



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

## Time Series Train/Validation/Test Split



Training Set	Validation Set	Test Set
<ul style="list-style-type: none"><li>• Fit parameters</li><li>• Largest portion</li></ul>	<ul style="list-style-type: none"><li>• Compare models</li><li>• Tune hyperparams</li></ul>	<ul style="list-style-type: none"><li>• <b>Held out</b></li><li>• Final metrics</li></ul>

## Definition 1 (Forecast Error Metrics)

Let  $y_t$  be actual,  $\hat{y}_t$  forecast:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_t (y_t - \hat{y}_t)^2}, \quad \text{MAE} = \frac{1}{n} \sum_t |y_t - \hat{y}_t|, \quad \text{MAPE} = \frac{100\%}{n} \sum_t \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

### When to Use Each

- **RMSE**: Penalizes large errors
- **MAE**: Robust to outliers
- **MAPE**: Scale-independent (%)

### Caution

- MAPE undefined when  $y_t = 0$
- Compare on **same** test set
- Report **out-of-sample** metrics

# 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:

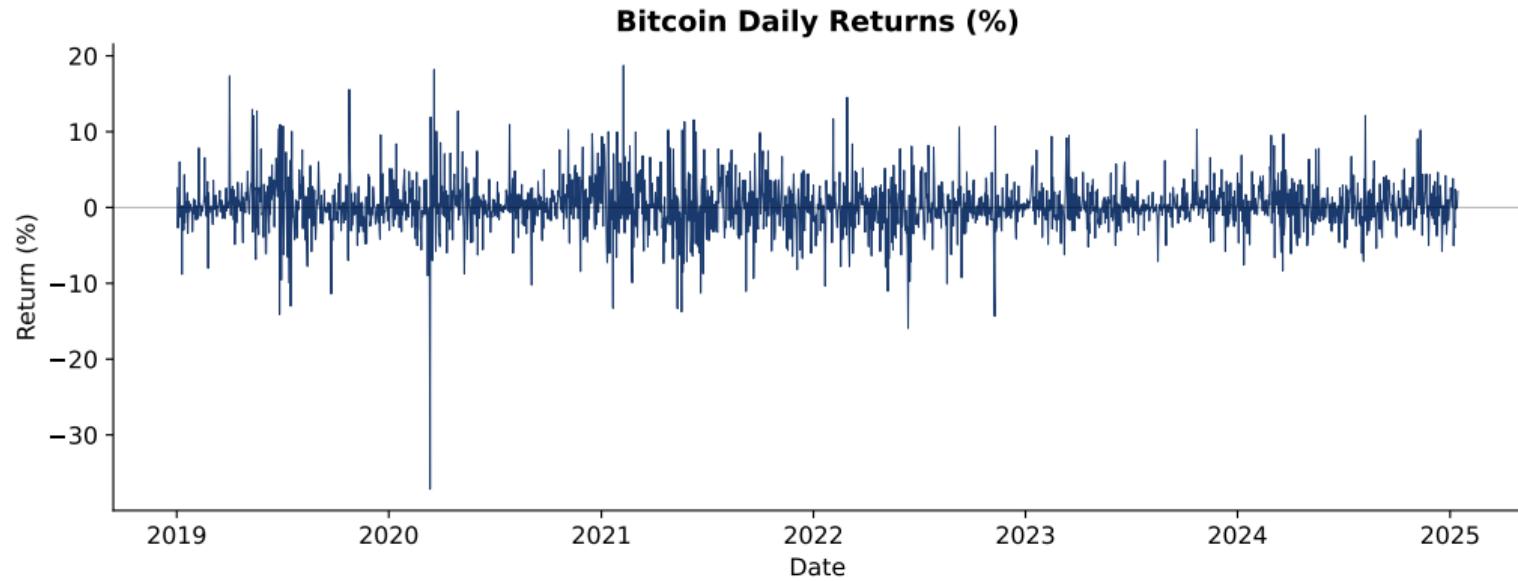
- **Unpredictable** in mean
- **Predictable** in variance

⇒ Focus on **volatility forecasting**

## Stylized Facts

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

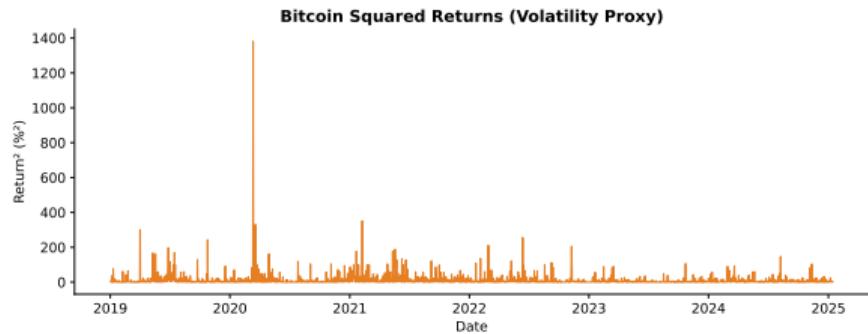
## Bitcoin: Volatility Clustering



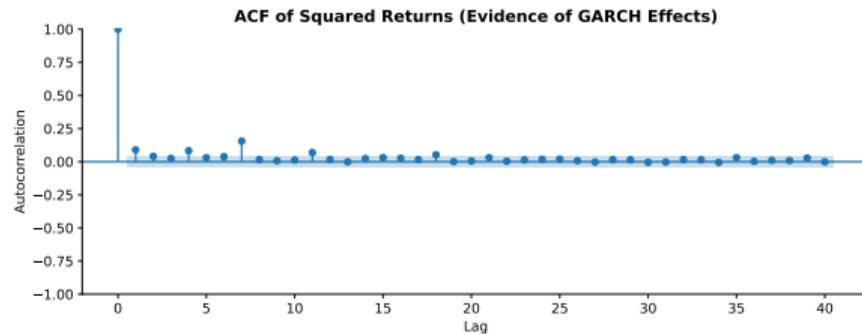
### Observation

Large returns tend to follow large returns, small follow small. This is **volatility clustering**—the phenomenon GARCH captures.

# Bitcoin: Evidence for GARCH



Squared returns  $r_t^2$  proxy for volatility  $\sigma_t^2$ . Spikes cluster together.



ACF bars exceed blue bands  $\Rightarrow$  significant autocorrelation at multiple lags.

## Why GARCH?

If  $r_t^2$  were white noise, ACF would be zero. Significant ACF means **past volatility predicts future volatility**—GARCH captures this!

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

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

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

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

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 (70%)	2019-01 to 2023-03	1,543
Validation (20%)	2023-03 to 2024-06	441
Test (10%)	2024-06 to 2025-01	221
<b>Total</b>	<b>2,205</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**.

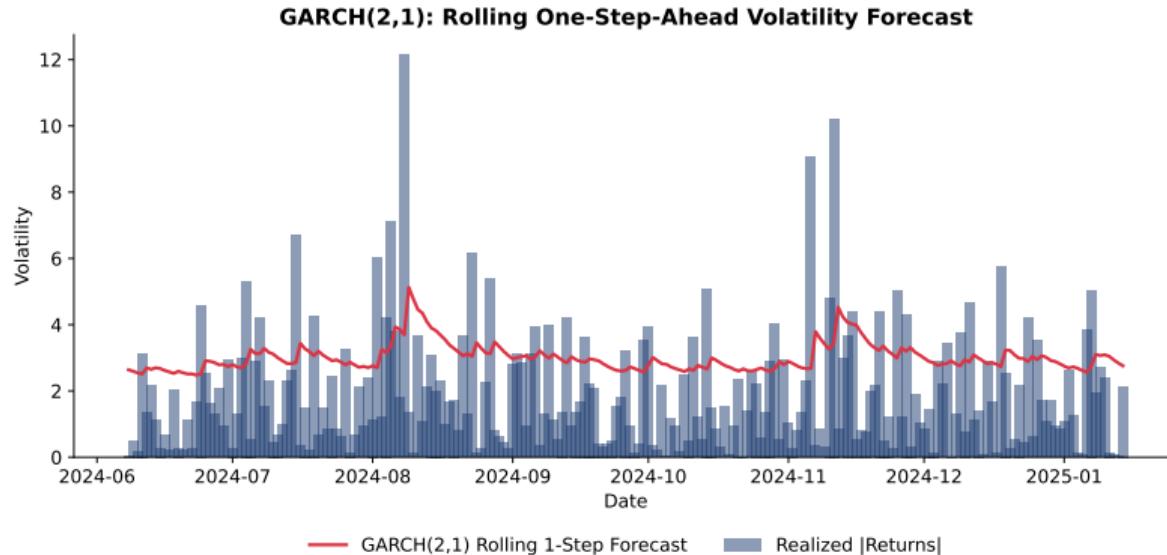
Model	AIC	BIC	Val MAE	Selection
GARCH(1,1)	6,994.8	7,020.6	<b>2.638</b>	Best
GARCH(2,1)	6,993.7	7,024.6	2.640	
GJR-GARCH(1,1)	6,983.7	7,014.6	2.669	
EGARCH(1,1)	—	—	—	Failed*

