Time Series Analysis – Quiz

Statistical Data Analysis

Question 1

A time series is decomposed into three components. If observed monthly ice cream sales show a long-term upward direction, peaks every July, and unpredictable day-to-day fluctuations, what are these three components called?

A. Mean, variance, and covariance

B. Trend, seasonality, and noise

C. Level, slope, and damping

D. Autocorrelation, partial autocorrelation, and lag

Question 1

A time series is decomposed into three components. If observed monthly ice cream sales show a long-term upward direction, peaks every July, and unpredictable day-to-day fluctuations, what are these three components called?

A. Mean, variance, and covariance

B. Trend, seasonality, and noise

C. Level, slope, and damping

D. Autocorrelation, partial autocorrelation, and lag

Answer: B

Every time series can be decomposed as Observed = Trend + Seasonal + Noise. The long-term upward direction is the trend, the regular July peaks are seasonality (repeating pattern), and the unpredictable daily fluctuations are noise (random component).

Question 2

A city's monthly electricity demand shows a clear repeating pattern: high in January and July (heating/cooling), low in April and October. The seasonal period s for this monthly data is:

A. 2, because there are two peaks per year

B. 6, because there are 6 months between peaks

C. 12, because the pattern repeats every 12 months

D. 4, because the pattern repeats quarterly

Question 2

A city's monthly electricity demand shows a clear repeating pattern: high in January and July (heating/cooling), low in April and October. The seasonal period s for this monthly data is:

A. 2, because there are two peaks per year

B. 6, because there are 6 months between peaks

C. 12, because the pattern repeats every 12 months

D. 4, because the pattern repeats quarterly

Answer: C

The seasonal period s refers to the number of time steps in one complete cycle. For monthly data with an annual pattern, $s = 12$ regardless of how many peaks occur within the year. This is the standard convention used in SARIMA notation, as shown in the lecture (e.g., quarterly GDP uses $s = 4$, monthly retail uses $s = 12$).

Question 3

A series has a constant mean of 20 and a constant variance of 4 over any time window. The ADF test yields a p-value of 0.002. What should you conclude?

A. The ADF test is inconclusive at this p-value

B. The series is non-stationary because the variance is too low

C. You need to difference the series before interpreting the ADF result

D. The series is stationary, supported by both visual stability and a significant ADF test

Question 3

A series has a constant mean of 20 and a constant variance of 4 over any time window. The ADF test yields a p-value of 0.002. What should you conclude?

A. The ADF test is inconclusive at this p-value

B. The series is non-stationary because the variance is too low

C. You need to difference the series before interpreting the ADF result

D. The series is stationary, supported by both visual stability and a significant ADF test

Answer: D

The ADF (Augmented Dickey-Fuller) test has as its null hypothesis that the series has a unit root (non-stationary). A low p-value (0.002 ¡ 0.05) rejects this null, indicating stationarity. Combined with the visual evidence of constant mean and variance, there is strong support for stationarity.

Question 4

You apply the ADF test and get $p = 0.45$, then the KPSS test and get $p = 0.01$. What do these results together indicate, and what should you do?

A. The series is non-stationary; you should apply differencing to achieve stationarity
B. Both tests agree the series is stationary; proceed to model fitting
C. The series is stationary because at least one test says so
D. The tests contradict each other; the series is likely trend-stationary and may need detrending

Question 4

You apply the ADF test and get p = 0.45, then the KPSS test and get p = 0.01. What do these results together indicate, and what should you do?

A. The series is non-stationary; you should apply differencing to achieve stationarity

B. Both tests agree the series is stationary; proceed to model fitting

C. The series is stationary because at least one test says so

D. The tests contradict each other; the series is likely trend-stationary and may need detrending

Answer: A

The ADF null hypothesis is 'unit root present' (non-stationary); p = 0.45 fails to reject it, suggesting non-stationarity. The KPSS null hypothesis is 'series is stationary'; p = 0.01 rejects it, also suggesting non-stationarity. Both tests agree the data is non-stationary, so you should apply differencing ($\Delta Y_t = Y_t - Y_{t-1}$) to make it stationary before modeling.

Question 5

You compute the ACF of a time series and find ACF(1) = 0.85, ACF(2) = 0.72, ACF(3) = 0.61, with values slowly decaying. The PACF shows a sharp cutoff after lag 2 (PACF(1) = 0.85, PACF(2) = 0.30, PACF(3) = 0.02). Which model is most appropriate?

