# Lecture 20 — Comprehensive Review, Problem Class, and Final Exam Preparation

Whole-course synthesis: theory map, core derivations, representative problems, and R recap

Jiajing Sun

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

Econometrics and Time Series Methods
Spring 2026

# Why Lecture 20 matters

Lecture 20 is not a filler lecture. It is where the whole course is compressed into a usable map.

1. We review the conceptual spine of Lectures 1–19.
2. We identify the derivations and definitions you should be able to reproduce cleanly.
3. We practice turning the course material into exam-style answers.

**Main goal**

By the end of today, you should not feel that the course is nineteen unrelated blocks. You should feel that the main models, tests, and algorithms fit into one coherent econometric workflow for dependent data.

# Where Lecture 20 fits in the course

- Lectures 1–5: univariate structure, forecasting, and nonstationarity basics.
- Lectures 6–10: multivariate dynamics, cointegration, and volatility.
- Lectures 11–16: nonparametrics, spectral methods, robust inference, and filtering.
- Lectures 17–19: state-space systems, continuous time, and high-frequency econometrics.
- **Lecture 20:** review, problem solving, and sample exam-style preparation.
- Lecture 21: final examination.

### Practical meaning

Today is where we turn a semester of material into a set of checkpoints, problem templates, and review priorities.

# Learning goals

By the end of the review class, you should be able to:

1. explain the role of stationarity, dependence, and innovations across the whole course;
2. recognize which model or test is appropriate for a given time-series problem;
3. reproduce the most important formulas without losing the intuition behind them;
4. organize a clear exam answer rather than writing disconnected formulas;
5. identify the most common mistakes in derivations, interpretation, and empirical workflow.

# Three-hour plan

### Hour 1

Foundational review: univariate structure, forecasting, nonstationarity, and the first key derivations.

### Hour 2

Intermediate and advanced review: multivariate models, volatility, nonparametrics, robust inference, filtering, and state-space logic.

### Hour 3

Sample exam-style problems, R recap, common mistakes, and final review checklist.

# Course architecture in one picture

$$\text{dependence} \Longrightarrow \text{representation} \Longrightarrow \text{estimation}$$
$$\Longrightarrow \text{inference} \Longrightarrow \text{signal extraction} \Longrightarrow \text{continuous-time extension}$$

- Chapter 2 gives the basic language of dependence.
- Chapters 3–4 add multivariate and conditional-variance structure.
- Chapters 5–7 add flexibility, robust inference, and filtering.
- Chapters 8–9 broaden the framework to nonstationarity and continuous time.

# Notation and dependence checklist

Across the whole course, five objects appear repeatedly:

$$\mu_t = \mathbb{E}(Y_t), \qquad \gamma(h) = \mathrm{Cov}(Y_t, Y_{t-h}), \qquad f(\lambda), \qquad \Omega, \qquad \mathcal{F}_t.$$

- $\gamma(h)$ is the time-domain view of dependence.
- $f(\lambda)$ is the frequency-domain view.
- $\Omega$ or long-run variance summarizes dependence in sums.
- $\mathcal{F}_t$ is the information set used for prediction and filtering.

# Weak stationarity and ergodicity

The baseline assumptions from the first lectures are:

$$\mathbb{E}(Y_t) = \mu, \qquad \mathrm{Cov}(Y_t, Y_{t-h}) = \gamma(h).$$

- Weak stationarity keeps first and second moments time-invariant.
- Ergodicity justifies replacing theoretical moments by sample moments.
- Mixing or other weak-dependence conditions support LLN and CLT arguments.

## Review question

Why is weak stationarity not enough by itself for sample averages to behave well? Because dependence may still be too strong without an ergodic or weak-dependence condition.

# Wold decomposition and innovations

For a purely nondeterministic covariance-stationary process,

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

with innovation sequence $\{\varepsilon_t\}$.

- Wold says linear dependence can always be represented as a one-sided MA($\infty$).
- ARMA models are finite-parameter restrictions on that representation.
- Forecasting theory is built around the innovation concept.

# AR($p$) causality and stationarity

For

$$\phi(L) Y_t = \varepsilon_t, \qquad \phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p,$$

causality and covariance stationarity require

$$\phi(z) \neq 0 \quad \text{for } |z| \leq 1.$$

- Root conditions are really statements about a one-sided MA($\infty$) representation.
- Companion-form stability is the multivariate analogue used later in VAR models.

