Week 3: Time Series Foundations

From Stationarity to ARIMA

Module Assessments: What and Why

A quick orientation so you see how the two courseworks connect to the module.

Assessment	Weight	When	What we're assessing
CW1	30%	Week 6 (6 March)	Responsible Data Science presentation
CW2	70%	Week 13	Applied Data Science report + scaffold notebook

Brief for CW1 is released this week; use it to choose a FinTech application and plan your 10-minute presentation.

CW1: Responsible Data Science in FinTech (30%)

Why this design: We want you to evaluate *data quality and bias* in a real FinTech context, not just describe a business model.

- **Task:** Choose a FinTech application; present (10 min) on context, data quality issues (selection bias, validity, reliability), and responsible practice.

Part I: Why Time Series?

Financial Data Is Temporal

Financial observations arrive **ordered in time**:

- Today's return may depend on yesterday's
- Volatility clusters: calm follows calm, storms follow storms
- Shocks persist but eventually decay
- Structural breaks violate assumptions

The challenge: Classical statistics assumes independence. Time series methods handle temporal dependence.

What Can We Actually Predict?

Before learning time series methods, we must ask: **where is the signal?**

Financial prediction divides into three distinct problems:

Problem	Target	Signal (R^2)	Methods
The Mean	Future returns	~1-2%	ARIMA rarely beats naive
The Variance	Volatility	~15-40%	GARCH family
The Cross-Section	Which assets	~5-15%	Factors, ML

Part II: Stationarity and Why It Matters

What Is Stationarity?

A **weakly stationary** series has:

1. Constant mean: $\mathbb{E}[Y_t] = \mu$ for all t
2. Constant variance: $\text{Var}(Y_t) = \sigma^2$ for all t
3. Autocovariance depends only on lag: $\text{Cov}(Y_t, Y_{t-k}) = \gamma_k$

Why it matters: If the process is changing over time, historical relationships may not hold going forward.

Example: SPY Prices (Non-Stationary)

Part III: Autocorrelation and PACF

The Partial Autocorrelation Function

PACF measures correlation at lag k **after controlling for intermediate lags** $(Y_{t-1}, \dots, Y_{t-k+1})$ (Brooks 2019 Ch 5). Equivalently, the *added contribution* of Y_{t-k} over an AR($k-1$) model (Tsay 2010 Ch 2).

- ACF: total correlation (includes indirect effects)
- PACF: unique contribution of lag k; for AR(p), sample PACF **cuts off at lag** p

Why it matters: Helps identify AR vs MA processes.

ACF vs PACF Interpretation

Process	ACF Pattern	PACF Pattern
AR(p)	Geometrically decaying	Cuts off after lag p
MA(q)	Cuts off after lag q	Geometrically decaying
ARMA(p,q)	Geometrically decaying	Geometrically decaying

For AR(p), PACF has p non-zero points then cuts off; for MA(q), ACF has q non-zero points then cuts off (Brooks 2019 Ch 5).

Part IV: AR, MA, and ARIMA Models

Autoregressive (AR) Models

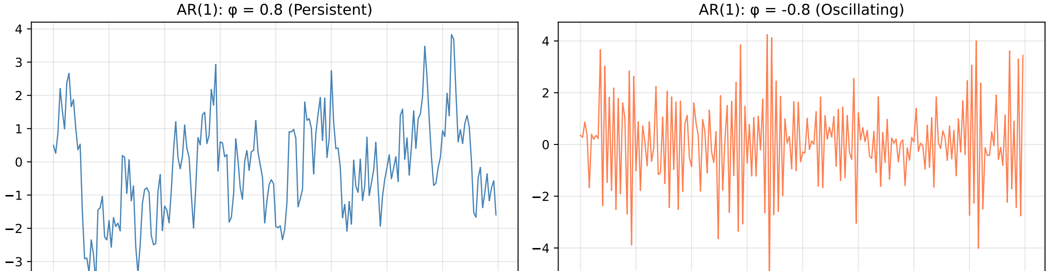
An **AR(p)** model uses past values to predict current value:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

Interpretation: For AR(1), ACF satisfies $\rho_k = \phi_1^k$: exponential decay (or alternating for $\phi_1 < 0$) (Tsay 2010 Ch 2). Stationarity requires characteristic roots < 1 in modulus.

