Distributed Control and Stochastic Analysis of Hybrid Systems
Supporting Safety Critical Real-Time Systems Design

WP8: Accident risk decomposition

BIAS AND UNCERTAINTY MODELLING IN ACCIDENT RISK ASSESSMENT

Mariken Everdij[1] and Henk Blom[1]

24th February 2005

Version: *0.6*
Task number: *8.4*
Deliverable number: *D8.4*
Contract: *IST-2001-32460 of European Commission*

[1] National Aerospace Laboratory NLR, NL

DOCUMENT CONTROL SHEET

Title of document: *Bias and uncertainty modelling in accident risk assessment*

Authors of document: *M.H.C. Everdij and H.A.P. Blom*

Deliverable number: *D8.4*

Contract: *IST-2001-32460 of European Commission*

Project: *Distributed Control and Stochastic Analysis of Hybrid Systems Supporting Safety Critical Real-Time Systems Design (HYBRIDGE)*

DOCUMENT CHANGE LOG

Version #	Issue Date	Sections affected	Relevant information
0.1	20 April 2004	All	Initial draft
0.2	17 June 2004	All	First draft
0.3	28 June 2004	All	Update
0.4	30 July 2004	All	Update
0.5	18 January 2005	All	Incorporation of CENA and NLR review
0.6	24 February 2005	2,3,4	Incorporation of Twente review

Version 1.0		Organisation	Signature/Date
Authors	M.H.C. Everdij	NLR	
	H.A.P. Blom	NLR	
Internal reviewers	P. Lezaud	CENA	
	S.H. Stroeve	NLR	
	A. Bagchi	TWEN	

Abstract

Accident risk assessment for complex safety critical operations such as encountered in air traffic management has to be done by an appropriate combination of stochastic analysis and Monte Carlo simulations. Within work package 8 of the HYBRIDGE project, novel sequential Monte Carlo simulation-based methods have been developed to estimate rare event probabilities. As part of that work package, this report studies bias and uncertainty modelling in rare event estimation. Part of this study is a literature review, and a comparison with the approach developed by [Everdij and Blom, 2002], which has been applied in many realistic applications, and appears to work well.

The bias and uncertainty modelling approach developed in this report is an extension of the [Everdij and Blom, 2002] approach. It uses the model-based accident risk decomposed as a sum of risk contributions (i.e. weighted conditional risks), together with the list of assumptions adopted during the modelling, including assumptions on the parameter values used in the model, and next compensates for the effect of these assumptions by adapting the weighted conditional risks with compensation factors, and next combining the results. The outcome of the bias and uncertainty modelling is an assessment of expected ('realistic') accident risk, together with a 95% credibility interval.

Contents

1. Introduction

This section introduces this report by giving its background as output of work package 8.4 of the HYBRIDGE project, by describing its objective, and by outlining its contents.

1.1 HYBRIDGE work package 8

Accident risk assessment for complex safety critical systems such as air traffic management has to be done by an appropriate combination of stochastic analysis and Monte Carlo simulations. Unfortunately, at this moment, the identification of an appropriate way to combine both approaches is more an art than a science, and the art tends to fall short when the scale of complexity increases and the control becomes distributed. The aim of work package 8 of the HYBRIDGE project is to develop novel methods for the decomposition of risk such that extremely low risk values can be assessed through a hierarchy of conditional Monte Carlo simulations. Work package 8 consists of four tasks:

- Task 8.1: Review existing risk decomposition and assessment methods, both analytical ones, Monte Carlo simulation approaches and combinations of these two. This review should distinguish between theory-based methods and heuristic methods.
- Task 8.2: Develop new risk decomposition and assessment methods. One of the key directions to be explored is the development of risk decomposition methods that make use of the fact that for strong Markov processes the Markov property holds true for stopping times.
- Task 8.3: Develop conditional Monte Carlo simulation techniques for accident risk assessment that make use of the risk decomposition developed in Task 8.2, and compare the new approach with the existing ones identified in Task 8.1.
- Task 8.4: Extend the risk decomposition approach with a recursive Bayesian estimation approach which enables the updating of the accident risk assessment while more and new information is coming available.

Tasks 8.1, 8.2 and 8.3 have been performed and reported in [HYBRIDGE D8.1], [HYBRIDGE D8.2] and [HYBRIDGE D8.3], respectively. The current report addresses Task 8.4, with emphasis on bias and uncertainty assessment, with in particular an extension of the approach reported in [Everdij and Blom, 2002].

1.2 Objective of the report

A typical accident risk assessment starts with developing a mathematical model of ‘reality’, after which accident risk is assessed using this model. During the modelling and accident risk assessment, assumptions are adopted for different reasons, such as:

- It is simply not possible to exactly copy ‘reality’ in all its nuances in the model. Minimal assumptions cover the boundaries within which the operation under consideration is studied or cover for aspects of reality that are unknown or uncertain.
- Often, during the model development and accident risk assessment more and new information on the operation is coming available. The incorporation of this new information sometimes requires re-doing the modelling all the way from the start, which takes a lot of effort. This is especially the case if the new information keeps coming in. To save this effort, the new information can be formulated by means of assumptions, the

effects of which are compensated for afterwards, after the model-based risk assessment has been completed.

Fortunately, it is not necessary to exactly copy ‘reality’ in the model. As long as the relevant aspects are modelled, minor details not influencing accident risk can often be omitted. Of course, the problem is to know this prior to performing risk assessments; in reality this is known in hindsight. Therefore, a more practical approach is to first use this mathematical model to cover the aspects of ‘reality’ that are believed to influence accident risk significantly, subsequently evaluate model-based accident risk, and next compensate for the effects of the model assumptions adopted. The compensation method is usually a bias and uncertainty assessment method. The outcome of the bias and uncertainty assessment is an assessment of expected (‘realistic’) accident risk, together with a credibility assessment of this expectation.

The aim of this report is to develop such bias and uncertainty assessment method, by extending the method developed in [Everdij and Blom, 2002]. The necessity for this extension rises from the fact that the method developed in [Everdij and Blom, 2002] compensates for assumptions adopted by using a multiplicative approach. In other HYBRIDGE WP8 tasks we saw that, to decompose risks that are evaluated through a hierarchy of conditional Monte Carlo simulations, accident risk is often written as a sum of weighted risks. And for the bias and uncertainty assessment of such sum, a multiplicative approach is less satisfactory.

1.3 Organisation of the report

This report is organised as follows:

- Section 1 introduces the current report.
- Section 2 outlines the bias and uncertainty assessment problem and literature on it.
- Section 3 develops, in a mathematical setting, the bias and uncertainty assessment method of [Everdij and Blom, 2002].
- Section 4 discusses and evaluates the assumptions adopted during the development of the method of Section 3.
- Section 5 provides the main contribution of this report: it considers accident risk as a sum of weighted risks instead of as a single accident risk value, and next develops a bias and uncertainty assessment method for this sum, by extending the method of Section 3.
- Section 6 discusses and evaluates the assumptions adopted during the development of the method of Section 5.
- Section 7 gives conclusions.
- Section 8 provides references.
- Appendix A gives the proofs of the theorems, lemmas and corollaries listed in Sections 3 and 5.
- Appendix B outlines properties of the lognormal distribution, which plays an important role in the bias and uncertainty assessment methods of Sections 3 and 5.
- Appendix C outlines approximations found in literature that consider sums of lognormal random variables; these sums play an important role in Section 5.

2. Problem definition and related literature

This section explains how a bias and uncertainty assessment makes part of an accident risk assessment (Subsection 2.1), gives an overview of literature on bias and uncertainty assessment in accident risk assessments (Subsection 2.2), gives a brief outline of the [Everdij and Blom, 2002] method (Subsection 2.3), and gives a comparison of that method with the literature (Subsection 2.4).

2.1 Bias and uncertainty in accident risk assessment

The accident risk assessment of a particular operation generally follows four tasks:
1 Development of a mathematical model;
2 Model-based accident risk assessment;
3 Bias and uncertainty assessment;
4 Realistic accident risk assessment: Combination of 2 and 3.

Task 1 involves the development of a stochastic dynamical *model* of the accident risk. Next, in Task 2, Monte Carlo simulations and/or mathematical analysis techniques are performed to assess this *model-based* accident risk. This model-based accident risk is only realistic if the model is a good representation of "reality". Since during the development of the stochastic dynamical model is it always necessary to adopt assumptions, the model will never be an exact replication of reality. Examples of these model assumptions are: numerical approximation assumptions, model structure assumptions, parameter value assumptions. Obviously, the model-based accident risk is not very useful if one has no idea of the effect on accident risk of these assumptions adopted. The bias and uncertainty assessment (Task 3) overcomes this problem by assessing each of the model assumptions adopted and estimating their effect on accident risk. After compensation of the combined effects (Task 4), one obtains an estimate of realistic accident risk, including an idea of the credibility of this estimate.

2.2 Literature on bias and uncertainty in accident risk assessment

The literature on bias and uncertainty in accident risk assessment covers several main sources, which provide relevant and complementary results regarding the uncertainty in accident risk assessment. These sources are outlined next, in historical sequence.

[Morgan and Henrion, 1990]
[Morgan and Henrion, 1990] explain (in their Chapter 4) that there are different types of quantity:
- Empirical quantities, which represent measurable properties of the real-world systems being modelled, e.g. fuel price;
- Defined constants, the value of which is certain, e.g. the number of days in July;
- Decision variables, also named control variables or policy variables, which are quantities over which the decision maker exercises direct control, e.g. plant size;
- Value parameters, which represent aspects of the preferences of the decision makers, e.g. discount rate;
- Index variables, which are used to define a location or cell in the spatial or temporal domain of the model, e.g. a particular year in a multi-year model;

- Model domain parameters, which specify the domain or scope of the system being modelled, e.g. time horizon;
- State variables, which are members of a minimal subset of variables from whose values it is possible to compute the values of all the model's variables;
- Outcome criteria, which are the variables used to rank or measure the desirability of possible outcomes, e.g. net profit.

Next, they discuss uncertainty in empirical quantities in terms of different kinds of sources from which it can arise, i.e. Statistical variation, Subjective judgement, Linguistic imprecision, Variability, Inherent randomness, Disagreement, Approximation, and discuss uncertainty about model form.

Chapters 5, 6 and 7 of [Morgan and Henrion, 1990] discuss human judgement about and with uncertainty, and how to use it. Here, they note that the psychology of judgement under uncertainty can lead to biased outcomes. The probability judgement of interviewed experts is driven by the ease with which they can think of previous occurrences of the event, or the ease with which they can imagine the event occurring. Also, biased results can be obtained if the experts are given particular background information in advance, for instance a base statistic.

Chapter 8 is most interesting given the objective of the current report; it is on the propagation and analysis of uncertainty. Various analytic and computational techniques are examined, e.g.,

- Sensitivity analysis, which computes the effect of changes in inputs on model predictions;
- Uncertainty propagation, which calculates the uncertainty in the model outputs induced by uncertainties in the inputs; and
- Uncertainty analysis, which compares the importance of the input uncertainties in terms of their relative contributions to uncertainties in the outputs.

Several basic concepts for each of these techniques are introduced, and next examined in more detail.

[Kumamoto and Henley, 1996]

[Kumamoto and Henley, 1996, Sections 11.6, 11.7 and 11.8] address the problem of how uncertainties in the component failures of fault trees are propagated through the AND gates and OR gates, to give uncertainty of the probability of occurrence of the top event in these fault trees. Kumamoto and Henley propose three methods for uncertainty propagation:

1. Monte Carlo propagation, in which component unavailabilities are sampled many times from probability distributions, and each of these unavailabilities are then propagated toward a top event. The resultant point values are used to evaluate the top event unavailability uncertainty. Additional techniques are proposed to deal with rare events.
2. Analytical moment propagation, in which the first and second moments of the summed or multiplied variables are evaluated and used to propagate the uncertainties. Various methods are proposed, amongst which Taylor-series expansion, orthogonal expansion, response surface methods. These methods work well for OR gates. An alternative approach is to describe the uncertainties in the component failures by Lognormal random variables (a random variable has a lognormal distribution if its natural logarithm has the normal distribution), because then their product is again Lognormal, such that at the AND gates random variables having these distributions can be multiplied.
3. Discrete probability algebra, in which continuous probability densities are approximated by a set of discrete probabilities. This method can be regarded as a deterministic version of Monte Carlo and is efficient for output functions with simple structures.

[Haimes, 1998]

[Haimes, 1998] gives an in depth overview of different types of bias and uncertainties. With respect to a formal risk assessment approach, three types of bias or uncertainties have to be handled:

- Parameter bias or uncertainty, referring to the lack of knowledge on the true values for the parameters used in the model.
- Model bias or uncertainty, referring to uncertainties in the general knowledge of the process: Models are simplified representations of real-world processes; as such, they must make certain assumptions concerning the true state of nature.
- Decision bias or uncertainty, referring to controversy or ambiguity concerning how to compare and weigh social objectives. Examples of sources for decision uncertainty are how to decide the social cost of risk, or how to decide what level of risk is acceptable.

Regarding non-parametric bias and uncertainties, [Haimes, 1998] explains that one should be aware that there is a large variety of bias and uncertainty types:

- Surrogate variables, referring to the use of particular variables instead of the real ones which are too difficult to assess;
- Excluded variables, referring to variables that are deemed insignificant in the model;
- Impact of abnormal situations, referring to e.g. simplification by aggregating numerous circumstances into a few broad categories, or using a model to represent a situation outside of its design;
- Approximation bias or uncertainty, covering the remaining types of uncertainty due to model generalisation, e.g. discrete distributions to represent continuous processes, or the limitation of finite runs in Monte Carlo analysis;
- Incorrect form, concerning the validity, or accuracy of the basic model being used;
- Bias or uncertainty due to disagreement, referring to e.g. conflicting data or expert opinion.

Regarding model parametric bias or uncertainties, [Haimes, 1998] studies the problem of optimising a (multi-objective) control function of both control variables and uncertain parameters, and shows that there is a trade-off for the decision maker between reduction in the optimality function and reduction in the uncertainty of model parameters. Some first-order approximation algorithms are proposed that handle optimality and sensitivity / uncertainty systematically and simultaneously. The algorithms are illustrated with some simple examples, where the functions to be optimised are polynomial in their variables.

[Hattis and Anderson, 1999], [Hammit and Shlyakhter, 1999], [Frey and Burmaster, 1999]

These are three journal papers, this first of which discusses issues like “What should be the implications of uncertainty”, “When is it reasonable and when is it bad to use a biased estimation procedure”, and “When and how does uncertainty matter”. The second paper discusses the expected value of information (EVI, a measure of the value of uncertainty reduction in making optimal decisions regarding alternatives), the probability of surprise, and tests for bias in estimating the EVI, and suggests procedures to guard against overconfidence and underestimation of the EVI when developing prior distributions and when combining distributions obtained from multiple experts. (Note that the EVI has also been discussed in [Morgan and Henrion, 1990], Chapter 12.) The third paper discusses methods for characterising variability and uncertainty, by comparing bootstrap simulation and a likelihood-based method.

[Ayyub, 2001]
[Ayyub, 2001] studies the problem of how to elicitate and combine expert opinions. Ayyub identifies three types of uncertainty metrics:
- Nonspecificity that results from imprecision connected with set sizes.
- Likelihood that results from various basis assignments represented by Entropy-like uncertainty measures.
- Fuzziness that results from vagueness.

Subsequently, [Ayyub, 2001] identifies the need for introducing uncertainty-based criteria for combining evidence from different experts:
- Minimum uncertainty criterion, where among alternative solutions the solution is selected that minimises uncertainty.
- Maximum uncertainty criterion, where all information available is taken into account such that it is certain that the real solution is within the uncertainty bounds evaluated.
- Uncertainty invariance, where all uncertainties are described in terms of the same scale and units, such that they can be combined.

[Cacuci, 2003]
[Cacuci, 2003] defines sensitivity of a measuring instrument as the change of response of a measuring instrument divided by the corresponding change in the stimulus. Uncertainty of measurement is defined as an interval within which the true value of a measured quantity would lie with a given probability. Uncertainty is defined with its limits and corresponding confidence probability, and can be expressed in absolute or relative form.

The author makes a distinction between local sensitivity analysis and global sensitivity analysis. The objective of local sensitivity analysis is to analyse the behaviour of the system responses locally around a chosen point or trajectory in the combined phase space of parameters and state variables. The objective of global sensitivity analysis is to determine all of the system's critical points in the state space formed by the parameters, state variables, and adjoint variables, and subsequently to analyse these critical points by local sensitivity analysis.

Next, [Cacuci, 2003] studies, in a mathematical formalism, local sensitivity and uncertainty analysis of both linear (Chapter IV) and non-linear (Chapter V) systems. The non-linear systems include operator responses and feedback. The sensitivities are obtained by calculating the first Gâteaux-differential of the system's response at the nominal value of the system's dependent variables and parameters. Two procedures are developed for calculating the sensitivities: The Forward Sensitivity Analysis Procedure (FSAP) and the Adjoint Sensitivity Analysis Procedure (ASAP). FSAP is concluded to be easier to develop and to implement than ASAP, but is only advantageous to employ if the number of different responses of interest exceeds the number of system parameters. For this reason, ASAP is usually the more practical choice.

[Ferson et al., 2003]
[Ferson et al., 2003] studies methods to incorporate Epistemic uncertainty in risk analysis. This is the type of uncertainty that results from the lack of knowledge about a system; alternative terms are Subjective uncertainty, or Ignorance. The development is focused on Dempster-Shafer evidential reasoning for unknown model parameter values only.

2.3 Bias and uncertainty assessment by [Everdij and Blom, 2002]

[Everdij and Blom, 2002] developed a method for assessing the bias and uncertainty in accident risk. This method considers all model assumptions adopted, including assumptions on the parameter values used in the model, and describes the effect of bias and uncertainty in these model assumptions propagating through the model, resulting in a bias and uncertainty assessment for accident risk. The mathematical theory behind the method uses 9 assumptions, named Bias and Uncertainty Model Assumptions. The method has been tested in many projects since. Further improvements have been studied in [Nurdin, 2002] for the parameter assumptions part and in [Stroeve et al, 2003] for the non-parameter assumptions part. A brief outline of the method is given below. The mathematical background is presented in Section 3.

The Bias and Uncertainty Assessment Method assesses the bias and uncertainty in model-based accident risk, with respect to realistic accident risk. It follows several steps:

1. Identify all model assumptions adopted. Usually, model assumptions of various types exist, such as parameter value assumptions, numerical approximation assumptions, model structure assumptions, assumptions due to non-coverage of identified hazards, etc. The first type (parameter value assumptions) is treated in step 2, the other types (together named non-parameter value assumptions) are treated in step 3.
2. Assess each model parameter value[2] on two aspects: 95% credibility interval for the real parameter value; and Risk log-sensitivity, expressed by the factor by which risk changes if the parameter value is changed by some normalised factor. From these assessments, a particular mathematical formula is used to find a 95% credibility interval around model-based accident risk, due to biases and uncertainties in the model parameter values.
3. Assess each non-parameter value assumption[3] separately on two aspects:
   - Did its introduction increase model-based risk with respect to realistic risk (i.e. is it a pessimistic assumption) or did it decrease model-based risk (i.e. is it an optimistic assumption)
   - By what factor did it increase or decrease risk. This factor is to be taken relative to all factors for assumptions already assessed.

   Both aspects are generally to be judged by operational experts. Next, model-based accident risk is compensated for all non-parameter value assumptions adopted, by using the assessed factors one by one to increase or decrease model-based accident risk. For example, if the first assumption was judged to be pessimistic by a factor 2, then model-based risk is divided by a factor 2 to compensate for this assumption (so that it comes closer to realistic risk). If the second assumption was judged to be pessimistic by a factor 1.5, taking account of the factor for the first assumption, then model-based risk is divided by an additional factor 1.5 to compensate for this second assumption.
4. The output of steps 2 and 3 are combined to obtain a 95% credibility interval for realistic accident risk, based on the model-based risk value, the non-parameter value assumption assessments and the parameter value assessments.

To save expensive computational time, steps 2 and 3 can be performed through qualitative assessments first (i.e. in terms of e.g. negligible, small, minor, significant, considerable, major), after which the most influential assumptions are re-assessed quantitatively.

---

[2] Some (hypothetical) examples of parameter value assumptions are: “Average airspeed of the aircraft at runway threshold equals 140 kts”, or “Average pilot reaction time in response to an emergency equals 2 seconds”.

[3] Some (hypothetical) examples of non-parameter value assumptions are: “Turbulence has no effect on collision risk”, or “The conflict avoidance instruction is not read back by the aircrew”.

This bias and uncertainty assessment approach compensates for bias and uncertainty in the model assumptions by multiplying model-based accident risk with compensation factors. Uncertainties in the parameter values are modelled by Lognormal variables[4], and propagation of these uncertainties through the model makes realistic accident risk (i.e. model-based risk compensated for the model assumptions adopted) also lognormal. And this yields that its expectation and 95% credibility interval can easily be assessed provided that the parameters of the lognormal distribution can be properly assessed. The lognormal distribution is a logical choice in case of multiplicative compensation of assumptions since the product (and the quotient) of two lognormal variables is again lognormal (compare: the sum (and the difference) of two Gaussian variables is again Gaussian).

2.4 Comparison with literature sources

The current subsection outlines how the techniques of [Everdij and Blom, 2002] compare to related material in the references of Subsection 2.2:

- [Morgan and Henrion, 1990] propose a few analytic techniques for sensitivity analysis and uncertainty propagation that have similarities to the method developed in [Everdij and Blom, 2002], with in particular the technique that compensates for bias and uncertainty by multiplication of a compensating factor. Main difference is that [Morgan and Henrion, 1990] pose the technique without proof or assumptions adopted, while [Everdij and Blom, 2002] give both.
- [Kumamoto and Henley, 1996], who consider fault trees, propose to use lognormal random variables to describe uncertainties in component failures at AND gates. The lognormal distribution also plays an important role in the [Everdij and Blom, 2002] method. [Kumamoto and Henley, 1996] explain that at OR gates the lognormal distribution is less appropriate since the distribution function of a sum of lognormals is unknown.
- [Haimes, 1998] mentions three types of bias and uncertainty, i.e. Parameter bias or uncertainty, Model bias or uncertainty, and Decision bias or uncertainty. The first two are considered by the [Everdij and Blom, 2002] method; the latter falls outside the scope of accident risk modelling.
- The other references in Subsection 2.2 do not present material directly related to the techniques in [Everdij and Blom, 2002].

The current HYBRIDGE study refines the results of [Everdij and Blom, 2002], and develops an extension to mitigate to a practically useful extent the OR gate problem mentioned by [Kumamoto and Henley, 1996].

[4] A random variable X is Lognormal if it can be written as $X = \exp(Z)$, or $Z = \ln X$, with Z Gaussian.

3. Factorisation model for bias and uncertainty in accident risk

This section presents the bias and uncertainty factorisation model as developed in [Everdij and Blom, 2002, Appendix III and IV], with some refinements by [Nurdin, 2002] added. The section gives a mathematical characterisation of how the model-evaluated expected accident risk for a particular application relates to the actual expected accident risk for that application.

The idea behind the characterisation and the set-up of this section is as follows:

The expected accident risk for a particular application is generally assessed by first developing a model, and next evaluating the accident risk based on this model. During the model development, always various assumptions are adopted, which may or may not be exactly according to reality. Therefore, model-based expected accident risk depends on the particular choice of assumptions, say $\overline{x}$. In particular, if we represent the assumptions by a random variable, say X, and regard accident risk as a function of this random variable, say $\rho(X)$, then the actual accident risk is equal to the expectation of the function, i.e. $E\{\rho(X)\}$, and the model-based accident risk is equal to the expectation of the function, conditional on the 'event' that the random variable is equal to the chosen model assumptions, i.e. $E\{\rho(X) \mid X = \overline{x}\} = \rho(\overline{x})$.

In this section (Subsection 3.6) we give a characterisation of actual accident risk in terms of model-based accident risk. The development of the characterisation is in three steps (Subsections 3.4 through 3.6) and uses several theorems, lemmas and corollaries. The proofs of these are given in Appendix A. Sometimes, these proofs depend on additional assumptions. To distinguish these additional assumptions from the risk model assumptions, we refer to these as *Bias and uncertainty model assumptions*. Some of the Bias and uncertainty model assumptions adopted may be quite reasonable, while others may be rather restrictive. This is discussed and evaluated in Section 4. In order to allow maximum flexibility in the use of the bias and uncertainty model, the current section will develop the same characterisation under four different subsets of Bias and uncertainty model assumptions. This way, in the end we can choose a subset of Bias and uncertainty model assumptions which appear to be least restrictive for a particular application, and still make use of the same characterisation of actual accident risk. The last table in this section gives an overview of all Bias and uncertainty model assumptions adopted, and shows which four subsets can be used to prove the characterisation of actual accident risk.

3.1 Notations

Throughout this section, we will use the following notational conventions, see e.g. [Mortensen, 1987]:

- Random variables are denoted by capital letters, for example X.
- Realisations or observations of random variables are denoted by the corresponding lowercase letters, for example x.
- The probability density function of a random variable X in point x is denoted by $p_X(x)$.
- The joint probability density function of two random variables X_1 in point x_1 and X_2 in point x_2 is denoted by $p_{X_1,X_2}(x_1, x_2)$.

- The conditional probability density function of random variable X_1 in point x_1 conditional on $X_2 = x_2$ is denoted by $p_{X_1|X_2}(x_1 | x_2)$.
- The expectation of a random variable X is denoted by $E\{X\}$.
- The conditional expectation of a random variable X, conditional on the observation $Y = y$ is denoted by $E\{X | Y = y\}$.
- Noting that $E\{X | Y = y\}$ is not a random variable, but is a function of the observed value y, i.e. $E\{X | Y = y\} = \psi(y)$, we also wish to discuss the random variable that is created by this operation without positing a specific outcome for our observation. In that case, we simply write $E\{X | Y\}$. Therefore, $E\{X | Y = y\}$ is not a random object, but $E\{X | Y\}$ is. Specifically, it will be a function of the random variable Y, that is $\psi(Y) = E\{X | Y\}$. See [Mortensen, 1987, page 44].

3.2 Problem definition

For a particular application (e.g. a particular old or newly developed air traffic management procedure), define:

- $\Re_{ACTUAL}$ as the expected actual number of collisions between aircraft per aircraft flight hour for the application.

