# Monte Carlo analysis used for Contingency estimating.

*Author's identification number:*

*Date of authorship: July 24, 2007*

## *TABLE OF CONTENTS:*

***LIST OF TABLES:***



***LIST OF FIGURES:***

***ABSTRACT:***

Risk management is a crucial component of effective project execution in general, and of project control in particular. If we do not identify risks, define the likelihood and impact of each and take action to mitigate the risks, our project plans and budgets might not be robust enough to withstand the inevitable unforeseen events. Effective and realistic plans and budgets are often based on the experience of project engineers, thus not explicitly defined and documented and therefore they can not be communicated and are not “controllable”.  This paper will show an explicit and documented risk analysis technique and a “live” model, based on the Crystal Ball[1)] software package that can be used in establishing contingency as part of a total project budget. It addresses the role of the project members and focuses on the Monte Carlo analysis to establish project contingency.

## ***INTRODUCTION***

*Why do we use Monte Carlo simulation?*
Current practice is that “single point estimates” are used along with a classification (class 2, class 3, etc). The understanding of the accuracy of +40% to -20%, +30% to -15%, +20% to – 10% or +10% to -5% (for resp. class 4, 3, 2 and 1 estimates) with a certainty of 60% (that implies a probability of over-/under-run of 20%) is often absent, resulting in confusion and misunderstanding when an estimate is actually “overrun”.
In present day estimating, Contingency is more and more considered to be a “**management decision**” instead of an estimating issue. In this context; Monte Carlo analysis can be used to better understand and explain the concepts of certainty and accuracy of estimates. With the visualized results it is simple to read at what “cost” the certainty can be raised, or in other words, what the extra contingency would be to reduce the probability of overrun from 20% to 10%.
Monte Carlo simulations also give an answer to the question regarding the “correct” contingency:

- Low contingency amount:
  - High probability of cost overrun
  - Driver for creative solutions
  - Probably concessions on quality/safety
- High contingency amount:
  - High probability of cost under run
  - Project is probably less feasible
  - Negative impact on cost efficiency
  - Cost under run may be used for other purposes
- Correct amount of contingency:
  - Probability of over- or under- run estimate is 50%

### Risk Management

The risk management processes described in this paper consists of two sub processes:

- Risk identification and impact evaluation
- Monte Carlo simulation

During the first sub-process all risks are identified by the project team in a Project Risk Assessment session. This session might result in different mitigation actions and covered by other controls. The risk(s) that should be covered by the estimated contingency the consequences for the estimate are quantified using risk factors with a specific probability of over- or under- run.
From previous projects estimate risk factors including low/high level range have been added to a risk check list that is used as a predefined list in the model.
The risk identification and impact evaluation is not a one time activity but should be (re-) done at appropriate moments during project execution (e.g. between each project phase).

The Monte Carlo simulation in the second sub-process uses the high and low levels, established during the previous step, per portion of the estimate. The resulting high risks can be evaluated in an efficient way by using Tornado or Sensibility charts. After analysis, a plan can be developed to increase the accuracy of scope and estimate.

Besides this, if conducted thoroughly, the process also gives early warnings and documents “weak” scope items (and thus estimate deficiencies). Improvements in scope and estimate can be made to increase the certainty.

## Introduction of the process and model

The risk identification and impact evaluation process is described in the first section of chapter **Risk analysis for contingency estimating**.
The model used for the Monte Carlo simulations is intended to capture the outcome of the risk identification and impact evaluation and it requires a certain breakdown of the estimate.
The cost breakdown should be done in such a way that the identified risks can be allocated to the estimate items. It might be necessary to recast the estimate before starting the Monte Carlo simulation.
As a typical example the estimated Equipment contains a large amount is for compressors with large commercial risks. In order to perform a risk analysis with emphasis on this aspect, the compressors should be extracted from the Equipment in a separate item.

The model has been developed based on the assumption that the user has basic statistical and adequate estimating knowledge.

