Lecture Notes: Class 2

SDE stuff

Consider the following general stochastic differential/integral equation:

$$
\begin{aligned}
dX_t &= \mu(t, X_t)dt + \sigma\left(t, X_t\right) dB_t \text{ or} \quad (1)\\
X_t &= X_0 + \int_0^t \mu\left(s, X_s\right) ds + \int_0^t \sigma\left(s, X_s\right) dB_s.
\end{aligned}
$$

This going to be our standard workhorse model, both for asset prices and state variables (e.g. interest rates) whose dynamics determine the asset prices. Before getting to some finance, we have some more background material to develop: tools we'll need down the road.

A. Solutions to SDEs: Existence and Uniqueness

Definition: Existence: X_t is called a **strong solution** to the SDE

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t \quad (2)$$

on $(\Omega, \mathcal{F}, \mathbb{P})$ with respect to B_t and X_0 and for all $0 \leq t < \infty$ if

1. $\{X_t\}_{t\geq 0}$ has continuous sample paths
2. X_t is adapted to $\mathcal{F}_t^B$
3. $\mathbb{P}\left[\left\{\int_0^T \left[|\mu\left(s\right)| + \sigma^2\left(s\right)\right] ds\right\} < \infty\right] = 1$ and
4. X_t satisfies the integral equation:

$$X_t = X_0 + \int_0^t \mu\left(s, X_s\right) ds + \int_0^t \sigma\left(s, X_s\right) dB_s$$

Definition: Uniqueness Suppose that there are two strong solutions to the SDE, equation 2. If

$$\mathbb{P}\left[X_t = \widetilde{X}_t \text{ for } 0 \leq t < \infty\right],$$

then **strong uniqueness** holds for the SDE.

The following theorem gives us conditions under which a unique solution exists:

Theorem 1 *If the coefficients μ and σ satisfy global Lipschitz and linear growth conditions,*

$$\begin{aligned} \left|\mu\left(t,x\right)-\mu\left(t,y\right)\right|^2+\left|\sigma\left(t,x\right)-\sigma\left(t,y\right)\right|^2 &\leq k^2\left|x-y\right|^2 \\ \left|\mu\left(t,x\right)\right|^2+\left|\sigma\left(t,x\right)\right|^2 &\leq k^2\left(1+\left|x\right|^2\right), \end{aligned}$$

and $X_0 = x_0$ is a nonrandom initial condition, then there exists a unique strong solution to (2) *that is uniformly bounded in $L^2(\mathbb{P})$.*

- Recall from ODE's that Lipshitz guarantees uniqueness but not existence. What sorts of functions satisfy these conditions? If f is continuously differentiable on [a,b] then it is Lipshitz. Bounded functions satisfy growth. There are weaker conditions for this (local growth and Lipshitz).
- Why does this matter? When writing down specific models, we need to be careful to make sure what we write down is well defined. There a noted paper that specified that the diffusion coefficient was equal to $\sigma(t, X_t) = \exp(\beta X_t)$, where X_t is a geometric Brownian motion. Oops.
- I've run into this problem in my own research. In my dissertation, I built non-parametric models for interest rates. The resulting estimates indicated the diffusion coefficient seemed to grow at squared rate. When doing simulations, I had fit a polynomial function (since

the estimates were nonparametric, to do simulations I needed to extrapolate off the support) to the estimated diffusion that contained cubic and quartic terms. Turns out, the resulting SDEs didn't have a solution, though when simulating everything looked ok.

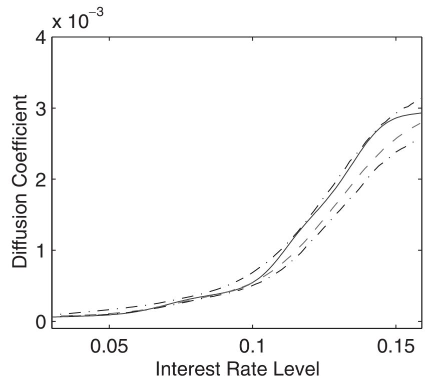

Figure: estimated diffusion coefficient from T-Bill data

- If we add a bound on the initial condition, we get bounds on higher moments also: If $E\left[(X_0)^{2m}\right] < \infty$, then there exists a C (depending only on m, k and T) for which

$$E\left[(X_t)^{2m}\right] \leq E\left[1 + (X_0)^{2m}\right] e^{Ct}.$$

  This is useful when dealing with nonlinear payoffs (like options or other derivatives) where you might need to insure that third or fourth moments are finite, to insure that things like skewness or kurtosis are finite.

Another key is that not only are the solutions continuous, the solutions to standard SDEs are Markov processes.

Definition: Markov process A stochastic process $\{X_t\}_{t\geq 0}$ adapted to a filtration $\mathcal{F}_t$ is a **Markov process** if the conditional distribution of X$_t$ given $\mathcal{F}_s$ is equal to the conditional distribution of X$_t$ given X$_s$, that is,

$$\mathbb{P}\left(X_t \leq y|\mathcal{F}_s\right) = \mathbb{P}\left(X_t \leq y|X_s\right).$$

If a process is Markov, then

$$\mathbb{E}^{\mathbb{P}}\Big[f(t, X_t)|\mathcal{F}_s\Big] = \mathbb{E}^{\mathbb{P}}\Big[f(t, X_t)|X_s\Big] = g(t, X_s),$$

which means the expectation is ONLY a function of the Markov state and not the entire path. This makes calculations feasible. This is a dimension reduction tool: the current state is a sufficient statistic for the entire historicaly path.

