1 Introduction

This Element is a pedagogical introduction to the technique of the change of numeraire in the martingale approach to option pricing, also known as arbitrage free pricing theory. Our intention is to present a reader-friendly explanation of the technique itself, and illustrate how it is applied in various fields of quantitative finance as the basis for building option valuation models. These applications include interest rate modeling, credit risk modeling, foreign exchange modeling, and others. We do not aspire to present an exhaustive list of such applications, or all the details of the markets, financial instruments, and models where martingale methods are particularly effective. Instead, we hope to provide the reader with enough background to be able to use the technique in his or her work.

The central concept of arbitrage free pricing is that of a *numeraire*. From the economic point of view, a numeraire is a tradeable asset conventionally used to quote prices of other tradeable goods. For example, currently most international commodities are traded using the US dollar as the numeraire. Historically, many other assets have been used as numeraires, such as certain metals (bronze, gold, silver), salt, agricultural products, and others. The choice of numeraire is dictated by convenience, and there is equivalence of asset prices expressed in terms of different numeraires. From the financial engineering perspective, a numeraire is a useful technical device that allows us to express certain financial transactions in conceptually transparent terms, and model their price processes as martingales. This point of view was first proposed in [10].

The key mathematical result that underlies this mechanism is *Girsanov's theorem*, which relates the change of the drift coefficient in an Ito process to a "change of variables" in the underlying probability measure. What is now known as Girsanov's theorem is actually a culmination of efforts by a number of mathematicians (notably Cameron, Martin, and Girsanov), who studied the effect of a change of variables in the measure on a probability space associated with a Brownian motion on the properties of a stochastic process on this space. It gives an explicit expression for the Radon–Nikodym derivative of such a change of variables in terms of the parameters of the stochastic process. A thorough and, at the same time, accessible mathematical presentation of stochastic calculus is presented in [21].

Girsanov's theorem plays a key conceptual role in arbitrage free pricing theory, which is the basis for the classical approach to derivatives pricing theory. The key tenet of this theory is *lack of arbitrage* (i.e. the inability to design portfolio strategies that generate riskless profit). Within this framework, lack of arbitrage opportunities is essentially equivalent to the ability of modeling asset prices as martingales. And while it is debatable whether the actual financial markets do not present arbitrage opportunities, arbitrage free pricing leads to a robust conceptual framework for derivatives modeling.

This Element is organized as follows. We start with an informal review of Girsanov's theorem, which is at the core of the arguments presented in the Element. This is followed by a brief review of the basic concepts of the arbitrage free pricing, just enough to cover our needs. We discuss the probabilistic concepts underlying this framework, but we do not delve into the nontrivial technical issues while doing so. For a full account of this theory, we refer to the original papers [15] and [16], as well as the thorough textbook presentation in [2] and [20]. Next, we discuss the technique of change of numeraire and derive the fundamental explicit relation between the drift coefficients corresponding to the same diffusion process under two different numeraires. This is followed by a number of important applications of the change of numeraire technique in interest rate models, FX quanto adjustments, credit risk modeling, mortgage-backed securities, and constant maturity swap (CMS) rates.

The intended style of exposition of the Element is conversational. We do not strive for mathematical rigor, using instead the "physics level" of mathematical reasoning. The theorems we quote do not always provide complete lists of technical assumptions, but focus instead on their financial relevance. On the finance side, we make technical simplifications as well. In the descriptions of fixed income instruments, we neglect various details, such as the delay between rate fixing dates and contract settlement dates, detailed definitions of the relevant day count fractions, business day conventions, precise conventions regarding various settlement days, and so on. In order to focus the discussion on the central concepts and not to obscure it with details, we also disregard the presence of some "bases" between various rates. Capturing these bases by mathematical models often presents mathematical challenges of their own.

2 Girsanov's Theorem

Before we move to finance, we start by reviewing the main mathematical result underlying much of formalism that follows, namely Girsanov's theorem.

The toy version of Girsanov's theorem is the following elementary calculation involving two Gaussian probability measures (or distributions) on the real axis. Consider two such measures:

$$d\mathsf{P}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big)dx, \tag{1}$$

and

$$dQ(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{(x-\nu)^2}{2\sigma^2} \Big) dx. \tag{2}$$

Note that P and Q may have different means, but they share the same variance. Then, manipulating the quadratic function in the exponential, we find that the following relation holds:

$$dQ(x) = L(x)dP(x), \tag{3}$$

where

$$L(x) = \exp\Big(\frac{(\nu-\mu)(x-\mu)}{\sigma^2} - \frac{1}{2}\frac{(\nu-\mu)^2}{\sigma^2} \Big). \tag{4}$$

Clearly,

$$\int_{-\infty}^{\infty} L(x)dP(x) = 1. \tag{5}$$

The function $L(x)$ is positive for all x and can suggestively be written as the "derivative" of measure Q with respect to measure P,

$$L(x) = \frac{dQ}{dP}(x), \tag{6}$$

known as the Radon–Nikodym derivative. One can think about the Radon–Nikodym derivative as an extension of the concept of the Jacobian of a change of variables familiar from calculus.

