Brownian Motion Market Model: From Basics

Author: Batuhan Koyuncu (degeneratefermion)

We start with the simplest stochastic market model.

- We assume the market price moves randomly, like a Brownian particle.

1. Price as a Dynamical Variable

Let

$$S(t)$$

be the price of an asset at time t.

Examples:

$$S(t) = \text{BTC price at time } t$$

or

$$S(t) = \text{stock price at time } t$$

We want to understand how

$$S(t)$$

evolves.

In deterministic mechanics one writes

$$\frac{dS}{dt} = v(S, t)$$

where v is a velocity field.

Markets are uncertain, so we need stochastic dynamics.

2. Brownian Motion

Introduce Brownian motion

$$W(t)$$

with properties:

Initial condition

$$W(0) = 0$$

Independent increments

For times

$$t_1 < t_2 < t_3 < t_4$$

the increments

$$W(t_2) - W(t_1)$$

and

$$W(t_4) - W(t_3)$$

are independent.

Independent increments

For times

$$t_1 < t_2 < t_3 < t_4$$

consider two non-overlapping increments:

$$X = W(t_2) - W(t_1)$$

and

$$Y = W(t_4) - W(t_3).$$

Brownian motion assumes these increments are **independent random variables**.

What does independence mean?

Two random variables

$$X$$

and

$$Y$$

are independent if knowing one gives no information about the other.

Mathematically, for any sets

$$A, B$$

independence means

$$P(X \in A,\ Y \in B) = P(X \in A)\, P(Y \in B).$$

So the joint probability factorizes.

Equivalent density form:

if

$$f_{X,Y}(x, y)$$

is the joint probability density, independence means

$$f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Applied to Brownian increments

Thus

$$W(t_2) - W(t_1)$$

is statistically independent from

$$W(t_4) - W(t_3)$$

whenever the intervals

$$[t_1, t_2]$$

and

$$[t_3, t_4]$$

do not overlap.

So future random shocks do not depend on past shocks.

Symbolically,

$$P(X \in A \mid Y \in B) = P(X \in A).$$

Conditional probabilities equal unconditional ones.

That is another definition of independence.

Consequence for correlations

Independence implies zero covariance:

$$\mathrm{Cov}(X, Y) = 0.$$

Since

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

we have

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

For Brownian increments,

$$\mathbb{E}[X] = \mathbb{E}[Y] = 0$$

so

$$\mathbb{E}[XY] = 0.$$

No correlation.

(Important: in general zero correlation does **not** imply independence, but independence does imply zero correlation.)

Physical interpretation

This is the Markov-type "memoryless" property.

The next fluctuation

$$dW(t)$$

does not remember previous fluctuations.

That is why Brownian motion is often called a random walk with no memory.

Gaussian increments

For small

$$\Delta t$$

we have

$$W(t+\Delta t) - W(t) \sim \mathcal{N}(0, \Delta t)$$

so

$$\mathbb{E}[\Delta W] = 0$$

and

$$\mathrm{Var}(\Delta W) = \Delta t.$$

Equivalent representation:

$$\Delta W = \sqrt{\Delta t} Z$$

with

$$Z \sim \mathcal{N}(0, 1).$$

3. Arithmetic Brownian Motion

First model:

$$dS = \mu dt + \sigma dW$$

Parameters are summarized in Table 1.

Symbol	**Meaning**
$S(t)$	Asset price at time t
μ	Drift parameter (average trend)
σ	Volatility parameter
dW	Brownian stochastic shock increment

Table 1. Fundamental parameters of the Brownian motion market model.

Discrete form:

$$S_{n+1} = S_n + \mu\Delta t + \sigma\sqrt{\Delta t} Z_n$$

with

$$Z_n \sim \mathcal{N}(0,1).$$

Interpretation:

$$\Delta S = \mu\Delta t + \sigma\sqrt{\Delta t} Z$$

so

$$\text{price change} = \text{trend} + \text{noise}.$$

4. Mean and Variance

Expectation:

$$\mathbb{E}[\Delta S] = \mu\Delta t$$

Variance:

$$\mathrm{Var}(\Delta S) = \sigma^2\Delta t.$$

Therefore

$$\mathbb{E}[S(t)] = S(0) + \mu t$$

and

$$\mathrm{Var}(S(t)) = \sigma^2 t.$$

Uncertainty grows as

$$\sigma\sqrt{t}.$$

Square-root law.

5. Problem of Arithmetic Brownian Motion

Solution:

$$S(t) = S(0) + \mu t + \sigma W(t)$$

can become negative.

But markets require

$$S(t) > 0.$$

Therefore this model is incomplete.

6. Geometric Brownian Motion

Instead model relative returns:

$$\frac{dS}{S} = \mu dt + \sigma dW$$

equivalently

$$dS = \mu S dt + \sigma S dW.$$

This is geometric Brownian motion.

Interpretation:

$$\text{return} = \text{drift} + \text{noise}.$$

7. Log-Price

Define

$$q(t) = \ln S(t).$$

Using Itô calculus:

$$dq = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW.$$

Define effective drift

$$a = \mu - \frac{1}{2}\sigma^2.$$

Then

$$dq = a dt + \sigma dW.$$

Now log-price follows ordinary Brownian motion with drift.

8. Solution

Integrating:

$$q(t) - q(0) = at + \sigma W(t).$$

Thus

$$q(t) = q(0) + at + \sigma W(t).$$

Since

$$q(0) = \ln S(0)$$

we get

$$\ln S(t) = \ln S(0) + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t).$$

Exponentiate:

$$S(t) = S(0) \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)\right].$$

Always positive.

9. Distribution

Since

$$W(t) \sim \mathcal{N}(0, t)$$

then

$$\ln S(t) \sim \mathcal{N}\left(\ln S(0) + \left(\mu - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right)$$

Therefore $S(t)$ is *lognormal*.

10. Simulation Formula

Discrete update:

$$q_{n+1} = q_n + \left(\mu - \frac{1}{2}\sigma^2\right) \Delta t + \sigma\sqrt{\Delta t} Z_n$$

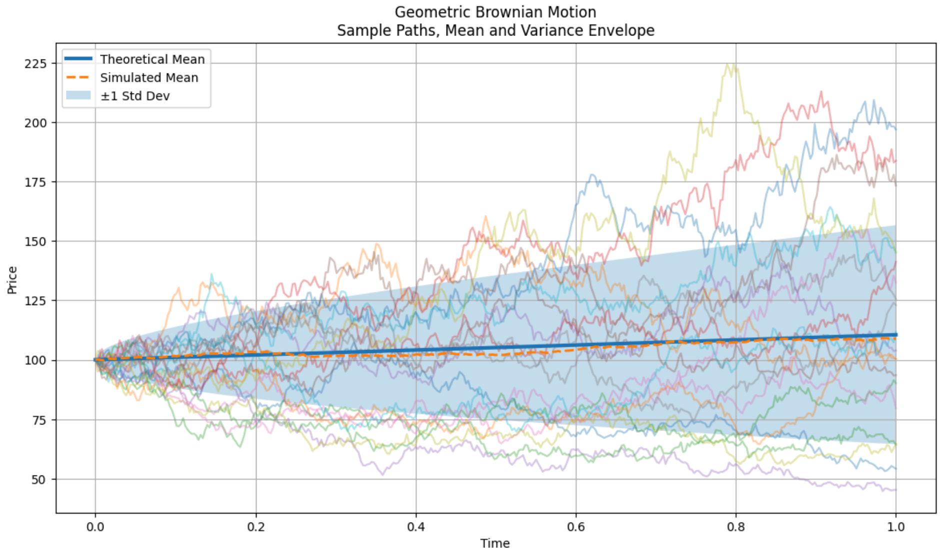
with

$$Z_n \sim \mathcal{N}(0, 1).$$

Equivalent price form:

$$S_{n+1} = S_n \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right) \Delta t + \sigma\sqrt{\Delta t} Z_n\right].$$

Fundamental Monte Carlo equation.




```
At t = T

Theoretical Mean:
110.51709180756477

Simulated Mean:
108.79803124765081

Theoretical Variance:
2119.266564001705

Empirical Variance:
1914.3203890973905
```

12. Expected Price

For geometric Brownian motion,

$$dS = \mu S dt + \sigma S dW$$

we found the exact solution

$$S(t) = S(0) \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)\right].$$

We now compute its expectation value.

Mean Price

Take expectation:

$$\mathbb{E}[S(t)] = \mathbb{E}\left[S(0) \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma W(t)\right)\right].$$

Factor out deterministic terms:

$$\mathbb{E}[S(t)] = S(0) e^{(\mu - \frac{1}{2}\sigma^2)t} \mathbb{E}\left[e^{\sigma W(t)}\right].$$

Now use

$$W(t) \sim \mathcal{N}(0, t).$$

For a Gaussian variable

$$X \sim \mathcal{N}(0, \nu^2)$$

one has

$$\mathbb{E}[e^{aX}] = e^{\frac{1}{2}a^2\nu^2}.$$

Using

$$a = \sigma$$

and

$$\nu^2 = t$$

gives

$$\mathbb{E}[e^{\sigma W(t)}] = e^{\frac{1}{2}\sigma^2 t}.$$

Therefore

$$\mathbb{E}[S(t)] = S(0) e^{(\mu - \frac{1}{2}\sigma^2)t} e^{\frac{1}{2}\sigma^2 t}.$$

The volatility terms cancel:

$$\boxed{\mathbb{E}[S(t)] = S(0)e^{\mu t}}$$

This is the mean price.

Median Price

The median corresponds to the center of the lognormal distribution.

Since

$$\ln S(t) \sim \mathcal{N}\left(\ln S(0) + \left(\mu - \frac{1}{2}\sigma^2\right)t, \sigma^2 t\right),$$

the median is

$$\boxed{\text{Median}[S(t)] = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t}}$$

Mean vs Typical Path

Observe:

Mean:

$$S(0)e^{\mu t}$$

Typical (median):

$$S(0)e^{(\mu - \frac{1}{2}\sigma^2)t}$$

Difference:

$$\frac{1}{2}\sigma^2.$$

This is the volatility correction.

Higher volatility lowers the typical growth rate.

Interpretation

Large upward fluctuations pull the mean upward.

Therefore:

$$\text{mean} > \text{median}$$

for

$$\sigma > 0.$$

This is a fundamental asymmetry of multiplicative stochastic growth.

Example

Suppose

$$\mu = 0.10$$

and

$$\sigma = 0.40.$$

Then

Mean growth:

$$e^{0.10} \approx 1.105$$

about

$$10.5\%$$

Median growth:

$$e^{0.10-\frac{1}{2}(0.4)^2} = e^{0.02} \approx 1.020$$

only

$$2\%$$

Huge difference.

Volatility eats typical growth.

Variance and Standard Deviation of Geometric Brownian Motion

We already found the exact solution of

$$dS = \mu S dt + \sigma S dW$$

to be

$$S(t) = S_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W(t)\right].$$

We now derive the variance and standard deviation.

Mean

Previously we showed

$$\mathbb{E}[S(t)] = S_0 e^{\mu t}.$$

Second Moment

To obtain the variance we first compute

$$\mathbb{E}[S(t)^2].$$

Square the solution:

$$S(t)^2 = S_0^2 \exp\left[2\left(\mu - \frac{1}{2}\sigma^2\right)t + 2\sigma W(t)\right].$$

Simplify:

$$S(t)^2 = S_0^2 \exp\left[(2\mu - \sigma^2)t + 2\sigma W(t)\right].$$

Take expectation:

$$\mathbb{E}[S(t)^2] = S_0^2 e^{(2\mu-\sigma^2)t}\,\mathbb{E}\left[e^{2\sigma W(t)}\right].$$

Using

$$W(t) \sim \mathcal{N}(0, t)$$

and the Gaussian identity

$$\mathbb{E}[e^{aX}] = e^{\frac{1}{2}a^2\nu^2}$$

with

$$a = 2\sigma$$

and

$$\nu^2 = t$$

gives

$$\mathbb{E}[e^{2\sigma W(t)}] = e^{2\sigma^2 t}.$$

Therefore

$$\mathbb{E}[S(t)^2] = S_0^2 e^{(2\mu-\sigma^2)t} e^{2\sigma^2 t}.$$

So

$$\mathbb{E}[S(t)^2] = S_0^2 e^{(2\mu+\sigma^2)t}$$

Variance

Definition:

$$\mathrm{Var}(S(t)) = \mathbb{E}[S(t)^2] - (\mathbb{E}[S(t)])^2.$$

Substitute:

$$\mathrm{Var}(S(t)) = S_0^2 e^{(2\mu+\sigma^2)t} - S_0^2 e^{2\mu t}.$$

Factor out

$$S_0^2 e^{2\mu t}.$$

Then

$$\boxed{\mathrm{Var}(S(t)) = S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)}$$

This is the exact variance.

Standard Deviation

The standard deviation is the square root:

$$\sigma_S(t) = \sqrt{\mathrm{Var}(S(t))}.$$

Thus

$$\sigma_S(t) = S_0 e^{\mu t}\sqrt{e^{\sigma^2 t} - 1}$$

This is the volatility band of the price process.

Short-Time Limit

For small

$$t$$

use

$$e^{\sigma^2 t} \approx 1 + \sigma^2 t.$$

Then

$$e^{\sigma^2 t} - 1 \approx \sigma^2 t.$$

Hence

$$\sigma_S(t) \approx S_0 e^{\mu t}\sigma\sqrt{t}.$$

So fluctuations grow like

$$\sqrt{t}.$$

Diffusive scaling again.

Summary

The moments of geometric Brownian motion are

Mean:

$$\mathbb{E}[S(t)] = S_0 e^{\mu t}$$

Variance:

$$\mathrm{Var}(S(t)) = S_0^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)$$

Standard deviation:

$$\sigma_S(t) = S_0 e^{\mu t}\sqrt{e^{\sigma^2 t} - 1}$$

Interpretation

The center of the probability distribution follows

$$S_0 e^{\mu t}$$

while uncertainty spreads according to

$$\sigma_S(t).$$

This is analogous to a drifting and spreading wave packet in statistical physics.

Deriving the Log-Price Equation Step by Step

We start from geometric Brownian motion:

$$dS = \mu S dt + \sigma S dW.$$

Equivalently,

$$\frac{dS}{S} = \mu dt + \sigma dW.$$

Now define the log-price

$$q = \ln S.$$

We want to find the stochastic differential equation for

$$dq.$$

1. Why Ordinary Calculus Is Not Enough

In ordinary calculus, one might write

$$dq = d(\ln S) = \frac{1}{S} dS.$$

This would give

$$dq = \mu dt + \sigma dW.$$

But this is incomplete in stochastic calculus.

The reason is that Brownian motion is very rough. Its increments scale as

$$dW \sim \sqrt{dt}.$$

Therefore

$$(dW)^2 \sim dt.$$

So second-order terms in

$$dW$$

cannot be ignored.

This is the key difference from ordinary calculus.

2. Itô Rules

In Itô calculus, we keep terms up to order

$$dt.$$

The basic multiplication rules are

$$(dt)^2 = 0,$$

$$dt\,dW = 0,$$

and

$$(dW)^2 = dt.$$

The logic is the following.

Since

$$dW \sim \sqrt{dt},$$

we have

$$(dW)^2 \sim dt.$$

So

$$(dW)^2$$

survives.

But

$$dt\,dW \sim dt^{3/2}$$

is much smaller than

$$dt$$

and is discarded.

Also,

$$(dt)^2$$

is even smaller and is discarded.

Therefore, in stochastic calculus,

$$(dW)^2$$

is the only second-order term that contributes at order

$$dt.$$

3. Itô's Lemma

Suppose a stochastic variable satisfies

$$dS = a(S,t)dt + b(S,t)dW.$$

Let

$$q = f(S, t)$$

be a differentiable function of

$$S$$

and

$$t.$$

Then Itô's lemma says

$$dq = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial S} dS + \frac{1}{2} \frac{\partial^2 f}{\partial S^2} (dS)^2.$$

This is like Taylor expansion, but with the important stochastic correction term

$$\frac{1}{2} \frac{\partial^2 f}{\partial S^2} (dS)^2.$$

In ordinary calculus this term would vanish.

In Itô calculus it can survive because

$$(dW)^2 = dt.$$

4. Apply Itô's Lemma to the Log-Price

Here our function is

$$f(S) = \ln S.$$

So

$$q = f(S) = \ln S.$$

The derivatives are

$$\frac{\partial f}{\partial t} = 0,$$

$$\frac{\partial f}{\partial S} = \frac{1}{S},$$

and

$$\frac{\partial^2 f}{\partial S^2} = -\frac{1}{S^2}.$$

From geometric Brownian motion,

$$dS = \mu S dt + \sigma S dW.$$

So we identify

$$a(S, t) = \mu S$$

and

$$b(S,t) = \sigma S.$$

5. Compute ((dS)^2)

We have

$$dS = \mu S dt + \sigma S dW.$$

Square it:

$$(dS)^2 = (\mu S dt + \sigma S dW)^2.$$

Expand:

$$(dS)^2 = \mu^2 S^2 (dt)^2 + 2\mu\sigma S^2 dt\, dW + \sigma^2 S^2 (dW)^2.$$

Using Itô rules,

$$(dt)^2 = 0,$$

$$dt\, dW = 0,$$

and

$$(dW)^2 = dt.$$

Therefore,

$$(dS)^2 = \sigma^2 S^2 dt.$$

6. Substitute into Itô's Lemma

Itô's lemma gives

$$dq = \frac{1}{S} dS + \frac{1}{2}\left(-\frac{1}{S^2}\right)(dS)^2.$$

Now substitute

$$dS = \mu S dt + \sigma S dW$$

and

$$(dS)^2 = \sigma^2 S^2 dt.$$

First term:

$$\frac{1}{S} dS = \frac{1}{S}(\mu S dt + \sigma S dW).$$

So

$$\frac{1}{S} dS = \mu dt + \sigma dW.$$

Second term:

$$\frac{1}{2}\left(-\frac{1}{S^2}\right)\sigma^2 S^2 dt = -\frac{1}{2}\sigma^2 dt.$$

Therefore,

$$dq = \mu dt + \sigma dW - \frac{1}{2}\sigma^2 dt.$$

Combine the deterministic drift terms:

$$dq = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW.$$

7. Final Result

The log-price obeys

$$dq = v dt + \sigma dW,$$

where

$$v = \mu - \frac{1}{2}\sigma^2.$$

Thus the effective drift of the log-price is

$$\mu - \frac{1}{2}\sigma^2.$$

The correction term

$$-\frac{1}{2}\sigma^2$$

is called the Itô correction.

It appears because the price process is both noisy and multiplicative.

8. Interpretation

The original price process is

$$dS = \mu S dt + \sigma S dW.$$

This says that price changes are proportional to the current price.

So the model is multiplicative.

But the logarithm transforms multiplicative growth into additive growth:

$$q = \ln S.$$

The resulting log-price dynamics are

$$dq = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW.$$

So in log-price space, the market behaves like a Brownian particle with drift

$$v = \mu - \frac{1}{2}\sigma^2$$

and diffusion strength

$$\sigma.$$

Consider again

$$dq = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW.$$

This is mathematically identical to diffusion with drift.

Drift Term

Define

$$v = \mu - \frac{1}{2}\sigma^2.$$

Then

$$dq = vdt + \sigma dW.$$

The deterministic piece

$$vdt$$

acts like constant velocity.

In mechanics this resembles a particle drifting under a uniform force.

Diffusion Term

The stochastic term

$$\sigma dW$$

causes random spreading.

Variance evolves as

$$\mathrm{Var}(q(t)) = \sigma^2 t.$$

Standard deviation:

$$\sigma\sqrt{t}.$$

Diffusive scaling.

The correspondence between financial quantities and statistical physics concepts is summarized in Table 2.

Finance	Statistical Physics
log-price q	particle position
drift v	constant force / velocity
volatility σ	diffusion coefficient
Brownian noise	thermal fluctuations
return distribution	probability density
ensemble of paths	statistical ensemble

Table 2. Mapping between financial stochastic processes and statistical physics analogues.

This is literally a diffusion process.

Fokker-Planck Equation

Let

$$P(q, t)$$

be the probability density for log-price.

The one-dimensional Fokker-Planck equation is:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial q}[A(q)P(q,t)] + \frac{1}{2}\frac{\partial^2}{\partial q^2}[B(q)P(q,t)] \,.$$

For our model,

$$A(q) = v$$

and

$$B(q) = \sigma^2.$$

Since both are constants,

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial q}[vP(q,t)] + \frac{1}{2}\frac{\partial^2}{\partial q^2}\left[\sigma^2 P(q,t)\right] \,.$$

Thus we obtain

$$\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial q} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial q^2}$$

This is the Fokker-Planck equation for the log-price under geometric Brownian motion.

It satisfies

$$\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial q} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial q^2}.$$

Terms:

First:

$$-v\frac{\partial P}{\partial q}$$

drift transport.

