STA 542: Introduction to Time Series Analysis

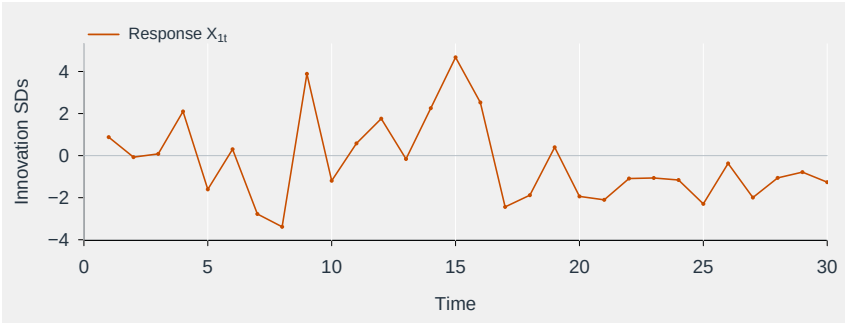
Lecture 11: Vector autoregression

Lasse Vuursteen

Duke
UNIVERSITY

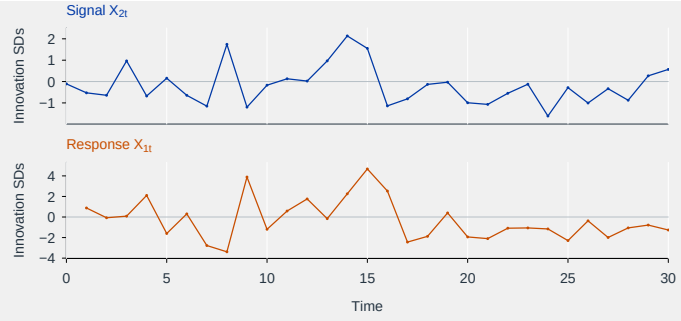
What can this series tell us about its next value?

Suppose only this response is observed.



Which features of its past would help predict the next observation?

Would observing a second series help?



Look for a delay between the signal and the response.

Would the predictive information have to work in both directions?

Each coordinate can depend on the preceding vector

For a zero-mean bivariate process, consider

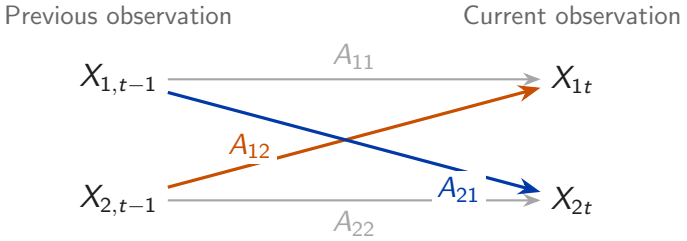
$$X_{1t} = A_{11}X_{1,t-1} + A_{12}X_{2,t-1} + Z_{1t},$$

$$X_{2t} = A_{21}X_{1,t-1} + A_{22}X_{2,t-1} + Z_{2t}.$$

The first index names the response coordinate. The second names the lagged predictor.

Separate AR(1) equations impose $A_{12} = A_{21} = 0$.

A row collects the predictors of one coordinate



The innovation Z_t supplies the new contribution at time t.

Vector white noise has no cross-time covariances

Vector white noise

$(Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z)$ has finite second moments and

$$\mathbb{E}[Z_t] = 0, \qquad \mathrm{Cov}(Z_t) = \Sigma_Z, \qquad \mathrm{Cov}(Z_t, Z_s) = 0 \quad (t \neq s).$$

- All pairs of coordinates are uncorrelated across distinct times.
- Coordinates at the same time may be correlated.

We take Σ_Z positive definite.

A VAR(1) uses one lag of the whole vector

Vector autoregression of order one

For $X_t \in \mathbb{R}^d$, $\mu \in \mathbb{R}^d$, and a real $d \times d$ matrix A,

$$X_t - \mu = A(X_{t-1} - \mu) + Z_t, \qquad (Z_t) \sim \mathrm{WN}_d(0, \Sigma_Z).$$

A causal solution depends only on current and past innovations.

Which matrices A allow a stationary causal solution?

Lagged dependence and simultaneous variation differ

Coefficient matrix A

A_{12} multiplies the preceding value $X_{2,t-1} - \mu_2$ in the equation for X_{1t}.

Innovation covariance Σ_Z

$(\Sigma_Z)_{12}$ measures covariance between Z_{1t} and Z_{2t} at the same time.

A correlated innovation does not by itself specify a lagged effect.

Repeated substitution exposes the old innovations

For $X_t - \mu = A(X_{t-1} - \mu) + Z_t$, iterate m times:

$$X_t - \mu = \underbrace{A^m(X_{t-m} - \mu)}_{\text{remote state}} + \underbrace{\sum_{j=0}^{m-1} A^j Z_{t-j}}_{\text{accumulated innovations}} .$$

Here $A^0 = I_d$.

A^j determines the contribution of an innovation j periods later.

Repeated substitution exposes the old innovations

For $X_t - \mu = A(X_{t-1} - \mu) + Z_t$, iterate m times:

$$X_t - \mu = \underbrace{A^m(X_{t-m} - \mu)}_{\text{remote state}} + \underbrace{\sum_{j=0}^{m-1} A^j Z_{t-j}}_{\text{accumulated innovations}} .$$

Here $A^0 = I_d$.

A^j determines the contribution of an innovation j periods later.

We need the remote state to vanish and the covariance sum to converge.

Along an eigenvector, powers reduce to scalar powers

An eigenvalue λ has a nonzero, possibly complex vector v with

$$Av = \lambda v \quad \Longrightarrow \quad A^j v = \lambda^j v.$$

The spectral radius is the largest eigenvalue modulus: $\rho(A) = \max_\lambda |\lambda|$.

