A Framework for Scenario Generation

Diane Reynolds

As risk measures and their underlying models grow more complex, risk managers rely increasingly on scenario-based methods. The complexity of the scenario-generation process typically leads to new challenges: ensuring that scenarios are generated consistently, and communicating the results to different audiences. To structure the process, five key questions are proposed that help define a scenario set. Next, a framework is introduced to describe, discuss and implement scenario generation simply and consistently. The framework separates the responses to the questions into smaller components and highlights their inter-relationships. The framework consists of four main components and several secondary components, allowing the process to be communicated effectively to disparate audiences.

Enterprise-wide risk management demands methodologies that can integrate the various risks, spanning both business units and geographic locations, faced by organizations today. Representing future uncertainty in terms of a set of scenarios is a particularly effective strategy for achieving this goal. The quality of the resulting analysis, however, depends on the ability to generate relevant scenarios, a task that grows increasingly complex with the proliferation of risk factors, models and sampling techniques. In this paper, the focus is on providing structure for the scenario-generation process, thereby allowing risk managers to deal more effectively with its complexity.

Mark-to-Future (MtF) is a scenario-based approach that measures and manages a variety of risks (Dembo et al. 2000). It provides flexibility in defining the scenarios of interest, the set of financial instruments, the portfolio hierarchy and the risk measures. All of these decisions are interdependent, and must be coordinated to ensure a sensible result. Consider, for example, using MtF to estimate the Value-at-Risk (VaR), based on historical data, for a large, diverse portfolio. Given the instruments in the portfolio and their respective pricing models, it is first necessary to identify a set of underlying risk factors for the portfolio. These risk factors might include interest rates, foreign exchange rates, or commodity prices. Once historical time series for all risk factors have been obtained, the data can be manipulated to produce a consistent set of scenarios. These scenarios act as input to the pricing models, which calculate scenario-dependent prices for all instruments. By combining the resulting prices with the portfolio position information, one obtains a profit-and-loss distribution for the portfolio, from which the VaR can be estimated.

As implied by the example above, scenarios are the basis of risk measurement in MtF. The more precisely the statistical scenarios span the set of possible future events, the more accurate are the risk measures calculated from the scenarios. Certainly, more accurate risk measures lead to more effective risk management.

Since many risk measures, such as VaR, are of a statistical nature, generating statistical scenarios is an important part of MtF, or any simulation-based methodology. Statistical scenarios are created by assuming that risk factors behave according to specific models, and then using these models to generate possible future outcomes. The models may range from simple historical approaches, which assume that previous risk factor changes recur in the future, to complex jump diffusion processes. The common feature is that a large number of scenarios are created and assumed to represent the set of all possible future events. While other types of scenarios, such as worst-case and sensitivity scenarios, also provide valuable insights into the true risks of the portfolio, this paper focuses only on the generation of statistical scenario sets.

Risk management has progressed from measuring market, credit, liquidity and other risks in isolation to measuring them jointly, thereby incorporating correlation and diversification effects. Proper joint measures require scenarios covering the set of all risk factors and full descriptions of the relationships among risk factors. This is the only way to produce a consistent view of the future, which leads to the consistent measurement of different types of risk. While the concept is simple (take all of the risk factors, estimate their inter-relationships and generate scenarios), practical problems abound.

Consider the joint measurement of market and credit risk, for example. The set of market risk factors can number in the thousands, while the number of counterparties often reaches the tens of thousands. As a result, the combined set of risk factors can quickly become unmanageable. Furthermore, the essential properties of the risk factors, such as historical trends, reporting frequencies and future expectations, may also differ substantially. Thus, accurately representing the evolution of the risk factors may involve a wide range of statistical methods.

A large number of risk factors with different properties complicates the task of generating statistical scenarios. The dynamic nature of the scenario-generation process presents a further challenge, namely, the system that produces scenarios must be flexible and extensible. As risk management expands in scope, new risk factors are continually introduced. Adding these risk factors to existing scenarios can be difficult, and often requires changes throughout the generation process.

New models for generating scenarios appear frequently. Some are extensive, dealing jointly with a variety of risk factors, while others focus on marginal distributions of a single risk factor. Ideally, when a new marginal model can be applied to a particular type of risk factor, it should be possible to simply substitute it for the existing model without affecting other risk factors included in the scenarios. Similarly, if a new joint model is proposed, it is more convenient to reuse as much of the current model and its implementation as possible than to undertake major changes to the existing process.

It is important to communicate the nature of a scenario set, because different audiences—from senior management to auditors, from the trader to the person implementing the scenario generation—must find a scenario set accessible. If not, the resulting risk measures may seem insufficient or irrelevant. Communication is simplified by allowing for different levels of abstraction when discussing a scenario set. Senior management may prefer a very high-level, non-technical description, while those who implement and maintain the scenario set need a thorough understanding of all technical details.

For example, the phrase: “a multi-step Monte Carlo scenario set in which the interest rates mean revert and the equities grow, over time” may sufficiently describe a scenario set for managerial purposes. In contrast, actually generating this scenario set requires a more detailed specification: "a multi-step quasi Monte Carlo method using an equally weighted variance-covariance (VCV) matrix for Canadian, American and Australian interest rates where each curve is represented by three components that mean revert, and American equities, adjusted for stock splits that grow over time."

The second description indicates, to some extent, the complexity of the models and risk factor relationships that typically underlie statistical scenarios. Explaining or understanding statistical scenario generation at a detailed level is often difficult for two main reasons. First, the models for individual risk factors and their joint behaviour are typically combined into one single, large model, making it hard to isolate their respective properties. Second, the calibration of this model is usually done in one long and involved process.

To address the issues identified above, this paper proposes to define statistical scenario generation in terms of a generic framework that provides levels of abstraction, segregates risk factors and models, and generally, structures the overall process. The framework breaks the process into a series of components, each comprising a small, manageable set of related decisions, which then can be explained and understood more easily. By combining components, complex scenario sets can be constructed in a piece wise fashion, rather than by trying to tackle the problem as a whole. This decomposition has other benefits as well. For example, separating the individual behaviours of the risk factors from their joint characteristics increases the flexibility in assigning models to the risk factors. The framework also divides naturally into several levels of abstraction, which facilitates the communication of scenario-related information.

The scenario-generation process begins with a specification of the types of scenarios that are required. For example, scenarios might be derived directly from historical time series or obtained using Monte Carlo methods. Thus, understanding the trade-offs between the various approaches is a critical element of the overall process. This paper identifies five key questions that not only help to make an informed decision in this regard, but that also direct the creation of the necessary framework components for generating the scenario set.

This paper is organized as follows. First, the decision-making process for creating scenario sets is reviewed and the five key questions that guide the process are derived. The framework components are then described, and related to the answers to these questions. The benefits of the framework are identified and illustrated by two examples, which construct scenario sets for VaR estimation and the calculation of credit exposure. Finally, the paper concludes with a brief summary of its findings.

