A Physicist's Perspective on Quantitative Finance

Adrià Garcés[1]

[1] *Departament de Física de la Matèria Condensada, Universitat de Barcelona, Barcelona, Spain*

E-mail: adria.garces@ub.edu

Abstract: This project aims to build a comprehensive collection of quantitative finance models from the perspective of a physicist. Each topic is explored through detailed notes, combining analytical tools from stochastic calculus, statistical mechanics, and numerical simulation. The goal is to bridge physical intuition and financial modeling by documenting a range of classical models while progressively incorporating real data analysis. All developments, explanations, and code are shared as a series of standalone "pills" in a structured GitHub repository. Acess the repository here.

Contents

1 Introduction

This repository gathers a growing collection of projects exploring key concepts in quantitative finance from a physicist's perspective. Each project includes theoretical notes, numerical simulations, and applications to real data, blending physical intuition with financial modeling. Topics range from stochastic processes and option pricing to interest rate models, portfolio theory, and statistical arbitrage. The notes are designed to be rigorous yet intuitive. Some background in differential equations, probability, and basic statistics is recommended, along with familiarity with Python and scientific computing tools.

Este repositorio reúne una colección creciente de proyectos que exploran conceptos clave de finanzas cuantitativas desde la perspectiva de un físico. Cada proyecto incluye apuntes teóricos, simulaciones numéricas y aplicaciones con datos reales, combinando intuición física con modelado financiero. Se abordan temas como procesos estocásticos, valoración de opciones, tipos de interés, teoría de carteras y arbitraje estadístico. Los apuntes buscan ser rigurosos pero accesibles; se recomienda tener conocimientos previos de ecuaciones diferenciales, probabilidad, estadística básica y cierta familiaridad con Python y herramientas de computación científica.

Aquest repositori recull una col·lecció creixent de projectes que exploren conceptes clau de finances quantitatives des del punt de vista d'un físic. Cada projecte combina apunts teòrics, simulacions numèriques i aplicacions a dades reals, unint la intuïció física amb el modelatge financer. Es tracten temes com processos estocàstics, preus d'opcions, tipus d'interès, teoria de carteres i arbitratge estadístic. Els apunts són rigorosos però accessibles; es recomana tenir nocions prèvies d'equacions diferencials, probabilitat i estadística bàsica, així com experiència amb Python i eines de càlcul científic.

Bu depo, bir fizikçinin bakış açısından nicel finansın temel kavramlarını keşfeden projelerden oluşan bir koleksiyondur. Her proje, teorik notlar, sayısal simülasyonlar ve gerçek verilere uygulamalar içerir; fiziksel sezgi ile finansal modellemeyi birleştirir. Ele alınan konular arasında stokastik süreçler, opsiyon fiyatlama, faiz oranı modelleri, portföy teorisi ve istatistiksel arbitraj yer alır. Notlar titiz ama sezgisel bir şekilde hazırlanmıştır. Diferansiyel denklemler, olasılık teorisi ve temel istatistik bilgisi ile Python ve bilimsel hesaplama araçlarına aşinalık önerilir.

2 Geometric Brownian Motion (GM)

2.1 Introduction

In finance, models capturing the dynamics of the price of an asset in market are essential – they both theoretical understanding and background and have practical applications.

Just like in physics stochastic models are used to understand the dynamics of physical observables of a system, i.e. magnetization of a material in a bath at temperature T which provides the system with fluctuations, modeling the price of an asset can also be done in a very similar way. The fluctuations (stochastic nature), in this case, is provided by a more complex environment that a single bath at temperature T – the interactions and interests of individuals who intervene in the market.

Brownian motion (BM) is the core of most stochastic models, in physics and everywhere else. However, it cannot directly be used to model asset prices, since nothing prevents a Brownian variable X_t to be negative, and asset prices are *strictly positive*. A modified version of Brownian motion which guarantees positiveness of asset prices is instead used.