Next, consider a stochastic model that can be used to estimate $\Re_{ACTUAL}$. During the development (or instantiation) of this stochastic model, assumptions have been made, and the model parameters have been given a value. To formalise this, define:

- $V = \{V_1, \ldots, V_{n_p}\}$ as the random variables for the set of parameters used in the stochastic model, where n_p is the number of parameters used,
- $\overline{v} = \{\overline{v}_1, \ldots, \overline{v}_{n_p}\}$ as the values that the respective parameters have been given in the stochastic model,
- $\overline{a} = \{\overline{a}_1, \ldots, \overline{a}_{n_a}\}$ as the set of non-parameter assumptions adopted during the instantiation of the stochastic model, where n_a is the number of non-parameter assumptions adopted,
- $A = (A_1, \ldots, A_{n_a})$ as a Boolean valued random vector of length n_a, where $A_i = 1$ if non-parameter assumption $\overline{a}_i$ holds true in the actual application and $A_i = 0$ if model assumption $\overline{a}_i$ does not hold true,
- $\mathbf{1} \triangleq (1, \ldots, 1)$ as a vector of length n_a, denoting the event that all non-parameter assumptions adopted during the instantiation of the stochastic model hold true in the actual operation.

With this, define:

- $\Re_{MODEL}$ as the model-based expected number of collisions between aircraft per aircraft flight hour, which is the result of the evaluation of the stochastic model instantiated for the particular application.

$\Re_{MODEL}$ is not necessarily equal to $\Re_{ACTUAL}$ since the non-parameter assumptions adopted for the evaluation of $\Re_{MODEL}$ and the choice for the parameter values may have created a bias away from $\Re_{ACTUAL}$. In addition, there may be uncertainty about the true bias. To make explicit that $\Re_{MODEL}$ depends on the parameter assumptions and on the non-parameter assumptions adopted, we define:

- A function $\rho(\alpha, v) \triangleq$ Expected number of collisions between aircraft per aircraft flight hour under non-parameter assumptions α and parameter assumptions v.

Then,

- $\Re_{MODEL} \triangleq \rho(1, \bar{v}) = E\{\rho(A,V) \mid A = 1, V = \bar{v}\}$.

With this, it follows that

- $\Re_{ACTUAL} = E\{\rho(A,V)\}$.

Then, the problem considered in this section is the following:

Problem definition:
Characterise $\Re_{ACTUAL} = E\{\rho(A,V)\}$ in terms of $\Re_{MODEL} = \rho(1,\bar{v})$ and properties of the stochastic model.

♦

In the remainder of this section, this problem is addressed as follows: First, Subsection 3.3 defines a term that is important to consider when studying changes in risk due to changes in parameter values, i.e. Log-Sensitivity. Next, Subsection 3.4 characterises $E\{\rho(1,V)\}$ in terms of $\Re_{MODEL} = \rho(1,\bar{v})$ and properties of V and properties of the model. Subsection 3.5 characterises $\Re_{ACTUAL}$ in terms of $E\{\rho(1,V)\}$ and properties of the non-parameter assumptions. Finally, Subsection 3.6 combines the results. The characterisation of $\Re_{ACTUAL}$ uses a few assumptions, which are named *Bias and uncertainty model assumptions*, to distinguish them from the formal model assumptions $(\bar{a}, \bar{v})$.

3.3 Log-sensitivity of accident risk

In [Morgan and Henrion, 1990] (page 174), sensitivity is defined as the rate of change of the output with respect to variation in an input. If we consider model-based risk as the output and the parameter values as the input, then in our case we have:

$$\text{Sensitivity}_i(v) \triangleq \frac{\partial \rho(1,v)}{\partial v_i}.$$

[Morgan and Henrion, 1990] also introduce the term Elasticity, which is a normalised sensitivity. In our case:

$$\text{Elasticity}_i(v) \triangleq \frac{\partial \rho(1,v)}{\partial v_i} \times \frac{v_i}{\rho(1,v)}.$$

Since accident risks come in very small values, we prefer to work in a natural logarithm domain, and we will define Log-Sensitivity to model-based risk of the i^{th} parameter value as follows:

$$\text{Log-Sensitivity}_i(v) \triangleq \frac{\partial \ln \rho(1,v)}{\partial \ln v_i}.$$

Note that

$$\frac{\partial \ln \rho(1,v)}{\partial \ln v_i} = \frac{1}{\rho(1,v)}\frac{\partial \rho(1,v)}{\partial \ln v_i} = \frac{1}{\rho(1,v)}\frac{\partial \rho(1,v)}{\frac{1}{v_i}\partial v_i} = \frac{v_i}{\rho(1,v)}\frac{\partial \rho(1,v)}{\partial v_i} = \text{Elasticity}_i(v).$$

Therefore, Log-Sensitivity and Elasticity are the same. We will continue to use the term Log-Sensitivity in favour of Elasticity, since mathematically it better captures the meaning.

Also note that the log-sensitivity is a function of v. In approximations, it is useful to consider it locally to be a constant function. And in particular, we will also study the log-sensitivity in the fixed point $v = \overline{v}$, i.e.

$$\beta_i \triangleq \text{Log-Sensitivity}_i(\overline{v}) = \left.\frac{\partial \ln \rho(1,v)}{\partial \ln v_i}\right|_{\ln v = \ln \overline{v}}.$$

3.4 Characterisation of $\rho(1,V)$ in terms of $\Re_{MODEL}$

In this subsection, $\rho(1,V)$ is characterised in terms of $\Re_{MODEL} = \rho(1,\overline{v})$. All proofs of theorems, lemmas and corollaries are given in Appendix A (most are from [Everdij and Blom, 2002], some are from [Nurdin, 2002] and some are new).

Definition 3.1:
For each $i = 1,\dots,n_p$, define a function φ_i by:

$$\varphi_i(x_i) \triangleq \frac{\rho(1,(\overline{v}_1,\dots,\overline{v}_{i-1},x_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))}{\rho(1,(\overline{v}_1,\dots,\overline{v}_{i-1},\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))}$$

♦

Definition 3.2:
For each $i = 1,\dots,n_p$, define a function φ_i ' by:

$$\varphi_i'(x_i, x_{i+1},\dots,x_{n_p}) \triangleq \frac{\rho(1,(\overline{v}_1,\dots,\overline{v}_{i-1},x_i,x_{i+1},\dots,x_{n_p}))}{\rho(1,(\overline{v}_1,\dots,\overline{v}_{i-1},\overline{v}_i,x_{i+1},\dots,x_{n_p}))}$$ ♦

Bias and uncertainty model assumption 3.1:
For any $i = 1,\dots,n_p$ and $x_i,\dots,x_{n_p}$, $\varphi_i(x_i) = \varphi_i'(x_i, x_{i+1},\dots,x_{n_p})$. ♦

Theorem 3.1:
Bias and uncertainty model assumption 3.1 holds true for all $v \in \mathbb{R}^{n_p}$ if and only if:

$$\rho(1,v) = \rho(1,\overline{v}) \times \prod_{i=1}^{n_p} \varphi_i(v_i) \tag{3.1}$$

♦

Theorem 3.1 yields that under Bias and uncertainty model assumption 3.1, for all $v \in \mathbb{R}^{n_p}$, $\rho(1,v)$ can be written as $\Re_{MODEL} = \rho(1,\overline{v})$ times the product of all $\varphi_i(v_i)$. Hence, in particular, the random variable $\rho(1,V)$ can be written as $\rho(1,V) = \rho(1,\overline{v}) \times \prod_{i=1}^{n_p} \varphi_i(V_i)$. In the remainder of this subsection, first (in Theorems 3.2, 3.3, 3.4 and 3.5), we characterise $\prod_{i=1}^{n_p} \varphi_i(V_i)$ by means of a probability distribution, which holds true under four different sets of Bias and uncertainty model assumptions; next (in Corollary 3.1), we combine those results with the result of Theorem 3.1 to characterise $\rho(\mathbf{1},V)$.

We start with the following bias and uncertainty model assumption:

Bias and uncertainty model assumption 3.2:
For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\varphi_i(v_i)=(v_i/\overline{v}_i)^{\beta_i}$, with β_i the log-sensitivity around $\overline{v}$ as defined in Subsection 3.3.

♦

Bias and uncertainty model assumptions 3.2 poses a direct form for $\varphi_i(v_i)$. Using this form, under a few more bias and uncertainty model assumptions on a probability distribution for V_i, we can directly determine a probability distribution for $\varphi_i(V_i)$:

Bias and uncertainty model assumption 3.3:
The random variables $V_1,\ldots,V_{n_p}$ are mutually independent. ♦

Bias and uncertainty model assumption 3.4:
For each $i=1,\ldots,n_p$, the expectation and variance of $\ln V_i$ exist and satisfy:
$E\{\ln V_i\}=\mu_i$ and $\mathrm{Var}\{\ln V_i\}=\sigma_i^2$. ♦

Bias and uncertainty model assumption 3.5:
Each V_i $(i=1,\ldots,n_p)$ is lognormally distributed. ♦
See Appendix B for some properties of the lognormal distribution. We note that if Bias and uncertainty model assumptions 3.4 and 3.5 both hold true, then V_i is lognormally distributed with parameters μ_i and σ_i^2. Notation: $V_i \sim \Lambda(\mu_i,\sigma_i^2)$.

Theorem 3.2:
Under Bias and uncertainty model assumptions 3.2, 3.3, 3.4, 3.5, $\prod_{i=1}^{n_p}\varphi_i(V_i)$ is lognormally distributed with parameters $\ln\widehat{B}$ and $\frac{1}{4}\widehat{U}$, i.e. $\prod_{i=1}^{n_p}\varphi_i(V_i)\sim\Lambda(\ln\widehat{B},\frac{1}{4}\widehat{U})$, with $\widehat{B}$ and $\widehat{U}$ as defined in Definition 3.3 below. ♦

Definition 3.3:

- $\widehat{B}\triangleq\prod_{i=1}^{n_p} b_i^{\beta_i}$
- $b_i\triangleq\exp(\mu_i)/\overline{v}_i$
- $\widehat{U}\triangleq\sum_{i=1}^{n_p}(\ln\ell_i^{|\beta_i|})^2$
- $\ell_i\triangleq\exp(2\sigma_i)$ ♦

We refer to b_i as the bias of the i^{th} parameter value, to ℓ_i as its uncertainty.

Under Bias and uncertainty model assumption 3.2, a direct form for $\varphi_i(v_i)$ is assumed, i.e. $\varphi_i(v_i) = (v_i / \overline{v}_i)^{\beta_i}$. Since this assumption might be restrictive, we will try to find the distribution for $\prod_{i=1}^{n_p} \varphi_i(V_i)$ also under some other bias and uncertainty model assumptions, which may (or may not) be less restrictive.

Lemma 3.1:
Under Bias and uncertainty model assumptions 3.3 and the assumption that $\varphi_i(\cdot)$ are Borel measurable, the random variables $\varphi_1(V_1),\ldots,\varphi_{n_p}(V_{n_p})$ are mutually independent.

♦

Bias and uncertainty model assumption 3.6:
The inverse of $\varphi_i(\cdot)$, denoted by $\varphi_i^{inv}(\cdot)$, exists for all *i*. ♦

Note that if Bias and uncertainty model assumption 3.6 holds, each $\varphi_i(\cdot)$ is Borel measurable.

Bias and uncertainty model assumption 3.7:
Each $\varphi_i(V_i)$ $(i = 1,\ldots,n_p)$ is lognormally distributed with parameters γ_i and θ_i^2.
Notation: $\varphi_i(V_i) \sim \Lambda(\gamma_i, \theta_i^2)$. ♦

Bias and uncertainty model assumption 3.8:
For $i = 1,\ldots,n_p$, the following two equations are satisfied
$E\{\ln \varphi_i(V_i)\} = \beta_i (E\{\ln V_i\} - \ln \overline{v}_i)$, and $E\{(\ln \varphi_i(V_i) - E\{\ln \varphi_i(V_i)\})^2\} = \beta_i^2 \mathrm{Var}\{\ln V_i\}$,
with β_i the log-sensitivity around $\overline{v}$ as defined in Subsection 3.3. ♦

Lemma 3.2:
Under Bias and uncertainty model assumptions 3.4, 3.5, 3.6, 3.7 and 3.8, for all $v \in \mathbb{R}^{n_p}$, $\varphi_i(v_i) = (v_i / \overline{v}_i)^{\beta_i}$ for $i = 1,\ldots,n_p$.

♦

Theorem 3.3:
Under Bias and uncertainty model assumptions 3.3, 3.4, 3.5, 3.6, 3.7 and 3.8, $\prod_{i=1}^{n_p} \varphi_i(V_i)$ is lognormally distributed with parameters $\ln \widehat{B}$ and $\frac{1}{4}\widehat{U}$, i.e. $\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\ln \widehat{B}, \frac{1}{4}\widehat{U})$. ♦

In Theorems 3.2 and 3.3, lognormal assumptions are used for V_i. In the remainder of this subsection, we make use of the Central Limit Theorem to relax these assumptions. First, introduce some definitions:

Definition 3.4:
- $H_i \triangleq \ln \varphi_i(V_i)$,
- $\gamma_i \triangleq E\{H_i\}$,

- $\Gamma \triangleq \sum_{i=1}^{n_p} \gamma_i$,
- $\theta_{i,n} \triangleq E\{| H_i - \gamma_i |^n\},\ n = 2,3,$
- $\Theta_n \triangleq \sqrt[n]{\sum_{i=1}^{n_p} \theta_{i,n}}$, $n = 2,3.$

♦

Subsequently, introduce Bias and uncertainty model assumptions 3.9 and 3.10:

Bias and uncertainty model assumption 3.9:
With Θ_n according to Definition 3.4, $\lim_{n_p \to \infty} \Theta_3 / \Theta_2 \to 0$. ♦

Bias and uncertainty model assumption 3.10:
The number of parameters n_p goes to infinity: $n_p \to \infty$ ♦

Then the following lemma can be proven:

Lemma 3.3:
Under Bias and uncertainty model assumptions 3.3, 3.6, 3.9 and 3.10, $\prod_{i=1}^{n_p} \varphi_i(V_i)$ is lognormally distributed with parameters Γ and Θ_2^2, i.e. $\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\Gamma, \Theta_2^2)$. ♦

The practical applicability of Lemma 3.3 depends on the feasibility of determining the parameters, Γ and Θ_2^2, and on the feasibility of the set of Bias and uncertainty model assumptions {3.3, 3.6, 3.9, 3.10}. This is discussed in Appendix IV of [Everdij and Blom, 2002]. There, no feasible way is found to evaluate Γ and Θ_2^2 in practise. Therefore, the result of Lemma 3.3 is not of direct practical use. Theorems 3.4 and 3.5 below find a way out:

Theorem 3.4:
Under Bias and uncertainty model assumptions 3.2, 3.3, 3.4, 3.9, 3.10, $\prod_{i=1}^{n_p} \varphi_i(V_i)$ is lognormally distributed with parameters $\ln \widehat{B}$ and $\frac{1}{4}\widehat{U}$, i.e. $\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\ln \widehat{B}, \frac{1}{4}\widehat{U})$. ♦

Theorem 3.5:
Under Bias and uncertainty model assumptions 3.3, 3.4, 3.6, 3.8, 3.9, 3.10, $\prod_{i=1}^{n_p} \varphi_i(V_i)$ is lognormally distributed with parameters $\ln \widehat{B}$ and $\frac{1}{4}\widehat{U}$, i.e. $\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\ln \widehat{B}, \frac{1}{4}\widehat{U})$. ♦

Note that the characterisations of Theorems 3.4 and 3.5 do not use Bias and uncertainty model assumptions 3.5 and 3.7, i.e. the assumptions that V_i and $\varphi_i(V_i)$ are both lognormal.

Finally, notice that we showed (in Theorems 3.2, 3.3, 3.4, and 3.5) that $\prod_{i=1}^{n_p}\varphi_i(V_i) \sim \Lambda(\ln \widehat{B}, \frac{1}{4}\widehat{U})$ holds true under four different sets of bias and uncertainty model assumptions. This result is combined with that of Theorem 3.1, to obtain the following corollary:

Corollary 3.1
Under Bias and uncertainty model assumptions {3.1, 3.2, 3.3, 3.4, 3.5}, or under {3.1, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8}, or under {3.1, 3.2, 3.3, 3.4, 3.9, 3.10}, or under {3.1, 3.3, 3.4, 3.6, 3.8, 3.9, 3.10}:

$$E\{\rho(\mathbf{1},V)\} = \rho(\mathbf{1},\overline{v}) \times \widehat{B} \times \exp(\frac{1}{8}\widehat{U}) \text{, and} \tag{3.2}$$

$$\Pr(\rho(\mathbf{1},V) \in \left[\rho(\mathbf{1},\overline{v}) \times \widehat{B} \times \exp(-\sqrt{\widehat{U}}), \rho(\mathbf{1},\overline{v}) \times \widehat{B} \times \exp(\sqrt{\widehat{U}})\right]) = 0.95 \tag{3.3}$$

♦

3.5 Relation between $E\{\rho(\mathbf{1},V)\}$ *and* $\Re_{ACTUAL}$

In this subsection, $\Re_{ACTUAL} = E\{\rho(A,V)\}$ is characterised in terms of $E\{\rho(\mathbf{1},V)\}$.

Definition 3.5:

$$\Psi \triangleq \frac{\Re_{ACTUAL}}{E\{\rho(\mathbf{1},V)\}}$$

♦

With Definition 3.5, we can write $\Re_{ACTUAL} = E\{\rho(\mathbf{1},V)\} \times \Psi$.

Next, the following theorem can be proven:

Theorem 3.6:

$$\Psi = \prod_{i=1}^{n_a}\left\{\Pr(A_i = 0 \mid A_1 = 1,\ldots,A_{i-1} = 1) \times \frac{E\{\rho(A,V) \mid A_1 = 1,\ldots,A_{i-1} = 1, A_i = 0\}}{E\{\rho(A,V) \mid A_1 = 1,\ldots,A_i = 1\}} + \right.$$
$$\left. + \Pr(A_i = 1 \mid A_1 = 1,\ldots,A_{i-1} = 1)\right\} \tag{3.4}$$

where the set $\{A_1,\ldots,A_{i-1}\}$ is empty if $i = 1$. ♦

Remark: Note that Equation (3.4) can be written in a shorthand notation as follows:

$$\Psi = \prod_{i=1}^{n_a}\left\{P_i Q_i + (1 - P_i)\right\},$$

with

- $P_i \triangleq \Pr(A_i = 0 \mid A_1 = 1,\ldots,A_{i-1} = 1)$, i.e. the probability of the non-applicability of assumption *i*, conditional on assumptions 1 through $i-1$ holding true;
- $Q_i \triangleq \frac{E\{\rho(A,V) \mid A_1 = 1,\ldots,A_{i-1} = 1, A_i = 0\}}{E\{\rho(A,V) \mid A_1 = 1,\ldots,A_i = 1\}}$, i.e. the effect on the risk bias if assumption *i* does not apply, conditional on assumptions 1 through $i-1$ holding true.

3.6 Results combined

In this subsection, we combine the results obtained in the previous two subsections. Use Definition 3.5 to write:

$$\Re_{ACTUAL} = E\{\rho(A,V)\} = E\{\rho(1,V)\} \times \Psi$$

and notice that the first factor on the right hand side of the above equation has been evaluated in Subsection 3.4, the second factor has been evaluated in Subsection 3.5. Combining these results yields:

Under Bias and uncertainty model assumptions {3.1, 3.2, 3.3, 3.4, 3.5}, or under {3.1, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8}, or under {3.1, 3.2, 3.3, 3.4, 3.9, 3.10}, or under {3.1, 3.3, 3.4, 3.6, 3.8, 3.9, 3.10}:

$$\Re_{ACTUAL} = E\{\rho(A,V)\} = \Psi \times \rho(1,\overline{v}) \times \widehat{B} \times \exp(\frac{1}{8}\widehat{U}) \text{, and} \tag{3.5}$$

$$\Pr(\Psi \times \rho(1,V) \in \left[\Psi \times \rho(1,\overline{v}) \times \widehat{B} \times \exp(-\sqrt{\widehat{U}}), \Psi \times \rho(1,\overline{v}) \times \widehat{B} \times \exp(\sqrt{\widehat{U}})\right]) = 0.95. \tag{3.6}$$

For easier reference, the Bias and uncertainty model assumptions used are repeated below in one table. The last four columns indicate which sets of Bias and uncertainty model assumptions are required to hold true (grey field) for Equations 3.5 and 3.6 to hold true. To be more specific: Equations 3.5 and 3.6 hold true under the set of Bias and uncertainty model assumptions {3.1, 3.2, 3.3, 3.4, 3.5} (these assumptions have a grey box in the column headed by “A”) or under Bias and uncertainty model assumptions {3.1, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8} (see the “B” column), or under Bias and uncertainty model assumptions {3.1, 3.2, 3.3, 3.4, 3.9, 3.10} (see the “C” column) or under Bias and uncertainty model assumptions {3.1, 3.3, 3.4, 3.6, 3.8, 3.9, 3.10} (see the “D” column).

Bias and uncertainty model assumption		**A**	**B**	**C**	**D**
3.1	For any $i=1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\varphi_i(x_i)=\varphi_i^{*}(x_i,x_{i+1},\ldots,x_{n_p})$.				
3.2	For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\varphi_i(v_i)=(v_i/\overline{v}_i)^{\beta_i}$, with β_i the log-sensitivity around $\overline{v}$ as defined in Subsection 3.3.				
3.3	The random variables $V_1,\ldots,V_{n_p}$ are mutually independent.				
3.4	For each $i=1,\ldots,n_p$, the expectation and variance of $\ln V_i$ exist and satisfy: $E\{\ln V_i\}=\mu_i$ and $\mathrm{Var}\{\ln V_i\}=\sigma_i^2$.				
3.5	Each V_i $(i=1,\ldots,n_p)$ is lognormally distributed.				
3.6	The inverse of $\varphi_i(\cdot)$, denoted by $\varphi_i^{inv}(\cdot)$, exists for all *i*.				
3.7	Each $\varphi_i(V_i)$ $(i=1,\ldots,n_p)$ is lognormally distributed with parameters γ_i and θ_i^2. Notation: $\varphi_i(V_i)\sim\Lambda(\gamma_i,\theta_i^2)$				
3.8	For $i=1,\ldots,n_p$, the following two equations are satisfied $E\{\ln\varphi_i(V_i)\}=\beta_i(E\{\ln V_i\}-\ln\overline{v}_i)$, and $E\{(\ln\varphi_i(V_i)-E\{\ln\varphi_i(V_i)\})^2\}=\beta_i^2\mathrm{Var}\{\ln V_i\}$, with β_i the log-sensitivity around $\overline{v}$ as defined in Subsection 3.3.				
3.9	With Θ_n according to Definition 3.4, $\lim_{n_p\to\infty}\Theta_3/\Theta_2\to 0$.				
3.10	The number of parameters n_p goes to infinity: $n_p\to\infty$.				

A: Grey fields indicate first set of Bias and uncertainty model assumptions under which Equations 3.5 and 3.6 hold true, i.e. {3.1, 3.2, 3.3, 3.4, 3.5};
B: Grey fields indicate second set of Bias and uncertainty model assumptions under which Equations 3.5 and 3.6 hold true, i.e. {3.1, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8};
C: Grey fields indicate third set of Bias and uncertainty model assumptions under which Equations 3.5 and 3.6 hold true, i.e. {3.1, 3.2, 3.3, 3.4, 3.9, 3.10};
D: Grey fields indicate fourth set of Bias and uncertainty model assumptions under which Equations 3.5 and 3.6 hold true, i.e. {3.1, 3.3, 3.4, 3.6, 3.8, 3.9, 3.10};

In Section 4, Bias and uncertainty model assumptions 3.1 through 3.10 are discussed on feasibility and practical applications.

4. Evaluation of Bias and uncertainty model assumptions 3.1 through 3.10

This section discusses and evaluates the bias and uncertainty model assumptions adopted in Section 3. Note that for Bias and uncertainty model assumptions 3.1 and 3.3 through 3.10, this was also done in Appendix IV of [Everdij and Blom, 2002]; an update of that discussion is included below in Subsection 4.1. Subsection 4.2 gathers the results of these evaluations and gives concluding remarks. Finally, Subsection 4.3 gives a set-up for the next step in the further development of the bias and uncertainty assessment modelling. This step is made in Section 5.

4.1 Evaluation of Bias and uncertainty model assumptions 3.1 through 3.10

This subsection discusses and evaluates Bias and uncertainty model assumptions 3.1 through 3.10. Based on its discussion, each bias and uncertainty model assumption is judged on its feasibility by a mark “0”, “–” or “– –”, where,

“0” denotes that the assumption usually (almost) holds true;
“–” denotes that the assumption may not always hold true, but is a logical choice and approximations may exist to reduce its effect;
“– –” denotes that the assumption is very restrictive.

Evaluation of Bias and uncertainty model assumption 3.1

3.1	For any $i = 1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\varphi_i(x_i) = \varphi_i^*(x_i, x_{i+1},\ldots,x_{n_p})$.

Bias and uncertainty model assumption 3.1 has great consequences. It states that for any $i = 1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\varphi_i(x_i) = \varphi_i^*(x_i, x_{i+1},\ldots,x_{n_p})$, where (from Definitions 3.1 and 3.2 respectively):

$$\varphi_i(x_i) \triangleq \frac{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},x_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},\overline{v}_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}, \text{ and } \varphi_i^*(x_i,x_{i+1},\ldots,x_{n_p}) \triangleq \frac{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},x_i,x_{i+1},\ldots,x_{n_p}))}{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},\overline{v}_i,x_{i+1},\ldots,x_{n_p}))}.$$

To catch the meaning of this assumption, let us first consider the case that it does not hold true. By using Definition 3.2 repeatedly, we obtain the exact expression $\rho(1,v) = \rho(1,\overline{v}) \times \prod_{i=1}^{n_p} \varphi_i^*(v_i, v_{i+1},\ldots,v_{n_p})$. This can be interpreted as follows: $\rho(1,v)$ is written as $\Re_{MODEL} = \rho(1,\overline{v})$ (i.e. model-based accident risk), which is next for each parameter multiplied by a *bias and uncertainty compensation factor* $\varphi_i^*(v_i, v_{i+1},\ldots,v_{n_p})$. This factor, which compensates for biases and uncertainties in the i^{th} parameter value, i.e. which compensates for variations and deviations of v_i away from $\overline{v}_i$, is not influenced by variations in the values of parameters 1 through *i*-1, but can be influenced by variations in the values of parameters $i+1$ through n_p.

