CONTROL AND FINANCIAL ENGINEERING

J.M. SCHUMACHER*

Abstract. The paper provides a review of the basics of financial engineering, with a few examples. We emphasize connections with control theory in a broad sense rather than with stochastic control theory in particular, and the reader is not assumed to be versed in stochastic processes. After a discussion of the main methods of financial risk management, a state-space framework for modeling financial markets is presented and used to explain crucial concepts of financial engineering such as absence of arbitrage, market completeness, hedging, and the Black-Scholes partial differential equation. Two brief case studies are presented: the construction of an indexed bond, and the hedging of long-term contracts for delivery of oil.

Key words. Risk management, Black-Scholes theory, hedging, complete markets, indexed bonds, oil futures.

AMS(MOS) subject classifications. 91B28, 93B50, 60H30

1. Introduction. Financial engineering (in the sense of risk management) is concerned with protecting economic activity against the adverse effects of financial risk factors such as movements in exchange rates, interest rates, and commodity prices. It has emerged in the past decades as a model-based discipline that provides rules for decisions to be taken on the basis of incoming observations. As such, financial engineering can be considered a neighboring field of control engineering. It will be one of the aims of this paper to highlight similarities both in purpose and in method between the two fields.

Another aim is to provide some tempering of the occasionally strong feelings, both positive and negative, that are provoked by the term "finance." Although there have been cases where the allure of the new financial theory has been used to deceive credulous bank managers, financial engineering cannot be compared with the countless schemes of playing the stock market that have been proposed. The basis of financial engineering is the insight that financial institutions, because of the access they have to large and liquid financial markets, are able to absorb certain risks for their customers. They can do this by balancing positions in such a way that sensitivity to the relevant risk factors is eliminated or at least substantially reduced. Key to this is the ability to determine the sensitivity of a given position to the risk factors that apply to it. In general it may be noted that risk is a major impediment to economic activity, so that there is a real economic benefit in methods that allow the transfer of the effects of uncertainties to parties that are able to implement effective risk-reducing strategies.

* Department of Econometrics and Operations Research, Tilburg University, P. O. Box 90153, 5000 LE Tilburg, the Netherlands.

Uncertainty is central to finance, and therefore it is no surprise that financial engineering studies are almost invariably based on on stochastic models. In large parts of control theory on the other hand, noise is ignored on the basis of the assumption that a control structure with good robustness properties in a deterministic setting will also behave well when noise is added, so that it doesn't seem necessary to carry the weight of explicitly introducing stochastic terms in the modeling phase. Of course, the classical field of stochastic control theory is alive and well and—not unexpectedly—has many applications in finance, which are reviewed elsewhere in this volume. This paper however is addressed to control theorists who are not necessarily au fait with stochastic processes. Although some of the models used in finance do require close familiarity with, for instance, the subtleties of Lévy processes, the essentials of financial engineering can be understood on the basis of a rudimentary knowledge of stochastic calculus rules. This paper includes a brief motivation of the stochastic calculus.

The paper is organized as follows. We begin in the next section with a discussion in general terms of financial risk management and the modeling of financial markets. Section 3 provides the most basic stochastic calculus rules together with some motivation. A state-space framework for the modeling of financial markets, akin to a standard model of control theory, is presented in Section 4 along with some specific examples of well-known models in finance. We emphasize similarities to the setting of control theory but we also discuss the specific features of financial models. Two specific examples of financial engineering are discussed in the following two sections. In Section 5, we use a model of bond markets together with inflation to investigate the possibility of constructing an inflation-indexed bond out of nominal bonds, using the fact that nominal bond prices are inflation-sensitive. Section 6 is concerned with the hedging of long-term contracts for oil delivery by actively trading short-term contracts. This is a famous case in finance due to the failure of an attempt by a US subsidiary of the German firm Metallgesellschaft in the early nineties to set up a scheme of this type, which resulted in a loss of well over a billion dollars for the mother company. Some concluding remarks are in Section 7.

2. Financial risk management. Financial risk arises due to fluctuations in markets (exchange rates, interest rates, energy prices, stock prices, and so on) as well as from general economic factors such as inflation and from other sources such as political developments. Our point of departure in this paper is not to try to predict such factors but rather to find ways of protection against any adverse movements that may take place. We shall concentrate on the effects of variables such as commodity prices or inflation rates that are measured in terms of real numbers; so we will not discuss the effects of political upheavals. Also, the models to be used in this paper will always assume that the relevant variables follow continuous paths, so that even small jumps do not occur. Continuous-time models in which real-valued variables follow continuous paths are typical for financial engineering, although substantial work has also been done on discrete-time models and on models that do allow jumps.

Two important methods of reducing risk are **diversification** and **hedging**. These two ideas are intertwined and it would be difficult to make a sharp distinction, but nevertheless one may say that while diversification focuses on the joint characteristics of assets, hedging concentrates on the risk factors behind asset prices. Both methods may be developed in a **static** as well as in a **dynamic** setting. Of course the dynamic settings are more close to control theory; however, to illustrate the two notions, let us consider some simple static examples.

First, let us consider a typical diversification scheme. Suppose that capital can be invested in two assets, one of which is more risky than the other, and suppose we try to minimize uncertainty. Is it then optimal to invest all capital in the safest asset? To make this problem precise, suppose that the values of the two assets at the end of the investment period (say, X_1 and X_2) are jointly normally distributed stochastic variables. Let the expected values of X_1 and X_2 be μ_1 and μ_2 respectively, and let their variances be σ_1^2 and σ_2^2, with $\sigma_1 < \sigma_2$; so the second asset is the more risky one. Finally, let the correlation coefficient of X_1 and X_2 be ρ. In this very simple setting, the investment problem may be formulated as the problem of finding two real numbers α_1 and α_2, with $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, and $\alpha_1 + \alpha_2 = 1$, such that the variance of $\alpha_1 X_1 + \alpha_2 X_2$ is minimized. This is a quadratic optimization problem that can readily be solved. It turns out that the optimal solution is to invest everything in the safest asset if

$$\rho \geq \frac{\sigma_1}{\sigma_2}.$$

If the above inequality is not satisfied, then the minimum is reached for a nontrivial combination of the two assets. In this case there is a **diversification benefit** from investing in both assets. Note that such a benefit is always present when the two assets are negatively correlated, and even, as long as σ_1 is positive, when they are uncorrelated. To illustrate the size of the effect in a numerical example, assume $\sigma_1 = 1$, $\sigma_2 = 2$, and $\rho = 0$. With these parameter values, the optimal solution is to invest 80% in the safe asset and 20% in the risky asset. The variance obtained with this allocation is $0.8^2 \cdot 1 + 0.2^2 \cdot 4 = 0.64 + 0.16 = 0.8$. If everything would have been invested in the safest asset, the variance of the portfolio value would have been 1.

There are many ways to take this simple story further, such as: consider more than two asset categories; do not only take variance into account, but also expected value; allow non-normal distributions and use risk measures other than variance; consider multiple periods or even continuous-time models, and allow reallocation through time; drop the restriction that portfolio weights should be nonnegative, so allow “short” positions; take liabilities into account; consider the model parameters such as expected values and correlations as uncertain rather than as exact, and take this into account in the optimization (robustness).

Now, for a simple example of "hedging", consider a pension fund that holds a bond portfolio to cover its future liabilities. Let r be the annual interest rate (assumed to be the same for all maturities). By a standard formula, the current value of the expected payments to be made is

$$P = \sum_{k=1}^{K} (1+r)^{-k} P_k$$

whereas the current value of the bond portfolio is

$$B = \sum_{k=1}^{K} (1+r)^{-k} B_k.$$

Here, of course, P_k is the sum of the anticipated pension payments in year k, whereas B_k is the nominal amount received from the bond portfolio in year k. We would like to have at least $B = P$. Note however that both the current value of assets B and the current value of liabilities P are sensitive to the interest rate r; so if the interest rate changes, the position of the pension fund will be affected. The effect may be positive or negative, but from a risk minimization point of view the trustees of the fund may prefer a situation in which the sensitivity of the fund's position to changes in the interest rate is minimized. This can be achieved by taking $B_k = P_k$ for all k; however, such a specific portfolio prescription may not be practical. To achieve first-order immunity, it is sufficient to have $dB/dr = dP/dr$, which is expressed by saying that the two portfolios should be duration matched. The terminology comes from the fact that the equality

$$\frac{1}{C}\frac{dC}{dr} = -\frac{\sum_{k=1}^{K} k(1+r)^{-(k+1)} C_k}{\sum_{k=1}^{K} (1+r)^{-k} C_k}$$

expresses the relative derivative with respect to r of the current value C of a series of cash flows $\{C_k\}$, up to a factor $-1/(1+r)$, as a weighted average of the times of payment.