Theorem 2 *If the drift and diffusion coefficients in the SDE*

$$dX_t = \mu\left(t, X_t\right) dt + \sigma\left(t, X_t\right) dB_t$$

satisfy conditions for the existence and uniqueness of a strong solution, then the solution X_t *is Markov with respect to* $\mathcal{F}_t = \sigma\left(\{X_s\}_{s \leq t}\right)$.

B. Ito's lemma

The primary tool we are going to use is Ito's lemma. This allows to evaluate the evolution of functions of Brownian motion and solutions to stochastic differential equations. Here is the general diffusion version:

Ito's Lemma: If $g : R_+ \times R \to R$ is twice continuously differentiable, then for

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t,$$

we have that

$$\begin{aligned} dg(t, X_t) = &\left[\frac{\partial g(t, X_t)}{\partial t} + \frac{\partial g(t, X_t)}{\partial X_t}\mu(t, X_t) + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 g(t, X_t)}{\partial X_t^2}\right] dt \\ &+ \sigma(t, X_t)\frac{\partial g(t, X_t)}{\partial X_t} dB_t. \end{aligned}$$

Or suppressing the arguments of g

$$dg = \left[\frac{\partial g}{\partial t} + \frac{\partial g}{\partial X_t}\mu(t, X_t) + \frac{1}{2}\sigma^2(t, X_t)\frac{\partial^2 g}{\partial X_t^2}\right] dt + \sigma(t, X_t)\frac{\partial g}{\partial X_t} dB_t.$$

There are many interesting examples and applications but before getting too deep, I want to highlight the main difference between Ito calculus and regular calculus, and then re-state Ito's lemma in maybe a more easily usable and memorable way.

- First, what is the difference between Ito and regular calculus? For this we can look at a simple example. Using the rules of normal calculus, we have the indefinite integral

  $$\int x dx = \frac{x^2}{2}.$$

  This is one of the 'standard' results. For stochastic calculus, we have that ($B_0 = 0$)

  $$\int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{t}{2},$$

which has an extra 't' term. This is the added bit from the Brownian motion, an extra term.

- There is another way to write Ito's lemma in terms of dX_t which is useful. Express Ito's lemma as

$$dg = \left[g_t + \frac{1}{2}\sigma^2(t, X_t)g_{xx}\right]dt + g_x dX_t.$$

Substituting in for dX_t, we have the same version as above

$$\begin{aligned} dg &= \left[g_t + \frac{1}{2}\sigma^2(t, X_t)g_{xx}\right]dt + g_x\left[\mu(t, X_t)dt + \sigma(t, X_t)dB_t\right] \\ &= \left[g_t + g_x\mu(t, X_t)dt + \frac{1}{2}\sigma^2(t, X_t)g_{xx}\right]dt + g_x\sigma(t, X_t)dB_t \end{aligned}$$

- There is another way to Ito's lemma, using some simple, ad-hoc, rules. The basic idea is just a Taylor series expansion of a total derivative:

$$df(t, X_t) = f_t dt + \frac{1}{2}f_{tt}(dt)^2 + f_x dX_t + \frac{1}{2}f_{xx}(dX_t)^2 + \frac{1}{2}f_{tx}(dt dX_t) + ...$$

How to, heuristically, calculate things like $dt \cdot dt$? Let's use the following multiplication rules:

$\cdot$	dt	dB_t
dt	0	0
dB_t	0	dt

Armed with these rules, we have

$$\begin{aligned} df &= f_t dt + \frac{1}{2}f_{tt}\cdot 0 + f_x dX_t + \frac{1}{2}f_{xx}\sigma^2(t, X_t)dt + \frac{1}{2}f_{tx}\cdot 0 \\ &= f_t dt + f_x dX_t + \frac{1}{2}f_{xx}\sigma^2(t, X_t)dt \\ &= \left[f_t + \mu(t, X_t)dt + \frac{1}{2}\sigma^2(t, X_t)f_{xx}\right]dt + f_x\sigma(t, X_t)dB_t, \end{aligned}$$

which "works" in that it gives the right answer. The key component in the calculation is that $dX_t \cdot dX_t = \sigma^2(t, X_t)dt$.

$\boxed{\text{Examples of Ito's lemma}}$

- Geometric Brownian motion

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

Consider the function $g(X_t) = \ln(X_t)$. The derivatives are

$$g_t(t,x) = 0,\ g_x(t,x) = \frac{1}{x},\ \text{and}\ g_{xx}(t,x) = -\frac{1}{x^2} \tag{3}$$

This implies that:

$$\begin{aligned} d\ln(X_t) &= \left[g_x(X_t)\mu(X_t) + \frac{1}{2}\sigma^2(X_t) g_{xx}(X_t) \right] dt + \sigma(X_t) g_x(X_t) dB_t \\ &= \left[\frac{1}{X_t}\mu X_t - \frac{1}{2}\sigma^2 X_t^2 \frac{1}{X_t^2} \right] dt + \sigma X_t \frac{1}{X_t} dB_t \\ &= \left[\mu - \frac{1}{2}\sigma^2 \right] dt + \sigma dB_t. \end{aligned}$$

