Introduction to Time Series Analysis

Lecture Notes - Fall 2025

Kedar Kulkarni

Contents

Preface

These notes are a personal compilation of concepts, proofs, examples, and intuition related to time series analysis, developed from my experience of teaching an *Introduction to Time Series Analysis* course to master's students in the Economics programme at Azim Premji University. They are intended as an introductory resource and a starting point for students to think through time series models and their underlying assumptions.

The primary objective of these notes is to build intuition and to explain, in simple and accessible terms, the basic concepts and foundational ideas required to understand time series analysis. The emphasis is on clarity and interpretation rather than technical sophistication alone. It is assumed that readers have prior exposure to a basic course in statistics and econometrics.

These notes are meant to be supplementary and should not be viewed as a substitute for standard textbooks. Students are strongly encouraged to consult established references to develop a deeper and more rigorous understanding of time series econometrics. In particular, three texts are recommended: Hamilton (2020) for advanced graduate students with strong mathematical training; Enders (2008) for a rigorous and detailed treatment of model specification and inference; and Levendis (2018) for students seeking an intuitive introduction with minimal reliance on advanced quantitative tools.

An accompanying set of supplementary materials providing `Stata` code along with intuitive explanations is available separately but is not included in this document.

I would like to thank Zico Dasgupta and Srinivas Raghavendra, my colleagues at Azim Premji University, for their valuable discussions, motivation, and theoretical support in designing and thinking through the course.

Large Language Models have been used to polish grammar and edit the text; however, none of the content has been generated by AI. All remaining errors are entirely my own.

License

Chapter 1

Introduction

1.1 Cross-Sectional Data

Much of students' prior exposure to econometrics is typically in the context of cross-sectional data. Such data can be thought of as a snapshot or still frame taken at a single point in time. Using cross-sectional data, we can study associations between variables and describe patterns in economic outcomes. However, it is difficult to understand how these relationships evolve over time or to make strong causal claims. While certain identifying assumptions allow causal interpretation, these are often fragile in cross-sectional settings due to issues such as omitted variable bias.

A key assumption underlying much of cross-sectional econometrics is that observations are independent and identically distributed (i.i.d.) [1]. This assumption has two components:

- **Independence:** One observation (for example, an individual's wage) is assumed to be unrelated to another's. This is plausible when individuals are randomly sampled from a large population.
- **Identical distribution:** All observations are assumed to be drawn from the same underlying data-generating process, such as the same wage distribution.

1.2 Towards Time Series

Cross-sectional analysis provides a snapshot of economic relationships; time series analysis, in contrast, is akin to observing a time-lapse video. Here, we follow how a variable evolves over time—day by day, month by month, or year by year. The central questions therefore shift from simply asking *"What is the relationship between X and Y?"* to also asking *"How does Y change over time?"* and *"Can past values help us predict the future?"*

Examples of time series data include:

[1] In practice, the i.i.d. assumption may be violated. For example, individuals from the same geographic region, occupation, or social network may have correlated outcomes. Nonetheless, this assumption often serves as a useful starting point.

- **Finance:** Daily stock prices, exchange rates, or asset returns.
- **Macroeconomics:** Quarterly GDP, monthly inflation, or annual unemployment rates.
- **Climate science:** Daily temperatures, monthly rainfall, or annual CO_2 emissions.
- **Health:** Weekly disease incidence or monthly hospital admissions.
- **Agriculture:** Crop yields recorded annually for the same region.

Unlike cross-sectional data, time series observations are typically **not independent**. The value of a variable today often depends on its own past values. For instance, today's stock price is likely to be close to yesterday's—a phenomenon known as *autocorrelation* or *serial correlation*.

The assumption of identical distribution is also problematic in time series contexts. Economic and natural systems evolve over time and are subject to trends, structural breaks, and seasonal patterns. As a result, observations from different periods may not be drawn from the same distribution. For example, GDP in 2020 is generated by a very different economic environment than GDP in 1980.

Time series analysis therefore requires tools that explicitly account for temporal dependence and for the possibility that the statistical properties of the data change over time.

1.3 Mathematical Notation

This section introduces the basic notational conventions used throughout these notes.

- Random variables are denoted by **capital letters**, such as X, Y, or Z, while their realizations are denoted by **lowercase letters**, such as x, y, or z.
- Time is indexed by t, and the value of a variable X at time t is written as X_t.
- Greek letters (such as α, β, and γ) denote **unknown parameters** to be estimated from data.
- Estimated quantities are denoted using a **hat**. For example, $\hat{\beta}$ represents an estimate of β.

We use the **lag operator**, denoted by L, to simplify expressions involving past values:

- $LX_t = X_{t-1}$,
- $L^2X_t = X_{t-2}$,
- $L^kX_t = X_{t-k}$.

We also define the **difference operator** Δ to represent changes over time:

- $\Delta X_t = X_t - X_{t-1} = (1 - L)X_t$,

- $\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2}$.

For notational convenience, we may write DX_t and D^2X_t to denote first and second differences, respectively. Differencing is a key transformation in time series analysis and is often used to remove trends and achieve stationarity.

1.4 Components of a Time Series

A time series can often be decomposed into distinct components that capture different types of variation. Understanding these components is useful for modeling, interpretation, and forecasting.

1.4.1 Typical Decomposition

A common additive decomposition of a time series X_t is

$$X_t = T_t + S_t + C_t + I_t,$$

where T_t denotes the trend, S_t the seasonal component, C_t the cyclical component, and I_t the irregular or noise component. In practice, cyclical and irregular movements are often difficult to separate, leading to the simpler representation

$$X_t = T_t + S_t + I_t.$$

When the magnitude of fluctuations depends on the level of the series, a multiplicative form is more appropriate:

$$X_t = T_t \cdot S_t \cdot I_t.$$

1.4.2 An Illustrative Example

Consider a synthetic time series generated using an additive structure:

$$X_t = T_t + S_t + I_t,$$

where the trend follows $T_t = 1 + 0.1t$, the seasonal component is given by $S_t = 1.6\sin(\pi t/6)$, and the irregular component follows an AR(1) process,

$$I_t = 0.7I_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1).$$

This construction produces a series with a steadily increasing trend, regular seasonal fluctuations, and persistent random shocks.

1.4.3 Forecasting the Series

Forecasting proceeds by predicting each component separately. The trend is extrapolated using its deterministic form, the seasonal component is computed directly from its periodic structure, and the irregular component is forecast recursively using its autoregressive dynamics. The h-step-ahead forecast of the series is therefore

$$\hat{X}_{t+h} = T_{t+h} + S_{t+h} + \hat{I}_{t+h}.$$

This approach provides a transparent framework for understanding how different features of a time series contribute to its future evolution.

Chapter 2

ARMA Model

2.1 Stationarity

2.1.1 The Intuition

Before we dive into autoregressive processes, it is important to discuss the idea of stationarity. Intuitively, a stationary time series is "well-behaved" in the sense that its statistical properties do not change over time. This makes it easier to model and forecast because patterns observed in the past are assumed to hold in the future.

Imagine trying to predict the temperature in a city:

- If the city's climate has been stable for decades, past data are likely to be informative about the future — this is closer to stationarity.
- If the city is experiencing rapid climate change, the patterns from the past may no longer apply — this is nonstationary.

In essence, for a process to be predictable using standard tools, we require that it does not "drift" unpredictably in its properties over time.

2.1.2 Formal Definition

A stochastic process $\{X_t\}$ is weakly stationary (or covariance stationary) if:

1. The mean is constant over time:

$$\mathbb{E}[X_t] = \mu \quad \text{for all } t$$

2. The variance is constant over time:

$$\text{Var}(X_t) = \sigma^2 \quad \text{for all } t$$

3. The covariance between two values depends only on the time gap (lag) between them, not on their absolute position in time:

$$\text{Cov}(X_t, X_{t-k}) = \gamma_k \quad \text{for all } t, \text{ and fixed } k$$

If any of these conditions fail, the series is nonstationary. Common violations include:

- A trend in the mean (mean is not constant)
- Changing volatility (variance is not constant)
- Changing correlation structure over time

2.1.3 Real-World Examples

- **Stationary:**
  - Daily temperature deviations from the monthly average
  - percentage change in stock prices over short periods
  - differences in interest rates.
- **Nonstationary:**
  - Absolute temperature readings over decades (due to global warming trend)
  - nominal GDP
  - Crop Yields

2.2 Autoregressions and Moving Averages (ARMA)

Suppose we are trying to predict tomorrow's sales. One extreme way to think about this is to assume that today's sales give us no useful information about tomorrow's sales. In other words, sales from day to day are completely unrelated.
Econometrically, this is like saying:

- There is no persistence in the data.
- Shocks to sales are entirely temporary, disappearing the very next day.
- Every day is "fresh" in terms of the sales number, determined only by random influences of that day.

This setup is equivalent to a purely random process (also called a *white noise process*), where:

$$X_t = \varepsilon_t$$

and

$$\varepsilon_t \sim \text{i.i.d. } N(0, \sigma^2)$$

This means:

$$\mathbb{E}[X_t] = 0$$

$$\text{Var}(X_t) = \sigma^2 \quad \text{(constant over time)}$$

$$\text{Cov}(X_t, X_{t-k}) = 0 \quad \text{for all } k \neq 0 \quad \text{(no autocorrelation)}$$

In such a world, past data does not help us forecast future values, because the process is driven entirely by unpredictable, independent shocks.
If we assume that

$$\varepsilon_t \overset{\text{i.i.d.}}{\sim} N(0, 1),$$

then the expected value of daily sales is

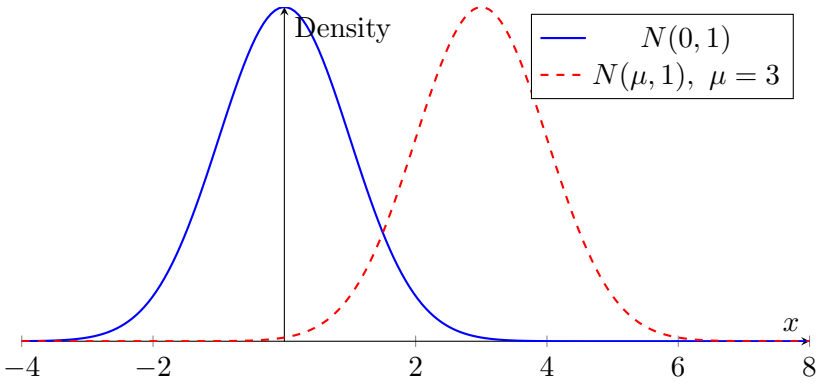
$$\mathbb{E}[X_t] = \mathbb{E}[\varepsilon_t] = 0.$$

This would imply that, on average, sales are zero, which is often unrealistic in practical applications. It is therefore better to write the model as

$$X_t = \mu + \varepsilon_t,$$

where μ is a constant representing the average level of sales and ε_t captures random deviations from this mean.

Illustration: The plot below compares the distribution of X_t in the two cases.



The blue curve shows a zero-mean normal distribution, while the red dashed curve shows the same distribution shifted to the right by $\mu = 3$.

2.2.1 The AR(1) Model

A simple extension of the constant mean model is the autoregressive model of order 1, or AR(1). In this model, the value of the series at time t depends linearly on its value at time $t-1$ and a

random error term.

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{2.1}$$

where:

- ϕ is the autoregressive parameter.
- ε_t is a white noise process, typically assumed to be i.i.d. $N(0, \sigma^2)$.

If we wish to allow the process to have a nonzero mean μ, we can write:

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t. \tag{2.2}$$

Here:

- μ is the constant (long-run mean) towards which the series reverts.
- When $|\phi| < 1$, the process is stationary with mean μ and variance $\sigma^2/(1-\phi^2)$.

Bias of OLS in the AR(1) model

Consider the AR(1) model

$$y_t = \rho\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} (0, \sigma^2), \tag{2.3}$$

and suppose the process is (weakly) stationary with $|\rho| < 1$.
The OLS estimator of ρ from the regression y_t on y_{t-1} (no intercept) is

$$\hat{\rho} \;=\; \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

Substitute the model (2.3) into the numerator:

$$\begin{aligned}\sum_{t=2}^{T} y_t y_{t-1} &= \sum_{t=2}^{T} (\rho y_{t-1} + \varepsilon_t) y_{t-1} \\ &= \rho \sum_{t=2}^{T} y_{t-1}^2 \;+\; \sum_{t=2}^{T} \varepsilon_t y_{t-1}.\end{aligned}$$

Therefore

$$\hat{\rho} = \rho \;+\; \frac{\sum_{t=2}^{T} \varepsilon_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}. \tag{2.4}$$

If the numerator and denominator were independent (or if we could replace the random denominator by its probability limit), then the second term would have expectation zero and $\hat{\rho}$ would be unbiased. However, in finite samples the numerator and denominator are *not* independent: both are functions of the same sequence $\{y_t\}$ (and hence of past shocks). This dependence produces a finite-sample bias.

Interpreting ϕ in the AR(1) model

Consider the AR(1) model above but in the deviation [1] form.

$$y_t - \mu = \varphi\,(y_{t-1} - \mu) + e_t,$$

where:

μ	is the long-run mean (equilibrium level),
φ	is the persistence parameter,
e_t	is a random shock with mean zero.

In simple words, the equation can be read as

$$\text{Deviation at time } t = \underbrace{\varphi \times \text{Deviation at time } t-1}_{\text{persistence from past}} + \underbrace{\text{new shock } e_t}_{\text{new information}}.$$

Interpretation of a one-time shock:

Table 2.1: Decay of a +10 shock in an AR(1) process with $\mu = 100$ and $\varphi = 0.6$

Period	**Deviation from mean**	y_t **value**
t	+10.000	110.000
$t+1$	+6.000	106.000
$t+2$	+3.600	103.600
$t+3$	+2.160	102.160
$t+4$	+1.296	101.296
$t+5$	+0.7776	100.7776
$t+6$	+0.4666	100.4666

Consider,

$$\mu = 100, \quad \varphi = 0.6.$$

Suppose in period t a shock $e_t = +10$ occurs and no further shocks occur afterwards ($e_{t+1} = e_{t+2} = \cdots = 0$).

Step 1: Immediate effect in period t

$$y_t = \mu + \varphi(y_{t-1} - \mu) + e_t.$$

If $y_{t-1} = \mu$, then $y_t = 100 + 0.6 \cdot 0 + 10 = 110$. The deviation from the mean is +10.

Step 2: Effect in period $t+1$ The AR(1) rule says 60% of the deviation from the mean carries over:

$$y_{t+1} - \mu = 0.6 \times (y_t - \mu) + 0 = 0.6 \times 10 = 6.$$

[1] see Appendix for a formal derivation

So the deviation shrinks from +10 to +6.

Step 3: Subsequent periods Each period keeps only 60% of the previous deviation:

$$y_{t+2} - \mu = 0.6 \times 6 = 3.6, \quad y_{t+3} - \mu = 0.6 \times 3.6 = 2.16, \quad \dots$$

For $\varphi = 0.6$, the decay pattern is geometric:

$$10,\ 6,\ 3.6,\ 2.16, \dots$$

meaning the shock's influence gets progressively smaller but never disappears instantly. If $\varphi = 0.6$, then 60% of the previous period's deviation from μ is carried forward, and the remaining 40% decays.

Causality in AR(1) Models:

The AR(1) process is *causal* if

$$|\varphi| < 1.$$

Under this condition, y_t can be written purely in terms of current and past shocks:

$$y_t = \mu + \sum_{j=0}^{\infty} \varphi^j e_{t-j}.$$

This representation shows that:

- y_t depends only on present and past innovations $(e_t, e_{t-1}, \dots)$, not on future shocks.
- The effect of a shock e_{t-j} decays geometrically at rate φ^j.
- The process is stable and does not explode over time.

$$y_t - \mu = \varphi\,(y_{t-1} - \mu) + e_t, \quad e_t \text{ i.i.d. with } \mathbb{E}[e_t] = 0.$$

Conditions on ϕ

$\phi < 1$ - Stationary, Causal: Deviations from the mean decay geometrically over time. The process has a finite long-run variance and is mean-reverting: today's shock has a diminishing influence (φ^h) h periods ahead. Example: $\varphi = 0.6$ means 60% of last period's deviation carries over.

$\phi = 1$ - Unit Root (Random Walk): No decay in deviations — shocks are *permanent*. A one-time shock changes the level of the process forever. Variance grows without bound over time, and there is no mean reversion. Forecasts are best guesses equal to the last observed value.

$\phi > 1$ - Explosive: Deviations grow over time instead of shrinking. Even small shocks eventually cause the process to diverge. The process is unstable; unless e_t exactly cancels past deviations, y_t moves further away from μ each period. Rare in real-world economic data, except during unstable regimes (e.g., hyperinflation).

Impulse Response Function (IRF)

Consider the mean-zero AR(1) process:

$$y_t = \phi y_{t-1} + e_t, \quad \mathbb{E}[e_t] = 0.$$

We can expand the process recursively:

$$\begin{aligned}
y_1 &= \phi y_0 + e_1, \\
y_2 &= \phi y_1 + e_2 = \phi(\phi y_0 + e_1) + e_2 = \phi^2 y_0 + \phi e_1 + e_2, \\
y_3 &= \phi y_2 + e_3 = \phi(\phi^2 y_0 + \phi e_1 + e_2) + e_3 \\
&= \phi^3 y_0 + \phi^2 e_1 + \phi e_2 + e_3, \\
&\vdots
\end{aligned}$$

By induction, for any horizon $t \geq 1$:

$$y_t = \phi^t y_0 + \sum_{i=1}^{t} \phi^{t-i} e_i.$$

From this representation, the impulse response function (IRF) k periods after the shock is $IRF(k) = \phi^k$. As we move farther away from X_0, its influence fades because it is multiplied by ϕ^t. A shock today, for example, like a sudden rise in oil prices affects the economy immediately, but over time, businesses and consumers adjust. The farther back in time the shock happened, the smaller its impact on today's outcome.

Forecasting

Suppose X_t evolves according to the AR(1) process:

$$X_t = 0.6X_{t-1} + e_t,$$

where $e_t \overset{\text{i.i.d.}}{\sim} N(0, 100)$. If we assume $X_0 = 100$ and the first four error terms are $20, -30, 10, 15$. Then the next four values of X_t are:

$$X_1 = 0.6X_0 + e_1 = 0.6(100) + 20 = 80,$$
$$X_2 = 0.6X_1 + e_2 = 0.6(80) - 30 = 18,$$
$$X_3 = 0.6X_2 + e_3 = 0.6(18) + 10 = 20.8,$$
$$X_4 = 0.6X_3 + e_4 = 0.6(20.8) + 15 = 27.48.$$

One-step-ahead forecast: We can compute the expected value of X_5 given past data:

$$\begin{aligned} E(X_5 \mid X_4, X_3, \dots) &= E(0.6X_4 + e_5 \mid X_4, X_3, \dots) \\ &= 0.6(27.48) + E(e_5 \mid X_4, X_3, \dots) \\ &= 16.488 + 0 = 16.488. \end{aligned}$$

Two-step-ahead forecast: Similarly, for X_6:

$$\begin{aligned} E(X_6 \mid X_4, X_3, \dots) &= E(0.6X_5 + e_6) \\ &= 0.6\,E(X_5) + E(e_6) \\ &= 0.6(16.488) + 0 = 9.8928. \end{aligned}$$

General formula

In general, for an AR(1) process:

$$E(X_{t+a} \mid X_t, X_{t-1}, \dots) = \phi^a X_t.$$

With $|\phi| < 1$, the multi-step-ahead forecasts decay geometrically toward the unconditional mean (which is 0 in this example).