This example, even if mathematically trivial, illustrates that hedging aims at obtaining insensitivity to certain risk factors by finding suitable combinations of assets that are all influenced by the same risk factors. This is a model-based activity because the dependence of asset values on risk factors needs to be modeled. For instance, it was assumed above that the interest rate is the same for all maturities. More advanced hedging schemes use models for the entire "term structure" of interest rates in which each maturity has its own interest rate associated to it. Hedging becomes even more powerful in a dynamic context, and the succes of the Black-Scholes model in finance is based in large part on its (idealized) assumption that hedging positions may in fact be adapted continuously in time.

A general picture of the structure of financial risk management is shown in Fig. 2. Observations typically include current asset prices; in-

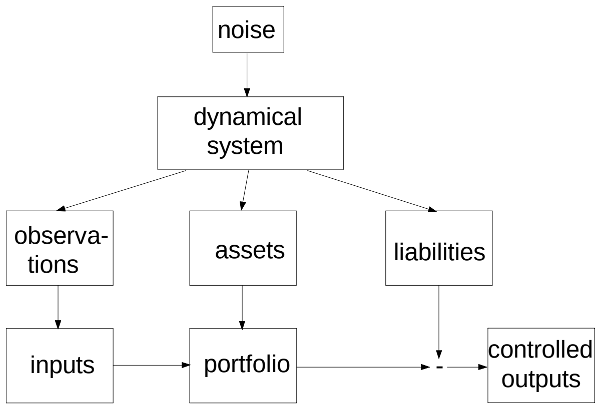

Fig. 1. *Structure of a model for financial risk management*

puts usually are portfolio weights. The controlled output is often a net value.

Financial models have a number of peculiar properties, which distinguish them from general models of control theory. A simple rule says that the value of a portfolio depends linearly on the portfolio weights. In control-theoretic terms, this property relates to the effects of inputs (portfolio weights) on controlled outputs. The rule must be satisfied in all cases where the basis assets are freely traded, because, if a portfolio would be worth more or less than the sum of the values of its parts, a simple profit could be made by either buying the parts and selling the portfolio or vice versa.[1] In control theory, one finds such cases where linearity is a matter of principle rather than of convenience also in certain physical systems, as a result of the laws of nature (for instance, conservation of mass).

A second important principle may be formulated as follows: no control strategy can produce a noise-free positive net value from a zero initial investment. This is an economic principle known as **absence of arbitrage**. It leads to a constraint on the way that asset prices depend on state variables; this will be discussed in more detail below.

There is also a certain constraint on input functions that needs to be imposed. This is most easily explained in discrete time. Let Y_t denote a vector[2] of asset prices and let u_t denote a vector of corresponding portfolio

[1]This is an example of an *arbitrage argument*; the validity (to a high extent) of such arguments is a keystone of mathematical reasoning in finance.

[2]The scalar case is trivial: if there is only a single asset, no portfolio changes can

weights. The control $\{u_t\}$ is said to be self-financing if for all t:

$$u_{t+1} \cdot Y_{t+1} = u_t \cdot Y_{t+1} \tag{2.1}$$

where the dot denotes vector product. This entails a single linear constraint on the vector of control inputs at each time t. One can rewrite the above using the forward difference operator:

$$\Delta(u_t \cdot Y_t) = u_t \cdot (\Delta Y_t).$$

This suggests the continuous-time version

$$d(u_t \cdot Y_t) = u_t \cdot dY_t \tag{2.2}$$

where d is an "infinitesimal forward difference"—in a sense to be discussed below.

One more specific feature of financial models concerns the way in which the output (net portfolio value) is connected to the inputs (portfolio weights). Let again Y_t denote a vector of asset prices, and let u_t be a vector of corresponding portfolio weights. The value of the portfolio is given by

$$V_t := u_t \cdot Y_t.$$

In discrete time, and under the condition (2.1), the change in portfolio value between time t and time $t + 1$ is given by

$$V_{t+1} - V_t = u_t \cdot (Y_{t+1} - Y_t)$$

or in Δ notation

$$\Delta V_t = u_t \cdot \Delta Y_t.$$

So the portfolio value at time T is given by

$$V_T = V_0 + \sum_{t=0}^{T-1} u_t \cdot \Delta Y_t. \tag{2.3}$$

This suggests that an analogous formula in continuous time should read

$$V_T = V_0 + \int_0^T u_t \cdot dY_t \tag{2.4}$$

but this needs to be interpreted with care, in order to give due weight to the observation that asset prices are highly irregular. The following section presents some modifications to the usual calculus rules which serve to express the effects of irregularity.

3. An ultrabrief introduction to stochastic calculus. We think of the vector of asset prices Y_t as dependent upon state variables and time, say $Y_t = \pi(t, X_t)$, where $\pi(t, x)$ is a smooth vector-valued function on (a subset of) R^{1+n}.[3] State equations will be used to provide the dynamics of

take place.

[3]The letter π is used here as a mnemonic for "price."

the state variable X_t. To write the controlled output $V_t = \int_0^t u_\tau \cdot dY_\tau$ as a function of inputs and states, it would be natural to use the chain rule

$$dY_t = \frac{\partial \pi}{\partial t} dt + \frac{\partial \pi}{\partial x} dX_t. \tag{3.1}$$

Consider in general the relation $y(t) = \phi(x(t))$. According to the usual chain rule, we can compute $y(t)$ approximately by

$$y(t) \approx y(0) + \sum \phi'(x(t_i))(x(t_{i+1}) - x(t_i)). \tag{3.2}$$

We evaluate the function $\phi'(x(\cdot))$ here at the left end of the interval $[t_i, t_{i+1}]$ rather than at any other point, because the intended application is to sums of the form (2.3) where it is essential that evaluation of the integrand takes place at the beginning of the time increment. Under this rule, the formula (3.2) may not provide a good approximation if $x(\cdot)$ is a highly irregular function of time. This is illustrated in Fig. 2, where the first panel shows a continuous but fairly irregular function, and the last panel shows the approximation error when the square of the given function is approximated by a sum of first-order differences on the grid with step size 1, following the formula (3.2). Actually, in this case where the function ϕ in (3.2) is

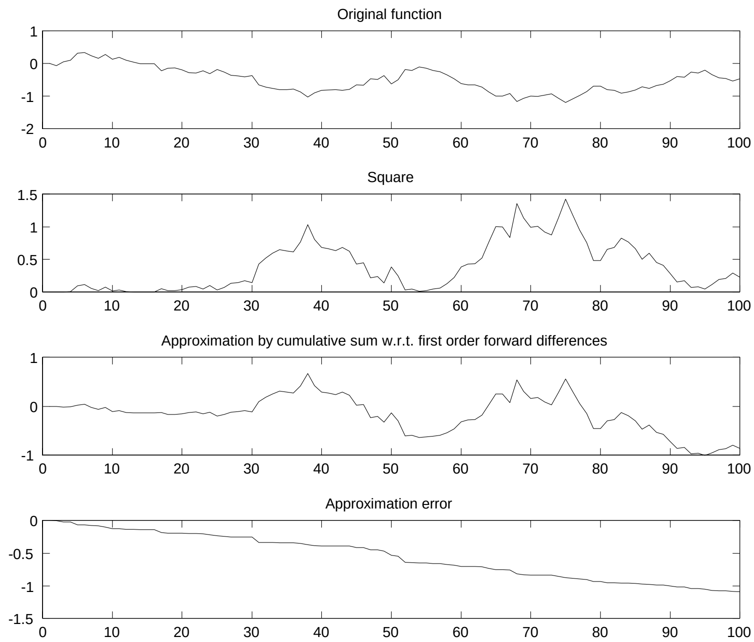

Fig. 2. *The need for the Itô correction term*

given by $\phi(x) = x^2$, the approximation error is easy to write down: the approximation is

$$\tilde{y}(t_j) = y(0) + \sum_{i=0}^{j-1} 2x(t_i)(x(t_{i+1}) - x(t_i))$$

whereas the exact formula is

$$y(t_j) = y(0) + \sum_{i=0}^{j-1} (x^2(t_{i+1}) - x^2(t_i))$$

so that the approximation error is given by

$$y(t) - \tilde{y}(t_j) = -\sum_{i=0}^{j-1} (x(t_{i+1}) - x(t_i))^2. \tag{3.3}$$

Asymptotically, if the function $x(t)$ has finite total variation, the sum at the right hand side of (3.3) must converge to zero when it is computed over finer and finer grids. However, if the function $x(t)$ remains irregular, in the sense that the sum at the right hand side of (3.3) is not small, at scales at which the approximation is computed (in the finance application: time scales corresponding to trading frequencies), then a second-order term should be included to get a good approximation.[4] Taking this down to infinitesimal scales, one is led to a new form of the chain rule: if $y(t) = \phi(x(t))$, then

$$dy = \phi'(x)\,dx + \tfrac{1}{2}\,\phi''(x)\,d[x,x] \tag{3.4}$$

where the term "$d[x,x]$" is an infinitesimal version of

$$\Delta[x,x](t) := (x(t + \Delta t) - x(t))^2.$$

We will also need a vector version:

$$dy = \frac{\partial \phi}{\partial x}(x)\,dx + \tfrac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2 \phi}{\partial x^2}(x)\,d[x,x]\right) \tag{3.5}$$

where now $d[x,x]$ is a matrix with entries of the form $d[x_i, x_j]$.