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

## Result

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

## Bitcoin: Final Test Set Evaluation



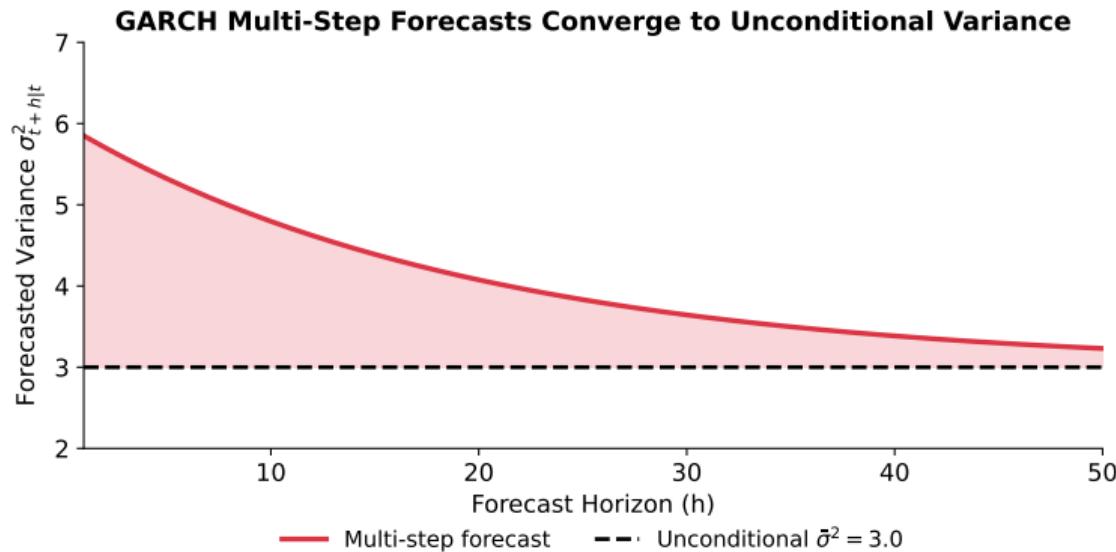
### Parameters

$\omega = 0.87$ ,  $\alpha = 0.09$ ,  $\beta = 0.84$   
 $\alpha + \beta = 0.93$  (high persistence)

### Test Performance

MAE = 1.82, RMSE = 2.14  
Forecast tracks realized volatility well.

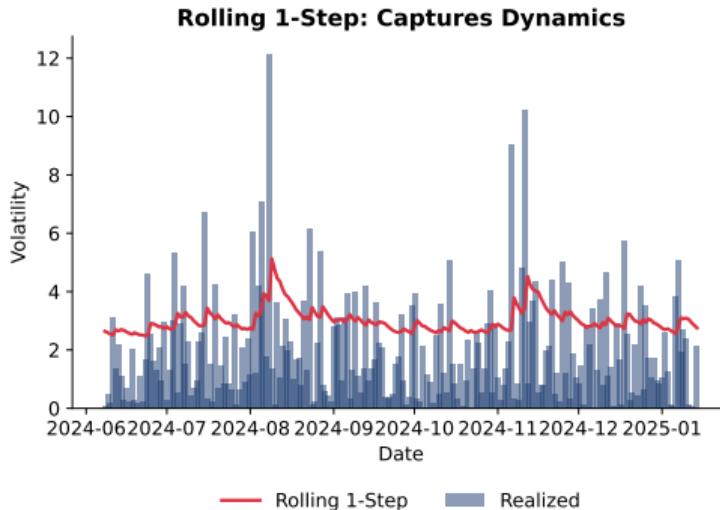
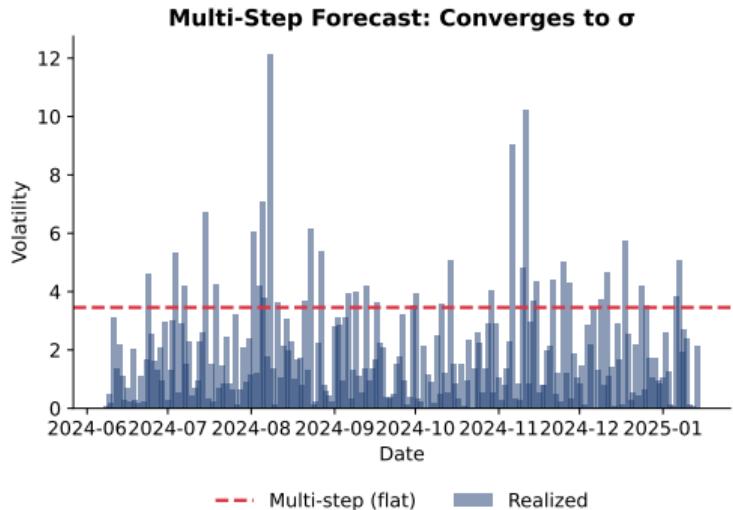
## GARCH: Multi-Step Forecasts Converge



### Key Insight

Multi-step forecasts converge to  $\bar{\sigma}^2 = \frac{\omega}{1-\alpha-\beta}$ . Solution: rolling one-step-ahead forecasts.

# GARCH: Rolling One-Step-Ahead Solution



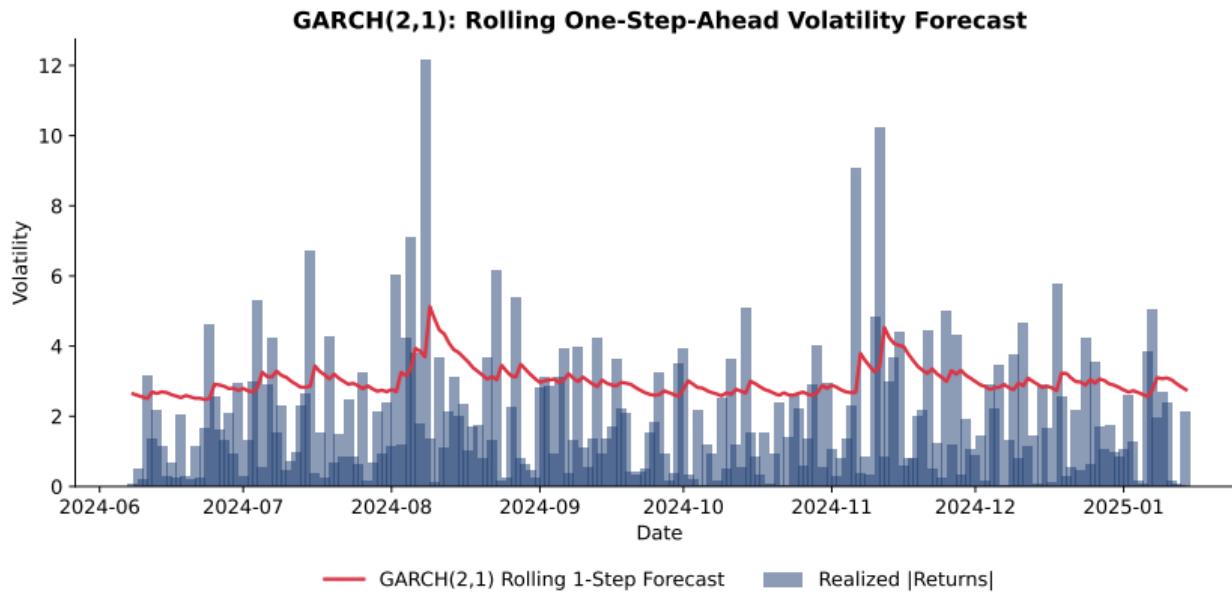
Multi-Step (Left)

Converges to  $\bar{\sigma}^2$  (flat)

Rolling 1-Step (Right)

Re-estimate at each  $t$  (dynamic)

## Bitcoin: GARCH Volatility Forecast (Test Set)



### Result

Rolling one-step-ahead GARCH(1,1) forecasts capture **dynamic volatility patterns**. Red line tracks realized volatility (blue area).