Next, consider Bias and uncertainty model assumption 3.1. Under this assumption, $\varphi_i^*(v_i, v_{i+1},\ldots,v_{n_p})$ is replaced by $\varphi_i(v_i)$. This means that the compensation factor for bias and

uncertainty in the i^{th} parameter is assumed to be no longer influenced by the variations in values given to the other parameters.
We can analyse this further, by evaluating the risk log-sensitivity, which, by Subsection 3.3, is defined as Log-Sensitivity$_i(v) \triangleq \frac{\partial \ln \rho(1,v)}{\partial \ln v_i}$. Using the exact expression $\rho(1,v) = \rho(1,\bar{v}) \times \prod_{i=1}^{n_p} \varphi_i\,(v_i, v_{i+1}, \ldots, v_{n_p})$, this can be further evaluated as

$$\text{Log-Sensitivity}_i(v) = \frac{\partial \ln[\rho(1,\bar{v}) \times \prod_{k=1}^{n_p} \varphi_k\,(v_k, v_{k+1}, \ldots, v_{n_p})]}{\partial \ln v_i} =$$

$$= \frac{\partial[\ln \rho(1,\bar{v}) + \sum_{k=1}^{n_p} \ln \varphi_k\,(v_k, v_{k+1}, \ldots, v_{n_p})]}{\partial \ln v_i} = \sum_{k=1}^{i} \frac{\partial \ln \varphi_k\,(v_k, v_{k+1}, \ldots, v_{n_p})}{\partial \ln v_i}.$$

Under Bias and uncertainty model assumption 3.1, this reduces to Log-Sensitivity$_i(v) = \frac{\partial \ln \varphi_i(v_i)}{\partial \ln v_i}$. Hence, in particular, under Bias and uncertainty model assumption 3.1, the log-sensitivity to risk of the i^{th} parameter value is not influenced by variations in the values given to the other parameters.

Another observation can be made from this log-sensitivity evaluation: Although in principle, risk log-sensitivity, as defined through Log-Sensitivity$_i(v) \triangleq \frac{\partial \ln \rho(1,v)}{\partial \ln v_i}$, is invariant w.r.t the vector of nominal parameter values $\bar{v}$, under Bias and uncertainty model assumption 3.1, Log-Sensitivity$_i(v)$ evaluates to $\frac{\partial \ln \varphi_i(v_i)}{\partial \ln v_i}$, in which $\varphi_i(v_i) \triangleq \frac{\rho(1,(\bar{v}_1, \ldots, \bar{v}_{i-1}, v_i, \bar{v}_{i+1}, \ldots, \bar{v}_{n_p}))}{\rho(1,\bar{v})}$ is implicitly also a function of $\bar{v}$. Therefore, under Bias and uncertainty model assumption 3.1, Log-Sensitivity$_i(v)$ is also a function of $\bar{v}$.

With these interpretations and evaluations in mind, some objections to Bias and uncertainty model assumption 3.1 can be summarised as follows:

1. Under Bias and uncertainty model assumption 3.1, the risk compensation factor due to bias and uncertainty of the i^{th} parameter value is not influenced by values given to other parameters, and in particular: the log-sensitivity to risk of the i^{th} parameter value is not influenced by variations in values given to other parameters;
2. Under Bias and uncertainty model assumption 3.1, the log-sensitivity to risk of the i^{th} parameter value can be influenced by the nominal value $\bar{v}$.

In practical applications, it appeared, however, (for a few parameters) that the risk log-sensitivity of the i^{th} parameter could be highly influenced by the value given to the j^{th} parameter and could also be influenced by the choice for $\bar{v}_i$.

In practical applications, we have made use of a safety conservative approximation, for example as follows: Consider the exact expression $\rho(1,v) = \rho(1,\bar{v}) \times \prod_{i=1}^{n_p} \varphi_i\,(v_i, v_{i+1}, \ldots, v_{n_p})$. Now, assume that for all v_2 through v_{n_p} we can find a worst case value, denoted by v_i^w for the i^{th} parameter, which is defined such that for all $i = 1, \ldots, n_p$ and $x_i, \ldots, x_{n_p}$, $\varphi_i\,(x_i, x_{i+1}, \ldots, x_{n_p}) \le \varphi_i\,(x_i, v_{i+1}^w, \ldots, v_{n_p}^w)$. Then $\rho(1,v) \le \rho(1,\bar{v}) \times \prod_{i=1}^{n_p} \varphi_i\,(v_i, v_{i+1}^w, \ldots, v_{n_p}^w)$. Next, the remainder of Section 3 reads the same, but with $\varphi_i(v_i)$ replaced by $\varphi_i'(v_i, v_{i+1}^w \ldots, v_{n_p}^w)$ and each occurrence of '$\rho(1,v)=$' and '$E\{\rho(1,V)\}=$' replaced by '$\rho(1,v)\leq$' and '$E\{\rho(1,V)\}\leq$', respectively. This approach gives a worst case estimate for accident risk.

Conclusion:
In general, Bias and uncertainty model assumption 3.1 is judged as a hard requirement ("– –").

Evaluation of Bias and uncertainty model assumption 3.2

3.2	For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\varphi_i(v_i)=(v_i/\overline{v}_i)^{\beta_i}$, with β_i the log-sensitivity around $\overline{v}$ as defined in Subsection 3.3.

According to Definition 3.1, $\varphi_i(v_i)\triangleq\dfrac{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},\overline{v}_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}$, hence Bias and uncertainty model assumption 3.2 states that $\dfrac{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},\overline{v}_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}=\left(\dfrac{v_i}{\overline{v}_i}\right)^{\beta_i}$, or

$$\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))=\rho(1,\overline{v})\times\left(\frac{v_i}{\overline{v}_i}\right)^{\beta_i}$$ or

$$\ln\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))=\ln\rho(1,\overline{v})+\beta_i(\ln v_i-\ln\overline{v}_i).$$

It can be easily seen that this assumption therefore simply implies a linearisation of $\ln\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))$ around $\ln\overline{v}_i$. We can conclude that this is a reasonable approximation, as long as $\ln\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))$ is (approximately) linear around $\ln\overline{v}_i$.

Unfortunately, in applications, $\ln\rho(1,v)$ is generally not a linear function of $\ln v_i$. This was observed as follows: if $\ln\rho(1,v)$ is linear then $\dfrac{\partial\ln\rho(1,v)}{\partial\ln v_i}$ is a constant. In that case, $\dfrac{\partial\ln\rho(1,v)}{\partial\ln v_i}$ can be approximated by $\dfrac{\ln\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},v_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))-\ln\rho(1,\overline{v})}{\ln v_i-\ln\overline{v}_i}$, which is equal to β_i. In applications, we tried to estimate β_i by using a particular factor $l_i\neq 1$ and evaluating $\overline{\beta}_i(l_i)=\dfrac{\ln\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},(l_i\times\overline{v}_i),\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))-\ln\rho(1,\overline{v})}{\ln(l_i\times\overline{v}_i)-\ln\overline{v}_i}$. It appeared that $\overline{\beta}_i$ was not a constant function of l_i. Note that [Nurdin, 2002] (in his Section 4.10) proposed as an alternative estimate of β_i to take the average over $\overline{\beta}_i(l_i)$ for different l_i.

However, in many applications, it does appear that $\ln\rho(1,v)$ is approximately a *piecewise linear* function of $\ln v_i$. This property is used in Section 5, to improve the effect of Bias and uncertainty model assumption 3.2.

Conclusion:
If, for all i, $\ln\rho(1,v)$ is a linear function of $\ln v_i$, then Bias and uncertainty model assumption 3.2 is judged a reasonable approximation ("0"). In other cases, Bias and uncertainty model

assumption 3.2 may not be completely accurate: "–", or even "– –", depending on the behaviour of $\ln \rho(1, v)$. In practice, there often are some relevant parameters for which "– –" applies.

Evaluation of Bias and uncertainty model assumption 3.3

3.3	The random variables $V_1, \ldots, V_{n_p}$ are mutually independent.

Two random variables *X* and *Y* are said to be independent if and only if their probability density functions are related by $p_{X,Y}(x, y) = p_X(x) p_Y(y)$, where $p_X(x)$ and $p_Y(y)$ are the marginal density functions of the random variables *X* and *Y* respectively, for all pairs (*x*,*y*). Hence, knowledge about *X* does not effect the knowledge about *Y* and vice versa.

In Bias and uncertainty model assumption 3.3, V_i models the random variable for the i^{th} parameter used in the stochastic model for accident risk. The expectation of V_i gives a measure for the bias in the value that the i^{th} parameter takes on in the model. The variance of V_i gives a measure for the uncertainty about this value. Hence Bias and uncertainty model assumption 3.1 can be interpreted as stating that for all *i* and *j*, the bias and uncertainty of the i^{th} parameter value should be independent of the bias and uncertainty of the j^{th} parameter value.

In general, the expectation and variance are determined from statistical data and expert knowledge about the value for the parameter. If these data are independent, this assumption holds true, or can be made to hold true. For example, if the i^{th} parameter is a function of the j^{th} parameter, then simply replace this j^{th} parameter by the function value, and delete it from the list of parameters.

<u>Conclusion:</u>
Bias and uncertainty model assumption 3.3 is a reasonable assumption ("0") since many difficulties can be eliminated.

Evaluation of Bias and uncertainty model assumption 3.4

3.4	For each $i = 1, \ldots, n_p$, the expectation and variance of $\ln V_i$ exist and satisfy: $E\{\ln V_i\} = \mu_i$ and $\text{Var}\{\ln V_i\} = \sigma_i^2$.

Bias and uncertainty model assumption 3.4 states that the expectation and variance of the natural logarithm of each parameter should exist. The expectation of the random variable $\ln V_i$ is defined as $E\{\ln V_i\} = \int_0^\infty \ln v \cdot p_{V_i}(v) dv$. It exists if $|\ln v| \cdot p_{V_i}(v)$ is integrable. A sufficient condition for this is that V_i is strictly positive and finite. A similar condition holds for the variance. This condition usually holds true or can be made to hold true; for example: for a parameter V_i that may take on both positive and negative values, introduce two parameters $V_i^+ > 0$ and $V_i^- > 0$, with $V_i = V_i^+ - V_i^-$.

<u>Conclusion:</u>
Bias and uncertainty model assumption 3.4 is judged as a reasonable assumption ("0").

Evaluation of Bias and uncertainty model assumption 3.5

3.5	Each V_i $(i = 1,\ldots,n_p)$ is lognormally distributed.

In a stochastic model for accident risk, usually various parameters are used, for example, standard look-ahead time for conflict detection, average duration of radar sweep, average time until failure of aircraft system (where the time until failure is described by some probability distribution), etc. Proper values need to be assigned to these parameters before the stochastic model can be (analytically or numerically) evaluated. Since statistical data is usually difficult to find, there exists some uncertainty about these values. For this reason, the parameters are random variables themselves; in this document they are denoted by $V_1,\ldots,V_{n_p}$. In the stochastic model, V_i has been given value $\overline{v}_i$, but the data or the experts may disagree on the best value for $\overline{v}_i$. The value $\overline{v}_i$ may be known up to a certain factor, for example, the real value for V_i could also be a factor 5 higher or lower than $\overline{v}_i$.

To capture these types of uncertainty in the parameter values, the lognormal distribution is very useful because of its key property of being multiplication-symmetric, i.e. for Y a $\Lambda(0,\sigma^2)$-distributed variable with probability density function p_Y and c a constant: $p_Y(1/c) = p_Y(c)$. Therefore, if a parameter is known up to a certain factor higher or lower than a certain median value, the lognormal distribution is a useful candidate to describe it. In [Kumamoto and Henley, 1996] this is formulated as follows: 'When the range or the confidence interval of a variable is expressed as a multiplicative rather than additive error factor, the lognormal is the proper distribution to describe the variable'. For properties of the lognormal distribution, see Appendix B of the current document.

There are a few cases in which a lognormal distribution is less appropriate:

- If the correct value for the parameter is in a 95% credibility interval that is weighted to one side. For example, the velocity of an average aircraft is with 95% credibility in an interval $[v_{low}, v_{high}]$, but the average value is very close to the upper end of this interval, far away from the multiplicative centre, i.e. far away from $\sqrt{v_{low} \times v_{high}}$.
- If for an application some parameters take on both positive and negative values. Lognormally distributed variables are always positive.

<u>Conclusion:</u>
Bias and uncertainty model assumption 3.5 is judged as a restrictive assumption, but in most cases it is a useful choice ("–").

Evaluation of Bias and uncertainty model assumption 3.6

3.6	The inverse of $\varphi_i(\cdot)$, denoted by $\varphi_i^{inv}(\cdot)$, exists for all *i*.

The inverse of a function φ_i exists if the function is injective, i.e. if $\varphi_i(x_1) = \varphi_i(x_2)$ then $x_1 = x_2$. Since the denominator of $\varphi_i(v_i)$ is a constant (equal to $\Re_{MODEL}$), the inverse exists under the following condition: if
$\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},x_1,\overline{v}_{i+1},\ldots,\overline{v}_{n_p})) = \rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},x_2,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))$ then $x_1 = x_2$.

Using $(A \Rightarrow B)$ is equivalent to $(\neg B \Rightarrow \neg A)$, yields that this Bias and uncertainty model assumption means: if we change the value for some parameter into any other value, then accident risk must change. This statement will usually hold for most factors modelled since they indirectly or indirectly influence safety in some way, albeit that the change of some parameter values may have a very small effect on accident risk. For the parameters that do not influence risk there is no harm expected since their accident risk log-sensitivity is zero.

Conclusion:
Bias and uncertainty model assumption 3.6 is a reasonable assumption (“0”).

Evaluation of Bias and uncertainty model assumption 3.7

3.7	Each $\varphi_i(V_i)$ $(i=1,\ldots,n_p)$ is lognormally distributed with parameters γ_i and θ_i^2. Notation: $\varphi_i(V_i) \sim \Lambda(\gamma_i,\theta_i^2)$.

Bias and uncertainty model assumption 3.7 states that $\varphi_i(V_i)$ is lognormally distributed, where φ_i is given by

$$\varphi_i(x_i) \triangleq \frac{\rho(1,(\overline{v}_1,\ldots,\overline{v}_{i-1},x_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}{\rho(1,\overline{v}_1,\ldots,\overline{v}_{i-1},\overline{v}_i,\overline{v}_{i+1},\ldots,\overline{v}_{n_p}))}.$$

Considering the reasoning in [Kumamoto and Henley, 1996] again, i.e.: 'When the range or the confidence interval of a variable is expressed as a multiplicative rather than additive error factor, the lognormal is the proper distribution to describe the variable', and noticing that $\varphi_i(x_i)$ denotes the factor by which expected actual accident risk changes if we use x_i instead of $\overline{v}_i$ as value for the i^{th} parameter, then a multiplicative Normal distribution seems a logical candidate distribution for $\varphi_i(V_i)$. However, if this distribution type applies for V_i (i.e. under Bias and uncertainty model assumption 3.5), it may not apply for $\varphi_i(V_i)$ also. Thus, this Bias and uncertainty model assumption 3.7 may be contradicting Bias and uncertainty model assumption 3.5.

Bias and uncertainty model assumption 3.7 has impact on the Risk log-sensitivity of each parameter. Since it is a rather restrictive assumption, a safety conservative approximation has been developed to reduce the impact of this bias and uncertainty model assumption (i.e. to determine risk log-sensitivity β_i, determine the change in risk x^{β_i} due to change x in parameter value, by trying x = ℓ_i and $x = 1/\ell_i$, and use the worst case result for β_i). If Bias and uncertainty model assumption 3.7 does hold true, then risk log-sensitivity can be determined by trying only x = ℓ_i, since x = $1/\ell_i$ will give the same result for β_i.

Conclusion:
Bias and uncertainty model assumption 3.7 is rather restrictive (“–”), especially if Bias and uncertainty model assumption 3.5 also holds true.

Evaluation of Bias and uncertainty model assumption 3.8

3.8	For $i=1,\ldots,n_p$, the following two equations are satisfied $E\{\ln\varphi_i(V_i)\} = \beta_i(E\{\ln V_i\} - \ln\overline{v}_i)$, and $E\{(\ln\varphi_i(V_i) - E\{\ln\varphi_i(V_i)\})^2\} = \beta_i^2\mathrm{Var}\{\ln V_i\}$, with β_i the log-sensitivity around $\overline{v}$ as defined in Subsection 3.3.

Bias and uncertainty model assumption 3.8 states that the expectation of the natural logarithm of the quotients defined at Bias and uncertainty model assumption 3.1 can be decomposed as a product of two components, the first of which is determined by the risk log-sensitivity and the second of which is determined by the expectation of the natural logarithm of the i^{th} parameter. The same holds true for the variance of the natural logarithm of the quotients.

If $E\{\ln V_i\} - \ln \overline{v}_i \neq 0$ then the assumption can only hold if
$\beta_i = \dfrac{E\{\ln \varphi_i(V_i)\}}{E\{\ln V_i\} - \ln \overline{v}_i} = \dfrac{\sqrt{E\{(\ln \varphi_i(V_i) - E\{\ln \varphi_i(V_i)\})^2\}}}{\sqrt{\text{Var}\{\ln V_i\}}}$, which, since
$E\{(\ln \varphi_i(V_i) - E\{\ln \varphi_i(V_i)\})^2\} = \text{Var}\{\ln \varphi_i(V_i)\}$, $E\{\ln V_i\} - \ln \overline{v}_i = E\{\ln V_i - \ln \overline{v}_i\}$ and
$\text{Var}\{\ln V_i\} = \text{Var}\{\ln V_i - \ln \overline{v}_i\}$, is equivalent to $\beta_i = \dfrac{E\{\ln \varphi_i(V_i)\}}{E\{\ln V_i - \ln \overline{v}_i\}} = \dfrac{\sqrt{\text{Var}\{\ln \varphi_i(V_i)\}}}{\sqrt{\text{Var}\{\ln V_i - \ln \overline{v}_i\}}}$.
In words, this means that the expectations of $\ln \varphi_i(V_i)$ and $\ln V_i - \ln \overline{v}_i$ have to be in the same proportion as their standard deviations, and this proportion is equal to the log-sensitivity.
Note that by definition $\beta_i = \left.\dfrac{\partial \ln \varphi_i(v_i)}{\partial \ln v_i}\right|_{\ln v = \ln \overline{v}}$, which can be rewritten as
$\beta_i = \left.\dfrac{\partial \ln \varphi_i(v_i)}{\partial(\ln v_i - \ln \overline{v}_i)}\right|_{\ln v - \ln \overline{v} = 0}$. This makes that Bias and uncertainty model assumption 3.8 states
that $\left.\dfrac{\partial \ln \varphi_i(v_i)}{\partial(\ln v_i - \ln \overline{v}_i)}\right|_{\ln v - \ln \overline{v} = 0} = \dfrac{E\{\ln \varphi_i(V_i)\}}{E\{\ln V_i - \ln \overline{v}_i\}} = \dfrac{\sqrt{\text{Var}\{\ln \varphi_i(V_i)\}}}{\sqrt{\text{Var}\{\ln V_i - \ln \overline{v}_i\}}}$.

The advantage of Bias and uncertainty model assumption 3.8 over Bias and uncertainty model assumption 3.7 is that there is no longer a possible conflict with Bias and uncertainty model assumption 3.5. However, the requirement is still strict, considering that the same parameter β_i is used both in the expression for the expectation of $\ln \varphi_i(V_i)$ and in the expression for the variance of $\ln \varphi_i(V_i)$. There is no reason for it to be less restrictive than Bias and uncertainty model assumption 3.2.

Conclusion:
Bias and uncertainty model assumption 3.8 is restrictive ("– –") for some parameters.

Evaluation of Bias and uncertainty model assumption 3.9

3.9	With Θ_n according to Definition 3.4, $\lim_{n_p \to \infty} \Theta_3 / \Theta_2 \to 0$.

According to Definition 3.4, $\Theta_n \triangleq \sqrt[n]{\sum_i \theta_{i,n}}$, with $\theta_{i,n} \triangleq E\{|H_i - \gamma_i|^n\}$, $H_i \triangleq \ln \varphi_i(V_i)$ and $\gamma_i \triangleq E\{H_i\}$. Moreover, $\varphi_i(x_i)$ denotes the factor by which expected actual accident risk changes if we use x_i instead of $\overline{v}_i$ as value for the i^{th} parameter. A sufficient condition for this Bias and uncertainty model assumption to hold true is if H_i are bounded and their means and variances are finite. Since, in the models that we use, the risk of an aircraft accident is always greater than zero, while on the other hand the number of aircraft involved is finite, there can never be an infinite number of aircraft accidents within an aircraft flight hour. Therefore $\ln \varphi_i(V_i)$ must be bounded, hence H_i is bounded, and its mean and variance must be finite.

Conclusion:
Bias and uncertainty model assumption 3.9 is a reasonable assumption (“0”).

Evaluation of Bias and uncertainty model assumption 3.10

3.10	The number of parameters n_p goes to infinity: $n_p \to \infty$.

For the Central Limit Theorem to be applicable, n_p must go to infinity: $n_p \to \infty$. However, for n_p sufficiently large but limited, the Central Limit Theorem already gives good results. A commonly used rule of thumb is that $n_p \approx 30$ is sufficient for the Central Limit Theorem (i.e. the version where sums of random variables are Gaussian distributed) to hold true, i.e. the number of parameters that the stochastic model uses must be at least 30. However, if the probability densities of the individual $\ln \varphi_i(V_i)$ are very different, $n_p \approx 30$ may not be sufficient. Moreover, the transformation to products of random variables may also require this number to be larger.

In typical applications in which we used the bias and uncertainty assessment method so far, the number of parameters used was at least 80. This number is expected to be large enough to obtain good results. At least, the error made will be insignificant, compared with errors made due to the adoption of other bias and uncertainty model assumptions.

Conclusion:
Bias and uncertainty model assumption 3.10 is a reasonable assumption (“0”).

4.2 Conclusion on the evaluation of Bias and uncertainty model assumptions 3.1 through 3.10

The last table in Section 3.4 shows that a practical characterisation of $\Re_{ACTUAL} = E\{\rho(A,V)\}$ can be found under four different sets of bias and uncertainty model assumptions. This table is repeated below, but the bias and uncertainty model assumption formulations have been replaced by the conclusion of the evaluation in the previous subsection.

Bias and uncertainty model assumption		**A**	**B**	**C**	**D**
3.1	Is judged as a hard requirement.	– –	– –	– –	– –
3.2	If $\ln \rho(1,v)$ is a linear function of $\ln v_i$, then Bias and uncertainty model assumption 3.2 is judged a reasonable approximation (“0”). In other cases, Bias and uncertainty model assumption 3.2 may not be completely accurate: “–”, or even “– –”, depending on behaviour of $\ln \rho(1,v)$. In practice, there often are some relevant parameters for which “– –” applies.	– –		– –	
3.3	Is a reasonable assumption since many difficulties can be eliminated.	0	0	0	0
3.4	Is judged as a reasonable assumption.	0	0	0	0
3.5	Is judged as a restrictive assumption, but in most cases it is a useful choice.	–	–		
3.6	Is a reasonable assumption.		0		0

3.7	Is rather restrictive, especially if Bias and uncertainty model assumption 3.5 also holds true		–		
3.8	Is restrictive for some parameters.		– –		– –
3.9	Is a reasonable assumption.			0	0
3.10	Is a reasonable assumption.			0	0

We can now compare the four sets of bias and uncertainty model assumptions with each other by counting the number of occurrences (#) of "– –", "–" and "0", and rank them on number of occurrences of "– –", "–" and "0". This is done in the table below.

Rank	**Set**	**# "– –"**	**# "–"**	**# "0"**
1	Set C	2	0	4
2	Set D	2	0	5
3	Set A	2	1	2
4	Set B	2	2	3

Based on this, sets C and D form the most logical explanation for the approximation developed in Section 3.

4.3 Accident risk as a sum of contributions

According to the previous subsection, the order of most favourable set to least favourable set seems to be C, D, A, B. We should note, however, that sets A and C have clear advantages in practical situations over sets B and D: under sets A and C, the risk log-sensitivity β_i is more easily measured. Therefore, it seems worth trying to improve sets A and C by an approach that works better for the bias and uncertainty model assumptions judged by "– –", which are 3.1 and 3.2.

The crux of these two assumptions is that they assume a log-linear approximation for accident risk, which is not always accurate. In particular, in an accident risk model, modelled risk typically is a weighted sum of conditional risk contributions, for example: $\Re_{MODEL} = \sum_k \Re_{MODEL}(k) \times \Pr_{MODEL}(k)$, where $\Re_{MODEL}(k)$ is the modelled accident risk conditioned on some particular model event type k, and $\Pr_{MODEL}(k)$ is the probability of occurrence of this event type k in the model. Now, it appears that whereas a log-linear approach for $\Re_{MODEL}$ is less suitable, a log-linear approach for $\Re_{MODEL}(k) \times \Pr_{MODEL}(k)$ works much better. This is due to the fact that when restricting to a particular condition k, many parameters do not play a role anymore, which makes $\Re_{MODEL}(k) \times \Pr_{MODEL}(k)$ a more smooth function of the parameter values.

Therefore, in Section 5, we extend the bias and uncertainty assessment modelling as developed in [Everdij and Blom, 2002], in order to cover accident risk that is written as a sum of risk contributions. We will restrict the extension to sets of Bias and uncertainty model assumptions analogous to sets A and C, but note that extensions to sets B and D are equally possible.

5. Weighted sum factorisation model for bias and uncertainty in accident risk

This section presents an extension of the bias and uncertainty factorisation model described in Section 3. The section gives a mathematical characterisation of how the model-evaluated expected accident risk for a particular application relates to the actual expected accident risk for that application, but where the model-evaluated accident risk is written as a weighted sum of conditional risks. Throughout this section, we use the notational conventions as proposed in Subsection 3.1.

5.1 Problem definition

Additionally to the definitions in Section 3, define

- An event sequence process $\{\kappa^{ij}_{\tau^{ij}}\}$ and let $\mathbf{K} = \{\kappa_1, \ldots, \kappa_{n_k}\}$ denote the set of events that the event sequence process $\{\kappa^{ij}_{\tau^{ij}}\}$ can take on, see [Blom et al, 2003].
- A function $\tilde{\rho}$ by $\tilde{\rho}(\alpha, v, \kappa) \triangleq$ the Expected number of collisions under condition $\{\kappa^{ij}_{\tau^{ij}} = \kappa\}$, between aircraft per aircraft flight hour under non-parameter assumptions α and parameter values v.
- A function $\tilde{\chi}$ by $\tilde{\chi}(\kappa, \alpha, v) \triangleq \Pr(\kappa^{ij}_{\tau^{ij}} = \kappa \mid A = \alpha, V = v)$, i.e. the probability of occurrence of the event $\{\kappa^{ij}_{\tau^{ij}} = \kappa\}$, under non-parameter assumptions α and parameter values v.

Then, $\rho(\alpha, v) = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(\alpha, v, \kappa)\tilde{\chi}(\kappa, \alpha, v)$ and in particular:

$\rho(\mathbf{1}, V) = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(\mathbf{1}, V, \kappa)\tilde{\chi}(\kappa, \mathbf{1}, V)$ and $E\{\rho(\mathbf{1}, V)\} = \sum_{\kappa \in \mathbf{K}} E\{\tilde{\rho}(\mathbf{1}, V, \kappa)\tilde{\chi}(\kappa, \mathbf{1}, V)\}$. This yields, using Definition 3.5, $\Re_{ACTUAL} = \Psi \times E\{\rho(\mathbf{1}, V)\} = \Psi \times \sum_{\kappa \in \mathbf{K}} E\{\tilde{\rho}(\mathbf{1}, V, \kappa)\tilde{\chi}(\kappa, \mathbf{1}, V)\}$.

