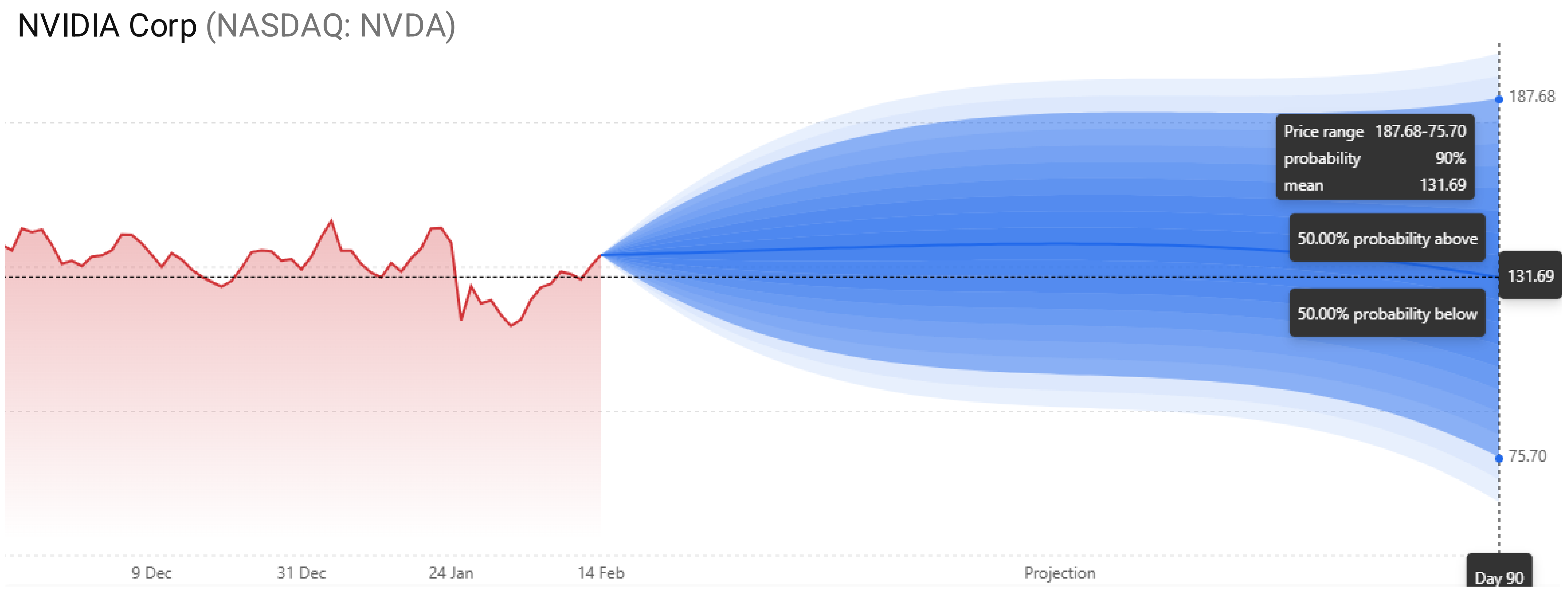
Time Series Modelling

Sukrit Mittal

Session goals:
- Understand temporal structure
- Build forecasting models
- Quantify risk & uncertainty

Session Agenda

- **Concepts**:
  - Time series, stationarity, prices vs returns
- **Dependence structure**:
  - Autocovariance, autocorrelation, partial autocorrelation
- **Models for the mean**:
  - AR, MA, ARMA, ARIMA



NVIDIA Corp (NASDAQ: NVDA)
187.68
Price range 187.68-75.70
probability 90%
mean 131.69
50.00% probability above
131.69
50.00% probability below
75.70
9 Dec
31 Dec
24 Jan
14 Feb
Projection
Day 90

Unrealistic Expectations: The Futility of Precisely Estimating a Stock's Expected Return

Sanjiv R. Das
Santa Clara University

Daniel Ostrov
Santa Clara University

December 12, 2023

...dedicated to the memory of Mark S. Joshi,
who worked to make results like these better known

What is a Time Series?

- A time series is simply a sequence of observations ordered in time.
- Formally, we often write this as $\{X_t\}_{t \in T}$, where t indexes time. In our context, T is usually discrete, such as:
  - Trading days (e.g., $t = 1,2,3, \ldots$).
  - Calendar months or quarters.
  - Intraday intervals (e.g., 5-minute bars).
- Examples from finance:
  - P_t: daily closing price of an index or stock.
  - r_t: daily log-return on that asset.
  - y_t: yield on a 10-year government bond.
  - s_t: credit spread or volatility index level.

Stochastic Processes & Notation

- A **stochastic process** is a collection of random variables indexed by some set, often time. A time series is therefore a specific type of stochastic process with time as the index.
- We will work with:
  - Discrete time: $t = 0,1,2, \ldots$
  - Real-valued processes: $X_t \in \mathbb{R}$ for each t.
- From a practical perspective, you can think of a stochastic process as "a random mechanism that generates a sequence of numbers over time."
- At each time point t, you observe a random variable X_t.
- Our modelling goal is to say something about the joint behavior of this entire sequence $\{\mathrm{X}_t\}$, not just about the marginal distribution of a single X_t.

Stationarity: Strict vs Weak

- A key concept in time series analysis is **stationarity**. Informally, stationarity means that the statistical properties of the process do not change over time.