2.1 Radon–Nikodym Derivative

We now consider a one-dimensional Brownian motion $W(t)$. Let the associated *probability space* be denoted by $(\Omega, \mathscr{F}, P)$, where Ω is the sample space, $\mathscr{F} = (\mathscr{F}_t)_{t\geq 0}$ is the filtered information set, and P is the probability measure. A sample $\omega \in \Omega$ is a realization (that is, a possible trajectory) of the Brownian motion, that is, a continuous function $\omega(t), t \geq 0$, with $\omega(0) = 0$. All possible sample paths up to time $t > 0$ form the information set $\mathscr{F}_t$, which represents the history of the process known at the time t.

The probability measure P is a Gaussian measure defined so that the following two conditions hold. For any $s < t$, the increment $\Delta W(s,t) = W(t) - W(s)$ has expected value 0,

$$E[\Delta W(s,t)] = 0, \tag{7}$$

and variance equal to the time increment:

$$E[\Delta W(s,t)^2] = t - s. \tag{8}$$

Furthermore, any two such increments $\Delta W(s_1, t_1)$ and $\Delta W(s_2, t_2)$, with $t_1 < s_2$, are assumed independent:

$$\mathsf{E}[\Delta W(s_1, t_1)\Delta W(s_2, t_2)] = 0. \tag{9}$$

In other words, a Brownian motion advances by an amount ΔW over time Δt according to the Gaussian probability distribution (1) with $\mu = 0$ and $\sigma = \sqrt{\Delta t}$.

By $\mathsf{E}[X] = \int_\Omega X(\omega)dP(\omega)$ (or $\mathsf{E}^P[X]$, whenever we want to be precise about the probability measure) we denote the expected value of a random variable X with respect to the measure P. We will also be concerned with the conditional expected value $\mathsf{E}[X|\mathscr{F}_t]$ of X given the information available up to time t. Interchangeably, whenever we want to save space, we will be denoting this quantity by $\mathsf{E}_t[X]$.

Consider now two probability measures P and Q on an abstract (not necessarily associated with a Brownian motion) probability space Ω. We say that Q is *absolutely continuous* with respect to P if probability zero events with respect to P are also probability zero under Q, that is, if it satisfies the following condition:

$$\text{for a set } A \subset \Omega, \text{ if } \mathsf{P}(A) = 0, \text{ then also } \mathsf{Q}(A) = 0. \tag{10}$$

An important classic theorem of real analysis states that if Q is absolutely continuous with respect to P, then there exists a positive function $L(\omega)$ on Ω, called the *Radon–Nikodym derivative*, or the *likelihood ratio*, such that $L(\omega)$ is integrable with respect to P, and

$$\mathsf{Q}(A) = \int_A L(\omega)d\mathsf{P}(\omega), \tag{11}$$

for all $A \subset \Omega$. In particular,

$$\int_\Omega L(\omega)d\mathsf{P}(\omega) = 1. \tag{12}$$

In general, not much more about L is known other than it exists. A quick and clever proof of this theorem, due to von Neumann, is presented in [25].

One can express relation (11) suggestively as

$$\frac{d\mathsf{Q}}{d\mathsf{P}}(\omega) = L(\omega). \tag{13}$$

In other words, the "volume element" $d\mathsf{Q}$ is always proportional to the "volume element" $d\mathsf{P}$, with the proportionality factor that is a positive function throughout the probability space.

Two probability measures Q and P are called *equivalent*, if Q is absolutely continuous with respect to P and P is absolutely continuous with respect to Q. This means that these two measures have exactly the same sets of measure zero (and, consequently, the same sets of measure one).

In the context of a Brownian motion, we can encode the dependence of the Radon–Nikodym derivative on the filtration by time. Namely, we consider the conditional expected value

$$L(\omega, t) = \mathsf{E}_t[L(\omega)], \tag{14}$$

which represents the knowledge of $L(\omega)$ based on the information available at time t. In the following, we will be suppressing the dependence of $L(\omega, t)$ on ω, and denote it simply by $L(t)$. In other words, $L(t)$ is a stochastic process.