# MA($q$) invertibility

For

$$Y_t = \theta(L)\varepsilon_t, \qquad \theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q,$$

invertibility requires

$$\theta(z) \neq 0 \quad \text{for } |z| \leq 1.$$

- Invertibility gives a unique AR($\infty$) representation.
- Without it, the same autocovariance structure can correspond to observationally equivalent parameterizations.

# ARMA identification logic

### AR signature

PACF cuts off; ACF tails off.

### MA signature

ACF cuts off; PACF tails off.

### ARMA signature

Neither ACF nor PACF cuts off cleanly, so identification relies on combined diagnostics, information criteria, and model comparison rather than one perfect visual rule.

# One-step forecasting and forecast error variance

For the innovation-based forecast,

$$\widehat{Y}_{t+1|t} = \mathbb{E}(Y_{t+1} \mid \mathcal{F}_t),$$

the forecast error is

$$e_{t+1|t} = Y_{t+1} - \widehat{Y}_{t+1|t}.$$

- In ARMA models, the one-step forecast error is the next innovation.
- Multi-step forecast MSE accumulates future innovation variance through the MA weights.

$$\mathrm{Var}(Y_{t+h} - \widehat{Y}_{t+h|t}) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \psi_j^2.$$

# Deterministic trend versus stochastic trend

$$Y_t = Q_p(t; \beta) + u_t, \qquad u_t \sim I(0)$$

versus

$$Y_t = Y_{t-1} + u_t.$$

- Trend-stationary series revert to a deterministic path after detrending.
- Difference-stationary series accumulate shocks permanently and require differencing.
- The same persistent-looking sample path can correspond to very different economic mechanisms.

# DF, ADF, and KPSS logic

The Dickey–Fuller regression rewrites

$$Y_t = \rho Y_{t-1} + u_t$$

as

$$\Delta Y_t = \pi Y_{t-1} + u_t, \qquad \pi = \rho - 1.$$

- DF / ADF test the null of a unit root.
- KPSS reverses the burden of proof and tests the null of stationarity.
- Using them together is often more informative than using either one alone.

# Review problem A: derive the AR(1) MA($\infty$) representation

If

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \qquad |\phi| < 1,$$

show that

$$Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$

## What a full answer should include

1. iterate the recursion repeatedly;
2. use $|\phi| < 1$ to justify the vanishing remainder;
3. explain why this is a one-sided Wold-type representation.

# Review problem B: interpret the ADF regression

Suppose the test regression is

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

- What is the null?
- Why are lagged differences added?
- How does the presence of $\alpha$ and $\beta t$ change the critical values?

## Short answer

The lagged differences absorb residual serial correlation, and deterministic components change the asymptotic experiment, so the reference distribution is no longer the same as in the no-drift, no-trend case.

# Hour 1 summary

1. Stationarity, ergodicity, and innovations are the language of the course.
2. AR / MA root conditions are about one-sided infinite representations.
3. Forecasting is organized around the information set and the innovation sequence.
4. Nonstationarity requires a decision between detrending and differencing.
5. DF / ADF / KPSS are diagnostic tools for that decision.

# VAR stability and companion form

For a VAR($p$),

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + u_t,$$

the companion matrix $F$ governs stability:

$$\mathcal{Y}_t = F\mathcal{Y}_{t-1} + \mathcal{U}_t.$$

- Stability requires all eigenvalues of $F$ to lie inside the unit circle.
- This is the multivariate analogue of AR root conditions.
- Once stable, the VAR has a VMA($\infty$) representation.

# Granger causality

Variable $X_t$ fails to Granger-cause $Y_t$ if lagged $X$'s do not help forecast $Y$ once lagged $Y$'s are already included.

$$H_0 : A_{12,1} = A_{12,2} = \cdots = A_{12,p} = 0$$

in the relevant VAR block notation.

- This is a predictive concept, not a structural-causal proof.
- The test is usually a Wald, LR, or F-type restriction test.

