Time-series,
Spring, 2026

ARIMA, Box–Jenkins process, model selection, forecasting

Faculty of DS & AI
Spring semester, 2026

Trong-Nghia Nguyen

Business AI Lab

Content

- **ARIMA**
- The Box-Jenkins Methodology
- Technique Controls & Diagnostics for ARIMA
- Forecasting & evaluation

ARIMA

Anatomy of the ARIMA (p, d, q) Model

ARIMA is an extension of the ARMA model specifically designed for forecasting non-stationary time series data. It applies an ARMA(p, q) process to a series that has been made stationary through differencing.

AR (p)		I (d)		MA (q)
Autoregressive	+	Integrated	+	Moving Average
Uses past values of the series itself to forecast the present.		Represents the number of differencing steps required to transform a non-stationary series into a stationary one.		Uses past forecast errors (shocks) to forecast the present.

ARIMA

The Prerequisite: Defining Weak Stationarity

To apply ARMA methodologies, a time series must exhibit weak stationarity, meaning its statistical properties remain constant over time.

Mean

The expected value is constant.

$$E[X_t] = \mu$$

Variance

The variance is constant.

$$\mathrm{Var}(X_t) = \sigma^2$$

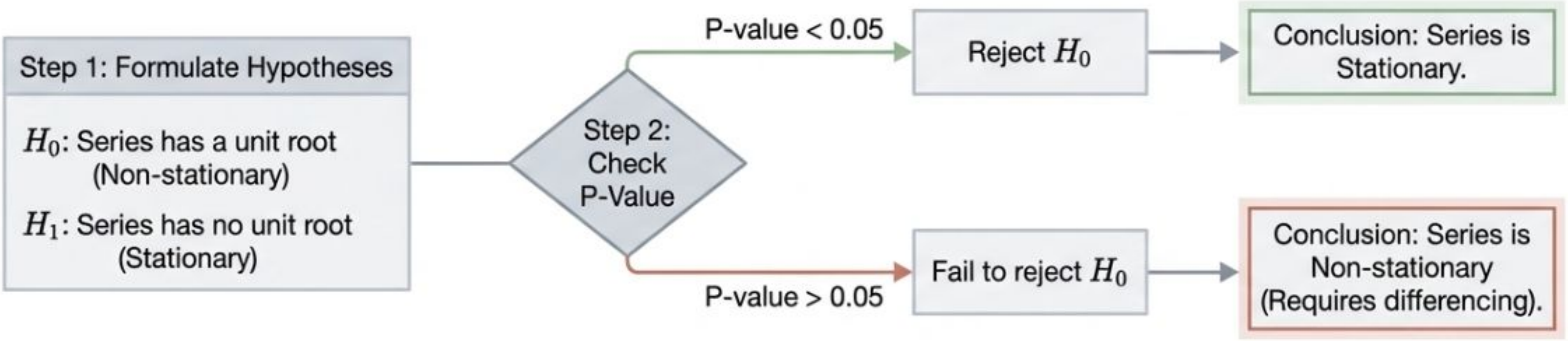
Autocovariance

Depends solely on the distance (lag) k, not on time t.

$$\mathrm{Cov}(X_t, X_{t-k})$$

Diagnosing Stationarity via the Augmented Dickey-Fuller (ADF) Test

The ADF test checks for the presence of a Unit Root, which indicates non-stationarity.



ARIMA

The "I" Component: Integration & Differencing (d)

Differencing removes stochastic trends to achieve stationarity. The parameter d represents the minimum number of differencing steps required.

1st Order Differencing

$$\nabla X_t = X_t - X_{t-1}$$

Use: Removes standard stochastic trend.

2nd Order Differencing

$$\nabla^2 X_t = \nabla(\nabla X_t) = X_t - 2X_{t-1} + X_{t-2}$$

Use: Applied only if 1st order fails to achieve stationarity.

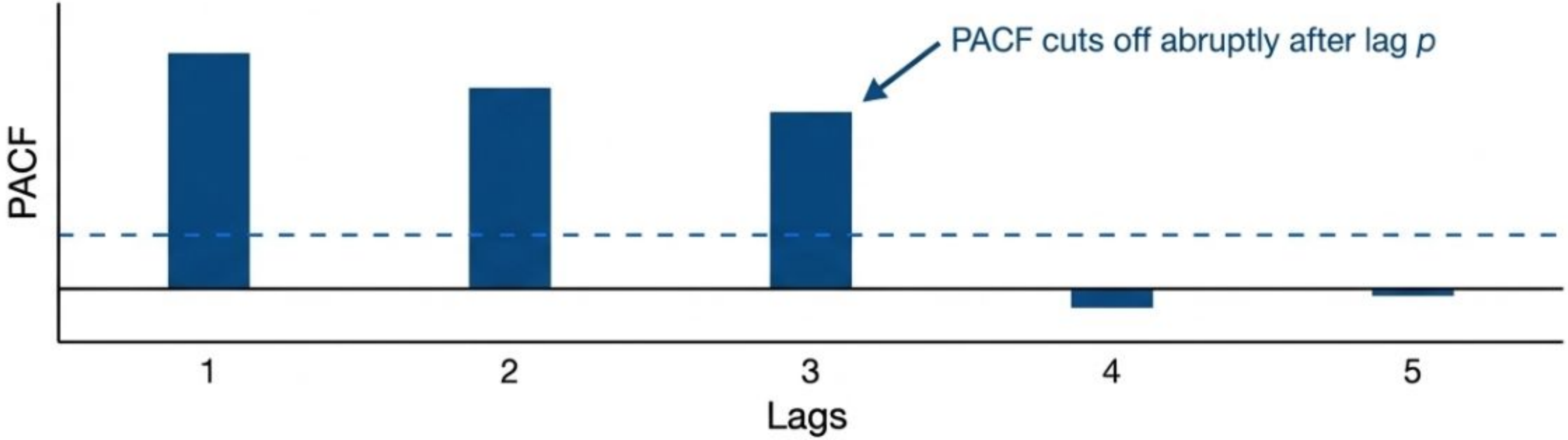
Selection Rule: Determine d using the ADF test.
In practice, d is typically 0, 1, or rarely 2.

ARIMA

The "AR" Component: Autoregressive (p)

Applied to the stationary series W_t, the Autoregressive process dictates that the current value is explained by p past values.

$$W_t = c + \phi_1 W_{t-1} + \cdots + \phi_p W_{t-p} + \varepsilon_t$$

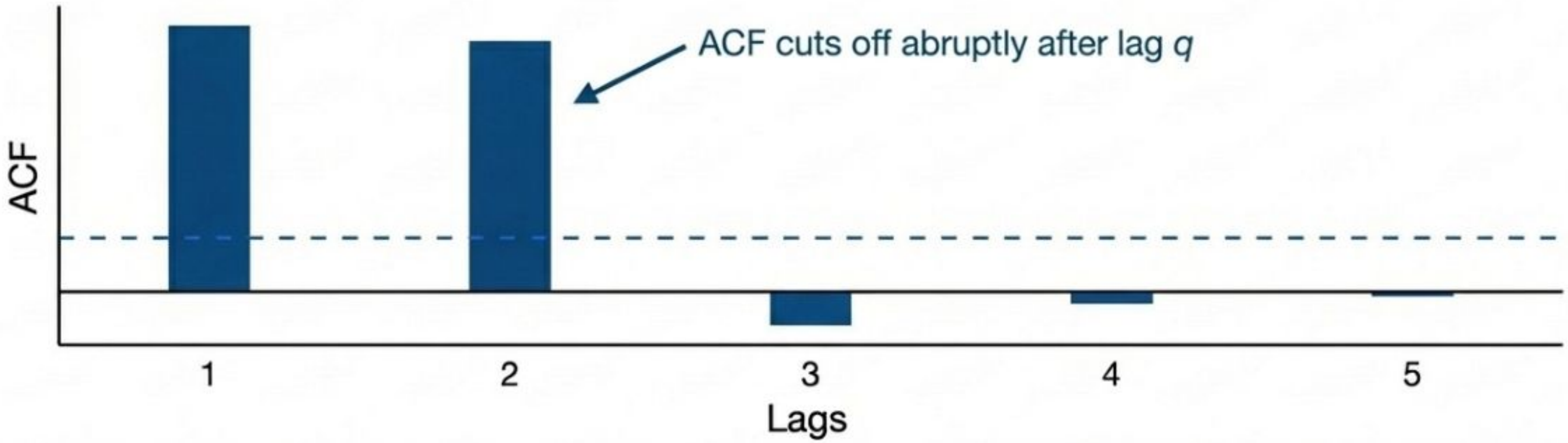


ARIMA

The "MA" Component: Moving Average (q)

The Moving Average process dictates that the current value of the stationary series W_t depends on the current shock (error) and q past shocks.

$$W_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$



ARIMA

Mathematical Formalization: The Backshift Operator (B)

To express ARIMA cleanly in mathematical and research contexts, we utilize the Backshift operator **B**, which shifts the time index back by one unit.