A. MA(2), because the PACF cuts off at lag 2

B. AR(2), because the PACF cuts off at lag 2 while the ACF decays gradually

C. ARMA(1,1), because both ACF and PACF decay

D. AR(1), because the ACF is highest at lag 1

Question 5

You compute the ACF of a time series and find ACF(1) = 0.85, ACF(2) = 0.72, ACF(3) = 0.61, with values slowly decaying. The PACF shows a sharp cutoff after lag 2 (PACF(1) = 0.85, PACF(2) = 0.30, PACF(3) = 0.02). Which model is most appropriate?

A. MA(2), because the PACF cuts off at lag 2

B. AR(2), because the PACF cuts off at lag 2 while the ACF decays gradually

C. ARMA(1,1), because both ACF and PACF decay

D. AR(1), because the ACF is highest at lag 1

Answer: B

The model identification table from the lecture states: for AR(p) models, the ACF decays gradually while the PACF cuts off at lag p. Here the ACF decays (0.85, 0.72, 0.61...) and the PACF cuts off sharply after lag 2 (PACF(3) is near zero). This signature pattern indicates AR(2).

Question 6

A fitted model's residual ACF shows: ACF(1) = 0.42, ACF(2) = 0.35, ACF(3) = 0.28, all well outside the confidence bands. What does this indicate?

A. The model is excellent because the residuals show strong autocorrelation

B. The residuals are white noise, confirming a good fit

C. The model has missed systematic patterns in the data and needs revision

D. This is expected for any ARIMA model and can be ignored

Question 6

A fitted model's residual ACF shows: ACF(1) = 0.42, ACF(2) = 0.35, ACF(3) = 0.28, all well outside the confidence bands. What does this indicate?

A. The model is excellent because the residuals show strong autocorrelation

B. The residuals are white noise, confirming a good fit

C. The model has missed systematic patterns in the data and needs revision

D. This is expected for any ARIMA model and can be ignored

Answer: C

Good residuals should resemble white noise, meaning all ACF values should be near zero and within the confidence bands. Significant autocorrelation in the residuals (ACF values of 0.42, 0.35, 0.28 outside the bands) indicates the model has failed to capture important patterns in the data. As the lecture states: 'Bad = pattern means model missed something.'

Question 7

In an AR(1) model $Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$, if $\phi_1 = 0.9$ and the current value is $Y_t = 100$ (ignoring the constant and noise), the best point forecast for Y_{t+1} is approximately:

A. 90

B. 100

C. 110

D. 81

Question 7

In an AR(1) model $Y_t = c + \phi_1 Y_{t-1} + \epsilon_t$, if $\phi_1 = 0.9$ and the current value is $Y_t = 100$ (ignoring the constant and noise), the best point forecast for Y_{t+1} is approximately:

A. 90

B. 100

C. 110

D. 81

Answer: A

With $\phi_1 = 0.9$ and ignoring the constant and noise term, the forecast is $Y_{t+1} = 0.9 \times 100 = 90$. This is exactly the numerical example given in the lecture: 'If $\phi = 0.9$ and $Y_t = 100$: $Y_{t+1} = 0.9 \times 100 + \epsilon \approx 90 + \text{noise}$.' A high ϕ means strong persistence.

Question 8

In an MA(1) model $Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1}$, a shock of $\epsilon = 10$ occurs at time $t = 1$ with $\theta_1 = 0.5$. What is the effect of this shock at times $t = 1$, $t = 2$, and $t = 3$?

A. 10, 10, 10 – the shock persists indefinitely

B. 10, 5, 2.5 – the shock decays exponentially

C. 5, 2.5, 0 – the shock is halved immediately

D. 10, 5, 0 – the shock affects only the current and next period

Question 8

In an MA(1) model $Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1}$, a shock of $\epsilon = 10$ occurs at time $t = 1$ with $\theta_1 = 0.5$. What is the effect of this shock at times $t = 1$, $t = 2$, and $t = 3$?

A. 10, 10, 10 – the shock persists indefinitely

B. 10, 5, 2.5 – the shock decays exponentially

C. 5, 2.5, 0 – the shock is halved immediately

D. 10, 5, 0 – the shock affects only the current and next period

Answer: D

In an MA(1) model, a shock affects the current period fully (effect $= 10$) and the next period scaled by θ_1 (effect $= 0.5 \times 10 = 5$), then disappears completely (effect $= 0$ at $t = 3$). This is exactly the numerical example from the lecture. Unlike AR models where effects persist, MA shocks fade quickly – the MA captures short-term effects only.

Question 9

An ARMA(1,1) model is written as $Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$. What ACF/PACF signature would you expect from this process?

