.008 & 1
\end{pmatrix}$$

Fig. 5: **Requested (C_1) and achieved ($\hat{C}_1$) correlation matrix**

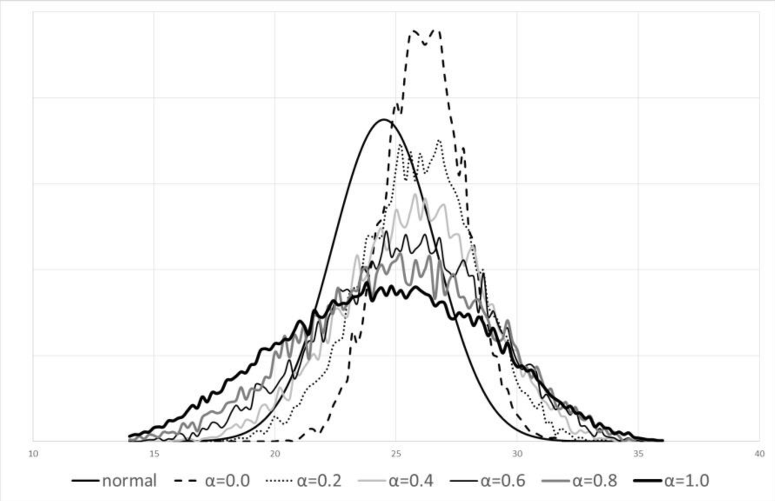
Monte Carlo simulation with correlations

To illustrate the influence of positive correlations first, we want to take a correlation matrix C_α (fig. 6) with values of α between 0 and 1, where - of course – the case α = 0 constitutes the already shown case of independency. This is an extreme example just to show the maximum impact.

$$C_\alpha = \begin{pmatrix} 1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1 & \alpha \\ \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & \alpha & 1 \end{pmatrix}$$

Fig. 6: **The matrix C_α**

The resulting densities of our example can be found in fig. 7. For the sake of comparison, the already known densities of fig. 3 are again included. It is obvious that with increasing values of α, the density becomes flatter, the tails get heavier, and the mode moves slightly a little to the left. This can also be seen if we compare the values of the mean, the standard deviation, and the VaR95 (fig. 8). We have to keep in mind that the calculation of the duration depends mostly on two operations: To build sums and maximums. If we look at two random variables and increase their correlation, this will not affect the mean of the sum, but the standard deviation will significantly increase, because there will be less compensations. If we look at the maximum of two variables, the increase of correlations will lead to a decrease of the mean of the maximum, because of less independency. The standard deviation of the maximum also depends on the relation of the means of the two random variables, but usually will not change that dramatically. Referring back to the results in fig. 8, we can postulate that the slight decrease in the mean is a consequence of the maximum operations, whereas the large increase of the standard deviation can be deduced from the summations. The resulting increase of the VaR95 is due to the fact that the decrease of the mean does not compensate the increase of the standard deviation.



source: authors' construction

Fig. 7: **The impact of positive correlations – case α**

	Mean	Std.Dev.	VaR95
PERT	***24.50***	***2.128***	***28.00***
α = 0.0	26.15	1.636	28.84
α = 0.1	26.04	2.039	29.34
α = 0.2	25.92	2.365	29.68
α = 0.3	25.80	2.651	30.05
α = 0.4	25.68	2.906	30.34
α = 0.5	25.55	3.138	30.55
α = 0.6	25.40	3.353	30.74
α = 0.7	25.25	3.553	30.90
α = 0.8	25.07	3.742	31.05
α = 0.9	24.86	3.924	31.18
α = 1.0	24.51	4.161	31.21

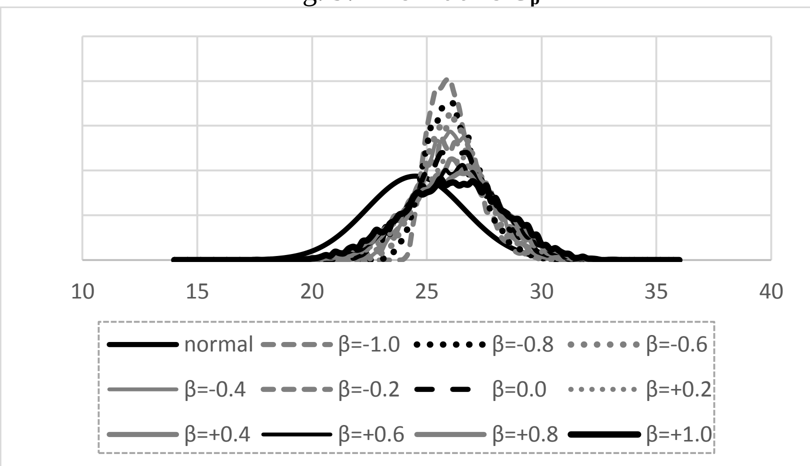
source: authors' construction

Fig. 8: **The means, standard deviations, and VaR95 depending on α**

In the second example we use the correlation matrix C_β (fig. 9) with values of β between -1 and +1. Looking at the results in fig. 10 and 11, we can detect almost the same behaviour of the mean, the standard deviation and the VaR95 as in the first example. The only obvious difference is the fact that the mean remains constant. This seems to be a result of the sparsity of the matrix C_β. The main statement that follows from both examples is: The increasing correlations lead to an increase of risk!