Examples

Calculate by hand the IRFs out to five periods for the following AR models: (a) $X_t = 0.5X_{t-1} + e_t$ and (b) $X_t = 0.10 + 0.5X_{t-1} + e_t$. Given the IRFs you calculated, do these all seem stationary? Why or why not?

Solution: (a)

$$X_t = 0.5X_{t-1} + e_t$$

For an AR(1) model, the IRF at horizon t is given by ϕ^t, where $\phi = 0.5$.

$$t = 0 : 1$$
$$t = 1 : 0.5$$
$$t = 2 : 0.25$$
$$t = 3 : 0.125$$
$$t = 4 : 0.0625$$
$$t = 5 : 0.03125$$

(b)

$$X_t = 0.10 + 0.5X_{t-1} + e_t$$

The AR coefficient is the same ($\phi = 0.5$), so the IRFs are identical to part (a). The constant term affects only the unconditional mean, not the propagation of shocks.

$$t = 0 : 1$$
$$t = 1 : 0.5$$
$$t = 2 : 0.25$$
$$t = 3 : 0.125$$
$$t = 4 : 0.0625$$
$$t = 5 : 0.03125$$

Stationarity. Both processes are stationary since $|\phi| = 0.5 < 1$. This is reflected in the IRFs, which decay geometrically to zero. The constant in (b) only shifts the unconditional mean and does not affect stationarity.

2.2.2 The AR(p) Model

An autoregressive process of order p is given by:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + e_t,$$

where $e_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$

Interpretation of AR(p) model

- **Dynamic dependence:** Each observation X_t depends linearly on the previous p lags. The coefficients $\phi_1, \ldots, \phi_p$ measure how much influence past values have on the present.
- **Role of coefficients:**
  - ϕ_1 captures short-run persistence (dependence on X_{t-1}).
  - Higher-order terms $\phi_2, \phi_3, \ldots, \phi_p$ capture medium-run or delayed effects.

  - The combined effect is richer dynamics than AR(1) — shocks may decay slowly, oscillate, or change sign depending on ϕ_i.

- **Impulse Response:** A one-unit shock e_t affects future values X_{t+h} through the sequence ψ_h. Unlike AR(1), where the decay is monotone (ϕ^h), in AR(p) the decay may be oscillatory if some roots of the characteristic polynomial are complex.

- **Forecasting:** Forecasts use a recursive substitution:

$$\hat{X}_{t+a|t} = \phi_1 \hat{X}_{t+a-1|t} + \cdots + \phi_p \hat{X}_{t+a-p|t}.$$

  The forecast path depends on the structure of $\phi_1, \ldots, \phi_p$. If coefficients imply high persistence, forecasts will stay closer to the current value. If coefficients imply oscillatory dynamics, forecasts will show cycles.

Examples:

Consider the model

$$X_t = 0.75X_{t-1} + 0.50X_{t-2} + 0.10X_{t-3} + e_t.$$

Now, forecast out from period four through period ten, given the initial values

$$X_1 = 5, \quad X_2 = -10, \quad X_3 = 15.$$

Graph these first ten observations on X_t. Does X_t appear to be mean-stationary?

Period	X_t	Type
X_1	5.00	Observed
X_2	−10.00	Observed
X_3	15.00	Observed
X_4	6.75	Forecast
X_5	11.5625	Forecast
X_6	13.5469	Forecast
X_7	16.6165	Forecast
X_8	20.3922	Forecast
X_9	24.9571	Forecast
X_{10}	30.5756	Forecast

Table 2.2: Observed values and forecasts of X_t for the AR(3) model

The sequence of forecasts grows rather than reverting to a fixed mean. Hence, the process is not mean-stationary.

2.2.3 The MA(1) Model

Model Structure

An MA(1) process is defined as

$$y_t = e_t + \theta e_{t-1}, \qquad e_t \sim \text{WN}(0, \sigma^2).$$

Here y_t is the observed series, e_t is white noise, and θ is the MA(1) coefficient.

Interpretation of θ

- θ measures how much of the previous period's shock e_{t-1} carries over into the current value y_t.
- If $\theta > 0$, positive shocks tend to increase the next observation as well, creating short-run momentum.
- If $\theta < 0$, positive shocks reduce the next observation, creating short-run mean reversion.
- The effect lasts only one period: beyond lag 1, past shocks have no impact.

Impulse Response Function (IRF)

The IRF traces the effect of a one-unit shock in e_t on future values of y_t:

$$IRF(0) = 1, \quad IRF(1) = \theta, \quad IRF(h) = 0 \quad \text{for } h \geq 2.$$

- At time t, the shock increases y_t by 1 unit.
- At time $t+1$, the shock continues to influence y_{t+1} by θ units.
- From $t+2$ onwards, the effect vanishes.

Thus, an MA(1) captures very short-run persistence in shocks: they matter contemporaneously and for one additional period, but no further.

MA(1) Model: Residuals (Innovations)

Consider the MA(1) model without constant:

$$y_t = e_t + \theta e_{t-1}, \quad e_t \sim \text{i.i.d. } (0, \sigma^2).$$

Given a consistent estimate $\hat{\theta}$, the residuals (or innovations) are obtained recursively as

$$\hat{e}_t = y_t - \hat{\theta}\, \hat{e}_{t-1}, \quad \text{with } \hat{e}_0 = 0.$$

Application to Growth Rate Data

We estimate the MA(1) coefficient as

$$\hat{\theta} \approx 0.9087.$$

The corresponding residuals (innovations) are shown below:

t	y_t (Growth Rate)	$\hat{e}_t$ (Innovation)
1	3.6551	3.6551
2	2.8890	-0.4325
3	5.8216	6.2146
4	7.1883	1.5408
5	-2.6711	-4.0713
6	-0.0554	3.6444
7	7.5348	4.2230
8	3.3318	-0.5058
9	6.3347	6.7944
10	5.0286	-1.1457

Table 2.3: Estimated innovations $\hat{e}_t$ for the MA(1) model.

Interpretation

- $\hat{e}_t$ represents the "new information" in period t not explained by lagged shocks.
- Large absolute values (e.g., $t = 3$, $t = 9$) indicate strong surprises relative to the MA(1) prediction.
- If the MA(1) is correctly specified, the sequence $\{\hat{e}_t\}$ should behave like white noise: no serial correlation and roughly constant variance.

Example: MA(1) Interpretation, IRF, and Forecasts

Question

Consider the MA(1) process

$$y_t \;=\; e_t \;+\; 0.60\, e_{t-1}, \qquad e_t \sim \mathrm{WN}(0, \sigma^2),\ \sigma = 2.0.$$

At the end of the sample (time T), the estimated residual (innovation) is $e_T = -1.5$.

1. Interpret the coefficient 0.60.
2. Derive the impulse response function (IRF).
3. Produce the forecasts $\hat{y}_{T+1|T}$, $\hat{y}_{T+2|T}$, and $\hat{y}_{T+3|T}$.

Solution

Interpretation. The coefficient 0.60 measures how a shock from one period ago influences the current value. A one-unit positive shock in period $t-1$ raises y_t by 0.60 units (in addition to the contemporaneous innovation e_t). Since $0.60 > 0$, shocks exhibit short-run *persistence* (momentum) that lasts exactly one period.

Impulse Response Function (IRF). By convention the IRF evaluates the effect of a one-unit shock at time t:

$$IRF(0) = 1, \qquad IRF(1) = 0.60, \qquad IRF(h) = 0 \text{ for } h \geq 2.$$

Thus, the shock has a unit impact contemporaneously, a spillover of 0.60 next period, and then vanishes.

Three-step forecasts. For an MA(1),

$$\hat{y}_{T+1|T} = 0.60\, e_T, \qquad \hat{y}_{T+h|T} = 0 \text{ for } h \geq 2.$$

Given $e_T = -1.5$:

$$\hat{y}_{T+1|T} = 0.60 \times (-1.5) = -0.90, \qquad \hat{y}_{T+2|T} = 0, \qquad \hat{y}_{T+3|T} = 0.$$

2.2.4 Moving Average Model: MA(p)

Model Structure

An MA(p) process is defined as

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_p e_{t-p}, \qquad e_t \sim \text{WN}(0, \sigma^2).$$

Here y_t is the observed series, e_t is white noise, and $\theta_1, \ldots, \theta_p$ are the moving average coefficients.

Interpretation of Coefficients

- Each coefficient θ_j measures the extent to which a shock from j periods ago still affects the current value y_t.
- For example, θ_1 captures spillover from the immediately previous shock, θ_2 from two periods ago, and so on.
- Positive values of θ_j imply short-run persistence (shocks carry over positively), while negative values imply short-run correction or mean reversion.
- The memory of the process lasts exactly p periods: shocks have no impact beyond lag p.

Impulse Response Function (IRF)

The IRF traces the effect of a one-unit shock at time t on future values:

$$IRF(0) = 1, \quad IRF(1) = \theta_1, \quad IRF(2) = \theta_2, \quad \cdots, \quad IRF(p) = \theta_p, \quad IRF(h) = 0 \text{ for } h > p.$$

- At the time of the shock ($h = 0$), the effect is one-for-one.
- In subsequent periods, the shock contributes according to the estimated θ_j values.
- After p periods, the effect disappears entirely.

2.2.5 AR vs. MA Processes: Illustrative Examples

Case	AR?	MA?	Reasoning
(a) Shock to GDP due to Global Financial Crisis	Yes	No	Large macro shocks persist over time as output takes many periods to adjust. Past GDP affects current GDP ⇒ autoregressive structure.
(b) Shock to cafeteria sales during a university fest	No	Yes	The shock is temporary and may spill over for a few days (residual impact). Once the event ends, no persistence remains ⇒ moving average structure.
(c) Shock to a financial stock because of a CEO controversy	Yes	Yes	Initial shock creates short-term "bounce" (MA component). If controversy changes fundamentals, persistence in returns arises ⇒ AR component as well.
(d) Weather shock affecting agriculture (e.g., unexpected rainfall)	No	Yes	Rainfall shock impacts output for one or two periods but quickly dies out. Short-memory dynamics ⇒ MA type.
(e) GDP growth series measured with statistical error	No	Yes	Underlying series is white noise, but measurement error contaminates adjacent observations. This induces an MA(1)-like structure.

2.2.6 ARMA

In many economic and financial time series, current outcomes are influenced both by their own past realizations and by past shocks. For instance, consider quarterly GDP growth rates in India.

On the one hand, growth often displays persistence: if growth was high in the last quarter, it is more likely to be high in the current quarter as well. This reflects an **autoregressive (AR)** component, where present values depend on past values.

On the other hand, unexpected shocks—such as a sudden rise in global oil prices, a monsoon shortfall, or a policy announcement by the Reserve Bank of India—can affect growth not only immediately but also in subsequent quarters as their effects work through supply chains, investment decisions, and consumer confidence. This reflects a **moving average (MA)** component, where present values depend on past shocks.

The ARMA model combines these two features, making it especially suitable for modeling and forecasting economic variables such as GDP growth, inflation, or exchange rates in the Indian context.

The ARMA(p,q) Model

An autoregressive moving average (ARMA) process of order (p, q) combines an autoregressive (AR) part of order p and a moving average (MA) part of order q. Formally, the process $\{y_t\}$ is written as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where

- $\{\varepsilon_t\}$ is a white-noise sequence with mean zero and variance σ^2,
- $\phi_1, \ldots, \phi_p$ are the autoregressive parameters,
- $\theta_1, \ldots, \theta_q$ are the moving average parameters.

An Intuitive ARMA(1,1) Example

Consider the ARMA(1,1):

$$y_t = \phi\, y_{t-1} + \varepsilon_t + \theta\, \varepsilon_{t-1}, \qquad \varepsilon_t \sim \mathrm{WN}(0, \sigma^2).$$

Intuitively, think of y_t as quarterly Indian GDP *growth*. Two forces drive it: (i) persistence from the last quarter ($\phi\, y_{t-1}$): strong demand or investment momentum tends to carry forward; (ii) shock propagation from last quarter ($\theta\, \varepsilon_{t-1}$): an unexpected shock (oil price jump, monsoon surprise, RBI policy shift) does not vanish instantly; it echoes for a while.

2.3 Model Selection

2.3.1 Autocovariance and Autocorrelation Functions

In time series analysis, two related concepts are used to study dependence across time: the autocovariance function (ACVF) and the autocorrelation function (ACF).

Autocovariance Function (ACVF)

For a weakly stationary process $\{X_t\}$ with mean μ, the k-th order autocovariance is defined as:

$$\gamma_k = \mathrm{Cov}(X_t, X_{t-k}) = E\big[(X_t - \mu)(X_{t-k} - \mu)\big].$$

- $\gamma_0 = \mathrm{Var}(X_t)$ is the variance.
- γ_k captures the raw linear dependence between X_t and X_{t-k}.
- By symmetry, $\gamma_{-k} = \gamma_k$.

Autocorrelation Function (ACF)

The autocorrelation at lag k is the normalized autocovariance:

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

- $\rho_0 = 1$ by definition.
- ρ_k is dimensionless and always lies between -1 and 1.
- The ACF is usually plotted to assess time dependence in data.

Relationship

- γ_k is the raw measure of dependence (in variance units).
- ρ_k is the standardized version, making it easier to compare across lags.
- Many theoretical derivations (e.g., Yule–Walker equations) are first expressed in terms of γ_k, and then converted to ρ_k.

2.3.2 Autocorrelation Functions: ACF and PACF

Autocorrelation Function (ACF). The autocorrelation function at lag h is defined as

$$\rho(h) = \frac{\mathrm{Cov}(y_t,\, y_{t-h})}{\mathrm{Var}(y_t)}.$$

It measures the linear dependence between y_t and its past values y_{t-h}.

Partial Autocorrelation Function (PACF). The partial autocorrelation at lag h, denoted $\phi(h)$, is the correlation between y_t and y_{t-h} *after controlling for* the intermediate lags $y_{t-1}, \ldots, y_{t-h+1}$. Formally, it is obtained from the regression

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_h y_{t-h} + u_t,$$

and $\phi(h)$ is the coefficient α_h.

Intuition.

- The ACF tells us how persistent the process is across different lags.
- The PACF isolates the "direct" contribution of lag h, filtering out the influence of intermediate lags.

2.3.3 Theoretical vs Empirical ACF/PACF

Theoretical ACF/PACF. For any ARMA(p, q) model, the population autocorrelation and partial autocorrelation functions have characteristic patterns:

- AR(p): PACF cuts off after lag p, ACF decays gradually.
- MA(q): ACF cuts off after lag q, PACF decays gradually.
- ARMA(p, q): both ACF and PACF decay gradually, with no finite cut-off.

Empirical ACF/PACF. In practice, we do not observe the true correlations but estimate them from finite samples:

$$\hat{\rho}(h) = \frac{\sum_{t=h+1}^{T}(y_t - \bar{y})(y_{t-h} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}.$$

- The sample ACF $\hat{\rho}(h)$ is noisy, especially at large lags.
- Confidence intervals (often $\pm 2/\sqrt{T}$) are used to judge statistical significance.

2.3.4 Model Selection via ACF/PACF

Step 1: Initial Identification. Compare the empirical ACF and PACF with the theoretical patterns to guess a plausible ARMA(p, q) structure.

- If PACF shows a sharp cut-off at lag p, suspect an AR(p).
- If ACF shows a sharp cut-off at lag q, suspect an MA(q).
- If both decay, an ARMA model is likely.

Step 2: Verification. Once a tentative order is chosen, estimate the model and check:

- Residuals: they should be approximately white noise (no remaining autocorrelation).
- Information criteria (AIC, BIC): to compare across models with different p, q.

2.3.5 AR(1): theoretical ACF, derivation and intuition

Model and stationarity

Consider the autoregressive AR(1) process with a constant:

$$y_t \;=\; a + \phi\, y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2).$$

If $|\phi| < 1$ the process is (weakly) stationary. Denote the unconditional mean by

$$\mu \;=\; \mathbb{E}[y_t] \;=\; \frac{a}{1-\phi}.$$

Center the process by defining $x_t \equiv y_t - \mu$. Then x_t satisfies

$$x_t \;=\; \phi\, x_{t-1} + \varepsilon_t,$$

with $\mathbb{E}[x_t] = 0$.

Autocovariance recursion

Define the autocovariances $\gamma_k = \mathrm{Cov}(x_t, x_{t-k})$ (so $\gamma_0 = \mathrm{Var}(x_t)$). Multiply the centered AR(1) equation by x_{t-k} and take expectations:

$$\mathbb{E}[x_t x_{t-k}] = \phi\, \mathbb{E}[x_{t-1} x_{t-k}] + \mathbb{E}[\varepsilon_t x_{t-k}].$$

For $k \geq 1$ the innovation ε_t is uncorrelated with x_{t-k}, so the last term is zero. Hence the recursion

$$\gamma_k \;=\; \phi\, \gamma_{k-1}, \qquad k \geq 1.$$

Iterating this recursion gives

$$\gamma_k \;=\; \phi^k \gamma_0, \qquad k = 0, 1, 2, \ldots$$

Autocorrelation function (ACF)

The autocorrelation at lag k is $\rho_k = \gamma_k / \gamma_0$. Using the result above,

$$\boxed{\rho_k \;=\; \phi^k} \quad (k = 0, 1, 2, \ldots).$$

Thus the theoretical ACF decays geometrically at rate $|\phi|$. In particular:

- $\rho_0 = 1$ by definition,
- $\rho_1 = \phi$ (so the first autocorrelation equals the AR(1) coefficient),
- ρ_k has the same sign as ϕ for all k, and $|\rho_k| \downarrow 0$ as $k \to \infty$ when $|\phi| < 1$.

Intuition

Each period the effect of a past shock is multiplied by ϕ. A shock ε_t affects x_t immediately (weight 1), then contributes ϕ to x_{t+1}, ϕ^2 to x_{t+2}, and so on. Therefore correlations between x_t and x_{t-k} shrink like powers of ϕ. If $|\phi|$ is close to 1, shocks persist (slow decay); if $|\phi|$ is small, dependence dies quickly.

Economic example: quarterly Indian GDP growth

Let y_t be quarterly GDP *growth* (in percent). An AR(1) reading

$$y_t = a + \phi\, y_{t-1} + \varepsilon_t$$

means current growth partly reflects last quarter's growth (momentum) plus a new shock (monsoon surprise, oil-price shock, policy change).

- If empirical estimation gives $\hat{\phi} = 0.4$, then the theoretical ACF implies

  $$\rho_1 \approx 0.4, \quad \rho_4 \approx 0.4^4 = 0.0256,$$

  i.e. correlation at one quarter is moderate but by one year (4 quarters) the serial dependence is negligible.

- If instead $\hat{\phi} = 0.8$, then

  $$\rho_1 \approx 0.8, \quad \rho_4 \approx 0.8^4 = 0.4096,$$

  indicating strong persistence: a high-growth quarter makes high growth more likely even a year later.

Thus, interpreting ϕ in Indian growth data tells us whether shocks and momentum are short-lived (small $|\phi|$) or persistent (large $|\phi|$).

Practical note on estimation

For an AR(1) the Yule–Walker identity gives $\phi = \rho_1$, so a simple moment-based estimator is the sample ACF at lag 1. In finite samples the sample ACF is noisy; more efficient estimates come from OLS or maximum likelihood (which coincide under Gaussianity). Always check the sample ACF plot: for a well-fitting AR(1) you should see a geometric-looking decay in the empirical ACF and a sharp drop in the PACF after lag 1.

