Advanced Time Series Analysis & Forecasting

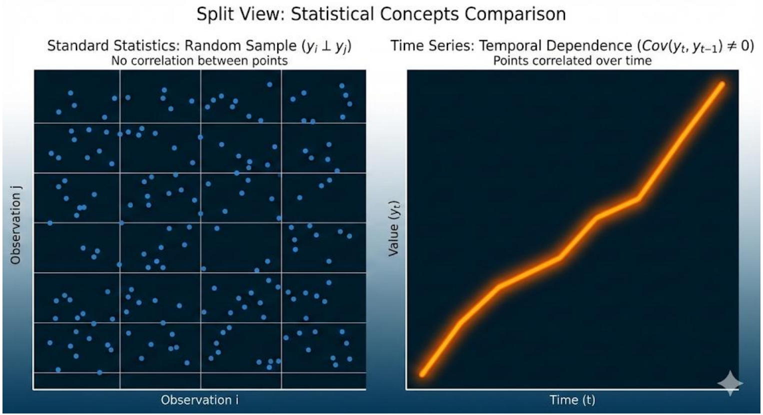
Lecture 9: Decomposition, ARIMA/ETS, Holidays & Cross-Validation

Assoc. Prof. Nguyen Binh Minh, Ph.D.

Topic: Advanced Time Series Analysis & Forecasting

- **Key Concepts:**
  - Decomposition (STL)
  - Exponential Smoothing (ETS)
  - ARIMA & Stationarity
  - Handling Holidays & Exogenous Variables
  - Time Series Cross-Validation

The Strategic Imperative of Temporal Analytics



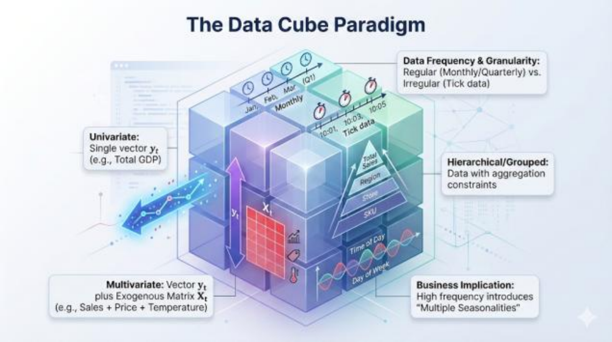
Why do we analyze time?

- Strategic Planning: Long-term trend analysis (e.g., Australian Tourism demand for 2030).
- Operational Efficiency: Short-term resource allocation (e.g., Rossmann store staffing for next Tuesday).
- Risk Management: Volatility modeling (ARCH/GARCH) - *Note: We focus on mean forecasting today.*

The Core Shift:

- Standard Statistics: $y_i \perp y_j$ (Independence assumption).
- Time Series: $Cov(y_t, y_{t-1}) \neq 0$ (Dependence is the signal).

The "Data Cube" Paradigm



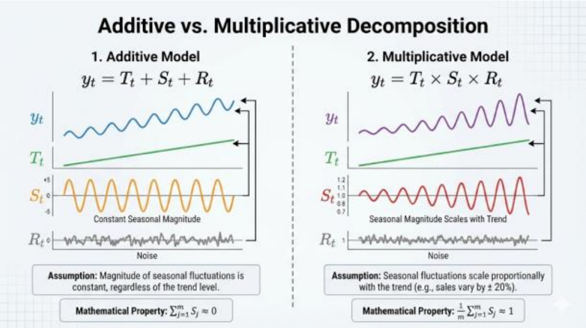
Structure of Business Time Series:

- Univariate: Single vector y_t (e.g., Total GDP).
- Multivariate: Vector $\mathbf{y_t}$ $plus\ Exogenous\ Matrix$ $\mathbf{X_t}$ (e.g., Sales + Price + Temperature).
- Hierarchical/Grouped: Data with aggregation constraints (e.g., Total Sales → Region → Store → SKU).

Data Frequency & Granularity:

- Regular (Monthly/Quarterly) vs. Irregular (Tick data).
- Business Implication: High frequency (Daily/Hourly) introduces "Multiple Seasonalities" (e.g., time of day AND day of week).

Components of a Time Series



The Generative Process: We assume the observed series y_t is a combination of unobserved components:

1. Trend-Cycle (T_t): The long-term direction and economic cycles.
2. Seasonality (S_t): Patterns repeating with fixed frequency m (e.g., $m = 12$ for monthly).
3. Remainder (R_t): The stochastic noise (residuals).

