The Johansen Procedure: Full Derivation

Companion notes for Unit 7

Natasha Kang

May, 2026

Table of contents

1 Overview

These notes fill in every step of the derivation underlying the Johansen procedure. The slides give the storyline; here we give the algebra. The target audience is a student who is comfortable with multivariate calculus and matrix algebra but has never seen profile-likelihood arguments on matrix parameters.

The argument has four moves, and every Johansen result on the slides is a consequence of one of them:

1. **Concentrate out short-run dynamics** by partialling — the FWL step. This reduces the VECM to a static-looking long-run regression in residualized variables.
2. **Profile out Σ_u.** With Gaussian errors, for any fixed Π, the MLE of Σ_u is the sample covariance of the residuals. Substituting back, the log-likelihood becomes $-\frac{T}{2}\log|\widehat{\Sigma}_u(\Pi)| + \text{const}$.
3. **Profile out α given β.** With $\Pi = \alpha\beta'$, $\widehat{\Sigma}_u(\alpha, \beta)$ is quadratic in α, so we can complete the square. The minimizer is $\hat{\alpha} = S_{01}\beta(\beta' S_{11}\beta)^{-1}$.
4. **Reduce the β-problem to a generalized eigenvalue problem.** After plugging $\hat{\alpha}$ in and using two standard determinant identities, the constrained minimization in β becomes the eigenproblem
$$S_{10}S_{00}^{-1}S_{01}\, v = \lambda\, S_{11}\, v.$$

Once we are at the eigenvalue problem, the LR rank tests fall out mechanically: every $\widehat{\Sigma}$ is a product of $(1 - \hat{\lambda}_i)$ factors, so the LR statistic is a sum of $\log(1 - \hat{\lambda}_i)$.

Throughout, $x_t \in \mathbb{R}^m$ is $I(1)$, the cointegrating rank is r, and we work conditionally on the lagged differences and any deterministic terms.

2 Setup and Notation

2.1 The VECM

The VECM representation of a VAR(p) in levels is

$$\Delta x_t \;=\; \Pi\, x_{t-1} \;+\; \sum_{i=1}^{p-1} \Gamma_i\, \Delta x_{t-i} \;+\; u_t, \qquad u_t \sim N(0, \Sigma_u). \tag{1}$$

Cointegration imposes $\text{rank}(\Pi) = r < m$, factored as $\Pi = \alpha\beta'$ with $\alpha, \beta \in \mathbb{R}^{m\times r}$ both of full column rank.

2.2 Partialling out short-run dynamics (FWL)

Let $W_t = (\Delta x'_{t-1}, \dots, \Delta x'_{t-p+1})'$. Define the FWL residuals

$$\begin{aligned} R_{0t} \;&=\; \Delta x_t \;-\; \hat{\mathbb{E}}[\Delta x_t \mid W_t], \\ R_{1t} \;&=\; x_{t-1} \;-\; \hat{\mathbb{E}}[x_{t-1} \mid W_t], \end{aligned}$$

where $\widehat{\mathbb{E}}[\cdot \mid W_t]$ denotes the linear projection onto the span of W_t (in practice, OLS).

By the Frisch–Waugh–Lovell theorem applied to (Equation 1), the OLS estimator of Π from the original system is identical to the OLS estimator from the residualized regression

$$R_{0t} \;=\; \Pi\, R_{1t} \;+\; u_t, \tag{2}$$

and the residuals u_t are the same. Under Gaussianity, the conditional log-likelihood for Π given the lagged differences factors as the marginal log-likelihood of (Equation 2) plus a Π-free piece. So all subsequent likelihood maximization can be done on (Equation 2) alone — this is what justifies treating the problem as a static regression in R_{0t}, R_{1t}.

2.3 Sample moments

Define

$$S_{ij} \;=\; \frac{1}{T}\sum_{t=1}^{T} R_{it}R'_{jt}, \qquad i,j \in \{0,1\}.$$

These are $m \times m$ matrices; S_{00} and S_{11} are symmetric positive-definite (with probability one, for T large enough), and $S_{10} = S'_{01}$.

3 The Gaussian Log-Likelihood

Stack the residuals as columns, $E = [u_1, \dots, u_T]$. Then $\sum_t u_t u'_t = EE'$. Under $u_t \sim N(0, \Sigma_u)$ i.i.d., the log-likelihood from (Equation 2) is

$$\ell(\Pi, \Sigma_u) \;=\; -\frac{T}{2}\log|\Sigma_u| \;-\; \frac{1}{2}\operatorname{tr}(\Sigma_u^{-1}EE') \;+\; \text{const},$$

where $E = E(\Pi)$ depends on Π through $u_t = R_{0t} - \Pi R_{1t}$.

