STA 542: Introduction to Time Series Analysis

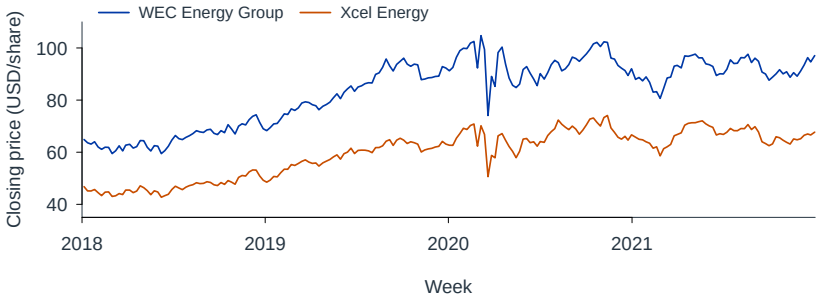
Lecture 12: From a pairs trade to matrix filters

Lasse Vuursteen

Duke
UNIVERSITY

Can relative prices suggest a trade?

Homework 5, Problem 2: WEC Energy Group and Xcel Energy.



Can a position use their price relation without forecasting either stock separately?

Real data: weekly Yahoo Finance Close prices in the Homework 5 extract; split adjusted, excluding dividends.

The gap measures departure from a price relation

Let W_t be the WEC price and X_t the Xcel price, in dollars per share. The supplied relation is

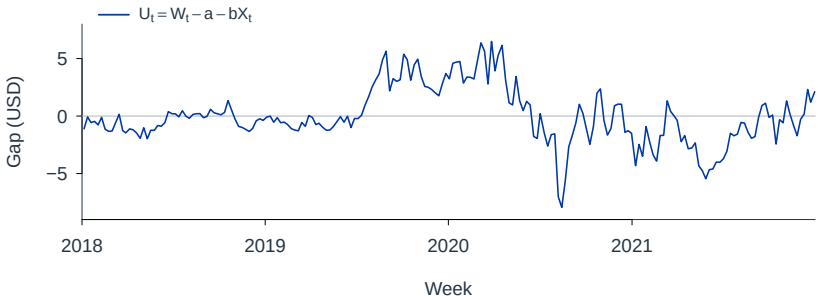
$$W_t = a + bX_t + U_t, \qquad a = 1.146277, \quad b = 1.385665.$$

At an observed pair (w_t, x_t), the gap is $u_t = w_t - a - bx_t$.

A positive gap places WEC above the price implied by its relation with Xcel.

Coefficients fitted on 209 weekly observations, 2018–2021; held fixed throughout the exercise.

The level relation produces an observable gap



The zero line is the supplied relation $w_t = a + bx_t$.

What additional assumption gives a forecast of the gap?

Real data: weekly Yahoo Finance Close prices in the Homework 5 extract; split adjusted, excluding dividends.

A stationary relation alone does not specify its forecast

The exercise also assumes the stationary gap satisfies

$$U_{t+1} = \phi U_t + Z_{t+1}, \qquad \phi = 0.876663.$$

For every t, Z_{t+1} has mean zero, finite positive variance, and is independent of the joint price history $\mathcal{H}_t = ((W_s, X_s) : s \leq t)$.

These assumptions determine how the current gap predicts a later gap.

Iterate the gap equation before taking expectations

For an integer $h \geq 1$,

$$U_{t+h} = \underbrace{\phi^h U_t}_{\text{known at time } t} + \underbrace{\sum_{j=1}^{h} \phi^{h-j} Z_{t+j}}_{\text{future innovations}} .$$

What remains after conditioning on $\mathcal{H}_t$?

Iterate the gap equation before taking expectations

For an integer $h \geq 1$,

$$U_{t+h} = \underbrace{\phi^h U_t}_{\text{known at time } t} + \underbrace{\sum_{j=1}^{h} \phi^{h-j} Z_{t+j}}_{\text{future innovations}} .$$

What remains after conditioning on $\mathcal{H}_t$?

The future terms have conditional mean zero, so $\mathbb{E}[U_{t+h} \mid \mathcal{H}_t] = \phi^h U_t$.

The expected change points toward zero

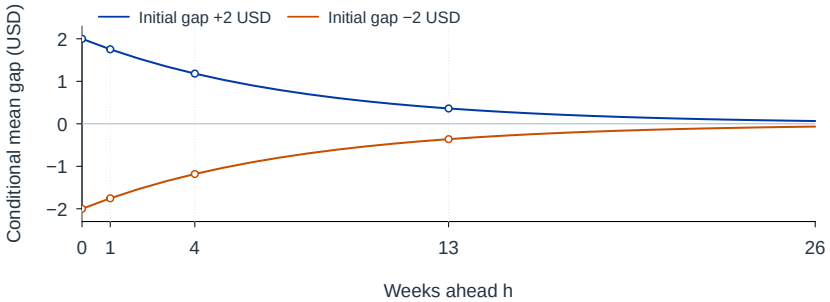
Subtract the current gap from its forecast:

$$\mathbb{E}[U_{t+h} - U_t \mid \mathcal{H}_t] = (\phi^h - 1)U_t.$$

Since $0 < \phi < 1$, the expected change has the opposite sign to U_t.

