Time Series Fundamentals

TA Session 1: Introduction to Stochastic Processes

Rafael Lincoln

Econometrics II – NYU PhD Program

Spring 2026

Today's Roadmap

1. **Introduction: Why Time Series?**
   - Objectives: Dynamic relationships & Forecasting
   - Structural vs. Reduced-form models
   - Stylized facts
2. **Introduction to Stochastic Processes**
   - Formal definitions: Filtrations, measurability
   - Models: DGP, correct specification
   - Key restrictions: Stationarity
   - Linear operators (lag & difference)
3. **Examples from Macroeconomics**
   - Monetary policy transmission
   - Samuelson's multiplier-accelerator
   - Euler equations

Part I: Introduction

Why Time Series?

What is a Time Series?

Definition 1

A **time series** is a sequence of data points indexed in time order, taken at successive equally spaced points in time.

How does it differ from other data structures?

- **Cross-sectional data**: No natural ordering of observations
- **Spatial data**: Ordering based on geographic location
- **Time series**: Natural temporal ordering with dependence structure

Key Challenge

We observe only *one realization* of history. Inference requires assumptions ensuring "the present is like the past."

Why Should We Care About Time Series?

Two Main Objectives:

1. **Dynamic Relationships** (Correlation and/or Causality)
   - *Causality*: Impact evaluation, policy analysis, price optimization
   - *Correlation*: Comovement between variables (balanced growth, asset pricing)

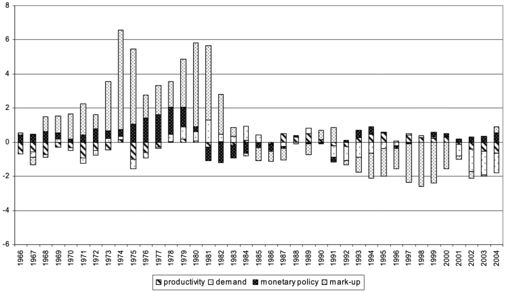
2. **Forecasting** future values based on scenarios
   - Macroeconomic forecasting (GDP, inflation, unemployment)
   - Asset allocation and risk quantification
   - *Nowcasting*: High-frequency monitoring of low-frequency variables

Example: Generate HD with structural model

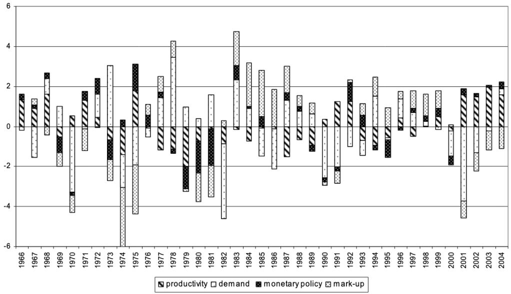
Smets & Wouters (2007)

Structural model ⇒ identified shocks ⇒ historical decomposition of inflation and output growth over time.

Annual inflation (GDP deflator - deviation from mean)



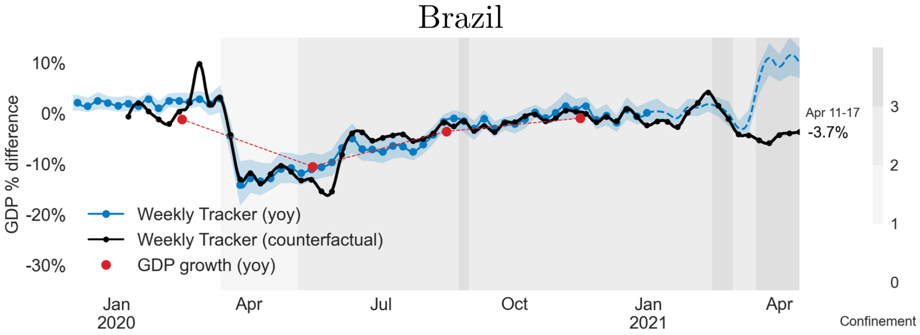
Annual per capita GDP growth (deviation from trend growth)



Example: Weekly Nowcasting

OECD Weekly Economic Tracker

A high-frequency "weekly tracker" uses weekly indicators to nowcast quarterly GDP growth in real time.



Example: Estimating a Phillips Curve

Consider the reduced-form Phillips curve:

$$\pi_t = \beta_0 + \beta_1 y_t + \beta_2 \pi_{t-1} + u_t$$

where π_t is inflation and y_t is the output gap.

Questions of Interest:

- What is the sign and size of β_1? (Inflation-output tradeoff)
- Is $\beta_2 \approx 1$? (Testing *adaptive expectations*: $\mathbb{E}[\pi_t \mid \mathcal{F}_{t-1}] = \pi_{t-1}$)

Key Insight

The presence of π_{t-1} creates *serial dependence* — standard cross-sectional methods don't apply directly!

Two Types of Models

Structural Models

- Behavioral functions with economic interpretation
- Causal relationships between variables
- Parameters have economic meaning
- Shocks are "structural" (e.g., monetary, fiscal)

Reduced-Form Models

- Correlations between variables
- Descriptive analysis
- No direct economic interpretation
- Useful for *forecasting*

Trade-off

Structural models answer "why?" but require stronger assumptions.
Reduced-form models answer "what?" and are useful for prediction.

