

# Mathematics of Continuous-Time Models

EBEN LAZARUS

MIT 15.472

*Fall 2022*

# Outline

Background

Probability

Stochastic Processes

Stochastic Integration

Itô Processes


SDEs

Conditional Moments

Martingale Measures

# Brief Background

## Why work in continuous time?

- ▶ Usual first answer: ~~More “realistic.”~~ 
  - ▶ Try trading continuously...
  - ▶ ... or programming a computer to store continuous-time data (or take a derivative)...
- ▶ Better answer: It's useful.
  - ▶ Often **simpler**, not more complex.
  - ▶ Can solve a range of problems that would require approximation or numerical solutions in discrete time (portfolio choice, option pricing, ICAPM, ...).
  - ▶ Even if you have no interest in theory, useful to speak the language.

# Overview

- ▶ Some technical overhead before we get to economic problems of interest.
- ▶ Want to give you a solid, reasonably self-contained background this week...
- ▶ ... and have you review with a short problem set before coming back for our main material.
- ▶ Shouldn't think of technical material as orthogonal to the economic questions we'll get to; some of the math was in fact originally developed specifically for this purpose (e.g., Bachelier, 1900).

# Overview

- ▶ Up first: Formalize the notion of gradual information revelation.
- ▶ Formal concepts: **filtration** and  **$\sigma$ -algebra**.
- ▶ With finite states, will be easy to think about filtrations intuitively. With infinite states and in continuous time (which is really where these concepts are useful), will become more abstract.
- ▶ Very quick review in class, but more slides included for reference. (Skipping anything with ↗.)
- ▶ Then continuous-time processes and stochastic integration.
- ▶ Duffie textbook has technical details & additional references; Back textbook gives a bit more economic intuition.
  - ▶ See syllabus for precise chapter references.

# Outline

Background

**Probability**

Stochastic Processes

Stochastic Integration

Itô Processes

SDEs

Conditional Moments

Martingale Measures

# Probability Space and $\sigma$ -Algebras

- ▶ Consider a set  $\Omega$  of states.

## Definition ( $\sigma$ -Algebra)

A  **$\sigma$ -algebra**  $\mathcal{F}$  on  $\Omega$  is a collection of subsets of  $\Omega$ , such that

1.  $\emptyset \in \mathcal{F}$ ,
2.  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$  (closed under complement),
3. for a countable set of subsets  $A_i$ ,  $A_i \in \mathcal{F} \Rightarrow \cup_i A_i \in \mathcal{F}$ .

- ▶ Elements of a  $\sigma$ -algebra are referred to as **events**.
- ▶ The  $\sigma$ -algebra represents the “fineness” of our information on  $\Omega$ .
  - ▶ E.g., take state  $\omega \in \Omega$  to refer to the wealth of some person as of today, and assume wealth can take on only 3 values:  $\Omega = \{\$100, \$200, \$300\}$ .
  - ▶ See their response to a survey question asking whether wealth  $\leq \$100$  or wealth  $> \$100$ . Can represent by setting  $\mathcal{F} = \{\emptyset, \$100, \{\$200, \$300\}, \Omega\}$ .

# Filtration

- ▶ Assume for now that time  $t$  is discrete, taking values in set  $\mathcal{T} = \{0, 1, \dots, T\}$ .

## Definition (Filtration)

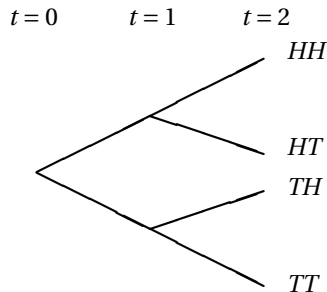
A **filtration**  $\mathbb{F}$  on  $\Omega$  is a collection of  $\sigma$ -algebras  $\mathcal{F}_t$ , for  $t \in \mathcal{T}$ , such that

$$\forall s > t, \quad \mathcal{F}_t \subset \mathcal{F}_s.$$

- ▶ Generally assume  $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_T = \mathcal{F}$ .
- ▶ When time is discrete and  $\Omega$  is finite, we can represent filtration by a tree.



## Constructing a Filtration $\leadsto$



- ▶ At  $t=0$ , “Nature” tosses two coins at once, and the outcomes are revealed one at a time in the next two periods,  $t=1, 2$ . (Observationally equivalent to coin tosses happening sequentially.)
- ▶  $\Omega = \{HH, HT, TH, TT\}$ ,  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ ,  $\mathcal{F}_1 = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}\}$ ,  
 $\mathcal{F}_2 = \mathcal{F} = 2^\Omega = \{\emptyset, \Omega, \{HH, HT\}, \{TH, TT\}, HH, HT, TH, TT\}$  = all subsets of  $\Omega$ .
- ▶ Information revealed over time, so  $\mathcal{F}_t$  (which enumerates possible events) becomes finer.

# Probability Measure

- ▶ We have states  $(\Omega)$  and collections of events  $(\mathcal{F})$ ; now define probabilities over those events.

## Definition (Probability Measure)

A **probability measure** is a function  $P: \mathcal{F} \rightarrow [0, 1]$  such that

1.  $P(\emptyset) = 0, P(\Omega) = 1$ ;
2. for a countable set of disjoint events  $A_i \in \mathcal{F}$ ,

$$P(\cup_i A_i) = \sum_i P(A_i).$$

- ▶ The triplet  $(\Omega, \mathcal{F}, P)$  is a **probability space**.
- ▶ In a dynamic setting, we add a filtration  $\mathbb{F}$  to a probability space (**filtered probability space**  $(\Omega, \mathcal{F}, \mathbb{F}, P)$ ).

## Constructing a Probability Space $\leadsto$

- ▶ Continue with the coin-flip example. Suppose the probability of heads is  $p$ . What's the right probability-space formalization of this idea?
- ▶ First, set  $P(\emptyset) = 0$ ,  $P(\Omega) = 1$ ,  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ .
- ▶ Next, considering the two sets

$A_H$  = the set of sequences beginning with  $H = \{HH, HT\}$

$A_T$  = the set of sequences beginning with  $T = \{TH, TT\}$

we set  $P(A_H) = p$ ,  $P(A_T) = 1 - p$ , and again  $\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}$ .