We can also use the 'box algebra' to get the right answer:

$$\begin{aligned} d\ln(X_t) &= g_x(X_t) dX_t + \frac{1}{2} g_{xx} dX_t dX_t \\ &= \frac{1}{X_t}\mu X_t dt + \sigma X_t \frac{1}{X_t} dB_t - \frac{1}{2}\frac{1}{X_t^2}(\sigma X_t dB_t)(\sigma X_t dB_t) \\ &= \left[\mu - \frac{1}{2}\sigma^2 \right] dt + \sigma dB_t. \end{aligned}$$

Thus, $\ln(X_t) = \ln(X_0) + (\mu - \frac{1}{2}\sigma^2)t + \sigma B_t$, or

$$X_t = X_0 \exp\Big((\mu - \frac{1}{2}\sigma^2)t + \sigma B_t \Big). \tag{4}$$

This is an explicit 'solution' of the SDE in terms of t and B_t.

- Brownian stochastic exponentials. Consider the function $M_t = g(t, B_t) = exp(B_t - \frac{1}{2}t)$. By direct calculation, for a Brownian motion, it is clear that $E^{\mathbb{P}}[M_t] = 1$. To see what Ito's lemma has to say, compute the derivatives. The time derivative is

$$\begin{aligned} g_t\left(t, B_t\right) &= \frac{\partial exp(B_t - \frac{1}{2}t)}{\partial t} = exp(B_t)\frac{\partial exp(-\frac{1}{2}t)}{\partial t} \\ &= exp\Big(B_t - \frac{1}{2}t\Big)(-\frac{1}{2}) \\ &= -\frac{g\left(t, B_t\right)}{2} \end{aligned} \tag{5}$$

  And the other partials are

$$g_x(t, x) = g\left(t, x\right) \text{ and } g_{xx}(t, x) = g(t, x).$$

  This implies that

$$\begin{aligned} dg\left(t, B_t\right) &= g(t, B_t)dB_t + \left[-\frac{1}{2}g\left(t, B_t\right) + \frac{1}{2}g(t, B_t)\right]dt \\ &= g(t, B_t)dB_t. \end{aligned} \tag{6}$$

  Or that

$$M_t = exp\Big(B_t - \frac{t}{2}\Big) = 1 + \int_0^t g(s, B_s)dB_s.$$

  It should be clear that g is a martingale. This is a specific case of a more general result: any well behaved integral against a Brownian motion is a martingale.

Interestingly, we appear to have 'solved' the stochastic differential equation. This last application of Ito's lemma you will literally do hundreds of times this term.

C. Solving Stochastic Differential Equations

We will now discuss SDE solutions a bit more formally. X_t is an analytical or closed form solution to

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dB_t$$

if there exists an analytical function f such that $X_t = f(B_t, B_0, t)$. There are only a handful of examples that can be solved.

Example 1 *Geometric Brownian Motion. Consider the following SDE:*

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

We are going to try and solve this by guessing a solution of the form $X_t = f(t, B_t)$ *and then match coefficients. Assuming that* f *is twice continuously differentiable, we have by Ito's lemma that*

$$dX_t = \left[f_t(t, B_t) + \frac{1}{2} f_{xx}(t, B_t)\right] dt + f_x(t, B_t)\, dB_t.$$

By matching the drift and diffusion coefficients, we assume that

$$\sigma f(t, x) = f_x(t, x) \quad \textit{and}$$
$$\mu f(t, x) = f_t(t, x) + \frac{1}{2} f_{xx}(t, x).$$

The first equation implies that $\frac{f_x}{f} = \sigma$ *a constant and the solution to this ordinary differential equation is*

$$f(t, x) = \exp(\sigma x + g(t))$$

where $g(t)$ *is an arbitrary function. Plugging this into the second equation, we have that*

$$\mu f(t, x) = f(t, x)\, g'(t) + \frac{1}{2} f(t, x)\, \sigma^2$$

which implies that $g\prime(t) = \mu - \frac{1}{2}\sigma^2$ *(by canceling the* f*'s). Thus the solution is given by:*

$$X_t = f(t, B_t) = \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma B_t\right).$$

This is extremely rare to be able to solve an SDE in closed form. There is one other case where you can come close to analytical solutions.

Example 2 *Ornstein-Uhlenbeck Process. Consider the following SDE:*

$$dX_t = -\alpha X_t dt + \sigma dB_t.$$

We guess a solution of the form:

$$X_t = a(t)\left[x_0 + \int_0^t b(s)\, dB_s\right]$$

and by Ito's lemma, we have that

$$\begin{aligned} dX_t &= a'(t)\left[x_0 + \int_0^t b(s)\, dB_s\right] dt + a(t)\, b(t)\, dB_t \\ &= \frac{a'(t)}{a(t)} X_t dt + a(t)\, b(t)\, dB_t \end{aligned}$$

where we've assumed that $a(t) > 0$ and $a(0) = 1$. Now we can match coefficients:

$$-\alpha = \frac{a'(t)}{a(t)} \text{ and } \sigma = a(t)\, b(t).$$

The first equation is solved by $a(t) = e^{-\alpha t}$ and the second is then $b(t) = \sigma e^{at}$. This implies that

$$\begin{aligned} X_t &= a(t)\left[x_0 + \int_0^t b(s)\, dB_s\right] \\ &= e^{-\alpha t}\left[x_0 + \int_0^t \sigma e^{as} dB_s\right] \\ &= e^{-\alpha t} x_0 + \int_0^t \sigma e^{-\alpha(t-s)} dB_s. \end{aligned}$$