A. ACF cuts off at lag 1, PACF decays

B. ACF decays, PACF cuts off at lag 1

C. Both ACF and PACF decay gradually

D. Both ACF and PACF cut off at lag 1

Question 9

An ARMA(1,1) model is written as $Y_t = \phi Y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$. What ACF/PACF signature would you expect from this process?

A. ACF cuts off at lag 1, PACF decays

B. ACF decays, PACF cuts off at lag 1

C. Both ACF and PACF decay gradually

D. Both ACF and PACF cut off at lag 1

Answer: C

The model identification table in the lecture clearly states that for ARMA models, both the ACF and PACF decay gradually. This is what distinguishes ARMA from pure AR (where only ACF decays and PACF cuts off) and pure MA (where only PACF decays and ACF cuts off). When you see both functions decaying, an ARMA model is appropriate.

Question 10

You have a trending series: 100, 102, 105, 109. After first differencing ($d = 1$), the resulting series is:

A. 100, 102, 105, 109 (unchanged)

B. 2, 3, 4

C. 1, 3, 4

D. 0.02, 0.03, 0.04

Question 10

You have a trending series: 100, 102, 105, 109. After first differencing ($d = 1$), the resulting series is:

A. 100, 102, 105, 109 (unchanged)

B. 2, 3, 4

C. 1, 3, 4

D. 0.02, 0.03, 0.04

Answer: B

Differencing computes $\Delta Y_t = Y_t - Y_{t-1}$. So: $102 - 100 = 2$, $105 - 102 = 3$, $109 - 105 = 4$. The resulting series (2, 3, 4) has the trend removed and is closer to stationary. This is the exact example from the lecture slide on differencing, illustrating how looking at changes rather than levels removes trends.

Question 11

You fit an ARIMA(2,1,0) model to a non-stationary series. Which statement correctly describes what this model does?

A. It differences the data once, then fits an AR(2) model to the differenced series
B. It applies two rounds of differencing, then fits an MA(1) model to the result
C. It fits a second-order autoregressive model to the original data without differencing
D. It differences the data once, then fits an MA(2) model to the differenced series

Question 11

You fit an ARIMA(2,1,0) model to a non-stationary series. Which statement correctly describes what this model does?

A. It differences the data once, then fits an AR(2) model to the differenced series

B. It applies two rounds of differencing, then fits an MA(1) model to the result

C. It fits a second-order autoregressive model to the original data without differencing

D. It differences the data once, then fits an MA(2) model to the differenced series

Answer: A

In ARIMA(p, d, q), $p = 2$ is the AR order (two past-value lags), $d = 1$ means one round of differencing, and $q = 0$ means no MA terms. So ARIMA(2,1,0) differences the data once to achieve stationarity, then fits an AR(2) model to the differenced series. The 'I' in ARIMA stands for 'Integrated,' referring to the differencing step.

Question 12

In the Box-Jenkins methodology, what is the correct sequence of steps?

A. Fit model, check ACF/PACF, difference, plot data, verify residuals, forecast

B. Check ACF/PACF, fit model, plot data, forecast, verify residuals

C. Difference the data, fit ARIMA(1,1,1), forecast, then check if residuals are acceptable

D. Plot data, test stationarity, difference if needed, check ACF/PACF, fit model, verify residuals, forecast

Question 12

In the Box-Jenkins methodology, what is the correct sequence of steps?

A. Fit model, check ACF/PACF, difference, plot data, verify residuals, forecast

B. Check ACF/PACF, fit model, plot data, forecast, verify residuals

C. Difference the data, fit ARIMA(1,1,1), forecast, then check if residuals are acceptable

D. Plot data, test stationarity, difference if needed, check ACF/PACF, fit model, verify residuals, forecast

Answer: D

The Box-Jenkins methodology follows a systematic sequence as presented in the lecture: Plot $\rightarrow$ Stationarity $\rightarrow$ Difference $\rightarrow$ ACF/PACF $\rightarrow$ Fit $\rightarrow$ Residuals $\rightarrow$ Forecast. You must first visualize the data, then ensure stationarity before identifying model orders from ACF/PACF patterns, and always verify residuals before trusting forecasts.

Question 13

A SARIMA model is written as SARIMA$(1,1,1)(1,1,1)_{12}$. What does the subscript 12 and the second set of parameters $(1,1,1)$ represent?