<u>The historical data referred to, the list of risk factors that is used and the choices for type of distribution have been derived from a broad range of projects from an owner organization [2] and are not necessarily applicable for other organizations.</u>

## Distributions

As the choice of distribution is fundamental for Monte Carlo simulations, a short justification must be provided.
In a situation where historical data indicate that only one typical distribution is required, this distribution should preferably be used for the Monte Carlo simulations. If this information is not available, then a predefined distribution can be chosen. The available distributions in Crystal Ball are shown in Figure 1: Distribution gallery

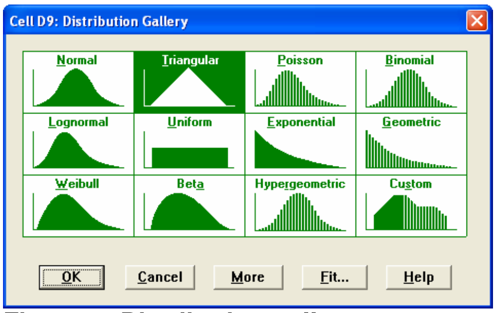


**Figure 1: Distribution gallery**

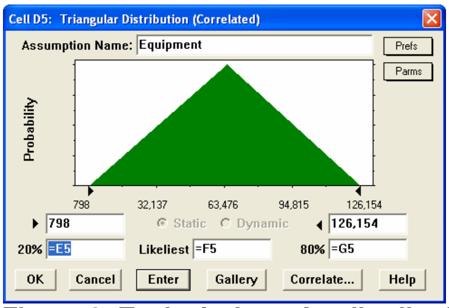


**Figure 2: Typical triangular distribution (Symmetrical)**

The following criteria should be taken into account when selecting a distribution:

- It shall use a <u>minimal</u> number of parameters
- It shall be able to represent <u>symmetrical</u> and <u>asymmetrical</u> distributions
- It shall allow mathematical analysis

The last criterion is less important because of modern ICT developments.

The pros/cons of the available distributions are listed bellow [3]:

- Normal: is always symmetrical and does not fit with expected cost deviations
- Log normal: although asymmetrical mean and standard deviation are needed to define boundaries

- Triangle (see Figure 2): is best when lack of data, extremely flexible, but less precise than normal distribution and therefore fits best with criteria.

**Estimate**

In this paper the following example estimate (Table 1) will be used. The highlighted values are used in the Monte Carlo simulation:

**Table 1: Example estimate**

| Estimate in EUR x 1.000 | | | | | | |
|---|---|---|---|---|---|---|
| | Core plant | Revamp existing plant | Feedstock storage and supply | Inter-connecting (utilities) | Product storage | Total |
| Equipment | 29.500 | 12.400 | 17.100 | 1.400 | 4.500 | 64.900 |
| Materials & Construction | | | | | | |
| Equipment erection | 1.700 | 1.000 | | 0.300 | 0.600 | 3.600 |
| Piping | 35.000 | 17.100 | 1.800 | 1.200 | 5.300 | 60.400 |
| Civil | 14.900 | 3.200 | 1.000 | 0.500 | 1.600 | 21.200 |
| Process Control | 8.500 | 0.900 | 1.000 | 0.100 | 0.300 | 10.800 |
| Electrical | 3.100 | 0.300 | 0.300 | 0.100 | 0.800 | 4.600 |
| Engineering | 27.300 | 15.100 | 3.800 | 1.400 | 6.900 | 54.500 |
| SUBTOTAL | 90.500 | 37.600 | 7.900 | 3.600 | 15.500 | 155.100 |
| Owner cost | 14.000 | 10.000 | 3.000 | 1.000 | 2.000 | 30.000 |
| GRAND TOTAL | 134.000 | 60.000 | 28.000 | 6.000 | 22.000 | 250.000 |

## ***RISK ANALYSES FOR CONTINGENCY ESTIMATING***

As mentioned before, the risk management process consists of two separate sub processes. In the first step, the risk identification and impact evaluation, the input data for the Monte Carlo simulation are defined. A project specific set of data is defined and in such cases it is recommended to involve the project team. Besides that, the risk identification and impact evaluation delivers an overview of the risk factors that have the highest impact on the total estimate.

