

# Lecture 21 — Final Revision of Advanced Time Series Methods

Whole-course synthesis: main ideas, core formulas, major theorems, empirical findings, and exam logic

Jiajing Sun

School of Economics and Management, University of Chinese Academy of Sciences

Econometrics and Time Series Methods  
Spring 2026

# What this final revision does

This final session is a **whole-course revision**, not a new topic.

- 1 It reviews the main content from univariate foundations to high-frequency econometrics.
- 2 It collects the core formulas and theoretical results that should be reproducible.
- 3 It highlights the main empirical findings and interpretation rules.
- 4 It shows how to turn the course into clear exam answers and research workflows.

## Target outcome

By the end, you should see the course as one integrated toolkit for dependent data.

# The course in one sentence

time dependence  $\implies$  models  $\implies$  estimation  
 $\implies$  valid inference  $\implies$  economic interpretation.

- **Representation:** how do we describe dependence?
- **Prediction:** how do we use past information?
- **Inference:** how do we account for serial dependence?
- **Signal extraction:** how do we separate trend, cycle, noise, and latent states?
- **Continuous time:** how do we handle financial data observed at very high frequency?

# Three-hour revision plan

## Hour 1: foundations

Stationarity, dependence, Wold representation, ARMA, forecasting, nonstationarity, unit roots, ADF, KPSS, trends, and random walks.

## Hour 2: systems, volatility, and flexible modeling

VAR, Granger causality, impulse responses, SVAR, cointegration, VECM, ARCH/GARCH, multivariate volatility, and nonparametric methods.

## Hour 3: robust inference, filtering, and continuous time

Spectrum, HAC, fixed- $b$ , self-normalization, bootstrap, GMM, deterministic filters, state-space models, Kalman filter, Brownian motion, diffusion models, realized volatility, and final exam logic.

# How to answer exam questions

A strong answer is usually built from five visible pieces:

model + formula + condition + interpretation + limitation

- 1 State the object: process, estimator, test, or model.
- 2 Write the relevant equation or statistic.
- 3 State the condition under which it is valid.
- 4 Explain the economic or statistical meaning.
- 5 Mention what the method does *not* prove.

# What is special about time-series data?

The central difference from cross-sectional data is **dependence over time**.

$$\text{Cov}(X_t, X_{t-h}) \neq 0 \quad \text{for some lags } h.$$

- Observations are ordered and past information matters.
- Sample averages can have larger or smaller variance than under independence.
- Forecasting, lag structure, and dynamic adjustment become central.
- Valid inference must handle serial correlation and heteroskedasticity.

## Main finding

Ignoring dependence usually leads to misleading standard errors and overconfident conclusions.

# Information, innovations, and prediction

Let  $\mathcal{F}_t$  denote information available at time  $t$ .

best forecast:  $\hat{X}_{t+h|t} = \mathbb{E}(X_{t+h} | \mathcal{F}_t)$ .

innovation:  $\varepsilon_t = X_t - \mathbb{E}(X_t | \mathcal{F}_{t-1})$ .

- An innovation is the new information arriving at time  $t$ .
- Forecast errors are built from future innovations.
- Many models are useful because they tell us how shocks propagate.

# Strict and weak stationarity

## Strict stationarity

The joint distribution of  $(X_{t_1}, \dots, X_{t_k})$  is unchanged by a common time shift.

## Weak stationarity

$$\mathbb{E}X_t = \mu, \quad \text{Var}(X_t) < \infty, \quad \text{Cov}(X_t, X_{t-h}) = \gamma(h).$$

- Strict stationarity is distributional.
- Weak stationarity is second-order.
- Most ARMA and spectral results are stated under weak stationarity.

# Autocovariance and autocorrelation

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}), \quad \rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

- $\gamma(h)$  measures dependence at lag  $h$ .
- $\rho(h)$  standardizes the scale and lies between  $-1$  and  $1$ .
- The ACF is a first diagnostic for persistence, seasonality, and mean reversion.
- PACF helps separate direct lag effects from indirect lag effects.

## Exam point

Always connect ACF/PACF patterns to model identification, not just to plots.

# Ergodicity, mixing, LLN, and CLT

Stationarity alone is not enough for sample averages to behave well.

- **Ergodicity:** time averages converge to population moments.
- **Mixing:** dependence weakens as observations become far apart.
- **LLN:**  $\bar{X}_T \rightarrow_p \mu$  under suitable dependence restrictions.
- **CLT:**  $\sqrt{T}(\bar{X}_T - \mu)$  has a normal limit with long-run variance.

$$\sqrt{T}(\bar{X}_T - \mu) \Rightarrow N(0, \Omega), \quad \Omega = \sum_{h=-\infty}^{\infty} \gamma(h).$$

# Long-run variance as a unifying object

$$\Omega = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h).$$

- Under independence,  $\Omega = \gamma(0)$ .
- Under serial dependence, all autocovariances contribute.
- HAC, fixed- $b$ , self-normalization, bootstrap, and GMM all react to this issue.