Univariate vs. Multivariate Models

Univariate: Series explained by its own past only

$$\pi_t = \beta_0 + \beta_2 \pi_{t-1} + u_t$$

Multivariate: Other variables also play a role

$$\begin{pmatrix} 1 & -\beta_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_t \\ y_t \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \alpha_0 \end{pmatrix} + \begin{pmatrix} \beta_2 & 0 \\ \alpha_1 & \alpha_2 \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

- Adding the output gap equation transforms it into a **Vector Autoregression (VAR)**
- Choice depends on the research question and data availability

Stylized Facts of Time Series

Time series data typically exhibit:

1. **Trend**: Long-term movement (growth, decline)
2. **Seasonality**: Periodic and regular fluctuations (quarterly, monthly)
3. **Cycles**: Medium-run fluctuations (business cycles) that are not strictly periodic (unlike seasonality)
4. **Irregularity**: Erratic fluctuations (noise)

Additional Features:

- **Heteroskedasticity**: Time-varying volatility (ARCH/GARCH effects)
- **Outliers**: Extreme observations
- **Structural breaks**: Changes in parameters over time

Implication

Models must account for these features to properly represent the DGP.

Part II (a): Stochastic Processes

Formal Definitions

Why Formal Definitions?

Goal: Provide rigorous foundations for time series analysis

Key Questions:

- What is a stochastic process mathematically?
- How do we relate the *observed* time series to the *underlying* process?
- Under what conditions can we learn from a single realization?

Building Blocks:

1. Probability spaces and measurability
2. Filtrations (information evolving over time)
3. Stochastic processes and their realizations

Measurable Functions

Definition 2 (Measurable Function)

A function $f : \Omega \to \widetilde{\Omega}$ between measure spaces $(\Omega, \mathcal{F})$ and $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ is $\mathcal{F}/\widetilde{\mathcal{F}}$-**measurable** if it preserves measurability:

$$\forall B \in \widetilde{\mathcal{F}}, \quad f^{-1}(B) \in \mathcal{F}$$

where $f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$.

Intuition: A measurable function “respects” the information structure. We can compute probabilities of events defined through f.

Why it matters: Random variables must be measurable functions for probabilities to be well-defined.

Filtered Probability Space

Definition 3 (Filtered Probability Space)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A **filtration** is an increasing sequence of σ-algebras:

$$\mathbb{F} := \{\mathcal{F}_t\}_{t \geq 0} \quad \text{with} \quad \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$$

The quadruple $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a **filtered probability space**.

Intuition:

- $\mathcal{F}_t$ represents information available at time t
- The sequence is *increasing*: questions answerable at time t remain answerable at $t+1$
- Information grows over time (we don't forget!)

Measurability of Stochastic Processes

Definition 4 (Adapted & Predictable Processes)

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space. A stochastic process $\{X_t\}_{t\geq 0}$ is:

- **$\mathbb{F}$-adapted** if X_t is $\mathcal{F}_t$-measurable for every t
- **$\mathbb{F}$-predictable** if X_t is $\mathcal{F}_{t-1}$-measurable for every t

Interpretation:

- **Adapted**: X_t is known given information up to and including time t (past *or* present)
- **Predictable**: X_t is known given information strictly before time t (past only)

Example

Stock price S_t is adapted (known at t), but not predictable (unknown at $t-1$).

The Natural Filtration

Definition 5 (X-adapted Filtration)

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\{X_t\}_{t\geq 0}$ a stochastic process. The **natural filtration** generated by X is:

$$\mathcal{F}_t^X = \sigma(X_0, X_1, \ldots, X_t)$$

The process X_t is $\mathcal{F}_t^X$-adapted by construction.

Intuition: $\mathcal{F}_t^X$ is the information we can extract from observing the process up to time t.

Common notation: $\mathcal{F}_{t-1}^X = \{X_{t-1}, X_{t-2}, \ldots\}$ is the "information set" at $t-1$.

Stochastic Process: Definition

Definition 6 (Stochastic Process)

A **stochastic process** $\{X_t\}_{t\geq 0}$ on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, which is $\mathbb{F}$-adapted, is an ordered sequence of random vectors:

$$\{X_t\}_{t\geq 0} = \{\mathbf{x}_t(\omega) : \omega \in \Omega,\ t \in \mathcal{T}\}$$

where $\mathbf{x}_t(\omega) \in \mathbb{R}^n$ for all $t \in \mathcal{T}$.

Two perspectives:

1. **Fix** t: $\mathbf{x}_t(\cdot)$ is a random variable mapping $\Omega \to \mathbb{R}^n$
2. **Fix** ω: $\mathbf{x}_\cdot(\omega)$ is a deterministic function of time — a **path** (or **realization**)

Time Series and Ensemble

Definition 7 (Time Series)

A **time series** $\{\mathbf{x}_t\}_{t=1}^{T}$ is a particular *path* (realization) of the stochastic process $\{X_t\}_{t\geq 0}$, which is its **generating mechanism**.

