ERM IN AN OPTIMAL CONTROL FRAMEWORK

Greg Taylor	Shaun Wang
Taylor Fry Consulting Actuaries Level 11, 55 Clarence Street Sydney NSW 2000 Australia	Georgia State University Robinson College of Business 11th floor, 35 Broad Street Atlanta, GA, 30303, USA
Phone: 61 2 9249 2901 Fax: 61 2 9249 2999 greg.taylor@taylorfry.com.au	Phone: 1-404-413-7486 shaunwang@gsu.edu

May 2013

Abstract

Much of ERM consists of qualitative discussion of the risks facing a business and controls over them. It is difficult to identify in the literature a clear body of theory to provide the foundation for the subject, integrating a business's objectives with its risk controls.

The present paper attempts this by formulation of ERM as an exercise in stochastic optimal control theory. Here there is a defined objective, which would usually include some aspect of profit, and a set of constraints (the risk controls). Optimal control theory provides a framework for balancing the one against the other, and also for considering whether or not particular risk controls are well advised.

The paper accepts the COSO (2004) definition of ERM, and its associated ERM Integrated Framework. After definitions, preliminary discussion and establishment of the control theory set-up, the paper is organised with one section for each item of the ERM Integrated Framework. Each of these sections interprets that item within the control theory model.

Keywords: enterprise risk management, ERM, optimal control theory, risk appetite.

1 Introduction

1.1 Background

The literature in **Enterprise Risk Management**, or **ERM**, has increased rapidly from about the turn of the century.

The origins of this can be found in the **Dynamic Financial Analysis ("DFA")** models of the 1990s, which in broad terms were simulation models usually encompassing an insurance company's insurance risk and market risk (e.g. Kaufmann, Gadner & Klett, 2001). Subsequently, these were expanded to ERM models that endeavoured to address all risks faced by the company, often with particular reference to operational risk.

The emphasis of DFA models had been one of micro-modelling, ever more minute detail of the insurance and asset management aspects of the insurance company. The models were detailed and precise.

The inclusion of operational risk added a dimension in which detailed and precise models were generally not readily available. As a consequence, papers on ERM have often tended to take on a more descriptive nature.

Nonetheless, the subject is widely perceived as of considerable importance. It receives impetus from financial regulators, some of whom administer prudential regulation that enshrines ERM.

1.2 Objectives of this paper

The descriptive nature of ERM has led to papers that contain lengthy lists of the risks faced by a business and discussion of how they may be controlled. The controls often come with a cost to the business or a restriction of its practices. The balance of these restrictions against the profit objectives is not always discussed in depth.

This may reflect the absence of a clear body of theory by means of which such balance may be investigated. Indeed, when we survey the ERM literature, we do not identify a foundational mathematical statement of objectives, a mathematical formulation of the ERM problem.

The first objective of the present paper is to define a recognisable mathematical framework within which ERM may be viewed, and then to provide a precise formulation of the problem posed by ERM. The intention is that the framework should readily accommodate all standard ERM concepts.

The entity to which ERM is applied here is the **risk business**. This is defined as any commercial enterprise whose outcomes are in some sense stochastic. This definition is sufficiently wide to include virtually all enterprises.

They might, but need not be, insurance companies. The "risk business" terminology is chosen so as to emphasise that the principles discussed here are of general applicability. Nonetheless, in view of the authors' backgrounds, illustrative examples will often be drawn from an insurance context.

2 Definition of ERM

2.1 COSO definition

The ERM landscape is littered with definitions of ERM. The Casualty Actuarial Society has its definition, for example, at http://www.casact.org/area/erm/frame.pdf. The Institute and Faculty of Actuaries and the Institute of Actuaries of Australia adopt the definition of Chapman (2006, pages 8-9).

The present paper will, however, adopt the definition given by the Committee of Sponsoring Organisations of the Treadway Commission (COSO), which is as follows:

> "Enterprise risk management is a **process**, effected by an entity's board of directors, management and other personnel, applied in **strategy setting** and **across the enterprise**, designed to identify **potential events** that may affect the entity, and **manage risk** to be **within its risk appetite**, to provide **reasonable assurance** regarding the **achievement of entity objectives**." (COSO, 2004)

Each of the bold passages in this definition represents a key component of it. The fact that such a high proportion of the definition is bold hints at its pithy nature. Indeed, it seems to us that this definition includes all the necessary components, and nothing more.

2.2 COSO integrated framework

As a step toward operationalising their definition of ERM, COSO (2004) produced their **ERM Integrated Framework**, which discusses ERM according to the following aspects:

- internal environment;
- objective setting;
- event identification;
- risk assessment;
- risk response;
- control activities;
- information and communication;
- monitoring.

Subsequent sections of this paper will consider these aspects within the framework established here.

3 ERM as a control process

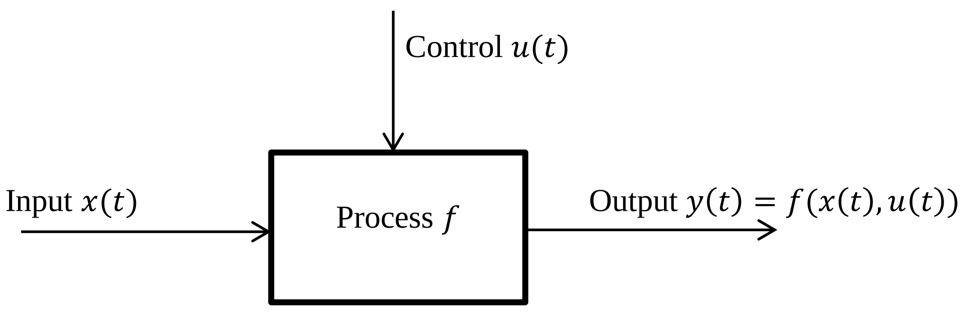
3.1 Optimal control theory

One of the paper's objectives, mentioned in Section 1.2, is to gain precision by framing a number of principal ERM concepts in mathematical language. The paper is founded on a proposition that ERM may be formulated as an exercise in optimal control. The next two sub-sections therefore take a digression into the rudiments (no more than that) of that theory.

3.1.1 Deterministic

Consider the system represented diagrammatically in Figure 3-1 in which some process transforms an **input** $x(t)$ at time t to an **output** $y(t)$. One may influence the input-output response by means of an external intervention $u(t, x(t))$, which is referred to as a **control**. The output is then $y(t) = f(x(t), u(t))$, depending on both input and control.

Figure 3-1 A deterministic control system



A simple example is given by a transistor amplifier. The transistor carries a current from the input, called the emitter, to the output, called the collector. It has a third terminal, called the base, which functions as the control. Changes in the voltage applied to the base affect the output voltage, the effect being linear over some ranges of the input and control.

If the input and control are held constant, then the output remains constant. If one now arranges that the output be substantially larger than the control, variations in the control produce a proportionate but amplified response in the output.

Returning now from this particular example to the general case, suppose that a performance measure is applied to the system in the form of a functional $J[y]$ of the entire output over the time interval $[0, T]$ up to some time horizon. One may wish to vary the control in such a way as to optimise system performance, i.e. maximise (or minimise) J.

As an example, the performance measure J might be the mean square error of the output's departure from some target value y_0, i.e.

$$J[y] = T^{-1} \int_0^T [y(t) - y^*]^2 \, dt = T^{-1} \int_0^T [f(x(t), u(t)) - y_0]^2 \, dt \quad (3.1)$$

and, for given input $x(t)$, the optimal control $u^*(t)$ is given by

$$u^*(t) = arg\, min_{u(t)} J[y] \quad (3.2)$$

In this set-up J is usually called the **objective function,** or **loss function** or **penalty function**.

