Chapter 3: ARIMA Models

Seminar

Seminar Outline

1. Review Quiz
2. Practice Problems
3. Worked Examples
4. Discussion Topics
5. Summary

Quiz 1: Integration Order

Question

A time series Y_t requires two differences to become stationary. What is its order of integration?

A) $I(0)$
B) $I(1)$
C) $I(2)$
D) Cannot be determined

Quiz 1: Integration Order

Question

A time series Y_t requires two differences to become stationary. What is its order of integration?

- A) $I(0)$
- B) $I(1)$
- C) $I(2)$
- D) Cannot be determined

Answer: C

By definition, $Y_t \sim I(d)$ means d differences are needed to achieve stationarity. Since two differences are required, $Y_t \sim I(2)$.

Quiz 2: Random Walk Properties

Question

For a random walk $Y_t = Y_{t-1} + \varepsilon_t$ with $\mathrm{Var}(\varepsilon_t) = \sigma^2$, what is $\mathrm{Var}(Y_t)$?

- A) σ^2
- B) $t \cdot \sigma^2$
- C) σ^2/t
- D) $\sigma^2/(1-\phi^2)$

Quiz 2: Random Walk Properties

Question

For a random walk $Y_t = Y_{t-1} + \varepsilon_t$ with $\text{Var}(\varepsilon_t) = \sigma^2$, what is $\text{Var}(Y_t)$?

- A) σ^2
- B) $t \cdot \sigma^2$
- C) σ^2/t
- D) $\sigma^2/(1-\phi^2)$

Answer: B

Since $Y_t = \sum_{i=1}^{t} \varepsilon_i$ (assuming $Y_0 = 0$), we have $\text{Var}(Y_t) = t \cdot \sigma^2$. The variance grows linearly with time – a key feature of non-stationarity.

Quiz 3: ADF Test Hypotheses

Question

In the Augmented Dickey-Fuller test, what is the null hypothesis?

- A) The series is stationary
- B) The series has a unit root
- C) The series has no autocorrelation
- D) The series is normally distributed

Quiz 3: ADF Test Hypotheses

Question

In the Augmented Dickey-Fuller test, what is the null hypothesis?

- A) The series is stationary
- B) The series has a unit root
- C) The series has no autocorrelation
- D) The series is normally distributed

Answer: B

The ADF test has H_0: unit root (non-stationary) vs H_1: stationary. We reject H_0 if the test statistic is sufficiently negative (below critical value).

Quiz 4: ARIMA Notation

Question

What does ARIMA(2,1,1) mean?

- A) AR(2) on differenced data with MA(1) errors
- B) AR(1) with 2 differences and MA(1)
- C) MA(2) with 1 difference and AR(1)
- D) 2 lags, 1 trend, 1 seasonal component

Quiz 4: ARIMA Notation

Question

What does ARIMA(2,1,1) mean?

- A) AR(2) on differenced data with MA(1) errors
- B) AR(1) with 2 differences and MA(1)
- C) MA(2) with 1 difference and AR(1)
- D) 2 lags, 1 trend, 1 seasonal component

Answer: A

ARIMA(p,d,q) means: p=AR order, d=differencing order, q=MA order. So ARIMA(2,1,1) has AR(2) and MA(1) components applied to the first-differenced series.

Quiz 5: Difference Operator

Question

What is $(1 - L)^2 Y_t$ expanded?

A) $Y_t - Y_{t-1}$
B) $Y_t - 2Y_{t-1} + Y_{t-2}$
C) $Y_t + 2Y_{t-1} + Y_{t-2}$
D) $Y_t - Y_{t-2}$

Quiz 5: Difference Operator

Question

What is $(1 - L)^2 Y_t$ expanded?

- A) $Y_t - Y_{t-1}$
- B) $Y_t - 2Y_{t-1} + Y_{t-2}$
- C) $Y_t + 2Y_{t-1} + Y_{t-2}$
- D) $Y_t - Y_{t-2}$

Answer: B

$(1 - L)^2 = 1 - 2L + L^2$, so $(1 - L)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$. This is the second difference of Y_t.