Definition 8 (Ensemble)

The set of all possible realizations of a stochastic process:

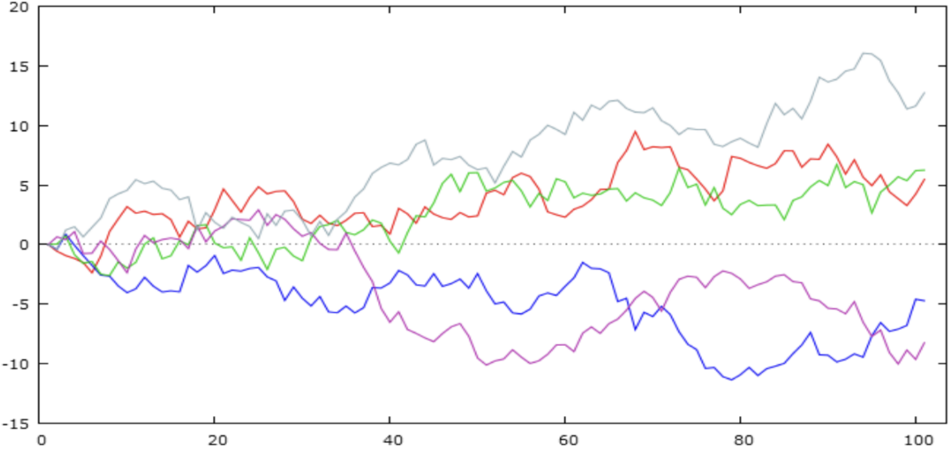
$$\{\{\mathbf{x}_t(\omega) : t \in \mathcal{T}\} : \omega \in \Omega\}$$

is called the **ensemble**.

Key Insight

In macroeconometrics, we observe **one** realization from the ensemble. We cannot observe alternative histories!

Time Series as a Realization of a Stochastic Process



The five time series in the figure are realizations of the same stochastic process!

Part II (b): Models for Stochastic Processes

DGP, Stationarity, and Structural/Reduced-Form Models

Data-Generating Process (DGP)

Definition 9 (Data-Generating Process)

The **DGP** of $\{X_t\}_{t\geq 0}$ is the probabilistic rule that generates the process over time. A convenient characterization is given by:

$$F(X_0) \quad \text{and} \quad \{\, F(X_t \mid \mathcal{F}^X_{t-1})\,\}_{t\geq 1},$$

i.e., the initial distribution and the collection of one-step-ahead conditional distributions given the information set.

Problem: This is an *infinite-dimensional* object!

- Analytically intractable
- Cannot estimate infinitely many parameters

Question: Can we represent it by something finite-dimensional?

Finite-Dimensional Distributions

Definition 10 (Finite-Dimensional Distribution)

Let $\{t_1, t_2, \ldots, t_s\} \subset \mathcal{T}$ be a finite set. The joint distribution:

$$F_{t_1,\ldots,t_s}(b_1, \ldots, b_s) = \mathbb{P}(\mathbf{x}_{t_1} \leq b_1, \ldots, \mathbf{x}_{t_s} \leq b_s)$$

The family of all such distributions is the **finite-dimensional distribution** of the process.

Theorem 11

Under general conditions, the probabilistic structure of $\{X_t\}_{t\geq 0}$ is ***completely specified*** *by its finite-dimensional distributions for all s and all choices of $\{t_1, \ldots, t_s\}$.*

Implication: We can work with finite-dimensional objects!

Time Series Model

Definition 12 (Time Series Model)

An econometric time series model is a family of functions:

$$\{\mathcal{M}(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, d_t; \psi) : \psi \in \Psi \subseteq \mathbb{R}^p\}$$

that aims to represent the true DGP, where:

- ψ is a p-dimensional parameter vector
- d_t represents possible structural breaks

Goal: Estimate $\hat{\psi}$ based on assumptions about $\mathcal{M}$.

Two flavors:

- Model for the *distribution*: $\mathcal{M}_D$
- Model for the *conditional expectation*: $\mathcal{M}_E$ (most common)

Correct Specification Axioms

Axiom: Correct Specification for Distribution

$$\mathcal{M}_D(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, d_t; \psi) = F(X_t \mid \mathcal{F}_{t-1}^X)$$

Axiom: Correct Specification for Expectation

$$\mathcal{M}_E(\mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, d_t; \psi) = \mathbb{E}[\mathbf{x}_t|\mathcal{F}_{t-1}^X]$$

In practice: We usually model only the conditional expectation

- Similar to CEF approach in cross-sectional econometrics
- Reduces dimensionality dramatically

Structural Model

Definition 13 (Structural Model)

Given data $\mathbf{z} = (\mathbf{y}, \mathbf{x})$ from the joint process $\{Y_t, X_t\}_{t \geq 0}$, a **structural model** for the conditional expectation is:

$$g(\mathbf{y}, \mathbf{x}, \mathbf{d}, \mathbf{u}; \psi_0) = 0$$

where $\mathbf{u}$ are **structural shocks** and $\mathbf{d}$ are structural breaks.