General formulation

For a more general formulation of these ideas, let $x(t), y(t), u(t)$ denote the system input, output and control, as above except that these are now m- and n- and r-vectors respectively. Further, define

$$X(t) = \{x(s): 0 \leq s \leq t\} \quad (3.3)$$

$$Y(t) = \{y(s): 0 \leq s \leq t\} \quad (3.4)$$

$$U(t) = \{u(s): 0 \leq s \leq t\} \quad (3.5)$$

and assume the system output takes the form

$$y(t) = f(X(t), U(t)) \quad (3.6)$$

where now $f: \Re^{m+r} \rightarrow \Re^n$, i.e. the output at any point of time may depend on the totality of the inputs and controls since initiation. Equation (3.6), describes the system dynamics and is referred to as the **state equation**.

In this formulation, assume that the objective function takes the form

$$J[y] = \int_0^T L[t, y(t)] \, dt \quad (3.7)$$

where $L: \Re^{n+1} \rightarrow \Re$ is called the **loss intensity function**. Although it is explicitly dependent on only t and $y(t)$, it is also dependent on the system controls through (3.6).

The optimal control $u^*(t)$ is still given by (3.2), though adapted to the new framework.

Terminal value

The objective function (3.7) implies that the behaviour of the system beyond the time horizon is irrelevant to the performance measure. Even the behaviour precisely at the horizon is irrelevant because it is assigned measure zero in the objective function.

Occasionally it may be desired that positive measure be assigned to the time horizon, and this may be done by adding a **terminal value** to the objective function thus:

$$J[y] = \int_0^T L[t, y(t)]\, dt + \Phi[T, y(T)] \qquad (3.8)$$

for suitable $\Phi: \Re^{n+1} \to \Re$.

Constraints

The system output may be subject to restrictions additional to those that arise from the optimisation of (3.8). It may be undesirable, for example, that the system response stray beyond the range $y_1 \le y(t) \le y_2$. An insistence that the response remain always within this range constitutes a constraint on the system.

The general form of constraint is $\mathcal{R}[Y(T)]$ for some relation $\mathcal{R}$, e.g.

$y_i(t) \le c_{(1)}, 0 \le t \le T$

$y_j(T) \le c_{(2)}$

$g\big(y_k(T)\big) \le c_{(3)}$ for some function $g(.)$

etc., where the $c_{(i)}$ are constants.

Summary

The summary statement of the control problem described above, including terminal value and constraints, is as follows:

$$u^*(t) = arg\, min_{u(t)} J[y] \qquad (3.2)$$

where

$$J[y] = \int_0^T L[t, y(t)]\, dt + \Phi[T, y(T)] \qquad (3.8)$$

$$y(t) = f(X(t), U(t)) \qquad (3.6)$$

and subject to constraints $\mathcal{R}_i[Y(T)], i = 1,2, \dots, I$.

The theory of such problems is called **optimal control theory**. See e.g. Speedy, Brown & Goodwin (1970).

3.1.2 Stochastic

All quantities in the discussion of Section 3.1.1 are deterministic. Consider the more general case in which the inputs $X(T)$ are stochastic, subject to measure $\mathcal{P}$. The performance measure (3.8) is then also stochastic and, in which case the optimisation (3.2) will be meaningful only if the stochastic functional $J[y]$ is mapped to a deterministic quantity.

This may be achieved by replacing (3.8) with the following:

$$K_{\mathcal{P}}\big[J[y]\big] = K_{\mathcal{P}}\left[\int_0^T L[t, y(t)]\,dt + \Phi[T, y(T)]\right] \tag{3.9}$$

where $K_{\mathcal{P}}: \mathbb{P} \times \Re \to \Re$, with $\mathcal{P} \in \mathbb{P}$, a family of probability measures. Thus, $K_{\mathcal{P}}$ is a mapping of J to a real number depending on the probability measure $\mathcal{P}$.

A simple example, and in fact the only one that will be required in this paper, is $K_{\mathcal{P}} = E_{\mathcal{P}}$, the expectation with respect to measure $\mathcal{P}$. In this case, (3.9) becomes

$$E_{\mathcal{P}}\big[J[y]\big] = E_{\mathcal{P}}\left[\int_0^T L[t, y(t)]\,dt + \Phi[T, y(T)]\right] \tag{3.10}$$

Constraints may still be written in the form $\mathcal{R}_i[Y(T)]$, as in Section 3.1.1, but must recognise that $Y(T)$ is now subject to measure $\mathcal{P}$. This fact can be made explicit by representing a constraint as $\mathcal{R}_i[Y(T); \mathcal{P}]$. Examples of such constraints would be:

$$E\big[y_j(T)\big] \le c_{(1)} \tag{3.11}$$

$$Prob\big[y_k(T) \le c_{(2)}\big] \le p_{(2)} \tag{3.12}$$

Thus the formulation of the control problem is now as follows:

$$u^*(t) = arg\,min_{u(t)}\, K_{\mathcal{P}}\big[J[y]\big] \tag{3.13}$$

where

$$K_{\mathcal{P}}\big[J[y]\big] = K_{\mathcal{P}}\left[\int_0^T L[t, y(t)]\,dt + \Phi[T, y(T)]\right] \tag{3.9}$$

$$y(t) = f(X(t), U(t)) \tag{3.6}$$

and subject to constraints $\mathcal{R}_i[Y(T); \mathcal{P}], i = 1,2, \dots, I$.

The theory of such problems is called **stochastic optimal control theory**.

Stochastic optimal control of dynamic systems

There is a sub-class of the problems defined above that is interesting because a specific equation defining their solutions exists. This is the sub-class of dynamic systems in which the measure $\mathcal{P}$ is defined by a stochastic process on $x(t)$, described by the stochastic differential equation (see e.g. Arnold (1974))

$$dx(t) = g\Big(t, x(t), u\big(t, x(t)\big)\Big)\,dt + h\Big(t, x(t), u\big(t, x(t)\big)\Big)\,dW(t) \tag{3.14}$$

where g, h are suitable functions and $W(t)$ is a standard Wiener process.

Equation (3.14) describes how the process $x(t)$ evolves over time. The first member on the right side is deterministic and the second member adds a stochastic component. The process is initialised by a given value of $x(0)$. For subsequent tractability, only **Markov control functions**, i.e. depending on just the current state of the system and not its prior history, are considered here.

Consider the case of an expectation objective function, as in (3.10), and suppose that at time s ($0 \le s \le T$) it takes the form

$$K[s,x] = E\left[\int_s^T L[t, x(t), u(t, x(t))]\, dt + \Phi[T, y(T)]\right] \tag{3.15}$$

This form will be compatible with (3.10) in certain cases. Although the integral in (3.10) is expressed in terms of $y(t)$, this may be re-expressed in terms of $X(t)$, by (3.6), and in certain cases this will be consistent with (3.15).

For completeness, the solution of the optimisation (3.13) will now be given, though it will not be used in the remainder of the paper. This matter will be discussed further in Section 3.2.4.

By Bellman's principle, $u^*(t, x(t))$ will be optimal for all $t \in [0, T]$ if it optimises $K[s, x]$ for each $s \in [0, T]$ and the solution is given by the **Hamilton-Jacobi-Bellman equation**:

$$\min_u\{\mathcal{A}K[s,x] + L[s, x(s), u(s, x(s))]\} = 0 \text{ for all s such that } 0 \le \text{s} \le \text{T} \tag{3.16}$$

subject to the end condition

$$K[T, x] = \Phi[T, y(T)] \tag{3.17}$$

where the operator $\mathcal{A}$ is defined as

$$\mathcal{A} = \frac{\partial}{\partial s} + \sum_{i=1}^{m} g_i(s,x,u)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}[h(s,x,u)h^T(s,x,u)]_{ij}\frac{\partial^2}{\partial x_i \partial x_j} \tag{3.18}$$

Here x_i, g_i denote the i-th components of x, g respectively (recall that $x(t)$ is m-dimensional) and $[hh^T]_{ij}$ denotes the (i,j) element of the $m \times m$ matrix hh^T.

One may be inclined to question how one has profited from this solution. It appears that one complicated problem, minimisation of (3.15), has been swapped for another, (3.16). In fact, the minimand is transformed from an integral in (3.15) to a differential quantity in (3.16) and there are well known special cases in which this enables closed form solutions. Nonetheless, many cases of (3.16) are not capable of explicit solution and numerical solutions are required.