Step 1: the risk identification and impact evaluation consists of the following main elements:

1. Determine the level of the analysis
   The level (the project sections, scope items etc) on which the risk analysis should be performed has to be determined. The project sections and scope items are grouped into 10-25 elements which have equal distribution values ("low" and "high" level) and to which the same risk factors apply. There is no added value in running a Monte Carlo simulation on a dataset with less items and a wide spread of low and high levels or on a dataset with much more items and a small spread or equal low and high levels.

2. Determine the value per section/item
   This value should consist of:
   - The estimated bare value
     This value is built up as follows:
     - Determination of bare value per sub section of detailed item (f.i equipment item and bulk material (meters of specified pipe)
     - The determination of man hours involved per detailed item should be based on a calculation program and should be auditable.
     - Allowances to complete the documented scope such as:
       - Cut and scrap allowance (e.g. for piping)
       - MTO allowance.
         This allowance covers the incompleteness of the MTO (e.g. meters pipe, numbers of elbows, tons of steel, etc) due to the non-existence of the detailed drawings (e.g. iso's, meters of cable, m3 of concrete, etc) because of the early project phase and a lack of the detailed/accurate measuring in this stage of the project. In general: the MTO reflecting items of which the quantities have been counted on/from the PID's is less than after detail design.
   - The escalation per item, based on the schedule.
3. Determine the distribution of the values per factor
   Only items which are relevant for the situation will have to be determined. It might be that only a limited number of items appear to be relevant. For practical reasons a triangular shape will be sufficient.
   The project-team should be involved when determining the values.
   If better information becomes available another type of distribution might be used. This better information can also be retrieved from literature or based on in house statistics.
   The triangular distribution:
   For this paper and if no better information is available the triangular distribution with the 20 and the 80 % boundaries can/will be used.
   This distribution is used for all values. This "topping off" by means of the 20 and 80% values is done to eliminate the effect of the outliers which **might** occur. This results in a (triangular) distribution per item within acceptable ranges.
   The most likely value represents the mean of the distribution for symmetrical distribution.
   The 20 % boundary reflects the probability of under-running the most likely value
4. Discuss and agree on the factors (which are applicable and in what range?) with the project team.
5. Determine per section/item the applicable factor.
   It might be that for a specific section/scope item several factors apply. Those factors will be multiplied with each other.

In the second step this distribution of risk factors is used in a Monte Carlo simulation to calculate the contingency <u>amount</u> to be added to the sections of the bare estimate, which add up to the total contingency.

<u>Step 2: the risk Monte Carlo simulation consists of following steps</u>:

1. Recast estimate (prepare the correct cost breakdown)
2. Define risk factors (choose from a predefined set or add new factors)
3. Connect risk factors to estimate items
4. Define distribution of risk factors (adapt the model to the project specific data)
5. Run Monte Carlo simulation and analyze the output
6. Check the outcome with the definition of the estimate classes

**Recast the estimate**

The first step is the breakdown of the estimate in a relevant number of parts. The example estimate (Table 1, see highlighted values) is already formatted in this way. The maximum number of parts in this model is limited to 14. From a practical point of view this number is sufficient but it can easily be extended if necessary.
The separation of parts can be in sections, equipment type, discipline, labor/material, etc.
It is important that this separation supports the connection to risk factors in step 3. Parts of the estimate may correlate with each other and by connecting them to risk factors this correlation can be entered in the model as well.

**Table 2: Recast estimate values**

| | | |
|---|---|---|
| | **Project:** | Testcase |
| | **Projectnumber:** | B.123456 |
| | **Monte Carlo Risk Analysis performed by:** | |
| **No** | **Section** | **Most Likely [kEUR]** |
| 1 | Equipment Core Plant | 29,500 |
| 2 | Equipment Revamp existing plant | 12,400 |
| 3 | Equipment Feedstock storage and supply | 17,100 |
| 4 | Equipment Interconnecting (utilities) | 1,400 |
| 5 | Equipment Product storage | 4,500 |
| 6 | Core Plant | 90,500 |
| 7 | Revamp existing plant | 37,600 |
| 8 | Feedstock storage and supply | 7,900 |
| 9 | Interconnecting (utilities) | 3,600 |
| 10 | Product storage | 15,500 |
| 11 | Client costs | 30,000 |
| 12 | | |
| 13 | | |
| 14 | | |
| | Total | 250,000 |