Quiz 6: KPSS vs ADF

Question

How does the KPSS test differ from the ADF test?

- A) KPSS tests for seasonality, ADF tests for trends
- B) KPSS has stationarity as null, ADF has unit root as null
- C) KPSS is more powerful than ADF
- D) There is no difference

Quiz 6: KPSS vs ADF

Question

How does the KPSS test differ from the ADF test?

- A) KPSS tests for seasonality, ADF tests for trends
- B) KPSS has stationarity as null, ADF has unit root as null
- C) KPSS is more powerful than ADF
- D) There is no difference

Answer: B

The key difference is reversed hypotheses. KPSS: H_0 = stationary. ADF: H_0 = unit root. Using both tests together provides stronger evidence.

Quiz 7: Overdifferencing

Question

If $Y_t \sim I(1)$ and we compute $\Delta^2 Y_t$, what happens?

- A) We get a better stationary series
- B) We introduce artificial negative autocorrelation
- C) The variance decreases
- D) Nothing changes

Quiz 7: Overdifferencing

Question

If $Y_t \sim I(1)$ and we compute $\Delta^2 Y_t$, what happens?

- A) We get a better stationary series
- B) We introduce artificial negative autocorrelation
- C) The variance decreases
- D) Nothing changes

Answer: B

Overdifferencing creates an MA component at the invertibility boundary. If $\Delta Y_t = \varepsilon_t$, then $\Delta^2 Y_t = \varepsilon_t - \varepsilon_{t-1}$, which is MA(1) with $\theta = -1$ (non-invertible).

Quiz 8: Forecast Variance

Question

For an ARIMA(0,1,0) model (random walk), how does forecast variance behave as horizon h increases?

- A) Stays constant
- B) Decreases to zero
- C) Grows linearly with h
- D) Converges to a finite limit

Quiz 8: Forecast Variance

Question

For an ARIMA(0,1,0) model (random walk), how does forecast variance behave as horizon h increases?

- A) Stays constant
- B) Decreases to zero
- C) Grows linearly with h
- D) Converges to a finite limit

Answer: C

For a random walk, $\text{Var}(Y_{T+h} - \hat{Y}_{T+h|T}) = h\sigma^2$. The forecast uncertainty grows without bound – a key characteristic of I(1) processes.

Problem 1: Unit Root Testing

Exercise

You have quarterly GDP data for 80 quarters. The ADF test (with constant and trend) gives a test statistic of -2.85. The 5% critical value is -3.41.

1. What is your conclusion about stationarity?
2. What would you do next?

Problem 1: Unit Root Testing

Exercise

You have quarterly GDP data for 80 quarters. The ADF test (with constant and trend) gives a test statistic of -2.85. The 5% critical value is -3.41.

1. What is your conclusion about stationarity?
2. What would you do next?

Solution

1. Since $-2.85 > -3.41$, we **fail to reject** H_0. The data appears to have a unit root (non-stationary).
2. Take the first difference ΔY_t and repeat the ADF test on the differenced series to confirm it is now stationary.

Problem 2: Model Identification

Exercise

After differencing a time series once, the ACF shows:

- Significant spike at lag 1 ($\rho_1 = 0.4$)
- All other lags insignificant

The PACF shows gradual decay.
What ARIMA model is suggested?

Problem 2: Model Identification

Exercise

After differencing a time series once, the ACF shows:

- Significant spike at lag 1 ($\rho_1 = 0.4$)
- All other lags insignificant

The PACF shows gradual decay.
What ARIMA model is suggested?

Solution

- ACF cuts off after lag 1 $\Rightarrow$ MA(1) component
- PACF decays $\Rightarrow$ Confirms MA structure
- Since we differenced once: $d = 1$

Suggested model: ARIMA(0,1,1) or IMA(1,1)

Problem 3: ARIMA Equation

Exercise

Write out the full equation for ARIMA(1,1,1):

$$(1 - \phi_1 L)(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$$

Expand this completely in terms of Y_t, Y_{t-1}, Y_{t-2}, etc.

