Econometrics

Practical Session 14

Introduction to Time Series and AR Models

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics

Theoretical Wrap-up

What is Time Series Data?

- **Time series** $y_1, y_2, \ldots, y_T$:
  - **ordered** observations
  - on the **same variable**
  - over T **evenly-spaced periods** (monthly, quarterly, annually, ...)
  - referred to the **same unit** (country, firm, individual, ...)
- **Purpose**: **predict** the present y_t and ultimately **forecast** y_{T+h} from past values patterns $y_{t-1}, y_{t-2}, \ldots$

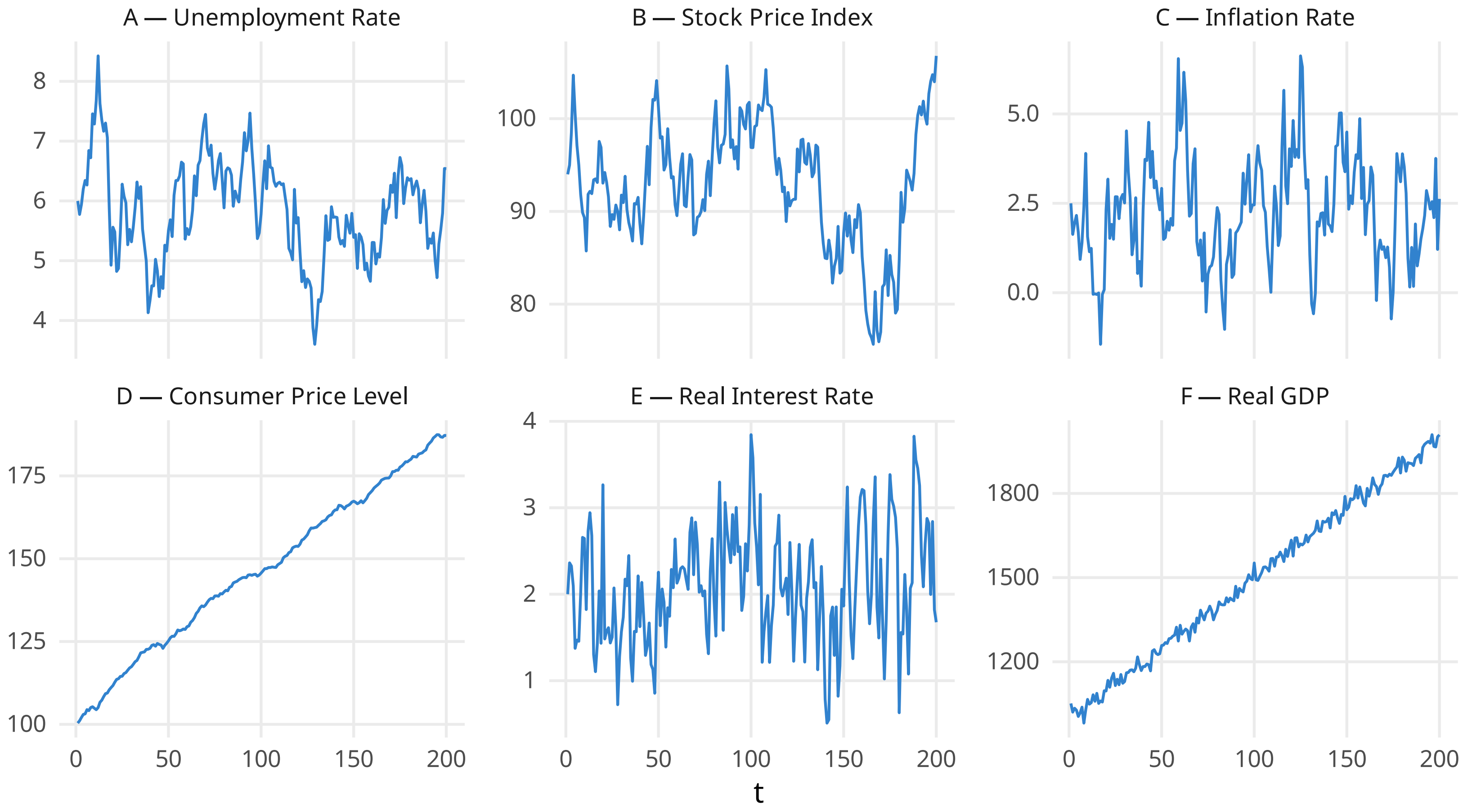
Key Concepts

1. **Autocorrelation** (*aka* serial correlation) $\longrightarrow$ $y_t = f(y_{t-1}, y_{t-2}, \ldots)$
2. **Stationarity** $\longrightarrow$ each y_t is a realization of the same distribution
   - Two types:
     1. **Strong** $\longrightarrow$ joint distribution of $(y_{s+1}, \ldots, y_{s+T})$ **independent from** s
     2. **Weak** $\longrightarrow$ usually enough:

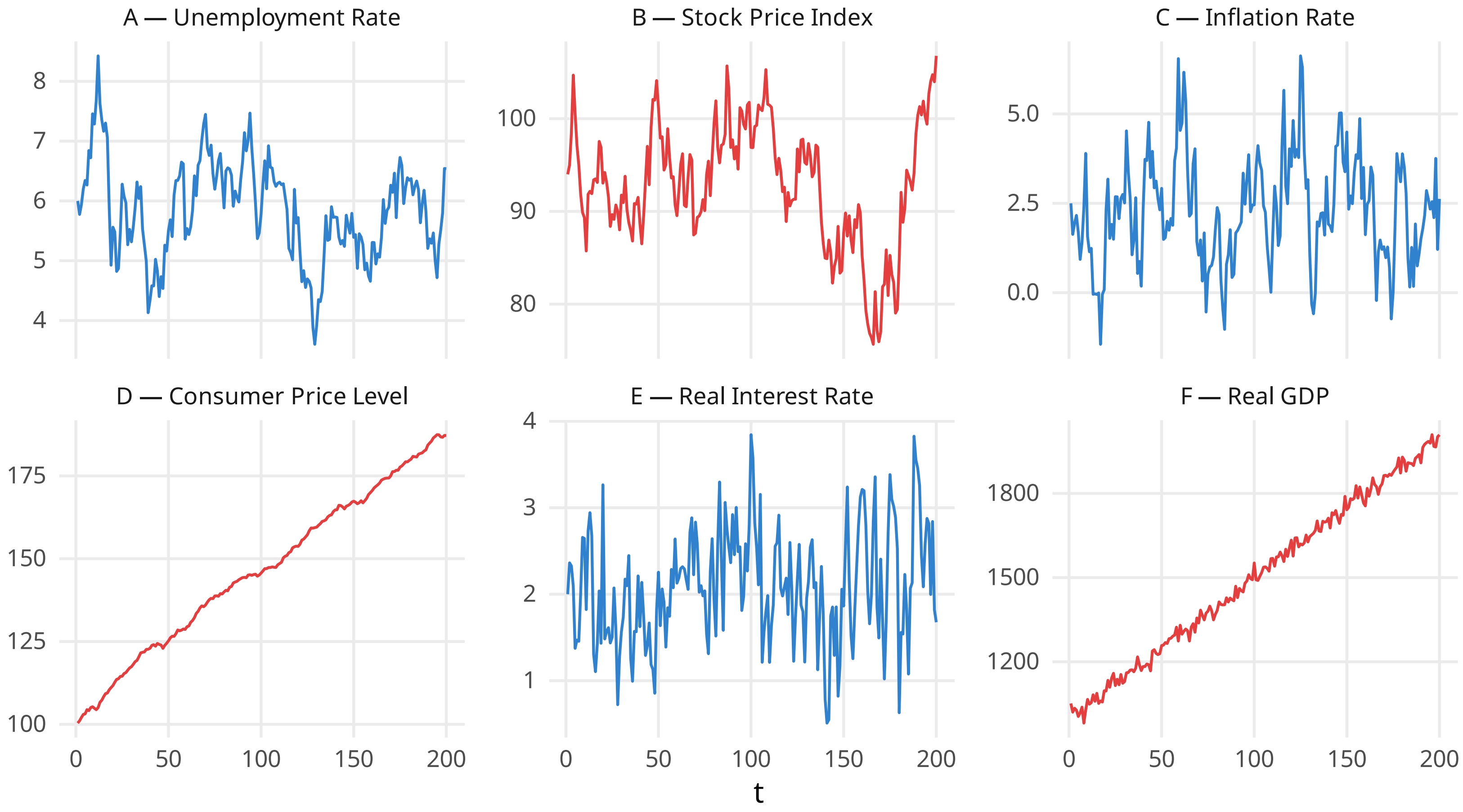
$$\mathbb{E}[y_t] = \mu, \quad \text{var}(y_t) = \sigma^2, \quad \text{cov}\left(y_t, y_{t-j}\right) = \gamma_j, \quad \forall t$$

   - Ensures that **historical patterns remain valid** for the future

Stationary or Non Stationary?



Stationary or Non Stationary?



Measuring Autocorrelation Under Stationarity

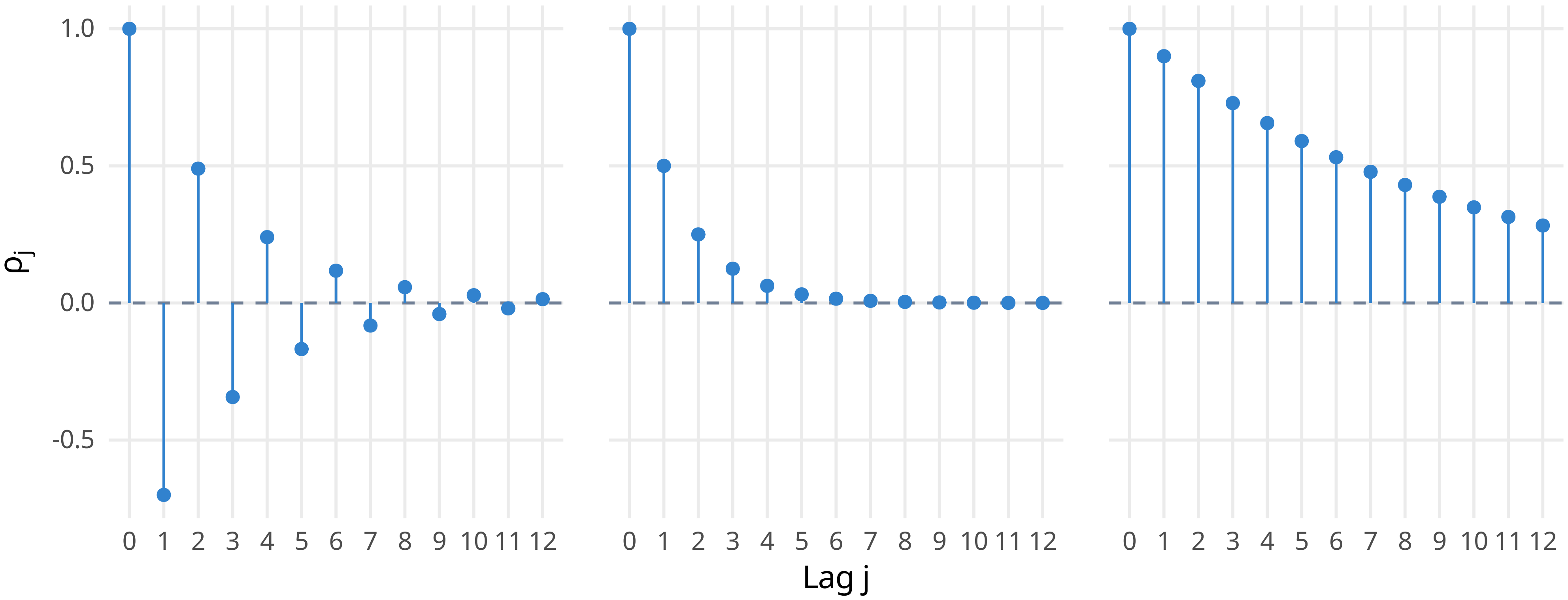
- The ***j*-th autocorrelation** measures the **linear dependence** between y_t and its j-th lag:
- Under (weak) stationarity, the **autocorrelation coefficient** is:

$$\rho_j = \frac{\gamma_j}{\sigma_y^2}, \quad \gamma_j = \mathbb{E}\big[(y_t - \mu)\big(y_{t-j} - \mu\big)\big]$$