Second:

$$\frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial q^2}$$

diffusion.

This is the advection-diffusion equation.

Gaussian Solution: Explicit Derivation

We want to solve the Fokker-Planck equation

$$\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial q} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial q^2}.$$

The initial condition is

$$P(q,0) = \delta(q - q_0).$$

This means that at time

$$t = 0$$

the log-price is known exactly:

$$q(0) = q_0.$$

1. Remove the Drift

The term

$$-v\frac{\partial P}{\partial q}$$

represents translation of the probability packet with velocity

$$v.$$

Define a moving coordinate

$$x = q - vt.$$

Equivalently,

$$q = x + vt.$$

Now define a new probability density

$$\rho(x,t) = P(q,t) = P(x+vt,t).$$

So

$$P(q,t) = \rho(q-vt,t).$$

2. Transform Derivatives

Since

$$x = q - vt,$$

we have

$$\frac{\partial x}{\partial q} = 1$$

and

$$\frac{\partial x}{\partial t} = -v.$$

First spatial derivative:

$$\frac{\partial P}{\partial q} = \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial q} = \frac{\partial \rho}{\partial x}.$$

Second spatial derivative:

$$\frac{\partial^2 P}{\partial q^2} = \frac{\partial^2 \rho}{\partial x^2}.$$

Time derivative:

$$\frac{\partial P}{\partial t} = \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x}\frac{\partial x}{\partial t}.$$

Thus

$$\frac{\partial P}{\partial t} = \frac{\partial \rho}{\partial t} - v\frac{\partial \rho}{\partial x}.$$

3. Substitute into the Fokker-Planck Equation

Original equation:

$$\frac{\partial P}{\partial t} = -v\frac{\partial P}{\partial q} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial q^2}.$$

Substitute the transformed derivatives:

$$\frac{\partial \rho}{\partial t} - v\frac{\partial \rho}{\partial x} = -v\frac{\partial \rho}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 \rho}{\partial x^2}.$$

The drift terms cancel:

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\sigma^2 \frac{\partial^2 \rho}{\partial x^2}.$$

So in the moving frame, the equation is the ordinary diffusion equation.

Define

$$D = \frac{1}{2}\sigma^2.$$

Then

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2}.$$

4. Initial Condition in the Moving Frame

At

$$t = 0,$$

we have

$$x = q.$$

Therefore

$$\rho(x, 0) = P(x, 0).$$

Using

$$P(q, 0) = \delta(q - q_0),$$

we get

$$\rho(x, 0) = \delta(x - q_0).$$

So the initial condition remains a delta peak centered at

$$q_0.$$

5. Solve the Diffusion Equation by Fourier Transform

Use the Fourier representation

$$\rho(x, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\rho}(k, t)e^{ikx}\,dk.$$

Then

$$\frac{\partial \rho}{\partial t} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\partial \tilde{\rho}}{\partial t}e^{ikx}\,dk.$$

Also,

$$\frac{\partial^2 \rho}{\partial x^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-k^2) \tilde{\rho}(k,t) e^{ikx} \, dk.$$

Substitute into

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2}.$$

We get

$$\frac{\partial \tilde{\rho}}{\partial t} = -Dk^2 \tilde{\rho}.$$

This is an ordinary differential equation in time.

6. Solve the Fourier-Space Equation

The equation is

$$\frac{\partial \tilde{\rho}}{\partial t} = -Dk^2 \tilde{\rho}.$$

Separate variables:

$$\frac{1}{\tilde{\rho}} \frac{\partial \tilde{\rho}}{\partial t} = -Dk^2.$$

Integrate:

$$\ln \tilde{\rho}(k,t) = -Dk^2 t + C(k).$$

Therefore

$$\tilde{\rho}(k,t) = \tilde{\rho}(k,0) e^{-Dk^2 t}.$$

7. Fourier Transform of the Initial Delta Function

Initial condition:

$$\rho(x,0) = \delta(x - q_0).$$

The Fourier transform is

$$\tilde{\rho}(k,0) = \int_{-\infty}^{\infty} \delta(x - q_0) e^{-ikx} \, dx.$$

Using the delta function property,

$$\int f(x) \delta(x - q_0) \, dx = f(q_0),$$

we get

$$\tilde{\rho}(k,0) = e^{-ikq_0}.$$

Thus

$$\tilde{\rho}(k,t) = e^{-ikq_0}e^{-Dk^2t}.$$

8. Transform Back to Position Space

Now

$$\rho(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ikq_0}e^{-Dk^2t}e^{ikx}\,dk.$$

Combine the exponentials:

$$\rho(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-Dk^2t}e^{ik(x-q_0)}\,dk.$$

This is a standard Gaussian integral:

$$\int_{-\infty}^{\infty} e^{-ak^2+ibk}\,dk = \sqrt{\frac{\pi}{a}}e^{-\frac{b^2}{4a}},$$

for

$$a > 0.$$

Here,

$$a = Dt$$

and

$$b = x - q_0.$$

Therefore

$$\rho(x,t) = \frac{1}{2\pi}\sqrt{\frac{\pi}{Dt}}\exp\left[-\frac{(x-q_0)^2}{4Dt}\right].$$

Simplify:

$$\rho(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\left[-\frac{(x-q_0)^2}{4Dt}\right].$$

9. Substitute Back (D=\frac12\sigma^2)

Since

$$D = \frac{1}{2}\sigma^2,$$

we have

$$4Dt = 4\left(\frac{1}{2}\sigma^2\right)t = 2\sigma^2 t.$$

Also,

$$\sqrt{4\pi Dt} = \sqrt{2\pi\sigma^2 t}.$$

Thus

$$\rho(x,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\exp\left[-\frac{(x-q_0)^2}{2\sigma^2 t}\right].$$

10. Return to the Original Coordinate

Recall

$$x = q - vt.$$

Therefore

$$x - q_0 = q - vt - q_0.$$

So

$$x - q_0 = q - q_0 - vt.$$

Hence

$$P(q,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}}\exp\left[-\frac{(q-q_0-vt)^2}{2\sigma^2 t}\right].$$

This is the desired solution.

11. Interpretation

The probability density is Gaussian in log-price.

Its mean is

$$\mathbb{E}[q(t)] = q_0 + vt.$$

Its variance is

$$\mathrm{Var}(q(t)) = \sigma^2 t.$$

Therefore:

- the center of the packet moves with velocity

$$v,$$

- the packet spreads with width

$$\sigma\sqrt{t}.$$

So the solution describes a Gaussian packet drifting and spreading, exactly like free diffusion.

Ensemble View

One trajectory means little.

We study

$$\{q_1(t), q_2(t), \dots, q_N(t)\}.$$

As

$$N \to \infty$$

ensemble statistics converge.

This is identical in spirit to statistical mechanics.

Financial Meaning

Drift:

$$v$$

models average market tendency.

Volatility:

$$\sigma$$

models uncertainty.

Brownian noise:

$$dW$$

models information shocks.

So market motion is

$$\text{drift} + \text{diffusion}.$$

Part 1: Replace Price Modelling by Return Modelling

We start from observed market prices

$$S_0, S_1, S_2, \ldots, S_N.$$

Here:

$$S_t$$

is the asset price at discrete time step

$$t.$$

For example, if we use daily data, then

$$S_t$$

is the closing price on day

$$t.$$

If we use hourly data, then

$$S_t$$

is the closing price of hour

$$t.$$

1. Why We Do Not Model Price Directly

The price itself is usually non-stationary.

For example, a stock can move from

$$100$$

to

$$200$$

over several years.

So the scale of the price changes.

A movement of

$$1$$

dollar is not equally important at all price levels.

If the price is

$$10,$$

then a move of

$$1$$

is

$$10\%.$$

If the price is

$$1000,$$

then a move of

$$1$$

is only

$$0.1\%.$$

So absolute price changes are not the natural object.

Instead, we study relative changes.

2. Simple Return

The simple return from time

$$t-1$$

to time

$$t$$

is

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}}.$$

Equivalently,

$$R_t = \frac{S_t}{S_{t-1}} - 1.$$

So if

$$S_{t-1} = 100$$

and

$$S_t = 105,$$

then

$$R_t = \frac{105-100}{100} = 0.05.$$

So the return is

$$5\%.$$

3. Log Return

For stochastic modelling, we usually prefer the log-return:

$$r_t = \ln \frac{S_t}{S_{t-1}}.$$

This can also be written as

$$r_t = \ln S_t - \ln S_{t-1}.$$

Define the log-price:

$$q_t = \ln S_t.$$

Then

$$r_t = q_t - q_{t-1}.$$

So log-returns are increments of log-price.

This is why log-returns are natural for Brownian models.

4. Relation Between Simple Return and Log Return

Since

$$R_t = \frac{S_t}{S_{t-1}} - 1,$$

we have

$$\frac{S_t}{S_{t-1}} = 1 + R_t.$$

Taking logarithm:

$$r_t = \ln(1 + R_t).$$

For small returns,

$$\ln(1 + x) \approx x.$$

Therefore, for small daily or hourly returns,

$$r_t \approx R_t.$$

Example:

If

$$R_t = 0.01,$$

then

$$r_t = \ln(1.01) \approx 0.00995.$$

So the difference is small.

5. Why Log Returns Are Better

Log returns are additive.

Suppose price evolves from

$$S_0$$

to

$$S_1$$

to

$$S_2.$$

The first log-return is

$$r_1 = \ln \frac{S_1}{S_0}.$$

The second log-return is

$$r_2 = \ln \frac{S_2}{S_1}.$$

Add them:

$$r_1 + r_2 = \ln \frac{S_1}{S_0} + \ln \frac{S_2}{S_1}.$$

Using

$$\ln a + \ln b = \ln(ab),$$

we get

$$r_1 + r_2 = \ln\left(\frac{S_1}{S_0} \frac{S_2}{S_1} \right).$$

The

$$S_1$$

cancels:

$$r_1 + r_2 = \ln \frac{S_2}{S_0}.$$

So multi-period log-return is just the sum of one-period log-returns.

In general,

$$\sum_{t=1}^{N} r_t = \ln \frac{S_N}{S_0}.$$

This is extremely useful.

6. Brownian Model in Discrete Form

From the continuous model

$$dq = vdt + \sigma dW,$$

we get the discrete approximation

$$q_{t+1} - q_t = v\Delta t + \sigma\sqrt{\Delta t} Z_t.$$

But

$$q_{t+1} - q_t = r_{t+1}.$$

Therefore,

$$r_{t+1} = v\Delta t + \sigma\sqrt{\Delta t} Z_t.$$

If we choose

$$\Delta t = 1$$

as one time step, then

$$r_t = v + \sigma Z_t.$$

This is the basic return model.

7. Return Model

The simplest statistical model is

$$r_t = \mu_r + \sigma_r Z_t,$$

where

$$Z_t \sim \mathcal{N}(0,1).$$

The parameters are summarized in Table 1.

Symbol	Meaning
r_t	log-return at time t
S_t	asset price at time t
μ_r	average return per time step
σ_r	return volatility per time step
Z_t	standard normal random variable

Table 1. Parameters of the basic discrete return model.

8. Estimate the Mean Return

Given observed returns

$$r_1, r_2, \ldots, r_N,$$

the empirical mean return is

$$\hat{\mu}_r = \frac{1}{N} \sum_{t=1}^{N} r_t.$$

This estimates the average return per time step.

For example, if we use daily data, then

$$\hat{\mu}_r$$

is the average daily log-return.

9. Estimate the Variance

The empirical variance is

$$\hat{\sigma}_r^2 = \frac{1}{N-1} \sum_{t=1}^{N} (r_t - \hat{\mu}_r)^2.$$

The empirical volatility is

$$\hat{\sigma}_r = \sqrt{\hat{\sigma}_r^2}.$$

This measures the typical size of random fluctuations.

10. Statistical Interpretation

The Brownian return model says

$$r_t \sim \mathcal{N}(\mu_r, \sigma_r^2).$$

That means the probability density of returns is

$$p(r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left[-\frac{(r-\mu_r)^2}{2\sigma_r^2}\right].$$

So the model assumes returns are Gaussian.

This is the first thing we will test with real market data.

11. What We Should Check Empirically

Once we compute returns, we should inspect:

$$\hat{\mu}_r,$$

$$\hat{\sigma}_r,$$

skewness,

$$\mathrm{Skew}(r),$$

and kurtosis,

$$\mathrm{Kurt}(r).$$

These tell us whether returns really behave like Gaussian Brownian noise.

12. Skewness

Skewness measures asymmetry.

It is defined as

$$\mathrm{Skew}(r) = \frac{\mathbb{E}[(r - \mu_r)^3]}{\sigma_r^3}.$$

If

$$\mathrm{Skew}(r) = 0,$$

the distribution is symmetric.

If

$$\mathrm{Skew}(r) < 0,$$

large negative returns are more dominant.

If

$$\mathrm{Skew}(r) > 0,$$

large positive returns are more dominant.

13. Kurtosis

Kurtosis measures tail heaviness.

It is defined as

$$\mathrm{Kurt}(r) = \frac{\mathbb{E}[(r - \mu_r)^4]}{\sigma_r^4}.$$

For a Gaussian distribution,

$$\mathrm{Kurt}(r) = 3.$$

Excess kurtosis is

$$\text{Excess Kurtosis} = \mathrm{Kurt}(r) - 3.$$

If excess kurtosis is positive, then returns have heavier tails than a Gaussian.

This is very common in real markets.

Part 2: Time-Dependent Volatility and Volatility Clustering

In Part 1 we assumed returns satisfy

$$r_t = \mu_r + \sigma_r Z_t$$

with

$$Z_t \sim \mathcal{N}(0,1),$$

and

$$\sigma_r = \text{constant}.$$

This is the Brownian assumption.

Real markets often violate this.

Volatility itself changes in time.

We therefore promote

$$\sigma_r \to \sigma_t.$$

Volatility becomes a dynamical variable.

1. What is Volatility Clustering?

Empirically:

large moves tend to be followed by large moves,

small moves tend to be followed by small moves.

This is called volatility clustering.

Even when raw returns have weak correlation:

$$\mathrm{Corr}(r_t, r_{t+\tau}) \approx 0,$$

absolute returns often satisfy

$$\mathrm{Corr}(|r_t|, |r_{t+\tau}|) > 0.$$

Often,

$$\mathrm{Corr}(|r_t|, |r_{t+\tau}|) \sim \tau^{-\beta}.$$

This indicates long memory in volatility.

Constant-volatility Brownian motion misses this.

2. Local Volatility Model

Instead of

$$r_t = \mu_r + \sigma_r Z_t$$

we write

$$r_t = \mu_t + \sigma_t Z_t.$$

Now:

- drift may vary,

$$\mu_t$$

- volatility may vary,

$$\sigma_t.$$

This is our first stochastic volatility model.

3. Rolling Volatility Estimator

Take a moving window of length

$$M.$$

Rolling mean:

$$\bar{r}_t = \frac{1}{M} \sum_{i=t-M+1}^{t} r_i.$$

Rolling variance:

$$\hat{\sigma}_t^2 = \frac{1}{M-1} \sum_{i=t-M+1}^{t} (r_i - \bar{r}_t)^2.$$

Rolling volatility:

$$\hat{\sigma}_t = \sqrt{\hat{\sigma}_t^2}.$$

Parameters:

Symbol	**Meaning**
r_t	log return
M	rolling window size
$\bar{r}_t$	rolling mean return
$\hat{\sigma}_t^2$	rolling variance estimate
$\hat{\sigma}_t$	rolling volatility

Table 2. Rolling volatility parameters.

5. Physics Analogy

Previously:

$$dq = vdt + \sigma dW.$$

Now:

$$dq = vdt + \sigma(t) dW.$$

The diffusion coefficient fluctuates.

This resembles diffusion in a changing thermal environment.

A Brownian particle whose temperature varies in time.

6. High-Volatility and Low-Volatility Regimes

Define median volatility:

$$\sigma_{med}.$$

Low-volatility regime:

$$\sigma_t < \sigma_{med}$$

High-volatility regime:

$$\sigma_t > \sigma_{med}.$$

Markets can switch between regimes.

This is the beginning of regime-switching theory.

8. Mean-Reverting Volatility Model

Instead of only estimating

$$\sigma_t,$$

we model it.

Assume

$$d\sigma = \kappa(\theta - \sigma)dt + \xi dW_\sigma.$$

Parameters:

Symbol	Meaning
κ	mean reversion strength
θ	long-run volatility level
ξ	volatility of volatility
W_σ	Brownian motion of volatility

Table 3. Mean-reverting volatility dynamics.

Then price obeys

$$dq = \mu dt + \sigma_t dW.$$

This is already close to stochastic-volatility models like Heston.

9. Discrete GARCH Model

A discrete volatility dynamics model:

$$r_t = \sigma_t z_t$$

with

$$z_t \sim N(0, 1).$$

Variance evolves:

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2.$$

This is GARCH(1,1).

Past shocks feed future volatility.

Exactly volatility clustering.

10. Trading Interpretation

A fixed stop-loss

$$S_{SL} = S_0(1 - \delta)$$

ignores volatility.

Instead use

$$S_{SL} = S_0 - k\sigma_t S_0.$$

where

$$k$$

is a volatility multiple.

Quiet markets:

smaller stops.

Volatile markets:

wider stops.

This is much more realistic.

Brownian Market Model with Real Data from Yahoo Finance

We will fetch market data using `yfinance`, then study:

1. price process S_t,
2. log-price $q_t = \ln S_t$,
3. log-return $r_t = q_t - q_{t-1}$,
4. mean return $\hat{\mu}$,
5. volatility $\hat{\sigma}$,
6. heteroscedasticity through rolling volatility,
7. Brownian/geometric Brownian simulation.

The basic model is

$$dS = \mu S dt + \sigma S dW.$$

In log-price form:

$$dq = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dW.$$

For discrete data:

$$r_t = \ln \frac{S_t}{S_{t-1}}.$$

Step 1: Fetch Data

We choose a ticker.

Examples:

$$\texttt{AAPL}$$

for Apple stock,

$$\texttt{MSFT}$$

for Microsoft,

$$\texttt{BTC-USD}$$

for Bitcoin.

We use hourly data:

$$\Delta t = 1 \text{ hour}.$$

Important: the Brownian model is theoretically more accurate in the limit

$$\Delta t \to 0.$$

For ordinary large steps such as 1 hour or 1 day, the model is only an approximation.

At larger Δt, jumps, news shocks, volatility clustering, and market microstructure effects become more visible.

	time	open	high	low	close	volum
411	2026-04-24 18:30:00+00:00	269.970001	270.380005	269.859985	270.089996	36231
412	2026-04-24 19:30:00+00:00	270.089996	271.480011	270.000000	271.059998	47443
413	2026-04-27 13:30:00+00:00	266.089996	268.359985	265.070007	267.359985	1021906
414	2026-04-27 14:30:00+00:00	267.390015	267.799988	266.619995	267.109985	325168
415	2026-04-27 15:30:00+00:00	267.140015	267.640015	266.660004	266.739990	155146

Step 2: Define Log-Price and Log-Return

The log-price is

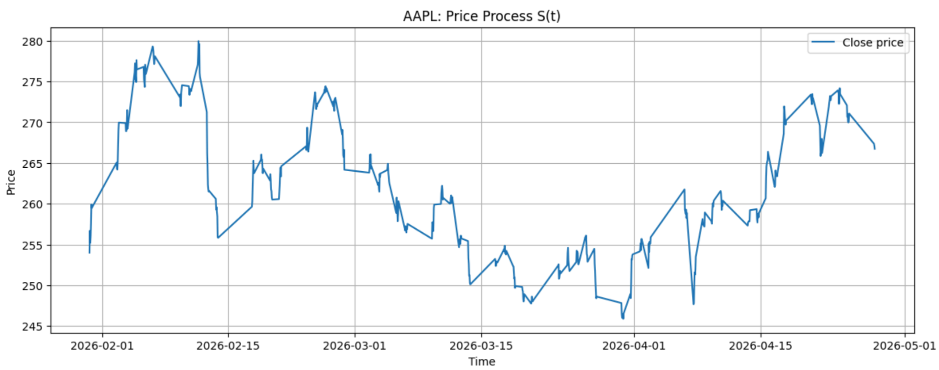
$$q_t = \ln S_t.$$

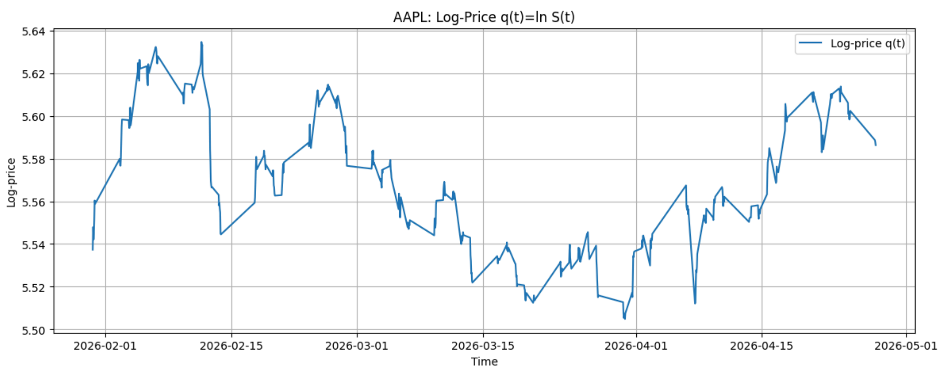
The log-return is

$$r_t = q_t - q_{t-1} = \ln \frac{S_t}{S_{t-1}}.$$

This is the main object for statistical modelling.

```
Ticker: AAPL
Number of candles: 416
Mean log-return per candle: 0.0001179279459390005
Volatility per candle: 0.006187308341622192
Variance per candle: 3.8282784514307565e-05
```



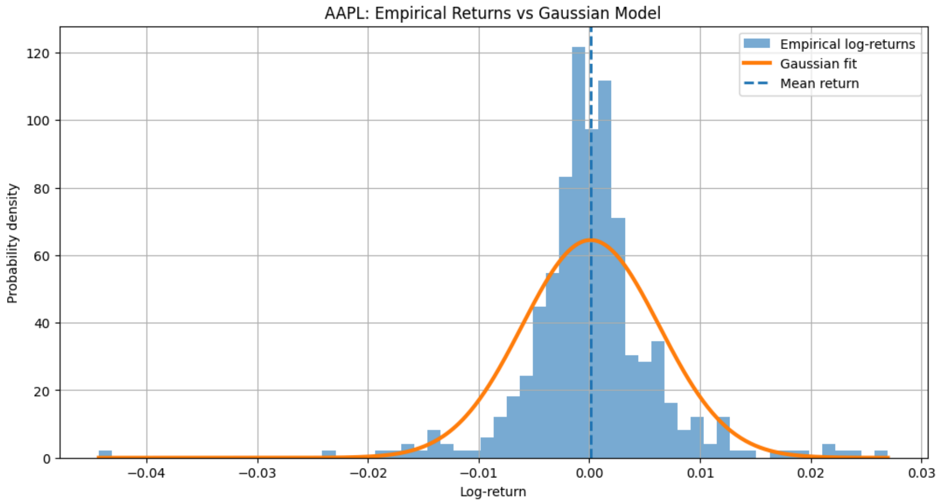


Step 3: Return Distribution

The Brownian model predicts approximately Gaussian log-returns:

$$r_t \sim \mathcal{N}(\mu, \sigma^2).$$

We compare empirical returns with a fitted Gaussian density.