2.2 Girsanov's Theorem in One Dimension

Consider now a *diffusion process* $X(t)$ defined by the following stochastic differential equation (SDE) in the sense of Ito:

$$dX(t) = \mu(X(t), t)dt + c(X(t), t)dW(t), \tag{15}$$

which starts at some initial value X_0. The coefficients $\mu(X(t), t)$ and $c(X(t), t)$ in (15) are called the *drift* and *diffusion* coefficients, respectively. A convenient way of thinking about a diffusion process is in terms of discretizing it over finite time increments and following sequentially each sample path through a Monte Carlo simulation. In this picture, $\mu(X(t), t)$ is the local mean of a Gaussian distribution, while $c(X(t), t)$ is the local standard deviation of that distribution.

In the light of the toy calculation in the beginning of this section, a natural question then arises: can we transform a diffusion process into a diffusion process with a different drift,

$$dX(t) = \widetilde{\mu}(X(t), t)dt + c(X(t), t)d\widetilde{W}(t), \tag{16}$$

by a change to an equivalent probability measure Q? In particular, can we make the new process a martingale?

Recall that a process $X(t)$ is a *martingale* if $\mathsf{E}^{\mathsf{Q}}[|X(t)|] < \infty$, for all t, and

$$X(s) = \mathsf{E}^{\mathsf{Q}}[X(t)|\mathscr{F}_s], \tag{17}$$

where $\mathsf{E}^{\mathsf{Q}}[\,\cdot\,|\mathscr{F}_s]$ denotes conditional expected value. In other words, given all information up to time s, the expected value of any future value of a martingale is the current value $X(s)$. The diffusion process $X(t)$ is a martingale, if the diffusion here is driftless (i.e. $\widetilde{\mu}(X(t), t) = 0$).

An affirmative answer to this question is provided by Girsanov's theorem. One might heuristically proceed like this. Write

$$\begin{aligned} dX(t) &= \widetilde{\mu}(X(t), t)dt + c(X(t), t)\Big(\frac{\mu(X(t), t) - \widetilde{\mu}(X(t), t)}{c(X(t), t)}\, dt + dW(t)\Big) \\ &= \widetilde{\mu}(X(t), t)dt + c(X(t), t)d\widetilde{W}(t), \end{aligned}$$

where

$$\begin{aligned}\widetilde{W}(t) &= W(t) + \int_0^t \frac{\mu(X(s),s) - \widetilde{\mu}(X(s),s)}{c(X(s),s)}\,ds \\ &\equiv W(t) - \int_0^t \theta(s)\,ds,\end{aligned} \tag{18}$$

and we would like to interpret $\widetilde{W}(t)$ as a new Brownian motion. The presence of the shift $\int_0^t \theta(s)ds$ creates, however, an issue, as it leads to violation of conditions (7) and (9). The calculation at the beginning of this section suggests that this can be remedied by changing to a different, but equivalent, probability measure.

Girsanov's theorem asserts that indeed, under some technical assumptions on the drift and diffusion coefficients, $\widetilde{W}(t)$ defined in (18) is a Brownian motion, provided that the probability measure is modified appropriately.

More specifically, let us define the positive stochastic process:

$$L(t) = \exp\Big(\int_0^t \theta(s)\,dW(s) - \frac{1}{2}\int_0^t \theta(s)^2 ds\Big). \tag{19}$$

Recall that since L is a stochastic process, we have suppressed its explicit dependence on ω. We now define the equivalent measure Q by the requirement that its Radon–Nikodym derivative with respect to P is given by

$$\frac{d\mathsf{Q}}{d\mathsf{P}}(t) = L(t). \tag{20}$$

Let us emphasize that the Radon–Nikodym derivative $L(t)$ has an explicit form, built out of the data that define the process $X(t)$.

THEOREM (Girsanov's theorem). *Assume that the following technical condition (Novikov's condition) holds:*

$$\mathsf{E}^{\mathsf{P}}\Big[\exp\Big(\frac{1}{2}\int_0^t \theta(s)^2 ds\Big)\Big] < \infty. \tag{21}$$

Then:

(i) The process $L(t)$ is a martingale under P. Furthermore, it satisfies the following stochastic differential equation:

$$dL(t) = L(t)\theta(t)dW(t). \tag{22}$$

(ii) The process $\widetilde{W}(t)$ is a Brownian motion under Q.