This is not closed form, but it is close. The same method allows us to solve

$$dX_t = \kappa(\theta - X_t)\, dt + \sigma dB_t$$

in closed form. Define $Y_t = X_t - \theta$ which implies that $dY_t = dX_t$ and

$$dY_t = -\kappa Y_t dt + \sigma dB_t.$$

By Ito's lemma,

$$
\begin{aligned}
d\left(e^{\kappa t} Y_t\right) &= \left(\kappa e^{\kappa t} Y_t + e^{kt} dY_t\right) \\
&= \left[\kappa e^{\kappa t} Y_t + -\kappa e^{kt} Y_t\right] dt + \sigma e^{kt} dB_t \\
&= \sigma e^{kt} dB_t
\end{aligned}
$$

which can be solved for

$$
\begin{aligned}
e^{\kappa t} Y_t &= Y_0 + \int_0^t \sigma e^{\kappa s} dB_s \\
Y_t &= e^{-\kappa t} Y_0 + \int_0^t \sigma e^{-\kappa(t-s)} dB_s
\end{aligned}
$$

and

$$
\begin{aligned}
X_t &= \theta + e^{-\kappa t}\left(X_0 - \theta\right) + \int_0^t \sigma e^{-\kappa(t-s)} dB_s \\
&= \theta\left(1 - e^{-\kappa t}\right) + e^{-\kappa t} X_0 + \int_0^t \sigma e^{-\kappa(t-s)} dB_s.
\end{aligned}
$$

The geometric Brownian motion and Ornstein-Uhlenbeck are about the only processes that can be computed in closed form. After that, one must use other methods. Like simulation.

D. Simulating Stochastic Differential Equationsbigskip

How to simulate a stochastic differential equation:

$$dX_t = \mu\left(t, X_t\right) dt + \sigma\left(t, X_t\right) dB_t?$$

The first algorithm is known as an Euler approximation. For a given partition, $t = t_0 < t_1 < ..., t_N = T$, define the discrete-time stochastic process

$$Y_{t_{n+1}} = Y_{t_n} + \mu\left(t_n, Y_{t_n}\right)\left(t_{n+1} - t_n\right) + \sigma\left(t_n, Y_{t_n}\right)\left(B_{t_{n+1}} - B_{t_n}\right)$$

where $Y_{n+1} \triangleq Y\left(t_{n+1}\right)$. Let $\Delta_n = t_{n+1} - t_n$ and let $\delta = \max \Delta_n$. It is easiest to assume that the discretization points are equally spaced, that is, $\tau_n = t_{n-1} + \Delta$ where $\Delta = \frac{T-t}{N}$. In this case, we have that

$$Y_{t_{n+1}} = Y_{t_n} + \mu\left(t_n, Y_{t_n}\right)\Delta + \sigma\left(t_n, Y_{t_n}\right)\left(B_{t_n+\Delta} - B_{t_n}\right).$$

Not surprisingly, as $\Delta \to 0$, we have that $Y_{t_n} \to X_{t_n}$ in an appropriate manner. In some cases, you can find explicit bounds on the speed of convergence.

The intuition for these approximations is simple. The solution to the SDE is given by:

$$X_{t+1} = X_t + \int_t^{t+1} \mu\left(s, X_s\right) ds + \int_t^{t+1} \sigma\left(s, X_s\right) dB_s.$$

By the definition of the integrals, we have that

$$\int_t^{t+1} \sigma\left(s, X_s\right) dB_s = \lim \sum_{j=0}^{n-1} \sigma\left(t_j, X_{t_j}\right)\left(B_{t_{j+1}} - B_{t_j}\right) \text{ and}$$
$$\int_t^{t+1} \mu\left(s, X_s\right) ds = \lim \sum_{j=0}^{n-1} \mu\left(t_j, X_{t_j}\right)\left(t_{j+1} - t_j\right)$$

so it is natural to approximate these as:

$$\tilde{X}_{t+1} = \tilde{X}_t + \sum_{j=0}^{n-1} \mu\left(t_j, \tilde{X}_{t_j}\right)\left(t_{j+1} - t_j\right) + \sum_{j=0}^{n-1} \sigma\left(t_j, \tilde{X}_{t_j}\right)\left(B_{t_{j+1}} - B_{t_j}\right).$$

If we take the smallest discretization interval, Δ, we have that

$$\tilde{X}_{t+\Delta} = X_t + \int_t^{t+\Delta} \mu\left(s, X_s\right) ds + \int_t^{t+\Delta} \sigma\left(s, X_s\right) dB_s$$

is approximated by

$$\tilde{X}_{t+\Delta} = \tilde{X}_t + \mu\left(t, \tilde{X}_t\right) \Delta + \sigma\left(t, \tilde{X}_t\right)\left(B_{t+\Delta} - B_t\right).$$

When will this be an accurate approximation? For example, if μ and σ are constant, we have that $\int_t^{t+\Delta} \mu ds = \mu\Delta$ and $\int_t^{t+\Delta} \sigma dB_s = \sigma\left(B_{t+\Delta} - B_t\right)$ and there is no discretization error. In general, the smoother (in terms of derivatives) the drift and diffusion coefficient are, the more accurate the Euler approximations.