## Summary

- ❶ Returns are stationary; prices are not
- ❷ GARCH(1,1) outperforms more complex variants
- ❸ High persistence ( $\alpha + \beta = 0.93$ )
- ❹ Volatility is predictable even when returns are not

## Limitations

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

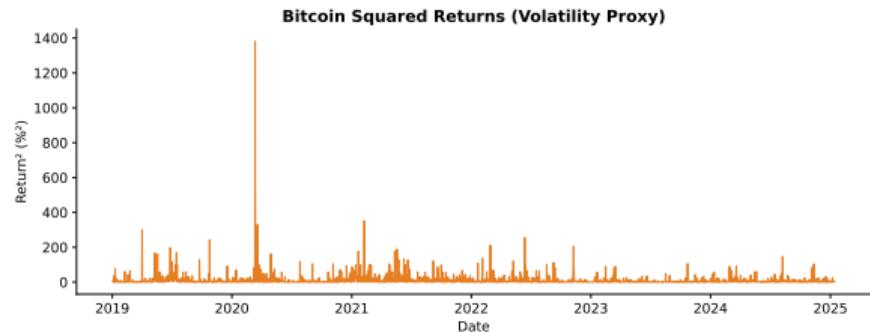
## Practical Implications

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

## Extensions

- Student-t innovations
- Realized volatility
- HAR models

# Bitcoin: GARCH Stylized Facts



Squared returns  $r_t^2$  as volatility proxy. Note the clustering of high-volatility periods.

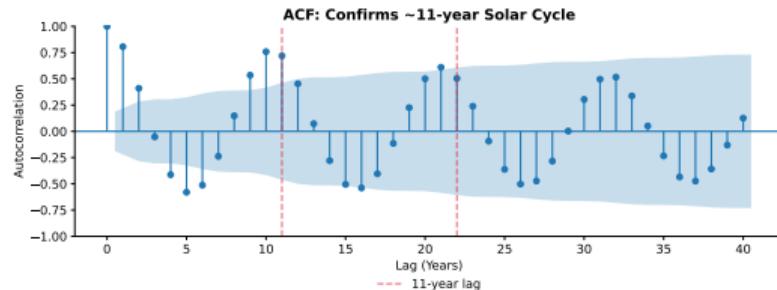
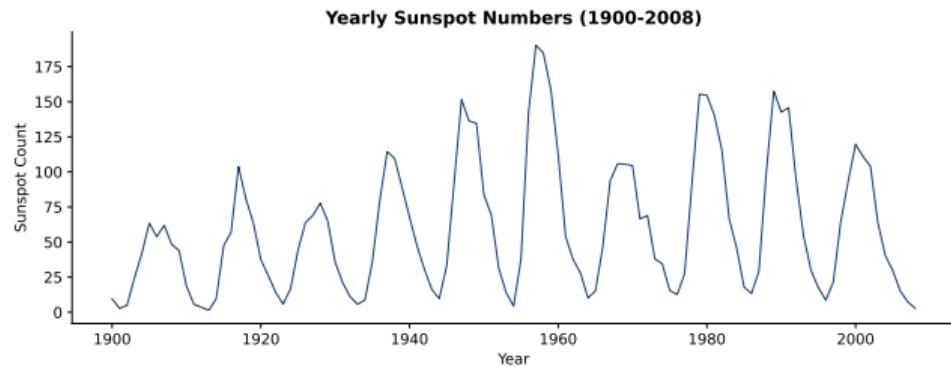
## Financial Stylized Facts

- ① **Volatility clustering:** Large moves follow large moves
- ② **Fat tails:** More extreme events than Normal predicts
- ③ **Leverage effect:** Negative returns → higher volatility
- ④ **Mean reversion:** Volatility returns to long-run level

## Why GARCH Works

GARCH captures facts 1 & 4. For fact 3, use GJR-GARCH or EGARCH. For fact 2, use Student-t innovations.

# Sunspots: The 11-Year Solar Cycle



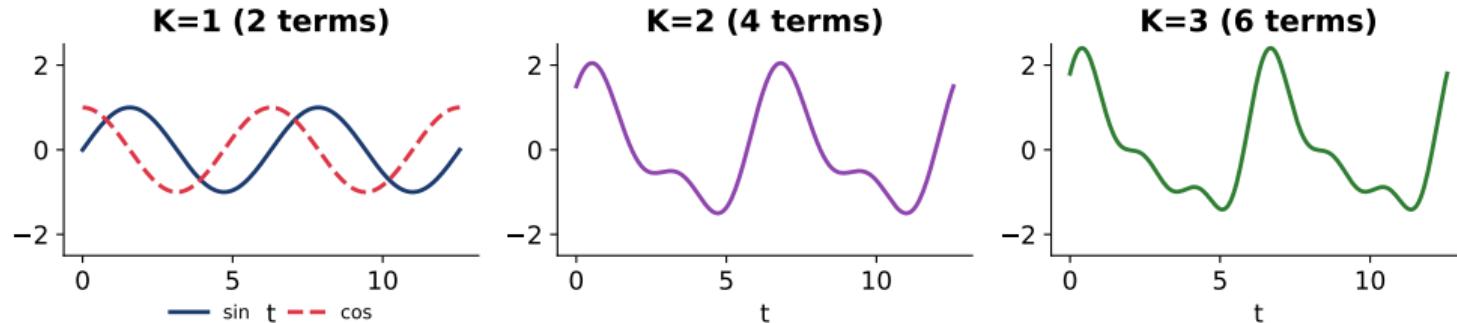
Dashed lines mark cycle peaks ( $\approx$  every 11 years). Amplitude varies.

## Challenge

SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>11</sub> requires estimating seasonal lags at 11, 22, 33... Too many parameters! **Solution:** Use Fourier terms instead.

# Fourier Terms for Seasonality

## Fourier Terms: More K = More Flexibility



### How It Works

Approximate any periodic pattern using sine and cosine waves:  $S_t = \sum_{k=1}^K [\alpha_k \sin\left(\frac{2\pi k t}{s}\right) + \beta_k \cos\left(\frac{2\pi k t}{s}\right)]$

### Key Insight

- $K = 1$ : Simple wave (2 params)
- $K = 3$ : Complex shape (6 params)
- For sunspots:  $s = 11$ ,  $K = 3$

# Sunspots: Model Selection

## Methodology

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

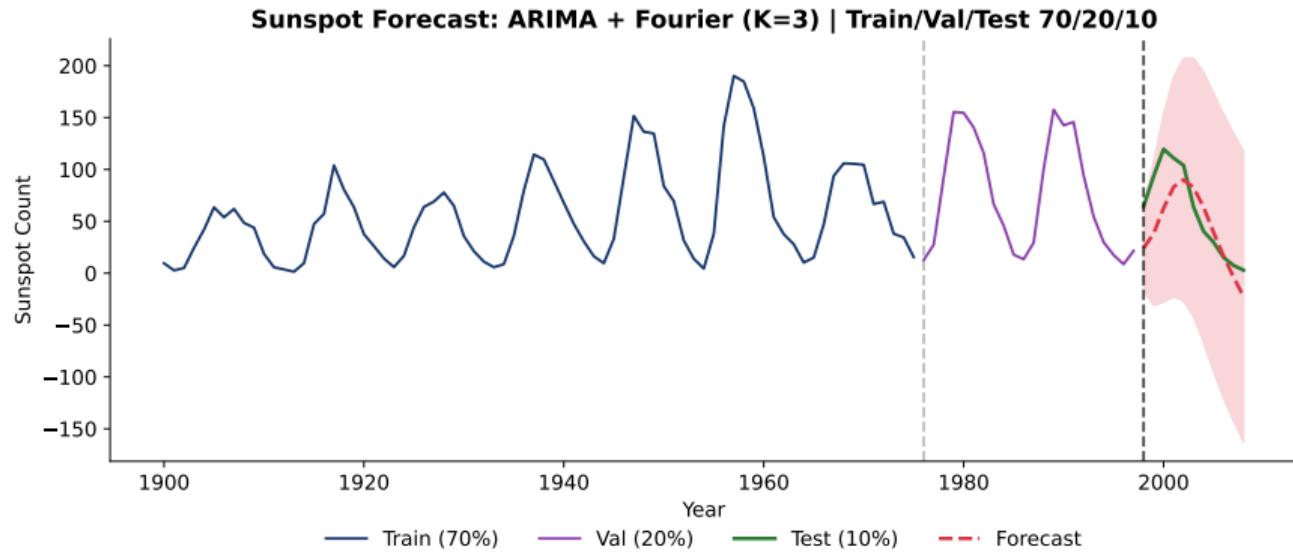
	Set	Period	N
Data Split	Training (70%)	1900–1975	76
	Validation (20%)	1976–1997	22
	Test (10%)	1998–2008	11
<b>Total</b>			<b>109</b>

	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: Forecast Results



### Model

ARIMA(2,0,1) + 3 Fourier terms captures the 11-year cycle dynamics.

### Test Performance

RMSE = 31.10, MAE = 25.83. The model tracks the overall cycle pattern.