It is convenient to use functions that are "highly irregular up to any scale," since these make it unnecessary to use scales explicitly in arguments. On the other hand, it does require a rather extensive mathematical framework to create a supply of such functions. The standard method is to use a probabilistic construction. To get a function that is irregular on a scale characterized by time step Δt, draw an independent random variable with distribution $N(0, \sigma^2 \Delta t)$ (normal with expectation 0 and variance $\sigma^2 \Delta t$, where σ is a constant) for each time step, form the cumulative sums,

[4] Whether the second-order approximation is good enough for a particular application is an empirical matter. For applications in finance it seems to work well.

and use linear interpolation to get a continuous function, say on an interval $[0, T]$.[5] The sum of the squared Δt-increments of this function on the interval $[0, T]$ will be approximately equal to $\sigma^2 T$ (the sum's expected value), and as the construction is repeated with smaller and smaller Δt, the approximation becomes better as a result of the law of large numbers. Consequently, repeating the exercise in Fig. 2 at smaller time scales will make the curve in the bottom panel look more and more like a straight line. In the limit, one arrives at what is called Brownian motion.

The development above is manifestly heuristic; for a complete account, see for instance [6]. The key property that we will be concerned with is that the infinitesimal second-order correction term corresponding to a Brownian motion W_t as constructed above (with $\sigma = 1$, i.e. a "standard" Brownian motion) is given by

$$d[W_t, W_t] \;=\; dt. \tag{3.6}$$

So, for instance (applying (3.4) with $\phi(x) = x^2$ and $x = W$):

$$d(W_t^2) = 2W_t dW_t + dt \tag{3.7}$$

so that we can write

$$W_t^2 = W_0^2 + 2\int_0^t W_s dW_s + t. \tag{3.8}$$

The paths of Brownian motions are "highly irregular functions" in a sense that has turned out to be useful in finance. By construction the collection of paths comes with a probability structure, so that we have a stochastic process rather than just a collection of functions. Moreover, one can form new processes with specific dynamic characteristics by using the Brownian motion as an "input" in stochastic differential equations of the type

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t. \tag{3.9}$$

Here, W_t denotes a vector of mutually independent standard Brownian motions, μ is a smooth vector-valued function, and σ is a smooth matrix-valued function; both μ and σ are defined on an appropriate domain of R^n. The equation (3.9) is understood in the sense of

$$X_t = X_0 + \int_0^t \mu(X_s)\, ds + \int_0^t \sigma(X_s)\, dW_s$$

where the rightmost integral is defined analytically (rather than formally as in (3.8)) as the limit in an appropriate sense[6] of sums of the form

[5] As the reader may have surmised, this is the way in which the function in the top panel of Fig. 2 was obtained.

[6] See for instance [8] for much more detailed information.

$\sum \sigma(t_i, X(t_i))(W(t_{i+1}) - W(t_i))$. Solutions of equations of the above form provide a rich source of processes that are useful in financial modeling.

An important theoretical use of the probabilistic structure is the following. Let f be a function defined on the vector space R^n, and let X be a vector-valued process given by the stochastic differential equation (3.9). One can then define for each $T \geq 0$ a new function $E_T^X f$ on R^n defined by

$$(E_T^X f)(x) = E[f(X_T) \mid X_0 = x]. \tag{3.10}$$

A straightforward way to compute an approximation of the value of this function at a given point $x \in \mathsf{R}^n$ it to compute a large number of discrete-time approximations of the sample paths of (3.9), note the value of f at the point reached at time T, and average the results. Clearly the operation will have a diffusive effect on the function that one starts with. Such behavior might also be described by means of a partial differential equation. For the process $Y_t := f(X_t)$ with X_t given by (3.9), the Itô formula (3.5) can be written in the form

$$dY = \left(\frac{\partial f}{\partial x}\mu + \tfrac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2 f}{\partial x^2}\sigma\sigma^\top \right) \right) dt + \frac{\partial f}{\partial x}\sigma dW$$

which suggests that for the function $\phi(t, x)$ defined by

$$\phi(t, x) = E[f(X_t) \mid X_0 = x] \tag{3.11}$$

one should have

$$\frac{\partial \phi}{\partial t}(t, x) = \frac{\partial \phi}{\partial x}(t, x)\mu(x) + \tfrac{1}{2}\,\mathrm{tr}\left(\frac{\partial^2 \phi}{\partial x^2}(t, x)\sigma(x)\sigma^\top(x) \right). \tag{3.12}$$

This can be indeed shown to hold; it is a classical relation between conditional expectations and partial differential equations. The stochastic representation (3.11) of solutions of PDEs of the form (3.12) that is obtained in this way is useful both analytically and numerically (Monte Carlo method).

4. Financial models.

4.1. State equations. Our basic model will be the following:

$$dX_t = \mu_X(X_t)dt + \sigma_X(X_t)dW_t \tag{4.1a}$$
$$Y_t = \pi_Y(t, X_t). \tag{4.1b}$$

In this model, all stochasticity derives from the driving Brownian motion process W_t, and we will consider the above equation typically on an interval $[0, T]$. The state process X_t takes values in an open subset $\mathcal{D}$ of R^n. The symbols μ_X and σ_X denote functions from $\mathcal{D}$ to R^n and to $\mathsf{R}^{n\times k}$, respectively. They should be such that the stochastic differential equation (4.1a) has a unique solution in the domain $\mathcal{D}$ given an initial condition X_0; these properties hold under fairly mild conditions (see for instance [7]). Prices of assets in the market are collected in the m-vector Y_t, which is related to the state vector X_t by the function $\pi_Y(t,x)$ from $[0,T]\times\mathcal{D}$ to R^m. It will be a standing assumption that $\pi_Y(t,x)\neq 0\in\mathsf{R}^m$ for all $(t,x)\in[0,T]\times\mathcal{D}$; in other words, it cannot happen that all asset prices are zero simultaneously.

On the basis of Itô's rule, we can write

$$dY_t \;=\; \mu_Y(t,X_t)dt+\sigma_Y(t,X_t)dW_t \tag{4.2}$$

where the functions $\mu_Y(t,x)$ and $\sigma_Y(t,x)$ can be expressed in terms of the data in (4.1). Specifically, each component Y_ℓ of the vector Y has associated to it a "drift function" $\mu_\ell(t,x)$ and a "volatility function" $\sigma_\ell(t,x)$ which are given by

$$\mu_\ell \;=\; \frac{\partial\pi_\ell}{\partial t}+\frac{\partial\pi_\ell}{\partial x}\mu_X+\tfrac{1}{2}\mathrm{tr}\,\frac{\partial^2\pi_\ell}{\partial x^2}\sigma_X\sigma_X^\top \tag{4.3a}$$

$$\sigma_\ell \;=\; \frac{\partial\pi_\ell}{\partial x}\sigma_X. \tag{4.3b}$$

Here, the arguments t and x have been suppressed to alleviate the notation; this will also often been done below. The gradient vector $\partial\pi_\ell/\partial x$ is defined as a row vector, and superscript T denotes transposition. The formula in (4.3a) may be written more explicitly as

$$\mu_\ell \;=\; \frac{\partial\pi_\ell}{\partial t}+\sum_{i=1}^{n}\frac{\partial\pi_\ell}{\partial x_i}(\mu_X)_i+\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2\pi_\ell}{\partial x_i\partial x_j}(\sigma_X\sigma_X^\top)_{ij}.$$

To simplify notation further, not only arguments of functions but also subscripts t for stochastic processes will be omitted, so that for instance (4.2) may be written as $dY=\mu_Y dt+\sigma_Y dW$ where it is understood that Y and W should have subscripts t and that μ_Y and σ_Y should be evaluated at (t,X_t).

It will be a standing assumption that all assets considered are "pure" assets, net of all costs and dividends. Moreover, it will be assumed that trading can take place continuously without transactions costs, that there is no restriction in taking short positions, and that lending and borrowing take place at the same interest rate. These are strong idealizations and it is certainly worthwile to investigate what happens when these assumptions are not met; however, by making these assumptions one is able to create a powerful theory which often serves well in practice as a first approximation.

4.2. Basic properties.

4.2.1. Conditions for absence of arbitrage. We want our model (4.1) to be such that it allows no arbitrage; that is, the model will be an equilibrium model in the sense that we assume that any occurring arbitrage opportunities have already been eliminated by market forces. The intuitive notion of absence of arbitrage has been formulated above as the nonexistence of any trading strategies that produce a riskless profit. One of the possible ways of formalizing this notion for the continuous-time model (4.1) is the following.