The fraction of the current gap predicted to close is $1 - \phi^h$.

How much of the gap is predicted to close?



At $h = 1, 4, 13$ weeks, the fractions are 12.33%, 40.93%, and 81.94%.

Calculated conditional means for initial gaps $+2$ and -2 dollars, with $\phi = 0.876663$.

Which conclusion came from which assumption?

Cointegration supplies a stationary linear combination of the price levels.

The AR(1) equation and the innovation assumption give the conditional forecast $\phi^h U_t$ and its rate of mean reversion.

There is still no model here for the future common price movement.

A short position gains when its price falls

Holding q shares from time t to $t + h$ earns the price gain

$$q(p_{t+h} - p_t).$$

Positive q means buying shares; negative q means selling borrowed shares.

The exercise permits fractional shares and holds each position fixed until exit.

Which side of the pair should we buy?

Suppose $u_t > 0$. WEC is above the supplied relation, and the model predicts a decrease in the gap.

Which stock should be bought, and which sold short?

Which side of the pair should we buy?

Suppose $u_t > 0$. WEC is above the supplied relation, and the model predicts a decrease in the gap.

Which stock should be bought, and which sold short?

Short WEC and buy Xcel. For $u_t < 0$, reverse the directions.

The share counts must still be chosen to cancel the common movement.

Write the gain for arbitrary holdings

For fixed signed quantities q_W, q_X, the price gain is

$$G_{t,h} = q_W(W_{t+h} - W_t) + q_X(X_{t+h} - X_t).$$

The relation $W_t = a + bX_t + U_t$ gives $W_{t+h} - W_t = b(X_{t+h} - X_t) + (U_{t+h} - U_t)$.

Substitute this identity to separate the common movement from the gap change.

Cancel the movement for which no forecast was supplied

$$G_{t,h} = \underbrace{(bq_W + q_X)(X_{t+h} - X_t)}_{\text{unspecified common movement}} + \underbrace{q_W(U_{t+h} - U_t)}_{\text{gap change}}.$$

Which choice makes the first term zero for every possible change in Xcel?

Cancel the movement for which no forecast was supplied

$$G_{t,h} = \underbrace{(bq_W + q_X)(X_{t+h} - X_t)}_{\text{unspecified common movement}} + \underbrace{q_W(U_{t+h} - U_t)}_{\text{gap change}}.$$

Which choice makes the first term zero for every possible change in Xcel?

Choose $q_X = -bq_W$. Then $G_{t,h} = q_W(U_{t+h} - U_t)$.

Choose the direction using the expected gap change

Under the hedge $q_X = -bq_W$,

$$\mathbb{E}[G_{t,h} \mid \mathcal{H}_t] = q_W(\phi^h - 1)U_t.$$

For $u_t \neq 0$, maximizing this quantity subject to $|q_W| = 100$ gives $q_W = -100\,\mathrm{sign}(u_t)$.

At $u_t = 0$, the exercise prescribes no position.

The gap determines both share positions

Since $100b = 138.5665$, the rule is

Entry gap	WEC shares	Xcel shares
$u_t > 0$	-100	$+138.5665$
$u_t < 0$	$+100$	-138.5665
$u_t = 0$	0	0

The ratio comes from cancelling the shared component, not from matching the dollar values of the two stock positions.

A larger gap produces a larger predicted price gain

For the chosen direction,

$$\mathbb{E}[G_{t,h} \mid \mathcal{H}_t] = 100(1 - \phi^h)|U_t|.$$

At $h = 13$, the predicted gain is about 81.94 dollars per dollar of absolute entry gap.

This is a conditional expectation; the realized change in the gap remains random.

The cash account finances the initial position

Write K_t for the initial cash balance. Choose

$$K_t = -q_W w_t - q_X x_t = -q_W(a + u_t).$$

Positive K_t is cash held; negative K_t is borrowing. At zero interest, the terminal stock-and-cash value equals $G_{t,h}$.

The initial net value is zero, so it cannot be used as a denominator for a return.

Fix the rule before using the later observations

Keep a, b, ϕ fixed at their training values. In the evaluation record,

$$t = 0, 13, 26, 39, 52, 65, 78, 91 \quad \longrightarrow \quad \text{exit at } t + 13.$$

Time zero is December 31, 2021; the later 104 weekly observations cover 2022–2023.

Use each entry's observed prices and gap. Close the old position before opening the next.

Exercise assumptions: execution at the displayed prices; price gains only, excluding dividends, interest, and trading costs.

