STA 542: Introduction to Time Series Analysis

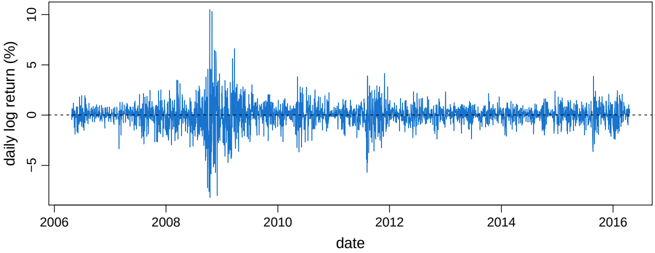
Lecture 2

Lasse Vuursteen

Duke University

Wed Aug 26, 2026

The observed record is finite; the model extend beyond it



- The plot contains 2,517 DJIA daily log returns, April 2006–April 2016. These numbers form one observed window $x_1, \ldots, x_T$.
- A process model also describes values before and after this window. Its equations are therefore required for every $t \in \mathbb{Z}$.

Real data: `astsa::djia`; returns shown in percent.

Writing a recursion does not yet define a process

Lecture 1 introduced three generating equations:

$$\text{MA(1):}\quad X_t = Z_t + \theta Z_{t-1},$$
$$\text{AR(1):}\quad X_t = \phi X_{t-1} + Z_t,$$
$$\text{GARCH(1,1):}\quad X_t = \sqrt{V_t} Z_t, \qquad V_t = \omega + \alpha X_{t-1}^2 + \beta V_{t-1}.$$

- The MA right-hand side already defines X_t from the driving sequence.
- AR and GARCH feed unknown process values back into themselves. We still have to find a joint process that satisfies the equations.

A process law must specify every finite window consistently

Recall that a process law consists of the joint CDFs

$$F_{t_1,\ldots,t_k}(c_1,\ldots,c_k) = \mathbb{P}(X_{t_1} \le c_1,\ldots,X_{t_k} \le c_k), \qquad t_1 < \cdots < t_k,$$

for every finite set of times.

- Marginalizing any window must reproduce the law already assigned to the smaller window.
- Choosing the CDFs separately creates infinitely many consistency constraints. Direct specification needs additional structure.

Gaussian laws reduce direct specification to two functions

Choose a mean function $\mu : \mathbb{Z} \to \mathbb{R}$ and a symmetric covariance kernel $\Sigma : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$. Require

$$(X_{t_1}, \ldots, X_{t_k}) \sim N\left(\begin{bmatrix} \mu(t_1) \\ \vdots \\ \mu(t_k) \end{bmatrix}, \left[\Sigma(t_i, t_j)\right]_{i,j=1}^{k} \right) \quad \text{for every } t_1 < \cdots < t_k.$$

- $\mu(t) = \mathbb{E}[X_t]$ and $\Sigma(s,t) = \mathrm{Cov}(X_s, X_t)$.
- Gaussian marginalization supplies consistency automatically.
- Validity still requires $a^\mathsf{T}[\Sigma(t_i, t_j)]_{i,j=1}^{k} a \geq 0$ for every window and every $a \in \mathbb{R}^k$.

Pairwise plausible covariances can be jointly impossible

Suppose $\Sigma : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ is given by

$$\Sigma(s,t) = \begin{cases} 1, & s = t, \\ 0.8, & |s-t| = 1, \\ 0, & |s-t| \geq 2. \end{cases}$$

Three consecutive times yield $\det[\Sigma(t_i, t_j)] = 1 - 2(0.8)^2 = -0.28$.

- The window matrix is not nonnegative definite.
- No process has these covariances. When correlations vanish beyond lag one, the neighbor correlation is at most $1/2$.

Some valid covariance kernels

For $v > 0$,

$$\Sigma(s,t) = v\mathbf{1}\{s = t\} \qquad \text{Gaussian white noise with variance } v,$$

$$\Sigma(s,t) = v\cos\{\lambda(s-t)\} \qquad \text{trigonometric process,}$$

$$\Sigma(s,t) = v\rho^{|s-t|}, \quad |\rho| < 1 \qquad \text{AR(1) with Gaussian white noise.}$$

- Each Σ is the covariance of a process already constructed, so every window matrix is nonnegative definite.
- A Gaussian model is any collection of valid pairs (μ, Σ).

A transition law can build later windows recursively

A time-homogeneous first-order Markov specification gives one conditional CDF

$$K(c \mid x) := \mathbb{P}(X_t \leq c \mid X_{t-1} = x), \qquad \text{the same for every } t.$$

- A law for X_1 and the kernel K determine the law of $(X_1, \ldots, X_T)$ by repeated conditioning.
- The linear Gaussian example uses $X_t \mid X_{t-1} = x \sim N(\phi x, \sigma^2)$.
- This construction starts at time 1. A process on $\mathbb{Z}$ has no first time.