The recast estimate parts are entered in columns **Section** and **Most Likely** (Table 2). Note: *This Table is showing only the relevant part of the total spreadsheet.*

**Define risk factors**

The definition of risks is facilitated by a predefined list of some 30 factors (Table 3) in the model. Keep in mind that these predefined risk factors are derived from historical data from an owner organization and that they are not necessarily applicable for other organizations.

**Table 3: List of risk factors, incl. distribution parameters**

| Risk factor ID | List of factors | Crystal Ball | Most Likely | 20% probability of underrun | 20% probability of overrun | # in risk analysis |
|---|---|---|---|---|---|---|
| 1 | Scope completeness | 1.05 | 1.05 | 100% | 110% | 11 |
| 2 | Plot plan clearance study | 1.00 | 1.00 | 99% | 101% | - |
| 3 | General lay-out | 1.00 | 1.00 | 99% | 101% | - |
| 4 | Non proven items in scope | 1.00 | 1.00 | 99% | 101% | - |
| 5 | Dependance on contractors (eg single source) | 1.00 | 1.00 | 99% | 101% | - |
| 6 | Worlwide procurement | 1.00 | 1.00 | 99% | 101% | - |
| 7 | Preferred vendors | 1.00 | 1.00 | 99% | 101% | - |
| 8 | Large equipment vendor selection | 1.00 | 1.00 | 97% | 103% | 3 |
| 9 | Contracting strategy | 1.00 | 1.00 | 99% | 101% | - |
| 10 | Availability of contractors during building period | 1.00 | 1.00 | 90% | 110% | 5 |
| 11 | Market conditions | 1.00 | 1.00 | 99% | 101% | - |
| 12 | Scale up factor | 1.00 | 1.00 | 99% | 101% | - |
| 13 | Revamp aspects | 1.00 | 1.00 | 90% | 110% | 3 |
| 14 | Re-use existing facilities | 1.00 | 1.00 | 99% | 101% | - |
| 15 | Location factor | 1.00 | 1.00 | 99% | 101% | - |
| 16 | Demolishing | 1.00 | 1.00 | 99% | 101% | - |
| 17 | Construction efficiency | 1.03 | 1.03 | 90% | 115% | 3 |
| 18 | Soil data | 1.00 | 1.00 | 99% | 101% | - |
| 19 | Duration construction | 1.00 | 1.00 | 99% | 101% | - |
| 20 | Construction rate | 1.03 | 1.03 | 95% | 110% | 3 |
| 21 | Permit requirements | 1.00 | 1.00 | 99% | 101% | - |
| 22 | Utility balance | 1.00 | 1.00 | 99% | 101% | - |
| 23 | Availability of utilities | 1.00 | 1.00 | 99% | 101% | - |
| 24 | Safety concept | 1.00 | 1.00 | 99% | 101% | - |
| 25 | Environmental aspects | 1.00 | 1.00 | 99% | 101% | - |
| 26 | Imports, exports, duties | 1.00 | 1.00 | 99% | 101% | - |
| 27 | Owner requirements | 1.00 | 1.00 | 99% | 101% | - |
| 28 | Section design | 1.00 | 1.00 | 99% | 101% | - |
| 29 | Product range | 1.00 | 1.00 | 99% | 101% | - |
| 30 | Effluent treatment | 1.00 | 1.00 | 99% | 101% | - |
| 31 | Confidence in estimating data (eg quantities) | 1.00 | 1.00 | 99% | 101% | - |
| 32 | Feedstock discussion | 1.00 | 1.00 | 99% | 101% | - |
| 33 | Team experience | 1.00 | 1.00 | 99% | 101% | - |
| 34 | | 1.00 | 1.00 | 99% | 101% | - |
| 35 | | 1.00 | 1.00 | 99% | 101% | - |
| 36 | | 1.00 | 1.00 | 99% | 101% | - |

The Low and high level of the individual risk factors are entered in columns **20% probability of under run** and **20% probability of overrun.** The value for **Most Likely** is calculated in the spreadsheet and the distribution values, per risk factor in the cells in column **Crystal Ball**, are taken over. Note: *The data in this Table are for a Test case only.*

The risk identification and impact evaluation (step 1) may result in additional factors that have not yet been added to the standard risk list and above shown predefined list and tune the Risk analyses for the actual project.

