ECON 112: Macroeconomic Data Analysis
Full Lecture Note Series

Integrated Notes from Lectures, Problem Sets, and Discussion Material

Contents

1 Course Road Map and Economic Motivation

These notes organize the full ECON 112 sequence into one coherent narrative. The course has one central goal:

> learn how to model macroeconomic time series so that statistical objects, forecasting objects, and causal interpretation objects are clearly separated.

The sequence proceeds in five layers.

1. **Statics to dynamics:** move from cross-sectional OLS logic to serial dependence and temporal propagation.
2. **Univariate dynamics:** MA, AR, ARMA, identification, and forecasting.
3. **Nonstationarity:** deterministic trends, stochastic trends, unit roots, and detrending strategies.
4. **Multivariate systems:** reduced-form VAR, structural VAR, impulse responses, and policy shock interpretation.
5. **Long-run structure:** spurious regression, cointegration, Engle-Granger, and error correction.

A recurring theme is econometric humility: a model can fit the data and still fail as causal evidence. The notes therefore keep theory, identification assumptions, and post-estimation interpretation tightly linked.

2 Foundations from Week 1 and Lecture 2

2.1 Data Objects and Moments

Let X be a random variable.

- Population mean: $\mu_X = \mathbb{E}[X]$.
- Population variance: $\sigma_X^2 = \mathrm{Var}(X) = \mathbb{E}[(X-\mu_X)^2]$.
- Covariance: $\sigma_{XY} = \mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$.
- Correlation: $\rho_{XY} = \sigma_{XY}/(\sigma_X\sigma_Y)$.

For i.i.d. sample $\{(x_i, y_i)\}_{i=1}^N$:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^N x_i, \qquad s_x^2 = \frac{1}{N}\sum_{i=1}^N (x_i-\bar{x})^2, \qquad s_{xy} = \frac{1}{N}\sum_{i=1}^N (x_i-\bar{x})(y_i-\bar{y}).$$

2.2 Linear Regression as a Projection

In the bivariate model

$$Y_i = \beta_0 + \beta_1 X_i + U_i,$$

OLS solves

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg\min_{b_0,b_1}\sum_{i=1}^N (Y_i - b_0 - b_1 X_i)^2.$$

Proposition 1 (Closed-form OLS coefficients)**.** *The OLS estimators are*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Proof. First-order conditions are

$$\sum_{i=1}^{N}(Y_i - b_0 - b_1 X_i) = 0, \qquad \sum_{i=1}^{N} X_i(Y_i - b_0 - b_1 X_i) = 0.$$

From the first equation, $b_0 = \bar{Y} - b_1 \bar{X}$. Substitute into the second equation:

$$\sum_i X_i \big[(Y_i - \bar{Y}) - b_1(X_i - \bar{X})\big] = 0.$$

Rearrange:

$$b_1 \sum_i (X_i - \bar{X})^2 = \sum_i (X_i - \bar{X})(Y_i - \bar{Y}).$$

Hence $b_1 = \hat{\beta}_1$, and then $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$. □

2.3 Why Gauss-Markov Matters for This Course

Classical assumptions imply OLS is BLUE in linear models:

- Zero conditional mean $\mathbb{E}[U_i \mid X_1, \ldots, X_N] = 0$,
- Homoskedasticity $\mathrm{Var}(U_i \mid X_1, \ldots, X_N) = \sigma^2$,
- No serial correlation $\mathbb{E}[U_i U_j \mid X_1, \ldots, X_N] = 0$ for $i \neq j$.

In macro time series, the second and third conditions often fail. That is the bridge into HAC errors, dynamic modeling, and explicit lag structure.

2.4 Inference Template

For any scalar parameter θ:

$$t = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}.$$

Approximate 95% confidence interval:

$$\hat{\theta} \pm 1.96\, SE(\hat{\theta}).$$

3 Time-Series Basics: What Changes Relative to Cross Sections

3.1 Definition and Weak Stationarity

A time series $\{y_t\}_{t \in \mathbb{Z}}$ is weakly stationary if:

1. $\mathbb{E}[y_t] = \mu$, constant in t,
2. $\mathrm{Var}(y_t) = \gamma(0) < \infty$, constant in t,
3. $\mathrm{Cov}(y_t, y_{t-j}) = \gamma(j)$, depends only on lag j, not calendar time.

Define ACF $\rho(j) = \gamma(j)/\gamma(0)$.

3.2 Lag Operator and Difference Operator

Let lag operator L satisfy $Ly_t = y_{t-1}$. Then

$$(1-L)y_t = y_t - y_{t-1} = \Delta y_t.$$

This notation makes ARMA algebra concise and is essential for unit-root and cointegration analysis.

3.3 White Noise as a Primitive Shock Process

A process $\{\varepsilon_t\}$ is white noise if:

$$\mathbb{E}[\varepsilon_t] = 0, \qquad \operatorname{Var}(\varepsilon_t) = \sigma_\varepsilon^2, \qquad \operatorname{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0 \ (j \neq 0).$$

Many macro models write observed variables as filters of white noise innovations.

3.4 Ergodicity and Why Time Averages Work

If $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$, then

$$\bar{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t \xrightarrow{p} \mu.$$

So one long realization can identify population moments under dependence, provided dependence decays sufficiently.

4 Moving-Average Dynamics

4.1 MA(1) and MA(q) Models

An MA(1):

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

A general MA(q):

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

Proposition 2 (Moments of MA(1))**.** *For* $y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$,

$$\mathbb{E}[y_t] = \mu, \qquad \operatorname{Var}(y_t) = \sigma_\varepsilon^2(1+\theta^2), \qquad \gamma(1) = \theta\sigma_\varepsilon^2, \qquad \gamma(j) = 0 \text{ for } |j| \geq 2.$$

Hence

$$\rho(1) = \frac{\theta}{1+\theta^2}, \qquad \rho(j) = 0 \ (|j| \geq 2).$$

Proof. Mean is immediate by zero mean of shocks. For variance:

$$\operatorname{Var}(y_t) = \operatorname{Var}(\varepsilon_t + \theta\varepsilon_{t-1}) = \operatorname{Var}(\varepsilon_t) + \theta^2 \operatorname{Var}(\varepsilon_{t-1}) + 2\theta \operatorname{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \sigma_\varepsilon^2(1+\theta^2).$$

For covariance at lag 1:

$$\operatorname{Cov}(y_t, y_{t-1}) = \operatorname{Cov}(\varepsilon_t + \theta\varepsilon_{t-1}, \varepsilon_{t-1} + \theta\varepsilon_{t-2}) = \theta \operatorname{Var}(\varepsilon_{t-1}) = \theta\sigma_\varepsilon^2.$$

For $|j| \geq 2$, no shock overlaps between y_t and y_{t-j}, so covariance is zero. □

Remark 1. MA models produce finite-memory dependence: the ACF cuts off after lag q. This is one key identification cue in Box-Jenkins workflows.