These approximations provide us with a way to operationalize the SDE's by simulating paths and computing moments of functionals. In fact, in many settings, the only feasible way to characterize the solutions is via simulation.

E. Connections between SDE's and PDE's

Consider first the case of a univarite diffusion model where we assume that X_t is the strong solution to

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dB_t.$$

We are going to want to calculate discounting expected values of the form,

$$f(t, X_t) = E_t^{\mathbb{P}}\left[e^{-\int_t^T r_s ds} g(X_T) | X_t\right],$$

which is a general representation of the value of an asset that have a payoff of $g(X_T)$ at time T. THis is just the usual discounted present value formula. It turns out that expectations of this form, which are Markov, solve partial differential equations. The following theorems connects expectations of payoffs with PDEs, getting a one step closer to the Black-Scholes formula. The first result has no discounting, the second has discouning, and the third additionally dividends.

Theorem 3 (Dynkin's Formula) *If f is twice continuously differentiable, $\frac{\partial f}{\partial x}$ is bounded and f satisfies the partial differential equation (for all t,x)*

$$f_t(t, x) + f_x(t, x)\, \mu(t, x) + \frac{1}{2}\sigma^2(t, x)\, f_{xx}(t, x) = 0$$

with boundary condition $f(x, T) = g(x)$, then

$$f(x, t) = E_t^{\mathbb{P}}[g(X_T) | X_t = x].$$

Proof. Apply Ito's lemma to the function f:

$$df(t, X_t) = \left[f_t + f_{xx}\mu(t, X_t) + \frac{1}{2}\sigma^2(t, X_t)\, f_{xx}\right] dt + f_x\sigma(t, X_t)\, dB_t.$$

By assumption the portion in brackets is equal to zero and we have that

$$g(X_T) = f(T, X_T) = f(t, X_t) + \int_t^T f_x(t, X_t)\, \sigma(t, X_t)\, dB_t$$

and taking expectations (since f_x is bounded), we have that

$$E_t^{\mathbb{P}}\left[g\left(X_T\right)|X_t\right] = f\left(t, X_t\right)$$

which is the result. ∎

Can this be extended? Yes, the first of which introduces discounting.

Theorem 4 (Feynman-Kac formula) *If f is twice continuously differentiable, $\frac{\partial f}{\partial x}$ is bounded and f satisfies the partial differential equation (for all t,x)*

$$f_t\left(t,x\right) + f_x\left(t,x\right)\mu\left(t,x\right) + \frac{1}{2}\sigma^2\left(t,x\right) f_{xx}\left(t,x\right) = r\left(x,t\right) f\left(x,t\right)$$

with boundary condition $f\left(x,T\right) = g\left(x\right)$, then

$$f\left(t,x\right) = E_t^{\mathbb{P}}\left[e^{-\int_t^T r(X_s,s)ds} g\left(X_T\right)|X_t = x\right].$$

The proof of this follows the proof of Dynkin's formula. Just apply Ito's lemma to the function, apply the condition in theorem, and integrate.

Proof. Apply Ito's lemma to the function $Y_s = f\left(s, X_s\right) e^{-\int_t^s r(X_u,u)du}$. This implies that

$$\begin{aligned}
dY_s &= \left[e^{-\int_t^s r(X_u,u)du}\right] df\left(s,X_s\right) + f\left(s,X_s\right) d\left[e^{-\int_t^s r(X_u,u)du}\right] \\
&= e^{-\int_t^s r(X_u,u)du}\left[\mu^f\left(s,X_s\right) ds + f_x\sigma\left(s,X_s\right) dB_s\right] - r\left(X_s,s\right) f\left(s,X_s\right) e^{-\int_t^s r(X_u,u)du} ds \\
&= e^{-\int_t^s r(X_u,u)du}\left\{\left(\mu^f\left(s,X_s\right) - r\left(X_s,s\right) f\left(X_s,s\right)\right) dt + f_x\left(t,X_s\right)\sigma\left(t,X_s\right) dB_s\right\}
\end{aligned}$$

where $\mu^f\left(t,X_t\right) = \left[f_t + f_x\mu\left(t,X_t\right) + \frac{1}{2}\sigma^2\left(t,X_t\right) f_{xx}\right]$. By assumption, the expression in parentheses is zero which implies that

$$dY_s = e^{-\int_t^s r(X_u,u)du} f_x\left(s,X_s\right)\sigma\left(s,X_s\right) dB_s.$$

Integrating, we have

$$Y_T = Y_t + \int_t^T e^{-\int_t^s r(X_u,u)du} f_x\left(s,X_s\right)\sigma\left(s,X_s\right) dB_s$$

and taking expectations implies that

$$f(t,x) = E_t^{\mathbb{P}}\left[e^{-\int_t^T r(X_s,s)ds} f(T, X_T) | X_t = x\right] = E_t^{\mathbb{P}}\left[e^{-\int_t^T r(X_s,s)ds} g(X_T) | X_t = x\right],$$

Since $Y_t = f(t, X_t)$.