Five key questions

Often, the first challenge in generating scenarios is to determine what type of scenario set to generate. Depending on the desired statistic or result, several options may be available. For example, VaR can be estimated from historical scenarios, single-step Monte Carlo scenarios or even multi-step Monte Carlo scenarios. In each case, there are benefits and drawbacks.

While historical scenarios provide impartial representations of historical risk factor distributions, the number of scenarios that can be produced may be limited by the amount of historical data that is available. Furthermore, historical scenarios include only events that have actually happened, and so they may not be representative of all events that could possibly happen in the future.

Monte Carlo scenarios overcome the obstacles faced by historical scenarios, but their use introduces new issues. Since Monte Carlo scenarios are samples from a statistical model, many samples (i.e., scenarios) may be required to adequately represent the model (assuming that the model itself accurately depicts reality). This increases the computation time required to evaluate the portfolio under Monte Carlo scenarios. The problem is further aggravated by multi-step Monte Carlo scenarios, which introduce a dimension of time into the scenarios (i.e., the portfolio must be evaluated at not just one, but a number of time points).

The number and type of scenarios generated can have a significant impact on the resulting VaR estimate. As such, a decision to generate historical or Monte Carlo scenarios brings several additional issues into consideration. If historical scenarios are selected, one must determine how to translate history into relevant predictions for the future. The return calculation, if any, is meant to rescale historical values based on today's information. In addition, since the accuracy of the statistical results depends on the number of scenarios, making efficient use of the available historical data is usually a significant concern. If Monte Carlo scenarios are selected, other issues surface. For example, how to ensure that two risk factors display the proper correlation when one has a normal distribution and the other has an empirical distribution? The combination of models to better reflect history, the calibration of the models, and the attempt to represent an entire (continuous) distribution with a limited number of scenarios are all important challenges. Clearly, the decision about which type of scenarios to generate in order to solve a problem is complex and multi-faceted, though an important one.

In order to simplify the decision-making process, it should not be viewed as a response to the single question: What scenarios are required? Rather, scenario sets can be more easily constructed by answering a series of simpler questions. The responses to the following five questions provide an outline of the scenario set that is to be generated:

1. What is the purpose of the scenario set?

The eventual use of the scenario set is critical in deciding how to create it. Historical or single-step Monte Carlo scenarios might be used to estimate VaR, while multi-step Monte Carlo scenarios are more appropriate for calculating credit exposures.

2. What risk factors must the scenario set include?

The list of risk factors that affect a portfolio's value must be identified and analysed. Proper understanding of the sources of risk and their quantification is essential to proper scenario modelling.

3. Do the risk factors need to be grouped or altered? If so, how should it be done?

In addition to listing risk factors, one must also decide whether the risk factors are acceptable in their current form, or whether they can be combined into smaller sets. For example, one technique for reducing the number of risk factors is principal components analysis.

4. What marginal distribution or process is most appropriate for each risk factor?

Once risk factors have been identified and the list analysed for possible reductions or omissions, one must determine the statistical properties of the risk factors. The statistical representations chosen for each risk factor must be consistent with an overall approach to statistical scenario generation. This point is discussed further in the presentation of the framework.

5. What are the technical considerations, such as run-time or memory?

This is an important practical question. The scenario set that fulfils the stated purpose must also be computationally tractable. Simulating the portfolio over the desired number of scenarios or trigger times may not fit into the processing time window. Different modelling decisions may reduce the number of scenarios required to achieve a specified accuracy, and, hence, allow the simulation to fit into the time window.

Answering the five questions above is only the first step in the scenario-generation process. The potential for large numbers of risk factors and models, along with the need for flexibility and extensibility, makes implementing a system for generating scenarios a challenging task. Formalizing and illustrating a framework that helps to structure and simplify this process is the subject of the remainder of this paper.

The framework

The framework is based on a series of components. Each component is defined, and then the components are linked together to create a specific set of scenarios. The strength of the framework is its simplicity.

First and foremost, the name of the scenario set should indicate its usage. The scenario-generation process itself is defined in several layers of abstraction. At the topmost level, there are only four main components. At the second level, each of the main components has, at most, three sub-components. Finally, the most detailed description includes a complete definition of each sub-component. This layered structure makes it easy to drill down into the details of the scenario-generation process, while still providing context for the overall process.

At the highest level, the framework consists of four main components: Blocks, Models, the Scenario Generator and the Scenario Set Definition (see Figure 1). Blocks are created from the set of risk factors affecting the portfolio. A **Block** is basically a group of risk factors with similar statistical properties (e.g., all foreign exchange rates that mean revert). A **Model** defines the distribution or evolutionary process for a risk factor, and also specifies a calibration method for obtaining all model parameters from historical (or other) data. Note that the Model does not specify how risk factors are related to each other. The **Scenario Generator** is a fully calibrated model for generating scenarios that link Models and Blocks, and defines relationships among risk factors. Finally, the **Scenario Set Definition** specifies the details of creating the actual scenario set, such as the number of scenarios, the trigger times (the future points of interest) and a description of the scenario set.

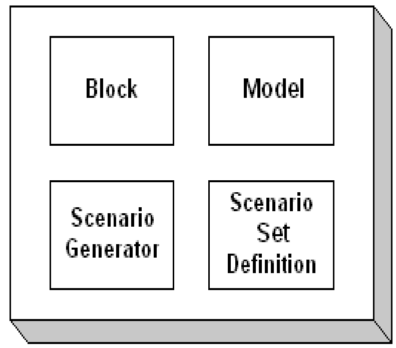

Figure 1: Top-level framework components

This top level of abstraction can be very useful in describing a scenario set in non-technical terms. For example, the description of a sample scenario set provided earlier—"a multi-step Monte Carlo scenario set in which the interest rates mean revert and the equities grow, over time"—includes only the main components, as outlined in Table 1.

Main component	Related description
Scenario Set Definition	"Multi step"
Scenario Generator	"Monte Carlo"
Block one	"Interest rates"
Block two	"Equities"
Model one	"Mean reversion"
Model two	"Growth"

Table 1: Example of top-level framework components

Figure 2 illustrates the second level of abstraction provided by the framework, which decomposes each main component into a set of up to three sub-components.