- **Strict stationarity:** A process $\{X_t\}$ is strictly stationary if the joint distribution of any finite subset $\left(X_{t_1}, \ldots, X_{t_k}\right)$ is invariant to shifts in time. That is, for all k, all time indices $t_1, \ldots, t_k$,and all shifts h:
$$\left(X_{t_1}, \ldots, X_{t_k}\right) \stackrel{d}{=} \left(X_{t_1+h}, \ldots, X_{t_k+h}\right)$$

- **Weak (or covariance) stationarity:** A process $\{X_t\}$ is weakly stationary if:
  - The mean is constant over time: $\mathbb{E}[X_t] = \mu$ for all t.
  - The variance is constant: $\mathrm{Var}(X_t) = \sigma^2$ for all t.
  - The covariance between values at times t and $t + h$ depends only on the lag h, not on the specific time t:
$$\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h).$$

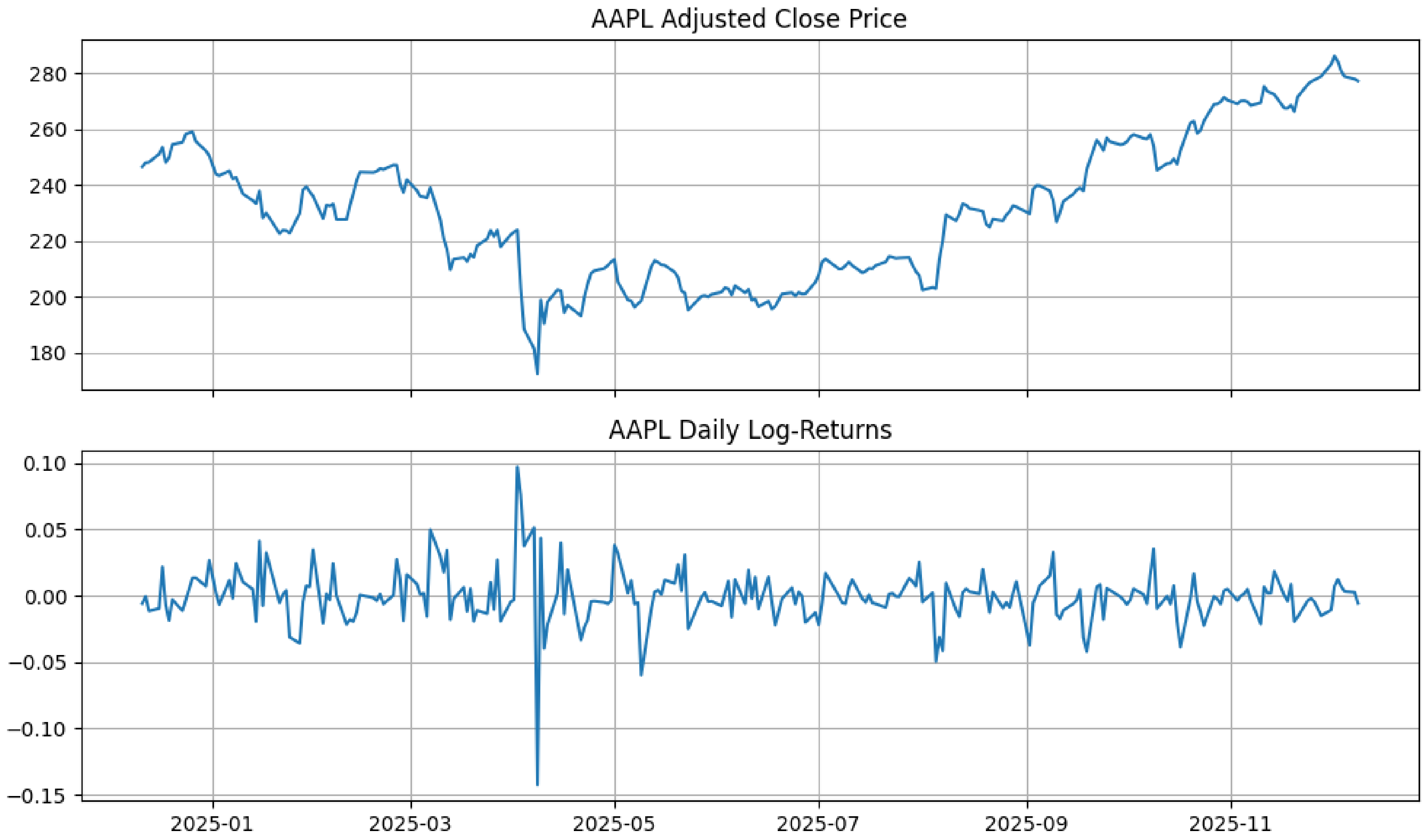
Why Stationarity Matters

- It ensures that quantities like the mean, variance, and autocorrelation are well-defined and stable over time.
- Many of the asymptotic results and estimation techniques rely on the process not "drifting" too much over time.
- In financial data, raw price series typically exhibit trends, structural breaks, and evolving volatility. Implying, they are not stationary.
- A big part of the modelling workflow is to transform the data so that the transformed series is "stationary enough" for these tools to be applicable.
  - For example, working with returns instead of prices directly.

Prices vs Returns

- Let P_t denote the price of an asset at time t. Then:
  - The **simple return** between $t-1$ and t is
$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$
  - The **log-return** (or continuously compounded return) is
$$r_t = \log P_t - \log P_{t-1} = \log\left(\frac{P_t}{P_{t-1}}\right)$$
- Why returns over prices?
  - Prices often show a clear upward (or downward) trend over long horizons, making them non-stationary. Returns typically fluctuate around a more stable mean, often near zero for short horizons (daily or weekly).
  - Log-returns have useful aggregation properties: *the log-return over multiple periods is the sum of the log-returns over subperiods*.