## Main finding

Time-series inference is largely about estimating, avoiding, or resampling the long-run variance.

# Wold representation

For a purely nondeterministic covariance-stationary process,

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty.$$

- This is the theoretical foundation for linear time-series models.
- The process is written as accumulated current and past innovations.
- ARMA models are finite-parameter approximations to this representation.

## Interpretation

The  $\psi_j$ 's are impulse responses in a univariate linear process.

# Linear processes and shock propagation

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.$$

$$\gamma(h) = \sigma_{\varepsilon}^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}.$$

- The coefficients  $\psi_j$  describe how shocks persist.
- Fast decay means weak persistence.
- Slow decay means high persistence and more difficult inference.
- Many “stylized facts” are statements about these propagation patterns.

# AR( $p$ ) models: stability and roots

$$X_t - c = \phi_1(X_{t-1} - c) + \cdots + \phi_p(X_{t-p} - c) + \varepsilon_t.$$

Lag-polynomial form:

$$\phi(L)X_t = c + \varepsilon_t, \quad \phi(z) = 1 - \phi_1z - \cdots - \phi_pz^p.$$

## Core theorem

The AR( $p$ ) process is causal and covariance-stationary if all roots of  $\phi(z) = 0$  lie outside the unit circle.

- Roots close to the unit circle imply high persistence.
- Roots on the unit circle imply nonstationarity.

## AR(1) as the benchmark

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t.$$

If  $|\phi| < 1$ ,

$$X_t - \mu = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, \quad \gamma(h) = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}, \quad \rho(h) = \phi^{|h|}.$$

- $\phi > 0$ : persistent mean reversion.
- $\phi < 0$ : alternating adjustment.
- $\phi \approx 1$ : near-unit-root behavior.

# MA( $q$ ) and invertibility

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}.$$

- MA models have finite shock memory.
- The ACF cuts off after lag  $q$ .
- Invertibility ensures innovations can be recovered from observed data.

## Invertibility condition

The roots of  $1 + \theta_1 z + \cdots + \theta_q z^q = 0$  should lie outside the unit circle.

# ARMA models

$$\phi(L)(X_t - \mu) = \theta(L)\varepsilon_t.$$

- AR part: persistent propagation of shocks.
- MA part: short-run correction and shock aggregation.
- Causality: AR roots outside the unit circle.
- Invertibility: MA roots outside the unit circle.

## Main finding

ARMA is not just curve fitting. It is a finite-dimensional way to approximate the Wold representation.

# Model identification: ACF, PACF, AIC, BIC

Model	ACF pattern	PACF pattern
AR( $p$ )	tails off	cuts off after $p$
MA( $q$ )	cuts off after $q$	tails off
ARMA	tails off	tails off

- ACF/PACF give preliminary identification.
- AIC often selects richer models; BIC is more parsimonious.
- Residual diagnostics check whether dependence remains.

# Forecasting with ARMA models

The optimal forecast under squared loss is

$$\hat{X}_{T+h|T} = \mathbb{E}(X_{T+h} | \mathcal{F}_T).$$

For a linear process,

$$X_{T+h} - \hat{X}_{T+h|T} = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \cdots + \psi_{h-1} \varepsilon_{T+1}.$$

$$\text{Var}(X_{T+h} - \hat{X}_{T+h|T}) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2.$$

## Interpretation

Forecast uncertainty comes from future shocks that are not in  $\mathcal{F}_T$ .

# Deterministic trends versus stochastic trends

## Trend-stationary model

$$Y_t = \alpha + \beta t + u_t, \quad u_t \text{ stationary.}$$

Shocks have temporary effects around a deterministic path.

## Difference-stationary model

$$Y_t = Y_{t-1} + \varepsilon_t, \quad \Delta Y_t = \varepsilon_t.$$

Shocks have permanent effects on the level.

## Main finding

The wrong trend specification changes both forecasts and economic interpretation.

# Unit roots and random walks

Random walk:

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

Random walk with drift:

$$Y_t = \mu + Y_{t-1} + \varepsilon_t.$$

- Variance grows with time.
- Shocks have permanent level effects.
- Standard stationary asymptotics do not apply.
- Differencing is often needed before ARMA modeling.

# ADF test: logic and specification

Typical ADF regression:

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} \sum_{j=1}^p \psi_j \Delta Y_{t-j} + u_t.$$

- Null: unit root,  $\rho = 0$ .
- Alternative: stationarity around included deterministic terms.
- Lagged differences absorb serial correlation.
- Critical values are nonstandard.