Goal of Analysis:

$$\text{Signal Extraction: } y_t \to \{T_t, S_t\}$$
$$\text{Forecasting: } \widehat{y_{t+h}} = f\left(\widehat{T_{t+h}}, \widehat{S_{t+h}}\right)$$

Additive vs. Multiplicative Decomposition

1. Additive Model:

$$y_t = T_t + S_t + R_t$$

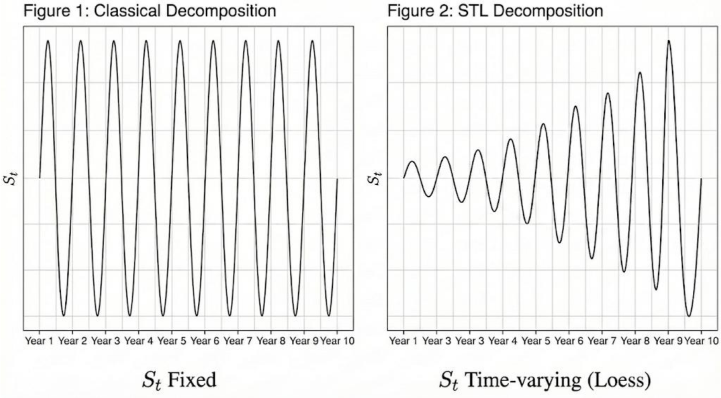
- Assumption: Magnitude of seasonal fluctuations is constant, regardless of the trend level.
- Mathematical Property: $\sum_{j=1}^{m} S_j \approx 0$

2. Multiplicative Model:

$$y_t = T_t \times S_t \times R_t$$

- Assumption: Seasonal fluctuations scale proportionally with the trend (e.g., sales vary by $\pm 20\%$, not $\pm \$100$).
- Mathematical Property: $\frac{1}{m}\sum_{j=1}^{m} S_j \approx 1.$

STL Decomposition (Seasonal-Trend with Loess)



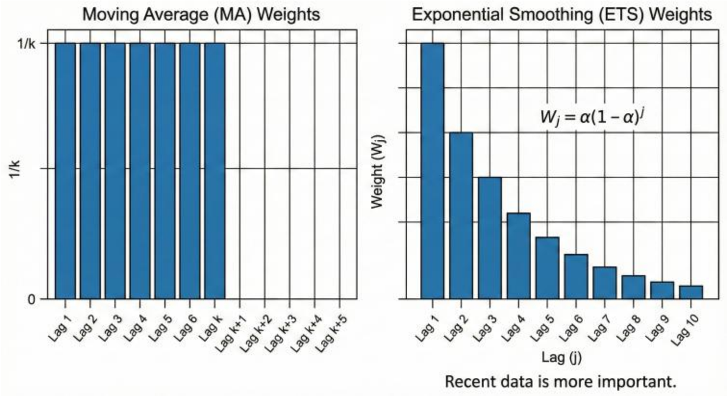
Why STL?

- Classical Decomposition: Assumes S_t is constant forever.
- STL: Allows S_t to change over time (e.g., people shopping more online in December now than 10 years ago).

Mechanism:

- Uses Loess (Locally Estimated Scatterplot Smoothing).
- Iterative process to estimate Trend and Seasonality.
- Robustness: Can handle outliers without distorting the trend.

The Philosophy of Exponential Smoothing



Moving Average vs. Exponential Smoothing

- **Simple Moving Average (MA):** Equal weights for last \$$k$\$ observations.

  $\text{Weight} = 1/k$ for all lags $1 \ldots k$.

- **Exponential Smoothing:** Weights decay exponentially as data gets older.
  - Recent data is more relevant for forecasting.

The Weighting Function:

$$\text{Weight}_j = \alpha(1-\alpha)^j$$

- Where $0 < \alpha < 1$ is the smoothing parameter.

Simple Exponential Smoothing (SES) - ETS(A,N,N)

Use Case: Data with no clear trend and no seasonality.

1. State Space Form:Forecast Equation: $y_{\widehat{t+h|t=}\ell_t}$ (Flat forecast).
2. Level Equation: $\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1}$

Interpretation: The new level ℓ_t is a weighted average of the new observation y_t and the previous level ℓ_{t-1}.

- y_t: New information.
- ℓ_{t-1}: Accumulated history.