- ▶ Finally, with  $A_{HH}$  = the sequence  $HH, \dots$ , we set  $P(A_{HH}) = p^2, P(A_{HT}) = p(1 - p), \dots$
- ▶ Since  $\mathcal{F} = \mathcal{F}_2$  is defined by  $\{\emptyset, \Omega, A_H, A_T, A_{HH}, A_{HT}, A_{TH}, A_{TT}\}$ , we've fully specified the probability measure  $P$  over  $\mathcal{F}$  (and it meets our definition).
  - ▶ But notice that along the way, we've also implicitly defined probability measures over the **sub- $\sigma$ -algebras**  $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2$  (which together make up the filtration), as each sub- $\sigma$ -algebra is itself a well-defined  $\sigma$ -algebra.

# Random Variables

- ▶ Have worked until now with discrete state spaces, but the formal tools we've been developing are really necessary/helpful for working with continuous spaces.
- ▶ For example, take  $\Omega = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ , where the first dimension of the state space represents financial wealth  $W$  and the second represents “human capital”  $H$  (and both are continuous).
  - ▶ How do you define the distribution of  $W$  conditional on  $H$ ? For any value of  $H$  we're conditioning on a zero-probability event, so does this even make sense?
  - ▶ Can't we just take  $f(W|H = h) = \lim_{\epsilon \rightarrow 0} f(W|H \in (h - \epsilon, h + \epsilon))$ ? Borel–Kolmogorov paradox: No!
  - ▶ Lots of similar/related problems arise in continuous spaces (e.g., Banach–Tarski paradox: can't construct well-behaved measure for *all* subsets of state space).
  - ▶ Resolution: Consider only certain subsets (open sets) and go directly to conditional expectation.

# Random Variables

## Definition (Borel $\sigma$ -algebra)

The Borel  $\sigma$ -algebra is the smallest  $\sigma$ -algebra containing all open subsets of a topological space  $\Omega$  (e.g.,  $\Omega = \mathbb{R}^d$ ).

## Definition (Measurability)

A function  $X: \Omega \rightarrow \mathbb{R}^d$  is  **$\mathcal{F}$ -measurable** iff for any set  $A$  in the Borel  $\sigma$ -algebra, the set  $\{\omega \in \Omega : X(\omega) \in A\}$  belongs to  $\mathcal{F}$ .

- ▶  $X$  is a *random variable* on the probability space  $(\Omega, \mathcal{F}, P)$ .
- ▶  $L^1$  denotes the set of integrable random variables (RVs for which we can take expectations):

$$\int_{\Omega} |X(\omega)| dP(\omega) < \infty$$

- ▶  $L^2$  denotes the set of square-integrable random variables (can take both mean and variance).

# Outline

Background

Probability

**Stochastic Processes**

Stochastic Integration

Itô Processes

SDEs

Conditional Moments

Martingale Measures

# Continuous-Time Filtrations and Martingales

- ▶ Assume that time is continuous and takes values in the set  $\mathcal{T} = [0, T]$ .
- ▶ Things we want to know:
  - ▶ How to define a continuous asset price process  $S_t$ ?
  - ▶ How to compute various properties of the price process, e.g., conditional moments (expected return, volatility ...)?
  - ▶ How to define a trading strategy  $\theta_t$  and compute a corresponding wealth process  $(\int \theta_t dS_t)$ ?
  - ▶ How to characterize continuous martingales and martingale measures?

# Basic Definitions

## Definition (Product $\sigma$ -Algebra)

The product algebra on  $\Omega \times \mathcal{T}$  is the algebra generated by the subsets of the form  $A \times B$  where  $A \in \mathcal{F}$  and  $B$  is in the Borel  $\sigma$ -algebra on  $\mathcal{T}$ .

## Definition (Stochastic Process)

A stochastic process is a function  $X: \Omega \times \mathcal{T} \rightarrow \mathbb{R}^d$  that is measurable w.r.t. the product algebra on  $\Omega \times \mathcal{T}$ .

- Easier to work with this definition than other “intuitive” definitions, e.g.  $X: \Omega \rightarrow (\mathbb{R}^d)^{\mathcal{T}}$ .

## Definition

Two stochastic processes  $X$  and  $Y$  are **versions** of each other iff for all  $t$ ,  $X_t = Y_t$  almost surely (“a.s.,” or with prob. 1).



# Basic Definitions

## Definition (Sample Path)

A **sample path** of a stochastic process is the function  $X(\omega, t)$  for a given  $\omega$ .

## Definition (Adapted Processes)

A stochastic process is **adapted** to the filtration  $\mathbb{F}$  iff for all  $t \in \mathcal{T}$ , the function  $X(\omega, t)$  is  $\mathcal{F}_t$ -measurable.

- ▶ An adapted stochastic process depends on the info set up to time  $t$  ( $\mathcal{F}_t$ ) so can't “see into the future.” Sometimes called “non-anticipating.” (What kinds of processes does this not work for?)
- ▶ We generally use  $X(t)$  or  $X_t$  instead of  $X(\omega, t)$ .
- ▶ Denote the set of one-dimensional adapted processes by  $\mathcal{L}$ .

# Conditional Expectation Given Available Information

## Definition (Conditional Expectation w.r.t. a $\sigma$ -Algebra)

Let  $s > t$ , so that the  $\sigma$ -algebra  $\mathcal{F}_t$  is smaller than  $\mathcal{F}_s$ . Let  $X_s$  be an  $L^1$   $\mathcal{F}_s$ -measurable random variable. A **conditional expectation** of  $X_s$  with respect to the  $\sigma$ -algebra  $\mathcal{F}_t$ ,  $E[X_s|\mathcal{F}_t]$ , is the (almost surely) uniquely defined  $L^1$   $\mathcal{F}_t$ -measurable random variable  $Y_t$  s.t. for all  $A \in \mathcal{F}_t$ ,

$$E[X_s \mathbb{1}_A] = E[Y_t \mathbb{1}_A],$$

or equivalently  $\int_A X_s dP = \int_A E[X_s|\mathcal{F}_t] dP$ .