Work the first trade from its entry prices

On December 31, 2021, $w_0 = 97.07$ and $x_0 = 67.70$. Using the unrounded recorded prices,

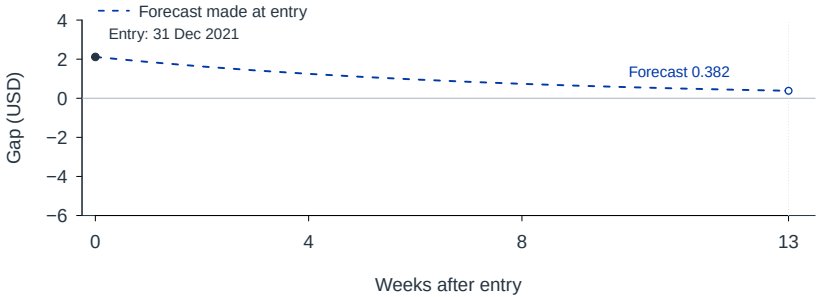
$$u_0 = w_0 - a - bx_0 \approx 2.114, \qquad q_W = -100, \quad q_X = 138.5665.$$

The cash balance is \$326.05 and the predicted 13-week price gain is \$173.23.

What gap does the model predict at the exit date?

Real data: weekly Yahoo Finance Close prices in the Homework 5 extract; split adjusted, excluding dividends.

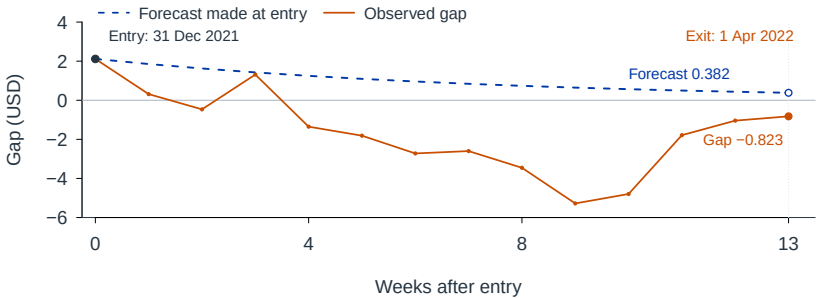
The first trade begins with a forecast toward zero



The forecast exit gap is $\phi^{13} u_0 \approx 0.382$.

Entry observed December 31, 2021; later values shown here are model conditional means.

The observed gap crosses zero during the first trade



At the prescribed exit, $u_{13} \approx -0.823$, giving $G_{0,13} = -100(u_{13} - u_0) = \$293.76$.

Real data: weekly Yahoo Finance Close prices in the Homework 5 extract; split adjusted, excluding dividends.

Both stock prices rose, yet the hedged position gained

At exit, WEC was \$101.13 and Xcel was \$72.75. The price gain, using the displayed rounded prices, is approximately

$$\underbrace{-100(101.13 - 97.07)}_{-\$406.00} + \underbrace{138.5665(72.75 - 67.70)}_{+\$699.76} = \$293.76.$$

The long Xcel position gained more than the short WEC position lost.

Real data: weekly Yahoo Finance Close prices in the Homework 5 extract; split adjusted, excluding dividends.

The first four trades include a loss

Each row uses the gap observed at that entry and the same fixed rule.

Trade	Entry–exit	Gap	WEC	Xcel	Predicted	Realized
1	31 Dec 21–01 Apr 22	2.114	−100	138.5665	173.23	293.76
2	01 Apr 22–01 Jul 22	−0.823	100	−138.5665	67.47	161.63
3	01 Jul 22–30 Sep 22	0.793	−100	138.5665	64.97	119.17
4	30 Sep 22–30 Dec 22	−0.399	100	−138.5665	32.68	−413.64

Real-data evaluation; holdings are shares, gaps are dollars, and gains are USD. Values rounded only for display.

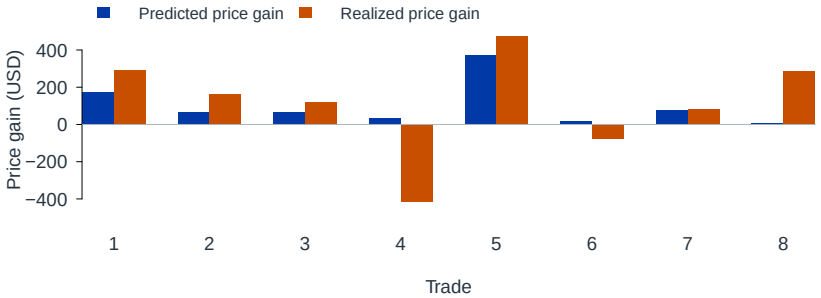
The remaining trades use newly observed entry gaps

The coefficients and 13-week holding period remain fixed.

