

Non-Stationary Time Series

EC 421, Set 9

Edward Rubin
25 February 2019

Prologue

Schedule

Last Time

Autocorrelation

Today

- Finish autocorrelation
- Brief introduction to nonstationarity
- In-class examples

Upcoming

- **Assignment** this week
- Office hours today.

R showcase

End of class.

Nonstationarity

Nonstationarity

Intro

Let's go back to our assumption of **weak dependence/persistence**

1. **Weakly persistent outcomes**—essentially, x_{t+k} in the distant period $t + k$ weakly correlates with x_t (when k is "big").

Nonstationarity

Intro

Let's go back to our assumption of **weak dependence/persistence**

1. **Weakly persistent outcomes**—essentially, x_{t+k} in the distant period $t + k$ weakly correlates with x_t (when k is "big").

We're essentially saying we need the time series x to behave.

Nonstationarity

Intro

Let's go back to our assumption of **weak dependence/persistence**

1. **Weakly persistent outcomes**—essentially, x_{t+k} in the distant period $t + k$ weakly correlates with x_t (when k is "big").

We're essentially saying we need the time series x to behave.

We'll define this *good behavior* as **stationarity**.

Nonstationarity

Stationarity

Requirements for **stationarity** (a *stationary* time-series process):

Nonstationarity

Stationarity

Requirements for **stationarity** (a *stationary* time-series process):

1. The **mean** of the distribution is independent of time, *i.e.*,

$$E[x_t] = E[x_{t-k}] \text{ for all } k$$

Nonstationarity

Stationarity

Requirements for **stationarity** (a *stationary* time-series process):

1. The **mean** of the distribution is independent of time, *i.e.*,

$$E[x_t] = E[x_{t-k}] \text{ for all } k$$

2. The **variance** of the distribution is independent of time, *i.e.*,

$$\text{Var}(x_t) = \text{Var}(x_{t-k}) \text{ for all } k$$

Nonstationarity

Stationarity

Requirements for **stationarity** (a *stationary* time-series process):

1. The **mean** of the distribution is independent of time, *i.e.*,

$$E[x_t] = E[x_{t-k}] \text{ for all } k$$

2. The **variance** of the distribution is independent of time, *i.e.*,

$$\text{Var}(x_t) = \text{Var}(x_{t-k}) \text{ for all } k$$

3. The **covariance** between x_t and x_{t-k} depends only on k —**not on t** , *i.e.*,

$$\text{Cov}(x_t, x_{t-k}) = \text{Cov}(x_s, x_{s-k}) \text{ for all } t \text{ and } s$$

Nonstationarity

Random walks

Random walks are a famous example of a nonstationary process:

Nonstationarity

Random walks

Random walks are a famous example of a nonstationary process:

$$x_t = x_{t-1} + \varepsilon_t$$

Nonstationarity

Random walks

Random walks are a famous example of a nonstationary process:

$$x_t = x_{t-1} + \varepsilon_t$$

Why?

Nonstationarity

Random walks

Random walks are a famous example of a nonstationary process:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \varepsilon_t$$

Why? $\text{Var}(\mathbf{x}_t) = t\sigma_\varepsilon^2$, which **violates stationary variance**.

Nonstationarity

Random walks

Random walks are a famous example of a nonstationary process:

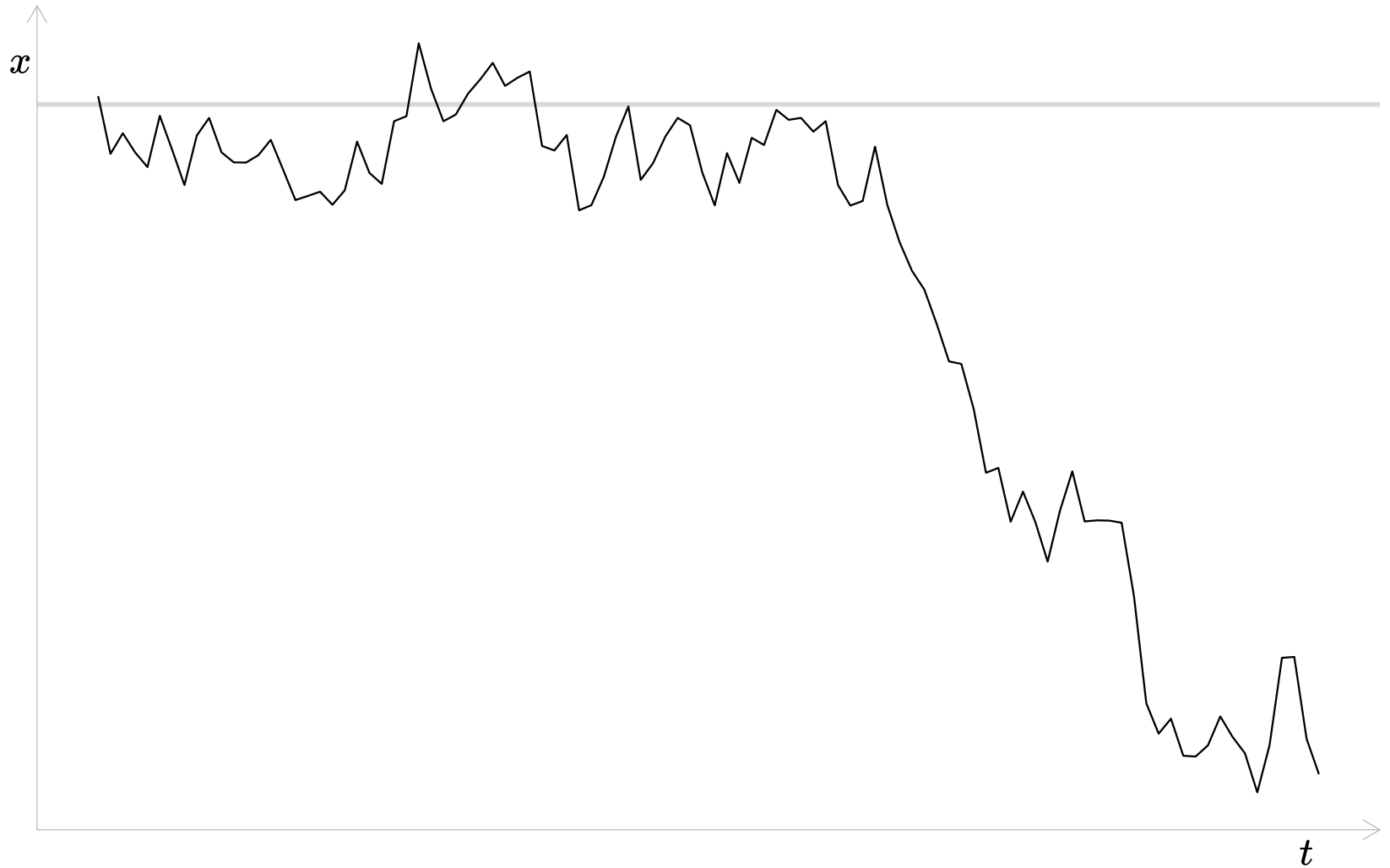
$$x_t = x_{t-1} + \varepsilon_t$$

Why? $\text{Var}(x_t) = t\sigma_\varepsilon^2$, which **violates stationary variance**.

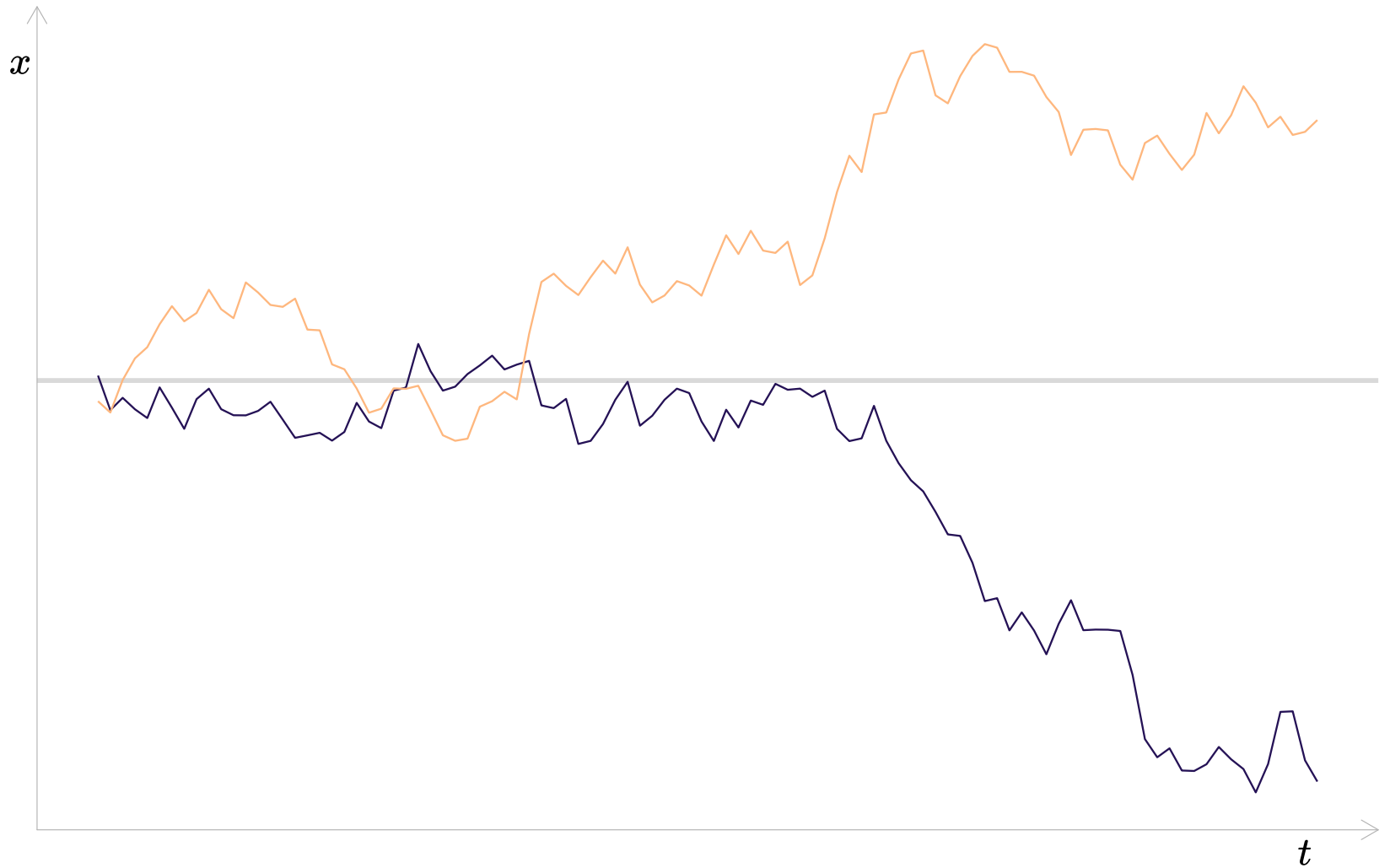
$$\begin{aligned}\text{Var}(x_t) &= \text{Var}(x_{t-1} + \varepsilon_t) \\ &= \text{Var}(x_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\ &= \text{Var}(x_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\ &\dots \\ &= \text{Var}(x_0 + \varepsilon_1 + \dots + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t) \\ &= \sigma_\varepsilon^2 + \dots + \sigma_\varepsilon^2 + \sigma_\varepsilon^2 + \sigma_\varepsilon^2 \\ &= t\sigma_\varepsilon^2\end{aligned}$$

Q: What's the big deal with this violation?

