Data Mining and Business Intelligence

Lecture 7: Time Series Properties and Related Diagnostics

Jing Peng
University of Connecticut

3/5/20

Agenda

- Midterm survey: https://uconn.co1.qualtrics.com/jfe/form/SV_aW5tRoxToVVdsNf
- Go over the quiz
- Lecture 6 recap
- Theoretical properties of time series
- Tests for stationarity and white noise
- SAS TSFS

Lecture 6 Recap

AR vs. MA Models

- AR(p): autoregressive model of order p

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

- MA(q): moving average model of order q

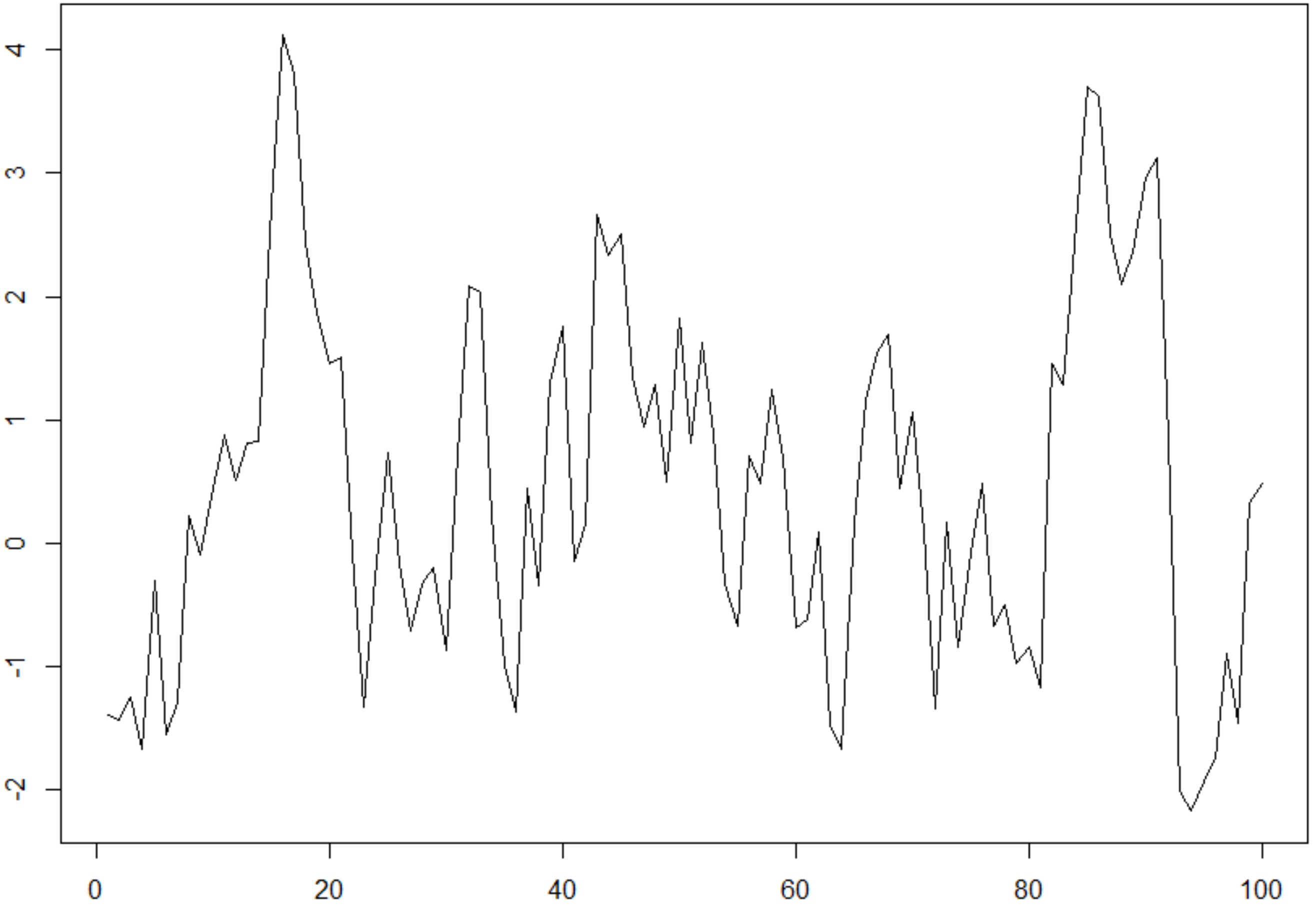
$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

- Examples for each process?

