

Frode Drevland

# Cost estimation under uncertainty



CONCEPT THEME PAMPHLET **NO.4**

### **About the author**

Dr. Frode Drevland is an associate professor at the Department of Civil and Environmental Engineering at the Norwegian University of Science and Technology (NTNU). He has worked extensively on the development of methods and tools for uncertainty analysis and uncertainty management related to capital costs.

This booklet is an English translation of Concept Theme Pamphlet No. 4, *Kostnadsestimering under usikkerhet*, originally published in 2013 (ISBN 978-82-93253-45-7) and primarily based on the author's work published in Concept Reports No. 10 and No. 13. In this booklet, the emphasis is on presenting, in a simple, practical manner, the theory and fundamental principles that the author has observed many practitioners struggle to understand, which often lead to errors in the estimates produced.

### **Frode Drevland**

#### ***Cost Estimation under Uncertainty***

Copyright © 2025 the author. All rights to the content of this publication belong to the author. This booklet is published with support from the Concept Research Programme.

### **Address**

Concept Research Programme  
Department of Civil and Environmental Engineering  
NTNU  
7491 Trondheim  
Norway

Published by Ex Ante Akademisk Forlag, Trondheim, December 2025.

ISSN (print): 1891-5620

ISSN (online): 1891-5655

ISBN (print): 978-82-8433-060-0

ISBN (online): 978-82-8433-061-7

Information about the Concept Research Programme: [www.ntnu.no/concept](http://www.ntnu.no/concept)



Norwegian University of  
Science and Technology

# 1 Introduction

Uncertainty is often defined as a lack of knowledge. Anyone who has been involved in the planning of investment projects knows that there is much one would like to know, but which is impossible to obtain in advance. This uncertainty, inherent in every project, makes it impossible to estimate precisely in advance how much a project will cost upon completion. Nevertheless, this is precisely what traditional deterministic estimation seeks to do. Each line item in the cost estimate is assigned a specific monetary value, without accounting for the uncertainty associated with it, and these are summed to produce an exact figure for the total project cost.

The alternative is to perform cost estimation under uncertainty, acknowledging that the world itself is uncertain. Here, the input values in the estimate are treated as uncertain. Such a cost estimate is called a *stochastic* (probability-based) cost estimate. The advantage of stochastic cost estimation is that it typically provides a significantly more realistic representation of costs and their associated uncertainty than traditional deterministic estimates. The disadvantage is that it places greater demands on the structure of the estimate, the execution of the estimation process, and on those who perform the work.

This booklet provides an introduction to cost estimation under uncertainty. The methodology presented here is equally well suited for use in large-scale public investment projects worth billions as for home bathroom renovations.

## Terminology note

In English-language practice, the term *risk analysis* is often used as an umbrella term for probability-based analyses of project cost outcomes; this is also the term used in official English translations of the Norwegian State Project Model (e.g. Circular R-108/25). In the Norwegian context, the corresponding concept is commonly referred to as *uncertainty analysis* (*usikkerhetsanalyse*), which is also the terminology used in the original Norwegian version of this booklet. For consistency with the source text, this translation therefore uses *uncertainty* and *uncertainty analysis* throughout.

## 2 Probability distributions – the way we express uncertainty

For most people, the most natural way to express uncertainty is to state a range. One says that a quantity is greater than  $X$  but less than  $Y$ . This works well in everyday life when discussing how much a carton of milk costs or similar quantities. However, for cost estimates that will serve as the basis for investments in the millions and billions, this is not sufficient. One can say that the cost of a road or a building will lie between  $X$  and  $Y$ , but it is clear to most that it is not equally probable that the cost will end up at  $Y$  as at some value between  $X$  and  $Y$ . To express both the range and the probabilities of uncertain values, it is most appropriate to use a probability distribution.

A probability distribution specifies the probabilities of different outcomes. Figure 1 shows the probability distribution for an ordinary six-sided die. There is a  $1/6$  chance for each outcome.

This distribution is an example of a discrete distribution. We can obtain 1 or 2 as the result of a die throw, but we cannot, for example, obtain 1.4. For use in cost estimation, it is more natural to employ continuous probability distributions; the most common are bell-shaped distributions such as the Normal, Gamma, and PERT.

Because probability distributions are central to cost estimation under uncertainty, one must be familiar with their terminology and properties.

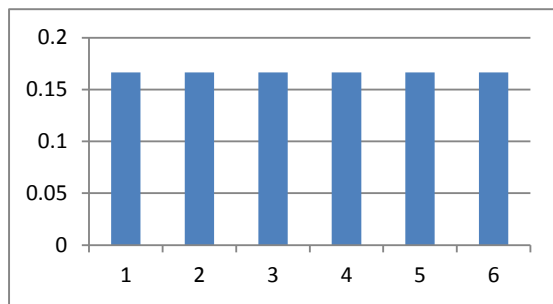


Figure 1: Probability distribution of a standard die

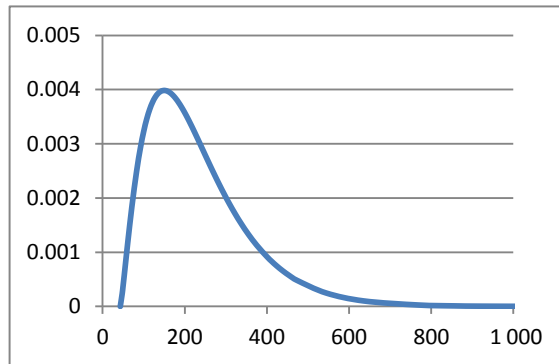


Figure 2: Example of a continuous probability distribution

### Percentiles and P-values

When using continuous probability distributions, it does not make sense to talk about the probability of a specific outcome. Instead, we can talk about the likelihood that the investment cost will be equal to or lower than a given number – in other words, the probability that a budget at this amount will hold.

An  $n$ -percentile is the value with an  $n$  per cent probability of not being exceeded. The 85th percentile, for example, indicates the value that, with 85% probability, will not be exceeded. In practice, percentiles are denoted as P85, P50, and so on.

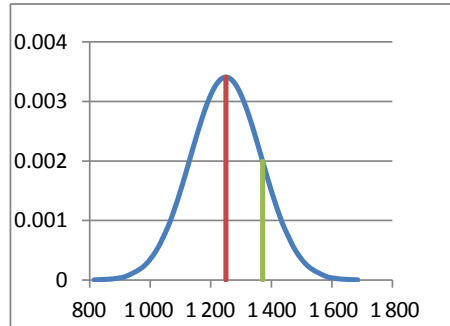


Figure 3: P50 (red) and P85 (green) percentiles indicated on a probability distribution (blue)

### The S-curve - the cumulative probability distribution

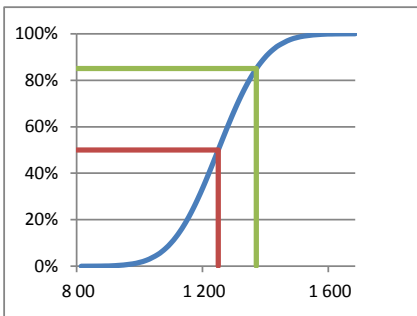


Figure 4: The cumulative probability distribution of a bell-shaped distribution

A challenge with continuous probability distributions is that they are not directly readable. If we look at Figure 3, the number on the  $y$ -axis has no direct meaning. If one wishes to read off a given P-value, one must take the integral of the distribution function to compute the area under the curve. This is not practical to do in most contexts, and one therefore usually prefers to use cumulative probability distributions to present estimation results.