In the following, let S_t be the price of an asset at time $t \in \mathbb{R}^+$. Positiveness of S_t is guaranteed if, for example, $S_t = S_0 e^{Z_t}$, where Z_t is a *stochastic* (random) function that depends on t and might generally depend on the previous history of S_t, verifying $Z_t|_{t=0} = 0$, so that $S_t|_{t=0} = S_0 > 0$. A way of achieving this is *geometric Brownian motion* (GBM), which guarantees positiveness of assets at any time t. Furthermore, GBM incorporates empirically observed phenomena such as the proportionality of an observable called returns, $R_t = dS_t/S_t$ – fluctuation of an assets price are usually relative to its current price, not absolute.

2.2 Theoretical Background

Modeling the stochastic evolution of an asset's price S_t guaranteeing positiveness and such that fluctuations are relative to the price locally in time can be done using, as mentioned in the introduction, GBM. We say the asset price S_t performs GBM if its statistical behavior can be drawn by trajectories satisfying the following stochastic differential equation (SDE),

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{2.1}$$

where W_t is the Wiener process[1], μ is a constant usually named *"percentage drift"* and σ is another constant usually referred to as *"percentage volatility"*. Before we keep going with the definition of interesting and practical observables in economics and finance we shall take a look at tools that come from physics that can help us understand the implications of this model.

2.2.1 The Fokker–Planck Equation

Directly from the SDE Eq. (2.1) the Fokker–Planck equation for the probability density function of finding the price to be $S_t = s$ at some time t is [**Ref**]

$$\frac{\partial}{\partial t} p_S(s,t) = -\mu \frac{\partial}{\partial s}[s p_S(s,t)] + \frac{1}{2}\sigma^2 \frac{\partial^2}{\partial s^2}[s^2 p_S(s,t)]. \tag{2.2}$$

Unfortunately, Eq. (2.2) is not linear and not readily solved. Using the change of variables $Z_t = \ln S_t$ ($z = \ln s$) we can find the Fokker–Planck equation not for Z_t, using that $p_Z(z,t) =$

[1]See [**?**] a brief introduction of what the Wiener process is.

$p_S(s = e^z, t)\, |\partial s/\partial z| = p_S(s = e^z, t)e^z$ and $\partial_s = e^{-z}\partial_z$ and $\partial_s^2 = e^{-2z}(\partial_z^2 - \partial_z)$, the FP equation becomes,

$$\frac{\partial p_Z(z,t)}{\partial t} = -\left(\mu - \frac{\sigma^2}{2}\right)\frac{\partial p_Z(z,t)}{\partial z} + \frac{1}{2}\sigma^2 \frac{\partial^2 p_Z(z,t)}{\partial z^2}. \tag{2.3}$$

Eq. (2.3) is the FP equation of a *Brownian* particle moving with velocity $v = -(\mu - \sigma^2/2)$ and diffusion coefficient $D = \sigma^2/2$. Note how in the case of a Brownian particle, the diffusion term comes from thermal fluctuations in the bath and its velocity does not depend on it, while here the drift term does depend on the volatility σ – we will see what the implications of this are in the following. Without loss of generality and setting the initial condition to be $p_S(s, 0) = \delta(s-1)$ so that $p_Z(z, 0) \sim \delta(z)$, the solution of Eq. (2.3) becomes

$$p_Z(z,t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(z - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t}\right). \tag{2.4}$$