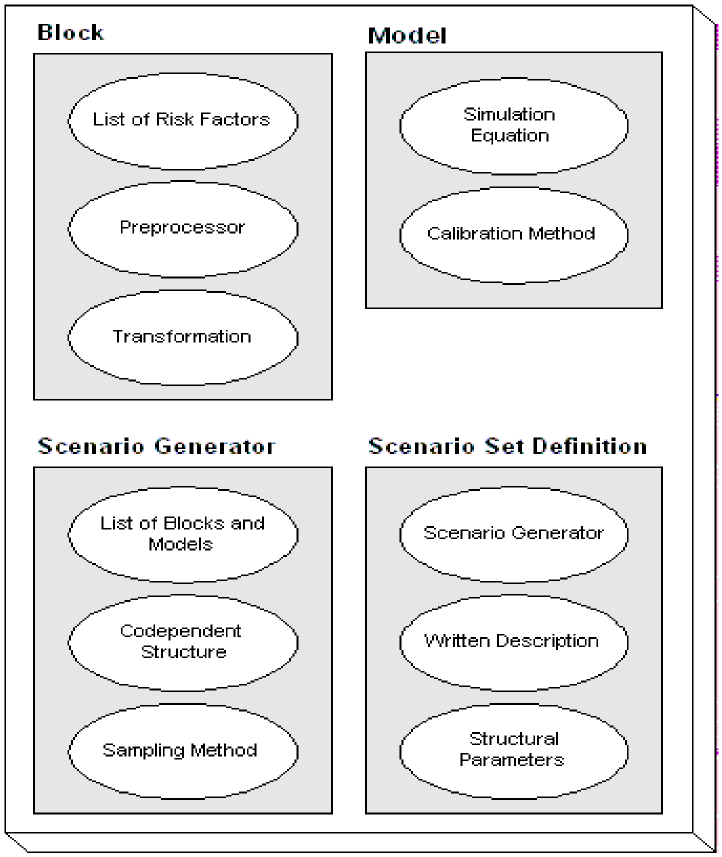

Figure 2: Second-level framework components

This second layer of detail is required to describe the scenario-generation process in technical terms. As such, it includes sub-components for processing and transforming risk factor data, estimating model parameters, specifying the relationships (i.e., codependence) among risk factors and generating the actual scenarios. The more detailed description of the sample scenario set—"a multi-step Quasi Monte Carlo method using an equally weighted variance-covariance matrix for Canadian, American and Australian interest rates where each curve is represented by three components that mean revert, and American equities, adjusted for stock splits and distributions, that grow over time"—includes both the main components, as shown in Table 1 and the sub-components, as shown in Table 2.

Sub-component	Related description
Codependent structure	"Variance-covariance"
Calibration method	"Equally weighted"
Sampling method	"Quasi Monte Carlo"
Transformations in the interest rate Blocks	"Three components"
Preprocessor in the equity Block	"Adjusted for stock splits and distributions"

Table 2: Example of second-level framework components

As is evident from these examples, the framework provides a natural and standard structure for describing the scenario-generation process to different audiences.

One noteworthy property of the framework is that there is no order to the definition of components: one can begin with the Scenario Generator, and work top-down to define the scenario-generation process, or begin with the risk factors and work bottom-up. The framework is meant to allow the definition of a scenario-generation process, rather than to provide an explicit procedure for creating the process; it is a group of components, not a series of steps.

Before describing each of the framework components in greater detail, the issue of data must be considered. Note that the framework itself does not address this issue directly. Although the availability of time series data facilitates the generation of scenarios, it is not an essential part of the framework. The required risk factor information, such as VCV matrices or reversion rates, for example, can be derived from time series data or simply obtained from external sources.

If desired, time series data may be tracked for each risk factor in a database, or some suitable repository. In addition to raw data, the database may contain information for categorizing risk factors. This information serves as the basis of the scenario-generation framework. Data may be requested to identify a risk factor, estimate parameters of a Monte Carlo model, create historical scenarios, or to address other needs in generating scenarios. The decision concerning which risk factors to track over time depends on the answer to the second of the five key questions. For ease of exposition, in the following description of the framework components, it is assumed that there is a database containing time series data for all relevant risk factors.

Block

A Block defines a group of related risk factors that share similar statistical properties, and are often of the same type (e.g., interest rates). These may be the original risk factors, in which case the Block is merely a grouping and data verification mechanism, or they may be complex functions of the original risk factors or underlying data, in which case the Block is responsible for creating the new risk factors and estimating their histories. In addition to grouping risk factors, a Block may also specify a set of operations to be performed on the risk factors by applying a transformation or a preprocessor. The answers to the second and third key questions determine the number and characteristics of the Blocks.

List of risk factors

All risk factors in a Block are simulated using the same Model, so it is important that they share the same statistical properties. Grouping risk factors, rather than treating them individually, also has several advantages when discussing or describing a scenario set. For example, "US interest rates mean revert," or "European stock indices grow" or "Californian electricity forwards follow a jump diffusion" all refer to groups of risk factors. Blocks allow these natural groupings to be carried into the description of scenario generation.

A further advantage of using Blocks becomes apparent when adding risk factors. In this case, incorporating a new risk factor that is similar to existing ones simply requires adding it to the appropriate Block. All of the links to the necessary Model, and the methods for parameter estimation (e.g., VCV calculations) are specified at the Block level rather than for each risk factor individually.

The other main benefit of grouping risk factors is the ability to consolidate them into a smaller group of abstract risk factors. This represents a transformation, a sub-component of the Block that is discussed further below.

Preprocessor

A **preprocessor** allows basic time series data from the database to be massaged into useful information for the simulation. For example, it is possible to track the high and low stock prices for the day in the database, and average them to get an estimate of the stock price for the day. A similar operation might involve the opening and closing stock prices or the bid and the ask quotes for foreign exchange rates. Separating the data collection and processing functions allows the precise use of the data to change when required. For example, tracking only the midpoint of the bid-ask spread limits the use of the data. In contrast, by tracking the bid and the ask prices separately, one can examine the midpoint, a weighted average, or the spread. This extra modelling freedom is advantageous in the ever-changing world of risk management.

Other examples of preprocessors include interpolation and data cleansing or checking routines. Rather than permanently updating the database with interpolated values, raw data can be stored and interpolation performed when necessary. Similarly, checks for outliers or negative values can be performed on demand.

Transformation

Transformations create abstract risk factors. An abstract risk factor is not directly observable in the market, but is usually created by combining the original risk factors in some manner. For example, the transformation principal components analysis is a VCV matrix-based technique that creates independent abstract risk factors (called principal components) by taking linear combinations of the original risk factors. The number of risk factors in the simulation can be reduced by retaining only those principal components that explain a significant amount of the VCV matrix. A reduction in the dimensionality of the simulation, combined with independence of risk factors, can significantly reduce the complexity of the modelling process.

Many transformations, other than principal components analysis, are possible. For example, an implied volatility surface might be represented by a set of equations. A transformation would then change the time series data for the surface into time series for the set of parameters of the equations. The set of possible transformations is limited to invertible functions. If a set of risk factors is transformed into a set of abstract risk factors, it must be possible to recover the original risk factors, although the recovery process may be approximate.

Model

A Model defines the future distribution or evolutionary process of a single (possibly abstract) risk factor. It does not account for any relations between risk factors, although all risk factors within a Block are assigned the same form of marginal distribution. The answer to the first key question determines the list of potential Models, while the answer to the fourth key question refines this list to include only Models actually used in generating the given scenario set. The Model includes both a simulation equation, and a method for calibrating it based on available information for each risk factor (usually historical data).

