Statistics of High Frequency Financial Data
FINM 33170 and STAT 33910

Per Mykland

Chicago, Winter 2021

Outline

Some Highlights of Stochastic Calculus

Information Sets, σ-fields

- Basic space: $(\Omega, \mathcal{F})$
- Ω is the set of all possible outcomes ω
- $\mathcal{F}$ is the collection of subsets $A \subseteq \Omega$ that will eventually be decidable (it will be observed whether they occured or not)

A collection $\mathcal{A}$ of subsets of Ω is a σ-*field* (sets that are decidable by a procedure of obseration) if

- $\emptyset, \Omega \in \mathcal{A}$: $\emptyset, \Omega$ are decidable ($\emptyset$ didn't occur, Ω did)
- if $A \in \mathcal{A}$, then $A^c = \Omega - A \in \mathcal{A}$: if A occurs, then A^c does not occur, and *vice versa*
- if $A_n, n = 1, 2, ...$ are all in $\mathcal{A}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$: if all the A_n are decidable, then so is the event $\cup_{n=1}^{\infty} A_n$ (the union occurs if and only if at least one of the A_i occurs)

Measurability of a Random Variable

- All random variables are thought to be a function of the basic outcome $\omega \in \Omega$:
  X is of the form $X(\omega)$
- A random variable X is called $\mathcal{A}$-*measurable* if the value of X can be decided on the basis of the information in $\mathcal{A}$.
  Formally, the requirement is that for all x, the set $\{X \leq x\} = \{\omega \in \Omega \ : \ X(\omega) \leq x\}$ be decidable ($\in \mathcal{A}$).
- The σ-field that has the same information as a random variable: $\mathcal{F}(X)$ is the smallest *sigma*-field making X measurable
- $\mathcal{F}(X)$ is the intersection of all σ-fields $\mathcal{A}$ so that X called $\mathcal{A}$-measurable. Exercise: verify that this is indeed a σ-field
- Ω and ω are usually “anonymous", but tie together the random variables in the same system

Evolution of Knowledge over Time

- The evolution of knowledge in our system is described by the *filtration* (or sequence of σ-fields) $\mathcal{F}_t$, $0 \leq t \leq T$. Here $\mathcal{F}_t$ is the knowledge available at time t. Since increasing time makes more sets decidable, the family $(\mathcal{F}_t)$ is taken to satisfy that if $s \leq t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$.
- Most processes will be taken to be *adapted* to $(\mathcal{F}_t)$: (X_t) is adapted to $(\mathcal{F}_t)$ if for all $t \in [0, T]$, X_t is $\mathcal{F}_t$-measurable. A vector process is adapted if each component is adapted.
- We define the filtration $(\mathcal{F}_t^X)$ *generated* by the process (X_t) as the smallest filtration to which X_t is adapted. By this we mean that for any filtration $\mathcal{F}'_t$ to which (X_t) is adapted, $\mathcal{F}_t^X \subseteq \mathcal{F}'_t$ for all t.

Wiener Processes

A Wiener process is Brownian motion relative to a filtration. Specifically,
The process $(W_t)_{0\leq t\leq T}$ is an $(\mathcal{F}_t)$-Wiener process if it is adpted to $(\mathcal{F}_t)$ and

- $W_0 = 0$;
- $t \rightarrow W_t$ is a continuous function of t;
- W has independent increments relative to the filtration $(\mathcal{F}_t)$: if $t > s$, then $W_t - W_s$ is independent of $\mathcal{F}_s$;
- for $t > s$, $W_t - W_s$ is normal with mean zero and variance $t - s$ (N(0,t-s)).

Brownian motion (W_t) is an $(\mathcal{F}_t^W)$-Wiener process.