■

One final extension (which is useful for dividends)

Theorem 5 *If f is twice continuously differentiable, $\frac{\partial f}{\partial x}$ is bounded and f satisfies the partial differential equation (for all t,x)*

$$f_t(t,x) + f_x(t,x)\mu(t,x) + \frac{1}{2}\sigma^2(t,x) f_{xx}(t,x) + h(t,x) = r(x,t) f(x,t)$$

with boundary condition $f(x,T) = g(x)$, then

$$f(x,t) = E_t^{\mathbb{P}}\left[e^{-\int_t^T r(X_s,s)ds} g(X_T) + \int_t^T e^{-\int_t^u r(X_s,s)ds} h(X_u, u)\, du | X_t = x\right].$$

F. Merton's replication argument

We have so far developed the tools of stochastic calculus. We have the building blocks: Brownian Motion and the Poisson Process, stochastic integrals and differential equations, and connections between PDE's and ODE's. Now we can finally use these tools to price assets. The payoff is in Merton's beautiful arbitrage replication argument.

Trading Strategies and Portfolios

We begin with the usual probability triple, $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$, and, unless otherwise noted, all random variables are defined on this space. Assume that there exists a Brownian motion, B_t, and that $\mathcal{F}_t = \sigma\left(B_s : s \leq t\right)$, the Brownian filtration.

To start, let's assume that S_t, the asset price, is in fact the Brownian motion, $S_t = B_t$. This model implies that information is generated by the realizations of a Brownian motion.

A **trading strategy** is an adapted process θ_t specifying at each state (ω) and time t the number of shares of the security held. The adapted restriction insures that your portfolio at time t depends solely on information up to and including time t. Thus, in general we write $\theta(t, \omega)$, although we typically just write θ_t.

The simplest example of a trading strategy is the constant strategy, $\theta_t = \theta$ for all t. If at time t, θ shares are purchased at price B_t and they are later sold at time s for B_s, this generates a trading (or capital) gain of $\theta\left(B_s - B_t\right)$. If we bought and held the shares for a short time-interval from t to $t + \Delta$, we would generate a trading gain of $\theta\left(B_{t+\Delta} - B_t\right)$. This suggests that the infinitesimal gain from holding θ_t shares is something like $\theta_t dB_t$. Where have we seen this before? That is why stochastic integrals make sense.

As a second example, suppose that $\theta(t,\omega) = \theta(t_{n-1})$ for $t \in [t_{n-1}, t_n)$ where $0 = t_0 < t_1 < ... < t_N = T$. Then the capital gains over a single interval are $\theta(t_{n-1})[B_{t_n} - B_{t_n - 1}]$ and the gains over the entire interval are

$$\sum_{n=1}^{N} \theta(t_{n-1})[B_{t_n} - B_{t_n - 1}]$$

which, if you recall, is just our definition of the stochastic integral, $\int_0^T \theta_s dB_s$, when the integrand is a "simple" function.

More generally, assume that the trading strategy is well behaved, that is, $\theta \in \mathcal{L}^2$ where

$$\mathcal{L}^2 = \left\{ \theta : \theta \text{ is adapted to } \mathcal{F}_t \text{ and } \int_0^T \theta_s^2 ds < \infty \right\}$$

and define the capital gains process,

$$G_t = \int_0^t \theta_s dB_s.$$

Intuitively, we hold θ_t units of the security from time t to $t+dt$. The portfolio must be formed only based on information available at time t, thus the restriction that the portfolio be adapted to $\mathcal{F}_t$. Since the stochastic integral is defined as

$$\lim_{\Delta \to 0} \sum_{j=0}^{n-1} \theta(t_{n-1})[B_{t_n} - B_{t_n - 1}],$$

we see that the portfolio is known before the Brownian increment is revealed.

From Lecture 1, we know that stochastic integrals are linear operators: that is, if ϕ_s and η_s are in $\mathcal{L}^2$, then for any a, b we have that

$$\int_0^t [a\phi_s + b\eta_s] dB_s = a \int_0^t \phi_s dB_s + b \int_0^t \eta_s dB_s.$$

This is typically how one assumes portfolio operate: if you buy twice as many securities, your portfolio gains are twice as large.

In this model, the asset price (B_t) is a martingale with respect to $(\mathcal{F}_t, \mathbb{P})$. Does this mean our portfolio is also a martingale? Not necessarily. When is this the case? If, for example, θ is bounded, then $\int_0^t \theta_s dB_s$ is a martingale. This is the analog to the betting example we discussed in Lecture 1. Recall the definition of a martingale, a continuous-time stochastic process, $\{X_s\}_{0\leq s\leq t})$, is a martingale relative to $(\mathcal{F}_t, \mathbb{P})$ if

1. X_t is adapted to $\mathcal{F}_t$ (information)
2. $E^{\mathbb{P}}|X_t| < \infty$ (integrability)
3. $E^{\mathbb{P}}[X_t|\mathcal{F}_s] = X_s$ for all $s < t$ (for almost all ω).

To verify these conditions, recall the class of processes for which we defined the stochastic integral: we defined $\int_0^T f(t)\, dB_t$ assuming that

1. $f(\cdot,\cdot)$ is measurable in $\mathcal{F}_t \times \mathcal{B}$
2. $f(t,\cdot)$ is adapted to $\mathcal{F}_t$
3. $E^{\mathbb{P}}\left[\int_0^T f^2(\omega,t)\, dt\right] = \int\int_0^T f^2(\omega,t)\, dt d\mathbb{P}(\omega) < \infty.$

We refer to the function space that contains these functions as $H^2[0,T]$. Under these conditions, we also argued that $E\left[\int_0^T f(t)\, dB_t\right]$ was a martingale. More generally, one can show that if $\theta_t \in H^2$, that $\int_0^t \theta_s dB_s$ is a martingale and that $var\left(\int_0^t \theta_s dB_s\right) = E\left(\int_0^t \theta_s^2 ds\right)$, the Ito isometry.