Step 4: Skewness and Kurtosis

For a Gaussian distribution:

$$\text{Skew} = 0$$

and

$$\text{Kurtosis} = 3.$$

If empirical kurtosis is much larger than 3, the distribution has fat tails.

This means extreme moves occur more often than Brownian motion predicts.

```
Skewness: -0.43647025896609287
Kurtosis: 12.04772298782651
Excess kurtosis: 9.04772298782651
Interpretation: returns have fatter tails than Gaussian.
```

Step 5: Rolling Volatility and Heteroscedasticity

The constant-volatility Brownian model assumes

$$\operatorname{Var}(r_t|\mathcal{F}_{t-1}) = \sigma^2.$$

This is homoscedasticity.

A more realistic model allows

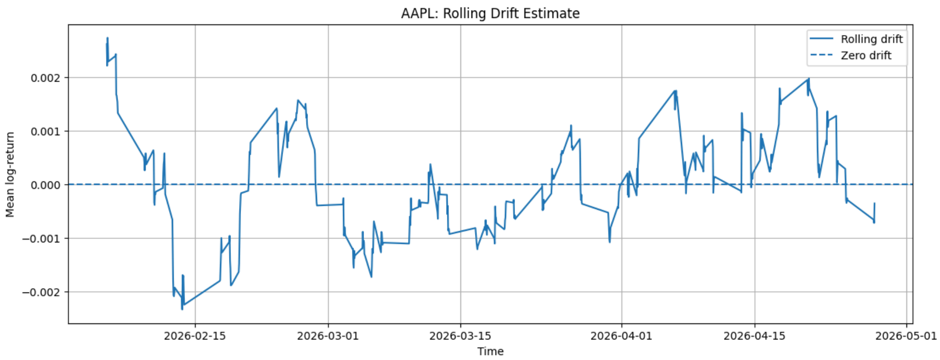
$$\operatorname{Var}(r_t|\mathcal{F}_{t-1}) = \sigma_t^2.$$

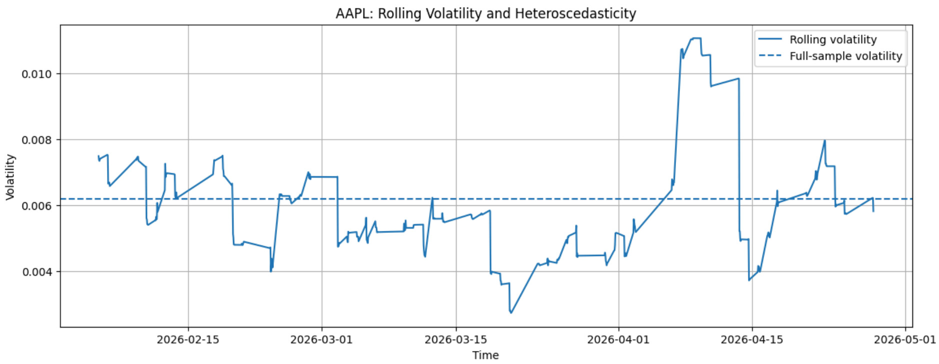
This is conditional heteroscedasticity.

We estimate

$$\sigma_t$$

using rolling volatility.





Step 6: Volatility Clustering

If Brownian motion with constant volatility were perfect, then

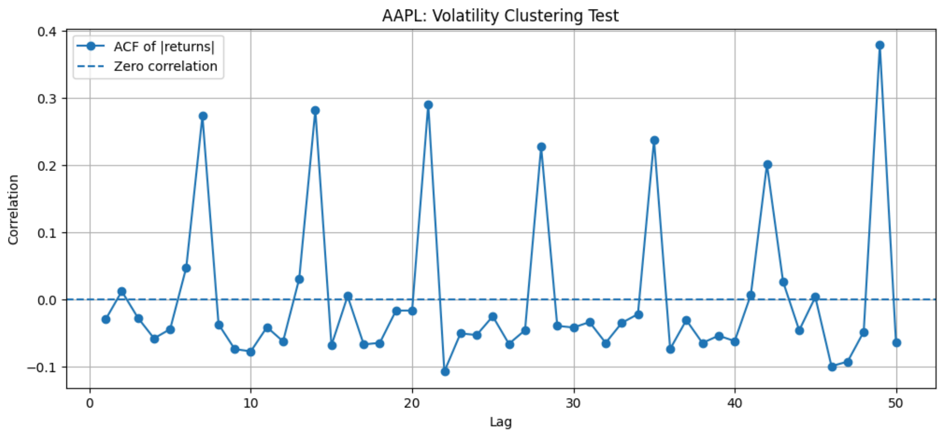
$$|r_t|$$

would have little memory.

But real markets often show

$$\mathrm{Corr}(|r_t|, |r_{t+\tau}|) > 0.$$

This is volatility clustering.



Step 7: Simulate Geometric Brownian Motion

We simulate the model

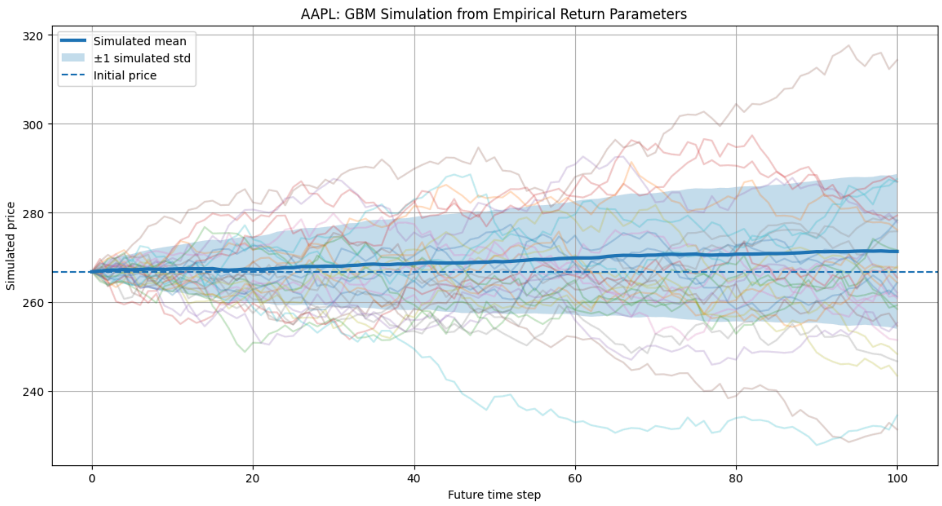
$$S_{n+1} = S_n \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t}Z_n\right].$$

Since our empirical estimates are already per candle, we set

$$\Delta t = 1.$$

Then:

$$S_{n+1} = S_n \exp\left[\left(\hat{\mu} - \frac{1}{2}\hat{\sigma}^2\right) + \hat{\sigma}Z_n\right].$$



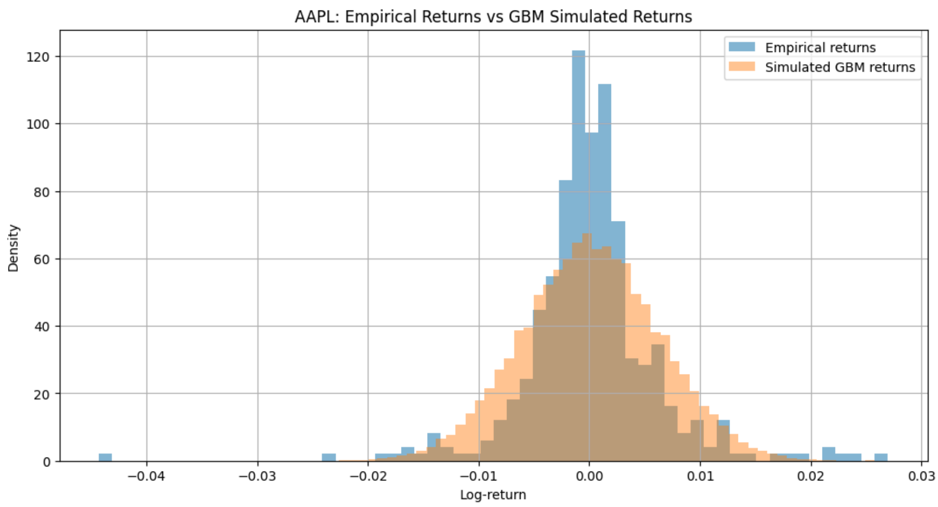
Step 8: Empirical vs Simulated Return Distribution

Now we compare:

1. empirical historical returns,
2. simulated Brownian returns.

If the Brownian model were perfect, both histograms would look similar.

Usually they do not, especially in the tails.




```
Empirical skewness: -0.43647025896609287
Simulated skewness: 0.005971028878978833

Empirical kurtosis: 12.04772298782651
Simulated kurtosis: 2.992726813144371
```

Step 9: Add Simple Spot Trading Boundaries

In spot trading there is no liquidation boundary.

But we can still define:

Stop-loss:

$$S_{SL}$$

Take-profit:

$$S_T.$$

For a long spot trade:

$$S_{SL} < S_0 < S_T.$$

The trade becomes a first-passage problem:

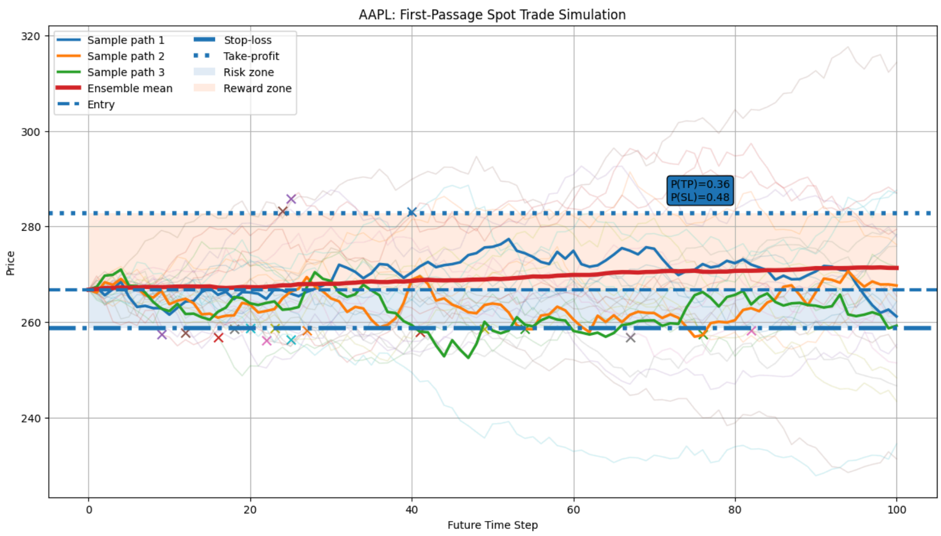
$$\tau_{SL} = \inf\{t : S(t) \leq S_{SL}\},$$

$$\tau_T = \inf\{t : S(t) \geq S_T\}.$$

A successful trade satisfies

$$\tau_T < \tau_{SL}.$$

```
stop_loss      0.483333
take_profit    0.360000
no_exit        0.156667
Name: proportion, dtype: float64
```



Interpretation

We have now built the full basic chain:

$$S_t \to q_t = \ln S_t \to r_t = q_t - q_{t-1}.$$

The simplest Brownian return model is

$$r_t = \mu + \sigma Z_t.$$

The more realistic heteroscedastic model is

$$r_t = \mu_t + \sigma_t Z_t.$$

The important empirical checks are:

$$\mathrm{Skew}(r),$$

$$\mathrm{Kurtosis}(r),$$

rolling volatility,

$$\sigma_t,$$

and autocorrelation of absolute returns:

$$\mathrm{Corr}(|r_t|, |r_{t+\tau}|).$$

The Brownian model is most mathematically justified when

$$\Delta t$$

is very small.

For ordinary data intervals like 1 hour or 1 day, it is an approximation.

At larger time steps, realistic effects appear:

- volatility clustering,
- fat tails,
- jumps,
- regime changes,
- time-dependent drift,
- time-dependent volatility.

So our final realistic spot-market form is

$$r_t = \mu_t + \sigma_t Z_t.$$

Part 4: Model Assumptions and Missing Realism Checks

Before upgrading to GARCH, Student-t noise, jump diffusion, or regime switching, we need to make the current model internally consistent.

So far our basic return model is

$$r_t = \mu_t + \sigma_t Z_t,$$

where

$$Z_t \sim \mathcal{N}(0, 1).$$

But before proceeding, we must add:

1. time-scale consistency,
2. train/test split,
3. stationarity diagnostics,
4. non-Gaussian diagnostics,
5. transaction costs,
6. benchmark comparison,
7. position sizing and account dynamics,
8. a clear assumptions table.

1. Time-Scale Consistency

The model depends on the time step

$$\Delta t.$$

For example:

$$\Delta t = 1 \text{ hour}$$

or

$$\Delta t = 1 \text{ day}.$$

If our data are hourly, then the estimated parameters

$$\hat{\mu}$$

and

$$\hat{\sigma}$$

are hourly parameters.

If our data are daily, they are daily parameters.

We must not mix hourly, daily, and annual parameters without rescaling.

For log returns, approximate scaling is:

$$\mu_{\text{annual}} \approx N \mu_{\Delta t},$$

and

$$\sigma_{\text{annual}} \approx \sqrt{N} \sigma_{\Delta t},$$

where

$$N$$

is the number of time steps per year.

```
Data interval: 1h
Steps per year used: 1638.0

Mean return per step: 0.00011792794593900005
Volatility per step: 0.006187308341622192

Annualized mean return: 0.1931659754480828
Annualized volatility: 0.25041405917886433
```

2. Train/Test Split

We should not estimate model parameters using the whole dataset and then test on the same dataset.

That creates look-ahead bias.

Instead:

$$\text{past data} \to \text{estimate parameters}$$

and

$$\text{future data} \to \text{test model.}$$

We split the data into:

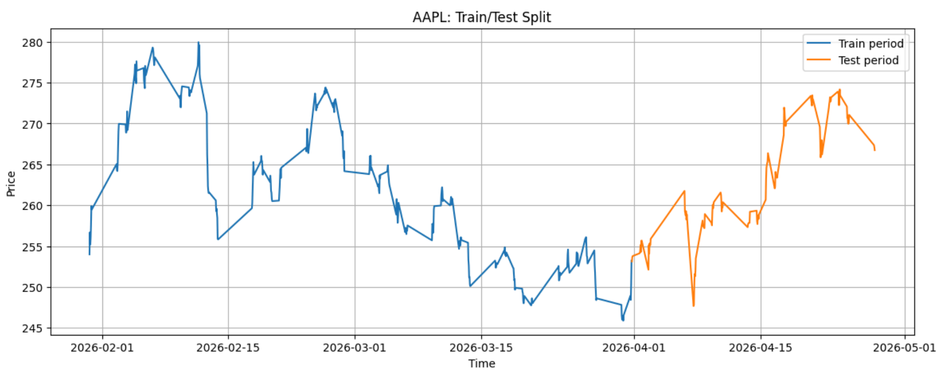
- training set,
- test set.

The training set is used to estimate

$$\mu, \sigma.$$

The test set is used to evaluate whether the model is realistic.

```
Train observations: 291
Test observations: 125

Training mean: -1.053731497428968e-05
Training volatility: 0.005740014235608447

Test mean: 0.0004159673512578337
Test volatility: 0.0071342425957522755
```



3. Stationarity Diagnostics

A stationary return process has statistical properties that do not change strongly with time.

For our simple model, this means roughly:

$$\mu_t \approx \text{stable}$$

and

$$\sigma_t \approx \text{stable}.$$

But real markets often have changing drift and changing volatility.

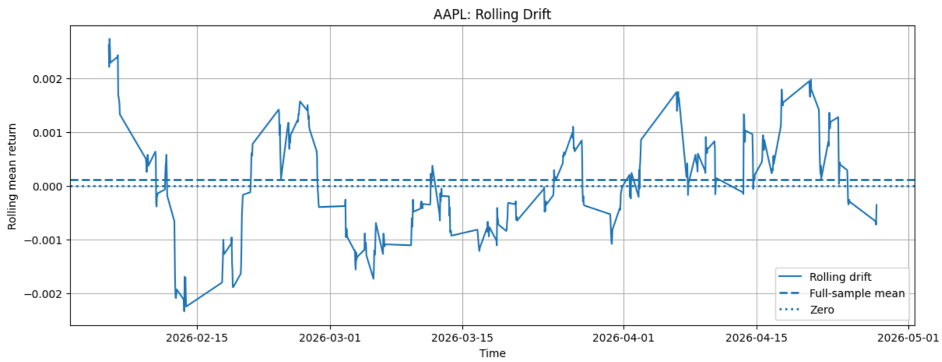
We inspect:

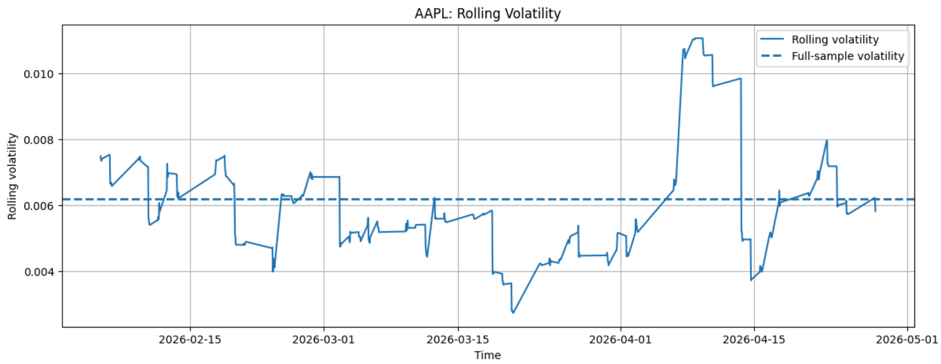
$$\hat{\mu}_t$$

rolling drift, and

$$\hat{\sigma}_t$$

rolling volatility.