Predictable Processes

For stochastic integrals, need concept of *predictable process*. "Predictable": one can forecast the value over infinitesimal time intervals. Basic example: "simple process", defined by break points $0 = s_0 = t_0 \leq s_1 < t_1 \leq s_2 < t_2 < \ldots \leq s_n < t_n \leq T$, and random variables $H^{(i)}$, observable (measurable) with respect to $\mathcal{F}_{s_i}$.

$$H_t = \begin{cases} H^{(0)} \text{ if } t = 0 \\ H^{(i)} \text{ if } s_i < t \leq t_i \end{cases} \tag{1}$$

At any time t, the value of H_t is known *before* time t (for $t \neq 0$).

Definition

More generally, a process H_t is predictable if it can be written as a limit of simple functions $H_t^{(n)}$. This means that $H_t^{(n)}(\omega) \to H_t(\omega)$ as $n \to \infty$, for all $(t, \omega) \in [0, T] \times \Omega$.

All adapted left continuous processes are predictable (càg, for *continue à gauche*) (JS, Proposition I.2.6 (p. 17)).

Stochastic Integrals

We here consider the meaning of the expression

$$\int_0^T H_t dX_t. \tag{2}$$

The ingredients are the integrand H_t, which is assumed to be predictable, and the integrator X_t, which will generally be a semi-martingale (to be defined below).
The expression (2) is defined for simple process integrands as

$$\sum_i H^{(i)}(X_{t_i} - X_{s_i}) \tag{3}$$

For predictable integrands H_t that are bounded and limits of simple processes $H_t^{(n)}$, the integral (2) is the limit in probability of $\int_0^T H_t^{(n)} dX_t$. This limit is well defined, *i.e.,* independent of the sequence $H_t^{(n)}$.

If X_t is a Wiener process, the integral can be defined for any predictable process H_t satisfying

$$\int_0^T H_t^2 dt \; < \; \infty. \tag{4}$$

It will always be the case that the integrator X_t is right continuous with left limits (*càdlàg*, for *continue à droite, limites à gauche*).
The integral process

$$\int_0^t H_s dX_s = \int_0^T H_s I\{s \leq t\} dX_s \tag{5}$$

can also be taken to be *càdlàg*. If (X_t) is continuous, the integral is then automatically continuous.

The Continuous Itô Processes

X_t is an Itô process relative to filtration $(\mathcal{F}_t)$ provided (X_t) is $(\mathcal{F}_t)$ adapted; and if there is an $(\mathcal{F}_t)$-Wiener process (W_t), and $(\mathcal{F}_t)$-adapted processes (μ_t) and (σ_t), with

$$\int_0^T |\mu_t| dt < \infty, \text{ and } \int_0^T \sigma_t^2 dt < \infty \tag{6}$$

so that

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s. \tag{7}$$

The process is often written on differential form:

$$dX_t = \mu_t dt + \sigma_t dW_t. \tag{8}$$

We note that the Itô process property is preserved under stochastic integration. If H_t is bounded and predictable, then

$$\int_0^t H_s dX_s = \int_0^t H_s \mu_s dt + \int_0^t H_s \sigma_s dW_s. \tag{9}$$

It is clear from this formula that predictable processes H_t can be used for integration w.r.t. X_t provided

$$\int_0^T |H_t \mu_t| dt < \infty \text{ and } \int_0^T (H_t \sigma_t)^2 dt < \infty. \tag{10}$$

Stochastic Integral as Trading Profit or Loss (P/L)

Suppose that X_t is the value of a security. Let H_t be the number of this stock that is held at time t. In the case of a simple process (1), this means that we hold $H^{(i)}$ units of X from time s_i to time t_i. The trading profit and loss (P/L) is then given by the stochastic integral

$$\sum_i H^{(i)}(X_{t_i} - X_{s_i}) \tag{11}$$

In this description, it is quite clear that $H^{(i)}$ must be known at time s_i, otherwise we would base the portfolio on future information. More generally, for predictable H_t, we similarly avoid using future information.