Definition 4.1. The model (4.1) is said to *admit no arbitrage* if there exists a scalar-valued function $\rho(t,x)$ such that for all $(t,x) \in \mathcal{D}$ the following holds: for all $\phi \in \mathrm{R}^m$ such that $\phi^\top \sigma_Y(t,x) = 0$,

$$\phi^\top \mu_Y(t,x) \;=\; \rho(t,x)\phi^\top \pi_Y(t,x). \tag{4.4}$$

The interpretation of the function $\rho(t,x)$ is that it represents the rate of return (at time t and in state x) on a riskless asset. Under absence of arbitrage there can be at most one such function, because if there would be two riskless assets with different returns, then an easy arbitrage opportunity would exist (borrow from one and lend to the other). The above definition extends this idea to combinations of risky assets that are constructed in such a way that, at least at the point (t,x), all risk is eliminated. The definition may be rephrased as: "any instantaneously riskless combination of assets generates the riskless return." Note that, indeed, the left hand side of (4.4) represents the absolute growth rate of the portfolio $\phi^\top Y$ while the right hand side is the riskless interest rate times the value of the same portfolio.

The above notion of absence of arbitrage was used in the original paper by Black and Scholes [1]. Actually this notion may be considered somewhat debatable, since it only refers to an arbitrage opportunity that exists at a single moment in time; one may wonder whether such short-lived opportunities can really be exploited. Fortunately it has turned out that if the above assumption is not satisfied it is in fact possible to devise strategies that operate on a nontrivial time interval and that still bring a riskless profit. The notion of absence of arbitrage can indeed be defined in various ways. One reformulation that is particularly useful in the state-space context is the following.

Theorem 4.1. The model (4.1) admits no arbitrage if and only if there exist a k-vector valued function $\lambda(t,x)$ and a scalar function $\rho(t,x)$ such that

$$\mu_Y - \rho\pi_Y = \sigma_Y\lambda. \tag{4.5}$$

Proof. If the relation (4.5) holds, then absence of arbitrage follows immediately from Def. 4.1. Conversely, let $\rho(t,x)$ be as in Def 4.1. Suppose that for a certain t and x the vector $\mu_Y(t,x) - \rho(t,x)\pi_Y(t,x)$ would not be in the range space of the matrix $\sigma_Y(t,x)$; then there would be an m-vector ϕ such that $\phi^\top(\mu_Y(t,x) - \rho(t,x)\pi_Y(t,x)) = 1$ while $\phi^\top\sigma_Y(t,x) = 0$, so that our model would not be arbitrage-free according to Def. 4.1. Therefore we can conclude that for all (t,x) there must be a k-vector $\lambda(t,x)$ such that $\mu_Y(t,x) - \rho(t,x)\pi_Y(t,x) = \sigma_Y(t,x)\lambda(t,x)$. □

Given the interpretation of ρ as the riskless interest rate, the left hand side of equation (4.5) can be viewed as a vector of excess returns. Correspondingly, the components of the vector λ can be viewed as market prices of risk which express the excess return, depending on t and x, that is required by the market per unit of volatility as represented by the components of the driving process W_t.

4.2.2. Market completeness. Theorem 4.1 states that absence of arbitrage holds in the model (4.1) if there exist functions ρ and λ that satisfy (4.5). In general there may be several such functions. If the equation (4.5) has a unique solution (ρ, λ), then the model (4.1) is said to constitute a complete market. Under the standing assumption that $\pi_Y(t,x) \neq 0 \in \mathsf{R}^m$ for all $(t,x) \in [0,T] \times \mathcal{D}$, it is equivalent to require only that the market price of risk λ is determined uniquely by (4.5). Another equivalent condition is that the matrix $[\sigma_Y(t,x) \quad \pi_Y(t,x)] \in \mathsf{R}^{m\times(k+1)}$ has full column rank for all (t,x). Obviously a necessary condition for this to happen is that the number of assets m exceeds the number of driving Brownian motions k. As will be seen below, market completeness is related to the problem of eliminating the risk associated with any liability that may be formulated in terms of the model (4.1).

4.2.3. Pricing. Suppose that the market described by (4.1) is free of arbitrage, and let us consider a fixed solution (λ, ρ) of (4.5). If C denotes an asset whose price can be expressed as a function of the time t and the state x in our model, then this function, which we denote by $\pi_C(t,x)$, must satisfy the equation

$$\mu_C - \rho\pi_C = \sigma_C\lambda \tag{4.6}$$

where μ_C and σ_C are determined from π_C as in (4.3). This is (a general form of) the Black-Scholes equation. More explicitly, by Itô's rule the equation (4.6) can be written as

$$\frac{\partial \pi_C}{\partial t} + \frac{\partial \pi_C}{\partial x}(\mu_X - \sigma_X\lambda) + \tfrac{1}{2}\mathrm{tr}\,\frac{\partial^2 \pi_C}{\partial x^2}\sigma_X\sigma_X^\top \; = \; \rho\pi_C. \tag{4.7}$$

In the frequently occurring situation in which the claim C concerns a contract that expires at a given time T and the value at time T is given by a formula in terms of the state variables, the above PDE may be used to compute its value at times preceding T.

Clearly there is a close similarity between equations (4.7) and (3.12), although there are some extra terms in (4.7), and there is a time reversal in the sense that $T-t$ rather than t is the relevant time variable for (4.7). In fact, a straightforward calculation shows that if one adds a new state K_t with initial value $K_0 = 1$ and dynamics given by

$$dK = -K(\rho dt + \lambda^\top dW) \tag{4.8}$$

and a new output function defined by

$$\tilde{\pi}_C(t,x,k) = k\pi_C(t,x) \tag{4.9}$$

then the equation analogous to (3.12) for the new output is exactly (4.7), up to a change of sign for the time derivative which is due to the time reversal mentioned above. Consequently, we can write

$$\pi_C(0,x) = E[K_T C_T \mid X_0 = x,\ K_0 = 1]. \tag{4.10}$$

The variable K_t is referred to as the pricing kernel.

4.2.4. Replication. Suppose that we have an arbitrage-free and complete market described by (4.1); so $[\sigma_Y \ \ \pi_Y]$ has full column rank, and we have

$$\mu_Y \ = \ [\sigma_Y \ \ \pi_Y]\begin{bmatrix}\lambda\\ \rho\end{bmatrix}.$$

Now introduce a new asset C with pricing function $\pi_C(t,x)$. Because the matrix $[\sigma_Y(t,x) \ \ \pi_Y(t,x)]$ has full column rank for all $(t,x) \in \mathcal{D}$, it follows from linear algebra that there exists an vector-valued function $\phi(t,x)$ such that

$$[\sigma_C(t,x) \ \ \pi_C(t,x)] \ = \ \phi^\top(t,x)[\sigma_Y(t,x) \ \ \pi_Y(t,x)] \tag{4.11}$$

Under absence of arbitrage, we have

$$\mu_C = [\sigma_C \ \ \pi_C]\begin{bmatrix}\lambda\\ \rho\end{bmatrix} = \phi^\top[\sigma_Y \ \ \pi_Y]\begin{bmatrix}\lambda\\ \rho\end{bmatrix} = \phi^\top\mu_Y. \tag{4.12}$$

Let V_t denote the value of the portfolio that is formed from the assets in Y by using the portfolio weights ϕ. Then

$$V_t = \phi^\top(t,X_t)\pi_Y(t,X_t) = \pi_C(t,X_t) = C_t$$

so the portfolio with weights ϕ "replicates" the asset C. Moreover, we have

$$\begin{aligned} dV \ &= \ dC \ = \ \mu_C dt + \sigma_C dW \ = \\ &= \ \phi^\top(\mu_Y dt + \sigma_Y dW) \ = \ \phi^\top dY \end{aligned}$$

which shows that the portfolio with weights ϕ is self-financing (see (2.2)).

