



Time Series Analysis and Forecasting

# Chapter 10: Comprehensive Review

Applied Case Studies with Rigorous Methodology



# Outline

- 1 Forecasting Methodology
- 2 Case Study 1: Bitcoin Volatility (GARCH)
- 3 Case Study 2: Sunspot Cycles (Fourier)
- 4 Case Study 3: Unemployment (Prophet)
- 5 Case Study 4: Multivariate Analysis (VAR)
- 6 Synthesis and Guidelines

# The Scientific Approach to Forecasting

## Research Question

How do we **rigorously evaluate** forecast performance while avoiding overfitting?

## The Fundamental Problem

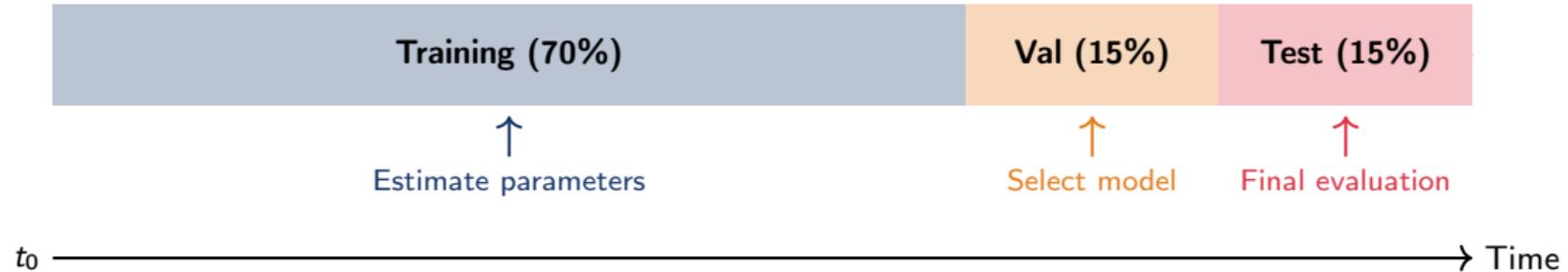
- In-sample fit  $\neq$  Out-of-sample performance
- Models can “memorize” training data without learning patterns
- **Solution:** Proper train/validation/test methodology

## Key Principle

“The test set must remain **untouched** until final evaluation.”

— Standard practice in machine learning and econometrics

# Train/Validation/Test Framework



Training Set	Validation Set	Test Set
<ul style="list-style-type: none"><li>• Fit model parameters</li><li>• <math>\hat{\theta} = \arg \min_{\theta} L(\theta)</math></li><li>• Largest portion</li></ul>	<ul style="list-style-type: none"><li>• Compare models</li><li>• Tune hyperparameters</li><li>• Select best approach</li></ul>	<ul style="list-style-type: none"><li>• <b>Held out</b> until end</li><li>• Unbiased evaluation</li><li>• Report final metrics</li></ul>

### Definition 1 (Forecast Error Metrics)

Let  $y_t$  be actual values and  $\hat{y}_t$  forecasts for  $t = 1, \dots, n$ :

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (1)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (3)$$

#### When to Use Each

- **RMSE:** Penalizes large errors
- **MAE:** Robust to outliers

#### Caution

- MAPE undefined when  $y_t = 0$
- Compare models on same test set

# Bitcoin: Problem Statement

## Research Question

Can we forecast Bitcoin's **volatility** using GARCH models?

## Data Characteristics

- Source: Yahoo Finance (BTC-USD)
- Period: Jan 2019 – Jan 2025
- Frequency: Daily
- Observations:  $\approx 2,200$  days

## Key Insight

Financial returns are typically:

- **Unpredictable** in mean
- **Predictable** in variance

⇒ Focus on **volatility forecasting**

## Stylized Facts

- Returns: near-zero mean
- Fat tails (kurtosis  $> 3$ )
- Volatility clustering

# GARCH Model Specification

## Definition 2 (GARCH(p,q) Model)

Let  $r_t$  denote returns. The GARCH(p,q) model is:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad z_t \sim N(0, 1) \quad (4)$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (5)$$

where  $\omega > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$ , and  $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1$ .