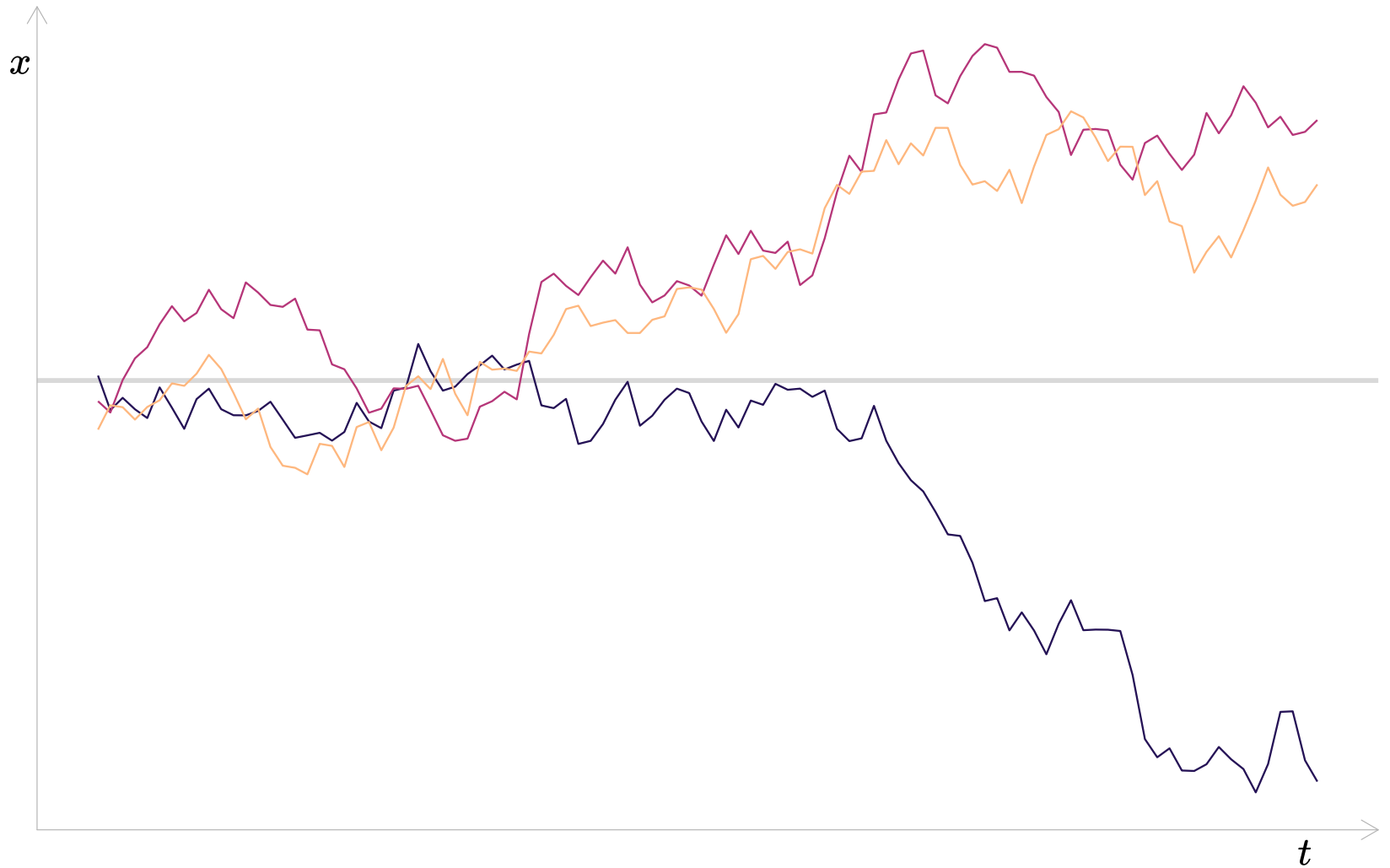
One 100-period random walk



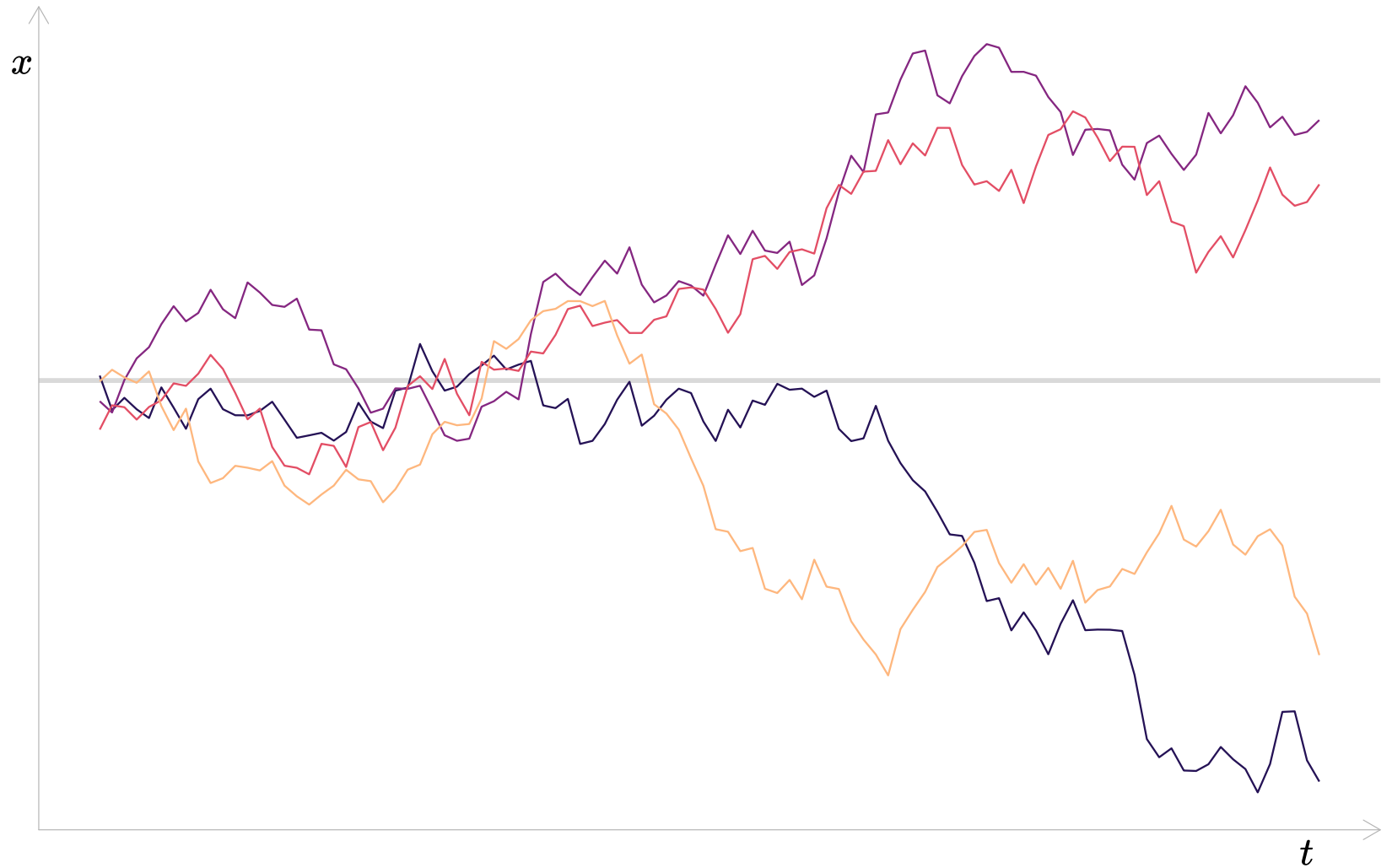
Two 100-period random walks



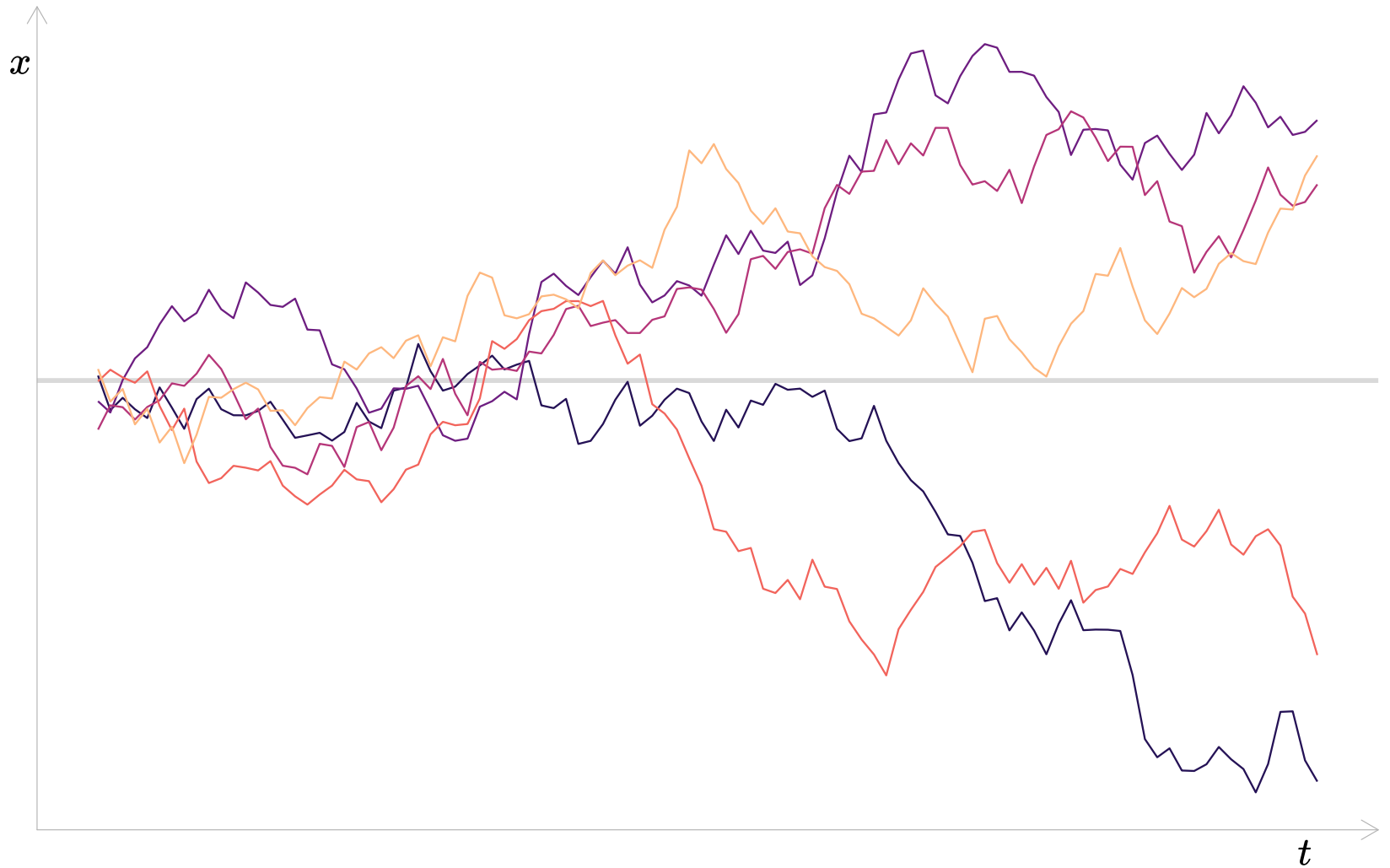