This – just as in the case of a Brownian particle moving with velocity v – means that $Z_t \sim \mathcal{N}(\mu_t, \sigma_t)$, with $\mu_t = (\mu - \sigma^2/2)t$ and $\sigma_t = \sigma\sqrt{t}$. Changing variable back again, the probability density function for an asset's price at finite time t becomes

$$\boxed{p_S(s,t) = \frac{1}{s\sqrt{2\pi\sigma^2 t}} \exp\left(-\frac{(\ln s - (\mu - \sigma^2/2)t)^2}{2\sigma^2 t}\right)}, \tag{2.5}$$

which then makes S_t to be *"lognormally"* distributed, that is $S_t \sim \text{Lognormal}(\mu_t, \sigma_t)$, also denoted as $\ln S_t \sim \mathcal{N}(\mu_t, \sigma_t)$. This follows the general way of defining GBM, since $S_t \sim e^{Z_t}$ where $Z_T \sim \mathcal{N}(\mu_t, \sigma_t)$. Note how the expected price of the asset exponentially increaes with time $\mathbb{E}[S_t] = \exp(\mu_t + \sigma_t^2/2) = e^{\mu t}$. This can also be directly seen using the Itô's prescription, for which the Brownian the price of the assets at time t is *not conditioned* to the Weiner process at the same time t (it does, however, in the case of the Stratonovich discretization scheme). In the Itô prescription,

$$\mathbb{E}[dS_t] = \mathbb{E}[\mu S_t dt] + \mathbb{E}[\sigma S_t dW_t] \Leftrightarrow d\mathbb{E}[S_t] \overset{\text{Itô}}{=} \mu\mathbb{E}[S_t]dt + \sigma\mathbb{E}[S_t]\mathbb{E}[dW_t] = \mu\mathbb{E}[S_t]dt. \tag{2.6}$$

so that again $\mathbb{E}[S_t] = e^{\mu t}$ when $S_0 = 1$. It is also important to note that if S_t follows GBM, then

$$X_t = \frac{Z_t - \mu_t}{\sigma_t} = \frac{\ln S_t - (\mu - \sigma^2/2)t}{\sigma\sqrt{t}} \sim \mathcal{N}(0,1). \tag{2.7}$$

This basically means histograms of either simulated or real data GBM collapses into a standard gaussian distribution with the proper change of variables. Note how when the initial price of the asset is $S_0 \neq 1$, then the only difference is that one has to add $-\ln S_0$ to the expected value, meaning

$$X_t = \frac{r_t - (\mu - \sigma^2/2)t}{\sigma\sqrt{t}} \sim \mathcal{N}(0,1), \tag{2.8}$$

where $r_t = \ln(S_t/S_0)$. The quantity r_t is called the logarithmic return and is widely used in econometrics.

2.2.2 The Binomial Approximation to GBM

The SDE in Eq. (2.1) can be used as a reference to study time–discrete models which accurately describe GBM. Taking time to be divided in intervals of length Δt, Eq. (2.1) can be written the following way

$$S_{t+\Delta t} - S_t = \mu S_t \Delta t + \sigma S_t \Delta W_t \Leftrightarrow \boxed{S_{t+\Delta t} = S_t\left(1 + \mu\Delta t + \sigma \Delta W_t\right)}, \tag{2.9}$$

where $\Delta W_t := W_{t+\Delta t} - W_t = \int_t^{t+\Delta t} dW_t \sim \mathcal{N}(0, \Delta t)$. In order to ease notation we now name times starting from t_0 with index $k \in \mathbb{Z}^+$, i.e. $t_k = t_0 + k\Delta t$ and we simply denote $S_{t_k} := S_k$. Then,

$$\begin{aligned} S_k &= S_{k-1}(1 + \mu\Delta t + \sigma\Delta W_k) \\ &= S_{k-2}(1 + \mu\Delta t + \sigma\Delta W_{k-1})(1 + \mu\Delta t + \sigma\Delta W_k) \\ &\;\vdots \\ &= S_0 \prod_{i=1}^{k}(1 + \mu\Delta t + \sigma\Delta W_i) =: S_0 L_1 L_2 \ldots L_k, \end{aligned} \tag{2.10}$$