The Heston Model

A popular model for volatility is due to Heston (1993). In this model, the process X_t is given by

$$dX_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma_t dW_t$$

$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dZ_t \ , \ \text{with} \tag{12}$$

$$Z_t = \rho W_t + (1 - \rho^2)^{1/2} B_t \tag{13}$$

where (W_t) and (B_t) are two independent Wiener processes, $\kappa > 0$, and $|\rho| \leq 1$. To assure that σ_t^2 does not hit zero, one must also require (Feller (1951)) that $2\kappa\alpha \geq \gamma^2$.

Conditional Expectations: Heuristic Definition

- X, Y random variables, $E|X| < \infty$
- Non-random conditional expectation: $f(y) = E(X|Y = y)$ (from undergraduate courses)
- Random conditional expectation given random variable Y: $E(X|Y) = f(Y)$
- Random conditional expectation given σ-field: if $\mathcal{A}$ is generated by Y, then $E(X|\mathcal{A}) = f(Y)$

Random conditional expectation is very useful in time dependent systems and large dimension. We shall see same phenomenon for Radon-Nikodyn derivatives, a.k.a. likelihood ratios, a.k.a. state price densities.

Heuristic definition runs into trouble for measure theoretic reasons, and also practically in time dependent systems and large dimension. But the heuristic is very useful to help you think.

Conditional Expectations: Official Definition

Theorem

Let $\mathcal{A}$ be a σ-field, and let X be a random variable so that $E|X| < \infty$. There is a $\mathcal{A}$-measurable random variable Z so that for all $A \in \mathcal{A}$,

$$EZI_A = EXI_A, \tag{14}$$

where I_A is the indicator function of A. Z is unique "almost surely", which is that if Z_1 and Z_2 satisfy the two criteria above, then $P(Z_1 = Z_2) = 1$.

We thus define

$$E(X|\mathcal{A}) = Z \tag{15}$$

where Z is given in the theorem. The conditional expectation is well defined "almost surely".

For further details and proof of theorem, see Section 34 (p. 445-455) of Billingsley (1995).

For a taste of the Proof

Show that $Z_1 = Z_2$ with probability 1.

[Hist: Consider set $A = \{\omega : \ Z_1(\omega) \geq Z_2(\omega)\}$, and so on.]

Properties of Conditional Expectations

- Linearity: for constant c_1, c_2:
$$E(c_1 X_1 + c_2 X_2 \mid \mathcal{A}) = c_1 E(X_1 \mid \mathcal{A}) + c_2 E(X_2 \mid \mathcal{A})$$
- Conditional constants: if Z is $\mathcal{A}$-measurable, then
$$E(ZX|\mathcal{A}) = ZE(X|\mathcal{A})$$
- Law of iterated expectations (iterated conditioning, tower property): if $\mathcal{A}' \subseteq \mathcal{A}$, then
$$E[E(X|\mathcal{A})|\mathcal{A}'] = E(X|\mathcal{A}')$$
- Independence: if X is independent of $\mathcal{A}$:
$$E(X|\mathcal{A}) = E(X)$$
- Jensen's inequality: if $g: \ x \to g(x)$ is convex:
$$E(g(X)|\mathcal{A}) \geq g(E(X|\mathcal{A}))$$

g is convex if $g(ax + (1-a)y) \leq ag(x) + (1-a)g(y)$ for $0 \leq a \leq 1$. Examples: $g(x) = e^x$, $g(x) = (x-K)^+$. Or g'' exists and is continuous, and $g''(x) \geq 0$.

Martingales

An $(\mathcal{F}_t)$ adapted process M_t is called a martingale if, for all t, $E|M_t| < \infty$, and if, for all $s < t$,

$$E(M_t|\mathcal{F}_s) = M_s. \tag{16}$$

- Central concept in our narrative.
- The concept of martingale applies equally to discrete and continuous time axis.
- A martingale is also known as a *fair game*, for the following reason. In a gambling situation, if M_s is the amount of money the gambler has at time s, then the gambler's expected wealth at time $t > s$ is also M_s.

