Chapter 6

Numeraires

In these notes we change the unit of account, which is the most useful single manoeuvre in derivative pricing and the least intuitive. Girsanov's theorem says a measure change moves drifts and cannot touch volatilities, and that asymmetry is what makes the technique work: choosing a numeraire chooses which quantity is a martingale, and the right choice makes a hard expectation trivial. We derive the forward and annuity measures that chapter 7 and chapter 15 need, and find that no-arbitrage assigns one price to each source of risk — the same projection that reappears as a hedge in chapter 23.

6.1 Change of Measure

Definition 6.1 (Radon-Nikodym Derivative)**.** Consider two equivalent probability measures $\mathbb{P}$ and $\hat{\mathbb{P}}$ on a measurable space (Ω, Σ). The Radon-Nikodym derivate $d\hat{\mathbb{P}}/d\mathbb{P} : \Omega \to \mathbb{R}$ is defined such that for any subset $A, \Omega \supseteqq A \in \Sigma$,

$$\int_A d\hat{\mathbb{P}} = \int_A (d\hat{\mathbb{P}}/d\mathbb{P}) d\mathbb{P}. \tag{6.1}$$

Suppose $(\Omega, \mathcal{F}, \mathcal{F}_t)$ is a filtered probability space, then note from the above definition we have

$$(d\hat{\mathbb{P}}/d\mathbb{P})|_{\mathcal{F}_t} = \mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{F}_t] \tag{6.2}$$

$(d\hat{\mathbb{P}}/d\mathbb{P})|_{\mathcal{F}_t}$ is also written as $(d\hat{\mathbb{P}}/d\mathbb{P})_t$ and is thus a martingale stochastic process (by iterated conditioning) in $\mathbb{P}$.

Example 6.1. Let's consider a simple example to better understand the Radon Nikodym derivative. Consider a die roll. We can assign multiple probability distributions to the outcomes.

The probability in $\hat{\mathbb{P}}$ of getting an odd number is $\int_{\omega \in \{1,3,5\}} d\hat{\mathbb{P}} = 1/2 + 1/8 + 1/32 = 21/32 = 3 * 1/6 + 3/4 * 1/6 + 3/16 * 1/6 = \int_{\omega \in \{1,3,5\}} (d\hat{\mathbb{P}}/d\mathbb{P}) d\mathbb{P}$

Theorem 6.2 (Abstract Bayes' Theorem)**.** *Let $\mathbb{P}$ and $\hat{\mathbb{P}}$ be two measures on measurable space $(\Omega, \mathcal{F})$. Let $\mathcal{G} \subset \mathcal{F}$ be another sigma algebra on Ω. Then for any $A \in \mathcal{G}$ and random variable X*

$$\mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}] = \frac{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})X|\mathcal{G}]}{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{G}]}. \tag{6.3}$$

ω	$\mathbb{P}$	$\hat{\mathbb{P}}$	$d\hat{\mathbb{P}}/d\mathbb{P}$
1	1/6	1/2	3
2	1/6	1/4	3/2
3	1/6	1/8	3/4
4	1/6	1/16	3/8
5	1/6	1/32	3/16
6	1/6	1/32	3/16

Proof. We show that for any $A \in \mathcal{G}$

$$\mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}]\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{G}] = \mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})X|\mathcal{G}]$$

Since the random variables involved are constant over A, we can check equality on integrals over A.

$$\begin{aligned}
\int_A \mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}]\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{G}]d\mathbb{P} &= \int_A \mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})\mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}]|\mathcal{G}]d\mathbb{P} && (\mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}] \text{ is } \mathcal{G} \text{ measurable}) \\
&= \int_A (d\hat{\mathbb{P}}/d\mathbb{P})\mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}]d\mathbb{P} && \text{(conditional expectation)} \\
&= \int_A \mathbb{E}^{\hat{\mathbb{P}}}[X|\mathcal{G}]d\hat{\mathbb{P}} && \text{(Radon-Nikodym)} \\
&= \int_A X d\hat{\mathbb{P}} && \text{(conditional expectation)} \\
&= \int_A (d\hat{\mathbb{P}}/d\mathbb{P})X d\mathbb{P} && \text{(Radon-Nikodym)} \\
&= \int_A \mathbb{E}^{\mathbb{P}}\left[(d\hat{\mathbb{P}}/d\mathbb{P})X \,\middle|\, \mathcal{G}\right] d\mathbb{P}.
\end{aligned}$$

The last step is the definition of conditional expectation. □

Remark. Taking $X = V_T$ where V_s is $\mathcal{F}_s$ adapted and $G = \mathcal{F}_t$ for $t < T$ in the above theorem, we get the very useful formula for measure change for conditional expectations on filtered spaces,

$$\mathbb{E}^{\hat{\mathbb{P}}}[V_T|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\frac{(d\hat{\mathbb{P}}/d\mathbb{P})_T}{(d\hat{\mathbb{P}}/d\mathbb{P})_t}V_T|\mathcal{F}_t] \tag{6.4}$$