Gains from Trading when Prices are Ito Processes

OK, so now we are almost there. Let's work with a single risky security, S_t, which solves

$$\begin{aligned} S_t &= S_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s \text{ or} \\ dS_t &= \mu_t dt + \sigma_t dB_t \end{aligned}$$

and a money market account, β_t which accrues interest continuously, $\beta_t = \beta_0 e^{\int_0^t r_s ds}$. Although the interest rate is potentially random, the money market is considered "locally riskless," since $d\beta_t = r_t\beta_t dt$ (note there is no dB_t term even if r_t solves a stochastic differential equation). In most of our examples will will further assume that the interest rate is constant, unless stochastic interest rates matter (fixed income section).

Example 3 *The Black-Scholes model will be our standard example throughout. This model has two assets and assumes that*

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$
$$d\beta_t = r\beta_t dt.$$

To make things easier (and clearer), we will work with strategies that do not allow for intermediate withdrawals or capital infusions. This implies all the gains are capital gains. So, for example, this doesn't allow intermediate consumption and there are no dividends. A **self-financing trading strategy** is a pair of processes, a_t and b_t, which represent the number of units of the money market and risky asset held in one's portfolio, such that

$$\pi_t = \underbrace{a_t S_t + b_t \beta_t}_{\text{Current value}} = \underbrace{a_0 S_0 + b_0 \beta_0}_{\text{Initial investment}} + \underbrace{\int_0^t a_s dS_s + \int_0^t b_s d\beta_t}_{\text{Capital gains}}$$

In the general case, we have that $dS_t = \mu_t dt + \sigma_t dB_t$ which implies (via linearity) that

$$a_t S_t + b_t \beta_t = \pi_0 + \int_0^t a_s \mu_s ds + \int_0^t a_s \sigma_s dB_s + \int_0^t b_s r_s \beta_s ds.$$

So now, for a given self-financing trading strategy in the stock and the bond, we know to represent the portfolio gains.

Merton's replication argument justifying Black-Scholes

Given the previous machinery, we can now use a self-financing strategy to price an option, ala Merton's no-arbitrage argument (1974, A Rational Theory of Option Pricing). This, to me, is the most fundamental asset pricing insight and argument. Assume that we have two assets to invest in, a risky asset following geometric Brownian motion, and a money market account (or bond) with a constant interest rate. Our goal is price a European call option, whose terminal payout is $Y_T = \max(0, S_T - K) = (S_T - K)^+$. That is, we want to find the price today, at time 0.

The argument relies on a couple of steps: (1) assume that the option is a smooth function of the underlying and time-to-maturity; (2) find a self-financing strategy in the stock on money market account such that $Y_T = a_T S_T + b_T \beta_T$ (we do this by guess and verify); (3) and then invoke no-arbitrage to show that $Y_0 = a_0 S_0 + b_0 \beta_0$.

So in simple steps, here is the argument:

1. Assume that $Y_t = C(S_t, t)$ where C is twice continuously-differentiable. This means we can apply Ito's lemma. This is not necessarily obvious and some math guys argue that this is an unjustified assumption as you are effectively assuming the answer. However, in the end, everyone agrees that this does hold, so we will ignore the technicalities.

2. Use Ito's lemma to characterize the dynamics of the option price:

$$dY_t = \mu_y(t)\, dt + \sigma_y(t)\, dB_t \tag{*}$$

   where

$$\begin{aligned} \mu_y(t) &= \frac{\partial C}{\partial t} + \mu S_t \frac{\partial C}{\partial S} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2} \\ \sigma_y(t) &= \sigma S_t \frac{\partial C}{\partial S}. \end{aligned}$$

3. Assume there exists a self-financing strategy (a, b) such that $Y_t = a_t S_t + b_t \beta_t$. If there is, then

$$\begin{aligned} dY_t &= a_t dS_t + b_t d\beta_t \\ &= [a_t \mu S_t + b_t \beta_t r]\, dt + a_t \sigma S_t dB_t. \end{aligned} \tag{$\dagger$}$$

   At this stage, why isn't the case that $dY_t = d(a_t S_t) + d(b_t \beta_t) = S_t da_t + a_t dS_t + ...$?

4. To find (a, b), match drift and volatility coefficients in $(*)$ and $(\dagger)$. Matching drifts gives

$$\frac{\partial C}{\partial t} + \mu S_t \frac{\partial C}{\partial S} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2} = a_t \mu S_t + b_t \beta_t r$$

   and matching diffusions gives

$$\sigma S_t \frac{\partial C}{\partial S} = a_t \sigma S_t.$$

   Solving the equation from equating diffusion coefficients,

$$a_t = \frac{\partial C(S_t, t)}{\partial S}.$$

   This term is often referred to as the option's Δ. It gives the number of shares of stock that must be bought or sold to replicate the option. Is this positive or negative? This "coefficient matching" procedure is justified in Duffie's textbook.