4 Profiling Out Σ_u

Fix Π and maximize over Σ_u. The derivative of the log-likelihood with respect to Σ_u^{-1} yields

$$\widehat{\Sigma}_u(\Pi) \;=\; \frac{1}{T}E(\Pi)E(\Pi)' \;=\; \frac{1}{T}\sum_{t=1}^{T}(R_{0t} - \Pi R_{1t})(R_{0t} - \Pi R_{1t})'. \tag{3}$$

Plugging this back in, the trace term simplifies: with $\widehat{\Sigma}_u := \widehat{\Sigma}_u(\Pi)$,

$$\operatorname{tr}\Big(\widehat{\Sigma}_u^{-1}EE'\Big) \;=\; \operatorname{tr}\Big((T\widehat{\Sigma}_u)^{-1}(T\widehat{\Sigma}_u)\Big) \;=\; \operatorname{tr}(I_m) \;=\; m,$$

which is constant in Π. (The factor T comes from $EE' = T\widehat{\Sigma}_u$.) So the profiled log-likelihood reduces to

$$\ell(\Pi) \;=\; -\frac{T}{2}\log\Big|\widehat{\Sigma}_u(\Pi)\Big| \;+\; \text{const}.$$

Consequence. The constrained MLE problem is equivalent to

$$\min_{\Pi:\operatorname{rank}(\Pi)=r} \;\Big|\widehat{\Sigma}_u(\Pi)\Big|.$$

This is a determinant minimization in a single matrix parameter Π of restricted rank.

5 Profiling Out α Given β

5.1 Quadratic-in-α form

Expand (Equation 3) using $\Pi = \alpha\beta'$:

$$\widehat{\Sigma}_u(\Pi) \;=\; S_{00} \;-\; \Pi S_{10} \;-\; S_{01}\Pi' \;+\; \Pi S_{11}\Pi'. \tag{4}$$

(Verify by writing $EE'/T = S_{00} - \Pi S_{10} - S_{01}\Pi' + \Pi S_{11}\Pi'$ from the definition of S_{ij}.)

Substitute $\Pi = \alpha\beta'$ and define

$$C \;:=\; S_{01}\beta, \qquad B \;:=\; \beta' S_{11}\beta.$$

Then

$$\widehat{\Sigma}_u(\alpha,\beta) \;=\; S_{00} \;-\; \alpha C' \;-\; C\alpha' \;+\; \alpha B\alpha'. \tag{5}$$

Here B is symmetric positive-definite (since S_{11} is and β has full column rank), so $B^{1/2}$ is well-defined.

5.2 Completing the square

The cross-terms in (Equation 5) are linear in α and the $\alpha B\alpha'$ term is quadratic. Complete the square:

$$\widehat{\Sigma}_u(\alpha,\beta) \;=\; \underbrace{S_{00} - CB^{-1}C'}_{=:\Sigma_*(\beta)} + (\alpha - CB^{-1})\,B\,(\alpha - CB^{-1})'. \tag{6}$$

This is straightforward to verify by expanding the second term: $\alpha B\alpha' - \alpha BB^{-1}C' - CB^{-1}B\alpha' + CB^{-1}BB^{-1}C' = \alpha B\alpha' - \alpha C' - C\alpha' + CB^{-1}C'$, which combines with $S_{00} - CB^{-1}C'$ to recover (Equation 5).

5.3 Why $\hat{\alpha} = CB^{-1}$ minimizes the determinant

Let $L := (\alpha - CB^{-1})B^{1/2}$, an $m \times r$ matrix. Then (Equation 6) reads

$$\widehat{\Sigma}_u(\alpha,\beta) \;=\; \Sigma_*(\beta) + LL'.$$

We need: $|\Sigma_* + LL'| \geq |\Sigma_*|$, with equality iff $L = 0$. This is the matrix-determinant lemma applied to a rank-r update:

$$|\Sigma_* + LL'| \;=\; |\Sigma_*| \cdot \left|I_r + L'\Sigma_*^{-1}L\right|. \tag{7}$$

Since Σ_* is positive-definite (Σ_* is the residual covariance from regressing R_{0t} on $\beta' R_{1t}$, which is positive-definite under non-collinearity), $L'\Sigma_*^{-1}L$ is positive semi-definite, so $|I_r + L'\Sigma_*^{-1}L| \geq 1$ with equality iff $L = 0$, i.e. $\alpha = CB^{-1}$. Therefore

$$\hat{\alpha}(\beta) \;=\; CB^{-1} \;=\; S_{01}\beta\,(\beta' S_{11}\beta)^{-1}. \tag{8}$$

Remark. *Equation (@eq-alphahat) has a transparent interpretation: $\hat{\alpha}(\beta)$ is the OLS coefficient from regressing R_{0t} on the r-dimensional regressor $\beta' R_{1t}$. This is exactly what FWL would tell you to do once β is fixed.*

5.4 The concentrated likelihood in β

Plug $\hat{\alpha}(\beta)$ back into (Equation 6): the second term vanishes, leaving

$$\widehat{\Sigma}_u(\beta) \;=\; \Sigma_*(\beta) \;=\; S_{00} \;-\; S_{01}\beta\,(\beta' S_{11}\beta)^{-1}\beta' S_{10}. \tag{9}$$