A. 12 AR lags and a second ARMA model for long-range dependence

B. A seasonal period of 12 time steps, with one seasonal AR lag, one seasonal differencing, and one seasonal MA term

C. 12 forecast steps ahead and backup parameter estimates

D. A rolling window of 12 observations for cross-validation

Question 13

A SARIMA model is written as SARIMA$(1,1,1)(1,1,1)_{12}$. What does the subscript 12 and the second set of parameters $(1,1,1)$ represent?

A. 12 AR lags and a second ARMA model for long-range dependence

B. A seasonal period of 12 time steps, with one seasonal AR lag, one seasonal differencing, and one seasonal MA term

C. 12 forecast steps ahead and backup parameter estimates

D. A rolling window of 12 observations for cross-validation

Answer: B

In SARIMA$(p, d, q)(P, D, Q)_s$, the first triplet (p, d, q) captures short-term (non-seasonal) patterns and the second triplet (P, D, Q) captures seasonal patterns at period s. Here $s = 12$ means the seasonal cycle is 12 time steps (e.g., monthly data with annual seasonality), $P = 1$ is one seasonal AR lag, $D = 1$ is one seasonal differencing, and $Q = 1$ is one seasonal MA term.

Question 14

You have quarterly GDP data ($s = 4$) with both a trend and strong seasonal pattern. After fitting SARIMA$(0, 1, 1)(0, 1, 1)_4$, the residual ACF shows all values within the confidence bands and no significant spikes. What should you conclude?

A. The model is underfitting because it has no AR terms

B. The model needs more seasonal differencing because GDP always trends upward

C. The residuals resemble white noise, indicating the model adequately captures both trend and seasonal dynamics

D. You should add more MA terms to improve the residuals further

Question 14

You have quarterly GDP data ($s = 4$) with both a trend and strong seasonal pattern. After fitting SARIMA$(0, 1, 1)(0, 1, 1)_4$, the residual ACF shows all values within the confidence bands and no significant spikes. What should you conclude?

A. The model is underfitting because it has no AR terms

B. The model needs more seasonal differencing because GDP always trends upward

C. The residuals resemble white noise, indicating the model adequately captures both trend and seasonal dynamics

D. You should add more MA terms to improve the residuals further

Answer: C

When residual ACF values all fall within the confidence bands with no significant spikes, the residuals behave like white noise. As stated in the lecture, good residuals should show no remaining pattern – 'Good = white noise ACF.' This means the model has successfully captured both the trend (via $d = 1$ differencing) and the seasonal pattern (via $D = 1$ seasonal differencing plus MA terms).

Question 15

In a GARCH(1,1) model $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, the parameters are $\alpha = 0.1$ and $\beta = 0.85$. What does $\alpha + \beta = 0.95$ tell us?

A. Volatility shocks are highly persistent, decaying slowly back to the long-run level

B. The model is non-stationary and will explode over time

C. The model perfectly explains 95% of variance in the data

D. The conditional variance is constant over time

Question 15

In a GARCH(1,1) model $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, the parameters are $\alpha = 0.1$ and $\beta = 0.85$. What does $\alpha + \beta = 0.95$ tell us?

A. Volatility shocks are highly persistent, decaying slowly back to the long-run level

B. The model is non-stationary and will explode over time

C. The model perfectly explains 95% of variance in the data

D. The conditional variance is constant over time

Answer: A

When $\alpha + \beta$ is close to 1 (but still less than 1), volatility shocks are highly persistent – they take a long time to decay back to the unconditional variance. The lecture states that with these typical values ($\alpha \approx 0.1$, $\beta \approx 0.85$), 'one shock keeps volatility elevated for 50+ periods.' The condition $\alpha + \beta < 1$ ensures stationarity of the variance process.

Question 16

In a GARCH(1,1) model with $\alpha = 0.1$, $\beta = 0.85$, and $\omega = 0.5$, a large shock produces $\epsilon_{t-1}^2 = 100$ and the previous conditional variance was $\sigma_{t-1}^2 = 20$. What is σ_t^2?

A. 10.5

B. 120.5

C. 17.0

D. 27.5

Question 16

In a GARCH(1,1) model with $\alpha = 0.1$, $\beta = 0.85$, and $\omega = 0.5$, a large shock produces $\epsilon_{t-1}^2 = 100$ and the previous conditional variance was $\sigma_{t-1}^2 = 20$. What is σ_t^2?

A. 10.5

B. 120.5

C. 17.0

D. 27.5

Answer: D

Applying the GARCH(1,1) formula: $\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 = 0.5 + 0.1 \times 100 + 0.85 \times 20 = 0.5 + 10 + 17 = 27.5$. The large shock ($\epsilon_{t-1}^2 = 100$) has substantially increased the conditional variance from 20 to 27.5, illustrating how GARCH captures the phenomenon that big shocks lead to elevated volatility.