Prices vs Returns



Visual Checks for Stationarity

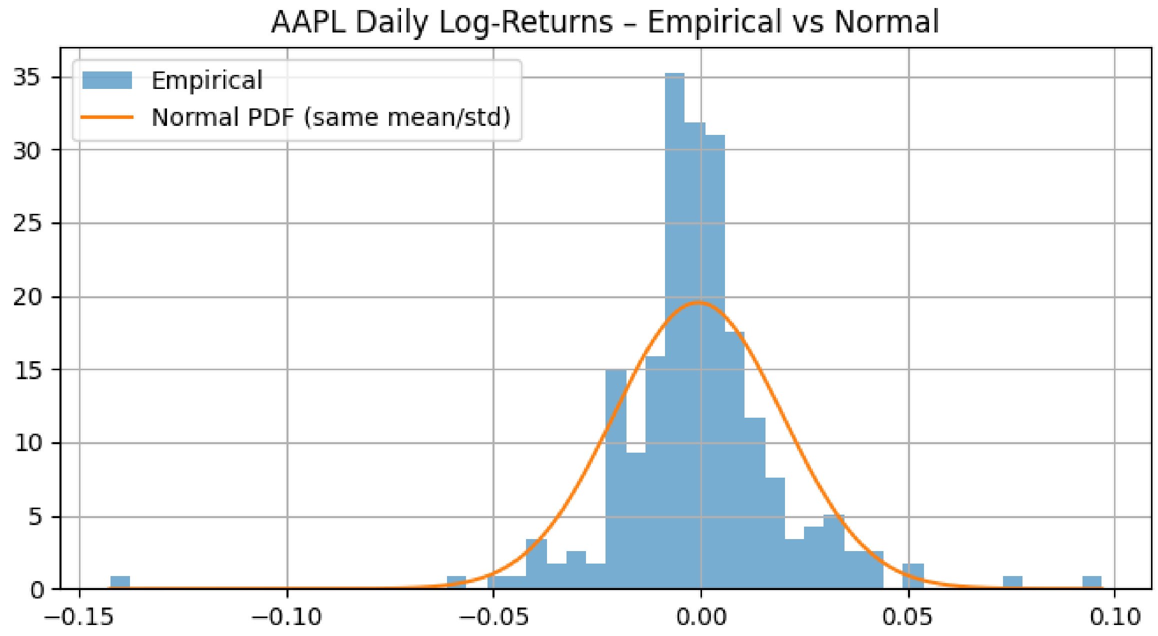
- Does the **mean level** seem roughly stable over time?
- Does the **variability** appear stable, or are there stretches of high and low volatility?
- Are there **obvious structural breaks**, such as crises or regime changes, where the series behaves very differently?

Distribution of Log-Returns

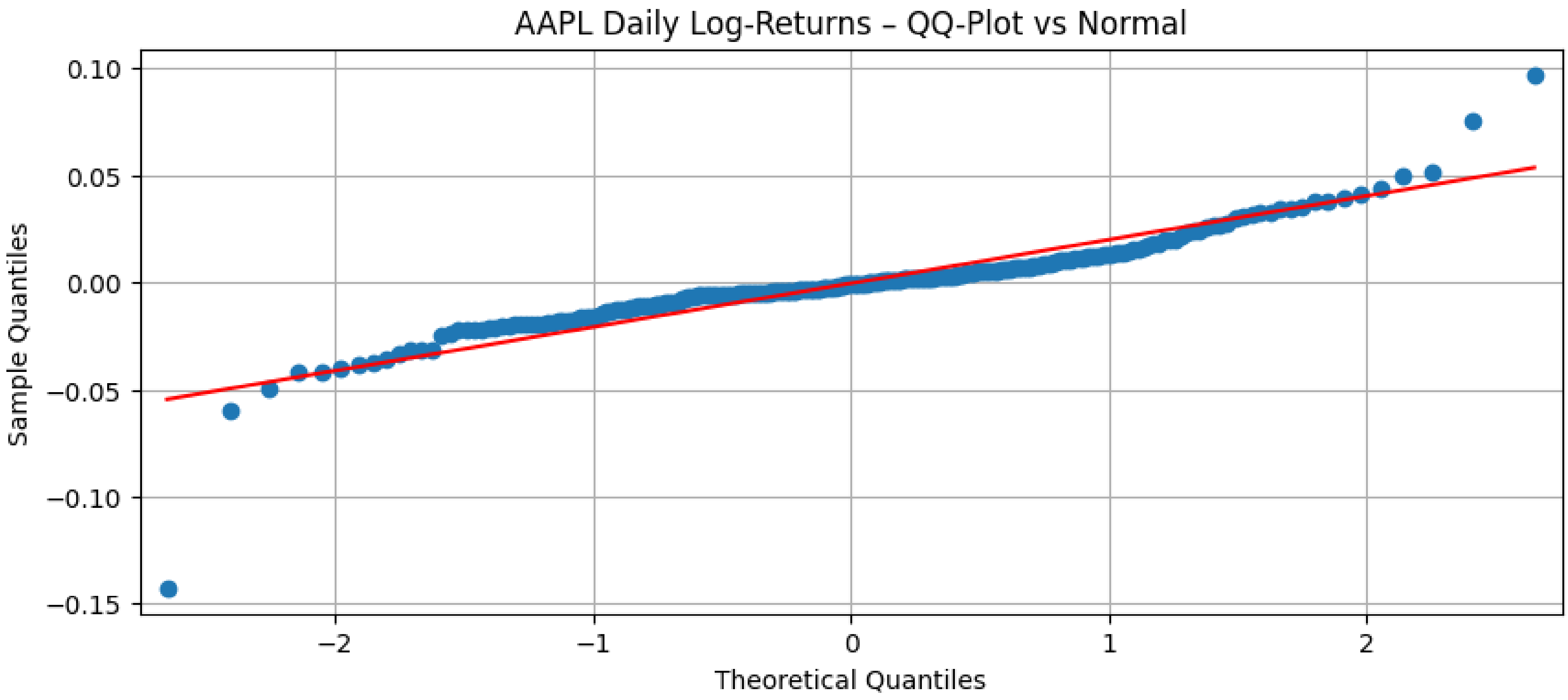
- Once we have a series of log-returns $\{r_t\}$,it is natural to examine its **empirical distribution**:
  - Construct a **histogram** or kernel density estimate of r_t.
  - Compare this empirical distribution to a **Normal distribution** with the same mean and variance.

In practice, financial return distributions often have the following features:

- **Fat tails (leptokurtosis):** Extreme positive or negative returns occur more often than a Normal model would predict.
- **Possible skewness:** The distribution might not be perfectly symmetric; losses can be more extreme or frequent than gains, or vice versa.