- ▶ The conditional expectation can be thought of as a projection on a relatively coarse information set (it is the solution to the problem  $\min_{\tilde{Y}_t} E[(X_s - \tilde{Y}_t)^2]$  s.t.  $\tilde{Y}_t$   $\mathcal{F}_t$ -measurable), and note that it is itself a RV.
  - ▶ Return to coin-flip example.  $\mathcal{F}_1 = \{\emptyset, \Omega, A_H, A_T\}$ , and set  $X_2 = \mathbb{1}_{HH}$ . Then  $[X_2|\mathcal{F}_1] = \frac{1}{2} \mathbb{1}_{A_H}$  works.
- ▶ Definition is not constructive; how do we know it's unique? Say  $Y_t, Y'_t$  both satisfy definition, and define  $A_\epsilon \equiv \{\omega : Y_t - Y'_t \geq \epsilon\}, \epsilon > 0$ . We have  $0 = \int_{A_\epsilon} (Y_t - Y'_t) dP \geq \epsilon P(A_\epsilon) \geq 0$ , so  $P(A_\epsilon) = 0$ .

# Conditional Expectation Given Available Information

## Definition (Conditional Expectation w.r.t. a $\sigma$ -Algebra)

Let  $s > t$ , so that the  $\sigma$ -algebra  $\mathcal{F}_t$  is smaller than  $\mathcal{F}_s$ . Let  $X_s$  be an  $L^1 \mathcal{F}_s$ -measurable random variable. A **conditional expectation** of  $X_s$  with respect to the  $\sigma$ -algebra  $\mathcal{F}_t$ ,  $E[X_s|\mathcal{F}_t]$ , is the (almost surely) uniquely defined  $L^1 \mathcal{F}_t$ -measurable random variable  $Y_t$  s.t. for all  $A \in \mathcal{F}_t$ ,

$$E[X_s \mathbb{1}_A] = E[Y_t \mathbb{1}_A],$$

or equivalently  $\int_A X_s dP = \int_A E[X_s|\mathcal{F}_t] dP$ .

- Nice thread from a few days ago by Ben Golub on conditional expectation:  
[https://twitter.com/ben\\_golub/status/1580349449481560064](https://twitter.com/ben_golub/status/1580349449481560064).

# Martingales

## Definition (Martingales)

Consider an adapted stochastic process  $X$  such that  $X_t$  is integrable for all  $t$ .

- ▶  $X$  is a **martingale** iff  $X_t = E(X_s | \mathcal{F}_t)$  for all  $s > t$ .
- ▶  $X$  is a **submartingale** iff  $X_t \leq E(X_s | \mathcal{F}_t)$  for all  $s > t$ .
- ▶  $X$  is a **supermartingale** iff  $X_t \geq E(X_s | \mathcal{F}_t)$  for all  $s > t$ .

# Stopping Times

## Definition (Stopping Times)

A **stopping time** is an  $\mathcal{F}$ -measurable function  $\tau : \Omega \rightarrow \mathcal{T}$  such that

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$$

- ▶ Function  $\tau$  is a stopping time if the “decision” it corresponds to (e.g., stop gambling after you’ve lost a certain amount of money) depends only on information available up to time  $t$ .

# Local Martingales

## Definition (Local Martingales)

An adapted stochastic process is a **local martingale** iff there exists an almost-surely increasing sequence  $\tau_n$  of stopping times that converge to  $T$  a.s., such that the “stopped” process  $X^n$  defined by  $X_t^n = X_{\min\{t, \tau_n\}}$  is a martingale for every  $n$ .

- ▶ A martingale is a local martingale, but the converse is not true.
- ▶ Local martingale intuitively is a martingale for a neighborhood around  $t$  but not necessarily universally. Will turn out to be quite useful for stochastic integration.

# Local Martingales vs. Martingales

## Example (Doubling Strategies)

Consider the gains from a “doubling” strategy, defined recursively as follows.

- ▶ For  $t \in [0, \frac{1}{2})$ ,  $X_t = 0$ .
- ▶ For  $t \in [\frac{1}{2}, \frac{3}{4})$ ,  $X_t = 1$  if a coin flip is heads and  $X_t = -1$  if it is tails. If the first coin flip is heads, then  $X_t = 1$  for  $t \in [\frac{3}{4}, \frac{7}{8})$ . Otherwise,  $X_t = 1$  if a second coin flip is heads, and  $X_t = -3$  if it is tails. If the first or second coin flip are heads, then  $X_t = 1$  for  $t \in [\frac{7}{8}, \frac{15}{16})$ . Otherwise,  $X_t = 1$  if a third coin flip is heads, and  $X_t = -7$  if it is tails, and so on.

The process  $X$  is a local martingale but not a martingale.

## Theorem

A local martingale that is bounded below is a supermartingale.

- ▶ Lower bound on wealth would make doubling strategy infeasible  $\implies$  wealth is then a supermartingale (non-increasing in expectation). Will use this to get absence of arbitrage.

# Brownian Motion

## Definition (Brownian Motion)

A **Brownian motion** is a stochastic process  $Z$  in  $\mathbb{R}^d$  such that

1.  $Z_0 = 0$  a.s.,
2. for all  $s > t$ ,  $Z_s - Z_t$  is normal with mean 0 and covariance matrix  $(s - t)I$ ,
3. for all  $0 \leq t_0 < t_1 < \dots < t_n \leq T$ , the random variables  $Z_{t_0}, Z_{t_1} - Z_{t_0}, \dots, Z_{t_n} - Z_{t_{n-1}}$ , are independent.

- ▶ **Brownian filtration:** The standard (or natural) filtration  $\mathbb{F}$  of a Brownian motion  $Z$  contains all the history of the Brownian motion up to time  $t$ .
  - ▶ Formally, the natural filtration of a stochastic process  $Z$  is defined by the collection of  $\sigma$ -algebras s.t.,  $\forall t$ ,  $\mathcal{F}_t$  is the smallest  $\sigma$ -algebra w.r.t. which  $Z$  is adapted (denoted  $\mathcal{F}_t = \sigma(Z_s, s \in [0, t])$ ).
- ▶ Brownian motion is a martingale w.r.t. its standard filtration: for  $s > t$ ,

$$E(Z_s | \mathcal{F}_t) = E(Z_t + (Z_s - Z_t) | \mathcal{F}_t) = Z_t + E(Z_s - Z_t | \mathcal{F}_t) = Z_t.$$



# Properties of Brownian Paths

- ▶ Denote a partition  $0 = t_0 < t_1 < \dots < t_n = T$  of  $[0, T]$ , by  $\mathcal{P}$ , and define the mesh of the partition as

$$\ell(\mathcal{P}) = \max_i |t_{i+1} - t_i|.$$

## Definition (Variation of a Function)

The **total variation** of a function  $f : \mathcal{T} \rightarrow \mathbb{R}$  is

$$V(f) = \sup_{\mathcal{P}} \sum_i |f(t_{i+1}) - f(t_i)|.$$

- ▶ Sample paths of a Brownian motion have infinite total variation.
- ▶ Brownian paths are nowhere differentiable.