Wiener Process as Martingales

Example

A Wiener process is a martingale. To see this, for $t > s$, since $W_t - W_s$ is N(0,t-s) given $\mathcal{F}_s$, we get that

$$\begin{aligned} E(W_t|\mathcal{F}_s) &= E(W_t - W_s|\mathcal{F}_s) + W_s \\ &= E(W_t - W_s) + W_s \text{ by independence} \\ &= W_s. \end{aligned} \tag{17}$$

Martingales; Representation by Final Value

M_t is a martingale for $0 \leq t \leq T$ if and only if one can write

$$M_t = E(X|\mathcal{F}_t) \text{ for all } t \in [0, T] \tag{18}$$

(only if by definition ($X = M_T$), if by Tower property). Note that for $T = \infty$ (which we do not consider here), this property may not hold. (For a full discussion, see Chapter 1.3.B (p. 17-19) of Karatzas and Shreve (1991).

Stochastic Integrals and Martingales

Example

If H_t is a bounded predictable process, and for any martingale X_t,

$$M_t = \int_0^t H_s dX_s \tag{19}$$

is a martingale.

Thus, any bounded trading strategy H in an asset M which is a martingale results in a martingale P/L.

To see this, consider first a simple process (1), for which $H_s = H^{(i)}$ when $s_i < s \le t_i$. For given t, if $s_i > t$, by the properties of conditional expectations,

$$\begin{aligned} E\left(H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_t\right) &= E\left(E(H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_{s_i})|\mathcal{F}_t\right) \\ &= E\left(H^{(i)}E(X_{t_i} - X_{s_i}|\mathcal{F}_{s_i})|\mathcal{F}_t\right) = 0, \end{aligned}$$

and similarly, if $s_i \le t \le t_i$, then

$$E\left(H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_t\right) = H^{(i)}(X_t - X_{s_i})$$

so that $E(M_T|\mathcal{F}_t) = E\left(\sum_i H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_t\right)$

$$= \sum_{i:t_i<t} H^{(i)}(X_{t_i} - X_{s_i}) + I\{t_i \le t \le s_i\}H^{(i)}(X_t - X_{s_i}) = M_t.$$

The result follows for general bounded predicable integrands by taking limits and using uniform integrability. For definition and results on uniform integrability, see Billingsley (1995).

Stopping Times and Local Martingales

Example of local martingale (see also Duffie (1996)):

$$X_t = \int_0^t \frac{1}{\sqrt{T-s}} dW_s \tag{20}$$

For $0 \leq t < T$, X_t is a zero mean Gaussian process with independent increments. We shall show below that

$$\mathrm{Var}(X_t) = \int_0^t \frac{1}{T-s} ds = \int_{T-t}^{T} \frac{1}{u} du = \log \frac{T}{T-t} \to \infty \text{ as } t \to T.$$

X_t is not defined at T. However, one can stop the process at a convenient time, as follows: Set, for $A > 0$,

$$\tau = \inf\{t \geq 0 : \; X_t = A\}. \tag{21}$$

One can show that $P(\tau < T) = 1$. Define a modified integral by

$$\begin{aligned} Y_t &= \int_0^t \frac{1}{\sqrt{T-s}} I\{s \leq \tau\} dW_s \\ &= X_{\tau \wedge t}, \text{ where } s \wedge t = \min(s, t). \end{aligned} \tag{22}$$

Trading Interpretation of Process Y_t

$$\text{Recall: } Y_t = \int_0^t \frac{1}{\sqrt{T-s}} I\{s \leq \tau\} dW_s = X_{\tau \wedge t}$$

Suppose that

- W_t is the value of a security at time t (the value can be negative, but that is possible for many securities, such as futures contracts).
- The short term interest rate is zero.

The process X_t comes about as the value of a portfolio which holds $1/\sqrt{T-t}$ units of the security W_t at time t. The process Y_t is obtained by holding this portfolio until such time that $X_t = A$, and then liquidating the portfolio.

In other words, we have displayed a trading strategy which starts with wealth $Y_0 = 0$ at time $t = 0$, and end with wealth $Y_T = A > 0$ at time $t = T$. In trading terms, this is an arbitrage. In mathematical terms, this is a stochastic integral w.r.t. a martingale, but the integral is no longer a martingale.