Theoretical Patterns of ACF and PACF

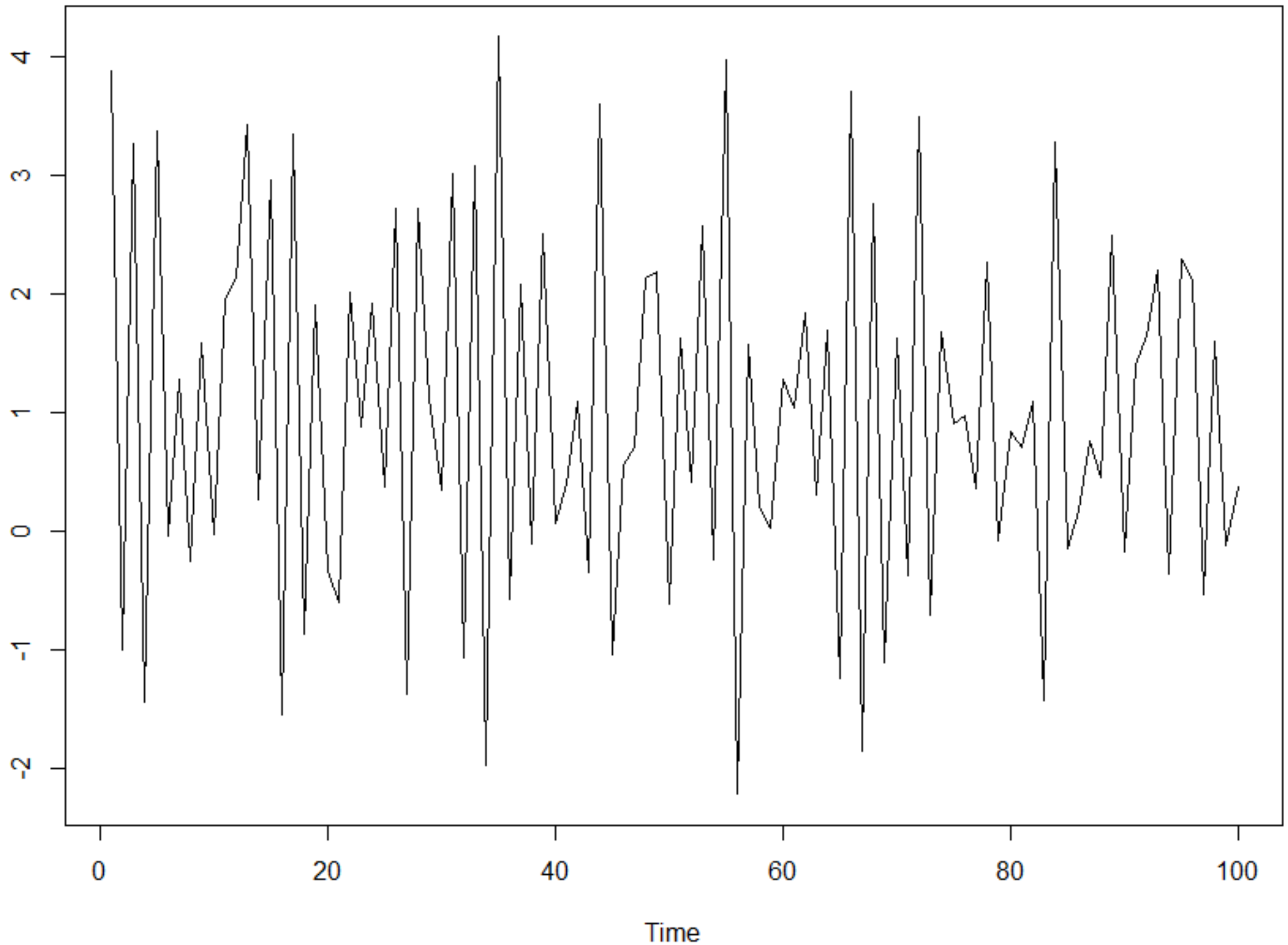
Type of Model	Typical Pattern of ACF	Typical Pattern of PACF
AR (p)	Decays exponentially or with damped sine wave pattern or both	**Cut-off after lags *p***
MA (q)	**Cut-off after lags *q***	Declines exponentially
ARMA (p,q)	Exponential decay	Exponential decay

Visual Inspection of Autocorrelation



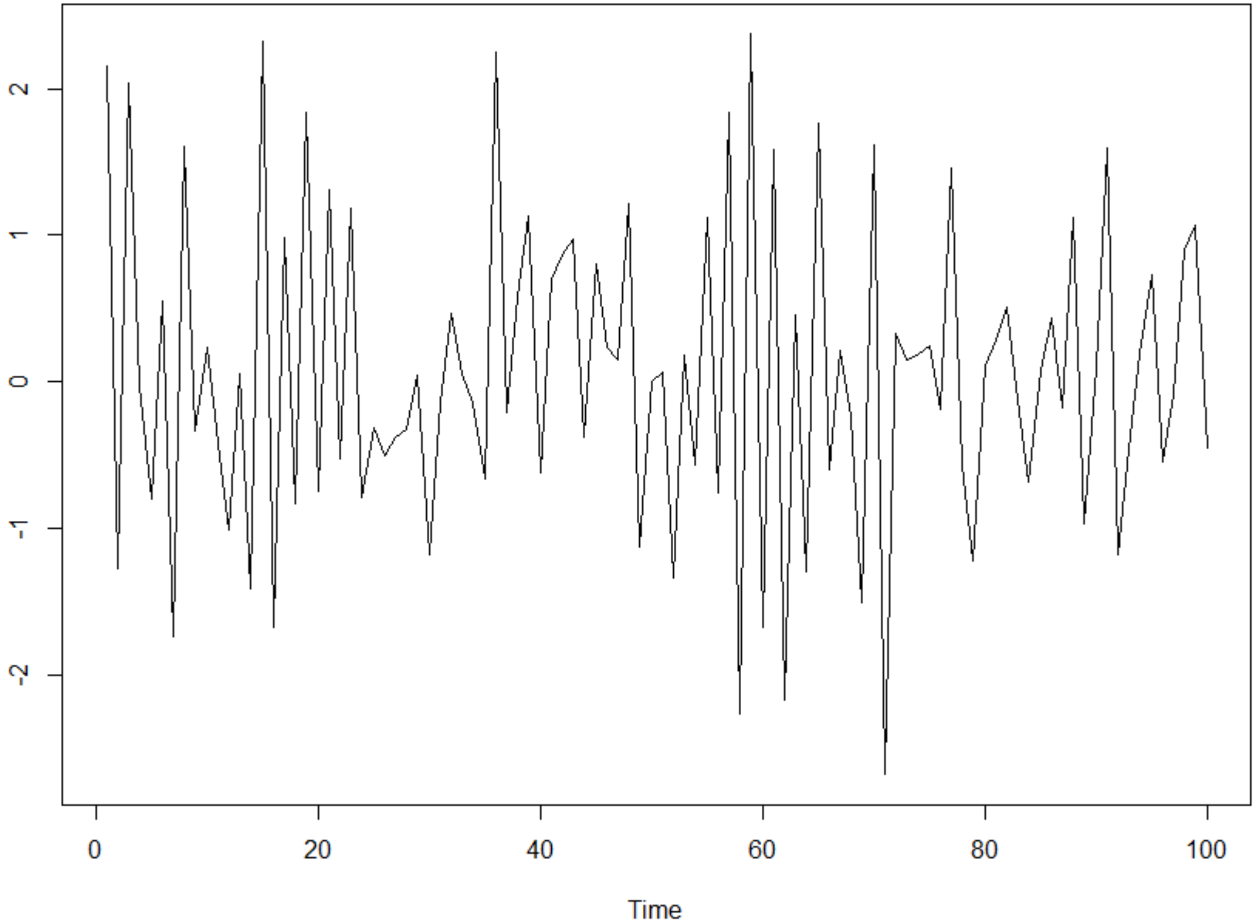
Positive or negative autocorrelation?