# Impulse responses and FEVD

If

$$Y_t = \sum_{j=0}^{\infty} \Psi_j u_{t-j},$$

then the impulse response at horizon $h$ is $\Psi_h$, or its orthogonalized version after a shock decomposition.

- IRFs trace dynamic propagation of shocks.
- FEVD attributes forecast error variance to different shocks.
- Identification matters for structural interpretation.

# Cointegration and VECM

If $Y_t$ is $I(1)$ but

$$\beta' Y_t$$

is $I(0)$, then the system is cointegrated and admits a VECM:

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + u_t.$$

- $\beta$ contains long-run relations.
- $\alpha$ contains adjustment speeds.
- Cointegration is long-run equilibrium with short-run deviations.

# SVAR identification

Reduced-form VAR shocks are mixtures of structural shocks. A structural model may impose

$$A_0 Y_t = A_1^* Y_{t-1} + \cdots + A_p^* Y_{t-p} + \varepsilon_t, \qquad \mathrm{Var}(\varepsilon_t) = I.$$

- Short-run restrictions constrain $A_0$.
- Long-run restrictions constrain cumulative responses.
- Sign or narrative restrictions are additional identification devices.

# ARCH and GARCH core recursion

For GARCH(1,1),

$$Y_t = \sigma_t \varepsilon_t, \qquad \sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2.$$

- Volatility clusters because large past shocks raise current conditional variance.
- A common covariance-stationary condition is $\alpha + \beta < 1$.
- QMLE is widely used even when Gaussianity is only an approximation.

# Asymmetric volatility

Leverage or asymmetry can be built in through models such as GJR-GARCH:

$$\sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \gamma Y_{t-1}^2 1\{Y_{t-1} < 0\} + \beta\sigma_{t-1}^2.$$

- Negative shocks can raise volatility more than positive shocks of the same size.
- The news impact curve is a compact way to visualize that asymmetry.

# Nonparametric density and regression basics

Two canonical estimators are:

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

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

- Bandwidth controls the bias–variance trade-off.
- Kernel choice matters less than bandwidth choice.

# Spectrum and periodogram

The spectral density is the Fourier transform of the autocovariance function:

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

The sample periodogram is

$$I_T(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} Y_t e^{-it\lambda_j} \right|^2 .$$

- The periodogram is informative but not itself a consistent spectral estimator.
- Smoothing is needed for consistency.

# HAC long-run variance estimator

The long-run variance is

$$\Omega = \sum_{h=-\infty}^{\infty} \Gamma(h),$$

estimated by a kernel-weighted truncated sum:

$$\widehat{\Omega} = \widehat{\Gamma}(0) + \sum_{h=1}^{b_T} k\left(\frac{h}{b_T}\right) \left[\widehat{\Gamma}(h) + \widehat{\Gamma}(h)'\right].$$

- Kernel choice and bandwidth choice matter.
- Positive semi-definiteness is an important practical issue.

# Fixed-$b$ intuition

Classical HAC asymptotics use

$$b_T \to \infty, \qquad \frac{b_T}{T} \to 0.$$

Fixed-$b$ instead keeps

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

- The limit law depends on $b$ and usually also on the kernel.
- The point is not nuisance-free inference.
- The payoff is a better first-order approximation and a framework for local power analysis.

# Self-normalization

Self-normalization replaces an external long-run variance estimator by a random normalizer built from the sample path itself.

- Shao-style SN often has robust size behavior but can be conservative.
- Adjusted-range SN is often more balanced and can be more robust in persistent or heavy-tailed settings.
- The historical ancestor is Student's $t$-statistic.

# Bootstrap under dependence

The IID bootstrap fails for time series because it destroys dependence.

- Parametric bootstrap: simulate from a fitted dynamic model.
- Block bootstrap: resample local dependence patterns.
- Subsampling: use statistics from shorter windows rather than resampled pseudo-series.

$$b \to \infty, \qquad \frac{b}{T} \to 0$$

is the familiar block-length requirement in the standard bootstrap regime.

# GMM with time-series moments

Given moments

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

the GMM estimator minimizes

$$\widehat{Q}_T(\theta) = \bar{g}_T(\theta)' W_T \bar{g}_T(\theta), \qquad \bar{g}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta).$$

- The optimal weight is the inverse long-run covariance of the moment process.
- The $J$-test checks overidentifying restrictions.