What should happen when every eigenvalue has modulus below one?

Stability is a condition on the whole matrix

If $\rho(A) < 1$, there are constants $C < \infty$ and $0 < q < 1$ such that

$$\|A^j v\|_2 \leq C q^j \|v\|_2 \qquad (j \geq 0,\ v \in \mathbb{R}^d).$$

This geometric bound is equivalent to $\rho(A) < 1$.

- Every initial direction eventually disappears.
- A contribution can grow at first before it decays.

Which matrix is stable?

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} -.8 & -.8 \\ .8 & -.8 \end{pmatrix}.$$

Predict first. Inspect powers or eigenvalues.

Which matrix is stable?

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} -.8 & -.8 \\ .8 & -.8 \end{pmatrix}.$$

Predict first. Inspect powers or eigenvalues.

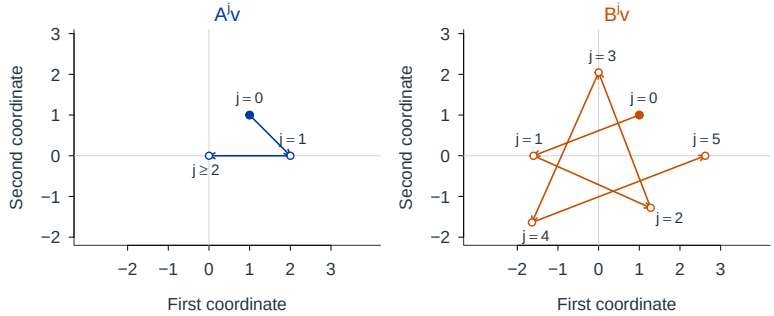
$A^2 = 0$, so $\rho(A) = 0$.

B has eigenvalues $-.8 \pm .8i$, so $\rho(B) = .8\sqrt{2} > 1$.

The sizes of individual entries do not settle stability.

Apply each matrix repeatedly to the same vector

Start with $v = (1, 1)^\top$. The arrows join successive values of $A^j v$ and $B^j v$.



A sends the vector to zero after two steps. B turns it while increasing its length.

Calculated paths, $j = 0, \ldots, 5$; equal axis scales; no new innovations.

Stable powers give a stationary causal solution

For $\rho(A) < 1$ and vector white noise with covariance Σ_Z,

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

The series converges in mean square and is the unique weakly stationary solution of the VAR(1) equation.

Each observation is a sum of current and past innovations, with coefficients that eventually decay.

Uncorrelated innovations make covariances add

Under $\rho(A) < 1$, the stationary representation gives the covariance $\Gamma(0) = \mathrm{Cov}(X_t)$:

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

- Distinct innovation times have zero cross-covariance.
- The geometric bound makes the sum finite.

The term $j = 0$ is Σ_Z, so $\Gamma(0)$ is positive definite.

Separate the newest innovation from the older ones

$$\Gamma(0) = \Sigma_Z + \sum_{j=1}^{\infty} A^j \Sigma_Z (A^j)^\top$$

$$= \Sigma_Z + A\Gamma(0)A^\top.$$

The covariance of the current observation is the new innovation covariance plus the propagated covariance of the previous observation.

The lag-one covariance carries the coefficient matrix

Recall $\Gamma(h) = \mathrm{Cov}(X_{t+h}, X_t)$. In the stationary causal solution,

$$\begin{aligned}\Gamma(1) &= \mathrm{Cov}\big(A(X_t - \mu) + Z_{t+1},\, X_t\big) \\ &= A\Gamma(0).\end{aligned}$$

The future innovation is uncorrelated with the current observation.

Iteration gives $\Gamma(h) = A^h\Gamma(0)$ for $h \geq 0$.

A matrix can rotate a contribution while shrinking it

Consider $X_t = AX_{t-1} + Z_t$ with

$$A = rR_\theta, \qquad R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad 0 < r < 1.$$

Take $(Z_t) \sim \mathrm{WN}_2(0, \sigma^2 I_2)$, with $\sigma^2 > 0$, and the stationary causal solution.

R_θ rotates counterclockwise by θ; r scales the length.

Where does an innovation in the first coordinate go?

For a single innovation $e_1 = (1, 0)^\top$, its contribution after j periods is

$$A^j e_1 = r^j R_{j\theta} e_1 = r^j \begin{pmatrix} \cos(j\theta) \\ \sin(j\theta) \end{pmatrix}.$$

The direction changes by θ each period. The magnitude is r^j.

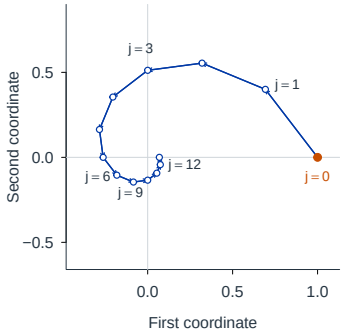
For the next figures, take $r = .8$ and $\theta = \pi/6$.

The effect moves between coordinates

Start with $(1, 0)^\top$.

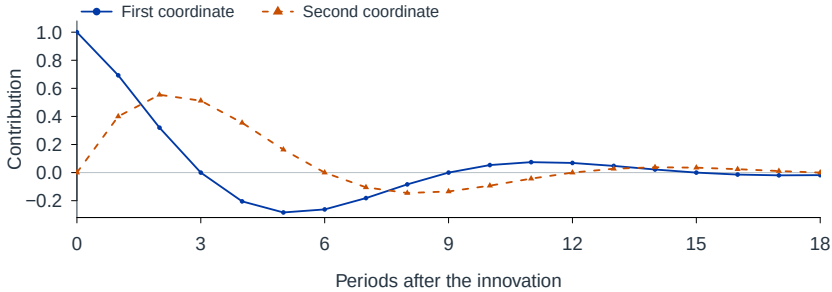
After three periods, the remaining contribution is entirely in the second coordinate.