# Sunspots: Key Takeaways

## When to Use Fourier Terms

- Seasonal period  $s$  is **long** (e.g., 11 years, 52 weeks)
- SARIMA would require too many seasonal lags
- Pattern is **smooth and periodic**
- Multiple cycles need to be captured

## Fourier vs SARIMA

	Fourier	SARIMA
Long seasons	✓	✗
Short seasons	OK	✓
Parameters	$2K$	Many
Flexibility	Fixed	Adaptive

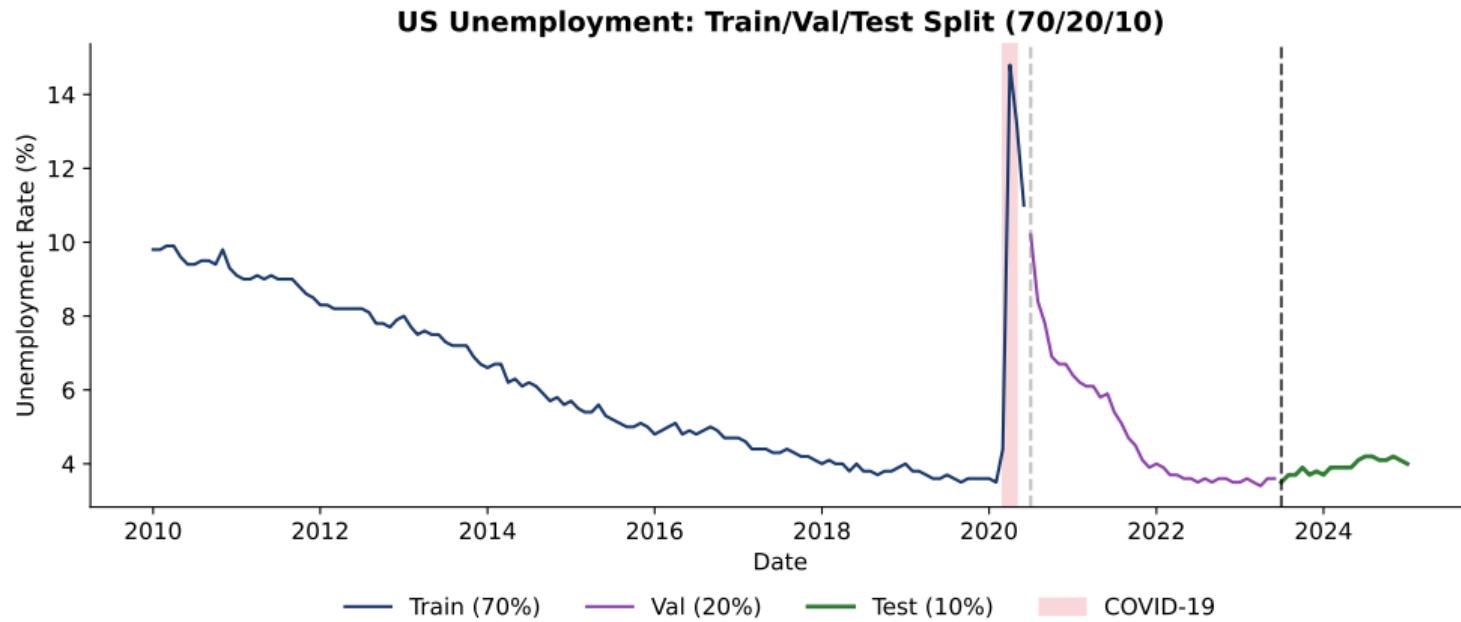
## Choosing K

Start with  $K = 1$ , increase until validation error stops improving. Too high  $K$  = overfitting.

## Applications

Climate cycles, business cycles, astronomical phenomena

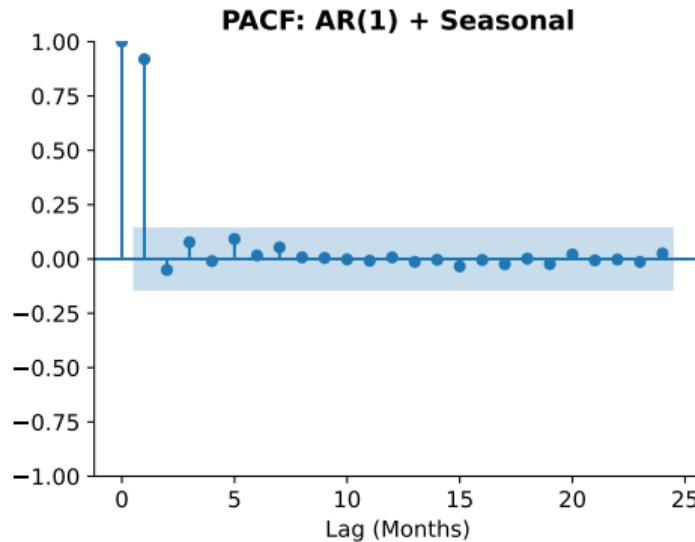
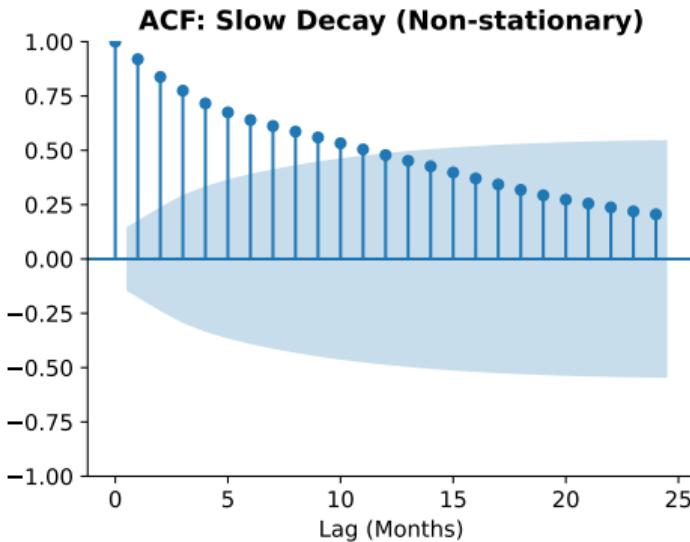
## Unemployment: Train / Validation / Test Split



### Methodology

**Training (70%):** Fit models. **Validation (20%):** Select best model. **Test (10%):** Final evaluation.

# Unemployment: Preliminary Analysis



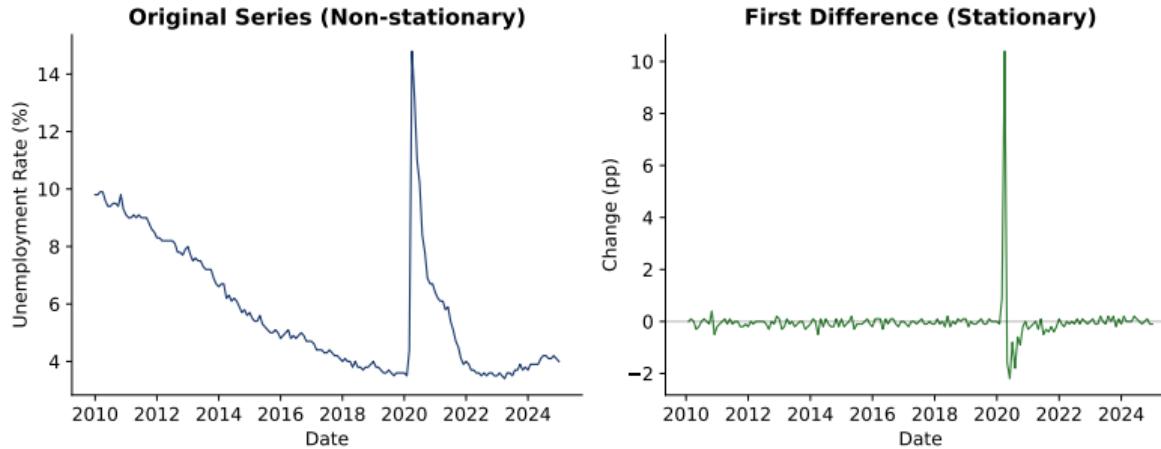
## ACF Interpretation

Slow decay  $\Rightarrow$  non-stationary series. Need differencing ( $d \geq 1$ ).

## PACF Interpretation

Significant spike at lag 1 suggests AR(1) component.  
Seasonal pattern at lag 12.

