Daniel Traian Pele

Bucharest University of Economic Studies
Faculty of Cybernetics, Statistics, and Economic Informatics
daniel.pele@csie.ase.ro

Academic Year 2025-2026

logo_csie.png

logo_ase.png

Outline

Course Overview: Methods Covered

Classical Methods

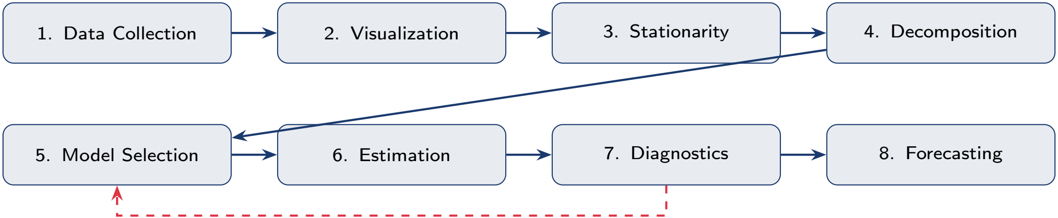
- Ch 1: Time Series Fundamentals
- Ch 2: ARMA Models
- Ch 3: ARIMA Models
- Ch 4: SARIMA Models
- Ch 5: GARCH Models

Advanced Methods

- Ch 6: VAR & Granger Causality
- Ch 7: Cointegration & VECM
- Ch 8: Modern Extensions
- Ch 9: Prophet & TBATS

Today: Apply ALL to Real Data!

The Complete Analysis Workflow



Key Principle

Model diagnostics may require returning to model selection (iterative process)

Real Datasets for This Chapter

S&P 500 Returns

- Daily financial data
- 2019-2024
- Volatility clustering
- ARIMA + GARCH

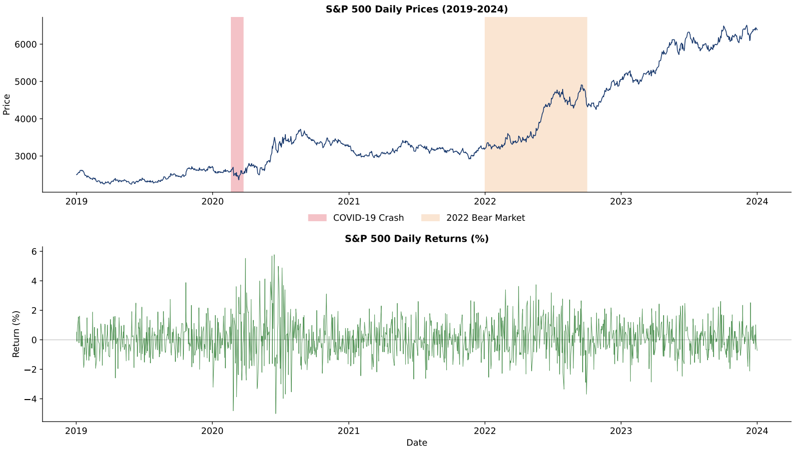
Air Passengers

- Monthly 1949-1960
- Classic dataset
- Trend + seasonality
- SARIMA vs Prophet

US Retail Sales

- Monthly 2018-2023
- FRED economic data
- COVID-19 impact
- Structural breaks

S&P 500: Data Overview



- **Data:** S&P 500 daily closing prices and returns (2019-2024)
- **Key events:** COVID-19 crash (March 2020), recovery, 2022 bear market

Step 1: Stationarity Testing

Augmented Dickey-Fuller Test

- H_0: Unit root (non-stationary)
- H_1: Stationary

Results on S&P 500:

Series	ADF Statistic	p-value
Prices	-1.23	0.66
Returns	-35.2	< 0.001

$\Rightarrow$ Prices: non-stationary
$\Rightarrow$ Returns: stationary

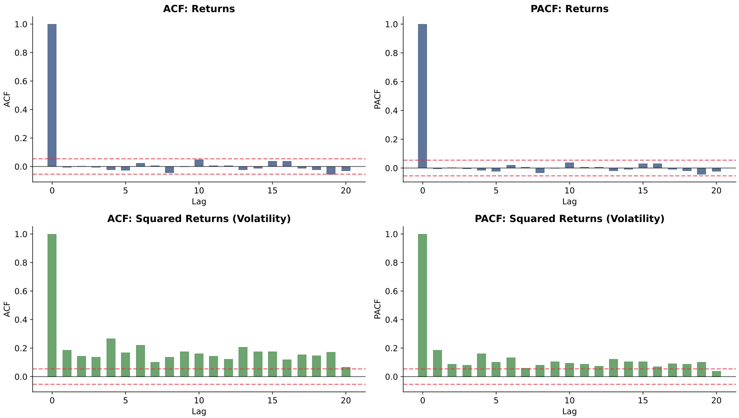
KPSS Test (Confirmation)

- H_0: Stationary
- H_1: Unit root

Prices: KPSS $= 4.21^{**}$
Returns: KPSS $= 0.08$

Both tests confirm: use returns!