As a learning organization it is important to have these new risk factors documented and added to the standard model. In this case the maximum number of factors in this model is 36, including the “standard” factors.

Risk factors should be independent. If this is not the case then mutually depending risk factors should be grouped in the model.

### Connect Risk factors to estimate items

Applicable risk factors are entered per estimate item. (See mark up in Table 4)

One estimate item can have multiple risk factors and risk factors can apply to multiple estimate items. At the right side of Table 3, in column **# in risk analysis**, it is shown how many times a risk factor is allocated. The marked risk factor ID 1 is 11 times allocated in the spreadsheet; indicated in Column 1.
There is no need to restrict the risk allocations in the columns to one single risk factor ID and for efficient review and communication the allocations of different risk factors can be grouped in one column.

**Table 4: Risk factors connected to estimate items (only 5 sets of risk factor columns shown!)**

| | | | Column 1 | | Column 2 | | Column 3 | | Column 4 | | Column 5 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Project:** | Testcase | | | | | | | | | | | | | |
| | **Projectnumber:** | B.123456 | | | | | | | | | | | | | |
| | **Monte Carlo Risk Analysis performed by:** | | | | | | | | | | | | | | |
| **No** | **Section** | **Most Likely [kEUR]** | Risk factor ID | Factor value | Risk factor ID | Factor value | Risk factor ID | Factor value | Risk factor ID | Factor value | Risk factor ID | Factor value | Total including contingency [kEUR] | Contingency, based on 50 % probability of underrun | Contingency factor |
| 1 | Equipment Core Plant | 29,500 | 1 | 1.05 | 8 | 1.00 | | 1.00 | | 1.00 | | 1.00 | 30,975 | 1,475 | 1.05 |
| 2 | Equipment Revamp existing plant | 12,400 | 1 | 1.05 | 8 | 1.00 | | 1.00 | | 1.00 | | 1.00 | 13,020 | 620 | 1.05 |
| 3 | Equipment Feedstock storage and supply | 17,100 | 1 | 1.05 | 8 | 1.00 | | 1.00 | | 1.00 | | 1.00 | 17,955 | 855 | 1.05 |
| 4 | Equipment Interconnecting (utilities) | 1,400 | 1 | 1.05 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | 1,470 | 70 | 1.05 |
| 5 | Equipment Product storage | 4,500 | 1 | 1.05 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | 4,725 | 225 | 1.05 |
| 6 | Core Plant | 90,500 | 1 | 1.05 | 10 | 1.00 | | 1.00 | 17 | 1.03 | 20 | 1.03 | 100,812 | 10,312 | 1.11 |
| 7 | Revamp existing plant | 37,600 | 1 | 1.05 | 10 | 1.00 | 13 | 1.00 | 17 | 1.03 | 20 | 1.03 | 41,884 | 4,284 | 1.11 |
| 8 | Feedstock storage and supply | 7,900 | 1 | 1.05 | 10 | 1.00 | | 1.00 | 17 | 1.03 | 20 | 1.03 | 8,800 | 900 | 1.11 |
| 9 | Interconnecting (utilities) | 3,600 | 1 | 1.05 | 10 | 1.00 | 13 | 1.00 | | 1.00 | | 1.00 | 3,780 | 180 | 1.05 |
| 10 | Product storage | 15,500 | 1 | 1.05 | 10 | 1.00 | 13 | 1.00 | | 1.00 | | 1.00 | 16,275 | 775 | 1.05 |
| 11 | Client costs | 30,000 | 1 | 1.05 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | 31,500 | 1,500 | 1.05 |
| 12 | | | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | - | - | 1.00 |
| 13 | | | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | - | - | 1.00 |
| 14 | | | | 1.00 | | 1.00 | | 1.00 | | 1.00 | | 1.00 | - | - | 1.00 |
| | Total | 250,000 | | | | | | | | | | | 271,197 | 21,197 | 108% |