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{s_y^2}, \quad \hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}(y_t - \overline{y})(y_{t-j} - \overline{y})$$

- **Correlogram (ACF plot):** plot of $\hat{\rho}_j$ vs. $j \longrightarrow$ allows to see the pattern

Measuring Autocorrelation Under Stationarity



The AR(1) Model: Autoregressive, Linear, 1 Lag

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}(0, \sigma^2)$$

- A **white noise** is an **innovation process** $\{\varepsilon_t\}$ such that:

$$\mathbb{E}[\varepsilon_t] = 0, \quad \text{var}(\varepsilon_t) = \sigma^2, \quad \text{cov}(\varepsilon_t, \varepsilon_s) = 0, \quad \forall s < t$$

- It allows to **apply OLS** to estimate the coefficients
  - β_0, β_1 have **no causal interpretation** $\longrightarrow$ purely predictive
  - **Prediction and inference** as before

The AR(1) Model: Autoregressive, Linear, 1 Lag

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{WN}\left(0, \sigma^2\right)$$

$|\beta_1| < 1$
Stationary

- Shocks temporary
- $\mathbb{E}[y_t] = \beta_0/(1 - \beta_1)$
- $\text{var}(y_t) = \sigma^2/\left(1 - \beta_1^2\right)$
- $\rho_j = \gamma_j/\gamma_0 = \beta_1^j$

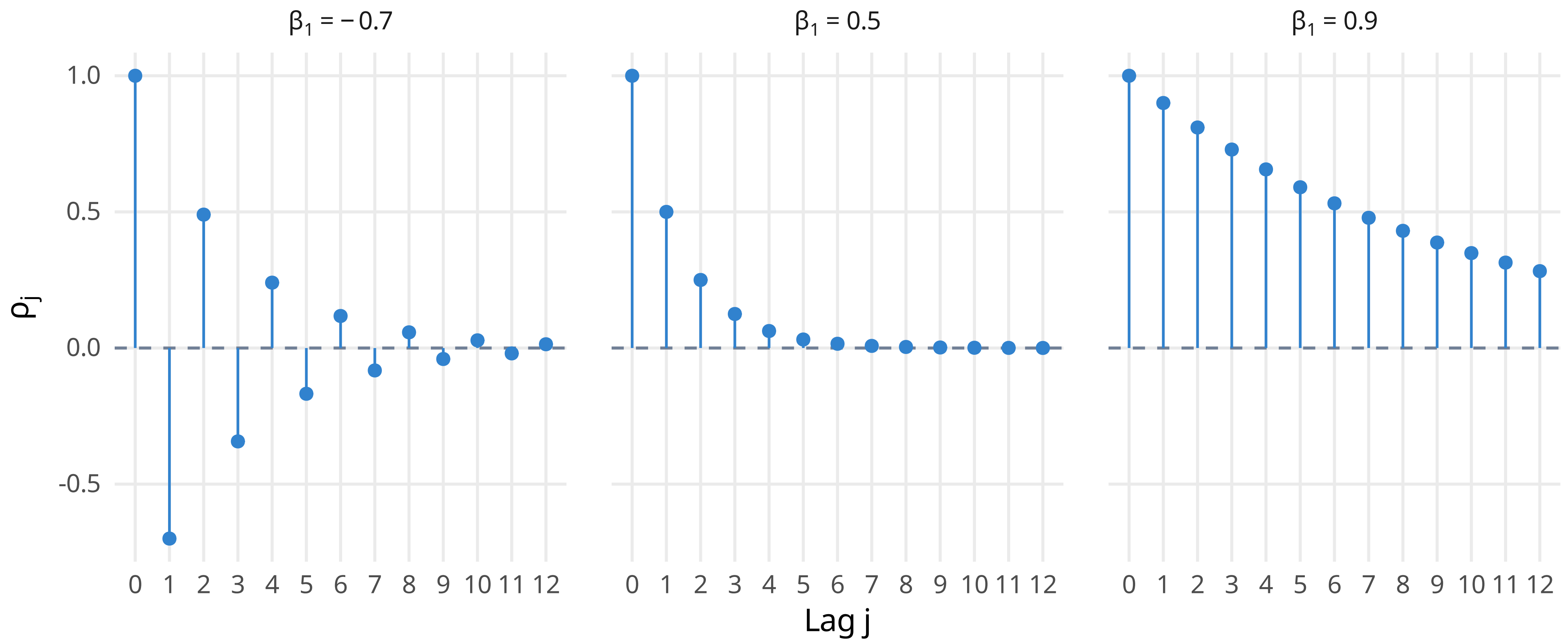
$\beta_1 = 1$
Unit root

- Shocks permanent
- $\mathbb{E}[y_t] = y_0 + t\beta_0$
- $\text{var}(y_t) = t\sigma^2 \to \infty$
- $\rho_j = \frac{\gamma_j}{\gamma_0} = 1$

