Chapter 4: SARIMA Models

Seminar

Seminar Outline

Quiz 1: Seasonal Differencing

Question

For monthly data with annual seasonality, what does the operator $(1 - L^{12})$ do?

- A) Takes 12 consecutive differences
- B) Computes $Y_t - Y_{t-12}$
- C) Averages over 12 months
- D) Removes the first 12 observations

Quiz 1: Seasonal Differencing

Question

For monthly data with annual seasonality, what does the operator $(1 - L^{12})$ do?

- A) Takes 12 consecutive differences
- B) Computes $Y_t - Y_{t-12}$
- C) Averages over 12 months
- D) Removes the first 12 observations

Answer: B

The seasonal differencing operator $(1 - L^{12})Y_t = Y_t - Y_{t-12}$ compares each observation with the same month in the previous year, removing annual seasonal patterns.

Quiz 2: SARIMA Notation

Question

What does SARIMA$(1,1,1) \times (1,1,1)_{12}$ represent?

- A) 12 different ARIMA models
- B) ARIMA with 12 AR and 12 MA terms
- C) ARIMA(1,1,1) with seasonal ARIMA(1,1,1) at period 12
- D) A model requiring 12 years of data

Quiz 2: SARIMA Notation

Question

What does SARIMA$(1, 1, 1) \times (1, 1, 1)_{12}$ represent?

- A) 12 different ARIMA models
- B) ARIMA with 12 AR and 12 MA terms
- C) ARIMA(1,1,1) with seasonal ARIMA(1,1,1) at period 12
- D) A model requiring 12 years of data

Answer: C

SARIMA$(p, d, q) \times (P, D, Q)_s$ combines regular ARIMA(p, d, q) with seasonal ARIMA(P, D, Q) at seasonal period s. Here we have both regular and seasonal AR(1), I(1), and MA(1) components with $s = 12$.

Quiz 3: The Airline Model

Question

The "airline model" refers to SARIMA$(0,1,1) \times (0,1,1)_{12}$. How many parameters does it have (excluding variance)?

- A) 2 parameters
- B) 4 parameters
- C) 6 parameters
- D) 12 parameters

Quiz 3: The Airline Model

Question

The "airline model" refers to SARIMA$(0,1,1) \times (0,1,1)_{12}$. How many parameters does it have (excluding variance)?

- A) 2 parameters
- B) 4 parameters
- C) 6 parameters
- D) 12 parameters

Answer: A

The airline model has only 2 parameters: θ_1 (regular MA coefficient) and Θ_1 (seasonal MA coefficient). Despite its simplicity, it captures many seasonal patterns remarkably well.

Quiz 4: ACF of Seasonal Data

Question

For monthly data with strong seasonality, where would you expect to see significant ACF spikes?

- A) Only at lag 1
- B) Only at lag 12
- C) At lags 12, 24, 36, ...
- D) Randomly distributed

Quiz 4: ACF of Seasonal Data

Question

For monthly data with strong seasonality, where would you expect to see significant ACF spikes?

- A) Only at lag 1
- B) Only at lag 12
- C) At lags 12, 24, 36, ...
- D) Randomly distributed

Answer: C

Seasonal data shows significant autocorrelation at the seasonal frequency and its multiples. For monthly data with annual seasonality, expect spikes at lags 12, 24, 36, etc., reflecting correlation between same months across years.

Quiz 5: Multiplicative Structure

Question

In SARIMA, what does “multiplicative structure” mean?

- A) The seasonal amplitude grows proportionally
- B) Regular and seasonal polynomials are multiplied
- C) We multiply the data by seasonal factors
- D) The model is estimated using multiplication

Quiz 5: Multiplicative Structure

Question

In SARIMA, what does "multiplicative structure" mean?

- A) The seasonal amplitude grows proportionally
- B) Regular and seasonal polynomials are multiplied
- C) We multiply the data by seasonal factors
- D) The model is estimated using multiplication

Answer: B

Multiplicative structure means the AR polynomials $\phi(L) \times \Phi(L^s)$ and MA polynomials $\theta(L) \times \Theta(L^s)$ are multiplied together, creating cross-terms that capture interactions between regular and seasonal dynamics.

