Problem Set 1

Econ 6376 — Time Series Econometrics

Parameters

- Feel free to work in groups, however your code and write ups must be your own.
- Submissions should be made through Blackboard and include two files:
  - Your write up as a single PDF, **no more than 10 pages**.
  - Your **organized, commented** R script that reproduces every result in your write up.
- Your submission is due by 11:59pm on the due date listed in Blackboard.
- Late submissions **WILL NOT** be accepted unless previously arranged.
- Make sure your answer is sufficient for a reader that is not familiar with the problem and/or time series to understand your solution.
- Treat your write up like a professional report. If you wouldn't turn it into your boss, don't turn it into me.

Grading

Total: 165 points.

- **Presentation quality: 20 points** — answers should be well organized, clearly written, and free of grammatical errors. Graphs should be clearly labeled, easy to read, and appropriately titled.
- **R script quality: 20 points** — code should be well commented, organized, reproducible top-to-bottom, and produce every figure and number in your write up.
- **Question 1: 25 points**
  - (a) 3 points — function correctness and reproducibility.
  - (b) 6 points — ACF definition, interpretation, and AR(1) derivation.
  - (c) 8 points — 4$×$2 grid, labeling, and per-series visual description.
  - (d) 5 points — 50-realization overlay and interpretation.
  - (e) 3 points — identification of distinguishing features between $\phi = 0.95$ and $\phi = 1.0$.
- **Question 2: 40 points**
  - (a) 30 points — three series, 10 points each (classification + evidence-based justification).
  - (b) 10 points — trend-stationarity workflow.

- **Question 3: 60 points**
  - (a) 15 points — Monte Carlo simulation implementation.
  - (b) 15 points — 2$×$2 grid of empirical distributions with theoretical overlays.
  - (c) 10 points — rejection-rate table for all four cells.
  - (d) 20 points — synthesis across the four cells.

Question 1 — Simulating and inspecting an AR(1) *(after Lecture 1)*

Consider the AR(1) process

$$y_t = \phi\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{iid}{\sim} N(0,1), \qquad y_0 = 0.$$

(a) Write a function `ar1_sim(phi, T)` that returns a length-T realization of this process. Set $T = 500$ throughout this question and use `set.seed(6376)` at the top of your R script.

(b) Define the autocovariance $\gamma(k)$ and autocorrelation $\rho(k)$ of a weakly stationary series. State (i) what $\rho(k)$ measures in plain language, (ii) the value of $\rho(0)$ and the bounds on $\rho(k)$, and (iii) derive the theoretical ACF of the AR(1) process, $\rho(k) = \phi^{|k|}$.

(c) For each of $\phi \in \{0,\ 0.5,\ 0.95,\ 1\}$, simulate one realization using your function. Produce a 4×2 grid of plots in which each row corresponds to one value of ϕ: the left panel shows the time series and the right panel shows its sample ACF out to lag 20. Label panels clearly and choose figure dimensions appropriate for a reader. In 1–2 sentences per row, describe what changes in the *visual character* of the series and of its ACF as ϕ increases.

(d) For $\phi = 0.95$, simulate **50 independent realizations** and overlay their time plots on a single figure (light-colored lines, no legend). In 2–3 sentences, explain what this picture illustrates about the relationship between a DGP and a sample path, and specifically why a single series can look "trending" even when the DGP is mean-reverting.

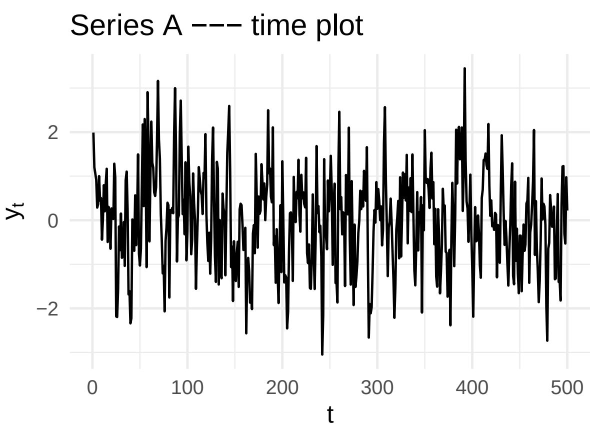
(e) The case $\phi = 1.0$ is the random walk. Without running any formal test, list two features of your plots in (c) and (d) that distinguish the $\phi = 1.0$ series from the $\phi = 0.95$ series, and explain why the distinction matters for later analysis.