Three 100-period random walks



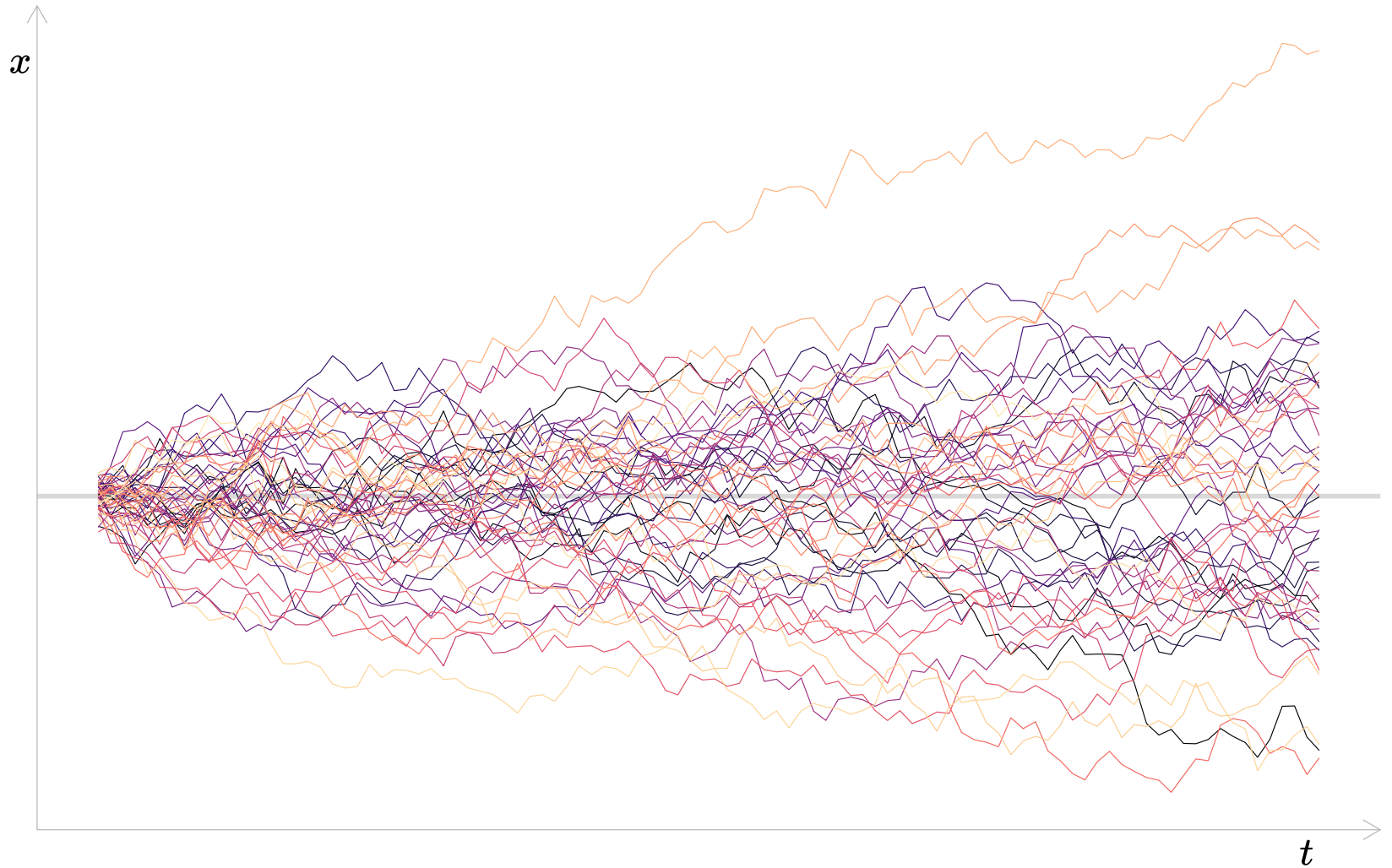
Four 100-period random walks



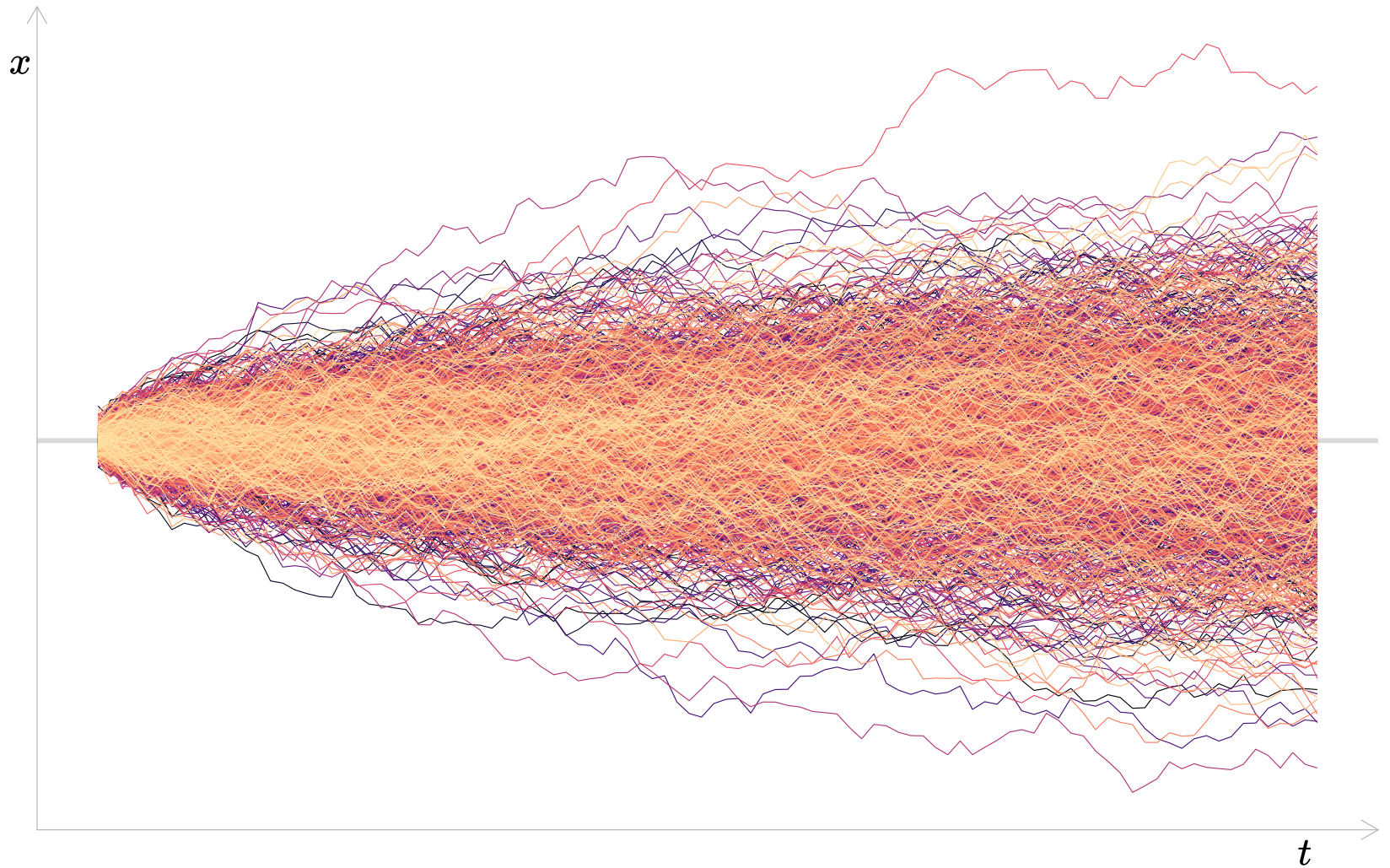
Five 100-period random walks



Fifty 100-period random walks



1,000 100-period random walks



Nonstationarity

Problem

One problem is that nonstationary processes can lead to **spurious** results.