5 Autoregressive Dynamics and Stability

5.1 AR(1) Core Results

An AR(1) with intercept:

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

If $|\phi| < 1$, stationary solution exists.

Proposition 3 (AR(1) unconditional moments)**.** *If* $|\phi| < 1$*, then*

$$\mathbb{E}[y_t] = \frac{c}{1-\phi}, \qquad \operatorname{Var}(y_t) = \frac{\sigma_\varepsilon^2}{1-\phi^2}, \qquad \rho(j) = \phi^j \textit{ for } j \geq 0.$$

Proof. Take expectations in $y_t = c + \phi y_{t-1} + \varepsilon_t$:

$$\mu = c + \phi\mu \Rightarrow \mu = \frac{c}{1-\phi}.$$

Center series: $\tilde{y}_t = y_t - \mu$, so $\tilde{y}_t = \phi\tilde{y}_{t-1} + \varepsilon_t$. Then

$$\operatorname{Var}(\tilde{y}_t) = \phi^2 \operatorname{Var}(\tilde{y}_{t-1}) + \sigma_\varepsilon^2 \Rightarrow \gamma(0) = \phi^2\gamma(0) + \sigma_\varepsilon^2 \Rightarrow \gamma(0) = \frac{\sigma_\varepsilon^2}{1-\phi^2}.$$

For lag j:

$$\gamma(j) = \operatorname{Cov}(\tilde{y}_t, \tilde{y}_{t-j}) = \phi \operatorname{Cov}(\tilde{y}_{t-1}, \tilde{y}_{t-j}) = \phi\gamma(j-1) \Rightarrow \gamma(j) = \phi^j\gamma(0).$$

Hence $\rho(j) = \phi^j$. □

5.2 AR(2) Stability and Characteristic Roots

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t.$$

Use lag polynomial:

$$\Phi(L)y_t = c + \varepsilon_t, \qquad \Phi(L) = 1 - \phi_1 L - \phi_2 L^2.$$

Stationarity requires roots of $\Phi(z) = 0$ lie outside the unit circle.

Remark 2. Equivalent statement in companion-matrix form: all eigenvalues of the state-transition matrix must have modulus less than one.

5.3 ARMA(p,q) and Identification Intuition

General ARMA:

$$\Phi(L)y_t = c + \Theta(L)\varepsilon_t,$$

with

$$\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p, \qquad \Theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q.$$

Heuristic signatures:

- AR(p): ACF tails off, PACF cuts at p.
- MA(q): ACF cuts at q, PACF tails off.
- ARMA: both tend to tail off.

6 Estimation Strategy: Box-Jenkins Logic

Lecture sequence emphasizes three disciplined stages.

6.1 Identification Stage

- Plot raw data; assess visible trend breaks and volatility episodes.
- Inspect ACF and PACF with confidence bands.
- Propose a short list of plausible low-order models.

6.2 Estimation Stage

- AR-only models: OLS often sufficient.
- ARMA models: MLE is standard because latent shock terms enter recursively.
- Choose lag orders using information criteria.

If SSR_k is residual sum of squares for model with k estimated dynamic parameters over sample size T:

$$\mathrm{AIC}(k) = \ln\left(\frac{SSR_k}{T}\right) + \frac{2(k+1)}{T}, \qquad \mathrm{BIC}(k) = \ln\left(\frac{SSR_k}{T}\right) + \frac{(k+1)\ln T}{T}.$$

BIC penalizes complexity more heavily, so it usually selects more parsimonious models.

6.3 Diagnostic Stage

- Residual plots and standardized residual plots.
- Residual ACF (Ljung-Box style logic): leftover serial structure means underfitting.
- Stability checks under sample perturbations.

6.4 Practical COVID-Era Weighting Idea

One lecture discusses downweighting extreme pandemic observations via variance scaling in shock terms. The point is not to "erase" data, but to prevent a short abnormal window from dominating low-frequency dynamics.

7 Forecasting Theory and Forecast Error Geometry

7.1 h-step Forecasts in AR Models

Under AR(1),

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

iterating expectations gives

$$\mathbb{E}_t[y_{t+h}] = \mu + \phi^h(y_t - \mu).$$

Proposition 4 (AR(1) forecast error variance)**.** *For AR(1) with* $|\phi| < 1$*,*

$$y_{t+h} - \mathbb{E}_t[y_{t+h}] = \sum_{j=0}^{h-1} \phi^j \varepsilon_{t+h-j},$$

so

$$\mathrm{Var}_t \left(y_{t+h} - \mathbb{E}_t[y_{t+h}] \right) = \sigma_\varepsilon^2 \sum_{j=0}^{h-1} \phi^{2j} = \sigma_\varepsilon^2 \frac{1 - \phi^{2h}}{1 - \phi^2}.$$

Proof. Write recursive substitution for y_{t+h} and separate terms measurable at time t from future shocks. Orthogonality of distinct innovations collapses all cross-terms in variance expansion. ☐

7.2 Random Walk Forecasting as a Benchmark

If

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

then

$$\mathbb{E}_t[y_{t+h}] = y_t, \qquad y_{t+h} - \mathbb{E}_t[y_{t+h}] = \sum_{j=1}^{h} \varepsilon_{t+j}, \qquad \mathrm{Var}_t(y_{t+h} - \mathbb{E}_t[y_{t+h}]) = h\sigma_\varepsilon^2.$$

Forecast uncertainty grows linearly with horizon, which is why long-horizon random-walk predictions are wide even when short-horizon fit looks excellent.

8 Nonstationarity, Unit Roots, and Detrending

8.1 Deterministic vs Stochastic Trend

Two different data-generating views:

$$\text{Trend-stationary (TS): } y_t = g(t) + c_t, \quad c_t \text{ stationary,}$$
$$\text{Difference-stationary (DS): } y_t = y_{t-1} + u_t \text{ (or near-unit-root AR).}$$

Why distinction matters:

- Detrending DS data with a polynomial can leave nonstationarity and create fake cycles.
- Differencing TS data removes trend but can overdifference and destroy low-frequency signal.

8.2 Dickey-Fuller Regression Derivation

Start from AR(1):

$$y_t = \alpha + \rho y_{t-1} + u_t.$$

Subtract y_{t-1}:

$$\Delta y_t = \alpha + \gamma y_{t-1} + u_t, \qquad \gamma = \rho - 1.$$

Test:

$$H_0 : \gamma = 0 \ (\rho = 1, \text{ unit root}), \qquad H_1 : \gamma < 0.$$

Distribution of the t-statistic under H_0 is nonstandard; use Dickey-Fuller critical values, not normal $N(0, 1)$ cutoffs.