2.3.6 Model Selection via AIC/BIC

When estimating time series or regression models, we often face the problem of choosing the model order (e.g., number of lags in AR, ARMA models, etc.). Two commonly used information criteria are:

$$\text{AIC (Akaike Information Criterion)} \quad \text{and} \quad \text{BIC (Bayesian Information Criterion)}.$$

AIC and BIC Criteria

For a model with log-likelihood value ℓ, number of estimated parameters k, and sample size n:

$$\text{AIC} = -2\ell + 2k,$$

$$\text{BIC} = -2\ell + k\ln(n).$$

- Both criteria trade off model fit (measured by the log-likelihood) against model complexity (measured by number of parameters).
- Lower AIC or BIC values indicate a preferred model.
- **Penalty difference:**
  - AIC penalizes complexity with $2k$.
  - BIC penalizes more heavily with $k\ln(n)$ (since $\ln(n) > 2$ for $n > 7$).

Practical Use

- AIC is more focused on predictive accuracy (often chooses larger models).
- BIC is more conservative, favoring simpler, more parsimonious models.
- In large samples, BIC tends to select the true model (if it is among the candidates).
- In practice, researchers often report both and check robustness.

Example (AR order selection)

Suppose we estimate AR(1), AR(2), AR(3) models on a dataset. We compute log-likelihoods and plug into AIC/BIC. Then, we choose the AR order with the lowest AIC/BIC.

Conclusion

- Both AIC and BIC balance fit vs. parsimony.
- Always compare across the same dataset and same dependent variable.
- Neither guarantees the "best" model; they are heuristics.

2.4 Appendix

2.4.1 Deviation-from-Mean AR(1) Model

Start with the AR(1) model with a non-zero mean μ:

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t \tag{2.5}$$

Expand the equation:

$$y_t = \phi y_{t-1} + (1 - \phi)\mu + \varepsilon_t \tag{2.6}$$

Now subtract the mean μ from both sides:

$$y_t - \mu = \phi y_{t-1} + (1 - \phi)\mu + \varepsilon_t - \mu \tag{2.7}$$

Group the terms:

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t \tag{2.8}$$

$$= \text{AR(1) in deviations from the mean} \tag{2.9}$$

This shows that the deviation from the mean, $y_t - \mu$, follows a zero-mean AR(1) process. It follows from equation (8) that when $E[y_t] = \mu = 0$, AR(1) model is $y_t = \phi y_{t-1} + \epsilon_t$.

2.4.2 Derivation of the ACF for AR(1)

Consider the zero-mean AR(1) process

$$X_t = \phi X_{t-1} + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t] = 0, \quad \mathbb{E}[\varepsilon_t \varepsilon_s] = \begin{cases} \sigma^2 & t = s, \\ 0 & t \neq s, \end{cases}$$

and assume $|\phi| < 1$ so the process is (weakly) stationary.

Preliminary facts. Two simple facts will be used below:

1. For any random variable Z, $\mathrm{Var}(Z) = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$. If $\mathbb{E}[Z] = 0$ then $\mathrm{Var}(Z) = \mathbb{E}[Z^2]$.
2. Weak (covariance) stationarity implies that the first two moments do not depend on time: $\mathbb{E}[X_t] = \mu$ (constant) and $\mathrm{Var}(X_t) = \gamma_0$ (constant for all t). Thus for a zero-mean stationary process ($\mu = 0$),

$$\mathbb{E}[X_{t-1}^2] = \mathrm{Var}(X_{t-1}) = \mathrm{Var}(X_t) = \gamma_0.$$

These justify replacing $\mathbb{E}[X_{t-1}^2]$ by γ_0 in the computations below.

Lag-1 covariance. By stationarity $\mathrm{Var}(X_t) = \mathrm{Var}(X_{t-1})$ and

$$\mathrm{Cov}(X_t, X_{t-1}) = \mathbb{E}[X_t X_{t-1}] - \mathbb{E}[X_t]\mathbb{E}[X_{t-1}].$$

With zero means this simplifies to $\mathbb{E}[X_t X_{t-1}]$. Substitute the AR(1) equation:

$$\mathrm{Cov}(X_t, X_{t-1}) = \mathbb{E}\big[(\phi X_{t-1} + \varepsilon_t)X_{t-1}\big] = \phi\,\mathbb{E}[X_{t-1}^2] + \mathbb{E}[\varepsilon_t X_{t-1}].$$

Since ε_t is uncorrelated with past X_{t-1}, the last term is zero. Using the preliminary facts,

$$\mathrm{Cov}(X_t, X_{t-1}) = \phi\,\mathbb{E}[X_{t-1}^2] = \phi\,\gamma_0,$$

where $\gamma_0 = \mathrm{Var}(X_t)$.

Autocorrelation at lag 1. By definition,

$$\rho_1 = \frac{\mathrm{Cov}(X_t, X_{t-1})}{\sqrt{\mathrm{Var}(X_t)\,\mathrm{Var}(X_{t-1})}} = \frac{\phi\,\gamma_0}{\gamma_0} = \phi.$$

Thus the lag-1 autocorrelation equals the AR coefficient: $\rho_1 = \phi$.

Higher-order autocovariances (recursion). Multiply the AR(1) equation by X_{t-k} and take expectations for $k \geq 1$:

$$\gamma_k \equiv \mathrm{Cov}(X_t, X_{t-k}) = \mathbb{E}[(\phi X_{t-1} + \varepsilon_t)X_{t-k}] = \phi\,\mathbb{E}[X_{t-1}X_{t-k}] + 0 = \phi\,\gamma_{k-1}.$$

Iterating this recursion yields

$$\gamma_k = \phi^k \gamma_0, \qquad k = 0, 1, 2, \ldots$$

Autocorrelation function. The autocorrelation at lag k is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k.$$

Hence the theoretical ACF of an AR(1) decays geometrically at rate $|\phi|$.

Variance (useful check). Solve for γ_0 by using $\gamma_0 = \phi\gamma_1 + \sigma^2$ (multiply the AR(1) equation by X_t and take expectations):

$$\gamma_0 = \phi(\phi\gamma_0) + \sigma^2 \quad \Longrightarrow \quad \gamma_0(1 - \phi^2) = \sigma^2 \quad \Longrightarrow \quad \gamma_0 = \frac{\sigma^2}{1 - \phi^2},$$

which is finite only when $|\phi| < 1$ (the stationarity condition).

Remark. If you denote the AR(1) coefficient by β instead of ϕ, as done in the Levendix Textbook, replace ϕ with β throughout; then $\rho_k = \beta^k$.

2.4.3 Derivation of the ACF for AR(2)

Consider the zero-mean AR(2) process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t] = 0, \quad \mathbb{E}[\varepsilon_t \varepsilon_s] = \begin{cases} \sigma^2 & t = s, \\ 0 & t \neq s. \end{cases}$$

Assume the process is (weakly) stationary.

Preliminary facts. As in the AR(1) derivation we use:

1. If $\mathbb{E}[Z] = 0$ then $\mathrm{Var}(Z) = \mathbb{E}[Z^2]$.
2. Stationarity implies $\mathbb{E}[X_t] = \mu$ (constant) and $\mathrm{Var}(X_t) = \gamma_0$ (constant for all t), so $\mathbb{E}[X_{t-k}^2] = \gamma_0$ for any fixed k.

Define autocovariances. Let $\gamma_k = \mathrm{Cov}(X_t, X_{t-k})$ for $k \geq 0$. In particular $\gamma_0 = \mathrm{Var}(X_t)$ and $\rho_k = \gamma_k/\gamma_0$ is the autocorrelation at lag k.

Yule–Walker equations. Multiply the AR(2) equation by X_{t-k} and take expectations. For $k \geq 2$ (and using $\mathbb{E}[\varepsilon_t X_{t-k}] = 0$ for $k \geq 1$) we obtain the recursion

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \qquad k \geq 2.$$

For the special cases $k = 1$ and $k = 0$ we get the two Yule–Walker equations

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1, \qquad (k = 1)$$

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2. \qquad (k = 0)$$

Solve for ρ_1 and ρ_2. From the $k = 1$ equation,

$$(1 - \phi_2)\,\gamma_1 = \phi_1\,\gamma_0 \quad \Longrightarrow \quad \gamma_1 = \frac{\phi_1}{1 - \phi_2}\,\gamma_0.$$

Divide by γ_0 to get the first autocorrelation:

$$\boxed{\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1}{1 - \phi_2}}.$$

Next compute γ_2 from the recursion (take $k = 2$):

$$\gamma_2 = \phi_1 \gamma_1 + \phi_2 \gamma_0 = \Big(\frac{\phi_1^2}{1 - \phi_2} + \phi_2\Big)\gamma_0.$$

Hence the second autocorrelation is

$$\boxed{\rho_2 = \frac{\gamma_2}{\gamma_0} = \phi_1\rho_1 + \phi_2 \;=\; \frac{\phi_1^2}{1-\phi_2} + \phi_2}.$$

General recursion for ρ_k. For all $k \geq 2$ divide the covariance recursion by γ_0 to obtain

$$\boxed{\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \qquad \rho_0 = 1,\ \rho_1 = \frac{\phi_1}{1-\phi_2}.}$$

Thus once ρ_0 and ρ_1 are known, all higher-order autocorrelations follow from this linear recurrence.

Variance (useful closure). Using $\gamma_1 = \rho_1\gamma_0$ and $\gamma_2 = \rho_2\gamma_0$ in the $k = 0$ Yule–Walker equation, we can solve for γ_0:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2 = \big(\phi_1\rho_1 + \phi_2\rho_2\big)\gamma_0 + \sigma^2,$$

so

$$\boxed{\gamma_0 \;=\; \frac{\sigma^2}{1 - \phi_1\rho_1 - \phi_2\rho_2}}.$$

(Plugging the expressions for ρ_1, ρ_2 expresses γ_0 purely in terms of ϕ_1, ϕ_2, σ^2.)

Stationarity condition (intuition). As with AR(1), stationarity requires that the effect of shocks die out as $h \to \infty$. Algebraically, this is equivalent to invertibility of the lag polynomial

$$\Phi(L) = 1 - \phi_1 L - \phi_2 L^2,$$

which holds if and only if the roots of $\Phi(z) = 0$ lie *outside* the unit circle (i.e. $|z_i| > 1$ for each root).

Interpretation.

- The formula $\rho_1 = \phi_1/(1-\phi_2)$ shows that for AR(2) the lag-1 correlation is not simply ϕ_1 (unlike AR(1)) — it reflects both direct persistence ϕ_1 and indirect two-step feedback via ϕ_2.
- The recursion $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}$ gives a clear computational recipe: compute ρ_1, ρ_2 as above, then iterate to get $\rho_3, \rho_4, \ldots$.
- Depending on the characteristic roots (real distinct, repeated, or complex conjugate), the ACF will show monotone decay, power-law times exponential (repeated root), or damped oscillations.

2.4.4 Example: AR(1) — $X_t = 0.5X_{t-1} + \varepsilon_t$

For an AR(1) with coefficient ϕ the autocorrelation function is

$$\rho_k = \phi^k, \qquad k = 0, 1, 2, \ldots$$

Hence with $\phi = 0.5$,

$$\boxed{\rho_k = 0.5^k}.$$

Numerical values:

$$\begin{aligned}
\rho_1 &= 0.5, \\
\rho_2 &= 0.25, \\
\rho_3 &= 0.125, \\
\rho_4 &= 0.0625, \\
\rho_5 &= 0.03125, \\
\rho_6 &= 0.015625, \text{ etc.}
\end{aligned}$$

Example: AR(2) — $X_t = 0.5X_{t-1} - 0.20X_{t-2} + \varepsilon_t$

The Yule–Walker recursion for autocorrelations is

$$\rho_0 = 1, \qquad \rho_1 = \phi_1 + \phi_2\rho_{-1} \text{ (use below)}, \qquad \rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} \quad (k \geq 2),$$

which for stationary processes reduces to the simple recursion

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, \qquad \rho_0 = 1,$$

with ρ_1 determined by the Yule–Walker pair of equations

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 \quad \Longrightarrow \quad (1 - \phi_2)\rho_1 = \phi_1, \\
\rho_2 &= \phi_1\rho_1 + \phi_2.
\end{aligned}$$

For $\phi_1 = 0.5,\ \phi_2 = -0.2$ we get

$$\rho_1 = \frac{\phi_1}{1 - \phi_2} = \frac{0.5}{1 - (-0.2)} = \frac{0.5}{1.2} = \frac{5}{12} \approx 0.4166667,$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 = 0.5 \cdot \frac{5}{12} - 0.2 = \frac{1}{120} \approx 0.0083333.$$

Thereafter use the recursion $\rho_k = 0.5\rho_{k-1} - 0.20\rho_{k-2}$. Numerically:

$$\begin{aligned}
\rho_1 &\approx 0.4166667,\\
\rho_2 &\approx 0.0083333,\\
\rho_3 &= 0.5 \cdot \rho_2 - 0.2 \cdot \rho_1 \approx -0.0791667,\\
\rho_4 &= 0.5 \cdot \rho_3 - 0.2 \cdot \rho_2 \approx -0.0412500,\\
\rho_5 &\approx -0.0047917,\\
\rho_6 &\approx 0.0058542,\\
\rho_7 &\approx 0.0038854,\\
\rho_8 &\approx 0.0007719,\\
\rho_9 &\approx -0.0003911,\\
\rho_{10} &\approx -0.0003499.
\end{aligned}$$

Interpretation.

- The AR(1) with $\phi = 0.5$ produces smooth geometric decay: correlations fall by a factor 0.5 each lag.
- The AR(2) displays small ρ_2 and then alternating small positive/negative values — a quickly damped, slightly oscillatory ACF. The initial $\rho_1 \approx 0.417$ shows modest persistence at lag 1, but beyond lag 2 the serial correlation is essentially negligible in magnitude (decaying toward zero quickly).

2.4.5 Derivation of the ACF for an ARMA(1,1)

Consider the (mean-zero) ARMA(1,1)

$$Y_t = \phi Y_{t-1} + \theta e_{t-1} + e_t, \qquad e_t \overset{iid}{\sim} (0, \sigma^2),$$

which is obtained from X_t by removing the mean. Let $\gamma(h) = \mathrm{Cov}(Y_t, Y_{t-h})$ and $\rho(h) = \gamma(h)/\gamma(0)$.

Basic Yule–Walker relations Multiply the defining equation by Y_{t-h} and take expectations. Using independence of e_t from the past one obtains

$$\gamma(h) = \phi\gamma(h-1) + \theta\,\mathbb{E}[e_{t-1}Y_{t-h}] \;+\; \mathbb{E}[e_tY_{t-h}].$$

Since $\mathbb{E}[e_tY_{t-h}] = 0$ for $h \geq 1$, and $\mathbb{E}[e_{t-1}Y_{t-h}] = 0$ for $h \geq 2$, we get the two useful relations

$$\boxed{\gamma(h) = \phi\,\gamma(h-1) \quad (h \geq 2)}$$

and for $h = 1$

$$\boxed{\gamma(1) = \phi\gamma(0) + \theta\sigma^2}$$

(indeed $\mathbb{E}[e_{t-1}Y_{t-1}] = \mathbb{E}[e_{t-1}^2] = \sigma^2$). For $h = 0$ one finds

$$\gamma(0) = \phi\gamma(1) + \big(1 + \phi\theta + \theta^2\big)\sigma^2,$$

which upon substituting $\gamma(1) = \phi\gamma(0) + \theta\sigma^2$ yields the closed form

$$\boxed{\gamma(0) = \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2}\,\sigma^2\,.}$$

Autocovariance at lag 1 Using $\gamma(1) = \phi\gamma(0) + \theta\sigma^2$ and the expression for $\gamma(0)$,

$$\gamma(1) = \phi\gamma(0) + \theta\sigma^2 = \gamma(0)\left(\phi + \frac{\theta\sigma^2}{\gamma(0)}\right).$$

Since $\gamma(0) = \sigma^2\frac{1+2\phi\theta+\theta^2}{1-\phi^2}$ we have

$$\frac{\theta\sigma^2}{\gamma(0)} = \theta \cdot \frac{1 - \phi^2}{1 + 2\phi\theta + \theta^2}.$$

Hence the lag-1 autocorrelation simplifies to the compact rational form

$$\boxed{\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \phi + \theta\frac{1 - \phi^2}{1 + 2\phi\theta + \theta^2} = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}\,.}$$

(You may check algebraically that the second and third expressions are equal.)

ACF for $h \geq 2$ From the recursion $\gamma(h) = \phi\gamma(h-1)$ for $h \geq 2$ we get, after dividing by $\gamma(0)$,

$$\boxed{\rho(h) = \phi\,\rho(h-1) \quad (h \geq 2)\,.}$$

Iterating this recursion gives the closed-form ACF for all $h \geq 2$:

$$\boxed{\rho(h) = \phi^{\,h-1}\rho(1) = \phi^{\,h-1}\,\frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2} \quad (h \geq 2)\,.}$$

Summary

- The variance $\gamma(0)$ (and hence the ACF) is finite only when $|\phi| < 1$ (stationarity).
- The ACF depends on ϕ and θ but not on σ^2 (since σ^2 cancels in $\rho(h)$).
- For the special case $\theta = 0$ the ARMA(1,1) reduces to AR(1) and $\rho(h) = \phi^h$.
- The MA(1) part affects only $\rho(0)$ and $\rho(1)$; for $h \geq 2$ the decay is purely geometric with rate ϕ.

Chapter 3

Stationarity and Invertibility

3.1 Introduction

A stochastic process $\{y_t\}$ is said to be **stationary** if its statistical properties do not change over time. Stationarity is crucial because many statistical methods and models (like ARMA) assume it. We typically work with the weaker notion of stationarity, called *covariance stationarity*.

3.1.1 Types of Stationarity

a) Mean Stationarity

A process $\{y_t\}$ is mean stationary if

$$\mathbb{E}[y_t] = \mu \quad \forall t.$$

This means the average level of the series is constant over time.

Example: Daily stock returns (mean close to zero).

Counterexample: GDP, which grows over time and thus has a time-varying mean.

b) Variance Stationarity

A process $\{y_t\}$ has variance stationarity if

$$\text{Var}(y_t) = \sigma^2 \quad \forall t.$$

This implies the volatility of the series does not change over time.

Example: White noise series $u_t \sim \text{iid}(0, \sigma^2)$.

Counterexample: A random walk, where $\text{Var}(y_t)$ increases with t.

c) Covariance Stationarity

A process $\{y_t\}$ is covariance stationary if:

$$\text{Cov}(y_t, y_{t+h}) = \gamma(h) \quad \text{depends only on lag } h, \text{ not on } t.$$

This means correlations between two points in time depend only on their distance apart, not on when they are observed.

Examples:

- If $h = 4$, then $\text{Cov}(y_1, y_5) = \text{Cov}(y_5, y_9) = \text{Cov}(y_{10}, y_{14})$.
- Similarly, $\text{Cov}(y_2, y_3) = \text{Cov}(y_7, y_8)$, because both involve lag $h = 1$.

Example Process: AR(1) $y_t = \phi y_{t-1} + u_t$, with $|\phi| < 1$. *Counterexample:* A trending series, where $\text{Cov}(y_1, y_5) \neq \text{Cov}(y_5, y_9)$.

Real-world Examples

- **Stationary:** Deviations of temperature from the seasonal average; daily stock returns; white noise.
- **Non-stationary:** GDP series (trend), stock prices (unit root), average temperature with climate change (trend).