Quiz 6: Seasonal vs Regular Differencing

Question

When would you apply both regular ($d = 1$) and seasonal ($D = 1$) differencing?

- A) When data has only a trend
- B) When data has only seasonality
- C) When data has both trend and seasonal non-stationarity
- D) Never – they cancel each other

Quiz 6: Seasonal vs Regular Differencing

Question

When would you apply both regular ($d = 1$) and seasonal ($D = 1$) differencing?

- A) When data has only a trend
- B) When data has only seasonality
- C) When data has both trend and seasonal non-stationarity
- D) Never – they cancel each other

Answer: C

Both types of differencing are needed when the series exhibits both a stochastic trend (regular unit root) and stochastic seasonality (seasonal unit root). For example, airline passenger data needs $(1 - L)(1 - L^{12})Y_t$ to achieve stationarity.

Problem 1: Expanding the Seasonal Difference

Exercise

Expand $(1-L)(1-L^{12})Y_t$ fully. What observations are involved?

Problem 1: Expanding the Seasonal Difference

Exercise

Expand $(1-L)(1-L^{12})Y_t$ fully. What observations are involved?

Solution

$(1-L)(1-L^{12}) = 1 - L - L^{12} + L^{13}$

Therefore: $(1-L)(1-L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$

Interpretation: This is the difference of differences:

- First seasonal difference: $Y_t - Y_{t-12}$ (this year vs last year)
- Then regular difference: $(Y_t - Y_{t-12}) - (Y_{t-1} - Y_{t-13})$

Problem 2: Airline Model Expansion

Exercise

Write out the full equation for the airline model SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$:
$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t$

Problem 2: Airline Model Expansion

Exercise

Write out the full equation for the airline model SARIMA$(0,1,1) \times (0,1,1)_{12}$:
$(1-L)(1-L^{12})Y_t = (1+\theta_1 L)(1+\Theta_1 L^{12})\varepsilon_t$

Solution

Expand the MA side: $(1+\theta_1 L)(1+\Theta_1 L^{12}) = 1 + \theta_1 L + \Theta_1 L^{12} + \theta_1\Theta_1 L^{13}$

Full model: $Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} = \varepsilon_t + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-12} + \theta_1\Theta_1\varepsilon_{t-13}$

Note: The cross-term $\theta_1\Theta_1 L^{13}$ is the multiplicative interaction between regular and seasonal MA components.

Problem 3: Parameter Count

Exercise

How many parameters (excluding σ^2) are in SARIMA$(2, 1, 1) \times (1, 0, 1)_4$?

Problem 3: Parameter Count

Exercise

How many parameters (excluding σ^2) are in SARIMA$(2,1,1) \times (1,0,1)_4$?

Solution

- Regular AR($p = 2$): $\phi_1, \phi_2 \Rightarrow 2$ parameters
- Regular MA($q = 1$): $\theta_1 \Rightarrow 1$ parameter
- Seasonal AR($P = 1$): $\Phi_1 \Rightarrow 1$ parameter
- Seasonal MA($Q = 1$): $\Theta_1 \Rightarrow 1$ parameter

Total: 5 parameters

Note: The differencing orders ($d = 1$, $D = 0$) don't add parameters – they're transformations applied to the data.

Problem 4: SARIMA Forecasting

Exercise

Given the airline model with $\theta_1 = -0.4$ and $\Theta_1 = -0.6$, and:

- $Y_T = 500,\ Y_{T-1} = 495,\ Y_{T-11} = 480,\ Y_{T-12} = 470$
- $\varepsilon_T = 5$, $\varepsilon_{T-11} = -3$, $\varepsilon_{T-12} = 2$

Forecast Y_{T+1}.

Problem 4: SARIMA Forecasting

Exercise

Given the airline model with $\theta_1 = -0.4$ and $\Theta_1 = -0.6$, and:

- $Y_T = 500,\ Y_{T-1} = 495,\ Y_{T-11} = 480,\ Y_{T-12} = 470$
- $\varepsilon_T = 5$, $\varepsilon_{T-11} = -3$, $\varepsilon_{T-12} = 2$

Forecast Y_{T+1}.