Nonstationarity

Problem

One problem is that nonstationary processes can lead to **spurious** results.

Defintion: Spurious

- not being what it purports to be; false or fake
- apprently but not actually valid

Nonstationarity

Problem

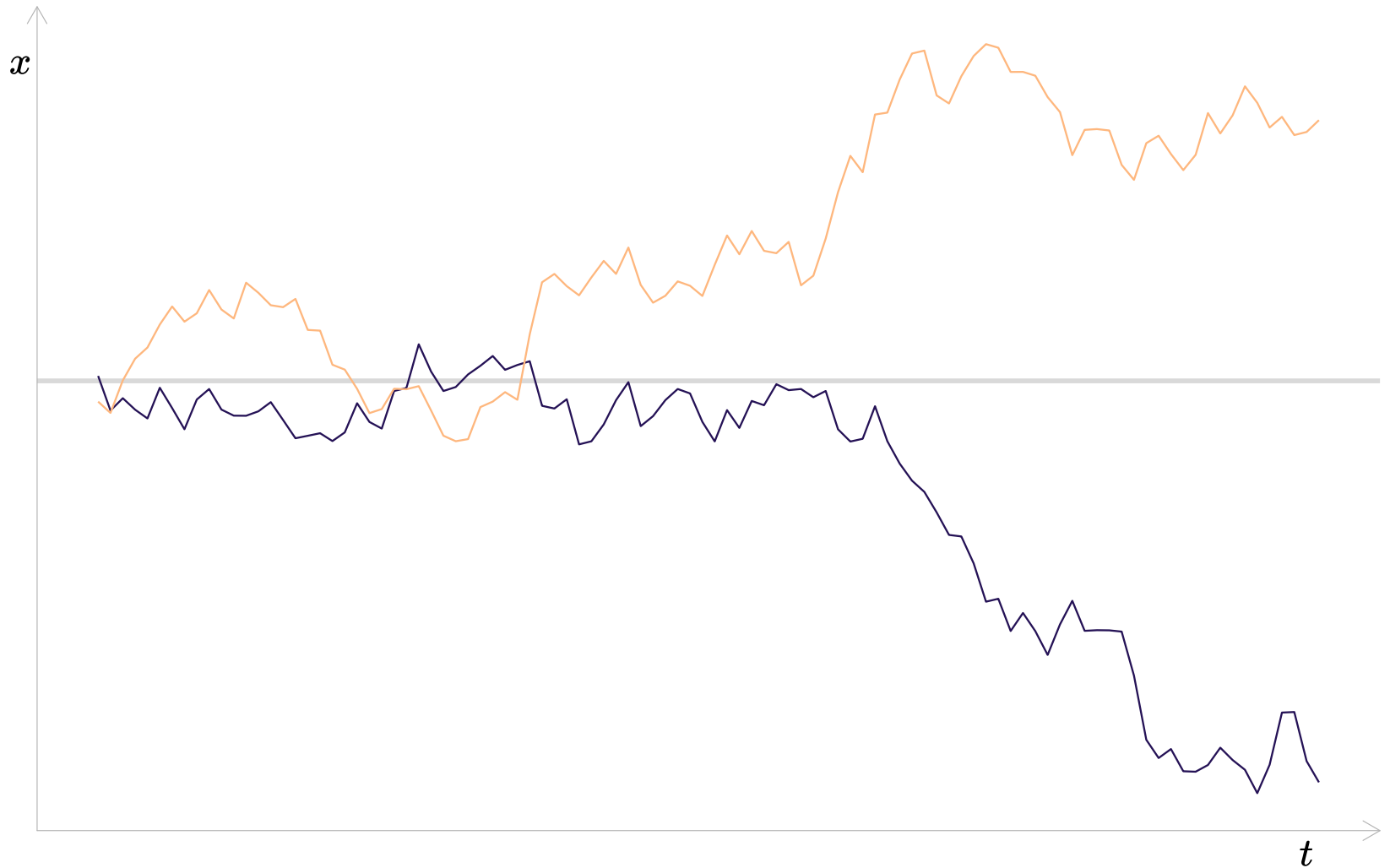
One problem is that nonstationary processes can lead to **spurious** results.