The MLE problem now reads

$$\min_{\beta\in\mathbb{R}^{m\times r},\,\beta\,\text{full column rank}} \left|\widehat{\Sigma}_u(\beta)\right|.$$

6 Reducing to a Generalized Eigenvalue Problem

6.1 Two determinant identities

We use:

Lemma (Determinant rewrite)**.** *For $A \in \mathbb{R}^{m\times m}$ invertible and $B \in \mathbb{R}^{m\times m}$,*

$$|A - B| \;=\; |A|\,\left|I - A^{-1}B\right|.$$

Lemma (Sylvester's determinant identity)**.** *For $U \in \mathbb{R}^{m\times r}$ and $V \in \mathbb{R}^{r\times m}$,*

$$|I_m + UV| \;=\; |I_r + VU|.$$

Apply the first lemma to (Equation 9) with $A = S_{00}$, $B = S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}$:

$$\left|\widehat{\Sigma}_u(\beta)\right| \;=\; |S_{00}| \cdot \left|I_m - S_{00}^{-1}S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}\right|.$$

Apply Sylvester's identity with $U = -S_{00}^{-1}S_{01}\beta$ and $V = (\beta' S_{11}\beta)^{-1}\beta' S_{10}$:

$$\left|I_m - S_{00}^{-1}S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}\right| \;=\; \left|I_r - (\beta' S_{11}\beta)^{-1}\beta' S_{10}S_{00}^{-1}S_{01}\beta\right|.$$

Combining,

$$\left|\widehat{\Sigma}_u(\beta)\right| \;=\; |S_{00}| \cdot \left|I_r - (\beta' S_{11}\beta)^{-1}\,\beta' S_{10}S_{00}^{-1}S_{01}\beta\right|. \tag{10}$$

6.2 Normalization and the reduced problem

The decomposition $\Pi = \alpha\beta'$ is identified only up to right multiplication of β by an invertible $r \times r$ matrix (since $\Pi = (\alpha M^{-1})(\beta M)' = \tilde{\alpha}\tilde{\beta}'$ for any $M \in GL_r$). Impose the normalization

$$\beta' S_{11}\beta \;=\; I_r. \tag{11}$$

Under this normalization, (Equation 10) collapses to

$$\left|\widehat{\Sigma}_u(\beta)\right| \;=\; |S_{00}| \cdot \left|I_r - \beta' A\beta\right|, \qquad A \;:=\; S_{10}S_{00}^{-1}S_{01}.$$

Note A is symmetric positive semi-definite. Minimizing $|\widehat{\Sigma}_u(\beta)|$ is equivalent (since $|S_{00}|$ is constant) to *maximizing* $|\beta' A\beta|$ subject to (Equation 11). This is the canonical-correlation extremal problem:

$$\boxed{\max_{\beta\in\mathbb{R}^{m\times r}} \;|\beta' A\beta| \quad \text{s.t.} \quad \beta' S_{11}\beta = I_r.} \tag{12}$$

6.3 First-order condition

Write the Lagrangian for $\log|\beta' A\beta|$ subject to the normalization, with multiplier matrix $\Lambda \in \mathbb{R}^{r\times r}$:

$$\mathcal{L}(\beta,\Lambda) \;=\; \log\bigl|\beta' A\beta\bigr| \;-\; \mathrm{tr}\Bigl(\Lambda\,(\beta' S_{11}\beta - I_r)\Bigr).$$

Differentiating $\log|\beta' A\beta|$: by Jacobi's formula, $d\log|M| = \mathrm{tr}(M^{-1}\,dM)$ for any invertible M. With $M(\beta) = \beta' A\beta$,

$$dM \;=\; (d\beta)' A\beta \;+\; \beta' A\,(d\beta).$$

By cyclicity of the trace and symmetry of A,

$$\mathrm{tr}\bigl(M^{-1} dM\bigr) \;=\; 2\;\mathrm{tr}\bigl(M^{-1}\beta' A\,(d\beta)\bigr) \;=\; 2\;\mathrm{tr}\bigl(\,A\beta M^{-1}\,(d\beta)'\,\bigr)^{\top}.$$

Differentiating the constraint similarly, the FOC reads

$$A\beta\,(\beta' A\beta)^{-1} \;=\; S_{11}\beta\,\Lambda$$

(up to a factor of 2 absorbed into Λ). Right-multiplying by $\beta' A\beta$ gives

$$A\beta \;=\; S_{11}\beta\,\Gamma, \qquad \Gamma \;:=\; \Lambda\,(\beta' A\beta). \tag{13}$$

6.4 Diagonalization and the generalized eigenvalue problem

The FOC (Equation 13) can be rotated. Replace β by βP for any $P \in GL_r$: then $A(\beta P) = S_{11}(\beta P)(P^{-1}\Gamma P)$, so we may pick P to diagonalize Γ. Choose P so that $P^{-1}\Gamma P = \mathrm{diag}(\lambda_1,\dots,\lambda_r)$, and write the columns of βP as $v_1,\dots,v_r$. Then (Equation 13) reads, column by column,

$$Av_i \;=\; \lambda_i\,S_{11}v_i, \qquad i = 1,\dots,r,$$

i.e.,

$$\boxed{S_{10}S_{00}^{-1}S_{01}\,v \;=\; \lambda\,S_{11}\,v.} \tag{14}$$

This is the **generalized eigenvalue problem** at the heart of Johansen's method. Equivalently, multiplying by S_{11}^{-1}, the eigenvalues are those of $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$.