$|\beta_1| > 1$
Explosive

- Rarely observed
- Shocks amplified
- $\mathbb{E}[y_t] \to \infty$

The AR(1) Model: Autoregressive, Linear, 1 Lag



Forecasting

- **In-sample fitted value:** $\hat{y}_t$ for $t = 1, \ldots, T$
- **Out-of-sample *h*-step-ahead forecast:**

$$\hat{y}_{T+h|T} = \mathbb{E}[y_{T+h} \mid y_T, y_{T-1}, \ldots]$$

- **Mean Squared Forecast Error (MSFE):**

$$\text{MSFE} = \mathbb{E}\left[\left(y_{T+h} - \hat{y}_{T+h|T}\right)^2\right]$$

- In time series analysis only **predictive accuracy** matters $\longrightarrow$ **minimise the MSFE**

Exercises

Exercise 1 | The Labor Share of Income in the US

The file `us_annual.csv` contains a dataset with annual US data from 1960 to 2024 for:

- `m_ell`: labor share of income — compensation of employees as a share of GDP
- `p_Y`: GDP price deflator (index, 2017 = 100)

a) Plot the labor share and a line with the sample mean. Does the series look stationary?

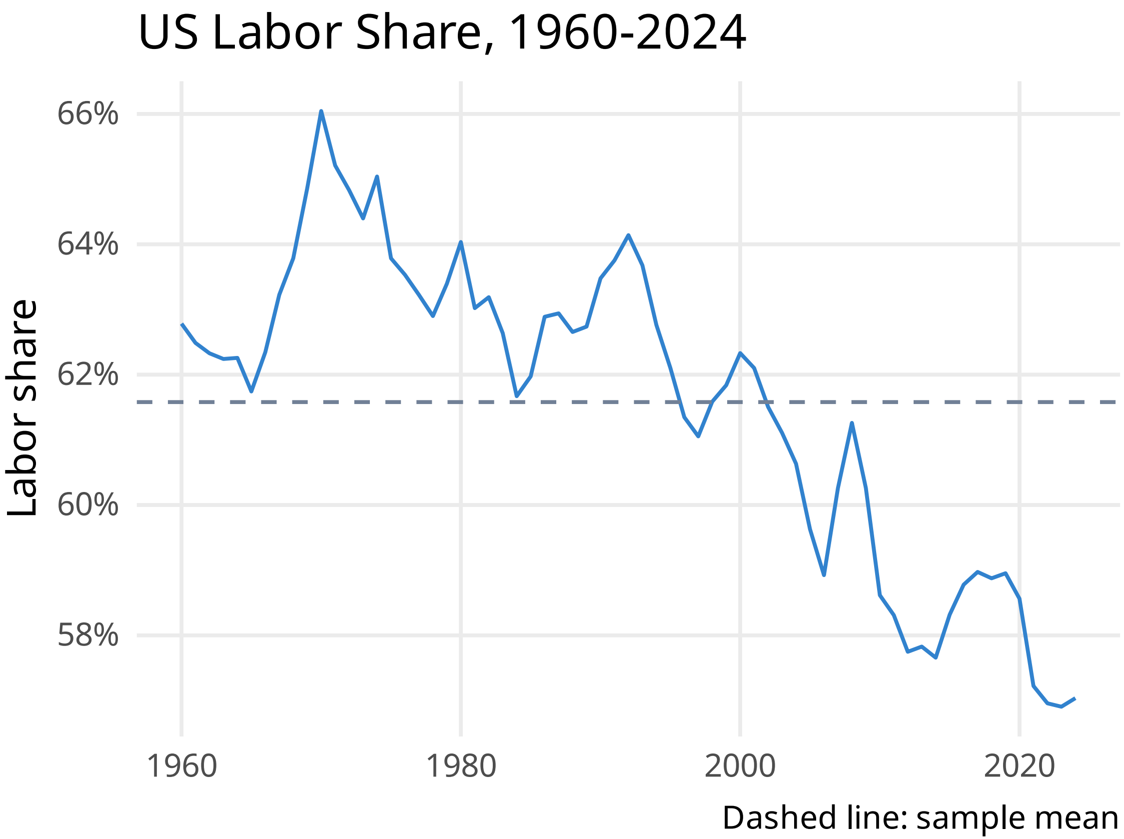
```r
source("aux/plotting.R")
library(tidyverse)
```

Exercise 1 | The Labor Share of Income in the US

```r
us <- read_csv("data/us_annual.csv")

us |>
  ggplot(aes(year, m_ell)) +
  geom_line(color = "#0061FF") +
  geom_hline(yintercept = mean(us$m_ell),
             linetype = "dashed", color = "gray40") +
  scale_y_continuous(labels = scales::percent_format(accuracy = 1))
+
  labs(x = NULL, y = "Labor share",
       title = "US Labor Share, 1960-2024",
       caption = "Dashed line: sample mean")
```