# Filtering in time and frequency domains

For a linear filter,

$$X_t = \sum_j a_j Y_{t-j}, \qquad f_X(\lambda) = |A(\lambda)|^2 f_Y(\lambda).$$

- Time-domain weights tell us how observations are combined.
- Gain and phase tell us how frequencies are treated.
- Moving averages are low-pass smoothers; differencing is a high-pass filter.

# State-space logic and the Kalman recursion

The linear Gaussian state-space model is

$$\alpha_t = F_t \alpha_{t-1} + \omega_t, \qquad y_t = H_t \alpha_t + \eta_t.$$

- Prediction: $\alpha_{t|t-1}$.
- Filtering: $\alpha_{t|t}$.
- Kalman gain: weight on the new information surprise.

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + K_t \nu_t.$$

# Brownian motion and diffusion bridge

Continuous-time models are built around

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

- Brownian motion is the canonical source of continuous-time randomness.
- Diffusions are dynamic models driven by Brownian increments.
- High-frequency econometrics uses quadratic-variation logic to estimate integrated variance and covariance.

# Hour 2 synthesis

1. VAR / VECM / SVAR organize multivariate dependence.
2. GARCH organizes conditional second-moment dynamics.
3. Spectrum, HAC, fixed-$b$, self-normalization, and bootstrap are all ways of handling dependent sums.
4. Filtering and state-space methods shift the focus from dependence as nuisance to signal extraction.
5. Continuous-time models extend the same econometric logic into a finer time scale.

# How to approach sample exam-style questions

A good exam answer usually has four parts:

1. define the model or testing problem clearly;
2. write the key formula correctly;
3. state the assumptions or identification conditions;
4. interpret the result economically or empirically.

**Avoid this**

Do not write formulas with no definitions, or intuition with no equations. Good answers combine both.

# Sample problem 1: univariate structure

Suppose $Y_t = \phi Y_{t-1} + \varepsilon_t$, $|\phi| < 1$.

1. Derive the MA($\infty$) representation.
2. Compute $\gamma(0)$ and $\gamma(1)$.
3. Explain why the ACF decays geometrically.

$$\gamma(0) = \frac{\sigma_\varepsilon^2}{1-\phi^2}, \qquad \gamma(h) = \phi^h \gamma(0).$$

# Sample problem 1: solution map

1. Iterate the recursion repeatedly.
2. Use $|\phi| < 1$ to eliminate the remainder term.
3. Express $Y_t$ as a weighted sum of current and past innovations.
4. Use the innovation orthogonality to derive autocovariances.

### What examiners like

A clean derivation plus one sentence explaining that geometric decay is the time-domain signature of a stationary AR(1).

# Sample problem 2: unit roots and trend

Suppose a series shows strong persistence. How would you decide whether to detrend or difference?

1. Write the trend-stationary and difference-stationary benchmark models.
2. Explain what DF / ADF and KPSS test.
3. State the economic meaning of choosing detrending versus differencing.

# Sample problem 2: solution map

- Detrending is appropriate when shocks are transitory around a deterministic path.
- Differencing is appropriate when shocks have permanent level effects.
- ADF and KPSS together provide a more balanced diagnosis than either test alone.

$$\Delta Y_t = \pi Y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta Y_{t-j} + u_t$$

is the core regression to discuss.

# Sample problem 3: VAR / VECM interpretation

You estimate a cointegrated two-variable system. Explain:

1. what $\beta' Y_t$ means;
2. what the loading matrix $\alpha$ means;
3. how a VECM combines short-run and long-run dynamics.

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta Y_{t-j} + u_t.$$

# Sample problem 3: solution map

- $\beta' Y_t$ is the long-run equilibrium error.
- $\alpha$ tells us which variables adjust and how quickly.
- $\Gamma_j$ governs short-run propagation in differences.

### Common mistake

Do not say that cointegration means the level variables are stationary. It means a particular linear combination is stationary.

# Sample problem 4: volatility and robust inference

Explain the difference between:

1. conditional variance dynamics in GARCH;
2. unconditional long-run variance in HAC inference;
3. self-normalization as an alternative to explicit long-run variance estimation.