Key Features:

- Shocks $\mathbf{u}$ have *economic interpretation* (monetary, fiscal, technology, ...)
- Parameters ψ_0 have *economic meaning* (elasticities, policy parameters, ...)
- Implies *causal* relationships

Reduced-Form Model

Definition 14 (Reduced-Form Model)

If $\{Y_t\}_{t\geq 0}$ has a unique solution as a function of past values and exogenous variables:

$$\mathbf{y}_t = h(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots, \mathbf{x}_t, \mathbf{v}_t, \mathbf{d}_t; \theta_0)$$

where $\theta_0 = m(\psi_0)$ and $\mathbf{v}_t = f(\mathbf{u}_t)$ are **reduced-form** parameters and shocks.

Key Features:

- Parameters and shocks *lose* economic interpretation
- Useful for **forecasting**: $\mathbb{E}[\mathbf{y}_t | \mathbf{x}_t, \mathcal{F}_{t-1}]$ is the best MSE predictor
- Easier to estimate (no simultaneity issues)

Trade-off

Reduced form: good for prediction, bad for policy analysis (Lucas critique!)

The Incidental Parameter Problem

Without restrictions: Suppose $Y_t \sim \mathcal{N}(\mu_t, \Gamma_{ts})$

We need to estimate:

$$\mathbb{E}(Y^T) = \begin{bmatrix} \mu_1' & \mu_2' & \cdots & \mu_T' \end{bmatrix}'$$

$$\text{Var}(Y^T) = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \cdots & \Gamma_{1T} \\ \Gamma_{21} & \Gamma_{22} & \cdots & \Gamma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{T1} & \Gamma_{T2} & \cdots & \Gamma_{TT} \end{bmatrix}$$

Problem

Number of parameters grows with $T \Rightarrow$ LLN and CLT fail!

Solution: Restrictions on the Process

Two types of restrictions:

1. **Time-homogeneity** (Stationarity)
   - All observations come from distributions with time-invariant features
   - Reduces: $\mu_t \to \mu$, $\Gamma_{ts} \to \Gamma_{|t-s|}$

2. **Memory restrictions** (Ergodicity/Mixing)
   - Each new observation contains "new information"
   - Enables consistent estimation from a single realization

With stationarity:

$$\mathbb{E}(Y^T) = \begin{bmatrix} \mu' & \mu' & \cdots & \mu' \end{bmatrix}', \quad \mathrm{Var}(Y^T) = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{T-1} \\ \Gamma_{-1} & \Gamma_0 & \cdots & \Gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix}$$

Weak Stationarity

Definition 15 (Weak Stationarity)

A stochastic process $\{Y_t\}$ is **weakly stationary** (covariance stationary, second-order stationary) if:

1. $\mathbb{E}[Y_t] = \mu$ for all $t \in \mathcal{T}$, where $\|\mu\| < \infty$
2. $\mathbb{E}[(Y_t - \mu)(Y_{t-h} - \mu)'] = \Gamma_h$ for all t, h, where $\|\Gamma_h\| < \infty$

Key Properties:

- Mean is constant over time
- Autocovariance depends only on the *lag* h, not on t
- Variance Γ_0 is constant

Lemma 16

For weakly stationary processes: $\Gamma_h = \Gamma'_{-h}$

Example: Autocovariance Matrices

For a bivariate process $Y_t = (Y_{1t}, Y_{2t})'$:

$$\Gamma_0 = \text{Var}(Y_t) = \begin{bmatrix} \text{Var}(Y_{1t}) & \text{Cov}(Y_{1t}, Y_{2t}) \\ \text{Cov}(Y_{1t}, Y_{2t}) & \text{Var}(Y_{2t}) \end{bmatrix}$$

$$\Gamma_1 = \text{Cov}(Y_t, Y_{t-1}) = \begin{bmatrix} \text{Cov}(Y_{1t}, Y_{1,t-1}) & \text{Cov}(Y_{1t}, Y_{2,t-1}) \\ \text{Cov}(Y_{2t}, Y_{1,t-1}) & \text{Cov}(Y_{2t}, Y_{2,t-1}) \end{bmatrix}$$

Interpretation:

- Diagonal: *auto*covariances of each component
- Off-diagonal: *cross*-covariances between components

Strong Stationarity

Definition 17 (Strong Stationarity)

A stochastic process $\{Y_t\}$ is **strongly stationary** (strictly stationary) if all finite-dimensional distributions are invariant to time shifts: for any $n \in \mathbb{N}$, any times $t_1, \ldots, t_n$, any shift $h \in \mathbb{Z}$, and any $y_1, \ldots, y_n$,

$$F_{t_1,\ldots,t_n}(y_1, \ldots, y_n) = F_{t_1+h,\ldots,t_n+h}(y_1, \ldots, y_n).$$

Relationship to Weak Stationarity:

- Strong stationarity $\Rightarrow$ Weak stationarity *if moments exist*
- Weak stationarity $\not\Rightarrow$ Strong stationarity (in general)
- For **Gaussian** processes: Weak $\Leftrightarrow$ Strong

Cauchy Counterexample

A Cauchy-distributed process can be strongly stationary but *not* weakly stationary (moments undefined!).