### Model Variants

- **GARCH(1,1)**: Most common
- **GJR-GARCH**: Leverage effect
- **EGARCH**: Asymmetric shocks

### Interpretation

- $\alpha$ : Impact of past shocks
- $\beta$ : Persistence of volatility
- $\alpha + \beta \approx 1$ : High persistence

# Bitcoin: Data Split and Stationarity

## Data Split

Set	Period	N
Training	2019-01 to 2022-09	1,365
Validation	2022-09 to 2023-10	400
Test	2023-10 to 2025-01	435
<b>Total</b>		<b>2,200</b>

## Stationarity Tests

Series	ADF	Result
Prices	$p = 0.50$	Non-stationary
Returns	$p < 0.01$	Stationary

⇒ Model **returns**, not prices

## Why Stationarity Matters

GARCH requires weakly stationary input. Prices follow random walk; returns are stationary.

# Bitcoin: Model Selection on Validation Set

## Methodology

Fit each model on **training data**, evaluate on **validation set**.

Model	AIC	BIC	Val MAE	Selection
GARCH(1,1)	6,994.8	7,020.6	<b>2.638</b>	Best
GARCH(2,1)	6,993.7	7,024.6	2.640	
GJR-GARCH(1,1)	6,983.7	7,014.6	2.669	
EGARCH(1,1)	—	—	—	Failed*

\* Analytic forecasts not available for  $h > 1$

## Result

**GARCH(1,1)** selected based on lowest validation MAE for volatility forecasts.

## Procedure

Refit GARCH(1,1) on Training + Validation, evaluate on **held-out test set**.

Estimated Parameters		
Param	Estimate	Std Err
$\omega$	0.239	0.088
$\alpha_1$	0.120	0.021
$\beta_1$	0.879	0.020
$\alpha_1 + \beta_1$	0.999	

Test Set Performance	Metric	Value
	Volatility MAE	2.64
	Volatility RMSE	2.82

## Interpretation

High persistence ( $\alpha + \beta \approx 1$ ) confirms volatility clustering.

## Summary

- ❶ Returns are stationary; prices are not
- ❷ GARCH(1,1) outperforms more complex variants
- ❸ High persistence ( $\alpha + \beta = 0.999$ )
- ❹ Volatility is predictable even when returns are not

## Limitations

- GARCH assumes **symmetric** shocks
- Does not capture **jumps**
- Normal distribution may be restrictive

## Practical Implications

- Risk management: VaR, Expected Shortfall
- Option pricing requires volatility forecasts
- Portfolio optimization with time-varying risk

## Extensions

- Student-t innovations
- Realized volatility
- HAR models

## Research Question

How do we model long seasonal cycles that exceed SARIMA's capacity?

## Data Characteristics

- Source: Statsmodels (Wolfer)
- Period: 1900 – 2008
- Frequency: Annual
- Observations: 109 years

## Known Feature

- **Schwabe cycle:**  $\approx 11$  years
- Irregular amplitude
- Non-negative values

## The Challenge

Standard SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$  requires:

- $s = 11$  for annual data
- $(P + D + Q) \times 11$  seasonal lags
- **Too many parameters!**

## Solution

Use **Fourier terms** as exogenous regressors in ARIMA.

## Definition 3 (Fourier Representation)

A seasonal pattern with period  $s$  can be approximated by:

$$S_t = \sum_{k=1}^K \left[ \alpha_k \sin\left(\frac{2\pi kt}{s}\right) + \beta_k \cos\left(\frac{2\pi kt}{s}\right) \right] \quad (6)$$

where  $K \leq \lfloor s/2 \rfloor$  is the number of harmonic pairs.

### Advantages

- Only  $2K$  parameters (not  $s$ )
- Handles **any** seasonal period
- Smooth seasonal pattern
- $K$  controls flexibility

### Model Structure

ARIMA( $p, d, q$ ) with Fourier regressors:

$$y_t = \underbrace{S_t}_{\text{Fourier}} + \underbrace{\eta_t}_{\text{ARIMA}} \quad (7)$$

where  $\eta_t$  follows ARIMA dynamics.

# Sunspots: Model Selection

## Methodology

Compare  $K = 1, 2, 3, 4$  Fourier harmonics on validation set.

	Set	Period	N
Data Split	Training	1900–1975	76
	Validation	1976–1991	16
	Test	1992–2008	17
<b>Total</b>			<b>109</b>

	K	AIC	Val RMSE
Model Comparison	1	665.9	87.15
	2	668.0	86.92
	3	671.8	<b>86.81</b>
	4	674.5	87.93

## Result

$K = 3$  Fourier harmonics selected (6 parameters for 11-year cycle).