3.2 ERM formulation

The present sub-section will be concerned with the formulation of ERM as an exercise in stochastic optimal control. Although the prior literature has not considered ERM in this light, it has hinted at it. For example, Orros & Smith (2012, Figure 8) include a diagram almost identical to Figure 3-1 but without mention of control theory.

3.2.1 State equation

The state equation of the control system described in Section 3.1.1 was given as (3.6), and this was said to describe the system dynamics, as illustrated in Figure 3-1. The adaptation of that diagram to the ERM context is as in Figure 3-2.

Figure 3-2 An ERM control system

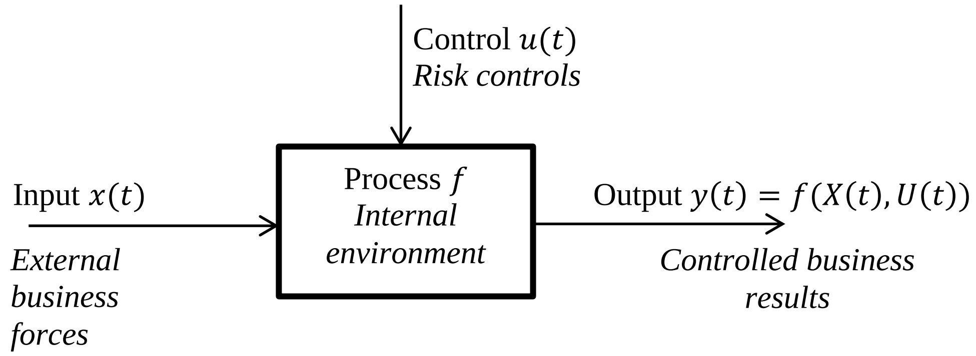


Figure 3-2 is intended to apply generically to businesses that are subject to risk. However, specific illustration may be given by reference to a general insurance company. Here,

- the **external business forces** would include natural hazards, regulation, the economy, competition, etc.;
- the **business results** would usually be financial, such as gross revenue, profit, dividend, economic value, etc.;
- the **internal environment** is the totality of the company's internal workings that transform the input forces to output results;
- the insurer's **risk controls** are those levers that its management may operate to modify this transform and so affect the output results, this being done in such a way as to influence the risk (discussed in Section 7) to which those results are subject.

Henceforth it will be assumed that all inputs and outputs are financial quantities.

3.2.2 Objective function

The objective function to be used in this paper was set out in (3.10). It will usually be possible to articulate the objectives of a risk business in this form. For example, the loss intensity function $L[t, y(t)]$ might denote the instantaneous financial loss at time t, discounted to time zero. Here, negative loss equals profit, and minimisation of loss equals maximisation of profit.

With this definition of $L[t, y(t)]$, the first of the two members on the right side of (3.10) represents the negative of expected discounted losses over the next T time periods. A terminal value $\Phi[T, y(T)]$ consistent with this would be some estimator of (the negative of) expected economic value of the business at time T, discounted to time zero.

3.2.3 Constraints

Constraints will generally fall into two categories:

- **externally** imposed restrictions on the way one may conduct one's business, e.g. regulatory restrictions; or
- restrictions that are imposed **internally** as "business rules", e.g. prohibitions on specific ranges of outputs as undesirable from a business viewpoint.

The general form of constraint was given as $\mathcal{R}_i[Y(T)]$ just after (3.10), and a couple of specific examples given in (3.11) and (3.12).

3.2.4 Discussion

Sections 3.2.1 to 3.2.3 have formulated ERM as a stochastic optimal control problem. The details of how this formulation relates to an ERM situation will be fleshed out in Sections 4 to 8.

The formulation is precise and, if the required inputs can be provided, a precise solution will be obtainable. However, a brief word on the practicalities of this is advisable at this early stage of the paper.

There are three categories of input to be considered:

(a) Business objectives, contributing to the definition of the objective function;
(b) Constraints;
(c) Probability measures, defining the statistical expectation required in (a) and (b).

There may be no difficulty in defining (a) and (b). Indeed, it seems a healthy exercise for a risk business to contemplate these aspects of its operation and articulate them as precisely as possible.

However, the probability measures required in (c) may be vague and elusive, as will be discussed in context in Sections 6.3 and 6.4. It should be recognised that unreliability of inputs has the potential to reduce the apparent precision of the above methodology to a travesty. It is not the suggestion of this paper that that methodology be pursued doggedly to the point of nonsensical conclusions, but rather that the situation be informed by any useful content of the theoretical framework.

The situation is similar to many others involving actuarial evaluation. Many of these may be formulated in precise mathematical terms but with inputs that are shrouded in uncertainty. The appropriate response is not to discard the theoretical foundations of the evaluation process as these form a valuable structure for the organisation of thought surrounding the problem.

The most useful approach might consist, for example, of a formal solution to that part of the total control problem for which inputs can be quantified with acceptable reliability; together with some scenario testing in the areas that are not so quantifiable; and perhaps some subjective assessment beyond that.

Within such a partial solution the purpose of the theoretical framework is the maximisation of the scope for formal solution, together with the provision of a structure to assist in bringing order to contemplation of the part of the problem less amenable to formal solution.

With the optimal control framework of Sections 3.2.1 to 3.2.3 in place, the remainder of the paper will be concerned with identification of the components of the COSO ERM Integrated Framework (Section 2.2) with aspects of the framework. These will be considered in an order that is logical from the standpoint of formulation of the control problem rather than in the order in which they appear in Section 2.2.

4 Event identification

The COSO (2004) description of this is as follows:

> "Internal and external events affecting achievement of an entity's objectives must be identified, distinguishing between risks and opportunities. Opportunities are channeled back to management's strategy or objective-setting processes".

The present section will consider the risks. These comprise the factors that materially affect the loss intensity function $L[t, x(t), u(t, x(t))]$ of (3.15) through the second argument $x(t)$. The effects through the third argument $u(t, x(t))$ will be considered in Section 7.

Events are typically considered according to major risk groups. For example, the "Prudential sourcebook for insurers" provided by the Financial Services Authority (see references) requires consideration of risks according to the following grouping:

- Credit risk;
- Market risk;
- Liquidity risk;
- Insurance risk;
- Operational risk;
- Group risk.

Sub-groups may be identified within each of these risk groups. The result can be a lengthy list of risks. For example, Tripp et al (2004) reproduce the list of just operational risks then adopted by the British Bankers Association (**"BBA"**). This consisted of a 3-tier hierarchy with about 140 entries at the third level.

These lists can be very specific, e.g. a single one of those third level entries was as follows:

Tier 1: People;
Tier 2: Employment law;
Tier 3: Wrongful termination.

With such extensive lists of effects on loss intensity L, there is an evident need for prioritisation in terms of materiality. In formulating L, one would naturally wish to apply greater precision to the inclusion of low-likelihood/high-impact events with a substantial effect on profit and/or capital (e.g. for an insurance company, hurricanes, earthquakes, tsunamis, etc.) than to the Tier 3 example just given.

This is not to trivialise the latter type of risk. **All** identified risks need to be accounted for in the ERM framework. As the framework proposed here is an evaluation framework, all risks need to be included in it. The issue is only the detail of their representation in L.

Usually, in broad terms, the detail of representation would increase with financial materiality. For example, low-likelihood/high-impact events with a substantial effect on profit and/or capital (in the case of an insurance company, hurricanes, earthquakes, tsunamis, etc.) may be represented in considerable detail. More minor risks, such as that of wrongful termination introduced above, may be represented in a simplified manner, say as a compound Poisson process with given frequency of occurrence and distribution of size of loss.

There may be some risks, particularly operational risks, which are of high financial materiality but whose risk processes are known only with such vagueness that they are incapable of representation in the loss function other than in simple terms. An example, again chosen the BBA list mentioned above, might be the following:

Tier 1: Systems;
Tier 2: Systems development and implementation;
Tier 3: Cost/time overruns.