## Unemployment: Stationarity Tests



Original: ADF  $p = 0.056$

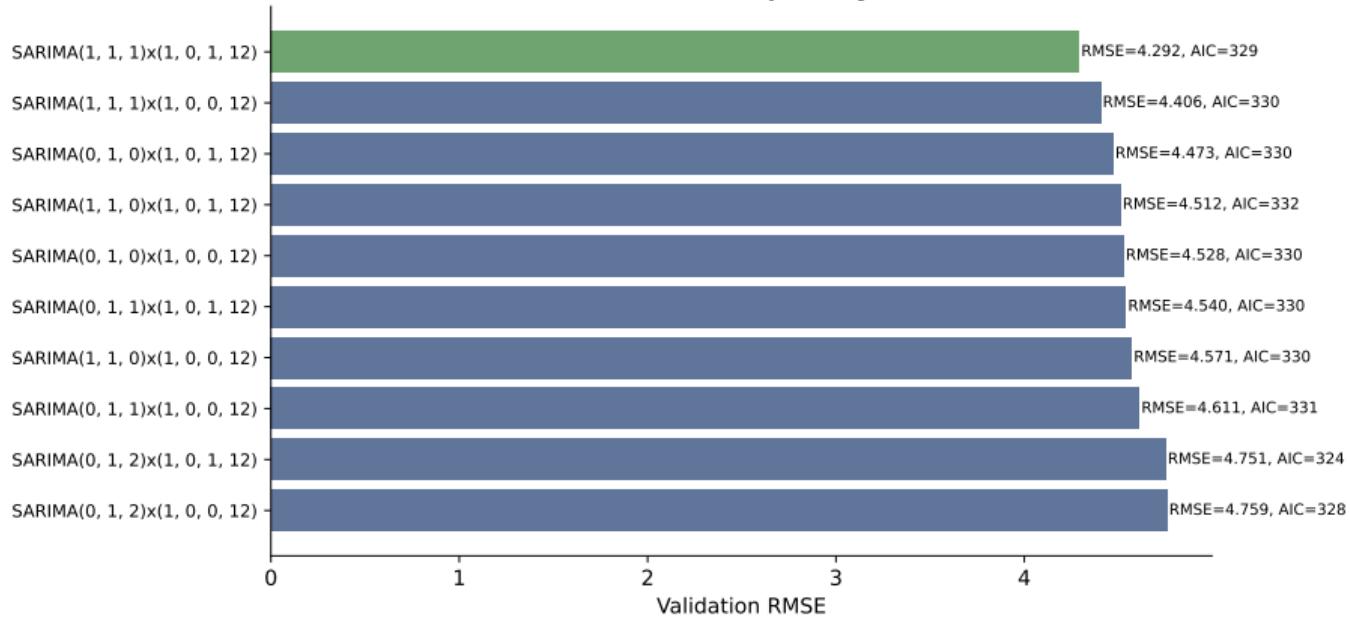
Non-stationary (slow ACF decay)

Differenced: ADF  $p < 0.001$

Stationary  $\Rightarrow$  use  $d = 1$

## Unemployment: Model Selection (Validation Set)

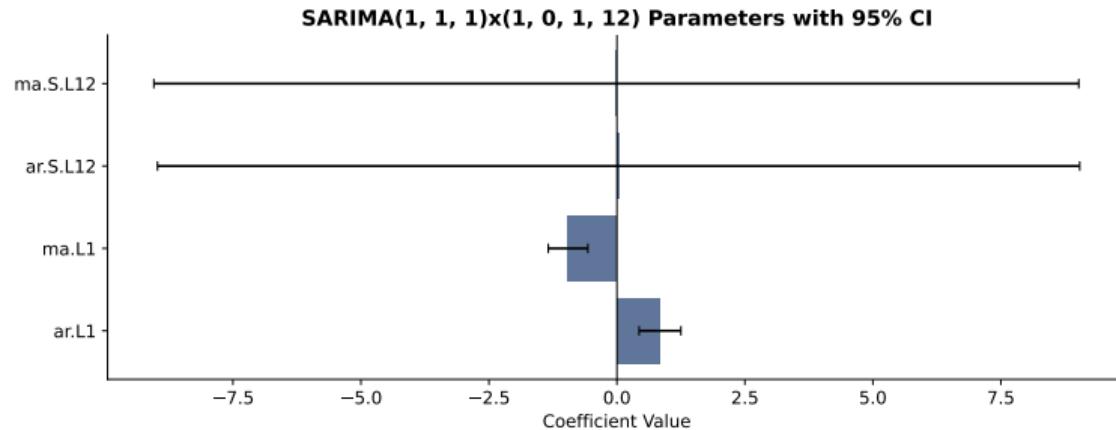
SARIMA Model Selection: Top 10 by Validation RMSE



Best: SARIMA(1,1,1)(1,0,0)<sub>12</sub>

Fit on training (70%), evaluate on validation (20%). Best model selected by lowest Val RMSE.

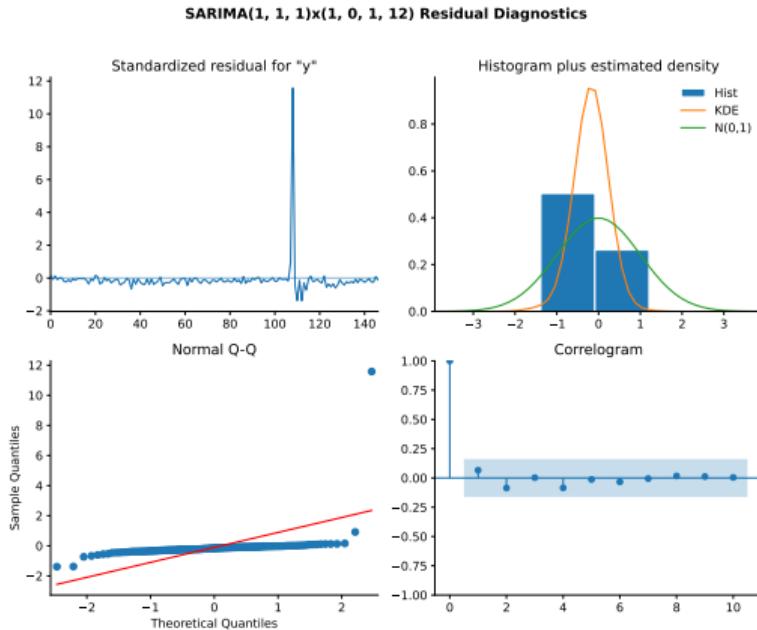
## Unemployment: SARIMA Parameters



SARIMA(1,1,1)(1,0,0)<sub>12</sub> fitted on Train+Val (2010-2019)

AR(1):  $\phi_1 = -0.86$ , MA(1):  $\theta_1 = 0.78$ , SAR(12):  $\Phi_1 = -0.08$  (n.s.)

# Unemployment: SARIMA Diagnostics



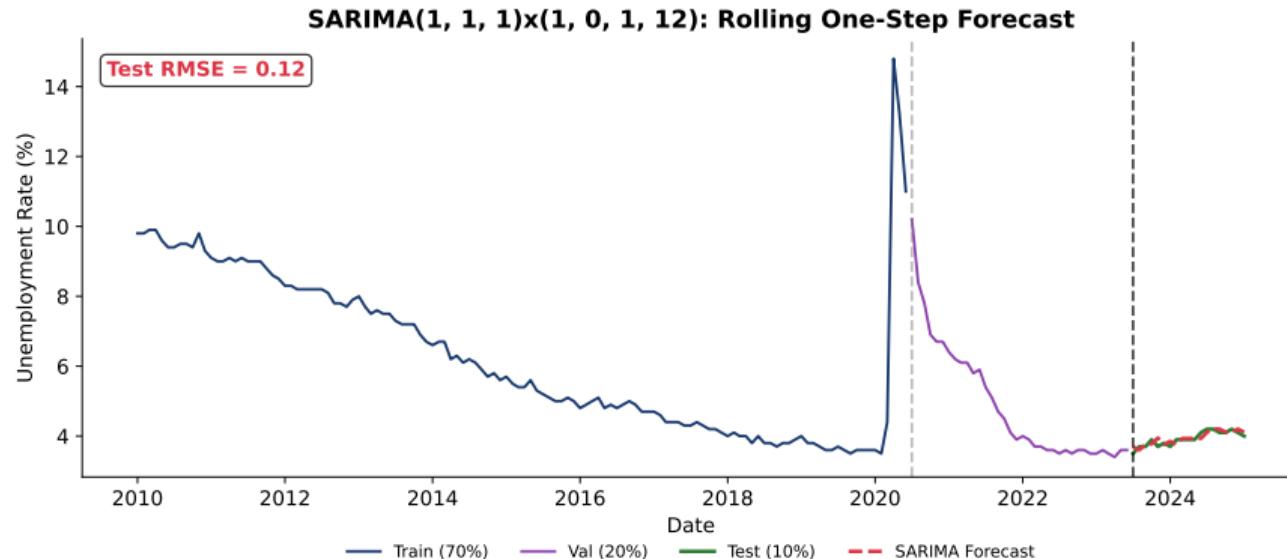
## Residuals

Std. residuals, histogram, ACF, Q-Q plot.

Ljung-Box  $p = 0.66$

No autocorrelation. Model well-specified.