Basic Definitions

$$BX_t = X_{t-1}$$

$$B^k X_t = X_{t-k}$$

Applying to Differencing

Demonstrating how the d component is translated. d-th order differencing is written as:

$$\nabla^d = (1 - B)^d$$

ARIMA

The General ARIMA (p, d, q) Equation

The AR Polynomial:
$(1 - \phi_1 B - \cdots - \phi_p B^p)$

The MA Polynomial:
$(1 + \theta_1 B + \cdots + \theta_q B^q)$

$$\phi(B)(1 - B)^d X_t = c + \theta(B)\,\varepsilon_t$$

The Integrated
(Differencing) Process

The raw time series and the
error term, respectively.

ARIMA

Practical Configurations of ARIMA

Model Configuration	Interpretation
ARIMA(1, 1, 0)	First-order differencing of a series that follows an AR(1) process.
ARIMA(0, 1, 1)	First-order differencing of a series that follows an MA(1) process.
ARIMA(1, 1, 1)	First-order differencing of a series that exhibits both AR(1) and MA(1) structures.

ARIMA

Parameter Selection Matrix

A synthesized reference for identifying the required parameters for building an ARIMA(p, d, q) model.

Integrated (d)

- Determines Stationarity.
- Identified via the ADF Test.
- (Minimize d to achieve P-value < 0.05).

Autoregressive (p)

- Determines reliance on past values.
- Identified via the PACF.
- (Look for the lag where the plot cuts off).

Moving Average (q)

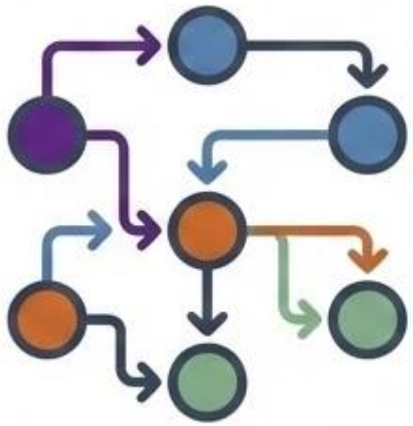
- Determines reliance on past shocks.
- Identified via the ACF.
- (Look for the lag where the plot cuts off).

Content

- ARIMA
- **The Box-Jenkins Methodology**
- Technique Controls & Diagnostics for ARIMA
- Forecasting & evaluation

The Box-Jenkins Methodology

1. Definition



The **Box-Jenkins** process is a systematic and iterative theoretical framework used to build, estimate, and test ARIMA models. Introduced by George Box and Gwilym Jenkins in 1970, this process establishes a scientific approach to finding the best fit for historical data through specific statistical evidence.

2. Core Purpose

Maximum extraction of data structure: Ensuring that the model has explained the entire pattern of the time series, transforming all predictable structures into parameters and leaving only 'White Noise' as the residual.

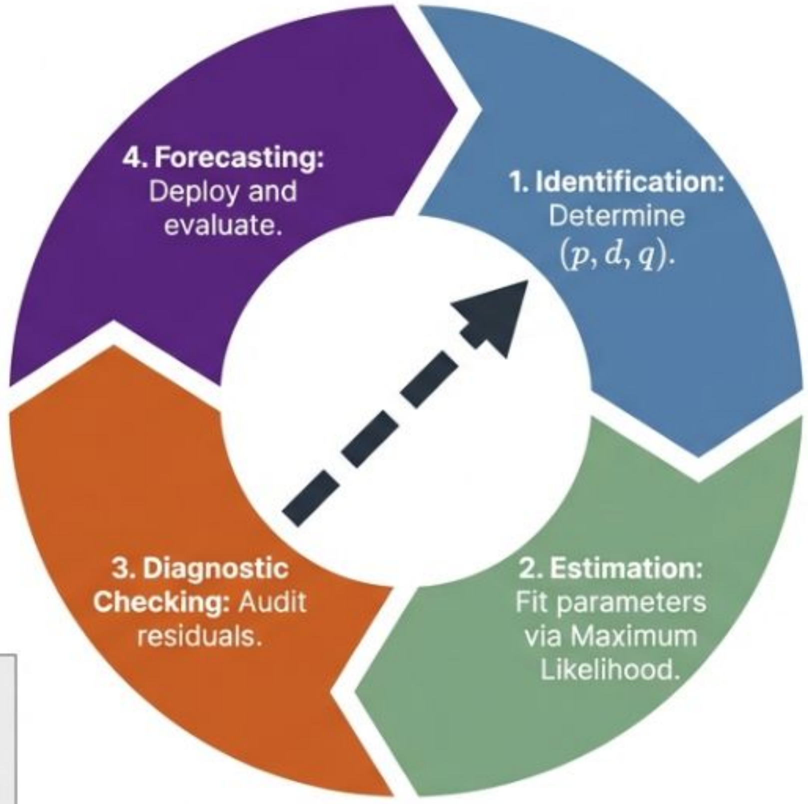
Parsimony: The goal is to find the simplest yet most efficient model (often based on minimizing the BIC index), avoiding overfitting and increasing generalizability.

Evidence-based validation: Instead of selecting parameters based on intuition, the process requires the use of objective statistical tools such as ACF/PACF plots and ADF unit root tests to determine the p, d, and q orders.

Reliability validation: The 'Diagnostic testing' phase acts as a final audit to ensure the model is structurally sound before being used for real-world decisions.

The Box-Jenkins Methodology

The 4-Stage Iterative Loop

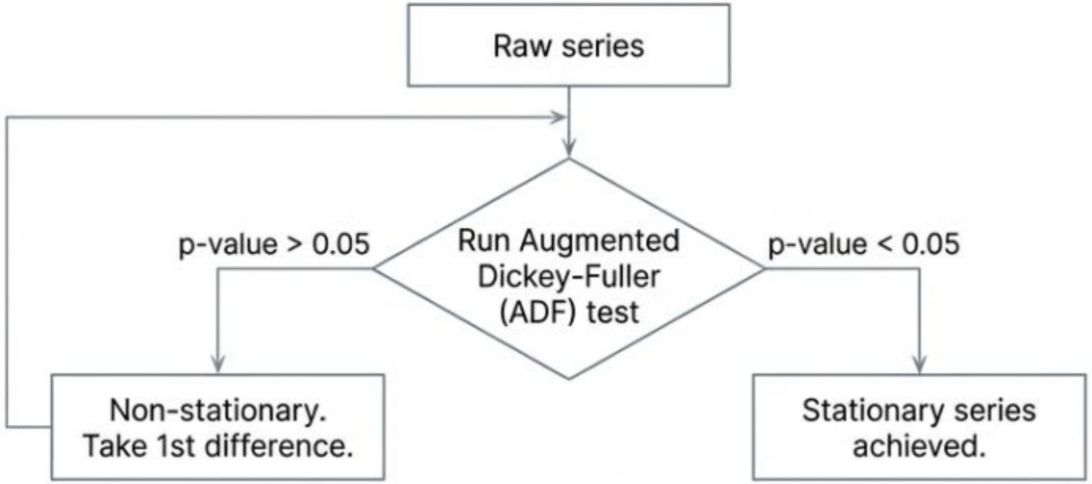


Introduced by Box & Jenkins (1970), this framework is a closed loop. If diagnostic checks fail at Stage 3, the modeler must return to Stage 1 to test a new configuration.