Representation of risks of this type more elaborately than by a compound Poisson process may well be considered over-engineering.

5 Internal environment

The COSO (2004) description of this is as follows:

> "The internal environment encompasses the tone of an organization, and sets the basis for how risk is viewed and addressed by an entity's people, including risk management philosophy and risk appetite, integrity and ethical values, and the environment in which they operate".

Much here is intangible but will be captured by the input-output relations of Figure 3-2. This amounts to a quantification of the loss intensity function $L[t, x(t), u(t, x(t))]$ of (3.15). Section 4 considered the different risks causing loss events $x(t)$. Quantification of L amounts to estimation of the effect of each on the objective function K (see (3.15)).

This effect can be considered as consisting of two components:

- the effect on K of a defined event, i.e. of fixed magnitude, under the relevant risk; and
- stochastic effects, consisting of:
    - the probability that the event in question occurs; and
    - the probabilities of any consequential effects involving other risks.

This decomposes the objective function into its deterministic and stochastic components respectively. The present section will consider the former, and Section 6 will consider the latter.

5.1 Independent risks

There may be some risks whose effect on K may be regarded as simple and discrete. Consider, for example, the effect of a major earthquake on an insurance company in the case where the loss intensity in $L[t, x(t), u(t, x(t))]$ in (3.15) is defined as amount of loss, discounted to time s.

Suppose that the loss to the insurer, net of any loss mitigation controls it may have in place (see Section 7), is $S. Then the contribution to L might be considered simply $S\ exp - \delta(t - s)$, where δ is the (continuous) discount rate, and this independent of all other risks considered within the vector $x(t)$.

A more detailed model of risk might envisage consequential effects of this loss, so that it would not be considered independent of other risks. For example, a sufficiently serious loss might cause liquidity concerns and lead to modification of asset allocation. This would create a link between the natural hazards risks and market risk.

It is evident from this example that the number of individual risks considered independent is likely to depend on the level of sophistication of the risk model. In broad terms, the greater the level of sophistication, the more links between different risks are likely to be recognised.

5.2 Dependency between risks

Recall that the input $x(t)$ was defined as an m-vector in Section 3.1.1. Let $x_i(t)$ denote the i-th component of $x(t)$. The risks $\{x_1, \ldots, x_m\}$ may not be independent. Consider, for example, the two operational risks from the BBA list set out in Table 5-1.

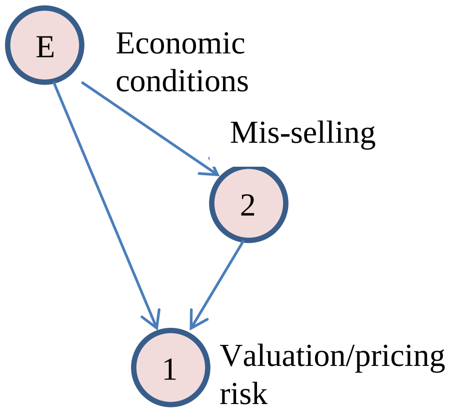
Table 5-1 Dependent operational risks

Risk	Tier 1	Tier 2	Tier 3
1	Process	Valuation/pricing	Model risk
2	People	Unauthorised activity/rogue trading/ employee misdeed	Illegal/aggressive selling tactics

If Risk 2 becomes manifest, there may be pressure to gild the lily in the pricing process, and even in the valuation process (Risk 1). Moreover, there may be a greater likelihood of Risk 2 becoming manifest in the time of an investment bubble, creating a three-way dependency between Risks 1 and 2 and economic conditions (market risk).

Recognition of dependency between risks requires some kind of map or diagram indicating the causality links between them. For example, the situation just described in connection with Table 5-1 would be represented by Figure 5-1.

Figure 5-1 Diagrammatic representation of dependency causality



These diagrams form the basis of causal models, which will be discussed further in Section 6.3.2. They are likely to be much larger and more complex than this when all risks are considered. As indicated at the end of Section 5.1, their size and complexity is likely to increase with increasing sophistication of the risk model generally.

6 Risk assessment

The COSO (2004) description of this is as follows:

> "Risks are analyzed, considering likelihood and impact, as a basis for determining how they should be managed. Risks are assessed on an inherent and a residual basis".

Put briefly, this consists of formulating the measure $\mathcal{P}$ introduced in Section 3.1.2 as associated with the risks defined in Section 4.

6.1 Inherent and residual risk

Inherent risk is that risk that resides in an organisation before the implementation of any risk controls. **Residual risk** is that risk that resides in the organisation after the implementation of risk controls.

The measure $\mathcal{P}$ can be viewed as taking one form, say $\mathcal{P}_{inherent}$, in the absence of any risk controls, but undergoing modification to a different measure $\mathcal{P}_{residual}$, in some sense less risky, after implementation of those controls. The present section will be concerned with the construction of $\mathcal{P}_{inherent}$. The conversion to $\mathcal{P}_{residual}$ will be discussed further in Section 7.

6.2 Likelihood and impact

As in Section 5.2, let the input $x(t) = [x_1(t), \dots, x_m(t)]$. It was assumed in Section 3.2.1 that each component is a financial quantity. Thus $x(t)$ may be characterised by an m-dimensional distribution function $F_X(x) = F_X(x_1, \dots, x_m)$.

It may be natural to construct $F_X(.)$ from the d.f.'s of individual risks (components of x). The d.f. of the r-th risk will be denoted by $F_{X_r}(x_r) = F_X(\infty, \ldots, \infty, x_r, \infty, \ldots, \infty)$.

A comment on this notation. Recall that $X(t)$ was defined by (3.3) as the accumulation of information $x(s)$ for s up to time t. For the present section, it will be convenient to let $X(t)$ denote the random variable whose observed value is $x(t)$. Similarly, for the present section, X_r will denote the random variable whose observed value is x_r.

It will usually be useful to consider each risk in terms of **frequency** and **severity**. Thus the cost of the r-th risk in the period of interest may be expressed as

$$X_r = \sum_{i=1}^{N_r} S_{ri} \tag{5.1}$$

where N_r is the frequency of losses from the r-th risk over the period and S_{ri} is the severity (i.e. amount) of the i-th of those losses.

The random variables N_r, S_{ri} will be characterised by d.f.'s F_{N_r}, F_{S_r}, assuming that the $S_{ri}, i = 1,2, \ldots$ are identically distributed and independent of N_r. In this case

$$F_{X_r}(x_r) = F_{S_r}^{N_r*} \tag{5.2}$$

where F^{n*} denotes the n-fold convolution of d.f. F.

Obvious examples of (5.2) are:

- $N_r \sim Poisson(\lambda_r)$, in which case X_r is compound Poisson; and
- $N_r \sim Binomial(1, \lambda_r)$, in which case X_r is compound binomial.

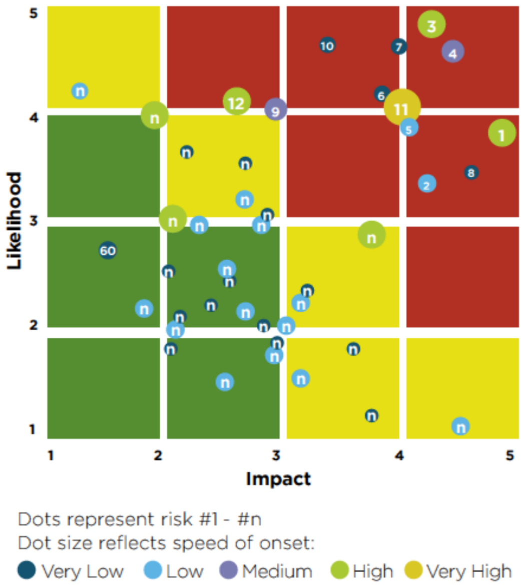
In each case $E[N_r] = \lambda_r$, referred to as the **expected frequency** of losses from the r-th risk. If μ_r denotes $E[S_{ri}]$, referred to as the **expected severity** of losses from the r-th risk, then

$$E[X_r] = \lambda_r \mu_r \tag{5.3}$$

More complex probability structures are possible, allowing for example for dependency between frequency and severity. However, for the estimation of the parameters involved in such structures, data requirements may often far exceed availability.