## Define distribution of risk factors

The values for **Most likely**, **20% probability of under run** ("Low level"), **20% probability of overrun** ("High level") and **Crystal Ball** (see Table 3) are entered in the spreadsheet.
In case historical data are available, these parameters for distribution can be derived from that data. This is the preferred option but often these historical data are not available and therefore it is not possible to calculate the values. In such cases, judgment is required, for example:

- Most likely: <1, =1 or >1
- Low level: 0.8, 0.9 or 0.95
- High level: 1.05, 1.10 or 1.2

These borders can be adjusted if better information is available.

In the columns right of Table 4 the result of the risk allocation to the estimate parts is shown. The column Total including Contingency is calculated as: Most likely x Factor value (Column1) x Factor value (Column2) …x Factor value (Column n).

## Run Monte Carlo simulation and analyze the output

During the Monte Carlo simulation the "bare" estimate items are multiplied by the relevant risk factors. The sum of the "bare" estimate items and the contingency is the total estimate that is shown in the graphic (Figure 3). The total distribution is defined by the distribution of the underlying risk factors.

For this example the Monte Carlo run is done for 1.000 trials.

After each simulation run the next graphic is generated. In order to align the certainty interval equal to the CE handbook definition the Certainty value has to be changed to 60%

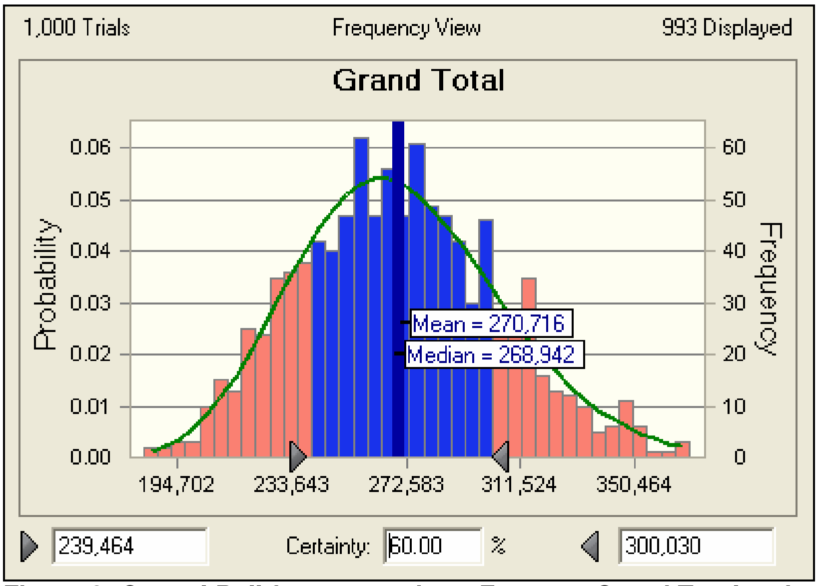


**Figure 3: Crystal Ball frequency chart, Forecast Grand Total estimate**

The data from the simulation run can be presented as cumulative distribution, statistical data and percentiles (Table 5)

**Table 5: Crystal Ball percentiles forecast Grand Total estimate**

Forecast: Grand Total

Edit View Forecast Preferences Help

1,000 Trials Percentiles View 993 Displayed

| Percentile | Forecast values |
|---|---|
| 0% | 183,405 |
| 10% | 227,110 |
| 20% | 239,464 |
| 30% | 251,518 |
| 40% | 259,717 |
| 50% | 268,942 |
| 60% | 278,017 |
| 70% | 287,616 |
| 80% | 300,030 |
| 90% | 315,098 |
| 100% | 395,186 |

From this table the total estimate amount including contingency can be seen. Normally the basis for an estimate is a 50% probability of over-/ or under-running. The total amount for this certainty is EUR 269 million and the contingency amount is EUR 19 million (8%). To improve the estimate to 80% certainty it needs EUR 50 million contingency (20%). The total estimate amount according the table is EUR 300 million ±.