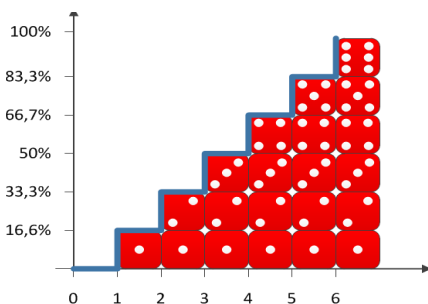


Figure 5: The cumulative probability distribution of a die

For bell-shaped probability distributions, the cumulative distribution curve takes on a characteristic S-shape, and it is therefore common to refer to this as an S-curve. Figure 4 shows how, for such a curve, it is possible to read off the P-values directly. Many struggle to understand the relationship between the ordinary probability distribution and the cumulative one. This is easier to understand when looking at a discrete distribution. Figure 5 shows the cumulative version of the die's distribution that was demonstrated in figure 1.

### Most likely value - mode

The *most likely value* (mode) is the peak of a probability distribution. This is the single value that is most likely.

### Median

The median is the point in a probability distribution at which half of the area under the curve lies to the left and half to the right. In other words, the median is identical to the 50th percentile, i.e., P50.

### Expected value - mean

The expected value is the centre of gravity of a probability distribution. It is the sum of all conceivable outcomes, each weighted by its respective probability. In statistics, the corresponding concept is the arithmetic mean – in other words, what most people associate with the term *average*. In probability theory, the expected value is denoted by  $E$  (Expected value).

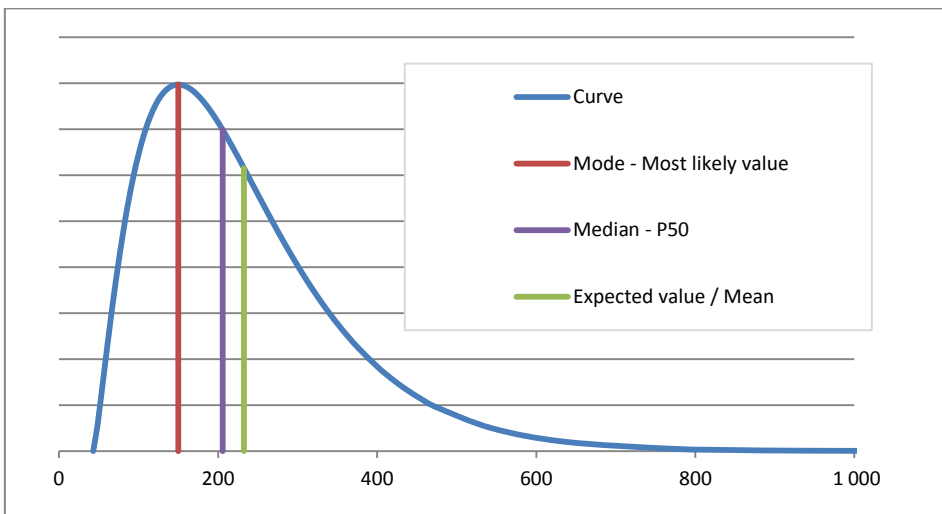


Figure 6: Most likely value, expected value, and median indicated on a right-skewed distribution

### The difference between median, most likely value, and expected value

For many, it can be challenging to distinguish among the median, the mode, and the expected value. For a symmetric probability distribution, these three values coincide; for skewed distributions, they differ. Investment costs typically follow a right-skewed distribution; therefore, it is crucial to understand these concepts clearly.

To illustrate the difference, we will use an example. Figure 6 shows how the three value concepts typically fall in a right-skewed distribution. However, it is easier to distinguish between them by examining a discrete probability distribution rather

than a continuous one. We will therefore examine the probability distribution of height for a group of men, as shown in Figure 7.

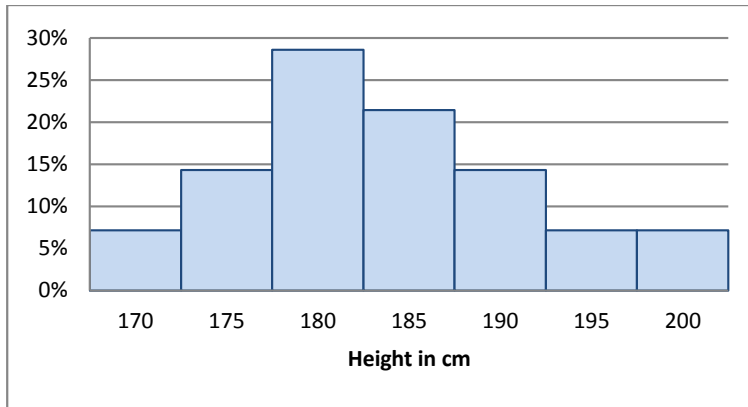


Figure 7: The probability distribution for the height of a group of men

If we order all men represented by this probability distribution from shortest to tallest, the median, most likely value, and expected value are shown in figure 8.

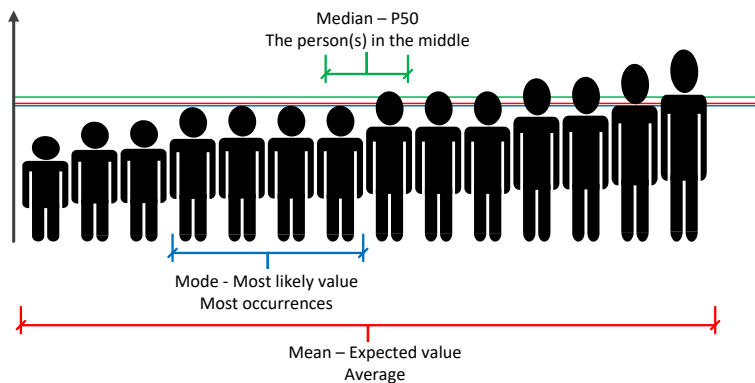


Figure 8: The difference between expected value, median, and most likely value

### Standard deviation

The standard deviation is the most common measure of the spread of a probability distribution. The standard deviation has the same units as the expected value (e.g., kroner or metres) and is typically of the same order of magnitude as the expected value.

The standard deviation is denoted by the Greek letter sigma and is mathematically defined as the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

Figure 9 shows the spread of values around the expected value in a Normal distribution. The Normal distribution is one of the most important distributions in statistics and probability theory because natural phenomena often follow a normal distribution. This applies, for example, to examination grades and life expectancy. The results of stochastic estimates will also typically be close to normally distributed (see subsection 6.2 for a more detailed explanation).

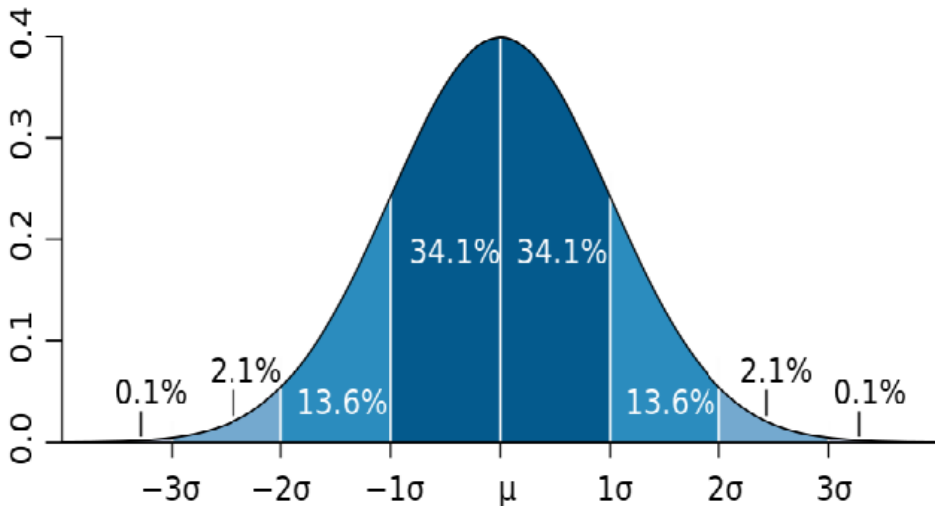


Figure 9: The spread of values in a Normal distribution around the expected value

### Variance

We usually use the standard deviation to quantify uncertainty, but it is defined in terms of the variance. The variance is the squared deviation from the expected value and is denoted  $\sigma^2$ . It is a quantity that is difficult to interpret in discussions of uncertainty, and, aside from specific calculations, there is no need to use it.