In ERM expected frequency is usually referred to as **likelihood**, and expected severity as **impact**, and they are represented in **likelihood-impact plots**, or **risk matrices**, with the appearance of Figure 6-1, which has been reproduced from Curtis & Carey (2012). Such diagrams evidently contain only expected values and no distributional information.

Figure 6-1 Likelihood-impact plot



ID	Risk	I	L	V	S
1	Supply chain disruption	4.8	3.7	3.8	4
2	Customer preference shift	4.1	3.3	3.5	2
3	Copper price rise >10%	4.3	4.7	2.3	4
4	Work stoppage > 1 week	4.4	4.5	4.1	3
5	Economic downturn	4.0	3.7	3.5	2
6	Supplier consolidation	3.8	4.2	3.2	1
7	Local competitors enter	3.9	4.5	3.6	1
8	New substitutes available	4.5	3.6	4.2	1
9	Cost of capital rise >5%	2.9	4.0	2.9	3
10	Tighter emission standards	3.4	4.6	2.9	1
11	FCPA violation	4.0	4.0	3.3	5
12	Exchange rate fluctuations	2.7	4.1	2.7	4
n	...	...	...	...	...
60	Impairment of assets	1.6	2.7	1.6	1

I = Impact L = Likelihood V = Vulnerability S = Speed of onset

The impacts of some risks may persist over time. For example, an event that damages reputation might cause a reduction in sales not only at the point of occurrence but for years subsequently. For such a risk event occurring at time t, the impact on sales revenue at time u $(\geq t)$ would be $S_{ri}(t)h(u-t)$, where $S_{ri}(t)$ is the initial severity and $h(u-t)$ is some decreasing function with $h(0)=1$, e.g. $h(u-t)=exp-(u-t)$.

6.3 Dependency between risks

The risks $\{X_1,\dots,X_m\}$ discussed in Section 6.2 may not be independent. An example of dependency was given at the commencement of Section 5.2.

In principle, quantification of the stochastic properties of $\{X_1,\dots,X_m\}$ requires an estimate of the joint d.f. $F_X(x_1,\dots,x_m)$. As in Section 6.2, it may be desirable to consider this distribution in terms of frequency and severity. In the case where these are independent, one then requires estimates of the joint d.f.'s $F_N(n_1,\dots,n_m)$ and $F_S(s_1,\dots,s_m)$ where n_r, s_r denote the frequency and severity respectively of the r-th risk.

These d.f.'s are complex structures and their construction may present difficulties. It is unlikely that the introduction of estimates of correlations will in itself be helpful. Risk management will usually be concerned with the occurrence of infrequent events, i.e. with the tails of the distributions of frequency and severity. The (Pearson) correlation coefficient is adapted to normal distributions and may serve poorly as a measure of dependency in the tails of the frequency and severity distributions. This is discussed further in Section 6.4.2.

6.3.1 Copulas

A copula may provide a more useful conversion of the set of marginal d.f.'s F_{N_r} into a joint d.f. F_N; and similarly marginal d.f.'s F_{X_r} into a joint d.f. F_X. A **copula** is a mapping $C: [0,1]^m \to [0,1]$ with the following properties (McNeil, Frey & Embrechts, 2005):

(a) $C(z)$ is non-decreasing in each component of z;
(b) $C(1, \dots 1, z_r, 1, \dots, 1) = z_r$ for each r (reproduction of marginals);
(c) C satisfies the **rectangle inequality**:

$$\sum_{j_1=1}^{2} \dots \sum_{j_m=1}^{2} (-1)^{j_1+\dots+j_m} C\left(z_{1,j_1}, \dots, z_{m,j_m}\right) \geq 0$$

where $z_{r,j_1} \leq z_{r,j_2}$ for each r.

The copula is used to construct the joint d.f. F_N as follows:

$$F_N(n_1, \dots, n_m) = C\left(F_{N_1}(n_1), \dots, F_{N_m}(n_m)\right) \tag{5.4}$$

where F_{N_r} is the frequency d.f. for the r-th risk. This $F_N(.)$ may be combined with (5.2) to yield $F_X(.)$ In the case of independence between each risk's frequency and severity.

When the joint d.f. is constructed from marginals in this way, it is a simple matter to incorporate long tailed marginals as required. Further, the copula may be chosen to reflect a chosen degree of tail dependence (see e.g. McNeil, Frey & Embrechts (2005, Sections 5.2 and 5.3)). This will be discussed further in Section 6.4.2.

It may be reasonable to assume independence between some of the marginals to which a copula is applied. The copula then factorises thus:

$$C(z_1, \dots, z_m) = C_1(z_1, \dots, z_s) C_2(z_{s+1}, \dots, z_m) \tag{5.5}$$

for s-dimensional and $(m - s)$-dimensional sub-copulas C_1 and C_2, when the sets of variates $\{Z_1, \dots, Z_s\}$ and $\{Z_{s+1}, \dots, Z_m\}$ are stochastically independent of each other.

6.3.2 Causal models

Another alternative might be the causal model foreshadowed in Section 5.2. This is the type of model in which each event is linked probabilistically to antecedents, and is discussed at some length by Corrigan & Luraschi (2013).

For a technical definition, consider a **directed graph** Γ, i.e. a collection of **nodes**, with **edges**, each connecting a pair of nodes, where each edge has a direction. For present purposes, Γ is required to be **acyclic**, i.e. there is no cycle of directed edges leading from a particular node ν back to ν.

A simple example of an acyclic directed graph appeared in Figure 5-1, in which node 1 can be influenced by either of nodes E or 2, but the occurrence of 2 is also influenced by E.

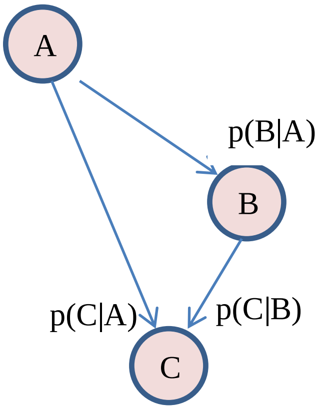
When the nodes of Γ are random variables, possibly latent, and the edges represent conditional dependencies between them, the graph becomes a **Bayesian network**. This is illustrated in Figure 6-2, which is a re-labelled form of Figure 5-1 in which:

- A denotes economic conditions (bubble conditions or not?);
- B denotes selling tactics (mis-selling or not?);
- C pricing aggression (under-pricing or not?).

Dependencies of the form $p(E|F)$ have been added, such that

$$Prob[C] = Prob[A]\{p(C|A) + p(B|A))p(C|B)\} \tag{5.6}$$

Figure 6-2 Example of a Bayesian network



6.4 Quantifiability of risks

6.4.1 Dependency structure

The estimation of $F_X(.)$ discussed in Sections 6.2 and 6.3 is evidently a high-dimensional problem. Little data may be available in respect of some risks, particularly operational risks. Questions may arise as to whether a meaningful estimate of $F_X(.)$, with all its inbuilt dependencies, can be obtained. Certainly, a structured approach to its estimation is desirable.

It will be useful to partition the full set of risks $\{X_1, \dots, X_m\}$ into subsets $\mathcal{S}_1, \dots, \mathcal{S}_q$ that may be reasonably assumed stochastically independent one from another. In graphic terms, the graph Γ of Section 6.3.2 decomposes into **disconnected sub-graphs** $\Gamma_1, \dots, \Gamma_q$, each Γ_j associated with $\mathcal{S}_j$. That is, $\Gamma = \Gamma_1 \cup \dots \cup \Gamma_q$ where, for any $v_j \in \Gamma_j$ and $v_k \in \Gamma_k$, there is no edge between v_j and v_k.