Visual Inspection of Autocorrelation



Positive or negative autocorrelation?

Visual Inspection of Autocorrelation



Positive or negative autocorrelation?

Theoretical Properties

Key Properties of Time Series

- How does the mean change over time?
- How does the variance change over time?
- How does the autocorrelation change over time?

Properties of AR Model (no need to memorize)

- Here we focus on AR(1)

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

- Mean: $E[Y_t] = \frac{\phi_0}{1-\phi_1}$
- Variance: $Var(Y_t) = \frac{\sigma^2}{1-\phi_1^2}$
- Autocorrelation: $ACF(k) = \phi_1^k$

https://online.stat.psu.edu/stat510/lesson/1/1.2

Properties of MA Model (no need to memorize)

- Here we focus on MA(1)

$$Y_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, \sigma^2)$$

- Mean: $E[Y_t] = \theta_0$
- Variance: $Var(Y_t) = (1 + \theta_1^2)\sigma^2$
- Autocorrelation: $ACF(1) = \frac{\theta_1}{1+\theta_1^2}$ and ACF(k)=0 for k>1

https://online.stat.psu.edu/stat510/lesson/2/2.1

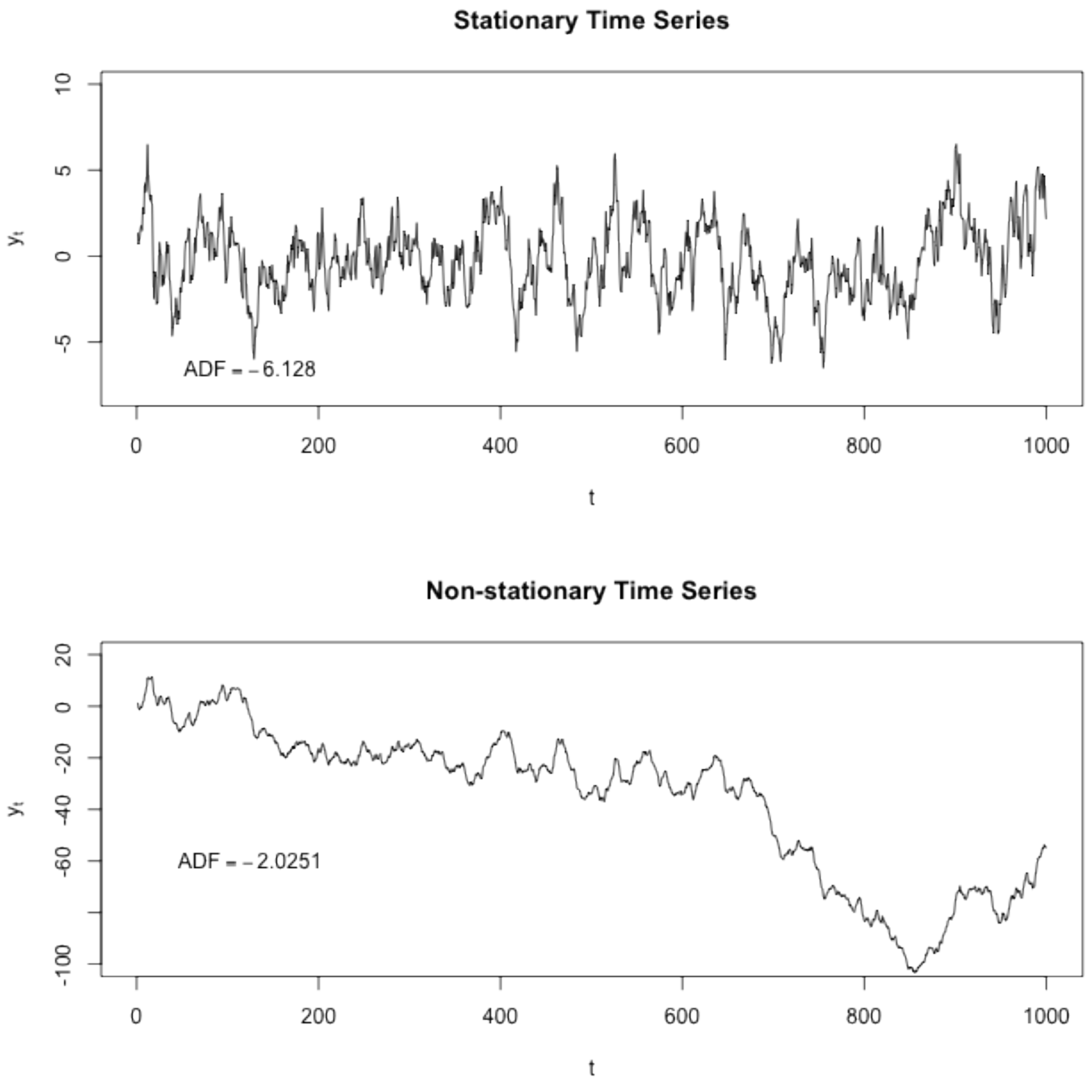
Properties of AR and MA Time Series

- How does the mean change over time?
- How does the variance change over time?
- How does the autocorrelation change over time?
- They are ALL constants that do not change over time!

Stationary Time Series

- Statistical Properties over Time
  - Constant mean (trend is flat)
  - Constant variance (variation is stable)
  - Constant autocorrelation (hard to tell from plots)
- AR and MA time series are both stationary
- A time series with trend or seasonality is NOT stationary

Stationary vs. Non-stationary time series



Source: https://en.wikipedia.org/wiki/Stationary_process#/media/File:Stationarycomparison.png