where we have defined $L_i = S_i/S_{i-1} = (1+\mu\Delta t+\sigma\Delta W_i)$. Note how since $\Delta W_i \sim \mathcal{N}(0, \Delta t)$, then $L_i \sim \mathcal{N}(1 + \mu\Delta t, \sigma^2\Delta t)$. This is fundamentally interesting – GBM can be sampled discrete time (with intervals of length Δ) by simply generating a sequence of jumps $L_i \sim \mathcal{N}(1 + \mu\Delta t, \sigma^2\Delta t)$ and then multiply them all together. Note how there is no correlation between two different jumps L_i, L_j with $j \neq i$. A sequence of this indepentdent identically distributed jumps $\{L_i\}_{i=1}^n$ generates a unique *"path"* for the evolution of the price of the asset, $S_n = S_0 \prod_{i=1}^n L_i$. This sets a clear pathway in order to simulate discrete time GBM without requiring the SDE, only using random numbers properly distributed, when Δt is small enough.

However, since $L_i \sim \mathcal{N}(1+\mu\Delta t, \sigma^2\Delta t)$ the jump take any value an a given time step, and this is not very comfortable when looking for analytical results. A very common an natural thing to do in order to simplify the problem, is restricting the possible values of the jumps to take two values (going up or down), $S_i \to S_{i+1} = uS_i$ (u from up) with some probability p and $S_i \to S_{i+1} = dS_i$ (d from down), with $u > 1 > d$, and then taking $ud = 1$ such that the composition of and up and down movement return to the price two steps before $S_{i+2} = udS_i = duS_i = S_i$. This is done this way such that it possesses *"recombining"* property, i.e. the evolution of the asset's price can be represented as depicted in Fig. (1).

The price of the asset at time n thus can be written $S_n = S_0 \prod_{i=1}^n X_i$ where $X_i \in \{u, d\}$ is a Bernoulli trial with probabilities $p(X_i = u) = p$, $p(X_i = d) = 1 - p$. This makes the asset's price to be Binomial distributed. Choosing $ud = 1$ convenient, but in order recover GBM the expected value and variance of a jump have to mapped. In the case of GBM, it can be shown that, for fixed S_t

$$\mathbb{E}[S_{t+\Delta t}|S_t] = S_t e^{\mu\Delta t} \tag{2.11}$$

$$\mathbb{E}[S_{t+\Delta t}^2|S_t] = S_t^2 e^{(2\mu+\sigma^2)t}, \tag{2.12}$$

while for the case of the Binomial approximation, since $S_{n+1} = uS_n$ with probability p, and $S_{n+1} = dS_n$ with probability $1 - p$,

$$\mathbb{E}[S_{n+1}|S_n] = S_n[up + d(1 - p)] \tag{2.13}$$

$$\mathbb{E}[S_{n+1}^2|S_n] = S_n^2[u^2p + d^2(1 - p)], \tag{2.14}$$

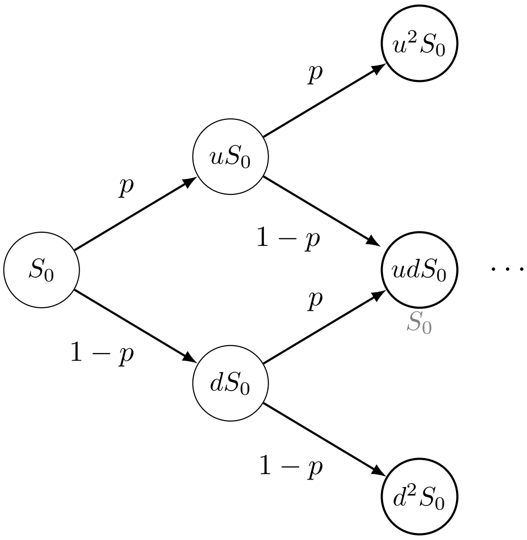

Figure 1: Recombining binomial evolution of an asset's price when $ud = 1$.