If a copula is used to describe dependency, the copula for $\{X_1, \dots, X_m\}$ may be modularised by an extension of (5.5):

$$C(z_1, \dots, z_m) = \prod_{j=1}^{q} C_j(\mathcal{S}_j) \tag{5.7}$$

For the purpose of parameterisation of the risk model, the subsets $\mathcal{S}_1, \dots, \mathcal{S}_q$ may each be considered in isolation.

6.4.2 Parameterisation

The use of copulas demands estimation of the associated marginals. For causal models, on the other hand, it is in their nature that some marginals will be implied by the conditional probabilities on which they depend. Even in the case of causal models, however, there will be certain root nodes for which marginal distributions are required.

Thus, parameterisation needs to be carried out in terms of:

- Marginal distributions; and
- Dependencies;
  - copulas; or
  - Bayesian network conditional probabilities.

Marginal distributions
One might commence with recognition of various broad groupings of risks, $\mathcal{G}_1, \dots \mathcal{G}_\mathrm{g}$, such as set out early in Section 4. Some of these may be capable of detailed modelling, viz.

- Credit risk;
- Market risk;
- Liquidity risk;
- Insurance risk.

For example, the interest rate component of market risk might be modelled by a detailed financial model such as that of Heath, Jarrow & Morton (1992). The insurance risk associated with technical liabilities might be estimated by a form of bootstrapping (e.g. Taylor, 2000).

The other risk groups mentioned in Section 4 (operational and group risks) may be more nebulous, but nonetheless extensive in scope and material in their financial implications. Selection of their marginals is therefore likely to rest on educated guesswork. Compound Poisson or compound binomial distributions are likely to play a prominent role.

Guesswork as to the severity distributions may be assisted by selecting from a low-dimensional family of suitably conservative distributions. An example would be the **generalised Pareto distribution** (Johnson, Kotz & Balakrishnan, 1994), whose d.f. takes the form

$$F_S(s;\mu,\sigma,\xi) = 1 - \left(1 + \frac{\xi(s-\mu)}{\sigma}\right)^{-1/\xi}, s > \mu \; for \; \xi \geq 0; \mu < s < \mu - \frac{\sigma}{\xi} for \; \xi < 0 \tag{5.8}$$

where μ, σ, ξ are respectively location, dispersion and shape parameters.

Copulas
To a large extent the modification of inherent to residual risk is concerned with removing the more extreme tail from the risk business's loss distribution. Thus, while a copula will

create a form of dependency between a set of marginal, greatest concern will lie with the **tail dependence** between them.

The formal definition of the **coefficient of upper tail dependence** between variates X_1, X_2 with d.f.'s F_1, F_2 respectively is (McNeil, Frey & Embrechts, 2005)

$$\lambda_u = \lim_{q \uparrow 1} Prob[X_1 > F_1^{-1}(q) | X_2 > F_2^{-1}(q)] \tag{5.9}$$

provided the limit exists, and where F^{-1} is understood here as the **generalised inverse** of F, i.e. $F^{-1}(q) = inf\{x: F(x) \geq q\}$.

If the marginals to which a copula is applied are long tailed, then the multivariate distribution generated by the copula is also long tailed. However, inappropriate choice of copula can generate inappropriate tail dependence. For example, $\lambda_u = 0$ (upper tails asymptotically independent) for the Gaussian copula, which may not be a property desired of a risk model.

An alternative for which λ_u is not necessarily zero is the **t-copula**, defined as (McNeil, Frey & Embrechts, 2005):

$$C_{\nu,\mathrm{P}}(u_1, \dots, u_m) = t_{\nu,\mathrm{P}}\left(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_m)\right) \tag{5.10}$$

where t_ν is the standard univariate t d.f. with ν degrees of freedom and $t_{\nu,\mathrm{P}}$ is the joint d.f. of a standardised multivariate t distribution with correlation matrix P.

In this case the parameterisation of the copula, once the number of degrees of freedom has been selected, is just as for the Gaussian case, selection of a correlation matrix. This similarity notwithstanding, however, the tail dependence of the two copulas may be very different.

Care will be required in the selection of the correlation matrix. It must be internally consistent, i.e. positive definite, and so correlations cannot be selected arbitrarily.

Bayesian network conditional probabilities

There is little of generality to be said here. One must simply make the best selection of conditional probabilities (see e.g. Figure 6-2) that can be achieved. Often this will be difficult in the face of little or no data but informed guesswork is preferable to avoidance of the problem.

One of the benefits of causal models is that correlation matrices need not be estimated. Dependencies are parameterised at the fundamental level of conditional probabilities and correlations, if ever required, are derivatives of these.

6.4.3 Stress testing

A stress test may be defined as follows. Consider a subset $\mathcal{S}_j$ of risks, as defined in Section 6.4.1, and let $\bar{\mathcal{S}}_j$ denote its complement, the set of all other risks. With a slight abuse of notation, regard $\mathcal{S}_j(t)$ as a vector of random variables at time t and Let s_j denote some baseline (normative) value of $\mathcal{S}_j(t)$.

Now consider two values of the expected loss intensity function appearing in (3.15), specifically $E_{\bar{\mathcal{S}}_j}\{L[t,x(t),u(t,x(t))]|\mathcal{S}_j(t)=s_j\}$ and $E_{\bar{\mathcal{S}}_j}\{L[t,x(t),u(t,x(t))]|\mathcal{S}_j(t)=s_j+\Delta_j\}$ where Δ_j is some suitable perturbation of $\mathcal{S}_j(t)$.

The stress test then consists of comparison of the two values of conditional expected loss intensity, measuring the effect of the perturbation.

Alternatively, one might examine two values of the objective function (3.15) itself, respectively

$$K\big[s,x|\mathcal{S}_j(t)=s_j\big]=E_{\bar{\mathcal{S}}_j}\left[\int_s^T L[t,x(t),u(t,x(t))]\,dt+\Phi\big[T,y(T)|\mathcal{S}_j(t)=s_j\big]\right] \tag{5.11}$$

and

$$K\big[s,x|\mathcal{S}_j(t)=s_j+\Delta_j\big]=E_{\bar{\mathcal{S}}_j}\left[\int_s^T L[t,x(t),u(t,x(t))]\,dt+\Phi\big[T,y(T)|\mathcal{S}_j(t)=s_j+\Delta_j\big]\right] \tag{5.12}$$

In this case, the stress test consists of comparison of the two values of the objective function.

Stress testing may be interesting as a means of assessing the degree of influence of particular variables, or sets thereof, on the manifestation of risk to the business. It is sometimes used as a last resort alternative to the quantification of the joint d.f. $F_{\mathcal{S}_j}$ of $\mathcal{S}_j$. The usual reason for this would be that the risks $\mathcal{S}_j$ were seen as so nebulous that their d.f. was seen as inestimable.

This may be a tempting course but it has its drawbacks. Chief among these is the standard criticism of scenario testing, that there is no probability associated with each scenario, no means of differentiating mild perturbations of the baseline from the most extreme.

As a consequence, there is likely to be difficulty in integrating stress testing of one set of risks with a genuine quantitative treatment of the others. Hence, in the event of apparent difficulty of estimation of $F_{\mathcal{S}_j}$, it may yet be desirable to associate a simple d.f. with $\mathcal{S}_j$ rather than evade the issue.

7 Risk response and control activities

The COSO (2004) description of **risk response** is as follows:

> "Management selects risk responses – avoiding, accepting, reducing, or sharing risk – developing a set of actions to align risks with the entity's risk tolerances and risk appetite".

The description of **control activities** is as follows:

> "Policies and procedures are established and implemented to help ensure the risk responses are effectively carried out".

Thus risk response is policy; control activities are implementation.

7.1 Nature of risk controls

The business results of a risk business are influenced by a variety of inherent risks $x_j(t)$, as discussed in Sections 4 to 6. Risk controls are actions by management aimed at modifying the distribution of those risks in such a way as to affect the objective function beneficially. The process is illustrated in Figure 3-2, and explained in Section 6.1 as the transformation of the measure $\mathcal{P}_{inherent}$ to $\mathcal{P}_{residual}$.