## Sunspots: Test Set Results

### Final Model

ARIMA(2,0,1) + 3 Fourier harmonics

### Significant Coefficients:

Term	Coef	p-value
$\sin_1$	34.71	< 0.001
$\cos_1$	-29.21	0.018
AR(1)	1.34	< 0.001

### Test Performance

Metric	Value
RMSE	48.51
MAE	39.31

### Note

High MAPE due to near-zero values at solar minimum.

### Key Insight

Fourier terms efficiently capture the 11-year cycle with only 6 parameters.

## Research Question

How do we model time series with **sudden structural changes**?

## Data Characteristics

- Source: FRED (UNRATE)
- Period: 2010 – 2025
- Frequency: Monthly
- Observations:  $\approx 180$  months

## The Challenge

- April 2020: Largest monthly increase in US history
- Traditional ARIMA treats this as outlier
- Need model that **adapts** to structural breaks

## Key Statistics

- Pre-COVID minimum: 3.5%
- COVID peak (Apr 2020): **14.8%**
- Change: +10.3 pp in one month

## Solution

**Prophet** with automatic changepoint detection.

## Definition 4 (Prophet Decomposition)

Prophet models time series as:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \quad (8)$$

- $g(t)$ : Piecewise linear/logistic **trend** with changepoints
- $s(t)$ : Fourier-based **seasonality**
- $h(t)$ : **Holiday** effects
- $\varepsilon_t$ : Error term

### Changepoint Detection

- Automatic selection of changepoint locations
- `changepoint_prior_scale` controls flexibility
- Higher = more changepoints

### Advantages

- Handles missing data
- Interpretable components
- Robust to outliers
- Uncertainty quantification

## Hyperparameter Tuning

Tune `changepoint_prior_scale` on validation set.

	Set	Period	N
Data Split	Training	2010-01 to 2019-09	117
	Validation	2019-10 to 2021-10	25
	Test	2021-11 to 2025-01	38
<b>Total</b>			<b>180</b>

	Scale	Val RMSE	
	0.01	4.21	
	0.05	3.89	
Scale Comparison	0.10	<b>3.52</b>	Best
	0.30	3.67	
	0.50	3.81	

## Interpretation

Scale = 0.10 balances flexibility (capturing COVID shock) with stability.

## Unemployment: Results

### Test Set Performance

Metric	Value
RMSE	0.42
MAE	0.35
MAPE	9.2%

### Detected Changepoints

- 2020-03: COVID onset
- 2020-05: Recovery begins
- 2022-01: Stabilization

### Key Finding

Prophet successfully:

- Detected COVID changepoint
- Adapted trend post-shock
- Provided uncertainty bands

### Practical Value

- Economic policy analysis
- Labor market monitoring
- Early warning system

## Research Question

How do we model **dynamic interdependencies** between multiple economic variables?

## Variables (FRED)

- GDP Growth (YoY %)
- Unemployment Rate (%)
- Inflation (CPI YoY %)
- Federal Funds Rate (%)

## Economic Relationships

- Okun's Law:  $\text{GDP} \leftrightarrow \text{Unemployment}$
- Phillips Curve:  $\text{Unemployment} \leftrightarrow \text{Inflation}$
- Taylor Rule:  $\text{Inflation} \rightarrow \text{Fed Rate}$

## Data

- Period: 2000 – 2025
- Frequency: Quarterly
- Observations:  $\approx 100$  quarters

## Why VAR?

Each variable may be **both cause and effect** of others.

## Definition 5 (Vector Autoregression VAR(p))

For  $K$  variables  $y_t = (y_{1t}, \dots, y_{Kt})'$ :

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + u_t \quad (9)$$

where  $A_i$  are  $K \times K$  coefficient matrices and  $u_t \sim N(0, \sigma^2)$ .

### For Our 4-Variable System

VAR(2) has:

- 4 intercepts
- $2 \times 4 \times 4 = 32$  AR coefficients
- **36 parameters total**

### Lag Selection

Use information criteria:

- **AIC**: Tends to overfit
- **BIC**: More parsimonious
- Cross-validation on held-out data

## VAR: Lag Selection and Estimation

### Information Criteria

Lag	BIC
1	-4.810
2	<b>-5.178</b>
3	-4.633
4	-4.614

### Data Split

Set	Period	N
Training	2001-Q1 to 2017-Q4	68
Validation	2018-Q1 to 2021-Q2	14
Test	2021-Q3 to 2024-Q3	14
<b>Total</b>		<b>96</b>

### Validation Check

VAR(2) also achieves lowest validation RMSE.