then the following system,

$$ud = 1 \tag{2.15}$$

$$up + d(1-p) = e^{\mu \Delta t} \tag{2.16}$$

$$u^2 p + d^2(1-p) = e^{(2\mu+\sigma^2)t}, \tag{2.17}$$

guarantees that the jumps are going to comply with the *"recombining property"* (esential for S_n to follow a Binomial distribution) and will have the same expected value and variance as the jumps provided by GBM[**?**]. The solution to this is a little bit involved but writes,

$$p = \frac{e^{\mu\Delta t} - d}{u-d} \tag{2.18}$$

$$u = \frac{e^{-\mu\Delta t}}{2}\left[1 + e^{\sigma^2\Delta t} + \sqrt{(1+e^{\sigma^2\Delta t})^2 - 4e^{2\mu\Delta t}}\right] \tag{2.19}$$

$$d = 1/u. \tag{2.20}$$

The expression for u is not very comfortable to work with. Since, however, we expect the discrete time model to resemble to GBM as $\Delta t \to 0$, a rather practical and approximated expressions for u and d can be shown be to,

$$u = e^{\sigma\sqrt{\Delta t}}, \quad d = e^{-\sigma\sqrt{\Delta t}}, \tag{2.21}$$

along with $p = (e^{\mu\Delta t} - d)/(u-d)$. This expressions for u, d and p were introduced by Cox, Ross, and Rubinstein [**Ref**] in a binomial model for option pricing. A simpler expression for u, d and p can be found by simply mapping the expected value and variance of X_i to the ones of $L_i \sim \mathcal{N}(1+\mu\Delta t, \sigma^2\Delta t)$.

2.3 Simulation

The pill "aa" found in the GitHub repository includes the numerical integration of the SDE as well as Monte Carlo (MC) simulations using the Binomial approximation and using real data of the price of an asset to test the hypothesis of the model.

2.3.1 Numerical integration of the SDE

The SDE of GBM for the price of the price of an asset as time evolves can be integrated numerical using the Euler–Maruyama method. For this case, given the initial condition S_0,

$$S_{t+\Delta t} = S_t \left(1 + \mu \Delta t + \sigma \sqrt{\Delta t} \chi_t \right), \tag{2.22}$$

where χ_t is sampled from a standard normal distribution, $\chi_t \sim \mathcal{N}(0,1)$. Without loss of generality and in order to be consistent with the initial condition we used in the theoretical background, we set $S_0 = 1$.