## Exam point

Always state whether the regression includes no constant, a constant, or constant plus trend.

# KPSS test: complementary logic

KPSS reverses the testing perspective.

- Null: stationarity around a level or deterministic trend.
- Alternative: unit-root-type nonstationarity.
- The statistic is based on partial sums of residuals.
- It uses a long-run variance estimator in the denominator.

## Main finding

ADF and KPSS are most useful together because they place the null hypothesis on opposite sides.

## Efficient markets and random-walk evidence

In financial applications, random-walk behavior is often connected to market efficiency.

- If prices incorporate information quickly, returns may be difficult to forecast.
- Price levels may look close to unit-root processes.
- Return predictability is usually weak and unstable.
- Evidence depends on horizon, sample period, transaction costs, and risk adjustment.

### Interpretation rule

A random-walk test is not a complete test of economic efficiency, but it is a useful diagnostic.

# Hour 1: main takeaways

- 1 Dependence changes the variance of estimators.
- 2 Stationarity makes stable second-order analysis possible.
- 3 Wold representation explains why ARMA models are natural.
- 4 AR roots govern stability and persistence.
- 5 Forecast errors come from future innovations.
- 6 Deterministic and stochastic trends imply very different economics.
- 7 ADF and KPSS should be interpreted as complementary tests.

## VAR models: the basic system

$$\mathbf{y}_t = \mathbf{c} + A_1 \mathbf{y}_{t-1} + \cdots + A_p \mathbf{y}_{t-p} + \mathbf{u}_t, \quad \mathbb{E}(\mathbf{u}_t \mid \mathcal{F}_{t-1}) = 0.$$

- VAR treats all variables as jointly dynamic.
- Each equation can be estimated by OLS under standard conditions.
- Stability is checked through the companion matrix.
- VAR is reduced-form unless additional identification is imposed.

# VAR stability and companion form

Write VAR( $p$ ) as VAR(1):

$$\mathbf{z}_t = \mathbf{a} + \mathcal{A}\mathbf{z}_{t-1} + \mathbf{e}_t.$$

## Core theorem

The VAR is stable if all eigenvalues of  $\mathcal{A}$  lie inside the unit circle.

- Stability implies a VMA( $\infty$ ) representation.
- Shocks have decaying effects.
- Impulse responses are well-defined.

# Granger causality

Variable  $x$  Granger-causes  $y$  if lagged  $x$ 's improve prediction of  $y$  conditional on other information.

$H_0$  : coefficients on lagged  $x$  in the  $y$  equation are zero.

- Tested by Wald, likelihood ratio, or  $F$ -type restrictions.
- It is about predictive content.
- It is not automatically structural causality.

## Impulse responses and forecast-error variance

For a stable VAR,

$$\mathbf{y}_t - \boldsymbol{\mu} = \sum_{j=0}^{\infty} \boldsymbol{\Phi}_j \mathbf{u}_{t-j}.$$

- $\boldsymbol{\Phi}_j$  traces the response after  $j$  periods.
- Orthogonalized IRFs depend on the decomposition of  $\boldsymbol{\Sigma}_u$ .
- FEVD asks how much forecast-error variance comes from each shock.
- Confidence bands are essential for interpretation.

### Main warning

Impulse responses are only as credible as the identification assumptions behind the shocks.

# SVAR identification

Reduced-form residuals are correlated:

$$\mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \Sigma_u.$$

Structural shocks satisfy:

$$A\mathbf{u}_t = B\boldsymbol{\varepsilon}_t, \quad \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = I.$$

- Short-run restrictions: contemporaneous zero restrictions.
- Long-run restrictions: restrictions on accumulated responses.
- Sign restrictions: economically motivated response signs.
- Recursive Cholesky identification is convenient but order-dependent.

# Cointegration

If  $\mathbf{y}_t$  is  $I(1)$ , cointegration means that some linear combination is stationary:

$$\beta' \mathbf{y}_t \sim I(0).$$

- Variables share a long-run equilibrium relation.
- Individual series can wander, but some spread is mean-reverting.
- Cointegration is central in macroeconomics and finance.

## Economic interpretation

Cointegration says that nonstationary variables are tied together by stable long-run forces.

## VECM representation

$$\Delta \mathbf{y}_t = \mathbf{c} + \alpha \beta' \mathbf{y}_{t-1} \Gamma_1 \Delta \mathbf{y}_{t-1} \cdots \Gamma_{p-1} \Delta \mathbf{y}_{t-p+1} \mathbf{u}_t.$$

- $\beta' \mathbf{y}_{t-1}$ : long-run equilibrium errors.
- $\alpha$ : adjustment speeds.
- $\Gamma_j$ : short-run dynamics.
- Cointegration rank determines how many equilibrium relations exist.