The Box-Jenkins Methodology

Stage 1: The Prerequisite of Stationarity

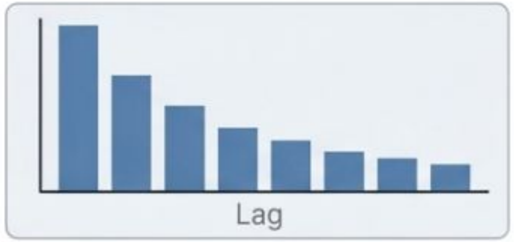
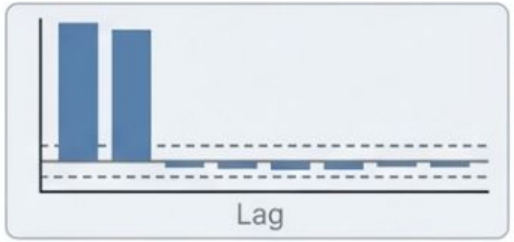
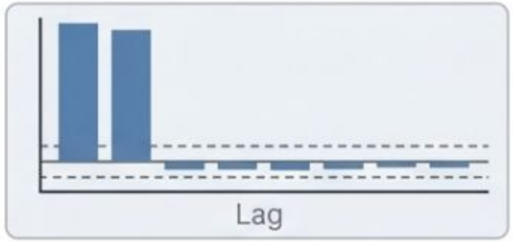
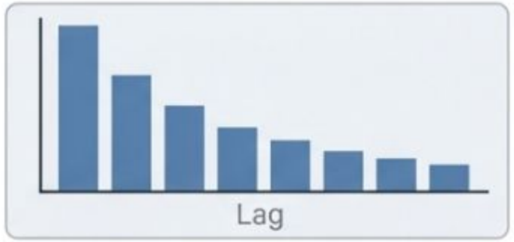
$$\overbrace{W_t}^{\text{Stationary Series}} = \overbrace{\nabla^d}^{\text{Differencing of order d}} \underbrace{X_t}_{\text{Raw Data}}$$



Rule of Thumb for d: Set d to the minimum number of differencing steps required to achieve $p < 0.05$.

The Box-Jenkins Methodology

Stage 1: Reading ACF & PACF Plots

	ACF Plot	PACF Plot
AR(p) Model	 Tailing off (exponential decay)	 Sharp cut-off after lag p
MA(q) Model	 Sharp cut-off after lag q	 Tailing off (exponential decay)

What if both tail off? Indicates a mixed ARMA(p,d,q) model.
Use small values for p and q (e.g., 1, 1, 1) without a clear single cut-off.

The Box-Jenkins Methodology

Stage 1: The Candidate Matrix

Do not select a single model yet. The output of the Identification stage is a shortlist.

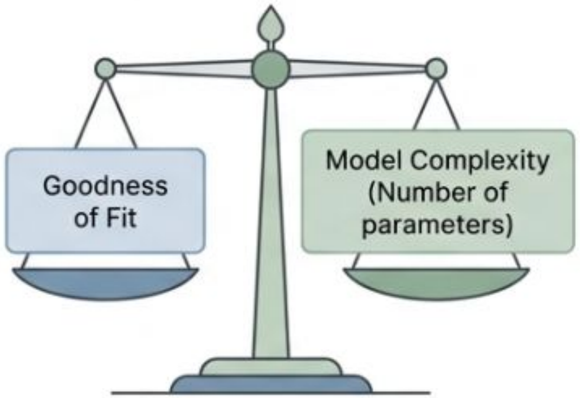
Candidate Shortlist
Candidate A: ARIMA (1, 1, 0)
Candidate B: ARIMA (0, 1, 1)
Candidate C: ARIMA (1, 1, 1)

The Box-Jenkins Methodology

Stage 2: Parameter Estimation

The Box-Jenkins Methodology

Stage 2: Model Selection



AIC (Akaike)

Balances fit and complexity. More flexible.

BIC (Bayesian)

Imposes a heavier penalty for extra parameters. Strictly prefers simplicity (parsimony).

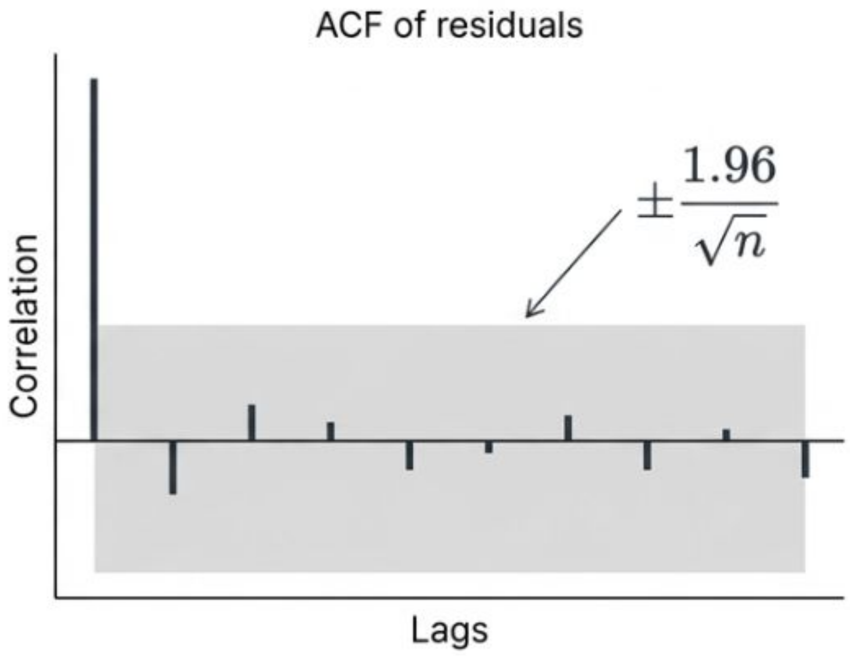
The Golden Rule: Select the candidate model with the lowest BIC (or lowest AIC for increased flexibility).

The Box-Jenkins Methodology

Stage 3: The Pursuit of White Noise

A valid model has extracted all temporal structure. The leftovers (residuals) must be White Noise.

- [ ] Mean equals zero ($\mu = 0$)
- [ ] Constant variance ($\sigma^2 = \text{const}$)
- [ ] Zero correlation across time



The Box-Jenkins Methodology

Stage 3: The Ljung-Box Test

Null Hypothesis (H_0): Residuals are white noise (no autocorrelation at lag k).

Expectation: We want a large p-value (> 0.05).

↓

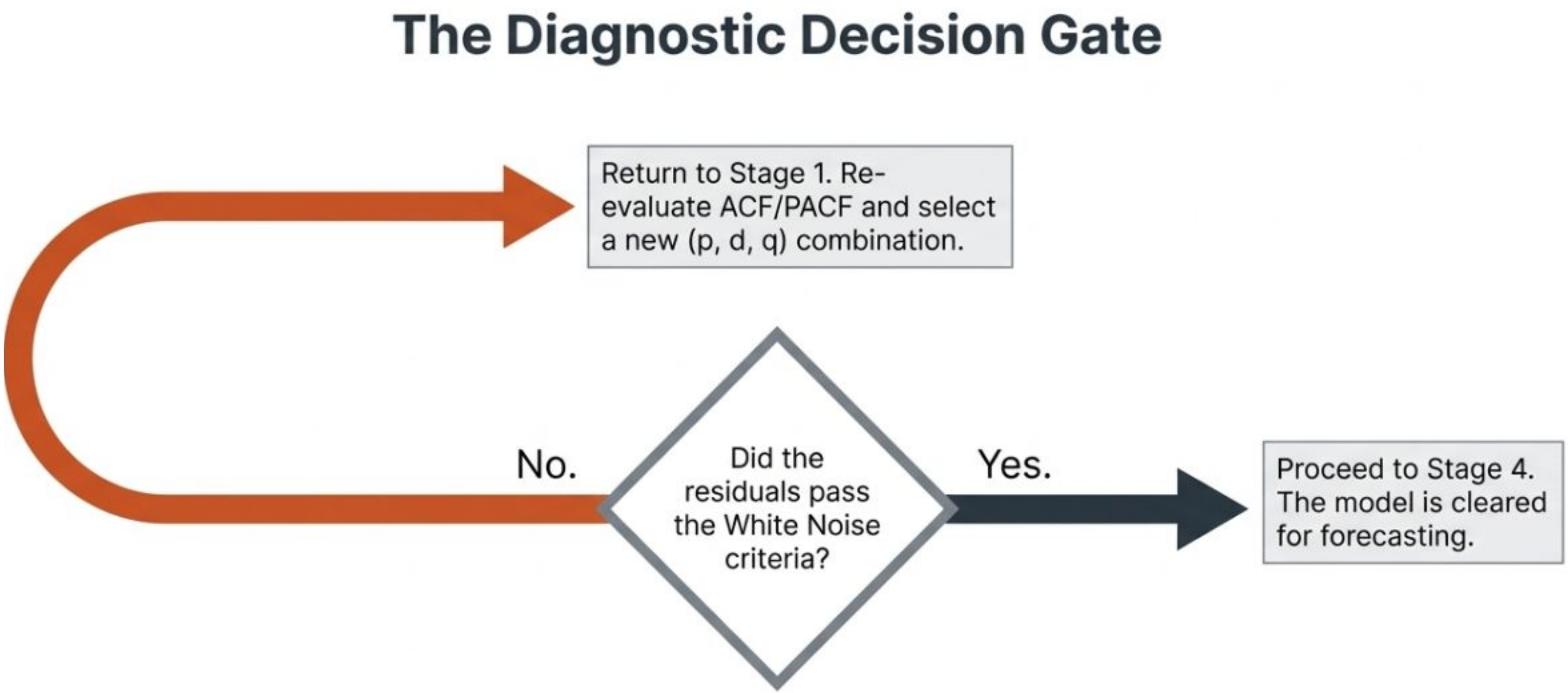
Result: Fail to reject H_0.