4.3. Examples.

4.3.1. The Black-Scholes model. Consider the following model which has one state variable, one uncertainty factor, and two assets denoted by S and B:

$$dX_t \ = \ \mu X_t dt + \sigma X_t dW_t \tag{4.13a}$$

$$S_t \ = \ X_t \tag{4.13b}$$

$$B_t \ = \ ce^{rt}. \tag{4.13c}$$

Here, μ, σ, and c are constants; both σ and c are supposed to be nonzero. The domain of the variable x is taken to be the set of positive numbers. In terms of the notation used earlier we have $\mathcal{D} = \{x \in \mathsf{R} \mid x > 0\}$, $n = 1$, $k = 1$, $m = 2$, and

$$\mu_X(t,x) = \mu x, \quad \sigma_X(t,x) = \sigma x, \quad Y = \begin{bmatrix} S \\ B \end{bmatrix}, \quad \pi_Y(t,x) = \begin{bmatrix} x \\ ce^{rt} \end{bmatrix}.$$

Furthermore, one easily computes

$$\mu_Y(t,x) \;=\; \begin{bmatrix} \mu x \\ cre^{rt} \end{bmatrix}, \quad \sigma_Y(t,x) \;=\; \begin{bmatrix} \sigma x \\ 0 \end{bmatrix}.$$

In particular, the matrix $[\sigma_Y \;\; \pi_Y]$ is given by

$$[\sigma_Y(t,x) \;\; \pi_Y(t,x)] \;=\; \begin{bmatrix} \sigma x & x \\ 0 & ce^{rt} \end{bmatrix}. \tag{4.14}$$

Clearly, the matrix that appears here is invertible for all $t \geq 0$ and $x > 0$, and so the equation (4.5) has a unique solution which is given by

$$\rho = r, \quad \lambda = \frac{\mu - r}{\sigma}.$$

The pricing equation for an asset C that is added to this model becomes

$$\frac{\partial \pi_C}{\partial t} + rx\frac{\partial \pi_C}{\partial x} + \tfrac{1}{2}\sigma^2 x^2 \frac{\partial^2 \pi_C}{\partial x^2} \;=\; r\pi_C. \tag{4.15}$$

This is the standard Black-Scholes equation. In the market defined by (4.13), the equation holds for any asset C whose value can be expressed in terms of the variables t and x. The standard example is of course the European call option with maturity T and strike price a, which satisfies $\pi_C(T,x) = \max(x - a, 0)$. Solving the PDE (4.15) with this boundary condition leads to the Black-Scholes formula for a call option.

In general, if sufficient boundary conditions for a derivative are supplied to determine the price function uniquely, a replication scheme can be set up as follows. The required portfolio weights are obtained from (4.11), which in the case at hand becomes

$$[\frac{\partial \pi_C}{\partial x}\sigma x \;\; \pi_C] \;=\; [\phi_S \;\; \phi_B] \begin{bmatrix} \sigma x & x \\ 0 & ce^{rt} \end{bmatrix}. \tag{4.16}$$

In particular, it follows that

$$\phi_S \;=\; \frac{\partial \pi_C}{\partial x}. \tag{4.17}$$

The component ϕ_B is then determined by the requirement that the portfolio should be self-financing, and it can be computed explicitly from (4.16).

4.3.2. An Asian option. The model of the previous subsection may have to be extended for other types of options, in particular for path-dependent options. As an example, consider a continuously sampled Asian option. Since such an option depends on the integral of the asset price over a certain period, a state variable has to be added to the model which will provide the information needed to determine the option price at expiry. The extended model may be written down as follows:

$$dX_1 = \mu X_1 dt + \sigma X_1 dW \quad (4.18a)$$
$$dX_2 = X_1 dt \quad (4.18b)$$
$$S = X_1 \quad (4.18c)$$
$$B_t = ce^{rt}. \quad (4.18d)$$

Eqn. (4.18b) implies that

$$(X_2)_T = \int_0^T (X_1)_t \, dt$$

if $(X_2)_0$ is taken to be zero. In terms of the model above, the Asian option with time of expiry T and strike a is defined by

$$\pi_C(T, x) = \max(x_2 - a, 0). \quad (4.19)$$

In our new model, we have the following data (letting Y consist of S and B as before): $\mathcal{D} = \{x \in \mathsf{R}^2 \mid x_1 > 0\}$, $n = 2$, $k = 1$, $m = 2$, and

$$\mu_X(t,x) = \begin{bmatrix} \mu x_1 \\ x_1 \end{bmatrix}, \quad \sigma_X(t,x) = \begin{bmatrix} \sigma x_1 \\ 0 \end{bmatrix},$$
$$Y = \begin{bmatrix} S \\ B \end{bmatrix}, \quad \pi_Y(t,x) = \begin{bmatrix} x_1 \\ ce^{rt} \end{bmatrix}.$$

We find

$$\mu_Y(t,x) = \begin{bmatrix} \mu x_1 \\ cre^{rt} \end{bmatrix}, \quad [\sigma_Y(t,x) \;\; \pi_Y(t,x)] = \begin{bmatrix} \sigma x_1 & x_1 \\ 0 & ce^{rt} \end{bmatrix} \quad (4.20)$$

and so we obtain $\rho = r$, $\lambda = (\mu - r)/\sigma$ as before. In fact it is easy to show in general that when state variables are added to a model for the purpose of pricing derivatives, and the original model already determined a complete market, then the riskless interest rate ρ and the market price of risk λ will be unchanged.

The partial differential equation for the pricing function of the Asian option C may now be written down from (4.6). After some minor manipulation, we find

$$\frac{\partial \pi_C}{\partial t} + rx_1 \frac{\partial \pi_C}{\partial x_1} + x_1 \frac{\partial \pi_C}{\partial x_2} + \tfrac{1}{2}\sigma^2 x_1^2 \frac{\partial^2 \pi_C}{\partial x_1^2} = r\pi_C. \quad (4.21)$$

This equation holds for any derivative whose price can be determined in terms of x_1 (the price of the underlying asset S), x_2 (the continuous average of the asset price), and t. For the Asian option with expiry T and strike a one should apply the boundary condition (4.19). It is possible to solve the resulting equations analytically, but the solution is quite a bit more complicated than in the case of a standard call option.

4.3.3. The Vasicek model. It is a well-known fact that interest rates depend on the period to which they apply—the ten-year rate is usually not the same as the six-month rate. This fact is expressed by saying that interest rates have a *term structure*. Interest rates are not themselves traded assets, but of course there is a one-to-one connection between the interest rate for period T and the price of a contract that will pay one unit of currency after period T (a *zero-coupon bond*).

A relatively simple model for interest rates may be constructed as follows. Consider the equations

$$dX_1 = a(b - X_1)dt + \sigma dW \quad (4.22a)$$

$$dX_2 = X_1 X_2 dt \quad (4.22b)$$

$$Y = X_2 \quad (4.22c)$$

where a and b are constant parameters. The interpretation of the model is that the variable X_1 represents the "short rate" (the interest rate that applies to loans of very short maturities), and X_2 is the value of the "money market account," which is a portfolio that is continually reinvested in short-maturity bonds. The parameter b represents a mean level of the short rate, and a determines the speed of mean reversion.

The model above does not define a complete market since the price of only one asset is given in terms of the state variables. Nevertheless, the riskless interest rate ρ is already determined; since $\pi_Y(t,x) = x_2$, $\mu_Y(t,x) = x_1 x_2$, and $\sigma_Y(t,x) = 0$, it follows from (4.5) that $\rho(t,x) = x_1$, in line with the interpretation of the model that was just given. One way to make the model complete would be to provide another traded asset together with a formula for its price in terms of t and x, but another way is to simply specify the market price of risk $\lambda(t,x)$. Assume for instance that $\lambda(t,x)$ is in fact constant; this constant λ then becomes a model parameter just like a, b, and σ.

The term structure of interest rates is determined by the prices of bonds of all maturities. There is no guarantee that bond prices can be written in terms of the state variable x_1 in the model above, since bonds are not defined in terms of the short rate. If one nevertheless assumes that the prices of bonds *are* functions of t and x_1 (this essentially comes down to the assumption that the movements of interest rates of all maturities can be described in terms of a model with a one-dimensional state variable), then one obtains what is commonly known as the *Vasicek model* [10] for the term structure of interest rates. To summarize, the Vasicek model consists of the equations (4.22) together with the assumptions that the market price of risk is constant and that bond prices can be written as functions of the time and of the short rate x_1.

Within the Vasicek model one can write down equations for bonds of different maturities. Let $\pi_T(t,x)$ denote the price, at time t and state x, of a bond that pays one unit of currency at time $T \geq t$. The general equation (4.5) becomes in the present case:

$$\frac{\partial \pi_T}{\partial t} + \frac{\partial \pi_T}{\partial x_1}[a(b - x_1) - \lambda\sigma] + \tfrac{1}{2}\frac{\partial^2 \pi_T}{\partial x_1^2}\sigma^2 \;=\; \pi_T x_1 \tag{4.23}$$

with the boundary condition

$$\pi_T(T,x) \;=\; 1. \tag{4.24}$$

This equation can be solved analytically.

5. Constructing an indexed bond.

5.1. A Gaussian market. More generally than (4.22a), one may consider equations of the "affine" form

$$dX = (FX + f)dt + GdW \tag{5.1a}$$

where F and G are constant matrices of sizes $n \times n$ and $n \times k$ respectively, and f is a constant n-vector. Models of this form are often used to describe the dynamic behavior of the term structure of interest rates, together with the assumptions that the riskless instantaneous return ρ and the market price of risk λ that arise in the Black-Scholes equation (4.6) are given by

$$\rho(t,x) = h^\top x, \quad \lambda(t,x) = \lambda \tag{5.1b}$$

where h is a constant vector of length n and λ is a constant vector of length k. So it is assumed that the riskless rate depends linearly on the state variables, and the risk premia are constant. In this section we adopt the affine model.