## Main findings from multivariate dynamics

- 1 VAR is powerful for prediction and dynamic interaction.
- 2 Granger causality is predictive, not automatically structural.
- 3 IRFs require careful identification and uncertainty quantification.
- 4 Cointegration prevents inappropriate differencing of long-run relations.
- 5 VECM separates long-run equilibrium correction from short-run dynamics.

# ARCH and GARCH motivation

Financial returns often have:

- weak linear autocorrelation in returns;
- strong autocorrelation in squared or absolute returns;
- volatility clustering;
- heavy tails and large shocks.

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t, \quad \mathbb{E}(z_t) = 0, \quad \text{Var}(z_t) = 1.$$

## Main finding

The mean may be hard to predict, but the variance is often highly predictable.

# ARCH and GARCH models

ARCH( $q$ ):

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2.$$

GARCH(1,1):

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2.$$

- $\alpha$ : reaction to recent shocks.
- $\beta$ : persistence in volatility.
- $\alpha + \beta < 1$ : covariance-stationary variance.
- $\alpha + \beta \approx 1$ : highly persistent volatility.

# Asymmetry, leverage, and IGARCH

- Negative financial shocks often increase future volatility more than positive shocks.
- Models such as GJR-GARCH and EGARCH allow sign-sensitive responses.
- IGARCH corresponds to near-permanent volatility persistence.
- QMLE remains useful when the conditional distribution is misspecified, under regularity conditions.

## Interpretation

Volatility models are not only forecasting tools. They encode financial risk dynamics.

## Volatility diagnostics

After fitting a volatility model, always check:

- 1 standardized residuals;
- 2 squared standardized residuals;
- 3 ARCH-LM effects left in residuals;
- 4 distributional tails;
- 5 parameter persistence and stability.

### Main warning

A fitted GARCH model is not enough. Diagnostics decide whether the model has removed the volatility dependence.

## Multivariate volatility

When multiple returns are involved, the object becomes a conditional covariance matrix:

$$H_t = \text{Var}(\mathbf{r}_t \mid \mathcal{F}_{t-1}).$$

- EWMA is simple and widely used in risk management.
- BEKK ensures positive definiteness but can be parameter-heavy.
- DCC separates conditional variances and dynamic correlations.
- The main practical issue is balancing flexibility and feasibility.

# Kernel density estimation

$$\hat{f}(x) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{x - X_t}{h}\right).$$

- $K$ : kernel shape.
- $h$ : bandwidth, the main smoothing parameter.
- Smaller  $h$ : lower bias, higher variance.
- Larger  $h$ : higher bias, lower variance.

## Main finding

Bandwidth choice matters more than the precise kernel in many applications.

# Nonparametric regression

Nadaraya–Watson estimator:

$$\hat{m}(x) = \frac{\sum_{t=1}^T K((x - X_t)/h) Y_t}{\sum_{t=1}^T K((x - X_t)/h)}.$$

- Useful for nonlinear conditional mean relationships.
- Local polynomial methods reduce boundary bias.
- Curse of dimensionality limits high-dimensional nonparametrics.
- In time series, dependence affects inference and bandwidth choice.

## Main findings from Hour 2

- 1 Systems matter when variables interact dynamically.
- 2 Identification is the difference between reduced-form shocks and structural shocks.
- 3 Cointegration protects long-run information.
- 4 Volatility is often more predictable than returns.
- 5 Nonparametric methods add flexibility but require careful smoothing choices.
- 6 Empirical interpretation must always follow the model's assumptions.

## Spectrum: dependence in the frequency domain

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h}.$$

- Low frequencies describe slow movements.
- High frequencies describe rapid fluctuations.
- Spectral peaks reveal cycles and periodic components.
- The zero-frequency spectrum is tied to the long-run variance.

$$\Omega = 2\pi f(0).$$

## Periodogram and spectral estimation

The periodogram is a sample analogue of frequency-specific variation:

$$I_T(\omega_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^T X_t e^{-i\omega_j t} \right|^2.$$

- Raw periodograms are noisy.
- Smoothing is needed for consistent spectral estimation.
- Frequency-domain logic connects naturally to HAC estimation.

# HAC long-run variance estimation

$$\hat{\Omega} = \hat{\gamma}(0) + 2 \sum_{h=1}^{b_T} k\left(\frac{h}{b_T}\right) \hat{\gamma}(h).$$

- $k(\cdot)$ : kernel or lag window.
- $b_T$ : bandwidth or truncation parameter.
- Classical consistency requires  $b_T \rightarrow \infty$  and  $b_T/T \rightarrow 0$ .
- Finite-sample performance depends heavily on bandwidth.

