Quiz 1, STATS/DATASCI 531/631 W25

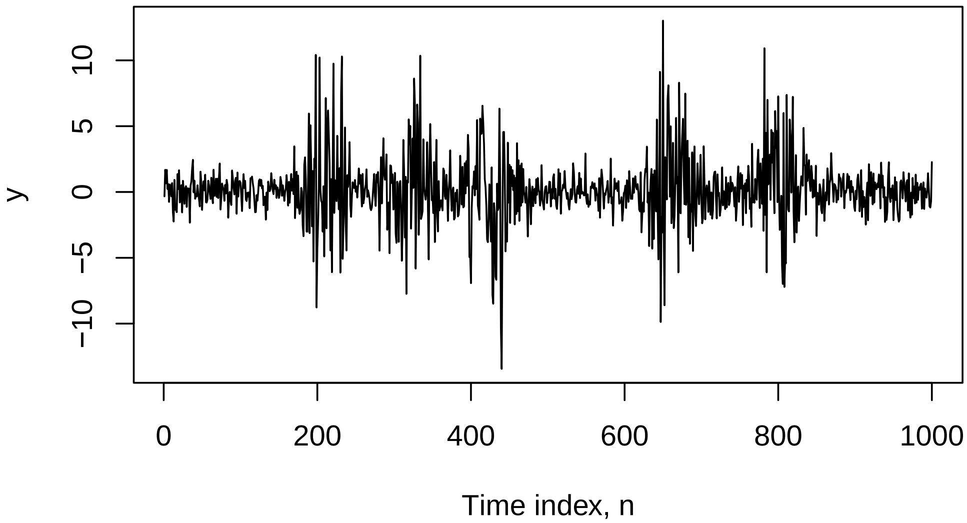
In class on 2/17, 2:30pm to 3:00pm

Name: **UMID:**

Instructions. You have a time allowance of 30 minutes. The quiz may be ended early if everyone is done. The quiz is closed book, and you are not allowed access to any notes. Any electronic devices in your possession must be turned off and remain in a bag on the floor.

For each question, circle one letter answer and provide some supporting reasoning.

Q1. Stationarity and unit roots.



Consider the time series plotted above. Which of the below is the most accurate statement about stationarity?

A. The plot shows that the data are clearly non-stationary. We could make a formal hypothesis test to confirm that, but it would not be insightful. To describe the data using a statistical model, we will need to develop a model with non-constant variance.

B. The sample variance is evidently different in different time intervals. However, we should not conclude that the underlying data generating mechanism is non-stationary before making a formal statistical test of equality of variances between the time regions that have lower sample variance and the regions that have higher sample variance. Visual impressions without a formal hypothesis test can be deceptive.

C. A model with randomly changing variance looks appropriate for these data. Since the variance for such a model is time-varying, the model must be non-stationary.

D. A model with randomly changing variance looks appropriate for these data. Despite the variance for such a model being time-varying, the model is stationary.

E. The sample variance is evidently different in different time intervals. An appropriate next step to investigate stationary would be to plot the sample autocorrelation function for different intervals to see if the dependence between time points is also time-varying.

Q2. Calculations for ARMA models

Let Y_n be an ARMA model solving the difference equation

$$Y_n = (1/4)Y_{n-2} + \epsilon_n + (1/2)\epsilon_{n-1}.$$

This is equivalent to which of the following:

A. $Y_n = (1/2)Y_{n-1} + \epsilon_n$

B. $Y_n = -(1/2)Y_{n-1} + \epsilon_n$

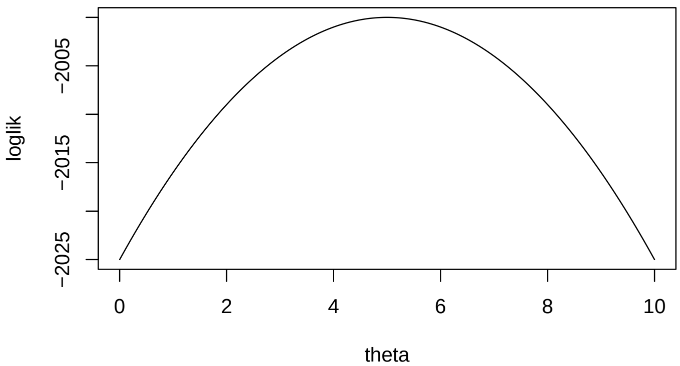
C. $Y_n = (1/2)Y_{n-2} - (1/16)Y_{n-4} + \epsilon_n + \epsilon_{n-1} + (1/4)\epsilon_{n-2}$

D. $Y_n = -(1/2)Y_{n-2} - (1/16)Y_{n-4} + \epsilon_n + \epsilon_{n-1} + (1/4)\epsilon_{n-2}$

E. None of the above

Q3. Likelihood-based inference for ARMA models

The R function `arima()` provides standard errors calculated using observed Fisher information. This question tests your understanding of what that means. Suppose a parametric model has a single parameter, θ, and the log-likelihood function when fitting this model to dataset is as follows:



What is the observed Fisher information (I_{obs}) for θ?

Hint 1. The observed Fisher information is accumulated over the whole dataset, not calculated per observation, so we don't have to know the number of observations, N.