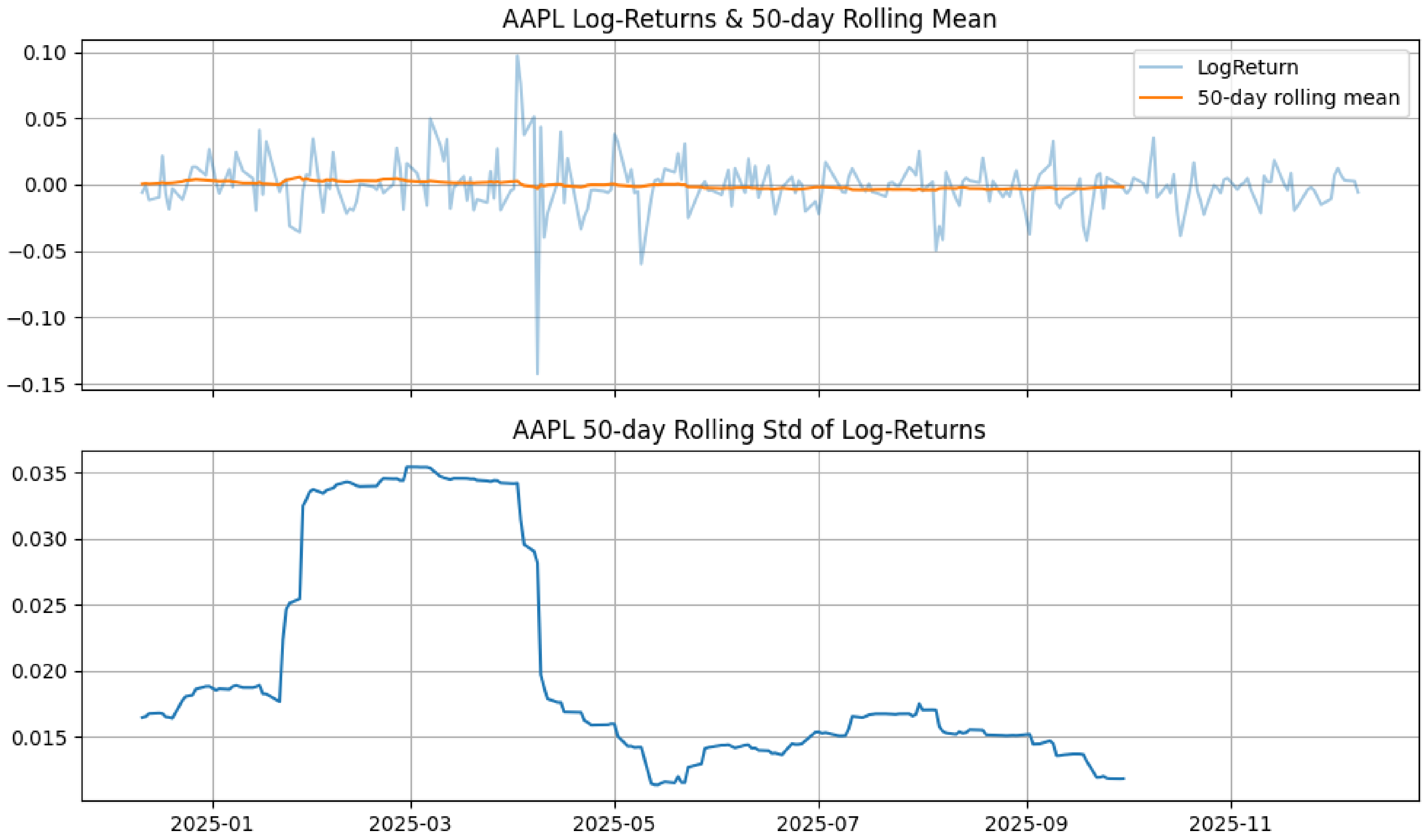
Distribution of Log-Returns



Volatility Clustering

- A hallmark of financial time series is **volatility clustering**: periods of high volatility tend to be followed by high volatility, and periods of low volatility by low volatility.
- If you look at a plot of daily returns, you might see:
  - Long stretches where returns are small and relatively calm.
  - Other stretches where returns are frequently large in magnitude, in both directions.
- Formally, this shows up as:
  - Weak or negligible autocorrelation in the returns themselves r_t.
  - Strong and persistent autocorrelation in **squared returns** r_t^2 or **absolute returns** $|r_t|$.

Volatility Clustering



AR(1) as a Simple Model

- A natural starting point for modelling temporal dependence is the **autoregressive model of order 1**, AR(1):
  $$X_t = \phi X_{t-1} + \eta_t,$$
  where:
  - ϕ is a parameter measuring the **persistence** of the series.
  - η_t is a white-noise innovation, typically assumed to have mean zero and constant variance.
- We can understand this model better by iterating it:
  - Substitute the expression for X_{t-1}:
    $$X_t = \phi(\phi X_{t-2} + \eta_{t-1}) + \eta_t = \phi^2 X_{t-2} + \phi\eta_{t-1} + \eta_t.$$
  - Continue this process; after k steps, we obtain:
    $$X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \eta_{t-j}.$$
- This shows that the current value X_t is a weighted sum of past shocks plus a term that depends on the value far in the past (X_{t-k}).

Stationarity Condition for AR(1)

- The AR(1) representation: $X_t = \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \eta_{t-j}$, is particularly informative as $k \to \infty$.
- If $|\phi| < 1$, then:
  - The term $\phi^k X_{t-k}$ tends to zero as $k \to \infty$, because ϕ^k decays geometrically.
  - We can then write: $X_t = \sum_{j=0}^{\infty} \phi^j \eta_{t-j}$.
- This expresses X_t as an **infinite moving average** (MA(∞)) of past innovations.
  - This kind of representation is central to the **Wold decomposition**, which states that any (zero-mean) covariance-stationary process can be written as an MA(∞) of white-noise shocks plus a deterministic component.
- Thus, the condition $|\phi| < 1$ is precisely the condition under which an AR(1) process is stationary and admits such an MA(∞) representation.

Variance of Stationary AR(1)

- For a stationary AR(1) process:
$$X_t = \phi X_{t-1} + \eta_t,$$
with $\mathbb{E}[\eta_t] = 0$ and $\mathrm{Var}(\eta_t) = \sigma_\eta^2$.
- In stationarity, the variance of X_t does not depend on t. Let $\mathrm{Var}(X_t) = \sigma_X^2$ .Then:
$$\sigma_X^2 = \mathrm{Var}(X_t) = \mathrm{Var}(\phi X_{t-1} + \eta_t).$$
- Assuming X_{t-1} and η_t are uncorrelated, we have:
$$\sigma_X^2 = \phi^2 \mathrm{Var}(X_{t-1}) + \mathrm{Var}(\eta_t) = \phi^2 \sigma_X^2 + \sigma_\eta^2.$$
- Rearranging:
$$\sigma_X^2 (1 - \phi^2) = \sigma_\eta^2 \;\Rightarrow\; \sigma_X^2 = \frac{\sigma_\eta^2}{1 - \phi^2}.$$
- This formula shows that as $|\phi|$ approaches 1, the variance of X_t becomes very large, reflecting the high persistence of shocks.