↓

Conclusion: Model is structurally sound.

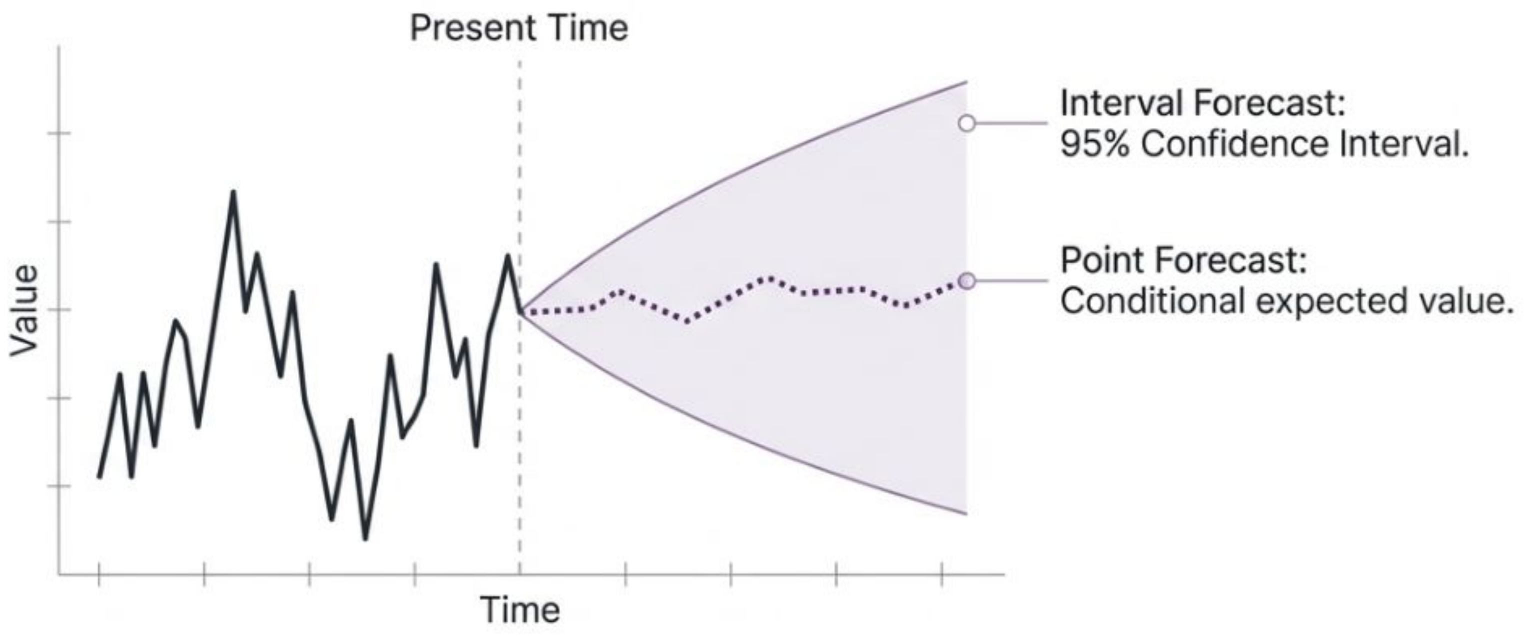
Warning: If ACF spikes exceed confidence bounds OR Ljung-Box p-value is very small, the model is inadequate.

The Box-Jenkins Methodology



The Box-Jenkins Methodology

Stage 4: Projecting the Future



Notice the expanding cone. The confidence interval widens as the forecast horizon (h) increases, reflecting the cumulative accumulation of error over time.

The Box-Jenkins Methodology

Stage 4: Evaluation Metrics

Compare forecasted values against actual values on a hold-out test set.

MAE (Mean Absolute Error)

Measures average magnitude of errors without considering direction. Highly interpretable.

RMSE (Root Mean Squared Error)

Squares the errors before averaging. Heavily penalizes large forecasting errors.

MAPE (Mean Absolute Percentage Error)

Expresses error as a percentage. Essential for comparing accuracy across series with entirely different scales.

The Box-Jenkins Methodology

The Box-Jenkins Master Summary

1. Identify	2. Estimate	3. Diagnose	4. Forecast
d (ADF test $< < 0.05$), p, q (ACF/PACF cut-offs). Output: Candidate shortlist.	Maximum Likelihood. Output: Parameters & IC scores. Rule: Lowest BIC.	Check residuals. Rule: Ljung-Box $p > 0.05$ (White Noise). Fail = Loop to 1.	Point & Interval forecasts. Evaluate via MAE, RMSE, MAPE.

Content

- ARIMA
- The Box-Jenkins Methodology
- **Technique Controls & Diagnostics for ARIMA**
- Forecasting & evaluation

Technique Controls & Diagnostics for ARIMA

A systematic pipeline filters candidate models into a single optimal choice

Finding the best fit using Maximum Likelihood.

Balancing fit against complexity via Information Criteria.

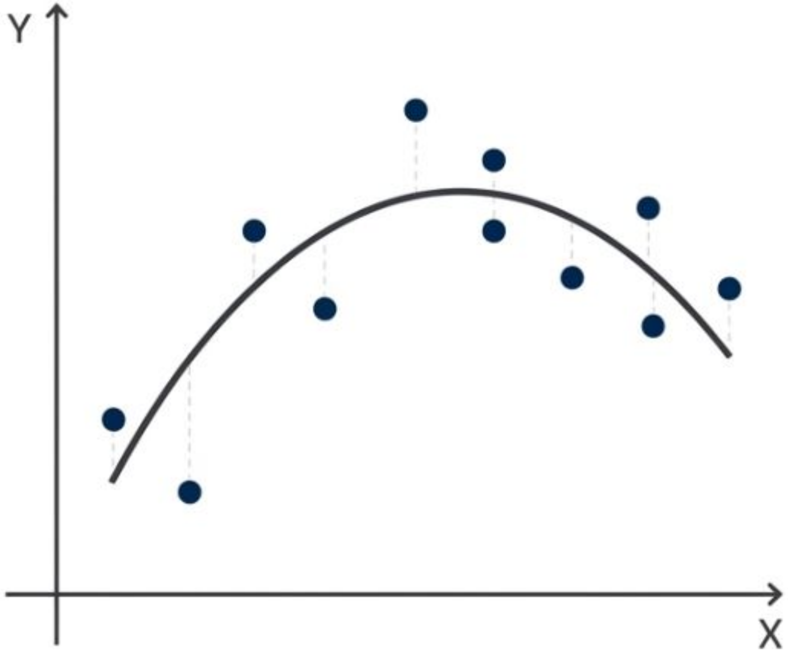
Ensuring no information is left behind using White Noise Tests.