5. How to find b_t? Since $Y_t = a_t S_t + b_t \beta_t$, we have b_t must satisfy

$$b_t = \frac{Y_t - \frac{\partial C}{\partial S} S_t}{\beta_t} = \frac{C_t - \frac{\partial C}{\partial S} S_t}{\beta_t}.$$

6. Now, given (a, b), we can substitute them into

$$\frac{\partial C}{\partial t} + \mu S_t \frac{\partial C}{\partial S} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2} = a_t \mu S_t + b_t \beta_t r$$

   to get

$$\frac{\partial C}{\partial t} + \boxed{\mu S_t \tfrac{\partial C}{\partial S}} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2} = \boxed{\tfrac{\partial C}{\partial S} \mu S_t} + \frac{C_t - \frac{\partial C}{\partial S} S_t}{\boxed{\beta_t}} \boxed{\beta_t} r.$$

Canceling the boxed terms and re-arranging, we have that

$$rC_t = \frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2}$$

which is the Black-Scholes PDE. With the boundary condition, $C_T = \max(0, S_T - K)$ we get the familiar Black-Scholes formula, which you verified in Homework #1. On your next homework, you will verify that the portfolio strategy is in fact, self-financing.

7. Why is this the no-arbitrage price? Suppose that $Y_0 > a_0 S_0 + b_0 \beta_0$, i.e., the option is more expensive than the replicating strategy. Then, we can sell one option contract for Y_0, enter into the trading strategy (a, b) proscribed above. The initial profit is positive: $Y_0 - a_0 S_0 + b_0 \beta_0 > 0$, $Y_T = C_T$, and the strategy is self-financing which means that no intermediate cash inflow is needed. Thus there is an arbitrage. A similar argument holds for $Y_0 < a_0 S_0 + b_0 \beta_0$, which shows that in fact, $Y_0 = a_0 S_0 + b_0 \beta_0 = C_0$

This was the argument that convinced the world that Black and Scholes were right. Recall their argument: they applied a continuous-time version of the CAPM and assumed the 'market' was the underlying stock and the asset being priced was the option. Somehow, despite the nonlinear payoffs of the option and the other seemingly counterfactual assumptions, they got the right answer. It was Merton's argument, however, that convinced everyone it was, in fact, correct.

The Feynman-Kac approach to Black-Scholes

Earlier, we connected PDE's with expectations. Recall Feynman-Kac: if

$$\boxed{r_t} C(x,t) = \frac{\partial C(x,t)}{\partial t} + \underline{\mu_t} \frac{\partial C(x,t)}{\partial x} + \underline{\sigma_t^2} \frac{1}{2} \frac{\partial^2 C(x,t)}{\partial x^2}$$

then

$$C(x,t) = E_{t,x}\left[e^{-\int_t^T \boxed{r_s} ds} g(X_T) \right]$$

and the dynamics of X_t are given by:

$$dX_t = \underline{\mu_t} dt + \underline{\sigma_t} dB_t.$$

This shows the connection between discounting, coefficients in parabolic PDE's and coefficients in SDE's.

In the case of Black-Scholes, we know that

$$\begin{aligned} rC_t &= \frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2} \\ C(S_T, T) &= g(S_T) \end{aligned}$$

which, with a little bit of creativity, suggests the following:

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_t dB_t \\ C(S_t, t) &= E\left[e^{-r(T-t)} g(S_T)\right]. \end{aligned}$$

NOtice that this isn't the original specification,

$$dS_t = \mu S_t dt + \sigma S_t dB_t.$$

It appears for the Feynman-Kac version to work, we need to have the drift equal to r. When would this be the case? If all (or the marginal investor) investors were risk-neutral, this would hold.

The interpretation of this comes from Cox and Ross (1976, JFE). In a risk-neutral world, all assets have the same expected return, the risk-free rate. Thus, if all investors were risk-neutral, we have that

$$E_t\left[\frac{dS_t}{S_t}\right] = rdt$$

which implies that

$$dS_t = rS_t dt + \sigma S_t dB_t$$

assuming volatility is constant. If this is the evolution for the stock price, then we have that

$$dC_t = \left[\frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S} + \frac{\sigma^2 S_t^2}{2} \frac{\partial^2 C}{\partial S^2}\right] dt + \sigma S_t \frac{\partial C}{\partial S} dB_t.$$

Using the Black-Scholes PDE to substitute for the drift, we have that

$$dC_t = rC_t dt + \sigma S_t \frac{\partial C}{\partial S} dB_t$$

which heuristically shows that

$$E_t \left[\frac{dC_t}{C_t} \right] = rdt.$$

Thus, it appears, that in the absence of arbitrage, all assets pay r. Is it actually the case that all assets pay the risk-free rate? No, however the result suggests that we can price this way. That is, when there is no arbitrage, replace the stock price drift by r and then price using the rule that all derivatives earn the risk-free rate. Cox and Ross (1976) suggest that this is the way pricing should be done - - replace μ with r and proceed as if the agents are risk-neutral. Why? It gets the right answer. That was not satisfying.

It turns out that there is a elegant argument explaining what is going on in the background. The short answer is that the two Brownian motions are different in the sense that they are defined under different probability measures. Thus, we next turn to the seminal papers of Harrison and Kreps (1978) and Harrison and Pliska (1981).