### Check the outcome with the definition of the estimate classes

Additional to the analysis of the Grand Total, in Frequency and Percentile view the outcome of the simulation has to be compared with the definition of the estimate classes. This check then determines the classification of the estimate. For the example the estimate would be Class 2, + 20% to - 10%

### Tornado

As the estimate is made up of different parts and many risk factors, it is difficult to have, or keep an overview on which risk factors do impact the total estimate the most.
With a Tornado diagram the variables in the model can be quickly pre-screened and selected for further analysis or more accurate risk assessment. In the Tornado diagram the risks that have the biggest impact on the total are presented, sorted by value and include minimal and maximal impact.

The following Tornado diagram has been generated for previous example (Figure 4):

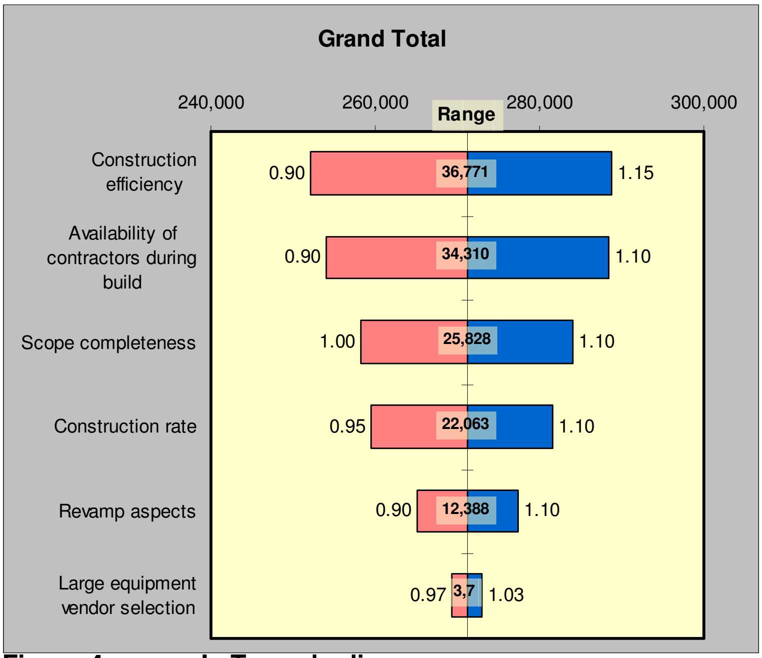


**Figure 4: example Tornado diagram**

Here you can see following items:

- Y axis: the 6 risk factors and the impact on the total estimate, sorted by value
- X-axis: the impact on the total estimate in chosen currency (EUR x 1.000)
- The values in the bars show the forecast value range across the variable tested within the testing range of 20 to 80%
- The values next to the bars show the Low level and High level (input) for the risk factors in %.

As this Tornado chart is the result after “internal” risk assessment reviews and impact evaluations the next step to improve the contingency could be the involvement of construction contractors. In this way the probability and impact for Construction efficiency, Contractor availability and the Construction rates could be more accurately be assessed and the “correct” project contingency could be estimated.

## ***CONCLUSION***

We have seen that present-day decision making for capital projects requires a better understanding of the risk exposure and the need for a documented risk assessment and control.

A separate sub-process to identify risks and to evaluate the risk impact is used to define the variables for the Monte Carlo simulation.

A “live” model for Monte Carlo simulations is described including basic considerations for recasting the estimate and how to build the simulation model.

In this paper a Monte Carlo simulation for an example capital project is presented and the output of that simulation has been analyzed. As a result the contingency amount, given a required probability, was estimated and decision makers are presented with a better understandable and visualized probable cost figure.

***References:***
1) Decisioneering, Cristal Ball Risk Analysis Software & Solutions
2) Owners estimating department: Monte Carlo risk analysis (April 12, 2002 and December 2003)

***Acknowledgements:***
3) Thiriez, H: Crystal Ball training February 8, 2001