Proof.

$$
\begin{aligned}
\mathbb{E}^{\hat{\mathbb{P}}}[V_T|\mathcal{F}_t] &= \frac{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})V_T|\mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{F}_t]} && \text{(abstract Bayes' theorem)}\\
&= \frac{\mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})V_T|\mathcal{F}_T]|\mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{F}_t]} && \text{(iterated conditioning)}\\
&= \frac{\mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{F}_T]V_T|\mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})|\mathcal{F}_t]} && \text{(V_T is $\mathcal{F}_T$ measureable)}\\
&= \frac{\mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})_T V_T|\mathcal{F}_t]}{(d\hat{\mathbb{P}}/d\mathbb{P})_t} && \text{(martigale property of Radon Nikodym derivative)}\\
&= \mathbb{E}^{\mathbb{P}}[\frac{(d\hat{\mathbb{P}}/d\mathbb{P})_T}{(d\hat{\mathbb{P}}/d\mathbb{P})_t}V_T|\mathcal{F}_t] && \text{($(d\hat{\mathbb{P}}/d\mathbb{P})_t$ is $\mathcal{F}_t$ measurable).}
\end{aligned}
$$

□

We consider processes until terminal time S i.e. $\mathcal{F} = \mathcal{F}_S$ and $(d\hat{\mathbb{P}}/\mathbb{P}) = (d\hat{\mathbb{P}}/\mathbb{P})_S$. We take a strictly positive martingale process to be the Radon-Nikodym derivative:

$$df_t = f_t\sigma(t)dW_t; \;\; f_t = e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds+\int_0^t \sigma(s)dW_s}. \tag{6.5}$$

Structure (A measure change moves the drift and cannot touch the volatility)**.** Before the theorem, the shape of what it says. Girsanov changes the drift of a process and leaves its diffusion coefficient exactly where it was.

That asymmetry explains several things at once, and stating it precisely matters, because the loose version is misleading in a way that bites later.

Quadratic variation is computed *pathwise*: $\langle M\rangle_t$ is a limit of sums of squared increments of the trajectory that actually occurred, so it is a random variable and not an expectation. Different paths have different quadratic variations — in a stochastic volatility model $\langle S\rangle_t = \int_0^t v_s S_s^2\, ds$, and a path that spent its life in a high-variance regime has a larger one than a path that did not. What is measure-invariant is therefore not a number but a *function of the path*: the map from trajectory to accumulated variance mentions no measure anywhere, so two equivalent measures assign the same quadratic variation to the same path. They disagree about which paths are likely, not about what each path's variance is. Equivalence is doing exactly one job in that sentence: the limit converges in probability, so $\langle M\rangle$ is defined only up to null sets, and equivalent measures are precisely those that agree on which sets those are.

That is why volatility can be estimated from one realisation and drift cannot. An observer sees a single trajectory; the sums of squared increments of *that* trajectory converge without reference to any measure, so no averaging over paths that did not happen is required. A statement about drift is the opposite kind of statement — it is a claim about the average over the paths one did not see — so it needs either many independent histories, of which there is one, or a long span of time. Chapter 21 measures the consequence: over a fixed window, sampling faster drives the volatility error to zero and leaves the drift error exactly where it was. It is also why chapter 22 can measure realised volatility and must argue about expected returns, and why every model in these notes is calibrated on volatilities and never on drifts.

It is also why the risk neutral measure exists at all. Pricing requires the drift to be one particular thing, the measure change is free to set it, and nothing about the observable roughness of the path has to be disturbed to do so.

One consequence deserves drawing out now, because collapsing it is the commonest way to misread the paragraph above. The quadratic variation of a given path is measure-free; the *distribution* of quadratic variation across paths is not. So

$$\mathbb{E}^{\mathbb{P}}\big[\langle S\rangle_T\big] \neq \mathbb{E}^{\mathbb{Q}}\big[\langle S\rangle_T\big]$$

is perfectly consistent with everything just said, and it is the variance risk premium that chapter 21 measures — the reason implied volatility exceeds realised on average. Girsanov leaves σ alone as a *coefficient* and does not leave the law of accumulated variance alone at all, because in a stochastic volatility model it changes the drift of v itself. The two statements to keep apart are that the realised variance of the path one saw is a fact, and that the variance one expects is a choice of measure.