Step 2: ACF/PACF Analysis of Returns



- **Returns:** Near white noise (weak autocorrelation)
- **Squared returns:** Strong persistence $\Rightarrow$ volatility clustering
- **Implication:** Need GARCH for volatility modeling

Step 3: ARIMA Model for Returns

Model Selection using AIC/BIC:

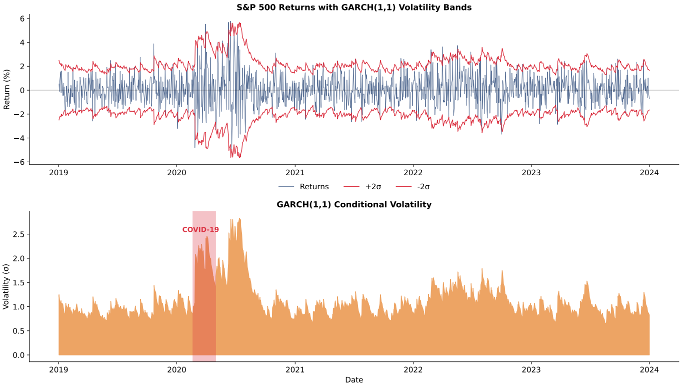
Model	AIC	BIC
ARIMA(0,0,0)	-8234	-8228
ARIMA(1,0,0)	-8236	-8224
ARIMA(0,0,1)	-8235	-8223
ARIMA(1,0,1)	**-8238**	**-8220**

Best model: ARIMA(1,0,1) or simple mean model

Key Insight

Stock returns are nearly unpredictable (efficient market hypothesis). The real story is in the **volatility**, not the mean!

Step 4: GARCH Model for Volatility



GARCH(1,1) Model:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

- High $\alpha + \beta \approx 0.98$ indicates strong volatility persistence
- COVID-19 period shows massive volatility spike

Financial Data: Summary of Approach

Raw Prices $\xrightarrow{\text{diff/log}}$ Log Returns $\xrightarrow{\text{mean eq.}}$ ARIMA(1,0,1) $\xrightarrow{\text{variance eq.}}$ GARCH(1,1)

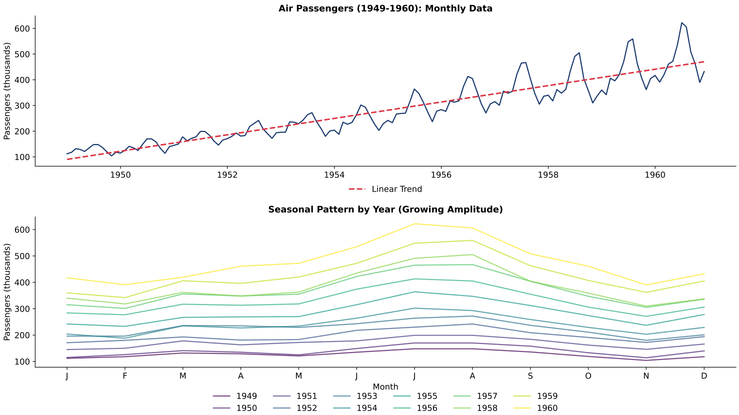
Key Findings:

- Returns are nearly unpredictable
- Volatility clusters significantly
- GARCH captures risk dynamics

Practical Use:

- Risk management (VaR)
- Option pricing
- Portfolio optimization

Air Passengers: The Classic Dataset



- **Data:** Monthly airline passengers (thousands), 1949-1960
- **Characteristics:** Upward trend + yearly seasonality + multiplicative pattern

Step 1: Decomposition Analysis

- **Trend:** Strong upward growth (aviation industry expansion)
- **Seasonal:** Peak in summer (July-August), trough in winter
- **Multiplicative:** Seasonal amplitude grows with level

Step 2: Making Data Stationary

Transformations Applied:

1. Log transformation (stabilize variance)
2. First difference (remove trend)
3. Seasonal difference (remove seasonality)

Series	ADF p-value
Original	0.99
Log	0.42
Log + Diff	0.07
Log + Diff + SDiff	< 0.01

SARIMA notation:

$$\text{SARIMA}(p, d, q)(P, D, Q)_s$$

For Air Passengers:

- $d = 1$ (first difference)
- $D = 1$ (seasonal difference)
- $s = 12$ (monthly seasonality)

Step 3: SARIMA Model Fitting

Model Selection:

Model	AIC
$(1,1,1)(1,1,1)_{12}$	1020.3
$(0,1,1)(0,1,1)_{12}$	1018.7
$(2,1,1)(0,1,1)_{12}$	**1017.1**
$(1,1,0)(1,1,0)_{12}$	1025.4

Best: SARIMA$(2,1,1)(0,1,1)_{12}$

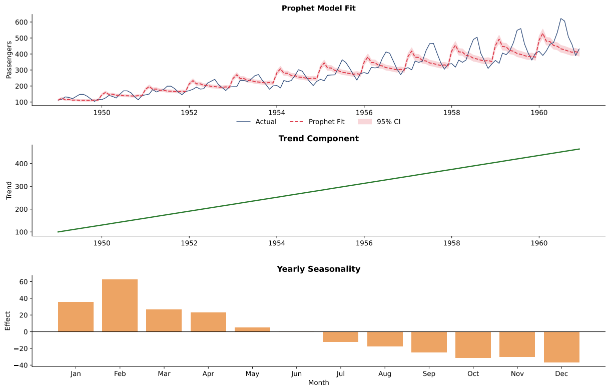
Model Equation:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})y_t$$
$$= (1 + \theta_1 B)(1 + \Theta_1 B^{12})\varepsilon_t$$

Diagnostics:

- Ljung-Box: p = 0.42
- Residuals: white noise

Step 4: Prophet as Alternative

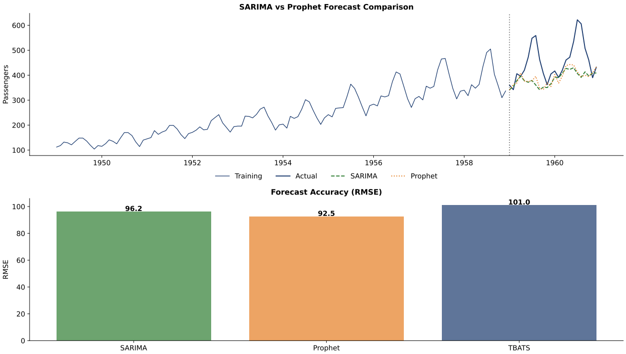


Prophet Model:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

- Trend $g(t)$: Piecewise linear with changepoints
- Seasonality $s(t)$: Fourier series (multiplicative mode)

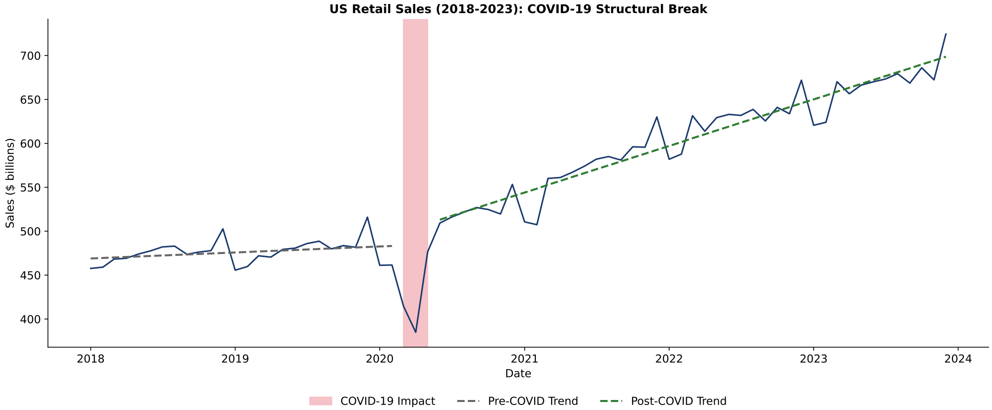
SARIMA vs Prophet: Comparison



Model	RMSE	MAE	MAPE (%)
SARIMA$(2,1,1)(0,1,1)_{12}$	18.2	14.5	3.1
Prophet (multiplicative)	21.4	17.2	3.7
TBATS	19.8	15.9	3.4

Conclusion: SARIMA slightly better for this classic, well-behaved dataset

US Retail Sales: COVID-19 Impact



- **Data:** US Retail Sales, monthly, 2018-2023 (FRED)
- **Challenge:** Major structural break in March-April 2020