Simulation equation

The **simulation equation** provides a formula for the marginal distribution of a single risk factor at one or more points in the future. This formula is applied to all risk factors in a given Block.

The evolution of a risk factor may be described by a discretized stochastic process. Common stochastic processes include Geometric Brownian Motion (GBM), mean reversion, jump diffusions and growth models. These types of processes have several desirable properties, most notably that the distribution of the risk factor at any point in time may be determined from the underlying process, and that the risk factors evolve in a consistent manner when examined at many points on the same timeline (i.e., the same scenario). Stochastic processes are commonly used in long-term simulations (i.e., periods of more than one month). The most significant drawbacks are that such processes typically assume that the risk factor values are normally or lognormally distributed at a given time, and that there is little or no autocorrelation of the risk factor values over time. Despite these issues, many portfolio credit-risk and asset-liability measures are based on multi-step scenarios derived from stochastic processes.

Instead of evolving the risk factor through time, one can simply assume that the level of a risk factor at a specific time follows a certain distribution. This allows the use of a wider range of marginal distributions than those available from stochastic processes (in particular, empirical distributions based on historical data can be implemented). Unfortunately, the price to be paid for better-fitting marginal models is a lack of time evolution in the scenario set; by choosing the distribution at one point in the future, only that time point may be simulated. Single-step scenarios, typically used in VaR calculations, are produced in this manner.

The choice of Model or simulation equation is a critical one; it determines how well the model represents reality. One desirable property of the process that is to be simulated is stationarity. In a stationary process, the model describes the trend in the data, leaving a random component that fluctuates around a specific level (typically zero). The fluctuations themselves may or may not be of constant magnitude, but should show no discernible trend. In financial time series, returns (absolute or relative) are much more likely to be stationary than prices. For example, the basis-point changes of interest rates might be more stationary than the rates themselves. For this reason, it is often preferable to model financial returns rather than financial prices.

Calibration method

A simulation equation typically contains parameters that allow it to be tailored to individual risk factors. Usually, this tailoring requires a method for estimating the parameters (i.e., calibrating the model) from historical data. Common **calibration methods** include least squares estimates for simple linear regressions and, more generally, maximum likelihood estimators. For example, suppose that a portfolio is exposed to the S&P 500 and the Nikkei 225 stock indices. Since both indices exhibit growth over time, a growth model for log returns is assigned to the Block containing these risk factors. However, using historical data and a particular calibration method, it is determined that the S&P 500 grows at 5% per annum, while the Nikkei 225 grows only at 3.5% per annum (i.e., the parameter of the growth model is different for the two indices).

Scenario Generator

A Scenario Generator is a fully specified, fully calibrated entity that can be used to generate scenarios. It comprises three sub-components: the list of Blocks and Models, the codependent structure and the sampling method. The decisions involved in defining these components are intertwined (e.g., the codependent structure limits the types of models that may be selected). The Scenario Generator completely describes the assumptions underlying a scenario set. Answers to all five key questions influence the specification of a Scenario Generator.

List of Blocks and Models

A Scenario Generator associates a Model with each Block. This is an essential conceptual step since it allows existing Blocks and Models to be combined in different ways to create different scenario-generation models. For example, a Block containing a set of FX rates might be associated with a mean reverting model in a multi-step Monte Carlo Scenario Generator, and with a lognormal model in a single-step Monte Carlo Scenario Generator.

Codependent structure

The **codependent structure** anchors the components of the framework by defining the relations between risk factors, and specifying how these relations are incorporated into the sampling methods to obtain scenarios. Defining a codependent structure can be challenging because of its complexity, and the fact that it is closely integrated with many other parts of the scenario-generation process.

The simplest example of a codependent structure is the VCV matrix of the risk factor returns. When the risk factors have a joint-normal distribution, the VCV matrix is sufficient to describe all of the relations between them. The risk factor correlations serve as the central input for translating random numbers into the required joint-normal distribution.

More complex codependent structures might involve more sophisticated translations containing many steps, or require inputs other than a VCV matrix. For example, a historical codependent structure specifies the relationship between risk factors by setting a common data period and time horizon for the returns. Returns on each risk factor are calculated over the same time period to produce a single instance or sample. Historical sampling methods draw from the resulting finite and discrete collection of instances to create scenarios that reflect the historical risk factor correlations.

Sampling method

The third sub-component of the Scenario Generator is the technique for drawing, or sampling, from the distribution. The **sampling method** is, formally, a process for selecting a finite number of instances of a variate from its assumed or known distribution. This is usually done in two steps as follows: first, a sample from the continuous uniform distribution between zero and one (inclusive) is drawn; then a translation is applied to transform the uniform sample to one from the desired distribution (e.g., normal). Selecting the sampling method is a key decision in the scenario-generation process because it determines the efficiency of the scenario set. Efficiency refers to the number of scenarios that are required to achieve a certain accuracy, defined by a confidence interval, for a selected risk measure. Note that a scenario set that is efficient for one measure (e.g., average P&L) may not be efficient for another (e.g., VaR).

The most common sampling method is pseudo-random sampling. One such technique is the linear congruential generator, which uses an equation of the form $x_{i+1} = (ax_i + b) \bmod c$ to generate successive random numbers (a, b and c are suitable constants and the initial number, x_0, is known as the seed). The simplicity, ease of implementation and resulting statistical properties make pseudo-random sampling a popular sampling method. However, this form of sampling does not usually produce the scenario set that most efficiently represents a distribution. A large body of research describes various alternative sampling methods, such as low-discrepancy sequences, stratified sampling and antithetic variates that are more effective in this regard (see, for example, Kreinin et al. 1998a and b).

Scenario Set Definition

A **Scenario Set Definition** contains the specifics of applying a Scenario Generator to a particular problem. For example, a multi-step Monte Carlo generator may be applied to a variety of market risk and credit risk problems by selecting different trigger times (future points of interest) or numbers of scenarios. The answer to the first key question determines the Scenario Set Definition. Such a definition consists of three sub-components: the selection of a Scenario Generator, a written description of the scenario set and the specification of various structural parameters.

Scenario Generator

Each **scenario set** is created by a single Scenario Generator, which must be specified in the Scenario Set Definition. The Scenario Generator determines how the scenarios are generated, while the structural parameters determine their form and number. Together, they produce a complete scenario set.

Written description

An important, non-technical part of the Scenario Set Definition is the written description, provided by the set's creator, of the intended usage and context of the scenario set. This information allows others to understand the applications of the scenario set in the proper context. No framework can ever replace those few key comments from the scenario set designer!

Structural parameters

The designer must also decide, in defining a scenario set, what form the scenarios take. Problem-specific parameters, such as the number of scenarios to be generated and the future dates that are of interest, are specified in the Scenario Set Definition. In addition, parameters for controlling the Scenario Generator can be provided. For example, identifying the seed of the random number generator ensures that the scenarios may be reproduced precisely.