Part II (c): Operations on Stochastic Processes

Linear Operators

Linear Operators

Definition 18 (Linear Operator)

Let X and Y be vector spaces over field F. The operator $T : X \to Y$ is **linear** if:

$$T(\alpha x_1 + \beta x_2) = \alpha T(x_1) + \beta T(x_2)$$

for all $x_1, x_2 \in X$ and $\alpha, \beta \in F$.

Context: Time series form a vector space. We can define useful linear operators on this space.

Key operators:

- Lag operator L
- Difference operator Δ

The Lag Operator

Definition 19 (Lag Operator)

For a stochastic process $\{\mathbf{z}_t\}$, the **lag operator** L is defined by:

$$L^j \mathbf{z}_t = \mathbf{z}_{t-j} \quad \forall j \in \mathbb{N}$$

Properties:

- L is a linear operator
- $L^0 = I$ (identity)
- Powers compose: $L^j L^k = L^{j+k}$

Proposition 1

For $|\alpha| < 1$*:*

$$(1 - \alpha L)^{-1} = 1 + \alpha L + \alpha^2 L^2 + \cdots = \sum_{j=0}^{\infty} \alpha^j L^j$$

The Difference Operator

Definition 20 (Difference Operator)

The **difference operator** Δ is defined by:

$$\begin{aligned}
\Delta \mathbf{z}_t &= (I - L)\mathbf{z}_t = \mathbf{z}_t - \mathbf{z}_{t-1} \\
\Delta^j \mathbf{z}_t &= (I - L)^j \mathbf{z}_t \quad \forall j \in \mathbb{N}_+ \\
\Delta_j \mathbf{z}_t &= (I - L^j)\mathbf{z}_t = \mathbf{z}_t - \mathbf{z}_{t-j}
\end{aligned}$$

Common uses:

- Δy_t: First difference (removes linear trend)
- $\Delta^2 y_t$: Second difference (removes quadratic trend)
- $\Delta_{12} y_t = y_t - y_{t-12}$: Seasonal difference (monthly data)

Key insight: Differencing can transform non-stationary series into stationary ones (unit root processes).

Lag Polynomials

Notation: We can write models compactly using **lag polynomials**:

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

Example: AR(p) model

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t$$

becomes

$$\Phi(L) y_t = u_t$$

Inversion: If roots of $\Phi(z) = 0$ are outside the unit circle:

$$y_t = \Phi(L)^{-1} u_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$$

This is the **MA(∞) representation**.

Part III: Examples

Applications to Macroeconomics

Why Examples Matter

Goal: Connect abstract definitions to concrete economic models

Key insight: Many structural macro models admit reduced-form time series representations

Examples we'll cover:

1. Monetary policy transmission (Phillips curve + Taylor rule)
2. Samuelson's multiplier-accelerator (1939)
3. Intertemporal optimization (Euler equations)

Each shows: **Structural model** $\rightarrow$ **Reduced form** (AR, VAR, ARMA)

Example 1: Monetary Policy Transmission

Structural model:

$$
\begin{aligned}
\pi_t &= \lambda y_t + \pi_t^e + u_{1t} && \text{(Phillips curve)} \\
y_t &= \gamma(i_{t-1} - \pi_t^e) + u_{2t} && \text{(IS curve)} \\
\pi_t^e &= \pi_{t-1} && \text{(Adaptive expectations)} \\
i_t &= i^* + \rho(\pi_t - \pi^*) && \text{(Taylor rule)}
\end{aligned}
$$

where:

- π_t: inflation, y_t: output gap, i_t: interest rate
- $\lambda \in (0, 1)$, $\gamma \in (-1, 0)$, $\rho \geq 0$
- $\mathbf{u}_t = (u_{1t}, u_{2t})' \sim \text{NID}(\mathbf{0}, \Omega)$

Monetary Policy: Reduced Form

Substituting the Taylor rule and expectations:

$$\pi_t = \lambda y_t + \pi_{t-1} + u_{1t}$$
$$y_t = \gamma(i^* - \rho\pi^*) + \gamma(\rho - 1)\pi_{t-1} + u_{2t}$$

In matrix form (VAR(1)):

$$\underbrace{\begin{bmatrix} 1 & -\lambda \\ 0 & 1 \end{bmatrix}}_{B} \begin{bmatrix} \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0 \\ \gamma(i^* - \rho\pi^*) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \gamma(\rho - 1) & 0 \end{bmatrix} \begin{bmatrix} \pi_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Since B is invertible, multiply by B^{-1}:

$$\mathbf{w}_t = \mathbf{c}_0 + C_1 \mathbf{w}_{t-1} + \mathbf{v}_t$$

where $\mathbf{v}_t = B^{-1}\mathbf{u}_t$ are *reduced-form* shocks.