## Kernels, bandwidth, and positive semidefiniteness

- Bartlett/Newey–West is common and simple.
- Quadratic spectral has attractive asymptotic MSE properties.
- Positive semidefinite estimators avoid invalid variance matrices.
- Bandwidth selection is a tuning problem, not a minor detail.

### Main finding

Good long-run variance estimation is not only algebra. It is also about tuning, finite samples, and the dependence strength in the data.

## Fixed- $b$ asymptotics

Classical HAC assumes  $b_T/T \rightarrow 0$ . Fixed- $b$  instead holds

$$\frac{b_T}{T} \rightarrow b \in (0, 1].$$

- The limiting distribution is nonstandard.
- It depends on the bandwidth ratio  $b$ .
- This keeps smoothing uncertainty visible in the asymptotic approximation.
- It is useful for local power and size analysis.

### Key point

The nuisance dependence on  $b$  is precisely why fixed- $b$  analysis is informative.

## Self-normalization

Self-normalization avoids direct long-run variance estimation by using random normalizers built from the data.

Example for a sample mean:

$$W_T^2 = T^{-2} \sum_{j=1}^T j^2 (\bar{X}_j - \bar{X}_T)^2.$$

$$\frac{T^{1/2}(\bar{X}_T - \mu_0)}{W_T} \Rightarrow \text{nonstandard pivotal limit.}$$

- Operationally attractive.
- Can be conservative depending on construction.
- Adjusted-range versions can be more balanced in some settings.

## Bootstrap and subsampling for dependent data

The bootstrap must preserve dependence.

- Moving block bootstrap resamples consecutive blocks.
- Stationary bootstrap randomizes block lengths.
- Sieve bootstrap fits a parametric approximation and resamples residuals.
- Wild bootstrap is useful for heteroskedastic regression settings.

### Main warning

Ordinary i.i.d. bootstrap usually breaks the time-series dependence structure.

# GMM with time-series moments

Moment condition:

$$\mathbb{E}[g_t(\theta_0)] = 0.$$

GMM estimator:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \bar{g}_T(\theta)' W_T \bar{g}_T(\theta), \quad \bar{g}_T(\theta) = \frac{1}{T} \sum_{t=1}^T g_t(\theta).$$

- GMM generalizes method of moments.
- Overidentification allows specification testing.
- Time-series GMM needs long-run covariance estimation.
- C-CAPM is a useful asset-pricing illustration.

## C-CAPM and Hansen–Singleton logic

Euler equation:

$$\mathbb{E}_t \left[ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{t+1} - 1 \right] = 0.$$

Moment condition with instruments  $z_t$ :

$$\mathbb{E} \left[ \left\{ \beta \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{t+1} - 1 \right\} z_t \right] = 0.$$

- Parameters are preference parameters.
- Instruments generate multiple moments.
- Overidentification tests ask whether the pricing restrictions fit the data.

## Main findings from robust inference

- 1 Serial dependence enters standard errors through the long-run variance.
- 2 HAC estimation depends on kernels and bandwidths.
- 3 Fixed- $b$  keeps bandwidth effects in the limit.
- 4 Self-normalization avoids direct LRV estimation but changes critical values.
- 5 Bootstrap is useful only if the resampling scheme respects dependence.
- 6 GMM is powerful when economic theory gives moment conditions.

# Filtering: time domain and frequency domain

A linear filter is

$$y_t = \sum_{j=-m}^m a_j x_{t-j}.$$

Frequency response:

$$A(\omega) = \sum_j a_j e^{-i\omega j}.$$

- Gain  $|A(\omega)|$ : how much each frequency is retained.
- Phase: how the timing of components is shifted.
- Low-pass filters retain slow movements.
- High-pass filters retain rapid movements.

## Common filters and empirical use

Filter	Retains	Typical use
Moving average	low frequency	smooth noise
First difference	high frequency	remove unit-root trend
HP filter	trend/cycle split	macro cycles
Band-pass	selected band	business-cycle frequencies
Kalman filter	model-based signal	latent states

### Main warning

Filters are not neutral. They impose a view of what counts as trend, cycle, and noise.

# State-space models

Measurement equation:

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t).$$

Transition equation:

$$\alpha_{t+1} = T_t \alpha_t + c_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t).$$

- $y_t$ : observed data.
- $\alpha_t$ : latent state.
- The model separates observation noise from state evolution.