Defintion: Spurious

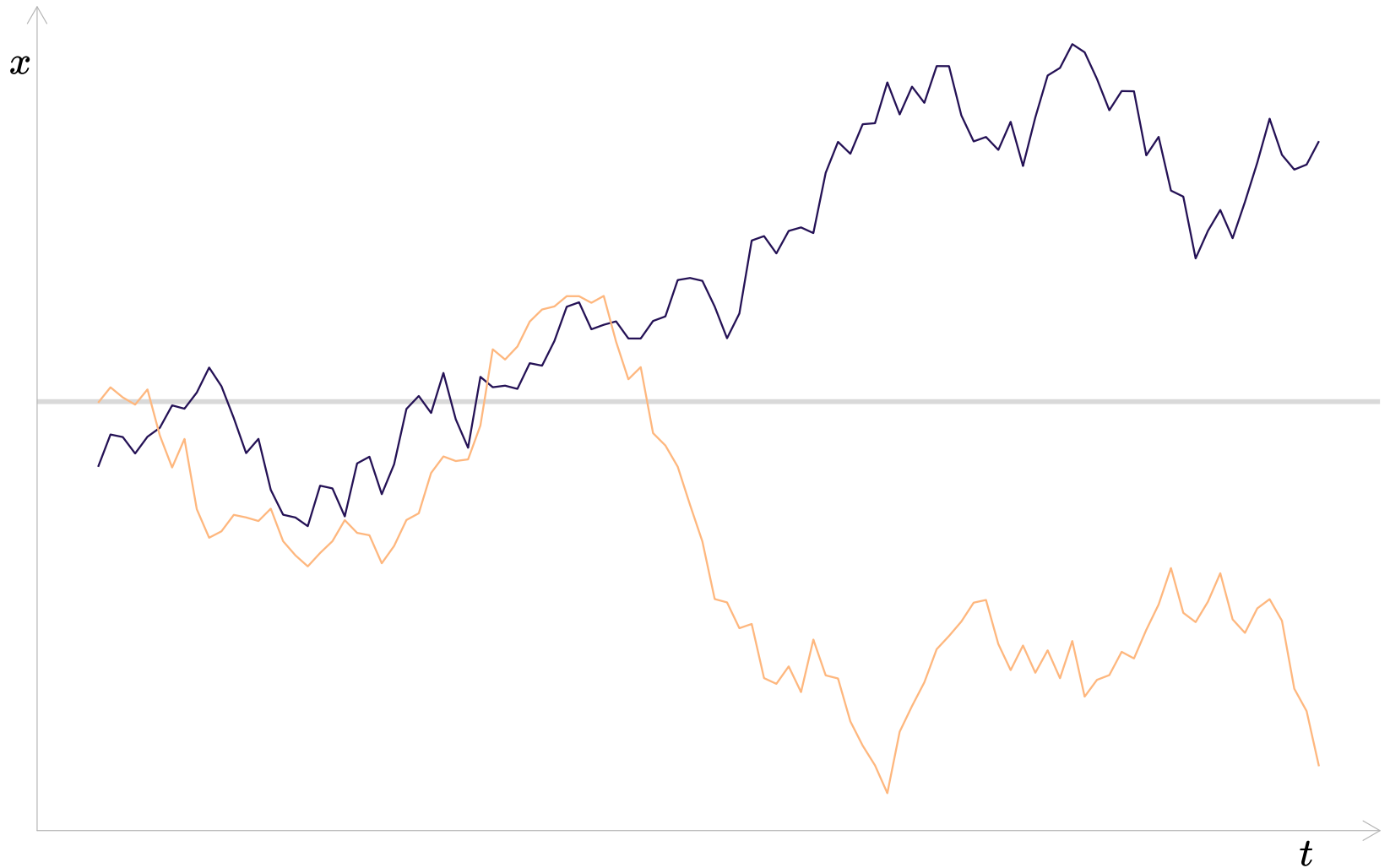
- not being what it purports to be; false or fake
- apprently but not actually valid

Back in 1974, Granger and Newbold showed that when they **generated random walks** and **regressed the random walks on each other**, **77/100 regressions were statistically significant** at the 5% level (should have been approximately 5/100).

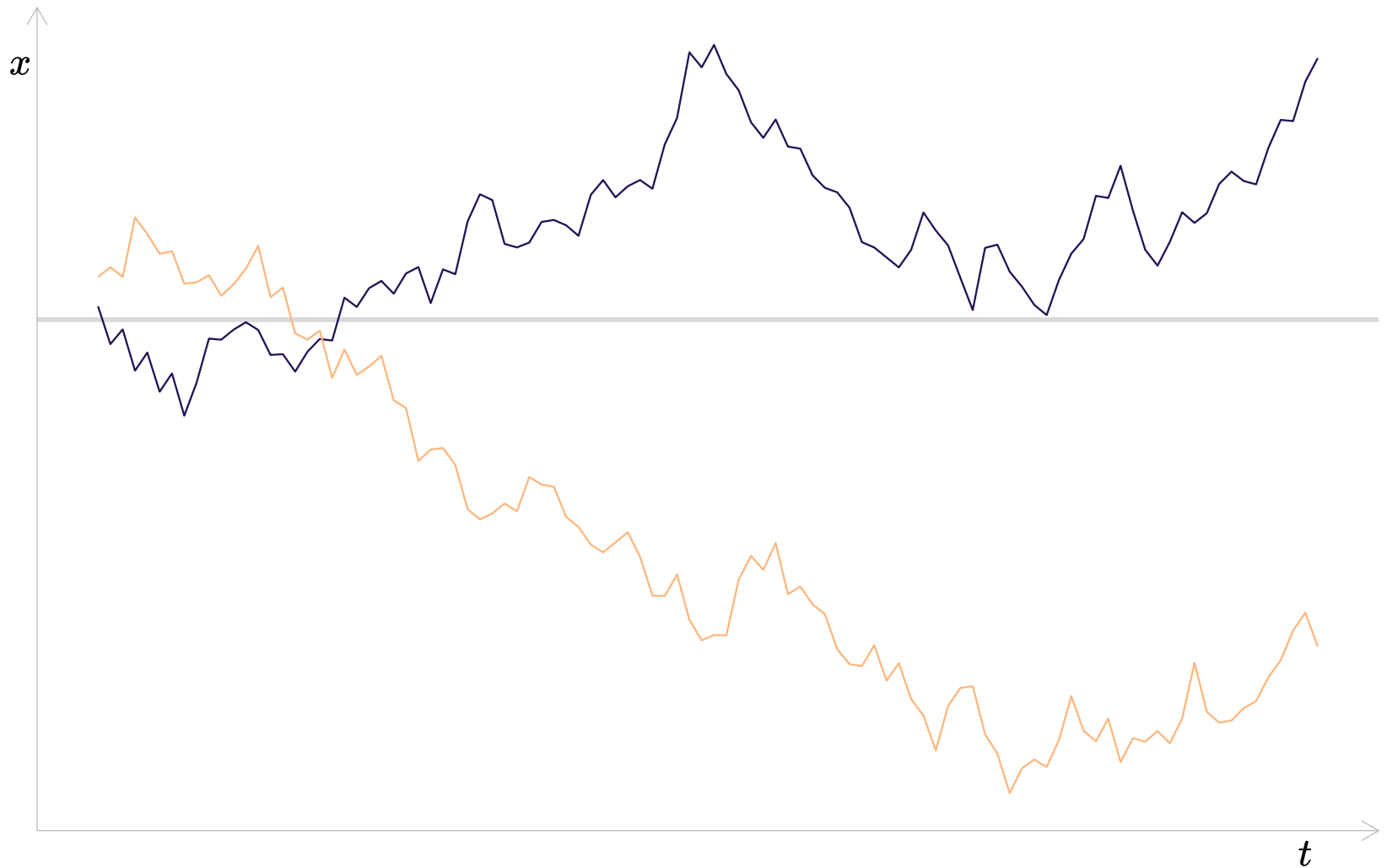
Granger and Newbold simulation example: t statistic ≈ -10.58



Granger and Newbold simulation example: t statistic ≈ -8.92



Granger and Newbold simulation example: t statistic ≈ -7.23



Nonstationarity

Problem

In our data, 74.6 percent of (independently generated) pairs reject the null hypothesis at the 5% level.