# Quadratic Variation

## Definition (Quadratic Variation)

The **quadratic variation** of a one-dimensional stochastic process  $X$  on  $[0, T]$  is

$$[X]_T \equiv \lim_{\ell(\mathcal{P}) \rightarrow 0} \sum_i |X(t_{i+1}) - X(t_i)|^2,$$

where the limit is taken in probability.

## Theorem

For Brownian motion processes  $Z$ , quadratic variation  $[Z]_T = T$  for all  $T \geq 0$ .

# Sketches of Proofs for Brownian Processes

## ► Quadratic variation: $\leadsto$

1. For partition  $\mathcal{P}$ , set  $\theta_i = (Z(t_{i+1}) - Z(t_i))^2 - (t_{i+1} - t_i)$ . By the independent normal increments property of Brownian motion,  $\theta_i$  is a sequence of independent RVs with zero mean.
2. Define  $\mathcal{Q}(\mathcal{P}) = \sum_i (Z(t_{i+1}) - Z(t_i))^2$  (so  $[Z]_T = \lim_{\ell(\mathcal{P}) \rightarrow 0} \mathcal{Q}(\mathcal{P})$ ). Then  $\mathcal{Q}(\mathcal{P}) - T = \sum_i \theta_i$ .
3. Since  $E[X^4] = 3$  for standard normal  $X$ , it can be shown that  $E(\mathcal{Q}(\mathcal{P}) - T)^2 = 2 \sum_i (t_{i+1} - t_i)^2 \leq 2\ell(\mathcal{P}) \sum_i (t_{i+1} - t_i) = 2\ell(\mathcal{P})T$ , so  $\lim_{\ell(\mathcal{P}) \rightarrow 0} E(\mathcal{Q}(\mathcal{P}) - T)^2 = 0$ .
4. Since  $E\mathcal{Q}(\mathcal{P}) = T$ , Markov's inequality implies  $[Z]_T = T$ .

## ► Total variation:

1. For a continuous process with finite total variation, its quadratic variation is zero.
2. Since Brownian processes have non-zero quadratic variation for  $T > 0$ , if we assume that such a process has finite total variation, we obtain a contradiction immediately.

## Discretization of the Brownian Motion $\leadsto$

- ▶ The increment  $dZ_t \equiv Z_{t+dt} - Z_t$  is normal with mean 0 and variance  $dt$ .
- ▶ We can represent a Brownian motion heuristically by a binomial tree, or a random walk, with increments following a binomial distribution:  $\pm\sqrt{dt}$ , each with probability 1/2.
- ▶ Donsker's invariance principle (or functional central limit theorem): As  $dt \rightarrow 0$ , the above random walk converges in distribution to the Brownian motion.
- ▶ In Monte Carlo simulations, replace Brownian motion with a discrete-time random walk.

# Outline

Background

Probability

Stochastic Processes

**Stochastic Integration**

Itô Processes

SDEs

Conditional Moments

Martingale Measures

# Motivation: Computing Gains from Trading

Consider a simplified securities market model:

- ▶ Assume that the riskless rate is 0.
- ▶ Stock price:  $Z$ . No dividends.
- ▶  $\theta_t$ : number of shares of the stock held at time  $t$ . Suppose that  $\theta_t$  changes only at a finite set of times,  $0 = t_0 < t_1 < \dots < t_n = T$ .
- ▶ Wealth at time  $T$  can be computed straightforwardly as

$$W_T = W_0 + \sum_{i=0}^{n-1} \theta_{t_i} (Z_{t_{i+1}} - Z_{t_i}).$$

- ▶ But how to compute wealth when  $\theta_t$  is a general adapted stochastic process? That is, we want to define

$$\int_0^T \theta_t dZ_t.$$

# Stochastic Integration

- ▶ The **Riemann integral** is

$$\int_0^T \theta_t dt = \lim_{\ell(\mathcal{P}) \rightarrow 0} \sum_i \theta_{t_i} (t_{i+1} - t_i),$$

and can be defined for a piecewise-continuous function  $\theta_t$ . (This is “standard” integration.)

- ▶ The Riemann integral can be extended to the **Stieltjes integral** (or Riemann–Stieltjes integral)

$$\int_0^T \theta_t dA_t = \lim_{\ell(\mathcal{P}) \rightarrow 0} \sum_i \theta_{t_i} (A_{t_{i+1}} - A_{t_i}),$$

where the function  $A_t$  is increasing. (We’ve been implicitly using this when integrating w.r.t. measure  $P$ .)

- ▶ Stieltjes integral can be further extended to the case where the function  $A_t$  has bounded variation.
- ▶ The problem: Sample paths of a Brownian motion have infinite variation!

# Stochastic Integration

## Definition (Simple Processes)

An adapted process  $\theta$  is **simple** iff there exists a partition  $0 = t_0 < t_1 < \dots < t_n = T$  of  $[0, T]$  such that  $\theta_t = \theta_{t_{i-1}}$  for  $t \in [t_{i-1}, t_i)$ .

- ▶ For a simple process, we define the stochastic integral by

$$\int_0^T \theta_t dZ_t = \sum_{i=0}^{n-1} \theta_{t_i} (Z_{t_{i+1}} - Z_{t_i}).$$

- ▶ So we're back to the case presented at the outset, but we're going to build up from here and take the limit of the above equation as we approximate general  $\theta_t$  better and better.
- ▶ Note that this is like a Stieltjes interval, but without taking the mesh to zero yet.
  - ▶ We're in fact going to take the mesh to zero; the difference will be that we obtain a limit in probability (actually slightly stronger), and not path-by-path. (That is, the limit we obtain will be a random variable that our sequence approximates well with prob. 1, but the limit and that RV are not necessarily equivalent for every zero-mass sample path. This is subtle; you just should know that it works!)