Solution

From the model: $Y_{T+1} = Y_T + Y_{T-11} - Y_{T-12} + \varepsilon_{T+1} + \theta_1\varepsilon_T + \Theta_1\varepsilon_{T-11} + \theta_1\Theta_1\varepsilon_{T-12}$
Setting $\mathbb{E}[\varepsilon_{T+1}] = 0$:
$\hat{Y}_{T+1} = 500 + 480 - 470 + 0 + (-0.4)(5) + (-0.6)(-3) + (-0.4)(-0.6)(2)$
$= 510 - 2 + 1.8 + 0.48 = \mathbf{510.28}$

Problem 5: Identifying Seasonal Period

Exercise

Match each data type with its typical seasonal period s:

1. Quarterly GDP data
2. Monthly retail sales
3. Weekly restaurant reservations
4. Daily electricity demand

Problem 5: Identifying Seasonal Period

Exercise

Match each data type with its typical seasonal period s:

1. Quarterly GDP data
2. Monthly retail sales
3. Weekly restaurant reservations
4. Daily electricity demand

Solution

1. Quarterly GDP: $s = 4$ (annual cycle over 4 quarters)
2. Monthly retail sales: $s = 12$ (annual cycle over 12 months)
3. Weekly restaurant reservations: $s = 7$ (weekly cycle) or $s = 52$ (annual)
4. Daily electricity demand: $s = 7$ (weekly pattern) or $s = 365$ (annual)

Note: Some series have multiple seasonal patterns (e.g., daily data may have weekly AND annual cycles).

Example: Monthly Retail Sales Analysis

Scenario

You have 5 years of monthly retail sales data showing clear December peaks and January troughs. Build an appropriate SARIMA model.

Step-by-step Approach

1. **Visual inspection**: Plot shows upward trend + strong December spikes
2. **Seasonal period**: Monthly data with annual pattern $\Rightarrow s = 12$
3. **Transformation**: Consider $\log(Y_t)$ if seasonal amplitude grows with level
4. **Differencing**: Try $(1 - L)(1 - L^{12})Y_t$ – check ACF/PACF
5. **Model selection**: Start with airline model, compare via AIC

Example: ACF/PACF Interpretation for Seasonal Data

Observed Patterns (after differencing)

- ACF: Significant at lags 1, 12; cuts off after lag 1 and lag 12
- PACF: Significant at lags 1, 12, 13; decays at multiples of 12

Interpretation

Regular component: ACF cuts off at 1 $\Rightarrow$ MA(1)
Seasonal component: ACF significant only at lag 12 $\Rightarrow$ seasonal MA(1)
Suggested model: SARIMA$(0, d, 1) \times (0, D, 1)_{12}$ – the airline model!

Alternative check: If PACF showed cutoff at seasonal lags instead of ACF, consider seasonal AR terms.

Example: Python Implementation

Fitting SARIMA in Python

```
from statsmodels.tsa.statespace.sarimax import SARIMAX
import pmdarima as pm

# Manual fit
model = SARIMAX(y, order=(0,1,1), seasonal_order=(0,1,1,12))
results = model.fit()
print(results.summary())

# Automatic selection
auto_model = pm.auto_arima(y, seasonal=True, m=12,
                           start_p=0, max_p=2,
                           start_q=0, max_q=2,
                           d=1, D=1,
                           trace=True)
```

Example: Interpreting SARIMA Output

Sample statsmodels Output

```
                          SARIMAX Results
==============================================================
Model:            SARIMAX(0,1,1)x(0,1,1,12)    AIC:     1348.52
                                                BIC:     1358.21
==============================================================
                 coef    std err        z      P>|z|
--------------------------------------------------------------
ma.L1         -0.4018      0.072    -5.58      0.000
ma.S.L12      -0.5521      0.081    -6.82      0.000
sigma2      1254.3201    142.856     8.78      0.000
```

Interpretation

- $\hat{\theta}_1 = -0.40$: Negative MA means positive shocks reduce next period's value
- $\hat{\Theta}_1 = -0.55$: Same-season correlation is captured
- Both coefficients significant ($p < 0.001$); $|\theta|, |\Theta| < 1$ – invertible

Discussion: Deterministic vs Stochastic Seasonality

Key Question

When should you use seasonal dummies vs SARIMA for seasonal data?