# Kalman filter recursion

Prediction:

$$a_{t|t-1} = T_{t-1}a_{t-1|t-1} + c_{t-1}.$$

Innovation:

$$v_t = y_t - Z_t a_{t|t-1} - d_t, \quad F_t = Z_t P_{t|t-1} Z_t' + H_t.$$

Update:

$$K_t = P_{t|t-1} Z_t' F_t^{-1}, \quad a_{t|t} = a_{t|t-1} + K_t v_t.$$

## Interpretation

The Kalman gain tells us how strongly new data revise the predicted state.

## Local level model and likelihood

$$y_t = \mu_t + \varepsilon_t, \quad \mu_{t+1} = \mu_t + \eta_t.$$

Kalman filter gives innovations  $v_t$  and variances  $F_t$ . Gaussian log-likelihood:

$$\ell = -\frac{1}{2} \sum_{t=1}^T \left[ \log(2\pi) + \log F_t + \frac{v_t^2}{F_t} \right].$$

- Small state noise: smooth level.
- Large state noise: flexible level.
- Likelihood-based inference follows from prediction errors.

## Main findings from filtering and state space

- 1 Deterministic filters are chosen operators.
- 2 State-space filters are implied by probabilistic models.
- 3 Kalman filtering is recursive prediction plus correction.
- 4 The innovation sequence is both a diagnostic and a likelihood device.
- 5 Filtering output should always be interpreted through the model or filter used.

# Brownian motion

Standard Brownian motion  $W_t$  satisfies:

- 1  $W_0 = 0$ ;
- 2 independent increments;
- 3  $W_t - W_s \sim N(0, t - s)$  for  $t > s$ ;
- 4 continuous sample paths.

## Key property

Brownian paths are continuous but almost surely nowhere differentiable.

This is why continuous-time econometrics needs stochastic calculus.

# Diffusion processes

$$dX_t = \mu(X_t, t; \theta) dt + \sigma(X_t, t; \theta) dW_t.$$

- Drift  $\mu$ : local predictable movement.
- Diffusion  $\sigma$ : local random variation.
- Many asset-price and interest-rate models are diffusions.
- Discrete observations require transition densities or approximations.

## Diffusion likelihood

If  $X_t$  is Markov and observed at  $t_0, \dots, t_T$ ,

$$L(\theta) = \prod_{i=1}^T p_{\Delta_i}(x_i \mid x_{i-1}; \theta).$$

- Exact likelihood needs known transition density.
- Approximate likelihood is used when transition density is unavailable.
- Euler approximation is simple but may be inaccurate for large intervals.
- Simulation-based and expansion methods improve approximation.

# Realized volatility

For intraday returns  $r_{t,i}$ ,

$$RV_t = \sum_{i=1}^M r_{t,i}^2.$$

Under ideal conditions,

$$RV_t \rightarrow_p \int_t^{t+1} \sigma_s^2 ds.$$

- Realized volatility estimates integrated variance.
- More frequent data reduce discretization error in theory.
- Microstructure noise creates bias at very high frequency.

## Microstructure noise and high-frequency covariance

Observed log price:

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}.$$

- Noise biases naive realized volatility upward.
- Sparse sampling reduces noise but wastes information.
- TSRV, realized kernels, and pre-averaging are noise-robust ideas.
- Asynchronous trading creates covariance problems such as the Epps effect.

### Main finding

High frequency is powerful, but only if the estimator handles market microstructure.

## R workflow across the course

- 1 Plot the series; check transformations and outliers.
- 2 Inspect ACF/PACF and persistence.
- 3 Test stationarity or unit roots when needed.
- 4 Choose mean dynamics: ARMA, VAR, VECM, or state-space model.
- 5 Check residuals and volatility clustering.
- 6 Choose inference method: standard, HAC, fixed- $b$ , self-normalized, bootstrap, or GMM.
- 7 Interpret results with assumptions and limitations.

# Method selection map

Empirical problem	Natural tools
Single stationary series	ARMA, forecasting, diagnostics
Unit-root behavior	differencing, ADF, KPSS, ARIMA
Multiple stationary variables	VAR, Granger, IRF
Long-run relations	cointegration, VECM
Time-varying risk	ARCH/GARCH, multivariate volatility
Unknown nonlinear relation	kernels, local regression
Serially dependent inference	HAC, fixed- $b$ , SN, bootstrap
Latent trend/state	filters, state-space, Kalman
High-frequency prices	realized volatility, noise-robust estimators

## Formula sheet I: dependence and linear models

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}), \quad \rho(h) = \gamma(h)/\gamma(0).$$

$$\Omega = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h).$$

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad (\text{Wold / linear process}).$$

$$\phi(L)(X_t - \mu) = \theta(L)\varepsilon_t \quad (\text{ARMA}).$$

$$\mathbf{y}_t = \mathbf{c} + A_1 \mathbf{y}_{t-1} + \dots + A_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (\text{VAR}).$$