8.3 Augmented Dickey-Fuller (ADF)

For AR(p)-type dynamics, add lagged differences:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{j=1}^{p-1} \psi_j \Delta y_{t-j} + u_t.$$

Again, unit-root null is $\gamma = 0$. Trend term βt is included if deterministic trend is plausible.

Remark 3. An ADF rejection is evidence against a unit root, not proof that every modeling issue is solved. Structural breaks and regime shifts can distort ADF behavior.

9 Trends and Cycles: Competing Filters

Macroeconomic decomposition often writes

$$y_t = g_t + c_t,$$

with trend g_t and cycle c_t.

9.1 Deterministic Detrending

Estimate a polynomial trend by OLS and use residuals as cycle:

$$y_t = a_0 + a_1 t + \cdots + a_n t^n + e_t, \qquad \hat{c}_t = \hat{e}_t.$$

Simple and transparent, but fragile when true trend is stochastic.

9.2 HP Filter Criterion

HP chooses $\{g_t\}$ to minimize

$$\sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(g_{t+1} - g_t) - (g_t - g_{t-1}) \right]^2.$$

First term rewards fit; second penalizes curvature. Larger λ yields smoother trend.

9.3 Hamilton Filter Idea

Hamilton's approach regresses future values on a lag block and defines cycle as forecast error. For horizon h and lag length p:

$$y_{t+h} = a + b_1 y_t + b_2 y_{t-1} + \cdots + b_p y_{t-p+1} + v_{t+h},$$

then cycle at $t + h$ is $\hat{c}_{t+h} = \hat{v}_{t+h}$.

Interpretation: the cycle is the component that is not forecastable from standard lag information.

10 Endogeneity and Instrumental Variables

10.1 Why OLS Fails Under Simultaneity

Suppose structural equation is

$$y_t = \beta x_t + u_t,$$

but x_t and u_t are correlated due to reverse causality or omitted state variables. Then

$$\text{plim } \hat{\beta}_{OLS} = \beta + \frac{\text{Cov}(x_t, u_t)}{\text{Var}(x_t)},$$

so bias does not vanish asymptotically.

10.2 IV Conditions

An instrument z_t must satisfy:

- **Relevance:** $\text{Cov}(z_t, x_t) \neq 0$.
- **Exogeneity/Exclusion:** $\text{Cov}(z_t, u_t) = 0$ and no direct channel to y_t except through x_t.

10.3 2SLS Mechanics

First stage:

$$x_t = \pi_0 + \pi_1 z_t + r_t.$$

Second stage:

$$y_t = \beta_0 + \beta_1 \hat{x}_t + e_t.$$

Matrix form:

$$\hat{\beta}_{2SLS} = (X' P_Z X)^{-1} X' P_Z y, \quad P_Z = Z(Z'Z)^{-1} Z'.$$

Proposition 5 (Consistency sketch of IV)**.** *Under relevance and exogeneity,*

$$\text{plim } \hat{\beta}_{IV} = \beta.$$

Proof. In single-instrument case,

$$\hat{\beta}_{IV} = \frac{\frac{1}{T}\sum_t z_t y_t}{\frac{1}{T}\sum_t z_t x_t}.$$

Using $y_t = \beta x_t + u_t$,

$$\text{plim } \hat{\beta}_{IV} = \frac{\beta \mathbb{E}[z_t x_t] + \mathbb{E}[z_t u_t]}{\mathbb{E}[z_t x_t]} = \beta,$$

since $\mathbb{E}[z_t u_t] = 0$ and denominator nonzero by relevance. ☐

11 Why VAR Models Enter Macroeconometrics

Pre-VAR equation-by-equation macro models were criticized for two reasons:

1. contemporaneous endogeneity across macro variables,
2. overly restrictive dynamic channels imposed a priori.

VAR logic treats each variable as endogenous and uses lag structure to capture dynamic interaction before imposing structural identification.

11.1 Structural and Reduced Forms

A structural VAR(1) can be written

$$Bx_t = Ax_{t-1} + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'] = \Lambda \text{ diagonal.}$$

Premultiply by B^{-1}:

$$x_t = \Phi x_{t-1} + u_t, \qquad \Phi = B^{-1}A, \quad u_t = B^{-1}\varepsilon_t, \quad \Sigma_u = \mathbb{E}[u_t u_t'] = B^{-1}\Lambda B^{-1\prime}.$$

Reduced form is estimable directly; structural form needs extra restrictions.

12 Reduced-Form VAR Theory and Estimation

12.1 VAR(p) and Companion Form

$$x_t = c + A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t, \qquad u_t \sim (0, \Sigma_u).$$

Define state vector

$$X_t = (x_t', x_{t-1}', \ldots, x_{t-p+1}')'.$$

Then

$$X_t = \mathcal{C} + \mathcal{A} X_{t-1} + \mathcal{U}_t,$$

where $\mathcal{A}$ is the companion matrix.

Theorem 1 (VAR stability criterion)**.** *A VAR(p) is covariance stationary if and only if all eigenvalues of $\mathcal{A}$ lie strictly inside the unit circle.*

Proof. Iterate companion system:

$$X_t = \sum_{j=0}^{m-1} \mathcal{A}^j \mathcal{C} + \mathcal{A}^m X_{t-m} + \sum_{j=0}^{m-1} \mathcal{A}^j \mathcal{U}_{t-j}.$$

If spectral radius $\rho(\mathcal{A}) < 1$, then $\mathcal{A}^m \to 0$, geometric matrix series converges, and second moments are finite/time-invariant. If any eigenvalue modulus is ≥ 1, either persistence fails to die out or moments diverge, violating covariance stationarity. ☐

12.2 MA Representation and Impulse Propagation

Under stability:

$$x_t = \mu + \sum_{j=0}^{\infty} \Psi_j u_{t-j}, \qquad \Psi_0 = I.$$

Ψ_j matrices map reduced-form innovations into dynamic responses.

12.3 Equation-by-Equation OLS

Each equation in reduced-form VAR has same regressors (lagged endogenous variables), so OLS equation-by-equation is consistent and asymptotically efficient within this system class when shocks are conditionally mean zero and homoskedastic over time.

12.4 Lag-Length Selection

Use AIC/BIC on candidate p, check residual diagnostics and stability roots. In practice:

- start from moderate p_{max},
- downselect by IC + economic plausibility,
- verify all companion roots are inside unit circle.

13 VAR Post-Estimation: Forecasting, Causality, IRFs

13.1 VAR Forecasting

Given estimated $\hat{A}_1, \dots, \hat{A}_p$, forecasts are generated recursively by replacing future shocks with zero and feeding predicted lags forward.

13.2 Granger Causality as Predictive Exclusion

Variable x_t Granger-causes y_t if lagged x terms improve prediction of y conditional on lagged y and other controls.