Simulations on AR and MA Time Series

- Properties.R

- Show how the mean and variance change over time as we repeat the simulations

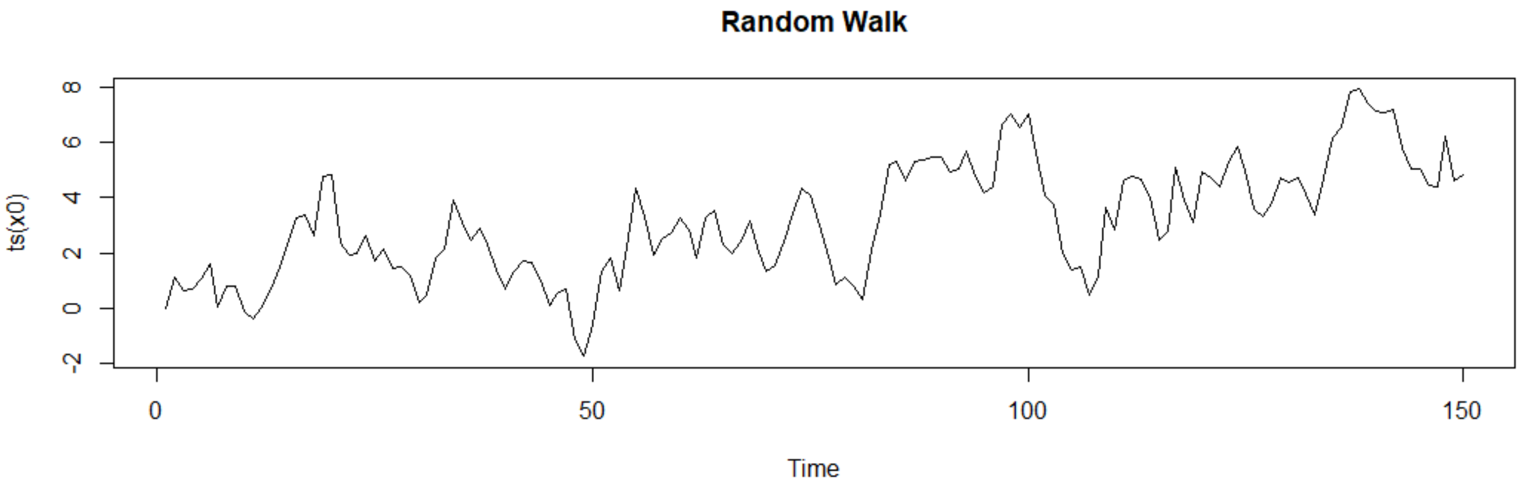
- Is AR(1) stationary when $\phi_1 \geq 1$? NO

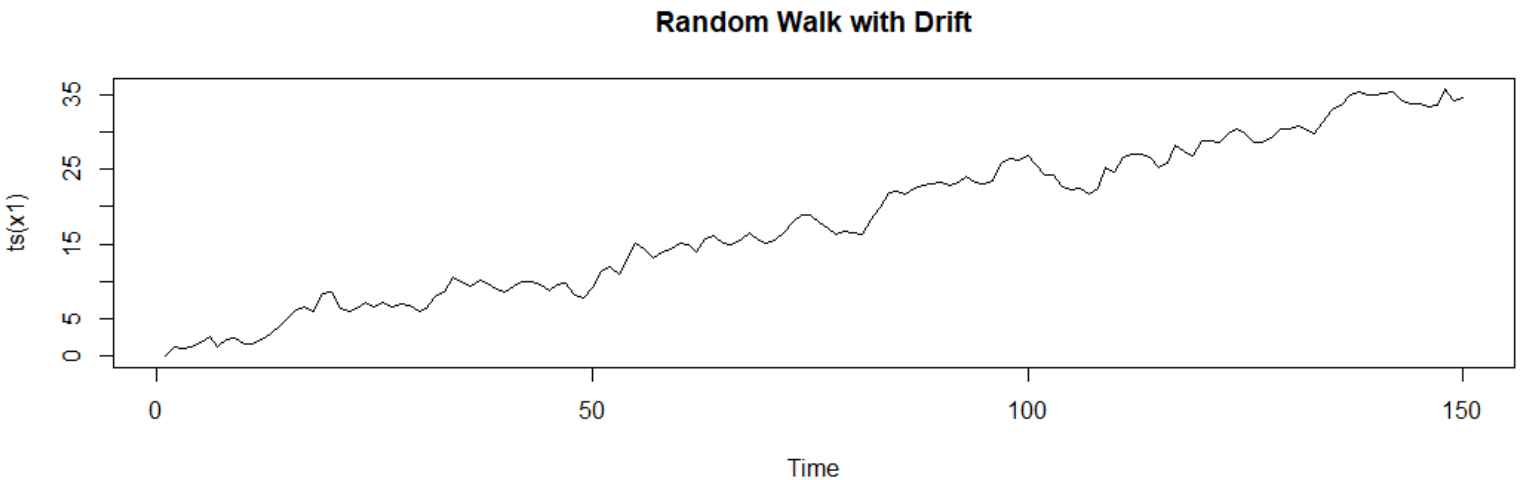
- Is MA(1) stationary when $\theta_1 \geq 1$? YES

Random Walk Models

- Random walk model: $Y_t = Y_{t-1} + \varepsilon_t$
  - The financial status of a gambler
  - Brownian motion
- Random walk with drift: $Y_t = \theta + Y_{t-1} + \varepsilon_t$
  - The financial status of a gambler
  - Brownian motion