The problem considered in the remainder of this section is the following:

Problem definition:
Characterise $\Re_{ACTUAL}$ in terms of $\tilde{\rho}(\mathbf{1}, \bar{v}, \kappa)$ and $\tilde{\chi}(\kappa, \mathbf{1}, \bar{v})$ and properties of the stochastic model.

♦

The remainder of this section is organised as follows: First, Subsection 5.2 discusses log-sensitivity, and in particular what this means in terms of the weighted sum expression above. Next, $\Re_{ACTUAL}$ is characterised in four steps:

- Subsection 5.3 the product $\tilde{\rho}(\mathbf{1}, V, \kappa) \times \tilde{\chi}(\kappa, \mathbf{1}, V)$ in terms of $\tilde{\rho}(\mathbf{1}, \bar{v}, \kappa)$ and $\tilde{\chi}(\kappa, \mathbf{1}, \bar{v})$ and properties of the stochastic model,
- Subsection 5.4 characterises $\rho(\mathbf{1}, V)$ in terms of $\tilde{\rho}(\mathbf{1}, V, \kappa) \times \tilde{\chi}(\kappa, \mathbf{1}, V)$ and properties of the stochastic model,

- Subsection 5.5 characterises $\rho(A,V)$ in terms of $E\{\rho(1,V)\}$ and properties of the non-parameter assumptions,
- Subsection 5.6 combines the results.

The proofs of all theorems, lemmas and corollaries can be found in Appendix A.

5.2 Log-sensitivity of accident risk

From Subsection 3.3, we have that the log-sensitivity of accident risk is defined by: Log-Sensitivity$_i(v) \triangleq \frac{\partial \ln \rho(1,v)}{\partial \ln v_i}$. In the current section, we use that $\rho(1,v)$ is written as a weighted sum: $\rho(1,v) = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)$. With this, we can further evaluate:

$$\text{Log-Sensitivity}_i(v) = \frac{1}{\rho(1,v)} \cdot \frac{\partial \rho(1,v)}{\partial \ln v_i} = \frac{1}{\rho(1,v)} \cdot \frac{\partial \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)}{\partial \ln v_i} =$$

$$\frac{1}{\rho(1,v)} \cdot \sum_{\kappa \in \mathbf{K}} \left[\tilde{\rho}(1,v,\kappa) \frac{\partial \tilde{\chi}(\kappa,1,v)}{\partial \ln v_i} + \tilde{\chi}(\kappa,1,v) \frac{\partial \tilde{\rho}(1,v,\kappa)}{\partial \ln v_i} \right].$$

We can also evaluate β_i, the log-sensitivity of $\rho(1,v)$ to the i^{th} parameter value in point $\bar{v}$ as defined in Subsection 3.3. First define $\beta_i(\kappa) \triangleq \left. \frac{\partial \ln \tilde{\rho}(1,v,\kappa)}{\partial \ln v_i} \right|_{\ln v = \ln \bar{v}}$ as the log-sensitivity of $\tilde{\rho}(1,v,\kappa)$ to the i^{th} parameter value in point $\bar{v}$, and $\eta_i(\kappa) \triangleq \left. \frac{\partial \ln \tilde{\chi}(\kappa,1,v)}{\partial \ln v_i} \right|_{\ln v = \ln \bar{v}}$ as the log-sensitivity of $\tilde{\chi}(\kappa,1,v)$ to the i^{th} parameter value in point $\bar{v}$. Hence

$\left. \frac{\partial \tilde{\rho}(1,v,\kappa)}{\partial \ln v_i} \right|_{\ln v = \ln \bar{v}} = \beta_i(\kappa)\tilde{\rho}(1,\bar{v},\kappa)$ and $\left. \frac{\partial \tilde{\chi}(\kappa,1,v)}{\partial \ln v_i} \right|_{\ln v = \ln \bar{v}} = \eta_i(\kappa)\tilde{\chi}(\kappa,1,\bar{v})$ and therefore

$$\beta_i = \text{Log-Sensitivity}_i(\bar{v}) = \frac{1}{\rho(1,\bar{v})} \cdot \sum_{\kappa \in \mathbf{K}} \left[\tilde{\rho}(1,\bar{v},\kappa)\eta_i(\kappa)\tilde{\chi}(\kappa,1,\bar{v}) + \tilde{\chi}(\kappa,1,\bar{v})\beta_i(\kappa)\tilde{\rho}(1,\bar{v},\kappa) \right] =$$

$$= \frac{1}{\rho(1,\bar{v})} \cdot \sum_{\kappa \in \mathbf{K}} \left[\tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})(\beta_i(\kappa) + \eta_i(\kappa)) \right].$$

5.3 Characterisation of the product $\tilde{\rho}(1,V,\kappa) \times \tilde{\chi}(\kappa,1,V)$

In this subsection, the product $\tilde{\rho}(1,V,\kappa) \times \tilde{\chi}(\kappa,1,V)$ is characterised in terms of $\tilde{\rho}(1,\bar{v},\kappa)$ and $\tilde{\chi}(\kappa,1,\bar{v})$.

Note that the characterisation of $\tilde{\rho}(1,V,\kappa) \times \tilde{\chi}(\kappa,1,V)$ is similar to the characterisation of $\tilde{\rho}(1,V)$ developed in Subsection 3.4. The main difference is that in Subsection 3.4, we proved this characterisation of $\tilde{\rho}(1,V)$ under four sets of bias and uncertainty model assumptions (referred to as A, B, C and D in Subsection 3.6), while in the current subsection, characterising $\tilde{\rho}(1,V,\kappa) \times \tilde{\chi}(\kappa,1,V)$, we drop two of these (i.e. we will drop sets similar to B and D). The reason for this limitation is simplicity only (sets similar to B and D lead to an analogous result).

Definition 5.1:
For each $i = 1, \ldots, n_p$, define a function $\tilde{\varphi}_i$ by:

$$\tilde{\varphi}_i(x_i, \kappa) \triangleq \frac{\tilde{\rho}(1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, x_i, \bar{v}_{i+1}, \ldots, \bar{v}_{n_p}), \kappa)}{\tilde{\rho}(1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, \bar{v}_i, \bar{v}_{i+1}, \ldots, \bar{v}_{n_p}), \kappa)}.$$

♦

Definition 5.2:
For each $i = 1, \ldots, n_p$, define a function $\tilde{\varphi}_i$' by:

$$\tilde{\varphi}_i'(x_i, x_{i+1}, \ldots, x_{n_p}, \kappa) \triangleq \frac{\tilde{\rho}(1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, x_i, x_{i+1}, \ldots, x_{n_p}), \kappa)}{\tilde{\rho}(1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, \bar{v}_i, x_{i+1}, \ldots, x_{n_p}), \kappa)}.$$

♦

Bias and uncertainty model assumption 5.1:
For any $i = 1, \ldots, n_p$ and $x_i, \ldots, x_{n_p}$, $\tilde{\varphi}_i(x_i, \kappa) = \tilde{\varphi}_i'(x_i, x_{i+1}, \ldots, x_{n_p}, \kappa)$. ♦

Lemma 5.1:
Under Bias and uncertainty model assumption 5.1, for all $v \in \mathbb{R}^{n_p}$:

$$\tilde{\rho}(1, v, \kappa) = \tilde{\rho}(1, \bar{v}, \kappa) \times \prod_{i=1}^{n_p} \tilde{\varphi}_i(v_i, \kappa).$$

♦

Definition 5.3:
For each $i = 1, \ldots, n_p$, define a function $\tilde{\xi}_i$ by:

$$\tilde{\xi}_i(x_i, \kappa) \triangleq \frac{\tilde{\chi}(\kappa, 1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, x_i, \bar{v}_{i+1}, \ldots, \bar{v}_{n_p}))}{\tilde{\chi}(\kappa, 1, \bar{v})}.$$

♦

Definition 5.4:
For each $i = 1, \ldots, n_p$, define functions $\tilde{\xi}_i$' by:

$$\tilde{\xi}_i'(x_i, x_{i+1}, \ldots, x_{n_p}, \kappa) \triangleq \frac{\tilde{\chi}(\kappa, 1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, x_i, x_{i+1}, \ldots, x_{n_p}))}{\tilde{\chi}(\kappa, 1, (\bar{v}_1, \ldots, \bar{v}_{i-1}, \bar{v}_i, x_{i+1}, \ldots, x_{n_p}))}.$$

♦

Bias and uncertainty model assumption 5.2:
For any $i = 1, \ldots, n_p$ and $x_i, \ldots, x_{n_p}$, $\tilde{\xi}_i(x_i, \kappa) = \tilde{\xi}_i'(x_i, x_{i+1}, \ldots, x_{n_p}, \kappa)$. ♦

Lemma 5.2:
Under Bias and uncertainty model assumption 5.2, for all $v \in \mathbb{R}^{n_p}$:

$$\tilde{\chi}(\kappa, 1, v) = \tilde{\chi}(\kappa, 1, \bar{v}) \times \prod_{i=1}^{n_p} \tilde{\xi}_i(v_i, \kappa).$$

♦

Bias and uncertainty model assumption 5.3:
For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\tilde{\varphi}_i(v_i,\kappa)=(v_i/\overline{v}_i)^{\beta_i(\kappa)}$, with $\beta_i(\kappa)$ the log-sensitivity of $\tilde{\rho}(1,v,\kappa)$ to the i^{th} parameter value around point $\overline{v}$ as defined in Subsection 5.2. ♦

Bias and uncertainty model assumption 5.4:
For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\tilde{\xi}_i(v_i,\kappa)=(v_i/\overline{v}_i)^{\eta_i(\kappa)}$, with $\eta_i(\kappa)$ the log-sensitivity of $\tilde{\chi}(\kappa,1,v)$ to the i^{th} parameter value around point $\overline{v}$ as defined in Subsection 5.2. ♦

Bias and uncertainty model assumption 5.5:
The random variables $V_1,\ldots,V_{n_p}$ are mutually independent. ♦

Bias and uncertainty model assumption 5.6:
For each $i=1,\ldots,n_p$, the expectation and variance of $\ln V_i$ exist and satisfy:
$E\{\ln V_i\}=\mu_i$ and $\mathrm{Var}\{\ln V_i\}=\sigma_i^2$. ♦

Bias and uncertainty model assumption 5.7:
Each V_i $(i=1,\ldots,n_p)$ is lognormally distributed. ♦

Theorem 5.1:
Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7} $\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)$ is lognormally distributed:

$$\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)\sim\Lambda(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa],\frac{1}{4}\widehat{G}_\kappa),$$

with $\widehat{B}_\kappa$, $\widehat{M}_\kappa$ and $\widehat{G}_\kappa$ as defined in Definition 5.5 below. ♦

Definition 5.5:
- $b_i \triangleq \exp(\mu_i)/\overline{v}_i$,
- $\widehat{B}_\kappa \triangleq \prod_{i=1}^{n_p} b_i^{\beta_i(\kappa)}$,
- $\widehat{M}_\kappa \triangleq \prod_{i=1}^{n_p} b_i^{\eta_i(\kappa)}$,
- $\ell_i \triangleq \exp(2\sigma_i)$,
- $\widehat{G}_\kappa \triangleq \sum_{i=1}^{n_p}(\ln \ell_i^{|\beta_i(\kappa)+\eta_i(\kappa)|})^2$. ♦

In Theorem 5.1, lognormal assumptions are used for V_i. In the remainder of this subsection, we make use of the Central Limit Theorem to relax these assumptions. First, introduce some definitions:

Definition 5.6:
- $H_i(\kappa) \triangleq \ln\tilde{\varphi}_i(V_i,\kappa)$
- $G_i(\kappa) \triangleq \ln\tilde{\xi}_i(V_i,\kappa)$

- $\gamma_i(\kappa) \triangleq E\{H_i(\kappa)\}$
- $\delta_i(\kappa) \triangleq E\{G_i(\kappa)\}$
- $\Gamma(\kappa) \triangleq \sum_{i=1}^{n_p} \gamma_i(\kappa)$
- $\mathrm{D}(\kappa) \triangleq \sum_{i=1}^{n_p} \delta_i(\kappa)$
- $\theta_{i,n}(\kappa) \triangleq E\{|H_i(\kappa) - \gamma_i(\kappa)|^n\}$, $n = 2,3$
- $\tau_{i,n}(\kappa) \triangleq E\{|G_i(\kappa) - \delta_i(\kappa)|^n\}$, $n = 2,3$
- $\Theta_n(\kappa) \triangleq \sqrt[n]{\sum_{i=1}^{n_p} \theta_{i,n}(\kappa)}$, $n = 2,3$
- $\mathrm{T}_n(\kappa) \triangleq \sqrt[n]{\sum_{i=1}^{n_p} \tau_{i,n}(\kappa)}$, $n = 2,3$

♦

Subsequently, introduce Bias and uncertainty model assumptions 5.8, 5.9 and 5.10:

Bias and uncertainty model assumption 5.8:
With Θ_n according to Definition 5.6, $\lim_{n_p \to \infty} \Theta_3(\kappa) / \Theta_2(\kappa) \to 0$. ♦

Bias and uncertainty model assumption 5.9:
With T_n according to Definition 5.6, $\lim_{n_p \to \infty} \mathrm{T}_3(\kappa) / \mathrm{T}_2(\kappa) \to 0$. ♦

Bias and uncertainty model assumption 5.10:
The number of parameters n_p goes to infinity: $n_p \to \infty$. ♦

Theorem 5.2:
Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10} $\tilde{\rho}(1, V, \kappa)\tilde{\chi}(\kappa, 1, V)$ is lognormally distributed:

$$\tilde{\rho}(1, V, \kappa)\tilde{\chi}(\kappa, 1, V) \sim \Lambda(\ln[\tilde{\rho}(1, \bar{v}, \kappa)\tilde{\chi}(\kappa, 1, \bar{v})\hat{B}_\kappa \hat{M}_\kappa], \frac{1}{4}\hat{G}_\kappa),$$

with $\hat{B}_\kappa$, $\hat{M}_\kappa$ and $\hat{G}_\kappa$ as defined in Definition 5.5. ♦

Finally, notice that, in Theorems 5.1 and 5.2, we found the same characterisation of $\tilde{\rho}(1, V, \kappa)\tilde{\chi}(\kappa, 1, V)$ by means of a probability distribution, under two different sets of bias and uncertainty model assumptions, i.e. {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7} and {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10}, i.e.:

$$\tilde{\rho}(1, V, \kappa)\tilde{\chi}(\kappa, 1, V) \sim \Lambda(\ln[\tilde{\rho}(1, \bar{v}, \kappa)\tilde{\chi}(\kappa, 1, \bar{v})\hat{B}_\kappa \hat{M}_\kappa], \frac{1}{4}\hat{G}_\kappa).$$

5.4 Characterisation of $\rho(1, V)$ in terms of $\tilde{\rho}(1, V, \kappa) \times \tilde{\chi}(\kappa, 1, V)$

In the previous subsection we found, under different sets of bias and uncertainty model assumptions, i.e. {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7} and {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10}, that the product of $\tilde{\rho}(1, V, \kappa)$ and $\tilde{\chi}(\kappa, 1, V)$ is lognormal:

$$\tilde{\rho}(1, V, \kappa)\tilde{\chi}(\kappa, 1, V) \sim \Lambda(\ln[\tilde{\rho}(1, \bar{v}, \kappa)\tilde{\chi}(\kappa, 1, \bar{v})\hat{B}_\kappa \hat{M}_\kappa], \frac{1}{4}\hat{G}_\kappa).$$

Next, in order to characterise $\rho(1,V) = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)$ we need to know the density or distribution of a sum of lognormally distributed variables.

There is no closed form known of the density or distribution function of a sum of lognormal random variables. Several approximations exist, see Appendix C. Below, we use Fenton-Wilkinson's method (Appendix C.4), which is relatively simple, and still performs rather well (see Appendix C.8).

Fenton-Wilkinson's method

Suppose X_i is lognormal: $X_i \sim \Lambda(\mu_{Z_i}, \sigma_{Z_i}^2)$, hence $Z_i = \ln X_i \sim N(\mu_{Z_i}, \sigma_{Z_i}^2)$ is Gaussian. Then the sum $X = X_1 + \cdots + X_n$ is approximated by another lognormal variable: $X' \sim \Lambda(\mu_W, \sigma_W^2)$, where $\mu_W = 2\ln u_1 - \frac{1}{2}\ln u_2$ and $\sigma_W^2 = \ln u_2 - 2\ln u_1$ and where

$$u_1 = \sum_{i=1}^{n} \exp(\mu_{Z_i} + \sigma_{Z_i}^2 / 2),$$

$$u_2 = \sum_{i=1}^{n} \exp(2\mu_{Z_i} + 2\sigma_{Z_i}^2) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \exp(\mu_{Z_i} + \mu_{Z_j})\exp(\frac{1}{2}(\sigma_{Z_i}^2 + \sigma_{Z_j}^2 + 2r_{ij}\sigma_{Z_i}\sigma_{Z_j})),$$

where $r_{ij} = \dfrac{E\{(Z_i - \mu_{Z_i})(Z_j - \mu_{Z_j})\}}{\sigma_{Z_i}\sigma_{Z_j}}$ is the correlation coefficient of Z_i and Z_j.

♦

Note that

- The term $\sigma_{Z_i}^2 + \sigma_{Z_j}^2 + 2r_{ij}\sigma_{Z_i}\sigma_{Z_j}$ can be written as $\sigma_{Z_i}^2 + \sigma_{Z_j}^2 + 2E\{(Z_i - \mu_{Z_i})(Z_j - \mu_{Z_j})\}$, which is equal to the variance of $Z_i + Z_j$.
- The correlation coefficient r_{ij} of Z_i and Z_j is in the interval $[-1,1]$. It is equal to zero if Z_i and Z_j are independent and it is equal to +1 or –1 if there is an exact linear relation between Z_i and Z_j (i.e. there exist a and b such that $Z_i = a_{ij}Z_j + b_{ij}$; if a_{ij} is positive then r_{ij} =1; if a_{ij} is negative then r_{ij} = –1). In particular, if Z_i and Z_j are independent then $\sigma_{Z_i}^2 + \sigma_{Z_j}^2 + 2r_{ij}\sigma_{Z_i}\sigma_{Z_j}$ reduces to $\sigma_{Z_i}^2 + \sigma_{Z_j}^2$.
- This method is only an approximation, therefore we pose another bias and uncertainty model assumption:

Bias and uncertainty model assumption 5.11:

The error made with using the Fenton-Wilkinson method for approximating the expectation and 95% credibility interval of the sum $\sum_{\kappa \in \mathbf{K}} \tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)$ is zero.

♦

Definition 5.7:

- $\widehat{G}_{\kappa\kappa'} \triangleq \sum_{i=1}^{n_p} (\beta_i(\kappa) + \eta_i(\kappa))(\beta_i(\kappa') + \eta_i(\kappa'))(\ln \ell_i)^2$

with ℓ_i as in Definition 5.5.

♦

Theorem 5.3:
Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.11} and {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10, 5.11}, the following holds true:

$\rho(1,V) \sim \Lambda(\mu_W, \sigma_W^2)$, with

$\mu_W = 2\ln u_1 - \frac{1}{2}\ln u_2$,

$\sigma_W^2 = \ln u_2 - 2\ln u_1$,

$$u_1 = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa \times \exp(\frac{1}{8}\widehat{G}_\kappa),$$

$$u_2 = \sum_{\kappa \in \mathbf{K}} (\tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa)^2 \times \exp(\frac{1}{2}\widehat{G}_\kappa) +$$

$$+2 \sum_{\substack{\kappa,\kappa' \in \mathbf{K} \\ \kappa' > \kappa}} \tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa \tilde{\rho}(1,\bar{v},\kappa')\tilde{\chi}(\kappa',1,\bar{v})\widehat{B}_{\kappa'} \widehat{M}_{\kappa'} \times \exp(\frac{1}{8}(\widehat{G}_\kappa + \widehat{G}_{\kappa'}) + \frac{1}{4}\widehat{G}_{\kappa\kappa'})$$

with $\widehat{B}_\kappa$, $\widehat{M}_\kappa$ and $\widehat{G}_\kappa$ defined in Definition 5.5 and $\widehat{G}_{\kappa\kappa'}$ defined in Definition 5.7.

♦

Corollary 5.1:
Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.11} and {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10, 5.11}, the following holds:

$E\{\rho(1,V)\} = u_1$ and

$$\Pr(\rho(1,V) \in \left[\frac{u_1^2}{\sqrt{u_2}}\exp(-2\sqrt{\ln\frac{u_2}{u_1^2}}), \frac{u_1^2}{\sqrt{u_2}}\exp(2\sqrt{\ln\frac{u_2}{u_1^2}})\right]) = 0.95,$$ with

$$u_1 = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa \times \exp(\frac{1}{8}\widehat{G}_\kappa),$$

$$u_2 = \sum_{\kappa \in \mathbf{K}} (\tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa)^2 \times \exp(\frac{1}{2}\widehat{G}_\kappa) +$$

$$+2 \sum_{\substack{\kappa,\kappa' \in \mathbf{K} \\ \kappa' > \kappa}} \tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa \tilde{\rho}(1,\bar{v},\kappa')\tilde{\chi}(\kappa',1,\bar{v})\widehat{B}_{\kappa'} \widehat{M}_{\kappa'} \times \exp(\frac{1}{8}(\widehat{G}_\kappa + \widehat{G}_{\kappa'}) + \frac{1}{4}\widehat{G}_{\kappa\kappa'})$$

with $\widehat{B}_\kappa$, $\widehat{M}_\kappa$ and $\widehat{G}_\kappa$ defined in Definition 5.5 and $\widehat{G}_{\kappa\kappa'}$ defined in Definition 5.7.

♦

5.5 Relation between $E\{\rho(1,V)\}$ and $\Re_{ACTUAL} = E\{\rho(A,V)\}$

In this subsection, $\Re_{ACTUAL} = E\{\rho(A,V)\}$ is characterised in terms of $E\{\rho(1,V)\}$.

Definition 5.8:

$$\Psi \triangleq \frac{\Re_{ACTUAL}}{E\{\rho(1,V)\}}$$

♦

Theorem 5.4:

$$\Psi = \prod_{i=1}^{n_a} \left\{ \Pr(A_i = 0 \mid A_1 = 1, \ldots, A_{i-1} = 1) \times \frac{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{i-1} = 1, A_i = 0\}}{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_i = 1\}} + \right.$$
$$\left. + \Pr(A_i = 1 \mid A_1 = 1, \ldots, A_{i-1} = 1) \right\}$$

where the set $\{A_1, \ldots, A_{i-1}\}$ is empty if $i = 1$. ♦

5.6 Results combined

In this subsection, we combine the results obtained in the previous two subsections. Use Definition 5.8 to write:

$$\Re_{ACTUAL} = E\{\rho(A,V)\} = E\{\rho(\mathbf{1},V)\} \times \Psi$$

and notice that the first factor on the right hand side of the above equation has been evaluated in Subsection 5.4, the second factor has been evaluated in Subsection 5.5. Combining these results yields:

Corollary 5.2

Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.11} and {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10, 5.11}, the following holds:

$$\Re_{ACTUAL} = E\{\rho(A,V)\} = \Psi \times u_1, \qquad (5.1)$$

and

$$\Pr(\Psi \times \rho(\mathbf{1},V) \in \left[\Psi \times \frac{u_1^2}{\sqrt{u_2}} \exp(-2\sqrt{\ln \frac{u_2}{u_1^2}}), \Psi \times \frac{u_1^2}{\sqrt{u_2}} \exp(2\sqrt{\ln \frac{u_2}{u_1^2}}) \right]) = 0.95 \qquad (5.2)$$

with

$$u_1 = \sum_{\kappa \in \mathbf{K}} \tilde{\rho}(\mathbf{1}, \overline{v}, \kappa) \tilde{\chi}(\kappa, \mathbf{1}, \overline{v}) \widehat{B}_\kappa \widehat{M}_\kappa \times \exp(\frac{1}{8} \widehat{G}_\kappa),$$

$$u_2 = \sum_{\kappa \in \mathbf{K}} (\tilde{\rho}(\mathbf{1}, \overline{v}, \kappa) \tilde{\chi}(\kappa, \mathbf{1}, \overline{v}) \widehat{B}_\kappa \widehat{M}_\kappa)^2 \times \exp(\frac{1}{2} \widehat{G}_\kappa) +$$
$$+ 2 \sum_{\substack{\kappa, \kappa' \in \mathbf{K} \\ \kappa' > \kappa}} \tilde{\rho}(\mathbf{1}, \overline{v}, \kappa) \tilde{\chi}(\kappa, \mathbf{1}, \overline{v}) \widehat{B}_\kappa \widehat{M}_\kappa \tilde{\rho}(\mathbf{1}, \overline{v}, \kappa') \tilde{\chi}(\kappa', \mathbf{1}, \overline{v}) \widehat{B}_{\kappa'} \widehat{M}_{\kappa'} \times \exp(\frac{1}{8}(\widehat{G}_\kappa + \widehat{G}_{\kappa'}) + \frac{1}{4} \widehat{G}_{\kappa\kappa'})$$

with $\widehat{B}_\kappa$, $\widehat{M}_\kappa$ and $\widehat{G}_\kappa$ defined in Definition 5.5 and $\widehat{G}_{\kappa\kappa'}$ defined in Definition 5.7.

♦

For easier reference, the Bias and uncertainty model assumptions used are repeated below. The last two columns indicate which sets of bias and uncertainty model assumptions are required to hold true (grey field) for Equations 5.1 and 5.2 to hold true.