Holt's Linear Trend Method - ETS(A,A,N)

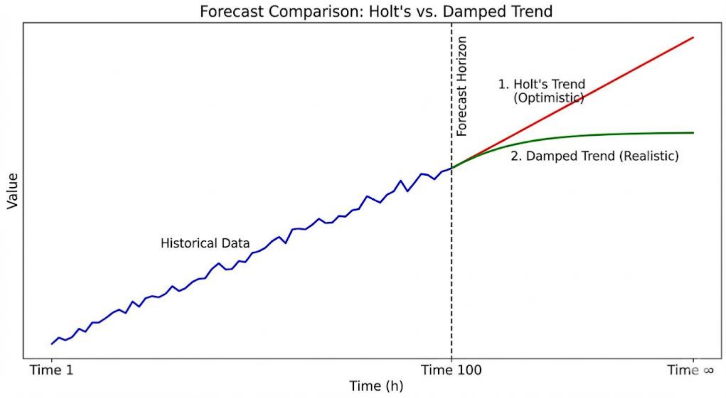
Use Case: Data with a trend but no seasonality. Two State Variables:

1. Level (ℓ_t)
2. Trend/Slope (b_t)

Equations:

- Forecast: $\widehat{y_{t+h|t=\ell_t}+hb_t}$
- Level: $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$
- Trend: $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$

Damped Trend Method - ETS(A,Ad,N)



The Business Reality: Trends rarely continue linearly forever. Competition and market saturation dampen growth.

Damping Parameter (ϕ):

- $0 < \phi < 1$
- Forecast: $y_{t+h|t}=\ell_t+(\widehat{\phi+\phi}^2+\cdots+\phi^h)b_t$

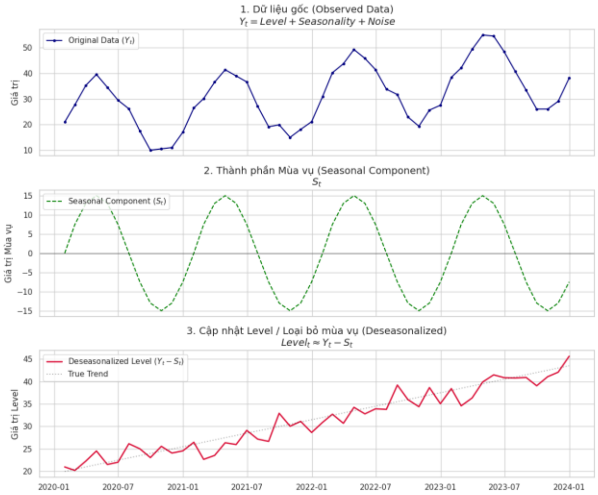
Behavior: The forecast trend flattens out into a horizontal line as $h \to \infty$. This is often the most accurate model for long-term business forecasting (e.g., sales budgeting).

Holt-Winters Seasonal Method (Additive) - ETS(A,A,A)

Use Case: Trend + Constant Seasonality. Three State Variables: Level (ℓ), Trend (b), Seasonal (s).

Recursive Equations:

1. Forecast: $y_{t+h|t=\widehat{\ell_t+hb_t}+s_{t-m+k}}$
2. Level: $\ell_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1})$
3. Trend: $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$
4. Seasonal:$s_t = \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$



Holt-Winters Seasonal Method (Multiplicative) - ETS(M,A,M)

Use Case: Trend + Proportional Seasonality (e.g., Air Passengers).

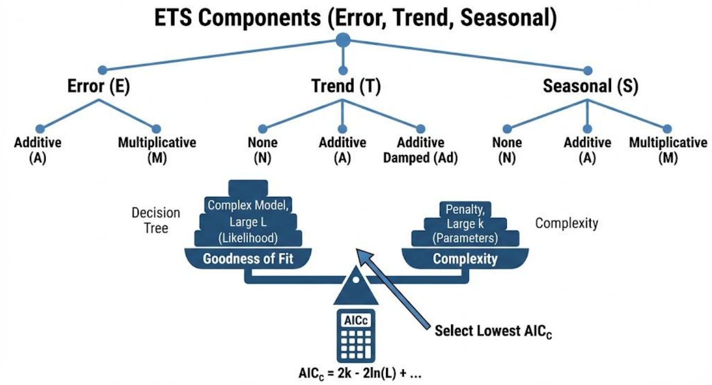
Equations:

- Forecast: $y_{t+h|t=(\widehat{\ell_t+hb_t})\times s_{t-m+k}}$
- Level: $\ell_t = \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1})$
- Seasonal: $s_t = \gamma \frac{y_t}{\ell_{t-1}+b_{t-1}} + (1-\gamma)s_{t-m}$

Key Difference:

- Additive uses Subtraction ($y_t - s$).Multiplicative uses Division (y_t/s).