Autocovariance Function (ACVF)

- For a weakly stationary process $\{X_t\}$ with mean μ, the **autocovariance function (ACVF)** is defined as:

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) = \mathbb{E}[(X_t - \mu)(X_{t+h} - \mu)] \; \forall \; h = 0, \pm 1, \pm 2, \ldots$$

- Key points:
  - Because of stationarity, $\gamma(h)$ depends only on lag h, not on the specific time t.
  - $\gamma(0) = \mathrm{Var}(X_t)$ is the variance.
  - $\gamma(h)$ describes how much information about X_{t+h} is contained in X_t.

- The autocovariance function is fundamental for understanding the dependence structure and for specifying and analyzing linear time series models.

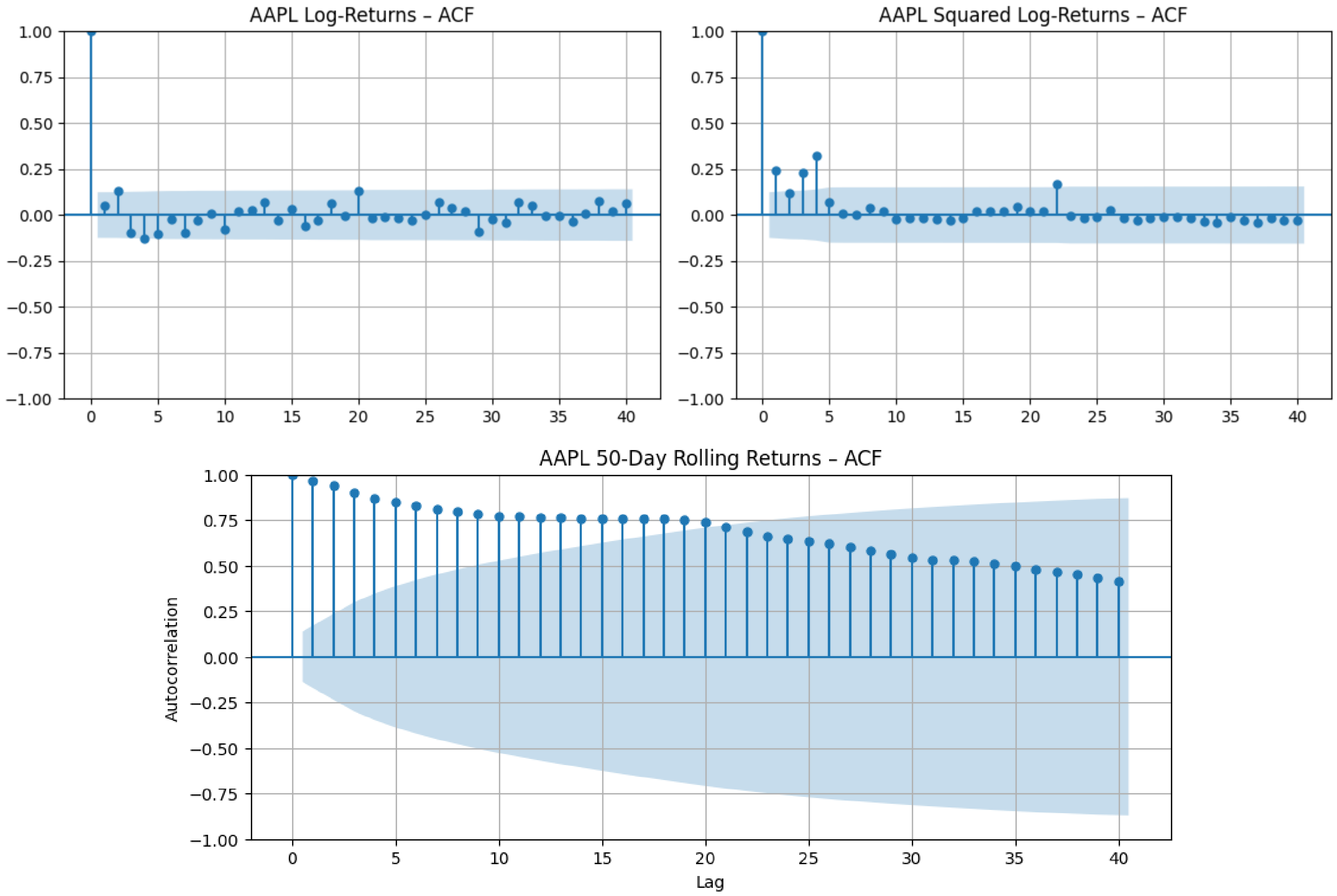
Autocorrelation Function (ACF)

- The **autocorrelation function** normalizes the autocovariance so that it lies in $[-1,1]$. It is defined as:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \qquad \forall h = 0, \pm 1, \pm 2, \dots$$

- Properties:
  - $\rho(0) = 1$)perfect correlation with itself).
  - $|\rho(h)| \leq 1$ for all h.
- For a weakly stationary process, $\rho(h)$ depends only on the lag h.
- In practice,
  - We work with the **sample ACF** $\hat{\rho}(h)$,computed from observed data and plot it as a function of h.
  - The pattern of significant spikes and decay in the ACF plot provides powerful hints about the underlying time series model.