Monetary Policy: Univariate Representation

Isolating inflation (since y_{t-1} doesn't appear in π_t equation):

$$\pi_t = \lambda\gamma(i^* - \rho\pi^*) + [1 + \lambda\gamma(\rho - 1)]\pi_{t-1} + u_{1t} + \lambda u_{2t}$$

This is an **AR(1)** for inflation:

$$\pi_t = \phi_0 + \phi_1 \pi_{t-1} + v_{1t}$$

Key observations:

- Same DGP, different information sets
- VAR uses (π_{t-1}, y_{t-1}); AR(1) uses only π_{t-1}
- Structural shocks $\mathbf{u}_t$ become reduced-form shocks $\mathbf{v}_t$

Example 2: Samuelson's Multiplier-Accelerator (1939)

Historical context: First mathematical model after Keynes' *General Theory*

Structural model:

$$
\begin{aligned}
y_t &= c_t + i_t + g && \text{(National income identity)} \\
c_t &= \alpha y_{t-1} + \varepsilon_t && \text{(Consumption function)} \\
i_t &= \beta(c_t - c_{t-1}) && \text{(Accelerator)}
\end{aligned}
$$

Reduced form: Substitute to get

$$y_t = g + \alpha(1+\beta)y_{t-1} - \beta\alpha y_{t-2} - \beta\varepsilon_{t-1} + (1+\beta)\varepsilon_t$$

This is an **ARMA(2,1)**:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta u_{t-1} + u_t$$

where $u_t = (1+\beta)\varepsilon_t$ and $\theta = -\beta/(1+\beta)$.

Example 3: Euler Equation from Optimization

Consumer's problem:

$$\max_{c_{t+i}, A_{t+i}} \mathbb{E}\left[\sum_{i=0}^{\infty}(1+\delta)^{-i}U(c_{t+i})\middle|\mathcal{F}_t\right]$$

subject to budget constraint and transversality condition.

With CRRA utility $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$:

Euler equation:

$$\mathbb{E}\left[\frac{1+r}{1+\delta}c_{t+1}^{-\gamma} - c_t^{-\gamma}\middle|\mathcal{F}_t\right] = 0$$

Parameters of interest:

- δ: Discount rate (patience)
- γ: Risk aversion (intertemporal substitution)

This is a *conditional moment restriction* — basis for GMM estimation!

Example 3: Structural to Reduced Form

- We can map the structural to a reduced form model under assumptions:
  1. **Assumption (1)**: $\delta = r$
  2. **Assumption (2)**: $y_t = \alpha y_{t-1} + \varepsilon_t$
- Under Assumption (1) and (2),

$$c_t = r(1+r)A_{t-1} + r\frac{1+r}{1+r-\alpha}y_t + \varepsilon_t$$

  where A_{t-1} is beginning of period assets, and ε_t is an autocorrelated error
- The above equation is a reduced form model for consumption

Common Patterns

What we've seen:

Example	**Structural**	**Reduced Form**
Monetary policy	4 equations	VAR(1) or AR(1)
Samuelson	3 equations	ARMA(2,1)
Euler equation	FOC	Moment restriction

Key insight: Economic models often have *autoregressive* reduced forms

- AR/MA/ARMA capture stylized facts: persistence, mean reversion, cycles
- VARs capture multivariate dynamics
- Euler equations lead to GMM/rational expectations estimation

Summary

Today we covered:

1. **Introduction**
   - Time series: dynamic relationships + forecasting
   - Structural vs. reduced-form models
   - Stylized facts
2. **Stochastic Processes**
   - Filtrations, measurability, DGP
   - Stationarity (weak, strong) — reduces parameters
   - Lag and difference operators
3. **Examples**
   - Monetary transmission $\rightarrow$ VAR
   - Samuelson $\rightarrow$ ARMA
   - Euler equation $\rightarrow$ GMM

Appendix: Ergodicity & Mixing, Martingales

Preview: What's Next

Section 7.4: Reduced-Form Models in Detail

- AR(p): Definition, stationarity, estimation
- MA(q): Definition, invertibility
- ARMA(p, q): Combining both

Section 7.5: Vector Autoregressions

- VAR(p) estimation and inference
- Impulse response functions
- Structural identification

Section 7.6: Estimation Methods

- OLS for AR models
- Maximum likelihood
- GMM for Euler equations

Appendix

Appendix: Ergodicity & Mixing

Section 7.2.2 (cont.)

α-Mixing Coefficient

Definition 21 (α-Mixing Coefficient)

Let $\mathcal{G}$ and $\mathcal{H}$ be σ-subfields of $\mathcal{F}$. The α-**mixing coefficient** is:

$$\alpha(\mathcal{G}, \mathcal{H}) = \sup_{G \in \mathcal{G}, H \in \mathcal{H}} |\mathbb{P}(G \cap H) - \mathbb{P}(G)\mathbb{P}(H)|$$

Intuition: Measures how "far from independent" two σ-algebras are.

- $\alpha = 0$: $\mathcal{G}$ and $\mathcal{H}$ are independent
- $\alpha > 0$: Some dependence exists

In time series context:

$$\mathcal{G} = \mathcal{F}_{-\infty}^{0} = \sigma(\ldots, X_{-1}, X_0), \quad \mathcal{H} = \mathcal{F}_m^{+\infty} = \sigma(X_m, X_{m+1}, \ldots)$$

$\mathcal{G}$ = "past", $\mathcal{H}$ = "distant future"

α-Mixing Condition

Definition 22 (α-Mixing Condition)

A sequence $\{X_t\}_{t=-\infty}^{\infty}$ is α-**mixing** (strong mixing) if:

$$\lim_{m\to\infty} \alpha_m = 0$$

where

$$\alpha_m = \sup_t \alpha(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+m}^{\infty})$$

Intuition: As the gap m increases, past and future become *asymptotically independent*.