3.1.2 Why Do We Care About Stationarity?

Stationarity is a fundamental concept in time series analysis. Many statistical methods assume that the underlying data-generating process is stationary. Here are the main reasons why stationarity matters:

1. **Well-understood properties:** Stationary processes are better understood than non-stationary ones. We know how to estimate them efficiently, and the sampling distributions of estimators are well-behaved.
2. **Valid inference:** The test statistics for non-stationary processes often do not follow the usual distributions (e.g., t- or F-distributions). This invalidates standard hypothesis testing. By knowing how a process is non-stationary, we can apply the necessary corrections (e.g., differencing, detrending).
3. **Avoiding spurious regression:** If we regress two unrelated integrated (non-stationary) processes on each other, we may obtain apparently strong relationships, even when no true relationship exists. This problem, known as *spurious regression*, arises because both series share a common time trend.

4. **Illustration of spurious regression:** Suppose X_t and Y_t are both trending upward due to time, but are otherwise unrelated. A regression of Y_t on X_t is likely to produce high R^2 and significant t-statistics, falsely suggesting a causal relationship.

5. **Historical reference:** This issue was highlighted by Granger and Newbold (1974), who showed that regressions involving unrelated non-stationary series can misleadingly indicate strong associations.

Key Takeaway

Ensuring stationarity (or applying the right transformations) is essential for:

- Making valid statistical inferences,
- Avoiding false correlations,
- Building reliable time series models.

3.1.3 The Unit Circle, Complex Plane, and Stationarity

The complex plane. The complex plane is a way to represent numbers that have both a real part and an imaginary part:

$$z = a + bi, \quad a, b \in \mathbb{R}, \quad i = \sqrt{-1}.$$

The horizontal axis represents the real part (a), and the vertical axis represents the imaginary part (b). Every complex number is a point in this plane.

The unit circle. The *unit circle* is the set of all complex numbers z such that

$$|z| = \sqrt{a^2 + b^2} = 1.$$

It is a circle of radius 1 centered at the origin in the complex plane.

Why complex numbers appear. When we analyze AR processes, we form the *characteristic polynomial*. For example, in an AR(2):

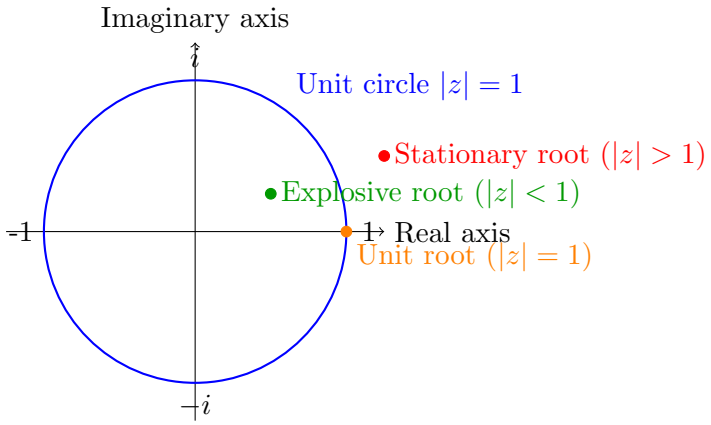
$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t, \quad \phi(L) = 1 - \phi_1 L - \phi_2 L^2.$$

We set $L = z$ and solve $\phi(z) = 0$. The solutions z^* can be real or complex. These roots capture how past values influence the present in the long run.

Why do we care about $|z^*| > 1$?

- If a root z^* lies **outside the unit circle** ($|z^*| > 1$), then shocks die out over time $\rightarrow$ stationary process.

- If a root lies **on the unit circle** ($|z^*| = 1$), shocks persist forever $\rightarrow$ unit root, non-stationary process.
- If a root lies **inside the unit circle** ($|z^*| < 1$), shocks amplify $\rightarrow$ explosive process.



3.1.4 Restrictions on AR(1) Coefficients

Start with the zero-mean AR(1) process

$$x_t = \beta x_{t-1} + e_t, \qquad e_t \sim \text{WN}(0, \sigma_e^2).$$

Define the lag (backshift) operator L by $Lx_t = x_{t-1}$. Then

$$\beta x_{t-1} = \beta L x_t,$$

so the AR(1) can be written in operator form as

$$x_t - \beta L x_t = e_t.$$

Factor out x_t to obtain

$$(1 - \beta L)\, x_t = e_t.$$

The polynomial in the lag operator,

$$\phi(L) \;=\; 1 - \beta L,$$

is called the *AR characteristic polynomial*. To find its characteristic root we replace the operator L by a complex number z and solve $\phi(z) = 0$:

$$\phi(z) = 1 - \beta z = 0 \quad \Longrightarrow \quad z^* = \frac{1}{\beta}.$$

The AR(p) stationarity condition requires all characteristic roots to lie *outside* the unit circle

in the complex plane. For AR(1) this means

$$|z^*| > 1 \quad \Longleftrightarrow \quad \left|\frac{1}{\beta}\right| > 1 \quad \Longleftrightarrow \quad |\beta| < 1.$$

Thus the AR(1) process has a single characteristic root $z^* = 1/\beta$, and it is covariance stationary if and only if $|\beta| < 1$.

3.1.5 Restrictions on the AR(2) Process

Consider the AR(2) model

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t, \qquad e_t \sim \text{WN}(0, \sigma_e^2).$$

Step 1. Lag operator form Using the lag operator L, we write

$$(1 - \phi_1 L - \phi_2 L^2) x_t = e_t.$$

The autoregressive characteristic polynomial is

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2.$$

Step 2. Characteristic equation Replace L with a complex number z to obtain

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0.$$

This quadratic equation has two roots, say z_1 and z_2.

Step 3. Stationarity condition The AR(2) process is stationary if and only if

$$|z_1| > 1 \quad \text{and} \quad |z_2| > 1.$$

Step 4. Inequalities for stationarity These root restrictions translate into the following parameter conditions:

$$\begin{aligned} \phi_2 &< 1, \\ \phi_2 &> -1, \\ \phi_1 + \phi_2 &< 1, \\ \phi_2 - \phi_1 &< 1. \end{aligned}$$

Step 5. Interpretation

- $\phi_2 < 1$ ensures the second lag does not generate explosive persistence.
- $\phi_2 > -1$ ensures oscillations do not grow unbounded.

- $\phi_1 + \phi_2 < 1$ controls the combined effect of both lags.
- $\phi_2 - \phi_1 < 1$ prevents the process from diverging when lags offset each other.

3.1.6 Restrictions on the AR(p) Process (Stationarity / Stability)

Consider the AR(p) process

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + e_t, \qquad e_t \sim \text{WN}(0, \sigma_e^2).$$

Characteristic polynomial and root condition Form the autoregressive characteristic polynomial in the lag operator L:

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p.$$

Replace L by a complex scalar z and solve the characteristic equation

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0.$$

Let the solutions (possibly complex) be $z_1^*, \ldots, z_p^*$.
Stationarity (stability) condition: $|z_i^*| > 1 \; \forall \, i = 1, \ldots p$. Equivalently, no root of $\phi(z) = 0$ lies on or inside the unit circle.

3.1.7 Examples: Stationarity of AR Processes

Which of the following AR processes are stationary, and why?

1. $X_t = 1.10 X_{t-1} + e_t$
2. $Y_t = 0.70 Y_{t-1} + 0.10 Y_{t-2} + e_t$
3. $Z_t = 0.80 Z_{t-1} + 0.30 Z_{t-2} + e_t$
4. $W_t = -0.80 W_{t-1} + 0.30 W_{t-2} + e_t$

Solutions:

- Process (1) is **not stationary** because $\beta_1 = 1.10 > 1$.
- Process (2) is **stationary** because:
  - Its lead coefficient is less than one in absolute value ($|\beta_1| = 0.70 < 1$).
  - Its coefficients add up to less than one ($\beta_1 + \beta_2 = 0.70 + 0.10 = 0.80 < 1$).
  - The coefficients are less than one unit apart ($\beta_2 - \beta_1 = 0.10 - 0.70 = -0.60 < 1$).
- Process (3) is **not stationary** as the coefficients add to more than one ($\beta_1 + \beta_2 = 0.80 + 0.30 = 1.10 > 1$).

- Process (4) is **not stationary**. While the first two conditions are met:
  - $|\beta_1| = 0.80 < 1$,
  - $\beta_2 - \beta_1 = 0.30 - (-0.80) = -0.50 < 1$,

  the third condition fails:

$$\beta_2 - (-\beta_1) = 0.30 - (-0.80) = 1.10 > 1.$$

Worked example: AR(2) — characteristic roots and their reciprocals

For an AR(2) process

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + e_t,$$

the characteristic polynomial is

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 = 0.$$

Divide by z^2 and set $u = z^{-1}$ to obtain the *inverse* characteristic equation

$$u^2 - \phi_1 u - \phi_2 = 0,$$

whose roots are $u_i = 1/z_i$. We solve both quadratics below.

(a) $X_t = 0.50X_{t-1} + 0.10X_{t-2} + e_t$.

1. Characteristic equation:

$$1 - 0.50z - 0.10z^2 = 0.$$

   Rearranged (standard quadratic form $az^2 + bz + c = 0$):

$$0.10z^2 + 0.50z - 1 = 0.$$

2. Solve by quadratic formula $z = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$:

$$a = 0.10, \quad b = 0.50, \quad c = -1.$$

   Discriminant:

$$b^2 - 4ac = 0.50^2 - 4(0.10)(-1) = 0.25 + 0.40 = 0.65.$$

$$z = \frac{-0.50 \pm \sqrt{0.65}}{2(0.10)} = \frac{-0.50 \pm 0.806225}{0.20}.$$

   Hence

$$z_1 = \frac{-0.50 + 0.806225}{0.20} \approx \frac{0.306225}{0.20} \approx 1.531125,$$

$$z_2 = \frac{-0.50 - 0.806225}{0.20} \approx \frac{-1.306225}{0.20} \approx -6.531125.$$

Both roots have $|z_i| > 1$, so the process is stationary.

3. Inverse characteristic equation (set $u = z^{-1}$):

$$u^2 - 0.50u - 0.10 = 0.$$

Solve:

$$u = \frac{0.50 \pm \sqrt{0.50^2 + 4(0.10)}}{2} = \frac{0.50 \pm \sqrt{0.65}}{2}.$$

Numerically:

$$u_1 = \frac{0.50 + 0.806225}{2} \approx 0.6531125, \qquad u_2 = \frac{0.50 - 0.806225}{2} \approx -0.1531125.$$

4. Check reciprocals:

$$1/z_1 \approx 1/1.531125 \approx 0.6531125 = u_1, \qquad 1/z_2 \approx 1/(-6.531125) \approx -0.1531125 = u_2.$$

3.1.8 Restrictions on MA(1) Coefficients

Consider the zero-mean MA(1) process

$$x_t = e_t + \theta e_{t-1}, \qquad e_t \sim \mathrm{WN}(0, \sigma_e^2).$$

In lag operator form,

$$x_t = (1 + \theta L)e_t.$$

The polynomial in the lag operator,

$$\theta(L) = 1 + \theta L,$$

is called the *MA characteristic polynomial*. To find its characteristic root we replace L by a complex number z and solve $\theta(z) = 0$:

$$\theta(z) = 1 + \theta z = 0 \quad \Longrightarrow \quad z^* = -\frac{1}{\theta}.$$

The MA(q) *invertibility* condition requires all roots of the MA polynomial to lie outside the unit circle. For MA(1) this means

$$|z^*| > 1 \quad \Longleftrightarrow \quad \left|-\frac{1}{\theta}\right| > 1 \quad \Longleftrightarrow \quad |\theta| < 1.$$

Thus the MA(1) process is always covariance stationary (no restriction), but is invertible if and only if $|\theta| < 1$.

3.1.9 Restrictions on MA(2) Coefficients

Consider the zero-mean MA(2) process

$$x_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}, \qquad e_t \sim \text{WN}(0, \sigma_e^2).$$

In lag operator form,

$$x_t = (1 + \theta_1 L + \theta_2 L^2) e_t.$$

The MA characteristic polynomial is

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2.$$

To obtain the characteristic roots, replace L by z and solve

$$1 + \theta_1 z + \theta_2 z^2 = 0.$$

The process is covariance stationary for all values of θ_1, θ_2. For invertibility, both roots of the quadratic equation must lie outside the unit circle in the complex plane:

$$|z_1| > 1, \quad |z_2| > 1.$$

This ensures that the MA(2) has a unique AR(∞) representation.

3.1.10 Restrictions on MA(q) Coefficients

Consider the zero-mean MA(q) process

$$x_t = e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}, \qquad e_t \sim \text{WN}(0, \sigma_e^2).$$

In lag operator form,

$$x_t = \theta(L) e_t, \qquad \theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q.$$

The MA characteristic polynomial is

$$\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q.$$

To find the characteristic roots, replace L by z and solve

$$1 + \theta_1 z + \cdots + \theta_q z^q = 0.$$

The MA(q) process is always covariance stationary (no restrictions). For invertibility, all roots

$\{z_1, \ldots, z_q\}$ of the MA polynomial must lie outside the unit circle:

$$|z_j| > 1 \quad \text{for all } j = 1, \ldots, q.$$

This guarantees a unique AR(∞) representation for the MA(q) process.

3.1.11 Invertibility of MA Processes

Consider the zero-mean MA(1) process

$$x_t = e_t + \theta e_{t-1}, \qquad e_t \sim \text{WN}(0, \sigma_e^2).$$

In lag operator form:

$$x_t = (1 + \theta L)e_t, \qquad \theta(L) = 1 + \theta L.$$

Definition (Invertibility).

An MA process is *invertible* if it can be written equivalently as an infinite-order AR process, i.e.

$$e_t = \psi(L)x_t, \qquad \psi(L) = 1 + \psi_1 L + \psi_2 L^2 + \cdots,$$

where the series converges absolutely. This allows us to recover the innovations uniquely from the observed data.

Example: MA(1) with $\theta = 0.5$ (invertible).

Start with

$$e_t = x_t - 0.5e_{t-1}.$$

Substituting recursively,

$$e_t = x_t - 0.5x_{t-1} + 0.25e_{t-2} = x_t - 0.5x_{t-1} + 0.25x_{t-2} - 0.125x_{t-3} + \cdots$$

so that

$$e_t = \sum_{j=0}^{\infty} (-0.5)^j x_{t-j}.$$

Since $|-0.5| < 1$, this infinite AR(∞) representation converges, and the process is invertible.

Example: MA(1) with $\theta = 2$ (not invertible).

Now

$$e_t = x_t - 2e_{t-1}.$$

Iterating,

$$e_t = x_t - 2x_{t-1} + 4x_{t-2} - 8x_{t-3} + \cdots$$

with coefficients $(-2)^j$ that diverge since $|-2| > 1$. Hence the process is not invertible.

General Condition.

For an MA(q) process

$$x_t = (1 + \theta_1 L + \cdots + \theta_q L^q)e_t,$$

invertibility requires all roots of the MA characteristic polynomial

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

to lie *outside* the unit circle in the complex plane. Equivalently, the parameters must be such that the infinite AR representation converges.

3.1.12 Restrictions on ARMA (1,1) Process

Consider the ARMA(1,1)

$$X_t = c + \phi X_{t-1} + \theta e_{t-1} + e_t, \qquad e_t \overset{iid}{\sim} (0, \sigma^2).$$

Assume $|\phi| < 1$ so the process is stationary and let $\mu = \mathbb{E}[X_t] = \dfrac{c}{1-\phi}$. Define the mean-zero series

$$Y_t = X_t - \mu,$$

so that

$$\boxed{Y_t = \phi Y_{t-1} + \theta e_{t-1} + e_t}$$

and $\gamma(h) = \mathrm{Cov}(Y_t, Y_{t-h})$.

General step ($h \geq 1$). Multiply the equation for Y_t by Y_{t-h} and take expectations:

$$\gamma(h) = \phi\,\gamma(h-1) + \theta\,\mathbb{E}[e_{t-1}Y_{t-h}] + \mathbb{E}[e_t Y_{t-h}].$$

Because e_t is independent of the past, $\mathbb{E}[e_t Y_{t-h}] = 0$ for all $h \geq 1$. Also, for $h \geq 2$ the shock e_{t-1} is independent of Y_{t-h} (since Y_{t-h} depends only on $e_{t-h}, e_{t-h-1}, \ldots$), so $\mathbb{E}[e_{t-1}Y_{t-h}] = 0$ for $h \geq 2$. Hence

$$\boxed{\gamma(h) = \phi\,\gamma(h-1) \quad \text{for } h \geq 2.}$$

Lag $h = 1$. For $h = 1$ we have

$$\gamma(1) = \phi\gamma(0) + \theta\,\mathbb{E}[e_{t-1}Y_{t-1}] + \underbrace{\mathbb{E}[e_t Y_{t-1}]}_{=0}.$$

Now $Y_{t-1} = \phi Y_{t-2} + \theta e_{t-2} + e_{t-1}$, so $\mathbb{E}[e_{t-1}Y_{t-1}] = \mathbb{E}[e_{t-1}^2] = \sigma^2$. Thus

$$\boxed{\gamma(1) = \phi\gamma(0) + \theta\,\sigma^2\ .}$$

Lag $h = 0$. Multiply the equation for Y_t by Y_t and take expectations:

$$\gamma(0) = \phi\gamma(1) + \theta\,\mathbb{E}[e_{t-1}Y_t] + \mathbb{E}[e_tY_t].$$

Compute the two expectations:

$$\mathbb{E}[e_tY_t] = \mathbb{E}[e_t(\phi Y_{t-1} + \theta e_{t-1} + e_t)] = \sigma^2,$$

and

$$\mathbb{E}[e_{t-1}Y_t] = \mathbb{E}\big[e_{t-1}(\phi Y_{t-1} + \theta e_{t-1} + e_t)\big] = \phi\mathbb{E}[e_{t-1}Y_{t-1}] + \theta\mathbb{E}[e_{t-1}^2] = \phi\sigma^2 + \theta\sigma^2 = (\phi+\theta)\sigma^2.$$

Therefore

$$\gamma(0) = \phi\gamma(1) + \theta(\phi+\theta)\sigma^2 + \sigma^2 = \phi\gamma(1) + \big(1 + \phi\theta + \theta^2\big)\sigma^2.$$

(Equivalently one can write the additive term as $1 + \theta(\phi+\theta)$.)

Closed form for $\gamma(0)$. Substitute $\gamma(1) = \phi\gamma(0) + \theta\sigma^2$ into the last equation:

$$\gamma(0) = \phi\big(\phi\gamma(0) + \theta\sigma^2\big) + (1 + \phi\theta + \theta^2)\sigma^2 = \phi^2\gamma(0) + \big(1 + 2\phi\theta + \theta^2\big)\sigma^2.$$

Rearrange to obtain the familiar closed form:

$$\boxed{\gamma(0) = \frac{1 + 2\phi\theta + \theta^2}{1 - \phi^2}\,\sigma^2\ .}$$

Summary (Yule–Walker conditions for ARMA(1,1)).

$$\boxed{\begin{aligned}
&\gamma(0) = \phi\gamma(1) + \big(1 + \phi\theta + \theta^2\big)\sigma^2,\\
&\gamma(1) = \phi\gamma(0) + \theta\sigma^2,\\
&\gamma(h) = \phi\gamma(h-1), \qquad h \geq 2.
\end{aligned}}$$

Dividing by $\gamma(0)$ gives the autocorrelation recursion:

$$\rho(1) = \phi + \frac{\theta\sigma^2}{\gamma(0)}, \qquad \rho(h) = \phi\rho(h-1)\ (h \geq 2),$$

with $\gamma(0)$ as above.

These equations hold under stationarity ($|\phi| < 1$). They are the ARMA(1,1) analogue of the usual Yule–Walker relations for AR models.

3.2 Non-Stationarity in Time Series

So far, we have focused on time series whose statistical properties, particularly the mean and variance, do not change over time. This assumption is often unrealistic in practice. Many economic and financial variables—such as GDP, price indices, and stock market indices—exhibit persistent growth or decline, making them inherently non-stationary.