ETS Taxonomy & Model Selection



The ETS (Error, Trend, Seasonal) Framework: We can mix and match components.

- Error: Additive (A), Multiplicative (M)
- Trend: None (N), Additive (A), Damped (Ad)
- Seasonal: None (N), Additive (A), Multiplicative (M)

Automatic Selection via AICc: We fit multiple models and select the one with the lowest AICc (Corrected Akaike Information Criterion).

$$AICc = -2\ln(L) + 2k + \frac{2k(k+1)}{T-k-1}$$

- Balances goodness of fit (L) vs. complexity (k).

ARIMA - Introduction

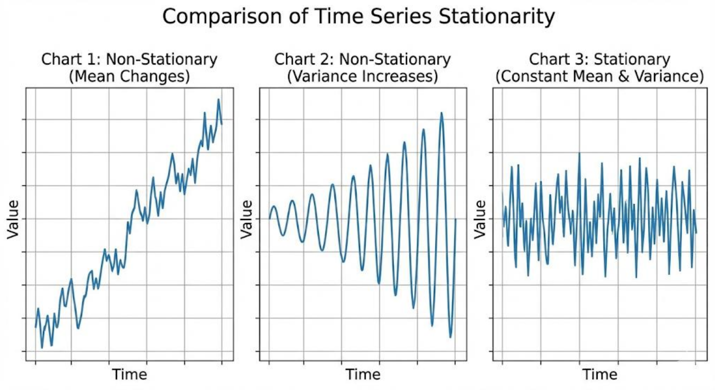
ARIMA: AutoRegressive Integrated Moving Average.

- Philosophy: Describes the autocorrelations in the data.
- Contrast to ETS: ETS describes Trend/Seasonality; ARIMA describes the correlations between lags.

The Three Pillars (p, d, q):

- AR (p): Regression on past values.
- I (d): Differencing to achieve stationarity.
- MA (q): Regression on past errors.

Stationarity (The Prerequisite)



Definition of Stationarity: A time series is stationary if its statistical properties do not change over time.

- Constant Mean (μ).
- Constant Variance (σ^2).
- Autocovariance depends only on lag k, not time t.

Visual Test: No trend, no seasonality, looks like "white noise" with constant amplitude.

Augmented Dickey-Fuller (ADF) Test

Objective: Rigorous statistical test for stationarity (Unit Root). Hypotheses:

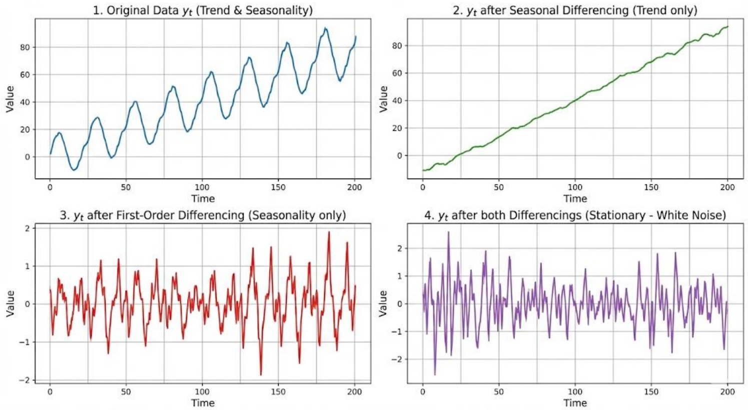
- H_0: Series has a unit root (Non-Stationary) $\rightarrow p > 0.05$
- H_a: Series is stationary. $\implies p < 0.05$

Test Equation:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum \delta_i \Delta y_{t-i} + \epsilon_t$$

We test if coefficient $\gamma = 0$ (which implies unit root).

Differencing (The 'I' in ARIMA)



First Differencing ($d = 1$):

$$y'_t = y_t - y_{t-1}$$

- Removes linear trend.