Interaction of framework components

To gain a better understanding of how the framework components interact, consider the case of a single-step Monte Carlo Scenario Generator. Suppose that there are two Blocks, each associated with a different Model, and that all of the components have been specified and calibrated as necessary. Figure 3 shows the flow of control that produces the actual scenarios.

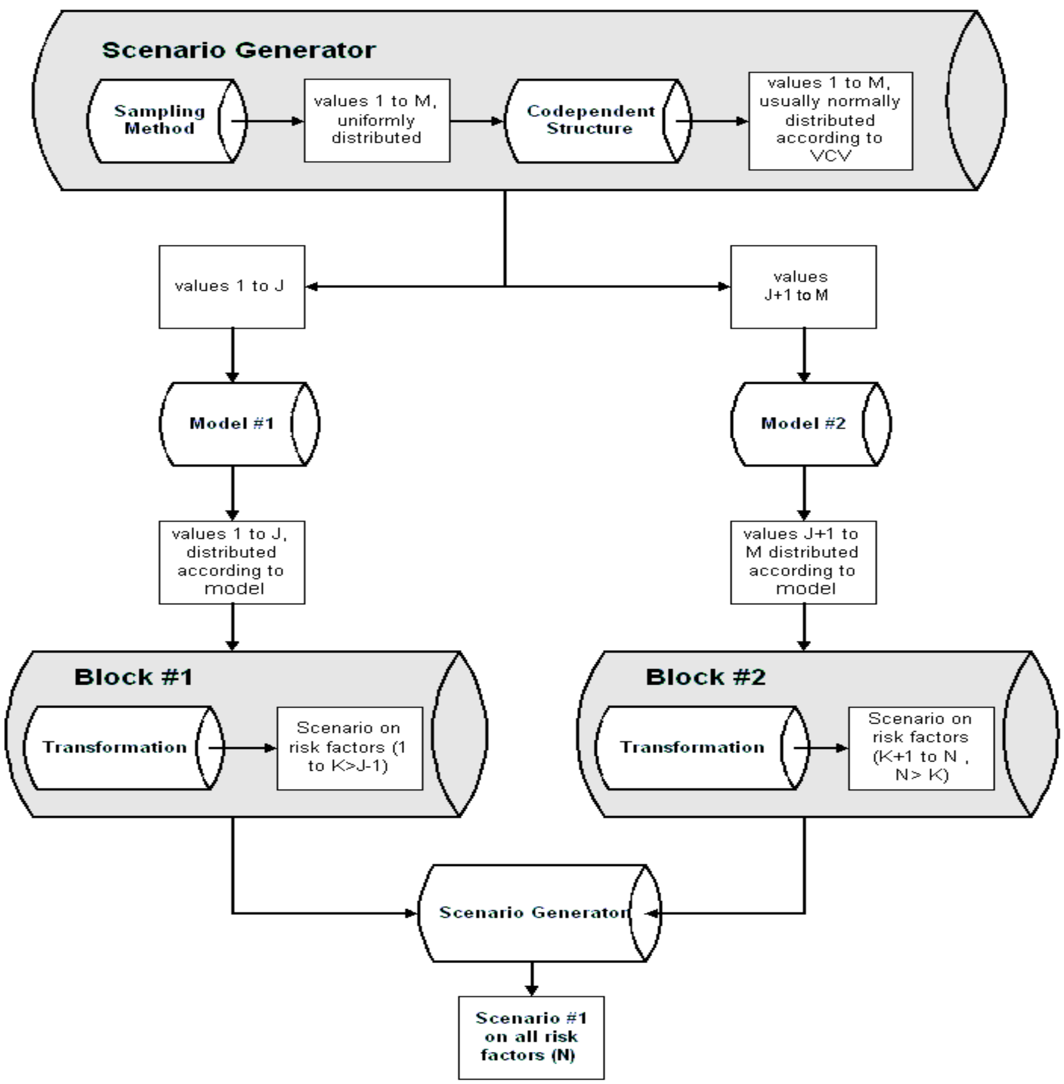

Figure 3: Data flow for scenario generation

This process can be viewed as consisting of seven steps, which must occur in sequence (see Table 3). Note that this sequence does not need to be followed when creating the components themselves, which can be defined in any order.

Strengths of the framework

The framework addresses some of the main issues in scenario generation: different audiences need information about scenarios; models are inherently complex; new models emerge and need to be incorporated; and large numbers of ever-changing risk factors must be managed. The aspects of the framework that address these problems, and the corresponding benefits, are summarized in Table 4.

The framework naturally provides several layers of abstraction: a description of its four main components (see Figure1); a more detailed description including the sub-components

Step	Associated component
1. Determine any required parameters through calibration.	All components may need to be queried.
2. Draw a sample from the uniform distribution on the interval [0,1].	Sampling method
3. Translate the sample to the proper joint distribution.	Codependent structure
4. Allocate portions of the sample to each Block.	Scenario Generator
5. Translate the sample to the proper marginal distribution.	Simulation Model
6. Transform the sample into risk factor space.	Transformation
7. Collate the results from each Block to create a scenario.	Scenario Generator

Table 3: Steps for producing scenarios

Aspects	Benefits
Levels of abstraction	Manages large numbers of risk factors and complex models Eases communication and understanding Reduces perceived complexity
Modularity and components	Allows the easy addition of risk factors Facilitates adoption of new mathematical techniques (models, transformations, preprocessors, and so on).
Separation of model from risk factors	Facilitates reuse Reduces duplication
Separation of marginal and joint distributions	Allows more modelling flexibility Provides clearer model description Facilitates reuse
Grouping of risk factors into Blocks	Simplifies reassignment of models Provides an important abstraction when the number of risk factors is large

Table 4: Aspects of the framework and their benefits

(see Figure 2); and, as will become apparent in the examples, a full description, including the details of each component. These levels of abstraction facilitate communication by allowing the audience to focus on the most appropriate level of information.

Abstraction is essential when attempting to deal with thousands of risk factors. Dealing with each one individually may be infeasible, but, by grouping risk factors, a tractable solution can be found. The complexity of each model can also be abstracted by using the framework. This allows greater accessibility to the generation process, while providing the technical details to those who require them.

The modularity of the framework promotes the reusability of components, which may be defined once and used repeatedly. The exchangeable nature of components in the framework allows for new models or risk factors to replace existing ones without disrupting the overall process.

Separating the model from the risk factors to which it applies permits a flexible implementation. Namely, a model can be easily applied to other risk factors, which promotes reuse and reduces duplication.

The separation of the marginal distributions from the joint distribution is a conceptual one. In fact, any combination of marginal models must be consistent with the joint distribution. This is a significant limitation, which must be considered at many stages of the scenario-generation process. For example, based on the availability of codependent structures (i.e., joint distributions), it may be necessary to restrict the set of possible models. However, the conceptual separation of risk factor behaviour from risk factor interactions is advantageous, since it permits the reuse of marginal models across joint distributions, and allows for the two to be described separately.