Its magnitude is $.8^3 = .512$.



Calculated contribution of one innovation: $r = .8$, $\theta = \pi/6$.

Both coordinate contributions eventually disappear

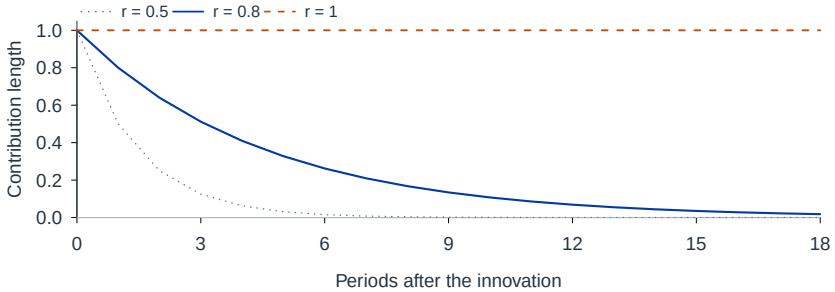


Initially, the first coordinate contribution falls while the second rises.

An observed path also receives new innovations at every time.

Calculated contribution $A^j(1,0)^\top$: $r = .8$, $\theta = \pi/6$.

The radius controls how long a contribution persists



The eigenvalues are $re^{i\theta}$ and $re^{-i\theta}$, so $\rho(A) = r$.

Calculated magnitudes $\|A^j e_1\|_2 = r^j$ at three values of r.

Zero contemporaneous correlation leaves room for lagged dependence

For the stationary rotation model with $\mathrm{Cov}(Z_t) = \sigma^2 I_2$,

$$\Gamma(0) = \frac{\sigma^2}{1 - r^2} I_2, \qquad \Gamma(h) = \frac{\sigma^2 r^h}{1 - r^2} R_{h\theta} \quad (h \geq 0).$$

Because $R_{j\theta} R_{j\theta}^\top = I_2$, the variance sum is geometric.

Current coordinates are uncorrelated. Cross-coordinate covariances at other lags can be nonzero.

At radius one, past innovations do not fade

Set $r = 1$, start at $X_0 = 0$, and take iid $Z_t \sim N_2(0, \sigma^2 I_2)$.

$$X_t = R_\theta X_{t-1} + Z_t, \qquad \mathrm{Cov}(X_t) = t\sigma^2 I_2.$$

Every innovation retains its magnitude, so variance accumulates.

If $R_\theta = I_2$, this is an ordinary bivariate random walk. What changes when the rotation is nontrivial?

Ordinary differences still contain the growing state

At $r = 1$, with the preceding initialization and innovations,

$$\nabla X_t = (R_\theta - I_2)X_{t-1} + Z_t.$$

If $R_\theta \neq I_2$, the difference variance grows with t. Hence the ordinary differences are not $I(0)$, and X_t is not $I(1)$ under our definition.

A time-varying filter can undo the rotation

Keep $r = 1$, $X_0 = 0$, and iid $Z_t \sim N_2(0, \sigma^2 I_2)$. Define $Y_t := R_{-t\theta} X_t$. Then

$$Y_t - Y_{t-1} = R_{-t\theta} Z_t.$$

The rotated innovations remain iid $N_2(0, \sigma^2 I_2)$. Thus (Y_t) is an ordinary bivariate random walk.

The filter changes with t: it uses a rotating coordinate system.

Now suppose the parameters are unknown

Observe $X_1, \ldots, X_T$ from the stationary causal model

$$X_t - \mu = A(X_{t-1} - \mu) + Z_t, \qquad \rho(A) < 1.$$

The innovations are vector white noise with positive-definite covariance Σ_Z.

We want to estimate μ, A, and Σ_Z. Can the mean and covariances identify them?

Yule–Walker equations identify the coefficients

The stationary covariance identities give

$$\Gamma(1) = A\Gamma(0) \quad \Longrightarrow \quad A = \Gamma(1)\Gamma(0)^{-1},$$

$$\Gamma(0) = A\Gamma(0)A^\top + \Sigma_Z \quad \Longrightarrow \quad \Sigma_Z = \Gamma(0) - A\Gamma(0)A^\top.$$

$\Gamma(0)$ is invertible because its covariance sum includes $\Sigma_Z > 0$.

The inverse multiplies on the right.

Replace population covariances by sample covariances

For random observations $X_1, \ldots, X_T$, let $\bar{X} = T^{-1} \sum_{t=1}^{T} X_t$ and define

$$\widehat{\Gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_{t+h} - \bar{X})(X_t - \bar{X})^\top, \qquad h = 0, 1.$$

Each summand is a $d \times d$ matrix.

We use divisor T at both lags and keep the future coordinate on the left.