Order the eigenvalues $1 \ge \hat\lambda_1 \ge \hat\lambda_2 \ge \cdots \ge \hat\lambda_m \ge 0$ and let $v_1,\dots,v_m$ be the corresponding eigenvectors normalized so $v_i' S_{11} v_j = \delta_{ij}$.

6.5 Picking the maximizer

Under (Equation 11), $\beta' A\beta$ has eigenvalues equal to a subset of $\hat\lambda_1,\dots,\hat\lambda_m$ (whichever eigenvectors β spans). To *maximize* $|\beta' A\beta| = \prod_{i=1}^{r}\mu_i$ (with μ_i the eigenvalues of $\beta' A\beta$), pick the largest r eigenvalues:

$$\hat\beta_r \;=\; [v_1, v_2, \dots, v_r], \qquad \hat\beta_r' A\hat\beta_r \;=\; \mathrm{diag}(\hat\lambda_1,\dots,\hat\lambda_r).$$

The corresponding $\hat\alpha$ comes from (Equation 8): $\hat\alpha_r = S_{01}\hat\beta_r$ (the normalization makes the $(\beta' S_{11}\beta)^{-1}$ factor disappear).

6.6 Canonical correlation analysis: the basic eigenproblem

Before connecting to Johansen, we derive the CCA eigenproblem from scratch.

Setup. Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be random vectors with $\Sigma_{XX} = \mathrm{Var}(X)$, $\Sigma_{YY} = \mathrm{Var}(Y)$ (both positive definite), and $\Sigma_{XY} = \mathrm{Cov}(X, Y)$. The first canonical correlation is the largest correlation achievable between linear combinations of X and Y:

$$\rho_1 \;=\; \max_{a,\,b} \; \frac{a'\Sigma_{XY}b}{\sqrt{a'\Sigma_{XX}\,a}\,\sqrt{b'\Sigma_{YY}b}}.$$

The objective is invariant to rescaling a and b, so the program is equivalent to

$$\max_{a,\,b} \; a'\Sigma_{XY}b \quad \text{s.t.} \quad a'\Sigma_{XX}\,a = 1, \;\; b'\Sigma_{YY}b = 1. \tag{15}$$

Lagrangian and FOCs. With multipliers μ, ν,

$$\mathcal{L}(a, b, \mu, \nu) \;=\; a'\Sigma_{XY}b \;-\; \tfrac{\mu}{2}\big(a'\Sigma_{XX}\,a - 1\big) \;-\; \tfrac{\nu}{2}\big(b'\Sigma_{YY}b - 1\big).$$

Setting $\partial\mathcal{L}/\partial a = 0$ and $\partial\mathcal{L}/\partial b = 0$ gives

$$\Sigma_{XY}b \;=\; \mu\,\Sigma_{XX}\,a, \qquad \Sigma_{YX}\,a \;=\; \nu\,\Sigma_{YY}b. \tag{16}$$

Multipliers equal the correlation. Pre-multiplying the first FOC by a' and the second by b', and using the constraints $a'\Sigma_{XX}a = b'\Sigma_{YY}b = 1$:

$$a'\Sigma_{XY}b \;=\; \mu, \qquad b'\Sigma_{YX}\,a \;=\; \nu.$$

The two left-hand sides are equal (a scalar equals its transpose), so $\mu = \nu$. Call this common value ρ — it is precisely the correlation we are maximizing.

Eliminate a. From the first FOC, $a = \rho^{-1}\Sigma_{XX}^{-1}\Sigma_{XY}b$. Substituting into the second FOC,

$$\Sigma_{YX}\,\rho^{-1}\Sigma_{XX}^{-1}\Sigma_{XY}b \;=\; \rho\,\Sigma_{YY}b \quad\Longleftrightarrow\quad \Sigma_{YY}^{-1}\Sigma_{YX}\,\Sigma_{XX}^{-1}\,\Sigma_{XY}b \;=\; \rho^2\,b. \tag{17}$$

Conclusion. The squared canonical correlations ρ_i^2 are the eigenvalues of $\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}$, with the canonical Y-directions b_i as the corresponding eigenvectors; the matching a_i are recovered from (Equation 16). Subsequent canonical pairs are obtained by adding orthogonality constraints $\mathrm{Cov}(a_i'X, a_j'X) = \mathrm{Cov}(b_i'Y, b_j'Y) = 0$ for $i \neq j$, which leads to the same eigenvalue problem; the remaining ρ_i^2 are simply the next-largest eigenvalues.