Corollary 23 (For Gaussian Random Variables)

A useful mixing condition is ***absolute summability****:*

Transformations and Invariant Events

Definition 24 (Transformation)

A **transformation** $T : \Omega \to \Omega$ is a 1-1 measurable mapping of outcomes.

Definition 25 (Measure-Preserving Transformation)

T is **measure-preserving** if $\mathbb{P}(TE) = \mathbb{P}(E)$ for all $E \in \mathcal{F}$.

Example: Shift Transformation
The **back-shift operator** T: $X_t(T\omega) = X_{t+1}(\omega)$

$$\begin{aligned} \omega &= (\ldots, x_1, x_2, x_3, \ldots) \\ T\omega &= (\ldots, x_2, x_3, x_4, \ldots) \\ T^2\omega &= (\ldots, x_3, x_4, x_5, \ldots) \end{aligned}$$

Under strict stationarity, the shift transformation is measure-preserving.

Ergodicity

Definition 26 (Ergodicity)

A strictly stationary sequence $\{X_t(\omega)\}$ is **ergodic** if the probability of every invariant event is either 0 or 1.

Invariant event: E such that $TE = E$ (unchanged by time shift).

Intuition: If an invariant event has probability strictly between 0 and 1, there exist "parallel universes" that never communicate with each other.

Theorem 27 (Ergodic Theorem)

Let $\{X_t\}$ be weakly stationary and ergodic with $\mathbb{E}[|X_t|] < \infty$. Then:

$$\frac{1}{T}\sum_{t=1}^{T} X_t(\omega) \xrightarrow{a.s.} \mathbb{E}[X_t]$$

Example: Ergodicity Failure (I)

Example: Simple two-sequence universe

Let $\Omega = \{\omega_1, \omega_2\}$ where:

$$\omega_1 = (\ldots, 1, 1, 1, \ldots), \quad \omega_2 = (\ldots, 0, 0, 0, \ldots)$$

with $\mathbb{P}(\omega_1) = p \in (0, 1)$ and $\mathbb{P}(\omega_2) = 1 - p$.

Define $X_t(\omega_1) = 1$ and $X_t(\omega_2) = 0$ for all t.

Analysis:

- Both ω_1 and ω_2 are invariant under the shift T
- $\mathbb{P}(\omega_1) = p \in (0, 1)$ — not 0 or 1!
- $\Rightarrow$ The process is **not ergodic**

If we happen to draw ω_1, we see only 1's forever and learn nothing about ω_2.

Example: Ergodicity Failure (II)

Example: Random intercept

$$X_t = U_t + Z, \quad t \in \mathbb{Z}$$

where $\{U_t\} \sim$ i.i.d. Uniform$[0,1]$ and $Z \sim N(0,1)$, independent of U_t.

What happens?

$$\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = \bar{U}_n + Z \xrightarrow{a.s.} \frac{1}{2} + Z$$

The sample mean converges to a *random variable*, not a constant!

Why? The autocovariance:

$$\mathrm{Cov}(X_t, X_{t+h}) = \mathrm{Var}(Z) = 1 \quad \forall h \neq 0$$

Dependence is *too strong* — it never dies out!

Lesson

Without ergodicity, time averages $\neq$ population averages.

Appendix: Martingales

Section 7.2.4

Doob-Dynkin Lemma

Lemma 28 (Doob-Dynkin)

Let $X : \Omega \to \mathbb{R}^n$ and $Y : \Omega \to \mathbb{R}^m$ be random vectors. Then Y is $\mathcal{F}^X$-measurable if and only if there exists a measurable function $g : \mathbb{R}^n \to \mathbb{R}^m$ such that:

$$Y = g \circ X$$

Corollary 29

We always have:

$$\mathbb{E}[Y|\mathcal{F}^X] = \mathbb{E}[Y|X] = g \circ X$$

for some measurable g.

Intuition: Conditional expectations are functions of the conditioning variables.

Martingale Process

Definition 30 (Martingale)

A sequence $\{Y_t\}_{t\in\mathbb{N}}$ is a **martingale** w.r.t. $\{\mathcal{F}_t\}$ if:

1. $\mathbb{E}[\|Y_t\|] < \infty$ for all t
2. Y_t is $\mathcal{F}_t$-measurable
3. $\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = Y_{t-1}$ for all t

Intuition: The best forecast of tomorrow's value is today's value.