4. Non-Gaussian Diagnostics

The Brownian model assumes

$$r_t \sim \mathcal{N}(\mu, \sigma^2).$$

A Gaussian distribution has

$$\text{Skew} = 0$$

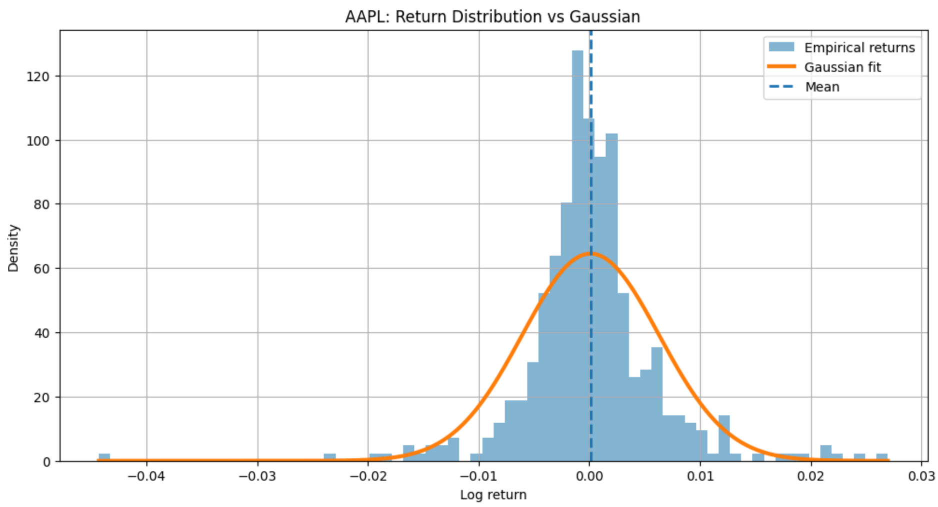
and

$$\text{Kurtosis} = 3.$$

If empirical kurtosis is larger than 3, returns have fat tails.

Fat tails mean extreme events occur more often than the Brownian model predicts.

```
Empirical skewness: -0.43647025896609287
Empirical kurtosis: 12.04772298782651
Empirical excess kurtosis: 9.04772298782651
```



5. Tail Diagnostics

A very useful diagnostic is to compare empirical tail probabilities with Gaussian tail probabilities.

For example:

$$P(|r-\mu|>2\sigma)$$

and

$$P(|r-\mu|>3\sigma).$$

For a Gaussian distribution, approximately:

$$P(|Z|>2)\approx 4.55\%,$$

and

$$P(|Z|>3)\approx 0.27\%.$$

If empirical probabilities are much larger, the market has fat tails.

```
Empirical P(|Z| > 2): 0.05542168674698795
Gaussian  P(|Z| > 2): 0.0455

Empirical P(|Z| > 3): 0.021686746987951807
Gaussian  P(|Z| > 3): 0.0027
```

6. Transaction Costs

A realistic spot trade must include fees and slippage.

Gross PnL is

$$\Pi_{\text{gross}} = Q(S_{\text{exit}} - S_0).$$

Net PnL is

$$\Pi_{\text{net}} = \Pi_{\text{gross}} - C_{\text{entry}} - C_{\text{exit}} - C_{\text{slippage}}.$$

A simple cost model is:

$$C_{\text{entry}} = fQS_0,$$

$$C_{\text{exit}} = fQS_{\text{exit}},$$

where

$$f$$

is the fee rate.

Slippage can be approximated as

$$C_{\text{slippage}} = sQ(S_0 + S_{\text{exit}}),$$

where

$$s$$

is the slippage rate.

7. Benchmark Comparison

A trading model should be compared against a simple benchmark.

For spot trading, the simplest benchmark is buy-and-hold.

Buy-and-hold return is

$$R_{\text{BH}} = \frac{S_{\text{final}} - S_{\text{initial}}}{S_{\text{initial}}}.$$

A strategy is meaningful only if it improves risk-adjusted performance compared to this benchmark.

```
Initial price: 254.0
Final price: 266.739990234375
Buy-and-hold return: 0.05015744186761811
Buy-and-hold return in percent: 5.015744186761811
```

8. Position Sizing

We define account equity:

$$A_t.$$

Let

$$\rho$$

be the fraction of account equity risked per trade.

Then planned risk is

$$R_t = \rho A_t.$$

If entry price is

$$S_0$$

and stop-loss is

$$S_{SL},$$

then the position size is

$$Q_t = \frac{R_t}{|S_0 - S_{SL}|}.$$

This converts a stochastic price model into a trading model.

```
Account equity: 10000.0
Risk fraction: 0.01
Planned risk: 100.0
Entry price: 266.739990234375
Stop-loss price: 258.73779052734375
Position size: 12.49656390256463
Position notional: 3333.3333333333326
```

9. First-Passage Spot Simulation with Costs

We now combine:

- Brownian simulated paths,
- stop-loss,
- take-profit,
- position sizing,
- fees,
- slippage,
- outcome probabilities.

For a long spot trade:

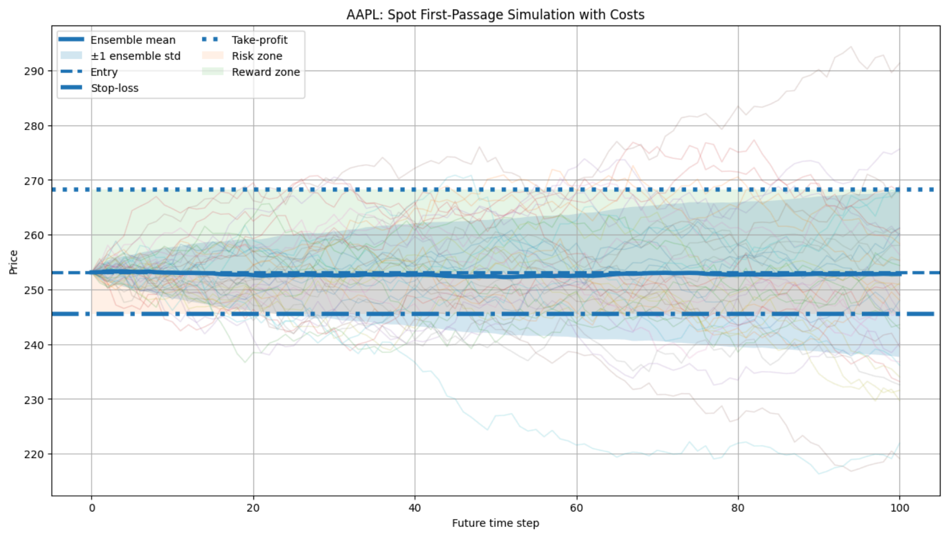
$$S_{SL} < S_0 < S_T.$$

The path exits when it first hits either boundary.

```
Entry: 253.1300048828125
Stop-loss: 245.53610473632813
Take-profit: 268.31780517578125
Quantity: 13.16846390800851
Planned risk: 100.0
```

	outcome	exit_price	exit_step	gross_pnl	entry_fee	exit_fee	slippage_
0	stop_loss	245.536105	100	-100.0	3.333333	3.233333	3.28
1	stop_loss	245.536105	36	-100.0	3.333333	3.233333	3.28
2	stop_loss	245.536105	40	-100.0	3.333333	3.233333	3.28
3	take_profit	268.317805	25	200.0	3.333333	3.533333	3.43
4	stop_loss	245.536105	12	-100.0	3.333333	3.233333	3.28

```
Outcome probabilities:
outcome
stop_loss      0.564
take_profit    0.250
no_exit        0.186
Name: proportion, dtype: float64

Mean net PnL: -9.759814560126626
Median net PnL: -109.85
Std net PnL: 129.71194765231175
Probability of profit: 0.358
Probability of loss: 0.642
```



10. Current Model Assumptions

Our current model is summarized in Table 5.

Assumption	**Meaning**
$r_t = \mu + \sigma Z_t$	basic Brownian return model
$Z_t \sim \mathcal{N}(0,1)$	Gaussian noise
μ, σ estimated on train data	avoid look-ahead bias
σ_t rolling estimate	diagnose heteroscedasticity
$S_{SL} < S_0 < S_T$	long spot first-passage boundaries
$\Pi_{\text{net}} = \Pi_{\text{gross}} - \text{costs}$	fees and slippage included
$Q = \frac{\rho A}{\lvert S_0 - S_{SL}\rvert}$	risk-based position sizing

Table 5. Current assumptions of the realistic spot-market model.

11. Final Status Before Further Upgrades

We have now fixed the most important missing pieces:

- time-scale consistency,
- train/test separation,
- rolling drift and volatility,
- heteroscedasticity diagnostics,
- Gaussianity diagnostics,
- tail diagnostics,
- transaction costs,
- benchmark comparison,
- position sizing,
- account-risk logic.

The current model is still simple:

$$r_t = \mu + \sigma Z_t.$$

But empirically we now know how to test its weaknesses.

The next mathematically natural upgrade is:

$$r_t = \mu + \sigma_t z_t,$$

with

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2.$$

This is the GARCH model.

It directly models conditional heteroscedasticity.

Part 5: GARCH Model — Conditional Heteroscedasticity

We now upgrade the return model.

Previously we used

$$r_t = \mu + \sigma Z_t,$$

where

$$\sigma$$

was constant.

This assumes homoscedasticity:

$$\mathrm{Var}(r_t | \mathcal{F}_{t-1}) = \sigma^2.$$

But real markets often show conditional heteroscedasticity:

$$\mathrm{Var}(r_t | \mathcal{F}_{t-1}) = \sigma_t^2.$$

The variance changes with time.

This leads naturally to GARCH.

1. From Returns to Residuals

Let

$$r_t$$

be the log-return at time step

$$t.$$

We decompose it into mean plus random shock:

$$r_t = \mu + \epsilon_t.$$

Here:

$$\mu$$

is the average return, and

$$\epsilon_t$$

is the unpredictable innovation.

So

$$\epsilon_t = r_t - \mu.$$

In the Brownian model, we assumed

$$\epsilon_t = \sigma Z_t,$$

with constant

$$\sigma.$$

In GARCH, we instead write

$$\epsilon_t = \sigma_t Z_t.$$

Now the shock amplitude changes with time.

2. Information Set

Let

$$\mathcal{F}_{t-1}$$

be all information available before observing

$$r_t.$$

For example:

$$\mathcal{F}_{t-1} = \{r_{t-1}, r_{t-2}, r_{t-3}, \ldots\}.$$

The conditional mean is

$$\mathbb{E}[r_t | \mathcal{F}_{t-1}] = \mu.$$

The conditional variance is

$$\mathrm{Var}(r_t | \mathcal{F}_{t-1}) = \sigma_t^2.$$

So GARCH is a model for the conditional variance.

3. ARCH Idea

Before GARCH, there is ARCH.

ARCH means Autoregressive Conditional Heteroscedasticity.

The idea is:

large shocks in the past increase volatility today.

The simplest ARCH(1) model is

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2.$$

Terms:

$$\omega$$

is the baseline variance.

$$\alpha\epsilon_{t-1}^2$$

is the effect of the previous squared shock.

If yesterday had a large return shock,

$$\epsilon_{t-1}^2$$

is large, so today's variance

$$\sigma_t^2$$

increases.

4. GARCH(1,1)

GARCH means Generalized ARCH.

The most common model is GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

This is the central equation.

It says today's variance depends on:

1. a constant baseline,

$$\omega,$$

2. yesterday's squared shock,

$$\epsilon_{t-1}^2,$$

3. yesterday's variance,

$$\sigma_{t-1}^2.$$

5. Meaning of Each Parameter

The GARCH(1,1) parameters are summarized in Table 6.

Symbol	**Meaning**
r_t	log-return at time t
μ	conditional mean return
ϵ_t	return innovation, $\epsilon_t = r_t - \mu$
Z_t	standardized random shock
σ_t	conditional volatility
σ_t^2	conditional variance
ω	baseline variance level
α	reaction to recent shocks
β	persistence of past volatility
$\mathcal{F}_{t-1}$	information available before time t

Table 6. Parameters of the GARCH(1,1) model.

6. Full GARCH Return Model

The full model is

$$r_t = \mu + \epsilon_t,$$

with

$$\epsilon_t = \sigma_t Z_t,$$

and

$$Z_t \sim \mathcal{N}(0,1).$$

The variance evolves as

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

Therefore:

$$r_t = \mu + \sigma_t Z_t.$$

But unlike Brownian motion,

$$\sigma_t$$

is not constant.

7. Why Squared Returns?

Volatility is about magnitude, not direction.

A large positive shock and a large negative shock both indicate turbulence.

Therefore GARCH uses

$$\epsilon_{t-1}^2.$$

For example:

$$\epsilon_{t-1} = 0.03$$

and

$$\epsilon_{t-1} = -0.03$$

both give

$$\epsilon_{t-1}^2 = 0.0009.$$

Both increase future volatility equally in symmetric GARCH.

8. Parameter Constraints

To make the model mathematically meaningful, we require

$$\omega > 0,$$

$$\alpha \geq 0,$$

and

$$\beta \geq 0.$$

These ensure

$$\sigma_t^2 > 0.$$

For covariance stationarity, we usually require

$$\alpha + \beta < 1.$$

If

$$\alpha + \beta$$

is close to 1, volatility is highly persistent.

This is common in financial markets.

9. Long-Run Variance

Take expectation of

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

In stationarity,

$$\mathbb{E}[\sigma_t^2] = \mathbb{E}[\sigma_{t-1}^2].$$

Also,

$$\mathbb{E}[\epsilon_{t-1}^2] = \mathbb{E}[\sigma_{t-1}^2].$$

Let

$$\bar{\sigma}^2 = \mathbb{E}[\sigma_t^2].$$

Then

$$\bar{\sigma}^2 = \omega + \alpha\bar{\sigma}^2 + \beta\bar{\sigma}^2.$$

So

$$\bar{\sigma}^2 = \omega + (\alpha + \beta)\bar{\sigma}^2.$$

Move terms:

$$(1 - \alpha - \beta)\bar{\sigma}^2 = \omega.$$

Therefore

$$\bar{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta}.$$

This exists only if

$$\alpha + \beta < 1.$$

10. Physical Interpretation

Brownian motion with constant volatility is like diffusion at constant temperature:

$$dq = \mu dt + \sigma dW.$$

GARCH is like diffusion in a medium whose temperature changes depending on recent fluctuations.

Large recent shocks heat the system:

$$\epsilon_{t-1}^2 \uparrow \Rightarrow \sigma_t^2 \uparrow.$$

Then volatility cools slowly through the persistence term:

$$\beta\sigma_{t-1}^2.$$

So GARCH is a discrete-time model of volatility memory.

11. GARCH vs Brownian Model

Brownian model:

$$r_t = \mu + \sigma Z_t.$$

GARCH model:

$$r_t = \mu + \sigma_t Z_t.$$

Brownian:

$$\mathrm{Var}(r_t|\mathcal{F}_{t-1}) = \sigma^2.$$

GARCH:

$$\mathrm{Var}(r_t|\mathcal{F}_{t-1}) = \sigma_t^2.$$

So Brownian motion is homoscedastic.

GARCH is conditionally heteroscedastic.

12. What GARCH Captures

GARCH captures:

- volatility clustering,
- persistent high-volatility regimes,
- persistent low-volatility regimes,
- time-varying risk,
- conditional heteroscedasticity.

But basic GARCH does not capture:

- jumps,
- asymmetric leverage effects,
- fat-tailed shocks unless we replace Gaussian noise,
- regime switching,
- structural breaks.

13. Next Code Step

In the next code cells we will:

1. take empirical log returns,

$$r_t,$$

2. fit a GARCH(1,1) model,
3. extract estimated conditional volatility,

$$\hat{\sigma}_t,$$

4. compare it with rolling volatility,
5. simulate returns using the fitted GARCH dynamics.

14. Prepare Returns for GARCH

The GARCH model is usually fitted to returns in percent scale.

We define

$$r_t = 100 \left(\ln S_t - \ln S_{t-1} \right).$$

The factor

$$100$$

is only a numerical convention.

It makes the return values easier for the optimizer to handle.

```
Number of returns: 415
Mean return in percent: 0.011792794593900005
Volatility in percent: 0.6187308341622196
```

15. Fit GARCH(1,1)

We fit the model

$$r_t = \mu + \epsilon_t,$$

with

$$\epsilon_t = \sigma_t Z_t,$$

and

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2.$$

The model estimates:

$$\mu, \omega, \alpha, \beta.$$

```
                     Constant Mean - GARCH Model Results
==============================================================================
=================
Dep. Variable:             log_return   R-squared:
0.000
Mean Model:             Constant Mean   Adj. R-squared:
0.000
Vol Model:                      GARCH   Log-Likelihood:
-387.157
Distribution:                  Normal   AIC:
782.314
Method:            Maximum Likelihood   BIC:
798.427
                                        No. Observations:
415
Date:                Mon, Apr 27 2026   Df Residuals:
414
Time:                        16:03:46   Df Model:
1
                                  Mean Model
==============================================================================
================
                 coef    std err          t      P>|t|
95.0% Conf. Int.
------------------------------------------------------------------------------
----------------
mu         6.5390e-03  3.352e-02      0.195      0.845 [-5.91
6e-02,7.224e-02]
                               Volatility Model
==============================================================================
================
                 coef    std err          t      P>|t|
95.0% Conf. Int.
------------------------------------------------------------------------------
----------------
omega          0.0110  2.030e-02      0.540      0.589 [-2.88
2e-02,5.076e-02]
alpha[1]       0.0000  2.301e-02      0.000      1.000 [-4.50
9e-02,4.509e-02]
beta[1]        0.9694  7.665e-02     12.648  1.146e-36      [
0.819,  1.120]
==============================================================================
================

Covariance estimator: robust
```

16. Extract Estimated Parameters

After fitting, we extract

$$\hat{\mu},$$

$$\hat{\omega},$$

$$\hat{\alpha},$$

and

$$\hat{\beta}.$$

The persistence of volatility is measured by

$$\hat{\alpha} + \hat{\beta}.$$

If

$$\hat{\alpha} + \hat{\beta} \approx 1,$$

volatility shocks decay slowly.

```
Estimated mu: 0.006538997416395127
Estimated omega: 0.010971939264581582
Estimated alpha: 0.0
Estimated beta: 0.9694323577771207
Volatility persistence alpha + beta: 0.9694323577771207

Long-run variance: 0.35893966517212383
Long-run volatility: 0.5991157360411458
```

17. Conditional Volatility

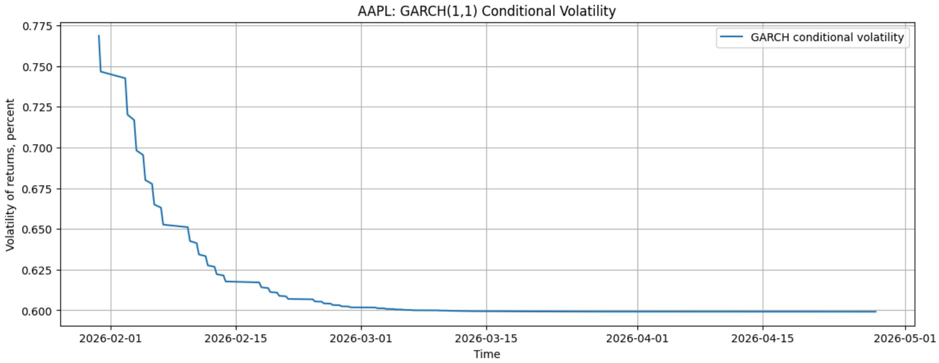
The fitted GARCH model gives an estimated conditional volatility:

$$\hat{\sigma}_t.$$

This is the model-implied volatility at time

$$t.$$

It should rise after large shocks and decay gradually afterward.



18. Compare Rolling Volatility and GARCH Volatility

Rolling volatility is a direct empirical estimate:

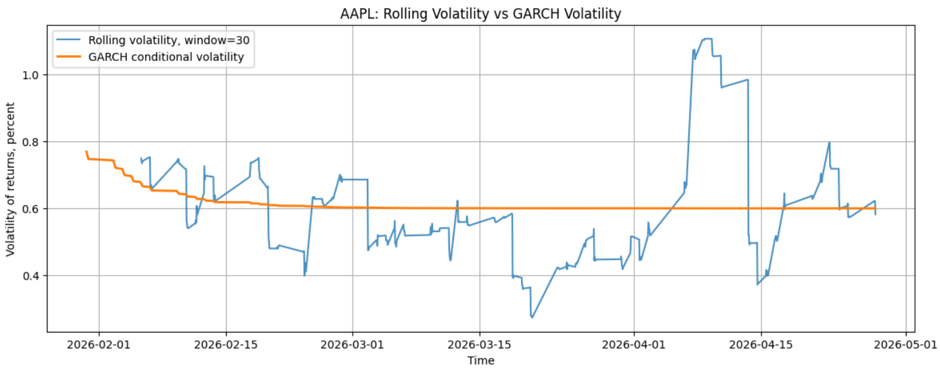
$$\hat{\sigma}_t^{\text{rolling}} = \sqrt{\frac{1}{M-1} \sum_{i=t-M+1}^{t} (r_i - \bar{r}_t)^2}.$$

GARCH volatility is model-based:

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}\hat{\epsilon}_{t-1}^2 + \hat{\beta}\hat{\sigma}_{t-1}^2.$$

Rolling volatility looks backward over a fixed window.

GARCH volatility evolves recursively.