2.3 Girsanov's Theorem in Many Dimensions

We have stated Girsanov's theorem for the case of a one-dimensional Brownian motion. This assumption is not really essential and, using a bit of linear algebra, one can easily formulate a version of Girsanov's theorem for an arbitrary multidimensional Brownian motion.

Consider an n-dimensional Brownian motion $W(t) = \big(W_1(t), \ldots, W_n(t)\big)$, whose components are uncorrelated one-dimensional Brownian motions,

$$dW_i(t)dW_j(t) = 0, \text{ for } i \neq j.$$

This assumption is not a real loss of generality: if we work with an n-dimensional Brownian motion whose components are correlated with a constant correlation matrix ρ,

$$dW_i(t)dW_j(t) = \rho_{ij}dt, \text{ for } i \neq j,$$

then we decompose each $W_j(t)$ in terms of the independent components and redefine the diffusion coefficients appropriately. Consider an n-dimensional diffusion process $X(t) = \big(X_1(t), \ldots, X_n(t)\big)$, given by the system of SDEs:

$$dX(t) = \mu(X(t), t)dt + c(X(t), t)dW(t), \tag{23}$$

where the drift $\mu(X(t), t)$ is a vector that takes values in $\mathbb{R}^n$, and the $n \times n$ matrix of diffusion coefficients $c(X(t), t)$ is assumed to be nonsingular.

Now, define the stochastic process:

$$L(t) = \exp\Big(\int_0^t \theta(s)^{\mathrm{T}}\, dW(s) - \frac{1}{2}\int_0^t \theta(s)^{\mathrm{T}}\theta(s)ds\Big), \tag{24}$$

where

$$\theta(t) = c(X(t), t)^{-1}\big(\widetilde{\mu}(X(t), t) - \mu(X(t), t)\big), \tag{25}$$

and where $^{\mathrm{T}}$ denotes transpose. Let $\widetilde{W}(t)$ be the shifted vector-valued Brownian motion:

$$\widetilde{W}(t) = W(t) - \int_0^t \theta(s)\, ds. \tag{26}$$

THEOREM (Girsanov's theorem – multidimensional version). *Assume that the following technical condition (Novikov's condition) holds:*

$$\mathsf{E}^{\mathsf{P}}\Big[\exp\Big(\frac{1}{2}\int_0^t \theta(s)^{\mathrm{T}}\theta(s)ds\Big)\Big] < \infty. \tag{27}$$

Then:

(i) *The process* $L(t)$ *is a martingale under* P. *Furthermore, it satisfies the following stochastic differential equation*:

$$dL(t) = L(t)\theta(t)^{\mathrm{T}} dW(t). \tag{28}$$

(ii) *The process* $\widetilde{W}(t)$ *is a Brownian motion under* Q.

3 Arbitrage Asset Pricing in a Nutshell

3.1 Frictionless Market Models

We consider a model of a *frictionless* financial market that consists of n (risky) assets $I_1, \ldots, I_n$. By frictionless we mean that:

(i) each of the assets is *liquid* (i.e. at each time any bid or ask order, regardless of size, can be immediately executed),
(ii) there are no *transaction costs* (i.e. each bid and ask order for each security at time t is executed at the same price level and without a commission),
(iii) executing an order, regardless of its size, leaves no *impact* on the market,
(iv) there is no *counterparty risk* (i.e. market participants do not default on fulfilling the terms of a trade).

This is obviously a gross oversimplification of reality, and much work has been done to relax these assumptions. From the conceptual point of view, however, it leads to a profound and workable framework of asset pricing theory. This is very much like the formulation of Newtonian gravity in vacuum, which neglects friction and energy dissipation.

We model the price processes of these assets by a vector of continuous time stochastic processes $S(t) = (S_1(t), \ldots, S_n(t))$, where $S_i(t)$ denotes the price process of asset I_i at time t. We emphasize that these processes represent market observable asset prices, and not merely some convenient state variables. We assume that each price process is a diffusion process. In other words, there is an underlying probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathsf{P})$ generated by a multidimensional Brownian motion $W(t) = (W_1(t), \ldots, W_d(t))$, which drives the price process,

$$dS(t) = \mu(S(t), t)dt + c(S(t), t)dW(t). \tag{29}$$

Here $\mu = (\mu_1, \ldots, \mu_n)$ and $c = (c_{ij})_{1\leq i\leq n, 1\leq j\leq d}$ are the vector of drift coefficients and the matrix of diffusion coefficients, respectively.

In order to develop an intuition for the concepts explained below, we recall the basic example from the world of equity derivatives.