Random Walk vs. Random Walk with Drift





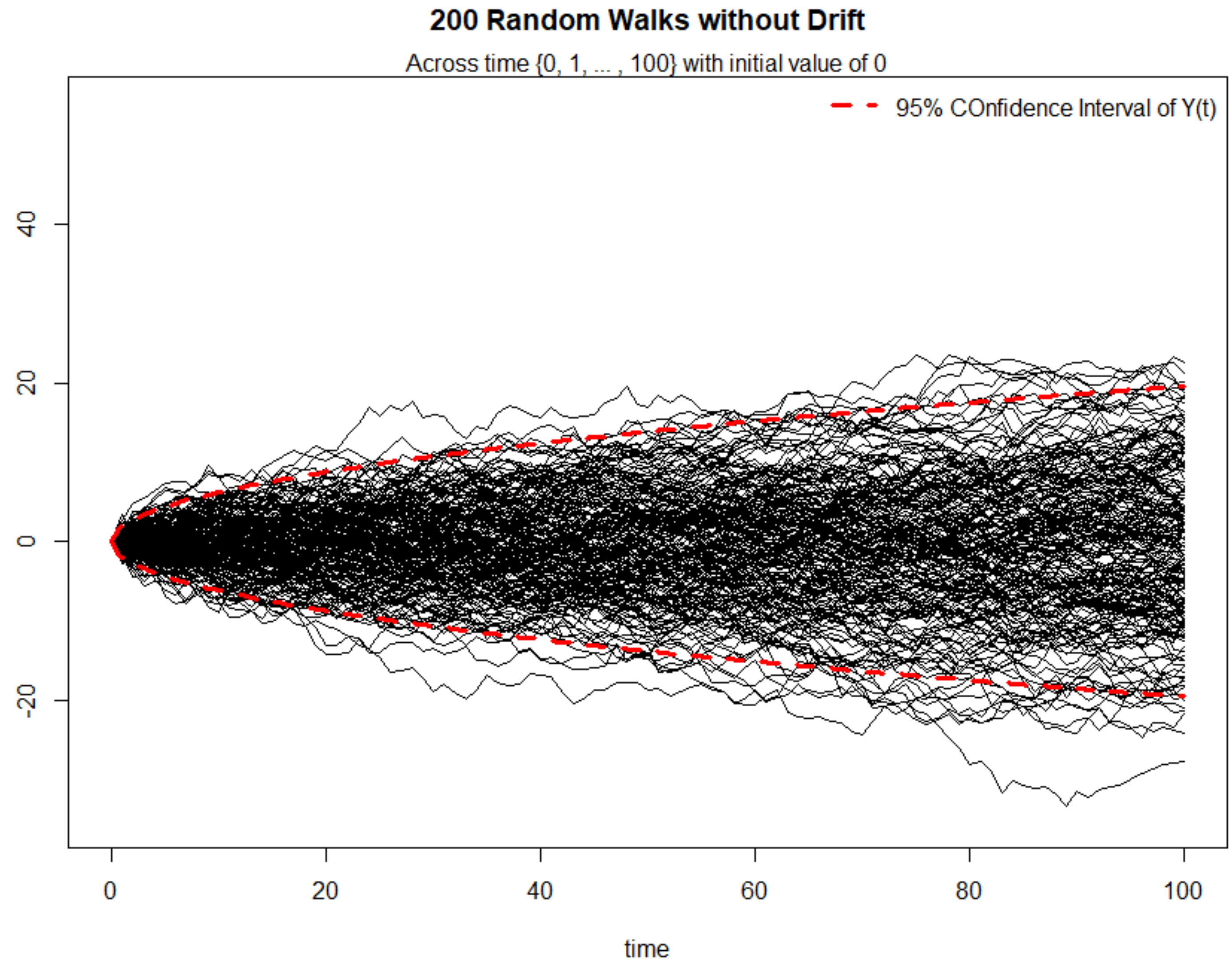
properties.R

Observations from Simulations

- Is a random walk model stationary?
  - Mean stable?
  - Variance stable?

- Is a random walk model with drift stationary?
  - Mean stable?
  - Variance stable?

Mean and Variance of Random Walk over Time



Theoretical Properties of Random Walk Models

- Random walk with drift: $Y_t = \theta + Y_{t-1} + \varepsilon_t$
  - Mean: $E[Y_t] = \theta t$
  - Variance: $Var(Y_t) = t\sigma^2$
- More technical details
  - https://www.youtube.com/watch?v=ouahL4HbwBE
  - https://www.youtube.com/watch?v=Nxcqi7UZenc

Stationarity Tests

Augmented Dickey-Fuller Unit Root Test

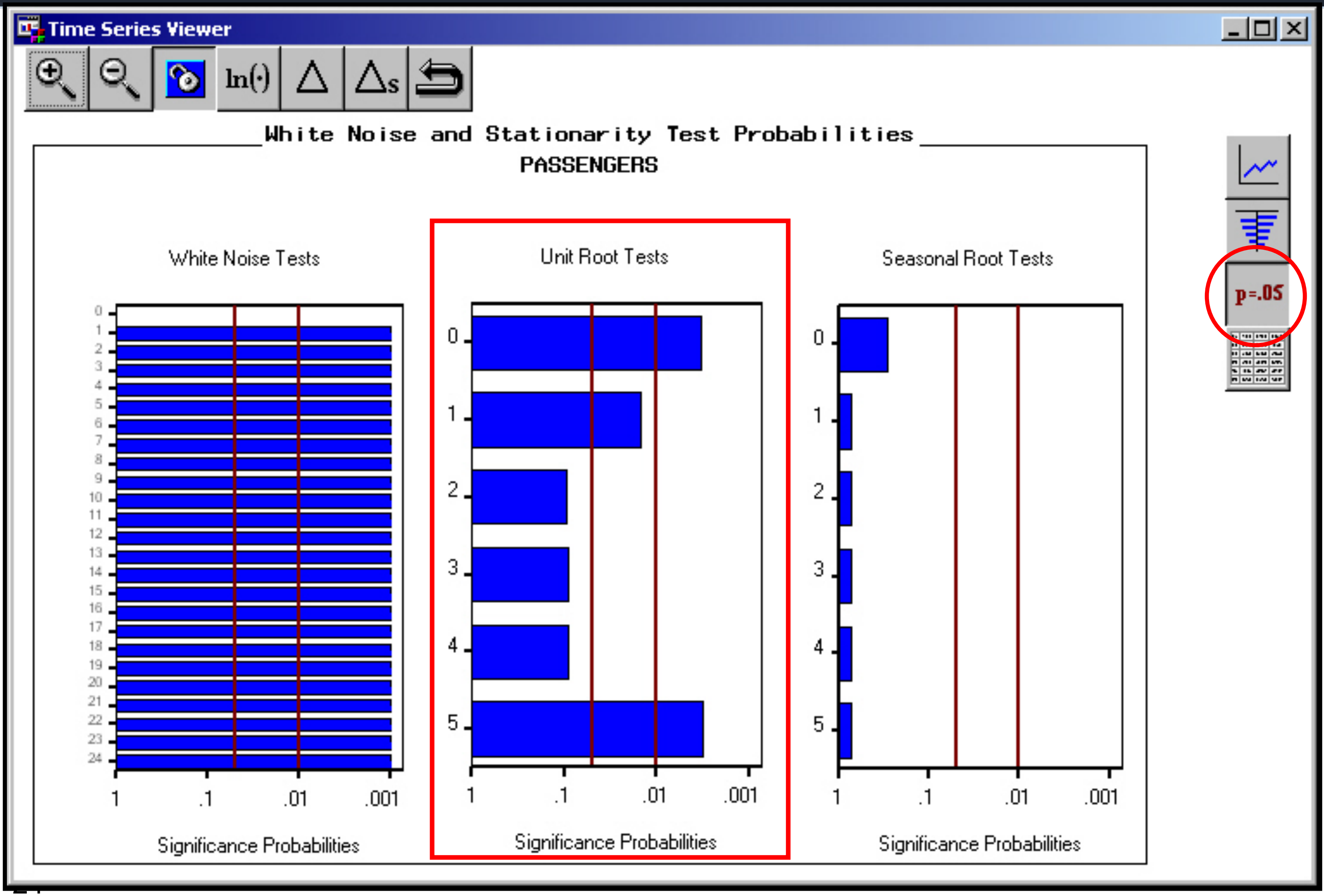
- Boundary condition: AR(1) model with $\phi_1 = 1$ is not stationary