Theorem 6.3 (Girsanov Theorem)**.** *If $W_{1,t}$ and $W_{2,t}$ are (possibly correlated) Brownian motions in $\mathbb{P}$ and the Radon-Nikodym derivative $(d\hat{\mathbb{P}}/d\mathbb{P})_t = f_t$ is given by $f_t = e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds + \int_0^t \sigma(s)dW^{\mathbb{P}}_{2,s}}$ then $W_{1,t} - \int_0^t \sigma(s)dW_{1,s}dW_{2,s}$ is a Brownian motion in $\hat{\mathbb{P}}$, where $dW_{1,s}dW_{2,s}$ is the covariation rate of W_1 and W_2 (zero if they are independent, $\rho\, ds$ if correlated at ρ, and ds in the case $W_1 = W_2$).*

Proof. We show that $X_t = W_{1,t} - \int_0^t \sigma(s)dW_{1,s}dW_{2,s}$ follows normal distribution with mean 0 and variance t in $\hat{\mathbb{P}}$. Other required properties can be verified easily. We show that the moment generating function of X_t is same as that of normal distribution with mean 0 and variance t.

$$\begin{aligned}
\mathbb{E}^{\hat{\mathbb{P}}}[e^{-yX_t}] &= \mathbb{E}^{\mathbb{P}}[(d\hat{\mathbb{P}}/d\mathbb{P})_t e^{-yX_t}] \\
&= \mathbb{E}^{\mathbb{P}}[e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds + \int_0^t \sigma(s)dW_{2,s}} e^{-yX_t}] \\
&= e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds}\mathbb{E}^{\mathbb{P}}[e^{\int_0^t \sigma(s)dW_{2,s} - yX_t}] \\
&= e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds}\mathbb{E}^{\mathbb{P}}[e^{\int_0^t \sigma(s)dW_{2,s} - y(W_{1,t} - \int_0^t \sigma(s)dW_{1,s}dW_{2,s})}] \\
&= e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds}\mathbb{E}^{\mathbb{P}}[e^{\int_0^t (\sigma(s)dW_{2,s} - ydW_{1,s} + y\sigma(s)dW_{1,s}dW_{2,s})}].
\end{aligned}$$

Define the martingale $Z_{y,t} = \int_0^t(\sigma(s)dW_{2,s} - ydW_{1,s})$. Then $e^{Z_{y,t} - \frac{1}{2}\int_0^t dZ_{y,s}dZ_{y,s}}$ is a martingale as well, with

$$dZ_{y,s}dZ_{y,s} = (\sigma^2(s) + y^2)ds - 2y\sigma dW_{1,s}dW_{2,s}.$$

We then have

$$\begin{aligned}
\mathbb{E}^{\hat{\mathbb{P}}}[e^{-yX_t}] &= e^{-\int_0^t \frac{1}{2}\sigma^2(s)ds}\mathbb{E}^{\mathbb{P}}[e^{Z_{y,t} - \frac{1}{2}\int_0^t dZ_{y,s}dZ_{y,s} + \frac{1}{2}\int_0^t (\sigma^2(s)+y^2)ds}] \\
&= e^{\frac{1}{2}y^2 t}\mathbb{E}^{\mathbb{P}}[e^{Z_{y,t} - \frac{1}{2}\int_0^t dZ_{y,s}dZ_{y,s}}] \\
&= e^{\frac{1}{2}y^2 t}(e^{Z_{y,t} - \frac{1}{2}\int_0^t dZ_{y,s}dZ_{y,s}})|_{t=0} \\
&= e^{\frac{1}{2}y^2 t}.
\end{aligned}$$

□

Remark (What a change of measure does, and what it does not do)**.** The algebra above is dense enough to lose the idea in, and the idea is simple. A change of measure does not move anything. Every path the world could take is still available afterwards, ending where it always would have. What changes is how much each one counts.

The die in the example at the start of this chapter is the honest picture of it: the same six faces, reweighted. Girsanov is that operation performed on a continuum of paths rather than six outcomes, and the Radon-Nikodym derivative is the column of weights.

For the geometric Brownian motion of chapter 5 the weights can be written down. If the asset drifts at μ in the real world and at r under the risk neutral measure, the change of measure is governed by

$$\theta = \frac{\mu - r}{\sigma}, \tag{6.6}$$

the excess return per unit of risk, called the market price of risk. A path ending at S_T carries the weight

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\theta W_T - \frac{1}{2}\theta^2 T}, \tag{6.7}$$

which decreases in W_T and therefore in S_T: the risk neutral measure counts the good outcomes for less. That is the whole content of the phrase “removing the risk premium”, and (6.6) says exactly how much removing costs.

Structure (One price per risk, and no-arbitrage is what sets it)**.** The scalar $\theta = (\mu - r)/\sigma$ above is a ratio that is easy to read as bookkeeping. With more than one asset it stops being bookkeeping and becomes the fundamental theorem in coordinates.

Take n assets driven by d Brownian motions, with σ the matrix of loadings and $\mu - r\mathbf{1}$ the vector of excess returns. Girsanov shifts each Brownian motion by some θ_j, and asset i's drift falls by its own exposure to those shifts. Demanding that *every* asset end up drifting at r under *one* measure is therefore the linear system

$$\mu - r\mathbf{1} = \sigma\,\theta. \tag{6.8}$$

Read the indices. There is one θ_j per Brownian motion and not one per asset: a price of risk belongs to the *risk*, and each asset's excess return is forced to be its own loadings against those common prices. Two assets exposed to the same factor must reward that exposure at the same rate, whatever else differs between them.