The sample identities give moment estimators

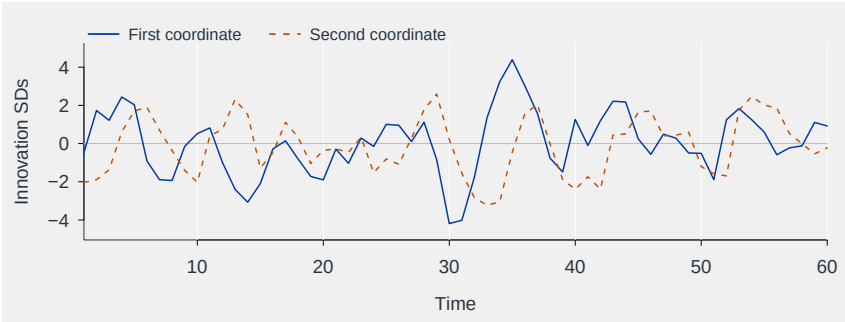
Provided $\widehat{\Gamma}(0)$ is invertible,

$$\widehat{\mu} = \bar{X},$$
$$\widehat{A} = \widehat{\Gamma}(1)\widehat{\Gamma}(0)^{-1},$$
$$\widehat{\Sigma}_Z = \widehat{\Gamma}(0) - \widehat{A}\widehat{\Gamma}(0)\widehat{A}^\top.$$

Convergence in probability of the sample moments to their population values gives consistency. Iid innovations with finite second moments provide a sufficient setting in the stable causal model.

Fit the model to one stationary simulated record

Use $A = .8R_{\pi/6}$ and iid $Z_t \sim N_2(0, I_2)$.



Initialize $X_0 \sim N_2(0, I_2/(1 - .8^2))$ independently of future innovations. Use all $T = 300$ observations for fitting; the plot shows the first 60.

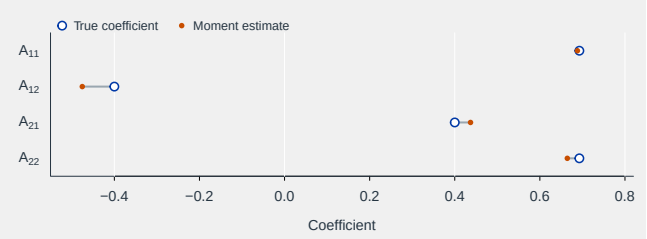
Simulated data: stationary Gaussian rotation, $T = 300$, $r = .8$, $\theta = \pi/6$, $\sigma^2 = 1$; seed 54211.

How close is the fitted matrix?

Substitute the realized record $x_1, \ldots, x_{300}$ into the estimators.

The fitted spectral radius is 0.816.

The true radius is .8.



Same simulated record; unrestricted moment fit.

Regression estimates one row at a time

The same model can be written as

$$X_t = (I_d - A)\mu + AX_{t-1} + Z_t.$$

Regress each current coordinate on an intercept and all lagged coordinates. The slopes form its row of A.

For $t = 2, \ldots, T$, the predictor–response pairs are $(X_1, X_2),\ (X_2, X_3), \ldots, (X_{T-1}, X_T)$.

The predictor sample ends at X_{T-1}; the response sample starts at X_2.

Why can least squares differ from the moment fit?

Both fits use the same lagged pairs. Fitting an intercept in least squares centers each side at its own sample mean:

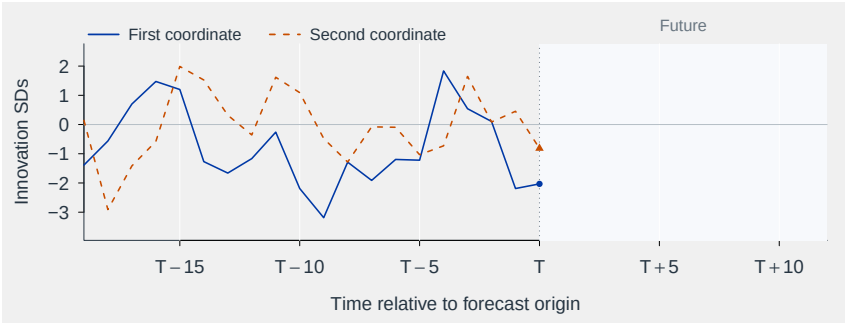
$$\bar{X}_{-} = \frac{X_1 + \cdots + X_{T-1}}{T-1} \quad \text{(predictors)},$$

$$\bar{X}_{+} = \frac{X_2 + \cdots + X_{T}}{T-1} \quad \text{(responses)}.$$

The moment fit centers both sides at the full-sample mean $\bar{X}$. Its $\widehat{\Gamma}(0)$ also includes X_T, whereas the least-squares predictor covariance uses only $X_1, \ldots, X_{T-1}$.

Where should the two coordinates go next?

Return to the same rotation record, now with the parameters treated as known.



Predict the direction of the next conditional mean for each coordinate.

Simulated data: stationary Gaussian rotation, $T = 300$, $r = .8$, $\theta = \pi/6$, $\sigma^2 = 1$; seed 54211.

Conditional means require assumptions about future noise

Consider a stationary causal VAR(1) with known parameters and *iid* innovations of mean zero and covariance Σ_Z. Let $\mathcal{H}_T = (X_1, \ldots, X_T)$. Causality makes this history depend only on innovations through T. The iid assumption makes future innovations independent of that history.

$$m_h := \mathbb{E}[X_{T+h} \mid \mathcal{H}_T].$$

m_h minimizes conditional expected squared Euclidean error.

Forward iteration separates the predictable part

$$X_{T+h} = \underbrace{\mu + A^h(X_T - \mu)}_{\text{known from the current observation}} + \underbrace{\sum_{j=0}^{h-1} A^j Z_{T+h-j}}_{\text{future innovations}}.$$

For integers $h \geq 1$, the second term has conditional mean zero.