A generating equation moves consistency into a solution problem

Fix a rule $g : \mathbb{R}^{1+q} \times \mathbb{R}^p \to \mathbb{R}$ and a noise specification $\mathcal{Q}$, a set of candidate process laws for (Z_t). Require

$$X_t = g\big(Z_t, Z_{t-1}, \ldots, Z_{t-q}, X_{t-1}, \ldots, X_{t-p}\big), \qquad t \in \mathbb{Z}.$$

- (Z_t) is the *driving sequence*; p and q count process lags and noise lags.
- The display is a collection of constraints on two processes. It may admit one solution, several solutions, or no solution.

A generating equation moves consistency into a solution problem

Fix a rule $g : \mathbb{R}^{1+q} \times \mathbb{R}^p \to \mathbb{R}$ and a noise specification $\mathcal{Q}$, a set of candidate process laws for (Z_t). Require

$$X_t = g\big(Z_t, Z_{t-1}, \ldots, Z_{t-q}, X_{t-1}, \ldots, X_{t-p}\big), \qquad t \in \mathbb{Z}.$$

- (Z_t) is the *driving sequence*; p and q count process lags and noise lags.
- The display is a collection of constraints on two processes. It may admit one solution, several solutions, or no solution.

Possible extension: have the rule depend on the time t ($g \equiv g_t$) – wise to avoid for now.

A solution is a joint process (X, Z)

Solution of a generating equation

A solution is a joint process $((X_t, Z_t))_{t\in\mathbb{Z}}$ such that

1. the law of (Z_t) belongs to $\mathcal{Q}$; and
2. with probability one, the equation holds at every $t \in \mathbb{Z}$.

- A law for X alone cannot be a solution because the equation also names Z.
- The model induced by the equation retains the X-process laws from all joint solutions.

Warm up: does a WN specification have a solution?

A process (Z_t) is $\mathrm{WN}(0, \sigma^2)$ if

$$\mathbb{E}[Z_t] = 0, \qquad \mathrm{Var}(Z_t) = \sigma^2, \qquad \mathrm{Cov}(Z_s, Z_t) = 0 \text{ for } s \neq t.$$

In this case, there is no generating equation. Does this specification have a solution?

Warm up: does a WN specification have a solution?

A process (Z_t) is $\mathrm{WN}(0, \sigma^2)$ if

$$\mathbb{E}[Z_t] = 0, \qquad \mathrm{Var}(Z_t) = \sigma^2, \qquad \mathrm{Cov}(Z_s, Z_t) = 0 \text{ for } s \neq t.$$

In this case, there is no generating equation. Does this specification have a solution?

- Yes: taking Z_t as i.i.d. $N(0, \sigma^2)$ draws gives a solution.
- Z_t i.i.d. with $\mathbb{P}(Z_t = -1) = \mathbb{P}(Z_t = 1) = 1/2$ also gives a solution.
- $\mathcal{Q}$ contains many laws; so the specification offers many solutions!

Warm up II: An MA(q) equation uses only noise lags

Recall the definition of MA(q): For coefficients $\theta_1, \ldots, \theta_q$,

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \qquad t \in \mathbb{Z},$$

with $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$.

Warm up II: An MA(q) equation uses only noise lags

Recall the definition of MA(q): For coefficients $\theta_1, \dots, \theta_q$,

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \qquad t \in \mathbb{Z},$$

with $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$.

- $q = 0$ recovers $X_t = Z_t$: the white noise specification itself!
- Once a driving sequence (Z_t) is given,

  $$Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

  is defined for every t. Setting X_t equal to that expression constructs a joint process (X, Z) directly. This works for any $q \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$.
- Given the noise law for (Z_t), the solution for (X_t) is fully determined.

The noise specification determines how much the equation says

The assumptions used in the course form a strict nesting:

$$\{\text{i.i.d. } N(0,\sigma^2)\} \subsetneq \{\text{i.i.d., mean } 0, \text{ variance } \sigma^2\} \subsetneq \{\mathrm{WN}(0,\sigma^2)\}.$$

- The left set contains one driving-process law: a *complete* specification.
- The other sets contain many laws: *partial* specifications. Equal first two moments need not imply equal likelihoods, or even equal existence conclusions.

Autoregressive feedback leaves the process unresolved

Recall the definition of an AR(1) process:

$$X_t = \phi X_{t-1} + Z_t, \qquad 0 < |\phi| < 1, \qquad (Z_t) \sim \mathrm{WN}(0, \sigma^2).$$

- The right-hand side still contains the unknown X_{t-1}.
- Running forward would require an initial value. On $\mathbb{Z}$, every proposed initial time has an earlier predecessor.
- We instead iterate the equation backward and inspect the remainder.