For equation

$$y_t = a + \sum_{j=1}^{p} b_j y_{t-j} + \sum_{j=1}^{p} c_j x_{t-j} + e_t,$$

test

$$H_0 : c_1 = \cdots = c_p = 0.$$

Rejection means predictive content, not automatic structural causality.

13.3 Reduced-Form Impulse Responses

Given MA representation $x_t = \mu + \sum_j \Psi_j u_{t-j}$, response of component k at horizon h to one-unit reduced-form innovation in component m is

$$IRF^{RF}_{k,m}(h) = e_k' \Psi_h e_m.$$

These depend on contemporaneous covariance across innovations and therefore usually do not have clean causal interpretation.

14 From Reduced-Form to Structural Interpretation

14.1 The Identification Problem

Need map

$$u_t = S\varepsilon_t, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'] = I,$$

so

$$\Sigma_u = SS'.$$

Σ_u has $n(n+1)/2$ distinct moments, while general S has n^2 entries. Extra restrictions are required.

14.2 Cholesky Identification

Choose lower-triangular S from Cholesky factorization of Σ_u. This imposes recursive contemporaneous ordering:

- earlier-ordered variables can affect later ones within period,
- later variables cannot affect earlier ones within period.

Proposition 6 (Existence and uniqueness condition)**.** *If Σ_u is symmetric positive definite, there exists a unique lower-triangular matrix S with positive diagonal such that $\Sigma_u = SS'$.*

14.3 Structural IRFs

With $u_t = S\varepsilon_t$, MA form becomes

$$x_t = \mu + \sum_{j=0}^{\infty} \Psi_j S \varepsilon_{t-j}.$$

Structural IRF:

$$IRF_{k,m}^{SVAR}(h) = e_k' \Psi_h S e_m.$$

These are interpretable as responses to orthogonalized structural shocks under maintained identifying assumptions.

14.4 Design Discipline for Policy Interpretation

A practical causal workflow used in lecture applications:

1. Define policy variable and activity variable clearly (for example short rate and output growth).
2. Estimate stable reduced-form VAR.
3. State and defend ordering restrictions.
4. Compute structural IRFs with uncertainty bands.
5. Check sign, timing, and persistence against economic mechanism.
6. Stress test with alternative orderings and lag lengths.

15 Monetary Policy Application Framework

Lecture material uses a monetary policy application to demonstrate how reduced-form dynamics become policy-shock narratives.

15.1 Core Variables and Interpretation

Typical bivariate setup:

$$\begin{pmatrix} g_t \\ i_t \end{pmatrix} = c + \sum_{j=1}^{p} A_j \begin{pmatrix} g_{t-j} \\ i_{t-j} \end{pmatrix} + u_t,$$

with g_t output growth and i_t short-rate policy indicator.

Under recursive ordering (g_t, i_t):

- policy can react contemporaneously to activity,
- activity does not react contemporaneously to policy within the same observation window.

This turns one orthogonalized innovation into a “monetary policy shock” series.

15.2 Case Study A: Dynamic Phillips-Curve System

A reduced-form VAR in inflation and unemployment highlights a common identification trap: strong reduced-form predictability may still be compatible with many structural stories. Structural claims require explicit orthogonalization assumptions and robustness checks.

15.3 Case Study B: Assigned-Country Monetary Transmission

A disciplined writeup template:

1. Data construction and transformation choices (levels, growth rates, logs).
2. Stability and lag selection evidence.
3. Chosen identification ordering and economic rationale.
4. IRF interpretation at short, medium, and long horizons.
5. Sensitivity to alternative ordering and sample windows.

16 Spurious Regression and Cointegration

16.1 Why Spurious Regression Happens

Regressing one random walk on another independent random walk often yields:

- high R^2,
- apparently significant t-statistics,
- economically meaningless relationship.

The issue is shared low-frequency wandering, not stable conditional mean structure.

Proposition 7 (Random-walk covariance growth)**.** *Let*

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2), \quad y_0 \text{ fixed.}$$

Then for $j \geq 0$,

$$\operatorname{Cov}(y_t, y_{t-j}) = (t-j)\sigma_\varepsilon^2.$$

Proof. Write

$$y_t = y_0 + \sum_{k=1}^{t} \varepsilon_k, \qquad y_{t-j} = y_0 + \sum_{k=1}^{t-j} \varepsilon_k.$$

After removing constants and expanding covariance,

$$\operatorname{Cov}(y_t, y_{t-j}) = Cov\left(\sum_{k=1}^{t} \varepsilon_k, \sum_{m=1}^{t-j} \varepsilon_m\right) = \sum_{k=1}^{t-j} \operatorname{Var}(\varepsilon_k) = (t-j)\sigma_\varepsilon^2.$$

□

16.2 Cointegration Definition

If y_t and z_t are each $I(1)$, but there exists nonzero vector $\beta = (1, -\theta)'$ such that

$$\beta' \begin{pmatrix} y_t \\ z_t \end{pmatrix} = y_t - \theta z_t$$

is $I(0)$, then (y_t, z_t) are cointegrated.

Economic meaning: variables share a stochastic trend, while some long-run equilibrium relation remains stationary.

16.3 Superconsistency Intuition

When cointegration is true, OLS estimate of θ in long-run regression converges faster than standard $\sqrt{T}$ rate (roughly T-rate). But finite-sample inference still needs specialized critical values in residual-based tests.

17 Engle-Granger Procedure and Error Correction

17.1 Two-Step Engle-Granger Workflow

1. Verify both series appear $I(1)$.
2. Estimate long-run relation by OLS:

$$z_t = \alpha + \theta y_t + e_t.$$

3. Test residual $\hat{e}_t$ for stationarity with residual-based ADF and Engle-Granger critical values.
4. If residual is stationary, estimate ECM.

17.2 ECM Derivation

Suppose long-run equilibrium error is

$$\nu_t = z_t - \theta y_t,$$

and follows AR(1):

$$\nu_t = \rho \nu_{t-1} + \eta_t, \qquad |\rho| < 1.$$

Then

$$\nu_t - \nu_{t-1} = (\rho - 1)\nu_{t-1} + \eta_t.$$

Since $\nu_t = z_t - \theta y_t$:

$$\Delta z_t - \theta \Delta y_t = (\rho - 1)(z_{t-1} - \theta y_{t-1}) + \eta_t.$$

Rearrange:

$$\Delta z_t = \theta \Delta y_t + \lambda (z_{t-1} - \theta y_{t-1}) + \eta_t, \qquad \lambda = \rho - 1 < 0.$$

Remark 4. The negative sign on λ is economically critical: if system drifts above equilibrium, correction term pushes it back down, and vice versa.

17.3 Decision Rule: VAR in Levels, Differences, or VECM

For variables $I(0)$: estimate VAR in levels.

For variables $I(1)$ and not cointegrated: difference and estimate VAR in differences.