Technique Controls & Diagnostics for ARIMA

Maximum Likelihood estimation determines the optimal parameters

$$L$$

- **Core Concept:** ARIMA models use the **Maximum Likelihood (ML)** method to estimate the AR and MA parameters.
- **The Likelihood Function (L):** Measures the exact probability that our observed data was generated by the chosen set of parameters.

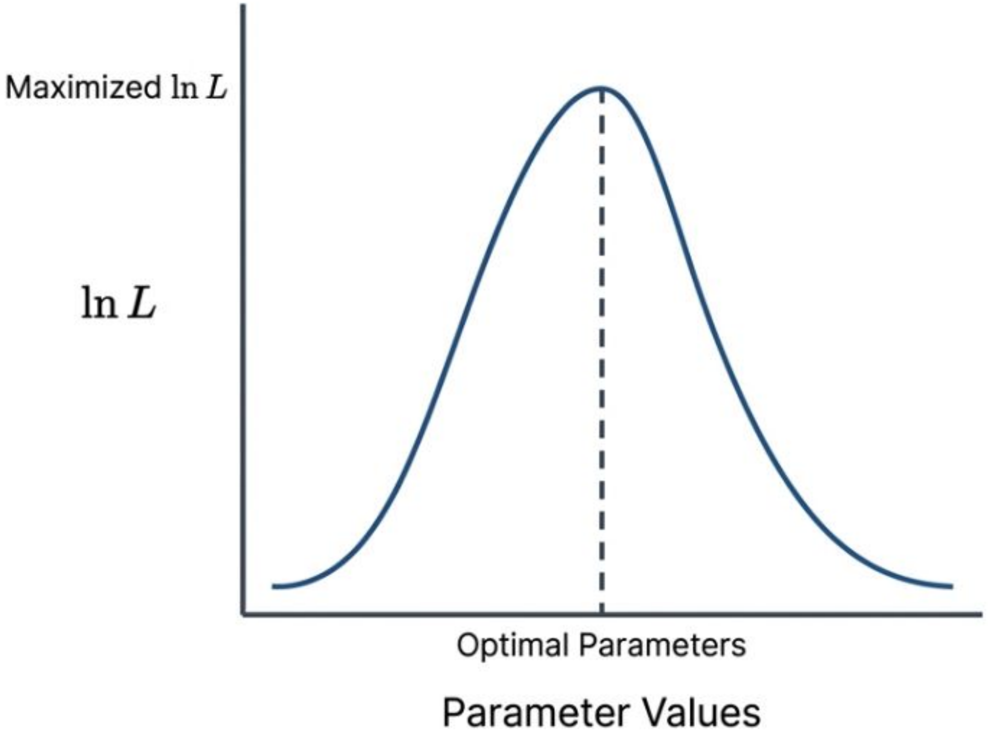


Technique Controls & Diagnostics for ARIMA

Maximizing the Log-Likelihood ensures the tightest mathematical fit

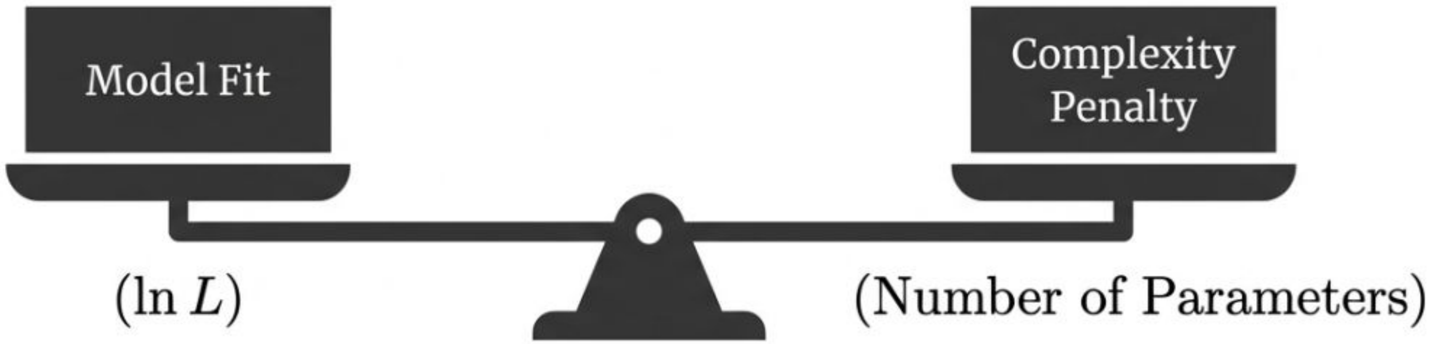
The Objective: Find the parameter set that strictly maximizes the Log-likelihood ($\ln L$).

The Rule: A higher Log-likelihood value mathematically proves the model is a closer fit to the observed data.



Technique Controls & Diagnostics for ARIMA

Model selection requires balancing accuracy with structural parsimony



When comparing multiple candidate models, raw fit is not enough.

The Parsimony Principle: We must use Information Criteria to objectively balance how well the model fits against how complex it is.

Technique Controls & Diagnostics for ARIMA

Information criteria objectively score models based on their efficiency

Akaike Information Criterion (AIC)	Bayesian Information Criterion (BIC)
$\mathrm{AIC} = -2\ln L + 2k$	$\mathrm{BIC} = -2\ln L + k\ln n$

Legend: k = Number of estimated parameters (AR/MA coefficients, constant, error variance)
n = Sample size (number of observations used for estimation)
ln L = Maximized Log-Likelihood

Technique Controls & Diagnostics for ARIMA

BIC applies a heavier penalty to complexity to favor simpler models

AIC Penalty Multiplier → $$2 > \ln n$$ ← BIC Penalty Multiplier

- **The Penalty Logic:** BIC punishes complexity significantly harder than AIC because $\ln n > 2$ for almost all real-world samples (where $n > 8$).
- **The Decision Rule:** Because it favors structurally parsimonious models, prioritize the model with the lowest BIC as the primary choice for long-term forecasting.

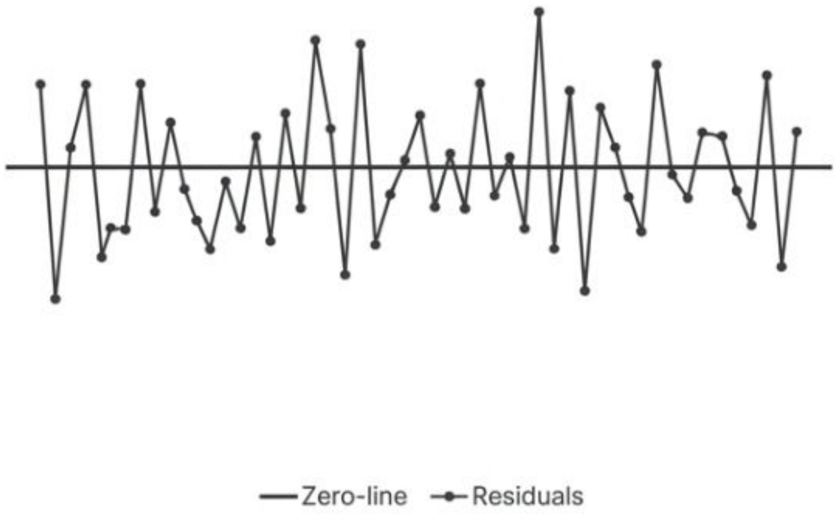
Technique Controls & Diagnostics for ARIMA

A successful model reduces all remaining residuals to purely White Noise

The Goal of Diagnostics: We isolate the model's residuals after fitting.