- Lesson for trading: some condition has to be imposed to make sure that a trading strategy in a martingale cannot result in arbitrage profit.
- The most popular approach to this is to require that the traders wealth at any time cannot go below some fixed amount $-K$. This is the so-called credit constraint. (So strategies are required to satisfy that the integral never goes below $-K$).
- This constraint does not quite guarantee that the integral w.r.t. a martingale is a martingale, but it does prevent arbitrage profit. The technical result is that the integral is a *super-martingale* (see the next section).

Stopping Times

For the purpose of characterizing the stochastic integral, we need the concept of a *local martingale.* For this, we first need to define:

Definition

A stopping time is a random variable τ satisfying $\{\tau \leq t\} \in \mathcal{F}_t$, for all t.

The requirement in this definition is that we must be able to know at time t wether τ occurred or not. The time (21) given above is a stopping time. On the other hand, the variable $\tau = \inf\{t : \; W_t = \max_{0 \leq s \leq T} W_s\}$ is not a stopping time. Otherwise, we would have a nice investment strategy.

Local Martingales

Definition

A process M_t is a local martingale for $0 \leq t \leq T$ provided there is a sequence of stopping times τ_n so that
(i) $M_{\tau_n \wedge t}$ is a martingale for each n; and
(ii) $P(\tau_n \to T) = 1$ as $n \to \infty$.

The basic result for stochastic integrals is now that the integral with respect to a local martingale is a local martingale, cf. result I.4.34(b) (p. 47) in JS.
Note: some variations over the concept of local martingale are in use.

Semimartingales

X_t is a semimartingale if it can be written

$$X_t = X_0 + M_t + A_t, 0 \leq t \leq T, \tag{23}$$

where X_0 is $\mathcal{F}_0$-measurable, M_t is a local martingale, and A_t is a process of finite variation, *i.e.*,

$$\sup \sum_i |A_{t_{i+1}} - A_{t_i}| < \infty, \tag{24}$$

where the supremum is over all grids $0 = t_0 < t_1 < ... < t_n = T$, and all n.

In particular, a continuous Itô process is a semimartingale, with

$$M_t = \int_0^t \sigma_t dW_t \text{ and } A_t = \int_0^t \mu_t dt. \tag{25}$$

A *supermartingale* is semimartingale for which A_t is nonincreasing. A *submartingale* is a semimartingale for which A_t is nondecreasing.

Quadratic Variation of a Semimartingale

- Grid of observation times: $\mathcal{G} = \{t_0, t_1, ..., t_n\}$ with $0 = t_0 < t_1 < ... < t_n = T$
- Maximum discrepancy: $\Delta(\mathcal{G}) = \max_{1 \le i \le n}(t_i - t_{i-1})$.
- *Quadratic variation* of process X relative to grid $\mathcal{G}$

$$[X, X]_t^{\mathcal{G}} = \sum_{t_{i+1} \le t} (X_{t_{i+1}} - X_{t_i})^2. \tag{26}$$

- Quadratic covariation of X and Y:

$$[X, Y]_t^{\mathcal{G}} = \sum_{t_{i+1} \le t} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}). \tag{27}$$

Theorem

For any semimartingale, there is a process $[X, Y]_t$ so that

$$[X, Y]_t^{\mathcal{G}} \xrightarrow{p} [X, Y]_t \text{ for all } t \in [0, T], \text{ as } \Delta(\mathcal{G}) \to 0. \tag{28}$$

The limit is independent of the sequence of grids $\mathcal{G}$.

The result follows from Theorem I.4.47 (p. 52) in JS. The t_i can even be stopping times.

- For a continuous Itô process,

$$[X,X]_t = \int_0^t \sigma_s^2 ds. \tag{29}$$

(Cf Thm I.4.52 (p. 55) and I.4.40(d) (p. 48) of JS.) In particular, for a Wiener process W, $[X,X]_t = \int_0^t 1 ds = t$.