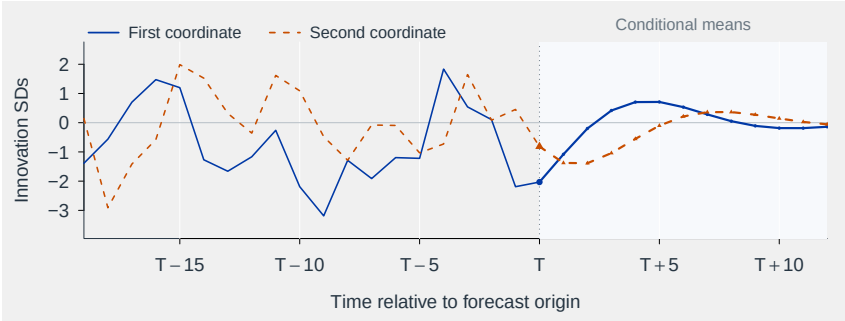
Forward iteration separates the predictable part

$$X_{T+h} = \underbrace{\mu + A^h(X_T - \mu)}_{\text{known from the current observation}} + \underbrace{\sum_{j=0}^{h-1} A^j Z_{T+h-j}}_{\text{future innovations}}.$$

For integers $h \geq 1$, the second term has conditional mean zero.

Thus $\widehat{X}_{T+h|T} = \mu + A^h(X_T - \mu)$.

The forecast rotates and decays toward the mean



With $\mu = 0$, the forecast is $A^h x_T = .8^h R_{h\pi/6} x_T$.

These are conditional means. Actual future observations receive new innovations.

Simulated data: stationary Gaussian rotation, $T = 300$, $r = .8$, $\theta = \pi/6$, $\sigma^2 = 1$; seed 54211.

Only future innovations remain in the forecast error

For the known-parameter iid model,

$$V_h := \mathrm{Cov}(X_{T+h} - \widehat{X}_{T+h|T} \mid \mathcal{H}_T) = \sum_{j=0}^{h-1} A^j \Sigma_Z (A^j)^\top.$$

At horizon one, $V_1 = \Sigma_Z$.

Each additional horizon includes another propagated innovation. The total mean squared Euclidean error is $\mathrm{tr}(V_h)$.

The first coordinate's forecast uses the second

In the zero-mean rotation model,

$$\widehat{X}_{1,T+1|T} = r\cos\theta\, X_{1T} - r\sin\theta\, X_{2T}.$$

For $\theta = \pi/6$, both current coordinates enter.

Contemporaneous zero correlation does not remove this lagged predictive relation.

Return to the signal and its delayed response

Let (Z_t) be iid $N_2(0, I_2)$ and, for every $t \in \mathbb{Z}$, define

$$X_{2t} = Z_{2t}, \qquad X_{1t} = cX_{2,t-1} + Z_{1t}, \qquad c \neq 0.$$

The signal is the second coordinate. The response is the first.

Each coordinate is an iid sequence, but the pair has a lagged dependence. The opening figures use $c = 2$.

Which history is available when predicting the response?

The target is $X_{1,t+1}$. Compare

$$\mathcal{H}_t^{(1)} = (\ldots, X_{1,t-1}, X_{1t}),$$

$$\mathcal{H}_t = (\ldots, X_{t-1}, X_t).$$

The first history contains the response only. The second also contains the signal.

Both histories end at the same time t.

The response's own history leaves both shocks unknown

The next response is $X_{1,t+1} = cZ_{2t} + Z_{1,t+1}$. Both shocks are independent of $\mathcal{H}_t^{(1)}$.

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t^{(1)}] = 0, \qquad \mathbb{E}[(X_{1,t+1} - 0)^2] = 1 + c^2.$$

Earlier responses contain signal shocks only through time $t - 1$.

No amount of response history reveals the fresh signal Z_{2t}.

Observing the signal reveals one of those shocks

The joint history contains $X_{2t} = Z_{2t}$. What is now predictable in $X_{1,t+1} = cZ_{2t} + Z_{1,t+1}$?

Observing the signal reveals one of those shocks

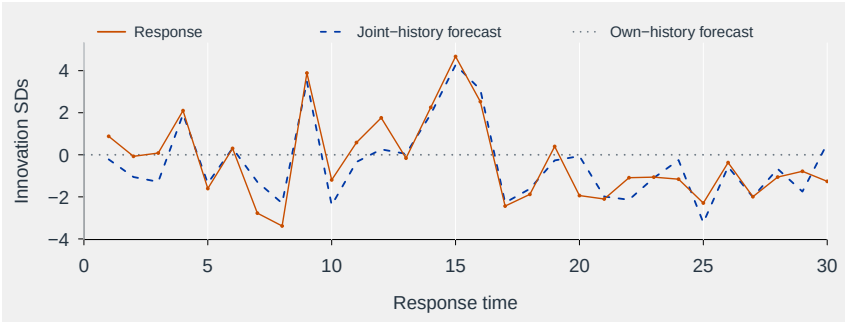
The joint history contains $X_{2t} = Z_{2t}$. What is now predictable in $X_{1,t+1} = cZ_{2t} + Z_{1,t+1}$?

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t] = cX_{2t}, \qquad \mathbb{E}[(X_{1,t+1} - cX_{2t})^2] = 1.$$

The mean squared prediction error falls by c^2.

Only the future response noise remains unknown.

The joint forecast follows the delayed signal



At each response time t, use information only through $t-1$. For $c=2$, the two theoretical MSEs are 5 and 1.

Simulated data: 30 responses with independent Gaussian noise; seed 54213.