## Formula sheet II: inference and state space

$$\widehat{\Omega} = \widehat{\gamma}(0) + 2 \sum_{h=1}^{b_T} k(h/b_T) \widehat{\gamma}(h).$$

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}} \bar{g}_T(\theta)' W_T \bar{g}_T(\theta).$$

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad \alpha_{t+1} = T_t \alpha_t + c_t + R_t \eta_t.$$

$$v_t = y_t - Z_t a_{t|t-1} - d_t, \quad a_{t|t} = a_{t|t-1} + K_t v_t.$$

$$dX_t = \mu(X_t, t; \theta) dt + \sigma(X_t, t; \theta) dW_t.$$

## Major theorem recap: what you should be able to state

In the final revision, do not only remember theorem names. You should be able to state:

- 1 the object the theorem applies to;
- 2 the main regularity or stability condition;
- 3 the conclusion of the theorem;
- 4 the econometric meaning of the conclusion.

### Revision rule

For every theorem, revise it as:

assumptions  $\implies$  mathematical conclusion  $\implies$  econometric use.

# Theorem 1: Wold representation

## Statement

Let  $\{X_t\}$  be a zero-mean, covariance-stationary, purely nondeterministic process. Then it has the one-sided linear representation

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

where  $\{\varepsilon_t\}$  is a white-noise innovation sequence.

- **Condition:** stationarity plus pure nondeterminism.
- **Conclusion:** the process is accumulated current and past innovations.
- **Use:** justifies linear-process and ARMA thinking.

## Theorem 2: AR causality and stationarity

Consider

$$\phi(L)X_t = \varepsilon_t, \quad \phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p.$$

### Statement

If all roots of  $\phi(z) = 0$  lie outside the unit circle, then the AR( $p$ ) process is causal and has the stationary MA( $\infty$ ) representation

$$X_t = \phi(L)^{-1} \varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty.$$

- **Condition:** AR roots outside the unit circle.
- **Use:** shocks have stable, decaying effects.

## Theorem 3: MA invertibility

Consider

$$X_t = \theta(L)\varepsilon_t, \quad \theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

**Statement**

If all roots of  $\theta(z) = 0$  lie outside the unit circle, then the MA( $q$ ) process is invertible:

$$\varepsilon_t = \theta(L)^{-1} X_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}, \quad \sum_{j=0}^{\infty} |\pi_j| < \infty.$$

- **Condition:** MA roots outside the unit circle.
- **Use:** innovations can be recovered from observed data.

## Theorem 4: VAR stability and VMA representation

For a VAR( $p$ ),

$$\mathbf{y}_t = \mathbf{c} + A_1 \mathbf{y}_{t-1} + \cdots + A_p \mathbf{y}_{t-p} + \mathbf{u}_t,$$

write the companion form

$$\mathbf{z}_t = \mathbf{a} + \mathcal{A} \mathbf{z}_{t-1} + \mathbf{e}_t.$$

### Statement

If every eigenvalue of  $\mathcal{A}$  has modulus strictly less than one, then the VAR is stable and has the VMA( $\infty$ ) representation

$$\mathbf{y}_t - \boldsymbol{\mu} = \sum_{j=0}^{\infty} \Phi_j \mathbf{u}_{t-j}.$$

- **Use:** impulse responses and forecast-error variances are well-defined.

## Theorem 5: Granger noncausality as restrictions

Suppose  $\mathbf{y}_t = (y_t, x_t, \mathbf{w}_t')'$  follows a stable VAR( $p$ ).

### Statement

$x_t$  does not Granger-cause  $y_t$ , conditional on  $\mathbf{w}_t$ , if and only if all coefficients on lagged  $x$ 's in the  $y$ -equation are zero:

$$H_0 : \phi_{yx,1} = \phi_{yx,2} = \dots = \phi_{yx,p} = 0.$$

This is tested as a joint linear restriction, often using a Wald, likelihood-ratio, or  $F$ -type test.

- **Meaning:** no additional predictive content from lagged  $x$ .
- **Warning:** predictive noncausality is not structural noncausality.

## Theorem 6: Cointegration and VECM representation

Let  $\mathbf{y}_t$  be an  $I(1)$  vector with cointegration rank  $r$ , where  $0 < r < n$ .

### Statement

If  $\Pi$  has rank  $r$ , then it can be factorized as

$$\Pi = \alpha\beta'$$

where  $\beta'\mathbf{y}_t$  gives  $r$  stationary cointegrating relations. The VAR in levels can be written as the VECM

$$\Delta\mathbf{y}_t = \mathbf{c} + \alpha\beta'\mathbf{y}_{t-1} \sum_{j=1}^{p-1} \Gamma_j \Delta\mathbf{y}_{t-j} \mathbf{u}_t.$$

- $\beta$ : long-run equilibrium relations.
- $\alpha$ : adjustment speeds toward equilibrium.