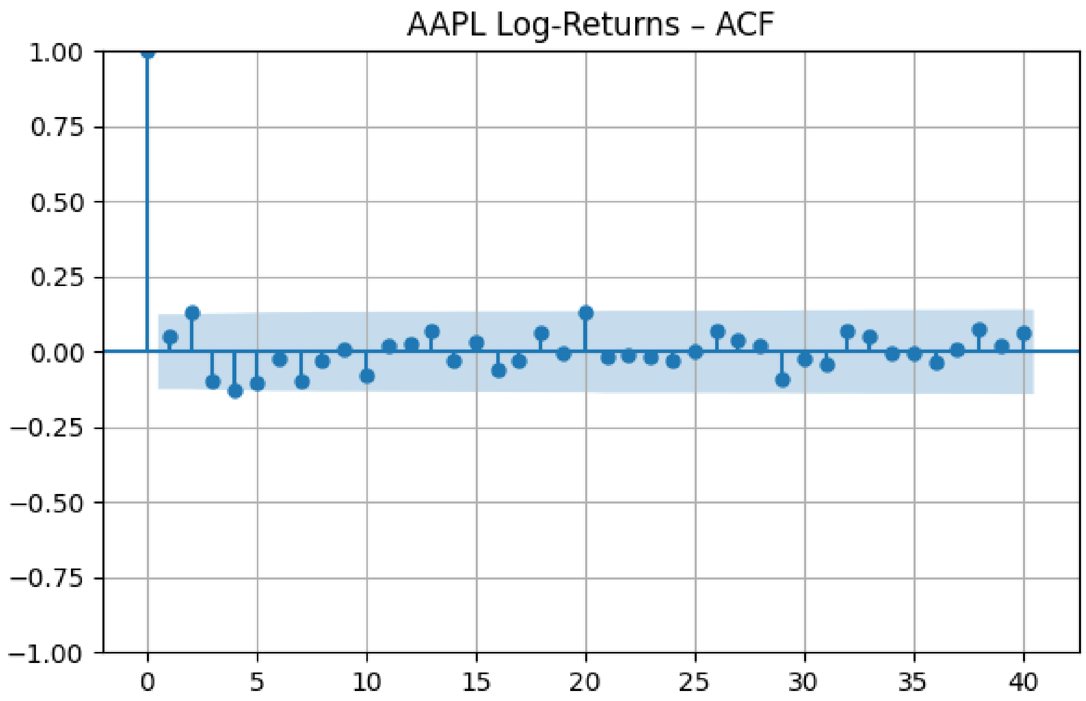
Autocorrelation Function (ACF)

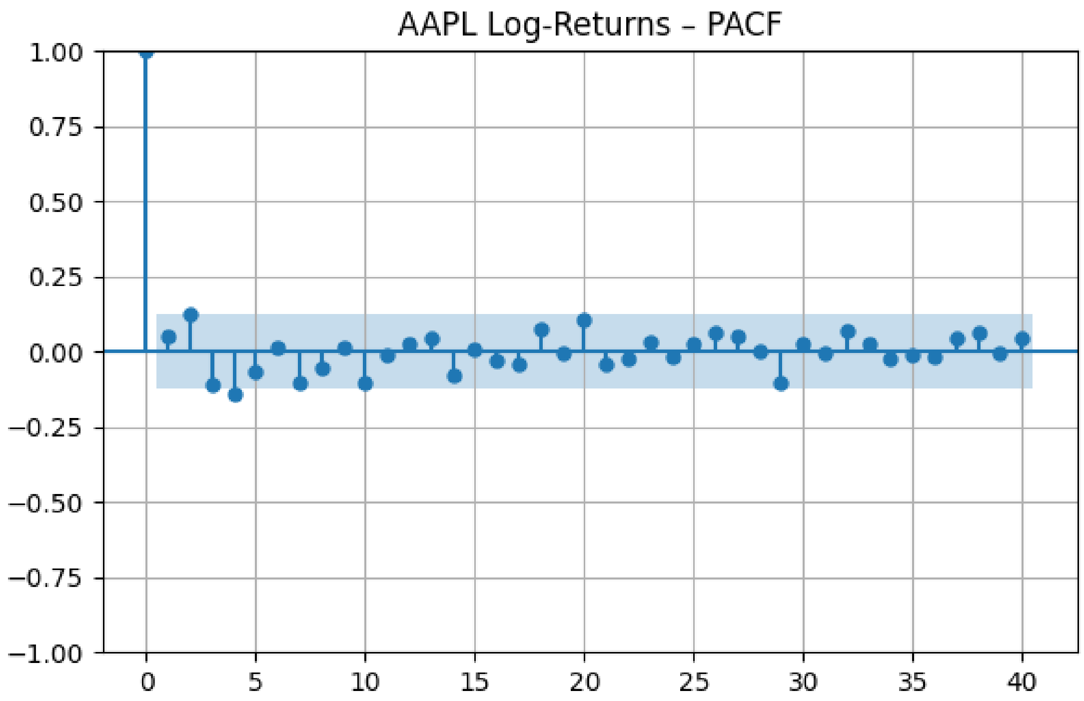


Partial Autocorrelation Function (PACF)

- The **partial autocorrelation function (PACF)** measures the correlation between X_t and X_{t+h} after removing the linear influence of the intermediate lags $X_{t+1}, \ldots, X_{t+h-1}$.
- Intuitively:
  - The ACF at lag h mixes **direct** and **indirect** effects. For example, X_t and X_{t+2} may be correlated partly because both correlate with X_{t+1}.
  - The PACF at lag h isolates the **direct** component of the relationship between X_t and X_{t+h}.
- In practice, the PACF at lag h can be obtained by:
  - Fitting a regression of X_t on $X_{t-1}, X_{t-2}, \ldots, X_{t-h}$ and taking the coefficient of X_{t-h}.
  - Or using standard algorithms implemented in statistical software.
- PACF plots are especially useful for identifying the order of autoregressive models.

Partial Autocorrelation Function (PACF)





ACF/PACF Heuristics

- The shapes of the ACF and PACF can be used as **heuristics** for choosing between AR, MA, and ARMA models:
- For an **AR(p)** process:
  - The **PACF** typically shows significant spikes up to lag p and then drops to near zero ("cuts off").
  - The **ACF** tends to decay gradually in magnitude as lag increases.
- For an **MA(q)** process:
  - The **ACF** typically shows significant spikes up to lag q and then cuts off.
  - The **PACF** tends to decay gradually.
- For an **ARMA(p, q)** process:
  - Both ACF and PACF usually show a more gradual decay rather than abrupt cutoff, making them harder to distinguish by eye.
- These are not hard rules, but they provide a starting point. In practice, you combine ACF/PACF patterns with information criteria and diagnostic checks to choose a sensible model.