According to formula (4.10), the price of a zero-coupon bond that pays one unit of currency at time T is given by

$$P(0,T) \;=\; EK_T \tag{5.2}$$

where K_t is the pricing kernel process, which is determined by

$$dK = -K(\rho dt + \lambda^\top dW), \quad K_0 = 1. \tag{5.3}$$

By the Itô formula (3.4), we have

$$d(\log K) = -(\rho + \tfrac{1}{2}\lambda^\top\lambda)dt - \lambda^\top dW. \tag{5.4}$$

Since it has been assumed that λ is constant and that ρ depends linearly on the state variable x, it follows that for any given T the random variable $\log K_T$ follows a normal distribution. Let m_T and v_T denote the mean and the variance, respectively, of $\log K_T$; then, according to a well-known formula,

$$EK_T \;=\; \exp(m_T + \tfrac{1}{2}v_T). \tag{5.5}$$

To compute the quantities m_T and v_T we proceed as follows. Define matrices $\tilde{F}$ and $\tilde{G}$ of sizes $(n+1)\times(n+1)$ and $(n+1)\times k$ respectively, and vectors $\tilde{f}$ and vector $\tilde{h}$ both of length $n+1$, as follows:

$$\begin{aligned} \tilde{F} &= \begin{bmatrix} 0 & -h^\top \\ 0 & F \end{bmatrix}, \quad \tilde{G} = \begin{bmatrix} -\lambda^\top \\ G \end{bmatrix}, \\ \tilde{f} &= \begin{bmatrix} -\frac{1}{2}\lambda^\top\lambda \\ f \end{bmatrix}, \quad \tilde{h} = [\,1\;\;0\;\cdots\;0\,]^\top. \end{aligned} \tag{5.6}$$

Then we can write

$$d\begin{bmatrix} \log K \\ X \end{bmatrix} = \tilde{F}\begin{bmatrix} \log K \\ X \end{bmatrix} dt + \tilde{f}dt + \tilde{G}dW. \tag{5.7}$$

The quantities m_T and v_T are given by standard formulas:

$$m_T \;=\; \tilde{h}^\top \left(e^{\tilde{F}T}\begin{bmatrix} 0 \\ X_0 \end{bmatrix} + \int_0^T e^{\tilde{F}(T-t)}\tilde{f}dt \right) \tag{5.8}$$

$$v_T \;=\; \tilde{h}^\top \left(\int_0^T e^{\tilde{F}(T-t)}\tilde{G}\tilde{G}^\top e^{\tilde{F}^{>}(T-t)}dt \right)\tilde{h}. \tag{5.9}$$

Denoting the pricing function of the bond with maturity T by π_T, we find that this function is of the form

$$\pi_T(0,x) \;=\; \exp(a(T) + b(T)x) \tag{5.10}$$

where the scalar function $a(\cdot)$ and the row vector function $b(\cdot)$ can be inferred from (5.8) and (5.9). More generally we can write

$$\pi_T(t,x) \;=\; \exp(a(T-t) + b(T-t)x). \tag{5.11}$$

5.2. A sufficient condition for completeness. The volatility of the price of the T-bond can be computed from the general formula

$$\sigma_T(t,x) \;=\; \frac{\partial \pi_T}{\partial x}(t,x)\sigma_X(t,x). \tag{5.12}$$

In our present case we have

$$\frac{\partial \pi_T}{\partial x}(t,x) = \pi_T(t,x)b(T-t)$$

and $\sigma_X = G$. Moreover, the function $b(t)$ can be expressed in terms of the original parameters of the model (5.1) by

$$b(t) \;=\; -h^\top \int_0^t e^{Fs} ds. \tag{5.13}$$

Therefore we find

$$\sigma_T(t,x) \;=\; -\pi_T(t,x)h^\top \left(\int_0^{T-t} e^{Fs}\, ds \right) G. \tag{5.14}$$

The given market is complete in terms of bonds if we can find $k+1$ maturities $T_1, \ldots, T_{k+1}$ such that the matrix of size $(k+1)\times(k+1)$ with rows $[\pi_{T_i} \;\; \sigma_{T_i}]$ $(i = 1, \ldots, k+1)$ is invertible. Because of the relation (5.14) and the fact that, for all T, t, and x, the number $\pi_T(t,x)$ is positive (as is seen from (5.11)), the invertibility condition holds if and only if the matrix

$$\begin{bmatrix} 1 & -h^\top \left(\int_0^{T_1 - t} e^{Fs}\, ds \right) G \\ \vdots & \vdots \\ 1 & -h^\top \left(\int_0^{T_{k+1} - t} e^{Fs}\, ds \right) G \end{bmatrix}$$

is invertible for each $t \in [0,T]$. This in turn is true if and only if the matrix above has no nonzero null vectors, that is, if the relation

$$\begin{bmatrix} 1 & -h^\top \left(\int_0^{T_1 - t} e^{Fs}\, ds \right) G \\ \vdots & \vdots \\ 1 & -h^\top \left(\int_0^{T_{k+1} - t} e^{Fs}\, ds \right) G \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_k \end{bmatrix} \;=\; 0 \tag{5.15}$$

(in which $w_0, \ldots, w_k$ are real numbers) implies that $w_0 = \cdots = w_k = 0$. If we write $w := [w_1 \;\cdots\; w_k]^\top$, the above relation can also be written

$$h^\top \left(\int_0^{T_i - t} e^{Fs}\, ds \right) Gw \;=\; w_0 \quad (i = 1, \cdots, k+1). \tag{5.16}$$

This means that the function of time defined by $c(\tau) = h^\top (\int_0^\tau e^{Fs}\, ds)Gw$ takes the same value w_0 at the $k+1$ different instants $T_1 - t, \ldots, T_{k+1} - t$. Since $c(\cdot)$ is a continuously differentiable function, this implies that the derivative $\dot{c}(\tau) = h^\top e^{F\tau}Gw$ is 0 at k different instants. Therefore, a sufficient condition for the market (5.1) to be complete in terms of bonds is that for any set of k different time instants $t_1, \ldots, t_k$ the $k \times k$ matrix

$$M(t_1, \ldots, t_k) := \begin{bmatrix} h^\top e^{Ft_1} \\ \vdots \\ h^\top e^{Ft_k} \end{bmatrix} G \tag{5.17}$$

is invertible.

5.3. A model for inflation. As a specific case of (5.1), consider the following model in which the state space dimension n is 3 and the number of sources of uncertainty k is 2:

$$dX_1 = X_2 dt \tag{5.18a}$$

$$dX_2 = \alpha(\overline{X}_2 - X_2)dt + \sigma_2 dW \tag{5.18b}$$

$$dX_3 = \beta(\overline{X}_3 - X_3)dt + \sigma_3 dW \tag{5.18c}$$

with

$$\rho(t, x) = x_2 + x_3, \quad \lambda(t, x) = \lambda \tag{5.18d}$$

where $\overline{X}_2$, $\overline{X}_3$, α, and β are positive constants, W is a 2-dimensional standard Brownian motion, σ_2 and σ_3 are constant row vectors of length 2, and λ is a constant column vector of the same length. This is a simplified version of a model used by Brennan and Xia [3]. It will be assumed that the vectors σ_2 and σ_3 are independent; otherwise the model could be rewritten in a form in which there is only one driving Brownian motion. The interpretation of the state variables is as follows: X_1 denotes the logarithm of inflation; X_2 is the rate of inflation; and X_3 is the real interest rate (that is, the interest rate after correction for inflation).

One has to verify that these interpretations are consistent with each other and with the definition of ρ. First, if we write $L := \exp(X_1)$ for the inflation index, then

$$dL = (\exp X_1)dX_1 = X_2 L dt \tag{5.19}$$

so that indeed X_2 is the rate of inflation. Moreover, the value of a money market account in nominal terms is equal to $M_t = L_t R_t$ where R_t satisfies $dR = X_3 R dt$; therefore, by Itô's formula (3.5),

$$dM = LdR + RdL = (X_2 + X_3)LRdt = (X_2 + X_3)Mdt \tag{5.20}$$

which confirms the interpretation of M. The model is of the form (5.1) with

$$F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\alpha & 0 \\ 0 & 0 & -\beta \end{bmatrix}, \quad f = \begin{bmatrix} 0 \\ \alpha\overline{X}_2 \\ \beta\overline{X}_3 \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ \sigma_2 \\ \sigma_3 \end{bmatrix}, \quad h = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}. \tag{5.21}$$

The constants $\overline{X}_2$ and $\overline{X}_3$ denote long-term equilibrium levels for inflation rate and real interest rate respectively; the speed at which the actual inflation and interest rates tend to return to these levels is determined by the constants α and β.