## Unemployment: SARIMA Rolling Forecast



### Problem: Structural Break

Rolling one-step-ahead forecast (re-estimate at each  $t$ ): **Test RMSE = 0.12.**

## Definition 3 (Prophet Decomposition)

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2)$$

where  $g(t)$  = trend,  $s(t)$  = seasonality,  $h(t)$  = holidays,  $\sigma^2$  = noise variance (estimated).

### Changepoint Detection

- Automatic location selection
- `changepoint_prior_scale` controls flexibility

### Advantages

- Handles missing data
- Interpretable components
- Robust to outliers

# Unemployment: Model Tuning

## Hyperparameter Tuning

Tune `changepoint_prior_scale` on validation set.

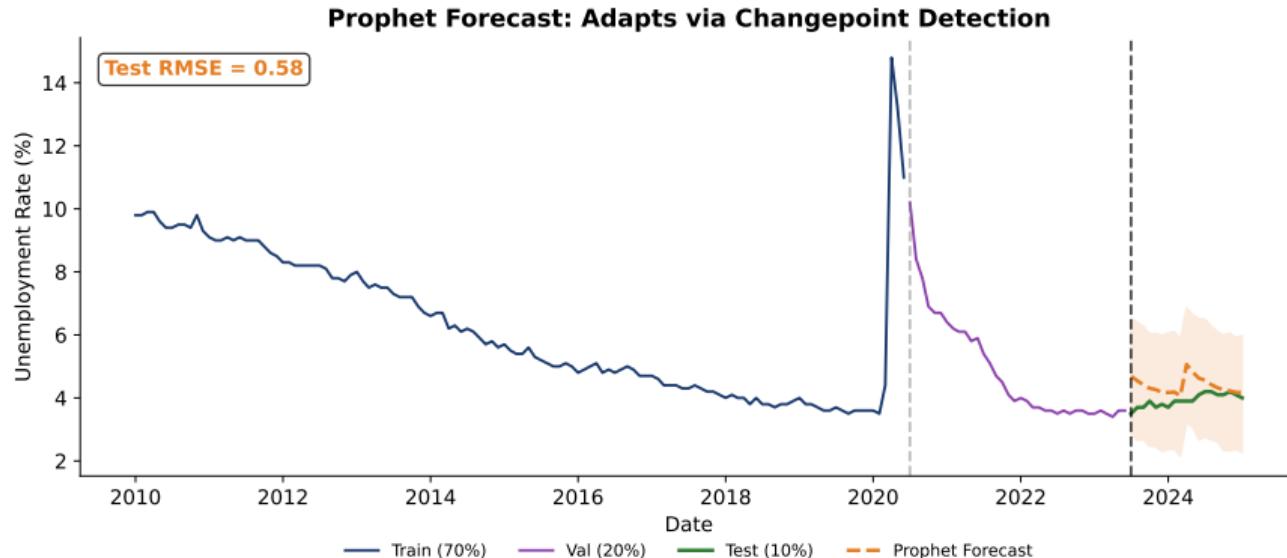
Data Split		
Set	Period	N
Training (70%)	2010-01 to 2020-06	126
Validation (20%)	2020-07 to 2023-06	36
Test (10%)	2023-07 to 2025-01	19
<b>Total</b>		<b>181</b>

Scale Comparison	Scale	Val RMSE	
	0.01	4.21	
	0.05	3.89	
	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.

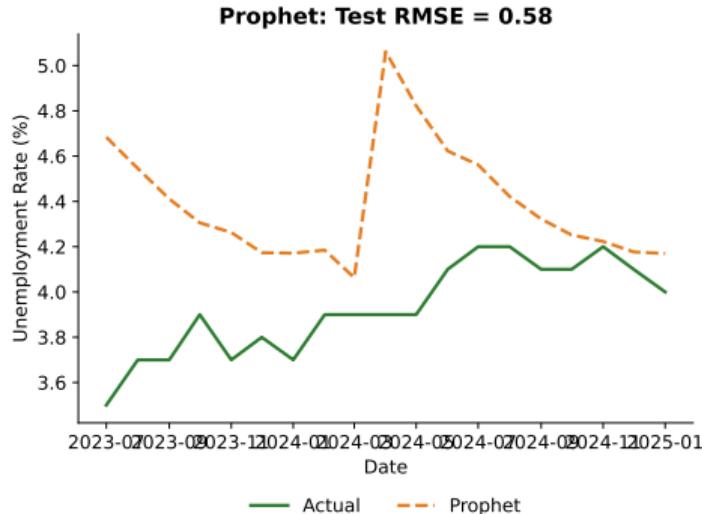
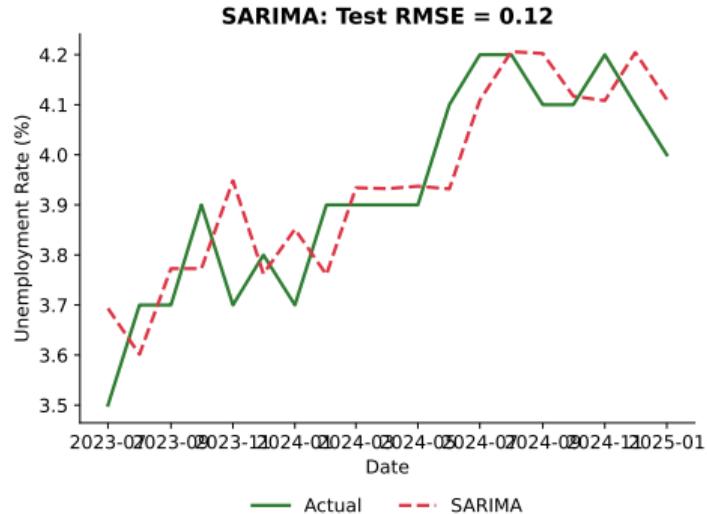
# Unemployment: Prophet Forecast Results



## Key Finding

Prophet adapts via changepoint detection. **Test RMSE = 0.58.**

# Unemployment: SARIMA vs Prophet Comparison



SARIMA: RMSE = 0.12

Rolling forecast performs well.

Prophet: RMSE = 0.58

Higher error due to structural break.

# Prophet: When to Use It

## Ideal Use Cases

- Business data with **holidays**
- **Missing values** present
- Need **interpretable** components
- Forecasts with **uncertainty bands**

## Prophet vs ARIMA

	Prophet	ARIMA
Changepoints	✓	✗
Missing data	✓	✗
Holidays	✓	✗
Speed	Fast	Moderate
Interpretable	✓	✗

## Caveat: Structural Breaks

Prophet handles breaks via changepoints, but **SARIMA outperformed** it on unemployment (0.12 vs 0.58).

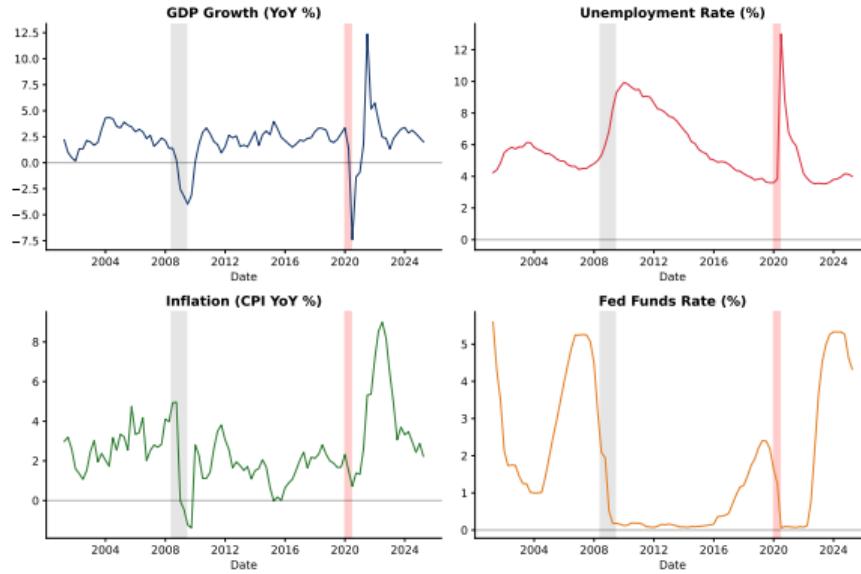
Always validate!

## Key Parameters

`changepoint_prior_scale`: flexibility

`seasonality_prior_scale`: smoothness

# VAR: Multivariate Economic Data



## Economic Relationships

**Okun's Law:**  $\text{GDP} \leftrightarrow \text{Unemployment}$ .

**Phillips Curve:**  $\text{Unemployment} \leftrightarrow \text{Inflation}$ .

## Why VAR?

Each variable is both cause and effect. VAR captures these feedback loops.

## Definition 4 (Vector Autoregression VAR(p))

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

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

where  $A_i$  are  $K \times K$  coefficient matrices,  $u_t \sim N(0, \Sigma)$ ,  $\Sigma$  = covariance matrix.