6.7 Connection to canonical correlations

The eigenvalues $\hat{\lambda}_i$ have a clean statistical interpretation: they are the squared sample canonical correlations between R_{0t} and R_{1t}. We show this in three steps.

Step 1: the CCA eigenproblem. From the previous subsection (equation Equation 17), the squared canonical correlations ρ_i^2 between random vectors X and Y are the eigenvalues of

$$\Sigma_{YY}^{-1}\,\Sigma_{YX}\,\Sigma_{XX}^{-1}\,\Sigma_{XY}. \tag{18}$$

Step 2: match the sample moments. Set $X = R_{0t}$ and $Y = R_{1t}$. The sample moments $S_{ij} = T^{-1}\sum_t R_{it}R_{jt}'$ are the empirical analogs of Σ_{ij}, with the obvious correspondences $S_{00} \leftrightarrow \Sigma_{XX}$,

$S_{11} \leftrightarrow \Sigma_{YY}$, $S_{01} \leftrightarrow \Sigma_{XY}$, $S_{10} \leftrightarrow \Sigma_{YX}$. Substituting into (Equation 18), the sample squared canonical correlations $\hat{\rho}_i^2$ are the eigenvalues of

$$S_{11}^{-1}\, S_{10}\, S_{00}^{-1}\, S_{01}.$$

Step 3: identify with the Johansen eigenvalues. Multiplying (Equation 14) through by S_{11}^{-1} gives

$$S_{11}^{-1}\, S_{10}\, S_{00}^{-1}\, S_{01}\, v \;=\; \lambda\, v,$$

which is identical to the eigenproblem in Step 2 with $\lambda \equiv \rho^2$. Hence

$$\hat{\lambda}_i \;=\; \hat{\rho}_i^2, \qquad i = 1, \ldots, m.$$

Implications.

- $0 \le \hat{\lambda}_i \le 1$, since squared correlations are bounded by one.
- $\hat{\lambda}_i \approx 0$: no linear long-run relation in direction i (the i-th canonical pair is essentially uncorrelated).
- $\hat{\lambda}_i$ bounded away from 0: a genuine cointegrating relation exists in direction i.

Why a large $\hat{\lambda}_i$ implies cointegration. Recall that in the sample,

$$\hat{\lambda}_i \;=\; \max_{a,\,b}\; \frac{(a' S_{01}\, b)^2}{(a' S_{00}\, a)\,(b' S_{11}\, b)}.$$

Suppose b is a *non*-cointegrating direction, so $b' R_{1t}$ remains $I(1)$. The three sample moments scale as

- $a' S_{00}\, a \;=\; O_p(1)$ — sample variance of a stationary series, converging in probability to $a' \Sigma_{00}\, a$.
- $b' S_{11}\, b \;=\; O_p(T)$ — for $I(1)$ data, $\mathrm{Var}(b' R_{1t}) = O(t)$, so the sample average $T^{-1} \sum_t (b' R_{1t})^2 = O_p(T)$.
- $a' S_{01}\, b \;=\; O_p(1)$ — the only non-obvious rate; we derive it next.

The cross-moment rate. Take the canonical scalar example: let $u_t \sim$ i.i.d.$(0, \sigma_u^2)$ play the role of $a' R_{0t}$, and let $z_t = z_0 + \sum_{s=1}^{t} \varepsilon_s$ with $\varepsilon_s \sim$ i.i.d.$(0, \sigma_\varepsilon^2)$ play the role of $b' R_{1t}$, with u_t uncorrelated with past ε. Define $C_T := \sum_{t=1}^{T} u_t\, z_{t-1}$. Then:

- Each summand has mean zero (since $u_t \perp z_{t-1}$) and $\mathrm{Var}(u_t z_{t-1}) = \sigma_u^2 \cdot \mathrm{Var}(z_{t-1}) = \sigma_u^2\, \sigma_\varepsilon^2\, (t - 1) = O(t)$.
- Cross-terms vanish: for $s < t$, $E[u_s z_{s-1}\, u_t z_{t-1}] = E[u_s z_{s-1}\, z_{t-1}\, E(u_t \mid \mathcal{F}_{t-1})] = 0$ by the martingale-difference property of u_t.
- Hence $\mathrm{Var}(C_T) = \sum_{t=1}^{T} O(t) = O(T^2)$, so $C_T = O_p(T)$ and $T^{-1} C_T = O_p(1)$.

More generally, by the functional central limit theorem,

$$T^{-1} C_T \;\xrightarrow{d}\; \sigma_u\, \sigma_\varepsilon \int_0^1 W(r)\, dB(r),$$

a stochastic integral that is non-degenerate but bounded in probability — confirming $a' S_{01}\, b = O_p(1)$.