19. Standardized Residuals

GARCH decomposes returns as

$$r_t = \mu + \sigma_t Z_t.$$

Therefore the standardized residual is

$$\hat{Z}_t = \frac{r_t - \hat{\mu}}{\hat{\sigma}_t}.$$

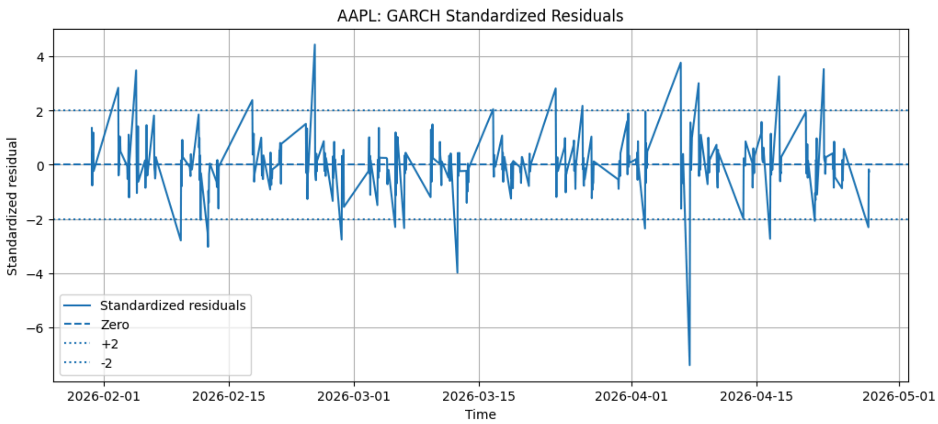
If the GARCH model is good and the Gaussian noise assumption is reasonable, then

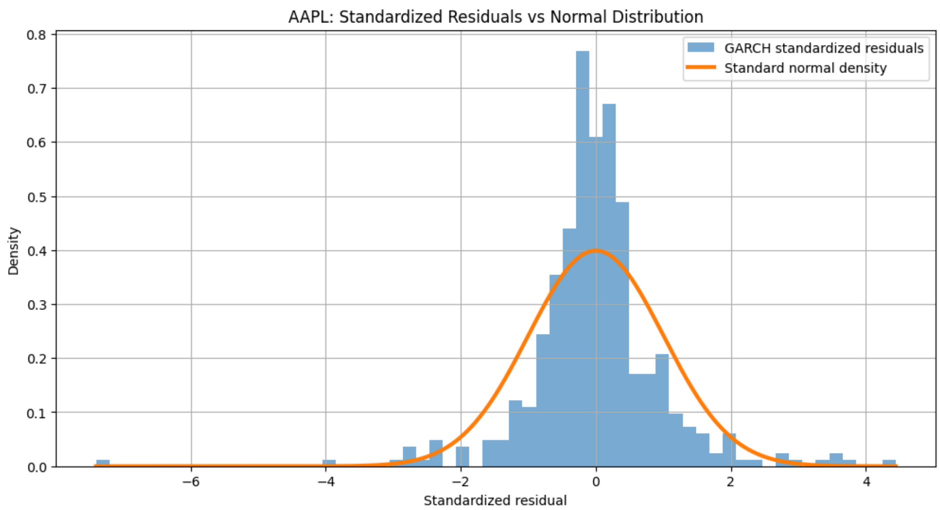
$$\hat{Z}_t$$

should look approximately like

$$\mathcal{N}(0, 1).$$

```
Standardized residual mean: 0.0038720441813337356
Standardized residual std: 1.005288386110056
Standardized residual skewness: -0.5642647519738831
Standardized residual kurtosis: 12.644548390433918
```





20. Simulate GARCH Returns

We now simulate returns using the fitted model.

The recursion is

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2.$$

Then

$$\epsilon_t = \sigma_t Z_t,$$

and

$$r_t = \mu + \epsilon_t.$$

Since our fitted returns are in percent, the simulated returns are also in percent.

```
Simulated GARCH returns shape: (300, 300)
Simulated GARCH volatility shape: (300, 300)
```

21. Convert Simulated Returns to Prices

We convert percent log-returns back to decimal log-returns:

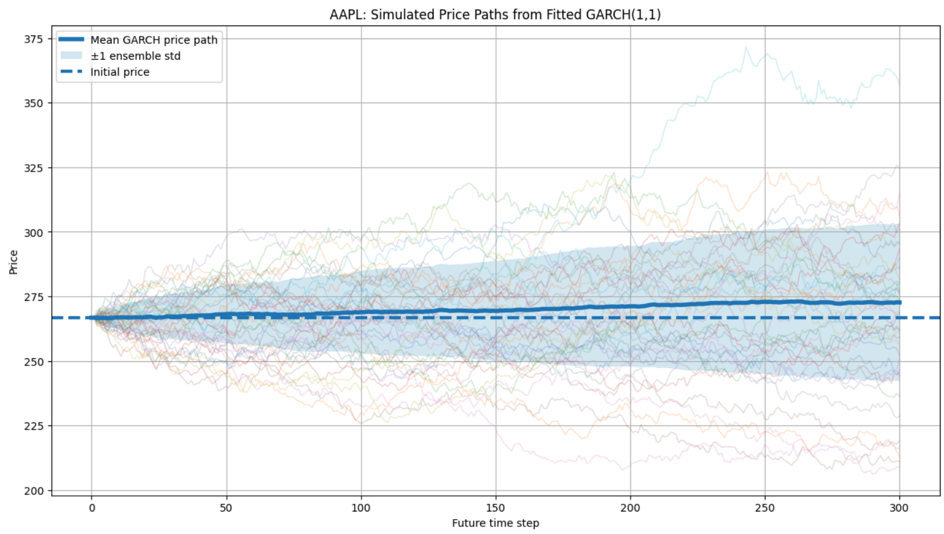
$$r_t^{\text{decimal}} = \frac{r_t^{\text{percent}}}{100}.$$

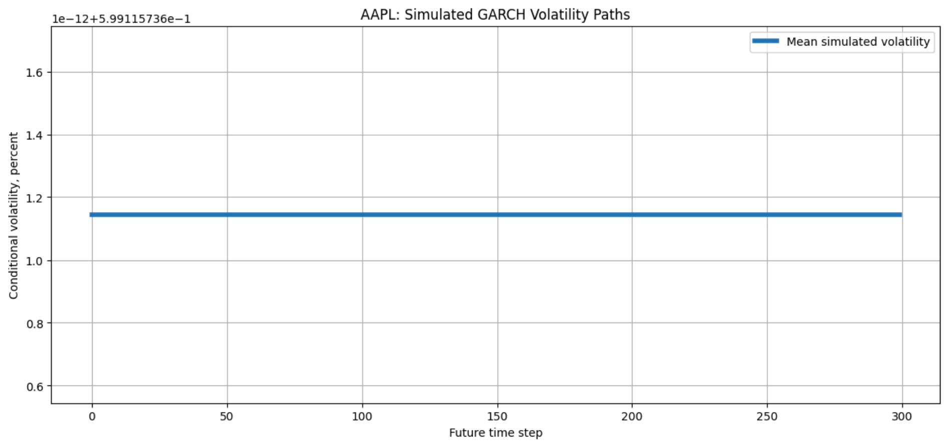
Then log-price evolves as

$$q_t = q_0 + \sum_{i=1}^{t} r_i.$$

Price is recovered by

$$S_t = e^{q_t}.$$





22. Compare Brownian and GARCH Conceptually

The Brownian model assumes

$$r_t = \mu + \sigma Z_t.$$

The GARCH model assumes

$$r_t = \mu + \sigma_t Z_t.$$

The difference is not the noise variable

$$Z_t.$$

The difference is that the amplitude of noise changes over time:

$$\sigma \to \sigma_t.$$

Therefore GARCH captures conditional heteroscedasticity:

$$\mathrm{Var}(r_t | \mathcal{F}_{t-1}) = \sigma_t^2.$$

23. Interpretation of the Fitted Model

After fitting, inspect:

$$\alpha$$

reaction to new shocks,

$$\beta$$

volatility persistence,

and

$$\alpha + \beta$$

total persistence.

If

$$\alpha$$

is large, volatility reacts strongly to recent shocks.

If

$$\beta$$

is large, volatility decays slowly.

If

$$\alpha + \beta \approx 1,$$

volatility is highly persistent.

This is common in market data.

24. What We Improved

We moved from

$$r_t = \mu + \sigma Z_t$$

to

$$r_t = \mu + \sigma_t Z_t.$$

The old model had constant variance:

$$\mathrm{Var}(r_t | \mathcal{F}_{t-1}) = \sigma^2.$$

The GARCH model has dynamic conditional variance:

$$\mathrm{Var}(r_t | \mathcal{F}_{t-1}) = \sigma_t^2.$$

This captures volatility clustering and makes the stochastic market model more realistic.

The next natural improvement is to replace Gaussian shocks by Student-t shocks:

$$Z_t \sim t_\nu.$$

That allows the model to capture fat tails better.

Part 6: Beyond Gaussian GARCH — Fat Tails and Heavy-Tailed Markets

Our GARCH model assumed

$$Z_t \sim \mathcal{N}(0, 1).$$

So returns satisfy

$$r_t = \mu + \sigma_t Z_t.$$

Volatility was stochastic,

but shocks were still Gaussian.

This is often unrealistic.

Empirical markets often exhibit:

- extreme events more frequent than Gaussian theory predicts,
- heavy tails,
- crash outliers,
- excess kurtosis.

To model this we replace Gaussian noise.

1. Student-t Innovations

Instead of

$$Z_t \sim N(0,1),$$

assume

$$Z_t \sim t_\nu.$$

This is a Student-t distribution with

$$\nu$$

degrees of freedom.

Then

$$r_t = \mu + \sigma_t Z_t.$$

The volatility dynamics can remain GARCH:

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2.$$

Only the innovation law changes.

2. Why Student-t?

Gaussian tails decay as

$$e^{-x^2/2}.$$

Very fast.

Student-t tails decay much slower:

$$P(|Z| > x) \sim x^{-\nu}.$$

Power-law tails.

This is much closer to empirical markets.

3. Role of Degrees of Freedom

Parameter

$$\nu$$

controls tail thickness.

Large

$$\nu$$

approaches Gaussian:

$$\nu \to \infty \Rightarrow t_\nu \to N(0,1).$$

Small

$$\nu$$

means fatter tails.

Typical finance estimates:

$$4 < \nu < 10.$$

```
                        Constant Mean - GARCH Model Results
==============================================================
=======================
Dep. Variable:                   log_return   R-squared:
0.000
Mean Model:                   Constant Mean   Adj. R-squared:
0.000
Vol Model:                            GARCH   Log-Likelihood:
-311.454
Distribution:      Standardized Student's t   AIC:
632.907
Method:                  Maximum Likelihood   BIC:
653.049
                                              No. Observation
s:                    415
Date:                      Mon, Apr 27 2026   Df Residuals:
414
Time:                              16:03:52   Df Model:
1
                                   Mean Model
==============================================================
=================
                  coef    std err          t      P>|t|
95.0% Conf. Int.
--------------------------------------------------------------
-----------------
mu         -5.9707e-03  1.917e-02     -0.311      0.755 [-4.3
55e-02,3.161e-02]
                                Volatility Model
==============================================================
================
                 coef    std err          t      P>|t|
95.0% Conf. Int.
--------------------------------------------------------------
----------------
omega      3.9182e-03  2.648e-03      1.480      0.139 [-1.27
2e-03,9.108e-03]
alpha[1]       0.0000  8.061e-03      0.000      1.000 [-1.58
0e-02,1.580e-02]
beta[1]        0.9892  8.132e-03    121.652      0.000      [
0.973,  1.005]
                               Distribution
==============================================================
===========
                 coef    std err          t      P>|t|  95.0%
Conf. Int.
--------------------------------------------------------------
-----------
nu             2.7003      0.209     12.939  2.703e-38 [  2.2
91,  3.109]
==============================================================
===========

Covariance estimator: robust
```

4. Compare Gaussian GARCH vs t-GARCH

We compare:

Gaussian GARCH:

$$Z_t \sim N(0,1)$$

versus

Student-t GARCH:

$$Z_t \sim t_\nu.$$

The heavy-tailed model often fits likelihood much better.

```
Gaussian GARCH AIC: 782.314217447229
Student-t GARCH AIC: 632.9072531907291

Gaussian GARCH BIC: 798.4273315281517
Student-t GARCH BIC: 653.0486457918826
```

Lower AIC/BIC generally means better fit.

Often t-GARCH wins.

That indicates fat-tailed shocks are needed.

```
Estimated degrees of freedom v = 2.7002810484702073
```

If

$$\nu$$

comes out small,

tails are very heavy.

That is a direct quantitative estimate of tail risk.

Part 7: Leverage Effect — Asymmetric Volatility

Standard GARCH is symmetric.

It treats

$$+5\%$$

and

$$-5\%$$

shocks identically.

But real markets often show:

negative shocks raise future volatility more than positive shocks.

This is called leverage effect.

GJR-GARCH Model

Upgrade:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \gamma I_{t-1}\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

where indicator:

$$I_{t-1} = \begin{cases} 1 & \epsilon_{t-1} < 0 \\ 0 & \epsilon_{t-1} \geq 0 \end{cases}$$

So negative shocks contribute extra:

$$\gamma.$$

```
                      Constant Mean - GJR-GARCH Model Results
=============================================================
=======================
Dep. Variable:                   log_return   R-squared:
0.000
Mean Model:                   Constant Mean   Adj. R-squared:
0.000
Vol Model:                        GJR-GARCH   Log-Likelihood:
-311.454
Distribution:      Standardized Student's t   AIC:
634.907
Method:                  Maximum Likelihood   BIC:
659.077
                                              No. Observation
s:                   415
Date:                    Mon, Apr 27 2026   Df Residuals:
414
Time:                            16:03:53   Df Model:
1
                                  Mean Model
=============================================================
=================
                  coef    std err          t      P>|t|
95.0% Conf. Int.
-------------------------------------------------------------
-----------------
mu         -5.9677e-03  1.914e-02     -0.312      0.755 [-4.3
47e-02,3.154e-02]
                               Volatility Model
=============================================================
=================
                  coef    std err          t      P>|t|
95.0% Conf. Int.
-------------------------------------------------------------
-----------------
omega       3.9188e-03  2.585e-03      1.516      0.130 [-1.1
48e-03,8.985e-03]
alpha[1]    1.9407e-11  2.017e-02  9.621e-10      1.000 [-3.9
54e-02,3.954e-02]
gamma[1]   -9.7920e-11  2.554e-02 -3.833e-09      1.000 [-5.0
07e-02,5.007e-02]
beta[1]         0.9892  9.533e-03    103.768      0.000
[  0.971,  1.008]
                              Distribution
=============================================================
===========
                 coef    std err          t      P>|t|  95.0%
Conf. Int.
-------------------------------------------------------------
-----------
nu             2.7002      0.232     11.635  2.751e-31 [  2.2
45,  3.155]
=============================================================
===========

Covariance estimator: robust
```

If

$$\gamma > 0$$

negative shocks generate more volatility than positive shocks.

Crash asymmetry captured.

Part 8: Jump Diffusion

Even heavy-tailed GARCH may miss sudden discontinuous jumps.

Add jumps:

$$dq = \mu dt + \sigma dW + JdN_t.$$

Terms:

Brownian diffusion:

$$\sigma dW$$

Jump component:

$$JdN_t$$

where

$$N_t$$

is Poisson process.

Poisson process satisfies

$$P(dN_t = 1) = \lambda dt$$

and

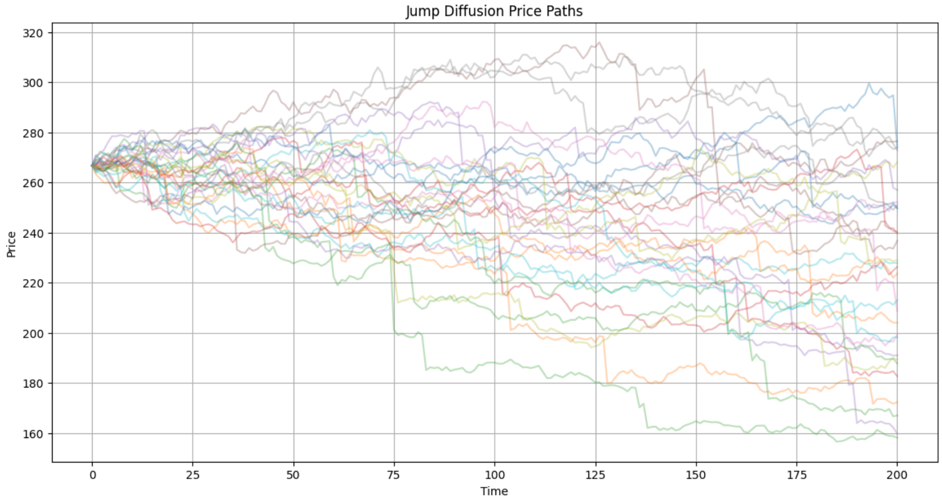
$$P(dN_t = 0) = 1 - \lambda dt.$$

Parameter

$$\lambda$$

is jump intensity.

Expected jumps per unit time.



Now rare crash jumps can occur.

Brownian model:

continuous paths.

Jump diffusion:

discontinuous paths.

Part 9: Hidden Regime Models

Markets may switch between hidden states:

$$s_t \in \{1, 2, \dots, K\}$$

Example:

1. bull low volatility
2. bear high volatility

Then:

$$r_t = \mu_{s_t} + \sigma_{s_t} Z_t.$$

Parameters depend on regime.

Transition probabilities:

$$P(s_t = j | s_{t-1} = i) = P_{ij}.$$

This is a Markov chain.

Now both drift and volatility switch.

Model Hierarchy Built So Far

Level 0:

Geometric Brownian motion

$$r_t = \mu + \sigma Z_t$$

Level 1:

State-dependent drift and volatility

$$r_t = \mu_t + \sigma_t Z_t$$

Level 2:

GARCH

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

Level 3:

t-GARCH

$$Z_t \sim t_\nu$$

Level 4:

Asymmetric GARCH

$$\gamma$$

captures leverage effect.

Level 5:

Jump diffusion

$$dq = \mu dt + \sigma dW + J dN$$

Level 6:

Regime switching

$$(\mu, \sigma) \to (\mu_{s_t}, \sigma_{s_t})$$

At this point we are approaching genuine quantitative finance.

Part 10: Meeting Theoretical Physics — Langevin, Fokker-Planck, and Path Integrals

We now reinterpret the market model using theoretical physics.

The central variable is the log-price

$$q(t) = \ln S(t).$$

The simplest stochastic market model is

$$dq = \mu dt + \sigma dW_t.$$

This is mathematically the same as overdamped Langevin dynamics.

In physics language:

$$\text{market log-price} \leftrightarrow \text{Brownian particle position}.$$

1. Langevin Form

Write

$$\frac{dq}{dt} = \mu + \sigma\eta(t),$$

where

$$\eta(t)$$

is Gaussian white noise.

It satisfies

$$\mathbb{E}[\eta(t)] = 0,$$

and

$$\mathbb{E}[\eta(t)\eta(t')] = \delta(t - t').$$

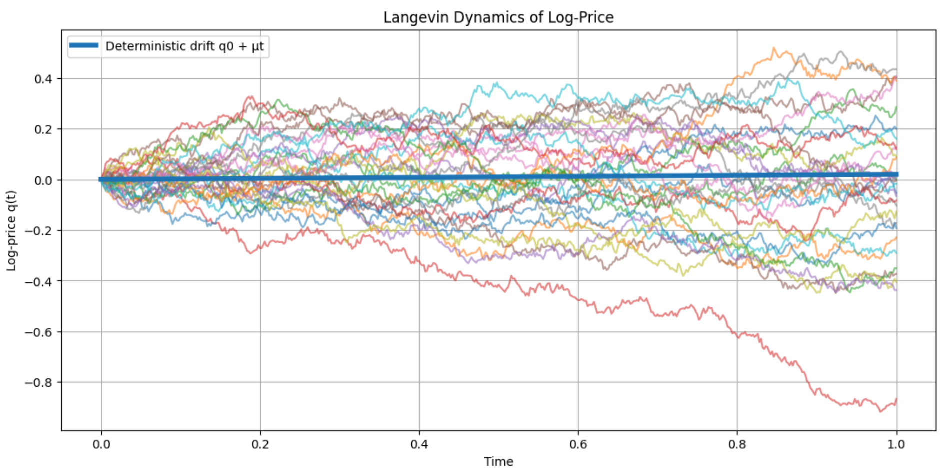
So the market is treated as a noisy dynamical system:

$$\text{deterministic drift} + \text{random fluctuations}.$$

Parameters:

Symbol	Meaning
$q(t)$	log-price coordinate
μ	drift velocity
σ	noise amplitude / volatility
$\eta(t)$	Gaussian white noise
W_t	Brownian motion

Table 7. Langevin interpretation of the market model.