Bias and uncertainty model assumption		Ã	C̃
5.1	For any $i = 1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\tilde{\varphi}_i(x_i,\kappa) = \tilde{\varphi}_i'(x_i, x_{i+1},\ldots,x_{n_p},\kappa)$.		
5.2	For any $i = 1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\tilde{\xi}_i(x_i,\kappa) = \tilde{\xi}_i'(x_i, x_{i+1},\ldots,x_{n_p},\kappa)$.		
5.3	For all $i = 1,\ldots,n_p$, and all $v \in \mathbb{R}^{n_p}$, $\tilde{\varphi}_i(v_i,\kappa) = (v_i/\overline{v}_i)^{\beta_i(\kappa)}$, with $\beta_i(\kappa)$ the log-sensitivity of $\tilde{\rho}(\mathbf{1}, v, \kappa)$ to the i^{th} parameter value around point $\overline{v}$.		
5.4	For all $i = 1,\ldots,n_p$, and all $v \in \mathbb{R}^{n_p}$, $\tilde{\xi}_i(v_i,\kappa) = (v_i/\overline{v}_i)^{\eta_i(\kappa)}$, with $\eta_i(\kappa)$ the log-sensitivity of $\tilde{\chi}(\kappa, \mathbf{1}, v)$ to the i^{th} parameter value around point $\overline{v}$.		
5.5	The random variables $V_1,\ldots,V_{n_p}$ are mutually independent.		
5.6	For each $i = 1,\ldots,n_p$, the expectation and variance of $\ln V_i$ exist and satisfy: $E\{\ln V_i\} = \mu_i$ and $\mathrm{Var}\{\ln V_i\} = \sigma_i^2$.		
5.7	Each V_i $(i = 1,\ldots,n_p)$ is lognormally distributed.		
5.8	With Θ_n according to Definition 5.6, $\lim_{n_p \to \infty} \Theta_3(\kappa)/\Theta_2(\kappa) \to 0$.		
5.9	With T_n according to Definition 5.6, $\lim_{n_p \to \infty} \mathrm{T}_3(\kappa)/\mathrm{T}_2(\kappa) \to 0$.		
5.10	The number of parameters n_p goes to infinity: $n_p \to \infty$.		
5.11	The error made with using the Fenton-Wilkinson method for approximating the expectation and 95% credibility interval of the sum $\sum_{\kappa \in \mathbf{K}} \tilde{\rho}(\mathbf{1}, V, \kappa)\tilde{\chi}(\kappa, \mathbf{1}, V)$ is zero.		

Ã : Grey fields indicate first set of Bias and uncertainty model assumptions under which Equations 5.1 and 5.2 hold true, i.e. {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.11};
C̃ : Grey fields indicate second set of Bias and uncertainty model assumptions under which Equations 5.1 and 5.2 hold true, i.e. {5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.8, 5.9, 5.10, 5.11};

In Section 6, these bias and uncertainty model assumptions are evaluated on feasibility.

6. Evaluation of Bias and uncertainty model assumptions 5.1 through 5.11

This section first, in Subsection 6.1, discusses and evaluates the bias and uncertainty model assumptions adopted in Section 5. Subsection 6.2 gathers the results of these evaluations and gives concluding remarks. Finally, Subsection 6.3 compares the results of Subsection 6.2 with those of Subsection 4.2.

6.1 Evaluation of Bias and uncertainty model assumptions 5.1 through 5.11

This subsection discusses and evaluates Bias and uncertainty model assumptions 5.1 through 5.11. As was done in Section 4, based on its discussion, each bias and uncertainty model assumption is judged on its feasibility by a mark “0”, “–” or “– –”, where,

“0” denotes that the assumption usually (almost) holds true;

“–” denotes that the assumption may not always hold true, but is a logical choice and approximations may exist to reduce its effect;

“– –” denotes that the assumption is very restrictive.

In several cases, we refer to the discussion on Bias and uncertainty model assumptions 3.1 through 3.10, which have been discussed in Section 4.

Evaluation of Bias and uncertainty model assumptions 5.1 and 5.2

5.1	For any $i=1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\tilde{\varphi}_i(x_i,\kappa)=\tilde{\varphi}_i\,'(x_i,x_{i+1},\ldots,x_{n_p},\kappa)$.
5.2	For any $i=1,\ldots,n_p$ and $x_i,\ldots,x_{n_p}$, $\tilde{\xi}_i(x_i,\kappa)=\tilde{\xi}_i\,'(x_i,x_{i+1},\ldots,x_{n_p},\kappa)$.

These assumptions are discussed together.

A similar reasoning applies as for Bias and uncertainty model assumption 3.1. However, the κ-dependent approximation made through Bias and uncertainty model assumption 5.1 relaxes the objections that applied against Bias and uncertainty model assumption 3.1.

First, consider the case that Bias and uncertainty model assumptions 5.1 and 5.2 do not hold true. By using Definitions 5.3 and 5.4 repeatedly, we obtain the exact expressions $\tilde{\rho}(\mathbf{1},v,\kappa)=\tilde{\rho}(\mathbf{1},\overline{v},\kappa)\times\prod_{i=1}^{n_p}\tilde{\varphi}_i\,'(v_i,v_{i+1},\ldots,v_{n_p},\kappa)$ and $\tilde{\chi}(\kappa,\mathbf{1},v)=\tilde{\chi}(\kappa,\mathbf{1},\overline{v})\times\prod_{i=1}^{n_p}\tilde{\xi}_i\,'(v_i,v_{i+1},\ldots,v_{n_p},\kappa)$. The factors $\tilde{\varphi}_i\,'(v_i,v_{i+1},\ldots,v_{n_p},\kappa)$ and $\tilde{\xi}_i\,'(v_i,v_{i+1},\ldots,v_{n_p},\kappa)$ compensate conditional risk and the probability of the condition for biases and uncertainties in the i^{th} parameter value. These factors are not influenced by variations in the values of parameters 1 through i-1, but can be influenced by variations in the values of parameters $i+1$ through n_p.

Next, consider Bias and uncertainty model assumptions 5.1 and 5.2 hold true. Under these assumptions, $\tilde{\varphi}_i\,'(v_i,v_{i+1},\ldots,v_{n_p},\kappa)$ is replaced by $\tilde{\varphi}_i(v_i,\kappa)$ and $\tilde{\xi}_i\,'(v_i,v_{i+1},\ldots,v_{n_p},\kappa)$ is replaced by $\tilde{\xi}_i(v_i,\kappa)$. This means that the compensation factor for bias and uncertainty in the i^{th} parameter is assumed to be no longer influenced by the variations in values given to the other parameters. In particular, following a reasoning as in the discussion on Bias and uncertainty model assumption 3.1, the log-sensitivities of $\tilde{\rho}(\mathbf{1},v,\kappa)$ and $\tilde{\chi}(\kappa,\mathbf{1},v)$ will not be influenced by variations in the values given to the other parameters, and they will be influenced by value $\bar{v}$.

With these interpretations and evaluations in mind, the objections to Bias and uncertainty model assumptions 5.1 and 5.2 are similar to those to Bias and uncertainty model assumption 3.1. However, the impact of these objections is much smaller now:

In practise, it appears that, given a certain condition κ, both $\tilde{\rho}(\mathbf{1},v,\kappa)$ and $\tilde{\chi}(\kappa,\mathbf{1},v)$ are influenced by much fewer parameter values than $\rho(\mathbf{1},v)$ is. For example, under a condition in which all technical systems are operational, parameters related to the frequency of failure of such systems do not influence the conditional risk. Therefore, the approximations $\tilde{\varphi}_i(x_i,\kappa)=\tilde{\varphi}_i'(x_i,x_{i+1},\ldots,x_{n_p},\kappa)$ and $\tilde{\xi}_i(x_i,\kappa)=\tilde{\xi}_i'(x_i,x_{i+1},\ldots,x_{n_p},\kappa)$ are much better than the approximation $\varphi_i(x_i)=\varphi_i'(x_i,x_{i+1},\ldots,x_{n_p})$. In particular, the log-sensitivity of conditional risk and the log-sensitivity of the probability of the condition to the i^{th} parameter are in general not much influenced by values given to the other parameters.

<u>Conclusion:</u>
Bias and uncertainty model assumptions 5.1 and 5.2 are judged restrictive, but more reasonable than Bias and uncertainty model assumption 3.1 ("–").

Evaluation of Bias and uncertainty model assumptions 5.3 and 5.4

5.3	For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\tilde{\varphi}_i(v_i,\kappa)=(v_i/v_i)^{\beta_i(\kappa)}$, with $\beta_i(\kappa)$ the log-sensitivity of $\tilde{\rho}(\mathbf{1},v,\kappa)$ to the i^{th} parameter value around point $\bar{v}$ as defined in Subsection 5.2.
5.4	For all $i=1,\ldots,n_p$, and all $v\in\mathbb{R}^{n_p}$, $\tilde{\xi}_i(v_i,\kappa_k)=(v_i/\bar{v}_i)^{\eta_i(\kappa_k)}$, with $\eta_i(\kappa)$ the log-sensitivity of $\tilde{\chi}(\kappa,\mathbf{1},v)$ to the i^{th} parameter value around point $\bar{v}$ as defined in Subsection 5.2.

These assumptions are discussed together.

Similar as for Bias and uncertainty model assumption 3.2, discussed in Subsection 4.1, we find that, since $\tilde{\varphi}_i(v_i,\kappa)\triangleq\dfrac{\tilde{\rho}(\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},v_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}),\kappa)}{\tilde{\rho}(\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},\bar{v}_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}),\kappa)}$ and

$\tilde{\xi}_i(v_i,\kappa)\triangleq\dfrac{\tilde{\chi}(\kappa,\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},v_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}))}{\tilde{\chi}(\kappa,\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},\bar{v}_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}))}$, Bias and uncertainty model assumptions 5.3 and 5.4 state that $\ln\tilde{\rho}(\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},v_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}),\kappa)=\ln\tilde{\rho}(\mathbf{1},\bar{v},\kappa)+\beta_i(\kappa)(\ln v_i-\ln\bar{v}_i)$ and

$\ln\tilde{\chi}(\kappa,\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},v_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}))=\ln\tilde{\chi}(\kappa,\mathbf{1},\bar{v})+\eta_i(\kappa)(\ln v_i-\ln\bar{v}_i)$.

These assumptions therefore imply linearisations of $\ln\tilde{\rho}(\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},v_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}),\kappa)$ and $\ln\tilde{\chi}(\kappa,\mathbf{1},(\bar{v}_1,\ldots,\bar{v}_{i-1},v_i,\bar{v}_{i+1},\ldots,\bar{v}_{n_p}))$ around $\ln\bar{v}_i$. These linearisations are accurate if $\ln\tilde{\rho}(\mathbf{1},v,\kappa)$ and $\ln\tilde{\chi}(\kappa,\mathbf{1},v)$ are linear functions of $\ln v_i$.

As we saw in Subsection 4.1, unfortunately, in applications, $\ln\rho(\mathbf{1},v)$ is generally not a linear function of $\ln v_i$. However, in many applications, for many parameters, it does appear that

$\ln \rho(1,v)$ is approximately a *piecewise linear* function of $\ln v_i$. It appears that in the sum $\rho(1,v)=\sum_{\kappa\in\mathbf{K}}\tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)$ for some values of v the highest contribution comes from one particular κ condition, while for other values of v the highest contribution comes from another κ condition. If the product $\tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)$ is approximated by a log-linear function, then the sum $\rho(1,v)=\sum_{\kappa\in\mathbf{K}}\tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)$ becomes piecewise log-linear. Therefore, the log-linear approximation for $\tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)$ gives results that are conform the observations made in applications. In fact, in applications, for many parameters, $\tilde{\rho}(1,v,\kappa)\tilde{\chi}(\kappa,1,v)$ indeed shows log-linear properties.

And this yields that the functions $\ln \tilde{\rho}(1,v,\kappa)$ and $\ln \tilde{\chi}(\kappa,1,v)$ are approximately linear functions of $\ln v_i$, and $\beta_i(\kappa)=\left.\frac{\partial \ln \tilde{\rho}(1,v,\kappa)}{\partial \ln v_i}\right|_{\ln v=\ln \bar{v}}$ and $\eta_i(\kappa)=\left.\frac{\partial \ln \tilde{\chi}(\kappa,1,v)}{\partial \ln v_i}\right|_{\ln v=\ln \bar{v}}$ are quite accurate approximations of $\frac{\partial \ln \tilde{\rho}(1,v,\kappa)}{\partial \ln v_i}$ and $\frac{\partial \ln \tilde{\chi}(\kappa,1,v)}{\partial \ln v_i}$, respectively.

<u>Conclusion:</u>
Bias and uncertainty model assumptions 5.3 and 5.4 are judged more accurate than Bias and uncertainty model assumption 3.2 (“–”).

Evaluation of Bias and uncertainty model assumption 5.5

5.5	The random variables $V_1,\ldots,V_{n_p}$ are mutually independent.

Is equal to Bias and uncertainty model assumption 3.3.

<u>Conclusion:</u>
Bias and uncertainty model assumption 5.5 is judged as a reasonable assumption (“0”).

Evaluation of Bias and uncertainty model assumption 5.6

5.6	For each $i=1,\ldots,n_p$, the expectation and variance of $\ln V_i$ exist and satisfy: $E\{\ln V_i\}=\mu_i$ and $\mathrm{Var}\{\ln V_i\}=\sigma_i^2$.

Is equal to Bias and uncertainty model assumption 3.4.

<u>Conclusion:</u>
Bias and uncertainty model assumption 5.6 is judged as a reasonable assumption (“0”).

Evaluation of Bias and uncertainty model assumption 5.7

5.7	Each V_i $(i=1,\ldots,n_p)$ is lognormally distributed.

Is equal to Bias and uncertainty model assumption 3.5.

<u>Conclusion:</u>
Bias and uncertainty model assumption 5.7 is judged as a restrictive assumption, but in most cases it is a useful choice (“–”).

Evaluation of Bias and uncertainty model assumption 5.8

5.8	$\lim_{n_p \to \infty} \Theta_3(\kappa)/\Theta_2(\kappa) \to 0$.

The same reasoning applies as for Bias and uncertainty model assumption 3.9.

Conclusion:
Bias and uncertainty model assumption 5.8 is judged as a reasonable assumption ("0").

Evaluation of Bias and uncertainty model assumption 5.9

5.9	$\lim_{n_p \to \infty} \mathrm{T}_3(\kappa)/\mathrm{T}_2(\kappa) \to 0$

The same reasoning applies as for Bias and uncertainty model assumption 3.9.

Conclusion:
Bias and uncertainty model assumption 5.9 is judged as a reasonable assumption ("0").

Evaluation of Bias and uncertainty model assumption 5.10

5.10	The number of parameters n_p goes to infinity: $n_p \to \infty$.

Is equal to Bias and uncertainty model assumption 3.10.

Conclusion:
Bias and uncertainty model assumption 5.10 is judged as a reasonable assumption ("0").

Evaluation of Bias and uncertainty model assumption 5.11

5.11	The error made with using the Fenton-Wilkinson method for approximating the expectation and 95% credibility interval of the sum $\sum_{\kappa \in \mathbf{K}} \tilde{\rho}(\mathbf{1},V,\kappa)\tilde{\chi}(\kappa,\mathbf{1},V)$ is zero.

See the discussion in Appendix C.8.

Conclusion:
Bias and uncertainty model assumption 5.11 is judged as a reasonable assumption ("0").

6.2 Conclusion on the evaluation of Bias and uncertainty model assumptions 5.1 through 5.11

The last table in Section 5.5 shows that a practical characterisation of $\Re_{ACTUAL} = E\{\rho(A,V)\}$ can be found under two different sets of bias and uncertainty model assumptions. This table is repeated below, but the bias and uncertainty model assumption formulations have been replaced by the conclusion of the evaluation in this section.

Bias and uncertainty model assumption		$\tilde{A}$	$\tilde{C}$
5.1	Is judged restrictive, but more reasonable than Bias and uncertainty model assumption 3.1.	–	–
5.2	Is judged restrictive, but more reasonable than Bias and uncertainty model assumption 3.1.	–	–
5.3	Is judged restrictive, but more accurate than Bias and uncertainty	–	–

	model assumption 3.2.		
5.4	Is judged restrictive, but more accurate than Bias and uncertainty model assumption 3.2.	–	–
5.5	Is judged as a reasonable assumption.	0	0
5.6	Is judged as a reasonable assumption.	0	0
5.7	Is judged as a restrictive assumption, but in most cases it is a useful choice.	–	
5.8	Is judged as a reasonable assumption.		0
5.9	Is judged as a reasonable assumption.		0
5.10	Is judged as a reasonable assumption.		0
5.11	Is judged as a reasonable assumption.	0	0

It appears that there are no occurrences of "– –" in the table, hence the applicable sets of bias and uncertainty model assumptions usually hold true, or approximations exist to make them hold true. We can now compare the two sets of bias and uncertainty model assumptions with each other by counting the number of occurrences (#) of "– –", "–" and "0" and rank them. This is done in the table below.

Rank	Set	# "– –"	# "–"	# "0"
1	Set $\tilde{C}$	0	4	6
2	Set $\tilde{A}$	0	5	3

Based on this, set $\tilde{C}$ forms the most logical explanation for the approximation developed in Section 5.

6.3 Comparison with Section 4 conclusion

Finally, we can compare the results of Subsection 6.2 with the results of Subsection 4.2, by comparing the sets of bias and uncertainty model assumptions under which the approximations developed in Sections 5 and 3 hold true. Subsection 4.2 gathered evaluations of four sets of bias and uncertainty model assumptions, A, B, C, and D. Subsection 6.2 gathered evaluations on two sets $\tilde{A}$ and $\tilde{C}$. Listed together, and ranked by counting the number of occurrences (#) of "– –", "–" and "0", this yields the following table:

Rank	Set	# "– –"	# "–"	# "0"
1	Set $\tilde{C}$	0	4	6
2	Set $\tilde{A}$	0	5	3
3	Set C	2	0	4
4	Set D	2	0	5
5	Set A	2	1	2
6	Set B	2	2	3

It can be easily seen that set $\tilde{A}$ compares with set A and set $\tilde{C}$ compares with set C. The differences are that:

- Some assumptions in sets A and C have been replaced by two similar assumptions in sets $\tilde{A}$ and $\tilde{C}$.

- Sets $\tilde{A}$ and $\tilde{C}$ hold an additional assumption (i.e. Bias and uncertainty model assumption 5.11).

This makes that set $\tilde{A}$ contains 8 assumptions, whereas the comparable set A contains only 5 assumptions; set $\tilde{C}$ contains 10 assumptions whereas the comparable set C contains only 6 assumptions. On the other hand, in sets A and C there are two occurrences of "– –", while in sets $\tilde{A}$ and $\tilde{C}$ there is none. This latter factor adds a lot of credibility to the method of Section 5, compared to the method of Section 3.

7. Conclusions

Accident risk assessment for complex safety critical operations such as encountered in air traffic management has to be done by an appropriate combination of stochastic analysis and Monte Carlo simulations. Within work package 8 of the HYBRIDGE project, novel sequential Monte Carlo simulation-based methods have been developed to estimate rare event probabilities. As part of that work package, this report studied bias and uncertainty modelling in rare event estimation. Part of this study was a literature review, and a comparison with the approach developed by [Everdij and Blom, 2002].

The bias and uncertainty modelling approach developed in [Everdij and Blom, 2002] uses the model-based accident risk, together with the list of assumptions adopted during the modelling, including assumptions on the parameter values used in the model, and next compensates for the effect of these assumptions by adapting model-based accident risk with compensation factors. The outcome of the bias and uncertainty modelling is an assessment of expected ('realistic') accident risk, together with a 95% credibility interval. In HYBRIDGE WP2 tasks [HYBRIDGE D2.2] we saw that, to understand where accident risk comes from, it is often written as a sum of weighted conditional risks. And for the bias and uncertainty assessment of such sum, the approach of [Everdij and Blom, 2002] needed to be extended. This extension was developed in Section 5 of the current report. As for the earlier method, the outcome of the extended bias and uncertainty modelling is an assessment of expected ('realistic') accident risk, together with a 95% credibility interval. It expected that the extension works similarly well, except that it is more demanding on expert input time (the approach is not essentially more difficult).

The contribution of this report to existing literature can be summarised as follows:

- To the best of our knowledge, in literature there is no complete bias and uncertainty modelling that considers both parameter value assumptions and non-parameter value assumptions in an integrated way. Both types play an important role in accident risk assessments. A few references identify the different types of assumptions, but do not develop a method to deal with them. Other references consider the propagation of uncertainties in parameter values, often limited to specific models such as fault trees.
- During the development of the approximations reported in Sections 3 and 5, the assumptions adopted have been carefully gathered, under the name Bias and uncertainty model assumptions. Although approximations similar to the one in Sections 3 and 5 have been (partly) reported in other literature (for example, [Morgan and Henrion, 1990] consider uncertainty propagation which is also based on a product of lognormal contributions), these references do not explicitly gather all assumptions adopted during the development of these approximations. This makes it difficult to discuss and evaluate the feasibility of these approximations. For the approximations developed in Sections 3 and 5, the bias and uncertainty model assumptions adopted have been discussed and evaluated in Sections 4 and 6 respectively.
- The primary contribution of this report was the extension of the [Everdij and Blom, 2002] modelling to the treatment of risks that are written as sums of weighted conditional risks. It appears that more bias and uncertainty model assumptions needed to be adopted to develop the approximations of this extension, with respect to the approximations in [Everdij and Blom, 2002], but that (see Section 6) as a whole, these assumptions are generally more feasible.

8. References

[Abu-Dayya and Beaulieu, 1994] A.A. Abu-Dayya and N.C. Beaulieu, *Outage probabilities in the presence of correlated lognormal interferers*, IEEE Transactions on Vehicular Technology, Vol. 43, No. 1, pp. 164-173, February 1994.

[Ayyub, 2001] B.M. Ayyub, *Elicitation of expert opinions for uncertainty and risks*, CRC Press, Boca Raton, Florida, 2001.

[Blom et al, 2003] H.A.P. Blom, G.J. Bakker, M.H.C. Everdij, M.N.J. Van der Park, *Collision risk modelling of air traffic*, Proceedings European Control Conference, Cambridge, UK, September 2003.

[Cacuci, 2003] D.G. Cacuci, *Sensitivity and uncertainty analysis, Volume I: theory*, Chapman and Hall/CRC, 2003.

[Everdij and Blom, 2002] M.H.C. Everdij and H.A.P. Blom, *Bias and uncertainty in accident risk assessment*, NLR report CR-2002-137, National Aerospace Laboratory NLR, 2002.

[Feller, 1971] W. Feller, *An introduction to probability and its applications*, Vol. 2, John Wiley and Sons, New York, 1971.

[Fenton, 1960] L.F. Fenton, *The sum of lognormal probability distributions in scatter transmission systems*, IRE Transactions, Vol. CS-8, pp. 57-67, March 1960.

[Ferson et al., 2003] S. Ferson, V. Kreinovich, L. Ginzburg, D.S. Myers, K. Sentz, *Constructing probability boxes and Dempster-Shafer structures*, Report SAND 2002-4015, Sandia National Laboratories, January 2003

[Frey and Burmaster, 1999] H.C. Frey and D.E. Burmaster, *Methods for characterizing variability and uncertainty: Comparison of bootstrap simulation and likelihood-based approaches*, Risk Analysis, Vol. 19, No. 1, 1999.

[Haimes, 1998] Y.Y. Haimes, *Risk modeling, assessment, and management*, Wiley series in system engineering, A.P. Sage, Series Editor, Wiley and Sons, Inc, New York, NY, 1998.

[Hammit and Shlyakhter, 1999] J.K. Hammit, A.I. Shlyakhter, *The expected value of information and the probability of surprise*, Risk Analysis, Vol. 19, No. 1, pp. 135-152, 1999.

[Hattis and Anderson, 1999] D. Hattis and E.I. Anderson, *What should be the implications of uncertainty, variability, and inherent "biases"/"conservatism" for risk management decision-making*, Risk Analysis, Vol. 19, No. 1, pp. 109-130, 1999.

[Ho, 1995] C-L. Ho, *Calculating the mean and variance of power sums with two lognormal components*, IEEE Transactions on Vehicular Technology, Vol. 44, No. 4, pp. 756-762, November 1995.

[HYBRIDGE D2.2] H.A.P. Blom, G.J. Bakker, M.H.C. Everdij, M.N.J. Van der Park, *Stochastic analysis background of accident risk assessment for Air Traffic Management*, HYBRIDGE Work package 2.2 report, Version 1.1, 29th July 2003

[HYBRIDGE D8.1] J. Krystul, A. Bagchi, and H.A.P. Blom, *Risk decomposition and assessment methods*, HYBRIDGE Work package 8.1 report, Version: 1.0, 27th June 2003.

[HYBRIDGE D8.2] P. Lezaud, J. Krystul, H.A.P. Blom, *Accident risk assessment and Monte Carlo simulation methods*, HYBRIDGE Work package 8.2 report, Version 0.5, 21st June 2004

[HYBRIDGE D8.3] J. Krystul and H.A.P. Blom, *Monte Carlo simulation of rare events in hybrid systems*, HYBRIDGE Work package 8.3 report, Version 0.6, 15th June 2004

[Kumamoto and Henley, 1996] H. Kumamoto and E.J. Henley, *Probabilistic risk assessment and management for engineers and scientists*, IEEE, New York, NY, 1996.

[Leipnik, 1991] R.B. Leipnik, *On lognormal random variables: I – the characterisatic function*, J. Austral. Math. Soc. Ser. B, Vol 32, pp. 327-347, 1991.

[Morgan and Henrion, 1990] M.G. Morgan and M. Henrion, with a chapter by M. Small, *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*, Cambridge University Press, 1990.

[Mortensen, 1987] R.E. Mortensen, *Random signals and systems*, John Wiley & Sons, New York, NY, 1987.

[Nurdin, 2002] H.I. Nurdin, *Mathematical modelling of bias and uncertainty in accident risk assessment*, Final thesis report University of Twente, HYBRIDGE WP2 report, June 2002.

[Pirinen, 1998] P. Pirinen, *Sensitivity Analysis of Conditional Co-channel Interference in Cellular FDMA/TDMA Systems*. In: Leung K., Vojcic B. eds. (1999) Multiaccess, Mobility and Teletraffic for Wireless Communications: Volume 3, Kluwer Academic Publishers, pp. 171-185, 1998.

[Pirinen, 2000] P. Pirinen, *Conditional outage probability evaluation in WCDMA at high data rates*, IEEE 6th International Symposium on Spread-Spectrum Technology and Applications, NJIT, New Jersey, USA, 6-8 September 2000.

[Rudin, 1976] W. Rudin, *Principles of mathematical analysis*, McGraw-Hill Inc, Third edition, 1976.

[Safak, 1993] A. Safak, *Statistical analysis of the power sum of multiple correlated lognormal components*, IEEE Transactions on Vehicular Technology, Vol. 42, No. 1, pp. 58-61, February 1993.

[Schwartz and Yeh, 1982] S.C. Schwartz and Y.S. Yeh, *On the distribution function and moments of power sums with lognormal components*, Bell Systems Technology Journal, Vol. 61, No. 7, pp. 1441-1462, September 1982.

[Stroeve et al, 2003] S.H. Stroeve, H.A.P. Blom, M.N.J. Van der Park, M.H.C. Everdij, *Improved bias and uncertainty assessment in accident risk assessment*, NLR-Memorandum LL-2003-013, October 2003.

Appendix A: Proofs of the theorems listed in Sections 3 and 5

Proof of Theorem 3.1:

"⇒" part: (from [Everdij and Blom, 2002], proof of Theorem 1)**:**

Use Definition 3.2 to write:

$\rho(\mathbf{1},v)=\rho(\mathbf{1},(v_1,v_2,v_3,\dots,v_{n_p}))=\rho(\mathbf{1},(\overline{v}_1,v_2,v_3,\dots,v_{n_p}))\times\varphi_1'(v_1,v_2,v_3,\dots,v_{n_p})$.