Wold Decomposition (Intuition)

- The **Wold decomposition theorem** is a foundational result in time series analysis. It states that any zero-mean, covariance-stationary process $\{X_t\}$ can be represented as:

$$X_t = \sum_{j=0}^{\infty} \psi_j \, \eta_{t-j} + Z_t,$$

- where:
  - η_t is a **white-noise** process (uncorrelated innovations).
  - $\{\psi_j\}$ is a sequence of coefficients satisfying certain summability conditions.
  - Z_t is a deterministic component (for example, a perfectly predictable part); in many cases, we can assume $Z_t = 0$ or treat it separately.
- Interpretation:
  - Any stationary series can be decomposed into a (possibly infinite) **moving average** of past shocks plus a deterministic part.
  - This justifies the centrality of **linear** time series models and explains why ARMA models—finite-order approximations to this infinite MA—are natural building blocks.

Lag (Shift) Operator

- To write time series models compactly, we use the **lag operator** L, defined by:
  - $LX_t = X_{t-1}$
  - More generally, $L^k X_t = X_{t-k}$
- We can then define **polynomials in the lag operator**. For example, for autoregressive coefficients $\phi_1, \dots, \phi_p$,define:
$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$
- Using this notation, an ARMA model can be expressed compactly as:
$$\phi(L) X_t = \theta(L) \eta_t$$
- where $\theta(L)$ is a polynomial in L collecting the moving-average part. This notation makes it easier to discuss properties like **stationarity** and **invertibility** in terms of the roots of these polynomials.

AR(p) Model

- An **autoregressive model of order p**, AR(p), has the form:
$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \eta_t$$
where η_t is white noise.
- Using the lag operator:
$$\phi(L) X_t = \eta_t$$
with
$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$$
- Key idea:
  - X_t is regressed on its own past values.
  - The parameters $\phi_1, \ldots, \phi_p$ capture how past values influence the present.
- For stationarity, we require that the roots of the characteristic polynomial $\phi(z) = 0$ lie outside the unit circle in the complex plane. Intuitively, this ensures that shocks to the system die out over time rather than causing explosive behavior.

MA(q) Model

- A **moving average model of order q**, MA(q), has the form:

$$X_t = \eta_t + \theta_1 \eta_{t-1} + \theta_2 \eta_{t-2} + \cdots + \theta_q \eta_{t-q}$$

  where:
  - η_t is white noise.
  - $\theta_1, \ldots, \theta_q$ are parameters controlling how past shocks affect the current value.
- In lag-operator notation:

$$X_t = \theta(L)\eta_t,$$

  with

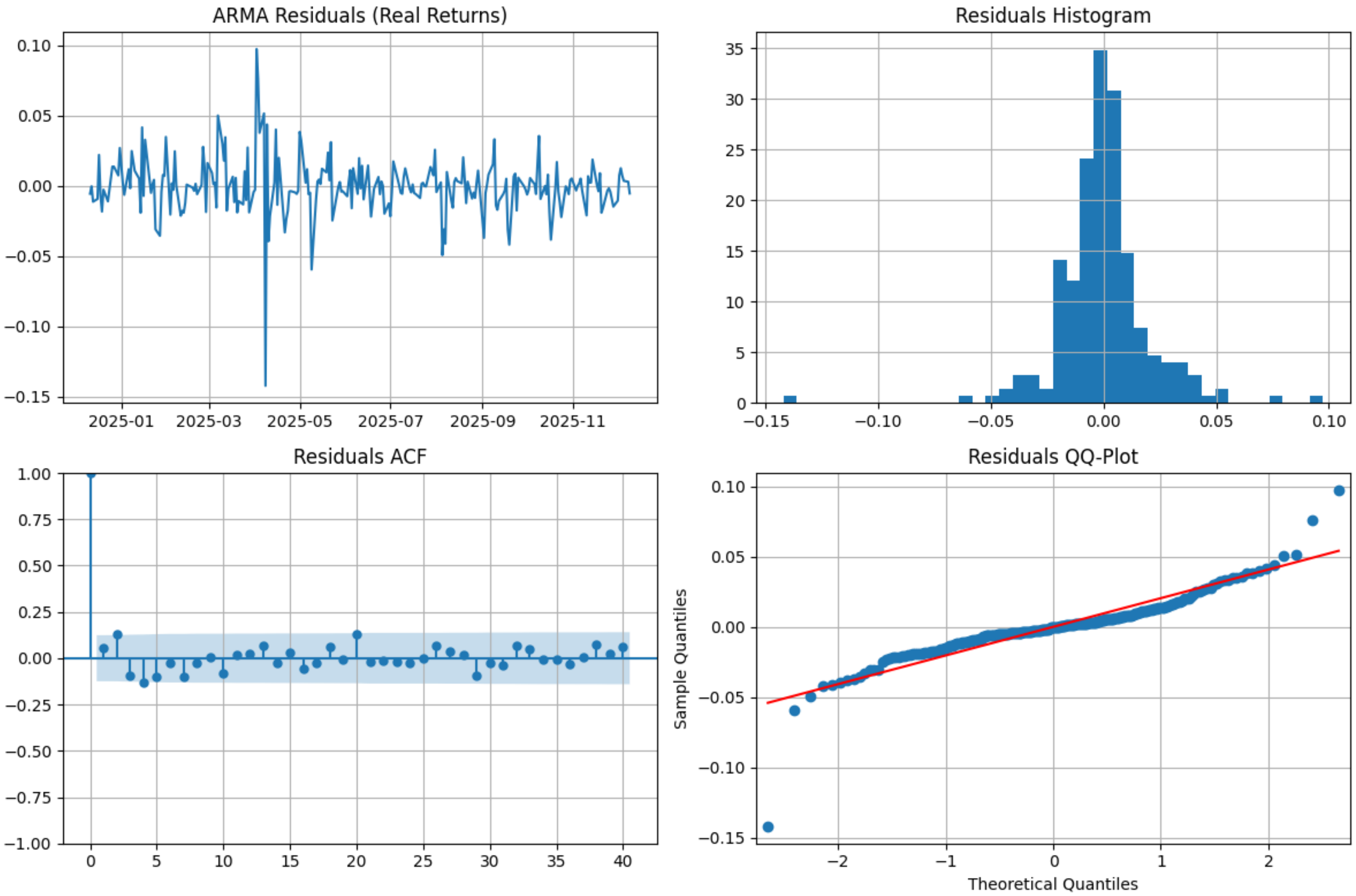
$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q.$$

- Unlike AR models, MA models express the series directly as a finite linear combination of current and past innovations.
- For **invertibility,** we require that the roots of $\theta(z) = 0$ lie outside the unit circle.