$$
C_\beta = \begin{array}{cc}
\begin{array}{cccccccccccc} A & B & C & D & E & F & G & H & I & J & K & L \end{array} & \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & \beta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & \beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & \beta & 0 & 0 & 0 & 0 \\
0 & 0 & \beta & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \beta & 0 \\
\beta & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \beta & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & \beta & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \beta \\
0 & 0 & 0 & 0 & 0 & \beta & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \beta & 0 & 1
\end{pmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \\ F \\ G \\ H \\ I \\ J \\ K \\ L \end{array}
\end{array}
$$

Fig. 9: **The matric C_β**



source: authors' construction

Fig. 10: **The impact of positive and negative correlations – case β**

β	Mean	Std.Dev.	VaR95
- 1.0	26.15	1.026	28.02
- 0.9	26.14	1.095	28.11
- 0.8	26.15	1.160	28.20
- 0.7	26.15	1.225	28.25
- 0.6	26.15	1.288	28.34
- 0.5	26.15	1.349	28.42
- 0.4	26.15	1.409	28.48
- 0.3	26.15	1.469	28.57
- 0.2	26.15	1.528	28.67
- 0.1	26.15	1.586	28.75
0.0	26.15	1.636	28.84

β	Mean	Std.Dev.	VaR95
0.0	26.15	1.636	28.84
0.1	26.14	1.703	28.93
0.2	26.14	1.761	29.02
0.3	26.14	1.817	29.10
0.4	26.15	1.874	29.20
0.5	26.15	1.931	29.31
0.6	26.15	1.986	29.38
0.7	26.15	2.041	29.47
0.8	26.15	2.098	29.55
0.9	26.15	2.152	29.62
1.0	26.15	2.201	29.69

source: authors' construction

Fig. 11: **The means, standard deviations, and VaR95 depending on β**

Conclusions and Remarks

It could be shown, that it is possible to introduce correlations in Monte Carlo simulation. This offers the possibility to make a model more realistic. And, although it is quite difficult to estimate complex correlation structures in practice, this offers a way to get an impression of the possible impacts of these correlations (sensitivity of the model). In our case we had in the first example amplitude in risk (measured in VaR95) from 28.84 to 31.21. In the second example we found 28.02 and 29.69. Therefore this influence cannot be neglected.

It should also be mentioned here that this way of implementing correlations in the Monte Carlo approach can not only be used for durations, but also for other parameters like costs.

References

Brandimarte (2014): *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics*, John Wiley & Sons, Hoboken, New Jersey, 2014

Fazar (1959): *Program Evaluation and Review Technique*, The American Statistician, Vol. 13, No. 2, 1959

Huynh/Lai/Soumare (2008): *Stochastic Simulation and Applications in Finance with MATLAB Programs*, John Wiley & Sons, Hoboken, New Jersey, 2008

Iman/Conover (1982): *A Distribution-Free Approach to Inducing Rank Correlation among Input Variables*, Communication in Statistics – Simulation and Computation 11(3), 1982

Kelley/Walker (1959): *Critical Path Planning and Scheduling*, Proceedings of the Eastern Joint IRE-AIEE-ACM Computer Conference, December 1-3, 1959

Kerzner (2009): *Project Management*, John Wiley & Sons, Hoboken, New Jersey, 2009

Mai/Scherer (2014): *Financial Engineering with Copulas Explained*, Pelgrave Macmillan Basingstoke, 2014

PMI (2010): *A guide to the project management body of knowledge*, (PMBOK®Guide) PMI 2010

Schelle/Ottmann/Pfeiffer (2006): *Project Manager*, GPM German Association for Project Management, 2006

Taylor (2010): *Introduction to Management Science*, Pearson, Prentice Hall. 2010

Tysiak (2015a): *Statistical Analyses in the Critical Field,* in: Project Management Development – Practice and Perspectives – Proceedings of the Fourth International Scientific Conference on Project Management in the Baltic Countries, April 16-17, 2015, Riga, University of Latvia 2015

Tysiak (2015b): *A Deeper Insight in Some Effects in Project Risk Management*, International Journal of Computing, 14 (1), 2015

Van Slyke (1963): *Monte Carlo Methods and the PERT Problem*, Operations Research 11 (5), 1963