Considerations

Seasonal dummies (deterministic):

- Fixed, repeating pattern each year
- Same December effect every year
- Appropriate when seasonality is stable

SARIMA (stochastic):

- Evolving seasonal pattern
- This year's December depends on last year's December
- Better when seasonal amplitude varies

Discussion: Log Transformation

Key Question

When should you take logarithms before fitting SARIMA?

Guidelines

Use log transformation when:

- Seasonal fluctuations grow with the level (multiplicative seasonality)
- Variance increases over time
- Data is strictly positive (prices, sales, counts)

Avoid log when:

- Seasonal pattern is additive (constant amplitude)
- Data contains zeros or negatives
- Already on a rate/ratio scale

Tip: Compare AIC of models with and without log transformation.

Discussion: Multiple Seasonalities

Challenge

Daily sales data may have both weekly (7-day) and annual (365-day) seasonal patterns. How do you handle this?

Approaches

1. **Nested SARIMA**: Model at shorter frequency, include longer as exogenous
2. **TBATS/BATS models**: Explicitly handle multiple seasonalities
3. **Fourier terms**: Add sin/cos terms for each seasonal frequency
4. **Prophet/similar**: Modern tools designed for multiple seasonalities

Note: Standard SARIMA handles only one seasonal period. For complex seasonality, consider specialized methods.

Discussion: Forecasting Seasonal Data

Key Question

What are the unique challenges of forecasting seasonal time series?

Challenges and Solutions

- **Horizon matters**: 12-month forecast means predicting a full cycle
- **Uncertainty grows**: Seasonal forecasts compound regular uncertainty
- **Turning points**: Capturing when seasons peak/trough
- **Structural breaks**: COVID-19 disrupted many seasonal patterns

Best practices:

- Use rolling-origin cross-validation
- Compare against seasonal naive benchmark
- Report forecast intervals, especially at seasonal horizons

Take-Home Exercises

1. **Theoretical**: Show that $(1-L)(1-L^4)$ can be written as $(1-L-L^4+L^5)$ and explain what this transformation does to quarterly data with annual seasonality.

2. **Computation**: For SARIMA$(1,0,0)\times(1,0,0)_4$ with $\phi_1 = 0.5$ and $\Phi_1 = 0.8$, write out the full AR polynomial and identify all non-zero coefficients.

3. **Applied**: Download monthly airline passenger data and:
   - Plot the series and identify trend/seasonality
   - Apply appropriate transformations
   - Fit the airline model and interpret coefficients
   - Generate 24-month forecasts with confidence intervals

4. **Comparison**: Fit both SARIMA$(0,1,1)\times(0,1,1)_{12}$ and SARIMA$(1,1,0)\times(1,1,0)_{12}$ to the airline data. Compare using AIC, BIC, and residual diagnostics. Which is preferred?

Exercise Solutions Hints

Hints

1. Expand $(1-L)(1-L^4) = 1 \cdot 1 - 1 \cdot L^4 - L \cdot 1 + L \cdot L^4 = 1 - L - L^4 + L^5$
2. AR polynomial: $(1-\phi_1 L)(1-\Phi_1 L^4) = 1 - 0.5L - 0.8L^4 + 0.4L^5$
3. For airline data:
   - Use log transformation (multiplicative seasonality)
   - Both $d = 1$ and $D = 1$ needed
   - Typical estimates: $\theta_1 \approx -0.4$, $\Theta_1 \approx -0.6$
4. The MA-based airline model typically fits better than pure AR seasonal model for this data (lower AIC).

Key Takeaways from This Seminar

Main Points

1. Seasonal differencing $(1 - L^s)$ removes stochastic seasonality
2. SARIMA notation: $(p, d, q) \times (P, D, Q)_s$ separates regular and seasonal
3. The airline model is surprisingly effective for many datasets
4. Multiplicative structure creates interaction terms
5. ACF/PACF show patterns at both regular and seasonal lags
6. Log transformation often needed for multiplicative seasonality

Next Steps

Chapter 5 will cover multivariate time series: VAR models, Granger causality, and cointegration.