- The process $[X,X]_t$ is usually referred to as the quadratic variation of the semimartingale (X_t). This is an important concept, as seen in the rest of the course. The theorem asserts that this quantity can be estimated consistently from data.
- The result follows from Theorem I.4.47 (p. 52) in JS.
- The t_i can even be stopping times.

Semimartingales with Jumps

- Most commonly studied discontinous semimartingale: $X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s + J_t$
- J_t is a process whose only quadratic variation comes from jumps: $[X, X]_t = \int_0^t \sigma_s^2 ds + [J, J]_t$ where $[J, J]_t = \sum_{0<s\leq t} \Delta X_s^2 = \sum_{0<s\leq t} \Delta J_s^2$ and $\Delta X_s = X_s - X_{s-}$
- Decomposition $J_t = J_t^{(1)} + J_t^{(2)}$
- $J_t^{(1)}$ has finitely many jumps at times that can be predictable
- $J_t^{(2)}$ is a “pure jump" Itô semimartingale
- Under condition (H) (Jacod and Protter (2012), p. 126), the “predictable" quadratic variation $\langle J^{(2)}, J^{(2)} \rangle_t$ is absolutely continous and locally bounded. This is a generalization of a compound Poisson process.

Properties of Quadratic Variation

(1) Bilinearity: $[X, Y]_t$ is linear in each of X and Y: so for example, $[aX + bZ, Y]_t = a[X, Y]_t + b[Z, Y]_t$.
(2) If (W_t) and (B_t) are two independent Wiener processes, then

$$[W, B]_t = 0. \tag{30}$$

Example

Suppose $Z_t = \rho W_t + (1 - \rho^2)^{1/2} B_t$, where W_t and B_t are independent Wiener processes.
One obtains from first principles that

$$\begin{aligned} [W, Z]_t &= \rho [W, W]_t + (1 - \rho^2)^{1/2} [W, B]_t \\ &= \rho t, \end{aligned} \tag{31}$$

since $[W, W]_t = t$ and $[W, B]_t = 0$.

Properties of Quadratic Variation (cont'd)

(3) For stochastic integrals over Itô processes X_t and Y_t,

$$U_t = \int_0^t H_s dX_s \text{ and } V_t = \int_0^t K_s dY_s, \tag{32}$$

then

$$[U, V]_t = \int_0^t H_s K_s d[X, Y]_s. \tag{33}$$

This is often written on "differential form" as

$$d[U, V]_t = H_t K_t d[X, Y]_t. \tag{34}$$

by invoking the same results that led to (29). For a rigorous statement, see Property I.4.54 (p.55) of Jacod and Shiryaev (2003) ("JS", from now on)

(4) For any Itô process X, $[X, t] = 0$.

Properties of Quadratic Variation (cont'd)

Example

(Leverage Effect in the Heston model). For the Heston model, recall

$$
\begin{aligned}
dX_t &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma_t dW_t \\
d\sigma_t^2 &= \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dZ_t \;, \text{ with} && (35) \\
Z_t &= \rho W_t + (1-\rho^2)^{1/2} B_t && (36)
\end{aligned}
$$

We have seen above that $[W, Z]_t = \rho t$, and so

$$
\begin{aligned}
d[X, \sigma^2] &= \gamma\sigma_t^2 d[W, Z]_t \\
&= \gamma\sigma^2 \rho dt. && (37)
\end{aligned}
$$

Properties of Quadratic Variation (cont'd)

(5) Invariance under discounting by the short term interest rate. The typical discount rate is the risk free short term interest rate r_t. Recall that $S_t = \exp\{X_t\}$. The discounted stock price is then given by

$$S_t^* = \exp\{-\int_0^t r_s ds\} S_t. \tag{38}$$

The corresponding process on the log scale is $X_t^* = X_t - \int_0^t r_s ds$, so that if X_t is given by (8), then

$$dX_t^* = (\mu_t - r_t)dt + \sigma_t dW_t. \tag{39}$$

The quadratic variation of X_t^* is therefore the same as for X_t.