Trade	Entry–exit	Gap	WEC	Xcel	Predicted	Realized
5	30 Dec 22–31 Mar 23	−4.535	100	−138.5665	371.60	472.97
6	31 Mar 23–30 Jun 23	0.194	−100	138.5665	15.93	−75.25
7	30 Jun 23–29 Sep 23	0.947	−100	138.5665	77.59	83.10
8	29 Sep 23–29 Dec 23	0.116	−100	138.5665	9.50	287.88

Real-data evaluation; holdings are shares, gaps are dollars, and gains are USD. Values rounded only for display.

Positive entry-time expectations still permit losses



There are 2 losing trades; the worst loses $413.64.

Real data: weekly Yahoo Finance Close prices in the Homework 5 extract; split adjusted, excluding dividends.

Add the gains, keeping the forecast times straight

Sum of forecasts made at the eight entries	\$812.96
Sum of realized price gains	\$929.62

Later entry gaps were not available at time zero.

These are dollar price gains at the stipulated position scale. No invested-capital return has been defined.

Future innovations remain in the trading gain

For the position chosen at entry,

$$G_{t,h} = \underbrace{q_W(\phi^h - 1)U_t}_{\text{conditional expected gain}} + \underbrace{q_W \sum_{j=1}^{h} \phi^{h-j} Z_{t+j}}_{\text{unpredictable contribution}} .$$

The second term has conditional mean zero, but can outweigh the first on a particular trade.

Dividends and trading costs would add further cash flows to this price-only calculation.

Beyond VAR(1)

What changes when we include more lags
and moving-average terms?

What if the most recent vector is not enough?

From here, $X_t \in \mathbb{R}^d$ is a vector observation with mean μ, and $Y_t = X_t - \mu$. A VAR(p) has

$$Y_t = A_1 Y_{t-1} + \cdots + A_p Y_{t-p} + Z_t, \qquad p \geq 1.$$

Each A_i is $d \times d$ and $(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$, with Σ_Z positive definite.

Can the VAR(1) stability argument still be used?

Store the previous observation as part of the state

For a VAR(2), stack the current and previous centered vectors:

$$\underbrace{\begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}}_{S_t} = \underbrace{\begin{pmatrix} A_1 & A_2 \\ I_d & 0 \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} Y_{t-1} \\ Y_{t-2} \end{pmatrix}}_{S_{t-1}} + \begin{pmatrix} Z_t \\ 0 \end{pmatrix}.$$

The first block predicts the new observation. The second copies the preceding one.

This is a VAR(1) in $\mathbb{R}^{2d}$.

The companion matrix includes every lag coefficient

Stack $S_t = (Y_t^\top, \ldots, Y_{t-p+1}^\top)^\top$. Then

$$F = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_d & 0 & \cdots & 0 & 0 \\ 0 & I_d & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & I_d & 0 \end{pmatrix}, \qquad S_t = FS_{t-1} + J^\top Z_t,$$

where $J = (I_d \ 0 \ \cdots \ 0)$ selects the first block: $Y_t = JS_t$.

Stability is determined by $\rho(F) < 1$, using all lag matrices together.

Stable companion powers give an innovation filter

If $\rho(F) < 1$,

$$S_t = \sum_{j \geq 0} F^j J^\top Z_{t-j}, \qquad Y_t = \sum_{j \geq 0} \underbrace{J F^j J^\top}_{\Psi_j} Z_{t-j}.$$

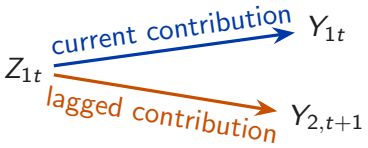
Only the first state block receives new noise; the remaining blocks copy past observations.

The padded noise covariance $J^\top \Sigma_Z J$ is singular for $p > 1$. The convergence argument needs only a finite positive-semidefinite covariance.

An innovation can enter another coordinate one period later

Consider the zero-mean model

$$Y_{1t} = Z_{1t}, \qquad Y_{2t} = Z_{2t} + Z_{1,t-1}, \qquad (Z_t) \sim \mathrm{WN}_2(0, I_2).$$



The same innovation enters two observations at different times.

A VMA combines a finite number of innovations

A vector moving average of order $q \geq 0$ is

$$Y_t = Z_t + C_1 Z_{t-1} + \cdots + C_q Z_{t-q}.$$

Here $Y_t = X_t - \mu$, each C_j is $d \times d$, and $(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$ with Σ_Z positive definite.

Every choice of the C_j gives a weakly stationary process, with $\mathbb{E}[Y_t] = 0$. There is no stability condition for this finite construction.

An entry specifies a coordinate and a lag

The entry $(C_j)_{ab}$ multiplies $Z_{b,t-j}$ in the equation for Y_{at}. In the preceding example,

$$C_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad (C_1)_{21} = 1.$$

Thus Z_{1t} contributes to $Y_{2,t+1}$ with coefficient one.

In general, Z_s contributes Z_s to Y_s and $C_j Z_s$ to Y_{s+j}, for $1 \leq j \leq q$. There is no term involving Z_s after time $s + q$.