## Stochastic Integral with Simple $\theta_t$ : Properties

- ▶ Setting  $\theta_t = 1 \forall t$ , we get  $\int_0^t dZ_s = Z_t - Z_0$ . (This seems good!)
- ▶ The stochastic process  $\int_0^t \theta_s dZ_s$  is adapted.
- ▶ The stochastic process  $\int_0^t \theta_s dZ_s$  is a martingale.
- ▶ The stochastic process  $\int_0^t \theta_s dZ_s$  is continuous.
- ▶ Linearity:

$$\int_0^t (\alpha \theta_s + \beta \phi_s) dZ_s = \alpha \int_0^t \theta_s dZ_s + \beta \int_0^t \phi_s dZ_s.$$

- ▶ Isometry:

$$\mathbb{E} \left( \int_0^t \theta_s dZ_s \right)^2 = \mathbb{E} \left( \int_0^t \theta_s^2 ds \right).$$

- ▶ Quadratic variation: For  $I_t = \int_0^t \theta_s dZ_s$ ,

$$[I]_t = \int_0^t \theta_s^2 ds.$$

# Integrability of Stochastic Processes: Notation

$$\mathcal{L}^1 = \left\{ X \in \mathcal{L} : \int_0^T |X_t| dt < \infty \text{ a.s.} \right\}$$

$$\mathcal{L}^2 = \left\{ X \in \mathcal{L} : \int_0^T X_t^2 dt < \infty \text{ a.s.} \right\}$$

$$\mathcal{H}^1 = \left\{ X \in \mathcal{L} : \mathbb{E} \left[ \left( \int_0^T X_t^2 dt \right)^{\frac{1}{2}} \right] < \infty \right\}$$

$$\mathcal{H}^2 = \left\{ X \in \mathcal{L} : \mathbb{E} \left( \int_0^T X_t^2 dt \right) < \infty \right\}$$

## Stochastic Integral: $\theta \in \mathcal{H}^2$

- ▶ For  $\theta \in \mathcal{H}^2$ , there exists a sequence of simple processes  $\theta^n$  such that

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \int_0^T (\theta_t^n - \theta_t)^2 dt \right) = 0.$$

- ▶ The sequence of stochastic integrals  $\int_0^T \theta_t^n dZ_t$  converges to a limit in  $L^2$ .
- ▶ The limit is independent of the particular sequence  $\theta^n$ . This limit is the stochastic integral

$$\int_0^T \theta_t dZ_t.$$

## Stochastic Integral: $\theta \in \mathcal{L}^2 \rightsquigarrow$

- ▶ It turns out that lots of interesting trading strategies don't fall into  $\mathcal{H}^2$ . Now allow  $\theta \in \mathcal{L}^2$ .
  - ▶ It's perhaps surprising that  $\int_0^T X_t^2 dt < \infty$  a.s. doesn't imply  $E\left(\int_0^T X_t^2 dt\right) < \infty$ , but you can come up with counterexamples with fat tails: e.g., assume  $\int_0^T X_t^2 dt$  is a RV that is equal to  $2^n$  with probability  $2^{-n}$ .

- ▶ For any  $n$ , define

$$\tau_n = \min\left(\inf\left\{t: \int_0^t \theta_s^2 ds = n\right\}, T\right)$$

and  $\theta_{n,t} = \theta_t 1_{t \leq \tau_n}$ . Then  $\theta_{n,t} \in \mathcal{H}^2$ , and we can define  $\int_0^t \theta_{n,s} dZ_s$  as above. (Intuitively, have defined integral up to stopping time  $\tau_n$ .)

- ▶ Since  $\theta \in \mathcal{L}^2$ ,  $\lim_{n \rightarrow \infty} \tau_n = T$ , a.s.
- ▶ We can then define

$$\int_0^t \theta_s dZ_s = \lim_{n \rightarrow \infty} \int_0^t \theta_{n,s} dZ_s$$

- ▶ The stochastic integral for a process in  $\mathcal{L}^2$  is not necessarily a martingale, but it is a local martingale.

# Outline

Background

Probability

Stochastic Processes

Stochastic Integration

**Itô Processes**

SDEs

Conditional Moments

Martingale Measures

# Itô Processes

## Definition (Itô Process)

A one-dimensional **Itô process** is a process  $S$  in  $\mathbb{R}$  such that

$$S_t = S_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dZ_s,$$

where the process  $\mu \in \mathcal{L}^1$  and the process  $\sigma \in \mathcal{L}^2$ .

- ▶ For an Itô process  $S$ , the processes  $\mu$  and  $\sigma$  are unique, in that any other processes are versions of them.
- ▶ Itô process in “differential” form:

$$dS_t = \mu_t dt + \sigma_t dZ_t$$

# Itô Processes

- ▶ If  $\mu, \sigma \in \mathcal{H}^2$ , then

$$\frac{d}{d\tau} E_t(S_\tau)|_{\tau=t} = \mu_t, \quad a.s.$$

and

$$\frac{d}{d\tau} \text{Var}_t(S_\tau)|_{\tau=t} = \sigma_t^2, \quad a.s.$$

- ▶ We refer to the process  $\mu$  as the **drift** process and to the process  $\sigma$  as the **diffusion** process.

# Multi-Dimensional Itô Processes ↗

## Definition (Multi-Dimensional Itô Process)

A multi-dimensional **Itô process** is a process  $S$  in  $\mathbb{R}^N$  such that

$$S_t = S_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dZ_s,$$

where the process  $Z$  is a Brownian motion in  $\mathbb{R}^d$ , the process  $\mu \in (\mathcal{L}^1)^N$  and the process  $\sigma \in (\mathcal{L}^2)^{N \times d}$ .

- Conditional covariance of the vector Itô Process is

$$\frac{d}{d\tau} \text{Cov}_t(S_\tau) |_{\tau=t} = \sigma_t \sigma_t', \quad a.s.,$$

where  $\text{Cov}_t$  is the covariance matrix, and  $\sigma_t'$  the transpose of the matrix  $\sigma_t$ .