### Relative standard deviation

The relative standard deviation indicates how large the standard deviation is in relation to the expected value. It is given in per cent:

$$\sigma_{\text{rel}} = \frac{\sigma}{E}$$

This makes sense when we are speaking of pure investment costs, but note that if the expected value is close to zero – for example, if the income and expenditure sides of a project balance—then this quantity is not meaningful. In such cases, it may be more relevant to calculate the uncertainty for the income and expenditure sides separately.

### Three-point estimate

A three-point estimate consists, as the name implies, of three estimates. All probability distributions that are commonly used in uncertainty analyses can be uniquely defined by means of three arbitrary points on the curve. One typically assesses a lower value, the most likely value (i.e., the peak of the curve), and an upper value. The lower and upper values are always symmetric percentiles on the distribution.

One usually uses either P10 and P90 or P01 and P99. In practice, P10 and P90 have proven to be better suited because most people cannot imagine how well or how poorly outcomes can actually vary in extreme cases. However, this depends both on the domain in which one operates, the level of detail used in the estimate, and other factors.

### Skewness

Probability distributions that are asymmetric are said to be skewed. A distribution where the peak lies to the left and has a longer tail to the right is called right skewed, and vice versa. As noted, investment costs typically follow a right skewed distribution.

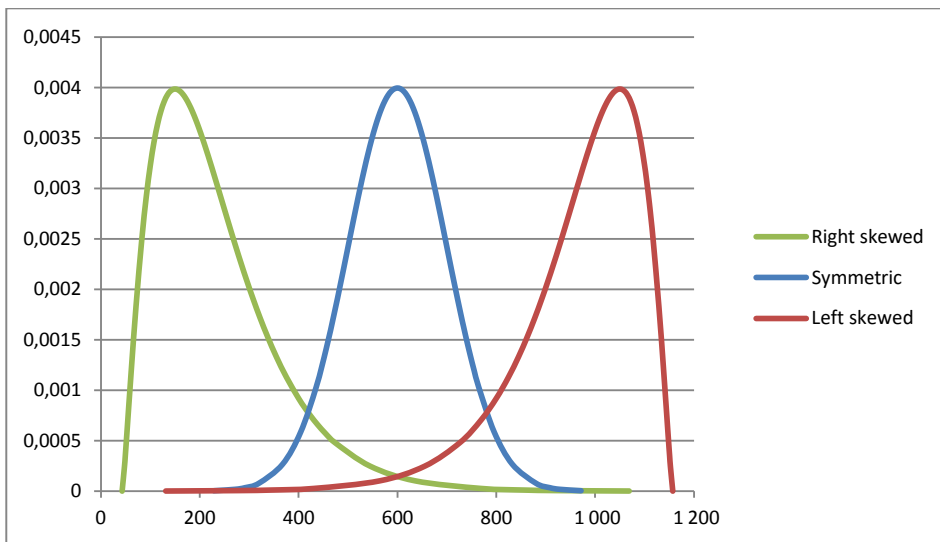


Figure 10: Probability distributions can be right skewed, left skewed, or symmetric

## 3 Cost estimate modelling

Building and structuring a cost estimate in stochastic cost estimation can be challenging. A stochastic estimate cannot be constructed in the same arbitrary manner as a deterministic cost estimate. This is mainly due to the principle of stochastic independence, which must be maintained for the result to be valid.

### 3.1 Stochastic independence

In a stochastic cost estimate, cost items are, by default, assumed to be stochastically independent. Stochastic independence means that two uncertain quantities or events do not influence each other in any way. A simple example is rolling dice. If you roll two dice in succession, the result of the second roll is completely independent of the first.

Assume we have an estimate with two cost items, A and B, each estimated to cost between 10 and 20 million NOK. If we suddenly receive information that makes us entirely certain that A will cost 20 million NOK, then we can only say that the cost items are stochastically independent if our expectation for what cost item B will cost does *not* change. If it *does* change, they are stochastically dependent and have what we call *covariation (correlation)* – the cost items vary in step.

This covariation must be accounted for in the estimate. This concerns the use of an accurate cost model; however, the phenomenon is often referred to as ‘calculating away’ uncertainty, as it is in the estimate that modelling errors become apparent.

The primary way to address covariation is to ensure it is absent in the first place. When building the cost breakdown structure, you should, as far as possible, ensure that the items are stochastically independent. It is often tempting to break down the estimate by contract structure or similar criteria, but this can be counterproductive if there is strong covariation. It is better to end up with a correct cost estimate result that later needs some work to be reorganised for project follow-up, than to end up with a structure that is perfect for follow-up but gives a misleading picture of the project’s uncertainty.

In addition to being deliberate about how you break down the estimate, we have two other methods for handling covariation. One is to extract the common elements across items into separate *uncertainty factors*. The other is to model covariation within the estimate explicitly. Both these methods are described later in this chapter.

#### **Top-down approach**

Traditional cost estimation typically employs a bottom-up approach. Here, the project is broken down to a detailed level. One starts at the bottom – at the nuts-and-bolts level – and sums upwards from there. The weakness of this approach is

the considerable risk of omissions, i.e. we fail to include everything the estimate is supposed to cover.

With a top-down approach, by contrast, one starts at the top and details down to a suitable level. If we estimate a single value for the entire project, then, by definition, nothing is omitted. The more we detail downwards, the greater the chance that things fall between two stools and are left out. Figure 11 illustrates the traditional bottom-up and top-down approaches.

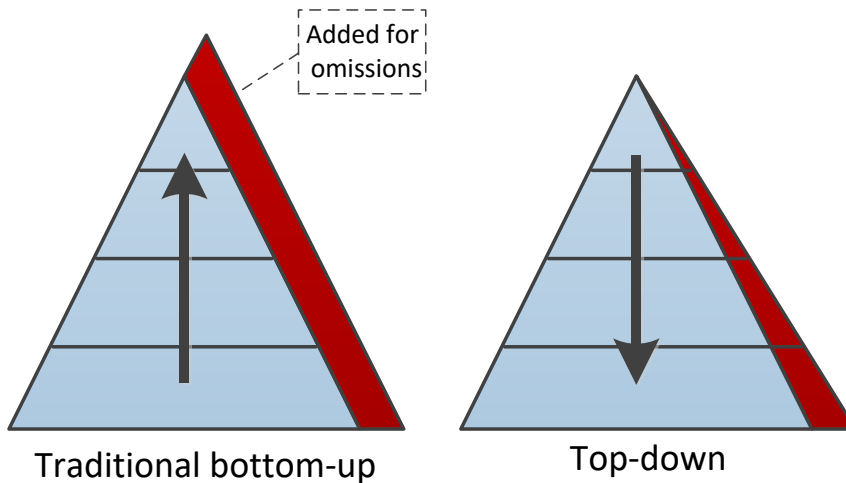


Figure 11: Traditional bottom-up approach versus a top-down approach.

In connection with stochastic cost estimates, there is also a challenge in controlling covariation in the estimates when using a very detailed breakdown. There will be dependencies that crisscross, which cannot be sensibly modelled in the estimate.

One approach is to first prepare a traditional estimate and then use an aggregated level of this as input to a stochastic cost estimate. This addresses the covariation-handling challenge, but it remains an unsatisfactory approach. The same degree of omission persists, and one often fails to work on the basis of a common set of assumptions throughout the entire estimating process.

### 3.2 Structure of the estimate

A stochastic cost estimate can, in theory, be constructed in various ways, especially if Monte Carlo simulation is used as the calculation method (see 6.3). The following describes certain principles for structuring and “building blocks” based on best practice. It is recommended to adhere to these unless one has a very good understanding of the underlying theory and firm control of what one is doing.