Individual ACFs miss this lagged relation

Return to $Y_{1t} = Z_{1t}$ and $Y_{2t} = Z_{2t} + Z_{1,t-1}$, with $\Sigma_Z = I_2$. Distinct scalar innovations are uncorrelated. Within either coordinate, observations at different dates use distinct scalar innovations.

Both individual ACFs are zero away from lag zero. But Y_{1t} and $Y_{2,t+1}$ share Z_{1t}. What is their covariance?

Individual ACFs miss this lagged relation

Return to $Y_{1t} = Z_{1t}$ and $Y_{2t} = Z_{2t} + Z_{1,t-1}$, with $\Sigma_Z = I_2$. Distinct scalar innovations are uncorrelated. Within either coordinate, observations at different dates use distinct scalar innovations.

Both individual ACFs are zero away from lag zero. But Y_{1t} and $Y_{2,t+1}$ share Z_{1t}. What is their covariance?

$$\text{Cov}(Y_{2,t+1}, Y_{1t}) = \text{Cov}(Z_{2,t+1} + Z_{1t}, Z_{1t}) = 1.$$

The cross-covariance reveals their lagged relation.

A matrix filter can change dimension as well as time

For input $V_t \in \mathbb{R}^d$, define

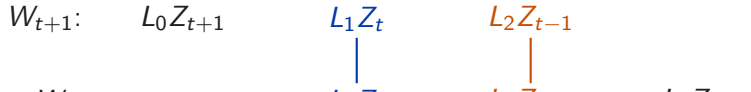
$$W_t = \sum_{j=0}^{m} L_j V_{t-j}, \qquad L_j \in \mathbb{R}^{r \times d}.$$

The output lies in $\mathbb{R}^r$. A weakly stationary input gives a weakly stationary output.

For a VMA, the input is white noise, $L_0 = I_d$, and $L_j = C_j$.

Which innovation times are shared?

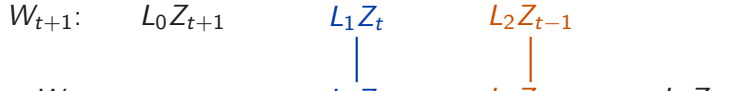
Take $W_t = L_0 Z_t + L_1 Z_{t-1} + L_2 Z_{t-2}$ with vector white noise.



Which terms contribute to $\text{Cov}(W_{t+1}, W_t)$?

Which innovation times are shared?

Take $W_t = L_0 Z_t + L_1 Z_{t-1} + L_2 Z_{t-2}$ with vector white noise.



Which terms contribute to $\text{Cov}(W_{t+1}, W_t)$?

Only the matching times t and $t-1$ survive.

Keep the future coefficient on the left

Using $\Gamma_W(h) = \mathrm{Cov}(W_{t+h}, W_t)$, the preceding example gives

$$\Gamma_W(1) = L_1 \Sigma_Z L_0^\top + L_2 \Sigma_Z L_1^\top.$$

For each shared innovation, take its coefficient in W_{t+1}, then its covariance, then the transpose of its coefficient in W_t.

Changing this order generally changes the matrix.

A finite innovation filter has a covariance cutoff

For $W_t = \sum_{j=0}^{m} L_j Z_{t-j}$,

$$\Gamma_W(h) = \sum_{j=0}^{m-h} L_{j+h} \Sigma_Z L_j^\top \quad (0 \le h \le m), \qquad \Gamma_W(h) = 0 \quad (|h| > m).$$

The shared innovation Z_{t-j} must occur in both sums, requiring $j \le m - h$.

For negative lags, $\Gamma_W(-h) = \Gamma_W(h)^\top$.

VARMA uses past observations and past innovations

For $Y_t = X_t - \mu$, a VARMA(p, q) equation is

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + Z_t + \sum_{j=1}^{q} C_j Z_{t-j}.$$

The coefficient matrices are $d \times d$ and $(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$ with Σ_Z positive definite.

Setting $q = 0$ gives a VAR; setting $p = 0$ gives a VMA.

Work a VARMA(1,1) by direct substitution

Take $Y_t = AY_{t-1} + Z_t + CZ_{t-1}$. Substituting once for Y_{t-1} gives

$$Y_t = Z_t + (A + C)Z_{t-1} + ACZ_{t-2} + A^2 Y_{t-2}.$$

Expanding Y_{t-2} contributes another $A^2 Z_{t-2}$. Thus the coefficient of Z_{t-2} is $A^2 + AC = A(A + C)$.

Repeated substitution gives candidate coefficients

Continuing suggests $Y_t = \sum_{j \geq 0} \Psi_j Z_{t-j}$, with

$$\Psi_0 = I_d, \qquad \Psi_j = A^{j-1}(A + C), \quad j \geq 1.$$

Each additional lag propagates the contribution by another factor of A.

Does this infinite sum converge to a stationary process?