Exercise 1 | The Labor Share of Income in the US



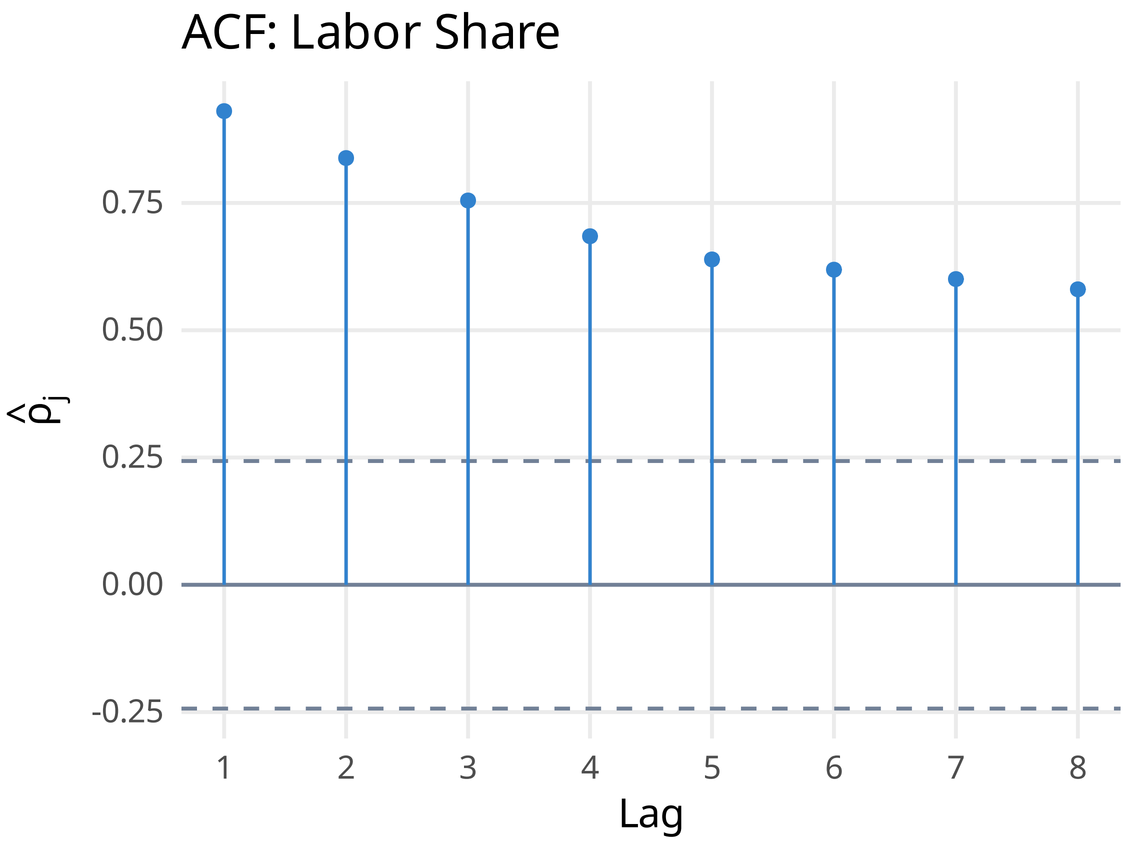
- Clear declining trend (~66% to ~57%)
- The mean is shifting, which is inconsistent with stationarity

Exercise 1 | The Labor Share of Income in the US

b) Plot the ACF of the Labor Share up to the 8th lag. What does it tell you about persistence?

```r
plot_acf(us$m_ell, lag.max = 8, title = "ACF: Labor Share")
```

Exercise 1 | The Labor Share of Income in the US



- $\hat{\rho}_1 \approx 0.93 \longrightarrow$ very high persistence
- The labor share barely changes year to year

Exercise 1 | The Labor Share of Income in the US

c) Should we trust forecasts based on patterns from the 1960s?

- A shifting mean casts doubt on stationarity
- The AR(1) can be a useful approximation but forecasts should be treated with caution

Exercise 1 | The Labor Share of Income in the US

d) Estimate an AR(1) for the Labor Share. What is $\hat{\beta}_1$? Is it statistically significant?

```
us_ell <- us |>
  mutate(m_ell_l1 = lag(m_ell)) |>
  filter(year <= 2023) |>
  drop_na()

ar1_ell <- lm(m_ell ~ m_ell_l1, data = us_ell)
summary(ar1_ell)
```

Exercise 1 | The Labor Share of Income in the US

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.004041   0.023149   0.175    0.862
m_ell_l1    0.991941   0.037479  26.466   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- $\hat{\beta}_1 \approx 0.99$ and highly significant (p-value ≈ 0)
- The labor share is extremely persistent; shocks decay very slowly

Exercise 1 | The Labor Share of Income in the US

e) What is the long-run mean of the labor share? What does it tell you?

- The long-run mean correspond to the expected value:

$$\mathbb{E}[m_\ell] = \widehat{m}_\ell = \frac{\hat{\beta}_0}{1 - \hat{\beta}_1} = 50.1\%$$

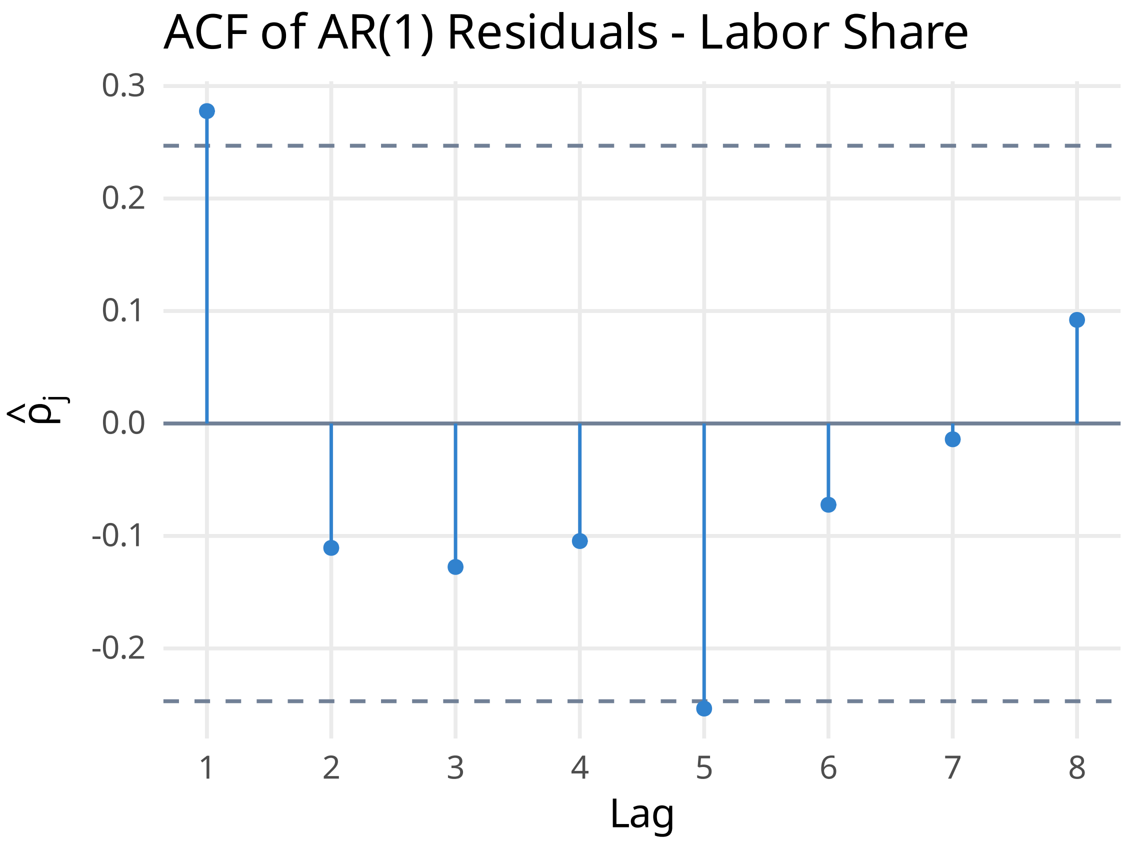
- Long-run mean ~50%, **below** the current level (~57%) $\longrightarrow$ consistent with the downward trend; the model expects further decline