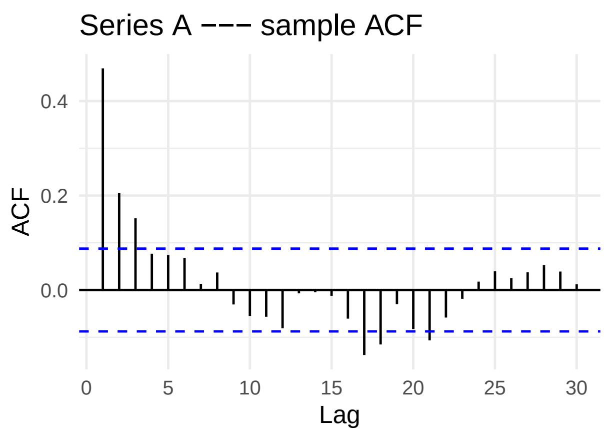
Question 2 — Diagnosing stationarity from output *(after Lecture 3)*

For each of the three series below, the following output has been provided: a time plot, the sample ACF, the Augmented Dickey-Fuller test statistic (no trend term), the KPSS test statistic (level specification), and the sample standard deviations of the series in levels and in first differences.

Note: KPSS is covered in the lecture notes but was not discussed in class. You are responsible for reading that section before answering.

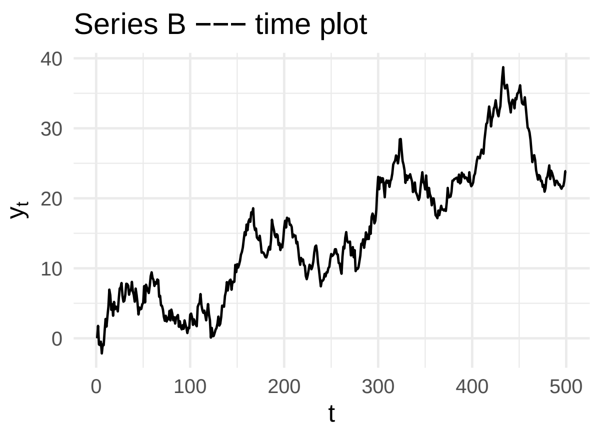
Series A

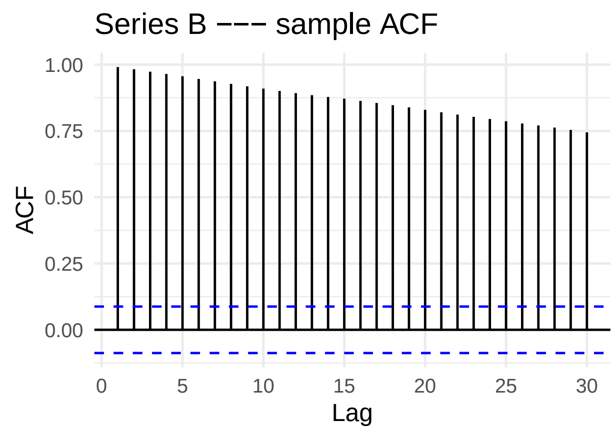




Statistic	Value
ADF statistic (no trend)	-6.618
KPSS statistic (level)	0.211
sd(levels)	1.111
sd(first differences)	1.142
Number of observations	500