2. From Log-Price to Price

Since

$$q(t) = \ln S(t),$$

we recover price by

$$S(t) = e^{q(t)}.$$

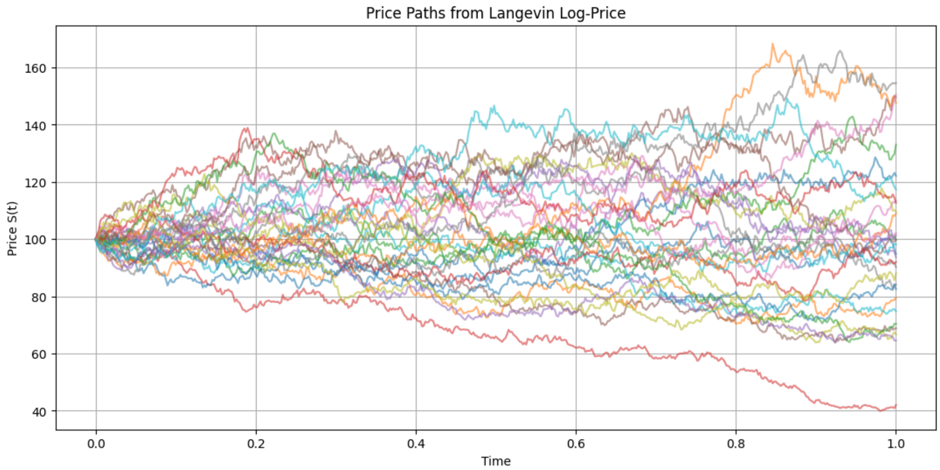
So additive Brownian motion in

$$q$$

becomes multiplicative stochastic motion in

$$S.$$

This is why geometric Brownian motion appears naturally.



3. Fokker-Planck Picture

The Langevin equation

$$dq = \mu dt + \sigma dW_t$$

has a corresponding Fokker-Planck equation for the probability density

$$P(q,t).$$

It is

$$\frac{\partial P}{\partial t} = -\mu \frac{\partial P}{\partial q} + \frac{1}{2}\sigma^2 \frac{\partial^2 P}{\partial q^2}.$$

This is the advection-diffusion equation.

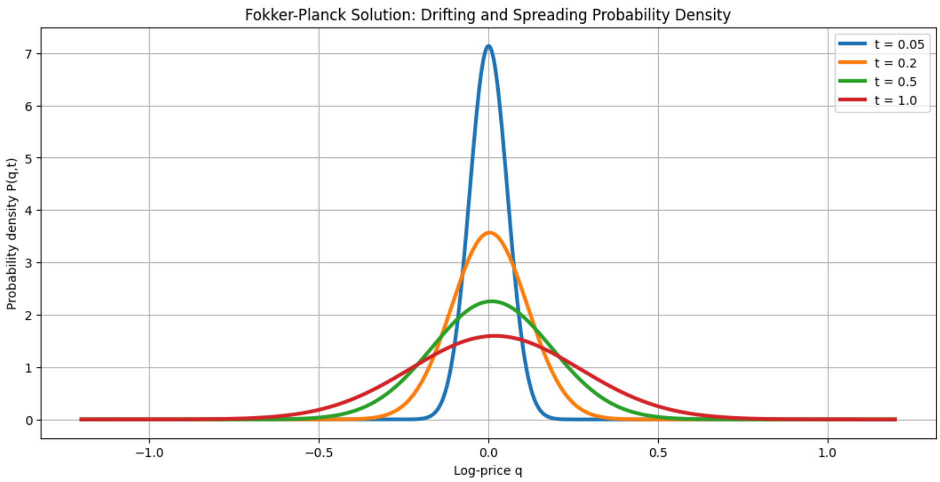
The solution starting from

$$P(q,0) = \delta(q - q_0)$$

is

$$P(q,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{(q - q_0 - \mu t)^2}{2\sigma^2 t} \right].$$

So the probability packet drifts and spreads.



4. Effective Potential and Drift

In physics, drift can be generated by a potential.

For overdamped Langevin dynamics:

$$dq = -\frac{\partial V}{\partial q} dt + \sigma dW_t.$$

So the drift is

$$\mu(q) = -\frac{\partial V}{\partial q}.$$

If

$$V(q)$$

decreases to the right, then the force pushes

$$q$$

to the right.

In market language:

- positive drift means upward tendency,
- negative drift means downward tendency,
- potential wells represent mean-reverting regions.

5. Example: Mean-Reverting Market

A simple potential well is

$$V(q) = \frac{1}{2}k(q - q_*)^2.$$

Then

$$-\frac{\partial V}{\partial q} = -k(q - q_*).$$

So the stochastic equation is

$$dq = -k(q - q_*)dt + \sigma dW_t.$$

This is the Ornstein-Uhlenbeck process.

It models mean reversion.

If

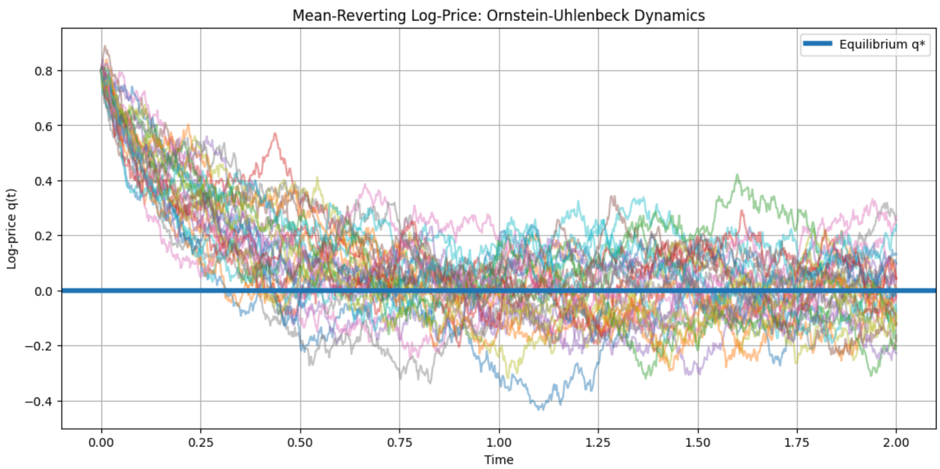
$$q > q_*,$$

the drift is negative.

If

$$q < q_*,$$

the drift is positive.



6. Trading Boundaries as Absorbing Walls

A stop-loss and take-profit can be treated like absorbing boundaries.

Let

$$q_{SL}$$

be the stop-loss boundary and

$$q_T$$

be the take-profit boundary.

Define hitting times:

$$\tau_{SL} = \inf\{t : q(t) \le q_{SL}\},$$

and

$$\tau_T = \inf\{t : q(t) \ge q_T\}.$$

A successful long trade satisfies

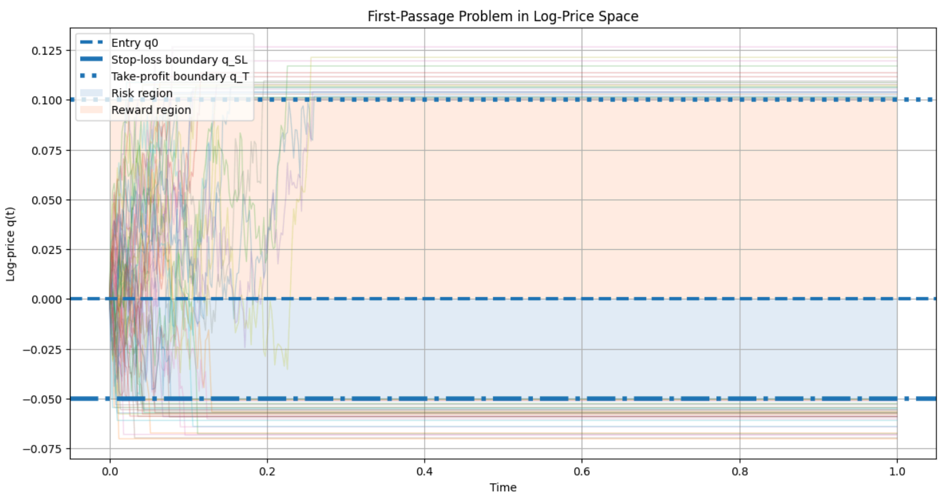
$$\tau_T < \tau_{SL}.$$

This is a first-passage problem.

In physics:

$$\text{trade outcome} = \text{which absorbing wall is reached first.}$$

```
stop_loss      0.61
take_profit    0.39
Name: proportion, dtype: float64
```



7. Action Functional: Path-Integral View

For Brownian motion with drift,

$$dq = \mu dt + \sigma dW_t,$$

a path

$$q(t)$$

has probability weight approximately

$$\mathcal{P}[q(t)] \propto \exp(-\mathcal{A}[q]),$$

where the action is

$$\mathcal{A}[q] = \frac{1}{2\sigma^2} \int_0^T (\dot{q} - \mu)^2 dt.$$

This is analogous to Euclidean path integrals in physics.

Most likely paths minimize the action.

Large deviations correspond to high-action trajectories.

8. Most Likely Path

For constant drift

$$\mu,$$

the action is minimized when

$$\dot{q} = \mu.$$

Therefore

$$q_{\text{cl}}(t) = q_0 + \mu t.$$

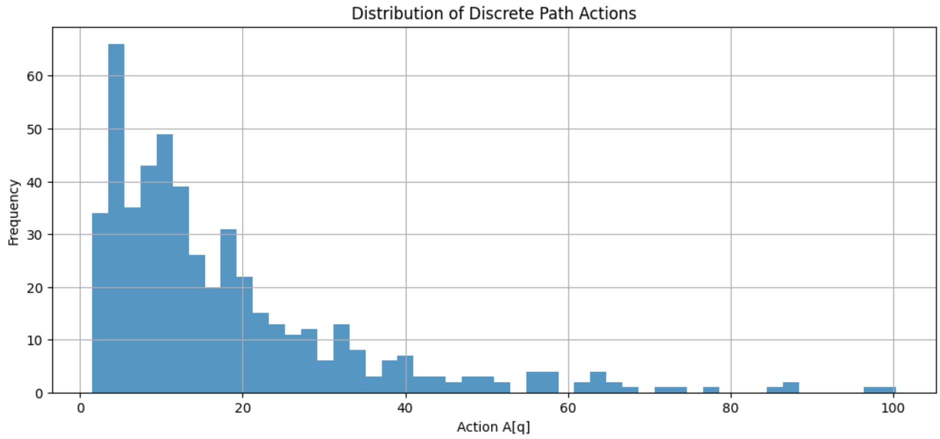
This is the classical path.

Stochastic fluctuations occur around it.

The volatility

$$\sigma$$

controls how strongly paths fluctuate around the classical path.




```
Mean action: 17.827643377627126
Minimum action: 1.5234777317445678
Maximum action: 100.30150855689601
```

9. Hamiltonian Interpretation

For the action

$$\mathcal{A}[q] = \frac{1}{2\sigma^2} \int (\dot{q} - \mu)^2 dt,$$

the corresponding Lagrangian is

$$L(q, \dot{q}) = \frac{1}{2\sigma^2}(\dot{q} - \mu)^2.$$

The conjugate momentum is

$$p = \frac{\partial L}{\partial \dot{q}} = \frac{\dot{q} - \mu}{\sigma^2}.$$

Solving for

$$\dot{q}$$

gives

$$\dot{q} = \mu + \sigma^2 p.$$

The Hamiltonian is

$$H = p\dot{q} - L.$$

Substitute:

$$H = p(\mu + \sigma^2 p) - \frac{1}{2\sigma^2}(\sigma^2 p)^2.$$

Therefore

$$H = \mu p + \frac{1}{2}\sigma^2 p^2.$$

This is the Hamiltonian of Brownian log-price dynamics.

10. Meaning of the Momentum

The momentum

$$p = \frac{\dot{q} - \mu}{\sigma^2}$$

measures deviation from the typical drift.

If

$$\dot{q} = \mu,$$

then

$$p = 0.$$

That is the most likely trajectory.

If

$$|\dot{q} - \mu|$$

is large, then

$$|p|$$

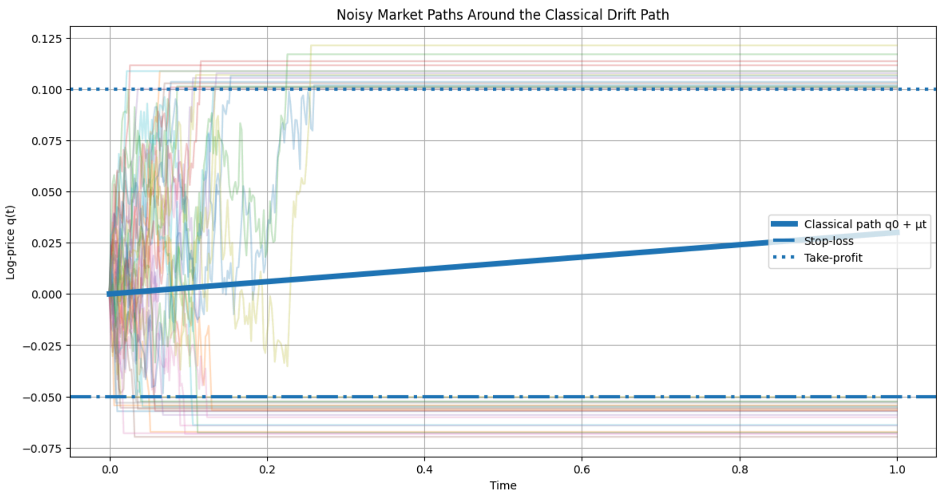
is large.

Such paths are rare.

So in this stochastic Hamiltonian picture,

$$p$$

measures large-deviation pressure away from the typical market path.



Summary: Finance as Statistical Field Theory of Paths

We started with market prices and arrived at physics language.

The correspondences are:

Market Model	**Theoretical Physics Interpretation**
$q(t) = \ln S(t)$	particle coordinate
$dq = \mu dt + \sigma dW$	Langevin equation
$P(q, t)$	probability density
Fokker-Planck equation	evolution equation for density
S_{SL}, S_T	absorbing boundaries
τ_{SL}, τ_T	first-passage times
$\mathcal{A}[q]$	Euclidean action of a path
$H(q, p)$	stochastic Hamiltonian

Table 8. Mapping between stochastic market analysis and theoretical physics.

The key physical picture is:

$$\text{market trajectory} = \text{random path weighted by an action.}$$

A trading rule is then a boundary-condition problem on this stochastic path ensemble.

Part 11: Optimal Trading and Prognosis Algorithms

Up to now we modeled the market.

Now we ask:

Can we build an algorithm that decides:

- buy,
- hold,
- sell

at time

$$t$$

using information

$$\mathcal{F}_t$$

to maximize returns?

This is a stochastic control problem.

State:

$$X_t$$

contains market information.

Control:

$$u_t$$

is trading action.

Goal:

maximize expected utility of wealth.

1. Predicting Next Return Sign

Suppose we want forecast:

$$\operatorname{sign}(r_{t+1}).$$

Define target variable

$$y_t = \begin{cases} 1 & r_{t+1} > 0 \\ 0 & r_{t+1} \le 0 \end{cases}$$

Binary prediction problem.

At time

$$t$$

we estimate

$$P(r_{t+1} > 0|\mathcal{F}_t).$$

Call it

$$p_t.$$

Then simple decision rule:

Buy if

$$p_t > 0.5.$$

Sell or stay out if

$$p_t < 0.5.$$

2. Features for Prognosis

Use information vector

$$X_t$$

for prediction:

$$X_t = (r_t, r_{t-1}, \sigma_t, m_t, V_t).$$

Possible components:

recent returns:

$$r_t, r_{t-1}, \ldots$$

volatility:

$$\sigma_t$$

momentum:

$$m_t = q_t - q_{t-k}$$

volume:

$$V_t$$

moving-average spread:

$$EMA_{20} - EMA_{50}.$$

	precision	recall	f1-score	support
0	0.49	0.62	0.55	53
1	0.50	0.37	0.43	54
accuracy			0.50	107
macro avg	0.50	0.50	0.49	107
weighted avg	0.50	0.50	0.49	107

Model estimates

$$p_t = P(r_{t+1} > 0 | \mathcal{F}_t).$$

This is a prognosis algorithm.

But maximizing accuracy alone does not maximize returns.

Trading objective is different.

3. Trading as Control Problem

Let wealth be

$$W_t.$$

Position:

$$u_t \in [-1, 1].$$

Examples:

$$u_t = 1$$

fully long.

$$u_t = 0$$

flat.

$$u_t = -1$$

short.

Wealth evolves:

$$W_{t+1} = W_t(1 + u_t r_{t+1}).$$

Control variable:

$$u_t.$$

Objective Functional

Maximize expected utility:

$$\max_{u_t} \mathbb{E}[U(W_T)].$$

Simple risk-neutral case:

$$U(W) = W.$$

Risk-averse:

$$U(W) = \log W.$$

or

$$U(W) = -\exp(-\gamma W).$$

Now this becomes stochastic optimal control.

Kelly Criterion

A first optimal solution is Kelly sizing.

If probability of gain is

$$p$$

and probability of loss is

$$q = 1 - p,$$

optimal fraction is

$$f^* = \frac{bp - q}{b}$$

where

$$b$$

is reward-to-risk ratio.

```
Kelly fraction = 0.3400000000000001
```

4. Decision Rule from Predicted Probability

Instead of merely

$$p_t > .5$$

we can use threshold:

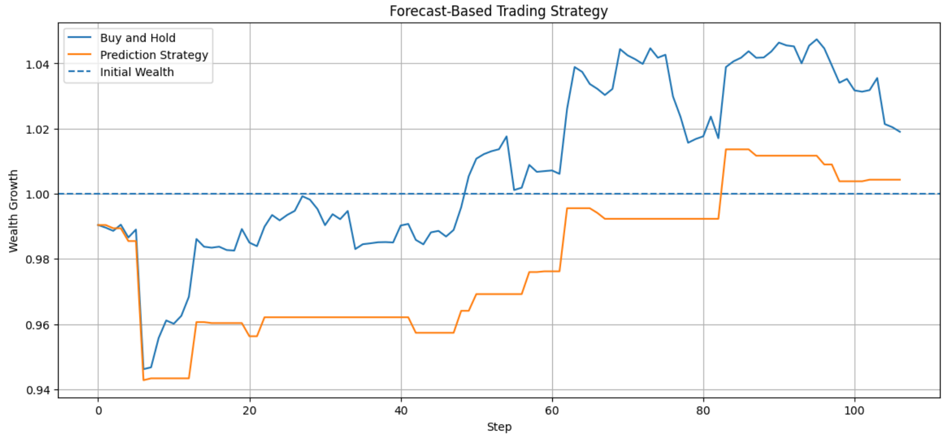
Buy only if

$$p_t > \theta.$$

For example

$$\theta = .55.$$

This filters weak signals.




```
Signal length: 107
Return length: 107
Final market wealth: 1.0190249857998128
Final strategy wealth: 1.0043205628935252
```

Now we are no longer merely forecasting.

We use

$$p_t$$

to generate control

$$u_t.$$

This turns prediction into optimal trading.

5. Hamilton-Jacobi-Bellman Equation

Define value function

$$V(W,t) = \max_u \mathbb{E}[U(W_T)|W_t = W].$$

Dynamic programming gives HJB equation:

$$0 = V_t + \max_u \left[\mu u W V_W + \frac{1}{2}\sigma^2 u^2 W^2 V_{WW}\right].$$

This is continuous-time optimal trading.

Optimal control from first-order condition:

$$u^* = -\frac{\mu V_W}{\sigma^2 W V_{WW}}.$$

This is the analog of an Euler-Lagrange solution for optimal portfolio control.

6. Statistical Physics Interpretation

Before:

we studied passive stochastic trajectories.

Now:

we apply control

$$u_t.$$

This is like steering a noisy dynamical system.

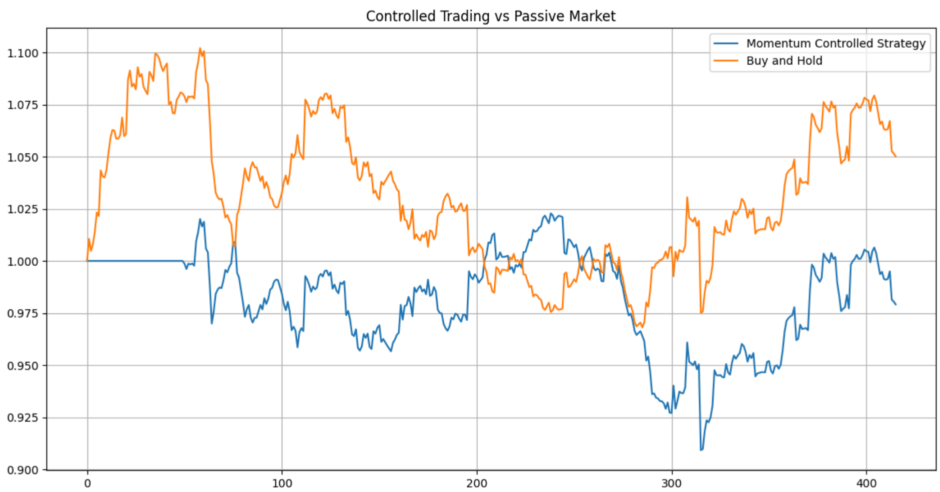
Market:

stochastic dynamics.