Nonstationarity

Problem

In our data, 74.6 percent of (independently generated) pairs reject the null hypothesis at the 5% level.

The point?

Nonstationarity

Problem

In our data, 74.6 percent of (independently generated) pairs reject the null hypothesis at the 5% level.

The point? If our disturbance is nonstationary, we cannot trust plain OLS.

Nonstationarity

Problem

In our data, 74.6 percent of (independently generated) pairs reject the null hypothesis at the 5% level.

The point? If our disturbance is nonstationary, we cannot trust plain OLS.

Random walks are only one example of **nonstationary processes**...

Random walk: $u_t = u_{t-1} + \varepsilon_t$

Nonstationarity

Problem

In our data, 74.6 percent of (independently generated) pairs reject the null hypothesis at the 5% level.

The point? If our disturbance is nonstationary, we cannot trust plain OLS.

Random walks are only one example of **nonstationary processes**...

Random walk: $u_t = u_{t-1} + \varepsilon_t$

Random walk with drift: $u_t = \alpha_0 + u_{t-1} + \varepsilon_t$

Nonstationarity

Problem

In our data, 74.6 percent of (independently generated) pairs reject the null hypothesis at the 5% level.

The point? If our disturbance is nonstationary, we cannot trust plain OLS.

Random walks are only one example of **nonstationary processes**...

Random walk: $u_t = u_{t-1} + \varepsilon_t$

Random walk with drift: $u_t = \alpha_0 + u_{t-1} + \varepsilon_t$

Deterministic trend: $u_t = \alpha_0 + \beta_1 t + \varepsilon_t$

Nonstationarity

A potential solution

Some processes are **difference stationary**, which means we can get back to our stationarity (good behavior) requirement by taking the difference between u_t and u_{t-1} .

Nonstationarity

A potential solution

Some processes are **difference stationary**, which means we can get back to our stationarity (good behavior) requirement by taking the difference between u_t and u_{t-1} .

Nonstationary: $u_t = u_{t-1} + \varepsilon_t$ (a random walk)

Nonstationarity

A potential solution

Some processes are **difference stationary**, which means we can get back to our stationarity (good behavior) requirement by taking the difference between u_t and u_{t-1} .

Nonstationary: $u_t = u_{t-1} + \varepsilon_t$ (a random walk)

Stationary: $u_t - u_{t-1} = u_{t-1} + \varepsilon_t - u_{t-1} = \varepsilon_t$

Nonstationarity

A potential solution

Some processes are **difference stationary**, which means we can get back to our stationarity (good behavior) requirement by taking the difference between u_t and u_{t-1} .

Nonstationary: $u_t = u_{t-1} + \varepsilon_t$ (a random walk)

Stationary: $u_t - u_{t-1} = u_{t-1} + \varepsilon_t - u_{t-1} = \varepsilon_t$

So if we have good reason to believe that our disturbances follow a random walk, we can use OLS on the differences, *i.e.*,

Nonstationarity

A potential solution

Some processes are **difference stationary**, which means we can get back to our stationarity (good behavior) requirement by taking the difference between u_t and u_{t-1} .

Nonstationary: $u_t = u_{t-1} + \varepsilon_t$ (a random walk)

Stationary: $u_t - u_{t-1} = u_{t-1} + \varepsilon_t - u_{t-1} = \varepsilon_t$

So if we have good reason to believe that our disturbances follow a random walk, we can use OLS on the differences, *i.e.*,

$$\begin{aligned}y_t &= \beta_0 + \beta_1 x_t + u_t \\y_{t-1} &= \beta_0 + \beta_1 x_{t-1} + u_{t-1} \\y_t - y_{t-1} &= \beta_1 (x_t - x_{t-1}) + (u_t - u_{t-1}) \\ \Delta y_t &= \beta_1 \Delta x_t + \Delta u_t\end{aligned}$$

Nonstationarity

Testing

Dickey-Fuller and augmented Dickey-Fuller tests are popular ways to test of random walks and other forms of nonstationarity.

Nonstationarity

Testing

Dickey-Fuller and augmented Dickey-Fuller tests are popular ways to test of random walks and other forms of nonstationarity.

Dickey-Fuller tests compare

$H_0: y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ with $|\beta_1| < 1$ (**stationarity**)

$H_a: y_t = y_{t-1} + \varepsilon_t$ (**random walk**)

Nonstationarity

Testing

Dickey-Fuller and augmented Dickey-Fuller tests are popular ways to test of random walks and other forms of nonstationarity.

Dickey-Fuller tests compare

$H_0: y_t = \beta_0 + \beta_1 y_{t-1} + u_t$ with $|\beta_1| < 1$ (**stationarity**)

$H_a: y_t = y_{t-1} + \varepsilon_t$ (**random walk**)

using a t test that $|\beta_1| < 1$.[†]

[†] People often just test $\beta_1 < 1$.

Fun with \mathbb{R}

In-class exercise

1. Download the dataset `fun_data.csv`.
2. Figure out the model for y_1 . (Which of the x variables caused y_1 ?)
3. Figure out the model for y_2 . (Which of the x variables caused y_2 ?)

Extra credit:

- Answers on `fun_answers.csv`: Should the variable be included?
- +1pt (-1pt) for each correct (incorrect) answer (T/F for each combination)
- You will get at least 2pts for submitting (`fun_answers.csv`)
- Points will be added on to your total homework score

Don't forget: nonlinearities, omitted variable bias, nonstationarity, etc.

Table of contents

Admin

1. Schedule
2. R showcase
3. Fun with R

Nonstationarity

1. Introduction
2. Random walks
3. The actual problem
4. A potential solution
5. Dickey-Fuller tests