Within the model (5.18) one can define a contract that will pay $L_T = \exp(X_1)_T$ units of currency at time T. This product may be called an *indexed bond*; it is a bond that is protected against inflation. Can we replicate such an indexed bond by means of ordinary bonds? This will certainly be the case if our market is complete in terms of bonds. To check this, compute the matrix $M(t_1, t_2)$ as in (5.17):

$$M(t_1,t_2) = \begin{bmatrix} 0 & e^{-\alpha t_1} & e^{-\beta t_1} \\ 0 & e^{-\alpha t_2} & e^{-\beta t_2} \end{bmatrix} \begin{bmatrix} 0 \\ \sigma_2 \\ \sigma_3 \end{bmatrix} = \begin{bmatrix} e^{-\alpha t_1} & e^{-\beta t_1} \\ e^{-\alpha t_2} & e^{-\beta t_2} \end{bmatrix} \begin{bmatrix} \sigma_2 \\ \sigma_3 \end{bmatrix}. \tag{5.22}$$

By assumption, the 2×2 matrix formed from σ_2 and σ_3 is invertible. It is easily verified that the 2×2 matrix of exponentials is invertible if and only if α is not equal to β. So we find that, if $\alpha \neq \beta$, the model (5.1) allows the construction of an indexed bond out of ordinary bonds. One can in fact use any three bonds of different maturities as long as these maturities exceed that of the indexed bond; alternatively, one can use bonds of shorter maturity and replace them by new bonds at the time of expiry ("rolling over").

6. The story of Metallgesellschaft. Metallgesellschaft AG, based in Frankfurt, is a large corporation doing business in metal, mining, and engineering. It has owned a US-based oil business known as MGRM (Metallgesellschaft Refining and Marketing). In 1992, MGRM set up a scheme in which it granted long-term contracts for delivery of oil to customers for a fixed price, covering periods up to ten years. The exposure to oil price risk was hedged by the purchase of short-term contracts ("futures") for which a liquid market exists. Under market conditions of 1993, the strategy required an enormous amount of cash input with no substantial income yet from oil deliveries. MG decided to stop the hedging scheme and wrote off about $1.5 billion.

The Metallgesellschaft case may be viewed as the Tacoma Narrows story of financial engineering.[7] The causes of the failure have been extensively discussed in the literature. Here we provide a brief discussion drawing upon the analysis by W. Bühler *et al.* in a recent paper [4].

[7] In 1940, construction was completed of a suspension bridge across the Narrows near Tacoma, WA, USA. Built according to a new and particularly elegant design, the bridge became known as "galloping Gertie" because of the tendency of the bridge deck to oscillate under the influence of the wind. On November 7, 1940, a strong and steady wind caused the bridge to collapse completely, at a cost of presumably less than $1.5 billion. Footage of the event can be found on the internet.

We start with formulating a model for the most important variable in our situation, namely the oil price. In a model with a single state variable, one may in fact use the oil price itself as the state variable and write the state equation in the form

$$dX_t = \mu_X(X_t)\,dt + \sigma_X(X_t)\,dW_t \tag{6.1}$$

where X_t denotes the oil price. It is natural to let the drift function $\mu_X(x)$ be positive when x is low, and negative when x is high, so as to express the empirical observation of the past decades that the oil price, while subject to continual changes, never strays very far from the historical average level. For instance one might take $\mu_X(x) = a(b-x)$ where b denotes a long-term average and a is a parameter expressing "speed of mean reversion." Under this model however there is a positive probability for the oil price to become negative. The following equation is suggested in [4]:

$$dX_t = \gamma(\Theta - \log X_t)X_t dt + \sigma X_t dW_t \tag{6.2}$$

where γ, Θ, and σ are constants.

Oil may be used for direct use as well as for investment. The Black-Scholes equation has been developed for financial assets rather than for consumption goods, and actually does not apply as such to the latter. A more general formulation which covers both the "investment" and the "consumption" case is the following (r is the interest rate, c represents storage cost per barrel per year):

$$\mu_X(x) - r\pi_X(x) - c \le \sigma_X(x)\lambda(x) \tag{6.3a}$$

$$\lambda(x) \ge 0 \tag{6.3b}$$

where for all x at least one of the inequalities is satisfied with equality. The above equations may be solved to find the market price of the risk of oil as an investment good:

$$\lambda(x) = \max\left(\frac{\mu_X(x) - r\pi_X(x) - c}{\sigma_X(x)},\, 0\right). \tag{6.4}$$

This expression is due to Bühler *et al.* [4], who obtained it from an equilibrium argument.

On the basis of the above model we can now discuss the price of contracts for oil delivery. In general, a *futures* contract with *maturity* T is a contract to deliver a given commodity, for instance one barrel of oil, at time T for a price F_T to be paid when delivery is made. So, in our present context where we assume a fixed interest rate, the market value at time t of a contract to deliver a barrel of oil at time T is $e^{-r(T-t)}F_T$. This value should satisfy the Black-Scholes equation. After some algebra, we find that the Black-Scholes equation for the T-futures price $F_T(x)$ becomes:

$$\mu_T(t,x) = \sigma_T(t,x)\lambda(x) \tag{6.5}$$

where

$$\mu_T = \frac{\partial F_T}{\partial t} + \frac{\partial F_T}{\partial x}\mu_X + \tfrac{1}{2}\,\frac{\partial^2 F_T}{\partial x^2}\,\sigma_X^2 \tag{6.6a}$$

$$\sigma_T = \frac{\partial F_T}{\partial x}\,\sigma_X \tag{6.6b}$$

with the final condition $F_T(T, x) = x$. This may be solved numerically in a straightforward way, for instance using a finite-difference approximation.

Various alternative models for oil futures have been formulated in the literature; see for instance [2, 9]. Depending on the model, one finds different pricing rules. Two alternatives to the rule implicit in (6.4–6.5) are obtained by replacing (6.4) by either

$$\lambda(x) = 0 \tag{6.7}$$

or

$$\lambda(x) = \frac{\mu_X(x) - r\pi_X(x) - c}{\sigma_X(x)}. \tag{6.8}$$

The first option corresponds to considering oil purely as a commodity rather than as an investment good. Under this assumption, the futures price is determined simply by the expected price at expiration of the contract, which can be computed on the basis of the assumed dynamics. One may therefore refer to the price obtained from (6.7) as the **expectation-based** price. On the other hand, assumption (6.8) results from viewing oil purely as an investment good; the futures price of a barrel of oil is determined by the cost of buying the barrel now and storing it until maturity. The cost of storage is easily computed:

$$\int_0^T ce^{-rt}dt \;=\; \frac{c}{r}(1 - e^{-rT})$$

where T is the time of expiration of the futures contract. Therefore the futures price at time t according to the arbitrage-based model is (taking into account that this is a price to be paid at time T)

$$F_T(t, x) = e^{r(T-t)}x + \frac{c}{r}(e^{r(T-t)} - 1). \tag{6.9}$$

One may indeed verify that this solves the partial differential equation obtained by inserting (6.8) and (6.6) into (6.5). The price (6.9) may be referred to as the **arbitrage-based** price. Obviously the pricing rule obtained from (6.4) is sort of a combination of the expectation-based and the arbitrage-based price; this is also evident if prices are computed in a particular case (see Fig. 6).[8] When oil prices are low the futures price is close

[8]The following parameter values were used: $\Theta = \log 20.5$, $\gamma = 2.5$, $\sigma = 0.35$, $r = 0.05$, $c = 4$ (dollars per barrel per year). These values have been suggested in [4].

to the arbitrage-based price, when oil prices are high the expectation-price dominates. Where the transition from one regime to the other takes place depends on the length of the futures contract; the results shown in Fig. 6 are for six-month futures. The rule based on (6.4) will be referred to as the **two-regime** pricing rule. In the concrete case of oil futures, the two-regime

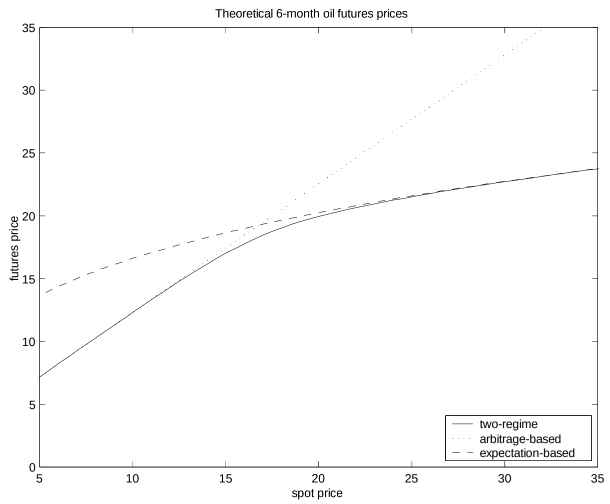

Fig. 3. *Six-month futures price of oil as a function of the price for immediate delivery according to three different models*

rule may be justified as follows. When oil prices are low with respect to the long-term average, the price is expected to rise and so the expectation-based model predicts that futures prices should be high. If that would indeed happen however, then market parties could make a riskless profit by selling futures contracts and storing the required oil until maturity. Therefore in this situation the arbitrage-based pricing rule dominates. If oil prices are higher than average, then according to the expectation-based theory the futures price is relatively low. An upward pressure on prices could be derived from an inversion of the arbitrage scheme used earlier, namely, buy futures contracts and sell stored oil. There is a physical limit to this scheme, however, due to the fact that at any time there is only so much oil that can be delivered immediately; also, companies may prefer to keep some oil in stock in order to meet fluctuations in demand. Therefore the proposed scheme does not push up prices and the futures price follows the expectation-based scheme. Empirical data are in fairly good agreement with the pattern shown by the two-regime rule, although there is certainly not an exact fit; this is indeed not to be expected from a model that is still quite simple.