Properties of Quadratic Variation (cont'd)

While this result remains true for certain other types of discounting (such as cost-of-carry), it is not true for many other types of discounting. For example, if one discounts by the zero coupon bond Λ_t maturing at time T, the discounted log price becomes $X_t^* = X_t - \log \Lambda_t$. Since the zero coupon bond will itself have volatility, we get

$$[X^*, X^*]_t = [X, X]_t + [\log \Lambda, \log \Lambda]_t - 2[X, \log \Lambda]_t. \qquad (40)$$

Variance and Quadratic Variation

Quadratic variation has a representation in terms of variance. The main result concerns martingales. For $E(X^2) < \infty$, define the conditional variance by

$$\mathrm{Var}(X|\mathcal{A}) = E((X - E(X|\mathcal{A}))^2|A) = E(X^2|\mathcal{A}) - E(X|\mathcal{A})^2. \quad (41)$$

and similarly

$$\mathrm{Cov}(X, Y|\mathcal{A}) = E\left((X - E(X|\mathcal{A}))(Y - E(Y|\mathcal{A}))|\mathcal{A}\right).$$

Theorem

Let M_t be a martingale, and assume that $E[M, M]_T < \infty$. Then, for all $s < t$,

$$\mathrm{Var}(M_t|\mathcal{F}_s) \;=\; E((M_t - M_s)^2|\mathcal{F}_s) \;=\; E([M, M]_t - [M, M]_s|\mathcal{F}_s).$$

This theorem is the beginning of something important: the left hand side relates to the central limit theorem, while the right hand side only concerns the law of large numbers. We shall see this effect in more detail in the sequel. (Proof: MZ2012, p. 130)

Spot quantities for Continuous Itô Processes

- $\lim_{h\downarrow 0} \frac{1}{h}\mathrm{Cov}(X_{t+h} - X_t, Y_{t+h} - Y_t|\mathcal{F}_t) = \frac{d}{dt}[X, Y]_t$
- Spot correlation:
  $\mathrm{cor}(X, Y)_t = \lim_{h\downarrow 0} \mathrm{cor}(X_{t+h} - X_t, Y_{t+h} - Y_t|\mathcal{F}_t)$
- Calculating spot correlation:
  $\mathrm{cor}(X, Y)_t = \frac{d[X,Y]_t/dt}{\sqrt{(d[X,X]_t/dt)(d[Y,Y]_t/dt)}}$
- In the Heston model,

$$\mathrm{cor}(X, \sigma^2)_t = \rho. \tag{42}$$

- In general, if $dX_t = \sigma_t dW_t + dt$-term, and $dY_t = \gamma_t dB_t + dt$-term, where W_t and B_t are two Wiener processes, then

$$\mathrm{cor}(X, Y)_t = \mathrm{sgn}(\sigma_t \gamma_t)\mathrm{cor}(W, B)_t. \tag{43}$$

Lévy's Theorem

Theorem

Suppose that M_t is a continuous $(\mathcal{F}_t)$-local martingale, $M_0 = 0$, so that $[M, M]_t = t$. Then M_t is an $(\mathcal{F}_t)$-Wiener process.

(Cf. Thm II.4.4 (p. 102) in JS.) More generally, from properties of normal random variables, the same result follows in the vector case: If $M_t = (M_t^{(1)}, ..., M_t^{(p)})$ is a continuous $(\mathcal{F}_t)$-martingale, $M_0 = 0$, so that $[M^{(i)}, M^{(j)}]_t = \delta_{ij} t$, then M_t is a vector Wiener process. (δ_{ij} is the Kronecker delta: $\delta_{ij} = 1$ for $i = j$, and $= 0$ otherwise.)