Which is a statement no-arbitrage can enforce, and does. If two assets on one Brownian motion offered different Sharpe ratios, sell the one paying less per unit of risk, buy the one paying more, in the ratio that cancels the exposure — the position has no randomness left and a positive drift. Solvability of (6.8) is exactly the absence of that trade, which is chapter 4's fundamental theorem written with matrices instead of measures:

Statement about (6.8)	**Statement about measures**
solvable	a risk neutral measure exists, so no arbitrage
uniquely solvable	it is unique, so the market is complete
unsolvable	an arbitrage, and the residual is the portfolio

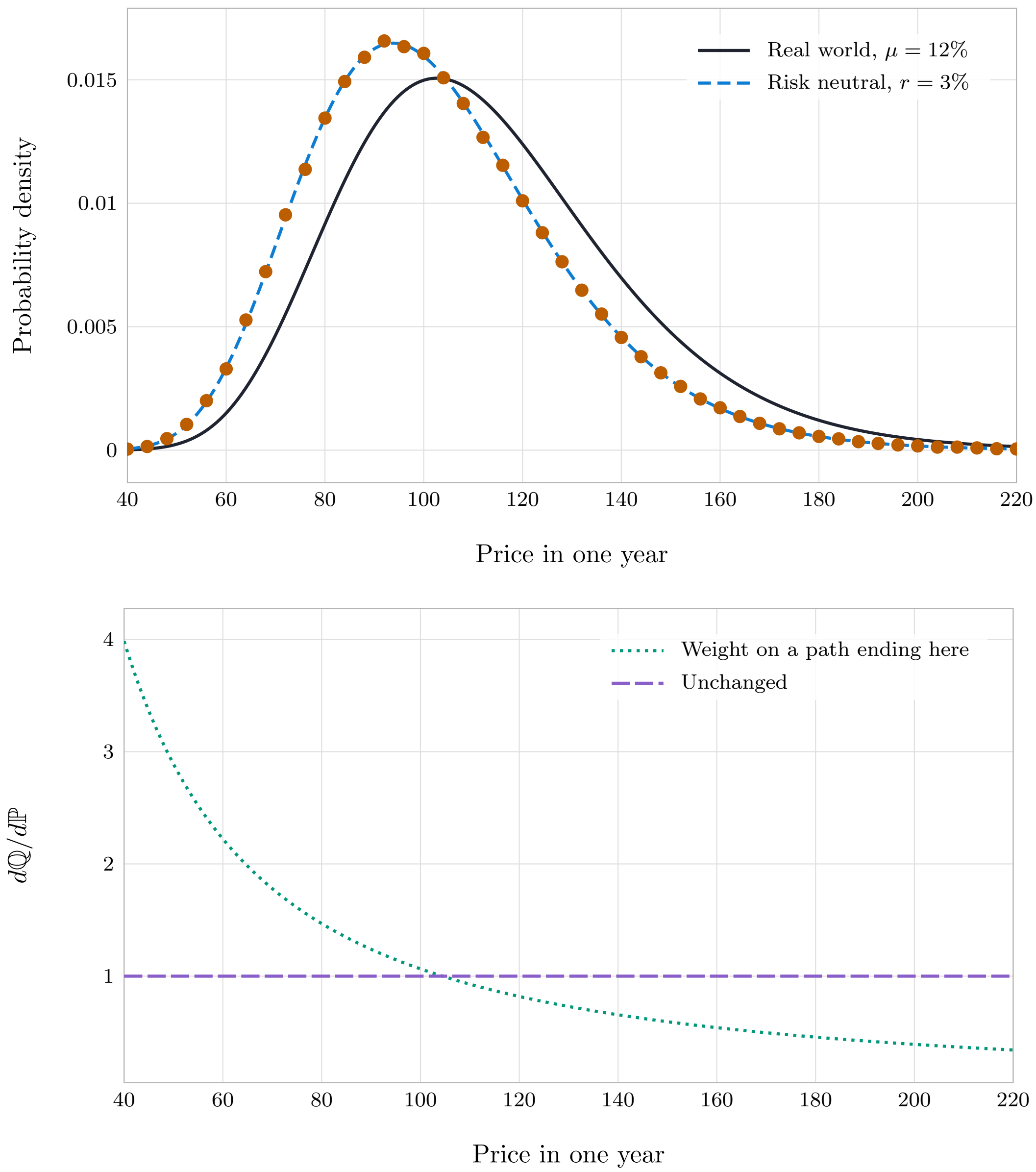

Figure 6.1: A change of measure, performed rather than described. The dotted sample is drawn under the real world measure at $\mu = 12\%$ and never redrawn; each draw is then multiplied by the weight (6.7). It lands on the risk neutral density at $r = 3\%$, which the sampling never saw. The second panel is the weight itself, tilting down across the outcomes. Drag μ onto the interest rate and the weight flattens onto one, the two densities merge, and the sample follows them — with no risk premium there is nothing to reweight, and the two measures coincide.