Problem 3: ARIMA Equation

Exercise

Write out the full equation for ARIMA(1,1,1):

$$(1 - \phi_1 L)(1 - L)Y_t = c + (1 + \theta_1 L)\varepsilon_t$$

Expand this completely in terms of Y_t, Y_{t-1}, Y_{t-2}, etc.

Solution

Expanding $(1 - \phi_1 L)(1 - L) = 1 - L - \phi_1 L + \phi_1 L^2 = 1 - (1 + \phi_1)L + \phi_1 L^2$:

$$Y_t - (1 + \phi_1)Y_{t-1} + \phi_1 Y_{t-2} = c + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Or equivalently:

$$Y_t = c + (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Problem 4: Forecast Calculation

Exercise

Given ARIMA(0,1,1): $\Delta Y_t = \varepsilon_t + 0.3\varepsilon_{t-1}$
At time T: $Y_T = 100$, $\hat{\varepsilon}_T = 2$, $\sigma^2 = 4$
Calculate:

1. $\hat{Y}_{T+1|T}$ (one-step forecast)
2. $\hat{Y}_{T+2|T}$ (two-step forecast)

Problem 4: Forecast Calculation

Exercise

Given ARIMA(0,1,1): $\Delta Y_t = \varepsilon_t + 0.3\varepsilon_{t-1}$
At time T: $Y_T = 100$, $\hat{\varepsilon}_T = 2$, $\sigma^2 = 4$
Calculate:

1. $\hat{Y}_{T+1|T}$ (one-step forecast)
2. $\hat{Y}_{T+2|T}$ (two-step forecast)

Solution

1. $\hat{Y}_{T+1|T} = Y_T + 0.3\hat{\varepsilon}_T = 100 + 0.3(2) = \mathbf{100.6}$
2. $\hat{Y}_{T+2|T} = \hat{Y}_{T+1|T} + 0.3 \cdot 0 = 100.6 + 0 = \mathbf{100.6}$
   (Future shocks $\varepsilon_{T+1}, \varepsilon_{T+2}$ are forecast as 0)

Problem 5: Confidence Intervals

Exercise

Continuing from Problem 4, calculate the 95% forecast intervals for $\hat{Y}_{T+1|T}$ and $\hat{Y}_{T+2|T}$.
Recall: $\sigma^2 = 4$, $\theta_1 = 0.3$

Problem 5: Confidence Intervals

Exercise

Continuing from Problem 4, calculate the 95% forecast intervals for $\hat{Y}_{T+1|T}$ and $\hat{Y}_{T+2|T}$.
Recall: $\sigma^2 = 4$, $\theta_1 = 0.3$

Solution

For IMA(1,1), the MA(∞) weights are $\psi_0 = 1$, $\psi_j = 1 + \theta_1$ for $j \geq 1$.
1-step: $\text{Var}(e_{T+1}) = \sigma^2 \psi_0^2 = 4$, so $SE = 2$
$100.6 \pm 1.96(2) = [\mathbf{96.68}, \mathbf{104.52}]$
2-step: $\text{Var}(e_{T+2}) = \sigma^2(\psi_0^2 + \psi_1^2) = 4(1 + 1.3^2) = 10.76$, $SE = 3.28$
$100.6 \pm 1.96(3.28) = [\mathbf{94.17}, \mathbf{107.03}]$

Example: Testing for Unit Root in Stock Prices

Scenario

You have daily closing prices for a stock over 500 days. You want to determine if prices follow a random walk.