Predictable Quadratic Variation

One can often see the symbol $\langle X, Y\rangle_t$.This can be called the predictable quadratic vartiation. Under regularity conditions, it is defined as the limit of $\sum_{t_i \le t} \mathrm{Cov}(X_{t_{i+1}} - X_{t_i}, Y_{t_{i+1}} - Y_{t_i} | \mathcal{F}_{t_i})$ as $\Delta(\mathcal{G}) \to 0$.

For continuous Itô processes, $\langle X, Y\rangle_t = [X, Y]_t$. For general semimartingales this equality does not hold. Also, except for Itô processes, $\langle X, Y\rangle_t$ cannot generally be estimated consistently from data without further assumptions. For example, If N_t is a Poisson process with intensity λ, then $M_t = N_t - \lambda t$ is a martingale. In this case, $[M, M]_t = N_t$ (observable), while $\langle M, M\rangle_t = \lambda t$ (cannot be estimated in finite time, though see discussion in Stoltenberg, M and Z, 2020).

For continuous semimartingales, The symbol $\langle X, Y\rangle_t$ is commonly used in the literature in lieu of $[X, Y]_t$

Itô's Formula for Continuous Itô processes

Theorem

Suppose that f is a twice continuously differentiable function, and that X_t is an Itô process. Then

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)d[X,X]_t. \tag{44}$$

Similarly, in the multivariate case, for $X_t = (X_t^{(1)}, ..., X_t^{(p)})$,

$$df(X_t) = \sum_{i=1}^{p} \frac{\partial f}{\partial x^{(i)}}(X_t)dX_t^{(i)} + \frac{1}{2}\sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x^{(i)} \partial x^{(j)}}(X_t)d[X^{(i)}, X^{(j)}]_t. \tag{45}$$

(Reference: Theorem I.4.57 in JS.)

Itô's Formula for Continuous Itô processes (cont'd)

We emphasize that (44) is the same as saying that

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)d[X,X]_s. \tag{46}$$

If we write out $dX_t = \mu_t dt + \sigma_t dW_t$ and $d[X,X]_t = \sigma_t^2 dt$, then equation (44) becomes

$$\begin{aligned} df(X_t) &= f'(X_t)(\mu_t dt + \sigma_t dW_t) + \frac{1}{2}f''(X_t)\sigma_t^2 dt \\ &= (f'(X_t)\mu_t + \frac{1}{2}f''(X_t)\sigma_t^2)dt + f'(X_t)\sigma_t dW_t. \end{aligned} \tag{47}$$

We note, in particular, that if X_t is a continuous Itô process, then so is $f(X_t)$.

Example

(Example of Itô's Formula: Stochastic Equation for a Stock Price.) We have so far discussed the model for a stock on the log scale, as $dX_t = \mu_t dt + \sigma_t dW_t$. The price is given as $S_t = \exp(X_t)$. Using Itô's formula, with $f(x) = \exp(x)$, we get

$$dS_t = S_t(\mu_t + \frac{1}{2}\sigma_t^2)dt + S_t\sigma_t dW_t. \tag{48}$$

Example

(Example of Itô's Formula: Proof of Lévy's Theorem.) Take $f(x) = e^{ihx}$, and go on from there. Left for the students. (Hard. See proof un Karatzas and Shreve (1991).)

Example of Itô's Formula: Genesis of the Leverage Effect

A case where quadratic covariation between a process and it's volatility can arise from basic economic principles. The following is the origin of the use of the word "leverage effect" to describe such covariation. This type of covaration can arise from many sources, and will later use the term leverage effect to describe the phenomenon broadly.

Suppose that the log value of a firm is Z_t, given as a GBM,

$$dZ_t = \nu dt + \gamma dW_t. \tag{49}$$

For simplicity, suppose that the interest rate is zero, and that the firm has borrowed C dollars (or euros, yuan, ...). If there are M shares in the company, the value of one share is therefore

$$S_t = (\exp(Z_t) - C)/M. \tag{50}$$

From the derivations on p. 132-134 in MZ (2012),

$$\mathrm{cor}(X, \sigma^2)_t = -1. \tag{51}$$

Next:
Estimation of volatility. Central limit theorems and stable convergence.