The objective function in form (3.10) is a statistical expectation, and it is taken with respect to the probability measure $\mathcal{P}$ associated with the business outputs $Y(T) = [Y_1(T), \ldots, Y_n(T)]$. To the extent that the effect of the change of $\mathcal{P} = \mathcal{P}_{inherent}$ to $\mathcal{P} = \mathcal{P}_{residual}$ is to decrease the objective function, the risk controls may be regarded beneficial and may be said to mitigate risk.

Risk controls may be considered in two categories that will be referred to here as:

- Prevention;
- Mitigation.

A **risk prevention control** seeks to reduce the likelihood of occurrence of a risk event or reduce its severity in the event that it does occur. An example might be separation of the pricing and valuation processes, and external peer review of valuations as a means of reducing the likelihood of under-valuation of liabilities in the situation envisaged in Table 5-1 and Figure 5-1.

A **risk mitigation control** seeks to modify the impact on business results of a risk event that has occurred with known severity. An example might be the purchase of excess of loss (**"XoL"**) reinsurance with deductible D. If F_S denotes the d.f. of the severity of a single event, then the modified d.f. in the presence of the XoL risk control is

$$\begin{aligned} F_{S(net)}(x) &= F_S(x), x < D \\ &= 1, \mathrm{x} \geq \mathrm{D} \end{aligned} \tag{6.1}$$

Thus, a risk prevention control seeks to eliminate or ameliorate the risk events themselves; a risk mitigation control does not seek to affect the events themselves but only the extent to which they impinge on business performance.

7.2 Cost of risk controls

Risk controls usually come at a cost, at least a short term cost. In the XoL example above, the expected benefit of the reinsurance to the cedant is $E_{F_S}[X] - E_{F_{S(net)}}[X]$. If the premium payable for the arrangement is P_{XoL}, then the (*a priori*) net cost to the cedant is $c_{XoL} = P_{XoL} - \left[E_{F_S}[X] - E_{F_{S(net)}}[X]\right]$.

Thus, the effect of the reinsurance is twofold:

- It shifts the expected profit of the business downward by an amount c_{XoL};
- It shortens the negative tail of the distribution of profit in accordance with (6.1).

Despite the downward shift in expected profit, the total effect of the arrangement on the objective function (3.10) may be favourable. This is particularly likely in the case of objective functions with longer time horizons. In the above reinsurance example, the effect of the reinsurance is to enhance the insurer's survival probability, thus increasing the expected value of profits over future years.

This may be illustrated by reference to (3.10). Suppose that this objective function is purely profit oriented, i.e.

$$L[t, y(t)] = -(exp\ -rt)S(t)\pi(y(t)) \tag{6.1}$$

(recall that L measures losses) where $S(t)$ is the probability of survival of the insurer to time t, $\pi(y(t))$ is the profit accruing per unit time from business results $y(t)$, and r is a rate of discount of profits.

Suppose further that

$$\Phi[T, y(T)] = -(exp\ -rT)S(T)\Pi(y(T)) \tag{6.1}$$

where $\Pi(y(T))$ is an estimate of future discounted profit at time T when business results are then $y(T)$.

Suppose that these results are based on the situation in which there is no reinsurance. If the above XoL reinsurance is now effected, the objective function will be increased by c_{XoL} at time zero, but $L[t, y(t)]$, $\Phi[T, y(T)]$ will shift by $\Delta L[t, y(t)], \Delta\Phi[T, y(T)]$ respectively.

The total change in objective function is then

$$\Delta E_{\mathcal{P}}\big[J[y]\big] = c + E_{\mathcal{P}}\left[\int_0^T \Delta L[t, y(t)]\, dt + \Delta\Phi[T, y(T)]\right] \tag{6.1}$$

If this is negative, then the reinsurance has improved the financial outlook of the insurer.

The kind of compromise between risk mitigation and its cost evident in the above example is typical of risk controls. It is necessary that both costs and benefits be recognised in the formulation of (3.10). Indeed, it may be reasonably argued that a voluntary risk control (there may also be involuntary risk controls, such as regulatory requirements) that does not improve the objective function requires serious consideration.

8 Objective setting

8.1 General discussion

The COSO (2004) description of **objective setting** is as follows:

> "Objectives must exist before management can identify potential events affecting their achievement. Enterprise risk management ensures that management has in place a process to set objectives and that the chosen objectives support and align with the entity's mission and are consistent with its risk appetite".

This is central to the risk management process as it involves a statement of the core of the problem to be solved. In mathematical terms, it consists of formulation of the objective function of the control problem.

The control problem was enunciated as (3.13), minimisation of the objective function (3.10), or its alternative form (3.15). Implementation requires a choice of the loss intensity function $L[t, x(t), u(t, x(t))]$, as briefly discussed in Section 3.2.2. The selection of risk controls that optimise the objective function is likely to be subject to constraints, as discussed in Section 3.2.3.

Formulation of L requires articulation of the risk business's objectives, as required by the COSO statement. It might, for example, seek to maximise discounted profit up to a defined time horizon; or it might seek to maximise share price at the horizon; etc.

The objectives may be circumscribed by constraints imposed internally or externally. External constraints might be regulatory, such as minimum capital requirements. Internal constraints may seek to limit volatility and disruption that could occur in the absence of those constraints. For example, limits may be placed on rates of growth or contraction of the business.

It is impossible to generalise too far in the discussion of a business's objective function and constraints as these will depend on the combination of management views and the impositions made by the external environment. However, the following formulation might be regarded as unremarkable for an insurance company:

Objective function

Maximise discounted expected profit over the interval $[0, \mathrm{T}]$ where T years is the time horizon; subject to the following constraints:

Constraints

$$\mathrm{Prob}[\textit{technical insolvency in } [0,\mathrm{T}]] < \mathrm{p}_1 \tag{7.1}$$

$$\mathrm{Prob}[\textit{revenue account in loss in any one year} \text{ in } [0,\mathrm{T}]] < \mathrm{p}_2 \tag{7.2}$$

$$\mathrm{Prob}[\textit{net cost of a single natural event} > k_1 \text{ in } [0,\mathrm{T}]] < \mathrm{p}_3 \tag{7.3}$$

$$\mathrm{Prob}\left[\textit{reduction in scale of operations} > k_2\% \textit{ in any one year} \text{ in } [0,\mathrm{T}]\right] < \mathrm{p}_4 \tag{7.4}$$

where the p_i, k_i are discretionary parameters.

Constraint (7.3) is concerned with reinsurance. Constraint (7.4) prevents practices, such as over-aggressive pricing, that might be profitable but highly disruptive operationally, particularly near the time horizon when future profitability has little weight.

Note that:

- The objective function, taken in isolation, encourages risk taking;
- Constraints (7.1)-(7.4) encourage risk mitigation.

A risk business's formulation of its control problem will often have a specific property in common with the above. It is that the objective function and constraints, taken in conjunction, provide a balance between risk taking behaviour and risk mitigation.

8.2 Risk appetite

A typical definition of risk appetite is given by Chapman (2006) as "the amount of risk a business is prepared to tolerate".

In quantitative terms this will be described by the risk mitigation requirements contained in the statement of the control problem, i.e. within the objective function and/or the constraints, most commonly the latter. In the example given in Section 8.1, the insurer's risk appetite is described by constraints (7.1)-(7.4). These are no more nor less than a statement of "the amount of risk a business is prepared to tolerate".

9 Monitoring

The COSO (2004) description of **monitoring** is as follows:

> "The entirety of enterprise risk management is monitored and modifications made as necessary. Monitoring is accomplished through ongoing management activities, separate evaluations, or both".

Since the present paper is intended to be quantitative, interest will be focused on the "evaluations". Each evaluation will be concerned with some aspect of performance of the risk business. A target, or set of targets, is established and data collected on the business's ongoing performance to test whether or not those targets are met.