- $\phi_1 = 0$: white noise; $\phi_1 = 1$: random walk; $\phi_1 < 0$: oscillating

Simulating an AR(1)



Part V: Forecasting and Validation

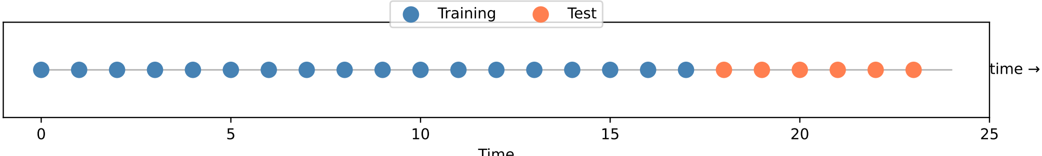
Simple Forecasting Methods (Baselines)

Before fitting complex models, establish baselines:

Method	Formula	When It Works
Naive	$\hat{y}_{t+h} = y_t$	Random walk, efficient markets
Mean	$\hat{y}_{t+h} = \bar{y}$	Mean-reverting, no trend
Drift	$\hat{y}_{t+h} = y_t + \frac{h}{T-1}(y_T - y_1)$	Trending series

Train/Test Splits for Time Series

Critical rule: Never shuffle time series data.



Part VI: Prediction Uncertainty

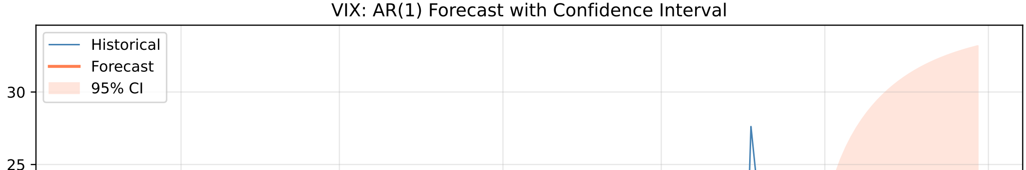
Confidence vs Prediction Intervals

Two types of uncertainty:

Type	What It Captures	Formula
Confidence Interval	Uncertainty about *mean*	$\hat{\mu} \pm t_{\alpha/2} \cdot SE(\hat{\mu})$
Prediction Interval	Uncertainty about *next value*	$\hat{y} \pm t_{\alpha/2} \cdot \sqrt{SE^2 + \sigma^2}$

Prediction intervals are **wider** because they include residual variance.

ARIMA Forecast with Intervals



Part VII: Beyond Stationarity

The Limitation of Classical Methods

Classical time series (ARIMA) assumes we can transform data to stationarity.

But what if:

- Structural breaks occur (COVID crash, policy changes)?
- Regimes shift over time?
- Non-linear dependencies exist?

Classical methods struggle when stationarity is fundamentally violated.

Sequence Learning: Beyond Stationarity

Classical Concept	Sequence Learning Extension	Key Advance
AR(p) process	Recurrent Neural Networks (RNN, LSTM, GRU)	Non-linear dependencies
Differencing for stationarity	Direct modelling of non-stationary sequences	No transformation required
ARIMA forecasts	Transformer architectures	Attention over

Part VIII: Key Takeaways

Summary: Five Core Concepts

1. **Stationarity** enables prediction (constant mean, variance, ACF)
2. **ACF/PACF** identify AR vs MA patterns
3. **ARIMA** combines differencing, AR, and MA
4. **Validation** requires time-aware splits (no shuffling)
5. **Sequence learning** extends beyond stationarity (preview)

Looking Ahead: Week 4 Volatility

Last week we modelled the *mean*. Next week we model *variance*:

- Returns are unpredictable → ACF near zero
- **But volatility is predictable** → ACF of squared returns shows persistence
- ARCH/GARCH models capture volatility clustering
- Connection: GARCH is to variance what ARIMA is to the mean

Directed Learning

- **Core**: Read Tsay (2010) Chapter 2 (linear time series) and Brooks (2019) Chapter 5 (univariate time series, Box–Jenkins); complete lab with Bloomberg data; experiment with ARIMA orders