Handling Structural Breaks

Option 1: Truncate Data

- Use only post-COVID data
- Pro: Clean, no breaks
- Con: Lose historical info

Option 2: Dummy Variables

- Add COVID indicator
- Pro: Uses all data
- Con: Complex in ARIMA

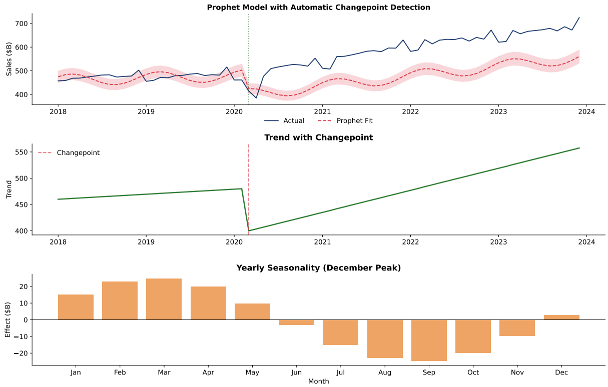
Option 3: Prophet with Changepoints

- Automatic detection
- Pro: Handles breaks naturally
- Con: May overfit

Recommendation

For COVID-impacted data, Prophet's changepoint detection or post-COVID truncation often works best.

Prophet for Retail Sales



Prophet Configuration:

- `changepoint_prior_scale = 0.1` (flexible for COVID)
- `seasonality_mode = 'multiplicative'`
- Automatic changepoint detection captures the break

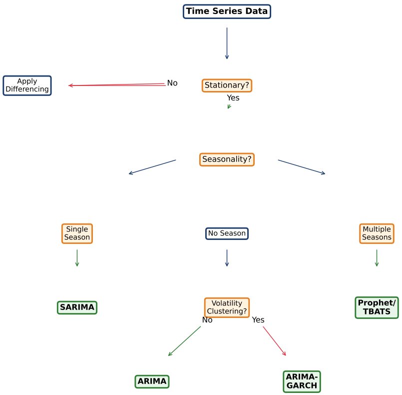
Model Comparison on Retail Sales

Model	RMSE ($B)	MAE ($B)	MAPE (%)
SARIMA (full data)	42.1	35.8	6.2
SARIMA (post-COVID)	18.3	14.7	2.4
Prophet (with breaks)	**15.2**	**12.1**	**2.0**
TBATS	19.4	15.8	2.6

Key Lesson

When data has structural breaks, traditional ARIMA struggles. Prophet's flexibility with changepoints makes it better suited for such scenarios.

Decision Framework



Model Selection Summary

Data Type	Characteristics	Recommended Model	Alternatives
Financial returns	No trend, volatility clustering	ARIMA-GARCH	EGARCH, GJR
Single seasonality	Trend + one seasonal period	SARIMA	ETS, Prophet
Multiple seasonality	Daily + weekly + annual	Prophet, TBATS	Dynamic regression
Structural breaks	COVID, regime changes	Prophet	Piecewise regression
Multiple series	Interdependencies	VAR, VECM	Factor models

Forecast Evaluation Metrics

Point Forecast Metrics:

RMSE (Root Mean Square Error):

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

MAE (Mean Absolute Error):

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

MAPE (Mean Absolute % Error):

$$\frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

When to Use Each:

- **RMSE**: Penalizes large errors more
- **MAE**: Robust to outliers
- **MAPE**: Scale-independent

Cross-Validation

Always use time series CV:

- Rolling window
- Expanding window
- Never shuffle!

Course Summary: Complete Toolkit

Understanding the Data

- Visualization first!
- Test for stationarity (ADF, KPSS)
- Identify seasonality patterns
- Check for structural breaks

Classical Models

- ARIMA: Non-seasonal data
- SARIMA: Single seasonality
- GARCH: Volatility modeling

Modern Approaches

- Prophet: Interpretable, handles breaks
- TBATS: Multiple seasonalities
- VAR/VECM: Multiple time series

Best Practices

- Always check diagnostics
- Use cross-validation
- Compare multiple models
- Domain knowledge matters!

Final Recommendations

1. **Start Simple**: Begin with visualization and basic statistics
2. **Test Assumptions**: Stationarity, normality, independence
3. **Iterate**: Model $\rightarrow$ Diagnose $\rightarrow$ Improve
4. **Compare**: Never rely on a single model
5. **Validate**: Out-of-sample testing is essential
6. **Communicate**: Clear visualizations and interpretations

Remember

"All models are wrong, but some are useful." — George Box

The goal is not perfect prediction, but useful insights and reasonable forecasts.

Thank You!

Questions?

Course Materials: `github.com/danpele/Time-Series-Analysis`

Contact: daniel.pele@csie.ase.ro

logo_csie.png logo_ase.png