A time series is said to be *non-stationary* if its mean, variance, or autocovariance depends on time. When this occurs, standard statistical methods that rely on stationarity assumptions cannot be applied directly. Nevertheless, non-stationary models such as the random walk and the random walk with drift have proven to be surprisingly effective forecasting tools, particularly in macroeconomic and financial applications.

3.2.1 Differencing and Integration

Differencing plays a role in time series analysis analogous to differentiation in calculus. By subtracting successive observations, differencing removes trends and stabilizes the mean of a series.

- If a series becomes stationary after first differencing, it is said to be integrated of order one, denoted $I(1)$.
- More generally, if a series must be differenced d times to achieve stationarity, it is integrated of order d, written $I(d)$.

Example

Consider the time series

$$\{y_t\} = \{2, 4, 6, 8, 10\}.$$

This series is non-stationary due to a deterministic upward trend. The first difference is

$$\Delta y_t = \{2, 2, 2, 2\},$$

which has a constant mean and variance. Hence, the original series is integrated of order one, or $I(1)$.

3.2.2 Canonical Non-Stationary Processes

We now examine four commonly encountered non-stationary processes and their integration properties.

Random Walk

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } (0, \sigma^2).$$

The variance of y_t increases linearly with time, implying that the process is non-stationary. First differencing yields

$$\Delta y_t = \varepsilon_t,$$

which is stationary. Hence, a random walk is an $I(1)$ process.

Random Walk with Drift

$$y_t = \mu + y_{t-1} + \varepsilon_t.$$

The mean of the series grows linearly over time, while the variance diverges. Differencing gives

$$\Delta y_t = \mu + \varepsilon_t,$$

which is stationary. Thus, this process is also $I(1)$.

Deterministic Trend Process

$$y_t = \alpha + \beta t + \varepsilon_t.$$

Non-stationarity arises from the deterministic trend in the mean, while the variance remains constant. Differencing removes the trend:

$$\Delta y_t = \beta + (\varepsilon_t - \varepsilon_{t-1}),$$

yielding a stationary process.

Random Walk with Drift and Deterministic Trend

$$y_t = \alpha + \beta t + y_{t-1} + \varepsilon_t.$$

Both the mean and variance depend on time. First differencing removes the unit root but leaves a deterministic trend. A second difference is required to achieve stationarity, implying that the process is $I(2)$.

Chapter 4

Unit Root Tests

4.1 Why Test for Unit Roots?

Many economic time series appear to trend over time, but visual inspection alone cannot distinguish between deterministic trends and stochastic trends. This distinction is critical, as it determines whether detrending or differencing is the appropriate transformation prior to modeling.

Unit root tests are statistical procedures designed to assess whether a time series contains a stochastic trend. In particular, they help determine whether shocks to a series have permanent or transitory effects—an issue central to forecasting, policy analysis, and statistical inference.

4.2 Illustrations of Non-Stationary Processes

The following figures illustrate the non-stationary processes discussed in the previous section and provide visual intuition for their distinct behavior.

4.2.1 Random Walk

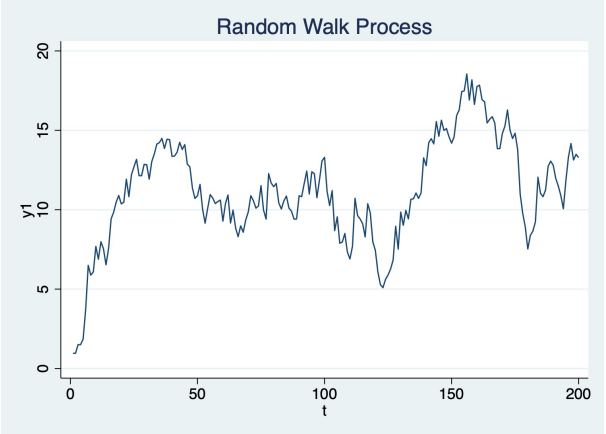


Figure 4.1: Random Walk Process

4.2.2 Random Walk with Drift

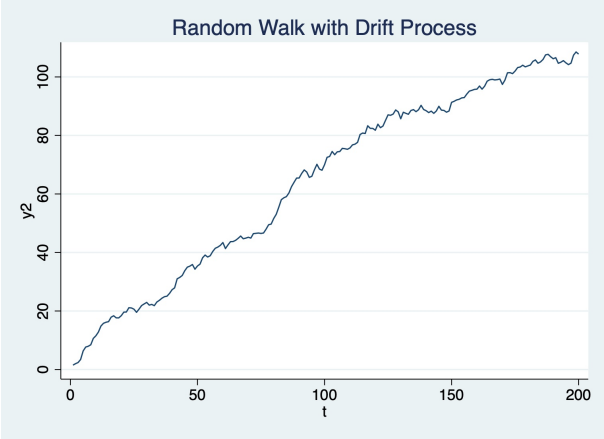


Figure 4.2: Random Walk with Drift

4.2.3 Deterministic Trend

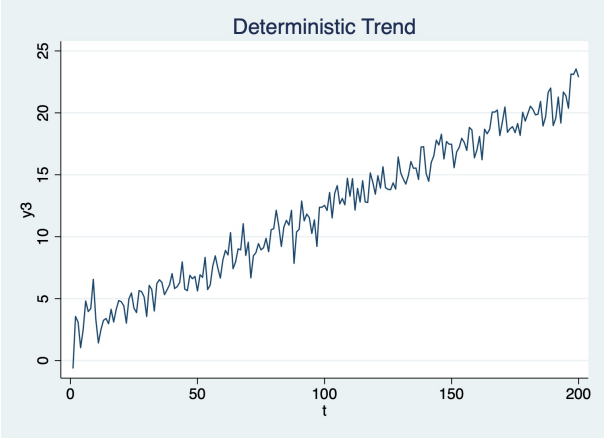


Figure 4.3: Deterministic Trend Process

4.2.4 Random Walk with Drift and Deterministic Trend

4.3 Dickey–Fuller Test

To formally assess whether a time series contains a unit root, we begin with the Dickey–Fuller (DF) test. The DF test is designed to test the null hypothesis that a series follows a random walk, implying non-stationarity.

Consider the simple autoregressive model:

$$y_t = \rho y_{t-1} + \varepsilon_t,$$

where ε_t is white noise. The series is non-stationary if $\rho = 1$. Subtracting y_{t-1} from both sides

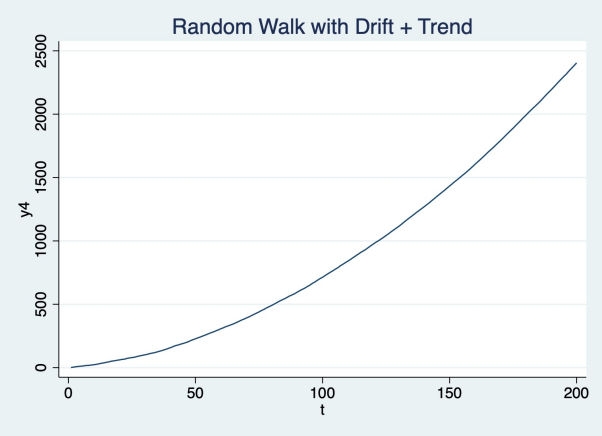


Figure 4.4: Random Walk with Drift and Trend

yields

$$\Delta y_t = (\rho - 1) y_{t-1} + \varepsilon_t.$$

Defining $\gamma = \rho - 1$, we can write:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t.$$

The hypotheses of the Dickey–Fuller test are:

$$H_0 : \gamma = 0 \qquad \text{(unit root; non-stationary)}$$
$$H_1 : \gamma < 0 \qquad \text{(stationary)}$$

If the null hypothesis is rejected, the series is deemed stationary. It is important to note that the distribution of the test statistic under the null is non-standard; therefore, critical values differ from those of the conventional t-test.

The basic Dickey–Fuller regression can be extended to allow for deterministic components:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \varepsilon_t,$$

or

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \varepsilon_t,$$

depending on whether a constant or a linear time trend is believed to be present.

Limitation of the Dickey–Fuller Test

The Dickey–Fuller test assumes that the error term ε_t is serially uncorrelated. In practice, this assumption is often violated, particularly in economic time series. When residuals exhibit autocorrelation, the DF test becomes invalid, motivating the need for a more general testing framework.

4.4 Augmented Dickey–Fuller Test

To address serial correlation in the error term, the Dickey–Fuller test is augmented by including lagged differences of the dependent variable. This leads to the Augmented Dickey–Fuller (ADF) test.

The ADF test is based on the following regression:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \varepsilon_t,$$

where:

- y_t is the series under investigation,
- $\Delta y_t = y_t - y_{t-1}$ is the first difference,
- t is an optional deterministic time trend,
- p is the number of lagged difference terms included to absorb serial correlation, and
- ε_t is a white-noise error term.

The hypotheses of the ADF test remain:

$$\begin{aligned} H_0 &: \gamma = 0 \qquad \text{(unit root; non-stationary)} \\ H_1 &: \gamma < 0 \qquad \text{(stationary)} \end{aligned}$$

If the estimated value of γ is sufficiently negative—so that the test statistic is less than the appropriate Dickey–Fuller critical value—we reject the null hypothesis and conclude that the series is stationary. Otherwise, we fail to reject the null, indicating the presence of a unit root.

4.4.1 Lag Length Selection

An important practical issue in implementing the ADF test is choosing the appropriate number of lagged difference terms (p). The goal is to include enough lags to ensure that the residuals of the test regression are approximately white noise.

Including too few lags may leave residual autocorrelation and bias the test, while including too many lags reduces the power of the test.

Two commonly used approaches to lag selection are:

1. **Visual inspection:** Examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to assess the degree of serial dependence.
2. **Information criteria:** Selecting the lag length that minimizes an information criterion such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC).

In `Stata`, the `varsoc` command can be used to compute AIC, HQIC, and BIC values for different lag lengths. The lag order that minimizes the BIC is often preferred, as it penalizes overfitting more strongly.

Chapter 5

Structural Breaks

5.1 Introduction

Macroeconomic relationships are rarely stable over long periods. Policy regimes, institutional frameworks, and external shocks can all alter the structural parameters that govern an economy's behavior. Such changes are known as *structural breaks*.

Detecting structural breaks is essential because ignoring them can lead to misleading inferences. For instance, a fiscal multiplier estimated under one policy regime may not hold after a regime change. Likewise, monetary transmission mechanisms can vary across inflation-targeting versus fixed-exchange-rate periods.

Structural break analysis therefore plays a vital role in macroeconomic policy evaluation, model stability testing, and forecasting accuracy. It allows economists to identify when relationships among key variables—such as output, inflation, and interest rates—undergo fundamental shifts.

5.2 Why Structural Breaks Matter in Macroeconomic Policy

Macroeconomic models often assume parameter stability. For example, the Phillips Curve relates inflation to unemployment, and the IS–LM framework relates output to interest rates. However, such relationships are contingent on institutional and policy environments.

When a central bank changes its policy rule (e.g., moving from discretionary to inflation targeting), or when a country undergoes a financial crisis or regime shift, the underlying data-generating process for key variables can change.

If econometric models ignore these breaks, estimated coefficients become biased and inconsistent. Policymakers relying on such models risk implementing ineffective or even counterproductive policies. Hence, testing for and accounting for structural breaks is a cornerstone of sound macroeconomic analysis.

5.3 The Chow Test: A Primer

The Chow test is a classical econometric test for the presence of a structural break at a *known* breakpoint. It is based on comparing the fit of two separate regressions—before and after the suspected break—with that of a pooled regression estimated over the full sample.

Consider a linear model:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{5.1}$$

Suppose there is a known breakpoint at observation T_b. The Chow test examines whether the parameters before and after T_b are equal:

$$H_0 : \beta_0^{(1)} = \beta_0^{(2)}, \quad \beta_1^{(1)} = \beta_1^{(2)} \tag{5.2}$$

Under the null hypothesis of parameter stability, the restricted (pooled) model is valid. The test statistic is:

$$F = \frac{(SSR_P - (SSR_1 + SSR_2))/k}{(SSR_1 + SSR_2)/(T_1 + T_2 - 2k)} \tag{5.3}$$

where:

- SSR_P is the sum of squared residuals from the pooled regression,
- SSR_1 and SSR_2 are the residual sums of squares from the subsample regressions before and after T_b,
- k is the number of estimated parameters.

Under H_0, $F - stat$ follows an $F \sim (k, T_1 + T_2 - 2k)$ distribution. If $F - stat$ exceeds the critical value, we reject the null hypothesis and conclude that a structural break exists at T_b.

5.3.1 Example: Testing for a Structural Change with the Chow Test

Suppose we want to study how **personal savings** (Y) in the United States are related to **disposable personal income** (X) using data from 1970 to 1995. A simple way to model this relationship is with a regression:

$$Y_t = \alpha_1 + \alpha_2 X_t + u_t$$

This model assumes that the relationship between savings and income has remained the *same* throughout the entire 26-year period. However, that may not be realistic. For instance, in 1982 the U.S. experienced a severe recession, with unemployment reaching nearly 10%. Such an event could have changed people's saving behavior.

To check whether this recession caused a **structural change**, we divide the data into two periods:

- Period 1: 1970–1981 (before the recession)
- Period 2: 1982–1995 (after the recession)

We then estimate three regressions:

$$\begin{aligned} \text{(1) Before 1982:} \quad & Y_t = \lambda_1 + \lambda_2 X_t + u_{1t}, \quad n_1 = 12 \\ \text{(2) After 1982:} \quad & Y_t = \gamma_1 + \gamma_2 X_t + u_{2t}, \quad n_2 = 14 \\ \text{(3) Entire period:} \quad & Y_t = \alpha_1 + \alpha_2 X_t + u_t, \quad n = 26 \end{aligned}$$

If there is **no structural change**, then the coefficients should be the same across both periods:

$$H_0 : \lambda_1 = \gamma_1 \quad \text{and} \quad \lambda_2 = \gamma_2$$

That is, the intercept and slope coefficients are the same across both subperiods — the regression relationship between Y_t and X_t did not change after 1981.

$$H_A : \lambda_1 \neq \gamma_1 \quad \text{and} \quad \lambda_2 \neq \gamma_2$$

That is, at least one of the coefficients (intercept or slope) differs across the two periods — implying a structural break after 1981.

However, if the 1982 recession did change the savings–income relationship, then at least one of these parameters will differ between the two subperiods. The **Chow test** formally checks this by comparing how well the combined regression (for all years) fits the data against how well the two separate regressions (before and after 1982) fit their respective data. If the two separate regressions fit *significantly better*, we conclude that a **structural break** occurred i.e. the relationship between savings and disposable income changed after 1982.

Table 5.1: Regression of Savings on Income

	Full Sample	**Pre-1981**	**Post-1981**	**Pooled (Chow)**
Income	0.0377***	0.0803***	0.0149	0.0803***
	(0.0042)	(0.0092)	(0.0084)	(0.0091)
D (Post-1981 dummy)	–	–	–	152.479***
	–	–	–	(33.092)
D × Income	–	–	–	-0.0655***
	–	–	–	(0.124)
Constant	62.423***	1.016	153.495***	1.016
	(12.761)	(10.604)	(32.712)	(10.524)
Observations	26	12	14	26
R-squared	0.767	0.902	0.207	0.882

Let RSS_p denote the residual sum of squares from the pooled regression (with dummy and interaction), and RSS_1 and RSS_2 denote the residual sum of squares from the separate regressions for the pre-1981 and post-1981 periods, respectively. The Chow test F-statistic is calculated as:

$$
\begin{aligned}
F &= \frac{(SSR_P - (SSR_1 + SSR_2))/k}{(SSR_1 + SSR_2)/(T_1 + T_2 - 2k)} \\
SSR_P &= 23248.3, \quad SSR_1 + SSR_2 = 11790.254, \quad T_1 = 12, \quad T_2 = 14, \quad k = 2 \\
\text{Numerator} &= \frac{23248.3 - 11790.254}{2} = \frac{11458.046}{2} = 5729.023 \\
\text{Denominator} &= \frac{11790.254}{12 + 14 - 4} = \frac{11790.254}{22} \approx 535.012 \\
F &= \frac{5729.023}{535.012} \approx 10.71
\end{aligned}
$$

Since $F_{cal} > F_{crit}$, we reject the null hypothesis and conclude that a structural break occurred around 1981 in the relationship between Savings and Income.

5.4 The Bai–Perron Test for Multiple Structural Breaks

While the Chow test is suitable for detecting a single structural break at a known point, most macroeconomic time series experience multiple breaks at *unknown* dates. The Bai–Perron test generalizes the Chow framework by allowing for multiple unknown breakpoints in linear regression models.

5.4.1 Model Framework

Consider the linear regression model:

$$y_t = x_t'\beta_i + u_t, \quad t = T_{i-1} + 1, \ldots, T_i, \quad i = 1, \ldots, m + 1 \tag{5.4}$$

where y_t is the dependent variable, x_t is a $(k \times 1)$ vector of regressors, β_i are regime-specific parameters, and u_t is an error term. The break dates $(T_1, T_2, \ldots, T_m)$ are unknown and to be estimated.

The null hypothesis assumes parameter stability:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{m+1} \tag{5.5}$$

against the alternative of m structural changes.

5.4.2 Estimation and Test Statistics

The Bai–Perron method estimates break dates by minimizing the global sum of squared residuals (SSR) across all possible segmentations subject to a minimum segment length, often expressed as a trimming parameter ε:

$$\{\hat{T}_1, \ldots, \hat{T}_m\} = \arg\min_{T_1, \ldots, T_m} \sum_{i=1}^{m+1} \sum_{t=T_{i-1}+1}^{T_i} (y_t - x_t'\hat{\beta}_i)^2 \tag{5.6}$$

where each $\hat{\beta}_i$ is estimated via OLS within its regime.

Bai and Perron proposed several test statistics:

- **SupF(m)**: tests the null of no structural break against the alternative of m breaks.
- **UDmax**: tests for the presence of at least one break (an overall test).
- **Sequential F-tests (SupF($l+1|l$))**: test whether adding another break significantly improves model fit.

These tests rely on comparing the decrease in SSRs between restricted (pooled) and unrestricted (segmented) models, analogous to the logic of the Chow test but generalized for multiple and unknown breaks.

5.4.3 Intuition

The Core Idea

The Bai-Perron test is a powerful method for detecting **multiple structural breaks** in time series data. The fundamental intuition is that a single regression model is inadequate when the underlying economic relationship has changed at several unknown points in time. Instead of assuming constant parameters, we allow the regression coefficients to change at multiple breakpoints that the algorithm discovers automatically.

Step 1: Problem Formulation

Consider a standard linear regression model:

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \quad t = 1, 2, \ldots, T \tag{5.7}$$

Where:

- Y_t: Dependent variable (e.g., sales, GDP growth)
- X_t: Independent variable(s)
- α, β: Regression parameters
- ε_t: Error term

The Bai Perron test checks for whether α and/or β change at unknown points during the sample period?

Step 2: The Segmentation Concept

Bai-Perron views the timeline as divisible into $m+1$ segments (where m is the number of breaks), with stable parameters within each segment.

- **No breaks** ($m = 0$): Single regression: $Y_t = \alpha + \beta X_t + \varepsilon_t$
- **One break** ($m = 1$): Two different regimes:

$$\begin{aligned} &\text{Segment 1: } Y_t = \alpha_1 + \beta_1 X_t + \varepsilon_t, \quad t = 1, \ldots, \tau_1 \\ &\text{Segment 2: } Y_t = \alpha_2 + \beta_2 X_t + \varepsilon_t, \quad t = \tau_1 + 1, \ldots, T \end{aligned}$$

- **Two breaks** ($m = 2$): Three different regimes, etc.