Step-by-step Approach

1. **Visual inspection**: Plot prices – likely shows trend
2. **ADF test on prices**: Expect to fail to reject H_0 (unit root)
3. **Take log returns**: $r_t = \ln(P_t/P_{t-1}) = \Delta \ln(P_t)$
4. **ADF test on returns**: Should reject H_0 (stationary)
5. **Conclusion**: Log prices are $I(1)$, returns are $I(0)$

Example: Box-Jenkins for Inflation Data

Scenario

Monthly inflation rates for 10 years. Build an ARIMA model.

Workflow

1. **Plot & test**: ADF suggests borderline – try both $d = 0$ and $d = 1$
2. **If** $d = 0$: Fit ARMA models, compare AIC
3. **If** $d = 1$: Examine ACF/PACF of ΔY_t
   - ACF: spike at lag 1, then cuts off
   - PACF: decays
   - $\Rightarrow$ Try ARIMA(0,1,1)
4. **Estimate**: Fit ARIMA(0,1,1), check coefficients
5. **Diagnose**: Ljung-Box on residuals (want $p > 0.05$)
6. **Compare**: AIC of ARIMA(0,1,1) vs ARMA(1,1) on levels

Example: Interpreting Python Output

statsmodels ARIMA Output

```
                          ARIMA Model Results
==============================================================
Dep. Variable:            D.y   No. Observations:    99
Model:              ARIMA(1,1,1)   AIC                  285.32
                                   BIC                  295.63
==============================================================
                  coef    std err       z      P>|z|
--------------------------------------------------------------
const           0.0521    0.048     1.085    0.278
ar.L1           0.4532    0.102     4.443    0.000
ma.L1          -0.2891    0.118    -2.450    0.014
sigma2          1.2340    0.176     7.011    0.000
```

Interpretation

- AR coefficient (0.45) is significant, MA coefficient (-0.29) is significant
- Constant (0.052) not significant – could set $c = 0$
- Check: $|\phi_1| < 1$ (stationary), $|\theta_1| < 1$ (invertible) – OK!

Discussion: Deterministic vs Stochastic Trends

Key Question

Why is it important to distinguish between deterministic and stochastic trends?

Discussion Points

- **Wrong treatment consequences**:
  - Detrending a unit root $\Rightarrow$ spurious stationarity
  - Differencing a trend-stationary $\Rightarrow$ overdifferencing
- **Economic interpretation**:
  - Deterministic trend: shocks are temporary
  - Stochastic trend: shocks have permanent effects
- **Policy implications**:
  - Does a recession permanently lower GDP, or does the economy return to trend?

Discussion: Model Selection Criteria

Key Question

When should you use AIC vs BIC for ARIMA model selection?

Considerations

- **AIC**: Minimizes prediction error, may overfit
  - Better for forecasting
  - Tends to select larger models
- **BIC**: Consistent model selection, more parsimonious
  - Better for identifying “true” model
  - Penalizes complexity more heavily
- **Practical advice**: Report both, prefer BIC if they disagree substantially

Discussion: Limitations of ARIMA

Key Question

What are the main limitations of ARIMA models?

Discussion Points

- **Linearity**: Cannot capture nonlinear dynamics
- **Constant variance**: Assumes homoskedasticity (no GARCH effects)
- **No structural breaks**: Parameters assumed constant
- **Univariate**: Ignores relationships with other variables
- **Symmetric**: Treats positive and negative shocks equally
- **Long-horizon forecasts**: Uncertainty grows rapidly

Extensions

These limitations motivate GARCH (volatility), VAR (multivariate), regime-switching models, etc.

Key Points from Today's Seminar

What We Covered

1. **Integration and differencing**: $I(d)$ processes require d differences
2. **Unit root testing**: ADF tests H_0: unit root; KPSS tests H_0: stationary
3. **ARIMA(p,d,q)**: Combines ARMA with differencing
4. **Model identification**: Use ACF/PACF patterns and information criteria
5. **Forecasting**: Point forecasts and growing confidence intervals

Next Seminar

Hands-on Python exercises with real economic data:

- Unit root testing with `statsmodels`
- Auto-ARIMA with `pmdarima`
- Forecasting and model diagnostics