Seasonal Differencing ($D = 1$):

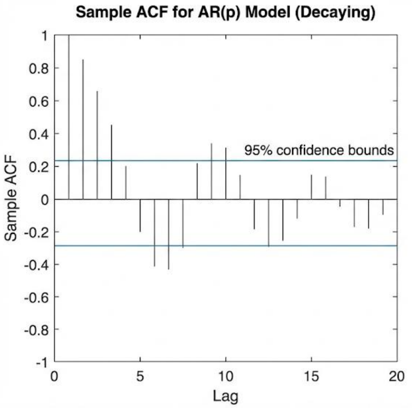
$$y'_t = y_t - y_{t-m}$$

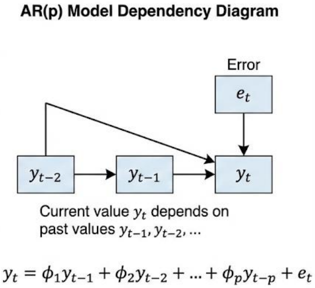
- Removes seasonality (e.g., subtract Jan 2023 from Jan 2024).

Backshift Notation (B):

$$By_t = y_{t-1}$$
$$y'_t = (1 - B)y_t$$

AutoRegressive Models - AR(p)





Concept: The current value is a linear combination of past p values.

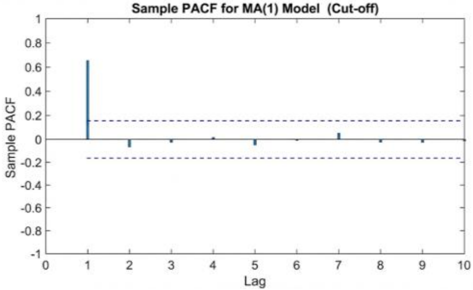
Equation:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

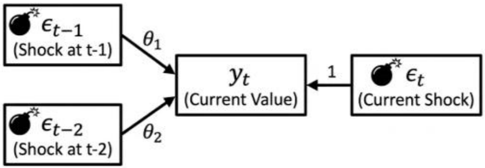
Using Backshift Operator:

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + \epsilon_t$$

Interpretation: System has "momentum." If sales were high yesterday, they are likely high today.



MA(2) Process: Current Value Determined by Past Shocks



$$y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

Moving Average Models - MA(q)

Concept: The current value is a linear combination of past q forecast errors (shocks). Equation:

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

Using Backshift Operator:

$$y_t = c + \left(1 + \theta_1 B + \cdots + \theta_q B^q\right) \epsilon_t$$

Interpretation: System is impacted by random shocks that dissipate over q periods.

Determining p and q: ACF & PACF

Diagnostic Plots:

- ACF (Autocorrelation Function): Correlation between y_t and y_{t-k}. (Direct + Indirect effects).
- PACF (Partial Autocorrelation): Correlation between y_t and y_{t-k} removing effects of lags $1 \ldots k-1$. (Direct effect only).

Rules of Thumb:

Model	ACF Pattern	PACF Pattern
AR(p)	Decays exponentially	**Cuts off** after lag p
MA(q)	**Cuts off** after lag q	Decays exponentially

ARIMA Model Selection Procedure (Box-Jenkins)

1. Plot data: Check for trend/seasonality.
2. Transform: Box-Cox (Logs) to stabilize variance.
3. Stationarize: Apply differencing (d, D) until ADF test passes.
4. Identify: Inspect ACF/PACF to guess p, q, P, Q.
5. Estimate: Use Maximum Likelihood Estimation (minimize AICc).
6. Diagnose: Check residuals. Must be white noise (Ljung-Box Test).

The Limit of Univariate Models

Scenario:

- You are forecasting Rossmann Sales.
- Easter is in March one year, April the next.
- Ramadan shifts by ~11 days/year.

Problem: Standard SARIMA assumes seasonality is fixed to the calendar month (e.g., "March effect"). It cannot handle Moving Holidays. SARIMA will look for a spike in March and miss the April Easter.

Dynamic Regression (ARIMAX / Reg-ARIMA)

Concept: Combine Linear Regression (for external factors) with ARIMA (for the time series error structure).

Equation:

$$y_t = \underbrace{\beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t}}_{\text{Regression}} + \eta_t$$

Key Insight: We forecast the residuals of the regression using ARIMA.

Modeling Holidays with Dummy Variables

Implementation Strategy: Create a binary column for every holiday.

- $x_{Easter,t} = 1$ if date t is Easter, else 0.

Window Effects: Holidays often have "lead-up" or "hangover" effects.

- Christmas: Sales spike 7 days before.
- New Year: Sales drop 2 days after.
- Solution: Create multiple dummies (e.g., Christmas_Lag_7, Christmas_Lead_2).

Facebook Prophet (Alternative Approach)

Why Prophet? Designed for business analysts to handle holidays/regressors easily.

- Additive Model:

- : Piecewise linear trend (handles structural breaks).
- $s(t)$: Fourier terms for seasonality.
- $h(t)$: Holiday effects (user-supplied list of dates).