Backward iteration exposes the missing initial condition

Repeated substitution gives, for every finite m,

$$X_t = Z_t + \phi Z_{t-1} + \cdots + \phi^m Z_{t-m} + \phi^{m+1} X_{t-m-1}.$$

- The finite noise sum is known once the driving sequence is given.
- The final term carries the unobserved state from the remote past.
- A solution with bounded second moments should make that remainder vanish in mean square when $|\phi| < 1$.

The backward calculation suggests an infinite-series candidate

Dropping the remainder suggests

$$X_t^\star = \sum_{j=0}^{\infty} \phi^j Z_{t-j}.$$

Two questions remain.

1. Does the random series converge?
2. If it converges, does its limit satisfy the AR equation?

A formal expression is not yet a random variable. We answer the questions in that order.

White-noise orthogonality makes the series converge

By the mean-square Cauchy criterion, a limit exists if the squared distance between late partial sums tends to zero. Here

$$\mathbb{E}\left[\left(\sum_{m<j\leq m'} \phi^j Z_{t-j}\right)^2\right] = \sigma^2 \sum_{m<j\leq m'} \phi^{2j} \longrightarrow 0.$$

- Zero cross-covariances remove all cross terms.
- The geometric tail vanishes because $|\phi| < 1$.
- The partial sums are Cauchy in mean square. The criterion therefore gives a finite-second-moment random variable $X_t^\star$.

Substitution verifies the candidate

$$\begin{aligned} \phi X^{\star}_{t-1} + Z_t &= \phi \sum_{j=0}^{\infty} \phi^j Z_{t-1-j} + Z_t \\ &= \sum_{j=0}^{\infty} \phi^j Z_{t-j} \\ &= X^{\star}_t. \end{aligned}$$

Thus $(X^\star, Z)$ is a solution of the AR(1) equation.

- Convergence establishes that $X^\star_t$ is defined.
- Substitution establishes that the defined process solves the equation. An infinite-series argument needs both steps.

Existence does not give uniqueness

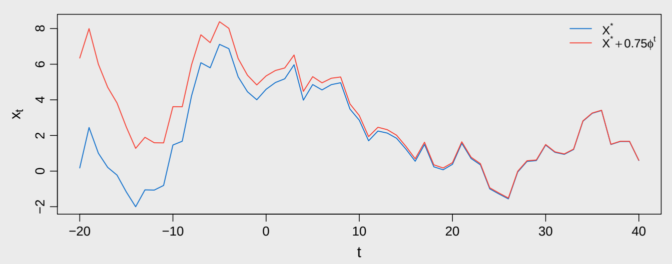
If W is any random variable defined jointly with the driving sequence,

$$X_t = X_t^\star + \phi^t W$$

also satisfies $X_t = \phi X_{t-1} + Z_t$.

- The added term solves the homogeneous equation $H_t = \phi H_{t-1}$.
- Conversely, the difference of any two solutions driven by the same Z has this form.
- The equation alone therefore admits an entire family of solutions.

The extra solution remembers the remote past



Simulated data: AR(1) with $\phi = 0.9$ and a common i.i.d. $N(0,1)$ driving sequence on $t = -20, \ldots, 40$. Blue starts from the invariant $N\big(0, 1/(1-\phi^2)\big)$ law and runs the recursion forward; red is the same path plus $0.75\phi^t$.

- Both paths satisfy $X_t = \phi X_{t-1} + Z_t$ with the same Z.
- Their difference fades forward and grows without bound backward.

Bounded second moments select one AR solution

Impose the selection criterion

$$\sup_{t \in \mathbb{Z}} \mathbb{E}[X_t^2] < \infty.$$

- $X^\star$ qualifies, with $\mathrm{Var}(X_t^\star) = \sigma^2/(1 - \phi^2)$.
- For nonzero W, the term $\phi^t W$ is unbounded in mean square as $t \to -\infty$.
- Hence $X^\star$ is the unique solution with bounded second moments. It will also be the unique weakly stationary solution.

Causality records the direction of dependence

A solution is *causal* when each X_t is a function of present and past driving noise sequence. The selected solution for $|\phi| < 1$ is causal:

$$X_t^\star = \sum_{j\geq 0} \phi^j Z_{t-j}. \qquad |\phi| > 1 : \quad X_t = -\sum_{j\geq 1} \phi^{-j} Z_{t+j}.$$

- The second series is the bounded solution when $|\phi| > 1$, but it depends on future noise.
- Boundedness and causality are distinct selection requirements.

Causal i.i.d. noise becomes an innovation

If (Z_t) is i.i.d., then Z_t is independent of the past noise that determines $X_{t-1}, X_{t-2}, \dots$. Consequently,

$$\mathbb{E}[Z_t \mid X_{t-1}, X_{t-2}, \dots] = 0, \qquad \mathbb{E}[X_{t-h} Z_t] = 0, \quad h \geq 1.$$

- Causality is doing the work: the past X values contain no future driving variables.
- White noise alone gives zero unconditional covariances. It does not give the displayed conditional statement.