For variables $I(1)$ and cointegrated: estimate VECM,

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + u_t,$$

where Π encodes long-run cointegration relations and adjustment speeds.

18 Additional Proof Compendium

18.1 Proof: Cointegration from a Common Stochastic Trend

Assume

$$x_t = \mu z_t + \epsilon_t^x, \qquad y_t = z_t + \epsilon_t^y, \qquad z_t = z_{t-1} + \eta_t,$$

where $\epsilon_t^x, \epsilon_t^y, \eta_t$ are stationary shocks. Then

$$x_t - \mu y_t = \epsilon_t^x - \mu \epsilon_t^y,$$

which is stationary. Hence x_t and y_t are cointegrated with vector $(1, -\mu)$.

18.2 Proof: h-step Forecast Error Variance for Random Walk

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

$$\mathbb{E}_t[y_{t+h}] = y_t, \quad y_{t+h} - y_t = \sum_{j=1}^{h} \varepsilon_{t+j}.$$

Therefore

$$\mathrm{Var}_t(y_{t+h} - y_t) = \sum_{j=1}^{h} \mathrm{Var}(\varepsilon_{t+j}) = h\sigma_\varepsilon^2.$$

18.3 Proof: Stationarity Condition for VAR(1)

Given $x_t = Ax_{t-1} + u_t$, iterate:

$$x_t = A^m x_{t-m} + \sum_{j=0}^{m-1} A^j u_{t-j}.$$

If all eigenvalues of A satisfy $|\lambda_i| < 1$, then $A^m \to 0$ and

$$x_t = \sum_{j=0}^{\infty} A^j u_{t-j}$$

exists with finite second moments. If any $|\lambda_i| \geq 1$, powers of A fail to vanish and covariance stationarity fails.

18.4 Proof: Mapping Structural and Reduced Innovations

Let structural system be

$$B_0 x_t = B_1 x_{t-1} + \varepsilon_t, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'] = I.$$

Premultiplying by B_0^{-1}:

$$x_t = A x_{t-1} + u_t, \quad A = B_0^{-1} B_1, \quad u_t = B_0^{-1} \varepsilon_t.$$

Hence

$$\Sigma_u = \mathbb{E}[u_t u_t'] = B_0^{-1} B_0^{-1\prime}.$$

Identification asks for a matrix factorization of Σ_u plus economically meaningful restrictions on B_0^{-1}.

19 Derivation and Proof Appendix I: Regression and IV Derivations

19.1 Matrix OLS from First Principles

Write the static linear model in matrix form:

$$y = X\beta + u, \qquad y \in \mathbb{R}^{T\times 1},\ X \in \mathbb{R}^{T\times k},\ \beta \in \mathbb{R}^{k\times 1}.$$

OLS solves

$$\hat\beta = \arg\min_b (y - Xb)'(y - Xb).$$

Theorem 2 (Matrix OLS estimator). *If X has full column rank k, then*

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

Proof. Expand objective:

$$Q(b) = y'y - 2b'X'y + b'X'Xb.$$

The gradient is

$$\nabla_b Q(b) = -2X'y + 2X'Xb.$$

Setting $\nabla_b Q(b) = 0$ gives normal equations $X'Xb = X'y$. Full rank implies $X'X$ invertible, hence $b = \hat\beta = (X'X)^{-1}X'y$. ☐

Proposition 8 (Conditional mean and variance). *Under $\mathbb{E}[u \mid X] = 0$, $\operatorname{Var}(u \mid X) = \sigma^2 I_T$:*

$$\mathbb{E}[\hat\beta \mid X] = \beta, \qquad \operatorname{Var}(\hat\beta \mid X) = \sigma^2 (X'X)^{-1}.$$

Proof. From $\hat\beta - \beta = (X'X)^{-1}X'u$,

$$\mathbb{E}[\hat\beta - \beta \mid X] = (X'X)^{-1}X'\mathbb{E}[u \mid X] = 0.$$

For variance:

$$\operatorname{Var}(\hat\beta \mid X) = (X'X)^{-1}X' \operatorname{Var}(u \mid X) X (X'X)^{-1} = \sigma^2 (X'X)^{-1}.$$

☐

19.2 Frisch-Waugh-Lovell (FWL) with Explicit Projection Algebra

Partition regressors into $X = [X_1\ X_2]$, where coefficient of interest is on X_1:

$$y = X_1\beta_1 + X_2\beta_2 + u.$$

Define residual-maker $M_2 = I - P_2$, $P_2 = X_2(X_2'X_2)^{-1}X_2'$.

Theorem 3 (FWL theorem)**.** *The OLS coefficient on X_1 from the full regression equals*

$$\hat\beta_1 = (X_1'M_2X_1)^{-1}X_1'M_2y.$$

Proof. Normal equation for β_1:

$$X_1'(y - X_1\beta_1 - X_2\beta_2) = 0.$$

From second block,

$$\hat\beta_2 = (X_2'X_2)^{-1}X_2'(y - X_1\beta_1).$$

Substitute into first block:

$$X_1'(I - X_2(X_2'X_2)^{-1}X_2')(y - X_1\beta_1) = 0.$$

That is $X_1'M_2(y - X_1\beta_1) = 0$, implying stated formula. ☐

19.3 Omitted Variable Bias in Full Matrix Form

Suppose true model is

$$y = X_1\beta_1 + X_2\beta_2 + u, \qquad \mathbb{E}[u \mid X_1, X_2] = 0,$$

but short regression omits X_2:

$$y = X_1\delta + v.$$

Proposition 9 (Multivariate OVB formula)**.** *If $T^{-1}X_1'X_1 \to Q_{11}$ nonsingular and $T^{-1}X_1'X_2 \to Q_{12}$, then*

$$\operatorname{plim} \hat\delta = \beta_1 + Q_{11}^{-1}Q_{12}\beta_2.$$

Proof.

$$\hat\delta = (X_1'X_1)^{-1}X_1'y = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + u).$$

So

$$\hat\delta = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'u.$$

The final term is $o_p(1)$ under exogeneity and LLN conditions, yielding the probability limit. ☐

19.4 2SLS as GMM and Asymptotic Variance

Moment condition for valid instruments $Z \in \mathbb{R}^{T\times m}$:

$$\mathbb{E}[z_tu_t] = 0, \qquad u_t = y_t - x_t'\beta.$$

2SLS equals one-step efficient linear IV under homoskedasticity:

$$\hat\beta_{2SLS} = (X'P_ZX)^{-1}X'P_Zy.$$

Proposition 10 (Asymptotic distribution of just-identified IV)**.** *For one endogenous regressor and one valid instrument,*

$$\sqrt{T}\,(\hat{\beta}_{IV} - \beta) \xrightarrow{d} \mathcal{N}\left(0,\ \frac{\operatorname{Var}(z_t u_t)}{\mathbb{E}[z_t x_t]^2}\right).$$

Proof. Write

$$\hat{\beta}_{IV} - \beta = \frac{T^{-1}\sum_t z_t u_t}{T^{-1}\sum_t z_t x_t}.$$

Denominator converges to nonzero $\mathbb{E}[z_t x_t]$ by relevance. Numerator obeys CLT under weak dependence and finite second moments, giving stated limit by Slutsky. ☐

20 Derivation and Proof Appendix II: Univariate Time-Series Proofs

20.1 Yule-Walker Equations for AR(p)

For covariance-stationary AR(p),

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t.$$

Let $\gamma(j) = \operatorname{Cov}(y_t, y_{t-j})$.