Drawn by `radon_nikodym` and `reweighted_histogram`.

The last row is constructive rather than merely a diagnosis. Solving (6.8) in the least squares sense leaves a residual orthogonal to every column of σ — so a portfolio holding those weights carries no exposure to any factor at all — while its excess return is the residual's own squared length, and therefore positive precisely when no set of prices fits. The failure hands over the trade. `factor_risk_prices` builds it and checks both halves.

This is also where chapter 8's term structure models get their λ. Every bond on the curve is driven by the same short rate, so every bond must reward that one risk at one rate, and forming a riskless combination of two maturities is the argument above with $n = 2$. The market price of interest rate risk is not an extra modelling assumption. It is (6.8) with one column.

6.2 Numeraire Measures

Definition 6.4 (Numeraire)**.** A Numeraire is a strictly positive price process of a tradable relative to which prices of all other tradables are expressed.

A savings account which earns interest at the instantaneous interest rate can be taken as a numeraire. The value of the savings account at any point is given by

$$A_t = e^{\int_0^t r(t)dt} = 1/D(t) \tag{6.9}$$

where $D(t)$ is the discount factor.

The fact that discounted trade prices are martingales in risk-neutral measure can then also be stated as: tradable prices in savings account (or money market) numeraire are martingales.

Consider another numeraire N_t. Since the numeraire is itself a price process, $D_t N_t$ is a martingale and we can take it as a Radon-Nikodym derivative, with a suitable normalization so that $\mathbb{E}^{\mathbb{P}}[(d\mathbb{N}/d\mathbb{P})] = 1$, giving

$$\mathbb{E}^{\mathbb{N}}[\frac{X_T}{N_T}|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\frac{(d\mathbb{N}/d\mathbb{P})_T}{(d\mathbb{N}/d\mathbb{P})_t}\frac{X_T}{N_T}|\mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[\frac{D_T N_T}{D_t N_t}\frac{X_T}{N_T}|\mathcal{F}_t] = \frac{1}{D_t N_t}\mathbb{E}^{\mathbb{P}}[D_T X_T|\mathcal{F}_t] = \frac{X_t}{N_t} \tag{6.10}$$

where $\mathbb{P}$ is the risk neutral measure and $\mathbb{N}$ is the measure corresponding to the numeraire N_t.

Remark (Prices in a numeraire are martingales in its measure)**.** The above equation shows that tradable prices with respect to a numeraire are martingales in the measure corresponding to the numeraire.

Risk neutral measure corresponds to the savings account (or money market) numeraire.

Definition 6.5 (T-forward measure)**.** If we take N_t to be the price of a riskless bond maturing at time T, the corresponding measure is the T-forward measure.

$$X_t/B(t,T) = \mathbb{E}^T[X_T/B(T,T)] = \mathbb{E}^T[X_T] \tag{6.11}$$

We see that the expectation of X_T in the T-forward measure gives the forward price $X_t/B(t,T)$ of the trade with expiration date T, hence the name T-forward measure. From the above equation we also notice that expiration T forward price process on an asset is martingale in T-forward measure.

Everything so far says which quantities are martingales under which measure, which is a statement about drifts being zero. The working question is usually the neighbouring one: given a process's drift under one numeraire, what is it under another. The answer follows from Girsanov and the density identified above, and it is short enough to record once and use thereafter — chapter 13 in particular does almost nothing else.

Theorem 6.6 (Change of numeraire, in drifts)**.** *Let A and B be numeraires, with $\mathbb{Q}^A$ and $\mathbb{Q}^B$ the corresponding measures, and let X be an Itô process. Then*

$$\mathrm{drift}^B(X) \;=\; \mathrm{drift}^A(X) \;+\; \frac{d}{dt}\Big\langle X,\, \ln \tfrac{B}{A}\Big\rangle_t. \tag{6.12}$$

In particular, for a strictly positive S the drift of $\ln S$ picks up $d\langle \ln S, \ln(B/A)\rangle$, and so does the relative drift in dS_t/S_t.

Proof. B is the price of a tradable and A is the numeraire, so B/A is a price expressed in units of A and is a $\mathbb{Q}^A$ martingale by the remark above. Normalising it,

$$Z_t \;=\; \frac{B_t/A_t}{B_0/A_0},$$

gives a strictly positive $\mathbb{Q}^A$ martingale starting at one, and it is exactly the density $(d\mathbb{Q}^B/d\mathbb{Q}^A)_t$ identified earlier in this section.