# Theorem 7: Dependent-data CLT and long-run variance

Let  $\{X_t\}$  be stationary with mean  $\mu$ , short-range dependence, and suitable moment/mixing conditions.

## Statement

$$\sqrt{T}(\bar{X}_T - \mu) \Rightarrow N(0, \Omega), \quad \Omega = \sum_{h=-\infty}^{\infty} \gamma(h) = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h).$$

- **Condition:** dependence must be weak enough for a CLT.
- **Conclusion:** the asymptotic variance is the long-run variance.
- **Use:** basis for HAC, GMM covariance estimation, and robust tests.

## Theorem 8: HAC consistency under small- $b$

For a sample autocovariance estimator  $\hat{\gamma}(h)$ , define

$$\hat{\Omega} = \hat{\gamma}(0) + 2 \sum_{h=1}^{b_T} k(h/b_T) \hat{\gamma}(h).$$

### Statement

Under stationarity, weak dependence, suitable kernel conditions, and the small- $b$  bandwidth condition

$$b_T \rightarrow \infty, \quad b_T/T \rightarrow 0,$$

the HAC estimator is consistent:

$$\hat{\Omega} \rightarrow_p \Omega.$$

- **Use:** standard normal or chi-squared asymptotic tests with serial correlation.

## Theorem 9: Fixed- $b$ and self-normalized limits

### Fixed- $b$ statement

If  $b_T/T \rightarrow b \in (0, 1]$ , the HAC-normalized statistic typically has a nonstandard limiting distribution:

$$\frac{\sqrt{T}(\bar{X}_T - \mu_0)}{\widehat{\Omega}_b^{1/2}} \Rightarrow \mathcal{L}_b,$$

where the limit depends on the bandwidth ratio  $b$ .

### Self-normalized statement

For a suitable recursive normalizer  $W_T$ ,

$$\frac{\sqrt{T}(\bar{X}_T - \mu_0)}{W_T} \Rightarrow \mathcal{L}_{\text{SN}},$$

where the limit is pivotal but nonstandard.

## Theorem 10: Kalman filter optimality

Consider the linear Gaussian state-space model

$$y_t = Z_t \alpha_t + d_t + \varepsilon_t, \quad \alpha_{t+1} = T_t \alpha_t + c_t + R_t \eta_t,$$

with Gaussian initial state and Gaussian disturbances.

### Statement

The Kalman filter recursively computes

$$a_{t|t-1} = \mathbb{E}(\alpha_t \mid y_1, \dots, y_{t-1}), \quad a_{t|t} = \mathbb{E}(\alpha_t \mid y_1, \dots, y_t),$$

and the corresponding minimum mean squared error covariance matrices.

- **Use:** optimal linear Gaussian filtering and likelihood construction from innovations.

# Major empirical findings checklist

- 1 Economic time series are usually serially dependent.
- 2 Macro levels often display strong persistence or unit-root behavior.
- 3 Financial returns are weakly predictable in mean but strongly predictable in volatility.
- 4 Multivariate conclusions depend on identification and lag specification.
- 5 Long-run relations should be handled with cointegration rather than naive differencing.
- 6 Robust inference can change conclusions when dependence is strong.
- 7 Filters and state-space models are useful but model-dependent.
- 8 High-frequency data require microstructure-noise corrections.

## Common exam mistakes

- Confusing strict stationarity with weak stationarity.
- Putting AR roots inside rather than outside the unit circle.
- Treating Granger causality as structural causality.
- Using a VAR in levels without thinking about unit roots or cointegration.
- Reporting GARCH estimates without checking standardized residuals.
- Writing HAC formulas without explaining the long-run variance.
- Saying fixed- $b$  removes nuisance parameters; it does not.
- Forgetting that Brownian motion is continuous but not differentiable.

# Final revision checklist

Before the final, make sure you can:

- 1 define every major model in one equation;
- 2 state the main condition for each model;
- 3 explain the empirical question each method answers;
- 4 interpret coefficients, shocks, test results, and diagnostics;
- 5 choose between alternative methods for a given dataset;
- 6 write at least one limitation for each conclusion.

## Last rule

After every formula, say what it means.

## Closing message

The whole course can be summarized as one discipline:

respect the time ordering, respect the dependence, and interpret cautiously.

- Good time-series analysis is not just estimation.
- It is model choice, diagnostic checking, robust inference, and responsible interpretation.
- The final exam rewards structure, clarity, and conceptual connections.

### Final message

Do not memorize isolated formulas. Learn the map that tells you when and why each formula is used.