Hint 2. Observations in time series models are usually not independent. Thus, the log-likelihood is not the sum of the log-likelihood for each observation. Its calculation will involve consideration of the dependence, and usually the job of calculating the log-likelihood is left to a computer.

Hint 3. The usual variance estimate for the maximum likelihood estimate, $\hat{\theta}$, is $\text{Var}(\hat{\theta}) \approx 1/I_{obs}$.

A: $I_{obs} = 2$

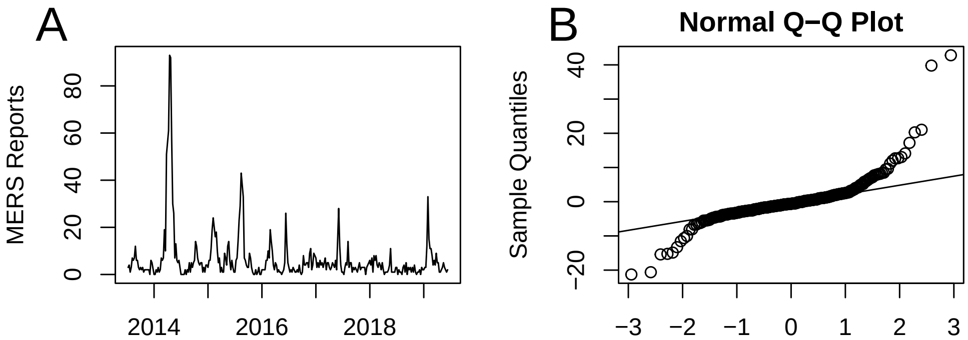
B: $I_{obs} = 1$

C: $I_{obs} = 1/2$

D: $I_{obs} = 1/4$

E: None of the above

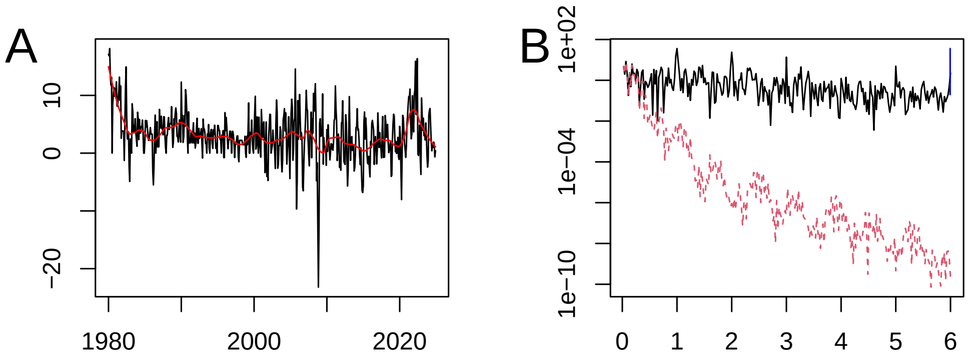
Q4. Interpreting diagnostics



(A) Weekly cases of Middle East Respiratory Syndome (MERS) in Saudi Arabia. (B) a normal quantile plot of the residuals from fitting an ARMA(2,2) model to these data using `arima()`. What is the best interpretation of (B)?

A: We should consider fitting a long-tailed error distribution, such as the t distribution.
B: The model is missing seasonality, which could be critical in this situation.
C: For using ARMA methods, these data should be log-transformed to make a linear Gaussian approximation more appropriate.
D: The normal quantile plot shows a long-tailed distribution, but this is not a major problem. We have over 300 data points, so the central limit theorem should hold for parameter estimates.
E: The normal quantile plot shows long tails, but with the right tail noticeably longer than the left tail. We should consider an asymmetric error distribution.
F: We should not interpret (B) before testing for stationarity. First run `adf.test()` and, if the null hypothesis is not rejected, recalculate (B) when fitting to the differenced data.

Q5. The frequency domain



The monthly US consumer price index (CPI) combines the price of a basket of products, such as eggs and bread and gasoline. (A) Annualized monthly percent inflation, i.e., the difference of log-CPI multiplied by 12×100 (black line); a smooth estimate via local linear regression (red line). (B) The periodogram of inflation and its smooth estimate. Which best characterizes the behavior of the smoother?

A: Cycles longer than 2 months are removed
B: Cycles shorter than 2 months are removed
C: Cycles longer than 2 year are removed
D: Cycles shorter than 2 year are removed
E: Cycles longer than (1/2) year are removed
F: Cycles shorter than (1/2) year are removed

Q6. Scholarship for time series projects

Four people in a team collaborate on a project. After the project is submitted, a reader identifies that part of the project is adapted from an unreferenced source, i.e., it has been plagiarized. The team worked using git and cooperates on tracking down the issue, and the commit history clearly reveals who wrote the problematic part of the project. What is the most appropriate course of action:

A. The guilty coauthor should be penalized heavily for poor scholarship, and the other coauthors should have a minor penalty for failing to check their colleague's work.

B. All coauthors should share the same penalty, since this is a team project and all coauthors share equal responsibility for the submitted report.

C. The guilty coauthor should be penalized heavily for poor scholarship. The other coauthors have demonstrated strong scholarship by following good transparent working practices that enabled this issue to get quickly resolved, so they should not receive any penalty.

D. It is necessary to collect more information before coming to a decision. For example, the team may argue that the source is well known to all readers so did not have to be cited.