The grouping of risk factors into Blocks is essential to the management of a large number of risk factors. It facilitates the addition of risk factors whose distributions are similar to those of existing risk factors (the new risk factor is simply added to the appropriate Block, with no other changes required). The grouping of risk factors allows them to be treated later as a single entity, which is an important abstraction. Finally, a Model may be associated with an entire set of risk factors rather than with each one individually, making it much easier to change from one Model to another.

Examples

The following examples illustrate the use of the framework. The first example shows its application to bootstrapping historical scenarios for VaR calculation, while the second generates multi-step Monte Carlo scenarios for credit exposure calculations. In both cases, it is necessary to simulate a large international portfolio containing a mixture of fixed income, equity and foreign exchange instruments and their derivatives to obtain the risk measure specified in each example. Analysis of the portfolio leads to the identification of 4,210 individual risk factors spanning 28 countries and 6 risk factor classes. Three years of data is available for all risk factors (ending September 30, 2001), but only the last year is relevant to risk analysis.

From the information provided, the task of generating appropriate scenarios, for either historical VaR analysis or credit exposure calculations, is daunting. The benefit of using the framework is that it provides a structured approach to solving the problem. The examples begin by answering the five key questions (the answers are assumed to reflect the experience of a typical risk manager), and then use those answers, in conjunction with the framework, to provide comprehensive—and comprehensible—descriptions of the scenario-generation process.

Ten-day historical VaR

In this example, the goal is to calculate a 10-day historical VaR for the portfolio described above. First, the five key questions are answered below.

1. The purpose of the scenario set is to calculate a 10-day historical VaR.

2. All risk factors affecting the portfolio must be included in the scenario set.

3. The risk factors need not be grouped, but, in this case, grouping is beneficial since several risk factors share statistical properties. No transformations of the risk factors are required.

4. The daily returns for each type of risk factor need to exhibit stationarity (a suitable return process might be determined through statistical analyses of the risk factors, assumed from previous experience, or based on industry literature).

5. The technical constraints require that, in order to minimize run-time, the scenario set must be efficient for VaR calculation(i.e., it should contain just enough scenarios to accurately estimate VaR).

These answers help to identify an appropriate method for generating scenarios and also guide the construction of the various framework components. The following sections describe this process in greater detail.

Scenario bootstrapping

Several techniques can be used to generate historical scenarios for calculating 10-day VaR. Perhaps the most straightforward method is to produce scenarios from disjoint 10-day periods in history. This avoids introducing correlations between scenarios, but if the amount of data is limited, it may be possible to obtain only a small number of scenarios (e.g., one year of data will produce about 25 scenarios). Conversely, one could use a rolling window approach, in which case two consecutive scenarios share nine days worth of data. This provides a larger number of scenarios (250 per year of data in this case), but the scenarios show a high degree of autocorrelation, which reduces the accuracy of the VaR estimate.

Scenario bootstrapping is a more sophisticated technique that is most useful when scenarios must span a relatively long horizon (e.g., 10 days), but only a limited amount of historical data (e.g., one year) is available. Scenario bootstrapping repeatedly samples daily changes (i.e., returns) in the risk factors in order to construct a scenario spanning a longer time horizon. In this case, a scenario of 10-day changes in risk factor levels is obtained by randomly selecting 10 daily changes from the past year and calculating their cumulative effect. The method assumes that day-to-day returns are independent, which is generally reasonable in practice. Although many financial time series do exhibit some degree of autocorrelation, the assumption that daily returns are independent is consistent with the efficient market hypothesis and the use of Brownian motion in multistep Monte Carlo models.

Blocks

In order to define the Blocks, the 4,210 risk factors are broken down in a logical fashion. This task proceeds in three stages as follows. First, each risk factor is assigned to a broader class, such as interest rates, foreign exchange rates, equities and several types of implied volatilities. Next, risk factors within a class are broken down by region. Typically, an international portfolio is subdivided into its regional components for management purposes, and so the division of risk factors by region is also quite natural. Finally, certain risk factors (called points) are grouped together to form logical curves for interest rates or logical surfaces for implied volatilities. This groups the 4,210 individual risk factors into only 676 composite risk factors (see Table 5), each of which is then modelled separately.

Table 6 contrasts the number of risk factors in each class before and after the grouping (into logical curves and surfaces). An analysis of the risk factors (discussed later on in the section on Models) indicates that all risk factors in the same class have the same return type.

.

Risk factor class	Geographical region	Description
Interest rates	North America	Two curves per currency; 25 points per curve; two currencies
	Europe	One curve per currency plus the Euro swap curve; average 15 points per curve; 15 currencies
	Asia-Pacific	One curve per currency; 18 points per curve; seven currencies
	South America	One curve per currency; eight points per curve; four currencies
Foreign exchange	All	Against USD; 28 in total
Equity indices	North America	195 sector indices; 180 in USD; 15 in CAD
	Europe	25 Euro-sector indices; 25 GBP sector indices
	Asia-Pacific	15 sector indices per currency; seven currencies
	South America	one equity index per currency; four currencies
Implied volatility–interest rate	North America	Swaption, cap and floor curves in USD; 25, 16 and 16 points, respectively
	Europe	Swaption, cap and floor curves in EUR; 25, 16 and 16 points, respectively
	Asia-Pacific and South America	None
Implied volatility–foreign exchange	All	250 surfaces; average 12 points per surface
Implied volatility–equity	North America	Surfaces for five main indices; 36 points per surface
	Europe	Surfaces for two main indices; 36 points per surface
	Asia-Pacific and South America	None

Table 5: Breakdown of risk factors by class and currency

In this case, it is sufficient to have one Block for each class of risk factor. Specifically, the Blocks are:

- Interest Rates (IR)
- Foreign Exchange (FX)
- Equity Indices (EQ)
- Implied Vols–IR
- Implied Vols–FX
- Implied Vols–EQ

Risk factor class	Number of risk factors	Number of points
Interest rates	31	498
Foreign exchange	28	28
Equity indices	354	354
Implied vols–IR	6	114
Implied vols–FX	250	3,000
Implied vols–EQ	7	216
Total	676	4,210

Table 6: Effect of logical grouping on numbers of risk factors

For all Blocks, the historical data requires preprocessing to ensure valid results. For instance, outliers are removed and linear interpolation is used to fill in any missing time series information. Note that the return type for the Foreign Exchange Block is logarithmic (see Table 7). Since one cannot take the logarithm of a negative number, an extra preprocessor is added to this Block to check for negative values. Negative values are removed and treated as missing data. As is typical of historical scenarios, no transformations are selected in the Blocks.

Models

Since scenario bootstrapping combines random daily returns, it is important that the daily returns for each risk factor be stationary over the period for which data is available. To achieve this, the manner of calculating daily returns varies depending on the risk factor. In this case, a separate Model is associated with each type of daily return.