Under $\varphi_i(x_i)=\varphi_i'(x_i,x_{i+1},\dots,x_{n_p})$ for i=1, this is equal to:

$\rho(\mathbf{1},(\overline{v}_1,v_2,v_3,\dots,v_{n_p}))\times\varphi_1'(v_1,v_2,v_3,\dots,v_{n_p})=\rho(\mathbf{1},(\overline{v}_1,v_2,v_3,\dots,v_{n_p}))\times\varphi_1(v_1)$.

In the same way, use Definition 3.2 and $\varphi_i(x_i)=\varphi_i'(x_i,x_{i+1},\dots,x_{n_p})$ alternately to write:

$$
\begin{aligned}
&\rho(\mathbf{1},v)=\rho(\mathbf{1},(\overline{v}_1,v_2,v_3,\dots,v_{n_p}))\times\varphi_1(v_1)=\\
&=\rho(\mathbf{1},(\overline{v}_1,\overline{v}_2,v_3,\dots,v_{n_p}))\times\varphi_2'(v_2,v_3,\dots,v_{n_p})\times\varphi_1(v_1)=\\
&=\rho(\mathbf{1},(\overline{v}_1,\overline{v}_2,v_3,\dots,v_{n_p}))\times\varphi_2(v_2)\times\varphi_1(v_1)=\\
&=\cdots=\\
&=\rho(\mathbf{1},(\overline{v}_1,\overline{v}_2,\overline{v}_3,\dots,\overline{v}_{n_p}))\times\varphi_{n_p}(v_{n_p})\times\cdots\times\varphi_2(v_2)\times\varphi_1(v_1)=\\
&=\rho(\mathbf{1},\overline{v})\times\prod_{i=1}^{n_p}\varphi_i(v_i)\,.
\end{aligned}
$$

"⇐" part (from [Nurdin, 2002], proof of Theorem 3.6).

Suppose for all v, $\rho(\mathbf{1},v)=\rho(\mathbf{1},\overline{v})\times\prod_{i=1}^{n_p}\varphi_i(v_i)$, then for any i,

$$\varphi_i'(x_i,x_{i+1},\dots,x_{n_p})=\frac{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_{i-1},x_i,x_{i+1},\dots,x_{n_p}))}{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_{i-1},\overline{v}_i,x_{i+1},\dots,x_{n_p}))}=$$

$$\frac{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))\times\prod_{j=1}^{i-1}\varphi_j(\overline{v}_j)\times\prod_{j=i}^{n_p}\varphi_j(x_j)}{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))\times\prod_{j=1}^{i}\varphi_j(\overline{v}_j)\times\prod_{j=i+1}^{n_p}\varphi_j(x_j)}=\frac{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))\times\prod_{j=1}^{i-1}\varphi_j(\overline{v}_j)\times\varphi_i(x_i)}{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))\times\prod_{j=1}^{i}\varphi_j(\overline{v}_j)}=$$

$$\frac{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))\times\prod_{j=1}^{i-1}\varphi_j(\overline{v}_j)\times\varphi_i(x_i)\times\prod_{j=i+1}^{n_p}\varphi_j(\overline{v}_j)}{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))\times\prod_{j=1}^{i}\varphi_j(\overline{v}_j)\times\prod_{j=i+1}^{n_p}\varphi_j(\overline{v}_j)}=\frac{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_{i-1},x_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))}{\rho(\mathbf{1},(\overline{v}_1,\dots,\overline{v}_i,\overline{v}_{i+1},\dots,\overline{v}_{n_p}))}=\varphi_i(x_i)\,.$$

♦

Proof of Theorem 3.2:

Under Bias and uncertainty model assumption 3.2, $\prod_{i=1}^{n_p}\varphi_i(V_i)=\prod_{i=1}^{n_p}(V_i/\overline{v}_i)^{\beta_i}$. Under Bias and uncertainty model assumptions 3.4 and 3.5, $V_i\sim\Lambda(\mu_i,\sigma_i^2)$. Under Bias and uncertainty model assumption 3.3, V_i are independent.

Use a property of the lognormal distribution: If $X_1,\dots,X_{n_p}$ are independent positive random variables, with $X_i\sim\Lambda(\varepsilon_i,\xi_i^2)$, and if $c_i>0$ and d_i are constants, $(i=1,\dots,n_p)$, then

$\prod_{i=1}^{n_p} c_i X_i^{d_i} \sim \Lambda(\sum_{i=1}^{n_p}(\ln c_i + d_i\varepsilon_i), \sum_{i=1}^{n_p} d_i^2 \xi_i^2)$. This yields:

$\prod_{i=1}^{n_p} \varphi_i(V_i) = \prod_{i=1}^{n_p}(V_i / \overline{v}_i)^{\beta_i} = \prod_{i=1}^{n_p}(\frac{1}{\overline{v}_i})^{\beta_i} \cdot V_i^{\beta_i} \sim \Lambda(\sum_{i=1}^{n_p}(-\beta_i \ln \overline{v}_i + \beta_i \mu_i), \sum_{i=1}^{n_p} \beta_i^2 \sigma_i^2)$.

Next, proceed to rewrite this, using Definition 3.3:

$\sum_{i=1}^{n_p} \beta_i(\mu_i - \ln \overline{v}_i) = \sum_{i=1}^{n_p} \beta_i(\ln \exp \mu_i - \ln \overline{v}_i) = \sum_{i=1}^{n_p} \beta_i(\ln b_i) = \sum_{i=1}^{n_p} \ln b_i^{\beta_i} = \ln \prod_{i=1}^{n_p} b_i^{\beta_i} = \ln \widehat{B}$, and

$\sum_{i=1}^{n_p} \beta_i^2 \sigma_i^2 = \sum_{i=1}^{n_p} \beta_i^2 \left(\frac{\ln \ell_i}{2}\right)^2 = \sum_{i=1}^{n_p} \frac{1}{4}(|\beta_i| \times \ln \ell_i)^2 = \frac{1}{4}\sum_{i=1}^{n_p}(\ln \ell_i^{|\beta_i|})^2 = \frac{1}{4}\widehat{U}$, hence

$\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\ln \widehat{B}, \frac{1}{4}\widehat{U})$

♦

Proof of Lemma 3.1 (from [Nurdin, 2002], proof of Lemma 3.5)**:**
It is only necessary to prove the statement for two arbitrary measurable functions $f(\cdot)$ and $g(\cdot)$, i.e. V_1 and V_2 are independent $\Rightarrow$ $f(V_1)$ and $g(V_2)$ are independent.

In probability theory, two sigma algebras $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ are said to be independent if for any set $A \in \mathcal{A}$ and $B \in \mathcal{B}$ we have that $P(A \cap B) = P(A)P(B)$. In a similar fashion two random variables X and Y are independent if the sigma-algebras $\sigma(X)$ and $\sigma(Y)$ are independent.

Now, if φ_i are measurable with respect to the Borel sigma-algebra $\Re$ of $\mathbb{R}$, then $(fV_1)^{-1}(\Re) = V_1^{-1}(f^{-1}(\Re)) \in \sigma(V_1)$ since $f^{-1}(\Re) \subset \Re$ and similarly $(fV_2)^{-1}(\Re) = V_2^{-1}(f^{-1}(\Re)) \in \sigma(V_2)$. Since V_1 and V_2 are independent we have algebras $\sigma(V_1)$ and $\sigma(V_2)$ are independent, hence so are $\sigma(fV_1) = (fV_1)^{-1}(\Re) \subset \sigma(V_1)$ and $\sigma(fV_2) = (fV_2)^{-1}(\Re) \subset \sigma(V_2)$.

♦

Proof of Lemma 3.2 (from [Everdij and Blom, 2002], proof of Lemma 2)**:**
Under Bias and uncertainty model assumption 3.6, $\varphi_i(\cdot)$ has an inverse $\varphi_i^{inv}(\cdot)$. This yields:

$$p_{\varphi_i(V_i)}(y) = \frac{d}{dy}\Pr(\varphi_i(V_i) \le y) = \frac{d}{dy}\Pr(V_i \le \varphi_i^{inv}(y)) = p_{V_i}(\varphi_i^{inv}(y))\frac{d}{dy}\varphi_i^{inv}(y) \quad \text{(A.1)}$$

Use that under Bias and uncertainty model assumptions 3.5 and 3.7, V_i and $\varphi_i(V_i)$ are both lognormally distributed, say $V_i \sim \Lambda(\mu_i, \sigma_i^2)$ and $\varphi_i(V_i) \sim \Lambda(\gamma_i, \theta_i^2)$, such that (A.1) yields

$$\frac{1}{y\theta_i\sqrt{2\pi}}\exp\left(-\frac{(\ln(y)-\gamma_i)^2}{2\theta_i^2}\right) = \frac{1}{\varphi_i^{inv}(y)\sigma_i\sqrt{2\pi}}\exp\left(-\frac{(\ln(\varphi_i^{inv}(y))-\mu_i)^2}{2\sigma_i^2}\right)\frac{d}{dy}\varphi_i^{inv}(y)$$

or,

$$\int \frac{d}{dy}\varphi_i^{inv}(y)dy = \int \frac{\varphi_i^{inv}(y)\sigma_i}{y\theta_i}\exp\left(\frac{(\ln(\varphi_i^{inv}(y))-\mu_i)^2}{2\sigma_i^2} - \frac{(\ln(y)-\gamma_i)^2}{2\theta_i^2}\right)dy . \quad \text{(A.2)}$$

This integral does not have a closed form solution. However, the fundamental theorem of calculus (e.g. Rudin, 1976; theorem 6.21) states that if the integral has a differentiable solution, then this solution is unique. Therefore, try a solution such that $\varphi_i(V_i) = \alpha_i V_i^{\beta_i}$, for particular α_i and β_i. We know that $\varphi_i(\overline{v}_i) = 1$, hence $\alpha_i = \overline{v}_i^{-\beta_i}$ and therefore $\varphi_i(V_i) = (V_i / \overline{v}_i)^{\beta_i}$. This gives $\varphi_i^{inv}(y) = \overline{v}_i \sqrt[\beta_i]{y}$ and $\frac{d}{dy}\varphi_i^{inv}(y) = \frac{\overline{v}_i}{y\beta_i}\sqrt[\beta_i]{y}$. Substitution into (A.2) yields that this solution satisfies the integral if there exists β_i such that both $\theta_i = \beta_i \sigma_i$ and $\gamma_i = \beta_i(\mu_i - \ln \overline{v}_i)$, which are the conditions of Bias and uncertainty model assumptions 3.4 and 3.8. (Note that this β_i exists if $\frac{\theta_i}{\sigma_i} = \frac{\gamma_i}{\mu_i - \ln \overline{v}_i}$).

♦

Proof of Theorem 3.3:
Using Lemma 3.2, we find that under Bias and uncertainty model assumptions 3.4, 3.5, 3.6, 3.7, 3.8, $\varphi_i(v_i) = (v_i / \overline{v}_i)^{\beta_i}$, which is equivalent to Bias and uncertainty model assumption 3.2. Therefore, use the proof of Theorem 3.2 by replacing Bias and uncertainty model assumption 3.2 by Bias and uncertainty model assumptions 3.4, 3.5, 3.6, 3.7, 3.8.

♦

Proof of Lemma 3.3 (from [Everdij and Blom, 2002], proof of Lemma 3)**:**
For the proof of Lemma 3.3, we use Lyapunov's Extended Central Limit Theorem (e.g. Feller, 1971):

Extended Central Limit Theorem (Lyapunov):
Suppose,
$H_1, \ldots, H_{n_p}$ are independent random variables with possibly different distributions,
$\gamma_i = E\{H_i\}$, $\theta_{i,2} = E\{(H_i - \gamma_i)^2\}$ and $\theta_{i,3} = E\{|H_i - \gamma_i|^3\}$, where $\theta_{i,3}$ must exist.
Moreover, let:

$$\Gamma = \sum_{i=1}^{n_p} \gamma_i \,,\quad \Theta_2 = \sqrt{\sum_{i=1}^{n_p} \theta_{i,2}} \,, \text{ and } \quad \Theta_3 = \sqrt[3]{\sum_{i=1}^{n_p} \theta_{i,3}} \,.$$

Then, if $\lim_{n_p \to \infty} \Theta_3 / \Theta_2 \to 0$, then for $n_p \to \infty$, $H_1 + \cdots + H_{n_p}$ is Normally (Gaussian) distributed with mean Γ and variance Θ_2^2. Notation: $H_1 + \cdots + H_{n_p} \sim N(\Gamma, \Theta_2^2)$ for $n_p \to \infty$.

♦

Now we get to the actual proof of Lemma 3.3:
According to Lemma 3.1, under Bias and uncertainty model assumptions 3.3 and 3.6, $\varphi_i(V_i)$ are independent random variables (with possibly different distributions). This yields that $H_i = \ln \varphi_i(V_i)$ are also independent random variables (with possibly different distributions). Lyapunov's Extended Central Limit Theorem, which holds under Bias and uncertainty model assumptions 3.3, 3.6, 3.9 and 3.10, gives that $\sum_{i=1}^{n_p} H_i \sim N(\Gamma, \Theta_2^2)$. Now, use $\ln(\exp(\sum_{i=1}^{n_p} H_i)) = \sum_{i=1}^{n_p} H_i \sim N(\Gamma, \Theta_2^2)$, and use that a variable is Lognormally distributed with parameters Γ and Θ_2^2 if its natural logarithm is Normally distributed with the same parameters. This yields:

$$\prod_{i=1}^{n_p} \varphi_i(V_i) = \prod_{i=1}^{n_p} \exp(H_i) = \exp(\sum_{i=1}^{n_p} H_i) \sim \Lambda(\Gamma, \Theta_2^2) \text{ if } n_p \to \infty .$$

♦

Proof of Theorem 3.4: (from [Nurdin, 2002], proof of Theorem 3.7.**)**
Using Lemma 3.3, under Bias and uncertainty model assumptions 3.3, 3.6, 3.9 and 3.10, we find $\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\Gamma, \Theta_2^2)$. Use Definition 3.4 and Bias and uncertainty model assumptions 3.4 and 3.8 and Definition 3.3 to obtain

$$\Gamma \triangleq \sum_{i=1}^{n_p} E\{\ln \varphi_i(V_i)\} = \sum_{i=1}^{n_p} \beta_i(\mu_i - \ln \overline{v}_i) = \ln \widehat{B} \text{ and}$$

$$\Theta_2^2 \triangleq \sum_{i=1}^{n_p} E\{(\ln \varphi_i(V_i) - E\{\ln \varphi_i(V_i)\})^2\} = \sum_{i=1}^{n_p} \beta_i^2 \sigma_i^2 = \frac{1}{4}\widehat{U}.$$

♦

Proof of Theorem 3.5:
Using Lemma 3.2, we find that under Bias and uncertainty model assumptions 3.4, 3.5, 3.6, 3.7, 3.8, $\varphi_i(v_i) = (v_i / \overline{v}_i)^{\beta_i}$, which is equivalent to Bias and uncertainty model assumption 3.2. Therefore, use the proof of Theorem 3.4 by replacing Bias and uncertainty model assumption 3.2 by Bias and uncertainty model assumptions 3.4, 3.5, 3.6, 3.7, 3.8.

♦

Proof of Corollary 3.1 (from [Everdij and Blom, 2002], combination of proof of Theorems 2, 3, 4, 5, and Corollaries 1, 2 and 3)**:**
From Theorem 3.1 we have $\rho(\mathbf{1}, V) = \rho(\mathbf{1}, \overline{v}) \times \prod_{i=1}^{n_p} \varphi_i(V_i)$. From Theorem 3.2 (under Bias and uncertainty model assumptions 3.2, 3.3, 3.4, 3.5), from Theorem 3.3 (under Bias and uncertainty model assumptions 3.3, 3.4, 3.5, 3.6, 3.7, 3.8), from Theorem 3.4 (under Bias and uncertainty model assumptions 3.2, 3.3, 3.4, 3.9, 3.10) and from Theorem 3.5 (under Bias and uncertainty model assumptions 3.3, 3.4, 3.6, 3.8, 3.9, 3.10) we have $\prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\ln \widehat{B}, \frac{1}{4}\widehat{U})$.

Use a property of the lognormal distribution: if $X \sim \Lambda(\varepsilon, \xi^2)$ and $a > 0$ then $aX \sim \Lambda(\ln a + \varepsilon, \xi^2)$. This gives

$$\rho(\mathbf{1}, V) = \rho(\mathbf{1}, \overline{v}) \times \prod_{i=1}^{n_p} \varphi_i(V_i) \sim \Lambda(\ln \rho(\mathbf{1}, \overline{v}) + \ln \widehat{B}, \frac{1}{4}\widehat{U}) = \Lambda(\ln[\rho(\mathbf{1}, \overline{v})\widehat{B}], \frac{1}{4}\widehat{U}).$$

Next, use another property of the Lognormal distribution: If $X \sim \Lambda(\varepsilon, \xi^2)$, then $E\{X\} = \exp(\varepsilon + \xi^2 / 2)$ and $\Pr(X \in [\exp(\varepsilon - 2|\xi|), \exp(\varepsilon + 2|\xi|) = 0.95$. This yields:

$$E\{\rho(\mathbf{1}, V)\} = E\{\rho(\mathbf{1}, \overline{v}) \times \prod_{i=1}^{n_p} \varphi_i(V_i)\} = \exp(\ln[\rho(\mathbf{1}, \overline{v})\widehat{B}] + \frac{1}{8}\widehat{U}) = \rho(\mathbf{1}, \overline{v}) \times \widehat{B} \times \exp(\frac{1}{8}\widehat{U}), \text{ and}$$

$$\Pr(\rho(\mathbf{1}, V) \in [\exp(\ln[\rho(\mathbf{1}, \overline{v})\widehat{B}] - 2\sqrt{\frac{1}{4}\widehat{U}}), \exp(\ln[\rho(\mathbf{1}, \overline{v})\widehat{B}] + 2\sqrt{\frac{1}{4}\widehat{U}})]) = 0.95, \text{ or}$$

$$\Pr(\rho(\mathbf{1}, V) \in [\widehat{B} \times \exp(-\sqrt{\widehat{U}}), \widehat{B} \times \exp(\sqrt{\widehat{U}})]) = 0.95.$$

♦

Proof of Theorem 3.6 (from [Everdij and Blom, 2002], proof of Theorem 6)**:**
Consider the first Boolean variable A_1, such that $A_1 = 1$ if non-parameter assumption $\overline{a}_1$ holds true and $A_1 = 0$ if $\overline{a}_1$ does not hold true. Next, use the law of total probability:

$$E\{\rho(A, V)\} = E\{\rho(A, V) \mid A_1 = 1\} \times \Pr(A_1 = 1) + E\{\rho(A, V) \mid A_1 = 0\} \times \Pr(A_1 = 0) \quad \text{(A.3)}$$

Here, $E\{\rho(A,V) \mid A_1 = 1\}$ denotes expected actual accident risk, conditional on non-parameter assumption $\bar{a}_1$ holding true, $\Pr(A_1 = 1)$ denotes the probability that non-parameter assumption $\bar{a}_1$ holds true, $E\{\rho(A,V) \mid A_1 = 0\}$ denotes expected actual accident risk, conditional on non-parameter assumption $\bar{a}_1$ not holding true and $\Pr(A_1 = 0)$ denotes the probability that non-parameter assumption $\bar{a}_1$ does not hold true.
Equation (A.3) can be rewritten into:

$$\frac{E\{\rho(A,V)\}}{E\{\rho(A,V) \mid A_1 = 1\}} = \Pr(A_1 = 0) \times \left\{ \frac{E\{\rho(A,V) \mid A_1 = 0\}}{E\{\rho(A,V) \mid A_1 = 1\}} + \frac{\Pr(A_1 = 1)}{\Pr(A_1 = 0)} \right\} \tag{A.4}$$

Next, consider the second Boolean variable A_2:

Use the following equality:

$$E\{\rho(A,V) \mid A_1 = 1\} = E\{\rho(A,V) \mid A_1 = 1, A_2 = 1\} \times \Pr(A_2 = 1 \mid A_1 = 1) + E\{\rho(A,V) \mid A_1 = 1, A_2 = 0\} \times \Pr(A_2 = 0 \mid A_1 = 1)$$

and write this into:

$$\frac{E\{\rho(A,V) \mid A_1 = 1\}}{E\{\rho(A,V) \mid A_1 = 1, A_2 = 1\}} = \Pr(A_2 = 0 \mid A_1 = 1) \times \left\{ \frac{E\{\rho(A,V) \mid A_1 = 1, A_2 = 0\}}{E\{\rho(A,V) \mid A_1 = 1, A_2 = 1\}} + \frac{\Pr(A_2 = 1 \mid A_1 = 1)}{\Pr(A_2 = 0 \mid A_1 = 1)} \right\} \tag{A.5}$$

Proceed in the same way with Boolean variables A_3 through A_{n_a}, the equation for the last Boolean variable being:

$$\frac{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a - 1} = 1\}}{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a} = 1\}} = \Pr(A_{n_a} = 0 \mid A_1 = 1, \ldots, A_{n_a - 1} = 1) \times$$
$$\times \left\{ \frac{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a - 1} = 1, A_{n_a} = 0\}}{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a} = 1\}} + \frac{\Pr(A_{n_a} = 1 \mid A_1 = 1, \ldots, A_{n_a - 1} = 1)}{\Pr(A_{n_a} = 0 \mid A_1 = 1, \ldots, A_{n_a - 1} = 1)} \right\} \tag{A.6}$$

If we multiply the left-hand sides of Equations (A.4) through (A.6), we get

$$\frac{E\{\rho(A,V)\}}{E\{\rho(A,V) \mid A_1 = 1\}} \times \frac{E\{\rho(A,V) \mid A_2 = 1\}}{E\{\rho(A,V) \mid A_1 = 1, A_2 = 1\}} \times \cdots \times \frac{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a - 1} = 1\}}{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a} = 1\}} = \frac{E\{\rho(A,V)\}}{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{n_a} = 1\}}$$

which equals Ψ. This yields that Ψ can be determined by multiplying all right-hand sides of Equations (A.4) through (A.6), resulting in

$$\Psi = \prod_{i=1}^{n_a} \Pr(A_i = 0 \mid A_1 = 1, \ldots, A_{i-1} = 1) \times$$
$$\times \left\{ \frac{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_{i-1} = 1, A_i = 0\}}{E\{\rho(A,V) \mid A_1 = 1, \ldots, A_i = 1\}} + \frac{\Pr(A_i = 1 \mid A_1 = 1, \ldots, A_{i-1} = 1)}{\Pr(A_i = 0 \mid A_1 = 1, \ldots, A_{i-1} = 1)} \right\}$$

Some rearranging of terms provides the result.

♦

Proof of Lemma 5.1:
This is analogous to the proof of Theorem 3.1.

♦

Proof of Lemma 5.2:
This is analogous to the proof of Theorem 3.1.

♦

Proof of Theorem 5.1:

Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4}, we have:

$\tilde{\rho}(1,V,\kappa)=\tilde{\rho}(1,\overline{v},\kappa)\times\prod_{i=1}^{n_p}\left(\frac{V_i}{\overline{v}_i}\right)^{\beta_i(\kappa)}$ and $\tilde{\chi}(\kappa,1,V)=\tilde{\chi}(\kappa,1,\overline{v})\times\prod_{i=1}^{n_p}\left(\frac{V_i}{\overline{v}_i}\right)^{\eta_i(\kappa)}$. This yields that:

$$\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)=\tilde{\rho}(1,\overline{v},\kappa)\times\prod_{i=1}^{n_p}(V_i/\overline{v}_i)^{\beta_i(\kappa)}\times\tilde{\chi}(\kappa,1,\overline{v})\times\prod_{i=1}^{n_p}(V_i/\overline{v}_i)^{\eta_i(\kappa)}=$$

$=\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\times\prod_{i=1}^{n_p}\left(\frac{V_i}{\overline{v}_i}\right)^{\beta_i(\kappa)+\eta_i(\kappa_k)}$. Under Bias and uncertainty model assumptions 5.6 and 5.7, $V_i\sim\Lambda(\mu_i,\sigma_i^2)$. Under Bias and uncertainty model assumption 5.5, V_i are independent.

Use a property of the lognormal distribution: If $X_1,\ldots,X_{n_p}$ are independent positive random variables, with $X_i\sim\Lambda(\varepsilon_i,\xi_i^2)$, and if $c_i>0$ and d_i are constants, $(i=1,\ldots,n_p)$, then $\prod_{i=1}^{n_p}c_iX_i^{d_i}\sim\Lambda(\sum_{i=1}^{n_p}(\ln c_i+d_i\varepsilon_i),\sum_{i=1}^{n_p}d_i^2\xi_i^2)$. This yields: $\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\times\prod_{i=1}^{n_p}\left(\frac{V_i}{\overline{v}_i}\right)^{\beta_i(\kappa)+\eta_i(\kappa)}\sim$

$$\sim\Lambda(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})]+\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))(\mu_i-\ln\overline{v}_i),\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))^2\sigma_i^2).$$

Next, proceed to rewrite this, using Definition 5.5:

$$\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))(\mu_i-\ln\overline{v}_i)=\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))(\ln\exp\mu_i-\ln\overline{v}_i)=\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))\ln b_i=$$

$=\sum_{i=1}^{n_p}\ln b_i^{\beta_i(\kappa)+\eta_i(\kappa)}=\ln\left[\prod_{i=1}^{n_p}b_i^{\beta_i(\kappa)+\eta_i(\kappa)}\right]=\ln\left[\prod_{i=1}^{n_p}b_i^{\beta_i(\kappa)}\prod_{i=1}^{n_p}b_i^{\eta_i(\kappa)}\right]=\ln[\widehat{B}_\kappa\widehat{M}_\kappa]$, and

$$\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))^2\sigma_i^2=\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))^2\left(\frac{\ln\ell_i}{2}\right)^2=\sum_{i=1}^{n_p}\frac{1}{4}(|\beta_i(\kappa)+\eta_i(\kappa)|\times\ln\ell_i)^2=$$

$=\frac{1}{4}\sum_{i=1}^{n_p}(\ln\ell_i^{|\beta_i(\kappa)+\eta_i(\kappa)|})^2=\frac{1}{4}\widehat{G}_\kappa$, which proves the result.