A complete Gaussian noise law supplies a likelihood

Under $Z_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and the selected stationary solution,

$$p_{\phi,\sigma^2}(x_1, \ldots, x_T) = f_{0,\,\sigma^2/(1-\phi^2)}(x_1) \prod_{t=2}^{T} f_{\phi x_{t-1},\,\sigma^2}(x_t),$$

where $f_{m,v}$ is the $N(m, v)$ density.

A complete Gaussian noise law supplies a likelihood

Under $Z_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ and the selected stationary solution,

$$p_{\phi,\sigma^2}(x_1, \ldots, x_T) = f_{0,\, \sigma^2/(1-\phi^2)}(x_1) \prod_{t=2}^{T} f_{\phi x_{t-1},\, \sigma^2}(x_t),$$

where $f_{m,v}$ is the $N(m, v)$ density.

The selected solution is the Gaussian process with $\mu(t) = 0$ and

$$\Sigma(s, t) = \frac{\sigma^2}{1 - \phi^2}\, \phi^{|s-t|}.$$

This is the AR(1) kernel from the direct-specification examples.

White noise alone supplies moments, not a likelihood

Every selected AR(1) solution driven by $\mathrm{WN}(0, \sigma^2)$ satisfies

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{1 - \phi^2}, \qquad \mathrm{Cov}(X_t, X_{t-h}) = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}.$$

White noise alone supplies moments, not a likelihood

Every selected AR(1) solution driven by $\mathrm{WN}(0, \sigma^2)$ satisfies

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{1 - \phi^2}, \qquad \mathrm{Cov}(X_t, X_{t-h}) = \frac{\sigma^2 \phi^{|h|}}{1 - \phi^2}.$$

- They do not determine a density. Maximum likelihood requires a complete noise law.

White noise alone supplies moments, not a likelihood

Every selected $\mathrm{AR}(1)$ solution driven by $\mathrm{WN}(0, \sigma^2)$ satisfies

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{1-\phi^2}, \qquad \mathrm{Cov}(X_t, X_{t-h}) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}.$$

- They do not determine a density. Maximum likelihood requires a complete noise law.
- However, these population identities support moment methods!

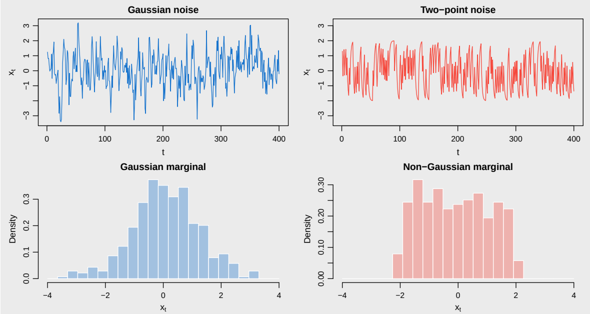
White noise alone supplies moments, not a likelihood

Every selected AR(1) solution driven by $\mathrm{WN}(0, \sigma^2)$ satisfies

$$\mathsf{Var}(X_t) = \frac{\sigma^2}{1-\phi^2}, \qquad \mathsf{Cov}(X_t, X_{t-h}) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}.$$

- They do not determine a density. Maximum likelihood requires a complete noise law.
- However, these population identities support moment methods!
- Learning the moments from one path still needs stationarity and an averaging condition.

The same AR moments can hide different process laws



- Both causal AR(1) processes use $\phi = 0.5$ and i.i.d. noise with mean zero and variance one, so their means and covariances agree.
- Gaussian and two-point noise produce different marginal shapes and different likelihoods.

Simulated data; seed 542, 100-observation burn-in. i.i.d. $N(0,1)$ versus i.i.d. $\mathbb{P}(Z_t = \pm 1) = 1/2$ driving noise.

At $\phi = 1$, boundedness selects no solution

The AR equation becomes the random walk

$$X_t = X_{t-1} + Z_t. \qquad X_t - X_s = \sum_{s<j\leq t} Z_j, \qquad \mathrm{Var}(X_t - X_s) = |t-s|\sigma^2.$$

- The homogeneous equation is $H_t = H_{t-1}$, so any two solutions differ by a constant random variable.
- No solution has second moments bounded over all $t \in \mathbb{Z}$.
- The selection criterion that worked for $|\phi| < 1$ leaves the entire family undistinguished.