Proposition 11 (Yule-Walker system)**.** *For $j \geq 1$:*

$$\gamma(j) = \phi_1\gamma(j-1) + \cdots + \phi_p\gamma(j-p),$$

and for $j = 0$:

$$\gamma(0) = \phi_1\gamma(1) + \cdots + \phi_p\gamma(p) + \sigma_\varepsilon^2.$$

Proof. Multiply centered AR(p) equation by $y_{t-j} - \mu$, take expectations, and use $\operatorname{Cov}(\varepsilon_t, y_{t-j}) = 0$ for $j \geq 1$. For $j = 0$, the innovation contributes $\operatorname{Var}(\varepsilon_t) = \sigma_\varepsilon^2$. ☐

20.2 AR(2) Stationarity Region in Parameter Space

For

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

the characteristic polynomial is $1 - \phi_1 z - \phi_2 z^2$.

Proposition 12 (Equivalent AR(2) stability inequalities)**.** *Covariance stationarity is equivalent to*

$$\phi_2 < 1, \qquad \phi_2 > -1, \qquad \phi_1 + \phi_2 < 1, \qquad \phi_2 - \phi_1 < 1.$$

Proof. Stationarity requires both roots of $1 - \phi_1 z - \phi_2 z^2 = 0$ outside unit circle. Mapping root-modulus constraints into (ϕ_1, ϕ_2)-space yields the triangular region above. This is algebraically equivalent to Schur stability of the AR(2) companion matrix. ☐

20.3 Infinite MA Representation for Stable ARMA

Let

$$\Phi(L)y_t = c + \Theta(L)\varepsilon_t, \qquad \Phi(0) = 1,$$

with all roots of $\Phi(z) = 0$ outside unit circle.

Theorem 4 (Wold-type linear representation)**.** *There exists absolutely summable* $\{\psi_j\}_{j\geq 0}$*,* $\psi_0 = 1$*, such that*

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \qquad \mu = \frac{c}{\Phi(1)}.$$

Proof. Stability implies $\Phi(L)^{-1} = \sum_{j\geq 0} \pi_j L^j$ with absolute summability. Then

$$y_t = \Phi(L)^{-1}c + \Phi(L)^{-1}\Theta(L)\varepsilon_t = \mu + \sum_{j\geq 0} \psi_j \varepsilon_{t-j},$$

where $\{\psi_j\}$ is convolution of $\{\pi_j\}$ and MA coefficients. □

20.4 Conditional Likelihood for ARMA(1,1)

For

$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t + \theta\varepsilon_{t-1}, \quad \varepsilon_t \sim iid\,\mathcal{N}(0, \sigma^2),$$

conditional on initial values and ε_0, Gaussian log-likelihood is

$$\ell(\mu, \phi, \theta, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \hat{\varepsilon}_t(\mu, \phi, \theta)^2.$$

MLE chooses parameters minimizing conditional innovation sum of squares, linking ARMA estimation to the Box-Jenkins estimation stage.

21 Derivation and Proof Appendix III: Forecasting as Projection

21.1 Projection Theorem in L2

Let $\mathcal{I}_t$ be sigma-field generated by observed history up to time t.

$$\hat{y}_{t+h|t} := \mathbb{E}[y_{t+h} \mid \mathcal{I}_t].$$

Theorem 5 (Orthogonality condition)**.** $\hat{y}_{t+h|t}$ *is the unique minimizer of*

$$\min_{g\in L^2(\mathcal{I}_t)} \mathbb{E}[(y_{t+h} - g)^2].$$

Moreover,

$$\mathbb{E}[(y_{t+h} - \hat{y}_{t+h|t})q] = 0 \quad \forall q \in L^2(\mathcal{I}_t).$$

Proof. Decompose any $g \in L^2(\mathcal{I}_t)$:

$$y_{t+h} - g = (y_{t+h} - \hat{y}_{t+h|t}) + (\hat{y}_{t+h|t} - g).$$

Square and take expectations. Cross term is zero by tower property:

$$\mathbb{E}[(y_{t+h} - \hat{y}_{t+h|t})(\hat{y}_{t+h|t} - g)] = 0,$$

since second factor is $\mathcal{I}_t$-measurable. Hence MSE is minimized uniquely at $g = \hat{y}_{t+h|t}$. □

21.2 General h-step Error Variance via MA Coefficients

If

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

then

$$\mathbb{E}_t[y_{t+h}] = \mu + \sum_{j=h}^{\infty} \psi_j \varepsilon_{t+h-j},$$

and forecast error is

$$e_{t+h|t} = y_{t+h} - \mathbb{E}_t[y_{t+h}] = \sum_{j=0}^{h-1} \psi_j \varepsilon_{t+h-j}.$$

Proposition 13 (Multi-step forecast variance)**.**

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

Proof. Apply orthogonality of innovations at different times. ☐

21.3 Direct vs Iterated Forecasting

Iterated forecasts recursively apply one-step law-of-motion; direct forecasts estimate separate model for each horizon h. Iterated forecasts are efficient under correct dynamic specification, while direct forecasts can be more robust under misspecification. This tradeoff is central when structural breaks (for example COVID period) are plausible.

22 Derivation and Proof Appendix IV: Unit Root and Detrending

22.1 Why Random Walks Are Nonstationary

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

$$y_t = y_0 + \sum_{j=1}^{t} \varepsilon_j, \qquad \mathrm{Var}(y_t) = t\sigma_\varepsilon^2.$$

Because variance depends on calendar time t, weak stationarity fails immediately.

22.2 ADF Regression from AR(p) in Levels

Start with

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} + u_t.$$

Subtract y_{t-1}:

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{j=1}^{p-1} \varphi_j \Delta y_{t-j} + u_t, \qquad \gamma = \rho - 1.$$

Unit root null is $H_0 : \gamma = 0$.

Remark 5. Lagged differences absorb serial correlation in u_t. If insufficient lags are included, the ADF t-statistic is size-distorted.