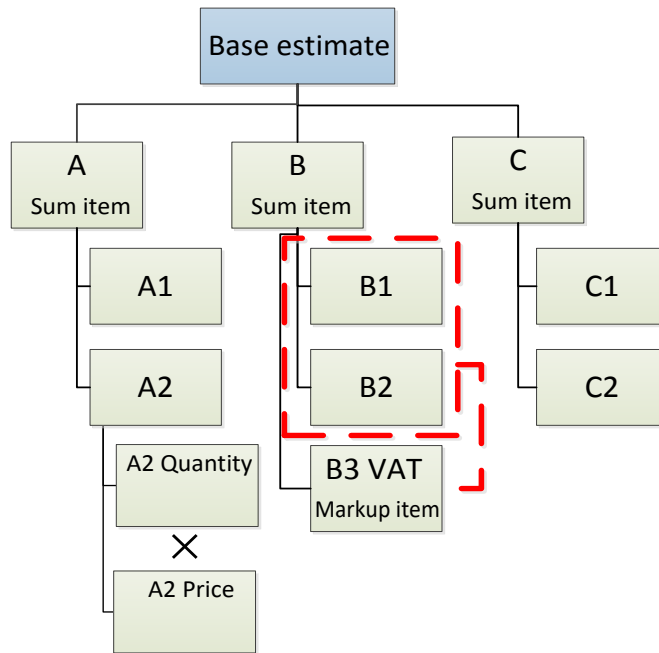


Figure 12: Example of breakdown of a base estimate into different item types.

### Base estimate

The base estimate is a hierarchical breakdown of what is to be delivered by the project, and one should here seek to choose a subdivision that, as far as possible, provides stochastic independence between the cost items. This is built up using sum items, ordinary cost items, and mark-up items.

### Sum items

All cost items that are not at the lowest level in the estimate are called sum items. They sum the cost items beneath them in the hierarchy.

### Ordinary cost items

Typically, the vast majority of cost items in an estimate are ordinary cost items. That is, cost items that are either a lump sum or a product of quantity and unit price.

### Mark-up items

Sometimes one wishes to include items calculated from other items. Examples include a markup for site setup or value-added tax (VAT). These items are calculated based on a three-point estimate of the percentage markup applied to the sum of the items on which they are based. Markup items are often at the same level in the estimate as the items from which they are calculated, but they may also be calculated from items elsewhere in the estimate.

### 3.3 Uncertainty factors

Even if one does an excellent job with the fundamental estimate structure, it will generally be quite impossible to ensure that all cost items are entirely independent. There will always be underlying conditions common to many, if not all, items; for example, market conditions. It is therefore common to extract these shared conditions into separate uncertainty factors that are calculated as a percentage markup on the relevant cost items.

In the overarching estimate structure, one often has a top item under which all the uncertainty factors are placed, as shown in the figure.

In the example in Figure 13, market conditions are separated as a separate uncertainty factor. For this uncertainty factor, one determines a baseline assumption about market conditions on the basis of which all items in the base estimate are priced. The uncertainty associated with this assumption is accounted for in the uncertainty factor.

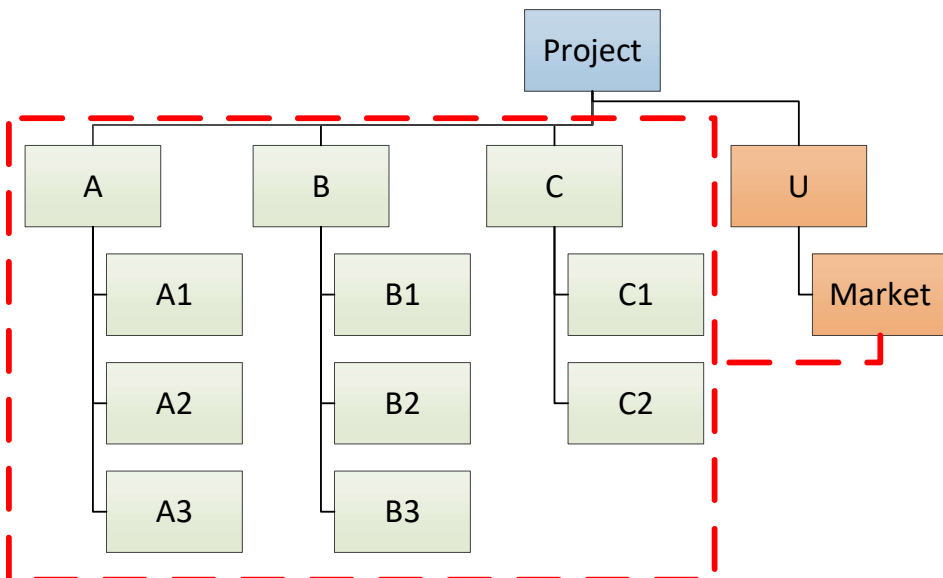


Figure 13: Uncertainty factor applied to the entire estimate

Uncertainty factors are specified as three-point estimates centred around zero or one, depending on practice. For example, (-10%, 0%, 15%) or (0.90, 1, 1.15). In most cases, the most likely value in the three-point estimate will be zero (or one). However, in certain cases, one may, during the estimating process, conclude that the most likely value differs from the baseline assumption used for the base estimate.

### 3.4 Covariation

In addition to having a suitable cost estimate structure and using uncertainty factors, when Monte Carlo simulation is employed, one can also capture covariation by explicitly incorporating it into the estimate. This entails stating the extent to which the input values in the estimate covary by specifying a pairwise covariation coefficient between 0 and 1. The number 0 indicates that the cost items are entirely independent, whereas 1 indicates that they are entirely dependent. Exactly how this is done is tool-dependent, but the most common is to use a so-called covariation matrix.

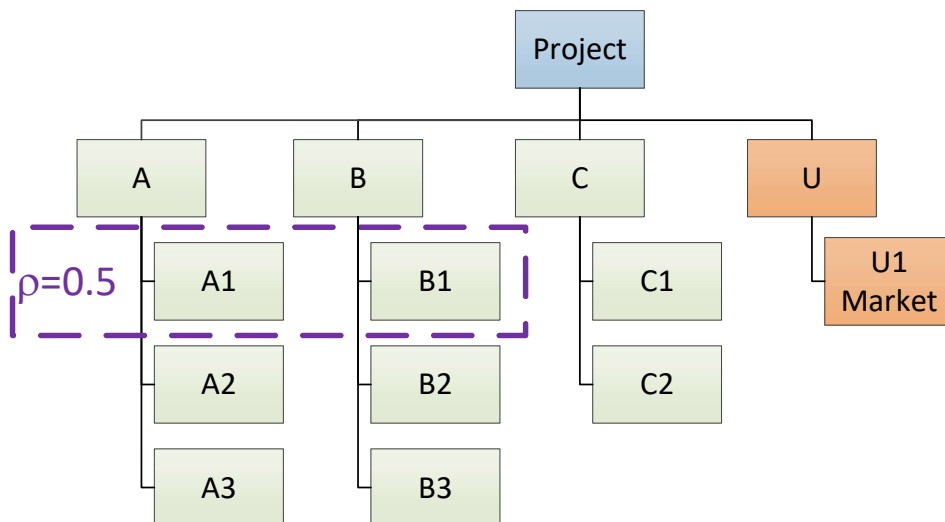


Figure 14: Modelling covariation.

The disadvantage of handling covariation in this way is that it is a relatively abstract concept, which can be harder for people to relate to than uncertainty factors. Moreover, covariation modelled in this way becomes largely invisible in the estimate compared with the use of uncertainty factors.

### 3.5 Event Items

The cost items discussed so far are all linked to estimation uncertainty. It is uncertain what the market price will be at the time of purchase, what the quantity will be, and similar factors. There may be underlying factors that push these amounts up or down, but none of the values are entirely dependent on a single event.

Uncertainty associated with single events is called event uncertainty. For example, whether a part of the project is to be built or not may depend on a political decision. In a cost estimate, we can address this type of uncertainty using event items. We must then state the probability of the event occurring and the cost consequence if it does.