Combining the three rates,

$$\hat{\lambda}_i \;\le\; \frac{(a' S_{01}\, b)^2}{(a' S_{00}\, a)\,(b' S_{11}\, b)} \;=\; \frac{O_p(1)}{O_p(1)\cdot O_p(T)} \;=\; O_p(T^{-1}) \;\xrightarrow{p}\; 0$$

along every non-cointegrating direction b. The only way to obtain a non-vanishing canonical correlation is to find a direction β_i such that $\beta_i' R_{1t}$ is itself stationary — equivalently, $\beta_i' x_{t-1}$ is $I(0)$, i.e., β_i is a cointegrating vector. In that case $\beta_i' S_{11}\,\beta_i = O_p(1)$ instead of $O_p(T)$, and $\hat{\lambda}_i$ stays bounded away from zero. Hence

$$\hat{\lambda}_i \overset{p}{\not\to} 0 \quad\Longleftrightarrow\quad \exists\,\beta_i :\; \beta_i' x_{t-1} \sim I(0).$$

Equivalently — and this is the error-correction reading — a nontrivial $\hat{\lambda}_i$ says that some linear combination $\beta_i' x_{t-1}$ predicts Δx_t, and a stationary predictor of a stationary outcome can only persist when $\beta_i' x_{t-1}$ is itself stationary, i.e., when the system has a long-run equilibrium that disequilibria correct toward.

Rank–eigenvalue link. The previous paragraph showed that each non-zero $\hat{\lambda}_i$ flags one cointegrating direction. To convert that into a *count* of cointegrating relations, link the eigenvalue count to rank(Π) via three observations.

1. Eigenvalues to rank of M. For a square matrix, the number of non-zero eigenvalues equals its rank. Hence

$$\#\{\text{non-zero eigenvalues of } M\} \;=\; \text{rank}(M).$$

2. Rank of M equals rank of Π. The OLS estimand from the (residualized) long-run regression is $\Pi = \Sigma_{01}\Sigma_{11}^{-1}$, so $\text{rank}(\Pi) = \text{rank}(\Sigma_{01})$ (since Σ_{11} is invertible). The CCA eigenproblem matrix is $M := \Sigma_{11}^{-1}\Sigma_{10}\Sigma_{00}^{-1}\Sigma_{01}$; both Σ_{00} and Σ_{11} are invertible, so $\text{rank}(M) = \text{rank}(\Sigma_{10}\,\Sigma_{00}^{-1}\,\Sigma_{01})$, and this equals $\text{rank}(\Sigma_{01})$ via the Gram-matrix identity $\Sigma_{10}\,\Sigma_{00}^{-1}\,\Sigma_{01} = (\Sigma_{00}^{-1/2}\Sigma_{01})'(\Sigma_{00}^{-1/2}\Sigma_{01})$ (a Gram matrix has the rank of its factor). Therefore $\text{rank}(M) = \text{rank}(\Pi)$.

3. Rank of Π *is the cointegrating rank.* By the VECM construction, $\text{rank}(\Pi) = r$ is the number of independent linear combinations $\beta_i' x_t$ that are stationary — i.e., the number of cointegrating relations.

Chaining the three steps,

$$\#\{\text{non-zero eigenvalues of } M\} \;=\; \text{rank}(M) \;=\; \text{rank}(\Pi) \;=\; \#\text{ cointegrating relations.}$$

The sample matrix $\widehat{M} = S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$ inherits this asymptotically: r of its eigenvalues remain $O_p(1)$ and the other $m - r$ are $O_p(T^{-1})$.

Rank dictionary. Combining the rank–eigenvalue link with the rate calculation, the relationship between $\text{rank}(\Pi) = r$ and the asymptotic behavior of $\hat{\lambda}_i$ is

- $r = 0$ ($\Pi = 0$, no cointegration, x_t purely $I(1)$): **all** m eigenvalues are $O_p(T^{-1})$ and vanish in probability — every direction b leaves $b' x_{t-1}$ as $I(1)$, so the variance mismatch applies uniformly;
- $0 < r < m$ (partial cointegration): exactly r eigenvalues are $O_p(1)$ — those associated with the r cointegrating directions — and the remaining $m - r$ are $O_p(T^{-1}) \to 0$;
- $r = m$ (x_t itself stationary): all S_{ij} are $O_p(1)$, so every $\hat{\lambda}_i$ is $O_p(1)$ and bounded away from zero.

In short, the **count** of non-vanishing eigenvalues equals the cointegrating rank.

Note, however, that at any finite T sampling noise keeps every $\hat{\lambda}_i$ strictly positive even when the population eigenvalue is zero (because $S_{01} = O_p(1)$ rather than exactly zero). The rank question therefore becomes a hypothesis-testing question: *how many $\hat{\lambda}_i$ are too large to be explained by sampling variation alone?* The trace and maximum-eigenvalue statistics derived below formalize precisely this.