Step 3: Finding Optimal Breakpoints

For a given number of breaks m, how do we find the best locations? Bai-Perron uses **global minimization of sum of squared residuals (SSR)**.

1. For each possible combination of breakpoints $\{\tau_1, \tau_2, \ldots, \tau_m\}$, estimate separate regressions for each segment
2. Calculate the total SSR across all segments:

$$SSR(\tau_1, \ldots, \tau_m) = \sum_{j=1}^{m+1} \sum_{t=\tau_{j-1}+1}^{\tau_j} (Y_t - \hat{\alpha}_j - \hat{\beta}_j X_t)^2 \tag{5.8}$$

   where $\tau_0 = 0$ and $\tau_{m+1} = T$
3. The optimal breakpoints minimize this total SSR:

$$\{\hat{\tau}_1, \ldots, \hat{\tau}_m\} = \arg \min_{\tau_1, \ldots, \tau_m} SSR(\tau_1, \ldots, \tau_m) \tag{5.9}$$

Step 4: Determining the Number of Breaks

Since we don't know m in advance, Bai-Perron uses **sequential testing**:

1. **Test** $H_0 : m = 0$ **vs** $H_1 : m = 1$
   - Use SupF test: $SupF = \max_\tau F(\tau)$
   - If significant, reject $H_0 \rightarrow$ at least one break exists
2. **Test** $H_0 : m = k$ **vs** $H_1 : m = k + 1$
   - Given k breaks, test each segment for an additional break
   - Use maximum of individual segment F-statistics
   - Stop when no significant additional break is found

5.4.4 Model Selection and Interpretation

The number of breaks, m, can be determined using sequential tests or information criteria such as the Bayesian Information Criterion (BIC) or the Schwarz criterion.

The estimated breakpoints $\hat{T}_i$ correspond to dates when the underlying relationship between variables changes significantly—for instance, following major policy reforms, crises, or regime shifts.

Once the breakpoints are identified, regime-specific coefficients $\hat{\beta}_i$ can be estimated to characterize each subperiod's dynamics.

5.4.5 Applications in Macroeconomics

In macroeconomic contexts, the Bai–Perron test is widely used to detect:

- Shifts in GDP growth or productivity trends,
- Changes in monetary policy regimes (e.g., pre- and post-inflation targeting),
- Breaks in fiscal policy response functions,
- Structural changes in exchange rate or trade relationships.

For example, applying the Bai–Perron test to GDP data may reveal breaks in the GDP series coinciding with major recessions or policy reforms.

5.4.6 Comparison with the Chow Test

The Chow test assumes a *known* single break date, while the Bai–Perron procedure allows for *multiple, unknown* breaks and optimally estimates both the number and timing of these changes. It thus provides a more flexible and data-driven approach to detecting structural instability in economic relationships.

5.4.7 Example: India GDP Series

GDP levels are almost always nonstationary (they have a stochastic trend). If you apply Bai–Perron directly to GDP levels, you risk spurious breaks — the algorithm may detect breaks that simply reflect long-run growth rather than actual structural changes. So, testing on GDP levels is meaningful only if you explicitly model the deterministic trend:

$$GDP_t = \beta_0 + \beta_1 t + \epsilon$$

and look for trend breaks (changes in slope/intercept). In empirical literature, if the aim is to test for structural changes in growth, then use of growth rates (first differences or log differences) is common. On the other hand, if testing for structural changes in potential output trend, then use GDP levels with a time trend.

Chapter 6

Vector Autoregressions (VAR)

6.1 Introduction

In the previous chapters, we have largely operated in a world of individual variables. We have considered modeling a single time series - such as unemployment rates, stock prices, inflation rates, GDP growth rates—using ARMA models. We learned to identify autoregressive components, moving average structures, and to test for stationarity. We also developed methods to detect structural breaks, those moments when the underlying data-generating process of a single series fundamentally changes.

Now, let's consider a concrete example that will serve as our running illustration throughout this chapter: the relationship between the unemployment rate (y_t) and the inflation rate (x_t). This relationship is classically captured by the Phillips Curve, which posits an inverse relationship between these two fundamental macroeconomic variables. The traditional time series analyst, armed with our univariate toolkit, might approach these two series separately. For unemployment, we would specify an autoregressive model:

$$y_t = \alpha_1 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_{1t} \tag{6.1}$$

where y_t represents the unemployment rate at time t. For inflation, we would similarly specify:

$$x_t = \alpha_2 + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \cdots + \gamma_p x_{t-p} + \varepsilon_{2t} \tag{6.2}$$

where x_t represents the inflation rate. In this approach, each series lives in its own self-contained universe, explained only by its own history, with any structural breaks modeled within each equation separately.

This separate-equations approach, while statistically sound for certain purposes, suffers from

a fundamental limitation: it ignores the economic reality that these variables interact with each other. Economic theory, particularly the Phillips Curve relationship, suggests that unemployment and inflation are intrinsically linked. Changes in inflation may influence future unemployment decisions, while labor market conditions may affect price-setting behavior and therefore inflation.

The error terms ε_{1t} and ε_{2t} in equations (1) and (2) must therefore contain not only pure random shocks but also the omitted effects of each variable on the other. When we model the series in isolation, we are essentially assuming that the past values of one variable contain no useful information for predicting the current value of the other variable, once we have accounted for its own history. This is a strong assumption—and one that often contradicts economic theory and intuition.

6.2 The Vector Autoregression

Vector Autoregression (VAR) represents a paradigm shift from this isolated approach. Rather than treating each variable as living in its own universe, VAR models recognize that economic variables typically exist within interconnected systems. The core insight is simple yet profound: to properly model one variable, we often need to include the history of other related variables.

Let us return to our unemployment-inflation example. The VAR framework would model these two series as a coupled system:

$$y_t = \alpha_1 + \phi_{11} y_{t-1} + \phi_{12} y_{t-2} + \phi_{13} x_{t-1} + \phi_{14} x_{t-2} + \varepsilon_{1t} \tag{6.3}$$

$$x_t = \alpha_2 + \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \phi_{23} x_{t-1} + \phi_{24} x_{t-2} + \varepsilon_{2t} \tag{6.4}$$

Notice the crucial difference: each equation now includes not only the variable's own lags but also the lags of the other variable in the system. The unemployment equation (3) includes lagged inflation terms (x_{t-1}, x_{t-2}), while the inflation equation (4) includes lagged unemployment terms (y_{t-1}, y_{t-2}).

6.2.1 Understanding the VAR Structure

The term "vector" in Vector Autoregression comes from the compact matrix representation of this system. Define a vector containing our two variables:

$$\mathbf{Z}_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} \tag{6.5}$$

We can then write our two-equation system as a single vector equation:

$$\mathbf{Z}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{Z}_{t-1} + \mathbf{\Phi}_2 \mathbf{Z}_{t-2} + \varepsilon_t \tag{6.6}$$

where $\mathbf{c} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$ is a vector of constants, $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are matrices of coefficients, and $\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$ is a vector of error terms.

This representation reveals the elegant symmetry of the VAR approach. We are essentially performing autoregression—hence "autoregression" in the name—but on a vector of variables rather than a single scalar variable.

6.2.2 The Economic Interpretation

Recall, the VAR framework:

$$y_t = \alpha_1 + \phi_{11} y_{t-1} + \phi_{12} y_{t-2} + \phi_{13} x_{t-1} + \phi_{14} x_{t-2} + \varepsilon_{1t}$$
$$x_t = \alpha_2 + \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \phi_{23} x_{t-1} + \phi_{24} x_{t-2} + \varepsilon_{2t}$$

What does this coupled system represent economically? Consider the unemployment equation (1). The coefficient γ_{11} measures how much last period's inflation affects current unemployment, holding constant unemployment's own history. Similarly, in the inflation equation (2), ϕ_{21} measures how much last period's unemployment affects current inflation.

The error terms ε_{1t} and ε_{2t} now represent true structural shocks—the portions of each variable that cannot be predicted from the joint history of both series. These are the innovations that potentially propagate through the system, affecting both variables in future periods.

This framework naturally captures the feedback effects that characterize most economic systems. For example, a shock to inflation (ε_{2t}) not only affects future inflation directly but also affects future unemployment through the cross-equation linkages, which in turn may feed back to affect inflation in subsequent periods.

6.2.3 Why VAR Matters: Three Fundamental Advantages

The VAR approach offers several important advantages over traditional single-equation methods:

- *Realistic Dynamics*: By allowing for cross-variable interactions, VAR models capture the rich interdependencies that characterize actual economic systems. The approach acknowledges that economic variables rarely evolve in isolation.
- *Forecasting Improvement:* Since each variable's equation includes more relevant information (the history of other related variables), VAR models typically produce better forecasts

than univariate models, particularly at medium-term horizons.

- *Structural Analysis:* The VAR framework provides a natural structure for analyzing how shocks propagate through an economic system—what we will later call impulse response analysis. We can trace out the dynamic effects of an unexpected change in one variable on the entire system over time.

As we proceed through this chapter, we will develop the technical details of VAR modeling—estimation, lag selection, inference, and interpretation. But the fundamental insight to carry forward is this: economic variables belong to the same universe, and modeling them as interacting partners rather than isolated individuals provides a more faithful representation of economic reality.

6.3 Estimating VAR Models

6.3.1 The Estimation Problem

Consider the VAR(p) model with K variables:

$$\mathbf{y}_t = \mathbf{c} + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_p \mathbf{y}_{t-p} + \varepsilon_t \tag{6.7}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})'$ is a $K \times 1$ vector of endogenous variables, $\mathbf{c}$ is a $K \times 1$ vector of constants, $\mathbf{\Phi}_i$ are $K \times K$ coefficient matrices, and ε_t is a $K \times 1$ white noise process with $\mathbb{E}[\varepsilon_t] = \mathbf{0}$ and $\mathbb{E}[\varepsilon_t \varepsilon_t'] = \mathbf{\Sigma}_\varepsilon$.

6.3.2 OLS Estimation

A remarkable property of VAR models is that **each equation can be estimated consistently using Ordinary Least Squares (OLS)**. This holds even though the equations are seemingly linked through the covariance matrix of the error terms. To see why, let's consider a bivariate VAR(2) system:

$$y_t = a_1 + \phi_{11} y_{t-1} + \phi_{12} x_{t-1} + \phi_{13} y_{t-2} + \phi_{14} x_{t-2} + e_{1t}, \tag{6.8}$$

$$x_t = a_2 + \phi_{21} y_{t-1} + \phi_{22} x_{t-1} + \phi_{23} y_{t-2} + \phi_{24} x_{t-2} + e_{2t}. \tag{6.9}$$

Let

$$Y_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix}, \quad e_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \quad z_t' = \begin{pmatrix} 1 & y_{t-1} & x_{t-1} & y_{t-2} & x_{t-2} \end{pmatrix}.$$

Define the coefficient matrix

$$B = \begin{pmatrix} a_1 & a_2 \\ \phi_{11} & \phi_{21} \\ \phi_{12} & \phi_{22} \\ \phi_{13} & \phi_{23} \\ \phi_{14} & \phi_{24} \end{pmatrix}.$$

Then equations (6.8)–(6.9) can be written compactly as:

$$Y_t = z_t' B + e_t.$$

Stacking Over Time

For $t = 3, \ldots, T$, stack all observations:

$$Y = \begin{pmatrix} Y_3' \\ Y_4' \\ \vdots \\ Y_T' \end{pmatrix}, \quad Z = \begin{pmatrix} z_3' \\ z_4' \\ \vdots \\ z_T' \end{pmatrix}, \quad E = \begin{pmatrix} e_3' \\ e_4' \\ \vdots \\ e_T' \end{pmatrix}.$$

The system in matrix form becomes:

$$\boxed{Y = ZB + E}$$

where

- Y is $(T-2) \times 2$,
- Z is $(T-2) \times 5$,
- B is 5×2,
- E is $(T-2) \times 2$.

OLS Estimation

Since both equations share the same regressors, ordinary least squares (OLS) applied equation-by-equation is equivalent to estimating the whole system jointly.
The OLS estimator for the coefficient matrix B is:

$$\boxed{\widehat{B} = (Z'Z)^{-1} Z'Y}.$$

The fitted values and residuals are:

$$\widehat{Y} = Z\widehat{B}, \qquad \widehat{E} = Y - Z\widehat{B}.$$

The estimated contemporaneous covariance matrix of the residuals is:

$$\widehat{\Sigma}_u = \frac{1}{T-2-k}\,\widehat{E}'\widehat{E},$$

where $k = 5$ is the number of regressors (including the constant).

Variance of the OLS Estimator

Under standard assumptions, the covariance of the vectorized estimator is:

$$\mathrm{Var}\big(\,\mathrm{vec}(\widehat{B})\big) = \Sigma_u \otimes (Z'Z)^{-1}.$$

For equation j, the covariance matrix of $\widehat{\beta}_j$ is:

$$\mathrm{Var}(\widehat{\beta}_j) = \sigma_{jj}(Z'Z)^{-1},$$

where σ_{jj} is the jth diagonal element of Σ_u.

6.3.3 Why OLS Works in VAR Systems

1. **Identical Regressors**: All equations have the same set of regressors (lagged values of all variables in the system)
2. **Efficiency**: While OLS per equation is consistent, the Seemingly Unrelated Regressions (SUR) estimator would be more efficient if the error terms across equations were correlated
3. **SUR = OLS**: In VAR models, since all equations have identical regressors, the SUR estimator reduces to equation-by-equation OLS
4. **Consistency**: Under standard regularity conditions, the OLS estimators are consistent and asymptotically normal
5. **Standard Errors**: The coefficients are identical in the two approaches (OLS and SUR). However, they differ in their standard errors because OLS assumes uncorrelated errors across equations. When shocks are correlated, standard errors must be adjusted accordingly. Stata's `var` and `varbasic` commands automatically handle this using the SUR framework. It can be shown that SUR estimates coincide with those from `var`.

6.3.4 The Full System Estimation

The entire VAR system can be compactly written as:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \tag{6.10}$$

where:

$$\mathbf{Y} = [\mathbf{y}_{p+1}, \mathbf{y}_{p+2}, \ldots, \mathbf{y}_T]' \tag{6.11}$$

$$\mathbf{B} = [\mathbf{c}, \mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p]' \tag{6.12}$$

$$\mathbf{E} = [\boldsymbol{\varepsilon}_{p+1}, \boldsymbol{\varepsilon}_{p+2}, \ldots, \boldsymbol{\varepsilon}_T]' \tag{6.13}$$

The multivariate OLS estimator is:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{6.14}$$

This provides the complete set of parameter estimates for the entire VAR system.

6.3.5 Error Properties in VAR Models

In a reduced-form VAR model, the error vector $\{u_t\}$ is assumed to satisfy:

$$E[u_t] = 0, \qquad E[u_t u_t'] = \Sigma_u, \qquad E[u_t u_s'] = 0 \quad \text{for } t \neq s.$$

Thus, the i.i.d. assumption is *not violated across time* — the error sequence $\{u_t\}$ is independent and identically distributed over t. However, the errors are *not independent across variables* at a given time period, since the components of u_t can be contemporaneously correlated:

$$\text{Cov}(u_{it}, u_{jt}) \neq 0 \quad \text{for some } i \neq j.$$

Why This is Not a Problem

This cross-equation contemporaneous correlation does not violate the classical estimation assumptions because:

- **Unbiasedness and Consistency:** Each equation in the VAR can be estimated using OLS without bias or inconsistency, since the regressors (lagged values of Y_t) are *predetermined* with respect to u_t:
$$E[z_t' u_t] = 0.$$

- **Efficiency:** Although the OLS estimators are unbiased and consistent, efficiency can be improved using system methods such as Seemingly Unrelated Regressions (SUR) or full-system GLS. However, in a standard VAR, each equation has the same set of regressors. In this special case, OLS and GLS yield identical coefficient estimates, i.e.,
$$\widehat{B}_{OLS} = \widehat{B}_{GLS}.$$

Therefore, contemporaneous correlation between equations does not violate the OLS assumptions, nor the i.i.d. assumption across time.

6.3.6 Practical Implementation

In practice, VAR estimation proceeds as follows:

1. Select the lag order p using information criteria (AIC, BIC, HQ)
2. For each equation $i = 1, \ldots, K$:
   - Regress y_{it} on a constant and p lags of all K variables
   - Store the coefficient estimates $\hat{\boldsymbol{\beta}}_i$
3. Collect the error terms and estimate the covariance matrix:

$$\hat{\boldsymbol{\Sigma}}_\varepsilon = \frac{1}{T - Kp - 1} \sum_{t=p+1}^{T} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t' \tag{6.15}$$

This straightforward OLS approach makes VAR models particularly attractive for empirical work, as they avoid the computational complexities of many other multivariate time series methods.

6.4 Example: Inflation and Unemployment

The relationship between inflation and unemployment has long been a central topic in macroeconomics, commonly captured by the concept of the *Phillips Curve*. Originally proposed by A.W. Phillips (1958), the Phillips Curve describes an inverse relationship between the rate of inflation and the rate of unemployment: when unemployment is low, inflation tends to rise, and when unemployment is high, inflation tends to fall. This relationship suggests a short-run trade-off faced by policymakers between stabilizing prices and promoting employment. Over time, however, economists have debated the stability of this trade-off, particularly after the experience of stagflation in the 1970s, which challenged the notion of a consistent inverse relationship. Modern interpretations often incorporate expectations, leading to the expectations-augmented Phillips Curve, which emphasizes that the long-run relationship between inflation and unemployment may be neutral once inflation expectations adjust.

VAR - Algebraic (scalar) form

The bivariate VAR(3) in scalar equations is

$$\begin{aligned}\text{unrate}_t = \alpha_1 &+ \beta_{11,1}\,\text{unrate}_{t-1} + \beta_{11,2}\,\text{unrate}_{t-2} + \beta_{11,3}\,\text{unrate}_{t-3} \\ &+ \beta_{12,1}\,\text{inflation}_{t-1} + \beta_{12,2}\,\text{inflation}_{t-2} + \beta_{12,3}\,\text{inflation}_{t-3} + e_{1t}, \qquad (6.16)\end{aligned}$$

$$\begin{aligned}\text{inflation}_t = \alpha_2 &+ \beta_{21,1}\,\text{unrate}_{t-1} + \beta_{21,2}\,\text{unrate}_{t-2} + \beta_{21,3}\,\text{unrate}_{t-3} \\ &+ \beta_{22,1}\,\text{inflation}_{t-1} + \beta_{22,2}\,\text{inflation}_{t-2} + \beta_{22,3}\,\text{inflation}_{t-3} + e_{2t}, \qquad (6.17)\end{aligned}$$

where $\{e_t\} = (e_{1t}, e_{2t})'$ is the reduced-form error vector.