Summable coefficient sequences define an infinite filter

Let $(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$ and let L_j be fixed $r \times d$ matrices. Assume

$$\sum_{j=0}^{\infty} |(L_j)_{ab}| < \infty \quad \text{for every } a = 1, \ldots, r, \quad b = 1, \ldots, d.$$

Then $W_t = \sum_{j \geq 0} L_j Z_{t-j}$ converges in mean square.

Each coefficient sequence is absolutely summable. Why does this control the random sum?

The omitted tail has vanishing mean square

Different innovation times are uncorrelated. For $n > m$,

$$\mathbb{E}\left[\left\|\sum_{j=m+1}^{n} L_j Z_{t-j}\right\|_2^2\right] = \sum_{j=m+1}^{n} \operatorname{tr}(L_j \Sigma_Z L_j^\top)$$

$$\leq \operatorname{tr}(\Sigma_Z) \sum_{j=m+1}^{n} \sum_{a,b} (L_j)_{ab}^2.$$

Each entry sequence is square summable, so the bound tends to zero as m grows. The partial sums converge in mean square.

The same matching rule gives the infinite covariance

Under the summability condition,

$$\Gamma_W(h) = \sum_{j=0}^{\infty} L_{j+h} \Sigma_Z L_j^\top, \qquad h \geq 0.$$

The mean is zero and the covariance depends only on the lag, so W is weakly stationary.

Negative lags still satisfy $\Gamma_W(-h) = \Gamma_W(h)^\top$.

Stable powers turn the candidate into a VARMA solution

For $\Psi_0 = I_d$ and $\Psi_j = A^{j-1}(A + C)$, the condition $\rho(A) < 1$ makes every entry sequence absolutely summable. The innovation series therefore defines a mean-zero weakly stationary process.

Since $\Psi_1 = A + C$ and $\Psi_j = A\Psi_{j-1}$ for $j \geq 2$,

$$Y_t - AY_{t-1} = Z_t + CZ_{t-1}.$$

Thus the constructed process satisfies the VARMA(1,1) equation.

Collect the coefficients into matrix polynomials

For a formal variable z, define

$$\Phi(z) = I_d - \sum_{i=1}^{p} A_i z^i, \qquad M(z) = I_d + \sum_{j=1}^{q} C_j z^j.$$

Moving the lagged observations to the left gives the AR minus signs. The lagged innovations retain the MA plus signs.

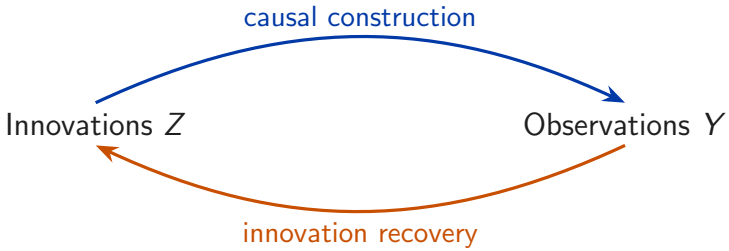
Replace powers by shifts to recover the VARMA equation

The backshift moves every coordinate together: $B^j Y_t = Y_{t-j}$. Replace each z^j in the matrix polynomials by B^j:

$$\Phi(B) Y_t = M(B) Z_t.$$

This is the original VARMA equation with the observation terms on the left and the innovation terms on the right.

Construction and recovery run in opposite directions



Both filters use current and past inputs. Recovery asks whether unobserved innovations can be calculated from observations.

Invert the AR polynomial on the left

If $\det \Phi(z) \neq 0$ for every complex z with $|z| \leq 1$, then

$$\Psi(z) = \Phi(z)^{-1} M(z) = \sum_{j \geq 0} \Psi_j z^j, \qquad Y_t = \sum_{j \geq 0} \Psi_j Z_{t-j}.$$

The coefficients are absolutely summable and $\Psi_0 = I_d$. The innovation filter therefore gives a causal weakly stationary solution.

Matrix order matters: $\Phi^{-1} M$ generally differs from $M \Phi^{-1}$.

Polynomial roots and companion eigenvalues are reciprocal

For a VAR(p), the zeros of $\det \Phi(z)$ are reciprocals of the nonzero companion eigenvalues.

$$|\lambda(F)| < 1 \quad \Longleftrightarrow \quad \text{corresponding root } |z| > 1.$$

For a scalar AR(1) with coefficient .8, $\Phi(z) = 1 - .8z$ has its root at 1.25.

The two stability checks describe the same condition.

Invertibility recovers innovations from observations

Also assume $\det M(z) \neq 0$ for every $|z| \leq 1$. Then

$$\Pi(z) = M(z)^{-1}\Phi(z) = \sum_{j \geq 0} \Pi_j z^j, \qquad Z_t = \sum_{j \geq 0} \Pi_j Y_{t-j}.$$

The inverse coefficients are absolutely summable and $\Pi_0 = I_d$. This is an *invertible* representation.