Being positive it can be written $dZ_t = Z_t\nu_t\, dW_t^A$, and Girsanov then says $W^B = W^A - \int \nu\, ds$ is a Brownian motion under $\mathbb{Q}^B$. So for any $dX_t = \mu_t^A\, dt + \sigma_t\, dW_t^A$, substituting $dW^A = dW^B + \nu\, dt$ gives

$$dX_t \;=\; \big(\mu_t^A + \sigma_t\nu_t\big)dt + \sigma_t\, dW_t^B.$$

It remains to recognise $\sigma\nu$. By Itô, $d\ln Z = \nu\, dW^A - \frac{1}{2}\nu^2 dt$, so $d\langle X, \ln Z\rangle = \sigma\nu\, dt$; and $\ln Z$ differs from $\ln(B/A)$ by a constant, which no covariation sees. That is (6.12).

For the last claim take $X = \ln S$. The relative drift follows because $dS/S = d\ln S + \frac{1}{2}d\langle \ln S\rangle$, and the quadratic variation term is pathwise and so the same under both measures. ☐

Remark (Reading the formula)**.** Three things stand out, since the formula gets used more often than it gets derived.

It depends on B/A alone. Neither numeraire matters on its own, only their ratio, which is why changing from A to B and back leaves every drift where it started.

A process uncorrelated with that ratio does not move at all. The adjustment is a covariation, so it vanishes exactly when the process and the numeraire ratio share no randomness, however volatile either happens to be.

And what it adjusts is a drift, never a volatility — the structure block at the head of this chapter, arriving as a formula.

Example 6.2 (Option Pricing)**.** To see how measure changes can be used in pricing, let's take the example of Black Scholes option pricing. For an option on stock with expiry at T and strike K, we have the valuation forumula

$$V_t = \frac{1}{D_t}\mathbb{E}^{\mathbb{P}}[D_T \max(S_T - K, 0)|\mathcal{F}_t]$$

where $\mathbb{P}$ is the risk neutral probability measure, D_t is the discount factor and S_t is the stock price which follows the Black Scholes diffusion

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t$$

where r is the riskfree interest rate and W_t is a Brownian motion in $\mathbb{P}$.

Denote by $\mathbb{I}_{S_T>K}$ the indicator variable which takes value 1 is $S_T > K$ and 0 otherwise. We then have

$$\begin{aligned} V_t &= \frac{1}{D_t}\mathbb{E}^{\mathbb{P}}[D_T\mathbb{I}_{S_T>K}(S_T-K)|\mathcal{F}_t] \\ &= \frac{1}{D_t}\mathbb{E}^{\mathbb{P}}[D_T\mathbb{I}_{S_T>K}S_T|\mathcal{F}_t] - \frac{1}{D_t}\mathbb{E}^{\mathbb{P}}[D_T\mathbb{I}_{S_T>K}K|\mathcal{F}_t] \end{aligned}$$

Taking $\mathbb{S}$ to be the measure corresponding to stock price as numeraire, and taking $\mathbb{T}$ to be the measure corresponding to T-expiry bond as numeraire, we have

$$\begin{aligned} V_t &= \frac{1}{D_t}\mathbb{E}^{\mathbb{S}}[D_T\frac{(d\mathbb{P}/d\mathbb{S})_T}{(d\mathbb{P}/d\mathbb{S})_t}\mathbb{I}_{S_T>K}S_T|\mathcal{F}_t] - \frac{1}{D_t}\mathbb{E}^{\mathbb{T}}[D_T\frac{(d\mathbb{P}/d\mathbb{T})_T}{(d\mathbb{P}/d\mathbb{T})_t}\mathbb{I}_{S_T>K}K|\mathcal{F}_t] \\ &= \frac{1}{D_t}\mathbb{E}^{\mathbb{S}}[D_T\frac{(d\mathbb{S}/d\mathbb{P})_t}{(d\mathbb{S}/d\mathbb{P})_T}\mathbb{I}_{S_T>K}S_T|\mathcal{F}_t] - \frac{1}{D_t}\mathbb{E}^{\mathbb{T}}[D_T\frac{(d\mathbb{T}/d\mathbb{P})_t}{(d\mathbb{T}/d\mathbb{P})_T}\mathbb{I}_{S_T>K}K|\mathcal{F}_t] \\ &= \frac{1}{D_t}\mathbb{E}^{\mathbb{S}}[D_T\frac{D_tS_t}{D_TS_T}\mathbb{I}_{S_T>K}S_T|\mathcal{F}_t] - \frac{1}{D_t}\mathbb{E}^{\mathbb{T}}[D_T\frac{D_t(B(t,T))}{D_TB(T,T)}\mathbb{I}_{S_T>K}K|\mathcal{F}_t] \\ &= S_t\mathbb{E}^{\mathbb{S}}[\mathbb{I}_{S_T>K}|\mathcal{F}_t] - KB(t,T)\mathbb{E}^{\mathbb{T}}[\mathbb{I}_{S_T>K}|\mathcal{F}_t] \end{aligned}$$

where $B(t,T)$ is the bond price with the diffusion

$$dB(t,T)/B(t,T) = rdt + \sigma_2 dW_{2,t}$$

where $W_{2,t}$ is another Brownian motion in $\mathbb{P}$ such that $dW_t dW_{2,t} = \rho dt$.