VAR(3) — Matrix Form

Define the 2×1 vector of endogenous variables:

$$Z_t = \begin{pmatrix} \text{unrate}_t \\ \text{inflation}_t \end{pmatrix}, \qquad u_t = \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}.$$

Let $\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}$, and define the 2×2 lag coefficient matrices as follows:

$$A_1 = \begin{pmatrix} \beta_{11,1} & \beta_{12,1} \\ \beta_{21,1} & \beta_{22,1} \end{pmatrix}, \qquad A_2 = \begin{pmatrix} \beta_{11,2} & \beta_{12,2} \\ \beta_{21,2} & \beta_{22,2} \end{pmatrix}, \qquad A_3 = \begin{pmatrix} \beta_{11,3} & \beta_{12,3} \\ \beta_{21,3} & \beta_{22,3} \end{pmatrix}.$$

Then the VAR(3) system can be written compactly as

$$\boxed{Z_t = \alpha + A_1 Z_{t-1} + A_2 Z_{t-2} + A_3 Z_{t-3} + u_t.}$$

6.4.1 Unit Root Tests

- **No constant (no drift)**
  Model:
  $$\Delta y_t = \alpha y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + u_t$$
  Hypotheses:
  $$H_0 : \alpha = 0 \quad \text{(unit root; random walk without drift)}$$
  $$H_1 : \alpha < 0 \quad \text{(stationary)}$$

- **With constant (drift)**
  Model:
  $$\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + u_t$$

Hypotheses:

$$H_0 : \alpha = 0 \quad \text{(unit root with drift)}$$

$$H_1 : \alpha < 0 \quad \text{(stationary around a constant)}$$

- **With constant and trend**

  Model:

$$\Delta y_t = \mu + \beta t + \alpha y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + u_t$$

  Hypotheses:

$$H_0 : \alpha = 0 \quad \text{(unit root (with/without drift))}$$

$$H_1 : \alpha < 0 \quad \text{(trend-stationary)}$$

ADF test statistics

Table 6.1: Augmented Dickey–Fuller test results

Variable	Model	ADF stat	1% crit	5% crit	p-value
Unrate	No constant	-0.851	-2.592	-1.950	–
	Constant (drift)	-2.802	-2.350	-1.654	0.0029
	Trend	-2.740	-4.019	-3.442	0.2197
Inflation	No constant	-1.174	-2.592	-1.950	–
	Constant (drift)	-2.301	-2.350	-1.655	0.0113
	Trend	-2.435	-4.020	-3.442	0.3614

Conclusions (comparison to 5% critical value)

- **Unrate**
  - No constant: ADF = -0.851, 5% critical = -1.950.
    Since $-0.851 > -1.950$, do **not** reject H_0 at 5% (nonstationary).
  - Constant (drift): ADF = -2.802, 5% critical = -1.654.
    Since $-2.802 < -1.654$, **reject** H_0 at 5% (stationary around a constant). (reported p-value = 0.0029)
  - Trend: ADF = -2.740, 5% critical = -3.442.
    Since $-2.740 > -3.442$, do **not** reject H_0 at 5% (not trend-stationary).
- **Inflation**
  - No constant: ADF = -1.174, 5% critical = -1.950.
    Since $-1.174 > -1.950$, do **not** reject H_0 at 5% (nonstationary).
  - Constant (drift): ADF = -2.301, 5% critical = -1.655.
    Since $-2.301 < -1.655$, **reject** H_0 at 5% (stationary around a constant). (reported p-value = 0.0113)

– Trend: ADF = -2.435, 5% critical = -3.442.
Since $-2.435 > -3.442$, do **not** reject H_0 at 5% (not trend-stationary).

Both *inflation* and *unemployment* are stationary around a constant.

6.4.2 Lag Order Selection Criteria

To determine the appropriate lag length for the VAR model including the variables *unrate* and *inflation*, the lag-order selection statistics were computed using the `varsoc` command in Stata. The sample covers the period from 1961Q2 to 2000Q4 with 159 observations. The results are presented in Table 6.2.

Table 6.2: VAR Lag Order Selection Criteria

Lag	LL	LR	df	p-value	FPE	AIC	HQIC	SBIC
0	-438.414				0.872882	5.5398	5.5555	5.5784
1	-35.406	806.02	4	0.000	0.005771	0.5208	0.5679	0.6366
2	3.858	78.528	4	0.000	0.003704	0.0773	0.1556	**0.2703**
3	10.5538	**13.392**	4	0.010	**0.003580**	**0.0433**	**0.1531**	0.3136
4	13.0203	4.9331	4	0.294	0.003651	0.0626	0.2037	0.4101

The information criteria (Final Prediction Error, Akaike Information Criterion, and Hannan–Quinn Information Criterion) all reach their minimum values at lag 3. Although the Schwarz Bayesian Information Criterion (SBIC) selects a shorter lag length of 2, the majority of the criteria—particularly AIC, HQIC, and FPE—suggest a lag order of three.

Based on the Akaike Information Criterion (AIC) and supporting evidence from FPE and HQIC, the optimal lag length for the VAR model is 3.

6.4.3 Vector Autoregression (VAR(3)) Estimation Results

A bivariate Vector Autoregression (VAR) model with three lags was estimated for *unrate* and *inflation* over the sample period 1961Q1–2000Q4. The model was estimated using 160 quarterly observations. The overall fit statistics suggest that the VAR(3) provides a good fit to the data.

- The coefficients on the first and second lags of both *unrate* and *inflation* are statistically significant in several cases, indicating that past values of these variables have important effects on their current values. Both series display clear evidence of persistence.

- In the **unrate equation**, the first lag of unemployment (`L1.unrate`) is large and highly significant (1.567, $p < 0.01$), confirming strong persistence in the unemployment rate. The second lag (`L2.unrate`) is negative and significant (-0.692, $p < 0.01$), suggesting partial mean reversion after accounting for the first lag. Among the inflation lags, only the first lag (`L1.inflation`) is marginally significant (0.154, $p = 0.037$), indicating that higher inflation in the previous quarter slightly increases the current unemployment rate. The other inflation lags are not statistically significant.

Table 6.3: Vector Autoregression Results: Unemployment Rate and Inflation (1961Q1–2000Q4)

	Coefficient	Std. Err.	z	P>z	[95% Conf. Interval]
Equation: unrate					
L1.unrate	1.567205	0.076961	20.36	0.000	[1.416364, 1.718046]
L2.unrate	-0.691751	0.134117	-5.16	0.000	[-0.954616, -0.428886]
L3.unrate	0.078435	0.075944	1.03	0.302	[-0.070413, 0.227284]
L1.inflation	0.153515	0.073620	2.09	0.037	[0.009223, 0.297807]
L2.inflation	-0.121067	0.089903	-1.35	0.178	[-0.297273, 0.055139]
L3.inflation	0.117145	0.077036	1.52	0.128	[-0.033843, 0.268134]
cons	0.121848	0.079681	1.53	0.126	[-0.034325, 0.278020]
Equation: inflation					
L1.unrate	-0.295985	0.080101	-3.70	0.000	[-0.452980, -0.138991]
L2.unrate	0.368962	0.139588	2.64	0.008	[0.095374, 0.642549]
L3.unrate	-0.115111	0.079042	-1.46	0.145	[-0.270031, 0.039809]
L1.inflation	0.676385	0.076623	8.83	0.000	[0.526207, 0.826563]
L2.inflation	0.120046	0.093570	1.28	0.200	[-0.063348, 0.303440]
L3.inflation	0.199405	0.080179	2.49	0.013	[0.042258, 0.356553]
cons	0.255973	0.082932	3.09	0.002	[0.093430, 0.418516]

- In the **inflation equation**, inflation shows strong own dynamics. The first lag of inflation (`L1.inflation`) is positive and highly significant (0.676, $p < 0.01$), and the third lag (`L3.inflation`) is also positive and significant (0.199, $p = 0.013$), pointing to persistence in inflation over multiple quarters. Regarding unemployment, the first lag (`L1.unrate`) has a negative and significant coefficient (-0.296, $p < 0.01$)), meaning that higher unemployment in the previous quarter tends to lower current inflation—consistent with a short-run Phillips curve effect. The second lag (`L2.unrate`) is positive and significant (0.369, $p < 0.01$), suggesting that the disinflationary impact of unemployment may reverse after about half a year.

- Overall, both the unemployment rate and inflation exhibit strong inertia. The results show a dynamic interaction between the two variables: inflation affects unemployment positively with a short lag, while unemployment tends to reduce inflation in the near term. These patterns are consistent with a lagged Phillips curve relationship where real and nominal dynamics adjust over time.

6.4.4 Granger Causality

Granger causality is a statistical concept used to test whether past values of one variable contain information that helps predict another variable, beyond what is already explained by its own past values. In other words, if including the lagged values of X improves the forecast accuracy of Y, we say that X Granger-causes Y. It does not imply true causal influence, but rather a predictive or temporal precedence relationship. Thus, Granger causality captures predictive causality, not necessarily structural or real-world causation.

In the VAR(3) framework:

- **Unemployment (unrate) Granger-causes Inflation** if the past values of unrate help predict inflation, beyond the information contained in past values of inflation.
- **Inflation Granger-causes Unemployment** if the past values of inflation help predict unrate, beyond the information contained in past values of unrate.

Formally, the hypotheses are defined as follows:

(a) Testing whether *Unemployment Rate* Granger-causes *Inflation*:

$$H_0 : \beta_{21,1} = \beta_{21,2} = \beta_{21,3} = 0 \quad \text{(unemployment rate does not Granger-cause inflation)}$$

$$H_1 : \text{At least one of } \beta_{21,i} \neq 0,\ i = 1, 2, 3 \quad \text{(unemployment rate Granger-causes inflation)}$$

(b) Testing whether Inflation Granger-causes Unemployment Rate:

$$H_0 : \beta_{12,1} = \beta_{12,2} = \beta_{12,3} = 0 \quad \text{(inflation does not Granger-cause unemployment rate)}$$

$$H_1 : \text{At least one of } \beta_{12,i} \neq 0,\ i = 1, 2, 3 \quad \text{(inflation Granger-causes unemployment rate)}$$

Interpretation

Testing for Granger causality involves estimating each equation separately and performing a joint significance test (typically an F-test or Wald test) on the coefficients of the lagged explanatory variables. If the null hypothesis is rejected, it implies that the lagged values of the corresponding variable contain predictive information for the dependent variable.

Remarks on Model Specification

- Granger causality does not imply true causation; it only reflects predictive relationships.
- Omitted variables can bias results:
  - If a third variable Z_t influences both unrate_t and inflation_t, omitting Z_t may yield *spurious* causality.
  - Conversely, excluding relevant dynamics can mask genuine predictive relationships.
- Economic theory should guide variable selection and lag length.

The Granger causality tests in Table 6.4 indicate statistically significant *bidirectional predictive relationships* between inflation and unemployment. The null hypothesis that inflation does not Granger-cause unemployment is strongly rejected ($\chi^2 = 23.48$, $p < 0.01$), implying that past values of inflation contain significant information for predicting the unemployment rate. Similarly, the null hypothesis that unemployment does not Granger-cause inflation is also rejected ($\chi^2 = 19.03$, $p < 0.01$), suggesting that lagged unemployment contributes to forecasting

inflation dynamics. Overall, these results point to a two-way Granger causal relationship between inflation and unemployment over the sample period 1961Q1–2000Q4, consistent with the dynamic interactions implied by the estimated VAR(3) model.

Table 6.4: Granger Causality Wald Tests: Inflation and Unemployment

Equation	**Excluded Variable**	**Chi-Square**	**df**	**p-value**
Unemployment Rate	Inflation	23.483	1	0.000
Unemployment Rate	All	23.483	1	0.000
Inflation	Unemployment Rate	19.032	1	0.006
Inflation	All	19.032	1	0.006

Notes: Null hypothesis — the excluded variable does not Granger-cause the dependent variable. Rejection of the null (small p-value) indicates predictive causality in the Granger sense. Results are based on the VAR(3) model estimated over 1961Q1–2000Q4.

6.4.5 Impulse Response Functions

We begin with the response of the unemployment rate to inflation shocks. The response starts close to zero, indicating that inflation shocks have little immediate impact on unemployment. Over subsequent quarters, unemployment rate rises steadily, reaching around 0.3–0.35 after roughly 10–12 quarters. Rather than oscillating, the increase is smooth and persistent, suggesting that inflation shocks have a lasting positive effect on unemployment.

The VAR results do not support the traditional short-run Phillips curve, which predicts that higher inflation reduces unemployment. Instead, the findings point to a reverse or long-run Phillips curve effect, where inflation shocks are followed by rising unemployment. This outcome may reflect policy tightening in response to inflation, upward shifts in inflation expectations, or cost-push shocks that raise prices while depressing output and employment. Overall, the results suggest that inflationary pressures ultimately lead to weaker labor market conditions rather than lower unemployment.adjustment frictions in wages and prices.

Next, the IRF for the response of inflation to an unemployment rate shock provides clear evidence consistent with the Phillips curve. A positive shock to unemployment leads to a noticeable and persistent decline in inflation, indicating that higher unemployment exerts downward pressure on prices. The response is negative from the outset and remains below zero throughout the forecast horizon, showing little sign of returning to baseline. This suggests that unemployment shocks have a lasting disinflationary effect rather than a short-lived one. Overall, the IRF implies a stable and persistent inverse relationship between unemployment and inflation, in line with the traditional Phillips curve intuition.

6.4.6 Forecast Error Variance Decomposition (FEVD)

The forecast error variance decomposition (FEVD) measures the proportion of the forecast variance of each endogenous variable that can be attributed to shocks in each variable within

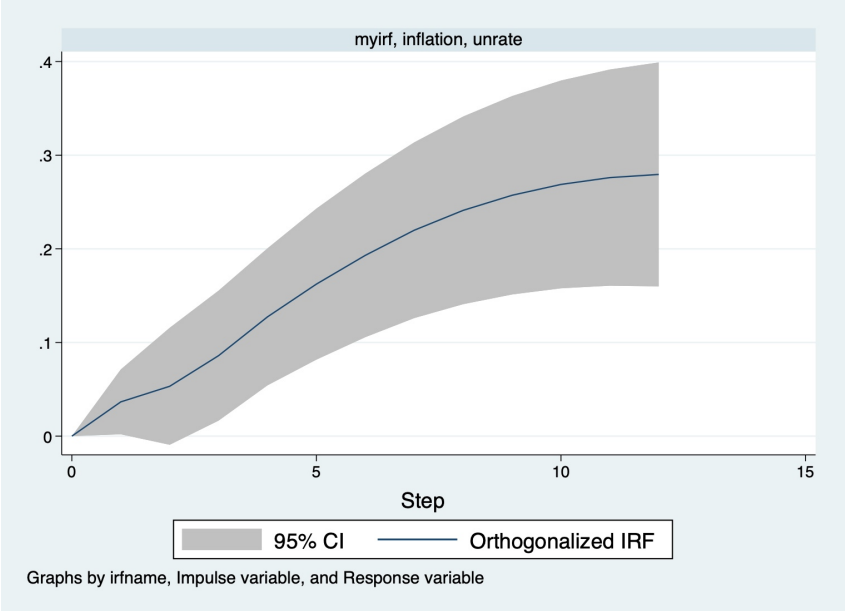


Figure 6.1: Impulse: *inflation*; Response: *unrate*

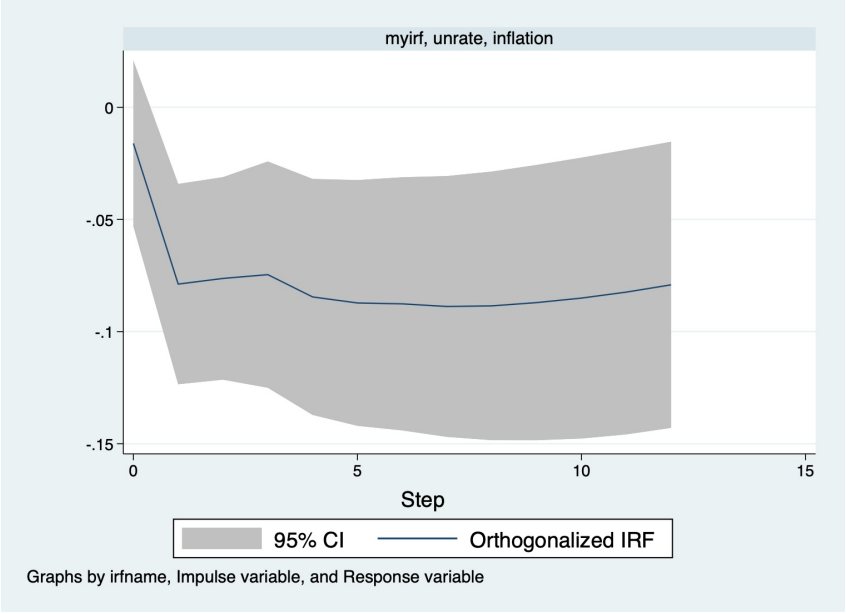


Figure 6.2: Impulse: *unrate*; Response: *inflation*

the system. Table 6.5 reports the FEVD for *unrate* and *inflation* at selected forecast horizons (steps).

Table 6.5: Forecast Error Variance Decomposition (FEVD)

Step	Response	Shock from unrate	Shock from inf	Lower (95%)	Upper (95%)
0	unrate	1.000	0.000	–	–
	inflation	0.000	1.000	–	–
1–3	unrate	0.988	0.111	-0.002	0.223
	inflation	0.012	0.889	0.777	1.001
4–6	unrate	0.926	0.173	0.006	0.339
	inflation	0.074	0.827	0.661	0.995
7–9	unrate	0.798	0.224	0.019	0.429
	inflation	0.202	0.776	0.571	0.981
10–12	unrate	0.658	0.264	0.030	0.498
	inflation	0.342	0.736	0.502	0.970

Interpretation

- At short horizons (1–3 quarters), the forecast variance of **unrate** is almost entirely explained by its own shocks (around 99% or more), with inflation contributing less than 1% to its variability.
- As the horizon lengthens, the role of inflation shocks in explaining unemployment variability rises steadily. By 6 quarters, inflation shocks account for roughly 17% of the forecast variance, reaching about 34% by the 12th quarter. This pattern indicates that inflation shocks exert an increasingly important and persistent influence on unemployment over time.
- For **inflation**, own shocks remain the dominant source of variation at all horizons—about 99% in the first quarter and still around 74% by the 12th quarter. However, the share of variance explained by unemployment shocks grows gradually from nearly zero initially to roughly 26% at longer horizons, reflecting a moderate feedback from labor market conditions to inflation dynamics.
- Overall, the FEVD results suggest that both series are primarily driven by their own innovations in the short run, but cross-variable effects become more pronounced as the forecast horizon increases. This reinforces the dynamic interaction between inflation and unemployment consistent with the Phillips curve mechanism—where each variable increasingly influences the other over time.

6.5 Appendix - OLS in Matrix Form

The Linear Regression Model is given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + ... + \beta_k X_{ik} + \epsilon_i; i = 1, 2, ...n$$

The linear regression model expressed in matrix form is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ 1 & X_{31} & X_{32} & \dots & X_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$Y = X\beta + \epsilon$$

where

- Y is an $n \times 1$ vector of dependent variables.

- X is an $n \times k$ matrix of independent variables.

- β is a $k \times 1$ vector of coefficients.

- ϵ is an $n \times 1$ vector of errors.

The sum squared of errors is given by

$$SSE = \epsilon'\epsilon = (Y - X\beta)'(Y - X\beta)$$

Taking the derivative with respect to β and setting it to zero:

$$\frac{\partial}{\partial \beta}(Y - X\beta)'(Y - X\beta) = -2X'Y + 2X'X\beta = 0$$

Solving for β:

$$X'X\beta = X'Y$$
$$\hat{\beta} = (X'X)^{-1}X'Y$$

Thus, the OLS estimator is $\hat{\beta} = (X'X)^{-1}X'Y$

Chapter 7

Error Correction Models

7.1 Introduction

In the previous chapter, we used VAR models to study how variables move together in the short run. VARs work well when all the variables are stationary i.e. when their means and variances don't change over time.

However, most economic variables, such as GDP, money supply, prices, are not stationary. They tend to grow over time. To deal with this, we usually take first differences (for example, using the change in GDP instead of the level of GDP). This makes the data stationary and allows us to use a VAR.