Exercise 1 | The Labor Share of Income in the US

f) Plot the ACF of the residuals of the AR(1) estimated above. Are there any significant spikes? Is the model well specified?

```
plot_acf(residuals(ar1_ell), lag.max = 8, title = "ACF of AR(1)
Residuals")
```

Exercise 1 | The Labor Share of Income in the US



- No significant spikes → residuals appear white noise
- AR(1) captures the autocorrelation structure well
- $\bar{R}^2 \approx 0.92$: the lagged value explains most of the variation in the current one

Exercise 1 | The Labor Share of Income in the US

g) Use the estimated model to forecast the labor share in 2024, using 2023 as the last observation. Compute the forecast error. How do you interpret it?

```
last_obs <- us_ell |> filter(year == 2023) |> pull(m_ell)
actual   <- us_ell |> filter(year == 2024) |> pull(m_ell)
b0_hat   <- coef(ar1_ell)["(Intercept)"]
b1_hat   <- coef(ar1_ell)["m_ell_l1"]
forecast <- b0_hat + b1_hat * last_obs
```

Exercise 1 | The Labor Share of Income in the US

```
Last observed (2023):    0.5691
Forecast for 2024:       0.5685
Actual (2024):           0.5704
Forecast error:          0.0019
```

- Error ≈ 0.002 ⟶ tiny, because high persistence means $y_t \approx y_{t-1}$
- Near-unit-root series are easy to forecast one step ahead, but this tells us little about the model's structural validity

Exercise 2 | US Inflation

The dataset also contains `p_Y`, the GDP price deflator (index, 2017 = 100).

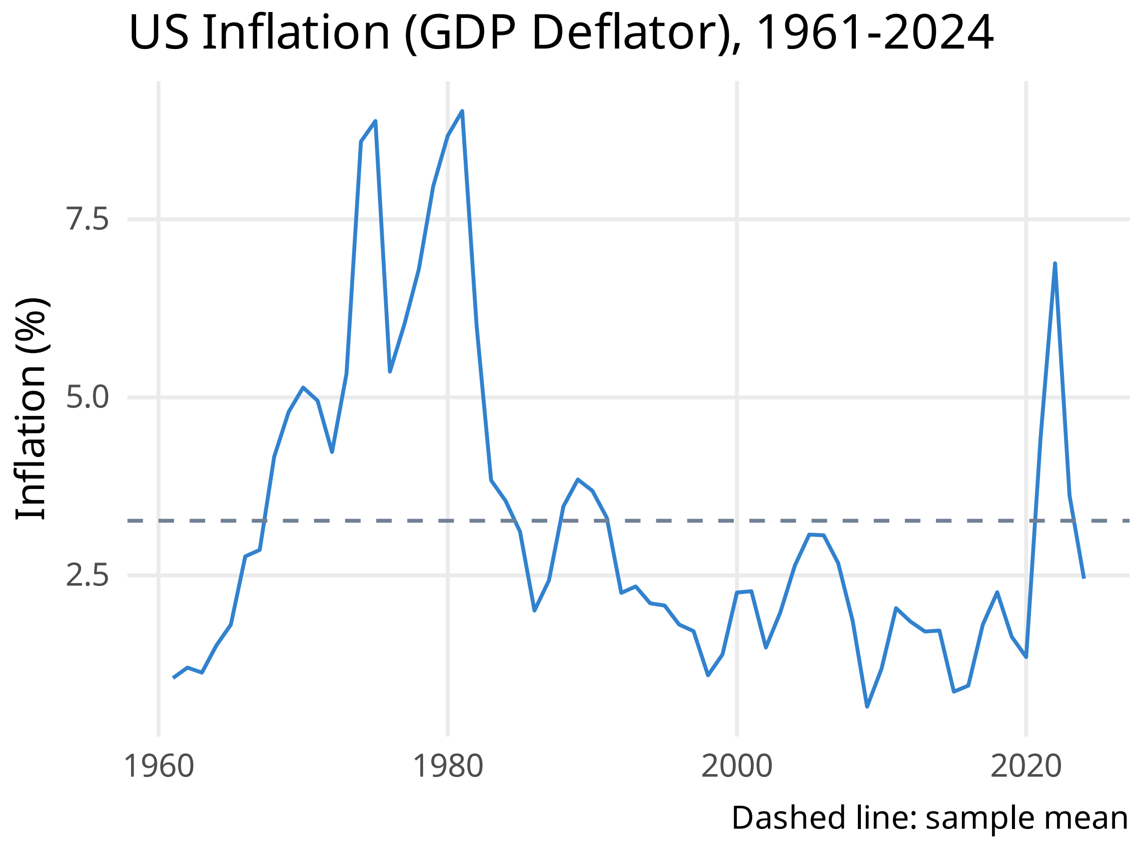
a) Compute the annual inflation rate as $\pi_t = 100 \times \left(\ln p_{Y,t} - \ln p_{Y,t-1}\right)$ and plot it with a line for the sample mean. Does the series look stationary?

```
us_pi <- us |>
  mutate(pi = 100 * (log(p_Y) - lag(log(p_Y)))) |>
  drop_na()
```