Granger causality compares forecasts from two histories

Granger causality in mean, at horizon one

For a bivariate stationary process with finite second moments, the second coordinate Granger-causes the first if, for some t,

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t] \neq \mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t^{(1)}] \quad \text{with positive probability.}$$

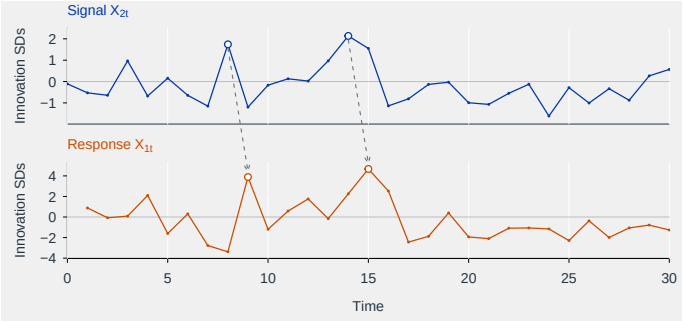
It does not Granger-cause the first if the forecasts agree with probability one for every t.

In the delayed-response model, X_2 Granger-causes X_1, but not conversely.



Clive Granger
Photo: Olaf Storbeck
CC BY-SA 2.0

The signal helps predict the next response



Each arrow spans one time step.

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t^{(1)}] = 0, \qquad \mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t] = 2X_{2t}.$$

Simulated data: 30 responses with independent Gaussian noise; seed 54213.

Different conditional means give a strict prediction gain

Write m_{own} and m_{joint} for the two conditional means. Then

$$\mathbb{E}[(X_{1,t+1} - m_{\text{own}})^2] - \mathbb{E}[(X_{1,t+1} - m_{\text{joint}})^2]$$
$$= \mathbb{E}[(m_{\text{joint}} - m_{\text{own}})^2].$$

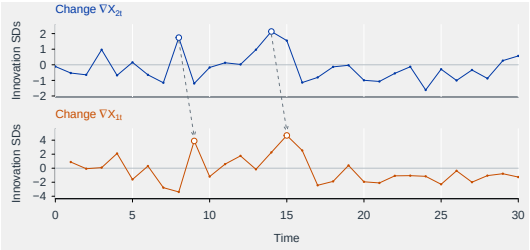
The improvement is positive exactly when the forecasts differ with positive probability.

For the delayed response, the right side is $\mathbb{E}[(cX_{2t})^2] = c^2$.

Let the changes follow a VAR(1)

For $\nabla X_t = X_t - X_{t-1}$, take the stationary causal model

$$\nabla X_t = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \nabla X_{t-1} + Z_t, \qquad Z_t \overset{\text{iid}}{\sim} N_2(0, I_2).$$

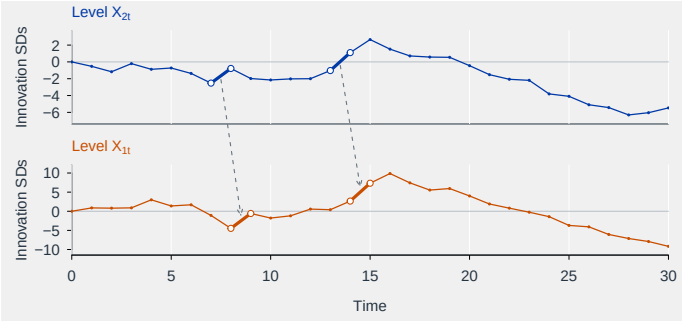


The latest change in X_2 predicts the next change in X_1:
$\mathbb{E}[\nabla X_{1,t+1} \mid \text{joint change history}] = 2\nabla X_{2t}$.

Simulated data: 30 responses with independent Gaussian noise; seed 54213.

Accumulating the changes gives the original levels

Set $X_0 = 0$ and $X_t = \sum_{s=1}^{t} \nabla X_s$ for $t \geq 1$.



Each arrow links two highlighted one-period movements.

Simulated data: 30 responses with independent Gaussian noise; seed 54213.

Predict the next level by adding the predicted change

Apply the same forecast comparison to these nonstationary levels. Use histories $\mathcal{H}_t^{(1)} = (X_{10}, \ldots, X_{1t})$ and $\mathcal{H}_t = (X_0, \ldots, X_t)$, with $t \geq 1$.

Since $X_{1,t+1} = X_{1t} + \nabla X_{1,t+1}$,

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t^{(1)}] = X_{1t},$$

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t] = X_{1t} + 2(X_{2t} - X_{2,t-1}).$$

The first coordinate's history is independent of the fresh Z_{2t}. The joint history reveals it as $X_{2t} - X_{2,t-1}$.

The second coordinate's own history and the joint history both forecast $X_{2,t+1}$ by X_{2t}.

A causal solution and iid noise justify the forecast

Consider $X_{t+1} = AX_t + Z_{t+1}$, with $\rho(A) < 1$ and iid mean-zero innovations with finite covariance Σ_Z. Use its stationary causal solution.

Observations through t depend only on innovations through t. Hence

$$\mathbb{E}[Z_{t+1} \mid \mathcal{H}_t] = 0,$$

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t] = A_{11}X_{1t} + A_{12}X_{2t}.$$

Outside this stability regime, a noncausal stationary solution may depend on future innovations.

A zero cross-coefficient removes the forecast gain

Keep this stationary causal model with iid mean-zero innovations. Then

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t] = A_{11}X_{1t} + A_{12}X_{2t},$$

$$\mathbb{E}[X_{1,t+1} \mid \mathcal{H}_t^{(1)}] = A_{11}X_{1t} + A_{12}\mathbb{E}[X_{2t} \mid \mathcal{H}_t^{(1)}].$$

If $A_{12} = 0$, the forecasts agree: X_2 does not Granger-cause X_1. This implication requires no Gaussian assumption.

For $A_{12} \neq 0$, a forecast gain also requires that X_{2t} is not already determined by X_1's complete history.