Pros/Cons vs. ARIMA:

- Pros: Handles outliers better, easy holiday integration, fast.
- Cons: Less rigorous statistical grounding than ARIMA.

The Golden Rule of Time Series Validation

Standard CV (K-Fold): Randomly shuffle data. Train on 80%, Test on 20%. Time Series Constraint: ILLEGAL!

- Shuffling destroys temporal dependence.
- "Peeking into the future": You cannot use data from 2024 to train a model predicting 2023.

Correct Approach: Time Series Cross-Validation (Walk-Forward Validation).

TSCV Strategy 1 - Expanding Window

Mechanism:

- Train: $$. $Forecast$: $T+1$.
- Train: $$. $Forecast$: $T+2$.
- Train: $$. $Forecast$: $T+3$.

Pros: Uses all available history.

Cons: Computationally expensive (retraining on growing data).

TSCV Strategy 2 - Rolling Window

Mechanism:

- Train: \$\$. $Forecast$: $T + 1$.
- Train: \$\$. $Forecast$: $T + 2$. (Drop index 1).

Pros:

- Constant training size.
- Adapts to "Structural Breaks" (forgets old, irrelevant history).

Evaluation Metrics - The Problem with MAPE

MAPE (Mean Absolute Percentage Error):

$$MAPE = mean\left(\left|\frac{y_t - \widehat{y_t}}{y_t}\right|\right) \times 100$$

Flaws:

1. Undefined at Zero: If $y_t = 0$ (e.g., store closed), MAPE = ∞.
2. Asymmetric: Penalizes positive errors more than negative ones.

The Gold Standard - MASE

MASE (Mean Absolute Scaled Error):

$$MASE = \frac{\mathrm{MAE}_{model}}{\mathrm{MAE}_{naive}}$$

- Numerator: MAE of your complex model.
- Denominator: MAE of the Seasonal Naive forecast on training data.

Interpretation:

- $MASE < 1$: Model is better than Naive (Good).
- $MASE > 1$: Model is worse than Naive (Useless).

Case Study 1 - Australian Domestic Tourism

Data: Quarterly visitor nights (1998-2016). Characteristics:

- Strong Trend (Growth).
- Strong Multiplicative Seasonality (Summer peaks grow larger as total tourists increase).
- Noise: Variable.

Model Selection:

- ARIMA: Requires Log transformation to handle multiplicative seasonality.
- ETS: ETS(M,A,M) (Multiplicative Error, Additive Trend, Multiplicative Seasonality) fits natively.
- Winner: ETS typically outperforms ARIMA on this specific dataset due to the clear seasonal shape.

Case Study 2 - Rossmann Store Sales

Data: 1,115 Drug Stores, Daily Sales. Complexity:

- Seasonality: Weekly (closed Sundays) + Annual (Christmas).
- Exogenous: Promo (Marketing), StateHoliday (Easter/Christmas), CompetitionOpen.
- Zeros: Closed days.

Approach:

- Naive: Fails (ignores Promos).
- Univariate ARIMA: Fails (cannot see Promo schedule).
- Dynamic Regression (ARIMAX):
  - y_t: Log(Sales).
  - x_t: Dummies for Promo, DayOfWeek, Holidays.
  - Errors: AR(1) or AR(2) to capture remaining serial dependence.

Summary of Methodologies

Feature	**ETS**	**ARIMA**	**Dynamic Reg / Prophet**
Focus	Trend & Seasonality	Autocorrelations	Exogenous Variables
Stationarity	Not Required	**Required** (d)	Required for residuals
Holidays	Difficult	Difficult	**Excellent**
Interpretation	Intuitive (Wgt Avg)	Abstract (Lags)	Regression Coefficients

The Forecasting Workflow

1. Visualize: Identify Trend, Seasonality, Outliers.
2. Pre-process: Log transform? Clean outliers?
3. Benchmark: Calculate MASE of Seasonal Naive.
4. Model: Fit ETS and ARIMA (Auto-selection via AICc).
5. Expand: Add Regressors if holidays/promos exist.
6. Validate: Time Series Cross-Validation (Rolling).
7. Select: Model with lowest Test MASE.

Final Thoughts & Resources

- Key Takeaway: "All models are wrong, some are useful." - George Box. Complex models (Deep Learning/LSTM) often fail against simple robust models (ETS) on noisy monthly data. Start simple.
- Required Reading:
  - Hyndman, R.J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed).
  - Makridakis Competitions (M4/M5) Findings.