Exercise 2 | US Inflation

```r
us_pi |>
  ggplot(aes(year, pi)) +
  geom_line(color = "#0061FF") +
  geom_hline(yintercept = mean(us_pi$pi),
             linetype = "dashed", color = "gray40") +
  labs(x = NULL, y = "Inflation (%)",
       title = "US Inflation (GDP Deflator), 1961-2024",
       caption = "Dashed line: sample mean")
```

Exercise 2 | US Inflation



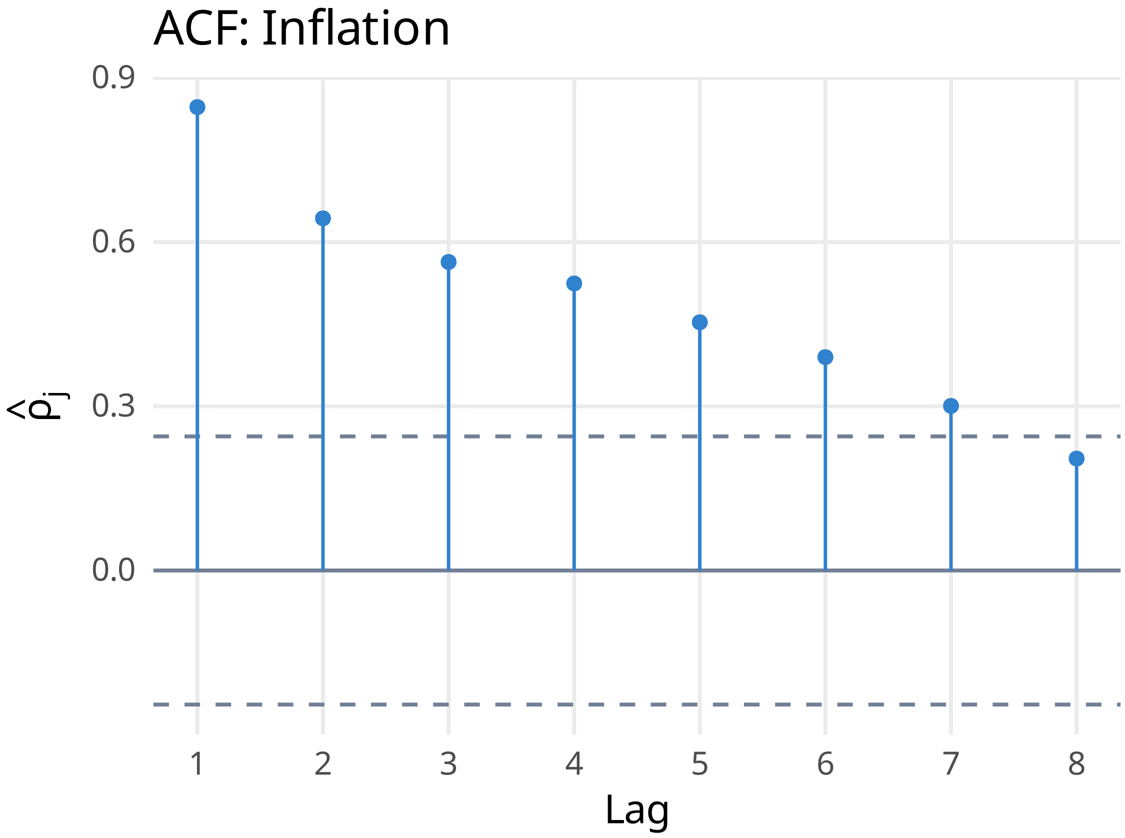
- Series fluctuates around a relatively stable mean → more consistent with stationarity than the labor share
- Clear inflation surge in the 1970s–80s and again in 2021–22

Exercise 2 | US Inflation

b) Plot the ACF of inflation up to the 8th lag. Is inflation persistent? How does $\hat{\rho}_1$ compare to the labor share?

```
plot_acf(us_pi$pi, lag.max = 8,
         title = "ACF: Inflation")
```

Exercise 2 | US Inflation



- $\hat{\rho}_1 \approx 0.85 \longrightarrow$ persistent, but clearly decaying
- Less persistent than the labor share ($\hat{\rho}_1 \approx 0.93$)
- More consistent with stationarity

Exercise 2 | US Inflation

c) Estimate an AR(1) for inflation on the pre-COVID sample (up to 2019). What is $\hat{\beta}_1$? Is it statistically significant?

```
us_pi_l <- us_pi |>
  mutate(pi_l1 = lag(pi)) |>
  drop_na()

ar1_pi <- lm(pi ~ pi_l1, data = us_pi_l |> filter(year <= 2019))
summary(ar1_pi)
```

Exercise 2 | US Inflation

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.36474    0.23273   1.567    0.123
pi_l1        0.89099    0.05947  14.982   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- $\hat{\beta}_1 \approx 0.89$ and statistically significant (p-value ≈ 0)
- Persistent but clearly below 1 $\longrightarrow$ inflation is stationary, unlike the labor share

Exercise 2 | US Inflation

CATÓLICA LISBON BUSINESS & ECONOMICS

d) What is the long-run inflation rate implied by the model? Is it economically plausible?

$$\mathbb{E}[\pi_t] = \hat{\pi}_t = \frac{\hat{\beta}_0}{1 - \hat{\beta}_1} = 3.3\%$$

- Long-run mean ≈ 3.3%: close to the pre-COVID sample average
- Consistent with the historical US inflation regime

Exercise 2 | US Inflation

e) Compute one-step-ahead forecasts for 2020–2023 using the model estimated in c). Also compute a naive forecast (the sample mean over 1961–2019). Which performs better?