## 3.6 Level of Detail in the Estimate

There are no clear rules for how detailed an estimate should be. However, there are some guidelines one should follow to secure the best possible result.

### **Keep control of stochastic dependence and covariation**

The main rule regarding the level of detail in the estimate is that one should never break the estimate down to a level of detail where one loses control of stochastic dependence and covariation.

### **Do not detail more than necessary for the purpose**

Another central principle is that one should never detail the estimate more than is required for the purpose of the estimate. If the primary aim is to determine the total project cost, one risks getting lost in the details if one breaks the estimate down to the nuts-and-bolts level. One cannot see the wood for the trees.

### **Do not detail more than there is a basis to do**

The level of detail in the estimate should always be adapted to the project's maturity level. If one prepares a very detailed estimate at an early stage, this entails making many assumptions about the final solution. One often lacks a basis for doing this in the early phase, thereby introducing uncertainty into the estimate that is not addressed.

### **Align with the level of experience data**

It is essential to be able to quantify the items in the estimate. That is to say, they must be at a level that harmonises with the empirical data that those who are to provide prices possess and can relate to.

## 4 Estimating uncertain quantities

In the estimates, all uncertain input values are given as probability distributions. The most common way to define these distributions is through three-point estimates derived from subjective expert judgements. Subjective expert judgements entail asking domain experts to assess values for the three-point estimates based on the information available and their best judgment.

An alternative to subjective expert judgements is to rely on hard statistical data. The disadvantage here is that one must have a very large number of observations on which to base the assessment. This makes such an approach more suitable for use in the insurance industry and similar sectors than for cost estimation in a large investment project. It often happens that someone crashes a car, but suspension bridges are seldom built in this country.

Although subjective expert judgements can sometimes appear to be 'guess-a-number', research has shown that, even when large amounts of statistical data are available, such judgements usually yield estimates that are as good as, or better than, those produced by purely statistical analysis.

The most common way to provide expert judgment is through group processes. This entails assembling a group of experts who have specific knowledge of the project, experience with similar projects, or specialist competence in areas relevant to estimating the project's cost. A facilitator then leads the group through a process to determine three-point estimates for the cost items in the estimate. A significant advantage of using group processes is that each participant contributes different perspectives, and the process enables the experts to gain insight into one another's experiences and reasoning.

## 5 The estimation process

Carrying out an estimation process can be a laborious affair, especially if one is to make use of group sessions. It is unfortunately not possible to cover this fully in this pamphlet. However, below we set out some main steps based on best practice.

### 1. Define the scope of the estimate and make assumptions

Before beginning the work on the estimate, one should ensure a clear delimitation of the project to be estimated. What lies within, and what lies outside, the project? Any assumptions used as the basis for the estimate must also be made explicit.

Note that, for the estimate to be valid, it is very important that the assumptions made are real. One cannot assume away uncertainty that the project is expected to handle.

### 2. Set up the structure for the base estimate

The base estimate is established in accordance with the principles given in chapter 3, but the cost items are not yet quantified. At this point, one should avoid detailing too far down. This can be done later if needed.

### 3. Identify uncertainty factors

After the structure of the base estimate has been established, it is time to identify the uncertainty factors, i.e., what is common and should be extracted from each cost item. Brainstorming is often used as a tool here. Many also use lists of standard uncertainty factors as a supplement. Relying entirely on such lists, however, is not advisable; one risks that the project-specific uncertainty is not captured.

Using brainstorming will typically yield a number of points that must be grouped into fewer factors. Again, the main rule is to ensure stochastic independence between the factors. At the same time, one should not have too few factors, as this often results in uncertainty that is not captured.

### 4. Define a baseline assumption for each uncertainty factor

When using uncertainty factors in the estimate, the idea is to separate this uncertainty from the base estimate. To accurately cost the base estimate, one must first establish baseline assumptions for each uncertainty factor. For example, if we choose to treat market conditions as an uncertainty factor, the uncertainty associated with this factor should reside within it, whereas all cost items in the estimate are priced based on a baseline assumption about market development.

### 5. Quantify the base estimate

When the baseline assumptions for each uncertainty factor are fixed, the base estimate can be quantified; that is, one goes into each cost item and specifies

a three-point estimate. One should always make a rigorous assessment of the spread of uncertainty for each cost item and avoid using template-like approaches, such as merely setting the extreme values to  $\pm 10\%$  of the most likely value.

### **6. Assess and quantify the uncertainty factors**

After the base estimate has been quantified, it is time to assess and quantify the uncertainty factors. This means assessing what may happen relative to the baseline assumption used in the base estimate and how many per cent up or down this will amount to.

### **7. Assess covariation**

Assuming that one uses a calculation method and a tool that allow modelling of covariation, this is the time to do it. After working through the entire estimate and quantifying the cost items, one can assess whether any dependencies between cost items remain unaccounted for and, if so, enter covariation into the estimate model.

### **8. Calculate**

Unless one uses pen and paper, this step usually consists simply of pressing a button and waiting while the simulation runs.

### **9. Assess the result**

After the estimate has been calculated, one must assess whether the result obtained is usable and reasonable. Is the uncertainty within the specified requirements, and perhaps even more importantly, is it large enough? There have been cases in which one has ended up with an uncertainty of under five per cent for projects assessed at the decision-making stage on financing, which is usually utterly unrealistic.

### **10. Refine**

If the estimate result is unacceptable, one must go back and do something. If the uncertainty is too high relative to the specified requirements, one should consider obtaining additional information to reduce uncertainty in the input values and, if necessary, to refine the estimate. Note that one should never split cost items, remove uncertainty factors, or take similar measures solely to reduce uncertainty in the estimate. Unless one possesses information indicating that this is the right thing to do, one has merely removed the uncertainty on paper. It has been 'calculated away', but the uncertainty has not disappeared from the project for that reason.

If, on the other hand, the total uncertainty proves to be too low relative to what is considered normal, the cause is likely that one has not sufficiently ensured stochastic independence and has not handled covariation adequately in the estimate. One must then go back and reassess the estimate structure, the use of uncertainty factors, and the modelling of covariation.

Another possible cause of low uncertainty may be that the three-point estimates for the cost items are too narrow, suggesting that the process for quantifying the cost items in the estimate has not been optimally conducted.

## 6 Calculating with uncertain quantities

There are two principal ways to calculate stochastic cost estimates: by using analytical methods and by using simulation methods.

When using analytical methods, the costs for each element are described by a mathematical expression. Each cost is then combined into a new expression that yields the distribution of possible outcomes. Modelling and computing a cost analysis using exact mathematical probability distributions and calculation methods is both very time-consuming and requires considerable mathematical proficiency. *Successive calculation* is an analytical method that uses approximation formulae that are sufficiently accurate for cost-estimation purposes. The methodology was originally developed as a computational tool for use in the so-called *Successive Principle* cost estimation methodology, hence the name.

Monte Carlo simulation is more of a 'brute-force' method compared with Successive calculation. Instead of calculating the estimate once, one does it several hundred to several thousand times. That is to say, one naturally has a computer run this. For each calculation run, a random value is drawn from the probability distribution of each item and used in the calculation. In other words, for each run, the computer 'throws dice' to determine which value it shall use for all the items and calculates the entire estimate based on these. After the simulation has finished, the computer has a statistical basis for preparing a probability distribution of the total sum and any other quantities one may wish to obtain.

### 6.1 Successive calculation versus Monte Carlo simulation

As of today in Norway, Monte Carlo simulation is the most common calculation method used in uncertainty analysis. The results obtained from Monte Carlo simulation and the Successive formulae are essentially the same, provided the same model structure is used. The advantage of Monte Carlo simulation is that one has greater freedom in how the model is built, which distribution functions are used, and the possibilities for correcting for covariation. But this, in turn, places much greater demands on the person who is to prepare the estimates. The exception here is specific tools, such as the Norwegian Public Roads Administration's *Anslag*, where a more rigid system for the estimate structure has been set up and the possibilities for choosing distribution functions, number of iterations, etc., have been limited in order to abstract away some of the complexity of the methodology.