Using Girsanov's theorem, we have

$$\begin{aligned} dW_t^{\mathbb{S}} &= dW_t^{\mathbb{P}} - \sigma dW_t^{\mathbb{P}}dW_t^{\mathbb{P}} & &= dW_t^{\mathbb{P}} - \sigma dt \\ dW_t^{\mathbb{T}} &= dW_t^{\mathbb{P}} - \sigma_2 dW_t^{\mathbb{P}}dW_{2,t}^{\mathbb{P}} & &= dW_t^{\mathbb{P}} - \sigma_2\rho dt \end{aligned}$$

as Brownian motions in the measures corresponding to S, and B(t,T) numeraires respectively.

$$\frac{dS_t}{S_t} = rdt + \sigma(dW_t^{\mathbb{S}} + \sigma dt) = rdt + \sigma(dW_t^{\mathbb{T}} + \sigma_2\rho dt) \tag{6.13}$$

where $dW_t^{\mathbb{S}} = dW_t - \sigma dt$ is a Brownian motion in $\mathbb{S}$ and $dW_t^{\mathbb{T}} = dW_t^{\mathbb{S}} - \sigma_2\rho dt$ is a Brownian motion in $\mathbb{T}$. Rearranging,

$$\frac{dS_t}{S_t} = (r+\sigma^2)dt + \sigma dW_t^{\mathbb{S}} = (r+\sigma\sigma_2\rho)dt + \sigma dW_t^{\mathbb{T}} \tag{6.14}$$

$$S_T = S_t e^{(r+\frac{1}{2}\sigma^2)\tau + \sigma W_\tau^{\mathbb{S}}} = S_t e^{\left(r+\sigma(\sigma_2\rho - \frac{1}{2}\sigma)\right)\tau + \sigma W_\tau^{\mathbb{T}}} \tag{6.15}$$

where $\tau = T - t$.

$$\begin{aligned}
\mathbb{E}^{\mathbb{S}}[\mathbb{I}_{S_T>K}|\mathcal{F}_t] &= \mathbb{E}^{\mathbb{S}}[\mathbb{I}_{S_t e^{(r+\frac{1}{2}\sigma^2)\tau+\sigma W_\tau^{\mathbb{S}}}>K}|\mathcal{F}_t] \\
&= \mathbb{E}^{\mathbb{S}}[\mathbb{I}_{(r+\frac{1}{2}\sigma^2)\tau+\sigma W_\tau^{\mathbb{S}}>\log(K/S_t)}|\mathcal{F}_t] \\
&= \mathbb{E}^{\mathbb{S}}[\mathbb{I}_{\sigma W_\tau^{\mathbb{S}}>\log(K/S_t)-(r+\frac{1}{2}\sigma^2)\tau}|\mathcal{F}_t] \\
&= \mathbb{E}^{\mathbb{S}}[\mathbb{I}_{\frac{1}{\sqrt{\tau}}W_\tau^{\mathbb{S}}>\frac{1}{\sigma\sqrt{\tau}}(\log(K/S_t)-(r+\frac{1}{2}\sigma^2)\tau)}|\mathcal{F}_t] \\
&= N(-\frac{1}{\sigma\sqrt{\tau}}(\log(K/S_t)-(r+\frac{1}{2}\sigma^2)\tau)) \\
&= N(d_1)
\end{aligned}$$

where N is the cumulative normal distribution, and $d_1 = \frac{1}{\sigma\sqrt{\tau}}(\log(S_t/K) + (r + \frac{1}{2}\sigma^2)\tau)$.

Similarly we have,

$$\begin{aligned}
\mathbb{E}^{\mathbb{T}}[\mathbb{I}_{S_T>K}|\mathcal{F}_t] &= \mathbb{E}^{\mathbb{T}}[\mathbb{I}_{S_t e^{\left(r+\sigma(\sigma_2\rho-\frac{1}{2}\sigma)\right)\tau+\sigma W_\tau^{\mathbb{T}}}>K}|\mathcal{F}_t] \\
&= \mathbb{E}^{\mathbb{T}}[\mathbb{I}_{\left(r+\sigma(\sigma_2\rho-\frac{1}{2}\sigma)\right)\tau+\sigma W_\tau^{\mathbb{T}}>\log(K/S_t)}|\mathcal{F}_t] \\
&= \mathbb{E}^{\mathbb{T}}[\mathbb{I}_{\sigma W_\tau^{\mathbb{T}}>\log(K/S_t)-\left(r+\sigma(\sigma_2\rho-\frac{1}{2}\sigma)\right)\tau}|\mathcal{F}_t] \\
&= \mathbb{E}^{\mathbb{T}}[\mathbb{I}_{\frac{1}{\sqrt{\tau}}W_\tau^{\mathbb{T}}>\frac{1}{\sigma\sqrt{\tau}}\left(\log(K/S_t)-\left(r+\sigma(\sigma_2\rho-\frac{1}{2}\sigma)\right)\tau\right)}|\mathcal{F}_t] \\
&= N\left(-\frac{1}{\sigma\sqrt{\tau}}\left(\log(K/S_t) - \left(r + \sigma(\sigma_2\rho - \frac{1}{2}\sigma)\right)\tau\right)\right) \\
&= N(d_2)
\end{aligned}$$

where N is the cumulative normal distribution, and $d_2 = \frac{1}{\sigma\sqrt{\tau}}(\log(S_t/K) + (r + \sigma\sigma_2\rho - \frac{1}{2}\sigma^2)\tau)$.

