Introduction to Modeling Time Series

Justin Offutt VP Quantitative Development

Fintech @ IU April 2025

What is a Stochastic Process?

- In probability theory, a stochastic process is defined over a probability space.

$$X_t : \Omega \to \mathbb{R}$$

  where $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space.
- A **stochastic process** is a collection of random variables indexed by a set T:

$$\{X_t\}_{t \in T}$$

- **Realization:** A single observed sequence from the process.

$$\text{Stochastic Process: } \{X_t\} \quad \rightsquigarrow \quad \text{Observed Data: } \{x_t\}$$

What is a Time Series?

- A time series is a stochastic process indexed by time:

$$\{X_t\}_{t=1}^{n}$$

- For example, if you record Apple's stock price every day, the actual prices are a realization of the stochastic process $\{A_t\}$.
- Time can be indexed:
  - **Discretely —** e.g., daily prices, where $t \in \{1, 2, \ldots\}$
  - **Continuously —** e.g., Brownian motion $\{W_t\}_{t \geq 0}$, where $t \in \mathbb{R}_+$
- Discrete example: Apple stock price over 4 days:

$$A_4 = \{198.78, 202.54, 196.23, 200.01\}$$

Stationarity vs Non-Stationarity

- **Strict Stationarity**: probability distribution doesn't change over time.
- **Weak Stationarity** (Second-order):
  - Constant mean: $\mathbb{E}[X_t] = \mu$
  - Constant variance: $\mathrm{Var}(X_t) = \sigma^2$
  - Constant autocovariance: $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$
- Non-stationarity: presence of trends, changing variance, or unit root.

White Noise

- A sequence of i.i.d. random variables.
- $\varepsilon_t \sim \mathrm{WN}(0, \sigma^2)$: mean 0, constant variance.
- No autocorrelation: $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t+h}) = 0$ for all $h \neq 0$.
- Often used as the error term in time series models to model "noise" (e.g low-volume uninformed retail trades).

Autoregressive Model (AR)

- $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$
- Captures momentum in data by apply regression on past values.
- Stationarity condition depends on roots of the characteristic equation (this is as simple as solving a quadratic).
- Partial autocorrelation function (aka PACF) cuts off after lag p.

ARIMA Model

- **ARIMA**(p, d, q): Autoregressive Integrated Moving Average.
- Useful for non-stationary series made stationary via differencing (up to some order of integration).
- Differencing operator: $\nabla X_t = X_t - X_{t-1}$
- Combines autoregression, differencing, and moving average components.

GARCH Model

- Generalized Autoregressive Conditional Heteroskedasticity.
- Captures volatility clustering: large changes followed by large changes.
- $\varepsilon_t = \sigma_t z_t, \quad z_t \sim \mathrm{WN}(0, 1)$
- $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$

Bottom Line

- Stochastic processes are built from probability spaces, a triplet whose members are $\{\Omega, \mathcal{F}, \mathbb{P}\}$: sample space (all possible worlds), event space ($\mathcal{P}(\Omega)$, or all things where probability is sensible), and probability measure respectively (a function that assigns a probability to any event $A \in \mathcal{P}$)
- Bottom line: time-series are a specific kind stochastic process whose index, $t \in T$, is ordered by time.
- All of these models are attempts of modeling underlying stochastic processes governing the behavior of realized time-series data.