Access to software is another point of criticism with Monte Carlo simulation. It is usually either relatively expensive or not available to the general public. An advantage of using the Successive formulae is that one can manage with an ordinary spreadsheet or even pen and paper.

Another disadvantage of Monte Carlo simulation is that it is often regarded as a black box. One puts some numbers in at one end and gets others out at the other,

but one neither sees nor understands what is happening inside. The methodology is therefore not well suited to providing an intuitive understanding of how calculations with uncertainty work. On the other hand, by using the Successive formulae, one can follow the calculations step by step and observe the results.

If one has not previously worked with cost estimation under uncertainty but is interested in trying it, it is advisable to start with the Successive formulae. You do not need to acquire or familiarise yourself with new software, and you will develop an intuitive understanding of how to calculate with uncertain quantities.

## 6.2 Successive calculation

In Successive calculation, one uses formulae based on expected values and standard deviations. The input values are given as Gamma distributions specified by a three-point estimate, and one computes each cost item's expected value and standard deviation from this.

Calculating with the expected value and standard deviation is possible because one assumes that the final estimate will be normally distributed. This assumption rests on the central limit theorem, which states that the sum of a large number of stochastic values with random distributions tends to be normally distributed. This means that even if one has distributions that are highly skewed taken individually, the result of summing them will be approximately normally distributed if one has a sufficiently large number of cost items.

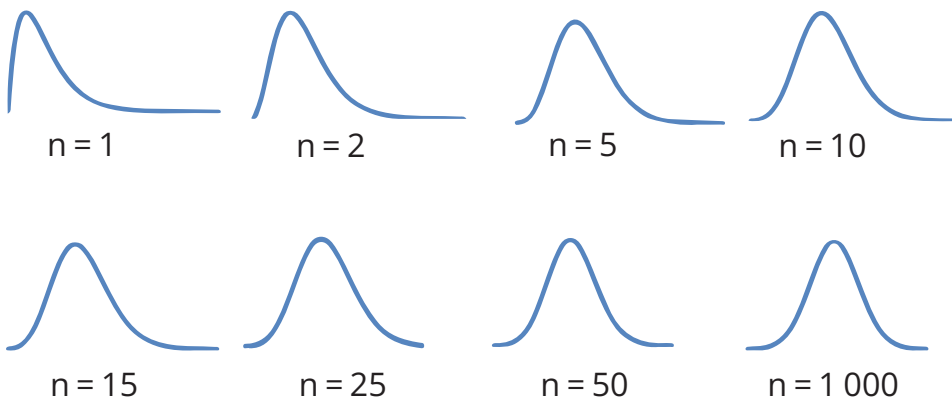


Figure 15: Change in curve shape when summing  $n$  identical items

Typically, approximately 30 cost items are considered sufficient. But this presupposes that the cost items are of the same order of magnitude and that there is more summation than multiplication in the estimate. (The central limit theorem applies to the summation of items, not to multiplication. With multiplication, the skewness increases.) For example, an estimate in which one has more uncertainty factors than cost items will typically not be normally distributed.

### **Formula set**

In Appendix A there is a complete set of formulae for Successive calculation with an explanation of when these are used, and worked examples for each of the formulas.

## **6.3 Monte Carlo simulation**

### **Monte Carlo tools**

To use Monte Carlo simulation as a calculation method, one must have suitable software available. There are certain specialised tools for cost estimation that use Monte Carlo simulation, for example the Norwegian Public Roads Administration's Anslag programme, but the most widespread are general Monte Carlo tools that run on top of Excel. The best known are @Risk and Crystal Ball.

### **Estimate setup**

Setting up estimates in such tools is usually straightforward. One sets up the estimate in an ordinary Excel spreadsheet, then specifies which cells contain uncertain values and which probability distribution to use, and which cells contain results to retain. Although one has considerable freedom to do as one wishes, it is recommended that one adhere to the guidelines provided in chapter 3.

### **Distribution functions**

When using Monte Carlo simulation, one must choose which distribution functions to use for each uncertain quantity in the estimate. One typically uses the same distribution type for all the cost items in the estimate. The main rule for selecting the distribution function is that it must be possible to specify it using a three-point estimate. Although a three-point estimate uniquely defines most distributions, the simulation tool used may not have built-in algorithms to convert the three-point estimate into the distribution's actual parameters. For example, the Gamma distribution, which is used in Successive calculation, cannot typically be specified using a three-point estimate. There is usually an option to specify using three percentiles, but specifying P50 is not the same as specifying the most likely value, which is what one uses in three-point estimates. It is therefore better to choose another distribution that actually allows us to use a three-point estimate as desired.

In most simulation tools, the PERT distribution will be the most suitable. One must simply be aware that it initially takes absolute minimum and maximum values as the endpoints. One must therefore specify that one instead uses the desired percentiles (for example, P10 and P90).

### **Iterations**

When we speak of iterations in connection with simulation, we mean full calculations of the estimate. In each iteration, the simulation engine draws new samples for all uncertain quantities and performs the calculations deterministically. After-

wards, all iterations are run, the simulation engine gathers all up.

It is essential to run enough iterations to obtain a reliable estimate. What is sufficient depends on the size and structure of the estimate; however, as a rule, 10,000 iterations are adequate. Most simulation tools also have the option to run the simulation until it converges.

### **Seed**

Simulation engines use a pseudorandom number generator seeded with a value. A given seed will always produce the same stream of random numbers. This means that if one runs a simulation twice in a row, one will obtain precisely the same result, even with only tens of iterations. It is therefore recommended that the simulation tool be set to use a random value as the seed. This makes it easier to determine whether the number of iterations is sufficient. A simulation of an estimate run twice with different seeds should give approximately identical results.

### 6.4 Complete calculation example

A complete calculation example illustrating how an estimate can be set up according to the methodology presented in this pamphlet is available for download from Concept’s website. There is both a version that uses Successive calculation and a version configured for Monte Carlo simulation with @Risk (the latter requires @Risk to be installed in addition to Excel).

Line item	P10	Likely	P90	Exp.val.	Std.dev.	Variance	Exp.val.	Variance	Std.dev.	Share
A Thing	140 000.00	150 000.00	165 000.00	152 066.12	9 881.42	97 642 519	152 066.12	97 642 519	9 881.42	3 %
<b>B Thingamajig</b>							<b>229 073.97</b>	<b>2 716 652 081</b>	<b>52 121.51</b>	
<b>B1 Labour</b>							<b>92 859.23</b>	<b>959 304 100</b>	<b>30 972.63</b>	27 %
B1a Hours	100.00	120.00	200.00	144.79	39.53	1 562				
B1b Hourly rate	500.00	600.00	800.00	641.32	118.58	14 061				
<b>B2 Parts</b>							<b>136 214.74</b>	<b>1 757 347 982</b>	<b>41 920.73</b>	50 %
B2a Cost	5 000.00	6 000.00	8 000.00	6 413.22	1 185.77	1 406 052				
B2b Quantity	15.00	20.00	28.00	21.24	5.14	26				
<b>Total base estimate</b>							<b>381 140.09</b>	<b>2 814 294 600</b>	<b>53 049.93</b>	
<b>Uncertainty factors</b>										
U1 Market	0.95	1.02	1.10	1.02	0.06	0.0035	9 197.76	657 997 154.88	25 651.46	19 %
U2 Project Management	0.98	1.00	1.02	1.00	0.02	0.0002	0.00	37 015 309.67	6 084.02	1 %
<b>Total uncertainty factors</b>							<b>9 197.76</b>	<b>695 012 465</b>	<b>26 363.09</b>	
<b>Total project</b>							<b>390 337.85</b>	<b>3 509 307 065</b>	<b>59 239.40</b>	<b>100 %</b>