Trader:

external control field.

Very much like controlled Langevin dynamics.



7. Hierarchy So Far

We now have:

Price dynamics:

$$dq = \mu dt + \sigma dW$$

Conditional heteroscedasticity:

GARCH.

Heavy tails:

t-GARCH.

Optimal prediction:

$$P(r_{t+1} > 0 | \mathcal{F}_t).$$

Optimal control:

$$u_t.$$

Hamilton-Jacobi-Bellman optimization:

$$\max E[U(W_T)].$$

This is now stochastic control theory for trading.

Part 12: Statistical Mechanics of Learning Meets Trading Strategy

We now combine two ideas:

1. A market model:

$$r_t = \mu_t + \sigma_t Z_t$$

2. A learning model:

$$\hat{y}_t = f_\theta(X_t)$$

where

$$X_t$$

is the market state and

$$\theta$$

are trainable parameters.

In the statistical mechanics of learning, learning can be interpreted as minimizing an energy-like cost function over a high-dimensional parameter space. Engel and Van den Broeck formulate learning rules through a learning error

$$E(J)$$

and associate a partition function

$$Z = \int d\mu(J) e^{-\beta E(J)}.$$

This gives a direct analogy:

$$\text{learning} \leftrightarrow \text{statistical mechanics of configurations.}$$

For trading, the “student” is our model, the “examples” are historical market states, and the “teacher” is the future market return.

1. Trading as Supervised Learning

At each time

$$t,$$

we observe a market state

$$X_t.$$

Examples:

$$X_t = (r_t, r_{t-1}, \sigma_t, m_t, V_t).$$

The target is future return direction:

$$y_t = \begin{cases} 1, & r_{t+1} > 0, \\ 0, & r_{t+1} \le 0. \end{cases}$$

The model learns a map

$$f_\theta : X_t \mapsto p_t,$$

where

$$p_t = P(r_{t+1} > 0 | X_t).$$

Then a simple strategy is:

buy if

$$p_t > \theta_{\text{buy}}.$$

2. Statistical Mechanics Dictionary

Trading / Machine Learning	Statistical Mechanics
θ	model parameters / microscopic configuration
X_t	input pattern / market state
y_t	teacher signal / target label
$L(\theta)$	energy function
$e^{-\beta L(\theta)}$	Boltzmann weight
β	inverse temperature
$\lambda \vert\vert\theta\vert\vert^2$	regularizing potential
test performance	generalization error

Table 9. Statistical mechanics interpretation of strategy learning.

3. Energy Function for a Trading Classifier

For binary classification, use cross-entropy loss:

$$E(\theta) = -\sum_{t=1}^{N} \left[y_t \ln p_t + (1 - y_t) \ln(1 - p_t) \right].$$

where

$$p_t = f_\theta(X_t).$$

Add regularization:

$$E_\lambda(\theta) = E(\theta) + \lambda ||\theta||^2.$$

The regularization term penalizes overly large weights.

This is analogous to adding an external confining potential in parameter space. Nielsen explains this as weight decay or L^2 regularization, where the network compromises between minimizing the data loss and keeping weights small.

4. Temperature and Overfitting

The Boltzmann distribution over models is

$$P(\theta) = \frac{1}{Z} e^{-\beta E(\theta)}.$$

where

$$Z = \int d\theta \, e^{-\beta E(\theta)}.$$

Interpretation:

- large β means low temperature,
- the learner strongly concentrates near low-loss configurations,
- small β means high temperature,
- the learner explores many parameter configurations.

Low temperature can overfit.

High temperature can underfit.

A good learner balances energy and entropy.

This is the same conceptual role as the bias-variance and model-complexity discussion in Biehl’s model-evaluation chapter.

5. Practical Learning Rule

We train parameters by stochastic gradient descent:

$$\theta_{k+1} = \theta_k - \eta \nabla_\theta E_{\text{batch}}(\theta_k).$$

Here:

$$\eta$$

is the learning rate.

Nielsen describes mini-batch stochastic gradient descent as estimating the full gradient from randomly selected training examples; the estimate is noisy but often sufficient to move in a direction that reduces the cost.

In statistical mechanics language:

SGD is noisy motion on an energy landscape.

```
Feature matrix shape: (366, 7)
Target shape: (366,)
```

6. Time-Ordered Train/Test Split

Financial data must not be randomly shuffled.

We use:

$$\text{past} \to \text{train}$$

and

$$\text{future} \to \text{test}.$$

This avoids look-ahead bias.

Random train/test splitting would leak temporal information.

```
Train size: 256
Test size: 110
```

7. Neural Strategy Model

We now train a shallow neural network:

$$p_t = f_\theta(X_t).$$

Architecture:

input layer:

$$X_t \in \mathbb{R}^d,$$

hidden layer:

$$h = \phi(WX + b),$$

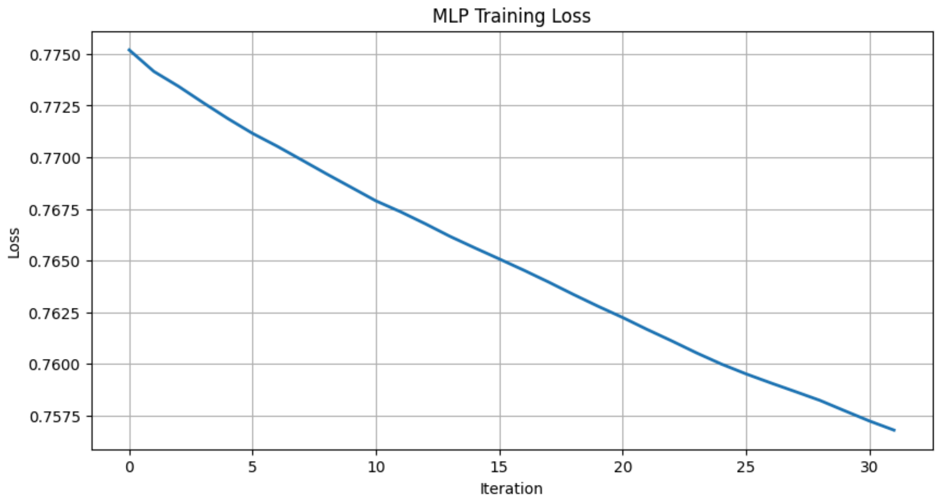
output:

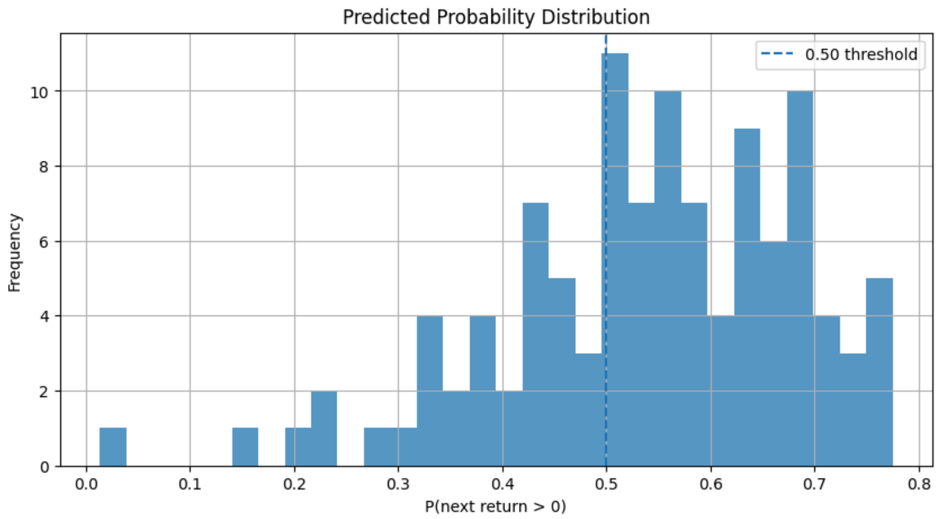
$$p_t = \sigma(w \cdot h + c).$$

The output is interpreted as:

$$p_t = P(r_{t+1} > 0 | X_t).$$

```
              precision    recall  f1-score   support

           0       0.49      0.34      0.40        50
           1       0.56      0.70      0.62        60

    accuracy                           0.54       110
   macro avg       0.52      0.52      0.51       110
weighted avg       0.53      0.54      0.52       110

ROC AUC: 0.5373333333333333

Training iterations: 32
Final loss: 0.7567989029745181
```





8. From Prediction to Trading Control

Prediction alone is not enough.

We define a trading signal:

$$u_t = \begin{cases} 1, & p_t > \theta_{\text{buy}}, \\ 0, & p_t \le \theta_{\text{buy}}. \end{cases}$$

For spot trading, we use:

$$u_t = 1$$

long,

$$u_t = 0$$

flat.

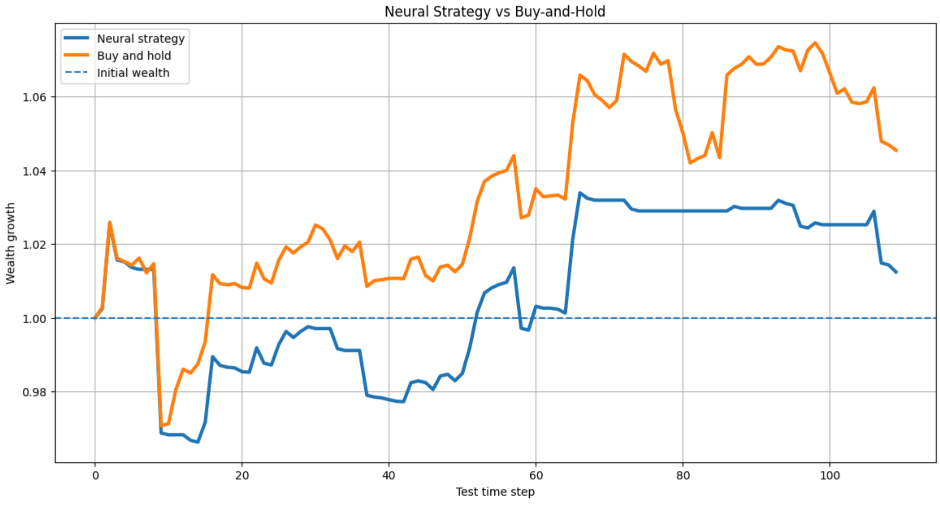
The strategy return is

$$R_t^{\text{strat}} = u_t r_{t+1}.$$

With transaction costs:

$$R_t^{\text{net}} = u_t r_{t+1} - c|u_t - u_{t-1}|.$$

```
Average signal: 0.5
Total turnover: 41
Final strategy wealth: 1.0124367578573992
Final buy-hold wealth: 1.0454446913235742
```



9. Energy Landscape Diagnostics

The neural network learns by minimizing loss.

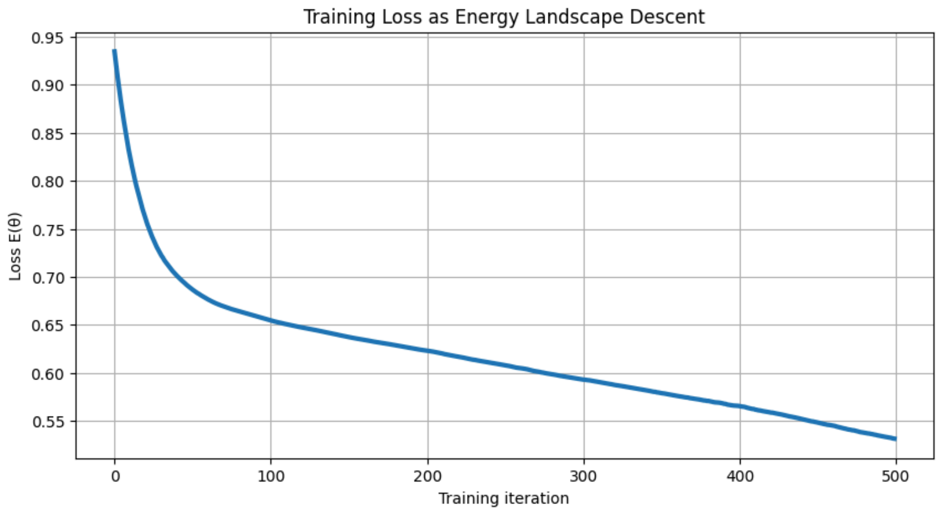
We can plot the training loss over iterations.

This is a numerical trace of motion down the energy landscape:

$$E(\theta_0) \to E(\theta_1) \to \cdots.$$

In theoretical physics language:

learning is relaxation in a high-dimensional rough energy landscape.



10. Generalization Error

Training performance is not enough.

We need generalization performance on unseen future data.

In trading:

$$\text{generalization error} = \text{failure to predict future market states.}$$

This is more severe than ordinary machine learning because market distributions are non-stationary.

Biehl emphasizes that supervised learning requires reliable estimates of performance in the working phase, not merely good training performance.

```
Strategy Sharpe: 0.274900717758535
Market Sharpe: 0.8768703684526555

Strategy max drawdown: -0.0577529767702466
Market max drawdown: -0.053827976279859024
```

11. Statistical Mechanics Interpretation of Trading Neural Networks

The trading model is now:

$$p_t = f_\theta(X_t).$$

The strategy is:

$$u_t = \Theta(p_t - \theta_{\text{buy}}).$$

The learning energy is:

$$E(\theta) = -\sum_t \left[y_t \ln p_t + (1 - y_t) \ln(1 - p_t)\right] + \lambda ||\theta||^2.$$

The financial objective is different:

$$J(\theta) = \mathbb{E}[R_{\text{net}}(\theta)] - \gamma \text{Risk}(\theta).$$

Therefore the classifier loss is only a proxy.

A model can have good classification accuracy but bad trading performance.

The physically correct goal is to minimize an effective free energy:

$$F_{\text{eff}}(\theta) = E_{\text{prediction}}(\theta) + \lambda ||\theta||^2 - \eta\, \mathbb{E}[R_{\text{net}}(\theta)] + \gamma\, \text{Risk}(\theta).$$

12. What We Have Built

We combined:

1. stochastic market states,

$$X_t = (r_t, \sigma_t, m_t, \dots),$$

2. supervised learning,

$$X_t \mapsto y_t,$$

3. neural-network energy minimization,

$$E(\theta),$$

4. statistical mechanics interpretation,

$$P(\theta) \propto e^{-\beta E(\theta)},$$

5. trading control,

$$u_t = \Theta(p_t - \theta_{\text{buy}}),$$

6. strategy evaluation,

$$\text{wealth, Sharpe, drawdown.}$$

This is the bridge:

$$\text{statistical mechanics of learning} + \text{stochastic finance} = \text{strategy learning as contro}$$

13. Next Natural Upgrade

The next step is to train directly on trading utility instead of classification accuracy.

Instead of minimizing

$$E_{\text{CE}}(\theta),$$

we define a differentiable strategy return:

$$u_t(\theta) = \sigma(a(p_t - \theta_{\text{buy}})),$$

and optimize

$$\mathcal{L}(\theta) = -\sum_t u_t(\theta) r_{t+1} + \lambda ||\theta||^2 + \gamma \text{DrawdownPenalty}.$$

Then the neural network is no longer just predicting.

It is directly learning a trading policy.

Part 13: Almost Real-Time Market Strategy Example

We now combine:

1. stochastic market features,
2. volatility estimation,
3. supervised learning,
4. statistical-mechanics interpretation,
5. buy/flat trading control.

We use recent market candles from Yahoo Finance.

The model observes a market state

$$X_t$$

and estimates

$$p_t = P(r_{t+1} > 0 | X_t).$$

Then the trading rule is:

$$u_t = \begin{cases} 1, & p_t > \theta, \\ 0, & p_t \le \theta. \end{cases}$$

where

$$u_t = 1$$

means long, and

$$u_t = 0$$

means flat.

This is a paper-trading simulation, not live order execution.

1. Fetch Recent Market Data

For a near-real-time experience, use short candles.

Examples:

- `AAPL` with `interval="5m"` during market hours,
- `MSFT`,
- `TSLA`,
- `BTC-USD` for a 24/7 market.

Yahoo Finance data may be delayed and should be treated as research data.

We use:

$$\Delta t = 5\ \mathrm{minutes}.$$

```
                          time          open          high           
low  \
1340 2026-04-27 15:40:00+00:00  77040.023438  77090.000000  7
7006.953125
1341 2026-04-27 15:45:00+00:00  77024.781250  77048.968750  7
6791.640625
1342 2026-04-27 15:50:00+00:00  76809.976562  76879.976562  7
6746.242188
1343 2026-04-27 15:55:00+00:00  76835.921875  76841.843750  7
6734.500000
1344 2026-04-27 16:00:00+00:00  76750.890625  76759.992188  7
6675.882812

             close     volume
1340  77040.460938  945451008
1341  76815.359375  237203456
1342  76831.898438  134148096
1343  76772.921875  114708480
1344  76746.242188  127668224

Latest candle time: 2026-04-27 16:00:00+00:00
Latest close: 76746.2421875
```

2. Build Stochastic Market Features

We define:

log-price:

$$q_t = \ln S_t$$

log-return:

$$r_t = q_t - q_{t-1}$$

short momentum:

$$m_t^{(5)} = \frac{S_t}{S_{t-5}} - 1$$

rolling volatility:

$$\sigma_t = \sqrt{\frac{1}{M-1} \sum_{i=t-M+1}^{t} (r_i - \bar{r}_t)^2.}$$

These form the market state vector:

$$X_t = (r_t, r_{t-1}, m_t, \sigma_t, V_t).$$

```
Any infinite values? False
Any NaN values? False
Feature matrix shape: (325, 8)
```

3. Time-Ordered Train/Test Split

We do not shuffle financial time series.

We train on the past and test on the future:

$$\text{past candles} \to \text{train}$$

$$\text{recent candles} \to \text{test}.$$

This imitates real-time deployment.

```
Train samples: 227
Test samples: 98
```

4. Train Prognosis Model

We use logistic regression.

It estimates:

$$p_t = P(r_{t+1} > 0 | X_t).$$

This is a simple probabilistic classifier.

Energy function:

$$E(w) = -\sum_t [y_t \ln p_t + (1 - y_t) \ln(1 - p_t)] + \lambda ||w||^2.$$

Statistical mechanics interpretation:

$$P(w) \propto e^{-\beta E(w)}.$$

```
              precision    recall  f1-score   support

           0       0.49      0.38      0.42        48
           1       0.51      0.62      0.56        50

    accuracy                           0.50        98
   macro avg       0.50      0.50      0.49        98
weighted avg       0.50      0.50      0.49        98

ROC AUC: 0.48291666666666666
```

5. Convert Prediction into Trading Control

Prediction:

$$p_t = P(r_{t+1} > 0 | X_t).$$

Trading rule:

$$u_t = \begin{cases} 1, & p_t > \theta, \\ 0, & p_t \leq \theta. \end{cases}$$

We use a conservative threshold:

$$\theta = 0.55.$$

This avoids trading on weak signals.