Formally, let X_t denote the risk factor value at the current time t and let X_i denote the risk factor value at some previous time i. The risk factor value X_s in scenario s is then calculated as shown in Table 7, in a manner consistent with the statistical properties of its respective class. For example, equity indices tend to change substantially in value over time, which makes many return types non-stationary. The choice of percentage change returns, in contrast, is relatively stationary for most equity indices.

From Table 7, it follows that five Models for return calculations must be created. Note that the Model specifies only the return type; the time horizon and data periods for all risk factors are specified simultaneously in the codependent structure, as described in the following section.

Scenario Generator

Models are assigned to Blocks as per Table 7, recognizing that Blocks and Models correspond to risk factor classes and return types, respectively.

Risk factor class	Return type	Calculation
Interest rates	Basis point shift	$X_s = X_t + 10{,}000^*\{(X_i - X_{i-1})/10{,}000\}$
Foreign exchange	Logarithmic change	$X_s = X_t^* \exp(\log(X_i/X_{i-1}))$
Equity indices	Percentage change	$X_s = X_t^*(X_i/X_{i-1})$
Implied vols–IR	Historical value	$X_s = X_i$
Implied vols–FX	Absolute change	$X_s = X_t + (X_i - X_{i-1})$
Implied vols–EQ	Percentage change	$X_s = X_t^*(X_i/X_{i-1})$

Table 7: Return types by risk factor class

Since scenarios are constructed by sampling actual daily returns for all risk factors simultaneously, a historical codependent structure is obtained. That is, the correlation between the changes in risk factors in the scenario set is directly determined from the historical data, since risk factor changes that coincided in the past are replicated in the scenario set. This example assumes that the historical data is taken from the one-year period between October 1, 2000 and September 30, 2001. Note, also, that scenario bootstrapping uses a one-day time horizon for return calculations, even though 10-day VaR will be estimated. Approximately 250 one-day returns are available since the returns of all risk factors are calculated from the same one-year period, and there are approximately 250 business days per year.

The sampling method is random, from a uniform distribution, with replacement. This means that a pool containing the 250 one-day returns (for all risk factors) is created. For each scenario, 10 daily returns are drawn from the pool and accumulated to obtain a 10-day return. All returns in the pool are equally likely to be selected and, since sampling occurs with replacement, the pool always contains the same number (250) of returns.

Scenario Set Definition

The Scenario Set Definition specifies the future dates that are of interest and the number of scenarios to be created. Since the objective is to obtain the 10-day VaR, a set of representative portfolio values is required at a single date only, 10 days in the future. The required number of scenarios depends in large part on the level at which VaR is to be calculated, and on the desired accuracy of the estimate. Assuming the 95% VaR is to be calculated, for instance, 500 scenarios may be deemed to provide sufficient accuracy. Finally, a description of the scenario set helps to explain its applicability. In this case, "500 bootstrapped scenarios from one year of data for calculating a 10-day historical VaR for the portfolio" is adequate for this purpose.

Note that the framework does not address issues such as the pricing of the portfolio under each scenario 10 days into the future, and the details of estimating 95% VaR.

Credit exposures

Now, suppose that the objective is to assess the credit riskiness of the portfolio on an annual basis for the next 10 years. This requires evolving risk factors over time in a consistent fashion. Since there is clearly insufficient historical data available in this case, multi-step Monte Carlo scenarios are appropriate for addressing this problem. This example, which illustrates a typical Monte Carlo model, again begins by answering the five key questions.

1. The purpose of the scenario set is to estimate credit exposures over a long time horizon.

2. All risk factors shown in Table 5 must be included in the analysis.

3. As indicated in the answer to the fifth key question, the number of risk factors needs to be reduced. Principal components analysis (PCA) and functional approximations are effective solutions to this problem.

4. Interest rates, foreign exchange rates and equity indices must each evolve over time according to an appropriate stochastic process. Moreover, interest rates and FX rates display mean reversion, while equity indices tend to grow over time. Implied volatilities of all types are assumed to be distributed randomly over time, but must remain positive.

5. One of the basic inputs to the Monte Carlo model is the variance-covariance (VCV) matrix. (The VCV matrix fits into the framework as part of the calibration of the codependent structure.) In this example, the number of risk factors in the overall VCV is too large; simply calculating the VCV matrix exceeds the time allotted to scenario generation in the overnight process. Unlike the previous example, which did not require the calculation of a VCV or the retention of many time steps per scenario, the number of risk factors is too high to make this scenario set practical.

Blocks

Partitioning the risk factors into Blocks and applying proper transformations are the keys to reducing the number of risk factors. Again, the first step in the process is the classification of the risk factors (Table 5). While it is natural to use the same breakdown as in the previous example, other classifications are of course also possible.

The breakdown of the entire list of risk factors into Blocks is guided by knowledge of the nature of the risk factors. Here, the assumption is that all risk factors within a class exhibit similar time series properties. In fact, many subclasses could exist. For example, it is possible to model risk factors from emerging markets differently from those relating to more established markets. Transformations can be used to reduce the number of points in composite risk factors. PCA is generally appropriate when applied to interest rate curves and FX-implied volatilities. Alternatively, functional representations of EQ-implied volatilities and IR-implied volatilities may be considered acceptable in long-term simulations such as the current example.

Interest rate Blocks

Preprocessing requirements for interest rate Blocks include interpolation, extrapolation and the elimination of both negative rates and negative forward rates. Other issues of data integrity can also be addressed at this stage. Historical data, while not used to produce scenarios directly as in the previous example, is still important for the calibration of the Monte Carlo models.

Replacing each interest rate curve with its three most significant principal components reduces the number of interest rate risk factors in the simulation from 498 to 93. To accomplish this, one Block is created containing each interest rate curve (note that grouping several interest rate curves into a single Block is not possible where curve-by-curve PCA is required). A PCA transformation, calibrated with an equally weighted VCV matrix based on one year of data, is applied to each Block. The VCV matrix needs only to include the risk factors in the Block; correlations between different interest rate curves are of no interest in the PCA process. It is then necessary to model the evolution of the three abstract risk factors for each of the 31 interest rate Blocks.

FX Block

A single Block is created for all 28 exchange rates. It is not necessary to perform any transformations in this case. Since daily returns of foreign exchange rates are lognormally distributed, preprocessors are used to filter extraordinary data, interpolate for missing values, calculate the average of the bid and ask to be used as the value, and check for negative values, which are removed and replaced by interpolated amounts. This preprocessing allows for more accurate model calibration in the scenario-generation process.

Equity Blocks

Separate Blocks are created to contain each of the North American, South American, European and Asian-Pacific equity indices. Risk factor data for all Blocks are preprocessed to remove outliers and interpolate missing values. Since 300 of 354 indices reside in North America or Asia-Pacific, PCA is performed on each of these two Blocks. For exposition purposes, assume that each Block can be represented in terms of 12 abstract risk factors (i.e., principal components).