In many cases the targets will simply be prescribed management actions and the monitoring will consist of completion of checklists. For example, Section 7.1 contemplated external peer review as one means of reducing the likelihood of under-valuation of liabilities in the situation envisaged in Table 5-1 and Figure 5-1. Either this occurs or it does not, and this may be monitored by means of a checklist.

Monitoring may be more demanding wherever a parametric model is involved. For example, the monitoring of insurance risk is likely to include comparison of claims experience, separately by line of business (and possibly sub-line) with targets that consist of expected frequencies and severities according to pricing models.

This type of monitoring is considered in detail by Taylor (2011), whose approach follows several major principles:

- Each parameter in the model under test generates one monitoring table.
- Monitoring is stochastic, i.e. each time it compares an observation with a target it gives the probability of a deviation from target as large as observed.
- This is achieved by the formulation of each comparison in terms of statistical hypothesis testing.

Table 9-1 illustrates hypothetically the monitoring of claim frequency against the variate, Driver Age, in the Motor line of business. "Statistical significance" of a ratio of value r in the table is defined as

$$Prob\left\{\left|\frac{Observed}{Target}-1\right|>r\middle|E\left[\frac{Observed}{Target}\right]=100\%\right\}$$

Table 9-1 Example of stochastic monitoring of claim frequency

Driver age	Claim frequency			
	Observed	**Target**	**Ratio: Observed/Target**	**Statistical significance**
	%	%	%	%
Under 25	17.6	18.4	96	21
25-30	13.0	11.9	109	6
30-40	11.0	10.1	109	2
40-50	11.4	10.8	106	4
50-65	11.0	10.5	105	8
Over 65	14.2	12.3	115	3
Total	**12.0**	**11.2**	**107**	**1**

Similar monitoring may be performed with respect to any other parametric model involved in the definition of the control problem, e.g. market risk, credit risk, etc.

10 Information and communication

The COSO (2004) description of **information and communication** is as follows:

> "Relevant information is identified, captured, and communicated in a form and timeframe that enable people to carry out their responsibilities. Effective communication also occurs in a broader sense, flowing down, across, and up the entity".

This aspect of risk management will not be discussed here. This is not to minimise its value or importance in any way. However, the present paper is concerned with the formulation of an ERM exercise as a well-posed mathematical problem. Whereas the elements of the COSO framework discussed in preceding sections (Sections 4 to 9) were all concerned with this formulation, the present section is concerned with the communication of it and its results within the business.

11 Conclusion

This paper has been sought to formulate the application of ERM to a risk business as a rigorous mathematical statement. This has been achieved by the interpretation of the application as a problem in stochastic optimal control theory.

This formulation has been set against the background of the COSO ERM integrated framework. All elements of that framework other than "information and communication" are found to be interpretable within the optimal control setting.

This has the benefit in some places of replacing erstwhile verbal definitions with precise quantitative statements. Just one example would be that of "risk appetite" (Section 8.2).

However, the main characteristic of the control theory formulation is that it integrates the organisations risk controls with its business objectives in a quantitative manner such that the effect of the former on the latter can be identified.

The business objectives are likely to be concerned largely with the generation of profit, which will usually require risk taking. The risk controls are concerned with risk mitigation and in that sense may constrain action directed toward achieving the business objectives. The formulation suggested here provides a disciplined structure within which the tension of these opposing forces may be examined and possibly resolved.

It should be said that this formulation of the problem does not establish a risk management framework. In COSO language, it does not prescribe the nature of the risk controls and their settings. The requirement remains for a risk officer to identify the risks to the business (Section 4) and the nature of the controls required (Section 7).

What the control theory formulation does do is to establish a formalised link between those risk controls and performance of the risk business in their presence. In doing so, it may assist with the settings to be selected for the controls by quantifying the interplay between them and business objectives.

A sanguine view of the risk management process would be that it proceeds exactly as implied by the formulation set out in Section 3.2. One selects an objective function (a representation of business objectives, as modified by some risk controls), selects constraints (additional risk controls), and then carries out optimisation of the objective function, resulting in forecasts of future optimised performance.

A world in which this occurred would certainly be a simple, neatly ordered and pure one. In practice, however, one might find that the forecast optimum performance appears poor, that one has mitigated risk to excess, and at the expense of future profitability.

It would then become necessary to embark on a sequence of experimental iterations in which the degree of risk mitigation was diminished and the consequences of the changes for performance noted. It would be necessary to continue this process until a reasonable balance between risk control and performance appeared to have been attained.

Parameterisation of the model established for optimal control will often present substantial difficulty, as discussed in Sections 6.4.2 and 6.4.3. This can be disheartening, and raise questions as to the meaningfulness of the exercise.

Nonetheless, a formal and disciplined model structure, parameterised by informed guesswork, with its limitations fully recognised, may be preferable to mere evasion of the issue. The structure recommended here is a stochastic one, and so provides a vehicle for the expression of uncertainty. If necessary, it can be expanded to a random effects model in which parameters are explicitly declared random drawings from some hyper-distribution.

One of the benefits of the formal structure is that it will permit investigation of the sensitivity of performance to variations in dubious parameters. The doubt one feels about the value chosen for a particular parameter might be partially assuaged if the influence of that parameter on ultimate outcomes can be shown to be limited.

There may be a tendency toward conservatism in the selection of parameters. This may be healthy in small doses. In larger measures, however, the accumulation of conservative margins may seriously compromise the model.

As noted earlier in this section, it is likely that risk controls would be varied in the model until the attainment of a suitable balance with forecast future performance. Accordingly, excessive conservatism in one area of the model is likely to result in compensation elsewhere.

This may lead to perverse results. For example, excessively conservative parameterisation of say the less tangible risks (e.g. fraud) may lead to weak mitigation of the more tangible (e.g. reinsurance). In this case, apparent conservatism in model parameterisation might actually result in increased risk.

References

Arnold L (1974). **Stochastic differential equations: theory and applications**. John Wiley & Sons, New York NY.

Chapman R J (2006). **Simple tools and techniques for enterprise risk management**. John Wiley & Sons Inc, New Jersey USA.

Corrigan J & Luraschi P (2013). **Operational risk modelling framework**. Milliman research report, Seattle USA.

COSO (2004). **Enterprise Risk Management -- Integrated Framework**. American Institute of Certified Public Accountants, NYC, USA.

Curtis P & Carey M (2012). **Risk assessment in practice**. Committee of Sponsoring Organizations of the Treadway Commission.

Financial Services Authority. **Prudential sourcebook for insurers**. At http://fsahandbook.info/FSA/html/handbook/INSPRU

Heath D, Jarrow R & Morton A (1992). Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation. **Econometrica**, 60(1), 77-105.

Johnson N L, Kotz S & Balakrishnan N (1994). **Continuous univariate distributions, Volume 1, 2nd ed**. John Wiley & Sons Inc, New York NY.

Kaufmann R, Gadner A & Klett R (2001). Introduction to dynamic financial analysis. **Astin Bulletin**, 31(1), 213-249.

McNeil A J, Frey R & Embrechts P (2005). **Quantitative risk management: concepts, techniques and tools**. Princeton University Press, Princeton NJ.

Orros GC & Smith J (2012). Enterprise risk management for health insurance from an actuarial perspective. **British Actuarial Journal**, 17(2), 259-314.

Speedy C B, Brown R F & Goodwin G C (1964). **Control theory: identification and optimal control**. Oliver & Boyd, Edinburgh UK.

Taylor G (2000). **Loss reserving: an actuarial perspective**. Kluwer Academic Publishers, Boston.

Taylor G (2011). A statistical basis for claims experience monitoring. **North American Actuarial Journal**, 15(4), 535-552.

Tripp M H, Bradley H L, Devitt R, Orros G C, Overton G L, Pryor LM & Shaw R A (2004). Quantifying operational risk in general insurance companies. **British Actuarial Journal**, 10(5), 919-1012.