Net strategy return:

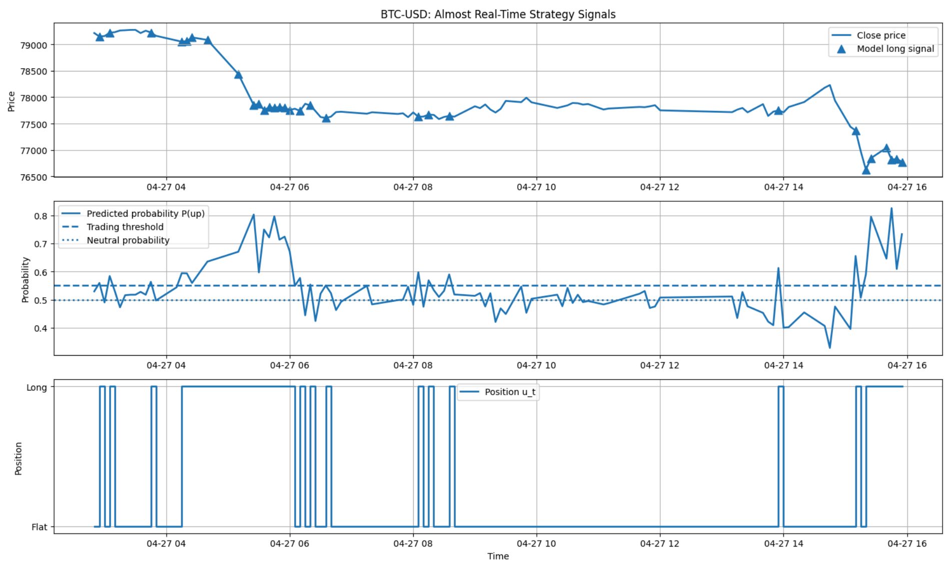
$$R_t^{net} = u_t r_{t+1} - c|u_t - u_{t-1}|.$$

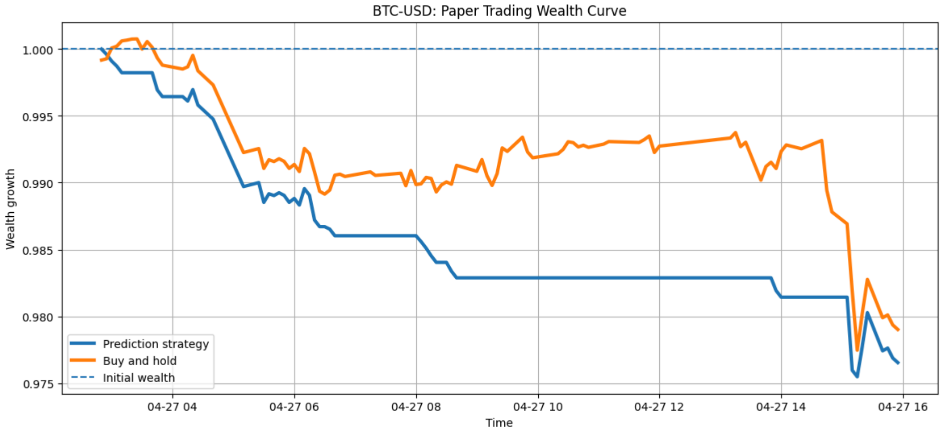
The term

$$c|u_t - u_{t-1}|$$

approximates transaction costs.

```
Latest probability of next return positive: 0.732839912184321
1
Latest signal: BUY / LONG
Final strategy wealth: 0.9765317534773483
Final market wealth: 0.9790093529433811
```





6. Almost Real-Time Interpretation

At the most recent candle, the model computes:

$$X_t$$

then estimates:

$$p_t = P(r_{t+1} > 0|X_t).$$

If

$$p_t > \theta,$$

the algorithm says:

$$u_t = 1.$$

Otherwise:

$$u_t = 0.$$

So the model behaves like an online decision system.

But it is still only paper trading.

```
Latest time: 2026-04-27 15:55:00+00:00
Latest close: 76772.921875

Predicted probability of positive next return: 0.732839912184
3211
Decision: BUY / LONG

Recent volatility vol20: 0.00248270751020715
Recent momentum mom5: -0.00350474207970064
Recent momentum mom10: -0.008598626031749879
```

7. Physics Interpretation

The classifier is a learned force field on market states.

Market state:

$$X_t$$

Prediction:

$$p_t = f_\theta(X_t)$$

Control:

$$u_t = \Theta(p_t - \theta).$$

So we have:

$$\text{market stochastic dynamics} + \text{learned decision field} = \text{controlled trading trajectory}$$

In statistical mechanics language, training minimizes an energy:

$$E(\theta)$$

while trading applies the learned configuration

$$\theta$$

to control wealth dynamics.

This algorithm is intentionally simple.

It does not yet include:

- robust walk-forward retraining,
- GARCH volatility features,
- Student-t heavy-tail modelling,
- stop-loss and take-profit execution,
- position sizing,
- slippage varying with liquidity,
- market regime detection,
- real broker execution.

But it combines the full chain:

$$S_t \to r_t \to X_t \to p_t \to u_t \to W_t.$$

Part 14: Making the Almost Real-Time Strategy More Realistic

The previous model used:

$$X_t \to p_t \to u_t$$

where

$$p_t = P(r_{t+1} > 0 | X_t)$$

and

$$u_t$$

is the trading decision.

Now we add:

1. walk-forward retraining,
2. GARCH volatility features,
3. Student-t heavy-tail diagnostics,
4. stop-loss and take-profit execution,
5. position sizing,
6. liquidity-dependent slippage,
7. regime detection,
8. paper-broker execution logic.

1. Walk-Forward Retraining

In real markets, we cannot train once and trust the model forever.

Instead we use walk-forward learning:

$$\text{train on past window} \to \text{predict next step} \to \text{move forward} \to \text{retrain}$$

At each time

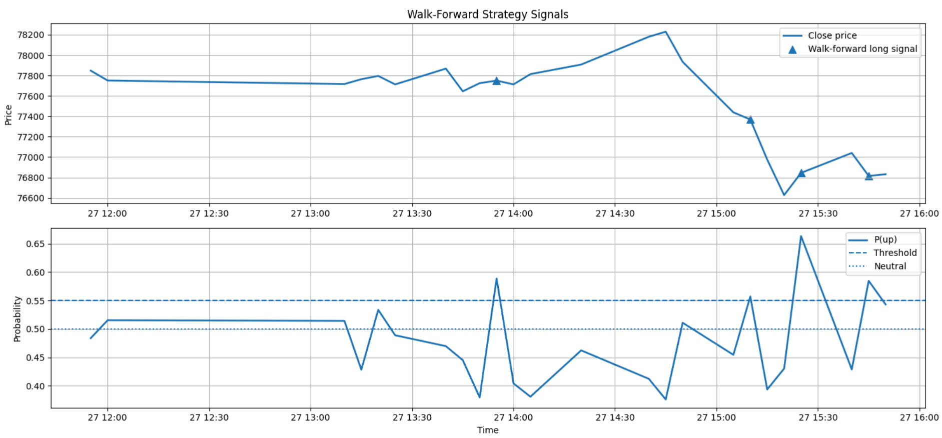
$$t,$$

the model only sees information before

$$t.$$

This avoids look-ahead bias.

	time	close	prob_up	signal
19	2026-04-27 15:20:00+00:00	76627.187500	0.430349	0
20	2026-04-27 15:25:00+00:00	76845.007812	0.663288	1
21	2026-04-27 15:40:00+00:00	77040.460938	0.428799	0
22	2026-04-27 15:45:00+00:00	76815.359375	0.584588	1
23	2026-04-27 15:50:00+00:00	76831.898438	0.543148	0



2. Add GARCH Volatility Feature

Previously we used rolling volatility:

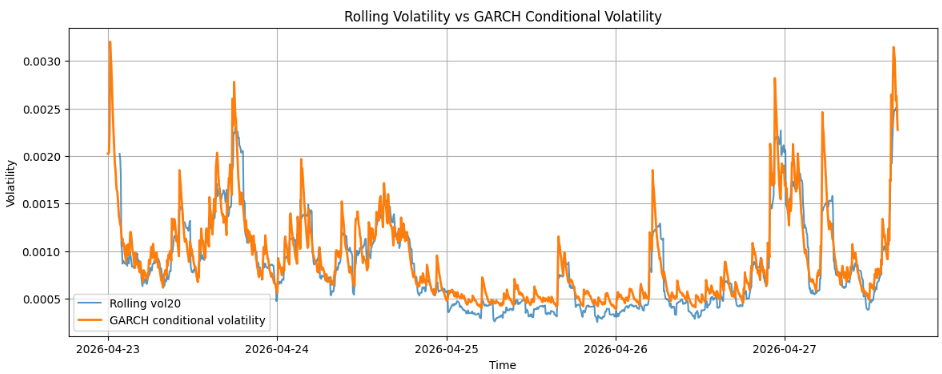
$$\sigma_t^{rolling}.$$

Now we add GARCH conditional volatility:

$$\sigma_t^{GARCH}.$$

This gives the model a learned estimate of conditional heteroscedasticity:

$$\mathrm{Var}(r_t|\mathcal{F}_{t-1}) = \sigma_t^2.$$



```
New feature matrix: (325, 9)
Target shape: (325,)
Any NaN? False
Any infinity? False
```

3. Student-t Heavy-Tail Model

Gaussian innovations assume thin tails:

$$Z_t \sim N(0,1).$$

But markets often have fat tails.

A Student-t model assumes:

$$Z_t \sim t_\nu.$$

The parameter

$$\nu$$

controls tail thickness.

Smaller

$$\nu$$

means heavier tails.

```
                        Constant Mean - GARCH Model Results
==============================================================
=======================
Dep. Variable:                  log_return   R-squared:
0.000
Mean Model:                  Constant Mean   Adj. R-squared:
0.000
Vol Model:                           GARCH   Log-Likelihood:
-1529.82
Distribution:      Standardized Student's t   AIC:
3069.64
Method:                 Maximum Likelihood   BIC:
3095.65
                                             No. Observation
s:                 1344
Date:                     Mon, Apr 27 2026   Df Residuals:
1343
Time:                             16:04:27   Df Model:
1
                                  Mean Model
==============================================================
================
                 coef    std err          t      P>|t|
95.0% Conf. Int.
--------------------------------------------------------------
----------------
mu             0.0141  1.456e-02      0.969      0.333 [-1.44
3e-02,4.266e-02]
                               Volatility Model
==============================================================
================
                 coef    std err          t      P>|t|
95.0% Conf. Int.
--------------------------------------------------------------
----------------
omega      6.4326e-03  4.734e-03      1.359      0.174 [-2.84
5e-03,1.571e-02]
alpha[1]       0.0869  2.628e-02      3.309  9.360e-04    [3.
545e-02,  0.138]
beta[1]        0.9131  2.832e-02     32.244 4.322e-228      [
0.858,  0.969]
                              Distribution
==============================================================
===========
                 coef    std err          t      P>|t|  95.0%
Conf. Int.
--------------------------------------------------------------
-----------
nu             3.7789      0.332     11.377  5.431e-30 [  3.1
28,  4.430]
==============================================================
===========

Covariance estimator: robust

Gaussian GARCH AIC: 3275.5257485105303
Student-t GARCH AIC: 3069.6353845352555
```

```
Estimated Student-t nu: 3.778909807024927
Gaussian GARCH AIC: 3275.5257485105303
Student-t GARCH AIC: 3069.6353845352555

Gaussian GARCH BIC: 3296.339370594863
Student-t GARCH BIC: 3095.652412140671
```

If Student-t GARCH has lower AIC or BIC, the data prefer heavy-tailed shocks.

This matters for strategy design because tail risk affects:

- stop-loss probability,
- drawdown risk,
- position sizing,
- threshold selection.

4. Stop-Loss and Take-Profit Execution

The model should not only say:

$$u_t = 1.$$

It should also define exits.

For a long spot trade:

$$S_{SL} < S_0 < S_T.$$

Stop-loss:

$$S_{SL} = S_0(1 - \delta_{SL}).$$

Take-profit:

$$S_T = S_0(1 + \delta_T).$$

A simple reward-risk rule is:

$$\delta_T = R_R \delta_{SL},$$

where

$$R_R$$

is the reward-to-risk ratio.

	entry_index	exit_index	entry_time	exit_time	entry_price	e
0	9	19	2026-04-27 13:55:00+00:00	2026-04-27 15:20:00+00:00	77750.000000	7697
1	20	23	2026-04-27 15:25:00+00:00	2026-04-27 15:50:00+00:00	76845.007812	7683

5. Position Sizing

We introduce account equity:

$$A_t.$$

Risk fraction per trade:

$$\rho.$$

Planned risk:

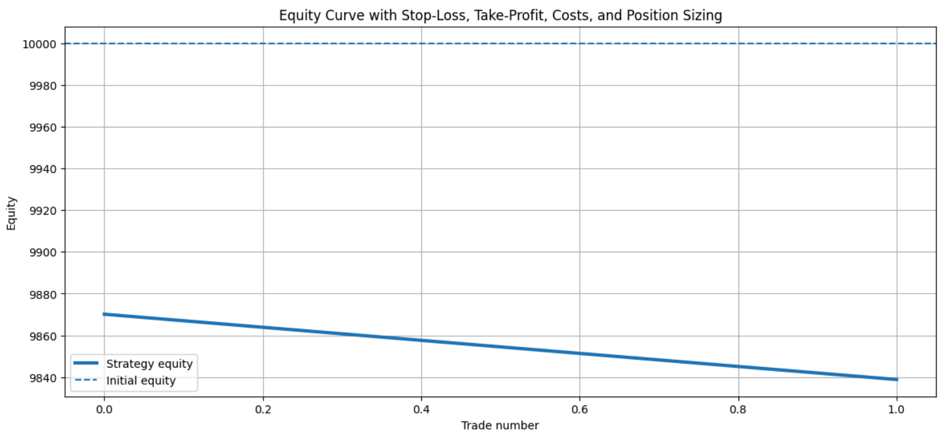
$$R_t = \rho A_t.$$

Position size:

$$Q_t = \frac{R_t}{|S_0 - S_{SL}|}.$$

This ensures every trade risks approximately the same fraction of capital.

	entry_index	exit_index	entry_time	exit_time	entry_price	e
0	9	19	2026-04-27 13:55:00+00:00	2026-04-27 15:20:00+00:00	77750.000000	7697
1	20	23	2026-04-27 15:25:00+00:00	2026-04-27 15:50:00+00:00	76845.007812	768




```
Number of trades: 2
Final equity: 9838.858277343681
Total net PnL: -161.14172265631905

outcome
stop_loss     1
time_exit     1
Name: count, dtype: int64
```

6. Liquidity-Dependent Slippage

Constant slippage is too simple.

Slippage should be worse when volume is low.

A simple model is:

$$s_t = s_0 \left(\frac{\bar{V}}{V_t} \right)^{\alpha}.$$

where:

$$s_0$$

is baseline slippage,

$$V_t$$

is current volume,

$$\bar{V}$$

is average volume,

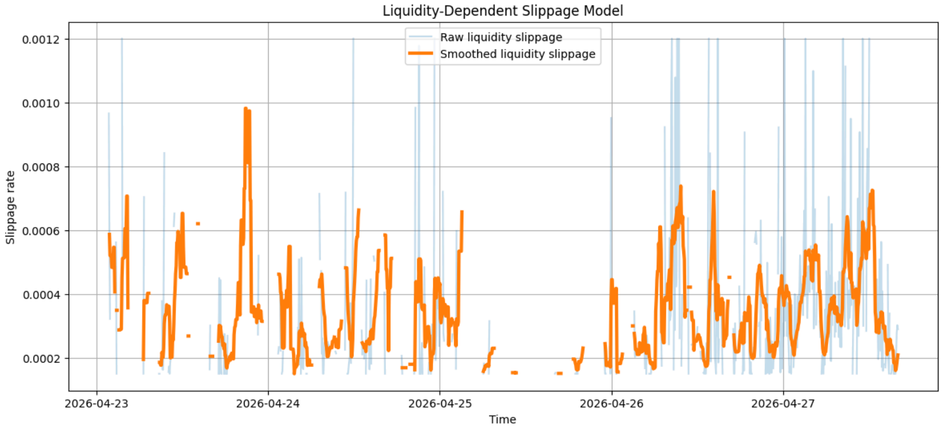
and

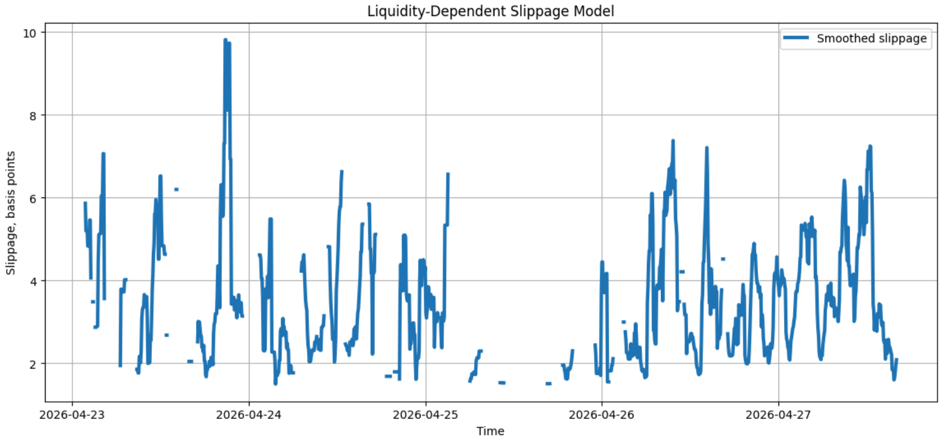
$$\alpha$$

controls sensitivity.

Low volume:

$$V_t < \bar{V} \Rightarrow s_t > s_0.$$





7. Market Regime Detection

We define regimes using trend and volatility.

Trend proxy:

$$m_t = \frac{S_t}{S_{t-k}} - 1.$$

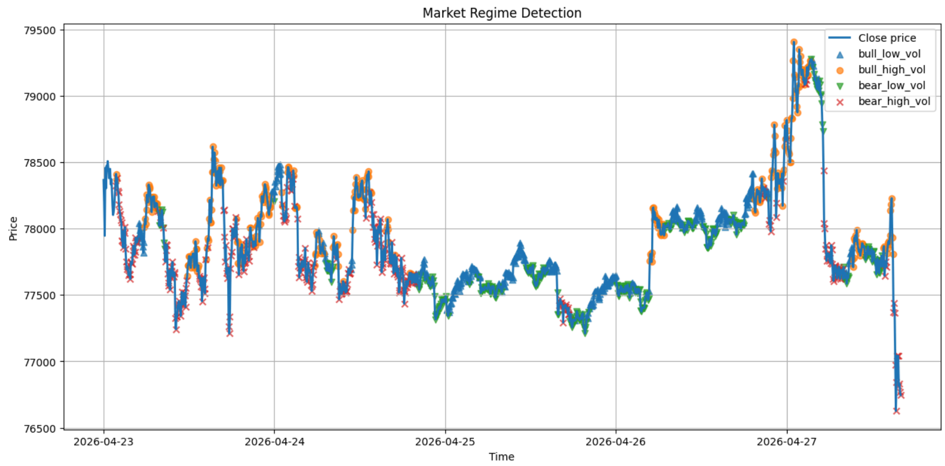
Volatility proxy:

$$\sigma_t.$$

Example regimes:

1. bull low-volatility,
2. bull high-volatility,
3. bear low-volatility,
4. bear high-volatility.

```
regime
bull_low_vol     344
bear_high_vol    342
bull_high_vol    320
bear_low_vol     319
Name: count, dtype: int64
```



8. Paper Broker Execution Layer

A broker execution layer translates signals into orders.

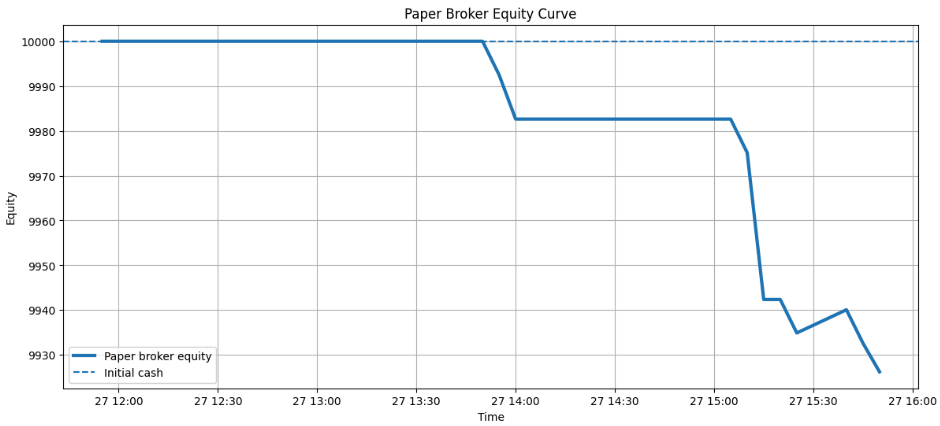
For research, we build a paper broker.

It tracks:

- cash,
- position,
- equity,
- entry price,
- fees,
- slippage,
- executed orders.

This is still simulation, not real broker execution.

	time	action	price	execution_price	quantity	fee
19	2026-04-27 15:20:00+00:00	hold	76627.187500	76627.187500	0.000000	0.000000
20	2026-04-27 15:25:00+00:00	buy	76845.007812	76883.430316	0.064691	4.973666
21	2026-04-27 15:40:00+00:00	sell	77040.460938	77001.940707	0.064691	4.981333
22	2026-04-27 15:45:00+00:00	buy	76815.359375	76853.767055	0.064701	4.972521
23	2026-04-27 15:50:00+00:00	sell	76831.898438	76793.482488	0.064701	4.968621




```
Final paper broker equity: 9926.231057028705
Number of buys: 4
Number of sells: 4
```

9. Final Integrated Model

We now have the chain:

$$S_t \to r_t \to X_t \to p_t \to u_t \to \text{order} \to W_t.$$

Where:

$$X_t$$

contains:

- recent returns,
- momentum,
- rolling volatility,
- GARCH volatility,
- volume,
- regime information.

The model includes:

- walk-forward retraining,
- heteroscedasticity,
- heavy-tail diagnostics,
- stop-loss and take-profit,
- position sizing,
- liquidity-dependent slippage,
- regime detection,
- paper broker execution.

This is now a research-level prototype for stochastic-learning-based spot strategy analysis.