And finally we have,

$$\begin{aligned}
V_t &= S_t N(d_1) - KB(t,T)N(d_2) \\
&= S_t N(d + \sigma\sqrt{\tau}) - KB(t,T)N(d + \sigma_2\rho\sqrt{\tau})
\end{aligned}$$

where $d = \frac{1}{\sigma\sqrt{\tau}}(\log(S_t/K) + (r - \frac{1}{2}\sigma^2)\tau)$.

Note that $\sigma\, dt$ and $\sigma_2\rho\, dt$ are the drift correction terms we get to stock price Brownian motion using Girsanov's theorem to change measure to stock and bond numeraires respectively.

Remark (The two terms are one event counted twice)**.** The derivation above deserves one more sentence: it explains something about the Black formula that is usually left as a curiosity.

$N(d_1)$ and $N(d_2)$ are both the probability of $\{S_T > K\}$ — the same event, the option finishing in the money. They are different numbers because they are computed under different measures: $N(d_1)$ under the measure in which the stock is the numeraire, $N(d_2)$ under the T-forward measure. Neither is "the" probability of exercise, and asking which one is misses that the question has no answer until a numeraire is named.

The gap between them is the measure change, and it is $\sigma\sqrt{\tau}$ worth of shift in the argument of N. It is therefore largest for long dated and volatile options and vanishes as $\sigma\sqrt{\tau} \to 0$, where there is no randomness for the reweighting to act on and the two measures agree.

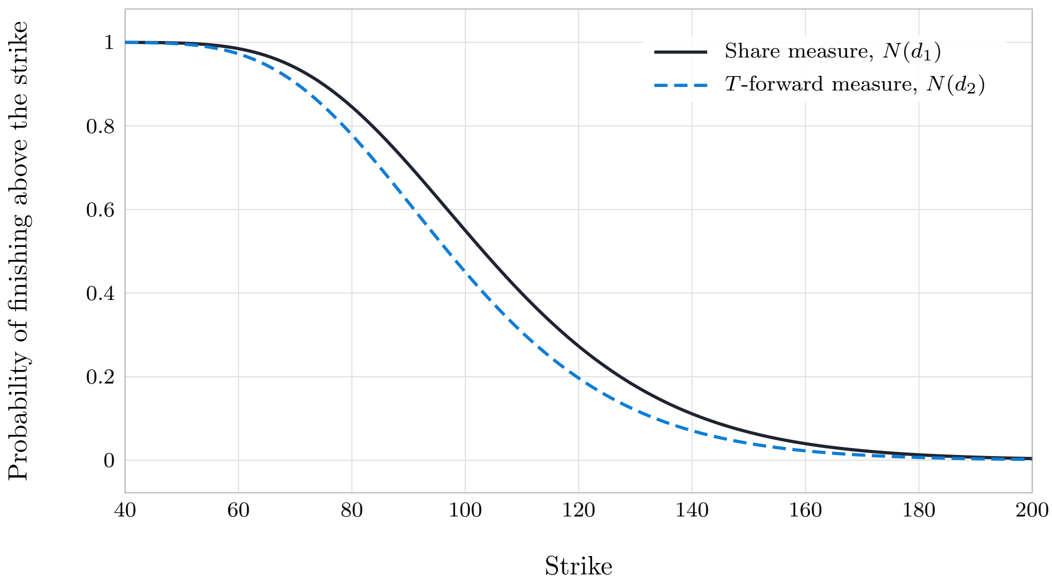

Figure 6.2: The probability that a one year option finishes in the money, at a 25% volatility, under the two measures the Black formula is assembled from. The share measure weights each path by the terminal price itself, so it counts the paths finishing high for more — and those are exactly the paths on which the option is exercised. It therefore assigns the higher probability at every strike. Both curves describe the same event; the market has not been consulted twice.
Drawn by `exercise_probabilities`.

References

- Girsanov, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability and its Applications*, 5(3), 285–301.

- Geman, H., El Karoui, N., & Rochet, J.-C. (1995). Changes of numeraire, changes of probability measure and option pricing. *Journal of Applied Probability*, 32(2), 443–458.

- Jamshidian, F. (1997). LIBOR and swap market models and measures. *Finance and Stochastics*, 1(4), 293–330.