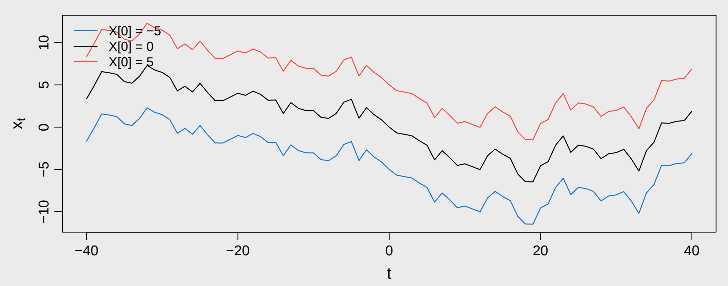
An example solutions to the random walk

Given a driving sequence (Z_t) and a random variable X_0, the equation $X_t = X_{t-1} + Z_t$ unfolds to

$$X_t = \begin{cases} X_0 + \sum_{j=1}^{t} Z_j & t > 0, \\ X_0 & t = 0, \\ X_0 - \sum_{j=t+1}^{0} Z_j & t < 0. \end{cases}$$

- Forward: $X_t = X_{t-1} + Z_t$ for $t \geq 1$.
- The special $t = 0$ check matters: $X_{-1} = X_0 - Z_0$, so indeed $X_0 = X_{-1} + Z_0$.
- Backward: $X_{-1} = X_0 - Z_0$, then $X_t = X_{t+1} - Z_{t+1}$ for $t \leq -2$.
- The resulting process is an *anchored* by its value at $t = 0$ (this is a choice, we could have chosen any other time).

A random walk is defined only up to its anchor and noise law



- The three paths use the same i.i.d. Gaussian increments and different values at $t = 0$.
- Their pairwise differences remain constant. The increment law does not determine the level law.

Some equations have no solution

Consider, for every $t \in \mathbb{Z}$,

$$x_t = 1 + a|x_{t-1}|, \qquad a \geq 0. \quad \Longrightarrow \quad x_t \geq 1 + a + \cdots + a^m \quad \text{for every } m.$$

- Any solution would satisfy $x_t \geq 1$, so backward substitution gives the displayed lower bound.
- The bound diverges as $m \to \infty$ if $a \geq 1$: No finite real sequence can satisfy the equation on all of $\mathbb{Z}$.
- The update rule does not guarantee a two-sided process.

GARCH(1,1) hides a random-coefficient recursion

Recall the GARCH(1,1) definition: $X_t = \sqrt{V_t} Z_t$ and $V_t = \omega + \alpha X_{t-1}^2 + \beta V_{t-1}$,

GARCH(1,1) hides a random-coefficient recursion

Recall the GARCH(1,1) definition: $X_t = \sqrt{V_t} Z_t$ and $V_t = \omega + \alpha X_{t-1}^2 + \beta V_{t-1}$, substitution gives

$$V_t = \omega + A_{t-1} V_{t-1}, \qquad A_t := \alpha Z_t^2 + \beta.$$

GARCH(1,1) hides a random-coefficient recursion

Recall the GARCH(1,1) definition: $X_t = \sqrt{V_t} Z_t$ and $V_t = \omega + \alpha X_{t-1}^2 + \beta V_{t-1}$, substitution gives

$$V_t = \omega + A_{t-1} V_{t-1}, \qquad A_t := \alpha Z_t^2 + \beta.$$

- Backward iteration multiplies random coefficients A_t (recognize the form

$$x_t = 1 + a|x_{t-1}|, \qquad a \geq 1. \quad \Longrightarrow \quad x_t \geq 1 + a + \cdots + a^m \quad \text{for every } m.$$

GARCH(1,1) hides a random-coefficient recursion

Recall the GARCH(1,1) definition: $X_t = \sqrt{V_t} Z_t$ and $V_t = \omega + \alpha X_{t-1}^2 + \beta V_{t-1}$, substitution gives

$$V_t = \omega + A_{t-1} V_{t-1}, \qquad A_t := \alpha Z_t^2 + \beta.$$

- Backward iteration multiplies random coefficients A_t (recognize the form

$$x_t = 1 + a|x_{t-1}|, \qquad a \geq 1. \quad \Longrightarrow \quad x_t \geq 1 + a + \cdots + a^m \quad \text{for every } m.$$

- Existence depends on whether those products shrink quickly enough.

Geometric growth determines GARCH existence

Suppose (Z_t) is i.i.d. and $\mathbb{E}[|\log(\alpha Z_1^2 + \beta)|] < \infty$. The backward candidate is

$$V_t = \omega\{1 + A_{t-1} + A_{t-1}A_{t-2} + \cdots\}.$$

- The necessary and sufficient condition required for existence is $\mathbb{E}[\log(\alpha Z_1^2 + \beta)] < 0$.[1]

[1] We will not prove this here; a proof is in Nelson (1990).