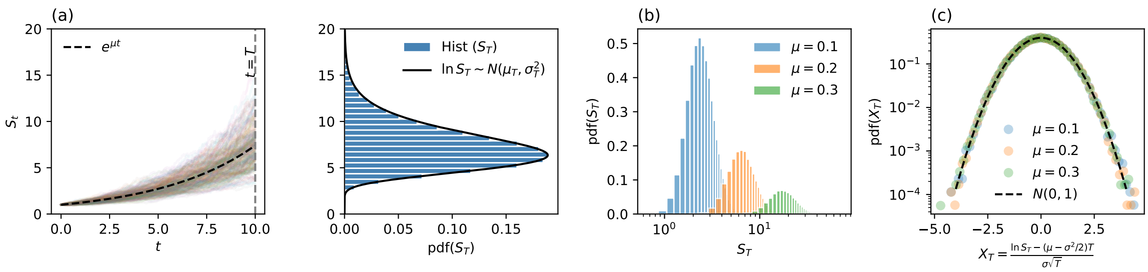

Figure 2: **Geometrical Brownian Motion.** (a) Stochastic evolution of $G = 250$ independent realizations of the evolution of an asset with initial price $S_0 = 1$ and $\mu = 0.1, \sigma = 0.1$. The black dashed line represents the expected average trajectory $S_t = S_0 e^{\mu t}$ and the vertical gray dashed line the final time of the simulation, $T = 10$ ($\Delta t = 10^{-2}$). On the right, the histogram of the values of $G = 10^5$ differentes trajectories at time $t = T$ (again $\Delta t = 10^{-2}$) is shown. The black line is the corresponding probability density function of the lognormal distribution with $\mu_t = (\mu - \sigma^2/2)T$ and $\sigma_T^2 = \sigma^2 T$; (b) Histograms of the price at time $T = 10$ of $G = 10^5$ for different values of μ for $\sigma = 0.1$ and $S_0 = 1$. The distributions look Gaussian as soon as the S_T axis is represented in logartihmic scale, as expected since $\ln S_T \sim \mathcal{N}(\mu_T, \sigma_T^2)$; (c) Histograms of the standarized variable $X_t = (\ln S_T - \mu_T)/\sigma_T$ for the same values of μ and σ in (b). As shown, all the distributions collapse into a standard normal distribution of 0 mean and unit variance.

2.3.2 Monte Carlo simulation with Binomial approximation

The Binomial approximation of GBM can be implemented using Monte Carlo (MC) dynamics. We start from an initial asset price $S_0 = 1$, without loss of generality, and make the price evolve following

$$S_{k+1} = \begin{cases} uS_k & \text{with probability } p \\ dS_k & \text{with probability } 1-p, \end{cases} \tag{2.23}$$

where again $u = e^{\sigma\sqrt{\Delta t}}$, $d = 1/u$ and $p = (e^{\mu\Delta t} - d)/(u - d)$ so that when Δt is small enough, then at time $t = n\Delta t$, $S_t \sim \text{Lognormal}(\mu_t, \sigma_t)$. Each MC step consists on generating a random number uniformly distributed between 0 and 1, $z \sim U(0,1)$, and given the price of the asset at time k, setting $S_{k+1} = uS_k$ if $z < p$ and $S_k = dS_k$ otherwise. As shown in Fig. (3,a) the dynamics followed by different realizations of the stochastic dynamics portrays the shape of the recombining tree depicted in Fig. (1). At the same time, a bigger time scales the qualitative behaviour of the trajectories simulated through the MC procedure of the Binomial approximation, Fig. (3,b),

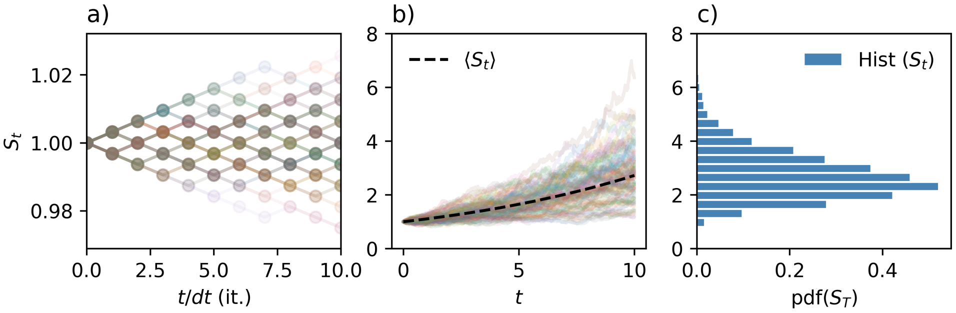

Figure 3: **Recombinin tree and MC dynamics.** (a) Evolution of the price of $G = 10^2$ different realizations over 10 MC steps with $S_0 = 1$ for $\mu = 0.1$, $\sigma = 0.1$, $\Delta t = 10^{-2}$, $u = e^{\sigma\sqrt{\Delta t}}$, $d = 1/u$ and $S_0 = 1$. (b) Evolution of the asset price of $G = 10^4$ different realization for the same values of $\mu, \sigma, \Delta t, p, u$ and d now at bigger time scales. The dynamics of the asset price now resembles more the one obtained through the direct numerical integration of the SDE. The dashed line is the noise average trajectory $\langle S_t = (1/N)\sum_{i=1}^{N} S_t^t$. (c) Histogram of the asset's price after time $T = 10$. The shape still resembles the one of a log–normally distributed asset price.