22.3 Deterministic Components and Test Specification

Three common deterministic choices:

Specification	Regression	Typical use case
No constant, no trend	$\Delta y_t = \gamma y_{t-1} + \cdots + u_t$	Mean-zero series around 0
Constant only	$\Delta y_t = \alpha + \gamma y_{t-1} + \cdots + u_t$	Nonzero mean, no linear trend
Constant and trend	$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \cdots + u_t$	Potential deterministic trend

Using the wrong deterministic specification can materially change inference near the unit-root boundary.

22.4 Spurious Regression Asymptotics: Core Mechanism

Let x_t and y_t be independent random walks. In levels regression

$$y_t = \alpha + \beta x_t + e_t,$$

both numerator and denominator in $\hat{\beta}$ are $O_p(T^2)$, so $\hat{\beta}$ converges to a nondegenerate random limit rather than zero. Standard t-ratios therefore do not follow standard normal reference behavior.

Remark 6. This is why high R^2 and low p-values in levels regressions with unit-root variables can be pure artifacts.

22.5 HP Filter in Matrix Form

Let $y = (y_1, \ldots, y_T)'$, $g = (g_1, \ldots, g_T)'$, and D denote second-difference matrix so that $(Dg)_t = g_{t+1} - 2g_t + g_{t-1}$. HP objective is

$$\min_g (y - g)'(y - g) + \lambda (Dg)'(Dg).$$

First-order condition:

$$(I + \lambda D'D)g = y.$$

Hence

$$\hat{g} = (I + \lambda D'D)^{-1} y, \qquad \hat{c} = y - \hat{g}.$$

This closed form makes clear HP is a linear smoother with frequency-dependent attenuation.

23 Derivation and Proof Appendix V: VAR Estimation and Inference

23.1 Stacked VAR Representation

For VAR(p),

$$x_t = c + A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t,$$

define regressor vector $w_t = (1, x'_{t-1}, \ldots, x'_{t-p})'$. Stack:

$$X = WB + U,$$

where $X \in \mathbb{R}^{T\times n}$, $W \in \mathbb{R}^{T\times k}$, $B \in \mathbb{R}^{k\times n}$. OLS estimator:

$$\hat{B} = (W'W)^{-1} W'X.$$

23.2 Why Equation-by-Equation OLS Equals SUR Here

Proposition 14. *In reduced-form VAR systems with identical regressors across equations, feasible GLS/SUR and equation-by-equation OLS yield identical coefficient estimates.*

Proof. SUR estimator can be written in vectorized form:

$$\mathrm{vec}(\hat{B}_{SUR}) = \Big[(I_n \otimes W)'(\Sigma_u^{-1} \otimes I_T)(I_n \otimes W)\Big]^{-1} (I_n \otimes W)'(\Sigma_u^{-1} \otimes I_T)\,\mathrm{vec}(X).$$

Because the regressor matrix is common across equations, Kronecker products simplify and Σ_u^{-1} factors out symmetrically, leaving $\mathrm{vec}(\hat{B}_{SUR}) = \mathrm{vec}(\hat{B}_{OLS})$. ☐

23.3 Asymptotic Normality of VAR OLS

Theorem 6. *Under covariance stationarity, finite fourth moments, and weak dependence:*

$$\sqrt{T}\ \mathrm{vec}(\hat{B} - B) \xrightarrow{d} \mathcal{N}\left(0,\ Q_W^{-1} \otimes \Sigma_u\right),$$

where $Q_W = \mathrm{plim}\ T^{-1}W'W$.

Proof. Use

$$\mathrm{vec}(\hat{B} - B) = \left(I_n \otimes (W'W)^{-1}W'\right)\mathrm{vec}(U).$$

Apply CLT to $T^{-1/2}\sum_t (w_t \otimes u_t)$, then Slutsky with $T^{-1}W'W \to Q_W$. ☐

23.4 Impulse Response Recursion

For stable VAR(p), MA coefficients satisfy

$$\Psi_0 = I, \qquad \Psi_h = \sum_{j=1}^{\min(h,p)} A_j \Psi_{h-j} \quad (h \geq 1).$$

This recursion is used directly in computational IRF routines.

23.5 Forecast Error Variance Decomposition (FEVD)

Under orthogonal structural shocks $u_t = S\varepsilon_t$, horizon-H forecast error for variable k has variance

$$\sum_{h=0}^{H-1} e_k' \Psi_h S S' \Psi_h' e_k.$$

Contribution of shock m:

$$\mathrm{FEVD}_{k,m}(H) = \frac{\sum_{h=0}^{H-1} (e_k' \Psi_h S e_m)^2}{\sum_{h=0}^{H-1} e_k' \Psi_h S S' \Psi_h' e_k}.$$

This is the variance-share analog of IRF analysis.

23.6 Granger Causality Test Statistic

Let unrestricted equation have RSS RSS_U and restricted one (imposing $c_1 = \cdots = c_p = 0$) have RSS_R. With $q = p$ restrictions and $T - k$ unrestricted degrees of freedom:

$$F = \frac{(RSS_R - RSS_U)/q}{RSS_U/(T-k)}.$$

Large F rejects predictive exclusion.

24 Derivation and Proof Appendix VI: Structural VAR Identification

24.1 Counting Restrictions Carefully

For n-variable SVAR with

$$u_t = S\varepsilon_t, \qquad \mathbb{E}[\varepsilon_t \varepsilon_t'] = I,$$

unknown S has n^2 free entries while $\Sigma_u = SS'$ provides $n(n+1)/2$ equations. Therefore required additional restrictions:

$$n^2 - \frac{n(n+1)}{2} = \frac{n(n-1)}{2}.$$

This count explains why identification choices are unavoidable, not optional.

24.2 Recursive (Cholesky) Identification as Zero Restrictions

Assume contemporaneous system

$$B_0 x_t = b + B_1 x_{t-1} + \cdots + B_p x_{t-p} + \varepsilon_t,$$

with B_0 lower triangular and ones on diagonal. Then

$$u_t = B_0^{-1}\varepsilon_t, \qquad \Sigma_u = B_0^{-1} B_0^{-1\prime}.$$

The zero pattern in B_0 corresponds to recursive within-period exclusion assumptions.

Proposition 15. *If Σ_u is positive definite and ordering is fixed, recursive short-run restrictions imply a unique B_0^{-1} with positive diagonal.*

Proof. Uniqueness follows from unique Cholesky factor S with positive diagonal such that $\Sigma_u = SS'$. Set $B_0^{-1} = S$ and invert. ☐

24.3 Long-Run Restrictions

In stable VAR, cumulative long-run impact matrix is

$$C(1) = \sum_{h=0}^{\infty} \Psi_h S = (I - A_1 - \cdots - A_p)^{-1} S.$$

Long-run neutrality assumptions set selected entries of $C(1)$ to zero, providing non-recursive identification alternatives.