Both displayed determinant conditions are sufficient for these constructions.

Add a lagged innovation to the rotation model

For the numerical example, use

$$A = .8 \begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix}, \quad C = \begin{pmatrix} .3 & .4 \\ 0 & -.2 \end{pmatrix}, \quad \Sigma_Z = \begin{pmatrix} 1 & .6 \\ .6 & 1 \end{pmatrix}.$$

The radii are $\rho(A) = .8$ and $\rho(C) = .3$, so both filters exist.

The innovation coordinates are correlated, and A and C do not commute.

The innovation coefficients give the stationary covariance

Use the stationary causal solution, with $\Psi_0 = I_2$ and $\Psi_j = A^{j-1}(A + C)$ for $j \geq 1$.

$$\Gamma(0) = \Sigma_Z + \sum_{k=0}^{\infty} A^k(A + C)\Sigma_Z(A + C)^\top (A^k)^\top.$$

The current innovation contributes Σ_Z. Each term in the sum is the variance contribution of one older innovation.

Separate the current innovation to find the lag-one covariance

Use $\Psi_1 = A + C$ and $\Psi_{j+1} = A\Psi_j$ for $j \geq 1$ in the infinite-filter covariance formula:

$$\begin{aligned}\Gamma(1) &= (A + C)\Sigma_Z + A\sum_{j=1}^{\infty} \Psi_j \Sigma_Z \Psi_j^\top \\ &= (A + C)\Sigma_Z + A\{\Gamma(0) - \Sigma_Z\} \\ &= A\Gamma(0) + C\Sigma_Z.\end{aligned}$$

The lagged innovation supplies the extra term $C\Sigma_Z$.

Evaluate those covariances for the rotation model

Summing the variance contributions through lag 200 gives

$$\Gamma(0) \approx \begin{pmatrix} 2.913 & 1.335 \\ 1.335 & 3.602 \end{pmatrix}, \qquad \Gamma(1) \approx \begin{pmatrix} 2.024 & 0.064 \\ 1.970 & 2.829 \end{pmatrix}.$$

The second matrix uses $\Gamma(1) = A\Gamma(0) + C\Sigma_Z$.

Calculated example: $A = .8R_{\pi/6}$, $C = \left(\begin{smallmatrix} .3 & .4 \\ 0 & -.2 \end{smallmatrix}\right)$; no simulation.

Rearrange the equation to recover an innovation

The VARMA(1,1) equation gives

$$Z_t = Y_t - AY_{t-1} - CZ_{t-1}.$$

If $\rho(C) < 1$, repeated substitution expresses the old innovation in terms of current and past observations.

Which lag-two matrix appears in that expansion: AC or CA?

Recovery reverses the product order

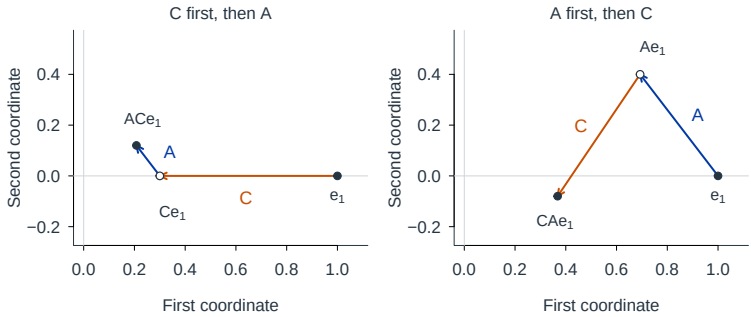
Substituting $Z_{t-1} = Y_{t-1} - AY_{t-2} - CZ_{t-2}$ gives

$$Z_t = Y_t - (A + C)Y_{t-1} + CAY_{t-2} + C^2 Z_{t-2}.$$

The next substitution supplies another $C^2 Y_{t-2}$, so $\Pi_2 = C^2 + CA$.

In general, $\Pi_0 = I_d$ and $\Pi_j = (-C)^j - (-C)^{j-1}A$ for $j \geq 1$.

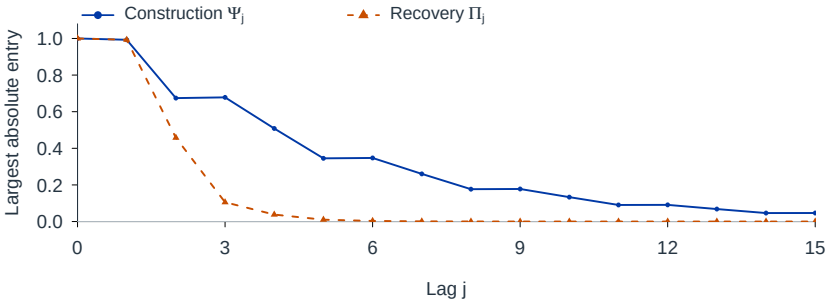
Changing the order changes the final vector



For $e_1 = (1, 0)^\top$, ACe_1 applies C first; CAe_1 applies A first.