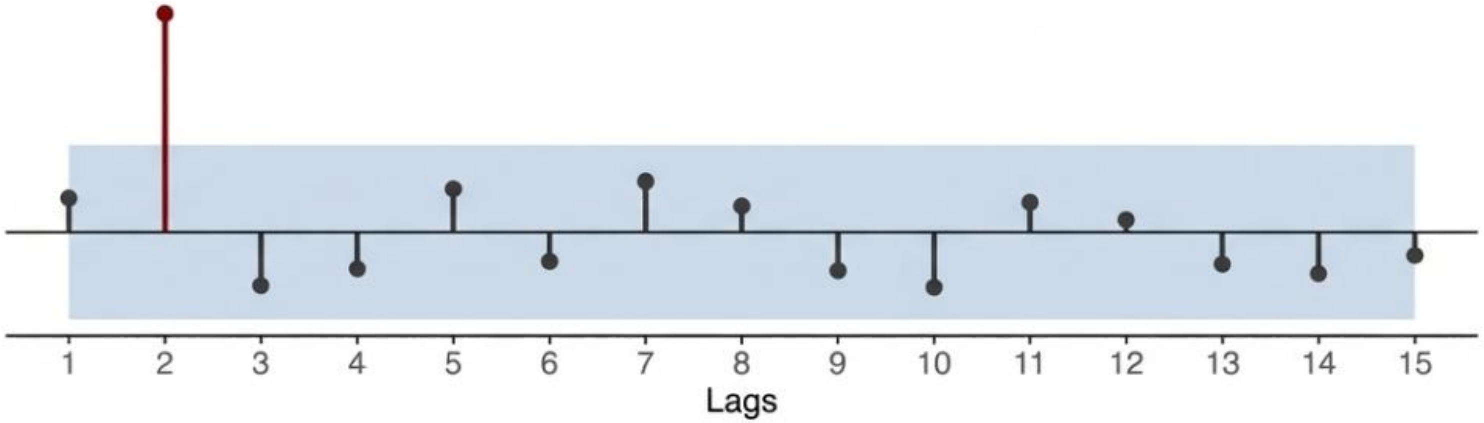
If the model has successfully captured all signals, structural patterns, and trends in the data, the residuals must qualify statistically as White Noise.

White Noise implies that no usable mathematical information remains in the leftovers.



Technique Controls & Diagnostics for ARIMA

Graphical control identifies trapped information via the ACF plot



- Confidence Threshold: All spikes must strictly fall within the $\pm$ 1.96 / $\sqrt{n}$ band (the shaded region).
- **Rejection Signs**: Spikes breaching the threshold at low lags (e.g., lag 2 above) prove the model failed to absorb all information. The model must be discarded, and the Identification phase must be repeated.

Technique Controls & Diagnostics for ARIMA

The Ljung-Box test provides a formal statistical check for autocorrelation

The Null Hypothesis (H0):

The residuals are completely uncorrelated.
(They are pure White Noise).

Visual graphical checks cannot replace formal, rigorous statistical testing.

The ultimate goal of the Ljung-Box test is to calculate if there is enough mathematical evidence to reject this Null Hypothesis.

Technique Controls & Diagnostics for ARIMA

Ljung-Box calculations require precise degrees of freedom corrections

Test Statistic (Q): Calculated based on the sum of squared sample correlations of the residuals.

Degrees of Freedom Correction: When testing an ARIMA model, the degrees of freedom must be manually corrected by $p + q$ (or $p + q + 1$ if a constant is included).

Why? This manual adjustment ensures the Chi-squared distribution is accurate for the resulting p-value calculation.

Test Statistic (Q)

↓

Degrees of Freedom Correction

$$model\_df = p + q$$

Technique Controls & Diagnostics for ARIMA

Interpret the Ljung-Box p-value to finalize model validation

p-value > 0.05	p-value < 0.05
Accept H0. The model is structurally adequate and mathematically cleared for forecasting.	Reject H0. The model is inadequate and has failed. You must reconfigure the parameters.

Technique Controls & Diagnostics for ARIMA

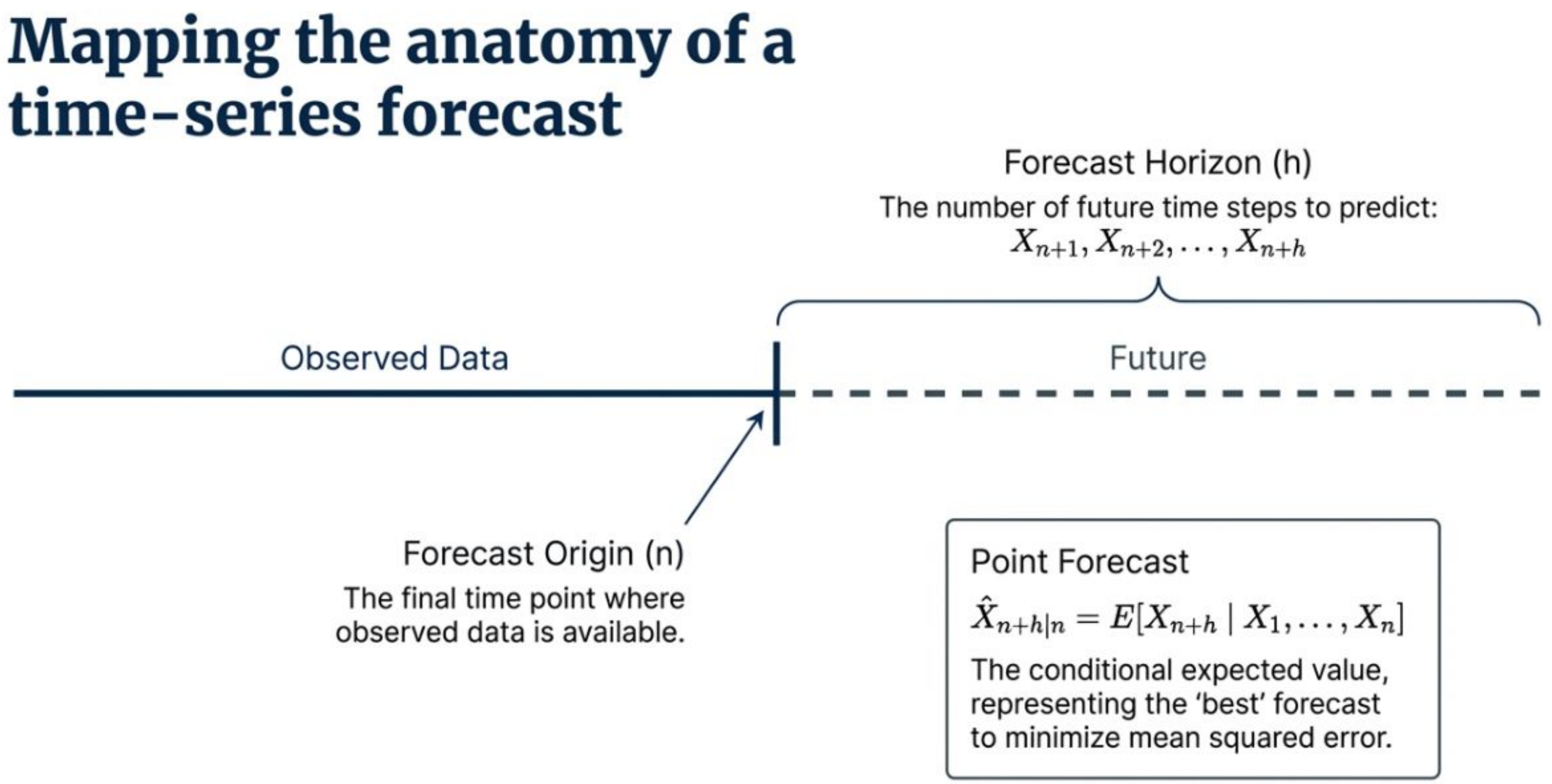
The definitive success criteria for technical control and diagnostics

Model Selection	BIC Score	Lowest value (Minimum)
Residual Mean	Summary Statistics	Approximately 0
Residual ACF	Graphical Plot	All spikes strictly within $\pm 1.96/\sqrt{n}$
Residual Correlation	Ljung-Box Test	p–value > 0.05

Content

- ARIMA
- The Box-Jenkins Methodology
- Technique Controls & Diagnostics for ARIMA
- **Forecasting & evaluation**

Forecasting & evaluation



Forecasting & evaluation

Two distinct structural approaches to forecasting the future

Multi-step Forecast

Fits the model once. Forecasts consecutively from a fixed origin.