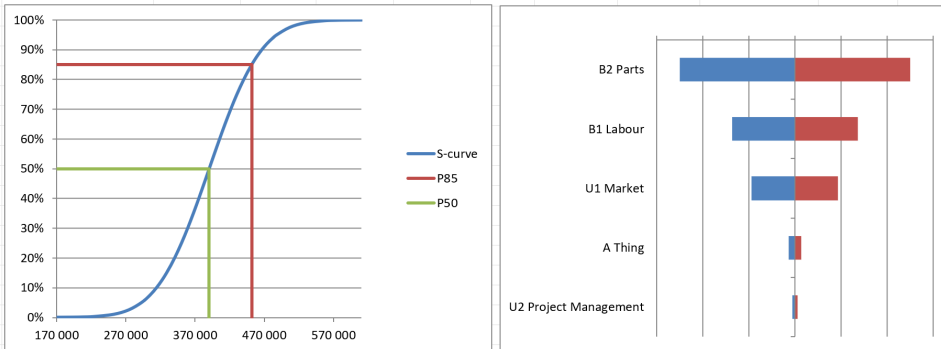


Figure 16: Screenshot of the Excel example that can be downloaded from Concept’s website

## 7 Estimate results

The main result of a stochastic cost estimate is a probability distribution of the project's total investment cost. It is common to present this graphically as an S-curve, as described in chapter 2. In addition, it is customary to extract relevant measures for centre and spread, typically the expected value, the median (P50), the standard deviation, and the relative standard deviation.

In many organisations, the funds appropriated to a project are based on a sum at a higher percentile than the centre. This is to increase the probability that the budget holds, even if the project incurs negative outcomes due to uncertainty — that is, an allowance for uncertainty is set aside in addition to the project is expected cost. For example, in large Norwegian government investment projects, it is common to use P85 (which is approximately the expected value plus one standard deviation). It is then also common that this figure is presented as part of the estimate result.

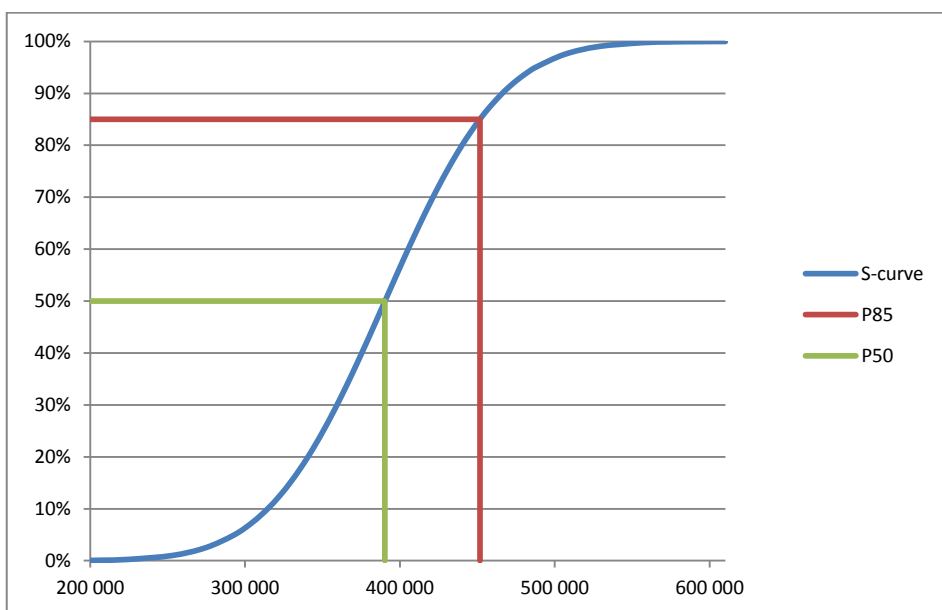


Figure 17: Example of an S-curve with P50 and P85

The S-curve and the measures provide an overview of what the project is estimated to cost and the uncertainty surrounding this estimate. In addition, it is desirable to identify which components of the estimate contribute most to the uncertainty, both to assess and prioritise measures. Furthermore, to determine where one should work, if necessary, to obtain further information if the uncertainty in the estimate is unacceptable.

A tornado diagram provides an overview of which cost items in the estimate con-

tribute most to the uncertainty and the share of the total uncertainty each contributes. How this is presented varies with which calculation method is used. In Monte Carlo tools, this usually emerges on the basis of a statistical analysis. When using Successive calculation, one computes a cost item's share of the total uncertainty by dividing the item's variance by the project's total variance.

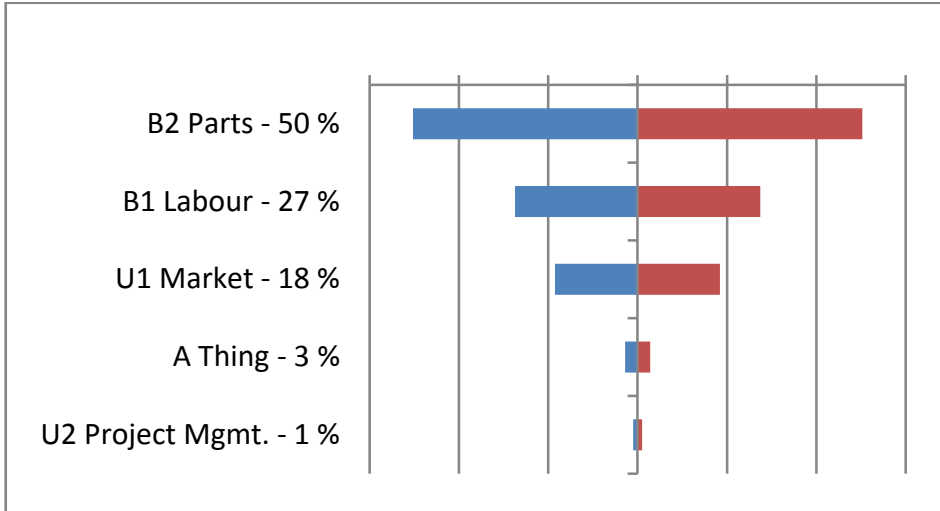


Figure 18: Example of a tornado diagram

Taken together, the S-curve and the tornado diagram provide a clear picture of the project's cost, the magnitude of uncertainty around this figure, and where this uncertainty is located within the project. They are therefore good tools for communicating estimate results to decision-makers, provided that these are trained in reading this type of information.

## Appendix A Successive calculation: formula set and worked example

### Calculating input values

In Successive calculation, all input values are assumed to follow a Gamma distribution and are specified using a three-point estimate. To obtain the expected value and standard deviation of these distributions for use in subsequent calculations, we apply the following formulae.

Table 1: Formulae for calculating the expected value and standard deviation from a three-point estimate

	Expected value	Standard deviation
<b>P1 / P99</b>	$E = \frac{P_1 + 2.9 \times M + P_{99}}{4.9}$	$\sigma = \frac{P_{99} - P_1}{4.6}$
<b>P10 / P90</b>	$E = \frac{P_{10} + 0.41 \times M + P_{90}}{2.41}$	$\sigma = \frac{P_{90} - P_{10}}{2.53}$

Where

$E$  = expected value

$\sigma$  = standard deviation

$M$  = most likely value

$P_n$  = the value at the  $n$ th percentile (quantile)

### Worked example

Assume we have an item in the estimate given by the three-point estimate (100, 200, 400), where the extreme values correspond to  $P_{10}$  and  $P_{90}$ . Then the calculations of the expected value and the standard deviation are:

$$E = \frac{P_{10} + 0.41 \times M + P_{90}}{2.41} = \frac{100 + 0.41 \times 200 + 400}{2.41} = 241.32$$

$$\sigma = \frac{P_{90} - P_{10}}{2.53} = \frac{400 - 100}{2.53} = 117.65$$

**Summation**

Summing expected values is direct addition. Summing standard deviations is done by summing variances and then taking the square root of the total variance.