Calculated example: $A = .8R_{\pi/6}$, $C = \left(\begin{smallmatrix} .3 & .4 \\ 0 & -.2 \end{smallmatrix}\right)$; no simulation.

The coefficient entries tend to zero in this example



The plot shows the largest absolute entry of Ψ_j and Π_j at each lag.

Calculated example: $A = .8R_{\pi/6}$, $C = \begin{pmatrix} .3 & .4 \\ 0 & -.2 \end{pmatrix}$; no simulation.

A finite record requires an initial innovation value

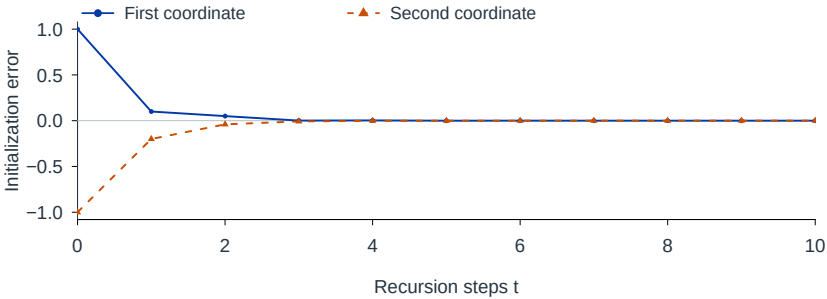
With observations from time zero onward, choose $\widetilde{Z}_0$ and compute

$$\widetilde{Z}_t = Y_t - AY_{t-1} - C\widetilde{Z}_{t-1}, \qquad t \geq 1.$$

Subtracting the true innovation equation gives $\widetilde{Z}_t - Z_t = -C(\widetilde{Z}_{t-1} - Z_{t-1})$.

Therefore $\widetilde{Z}_t - Z_t = (-C)^t(\widetilde{Z}_0 - Z_0)$.

The effect of initialization fades



The error tends to zero under $\rho(C) < 1$; it need not be zero at a finite time.

Calculated initialization error $(-C)^t(1,-1)^\top$ for the displayed lagged-innovation matrix C.

Several observed lags still give ordinary regressions

For a VAR(p) and the random sample $X_1, \ldots, X_T$, write $c = (I_d - \sum_i A_i)\mu$. The least-squares estimators minimize

$$\sum_{t=p+1}^{T} \left\| X_t - c - \sum_{i=1}^{p} A_i X_{t-i} \right\|_2^2 .$$

For realized data $x_1, \ldots, x_T$, fit d regressions, each on an intercept and all dp lagged coordinates.

Check stability using the fitted companion matrix.

The covariance equations also extend to several lags

In the causal VAR(p), with $\Gamma(h) = \text{Cov}(Y_{t+h}, Y_t)$,

$$\Gamma(h) = \sum_{i=1}^{p} A_i \Gamma(h-i), \qquad h = 1, \ldots, p.$$

Replace the covariance matrices by sample versions and solve for the lag matrices, provided the sample system is nonsingular.

The sample mean estimates μ; fitted residual covariances estimate Σ_Z.

Use the calculated covariances as equations for fitting

The stationary VARMA(1,1) covariance calculation gave

$$\Gamma(1) = A\Gamma(0) + C\Sigma_Z.$$

Now regard A, C, Σ_Z as unknown. Match the model formulas for $\Gamma(h)$ at several lags to their sample covariance matrices.

The equations involve AR coefficients, MA coefficients, and innovation covariances. They are generally nonlinear.

Candidate coefficients produce candidate residuals

For a candidate mean and matrices, compute

$$\widetilde{z}_t = (x_t - \mu) - A(x_{t-1} - \mu) - C\widetilde{z}_{t-1}.$$

Changing A, C, μ changes the earlier residuals as well as the current one.

Minimizing their squared lengths is therefore a nonlinear least-squares problem.

Likelihood needs a full innovation law

White-noise means and covariances do not specify a joint density. Assume i.i.d. innovation vectors with a specified joint density family, and let θ collect all model parameters.

$$L_T(\theta) = p_\theta(x_1, \ldots, x_T).$$

The stationary causal model determines this observed-record likelihood.

An arbitrary initialization of the recovery recursion does not by itself evaluate it exactly.

Use the same likelihood for estimation and prior updating

Maximum likelihood maximizes $L_T(\theta)$. With prior density π,

$$\pi(\theta \mid x_1, \ldots, x_T) \propto L_T(\theta)\pi(\theta).$$

The matrix model changes the likelihood calculation; the fitting principles are the same.

Continue with the companion and recovery calculations

- Build the VAR(2) companion matrix and check all its eigenvalues.
- Compare $A^2 + AC$ with $C^2 + CA$ in the rotation example.
- Explain why two initialized recovery sequences become close.

Recovered innovations will also support the later development of multivariate forecasts.