But there's a problem: When we take differences, we lose important long-run information. We can see how things change from one period to another, but we can't see the actual levels or long-term relationships between the variables. For example, if two variables move together over time differencing them separately ignores the fact that they might share a stable long-run relationship. This is where cointegration comes in. If two or more non-stationary variables move together in such a way that their combination is stationary, then they are said to be cointegrated.

Cointegration means there is a long-run equilibrium relationship linking these variables. Engle and Granger (1987) showed that instead of differencing everything, we can model both: short-run changes (how variables adjust each period), and long-run relationships (how they move together over time). The model that does this is called the Vector Error Correction Model (VECM).

7.2 Error Correction Mechanism (ECM) and Cointegration

When two variables, Y_t and X_t, are cointegrated, it means that even though each of them is non-stationary, there exists a stable long-run relationship between them in levels.

$$Y_t = \beta X_t + \varepsilon_t,$$

where the residual ε_t is stationary. This residual represents the deviation from the long-run equilibrium.

If Y_t and X_t are cointegrated, regressing Y_t on X_t is not spurious. When the two variables drift apart, there exists a mechanism that drives them back toward equilibrium. This is called an error correction mechanism (ECM).

Suppose X_t follows a random walk:

$$X_t = X_{t-1} + u_t,$$

and that Y_t is related to X_t through a long-run equilibrium:

$$Y_t = \beta X_t + \varepsilon_t.$$

According to **Granger**:

> If two variables are cointegrated, then there exists an equivalent **Error Correction Model (ECM)** representation.

The simplest ECM for Y_t and X_t can be written as:

$$\Delta Y_t = \gamma \, \Delta X_t - \alpha(Y_{t-1} - \beta X_{t-1}) + e_t,$$

or equivalently,

$$\Delta Y_t = \gamma \, \Delta X_t - \alpha \, \varepsilon_{t-1} + e_t,$$

where Δ denotes the first difference operator.

7.2.1 Interpretation

- ΔY_t and ΔX_t represent **short-run changes**.
- $(Y_{t-1} - \beta X_{t-1})$ is the **long-run equilibrium error**.
- α measures the **speed of adjustment** toward equilibrium: if $\alpha > 0$, then Y_t moves back toward equilibrium when it deviates in the previous period.
- γ captures the **short-run effect** of changes in X_t on Y_t.

The ECM states that: The change in Y_t depends on (a) the change in X_t and (b) the extent to which Y_{t-1} and X_{t-1} were out of equilibrium in the previous period. When there are multiple cointegrated variables, the ECM generalizes to a Vector Error Correction Model (VECM),

which captures both: a) short-run dynamics (in first differences), and b) long-run equilibrium relationships (in levels).

7.2.2 The Error Correction Mechanism and the Speed of Adjustment

In the Error Correction Model (ECM), the change in Y_t depends on two key components: the change in X_t and the *lagged error* from the long-run equilibrium relationship between Y_t and X_t. The term in parentheses, $(Y_{t-1} - \beta X_{t-1})$, represents how far the system was from equilibrium in the previous period i.e. the deviation or "error" from the cointegrating relationship. Because this term appears with a lag, the model assumes that the adjustment to disequilibrium takes at least one period to begin. This is intuitive: in most economic systems, when an imbalance occurs, it takes time for markets, firms, or agents to recognize it and adjust their behavior.

The ECM implies that when Y_{t-1} is *above* the level implied by its equilibrium relationship with X_{t-1} (so the error term is positive), the change in Y_t should decrease to restore balance. Conversely, when Y_{t-1} is *below* its equilibrium value (the error term is negative), Y_t tends to increase, pushing the system back toward equilibrium. This feedback mechanism ensures that deviations from the long-run relationship are corrected over time, preventing the variables from drifting apart indefinitely even though both are non-stationary in levels.

The parameter α — often called the **speed of adjustment coefficient** — quantifies how strongly and how quickly Y_t responds to disequilibrium. A higher value of α means faster adjustment: for example, if $\alpha = 0.5$, half of the previous period's disequilibrium is corrected in the next period. A smaller α implies slower adjustment, and if $\alpha = 0$, there is no correction mechanism at all, meaning the variables would drift apart without returning to equilibrium. Thus, α determines how efficiently the system restores its long-run balance following short-run shocks.

7.2.3 Deriving the Error Correction Model from an ARDL / VAR(1) equation

Start from the dynamic (ARDL / single-equation VAR(1)) specification:

$$Y_t = \alpha_{11} X_t + \lambda_1 Y_{t-1} + \lambda_2 X_{t-1} + e_t. \tag{7.1}$$

Subtract Y_{t-1} from both sides to express the equation in differences:

$$Y_t - Y_{t-1} = \alpha_{11} X_t + \lambda_1 Y_{t-1} + \lambda_2 X_{t-1} + e_t - Y_{t-1} \tag{7.2}$$

$$\Delta Y_t = \alpha_{11} X_t + (\lambda_1 - 1) Y_{t-1} + \lambda_2 X_{t-1} + e_t. \tag{7.3}$$

Add and subtract $\alpha_{11}X_{t-1}$ on the right-hand side of (7.3):

$$\Delta Y_t = \alpha_{11}X_t - \alpha_{11}X_{t-1} + (\lambda_1 - 1)Y_{t-1} + \lambda_2 X_{t-1} + \alpha_{11}X_{t-1} + e_t \tag{7.4}$$

$$= \alpha_{11}\Delta X_t + (\lambda_1 - 1)Y_{t-1} + (\lambda_2 + \alpha_{11})X_{t-1} + e_t. \tag{7.5}$$

Now define the convenient parameter reparametrizations

$$\alpha \equiv 1 - \lambda_1, \qquad \theta \equiv \lambda_2 + \alpha_{11}.$$

Note that $\lambda_1 - 1 = -\alpha$. Substituting these definitions gives

$$\Delta Y_t = \alpha_{11}\Delta X_t - \alpha Y_{t-1} + \theta X_{t-1} + e_t. \tag{7.6}$$

Factor the lagged terms to expose the cointegrating (long-run) combination. If we set $\beta \equiv \theta/\alpha$ (assuming $\alpha \neq 0$), then

$$\Delta Y_t = \alpha_{11}\Delta X_t - \alpha\Big(Y_{t-1} - \frac{\theta}{\alpha}X_{t-1}\Big) + e_t \tag{7.7}$$

$$= \alpha_{11}\Delta X_t - \alpha\big(Y_{t-1} - \beta X_{t-1}\big) + e_t. \tag{7.8}$$

This is the Error Correction Model (ECM) representation:

$$\boxed{\Delta Y_t = \alpha_{11}\Delta X_t - \alpha\big(Y_{t-1} - \beta X_{t-1}\big) + e_t}$$

where $(Y_{t-1} - \beta X_{t-1})$ is the lagged equilibrium error (the cointegrating residual) and α is the speed-of-adjustment coefficient.

7.3 The Engle–Granger Approach

Estimating an Error Correction Model (ECM) is conceptually straightforward. In the single-equation case, the ECM can be estimated using ordinary least squares (OLS) with standard regression tools (e.g., `reg` in Stata). Later, when we deal with systems of equations (Vector ECMs), we will use Stata's `vec` command. Before that, we begin with the simpler, single-equation case.

7.3.1 The ECM Representation

Suppose that two variables X_t and Y_t are cointegrated. A simple ECM for these variables can be written as:

$$\Delta Y_t = \delta + \gamma\,\Delta X_t - \alpha(Y_{t-1} - \beta X_{t-1}) + e_t, \tag{7.9}$$

where:

- Δ denotes the first difference operator,

- $(Y_{t-1} - \beta X_{t-1})$ represents the *long-run equilibrium error*, and
- α is the *speed-of-adjustment* parameter that measures how quickly the system returns to equilibrium after a shock.

7.3.2 Cointegration - Two Variables

Two or more variables are said to be **cointegrated** if:

1. Each variable is integrated of order one, $I(1)$, and
2. A linear combination of them is integrated of a lower order, typically $I(0)$.

In other words, even though X_t and Y_t may each be non-stationary in levels, there exists a linear combination such as:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t,$$

where the residual ε_t is stationary. This residual captures the long-run equilibrium relationship between Y_t and X_t.

7.3.3 The Engle–Granger Two-Step Procedure

Once we have established that the variables are cointegrated, the Engle–Granger (1987) approach proceeds in two simple steps:

1. **Estimate the cointegrating equation** and obtain the residuals:

   $$Y_t = \beta_0 + \beta_1 X_t + \hat{\varepsilon}_t.$$

   These residuals, $\hat{\varepsilon}_t$, measure the deviation from the long-run equilibrium.

2. **Estimate a short-run model in differences**, but include the lagged residuals as a regressor:

   $$\Delta Y_t = \delta + \gamma \, \Delta X_t - \alpha \, \hat{\varepsilon}_{t-1} + e_t.$$

   The lagged residual term, $\hat{\varepsilon}_{t-1}$, represents the previous period's disequilibrium, and the coefficient α captures how strongly Y_t adjusts to restore equilibrium.

A constant term can be included in the long-run regression, either inside the parentheses of the ECM or outside of it, as shown in Equation (7.9).

7.3.4 Important Practical Considerations

- In the second-stage ADF test on the residuals, **do not include a constant** if one was already included in the first-stage regression. (For example, in Stata, the command `reg Y X` automatically includes a constant.)

- **Unit root tests have low power**, particularly when the true process is close to having a unit root. This means cointegration tests can sometimes fail to reject the null hypothesis even when cointegration is present (see Elliott, 1998).
- The second-stage unit root test on $\hat{\varepsilon}_t$ requires **adjusted critical values**, since the test is based on *estimated* residuals. The appropriate critical values were tabulated by:
  - Engle and Yoo (1987),
  - MacKinnon (1991, 2010).

  These residual-based critical values differ from the standard DF/ADF tables, to account for the extra uncertainty introduced in the first-stage estimation.

7.3.5 Interpretation

If the residuals $\hat{\varepsilon}_t$ are stationary, then X_t and Y_t share a long-run equilibrium relationship. We can then estimate the ECM in Equation (7.9), which links the short-run dynamics (in differences) to the long-run equilibrium (in levels). The coefficient α in the ECM captures how strongly Y_t responds to deviations from this equilibrium, thereby correcting disequilibrium over time.

7.4 The Johansen Approach to Cointegration

The Engle–Granger residual-based tests of cointegration are intuitive and well-suited for testing a single cointegrating relationship between two variables. However, they are not designed to identify *multiple* cointegrating equations, which may arise when we are working with systems that include more than two variables. In such cases, a more general approach is needed. This approach was pioneered by **Helmut Johansen**.

7.4.1 Overview of Johansen's Contributions

Johansen developed a maximum likelihood framework for testing and estimating the number of cointegrating relationships in a system of equations. His key papers include:

- Johansen (1988): Developed the eigenvalue tests for cointegration in the case with no constants or seasonal dummies in the long-run cointegrating equations.
- Johansen (1991): Extended the framework to allow for constants, deterministic trends, and seasonal dummy variables — features that are crucial in empirical work and that affect the distribution of the test statistics.
- Johansen (1995b): Expanded the method to cases where the variables are integrated of order two, $I(2)$.
- Johansen (1994): Provided a non-technical summary of these results.

7.4.2 Cointegration and the Rank of the Matrix Π

The long-run (cointegrating) relationships between variables in a vector autoregressive (VAR) model can be represented in the matrix Π of the Vector Error Correction Model (VECM):

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \cdots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Pi Y_{t-k} + \varepsilon_t.$$

All of the cointegrating information in the system is contained in the term ΠY_{t-k}. The number of cointegrating equations is determined by the **rank** of the matrix Π, denoted r.

- If $r = 0$, then Π is a null matrix and the variables are not cointegrated. In this case, we estimate a VAR in first differences if the variables are $I(1)$.
- If $r = n$, then all variables are stationary and we estimate a VAR in levels.
- The interesting case is when $0 < r < n$: there are r linearly independent cointegrating relationships among the n variables.

7.4.3 Interpreting the Rank r

If there are n variables in the system, there can be at most $n - 1$ linearly independent cointegrating vectors. You cannot have n independent cointegrating vectors among n variables.
For example, consider two variables X_t and Y_t. If the residuals from the regression

$$Y_t = a + bX_t + e_t$$

are stationary, then X_t and Y_t are cointegrated. In that case, the residuals from the reverse regression,

$$X_t = a' + b'Y_t + e'_t,$$

will also be stationary. This is because one equation is simply a linear transformation of the other:

$$\frac{e_t}{b} = X_t + \frac{a}{b} - \frac{Y_t}{b}.$$

Thus, knowing one cointegrating equation implies the other; they are not independent.
Moreover, if each shock in the system follows a unit root process, then the variables cannot be cointegrated. In a two-variable system, we would need one variable (say X_t) to have a unit root, and the other (Y_t) to depend on it. If both are independent unit root processes, they will drift apart and will not be cointegrated.

7.4.4 Testing for the Number of Cointegrating Relationships

Johansen's framework allows for testing the number of cointegrating vectors (r) by examining the rank of the matrix Π. He proposed two likelihood ratio test statistics:

1. **The Maximum Eigenvalue Test** — tests the null hypothesis that the number of cointegrating vectors is r against the alternative of $r + 1$.

2. **The Trace Test** — tests the null hypothesis that the number of cointegrating vectors is at most r against the alternative that it is greater than r.

These are not merely two different statistics for the same hypothesis; they test different hypotheses and may lead to different empirical conclusions. In practice, this is quite common — the two tests can suggest different values of r. As always, it is good practice to choose the model specification that yields **economically meaningful** and interpretable results.

7.4.5 Summary

- The Engle–Granger method is limited to testing for one cointegrating relationship.
- The Johansen method generalizes this to multiple variables and multiple cointegrating vectors.
- The number of cointegrating relationships corresponds to the rank of Π.
- The Johansen approach provides two main tests: the trace test and the maximum eigenvalue test.
- These tests may lead to different conclusions, so researchers must rely on economic theory and model diagnostics to select the appropriate model.

7.5 Example - GDP and Consumption

$$\Delta \mathbf{y}_t = \alpha\, \beta'\, \mathbf{y}_{t-1} + \mathbf{c} + \varepsilon_t, \qquad \mathbf{y}_t = \begin{bmatrix} \log GDP_t \\ \log PCE_t \end{bmatrix}, \qquad \Delta \mathbf{y}_t = \begin{bmatrix} \Delta \log GDP_t \\ \Delta \log PCE_t \end{bmatrix}.$$

$$\underbrace{EC_{t-1}}_{\text{error–correction term}} = \beta' \mathbf{y}_{t-1} = \log GDP_{t-1} \; - \; 0.9636425 \; \log PCE_{t-1} \; - \; 70.06554.$$

$$\Delta \log GDP_t = \; \alpha_{\text{GDP}}\, EC_{t-1} \; + \; \mu_{\text{GDP}} \; + \; \varepsilon_{1t},$$

with the estimated coefficients:

$$\boxed{\Delta \log GDP_t = -8.160313 \; EC_{t-1} \; - \; 0.0173585 \; + \; \varepsilon_{1t}}$$

$$\Delta \log PCE_t = \; \alpha_{\text{PCE}}\, EC_{t-1} \; + \; \mu_{\text{PCE}} \; + \; \varepsilon_{2t},$$

with the estimated coefficients:

$$\boxed{\Delta \log PCE_t = -8.113909\ EC_{t-1}\ +\ 0.0174577\ +\ \varepsilon_{2t}}$$

Cointegration

The Johansen trace test indicates the presence of one long-run equilibrium relationship between real GDP and real personal consumption expenditures (PCE). The estimated cointegrating vector, normalized on $\log GDP$, is:

$$\log GDP_t - 0.964 \log PCE_t - 70.07 = 0.$$

The coefficient on $\log PCE$ is extremely precise, implying a very stable long-run proportionality between GDP and consumption. Since consumption is the largest component of GDP, an elasticity close to one is economically plausible.

Adjustment Toward Equilibrium

Both variables exhibit strong and statistically significant adjustment to disequilibrium:

$$\alpha_{\text{GDP}} = -8.16 \quad (p = 0.000), \qquad \alpha_{\text{PCE}} = -8.11 \quad (p = 0.000).$$

The negative signs indicate that shocks which push GDP and PCE away from their long-run equilibrium are corrected rapidly. If GDP is high relative to PCE, both GDP and PCE decline in the next period; if GDP is low relative to PCE, both rise. The magnitudes suggest extremely fast adjustment toward equilibrium.

Short-Run Dynamics

With only one lag in the VECM, the short-run dynamics are dominated by the error-correction mechanism. The own-lagged differences do not appear in the model, implying that nearly all short-run movements in GDP and PCE arise from the need to restore the long-run equilibrium. The R^2 values around 0.14 indicate modest explanatory power for short-run fluctuations.

Economic Meaning

In the short run, GDP and PCE respond primarily to deviations from their long-run relationship. Divergences between the two series are corrected quickly. In the long run, GDP and PCE move in near-proportional fashion, consistent with the national accounts identity in which consumption is the dominant component of GDP. Both variables are endogenous in the sense that they share the burden of restoring equilibrium.

7.6 Example - Quantity Theory of Money

$$\Delta \mathbf{y}_t = \alpha \beta' \mathbf{y}_{t-1} + \Gamma_1 \Delta \mathbf{y}_{t-1} + \mathbf{c} + \varepsilon_t,$$

where

$$\mathbf{y}_t = \begin{bmatrix} \ln M2_t \\ \ln CPI_t \\ \ln GDP_t \end{bmatrix}.$$

Cointegrating Relation

Johansen normalization (normalized on $\ln M2$) gives the cointegrating vector

$$\beta' \mathbf{y}_t = \ln M2_t - 0.973 \ln CPI_t - 1.277 \ln GDP_t + 9.706.$$

Thus the error–correction term is

$$EC_{t-1} = \ln M2_{t-1} - 0.973 \ln CPI_{t-1} - 1.277 \ln GDP_{t-1} + 9.706.$$

Error–Correction Model

Money Equation:

$$\Delta \ln M2_t = 0.00953\, EC_{t-1} + 0.00366.$$

Inflation Equation:

$$\Delta \ln CPI_t = 0.00772\, EC_{t-1} - 0.00121.$$

GDP Equation:

$$\Delta \ln GDP_t = 0.00796\, EC_{t-1} - 0.00320.$$

(Only the error–correction coefficients are shown because of one lag)

Cointegration Interpretation

The trace test indicates one cointegrating vector linking money, prices, and output. The estimated long-run relation is

$$\ln M2_t = 0.973 \ln CPI_t + 1.277 \ln GDP_t - 9.706.$$

Higher price levels or higher real output require higher money balances to maintain long-run equilibrium. Departures from this equilibrium are captured by the error–correction term.

Adjustment Toward Long-Run Equilibrium

The speed-of-adjustment coefficients are:

$$\alpha_{\text{M2}} = 0.00953 \qquad (p = 0.001)$$

$$\alpha_{\text{CPI}} = 0.00772 \qquad (p = 0.000)$$

$$\alpha_{\text{GDP}} = 0.00796 \qquad (p = 0.003)$$

All three adjustment coefficients are positive and statistically significant.

Interpretation:
All variables respond to disequilibrium in the long-run money–price–output relationship:

- When money is "too high" relative to CPI and GDP, all three variables adjust upward in the following period.
- The system restores equilibrium jointly; no variable is weakly exogenous.
- Adjustment speeds are small (0.7–1.0%), meaning equilibrium correction is gradual.

Economic Meaning

- The long-run relationship is consistent with a money-demand interpretation: higher prices and higher output require higher money balances.
- All three variables adjust significantly to the long-run error–correction term.
- Money, inflation, and output are jointly endogenous and influence each other in restoring long-run balance.