## Itô's Lemma: Motivation ↗

- ▶ We're going to be working a lot with stochastic integrals of Itô processes. (E.g., take  $S_t$  to be a price process, and again want to know how wealth accumulates over time.)
- ▶ It turns out stochastic integrals behave “unusually” relative to our standard rules of integration. For example, take  $S_t = Z_t$  (i.e., the Itô process is just a standard Brownian motion), and say we want to know  $\int_0^t Z_s dZ_s$ .
  - ▶ For example, imagine  $Z_t$  is the short rate (which follows a random walk), and you invest  $Z_t$  units in the short rate at every point in time. (Strange trading strategy, but will be useful for exposition.)
- ▶ If the usual rules of calculus applied, we could set  $dZ_t = \frac{dZ}{dt} dt$ , and then from the chain rule,  $Z_t dZ_t = \left( Z_t \frac{dZ}{dt} \right) dt = \frac{1}{2} \left( \frac{d}{dt} Z_t^2 \right) dt$ . This would suggest that  $\int_0^t Z_s dZ_s = \frac{1}{2} \int_0^t \frac{d}{ds} Z_s^2 ds = \frac{1}{2} Z_t^2$ .
- ▶ This turns out to be wrong in an important way. The time derivative  $\frac{dZ}{dt}$  doesn't exist. More importantly, we know this naive procedure gives us an incorrect answer: stochastic integrals must be local martingales, and  $\frac{1}{2} Z_t^2$  isn't (it grows in expectation smoothly over  $t$ ).
- ▶ We want some method to get the right answer without having to analytically solve for every stochastic integral. What we'll use is, in effect, a chain rule for stochastic calculus.

# Itô's Lemma

## Itô's Lemma (One-Dimensional Case)

Let  $S$  be an Itô process:

$$dS_t = \mu_t dt + \sigma_t dZ_t.$$

Then any process  $C$  defined as a function  $C_t = f(t, S_t)$ , with  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  twice continuously differentiable, is an Itô process with

$$\begin{aligned} dC_t &= f_t(t, S_t)dt + f_S(t, S_t)dS_t + \frac{1}{2}f_{SS}(t, S_t)\sigma_t^2 dt \\ &= \left( f_t(t, S_t) + f_S(t, S_t)\mu_t + \frac{1}{2}f_{SS}(t, S_t)\sigma_t^2 \right) dt + f_S(t, S_t)\sigma_t dZ_t, \end{aligned}$$

where  $f_x$  denotes  $\frac{\partial f}{\partial x}$ .

- **Notation:** Often write  $\mathcal{D}_{S,t}f(t, S_t) = f_t(t, S_t) + f_S(t, S_t)\mu_t + \frac{1}{2}f_{SS}(t, S_t)\sigma_t^2$ , so above becomes  $dC_t = \mathcal{D}_{S,t}f(t, S_t)dt + f_S(t, S_t)\sigma_t dZ_t$ .
- Multi-dimensional case:  $dC_t = f_t(t, S_t)dt + f_S(t, S_t)dS_t + \frac{1}{2}\text{trace}[\sigma_t \sigma_t' f_{SS}(t, S_t)] dt$ .

## Itô's Lemma: Heuristic Proof $\leadsto$

- Consider a second-order Taylor expansion of  $dC_t$  (more precisely, expand  $C_{t+dt} = f(t+dt, S_t + dS_t)$  around  $C_t = f(t, S_t)$ , and express  $dC_t \approx C_{t+dt} - C_t$  for  $dt \approx 0, dS_t \approx 0$ ):

$$dC_t = f_S(t, S_t) dS_t + f_t(t, S_t) dt \\ + \frac{1}{2} \left[ f_{SS}(t, S_t) (dS_t)^2 + \underbrace{2f_{St}(t, S_t) (dS_t dt) + f_{tt}(t, S_t) (dt)^2}_{o(dt)} \right].$$

- The first term in the brackets is

$$f_{SS}(t, S_t) (dS_t)^2 = f_{SS}(t, S_t) (\mu_t dt + \sigma_t dZ_t)^2 \\ = f_{SS}(t, S_t) \left[ \underbrace{\mu_t^2 (dt)^2 + 2\mu_t \sigma_t (dt dZ_t)}_{o(dt)} + \sigma_t^2 (dZ_t)^2 \right].$$

- The term  $(dZ_t)^2$  is of order  $dt$  (in fact  $(dZ_t)^2 = dt$ ), so it stays in. Why? Recall quadratic variation of Brownian motion.

## Itô's Lemma: Takeaways $\leadsto$

$$dC_t = f_t(t, S_t)dt + f_S(t, S_t)dS_t + \frac{1}{2}f_{SS}(t, S_t)\sigma_t^2 dt$$

- ▶ Gives us a chain rule for stochastic calculus; the new term is in red.
- ▶ The result tells us how to go from an Itô process  $S_t$  directly to a new process  $C_t = f(t, S_t)$  (in differential form), which saves an enormous amount of work.
- ▶ What's the intuition? Set drift  $\mu_t = 0$  and set  $f_t = 0$  (holding  $S_t$  fixed,  $C$  doesn't depend on  $t$ ). Then  $E[dC_t] = \frac{1}{2}f_{SS}(t, S_t)\sigma_t^2 dt$  expresses what amounts to a Jensen's inequality effect: if  $f$  is concave in  $S_t$ , then  $C_t$  is expected to fall on average due to variation in  $S_t$ . (Try for  $C_t = \log(S_t)$ .)
- ▶ Let's go back to our example:  $\int_0^t Z_s dZ_s$ . Set  $S_t = Z_t$ ,  $C_t = S_t^2/2 - t/2 = Z_t^2/2 - t/2$ , and apply Itô's Lemma to get  $dC_t = Z_t dZ_t + dt/2 - dt/2 = Z_t dZ_t$ . Then  $C_t = \int_0^t dC_s = \int_0^t Z_s dZ_s$ , and we know  $C_t = Z_t^2 - t/2$ , so  $\int_0^t Z_s dZ_s = Z_t^2 - t/2$ .
  - ▶ Same as before but with correction term  $-t/2$ , which turns the integral into a local martingale.