EXAMPLE (Black–Scholes model). In this classic model of equity derivatives, $S_1(t) = B(t)$ is the riskless money market account, and $S_2(t) = S(t)$ is a risky stock, with the dynamics given by

$$
\begin{aligned}
dB(t) &= rB(t)dt, \\
dS(t) &= \mu S(t)dt + \sigma S(t)dW(t).
\end{aligned} \tag{30}
$$

The rate of return r on the money market account is called the *riskless rate*, while μ is the rate of return on the risky asset. The financial reality is that there is no such thing as riskless rate (whose closest proxy, in the US dollar market, is the SOFR rate), but its presence in the Black–Scholes model helps one understand the general framework of risk neutral valuation.

3.2 Self-Financing Portfolios and No Arbitrage

A *portfolio* is specified by the vector of weights $w(t) = (w_1(t), \ldots, w_n(t))$, according to which capital is allocated to each asset at time t. We assume that $w(t)$ is a process adapted to the information set $\mathscr{F}_t$. In other words, all allocation decisions in the portfolio are made based on the information available up to time t. We also assume that the weights are nonnegative, and they add up to one.

The value process of the portfolio is given by

$$V(t) = w(t)^{\mathrm{T}} S(t). \tag{31}$$

A portfolio is *self-financing*, if

$$dV(t) = w(t)^{\mathrm{T}} dS(t), \tag{32}$$

or, equivalently,

$$V(t) = V(0) + \int_0^t w(s)^{\mathrm{T}} dS(s). \tag{33}$$

In other words, the price process of a self-financing portfolio does not allow for infusion or withdrawal of capital. It is entirely driven by price processes of the constituent instruments and their weights.

A fundamental assumption of arbitrage pricing theory is that financial markets (or at least, their models) are free of arbitrage opportunities.[1] An *arbitrage opportunity* arises if one can construct a self-financing portfolio such that:

(i) the initial value of the portfolio is zero, $V(0) = 0$,
(ii) with probability one, the portfolio has a nonnegative value at maturity, $\mathsf{P}(V(T) \geq 0) = 1$,
(iii) with a positive probability, the value of the portfolio at maturity is positive, $\mathsf{P}(V(T) > 0) > 0$.

[1] This assumption is, mercifully, violated frequently enough so that much of the financial industry can sustain itself exploiting the market's lack of respect for arbitrage freeness.

We say the model is *arbitrage free* if it does not allow arbitrage opportunities. As we shall see in the next section, requiring arbitrage freeness has important consequences for price dynamics.

3.3 Complete Markets

A random variable X on the probability space $(\Omega, (\mathscr{F}_t)_{t\geq 0}, \mathsf{P})$ is square integrable if $\mathsf{E}[X^2] < \infty$. In other words, a square integrable random variable has a well-defined variance. A financial market is called *complete* if each such random variable can be obtained as the terminal value of a self-financing trading strategy, that is, if there exists a self-financing portfolio such that $X = V(T)$.

This somewhat technical sounding condition has a natural interpretation. If we think about X as the price of a (possibly very complicated) contingent claim, market completeness means that such a claim can be replicated by way of a self-financing trading strategy. In practical terms it means that the price of a contingent claim is uniquely determined in terms of the replicating portfolio. Completeness is a convenient but restrictive assumption, and many widely used models violate it. We will discuss some of such models in what follows.

EXAMPLE (Black–Scholes model revisited). To illustrate this concept, we consider again the Black–Scholes model. Let $g(S(T))$ be a time T payoff. For example, $g(S(T)) = \max(S(T) - K, 0)$ corresponds to a call option struck at K. Let $v = v(x, t)$ be the solution to the terminal value problem:

$$\begin{aligned} \frac{\partial \varphi}{\partial t} + rx\frac{\partial v}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v}{\partial x^2} - rv = 0, \\ v(x,t) = g(x). \end{aligned} \tag{34}$$

This is the terminal value problem for the Black–Scholes (or backward Kolmogorov) equation arising in the Black–Scholes model, and which is a consequence of Ito's lemma. Then, the portfolio with weights

$$\begin{aligned} w_1(t) &= \frac{v(S(t),t) - S(t)\frac{\partial v(S(t),t)}{\partial x}}{B(t)}, \\ w_2(t) &= \frac{\partial v(S(t),t)}{\partial x} \end{aligned} \tag{35}$$

is the replicating portfolio for the payoff $g(S(T))$. Indeed, $w_2(t)$ is the option delta, which is the dynamically updated number of shares of the underlying asset, and $w_1(t)$ is the number of units of the riskless bond. The value of this portfolio is, as expected, equal to

$$V(t) = v(S(t), t). \tag{36}$$