as well as the histogram at time $T = 10$, Fig. (3,c) qualitatively resemble the ones obtained through the direct numerical integration of the SDE. The dashed line in Fig. (3) represents the trajectory averaged over the noise $\langle S_t \rangle = (1/N)\sum_{i=1}^{N} S_t^{(i)}$ and looks, exponential. In order to quantify how good the Binomial approximation is we perform simulations for different values of μ, keeping σ and Δt the same. Indeed, as seen in Fig. (4,a) $\log\langle S_t \rangle \sim \mu t$, recovering what we know from GBM. As also shown in Fig. (4,b-c), the lognormal distribution with parameters μ, σ and Δt perfectly fit the distributions at time $T = N\Delta t$ (N is the number of MC steps). The data can be made to collapse, again, into a standard normal distribution with the proper change of variables.

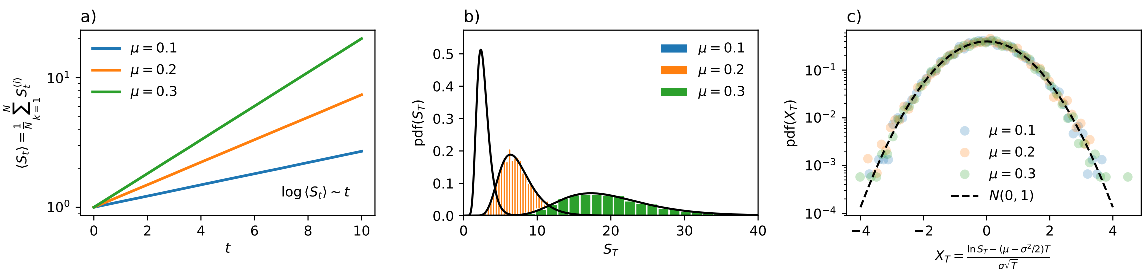

Figure 4: **Binomial approximation and GBM.** (a) Average trajectory of $G = 10^4$ different stochastic realizations for different values of μ and fixed $\sigma = 0.1$, $\Delta t = 10^{-2}$ and initial price $S_0 = 1$. The logartihmic scale of the price shows how $\log\langle S_t \rangle \sim \mu t$, as expected from GBM. (b) Same parameters, now showing the histogram obtained after a running time $T = 10$. The black lines are the probability density function obtained from a log–normal distribution with mean$\mu_T = (\mu - \sigma^2/2)T$ and variance $\sigma_T^2 = \sigma^2 T$ (no fitting parameters). (c) Collapse of the distributions in (b) into a single distribution. The black dashes line is a normal gaussian distribution with mean 0 and unit variance.

2.3.3 Real data from and GBM

In order to test the real applicability of the model studied we can directly take data from the real price of an asset of the market. In this case, we will take a look at the evolution of the price of an asset of ALPHABET INC. (GOOG).

Accessing financial data can be done using Python, by installing the library YahooFinance. A very basic tutorial on accessing data in found inside the pill of this topic, check out the GitHub repository . We particularly access data of the price of an asset between 2020 and 2025. We keep record only of the closing asset price of each day, and thus have data for every single (trading) day between 01/01/2020 and 01/01/2025. The daily return is computed as follows $r_{d,k} = \ln(S_k/S_{k-1})$ where S_k is the closing price of day k and S_{k-1} is the closing price of the previous day. As shown in Fig. (5,c) the distribution of logarithmic returns seems to be normally distributed, with numerical means and variace $\bar{r}$ and σ_r (how these are computed is specified in the pill). Interestingly enough, the distribution of daily log returns have *"fat"* tals, that diverge from the lognormal behaviour.