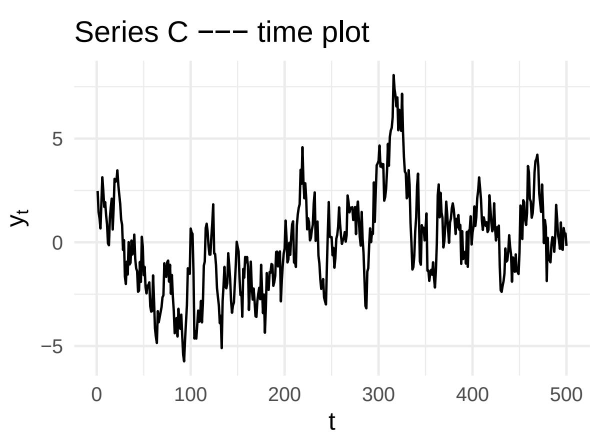
Series B

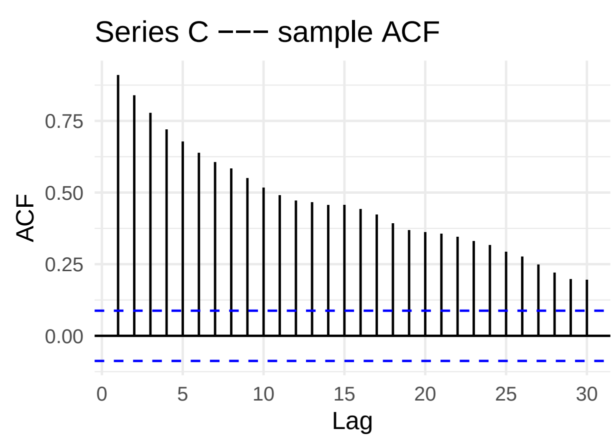




Statistic	Value
ADF statistic (no trend)	-2.669
KPSS statistic (level)	6.984
sd(levels)	9.546
sd(first differences)	1.021
Number of observations	500

Series C





Statistic	Value
ADF statistic (no trend)	-3.836
KPSS statistic (level)	2.090
sd(levels)	2.297
sd(first differences)	0.967
Number of observations	500

Questions

(a) For each of Series A, B, and C, classify the series as **stationary** or **non-stationary**. Justify your call using at least three pieces of the provided evidence. Note any evidence that appears to conflict with your classification and explain how you weighted it.

(b) Suppose you are handed a time series whose plot shows an apparent trend. Describe the steps you would take to determine whether it is **trend-stationary** or **non-stationary**. Be explicit about (i) what transformation you would apply to the series, (ii) what diagnostics you would run on the transformed series, and (iii) how you would map those diagnostic results back to a classification.

Question 3 — Spurious regression: a Monte Carlo study *(after Lecture 3)*

Granger and Newbold (1974) showed that regressions involving non-stationary time series can produce misleading results. You will reproduce the core finding via simulation and extend it to isolate what drives the problem.

Setup. Set $T = 500$, $N_{\text{sim}} = 10{,}000$, and $\alpha = 0.05$. Use `set.seed(6376)` at the top of your script. For each of the four cells below, perform N_{sim} iterations of the following procedure:

1. Generate two *independent* series $\{x_t\}$ and $\{y_t\}$ of length T under the DGPs specified for the cell. Use $N(0, 1)$ innovations throughout.
2. Estimate $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{e}_t$ by OLS.
3. Record the t-statistic on $\hat{\beta}_1$.

Cell	DGP for x_t	DGP for y_t
1	Random walk	Random walk
2	AR(1), $\phi = 0.7$	AR(1), $\phi = 0.7$
3	White noise ($\phi = 0$)	White noise ($\phi = 0$)
4	AR(1), $\phi = 0.7$	Random walk

(a) Implement the simulation. You are encouraged, but not required, to write a single function that takes the two DGPs as arguments and returns the vector of N_{sim} t-statistics.

(b) Plot the empirical distribution of the t-statistics for each cell as a 2×2 grid of histograms. On each panel, overlay a t-distribution with the appropriate degrees of freedom. Label axes and titles so a reader unfamiliar with the exercise can orient.

(c) For each cell, report the empirical rejection rate at the 5% level — the proportion of simulations in which $|t|$ exceeds the two-sided critical value. Present your results as a four-row table with columns: cell, DGP description, empirical rejection rate, and expected rejection rate under the classical linear regression assumptions.

(d) In 4–6 sentences, synthesize your findings across the four cells. Your answer must address: (i) what the white-noise cell (cell 3) confirms; (ii) why cell 2's rejection rate departs from 5% despite both series being stationary; (iii) what cell 4 demonstrates that cells 1–3 alone do not; and (iv) the practical implication for an analyst running a regression without having checked stationarity first.