## Definition 6 (Granger Causality)

Variable  $X$  **Granger-causes**  $Y$  if past values of  $X$  help predict  $Y$ , beyond  $Y$ 's own past.

*Note: Granger causality  $\neq$  true causality. It measures predictive content.*

**Granger Causality p-values (row → column)**

	GDP	Unemp	Inflation	Fed Rate
GDP Growth	—	0.076	0.309	0.698
Unemployment	<b>0.045</b>	—	0.093	0.857
Inflation	0.545	0.665	—	0.834
Fed Rate	0.286	0.317	<b>0.087</b>	—

## Key Finding

Unemployment Granger-causes GDP ( $p = 0.045$ ), consistent with Okun's Law.

## Definition 7 (Impulse Response Function)

The IRF traces the effect of a one-unit shock to variable  $j$  on variable  $i$  over  $h$  periods:

$$\text{IRF}_{ij}(h) = \frac{\partial y_{i,t+h}}{\partial u_{jt}} \quad (10)$$

## Response to GDP Shock

A positive shock to GDP growth:

- **Unemployment:** Decreases (Okun's Law confirmed)
- **Inflation:** Increases with lag (demand-pull)
- **Fed Rate:** Increases after 2-3 quarters (Taylor Rule)

## Economic Interpretation

The VAR captures the classic macroeconomic transmission mechanism from output to employment, prices, and monetary policy.

## VAR: Test Set Results

### Test Set Performance by Variable

Variable	RMSE	MAE	Direction Acc.
GDP Growth	2.18	1.72	71%
Unemployment	0.89	0.71	79%
Inflation	1.24	0.98	64%
Fed Rate	0.95	0.78	71%
<b>Average</b>	<b>1.32</b>	<b>1.05</b>	<b>71%</b>