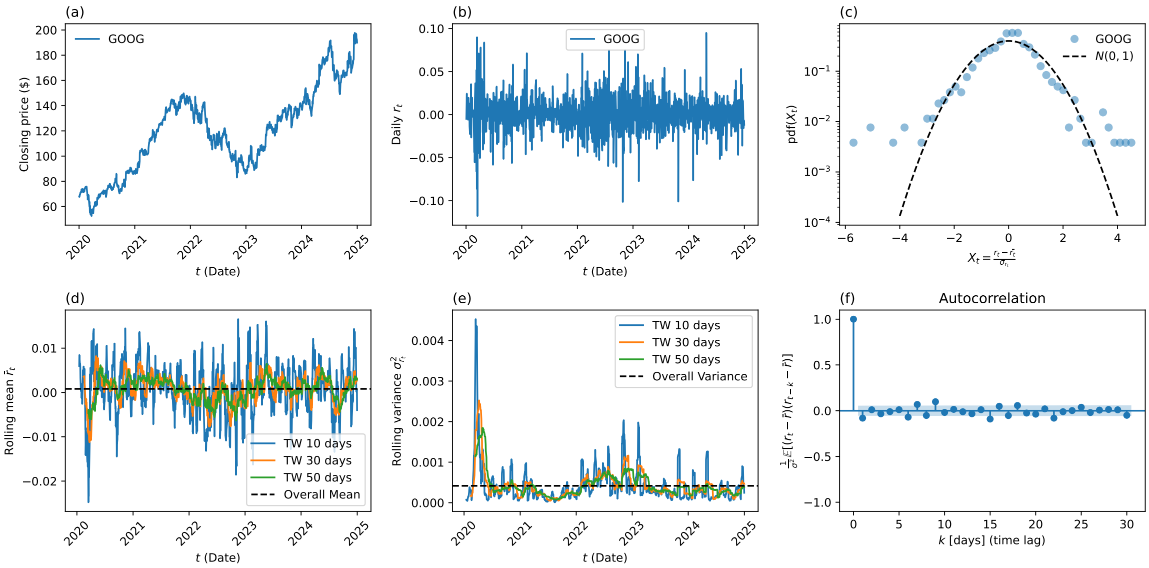

Figure 5: **Real data & GBM.** (a) evolution of the daily closing price of an asset of Alphabet Inc. between 01/01/2020 and 01/01/2025, (b) evolution of the logarithimic daily return $r_{d,k} = \log(S_k/S_{k-1})$ in the same time gap as the one in (a). (c) Distribution of the logartihmic daily returns. The dashed black line presents a normal distribution with mean 0 and unit variance. (d-e) Evolution of the mean and variance of the logarithmic daily returns over time windows of 10, 30 and 50 years along the 5 years of data. The dashed lines represent the overable average and variance. (f) Autocorrelation function of the logartihmic daily returns for different time lags k.

Key hypothesis of GBM include the fact that returns are independently and identically sampled, with constant mean and variance, e.g. recall that for GBM $L_k = S_k/S_{k-1} \sim \mathcal{N}(1 + \mu\Delta t + \sigma\Delta W_k)$. We put this to test by computing the average value and variance of the logarithmic daily return over different time windows of 10, 30 and 50 days over the 5 years of data. In Fig. (5,d-e) the average value over these time window gaps are shown. The mean logartihmic daily return seems to fluctuate homogeneously around the global mean value, wheres the variance presents peaks of activity in certain periods of time. While assuming the mean value of daily log–returns is more or less acceptable, the hypothesis does not seem to directly hold for the variance. In order to check the no–correlation betwen logartihmic daily returns we compute the autocorrelation function $C_k = (1/\sigma^2)\langle(r_t - \bar{r})(r_{t-k} - \bar{r})\rangle$, where k is called time lag. As shown in Fig. (5,f), logarithmic daily returns do not seem to be too correlated. Another way to compute autocorrelations in a tim series is by using the Ljung–Box test[2]. A direct implementation in Python (see pill) shows that indeed returns are not completely independent, meaning this hypothesis also break.

[2]Fin more information on what the Ljung–Box test is and what criteria is applied here.