In a Gaussian VAR(1), the direction is in the matrix

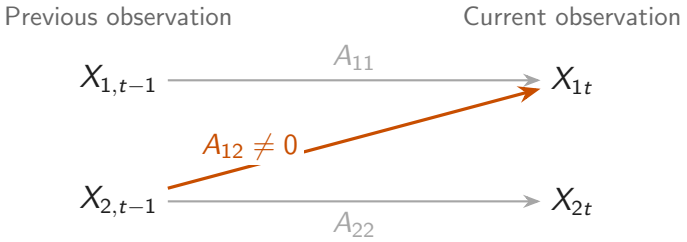
Now assume iid $N_2(0, \Sigma_Z)$ innovations with positive-definite Σ_Z. Keep the same stationary causal solution.

$$X_2 \text{ Granger-causes } X_1 \quad \iff \quad A_{12} \neq 0.$$

These additional assumptions ensure that X_{2t} is not determined by X_1's history, giving the converse.

The reverse direction is governed by A_{21}. For the delayed response, $A_{12} = c \neq 0$ and $A_{21} = 0$.

Which way does Granger causality run?



Keep this stationary causal Gaussian model, and take $A_{21} = 0$. Which coordinate Granger-causes the other?

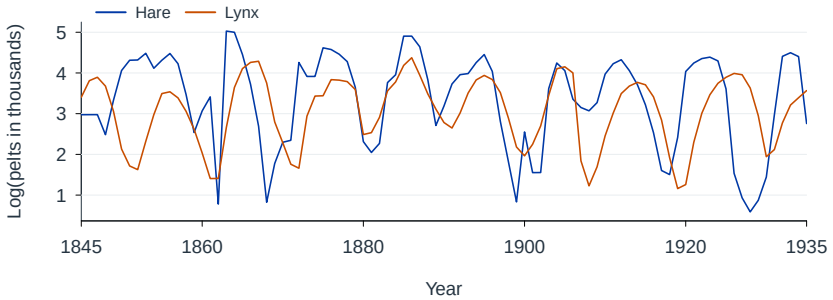
Predictive information does not establish an intervention effect

Granger causality concerns the specified observed histories and forecast target.

- The available variables and the model's lag structure matter.
- A fitted nonzero coefficient alone does not establish the process property.
- An intervention claim requires additional causal assumptions.

The delayed-response calculation works because its joint law is specified.

Return to the hare and lynx records



Can the current year's pair predict next year's log records?

The pelt records are proxies for animal populations.

Real data: Hudson's Bay Company pelt purchases, in thousands; `astsa::Hare` and `astsa::Lynx`, 1845–1935.

The paired equations form a VAR(1)

Let $Y_{H,t}$ and $Y_{L,t}$ be the logs of the hare and lynx records, measured in thousands of pelts before taking logs.

$$\begin{pmatrix} Y_{H,t+1} \\ Y_{L,t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} a_H \\ a_L \end{pmatrix}}_{a} + \underbrace{\begin{pmatrix} b_{HH} & b_{HL} \\ b_{LH} & b_{LL} \end{pmatrix}}_{A} \begin{pmatrix} Y_{H,t} \\ Y_{L,t} \end{pmatrix} + Z_{t+1}.$$

Assume $\mathbb{E}[Z_{t+1} \mid \mathcal{H}_t] = 0$ and $\mathrm{Cov}(Z_{t+1} \mid \mathcal{H}_t) = \Sigma_Z$, with Σ_Z fixed. Here $\mathcal{H}_t$ contains both log histories through year t.

What do the fitted cross-coefficients say?

Fit each row by least squares using the 90 annual transitions.

$$\widehat{a} = \begin{pmatrix} 2.002 \\ 0.230 \end{pmatrix}, \qquad \widehat{A} = \begin{pmatrix} 0.740 & -0.372 \\ 0.206 & 0.700 \end{pmatrix}.$$

In each equation, hold the species' own current log record fixed.

What do the fitted cross-coefficients say?

Fit each row by least squares using the 90 annual transitions.

$$\widehat{a} = \begin{pmatrix} 2.002 \\ 0.230 \end{pmatrix}, \qquad \widehat{A} = \begin{pmatrix} 0.740 & -0.372 \\ 0.206 & 0.700 \end{pmatrix}.$$

In each equation, hold the species' own current log record fixed.

A larger current lynx record lowers the fitted next log hare value.

A larger current hare record raises the fitted next log lynx value.

Will the fitted forecasts settle down?

The fitted matrix has eigenvalues

$$\lambda_{1,2} \approx 0.720 \pm 0.276\, i, \qquad |\lambda_1| = |\lambda_2| \approx 0.771.$$

What does this imply for repeated multiplication by $\widehat{A}$?

Will the fitted forecasts settle down?

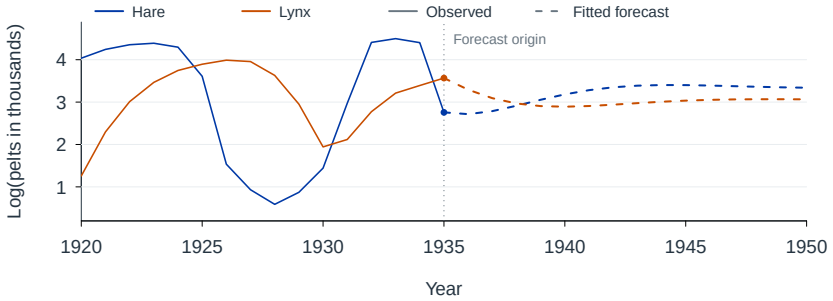
The fitted matrix has eigenvalues

$$\lambda_{1,2} \approx 0.720 \pm 0.276\, i, \qquad |\lambda_1| = |\lambda_2| \approx 0.771.$$

What does this imply for repeated multiplication by $\widehat{A}$?