Equal noise moments can give opposite existence conclusions

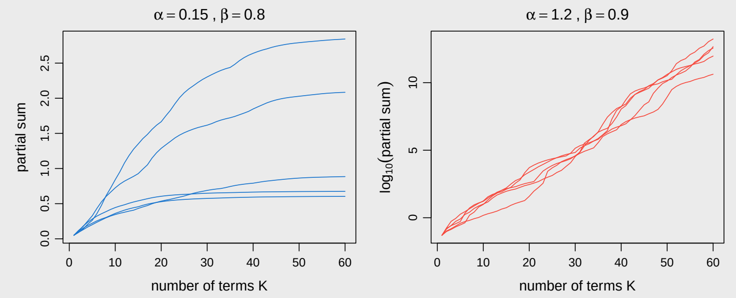
At $(\alpha, \beta) = (0.3, 0.75)$, consider two i.i.d. noise laws with mean zero and variance one:

$$\begin{aligned} Z_1 \sim N(0,1) \quad &: \quad \mathbb{E}[\log(0.3Z_1^2 + 0.75)] \approx -0.0074 < 0, \\ \mathbb{P}(Z_1 = -1) = \mathbb{P}(Z_1 = 1) = 1/2 \quad &: \quad \log(1.05) \approx 0.0488 > 0. \end{aligned}$$

Gaussian noise yields a finite positive solution; two-point noise does not. A partial mean-and-variance specification contains both candidate laws.


GARCH(1,1)
$X_t = \sqrt{V_t} Z_t$ and $V_t = \omega + \alpha X_{t-1}^2 + \beta V_{t-1}$, with $\mathbb{E}[\log(\alpha Z_1^2 + \beta)] > 0$
THE SOLUTION DOES NOT EXIST

In a simulation: the backward GARCH candidate can stabilize or explode



- Each curve uses one simulated Gaussian past. Increasing K means looking farther into that past; the horizontal axis is *look-back depth*, not time.
- On the left, $\mathbb{E}[\log A_1] \approx -0.0704$: the newly added products shrink and $S_{K,t}$ approaches a finite limit. On the right, $\mathbb{E}[\log A_1] \approx 0.5336$: the products grow and $S_{K,t}$ explodes. The right vertical axis is logarithmic.

The three equations now have concrete outcomes

Does a generating equation define a process?

MA(1):	any noise law gives a solution; X is determined by Z,
AR(1):	$\lvert\phi\rvert < 1$ selects a unique bounded (causal) solution,
GARCH:	existence depends on the noise law, at the same parameters.

The three equations now have concrete outcomes

Does a generating equation define a process?

MA(1): any noise law gives a solution; X is determined by Z,

AR(1): $|\phi| < 1$ selects a unique bounded (causal) solution,

GARCH: existence depends on the noise law, at the same parameters.

Existence, together with a selection rule when needed, is what it means for the process to be well defined.

When is a time series model well defined?

Recall: A time series model is a collection of process laws.
It is well defined when every law in the collection arises from a selected solution.

When is a time series model well defined?

Recall: A time series model is a collection of process laws.
It is well defined when every law in the collection arises from a selected solution. For

example:

$$\Big\{\mathcal{L}((X_t)_{t\in\mathbb{Z}}) : X_t = \phi X_{t-1} + Z_t \text{ for } t \in \mathbb{Z}, \\ |\phi| < 1, \quad (Z_t) \sim \mathrm{WN}(0, \sigma^2), \quad \sup_{t\in\mathbb{Z}} \mathbb{E}[|X_t|^2] < \infty\Big\}$$

- Boundedness selects one solution at each (ϕ, σ^2), so every member of the collection exists.

When is a time series model well defined?

Recall: A time series model is a collection of process laws.
It is well defined when every law in the collection arises from a selected solution. For example:

$$\Big\{\mathcal{L}((X_t)_{t\in\mathbb{Z}}) : X_t = \phi X_{t-1} + Z_t \text{ for } t \in \mathbb{Z}, \\ |\phi| < 1, \quad (Z_t) \sim \mathrm{WN}(0, \sigma^2), \quad \sup_{t\in\mathbb{Z}} \mathbb{E}[|X_t|^2] < \infty\Big\}$$

- Boundedness selects one solution at each (ϕ, σ^2), so every member of the collection exists.

Having a model does not make one path informative

Given a well defined time series model, the next natural question is: suppose that we were to observe a single path from this model, what can we learn about the law that generated it? For example, given the model

$$\Big\{\mathcal{L}((X_t)_{t\in\mathbb{Z}}) : X_t = \phi X_{t-1} + Z_t \text{ for } t \in \mathbb{Z}, \\ |\phi| < 1, \quad (Z_t) \sim \mathrm{WN}(0, \sigma^2), \quad \sup_{t\in\mathbb{Z}} \mathbb{E}[|X_t|^2] < \infty\Big\}$$

if we observe a path $x_1, \dots, x_T$ and T is large enough:

- Can we learn the parameters ϕ and σ^2?
- Can we learn the distribution of Z_t?
- And how large should T be?