The Johansen procedure is therefore canonical correlation analysis between residualized differences and residualized levels, and the rank tests below ask: *how many canonical correlations are significantly nonzero?*

7 Determinants in Eigenvalue Form

Two determinants enter the LR statistic: the unrestricted $|\widehat{\Sigma}_m|$ (full-rank Π) and the rank-r restricted $|\widehat{\Sigma}_r|$.

7.1 Unrestricted residual covariance

For full-rank Π, the unconstrained MLE is OLS: $\widehat{\Pi} = S_{01}S_{11}^{-1}$. Plugging into (Equation 4),

$$\widehat{\Sigma}_m \;=\; S_{00} \;-\; S_{01}S_{11}^{-1}S_{10} \;-\; S_{01}S_{11}^{-1}S_{10} \;+\; S_{01}S_{11}^{-1}S_{11}S_{11}^{-1}S_{10} \;=\; S_{00} - S_{01}S_{11}^{-1}S_{10}.$$

Compute $|\widehat{\Sigma}_m|$ in eigenvalue form. Pre- and post-multiply by $S_{00}^{-1/2}$:

$$S_{00}^{-1/2}\,\widehat{\Sigma}_m\,S_{00}^{-1/2} \;=\; I_m \;-\; S_{00}^{-1/2}S_{01}S_{11}^{-1}S_{10}S_{00}^{-1/2} \;=\; I_m - CC',$$

where $C = S_{00}^{-1/2}S_{01}S_{11}^{-1/2}$ as in the remark above. The eigenvalues of CC' equal those of $C'C = S_{11}^{-1/2}S_{10}S_{00}^{-1}S_{01}S_{11}^{-1/2}$, and this matrix is similar to $S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}$, whose eigenvalues are the Johansen eigenvalues $\hat{\lambda}_1, \dots, \hat{\lambda}_m$. Therefore

$$|I_m - CC'| \;=\; \prod_{i=1}^{m}(1 - \hat{\lambda}_i),$$

and taking determinants of both sides of the similarity transform,

$$\boxed{|\widehat{\Sigma}_m| \;=\; |S_{00}|\prod_{i=1}^{m}(1 - \hat{\lambda}_i).} \tag{19}$$

7.2 Rank-r restricted residual covariance

From (Equation 9) and the normalization (Equation 11),

$$\widehat{\Sigma}_r \;=\; S_{00} \;-\; S_{01}\hat{\beta}_r\hat{\beta}_r'S_{10}.$$

Pre- and post-multiply by $S_{00}^{-1/2}$:

$$S_{00}^{-1/2}\,\widehat{\Sigma}_r\,S_{00}^{-1/2} \;=\; I_m \;-\; C\,P_r\,C', \qquad P_r \;:=\; S_{11}^{1/2}\,\hat{\beta}_r\hat{\beta}_r'\,S_{11}^{1/2},$$

where the factorization absorbs the $S_{11}^{1/2}$ pieces into a clean projector. Specifically, set $W_r := S_{11}^{1/2}\hat{\beta}_r$, an $m \times r$ matrix. The normalization $\hat{\beta}_r' S_{11} \hat{\beta}_r = I_r$ gives $W_r' W_r = I_r$, so

$$P_r \;=\; W_r W_r'$$

is the *orthogonal projector* onto $\mathrm{span}(W_r) \subset \mathbb{R}^m$.

P_r projects onto the top-r eigenspace of $C'C$. From (Equation 14), the eigenvectors v_i of the generalized problem satisfy $C'C\,w_i = \hat{\lambda}_i\,w_i$ where $w_i = S_{11}^{1/2} v_i$ (verify: $S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}(S_{11}^{1/2} v_i) = S_{11}^{-1/2} A v_i = S_{11}^{-1/2} \cdot \hat{\lambda}_i S_{11} v_i = \hat{\lambda}_i (S_{11}^{1/2} v_i)$). The columns of W_r are exactly $w_1, \dots, w_r$, the top-r eigenvectors of $C'C$.

Eigenvalues of CP_rC'. For each w_i with $\hat{\lambda}_i > 0$, the vector Cw_i is an eigenvector of CC' with the same eigenvalue $\hat{\lambda}_i$. Now apply CP_rC' to Cw_i:

- If $i \le r$: $w_i \in \mathrm{span}(W_r)$, so $P_r w_i = w_i$, hence $CP_rC'(Cw_i) = CP_r(C'C)w_i = \hat{\lambda}_i\,Cw_i$.
- If $i > r$: $w_i \perp \mathrm{span}(W_r)$ (eigenvectors of distinct eigenvalues are orthogonal under the symmetric problem $C'Cw = \lambda w$), so $P_r w_i = 0$, hence $CP_rC'(Cw_i) = 0$.