The fitted recurrence is stable. Its forecasts converge to $\widehat{\mu} = (I_2 - \widehat{A})^{-1}\widehat{a}$ with damped oscillations.

Forecast the log records from 1935

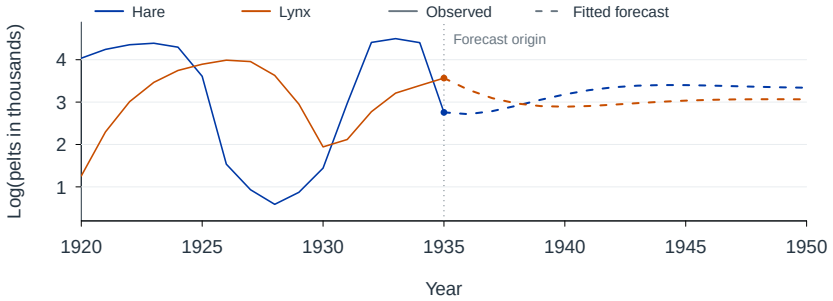


Iterate $\widehat{y}_{t+h|t} = \widehat{a} + \widehat{A}\widehat{y}_{t+h-1|t}$ from the last observed pair.

Does exponentiating a log forecast give mean pelts?

Real data: Hudson's Bay Company pelt purchases, in thousands; `astsa::Hare` and `astsa::Lynx`, 1845–1935.

Forecast the log records from 1935



Iterate $\widehat{y}_{t+h|t} = \widehat{a} + \widehat{A}\widehat{y}_{t+h-1|t}$ from the last observed pair.

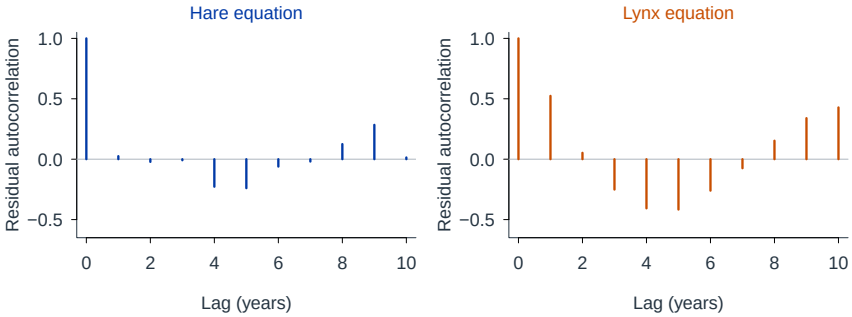
Does exponentiating a log forecast give mean pelts?

No. In general, $\mathbb{E}[e^{Y_{H,t+h}} \mid \mathcal{H}_t] \neq e^{\mathbb{E}[Y_{H,t+h}|\mathcal{H}_t]}$.

Real data: Hudson's Bay Company pelt purchases, in thousands; `astsa::Hare` and `astsa::Lynx`, 1845–1935.

Do the residuals leave predictive information unexplained?

Under the conditional-mean model, innovations are uncorrelated across years.

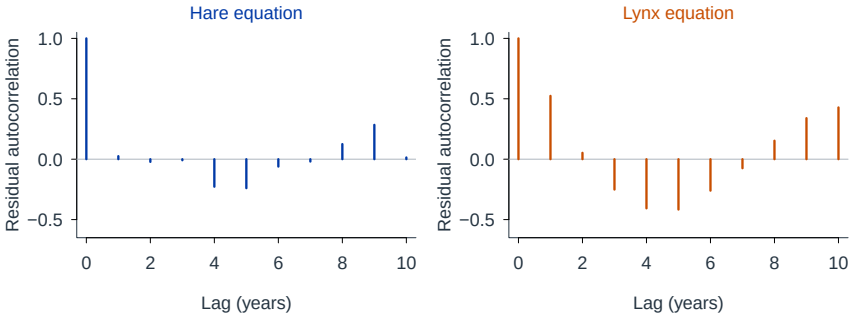


The lag-one residual correlations are 0.025 for hare and 0.524 for lynx.

Real data: Hudson's Bay Company pelt purchases, in thousands; `astsa::Hare` and `astsa::Lynx`, 1845–1935.

Do the residuals leave predictive information unexplained?

Under the conditional-mean model, innovations are uncorrelated across years.



The lag-one residual correlations are 0.025 for hare and 0.524 for lynx.

The lynx residuals suggest that earlier observations may help prediction.

Real data: Hudson's Bay Company pelt purchases, in thousands; `astsa::Hare` and `astsa::Lynx`, 1845–1935.

Which directions would the fitted model imply?

Suppose the fitted VAR(1) is the stationary causal model, with iid $N_2(0, \Sigma_Z)$ innovations and positive-definite Σ_Z.

$$\widehat{A}_{12} = -0.372 \neq 0, \qquad \widehat{A}_{21} = 0.206 \neq 0.$$

Which directions of Granger causality in mean follow for the log records?

Which directions would the fitted model imply?

Suppose the fitted VAR(1) is the stationary causal model, with iid $N_2(0, \Sigma_Z)$ innovations and positive-definite Σ_Z.

$$\widehat{A}_{12} = -0.372 \neq 0, \qquad \widehat{A}_{21} = 0.206 \neq 0.$$

Which directions of Granger causality in mean follow for the log records?

Each log series Granger-causes the other under these assumptions.

Would this identify the effect of changing an animal population?

Continue with longer horizons and more lags

- For the delayed response, does the extra history help two periods ahead?
- Fit and forecast the rotation and hare–lynx examples in the workbook.
- Next: several lags, recovering innovations, and joint prediction regions.