Examples:

- Random walk: $S_t = S_{t-1} + \varepsilon_t$ with $\mathbb{E}[\varepsilon_t|\mathcal{F}_{t-1}] = 0$
- Stock prices (under risk-neutral measure)
- Cumulative sum of fair bets

Martingale Difference Sequence

Definition 31 (Martingale Difference Sequence (MDS))

A sequence $\{Y_t\}_{t\in\mathbb{N}}$ is a **martingale difference sequence** w.r.t. $\{\mathcal{F}_t\}$ if:

1. $\mathbb{E}[\|Y_t\|] < \infty$ for all t
2. Y_t is $\mathcal{F}_t$-measurable
3. $\mathbb{E}[Y_t|\mathcal{F}_{t-1}] = 0$ for all t

Key property: MDS are **uncorrelated** over time!

Corollary 32

If $\{Y_t\}$ *is MDS, then* $\mathbb{E}[Y_tY_{t-j}] = 0$ *for all* $j \neq 0$.

Proof: $\mathbb{E}[Y_tY_{t-j}] = \mathbb{E}[\mathbb{E}[Y_t|\mathcal{F}_{t-1}]Y_{t-j}] = \mathbb{E}[0 \cdot Y_{t-j}] = 0$

Note: MDS is *weaker* than i.i.d. — allows for conditional heteroskedasticity!

MDS Properties

Theorem 33

Let $\{Y_t\}$ be MDS w.r.t. $\{\mathcal{F}_t\}$, and let $g_{t-1} = g(Y_{t-1}, Y_{t-2}, \dots)$ be measurable and integrable. Then:

1. *$\{Y_t g_{t-1}\}$ is also MDS*
2. *Y_t and g_{t-1} are uncorrelated*

Proof sketch:

$$\mathbb{E}[Y_t g_{t-1} | \mathcal{F}_{t-1}] = g_{t-1}\mathbb{E}[Y_t | \mathcal{F}_{t-1}] = g_{t-1} \cdot 0 = 0$$

Implication: Products of MDS with past functions remain MDS.
This is crucial for:

- Deriving variance formulas
- Establishing CLTs for time series
- Analyzing ARCH/GARCH models

Proof: Autocovariance Symmetry

Lemma 34

For a weakly stationary process: $\Gamma_k = \Gamma'_{-k}$

Proof.

Since the process is weakly stationary:

$$\mathbb{E}[Y_t] = \mathbb{E}[Y_{t-k}] = \mu \quad \forall k$$

Therefore:

$$\begin{aligned}\Gamma_k &= \mathbb{E}[(Y_t - \mu)(Y_{t-k} - \mu)'] \\ &= \mathbb{E}[(Y_{t-k} - \mu)(Y_t - \mu)']' \\ &= \Gamma'_{-k}\end{aligned}$$

□

Detailed Derivation: Monetary Transmission VAR

Starting from:

$$B\mathbf{w}_t = \mathbf{a}_0 + A_1\mathbf{w}_{t-1} + \mathbf{u}_t$$

With $B = \begin{bmatrix} 1 & -\lambda \\ 0 & 1 \end{bmatrix}$, we have $B^{-1} = \begin{bmatrix} 1 & \lambda \\ 0 & 1 \end{bmatrix}$

Multiplying through:

$$\mathbf{w}_t = B^{-1}\mathbf{a}_0 + B^{-1}A_1\mathbf{w}_{t-1} + B^{-1}\mathbf{u}_t$$

Reduced-form parameters:

$$\mathbf{c}_0 = \begin{bmatrix} \lambda\gamma(i^* - \rho\pi^*) \\ \gamma(i^* - \rho\pi^*) \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 1 + \lambda\gamma(\rho - 1) & 0 \\ \gamma(\rho - 1) & 0 \end{bmatrix}$$

$$\mathbf{v}_t = \begin{bmatrix} u_{1t} + \lambda u_{2t} \\ u_{2t} \end{bmatrix}$$

Additional Mixing Conditions

Beyond α-mixing, other conditions exist:

β**-mixing** (absolute regularity):

$$\beta_m = \sup_t \mathbb{E}\left[\sup_{H \in \mathcal{F}_{t+m}^{\infty}} |\mathbb{P}(H|\mathcal{F}_{-\infty}^{t}) - \mathbb{P}(H)|\right]$$

ρ**-mixing** (maximal correlation):

$$\rho_m = \sup_t \sup_{\substack{f \in L^2(\mathcal{F}_{-\infty}^{t}) \\ g \in L^2(\mathcal{F}_{t+m}^{\infty})}} |\mathrm{Corr}(f, g)|$$

Hierarchy:

$$\text{i.i.d.} \Rightarrow \beta\text{-mixing} \Rightarrow \alpha\text{-mixing}$$

$$\beta\text{-mixing} \Rightarrow \rho\text{-mixing}$$

Each has different implications for CLTs and moment inequalities.

References

Textbooks:

- Hamilton (1994): *Time Series Analysis*
- Brockwell & Davis (1991): *Time Series: Theory and Methods*
- Lütkepohl (2005): *New Introduction to Multiple Time Series Analysis*

Course Materials:

- Lecture Notes: Sections 7.1–7.3
- Cochrane (2005): *Time Series for Macroeconomics and Finance*

Classic Papers:

- Samuelson (1939): "Interactions between the multiplier analysis and the principle of acceleration"
- Sims (1980): "Macroeconomics and Reality"