# Sample problem 4: solution map

- GARCH models $\sigma_t^2$ as a dynamic latent conditional second moment.
- HAC estimates the long-run covariance of sums or moment conditions.
- Self-normalization uses an internal stochastic denominator built from the sample path.

## Good short sentence

GARCH is a model for volatility dynamics; HAC and self-normalization are tools for inference under dependence.

# Sample problem 5: filtering and state-space methods

Compare:

1. a deterministic moving-average filter;
2. a kernel smoother;
3. a Kalman filter in a state-space model.

- What is chosen directly by the analyst?
- What is implied by a probabilistic model?
- What changes in real time versus full-sample analysis?

# Sample problem 5: solution map

- Moving-average and kernel smoothers choose weights directly.
- The Kalman filter is implied by state and observation equations under Gaussian updating.
- Real-time filtering uses a smaller information set than smoothing with the full sample.

$$\hat{\alpha}_{t|t} = \hat{\alpha}_{t|t-1} + K_t v_t$$

is the key model-based update equation.

# R recap: ARMA and unit-root workflow

```
acf(x); pacf(x)
fit <- arima(x, order = c(p,0,q))
Box.test(residuals(fit), lag = 20, type = "Ljung")

library(urca)
adf <- ur.df(x, type = "trend", lags = 4)
kpss <- ur.kpss(x, type = "tau")
```

**What to remember**

Plot first, diagnose persistence second, estimate third, and always check residuals.

# R recap: VAR, VECM, and GARCH

```
library(vars)
varfit <- VAR(Y, p = 2, type = "const")
irf(varfit)

library(urca)
jo <- ca.jo(Y, type = "trace", ecdet = "const", K = 2)

library(rugarch)
spec <- ugarchspec()
gfit <- ugarchfit(spec, x)
```

# R recap: HAC, GMM, and filtering

```
library(sandwich)
NeweyWest(lmfit, lag = 4, prewhite = FALSE)

library(gmm)
gmm(g, x = data, t0 = theta0)

library(mFilter)
bkfilter(x, pl = 6, pu = 32, nfix = 12)
hpfilter(x, freq = 1600)
```

# R recap: state-space and high-frequency work

```
library(KFAS)
model <- SSModel(y ~ SSMtrend(1, Q = list(NA)), H = NA)
fit <- fitSSM(model, inits = c(0,0))
out <- KFS(fit$model)

library(highfrequency)
rv <- rCov(prices)
rk <- rKernelCov(prices)
```

# Common mistakes across the course

1. Confusing predictive causality with structural causality.
2. Differencing by default without asking whether detrending is more appropriate.
3. Treating filtered data as if they were raw data.
4. Reporting robust standard errors without explaining what dependence problem they address.
5. Writing down a model without stating the identifying or stationarity conditions.

# Formula checklist for final review

If you are revising efficiently, make sure you can reconstruct these without notes:

- AR($p$) and MA($q$) root conditions.
- Forecast MSE formula in MA($\infty$) form.
- DF / ADF regression.
- VAR companion form and VECM representation.
- GARCH(1,1) recursion.
- HAC estimator.
- GMM criterion and $J$-test.
- Kalman update equation.
- Diffusion SDE and quadratic variation identity.

# What a good answer should look like

A strong answer usually:

1. defines the object precisely;
2. writes the right formula cleanly;
3. states the condition under which the formula is valid;
4. explains the economic or statistical meaning in ordinary English.

**One-line rule**

The best answers are mathematically correct and verbally clear at the same time.

# Final questions to self-test

Before the exam, ask yourself whether you can answer these quickly:

1. Why does invertibility matter for MA models?
2. Why does cointegration not contradict nonstationary levels?
3. Why is fixed-$b$ not nuisance-free?
4. Why can a symmetric filter look better than a real-time filter?
5. Why does realized volatility fail under microstructure noise?

# Course closing and preview of Lecture 21

Lecture 21 is the final examination.

- Read the exam instructions carefully.
- Organize answers clearly, not just formula by formula.
- Use notation consistently.
- If you are unsure, explain your reasoning rather than writing disconnected symbols.

### Closing message

This course has really been about one big theme: how to model, estimate, forecast, and interpret dependent data without pretending they are IID. If that theme is clear, then the many separate topics already fit together.