♦

Proof of Theorem 5.2:

Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4}, we have:

$\tilde{\rho}(1,V,\kappa)=\tilde{\rho}(1,\overline{v},\kappa)\times\prod_{i=1}^{n_p}\left(\frac{V_i}{\overline{v}_i}\right)^{\beta_i(\kappa)}$, or $\ln\tilde{\rho}(1,V,\kappa)=\ln\tilde{\rho}(1,\overline{v},\kappa)+\sum_{i=1}^{n_p}\beta_i(\kappa)(\ln V_i-\ln\overline{v}_i)$, and

$\tilde{\chi}(\kappa,1,V)=\tilde{\chi}(\kappa,1,\overline{v})\times\prod_{i=1}^{n_p}\left(\frac{V_i}{\overline{v}_i}\right)^{\eta_i(\kappa)}$, or $\ln\tilde{\chi}(\kappa,1,V)=\ln\tilde{\chi}(\kappa,1,\overline{v})+\sum_{i=1}^{n_p}\eta_i(\kappa)(\ln V_i-\ln\overline{v}_i)$.

Under Bias and uncertainty model assumption 5.5, $\ln V_i$ are independent, hence $\ln\tilde{\rho}(1,V,\kappa)$ and $\ln\tilde{\chi}(\kappa,1,V)$ are both written as sums of independent random variables. For both, we can apply Lyapunov's Extended Central Limit Theorem (see the proof of Lemma 3.3). Under Bias and uncertainty model assumptions 5.5, 5.8 and 5.10, this yields, similar as in the proof of Lemma 3.3 and Theorem 3.4:

$\ln\tilde{\rho}(1,V,\kappa)=\ln\tilde{\rho}(1,\overline{v},\kappa)+\sum_{i=1}^{n_p}\beta_i(\kappa)(\ln V_i-\ln\overline{v}_i)\sim N(\ln\tilde{\rho}(1,\overline{v},\kappa)+\Gamma(\kappa),\Theta_2^2(\kappa))$, while under Bias and uncertainty model assumptions 5.5, 5.9 and 5.10, this yields,

$\ln\tilde{\chi}(\kappa,1,V)=\ln\tilde{\chi}(\kappa,1,\overline{v})+\sum_{i=1}^{n_p}\eta_i(\kappa)(\ln V_i-\ln\overline{v}_i)\sim N(\ln\tilde{\chi}(\kappa,1,\overline{v})+\mathrm{D}(\kappa),\mathrm{T}_2^2(\kappa))$.

Now, we know that the sum of two Gaussian random variables is again Gaussian, with an expectation equal to the sum of the expectations of the two Gaussians, and a variance equal to the sum of the variances of the two Gaussians plus twice the covariance of the two Gaussians. In other words, $\ln\tilde{\rho}(1,V,\kappa)+\ln\tilde{\chi}(\kappa,1,V)\sim$

$\sim N(\ln\tilde{\rho}(1,\overline{v},\kappa)+\Gamma(\kappa)+\ln\tilde{\chi}(\kappa,1,\overline{v})+D(\kappa),\Theta_2^2+\mathrm{T}_2^2+2\mathrm{Cov}(\ln\tilde{\rho}(1,V,\kappa),\ln\tilde{\chi}(\kappa,1,V)))$

Hence, what we need to find is the covariance of $\ln\tilde{\rho}(1,V,\kappa)$ and $\ln\tilde{\chi}(\kappa,1,V)$.

$$\begin{aligned}&\mathrm{Cov}(\ln\tilde{\rho}(1,V,\kappa),\ln\tilde{\chi}(\kappa,1,V))=\\&E\{(\ln\tilde{\rho}(1,V,\kappa)-E\{\ln\tilde{\rho}(1,V,\kappa)\})(\ln\tilde{\chi}(\kappa,1,V)-E\{\ln\tilde{\chi}(\kappa,1,V)\})\}=\\&E\{\ln\tilde{\rho}(1,V,\kappa)\ln\tilde{\chi}(\kappa,1,V)\}-E\{\ln\tilde{\rho}(1,V,\kappa)\}E\{\ln\tilde{\chi}(\kappa,1,V)\}\end{aligned}$$

Now, use $\ln\tilde{\rho}(1,V,\kappa)=\ln\tilde{\rho}(1,\overline{v},\kappa)+\sum_{i=1}^{n_p}\beta_i(\kappa)(\ln V_i-\ln\overline{v}_i)$ and

$\ln\tilde{\chi}(\kappa,1,V)=\ln\tilde{\chi}(\kappa,1,\overline{v})+\sum_{i=1}^{n_p}\eta_i(\kappa)(\ln V_i-\ln\overline{v}_i)$, and

$E\{\ln\tilde{\rho}(1,V,\kappa)\}=\ln\tilde{\rho}(1,\overline{v},\kappa)+\Gamma(\kappa)$ and $E\{\ln\tilde{\chi}(\kappa,1,V)\}=\ln\tilde{\chi}(\kappa,1,\overline{v})+\mathrm{D}(\kappa)$, with $\Gamma(\kappa)=\sum_{i=1}^{n_p}\beta_i(\kappa)E\{\ln V_i-\ln\overline{v}_i\}$ and $\mathrm{D}(\kappa)=\sum_{i=1}^{n_p}\eta_i(\kappa)E\{\ln V_i-\ln\overline{v}_i\}$ and the fact that under Bias and uncertainty model assumption 5.3, V_i are independent, hence $E\{\ln V_i\times\ln V_j\}=E\{\ln V_i\}\times E\{\ln V_j\}$ for $i\neq j$. This yields, after some straightforward evaluations: $\mathrm{Cov}(\ln\tilde{\rho}(1,V,\kappa),\ln\tilde{\chi}(\kappa,1,V))=$

$$\begin{aligned}&E\{\ln\tilde{\rho}(1,V,\kappa)\ln\tilde{\chi}(\kappa,1,V)\}-E\{\ln\tilde{\rho}(1,V,\kappa)\}E\{\ln\tilde{\chi}(\kappa,1,V)\}=\\&E\{\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)(\ln V_i-\ln\overline{v}_i)^2\}-\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)(E\{\ln V_i-\ln\overline{v}_i\})^2=\\&\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)\left(E\{(\ln V_i-\ln\overline{v}_i)^2\}-(E\{\ln V_i-\ln\overline{v}_i\})^2\right)=\end{aligned}$$

$\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)\mathrm{Var}\{\ln V_i-\ln\overline{v}_i\}=\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)\sigma_i^2$. Therefore, the variance of $\ln[\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)]$ is equal to $\Theta_2^2+\mathrm{T}_2^2+\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)\sigma_i^2$. Now, use that $\Theta_2^2=\sum_{i=1}^{n_p}\beta_i^2(\kappa)\sigma_i^2$ and $\mathrm{T}_2^2=\sum_{i=1}^{n_p}\eta_i^2(\kappa)\sigma_i^2$ to get $\Theta_2^2+\mathrm{T}_2^2+\sum_{i=1}^{n_p}\beta_i(\kappa)\eta_i(\kappa)\sigma_i^2=$

$$\sum_{i=1}^{n_p}(\beta_i^2(\kappa)+\eta_i^2(\kappa)+2\beta_i(\kappa)\eta_i(\kappa))\sigma_i^2=\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))^2\sigma_i^2=$$

$$\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))^2\frac{1}{4}(\ln\ell_i)^2=\frac{1}{4}\sum_{i=1}^{n_p}(\ln\ell_i^{|\beta_i(\kappa)+\eta_i(\kappa)|})^2=\frac{1}{4}\widehat{G}_\kappa$$

Therefore, $\ln[\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)]\sim N(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa],\frac{1}{4}\widehat{G}_\kappa)$ or,

$$\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)\sim\Lambda(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa],\frac{1}{4}\widehat{G}_\kappa)$$

♦

Proof of Theorem 5.3:
From Theorems 5.1 and 5.2, we have

$$\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)\sim\Lambda(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa],\frac{1}{4}\widehat{G}_\kappa)$$

Under Bias and uncertainty model assumption 5.11, we can use Fenton-Wilkinson's method by defining:

$$X_\kappa\triangleq\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)\sim\Lambda(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa],\frac{1}{4}\widehat{G}_\kappa)$$

$$Z_\kappa\triangleq\ln X_\kappa=\ln\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)\sim N(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa],\frac{1}{4}\widehat{G}_\kappa)$$

$$X\triangleq\rho(1,V)=\sum_{\kappa\in\mathbf{K}}\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)=\sum_{\kappa\in\mathbf{K}}X_\kappa\,.$$

This yields:

$$u_1=\sum_{\kappa\in\mathbf{K}}\exp(\mu_{Z_\kappa}+\sigma_{Z_\kappa}^2/2)=\sum_{\kappa\in\mathbf{K}}\exp(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa]+(\frac{1}{4}\widehat{G}_\kappa)/2)=$$

$$=\sum_{\kappa\in\mathbf{K}}\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa\times\exp(\frac{1}{8}\widehat{G}_\kappa)$$

$$u_2=\sum_{\kappa\in\mathbf{K}}\exp(2\mu_{Z_\kappa}+2\sigma_{Z_\kappa}^2)+2\sum_{\substack{\kappa,\kappa'\in\mathbf{K}\\ \kappa'>\kappa}}\exp(\mu_{Z_\kappa}+\mu_{Z_{\kappa'}})\exp(\frac{1}{2}(\sigma_{Z_\kappa}^2+\sigma_{Z_{\kappa'}}^2+2r_{\kappa\kappa'}\sigma_{Z_\kappa}\sigma_{Z_{\kappa'}}))=$$

$$=\sum_{\kappa\in\mathbf{K}}\exp(2\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa]+2\times\frac{1}{4}\widehat{G}_\kappa)+$$

$$+2\sum_{\substack{\kappa,\kappa'\in\mathbf{K}\\ \kappa'>\kappa}}\exp(\ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa]+\ln[\tilde{\rho}(1,\overline{v},\kappa')\tilde{\chi}(\kappa',1,\overline{v})\widehat{B}_{\kappa'}\widehat{M}_{\kappa'}])\times\exp(\frac{1}{2}C_{\kappa\kappa'})=$$

$$=\sum_{\kappa\in\mathbf{K}}(\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa)^2\times\exp(\frac{1}{2}\widehat{G}_\kappa)+$$

$$+2\sum_{\substack{\kappa,\kappa'\in\mathbf{K}\\ \kappa'>\kappa}}\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa\widehat{M}_\kappa\tilde{\rho}(1,\overline{v},\kappa')\tilde{\chi}(\kappa',1,\overline{v})\widehat{B}_{\kappa'}\widehat{M}_{\kappa'}\times\exp(\frac{1}{2}C_{\kappa\kappa'})\,.$$

Here, $C_{\kappa\kappa'}$ is equal to

$$C_{\kappa\kappa'}=\frac{1}{4}\widehat{G}_\kappa+\frac{1}{4}\widehat{G}_{\kappa'}+2\mathrm{Cov}\{\ln[\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)],\ln[\tilde{\rho}(1,V,\kappa')\tilde{\chi}(\kappa',1,V)]\}\,.$$

Under Bias and uncertainty model assumptions {5.1, 5.2, 5.3, 5.4, 5.5}, $\tilde{\rho}(1,V,\kappa) = \tilde{\rho}(1,\overline{v},\kappa) \times \prod_{i=1}^{n_p} \left(\frac{V_i}{\overline{v}_i}\right)^{\beta_i(\kappa)}$ and $\tilde{\chi}(\kappa,1,V) = \tilde{\chi}(\kappa,1,\overline{v}) \times \prod_{i=1}^{n_p} \left(\frac{V_i}{\overline{v}_i}\right)^{\eta_i(\kappa_k)}$ and V_i are independent, hence we can evaluate:

$$\mathrm{Cov}\{\ln[\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)], \ln[\tilde{\rho}(1,V,\kappa')\tilde{\chi}(\kappa',1,V)]\} =$$
$$E\Big\{(\ln[\tilde{\rho}(1,V,\kappa)\tilde{\chi}(\kappa,1,V)] - \ln[\tilde{\rho}(1,\overline{v},\kappa)\tilde{\chi}(\kappa,1,\overline{v})\widehat{B}_\kappa \widehat{M}_\kappa]) \times$$
$$\times(\ln[\tilde{\rho}(1,V,\kappa')\tilde{\chi}(\kappa',1,V)] - \ln[\tilde{\rho}(1,\overline{v},\kappa')\tilde{\chi}(\kappa',1,\overline{v})\widehat{B}_{\kappa'} \widehat{M}_{\kappa'}])\Big\} =$$
$$\sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))(\beta_i(\kappa')+\eta_i(\kappa'))\sigma_i^2 = \sum_{i=1}^{n_p}(\beta_i(\kappa)+\eta_i(\kappa))(\beta_i(\kappa')+\eta_i(\kappa'))(\frac{1}{2}\ln \ell_i)^2 = \frac{1}{4}\widehat{G}_{\kappa\kappa'},$$

with $\widehat{G}_{\kappa\kappa'}$ defined in Definition 5.7. This evaluation is lengthy but straightforward, and is omitted here.

♦

Proof of Corollary 5.1:
Follows immediately from Theorem 5.3, using that if $\rho(1,V) \sim \Lambda(\mu_W, \sigma_W^2)$ then

$E\{\rho(1,V)\} = \exp(\mu_W + \sigma_W^2/2) = \exp(2\ln u_1 - \tfrac{1}{2}\ln u_2 + \tfrac{1}{2}\ln u_2 - \ln u_1) = u_1$ and

$\Pr(\rho(1,V) \in [\exp(\mu_W - 2\sigma_W), \exp(\mu_W + 2\sigma_W)]) = 0.95$, hence

$$\Pr(\rho(1,V) \in [\exp(2\ln u_1 - \tfrac{1}{2}\ln u_2 - 2\sqrt{\ln u_2 - 2\ln u_1}), \exp(2\ln u_1 - \tfrac{1}{2}\ln u_2 + 2\sqrt{\ln u_2 - 2\ln u_1})]) = 0.95$$

or

$$\Pr(\rho(1,V) \in \left[\frac{u_1^2}{\sqrt{u_2}}\exp(-2\sqrt{\ln\frac{u_2}{u_1^2}}), \frac{u_1^2}{\sqrt{u_2}}\exp(2\sqrt{\ln\frac{u_2}{u_1^2}})\right]) = 0.95\,.$$

♦

Proof of Theorem 5.4:
Equal to the proof of Theorem 3.6

♦

Proof of Corollary 5.2:
Immediately follows from Corollary 5.1 and Theorem 5.4.

♦

Appendix B: Properties of the lognormal distribution

B.1 Definition of lognormal

X is lognormally distributed (notation $X \sim \Lambda(\mu,\sigma^2)$) if it can be written as
$X = \exp(Z)$, with Z Gaussian distributed (notation $Z \sim N(\mu,\sigma^2)$, hence $\ln X \sim N(\mu,\sigma^2)$).

Note:
In some literature, the definition of lognormal is given using a 10-base logarithm:
X is lognormal if it can be written as $X = 10^{Y/10}$, with Y Gaussian, hence $Y = 10 \times {}^{10}\log X$.
Since ${}^{10}\log X = \ln X / \ln 10$, we find that $Y = Z\frac{10}{\ln 10} \sim N(\mu\frac{10}{\ln 10},(\sigma\frac{10}{\ln 10})^2)$, with $\frac{10}{\ln 10} \approx 4.343$.
The Gaussian variable Y, and in particular its standard deviation, is commonly measured in dB (decibel). The standard deviation of Y is called the *decibel spread*.

B.2 Cumulative distribution function

$$\Pr(X \le x) = F_X(x) = \int_0^x \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right) dx.$$

This can be written as $F_X(x) = 1 - Q\left(\frac{\ln x-\mu}{\sigma}\right)$ or $F_X(x) = \frac{1}{2}\left(1+erf\left(\frac{\ln x-\mu}{\sigma\sqrt{2}}\right)\right)$,

where Q is the Gaussian tail function defined by $Q(z) = \frac{1}{\sqrt{2\pi}}\int_z^\infty \exp(-\frac{1}{2}t^2)dt$ and erf is the error function defined by $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$, hence $Q(z) = \frac{1}{2} - \frac{1}{2}erf\left(\frac{z}{\sqrt{2}}\right)$.

B.3 Complementary distribution function

$$\Pr(X > x) = \int_x^\infty \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right) dx = Q\left(\frac{\ln x-\mu}{\sigma}\right),$$ with Q as above.

B.4 Probability density function

$$f_X(x) = \frac{1}{\sigma x\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), \quad 0 \le x \le \infty$$

In the figure below, the probability density functions of some lognormal variables are plotted and compared with normal densities.

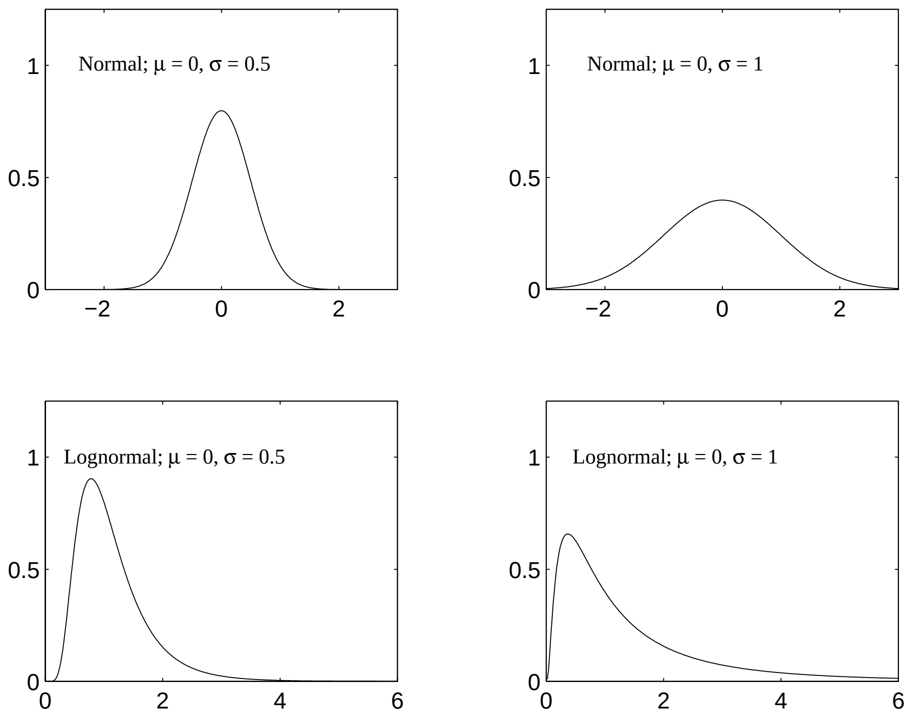

Figure B.1 *Example of Lognormal probability densities, compared with Normal probability densities.*

B.5 Characteristic function, moment generating function, raw moment, cumulant

The **characteristic function** $\phi_X(t)$ is defined as $\phi_X(t) = \int_{-\infty}^{\infty} \exp(itx) f_X(x)dx = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \alpha_X(k)$,

where $\alpha_X(k)$ is the k^{th} **raw moment about 0**, i.e. $\alpha_X(k) = E\{X^k\} = \int_{-\infty}^{\infty} x^k f_X(x)dx$ and $\alpha_X(0) \equiv 1$. It is easy to see that $\left.\frac{d^k}{dt^k}\phi_X(t)\right|_{t=0} = i^k \alpha_X(k)$ and $f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi_X(t)\exp(-itx)dt$.

The **moment generating function** $M_X(t)$ is defined as $E\{\exp(tX)\}$, or $M_X(t) = \int_{-\infty}^{\infty} \exp(tx) f_X(x)dx$. It is easy to see that $\left.\frac{d^k}{dt^k} M_X(t)\right|_{t=0} = \alpha_X(k)$.

The k^{th} **cumulant of** X is defined as $\chi_X(k) = \left.\frac{d^k}{dt^k}[\ln \phi_X(t)]\right|_{t=0}$.

For lognormal random variables, neither the **characteristic function** $\phi_X(t)$ nor the **moment generating function** $M_X(t)$ exist in closed form. For the former, the next best result is [Leipnik, 1991], who derived the characteristic function of a lognormal random variable in the form of a (demanding but) rapidly converging series of Hermite functions in a logarithmic variable.

The **raw moments** of lognormally distributed X can be obtained using the moment generating function of the Gaussian variable $Z \sim N(\mu, \sigma^2)$, with $X = \exp(Z)$. This moment

generating function of Z equals $M_Z(t) = \exp(t\mu + \frac{1}{2}t^2\sigma^2)$. Now use that $E\{X^k\} = E\{\exp(kZ)\} = M_Z(t)\big|_{t=k}$, then $\alpha_X(k) = \exp(k\mu + \frac{1}{2}k^2\sigma^2)$. The first four **cumulants** $\chi_X(k)$, $k = 1,\ldots,4$ can be written in terms of the first four raw moments $\alpha_X(k)$, $k = 1,\ldots,4$, as follows: $\chi_X(1) = \alpha_X(1)$, $\chi_X(2) = \alpha_X(2) - \chi_X^2(1)$,
$\chi_X(3) = \alpha_X(3) - 3\chi_X(2)\chi_X(1) - \chi_X^3(1)$,
$\chi_X(4) = \alpha_X(4) - 4\chi_X(3)\chi_X(1) - 3\chi_X^2(2) - 6\chi_X(2)\chi_X^2(1) - \chi_X^4(1)$.

B.6 Central Moments

- $E\{X\} = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$
- $E\{X - E(X)\}^2 = Var\{X\} = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$
- $E\{X - E(X)\}^3 = (\exp(\sigma^2) + 2)\sigma^3\sqrt{\exp(\sigma^2) - 1}$
- $E\{X - E(X)\}^4 = \left((\exp(\sigma^2))^4 + 2(\exp(\sigma^2))^3 + 2(\exp(\sigma^2))^2 - 3\right)\sigma^4$

B.7 Other general properties

- The typical value x_{typical}, corresponding to the maximum of the distribution, is $x_{\text{typical}} = \exp(\mu - \sigma^2)$.
- The median x_{median}, such that $\int_0^{x_{\text{median}}} f_X(x)dx = \int_{x_{\text{median}}}^{\infty} f_X(x)dx = \frac{1}{2}$ is $x_{\text{median}} = \exp(\mu)$.
- The coefficient of variation, $C \equiv \sqrt{Var\{X\}} / E\{X\}$, is $C = \sqrt{\exp(\sigma^2) - 1}$.
- Skewness, i.e. $E\{X - E(X)\}^3 / \sigma^3 = (\exp(\sigma^2) + 2)\sqrt{\exp(\sigma^2) - 1}$.
- Kurtosis, i.e. $E\{X - E(X)\}^4 / \sigma^4 = (\exp(\sigma^2))^4 + 2(\exp(\sigma^2))^3 + 2(\exp(\sigma^2))^2 - 3$.

B.8 Confidence intervals

$\Pr(X \in [\exp(\mu - k\sigma), \exp(\mu + k\sigma)]) = erf\left(\frac{k}{\sqrt{2}}\right)$, with $erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt$.

Typical values are:

$erf\left(\frac{1}{\sqrt{2}}\right) = 0.6827$; $erf\left(\frac{2}{\sqrt{2}}\right) = 0.9545$; $erf\left(\frac{3}{\sqrt{2}}\right) = 0.9973$;
$erf\left(\frac{1.96}{\sqrt{2}}\right) = 0.95$; $erf\left(\frac{2.5758}{\sqrt{2}}\right) = 0.99$;

Hence in particular: $\Pr(X \in [\exp(\mu - 1.96\sigma), \exp(\mu + 1.96\sigma)]) = 0.95$.

B.9 Product form

- If $X \sim \Lambda(\mu,\sigma^2)$ and $c>0$ then $cX^b \sim \Lambda(\ln c + b\mu, b^2\sigma^2)$
- If X_i are independent positive random variables, with $X_i \sim \Lambda(\mu_i,\sigma_i^2)$ and $c>0$ then $c\prod_{i=1}^{n} X_i^{b_i} \sim \Lambda(\ln c + \sum_{i=1}^{n} b_i\mu_i, \sum_{i=1}^{n} b_i^2\sigma_i^2)$.

B.10 Multiplicative Central Limit Theorem

Multiplicative Central Limit Theorem

If H_i are identically distributed pairwise independent *positive* random variables (not necessarily Lognormal), then $\prod_{i=1}^{n} H_i \sim \Lambda(nE\{\ln H_i\}, nVar\{\ln H_i\})$ asymptotically.

Multiplicative Extended Central Limit Theorem (Lyapunov)

Suppose, $H_1,\ldots,H_n$ are independent random variables with possibly different distributions (not necessarily Lognormal), and $\gamma_i = E\{H_i\}$, $\theta_i^2 = E\{(H_i-\gamma_i)^2\}$ and $\omega_i^3 = E\{|H_i-\gamma_i|^3\}$, where ω_i must exist.

Moreover, let: $\Gamma = \sum_{i=1}^{n}\gamma_i$, $\Theta^2 = \sum_{i=1}^{n}\theta_i^2$, and $\Omega^3 = \sum_{i=1}^{n}\omega_i^3$. Then, if $\lim_{n\to\infty}\Omega/\Theta \to 0$, then for $n\to\infty$, $\sum_{i=1}^{n} H_i \sim N(\Gamma,\Theta^2)$. Since $\sum_{i=1}^{n} H_i = \ln(\exp(\sum_{i=1}^{n} H_i)) = \ln(\prod_{i=1}^{n}\exp(H_i)) \sim N(\Gamma,\Theta^2)$, this yields that $\prod_{i=1}^{n}\exp(H_i) \sim \Lambda(\Gamma,\Theta^2)$.

Appendix C: Approximations that consider sums of lognormally distributed random variables

Suppose $X_1,\ldots,X_n$ are lognormally distributed random variables: $X_i \sim \Lambda(\mu_i,\sigma_i^2)$. In this appendix, we are looking for properties (e.g. density or distribution function) of $X=\sum_{i=1}^n X_i$, i.e. the sum of n lognormal variables. Since no closed form of the density or distribution function of X exists, not even if $X_1,\ldots,X_n$ are independent, several approximations have been developed in literature. Five of these approximations are presented in this appendix. In addition, the Central Limit Theorem and the Extended Central Limit theorem (these consider sums of generally distributed variables rather than specifically sums of lognormals) are applied to approximate X. In Appendix C.8, the approximations considered are compared, based on comparisons reported in literature.

C.1 Central Limit Theorem applied to lognormal variables

Suppose X_i are identically distributed pairwise independent lognormal variables, i.e. $X_i \sim \Lambda(\mu,\sigma^2)$, then $\sum_{i=1}^n X_i \sim N(nE\{X\}, n\mathrm{Var}\{X\})$, i.e. $\sum_{i=1}^n X_i \sim N(ne^{\mu+\frac{1}{2}\sigma^2}, ne^{2\mu+\sigma^2}(e^{\sigma^2}-1))$ asymptotically.