WE HAVE A WELL-DEFINED FAMILY OF TIME-SERIES LAWS.
SO ONE LONG OBSERVED PATH TELLS US WHICH LAW GENERATED IT, RIGHT?
IT TELLS US, RIGHT?

If every time has its own law, one path provides no repetition

Suppose

$$Z_t \overset{\text{iid}}{\sim} N(0,1), \qquad X_t = \mu_t + Z_t,$$

where $(\mu_t)_{t\in\mathbb{Z}}$ is an unrestricted sequence of unknown means.

If every time has its own law, one path provides no repetition

Suppose

$$Z_t \overset{\text{iid}}{\sim} N(0,1), \qquad X_t = \mu_t + Z_t,$$

where $(\mu_t)_{t\in\mathbb{Z}}$ is an unrestricted sequence of unknown means.

time	1	2	$\cdots$	T	$T+1$
unknown mean	μ_1	μ_2	$\cdots$	μ_T	μ_{T+1}
evidence	x_1	x_2	$\cdots$	x_T	none

- Each observation informs a different mean.
- Increasing T adds one observation and one new unknown.
- Nothing in the observed path constrains the next mean μ_{T+1}.

If every time has its own law, one path provides no repetition

Suppose

$$Z_t \stackrel{\text{iid}}{\sim} N(0,1), \qquad X_t = \mu_t + Z_t,$$

where $(\mu_t)_{t\in\mathbb{Z}}$ is an unrestricted sequence of unknown means.

time	1	2	$\cdots$	T	$T+1$
unknown mean	μ_1	μ_2	$\cdots$	μ_T	μ_{T+1}
evidence	x_1	x_2	$\cdots$	x_T	none

- Each observation informs a different mean.
- Increasing T adds one observation and one new unknown.
- Nothing in the observed path constrains the next mean μ_{T+1}.

To learn from one path, laws at different times must repeat or be related.

Stationarity makes shifted windows comparable

Stationarity

A process is *strictly stationary* if, for every finite window and every 'shift' $h \in \mathbb{Z}$,

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

It is *weakly stationary* if it has finite second moments and $\mathbb{E}[X_t] = \mu$, $\text{Cov}(X_t, X_{t+h}) = \gamma(h)$ for some function $\gamma : \mathbb{Z} \to \mathbb{R}$.

Stationarity makes shifted windows comparable

Stationarity

A process is *strictly stationary* if, for every finite window and every 'shift' $h \in \mathbb{Z}$,

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

It is *weakly stationary* if it has finite second moments and $\mathbb{E}[X_t] = \mu$, $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$ for some function $\gamma : \mathbb{Z} \to \mathbb{R}$.

- Strict stationarity compares full window distributions.

Stationarity makes shifted windows comparable

Stationarity

A process is *strictly stationary* if, for every finite window and every 'shift' $h \in \mathbb{Z}$,

$$(X_{t_1}, \dots, X_{t_k}) \overset{d}{=} (X_{t_1+h}, \dots, X_{t_k+h}).$$

It is *weakly stationary* if it has finite second moments and $\mathbb{E}[X_t] = \mu$, $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$ for some function $\gamma : \mathbb{Z} \to \mathbb{R}$.

- Strict stationarity compares full window distributions.
- Weak stationarity retains only means and covariances.

Stationarity makes shifted windows comparable

Stationarity

A process is *strictly stationary* if, for every finite window and every 'shift' $h \in \mathbb{Z}$,

$$(X_{t_1}, \ldots, X_{t_k}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_k+h}).$$

It is *weakly stationary* if it has finite second moments and $\mathbb{E}[X_t] = \mu$, $\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$ for some function $\gamma : \mathbb{Z} \to \mathbb{R}$.

- Strict stationarity compares full window distributions.
- Weak stationarity retains only means and covariances.
- Note: Strict stationarity is a stronger than weak stationarity.

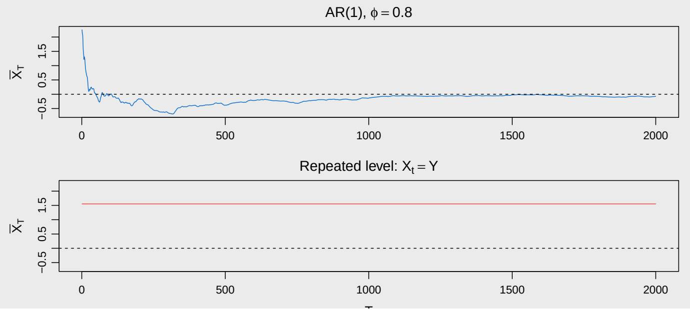
Stationarity alone does not create repeated information

Let Y have mean μ and variance $\eta^2 > 0$, and set

$$X_t = Y \quad \text{for every } t. \qquad \bar{X}_T = Y \quad \text{for every } T.$$

- Every shifted window has exactly the same law, so the process is strictly stationary.
- One path repeats one draw from the distribution of Y forever. Its sample mean does not approach μ.
- Stationarity supplies comparability across time, not fresh information.