The idea of the hedging scheme used by MGRM was that the prices of both long-term and short-term futures contracts are both sensitive to the spot price of oil. The value of a portfolio consisting of both short-term and long-term contracts can therefore be made insensitive to changes in the oil price if the two types of contracts are combined in appropriate proportions. The existing liquidity of the market for contracts for delivery of oil in three or six months can in this way be transferred to contracts for delivery in eight or ten years. The scheme developed by MGRM was successful in the sense that many customers were happy to enter into long-term contracts that guaranteed delivery of oil for a fixed price. Once the contracts were agreed upon, MGRM hedged them by short-term futures. Of course the short-term contracts used for hedging expire after some time, and then they need to be "rolled over" (replaced by new contracts). If the futures price is higher than the spot price, which is not usually the case but does happen sometimes, it is expensive to roll over a futures contract. Theoretically the required cash input is counterbalanced by the gain in value of the long-term contracts. It requires a strong belief in the theory however to maintain such a scheme when in the short run it drains the liquid assets of a company. In any case, at the end of 1993 the central board of Metallgesellschaft in Frankfurt decided to stop the cash outflow and to close all contracts at considerably less than the theoretical price. MG lost about 10% of its value in the operation.

Some noted economists blamed the MG board for lack of faith. However, even when in general one has confidence in the principles of financial mathematics, one may still debate which model to choose in a particular situation. The choice of a model may in fact have very significant consequences for the ensuing strategy. In the particular case of oil futures, the "hedge ratio" of a ten-year future (the number of short-term, say six-month, futures that should be bought for every long-term future sold in order to keep a position that is theoretically insensitive to oil price changes) depends strongly on whether one uses the arbitrage-based, expectation-based, or two-regime model; see Fig. 6.[9] Apparently the actions of MGRM were based on the arbitrage-based model, which calls for a hedge ratio of 1 irrespective of the maturities of the futures that are to be hedged and of the futures that are used for hedging. According to both the expectation-based and the two-regime model, the value of a ten-year future is not at all sensitive to changes of the current oil price; it is in fact completely determined by the model parameters[10] and neither of the two models suggests any hedging action for, say, about five years. So according to both alternative models MGRM should not have bought any short-term futures at all in 1993 to hedge its long-term commitments.

It should be noted that the use of an incorrect hedging strategy does not necessarily lead to a loss; it may actually produce a gain. This is

[9] The parameter values used here are the same as before.

[10] This implies, of course, that there are further robustness issues which however we do not discuss here.

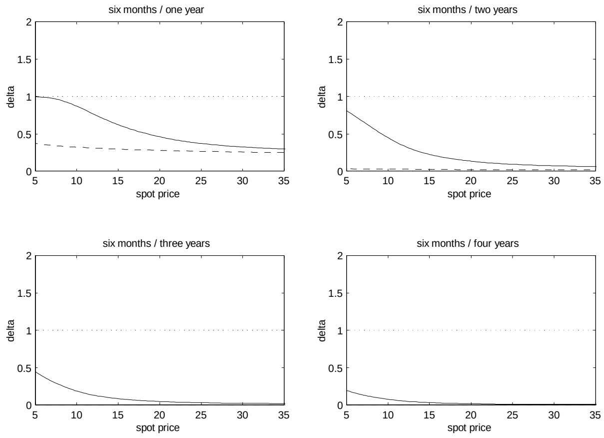

Fig. 4. *Ratios for hedging futures of several maturities by means of six-month futures as a function of the current oil price, according to three different models (dotted: arbitrage-based, drawn: two-regime, dashed: expectation-based)*

due to the fact that incorrect hedging creates a position that is speculative rather than risk-limited. Such positions can, under certain reward structures for company executives, be attractive to managers who see a possibility for large bonuses at limited risk for themselves. Therefore, the Metallgesellschaft case may have been one of those in which a company was punished heavily for lack of sophistication at the top level. These remarks aside, the question remains whether in principle one can transfer the liquidity of short-term oil futures to long-term futures by means of a hedging scheme. Experience has shown that it takes time for markets to mature, and innovations must take place in a stepwise manner. The attempt in 1992 by MGRM was a bridge too far.

7. Conclusions. During the 1980s and 1990s, in a gradual but not always smooth learning process, model-based thinking has gained acceptance in the finance industry. Academia has responded by the institution of programs in financial mathematics, which in turn have facilitated further developments in industry. While the origins of the new financial theory lie in the pricing of option contracts on stocks, the theory has been extended to contracts involving interest rates, commodities, and credit risk. The process continues as liquidity is transferred across markets, and in particular the interaction between financial and insurance markets is of current interest.

The new financial theory is based on the systematic investigation of actions that may be taken by the provider of a given contract to keep the ensuing liabilities in check. Taking decisions in the course of time on the basis of incoming information is therefore as central to financial engineering as it is to control engineering; dynamic modeling is the key to both fields. Financial models do have some specific features which distinguish them from models used in electrical or mechanical engineering; also, they are nearly always stochastic models, unlike many models used in control theory. Nevertheless, there is sufficient overlap in themes to think of control engineering and financial engineering as close neighbors. One of these themes that should be specifically mentioned here is the idea of robustness. Systematic thinking about robustness is fairly new in financial theory and is now developing along lines that are not unfamiliar to control theorists. Among other issues in finance where ideas from control theory could be helpful, one might mention noisy observations and the effects or regulatory constraints on financial markets.

In this paper we have focused attention on "complete markets," where the noise cancellation problem can be solved exactly, or in other words all risk generated by a specific liability can be eliminated. In such models the problem of valuation of contracts can be completely divorced from considerations concerning the risk preferences of market participants. Financial theory however has increasingly moved towards incomplete market models in which risk profiles may be changed by hedging, but always some risk remains; this is for instance a setting that has to be chosen in order to allow a discussion of interaction between financial markets and insurance markets. Valuation of contracts under such circumstances calls for a specification of the "acceptability" of positions. This may be done in a classical way by making use of utility functions, but the expected utility paradigm is not always able to explain empirical data and so it may be important to take into account "model uncertainty" as well. How to describe model uncertainty is a subject of current debate in financial mathematics, and not surprisingly some of the proposed frameworks are reminiscent of what has been developed in robust control theory (see for instance [5]). Developments in this area hold much promise for the future.

REFERENCES

[1] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, Journal of Political Economy, 81 (1973), pp. 637–659.

[2] M. J. Brennan and E. S. Schwartz, *Evaluating natural resource investments*, Journal of Business, 58 (1985), pp. 135–157.

[3] M. J. Brennan and Y. Xia, *Dynamic asset allocation under inflation*, Journal of Finance, 57 (2002), pp. 1201–1238.

[4] W. Bühler, O. Korn, and R. Schöbel, *Pricing and hedging of oil futures. A unifying approach*, Working Paper 00-01, Lehrstuhl für Finanzierung, Univ. Mannheim, 2000.

[5] H. Föllmer and A. Schied, *Robust preferences and convex measures of risk*, in Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann, K. Sandmann and P. Schönbucher, eds., Springer, Berlin, 2002, pp. 39–56.

[6] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus* (2nd ed.), Springer, New York, 1991.

[7] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications,* 5th ed., Springer, Berlin, 1998.

[8] P. E. Protter, *Stochastic Integration and Differential Equations: A New Approach*, Applications of Mathematics, 21, Springer, New York, 1999.
[9] E. S. Schwartz, *The stochastic behavior of commodity prices: Implications for valuation and hedging*, Journal of Finance, 52 (1997), pp. 923–973.
[10] O. A. Vasicek, *An equilibrium characterization of the term structure*, Journal of Financial Economics, 5 (1977), pp. 177–188.