C.2 Extended Central Limit Theorem (Lyapunov) applied to lognormal variables

Suppose, $X_1,\ldots,X_n$ are independent lognormal variables: $X_i \sim \Lambda(\mu_i,\sigma_i^2)$, and $\gamma_i = E\{X_i\} = \exp(\mu_i + \tfrac{1}{2}\sigma_i^2)$, $\theta_i^2 = E\{(X_i-\gamma_i)^2\} = e^{2\mu_i+\sigma_i^2}(e^{\sigma_i^2}-1)$ and $\omega_i^3 = E\{|X_i-\gamma_i|^3\}$, where ω_i must exist. Moreover, let: $\Gamma = \sum_{i=1}^n \gamma_i$, $\Theta^2 = \sum_{i=1}^n \theta_i^2$, and $\Omega^3 = \sum_{i=1}^n \omega_i^3$.

Then, if $\lim_{n\to\infty} \Omega/\Theta \to 0$, then for $n\to\infty$, $\sum_{i=1}^n X_i \sim N(\Gamma,\Theta^2)$, i.e.

$$\sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n e^{\mu_i+\frac{1}{2}\sigma_i^2}, \sum_{i=1}^n e^{2\mu_i+\sigma_i^2}(e^{\sigma_i^2}-1)\right).$$

C.3 Farley's method

Suppose $X_1,\ldots,X_n$ are independent identically distributed lognormal variables: $X_i \sim \Lambda(\mu,\sigma^2)$, then the distribution function of $X=\sum_{i=1}^n X_i$ is approximated by

$\Pr(X\le x) = F_X(x) \approx \left[1-Q\left(\frac{\ln x-\mu}{\sigma}\right)\right]^n$, where Q is the Gaussian tail function defined by

$Q(z) = \frac{1}{\sqrt{2\pi}}\int_z^\infty \exp(-\tfrac{1}{2}t^2)dt$, or $Q(z) = \tfrac{1}{2} - \tfrac{1}{2} erf\left(\frac{z}{\sqrt{2}}\right)$. Reference: e.g. [Beaulieu et al, 1995].

C.4 Fenton-Wilkinson's method

Suppose $X_i = \exp(Z_i)$ is lognormal, with Z_i Gaussian, and suppose we want to know the pdf of $X = X_1 + \cdots + X_n = \exp(Z_1) + \cdots + \exp(Z_n)$. This sum is approximated by another lognormal variable, i.e. $X = \exp(Z_1) + \cdots + \exp(Z_n) \cong \exp(W)$, where W is Gaussian.

In Fenton-Wilkinson's method (e.g. [Abu-Dayya and Beaulieu, 1994]), the mean μ_W and standard deviation σ_W of W are obtained by matching the first two raw moments of $X = \exp(W)$ with the first two raw moments of $X = \exp(Z_1) + \cdots + \exp(Z_n)$:

$$E\{X\} = E\{\exp(W)\} = E\{\exp(Z_1) + \cdots + \exp(Z_n)\} \Rightarrow$$

$$\Rightarrow \exp(\mu_W + \sigma_W^2 / 2) = \sum_{i=1}^{n} \exp(\mu_{Z_i} + \sigma_{Z_i}^2 / 2) =: u_1 .$$

$$E\{X^2\} = E\{\exp(2W)\} = E\{(\exp(Z_1) + \cdots + \exp(Z_n))^2\} \Rightarrow$$

$$\Rightarrow \exp(2\mu_W + 2\sigma_W^2) = \sum_{i=1}^{n} E\{(\exp(Z_i))^2\} + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E\{\exp(Z_i + Z_j)\} =$$

$$= \sum_{i=1}^{n} \exp(2\mu_{Z_i} + 2\sigma_{Z_i}^2) + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \exp(\mu_{Z_i} + \mu_{Z_j}) \exp(\tfrac{1}{2}(\sigma_{Z_i}^2 + \sigma_{Z_j}^2 + 2r_{ij}\sigma_{Z_i}\sigma_{Z_j})) =: u_2 .$$

where $r_{ij} = \dfrac{E\{(Z_i - \mu_{Z_i})(Z_j - \mu_{Z_j})\}}{\sigma_{Z_i}\sigma_{Z_j}}$ is the correlation coefficient of Z_i and Z_j.

Solving for μ_W and σ_W yields:

$$\mu_W = 2\ln u_1 - \tfrac{1}{2}\ln u_2$$

$$\sigma_W^2 = \ln u_2 - 2\ln u_1$$

Hence, X is lognormally distributed with parameters μ_W and σ_W^2, i.e. $X \sim \Lambda(\mu_W, \sigma_W^2)$.

C.5 Safak's extension of Schwartz and Yeh

The approximation proposed by [Schwartz and Yeh, 1982] uses an iterative procedure to compute the mean and variance of a sum of *independent* lognormal random variables. [Safak, 1993] proposed an extension of this approximation to sums of *correlated* lognormal variables. This extension is presented here.

Suppose $X_i = \exp(Z_i)$ is lognormal, with Z_i Gaussian, and suppose we want to know the mean and variance of $X = X_1 + \cdots + X_n = \exp(Z_1) + \cdots + \exp(Z_n)$. [Safak, 1993] provides exact analytical formulas for calculating the mean and variance of $\ln X$.

Write $W_n = \ln X = \ln(\sum_{i=1}^{n} e^{Z_i}) = \ln(\sum_{i=1}^{n-1} e^{Z_i} + e^{Z_n}) = \ln(e^{W_{n-1}} + e^{Z_n}) = \ln(e^{W_{n-1}} + e^{W_{n-1}+Y_n}) =$

$\ln(e^{W_{n-1}}(1 + e^{Y_n})) = W_{n-1} + \ln(1 + e^{Y_n})$, where $Y_n = Z_n - W_{n-1}$.

Then, the mean and variance of W_n are given by

$\mu_{W_n} = \mu_{W_{n-1}} + G_1(\sigma_{Y_n}, \mu_{Y_n})$ and

$$\sigma^2_{W_n} = \sigma^2_{W_{n-1}} - G_1^2(\sigma_{Y_n}, \mu_{Y_n}) + \frac{2(r_{W_{n-1}Z_n}\sigma_{Z_n} - \sigma_{W_{n-1}})\sigma_{W_{n-1}}}{\sigma^2_{Y_n}} G_3(\sigma_{Y_n}, \mu_{Y_n}) + G_2(\sigma_{Y_n}, \mu_{Y_n}),$$

where $\mu_{Y_n} = \mu_{Z_n} - \mu_{W_{n-1}}$ and $\sigma^2_{Y_n} = \sigma^2_{W_{n-1}} + \sigma^2_{Z_n} - 2r_{W_{n-1}Z_n}\sigma_{W_{n-1}}\sigma_{Z_n}$ are the mean and variance of $Y_n = Z_n - W_{n-1}$, where $r_{W_{n-1}Z_n}$ is the correlation coefficient of W_{n-1} and Z_n:

$$r_{W_{n-1}Z_n} = \frac{E\{(W_{n-1} - \mu_{W_{n-1}})(Z_n - \mu_{Z_n})\}}{\sigma_{W_{n-1}}\sigma_{Z_n}} = \sigma_{W_{n-2}} \frac{r_{W_{n-2}Z_n}}{\sigma_{W_{n-1}}} + \frac{r_{Z_{n-1}Z_n}\sigma_{Z_{n-1}} - r_{W_{n-2}Z_n}\sigma_{W_{n-2}}}{\sigma_{W_{n-1}}\sigma^2_{Y_{n-1}}} G_3(\sigma_{Y_{n-1}}, \mu_{Y_{n-1}}),$$

with $r_{W_1Z_n} = r_{Z_1Z_n}$, and where

$$G_1(\sigma,\mu) = \mu\Phi\left(\frac{\mu}{\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} + \sum_{k=1}^{\infty} C_k[F(\sigma,\mu,k) + F(\sigma,-\mu,k)],$$

$$G_2(\sigma,\mu) = (\mu^2+\sigma^2)\Phi\left(\frac{\mu}{\sigma}\right) + (\mu + \ln 4)\frac{\sigma}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} + 2\sum_{k=1}^{\infty} C_k(\mu - k\sigma^2)F(\sigma,\mu,k) +$$

$$+\sum_{k=2}^{\infty} B_{k-1}[F(\sigma,\mu,k) + F(\sigma,-\mu,k)],$$

$$G_3(\sigma,\mu) = \sigma^2\sum_{k=0}^{\infty}(-1)^k[F(\sigma,\mu,k) + F(\sigma,-\mu,k+1)],$$

$$F(\sigma,\mu,k) = e^{-k\mu + \frac{k^2\sigma^2}{2}}\Phi\left(\frac{\mu - k\sigma^2}{\sigma}\right),$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp(-\tfrac{1}{2}t^2)dt, \quad C_k = \frac{(-1)^{k+1}}{k}, \quad B_k = \frac{2(-1)^{k+1}}{k+1}\sum_{j=1}^{k}\frac{1}{j}.$$

C.6 Ho's extension of Schwartz and Yeh

[Ho, 1995] proposed a modification of Safak's extension of Schwartz and Yeh that significantly reduces the computational work and round-off error.

Suppose $X_i = \exp(Z_i)$ is lognormal, with Z_i Gaussian, and suppose we want to know the pdf of $X = X_1 + \cdots + X_n = \exp(Z_1) + \cdots + \exp(Z_n)$. Ho's extension of Schwartz and Yeh's approximation uses an iterative procedure to compute the mean and variance of X for any finite number of components. The basic equations are equal to those of [Safak, 1993]:

Write $W_n = \ln X = \ln(\sum_{i=1}^{n} e^{Z_i}) = \ln(\sum_{i=1}^{n-1} e^{Z_i} + e^{Z_n}) = \ln(e^{W_{n-1}} + e^{Z_n}) = \ln(e^{W_{n-1}} + e^{W_{n-1}+Y_n}) =$

$\ln(e^{W_{n-1}}(1+e^{Y_n})) = W_{n-1} + \ln(1+e^{Y_n})$, where $Y_n = Z_n - W_{n-1}$.

Then, the mean and variance of W_n are given by

$\mu_{W_n} = \mu_{W_{n-1}} + G_1(\sigma_{Y_n}, \mu_{Y_n})$ and

$$\sigma^2_{W_n} = \sigma^2_{W_{n-1}} - G_1^2(\sigma_{Y_n}, \mu_{Y_n}) + \frac{2(r_{W_{n-1}Z_n}\sigma_{Z_n} - \sigma_{W_{n-1}})\sigma_{W_{n-1}}}{\sigma^2_{Y_n}} G_3(\sigma_{Y_n}, \mu_{Y_n}) + G_2(\sigma_{Y_n}, \mu_{Y_n}),$$

where $\mu_{Y_n} = \mu_{Z_n} - \mu_{W_{n-1}}$ and $\sigma^2_{Y_n} = \sigma^2_{W_{n-1}} + \sigma^2_{Z_n} - 2r_{W_{n-1}Z_n}\sigma_{W_{n-1}}\sigma_{Z_n}$ are the mean and variance of $Y_n = Z_n - W_{n-1}$, where $r_{W_{n-1}Z_n}$ is the correlation coefficient of W_{n-1} and Z_n:

$$r_{W_{n-1}Z_n} = \frac{E\{(W_{n-1}-\mu_{W_{n-1}})(Z_n-\mu_{Z_n})\}}{\sigma_{W_{n-1}}\sigma_{Z_n}} = \sigma_{W_{n-2}}\frac{r_{W_{n-2}Z_n}}{\sigma_{W_{n-1}}} + \frac{r_{Z_{n-1}Z_n}\sigma_{Z_{n-1}} - r_{W_{n-2}Z_n}\sigma_{W_{n-2}}}{\sigma_{W_{n-1}}\sigma^2_{Y_{n-1}}} G_3(\sigma_{Y_{n-1}}, \mu_{Y_{n-1}}).$$

with $r_{W_1Z_n} = r_{Z_1Z_n}$.

The difference between [Safak, 1993] and [Ho, 1995] lies in the computation of the remaining functions:

$$G_1(\sigma_{Y_n}, \mu_{Y_n}) = A_0 + I_1,$$

$$G_2(\sigma_{Y_n}, \mu_{Y_n}) = I_3 + 2I_4 + \sigma^2_{Y_n} I_0 + \mu_{Y_n} A_0,$$

$$G_3(\sigma_{Y_n}, \mu_{Y_n}) = \sigma^2_{Y_n}(I_2 + I_0),$$

$$I_4 = \sigma^2_{Y_n}[f_{Y_n}(0)\ln 2 - I_5] + \mu_{Y_n} I_6,$$

$$A_0 = \frac{\sigma_{Y_n}}{\sqrt{2\pi}}\exp(-(\mu_{Y_n}/\sigma_{Y_n})^2/2) + \mu_{Y_n} I_0,$$

$$I_i = \int_0^1 h_i(v)v^{-1}dv,$$

$$h_i(v) = \begin{cases} \frac{1}{\sqrt{2\pi}}\exp(-(\ln v + \mu_{Y_n}/\sigma_{Y_n})^2/2) & i=0 \\ (f_{Y_n}(\ln v) + f_{Y_n}(-\ln v))\ln(1+v) & i=1 \\ (f_{Y_n}(\ln v) - f_{Y_n}(-\ln v))(1+v^{-1})^{-1} & i=2 \\ (f_{Y_n}(\ln v) + f_{Y_n}(-\ln v))\ln^2(1+v) & i=3 \\ -f_{Y_2}(-\ln v)\ln v \ln(1+v) & i=4 \\ f_{Y_n}(-\ln v)(1+v^{-1})^{-1} & i=5 \\ f_{Y_n}(-\ln v)\ln(1+v) & i=6 \end{cases}$$

$$f_{Y_n}(y) = \frac{1}{\sqrt{2\pi\sigma^2_{Y_n}}}\exp\left(-\frac{(y-\mu_{Y_n})^2}{2\sigma^2_{Y_n}}\right).$$

C.7 Cumulant Matching Approximation

The cumulant matching approximation approach divides the range of the complementary distribution function of X into regions and applies a lognormal based approximation in each region. Result is a piecewise lognormal distribution function. The approach presented below is by [Aby-Dayya and Beaulieu, 1994], for correlated X_i.

[Aby-Dayya and Beaulieu, 1994] assume that the random variable $X = X_1 + \cdots + X_n = \exp(Z_1) + \cdots + \exp(Z_n)$ can be approximated in each region by a reference lognormal variable $R \sim \Lambda(\mu_R, \sigma^2_R)$. The parameters μ_R and σ^2_R of the first reference distribution (the region of largest values of the complementary distribution function) are found by matching the first and second cumulants of R with the corresponding cumulants of X (see Appendix B for a definition of cumulant). The parameters of the second reference distribution are found by matching the second and third cumulants in a similar manner. If a third region is required, the third and fourth cumulants are matched as previously.

For the first region this yields:

$\sigma_R^2 = \sigma_{12}^2 = \ln(1 + \chi_X(2)/\chi_X^2(1))$ and $\mu_R = \mu_{12} = \ln\left(\frac{\chi_X(1)}{\sqrt{1 + \chi_X(2)/\chi_X^2(1)}}\right)$

For the second region this yields:

$\sigma_R^2 = \sigma_{23}^2 = \ln(A_1 + A_2 - 1)$ and

$\mu_R = \mu_{23} = \ln\left(\sqrt{\chi_X(2)/(e^{2\sigma_R^2} - e^{\sigma_R^2})}\right) = \frac{1}{2}\left(\ln \chi_X(2) - \sigma_R^2 - \ln(e^{\sigma_R^2} - 1)\right)$, where

$A_1 = \left\{\frac{g}{2}[1 + \sqrt{1 - 4/g}] - 1\right\}^{1/3}$, $A_2 = \left\{\frac{g}{2}[1 - \sqrt{1 - 4/g}] - 1\right\}^{1/3}$, $g = 4 + \chi_X^2(3)/\chi_X^3(2)$

For the third region this yields:

$\sigma_R^2 = \sigma_{34}^2 = \ln(s^2)$ and $\mu_R = \mu_{34} = \frac{\chi_X^{1/3}(3)}{s(s^6 - 3s^2 + 2)^{1/3}}$, where s is a positive solution to the following equation: $\frac{s^2 - 4s^6 - 3s^4 + 12s^2 - 13}{(s^6 - 3s^2 + 2)^{4/3}} = \frac{\chi_X(4)}{\chi_X^{4/3}(3)}$.

The cumulants $\chi_X(k)$ used in these expressions are written as a function of the raw moments $\alpha_X(k)$ as in Appendix B:

$\chi_X(1) = \alpha_X(1)$,

$\chi_X(2) = \alpha_X(2) - \chi_X^2(1)$,

$\chi_X(3) = \alpha_X(3) - 3\chi_X(2)\chi_X(1) - \chi_X^3(1)$,

$\chi_X(4) = \alpha_X(4) - 4\chi_X(3)\chi_X(1) - 3\chi_X^2(2) - 6\chi_X(2)\chi_X^2(1) - \chi_X^4(1)$.

The raw moments $\alpha_X(k)$ are determined as follows:

$$\alpha_X(k) = E\{X^k\} = \sum_{m \in M} \frac{k!}{r_1! r_2! \cdots r_n!} \exp(\mu_F + \frac{1}{2}\sigma_F^2),$$

where M is the set of all non-negative integer combinations of $\{r_1, r_2, \ldots, r_n\}$ such that $r_1 + r_2 + \cdots r_n = k$, $\mu_F = \sum_{j=1}^{n} r_j \mu_{Z_j}$ and $\sigma_F^2 = \sum_{j=1}^{n} r_j^2 \sigma_{Z_j}^2 + 2\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} r_j r_i \rho_{Z_j Z_i} \sigma_{Z_j} \sigma_{Z_i}$, where

$\rho_{Z_j Z_i} = \frac{E\{(Z_j - \mu_{Z_j})(Z_i - \mu_{Z_i})\}}{\sigma_{Z_j}\sigma_{Z_i}}$ is the correlation coefficient of Z_j and Z_i.

The value of the distribution argument *u*, where it is appropriate to shift from one reference lognormal distribution with parameters $\mu_{R(k-1)}$ and $\sigma_{R(k-1)}^2$ to the next reference lognormal distribution with parameters $\mu_{R(k)}$ and $\sigma_{R(k)}^2$ is given by $u = \left(\frac{\sigma_{R(k)} \ln \mu_{R(k-1)}}{\sigma_{R(k-1)} \ln \mu_{R(k)}}\right)^{1/(\sigma_{R(k)} - \sigma_{R(k-1)})}$.

In [Ho, 1995], this switching point is explained differently: Choose $\mu_{R(k)}$ and $\sigma_{R(k)}^2$ as parameters for the k^{th} region $(x_{k-1}, x_k]$ of the distribution of *X*. The switching point x_k is defined as $x_k = \dfrac{\sigma_X(k+1)\mu_X(k) - \sigma_X(k)\mu_X(k+1)}{\sigma_X(k+1) - \sigma_X(k)}$, which is obtained by solving $\dfrac{x_k - \mu_X(k)}{\sigma_X(k)} = \dfrac{x_k - \mu_X(k+1)}{\sigma_X(k+1)}$.

C.8 Comparisons of the different approximations

This subsection compares the different approximations presented in this appendix, based on evaluations reported in literature.

Note that in this comparison we mainly restrict to issues that are relevant to the application of our interest: we are looking for properties of $X = X_1, \dots, X_n$, with X_i lognormal (to be more specific: $n = n_k$ and $X_\kappa \triangleq \tilde{\rho}(1, V, \kappa) \times \tilde{\chi}(\kappa, 1, V) \sim \Lambda(\ln[\tilde{\rho}(1, \overline{v}, \kappa)\tilde{\chi}(\kappa, 1, \overline{v})\widehat{B}_\kappa \widehat{M}_\kappa], \frac{1}{4}\widehat{G}_\kappa)$, see Section 5), and in particular we are interested in two specific properties:

- The expectation of X .
- Values α_1 and (in particular) α_2 such that $\Pr(X > \alpha_2) = (1 - 0.95)/2 = 2.5 \times 10^{-2}$ and $\Pr(X < \alpha_1) = (1 - 0.95)/2 = 2.5 \times 10^{-2}$ (which makes $[\alpha_1, \alpha_2]$ a 95% credibility interval for X).

The former is not really the problem, since the expectation of a sum is equal to the sum of expectations of lognormals, and these individual expectations are easily obtained from the distributions of X_κ. The latter can be described by studying properties of the complementary distribution function (CDF, i.e. $\Pr(X > x)$) and the distribution function (DF, i.e. $\Pr(X \leq x)$).

The findings of the comparisons are summarised below:

- Many have tried to apply the (Extended) Central Limit Theorem to approximate a sum of lognormals by a Gaussian, and have concluded that this does not work very well for *n* small (moments grow exponentially with their order). Since in our applications, *n* is expected to take on values in the range 3 to 9, this is a reason to not use this approximation. Moreover, Monte Carlo simulations reveal that the sum of lognormals approaches another lognormal (rather than a Gaussian) rather well, which speaks in favour of the other approximations.
- The table below compares the approximations, based on the conditions on $X_1, \dots, X_n$:

Central Limit Theorem	X_i are identically distributed pairwise independent
Extended Central Limit Theorem	X_i are independent
Farley	X_i are identically distributed independent
Fenton-Wilkinson	X_i may be correlated
Safak's extension of Schwartz and Yeh	X_i may be correlated
Ho's extension of Schwartz and Yeh	X_i may be correlated
Cumulant Matching approach	X_i may be correlated

Hence, Farley and the Central Limit Theorem are the most restrictive: the terms in the sum need to be independent and identically distributed. The Extended Central Limit Theorem allows different distribution parameters, while the remaining approximations are least restrictive: X_i may be correlated. We note that in our application of interest, X_i are not expected to be identically distributed, and they may be correlated. Therefore, the first three approximations are probably not useful here.

- Two aspects why Farley's method may still be useful: Farley is said to be valid for large variances [Schwartz and Yeh, 1982]. For any variance, Farley gives a strict upperbound on the cumulative distribution function: The "≈" in Appendix C.3 can be replaced by "<" [Beaulieu et al, 1995].
- According to [Ho, 1995], the problem of Cumulants Matching method is that the ratio $\chi_X^2(3)/\chi_X^3(2)$ may be so large that the values of A_1 and A_2 are not available.
- [Abu-Dayya and Beaulieu, 1994] compare Fenton-Wilkinsons method, Safak's extension of Schwartz and Yeh, and a Cumulants Matching Approach. They conclude that Fenton-Wilkinsons may be the best to compute the CDF of sums of correlated lognormal variables, due to its accuracy and computational simplicity.
- [Ho, 1995] concludes that, compared with Fenton-Wilkinson and Cumulants Matching, Schwartz and Yeh yield better estimates of the mean and variance of $\ln X$. Fenton-Wilkinson better approximates the tail of the CDF.
- [Beaulieu et al, 1995] consider the CDF and the DF of a sum of *independent* lognormals, using Fenton-Wilkinson (F-W), Schwartz and Yeh (S-Y), Cumulants Matching (C-M) and Farley. Numerical results of application of these methods, for different values of *n* and the decibel spread (see Appendix B) are compared with Monte Carlo simulation results. The conclusions drawn on which method provides the best results depend on the value of the CDF and the DF. Since in our case we are interested in good approximations for values α_1 and (in particular) α_2 in CDF = $\Pr(X > \alpha_2)$ = 2.5×10^{-2} and DF = $\Pr(X \leq \alpha_1)$ = 2.5×10^{-2}, only those results are repeated here. See the tables below.

α_2 is best appr. by	decibel spread = 6 dB	decibel spread = 12 dB
n = 2	F-W and C-M are best; Farley holds second place; S-Y is not good	Farley is best; F-W and C-M hold second place; S-Y is not good
n = 6	F-W and C-M are best; S-Y is not good; Farley was not tested	Farley is best; F-W and C-M hold second place; S-Y is not good
n = 20	F-W and C-M are best; S-Y holds second place; Farley is not good	Not tested
n = 30	F-W and C-M are best; S-Y holds second place; Farley is not good	Not tested

α_1 is best appr. by	decibel spread = 6 dB	decibel spread = 12 dB
n = 2	S-Y is best; F-W and Farley hold second place; C-M not tested	S-Y and Farley are best; F-W is not good; C-M not tested
n = 6	S-Y are best; F-W holds second place; Farley is not good; C-M not tested	S-Y is best; Farley holds second place; F-W is not good; C-M not tested

- [Pirinen, 2000] evaluated the mean and variance of sums of multiple independent or correlated lognormal signals using Fenton-Wilkinson (F-W), Ho's version of Schwartz and Yeh (S-Y), and Monte Carlo simulation. It appeared that the mean and variance estimated using S-Y compared to the simulation results very well, but F-W gives lower estimates on the mean value and higher variance results than S-Y.
- According to [Pirinen, 2003], Schwartz and Yeh can be best applied when the range of the standard deviation of the individual components is between 4 and 12 dB. If all components in the summation are identically distributed, this approximation tends to underestimate the variance in the resulting sum. The error increases as a function of the number of added components.
- Several references, e.g. [Pirinen, 2003], [Schwartz and Yeh, 1982], report that Fenton-Wilkinson's method applied to the uncorrelated case is applicable if the standard deviations of the lognormal components are lower than 4 dB. In the correlated case, Fenton-Wilkinson is quite accurate at higher deviation values (up to 12 dB, according to [Abu-Dayya and Beaulieu, 1994]). Here, note that the dB (decibel) dimension is used if the definition of lognormal is given using a 10-base logarithm (see Appendix B). If we translate this to our application, we get σ_{Y_κ} < 4 dB or 12 dB, where $Y_\kappa = \ln X_\kappa \times \frac{10}{\ln 10}$ and $\frac{10}{\ln 10} = 4.34$. Since $X_\kappa \sim \Lambda(\ln[\tilde{\rho}(1,\bar{v},\kappa)\tilde{\chi}(\kappa,1,\bar{v})\widehat{B}_\kappa \widehat{M}_\kappa], \frac{1}{4}\widehat{G}_\kappa)$, we find that σ_{Y_κ} < 4 dB yields $4.34 \times \sqrt{\widehat{G}_\kappa / 4} < 4$, or $\widehat{G}_\kappa < 3.39$, and σ_{Y_κ} < 12 dB yields $4.34 \times \sqrt{\widehat{G}_\kappa / 4} < 12$, or $\widehat{G}_\kappa < 30.5$. Leafing through the bias and uncertainty results for different projects, in which we only applied the bias and uncertainty assessment method of Section 3, i.e., without the conditioning on the event sequence classification process, we find values for $\widehat{U}$ ranging from 1.7 to 9.0. (Translation: the upperbound divided by the lowerbound of a 95% credibility interval for collision risk is between 13 and 403). If we assume that $\widehat{G}_\kappa = 2\widehat{U}$ for all κ then $\widehat{G}_\kappa$ could have values between 4 and 20. Hence, Fenton-Wilkinson would not be applicable in the uncorrelated case, but it would generally be applicable in the correlated case.

Conclusion:

Taking into account all the comparisons of the approximations discussed above, we conclude that Fenton-Wilkinson's approximation is the most logical choice for our application. It is relatively simple to use, and still performs rather well.