In a normal cost estimate, all items that are not at the lowest level in the hierarchy are *sum items* and are calculated using these formulae.

Table 2: Summation of expected value and standard deviation

	Expected value	Standard deviation
<b>Two items</b>	$E_{\text{tot}} = E_A + E_B$	$\sigma_{\text{tot}} = \sqrt{\sigma_A^2 + \sigma_B^2}$
<b><i>n</i> items</b>	$E_{\text{tot}} = \sum_{i=1}^n E_i$	$\sigma_{\text{tot}} = \sqrt{\sum_{i=1}^n \sigma_i^2}$

**Worked example**

Assume we have two items *A* and *B* with the following expected values and standard deviations:

	<i>E</i>	$\sigma$
<i>A</i>	500	150
<i>B</i>	400	200

Then

$$E_{\text{Tot}} = E_A + E_B = 500 + 400 = 900$$

and

$$\sigma_{\text{Tot}} = \sqrt{\sigma_A^2 + \sigma_B^2} = \sqrt{150^2 + 200^2} = 250$$

**Multiplication**

Calculating the product of uncertain quantities is straightforward multiplication for the expected value. However, for the multiplication of standard deviations it becomes a little more involved, because here the expected values of the uncertain quantities also play a role, as illustrated in the figure to the right.

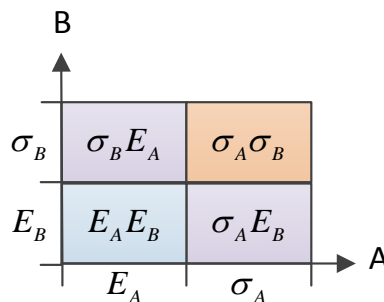


Figure 19: Multiplication of uncertain quantities

The formulae for the multiplication of standard deviations apply to 'hierarchical' multiplication. This applies, for example, when multiplying quantity and price. If, on the other hand, we have multiplication that goes outside the cost-estimate hierarchy, such as when using mark-up items or uncertainty factors, then one must take a few additional steps to include all of the uncertainty.

This is because, when we calculate with uncertain quantities, the following holds when calculating the total for the product of a sum of items and an uncertainty factor:

$$\text{Tot} = P \times (1 + f) \neq P + P \times f$$

The reason for this is that  $P$  does not represent a single value, but a variable that can take several values. That is, in the last expression where  $P$  appears twice,  $P$  is in reality not one and the same item, but rather two identical, yet stochastically independent, items.

In order to include all uncertainty for a factor  $f$ , we must therefore first calculate the total for  $P \times f$  and then subtract the uncertainty for  $P$  that is already included in the estimate.

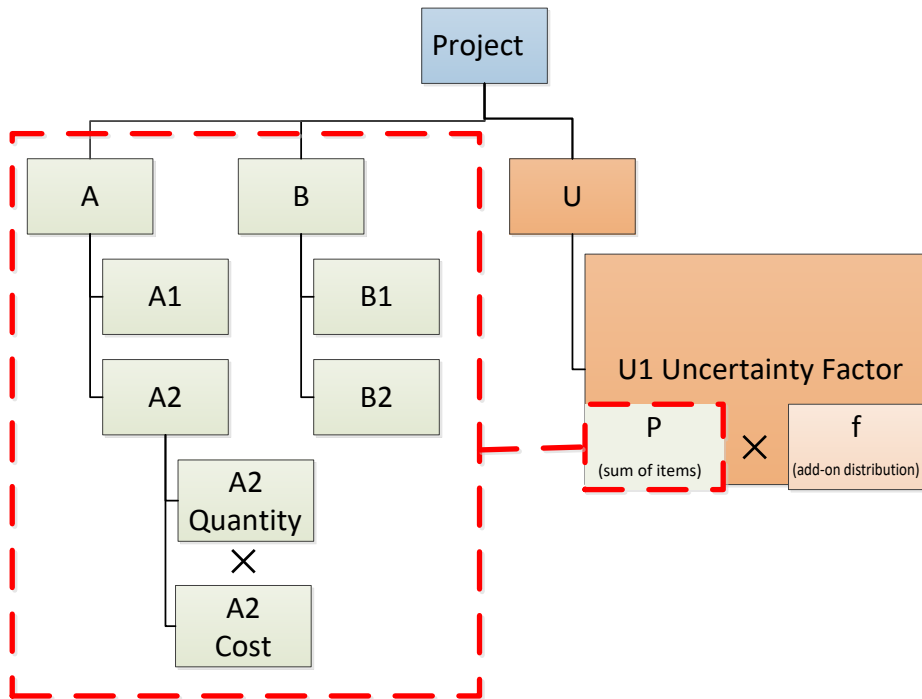


Figure 20: Hierarchical and non-hierarchical multiplication in the estimate

Table 3: Formulae for the multiplication of expected value and standard deviation

	Expected value	Standard deviation
<b>Multiplication</b>	$E_{Tot} = E_A \times E_B$	$\sigma_{Tot} = \sqrt{(\sigma_A E_B)^2 + (\sigma_B E_A)^2 + (\sigma_A \sigma_B)^2}$
<b>Multiplication of a factor and sum of items<sup>1</sup></b>	$E_F = E_P \times E_f$	$\sigma_F = \sqrt{(\sigma_f(1 + E_P))^2 + (\sigma_P E_f)^2 + (\sigma_P \sigma_f)^2 - \sigma_P^2}$

<sup>1</sup> “Sum of items” refers to a subtotal  $P$  in the estimate, and  $f$  is the factor (e.g., mark-up).

**Worked example - product of quantity and price**

Assume we have an item with quantity  $A$  and price  $B$  with expected value and standard deviation as given:

	$E$	$\sigma$
$A$	10	12
$B$	20	15

Then

$$E_{\text{Tot}} = E_A \times E_B = 10 \times 20 = 200$$

and

$$\begin{aligned} \sigma_{\text{Tot}} &= \sqrt{(\sigma_A E_B)^2 + (\sigma_B E_A)^2 + (\sigma_A \sigma_B)^2} \\ &= \sqrt{(12 \times 20)^2 + (15 \times 10)^2 + (12 \times 15)^2} = 335.41 \end{aligned}$$

**Worked example - product of sum of items and factor**

Assume we have a sum of items given by  $P$  and a factor  $f$ , where the probability distribution for the mark-up is given by  $f$ :

	$E$	$\sigma$
$P$	1000	100
$f$	0.02	0.10

Then

$$E_F = E_P \times E_f = 1000 \times 0.02 = 20$$

and

$$\sigma_F = 102.49$$

## Concept Theme Pamphlets

Concept Theme Pamphlets that have been translated to English.

<b>Pamphlet no.</b>	<b>Title</b>	<b>Author(s)</b>	<b>Published</b>
4	Cost estimation under uncertainty	Frode Drevland	2013 (EN 2025)
6	Process for cost estimation under uncertainty	Olav Torp Frode Drevland Kjell Austeng	2015 (EN 2025)

Concept Theme Pamphlets and other publications from the Concept Programme can be downloaded from: **[ntnu.no/concept](https://ntnu.no/concept)**

The Concept research program aims to develop know-how to help make more efficient use of resources and improve the effect of major public investments. The program is designed to follow up on the largest public projects over a period of several years, and help improve design and quality assurance of future public projects before they are formally approved.

The program is based at The Norwegian University of Science and Technology (NTNU), at the Department of Civil and Environmental Engineering. It cooperates with key Norwegian and international professional institutions and universities, and is financed by the Norwegian Ministry of Finance.



The Concept Research Program  
Department of Civil and Environmental Engineering, NTNU, N-7491 Trondheim, NORWAY

Information about om the Concept Research Program,: [www.ntnu.no/concept](http://www.ntnu.no/concept)

ISSN: 1891-5620 (print version)

ISSN: 1891-5655 (online version)

ISBN: 978-82-8433-060-0 (print version)

ISBN: 978-82-8433-061-7 (online version)