Question 17

In exponential smoothing (ETS), what does a high smoothing parameter α (close to 1) imply?

A. All past observations receive roughly equal weight

B. The forecast relies almost entirely on the long-term average

C. Weights decay slowly, giving substantial influence to distant past observations

D. Recent observations receive much more weight, and older observations are quickly discounted

Question 17

In exponential smoothing (ETS), what does a high smoothing parameter α (close to 1) imply?

A. All past observations receive roughly equal weight

B. The forecast relies almost entirely on the long-term average

C. Weights decay slowly, giving substantial influence to distant past observations

D. Recent observations receive much more weight, and older observations are quickly discounted

Answer: D

In exponential smoothing, weights on past observations decay exponentially. A high α means fast decay – recent data gets far more weight than older data. Conversely, a low α means slow decay, giving more influence to the distant past. The lecture states: 'High alpha = fast decay' and 'Recent data gets more weight – alpha controls decay speed.'

Question 18

You fit three ARIMA models to the same dataset and obtain AIC values of 320, 305, and 312 respectively. Which model should you select, and why?

A. Model 1 (AIC $=$ 320), because higher AIC indicates better predictive power

B. Model 2 (AIC $=$ 305), because lower AIC indicates a better balance of fit and complexity

C. Model 3 (AIC $=$ 312), because it represents a compromise between the other two

D. You cannot compare models using AIC; you must use R^2 instead

Question 18

You fit three ARIMA models to the same dataset and obtain AIC values of 320, 305, and 312 respectively. Which model should you select, and why?

A. Model 1 (AIC = 320), because higher AIC indicates better predictive power

B. Model 2 (AIC = 305), because lower AIC indicates a better balance of fit and complexity

C. Model 3 (AIC = 312), because it represents a compromise between the other two

D. You cannot compare models using AIC; you must use R^2 instead

Answer: B

The lecture states the rule clearly: 'Lower AIC/BIC = better model.' AIC (Akaike Information Criterion) balances goodness of fit against model complexity by penalizing additional parameters. Model 2 with AIC = 305 achieves the best trade-off. Both AIC and BIC penalize complexity, but AIC is preferred for prediction while BIC tends to select the true model.

Question 19

You forecast 10 periods ahead and compute the following errors against actual values: errors = [2, -3, 1, -2, 4, -1, 3, -2, 1, -3]. The MAE (Mean Absolute Error) is:

A. 0, because the positive and negative errors cancel out

B. 22

C. 2.2

D. 4.8

Question 19

You forecast 10 periods ahead and compute the following errors against actual values: errors = [2, -3, 1, -2, 4, -1, 3, -2, 1, -3]. The MAE (Mean Absolute Error) is:

A. 0, because the positive and negative errors cancel out

B. 22

C. 2.2

D. 4.8

Answer: C

MAE is the mean of the absolute values of the errors: MAE $= \frac{|2|+|-3|+|1|+|-2|+|4|+|-1|+|3|+|-2|+|1|+|-3|}{10} = \frac{2+3+1+2+4+1+3+2+1+3}{10} = \frac{22}{10} = 2.2$. Unlike the simple mean of errors (which would be zero here due to cancellation), MAE takes absolute values to measure the average magnitude of forecast errors.

Question 20

Why can you NOT use standard k-fold cross-validation (randomly shuffling data) to evaluate a time series forecasting model?

A. Random shuffling destroys the temporal order, allowing the model to train on future data and predict the past, which inflates accuracy

B. Time series datasets are always too small for k-fold CV

C. k-fold CV only works for classification problems, not regression

D. Cross-validation is unnecessary for time series because AIC is sufficient

Question 20

Why can you NOT use standard k-fold cross-validation (randomly shuffling data) to evaluate a time series forecasting model?

A. Random shuffling destroys the temporal order, allowing the model to train on future data and predict the past, which inflates accuracy

B. Time series datasets are always too small for k-fold CV

C. k-fold CV only works for classification problems, not regression

D. Cross-validation is unnecessary for time series because AIC is sufficient

Answer: A

The lecture highlights that you 'cannot shuffle TS data' because the temporal order is fundamental. Random shuffling would let the model see future observations during training and predict past values during testing, creating data leakage that inflates apparent accuracy. Instead, rolling origin (expanding window) cross-validation must be used: train up to time t, forecast horizon h, then move forward and repeat.