A stable AR path averages; a repeated level does not



- Both simulated processes are strictly stationary.
- The AR running mean moves toward its population mean 0; the repeated-level running mean remains equal to its single draw Y.

The second ingredient is ergodicity

Informally, ergodicity says that time averages along one path recover 'space averages' under the process law.

The second ingredient is ergodicity

Informally, ergodicity says that time averages along one path recover ‘space averages’ under the process law.

Mean ergodicity

A weakly stationary process is *mean ergodic* if

$$\mathbb{E}\left[(\bar{X}_T - \mu)^2\right] \longrightarrow 0.$$

- Then $\bar{X}_T$ converges to μ in mean square, and hence in probability.
- The repeated-level process fails: $\bar{X}_T = Y$ for every T.
- Note that $\mathbb{E}X_t = \mu$ under weak stationarity, so $\mathbb{E}\left[(\bar{X}_T - \mu)^2\right] = \mathrm{Var}(\bar{X}_T)$.

Covariance decay makes the mean learnable

Proposition

If a weakly stationary process satisfies $\gamma(h) \to 0$ as $|h| \to \infty$, then it is mean ergodic.

- Distant observations become uncorrelated, so the average along one path can still recover μ.

Proof: expand $\mathrm{Var}(\bar{X}_T)$ into a lag average

Proof. Weak stationarity gives $\mathrm{Cov}(X_s, X_t) = \gamma(t-s)$, so

$$\begin{aligned}\mathrm{Var}(\bar{X}_T) &= \frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T}\gamma(t-s)\\ &= \frac{1}{T}\left[\gamma(0) + 2\sum_{h=1}^{T-1}\left(1-\frac{h}{T}\right)\gamma(h)\right].\end{aligned}$$

Grouping pairs by lag $h = t - s$ produces the weights $1 - h/T$. It remains to show that this average tends to zero.

Proof: both terms of the lag average vanish

Proof (continued). The first term satisfies $\gamma(0)/T \to 0$. For the second, the weights obey $0 < 1 - h/T \le 1$, hence

$$\left| \frac{2}{T} \sum_{h=1}^{T-1} \left(1 - \frac{h}{T}\right) \gamma(h) \right| \le \frac{2}{T} \sum_{h=1}^{T-1} |\gamma(h)| \longrightarrow 0,$$

because averages of a sequence with $|\gamma(h)| \to 0$ also tend to zero. Therefore $\mathbb{E}[(\bar{X}_T - \mu)^2] = \mathrm{Var}(\bar{X}_T) \to 0$.

An MA mean is learnable: covariances vanish past lag q

Write $X_t = \sum_{j=0}^{q} \theta_j Z_{t-j}$ with $\theta_0 = 1$ and $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$. White-noise orthogonality yields

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} & |h| \le q, \\ 0 & |h| > q. \end{cases}$$

- For $|h| \le q$ the windows overlap, and $\gamma(h)$ is σ^2 times the inner product of the shared coefficients.
- For $|h| > q$ the windows are disjoint, so those terms vanish. Thus $\gamma(h) \to 0$, and the proposition applies.

A stable AR mean is learnable: covariances decay geometrically

The selected solution is $X_t^\star = \sum_{j\geq 0} \phi^j Z_{t-j}$ with $|\phi| < 1$. White-noise orthogonality gives, for $h \geq 0$,

$$\gamma(h) = \sigma^2 \sum_{k=0}^{\infty} \phi^{h+k} \phi^k = \frac{\sigma^2 \phi^h}{1-\phi^2}, \qquad \gamma(h) = \frac{\sigma^2 \phi^{|h|}}{1-\phi^2}.$$

- Then $|\gamma(h)| \to 0$ geometrically, so the proposition applies.

A random-walk level is outside the proposition

For the solution anchored at $X_0 = 0$,

$$X_t = \sum_{j=1}^{t} Z_j \ (t > 0), \qquad \mathrm{Var}(X_t) = |t|\sigma^2.$$

- The variance depends on t, so the level is not weakly stationary.
- The proposition does not apply. Its increments $X_t - X_{t-1} = Z_t$ are white noise, hence mean ergodic.

Next week: forecasting under a known process law

- Conditional means, forecast errors, and prediction intervals.
- What weak stationarity contributes to best linear prediction.
- Forecasting the future, interpolating missing values, and backcasting the past.

We return to what one path can reveal in more detail in Workbook 5!