# Outline

Background

Probability

Stochastic Processes

Stochastic Integration

Itô Processes

**SDEs**

Conditional Moments

Martingale Measures

# Stochastic Differential Equations

## Definition (SDE)

The Itô process  $X_t$  satisfies a stochastic differential equation

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dZ_t, \quad X_0 = x_0$$

if it satisfies

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

- ▶ Relative to general Itô process, have restricted  $\mu_t = \mu(t, X_t)$ ,  $\sigma_t = \sigma(t, X_t)$ , so that the process is now “self-contained.”
- ▶ To guarantee that there exists a unique solution to an SDE, one must impose restrictions on the coefficients (can't be too large or vary too much). This is similar to the case for ODEs.

# Common SDEs

## Example (Arithmetic Brownian Motion)

Solve the following SDE:

$$dX_t = \mu dt + \sigma dZ_t$$

$$X_t = X_0 + \mu t + \sigma Z_t$$

## Example (Geometric Brownian Motion)

Solve the following SDE:

$$dX_t = \mu X_t dt + \sigma X_t dZ_t$$

$$X_t = X_0 \exp\left((\mu - \sigma^2/2)t + \sigma Z_t\right)$$

# Numerical Solutions of SDEs ↗

- ▶ Except for a few special cases, SDEs do not have explicit solutions.
- ▶ The most basic and common method of approximating solutions of SDEs numerically is using the first-order Euler scheme.
- ▶ Use the grid  $t_i = i\Delta t$ .

$$\hat{X}_{i+1} = \hat{X}_i + \mu(t_i, \hat{X}_i) \Delta t + \sigma(t_i, \hat{X}_i) \sqrt{\Delta t} \tilde{\epsilon}_i,$$

where  $\tilde{\epsilon}_i$  are IID  $\mathcal{N}(0, 1)$  random variables.

- ▶ Higher-order schemes; e.g., Milstein scheme (includes second-order Itô term).



# Outline

Background

Probability

Stochastic Processes

Stochastic Integration

Itô Processes

SDEs

**Conditional Moments**

Martingale Measures

# Conditional Moments of Diffusion Processes

- ▶ We refer to the (possibly unknown) solution of an SDE as a **diffusion process**.
- ▶ Often need to compute conditional moments of diffusion processes:
  - ▶ expected returns and variances of returns on assets over finite time intervals;
  - ▶ using GMM to estimate a diffusion process from discretely sampled data;
  - ▶ prices of derivatives.
- ▶ One possible approach: formulate the problem as a partial differential equation. We'll see how this works by considering the Kolmogorov backward equation, a particular transformation of the diffusion process into a PDE that can sometimes be solved analytically.

## Kolmogorov Backward Equation

- ▶ Start with a diffusion process  $X$  with coefficients  $\mu(t, X)$  and  $\sigma(t, X)$ .
- ▶ Objective: compute a conditional expectation

$$f(t, X_t) = E[g(X_T) | X_t].$$

- ▶ Note that diffusion processes are Markov, so can condition on ( $\sigma$ -algebra generated by)  $X_t$  rather than full path from 0 to  $t$ .
- ▶ Suppose  $f(t, x)$  is a smooth function of  $t$  and  $x$ . By the law of iterated expectations,

$$f(t, X_t) = E_t[f(t + dt, X_{t+dt})] \Rightarrow E_t[df(t, X_t)] = 0.$$

- ▶ Using Itô's Lemma,  $E_t[df(t, X_t)] = (f_t(t, X_t) + f_X(t, X_t)\mu(t, X_t) + \frac{1}{2}f_{XX}(t, X_t)\sigma(t, X_t)^2) dt = 0$

## Kolmogorov Backward Equation

$$f_t + \mu(t, x)f_x + \frac{1}{2}\sigma(t, x)^2f_{xx} = 0, \text{ with boundary condition } f(T, x) = g(x).$$

## Application: Diffusion Estimation from Discrete Data ↗

Say we want to estimate mean-reverting diffusion process from discretely sampled data via MLE.

- ▶ Observe realizations  $X_n$ ,  $n = 0, 1, \dots, N$ .
- ▶ Diffusion coefficients known in parametric form:  $\mu(t, X; \theta)$  and  $\sigma(t, X; \theta)$ .
- ▶ In order to conduct MLE to estimate  $\theta$ , need to know transition density

$$p(X_n, X_{n+1}; \theta).$$

- ▶ Can solve for transition density (conditional on  $\theta$ ) using Kolmogorov backward equation and then maximize likelihood.

# Diffusion Estimation for Mean-Reverting Process $\leadsto$

## Example (Ornstein–Uhlenbeck Process)

$$dX_t = -\lambda(X_t - \bar{X}) dt + \sigma dZ_t$$

Let's say we want to solve not just for conditional expectations, but transition densities. To characterize the transition density, compute the characteristic function

$$\phi(x, w) = E \left[ e^{iwx_T} | X_0 = x \right], \quad i = \sqrt{-1}.$$

Define

$$f(t, x; w) = E \left[ e^{iwx_T} | X_t = x \right] \Rightarrow \phi(x, w) = f(0, x).$$

$f(t, x)$  satisfies the Kolmogorov backward equation

$$f_t - f_x \lambda(x - \bar{X}) + f_{xx} \frac{\sigma^2}{2} = 0, \quad f(T, x) = e^{iwx}.$$

Look for a solution of the form  $f(t, x) = \exp[a_0(t) + a_1(t)x]$ .

## Diffusion Estimation for Mean-Reverting Process (Continued) ⇆

### Example

The characteristic function of a Gaussian random variable  $\mathcal{N}(\mu, \sigma^2)$  is given by

$$\exp \left[ iw\mu - w^2 \frac{\sigma^2}{2} \right]$$

⇒ O-U transition density is Gaussian with mean

$$\bar{X} + (x - \bar{X})e^{-\lambda T}$$

and variance

$$\frac{\sigma^2}{2\lambda} \left( 1 - e^{-2\lambda T} \right).$$

In the discretely sampled MLE example,  $T$  can be interpreted as the time between successive observations of  $X_t$ .

# Outline

Background

Probability

Stochastic Processes

Stochastic Integration

Itô Processes

SDEs

Conditional Moments

**Martingale Measures**

# Martingale Representation

- ▶ As above, stochastic integrals are local martingales.
- ▶ The reverse is also true: all local martingales can be represented as stochastic integrals. This will be useful for representing payoffs from dynamic trading.