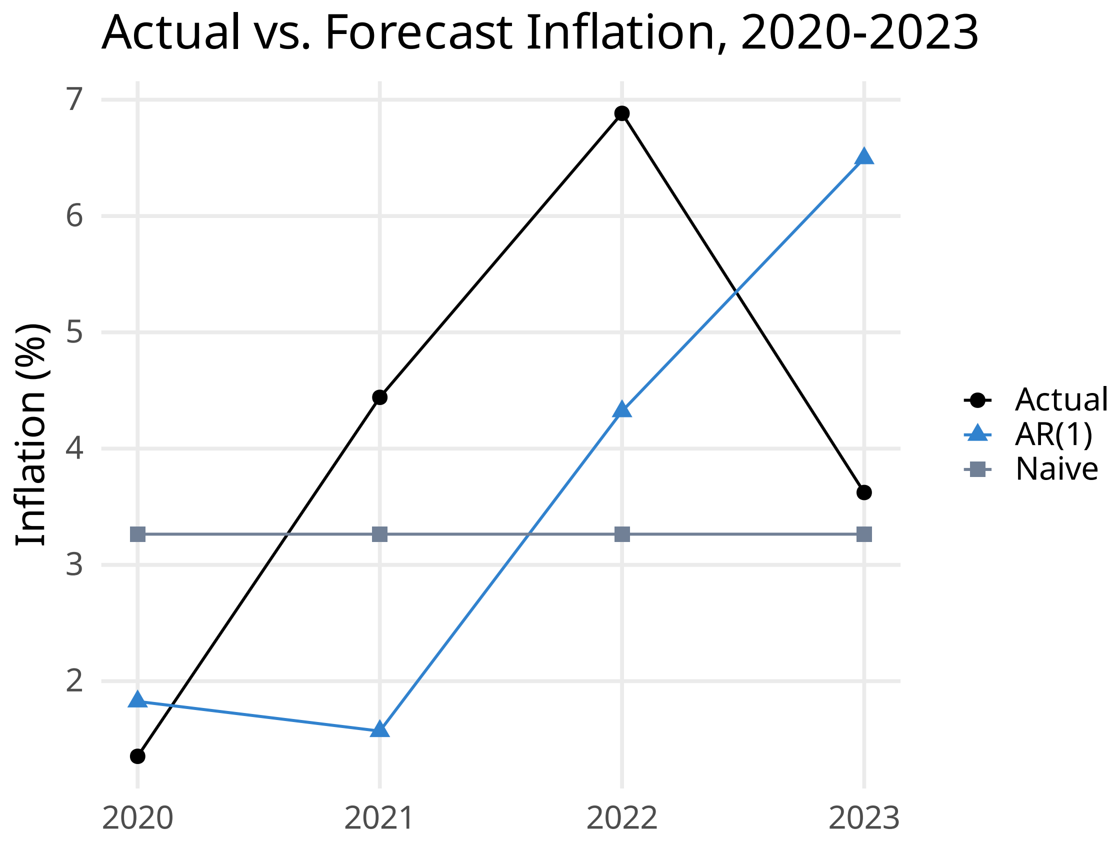
```
b0_pi      <- coef(ar1_pi)["(Intercept)"]
b1_pi      <- coef(ar1_pi)["pi_l1"]
naive_mean <- mean((us_pi_l |> filter(year <= 2019))$pi)
```

Exercise 2 | US Inflation

```r
(pi_forecast <- us_pi_l |>
  filter(year >= 2020, year <= 2023) |>
  mutate(
    `AR(1)` = b0_pi + b1_pi * pi_l1,
    Naive = naive_mean) |>
  select(year, Actual = pi, `AR(1)`, Naive))

pi_forecast |>
  summarise(
    MSFE_AR1  = mean((Actual - `AR(1)`)^2),
    MSFE_Naive = mean((Actual - Naive)^2)
  )
```

Exercise 2 | US Inflation



- Large errors in 2021–22 (+2.9 and +2.6 pp): the AR(1) was anchored on low pre-COVID inflation and adapted too slowly
- AR(1) MSFE ≈ 5.8 vs. naive MSFE ≈ 4.6: the naive benchmark **outperforms**
- Good in-sample fit does not guarantee good out-of-sample performance

Discussion

1. Which series would you rather forecast — the labor share or inflation — and why?

- The labor share has $\hat{\beta}_1 \approx 0.99$: one-step-ahead forecasts are trivially accurate ($\hat{y}_{t+1} \approx y_t$), but this is just inertia, not genuine predictability
- Inflation has $\hat{\beta}_1 \approx 0.89$: more genuine mean-reversion, but COVID showed the model can break badly under structural shifts
- A near-unit-root series is **easy** to forecast one step ahead but **dangerous** over longer horizons — uncertainty accumulates fast
- Neither is unconditionally better $\longrightarrow$ it depends on the horizon, the purpose, and whether you trust stationarity

Discussion

2. Think about the Portuguese electricity consumption data you will use for the research project. Do you expect it to be stationary?

- Electricity has a strong **weekly cycle** (weekdays vs. weekends) and **seasonal cycle** (winter vs. summer) $\longrightarrow$ the raw series almost certainly looks non-stationary
- But these cycles are **predictable and regular** $\longrightarrow$ different from the drifting mean of the labor share
- A first step before modeling: plot the series and its ACF, and ask whether the patterns are stable over time