24.4 External Instrument (Proxy-SVAR) Logic

If observable proxy m_t is correlated with one structural shock ε_t^p and orthogonal to others:

$$\mathrm{Cov}(m_t, \varepsilon_t^p) \neq 0, \qquad \mathrm{Cov}(m_t, \varepsilon_t^j) = 0 \ (j \neq p),$$

then covariance between proxy and reduced-form residuals identifies one column of S up to scale. This extends causal design beyond ordering-based identification.

25 Derivation and Proof Appendix VII: Cointegration and VECM

25.1 Deriving VECM from VAR(p) in Levels

Start from

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t.$$

Subtract x_{t-1} and rearrange:

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + u_t,$$

where

$$\Pi = \sum_{j=1}^{p} A_j - I, \qquad \Gamma_j = - \sum_{m=j+1}^{p} A_m.$$

25.2 Rank of Pi and Economic Interpretation

- $\operatorname{rank}(\Pi) = 0$: no cointegration, pure differenced VAR.
- $0 < \operatorname{rank}(\Pi) = r < n$: r cointegration relations, $\Pi = \alpha\beta'$.
- $\operatorname{rank}(\Pi) = n$: levels are stationary.

When $0 < r < n$, columns of β are long-run equilibrium vectors and α contains speeds of adjustment.

25.3 Error-Correction Stability and Sign Restrictions

For scalar ECM

$$\Delta y_t = \kappa(z_{t-1} - \theta y_{t-1}) + \xi_t,$$

dynamic stability requires $\kappa < 0$. If $z_{t-1} - \theta y_{t-1} > 0$, a negative κ lowers y_t, moving system back to equilibrium.

Proposition 16 (Half-life under AR(1) correction)**.** *If equilibrium error follows* $e_t = \rho e_{t-1} + \eta_t$, $|\rho| < 1$*, then deviation half-life is*

$$h_{1/2} = \frac{\log(0.5)}{\log|\rho|}.$$

Proof. By recursion, $e_{t+h} = \rho^h e_t + \sum_{j=1}^{h} \rho^{h-j} \eta_{t+j}$. Ignoring future shocks for deterministic decay of a given deviation, set $|\rho|^h = 0.5$. ☐

25.4 Superconsistency of Cointegrating OLS: Proof Sketch

Suppose $z_t = \theta y_t + e_t$, with $y_t \sim I(1)$, $e_t \sim I(0)$. Then

$$\hat{\theta} - \theta = \frac{\sum_t y_t e_t}{\sum_t y_t^2}.$$

Orders:

$$\sum_t y_t^2 = O_p(T^2), \qquad \sum_t y_t e_t = O_p(T).$$

Therefore

$$\hat{\theta} - \theta = O_p(T^{-1}),$$

faster than standard $T^{-1/2}$ rate.

26 Derivation and Proof Appendix VIII: Causal Design Case Studies

26.1 Case Study 1: Recursive Monetary Policy Shock Design

Consider $x_t = (g_t, \pi_t, i_t)'$, where g_t is output growth, π_t inflation, i_t policy rate. Reduced-form VAR estimated first. Structural interpretation uses ordering (g_t, π_t, i_t) or (π_t, g_t, i_t), each embedding different within-period information assumptions.

Design checklist for credible causal interpretation:

1. Verify reduced-form stability.
2. State exact contemporaneous exclusion assumptions implied by ordering.
3. Report IRF sign, peak timing, and persistence.
4. Show robustness to alternative lag length and ordering.
5. Distinguish "policy shock" from systematic policy-rule component.

26.2 Case Study 2: Narrative/External-Instrument Monetary Shock

When available, narrative shock series m_t can be used as external instrument for policy equation innovation u_t^i:

$$\mathrm{Cov}(m_t, u_t^i) \neq 0, \qquad \mathrm{Cov}(m_t, u_t^j) = 0 \ (j \neq i).$$

Then IRFs can be constructed from identified policy-shock column without relying solely on recursive ordering.

26.3 Case Study 3: Cointegrated Macro Relationship

Suppose long-run theory predicts $c_t - \lambda y_t$ stationary (for example consumption and income). A standard empirical sequence:

1. test each variable for $I(1)$,
2. estimate cointegrating relation,
3. test residual stationarity with Engle-Granger critical values,
4. estimate ECM or VECM and verify negative adjustment coefficients,
5. evaluate short-run dynamics separately from long-run equilibrium.

This prevents conflating transient co-movement with equilibrium economics.

27 Integrated Workflow for Empirical Projects

27.1 Step-by-Step Checklist

1. **Transform data consistently:** levels vs logs vs growth rates.
2. **Diagnose stationarity:** visuals + ADF + economic theory.

3. **Select model class:** ARMA, VAR, or VECM depending on integration/cointegration properties.
4. **Estimate parsimoniously:** choose lags by BIC/AIC with stability checks.
5. **Validate residuals:** serial correlation, volatility behavior, structural breaks.
6. **Interpret carefully:** prediction claims, reduced-form dependence, and structural claims are different layers.
7. **Stress test:** re-estimate under alternative lags/orderings/sample windows.

27.2 Frequent Mistakes to Avoid

- Treating Granger causality as fully structural causality.
- Running levels VAR on nonstationary noncointegrated data.
- Forgetting that Cholesky identification depends on variable ordering.
- Using standard normal critical values for residual-based cointegration tests.
- Reporting one preferred specification without robustness table.

28 Linear Algebra Appendix for VAR Work

28.1 Matrix Inverse for 2x2

For

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad ad - bc \neq 0,$$

$$M^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

This appears directly in two-variable structural-to-reduced transformations.

28.2 Eigenvalues and Stability

For matrix A, eigenvalues solve

$$\det(A - \lambda I) = 0.$$

In VAR practice, root plots and eigenvalue magnitudes provide the operational stability check.

28.3 Positive Definiteness and Cholesky

Symmetric Σ is positive definite if

$$v'\Sigma v > 0 \quad \forall v \neq 0.$$

Then unique lower-triangular L with positive diagonal exists such that $\Sigma = LL'$.

29 Compact Recap of the Full Course

1. OLS teaches projection logic and where exogeneity fails.
2. ARMA teaches persistence, mean reversion, and forecasting mechanics.
3. Unit-root analysis separates stochastic trend from cyclical fluctuation.
4. VAR teaches multivariate dependence with minimal theory restrictions.
5. SVAR adds identifying assumptions to turn innovations into structural shocks.
6. Cointegration and ECM recover long-run equilibrium with short-run adjustment.

The main intellectual payoff is the ability to distinguish three valid but different statements:

- **Predictive statement:** one variable helps forecast another.
- **Statistical dependence statement:** shocks co-move through reduced-form covariance.
- **Causal statement:** under explicit structural restrictions, a policy shock moves outcomes in a specific dynamic pattern.