- Unit Root Test: null hypothesis $\gamma = 0$ against the alternative $\gamma < 0$ (one-sided test)

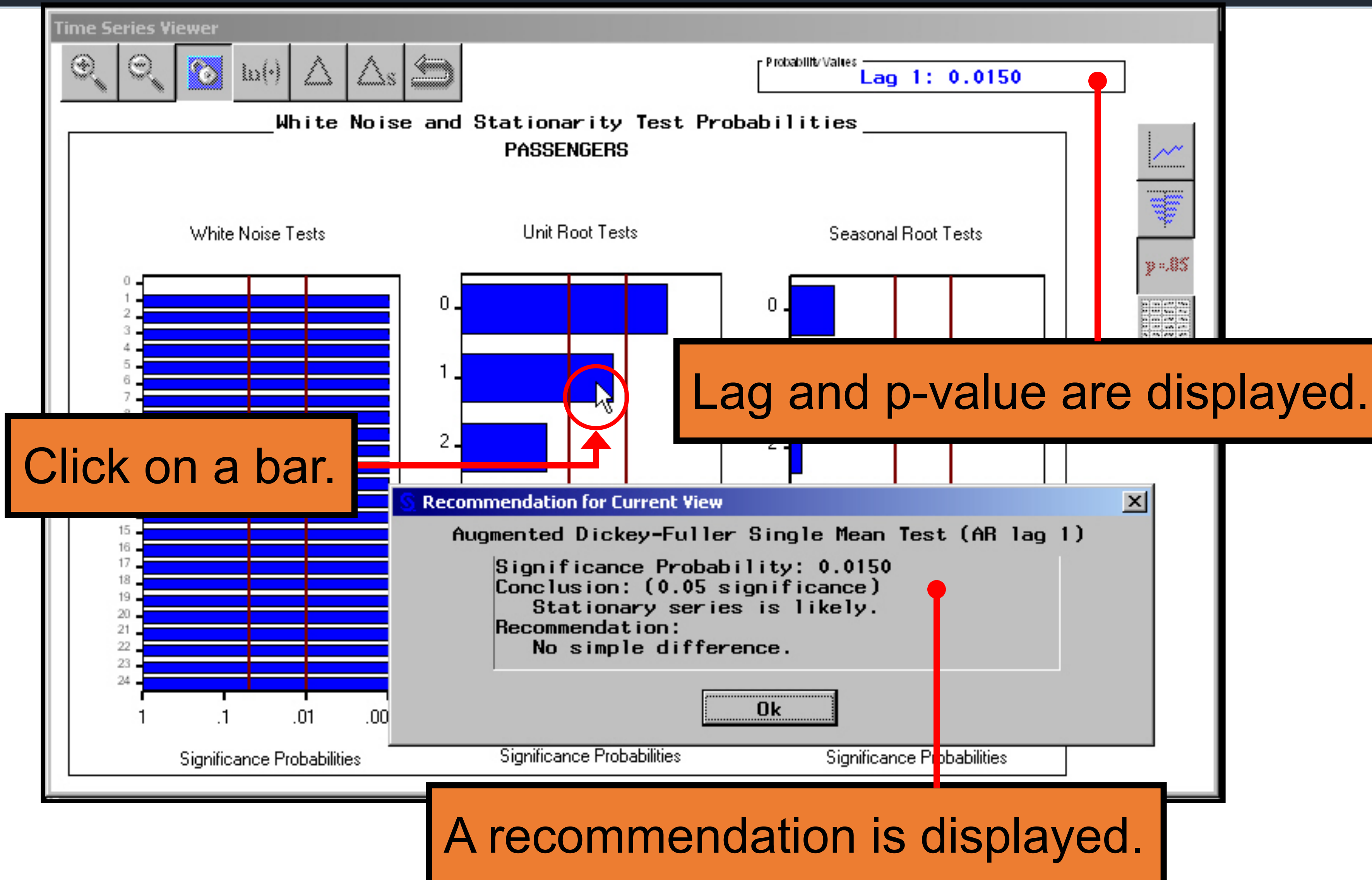
$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

  - The null hypothesis $\gamma = 0$ implies a unit root ($\phi_1 = 1$) and hence acceptance of null suggests nonstationarity
  - Test statistics are available for lags 0 through at most 5. The six tests are not independent, but they may yield conflicting results

Stationarity Tests (Trend)