ARMA(p, q) Model

- An **ARMA(p, q)** model combines autoregressive and moving average components:
$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \eta_t + \theta_1 \eta_{t-1} + \cdots + \theta_q \eta_{t-q}.$$
- Equivalently, in lag-operator form:
$$\phi(L) X_t = \theta(L) \eta_t$$
  where:
  - $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$,
  - $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$.
- Interpretation:
  - The **AR part** describes how past values influence the current value.
  - The **MA part** describes how past shocks influence the current value.
- ARMA models are flexible yet relatively parsimonious, making them the standard workhorse for modelling **stationary mean dynamics** in time series.

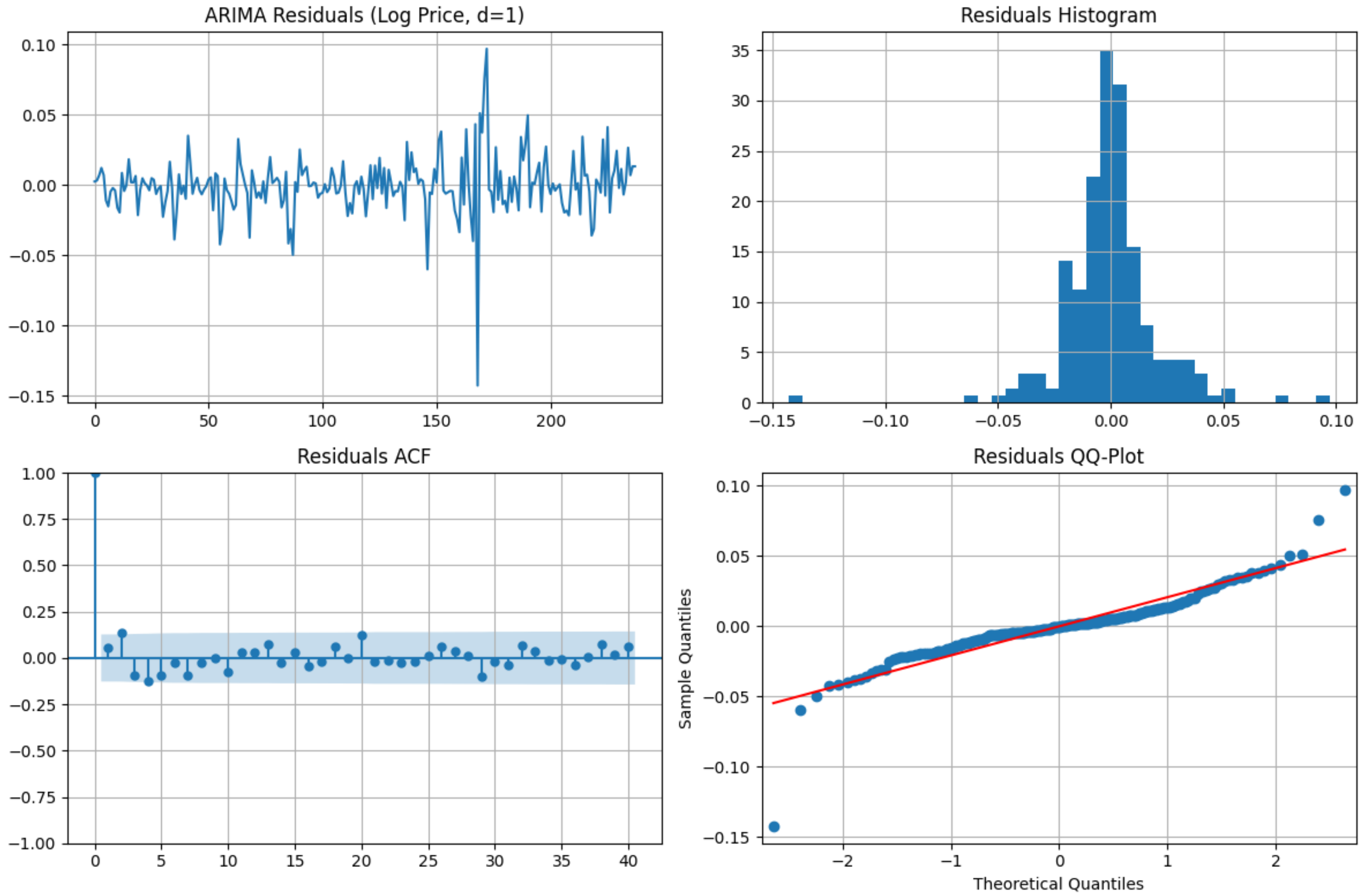
ARMA(p, q) Model



From ARMA to ARIMA

- ARMA models assume the underlying series is stationary. However, many real-world series, especially in macroeconomics and some financial contexts, are non-stationary due to trends or other persistent features. The **ARIMA(p, d, q)** model extends ARMA by allowing for **differencing** to remove non-stationarity:
  - Let Y_t be the original, possibly non-stationary series.
  - Define the first difference:
  $$\Delta Y_t = Y_t - Y_{t-1}$$
  - Higher differences are defined recursively: $\Delta^2 Y_t = \Delta(\Delta Y_t)$,etc.
- An ARIMA(p, d, q) model assumes that the **d-th difference** of Y_t ,namely $\Delta^d Y_t$, follows an ARMA(p, q) model.
- In many financial applications:
  - Raw prices $P_t$are non-stationary.
  - Log-returns or first differences $\Delta \log P_t$ are used to achieve stationarity, leading effectively to an ARIMA model with $d = 1$.

ARIMA Model



ARIMA Residuals (Log Price, d=1)
Residuals Histogram
Residuals ACF
Residuals QQ-Plot
Sample Quantiles
Theoretical Quantiles

Thank you!