### 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

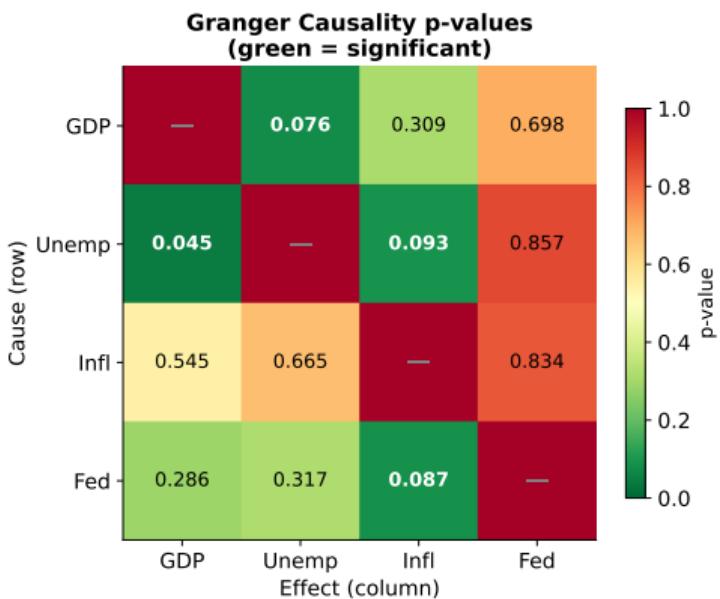
### Data Split

Set	Period	N
Training (70%)	2001-Q1 to 2017-Q4	67
Validation (20%)	2018-Q1 to 2022-Q4	20
Test (10%)	2023-Q1 to 2025-Q1	10
<b>Total</b>		<b>97</b>

### Validation Check

VAR(2) also achieves lowest validation RMSE.

# Granger Causality Analysis



## What is Granger Causality?

$X$  Granger-causes  $Y$  if past  $X$  improves prediction of  $Y$  beyond past  $Y$  alone.

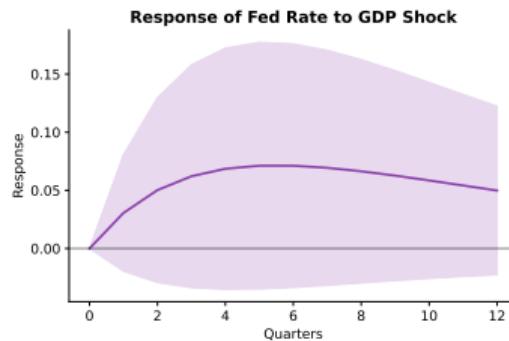
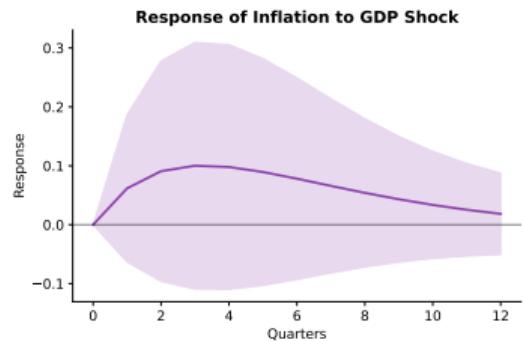
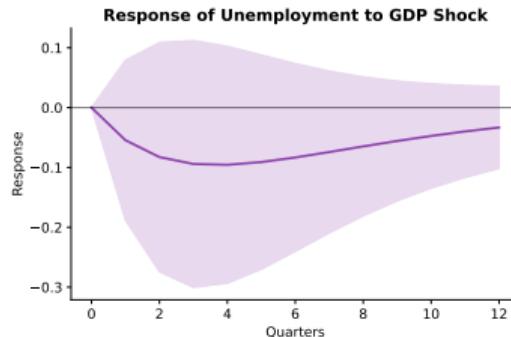
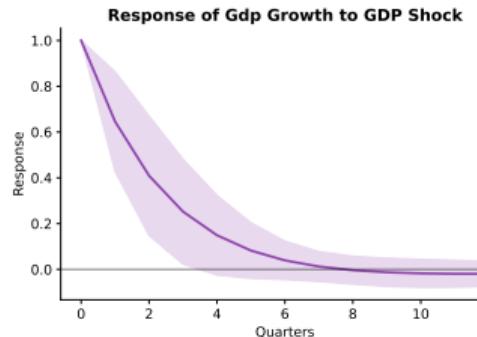
*Warning:* “Granger causality”  $\neq$  true causality!

## Economic Findings

- Unemp  $\rightarrow$  GDP ( $p = 0.045$ ): Okun's Law
- Fed  $\rightarrow$  Inflation ( $p = 0.087$ ): Monetary policy works

Green cells:  $p < 0.10$  (significant). Read: row causes column.

# Impulse Response Functions (IRF)



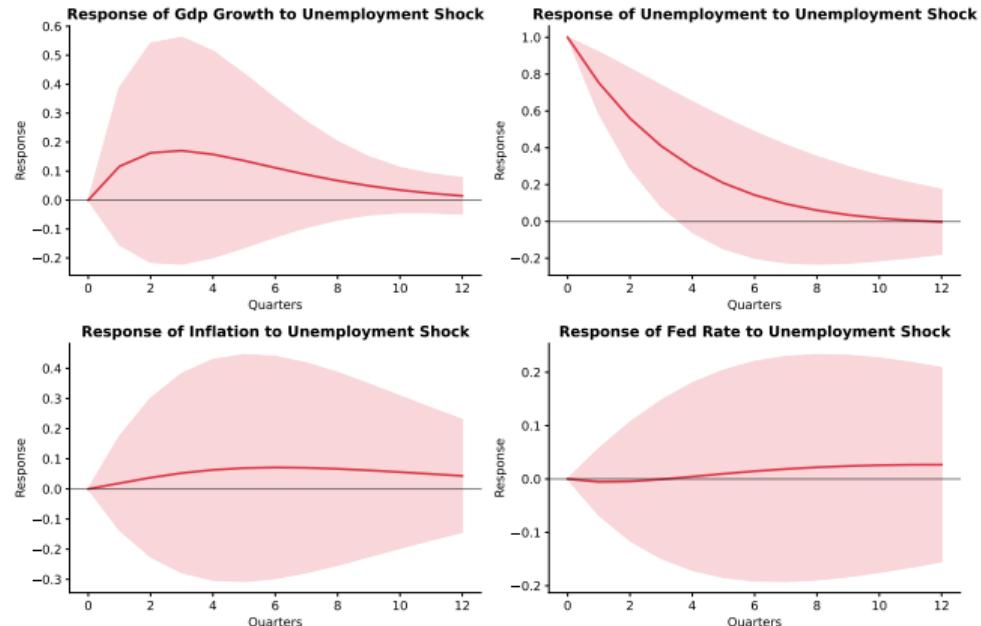
## What is IRF?

Shows how a 1-unit shock to one variable affects others over time.

## GDP Shock Effects

- **Unemp** ↓: Okun's Law
- **Inflation** ↑: Demand-pull
- **Fed Rate** ↑: Taylor Rule

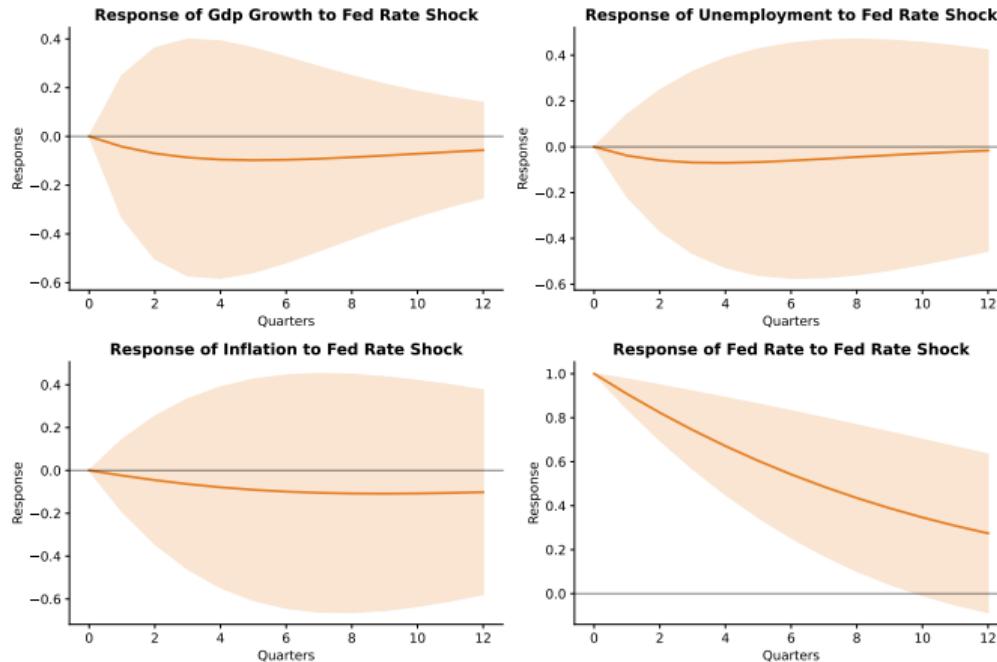
# IRF: Unemployment Shock



## Effects

↑ Unemp  $\Rightarrow$  ↓ GDP (Okun), ↓ Inflation (Phillips), Fed cuts rates.