Implied volatility Blocks

Each of the 263 implied volatility surfaces is assigned its own Block, and standard preprocessing removes outliers and interpolates missing values. A transformation is applied to each Block in order to reduce the number of risk factors. PCA is used to represent FX implied volatility surfaces in terms of two abstract risk factors, thereby reducing the number of associated risk factors from 3,000 to only 500. For interest rate and equity implied volatilities, a functional transformation requiring two and three parameters, respectively, is used to parametrize each volatility surface. The parameters effectively become the abstract risk factors that must be simulated.

The above transformations reduce the number of risk from 4,210 to 732 (see Table 8). This represents an 82% decrease in the number of risk factors, and a 97% decrease in the size of the VCV matrix required to implement the codependent structure. Note that the largest VCV matrix that was required to perform PCA was of size 195 by 195 (for the North American equity indices). The resulting computational effort is insignificant in comparison to the overall savings in calculations.

Risk factor type	Number of abstract risk factor (points)
Interest rates	93
Foreign exchange	28
Equity indices	78
Implied vols–IR	12
Implied vols–FX	500
Implied vols–EQ	21
Total	732

Table 8: Number of abstract risk factor points by risk factor type

Models

The evolution of the risk factors is determined by appropriate stochastic processes in this case.

Foreign exchange rate returns mean revert and follow a lognormal distribution. The process is calibrated using least squares estimates for the rates and levels of reversion based on two years of historical data.

Interest rate returns also mean revert and are lognormally distributed. However, by performing PCA, normally distributed principal components were created (they also mean revert). The normal mean reverting process for the principal components is calibrated using maximum likelihood estimates for the rates and levels of reversion based on three years of historical data.

Equity indices grow over time and are typically lognormally distributed. Growth rates are estimated using the least squares estimates of the slope of the log value of the index based on three years of data.

Principal components used to represent the North American and Asian equity indices are modelled using a normal distribution plus a growth component. Calibration similar to that for the natural equity index model is employed.

Implied volatilities are randomly distributed through time, without any visible trend. A Brownian Motion model is created for the foreign exchange implied volatility principal components. The parameters derived for both interest rate and equity implied volatility are found to be lognormally distributed, and a Geometric Brownian Motion model is created for these risk factors. These models have no parameters beyond the VCV matrix that is calculated as part of the codependent structure calibration.

Scenario Generator

The Scenario Generator associates Blocks and Models as per Table 9. The codependent structure is a simple VCV matrix of dimensions 732 by 732 calculated on an equally weighted basis over three years of data. Pseudo-random sampling is used to produce 10 annual samples for each scenario. Figure 3 shows the flow of the random samples through the other components of the framework to produce a scenario. The steps of scenario generation are similar to those outlined in Table 3 for the single-step Monte Carlo case.

Scenario Set Definition

Suppose that 1,000 scenarios are judged to be sufficient for calculating the credit exposure to each counterparty. In this case, the Scenario Set Definition specifies that 1,000 scenarios be generated for all possible risk factors at annual trigger times for a 10-year period. A suitable description of the resulting scenario set might be “1,000 annual scenarios for 10 years based on multi-step Monte Carlo processes specific to each type of underlying factor.”

Note that further simulation is required to induce changes in counterparty creditworthiness under each market scenario in order to measure credit exposures. This processing, while necessary, is beyond the scope of scenario generation considered in this paper.

Block (by risk factor type)	Model
Interest rates	Normal with mean reversion
Foreign exchange	Lognormal with mean reversion
Equity indices (North American and Asian)	Normal with growth
Equity indices (European and South American)	Lognormal with growth
Implied vols–IR	Geometric Brownian Motion
Implied vols–FX	Brownian Motion
Implied vols–EQ	Geometric Brownian Motion

Table 9: Blocks and Models

Assessing the benefits

The preceding examples show how the five key questions and the framework help to organize a complex process into a more manageable set of tasks. Despite the large number of diverse risk factors, the framework allows the scenario-generation process to be described clearly and concisely in a modular fashion. This makes it easier to communicate the nature of the scenario set to various audiences.

By using the components in a consistent manner, it is possible to identify quickly the impact of changing an existing risk factor, model or transformation. In the case of historical bootstrapping, for instance, modifying the return type for implied volatilities on equities from percentage change to historical value merely requires attaching the implied volatility–EQ Block to the historical value Model, rather than the percentage change Model, in the Scenario Generator.

As an illustration of the framework's extensibility, observe that adding a new risk factor to an existing class of risk factors is as simple as including it in the appropriate Block. This task can be clarified through rules for establishing membership in each Block, such as, "all foreign exchange risk factors belong to the FX Block."

Note that the codependent structure does not necessarily require updating after adding a new risk factor, since the framework contains only a description of how to calculate the codependencies, rather than the structure itself. For example, the codependent structure might contain a formula for estimating a VCV, but not the actual matrix itself.

Incorporating a new risk factor category is also straightforward. This involves creating a new Block containing the new risk factors, along with any required preprocessors or transformations, and then associating the Block with a new or existing Model in the Scenario Generator. Similarly, new Models can be created and then attached to Blocks, as necessary, in the Scenario Generator.

Conclusions

The process of generating statistical scenarios is based on the answers to five key questions. The answers are used to decompose the process into a series of identifiable, reusable and self-contained components within the scenario-generation framework. The framework relates the thought process and goals of the risk manager directly to the scenario-generation process.

The framework addresses some of the most significant issues in scenario generation: varied audiences need information about scenarios; models are inherently complex; new models need to be incorporated as they emerge; and large numbers of ever-changing risk factors must be managed. To deal with these issues, the framework provides several levels of abstraction; a modular structure; a separation of risk factors and models; and a separation of the joint and marginal distributions. The strengths of the framework include its flexibility, extensibility and usefulness in explaining large-scale scenario-generation processes clearly and concisely.

By breaking down the scenario-generation process into four main components, each with a limit of three sub-components, the process becomes more transparent. Each component of the generator can be examined until it is understood at a sufficient level of detail. Once each component is understood, the overall picture presented by a scenario set is just a tiny step—not a leap of faith—away.

References

Dembo, R., A. Aziz, D. Rosen and M. Zerbs, 2000, *Mark-to-Future: A Framework for Measuring Risk and Reward*, Toronto, ON: Algorithmics Publications.

Kreinin, A., L. Merkoulovitch, D. Rosen and M. Zerbs, 1998a, "Measuring portfolio risk using quasi-Monte Carlo methods," *Algo Research Quarterly,* 1(1): 17–25.

Kreinin, A., L. Merkoulovitch, D. Rosen and M. Zerbs, 1998b, "Principal component analysis in quasi- Monte Carlo simulation," *Algo Research Quarterly,* 1(2): 21–29.