The non-zero eigenvalues of CP_rC' are therefore exactly $\hat{\lambda}_1, \dots, \hat{\lambda}_r$, and the remaining $m - r$ eigenvalues are zero. Hence the eigenvalues of $I_m - CP_rC'$ are $1 - \hat{\lambda}_1, \dots, 1 - \hat{\lambda}_r, 1, \dots, 1$, and

$$|I_m - CP_rC'| \;=\; \prod_{i=1}^{r}(1 - \hat{\lambda}_i).$$

Putting it together,

$$\boxed{|\hat{\Sigma}_r| \;=\; |S_{00}| \prod_{i=1}^{r}(1 - \hat{\lambda}_i).} \tag{20}$$

8 The LR Rank Tests

8.1 The likelihood ratio

The LR statistic for $H_0 : \mathrm{rank}(\Pi) \le r$ against the unrestricted alternative is $\mathrm{LR}(r) = -2(\ell_r - \ell_m)$, where ℓ_r and ℓ_m are the maximized log-likelihoods under the rank-r restriction and the unrestricted model, respectively. Substituting the profiled log-likelihood $\ell(\Pi) = -\frac{T}{2}\log|\hat{\Sigma}_u(\Pi)| + \mathrm{const}$ from Section 4 — with $\hat{\Sigma}_r := \hat{\Sigma}_u(\hat{\Pi}_r)$ and $\hat{\Sigma}_m := \hat{\Sigma}_u(\hat{\Pi}_{\mathrm{OLS}})$ — the additive constant cancels, leaving

$$\mathrm{LR}(r) \;=\; -2(\ell_r - \ell_m) \;=\; -2\left(-\frac{T}{2}\right)(\log|\hat{\Sigma}_r| - \log|\hat{\Sigma}_m|) \;=\; T\log\frac{|\hat{\Sigma}_r|}{|\hat{\Sigma}_m|} \;=\; -T\log\frac{|\hat{\Sigma}_m|}{|\hat{\Sigma}_r|}.$$

Combining (Equation 19) and (Equation 20), the $|S_{00}|$ factors cancel:

$$\frac{|\hat{\Sigma}_m|}{|\hat{\Sigma}_r|} \;=\; \prod_{i=r+1}^{m}(1 - \hat{\lambda}_i),$$

so

$$\mathrm{LR}(r) \;=\; -T\sum_{i=r+1}^{m}\log(1 - \hat{\lambda}_i).$$

8.2 Trace test

Compares rank r to the unrestricted ($r = m$) model:

$$\mathrm{LR}_{\text{trace}}(r) \;=\; -T \sum_{i=r+1}^{m} \log(1 - \hat{\lambda}_i).$$

Tests $H_0 : \operatorname{rank}(\Pi) \le r$ against $H_1 : \operatorname{rank}(\Pi) > r$. Under H_0, all of $\hat{\lambda}_{r+1}, \dots, \hat{\lambda}_m$ should be near zero in population.

8.3 Maximum eigenvalue test

Compares rank r to rank $r + 1$:

$$\mathrm{LR}_{\max}(r, r+1) \;=\; -T \log(1 - \hat{\lambda}_{r+1}).$$

Tests $H_0 : \operatorname{rank}(\Pi) = r$ against $H_1 : \operatorname{rank}(\Pi) = r + 1$.

Both statistics have non-standard limiting distributions (functionals of vector Brownian motion); critical values are tabulated in Johansen (1995) and depend on the deterministic specification (constant, trend, restricted constant, etc.).

9 Summary of the Full Argument

The four-move structure can be written as a single chain of equivalences:

$$\begin{aligned}
\max_{\Pi:\operatorname{rank}\Pi = r} \ell(\Pi, \Sigma_u) &\overset{\text{profile } \Sigma_u}{\iff} \min_{\operatorname{rank}\Pi = r} |\widehat{\Sigma}_u(\Pi)| \\
&\overset{\Pi = \alpha\beta'}{\iff} \min_{\alpha,\beta} |\widehat{\Sigma}_u(\alpha, \beta)| \\
&\overset{\text{profile } \alpha}{\iff} \min_{\beta} |\widehat{\Sigma}_u(\beta)| \\
&\overset{\text{determinant identities}}{\iff} \max_{\beta:\beta' S_{11} \beta = I_r} |\beta' A \beta| \\
&\overset{\text{Lagrangian + Jacobi}}{\iff} \text{top-}r \text{ generalized eigenvectors of } S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}.
\end{aligned}$$

The eigenvalues $\hat{\lambda}_i$ produced by the eigenvalue problem parameterize every quantity of interest: $\hat{\beta}_r$ is the top-r eigenvectors, $\widehat{\Sigma}_r$ is $|S_{00}|$ times the product of $(1 - \hat{\lambda}_i)$ for $i \le r$, and the LR rank statistics are sums of $\log(1 - \hat{\lambda}_i)$.

10 References (selected)

- Johansen, S. (1988). "Statistical analysis of cointegration vectors." *Journal of Economic Dynamics and Control* 12, 231–254.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer. (Chapter 6 on cointegration; clear treatment of the FWL step and the canonical-correlation interpretation.)