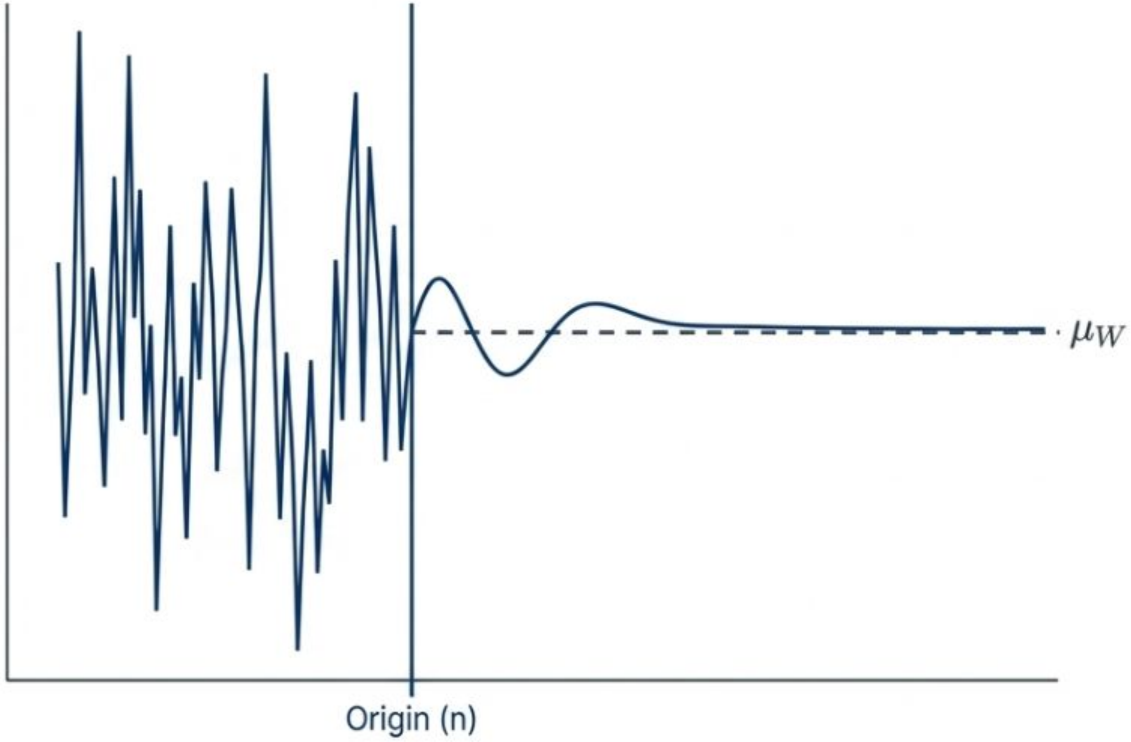
Rolling One-step-ahead Forecast

Updates continuously. Forecasts one step, absorbs new data, and repeats.

The choice of method depends entirely on the specific operational use case and testing requirements.

Forecasting & evaluation

Multi-step forecasting naturally drifts toward the unconditional mean



The ARIMA model is fitted once on the training set. It then forecasts h steps consecutively from the fixed origin.

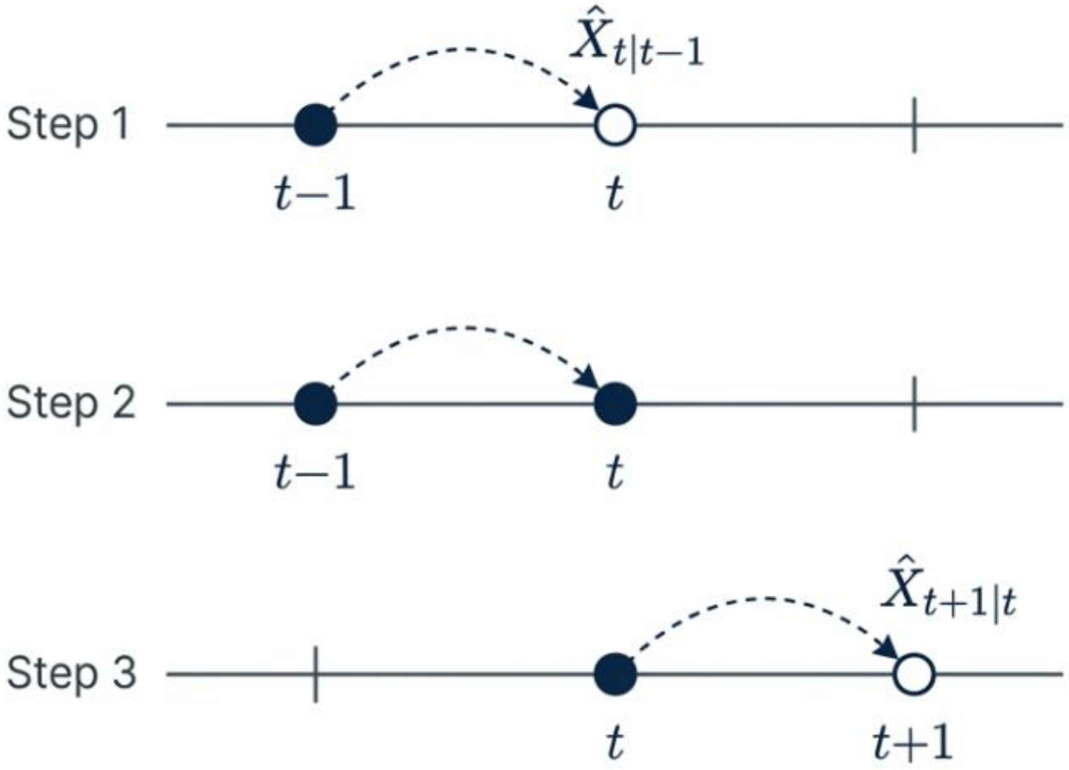
Mean Reversion

For stationary ARMA processes, as the horizon h increases, the forecast approaches the unconditional mean (μ_W). It reflects long-term average trends rather than tracking short-term actual fluctuations.

Optimal for fixed-period planning from the present moment (e.g., locking in a 12-month budget).

Forecasting & evaluation

Rolling one-step-ahead forecasts update iteratively with new data



Process

At each time t in the test set, predict the immediate next step. Update the origin with the actual observed value at t, then predict t+1.

Advantage

Yields higher accuracy by utilizing the most recent information, preventing error accumulation over multiple consecutive steps.

Use Case

The standard method for assessing a model's tracking ability against reality and comparing chart performance.

Point forecasts require intervals to quantify inherent uncertainty

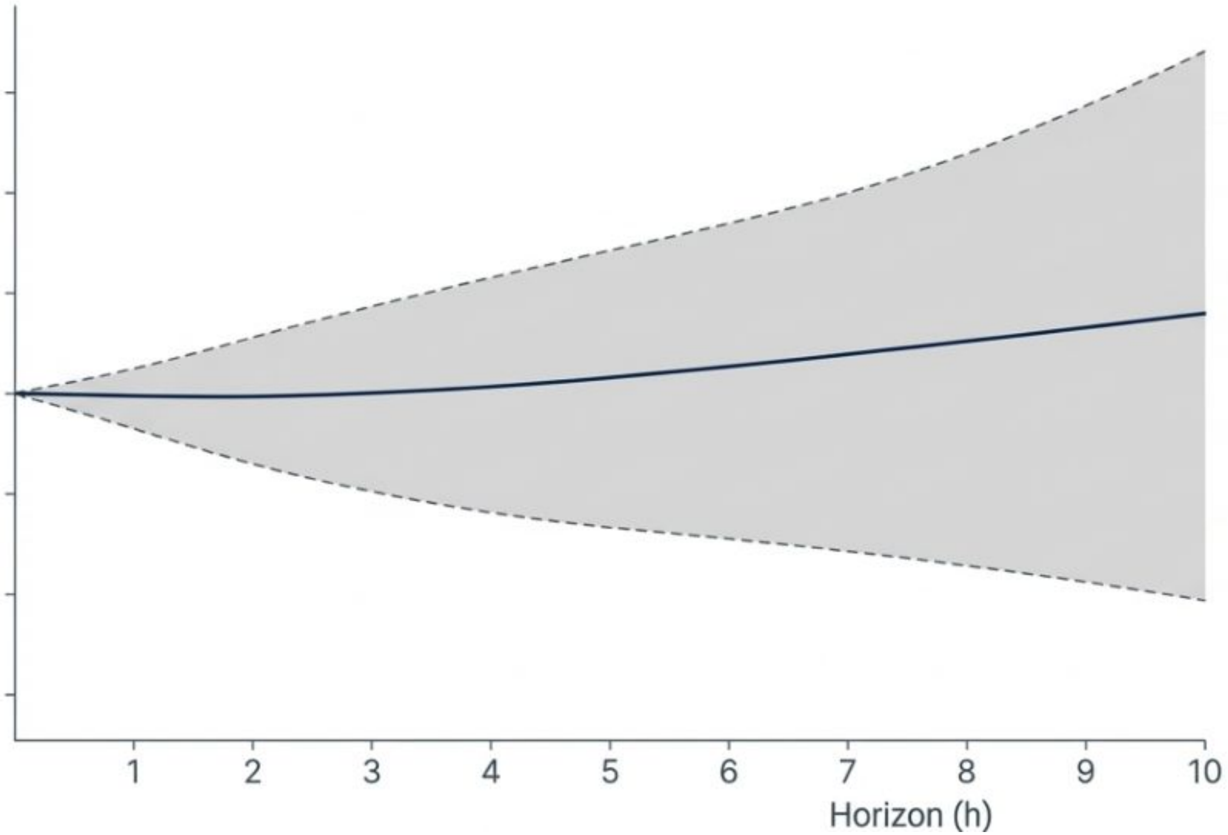
A forecast is never an absolute certainty. 95% forecast intervals are required to measure boundary expectations.