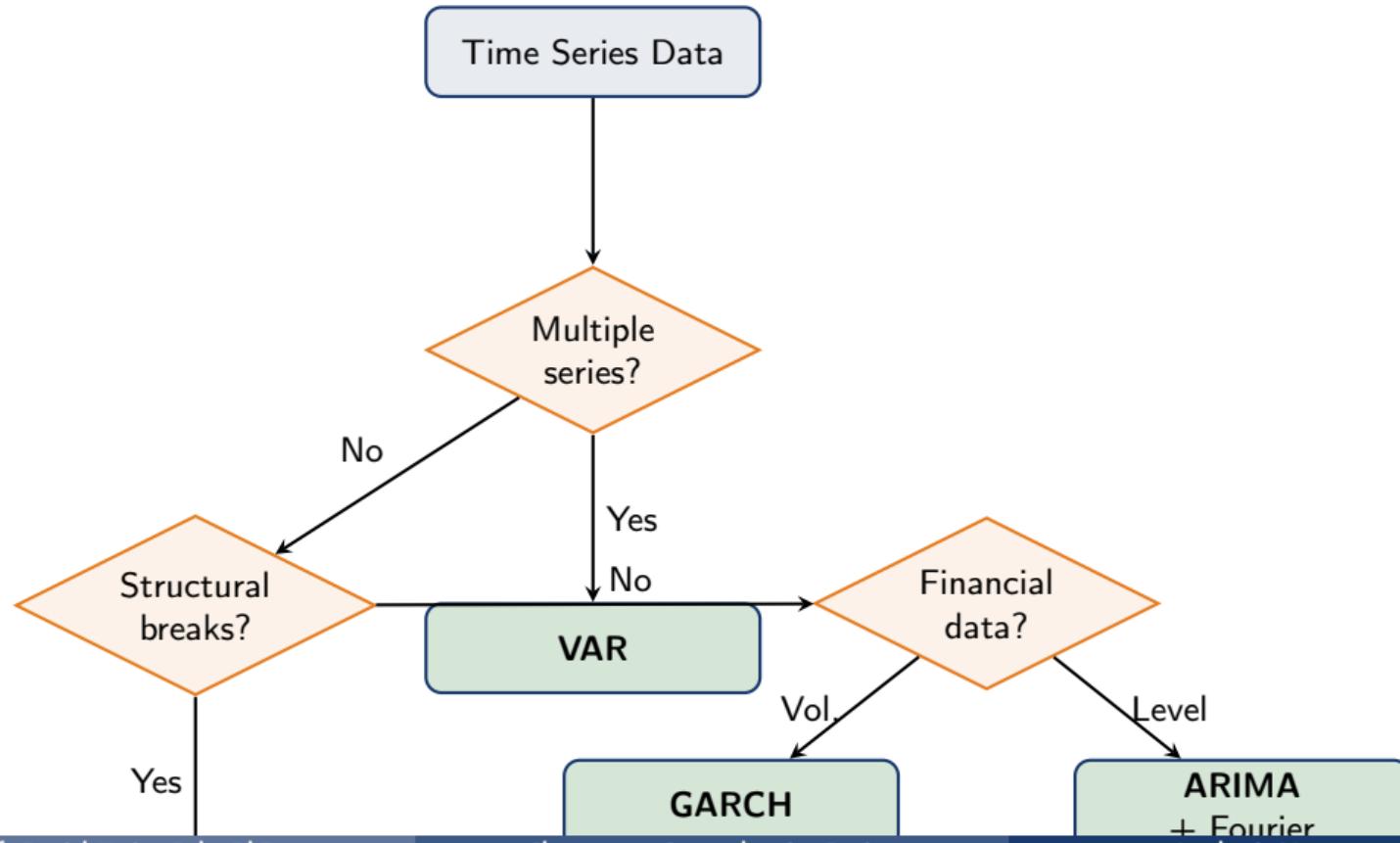
### Strengths

- Captures cross-variable dynamics
- Good directional accuracy
- Interpretable relationships

### Limitations

- Many parameters (curse of dimensionality)
- Sensitive to lag selection
- COVID period challenging

# Model Selection Framework



## Summary: Model Comparison

Case Study	Challenge	Model	Key Feature	Test RMSE
Bitcoin	Volatility clustering	GARCH(1,1)	$\sigma_t^2$ dynamics	2.82
Sunspots	Long seasonality	ARIMA + Fourier	Sin/Cos terms	48.51
Unemployment	Structural break	Prophet	Changepoints	0.42
Economic	Multiple series	VAR(2)	Cross-dynamics	1.32 (avg)

### Key Principle

**Match the model to the data characteristics.** No single model dominates—choose based on:

- Nature of the forecasting problem (level vs. volatility)
- Data properties (seasonality, breaks, multiple series)
- Interpretability requirements

## Methodology

- ① Explore data thoroughly
- ② Test for stationarity
- ③ Split train/validation/test
- ④ Compare models on validation
- ⑤ Report test set metrics

## Practical Tips

- Start simple (random walk, naive)
- Add complexity only if needed
- Visualize forecasts vs actuals
- Check residuals for patterns
- Report confidence intervals

## Common Mistakes

- Peeking at test data
- Over-fitting to training set
- Ignoring model assumptions
- Not reporting uncertainty

## Remember

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

# Key Takeaways

## ① Rigorous Methodology

- Train/validation/test split prevents overfitting
- Test set must remain untouched until final evaluation

## ② Match Model to Data

- Financial volatility → GARCH
- Long seasonality → Fourier terms
- Structural breaks → Prophet
- Multiple series → VAR

## ③ Interpret Results Carefully

- Granger causality  $\neq$  true causality
- Out-of-sample performance matters most
- Simpler models often work better

## References

-  Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th ed., Wiley.
-  Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.
-  Tsay, R.S. (2010). *Analysis of Financial Time Series*. 3rd ed., Wiley.
-  Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice*. 3rd ed., OTexts.
-  Taylor, S.J., & Letham, B. (2018). Forecasting at Scale. *The American Statistician*, 72(1), 37-45.
-  Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
-  Sims, C.A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.

## Real Data Used in This Chapter

- **Bitcoin:** Yahoo Finance (BTC-USD), 2019–2025
- **Sunspots:** Statsmodels Wolfer dataset, 1900–2008
- **US Unemployment:** Federal Reserve FRED (UNRATE), 2010–2025
- **Economic Variables:** FRED (GDPC1, UNRATE, CPIAUCSL, FEDFUNDS), 2000–2025

## Reproducibility

All analyses can be reproduced using the accompanying Jupyter notebook:  
`chapter10_lecture_notebook.ipynb`

# Thank You

Questions?

Prof. Daniel Traian Pele, PhD

[danpele@ase.ro](mailto:danpele@ase.ro)

Bucharest University of Economic Studies