Interacting with Stationarity Tests



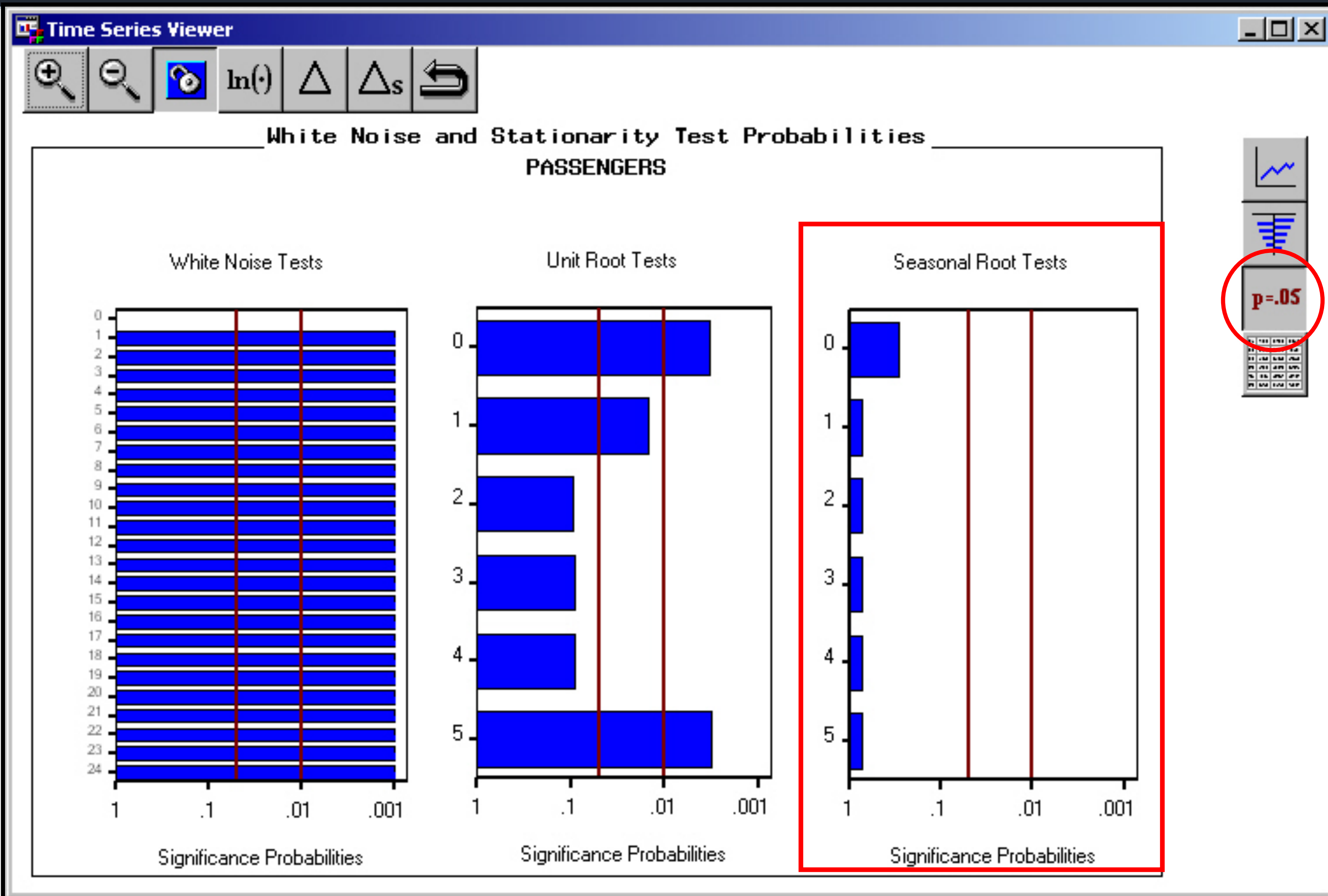
Diagnosing Trend

- A time series having a trend component usually exhibits the following
  - a time series plot that is trending up, down, or in a deterministic fashion
  - a highly significant ACF, PACF, and IACF at lag 1
  - an ACF with many significant lags decaying slowly from lag 1
  - an ACF with few significant values after first differencing is applied
  - unit root tests that are not significant but become significant when a first difference is applied

Diagnosing Seasonality

- A time series with a seasonal component having a period S usually exhibits the following:
  - a time series plot that has repetitive behavior every S time units
  - significant ACF, PACF, and IACF values at lag S
  - an ACF with significant values at lags that are multiples of S
  - seasonal root tests that are not significant but become significant when a difference of order S is applied

Stationarity Tests (Seasonality)



White Noise Test

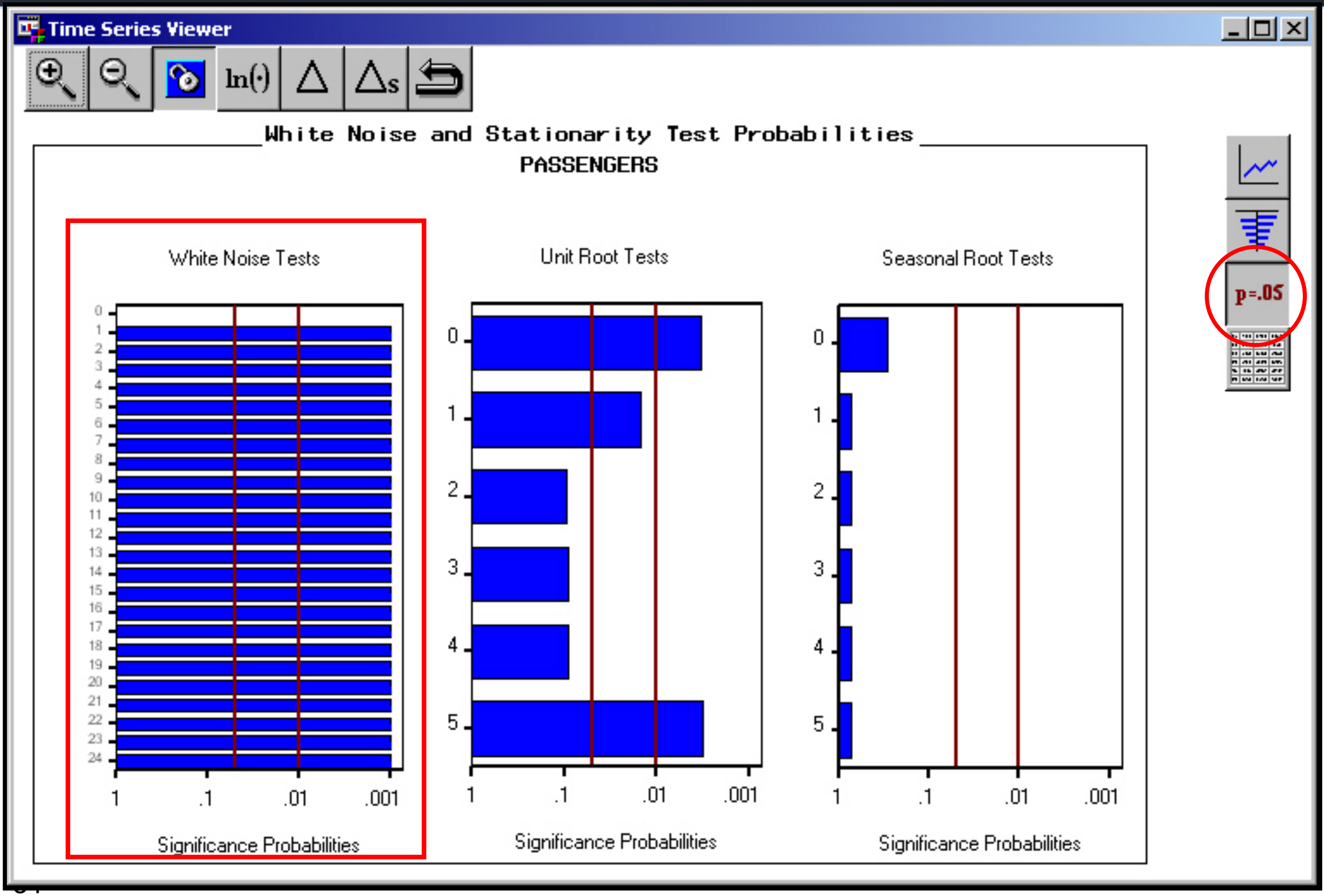
The Ljung-Box Chi-Square Test for White Noise