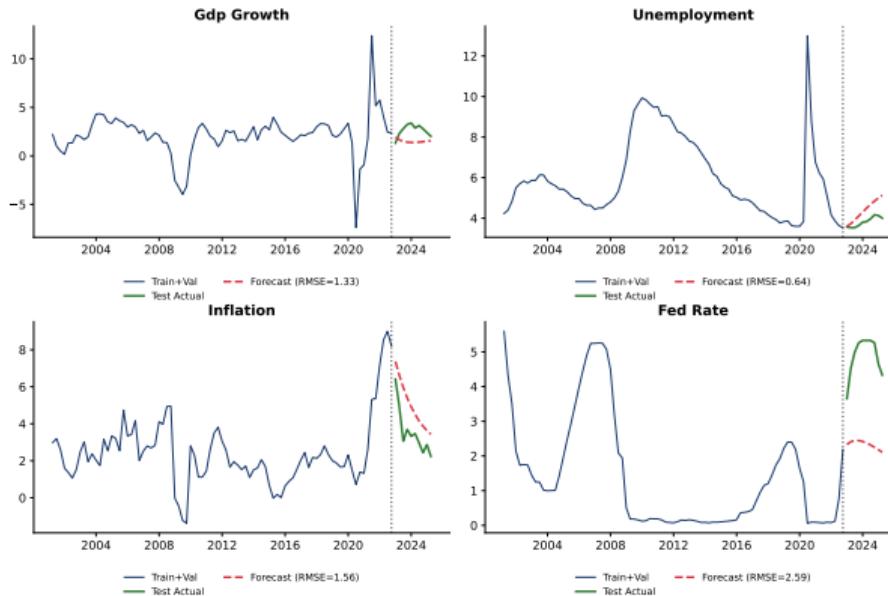
# IRF: Fed Rate Shock



## Monetary Policy

Rate hike  $\Rightarrow$  GDP  $\downarrow$ , Unemployment  $\uparrow$ , Inflation  $\downarrow$ .

## VAR: Forecast (Train/Val/Test)



### Rolling One-Step-Ahead Forecast

VAR captures GDP-Unemployment dynamics. COVID shock visible in test period.

## VAR: Test Set Results

### Test Set Performance by Variable

Variable	RMSE	MAE	Direction Acc.
GDP Growth	1.33	0.99	50%
Unemployment	0.64	0.52	50%
Inflation	1.56	1.12	60%
Fed Rate	2.59	2.45	80%
<b>Average</b>	<b>1.53</b>	<b>1.27</b>	<b>60%</b>

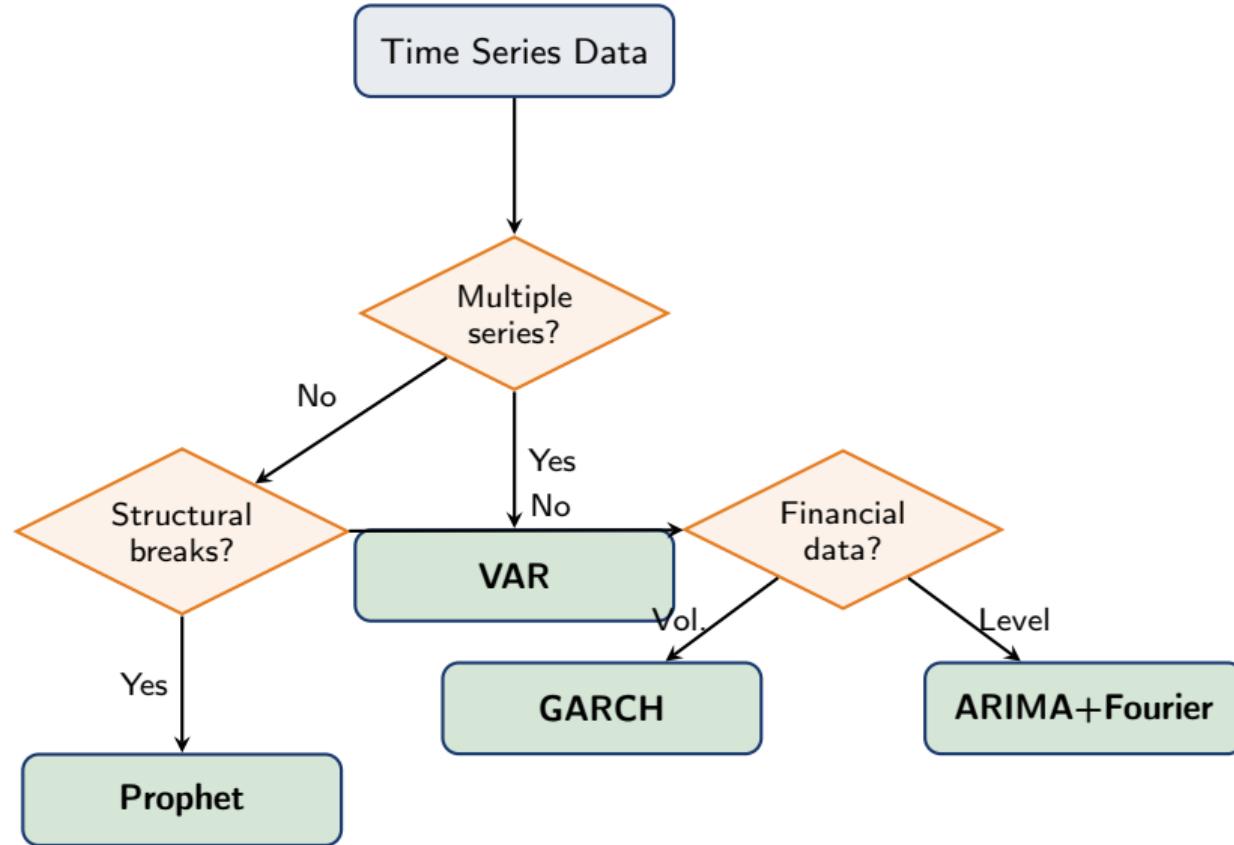
### Strengths

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

### Limitations

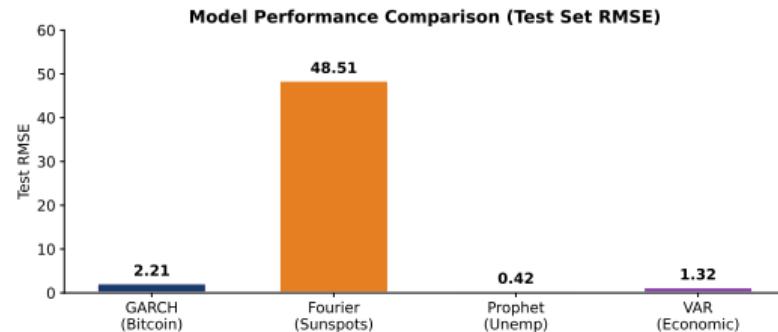
- Many parameters (curse of dimensionality)
- Sensitive to lag selection
- COVID period challenging

# Model Selection Framework



## Summary: Model Comparison

Case	Challenge	Model	RMSE
Bitcoin	Volatility	GARCH	2.15
Sunspots	Seasonality	Fourier	31.10
Unemp	Break	SARIMA	0.12
Economic	Multi-var	VAR	1.53



### 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

# Comprehensive Model Comparison

Feature	GARCH	Fourier	Prophet	VAR
Target	Volatility	Level	Level	Multiple
Seasonality	No	Yes (long)	Yes (multi)	No
Structural breaks	No	No	Yes	No
Multiple series	No	No	No	Yes
Interpretable	Medium	High	High	High
Parameters	Few	2K	Auto	Many
Missing data	No	No	Yes	No
Best for	Finance	Cycles	Business	Macro

## Our Results

- GARCH: MAE=1.82 (volatility)
- Fourier: RMSE=31.10 (cycles)
- SARIMA: RMSE=0.12 (breaks)
- VAR: Avg RMSE=1.53 (multi)

## Key Insight

Each model excels in its domain. The art is matching the model to the data characteristics.

# Best Practices for Applied Forecasting

## 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