Future Shocks (ε_{n+h})
Unpredictable white noise inherent to future events.

Parameter Estimation ($\hat{\phi}, \hat{\theta}$)
Slight deviations and inaccuracies in the fitted model parameters.

Forecasting & evaluation

The ‘Expanding Cone’ reflects exponentially growing uncertainty



Key Takeaway

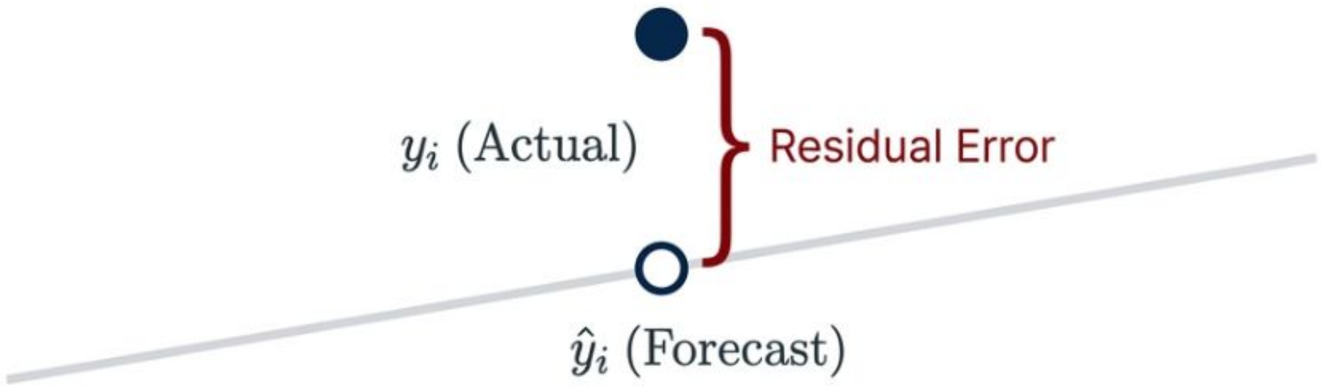
The further into the future the forecast extends, the wider the interval boundaries become.

Context

This visually proves that long-term predictions carry inherently higher risk and mathematical uncertainty than short-term predictions.

Forecasting & evaluation

Evaluating accuracy relies on measuring out-of-sample residuals



Core Concept

To determine model validity, we isolate the test set and rigorously compare actual observed values (y_i) against our model's predicted values ($\hat{y}_i$).

Next Steps

We quantify this gap using three standard mathematical metrics.

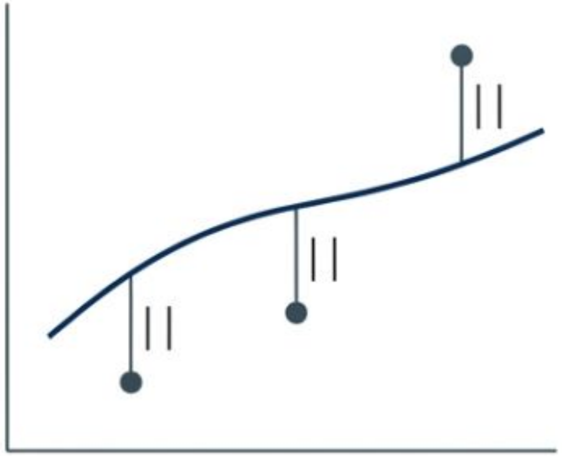
Forecasting & evaluation

Mean Absolute Error (MAE) captures the average magnitude

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Meaning

Calculates the average absolute deviation of forecasts from actuals.



Characteristics

Remains in the identical unit to the original dataset.
Highly robust against isolated outlier values.

Forecasting & evaluation

Root Mean Square Error (RMSE) heavily penalizes large variances

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Meaning

Measures the standard deviation of the forecasting residuals.

Characteristics

The squaring step strictly punishes large errors. Consequently, it is the most universally adopted metric for model optimization.

Forecasting & evaluation

Mean Absolute Percentage Error (MAPE) enables cross-scale comparison

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{|y_i|}$$

Meaning

Expresses the forecasting error as a standardized percentage.

Characteristics

- Scale-independent, allowing comparison across completely different datasets.

⚠ Critical Caveat: Mathematically fails or deeply misleads if actual observed values are exactly zero or near-zero.

Forecasting & evaluation

Selecting the appropriate metric based on data characteristics

	MAE	RMSE	MAPE
Unit	Native	Native	Percentage
Outlier Sensitivity	Low	High (Penalizes)	Skews heavily on zero
Primary Utility	General baseline accuracy.	Primary metric for model optimization.	Comparing structurally different datasets.

Forecasting & evaluation

Final Implementation Checklist: Preparation and Fitting

1

Data Splitting

Isolate the dataset chronologically into a Training set (typically 85-90%) and a holdout Test set (10-15%).

2

Model Fitting

Take the optimal ARIMA structure identified during Phase 3 and fit its parameters exclusively on the Training set.

3

4

Forecasting & evaluation

Final Implementation Checklist: Generation and Evaluation

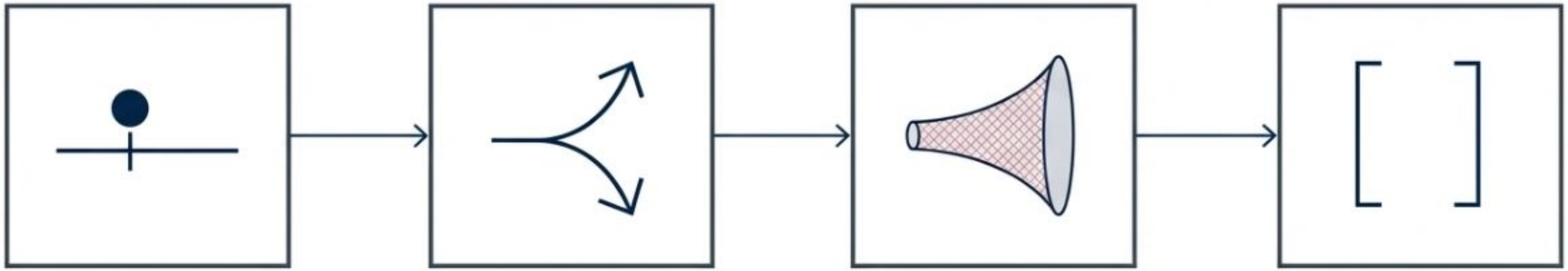
3. **Generate & Calculate**
   Execute predictions across the Test set horizon. Compute MAE and RMSE to quantify the out-of-sample error.

4. **Visual Assessment**
   Plot the point forecasts overlaying the actuals, enveloped by the 95% confidence intervals, to visually confirm model logic and tracking capability.

Validation Complete

Forecasting & evaluation

Phase 4 Summary: From Origin to Validation



Define Parameters

Establish Forecast Origin (n) and Horizon (h).

Select Method

Choose between Multi-step (long-term trend) or Rolling (short-term tracking).

Expect Uncertainty

Respect the expanding cone of variance driven by shocks and parameter estimation.

Rigorous Validation

Measure out-of-sample residuals strictly via RMSE, RMSE, MAE, or MAPE before finalizing the model.

Thank you