- A *white noise* time series in which the observations are i.i.d. normal

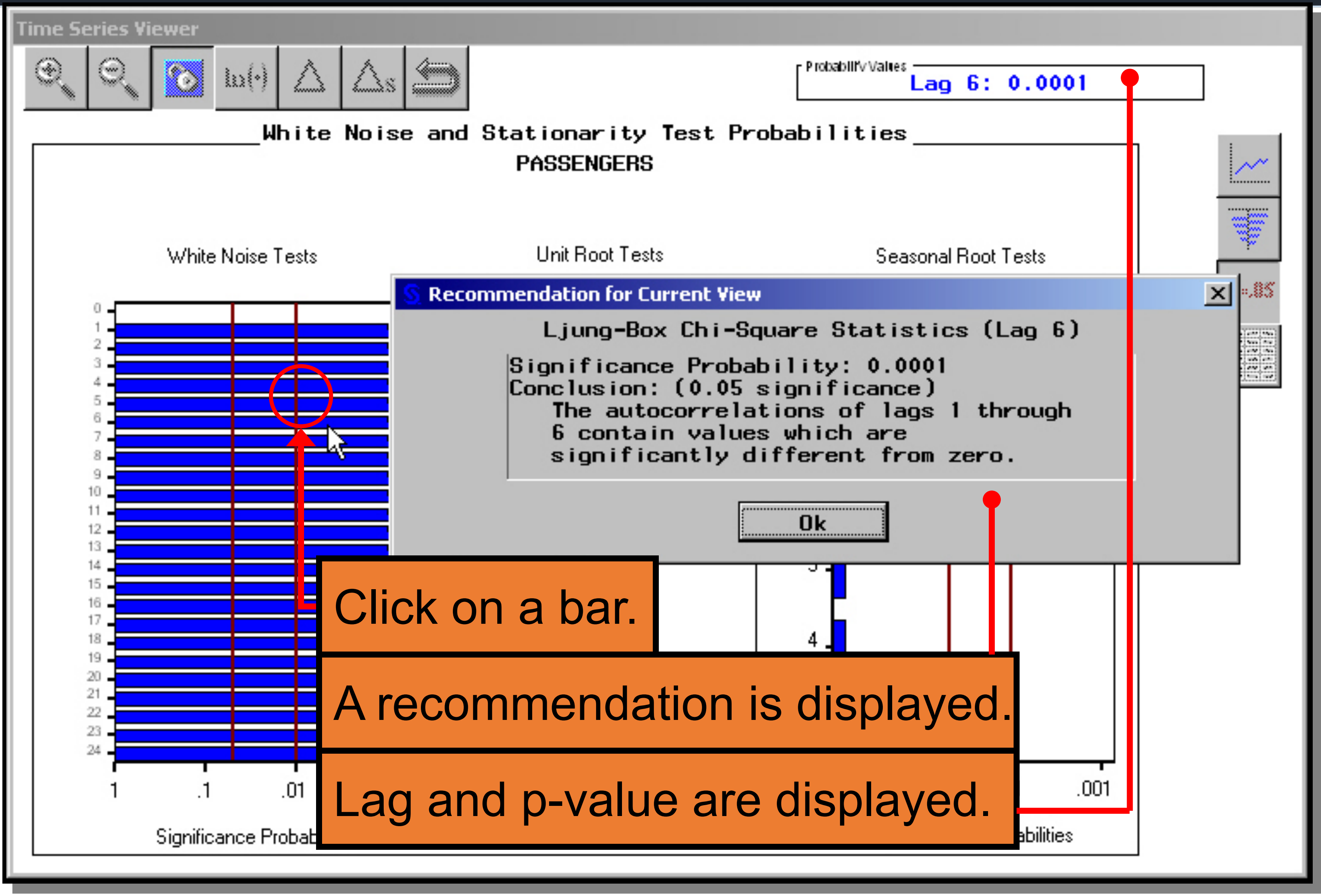
- The Ljung-Box test may be applied to the original series (or to the residuals after fitting a model)

- Reject the null that the series is white noise if $p<0.05$

White Noise Tests



Interacting with White Noise Tests



Take Away

- White Noise Test: insignificant p-values (larger than 0.05) indicates white noise because the NULL is white noise

- Unit/Seasonal Root Test: significant p-values (less than 0.05) suggests stationarity because the NULL indicates nonstationarity

Readings

- https://online.stat.psu.edu/stat510/lesson/1/1.2
- https://online.stat.psu.edu/stat510/lesson/2/2.1
- Forecasting chapter 1
- http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#tfintro_toc.htm