## Martingale Representation Theorem

If  $X$  is a local martingale adapted to  $\mathbb{F}$ , then there exists a process  $\theta \in (\mathcal{L}^2)^d$  such that

$$X_t = X_0 + \int_0^t \theta_s dZ_s.$$

If  $X_t$  is a martingale with finite variance for any  $t$ , then martingale representation can be achieved with  $\theta \in (\mathcal{H}^2)^d$ .

- ▶ Tells us that every (local) martingale can be written as a stochastic integral, so the Brownian motion “spans” all the (local) martingales.
- ▶ Further, all martingales adapted to filtration generated by Brownian motion have continuous sample paths. Informally, the Brownian filtration represents a world with continuous information arrival.



## Changes of Measure: Some Basic Results

- ▶ Consider a probability space  $(\Omega, \mathcal{F}, P)$ .
- ▶  $Y > 0$  is  $\mathcal{F}$ -measurable, and  $E^P(Y) = 1$  (where  $E^P$  denotes the expectation under  $P$ , which we've previously “hidden” since it's clear what measure we're talking about).
- ▶ Starting from  $P$ , define a new probability measure  $Q$  by

$$Q(A) = E(1_A Y), \quad \forall A \in \mathcal{F}.$$

- ▶  $Y = dQ/dP$  is the **Radon-Nikodym derivative** of  $Q$  w.r.t.  $P$ .
- ▶ We denote by  $E^Q(X)$  the expectation under measure  $Q$ .
- ▶  $P$  and  $Q$  agree on the sets of measure zero, a notion referred to as equivalence.

### Equivalent Measures

$P$  and  $Q$  are **equivalent** if  $P(A) = 0 \Leftrightarrow Q(A) = 0, \forall A \in \mathcal{F}$ .

$P$  and  $Q$  are equivalent iff  $Y > 0, P$  a.s.

# Conditional Expectation under $Q$

## Exercise

Show that the following are true:

1. For  $\forall A \in \mathcal{F}$ ,

$$E(Y|A) = \frac{Q(A)}{P(A)}$$

2. For an  $\mathcal{F}$ -measurable function  $X$ ,

$$E^Q(X) = E(XY)$$

3. If  $P$  and  $Q$  are equivalent,

$$E^Q(X|A) = \frac{E(XY|A)}{E(Y|A)}$$

# Novikov's Condition

## Definition (Novikov's Condition)

A process  $\eta \in (\mathcal{L}^2)^d$  satisfies Novikov's condition iff

$$\mathbb{E} \left[ \exp \left( \frac{1}{2} \int_0^T \|\eta_t\|^2 dt \right) \right] < \infty.$$

## Theorem (Exponential Martingales)

If the process  $\eta \in (\mathcal{L}^2)^d$  satisfies Novikov's condition, then

$$\xi_t^\eta = \exp \left( - \int_0^t \eta'_s dZ_s - \frac{1}{2} \int_0^t \|\eta_s\|^2 ds \right)$$

is a martingale.

# Girsanov's Theorem: Setup

- ▶ Assume process  $\eta$  satisfies Novikov's condition.
- ▶  $\xi^\eta > 0$  is a martingale,  $\xi_0^\eta = 1 \Rightarrow E(\xi_T^\eta) = 1$ .
- ▶ Define a probability measure  $Q^\eta$  by

$$Q^\eta(A) = E(1_A \xi_T^\eta),$$

i.e. by  $dQ^\eta/dP = \xi_T^\eta$ .

- ▶  $P$  and  $Q^\eta$  are equivalent.

# Girsanov's Theorem

## Theorem (Girsanov's Theorem)

- ▶ Assume  $\eta \in (\mathcal{L}^2)^d$  is such that  $\xi^\eta$  is a martingale. Then the process  $Z^\eta$  defined by

$$Z_t^\eta = Z_t + \int_0^t \eta_s ds,$$

is a Brownian motion under  $Q^\eta$ .

- ▶  $Z^\eta$  has a martingale representation property under  $Q^\eta$ : for any local  $Q^\eta$ -martingale  $M_t$  adapted to  $\mathbb{F}$ , there exists a process  $\phi \in (\mathcal{L}^2)^d$ , such that

$$M_t = M_0 + \int_0^t \phi_s dZ_s^\eta.$$

- ▶ Both parts are important, but the first part is usually emphasized: tells us that by adding drift  $\eta_t$  to Brownian motion under  $P$ , we get a Brownian motion under the new measure  $Q^\eta$ .
- ▶ Will be crucial for risk-neutral pricing:  $Q^\eta$  will be closely related to risk-neutral measure.

# Girsanov's Theorem: Implication

## Theorem

Consider an Itô process  $S$  in  $\mathbb{R}^N$ ,

$$S_t = S_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dZ_s$$

and processes  $v \in (\mathcal{L}^1)^N$  and  $\eta \in (\mathcal{L}^2)^d$  such that

$$\sigma_t \eta_t = \mu_t - v_t.$$

If the process  $\xi^\eta$  is a martingale, then  $S$  is an Itô process under  $Q^\eta$ , and

$$S_t = S_0 + \int_0^t v_s ds + \int_0^t \sigma_s dZ_s^\eta.$$

**Diffusion Invariance Principle:** Under any equivalent probability measure, the diffusion part of the process  $S$  stays the same, and only the drift changes.

# Girsanov's Theorem: Implication

## Theorem

Consider an Itô process  $S$  in  $\mathbb{R}^N$ ,

$$S_t = S_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dZ_s$$

and processes  $\nu \in (\mathcal{L}^1)^N$  and  $\eta \in (\mathcal{L}^2)^d$  such that

$$\sigma_t \eta_t = \mu_t - \nu_t.$$

If the process  $\xi^\eta$  is a martingale, then  $S$  is an Itô process under  $Q^\eta$ , and

$$S_t = S_0 + \int_0^t \nu_s ds + \int_0^t \sigma_s dZ_s^\eta.$$

Can see where we're headed: Will pick  $\nu_t = 0 \iff \mu_t - \sigma_t \eta_t = 0$ , so discounted return process will be a martingale under  $Q^\eta$ .