Elements of Econometrics.
Lecture 20.
Nonstationary Time Series.

FCS, 2022-2023

ASSUMPTIONS FOR MODEL C: WEAK PERSISTENCE

ASSUMPTIONS FOR MODEL C

C.1 The model is linear in parameters and correctly specified

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t$$

C.2 The time series for the regressors are (at most) weakly persistent

…

For time series which are "strongly persistent" in time or non-stationary, like random walks (see below) it is very typical to get "spurious regressions" if trying to apply the OLS: the variables seem to be highly related where there is no actual relationships between them.

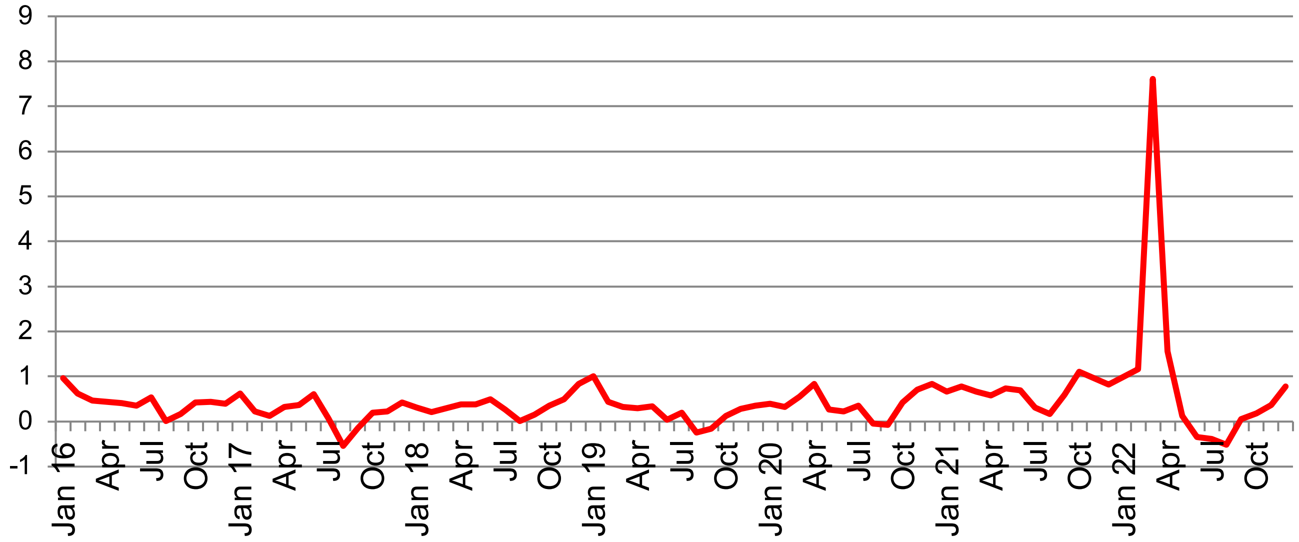
Stationary Stochastic Processes

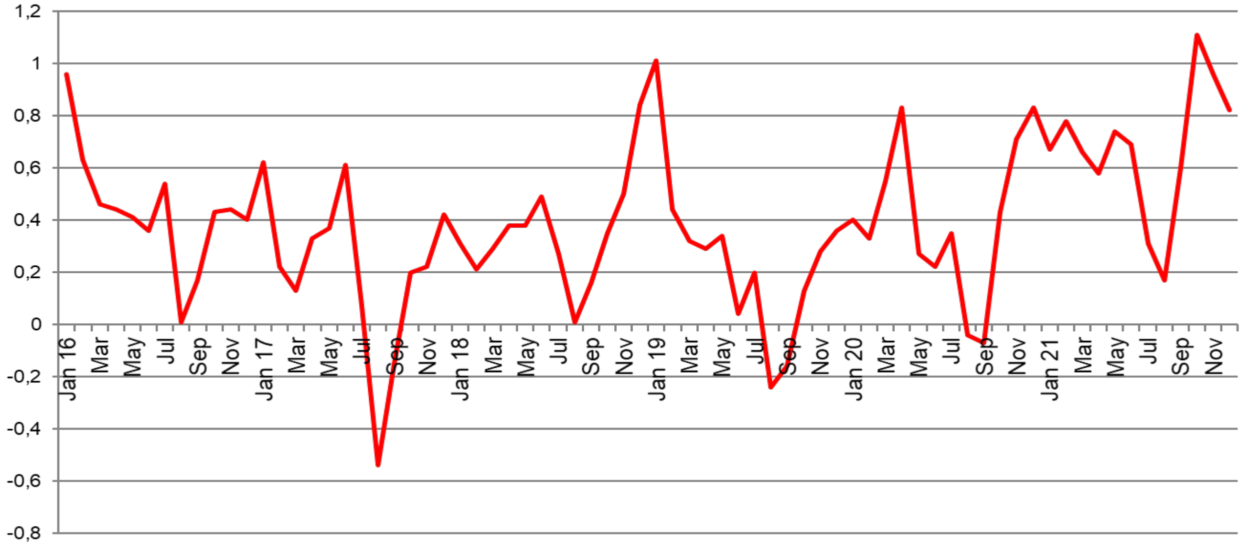
Stationarity (strong stationarity) of a stochastic process X_t is observed if the joint distribution of $X_{t1},X_{t2},\ldots,X_{tm}$ is identical to the joint distribution of $X_{t1+t},X_{t2+t},\ldots,X_{tm+t}$ for any $m,t,t_1,\ldots,t_m$.

A stochastic process is weakly stationary (or covariance stationary) if $E(X_t)$ is constant, $Var(X_t)$ is constant, and for any $t,s \geq 1$, $Cov(X_t, X_{t+s})$ depends only on s and not on t

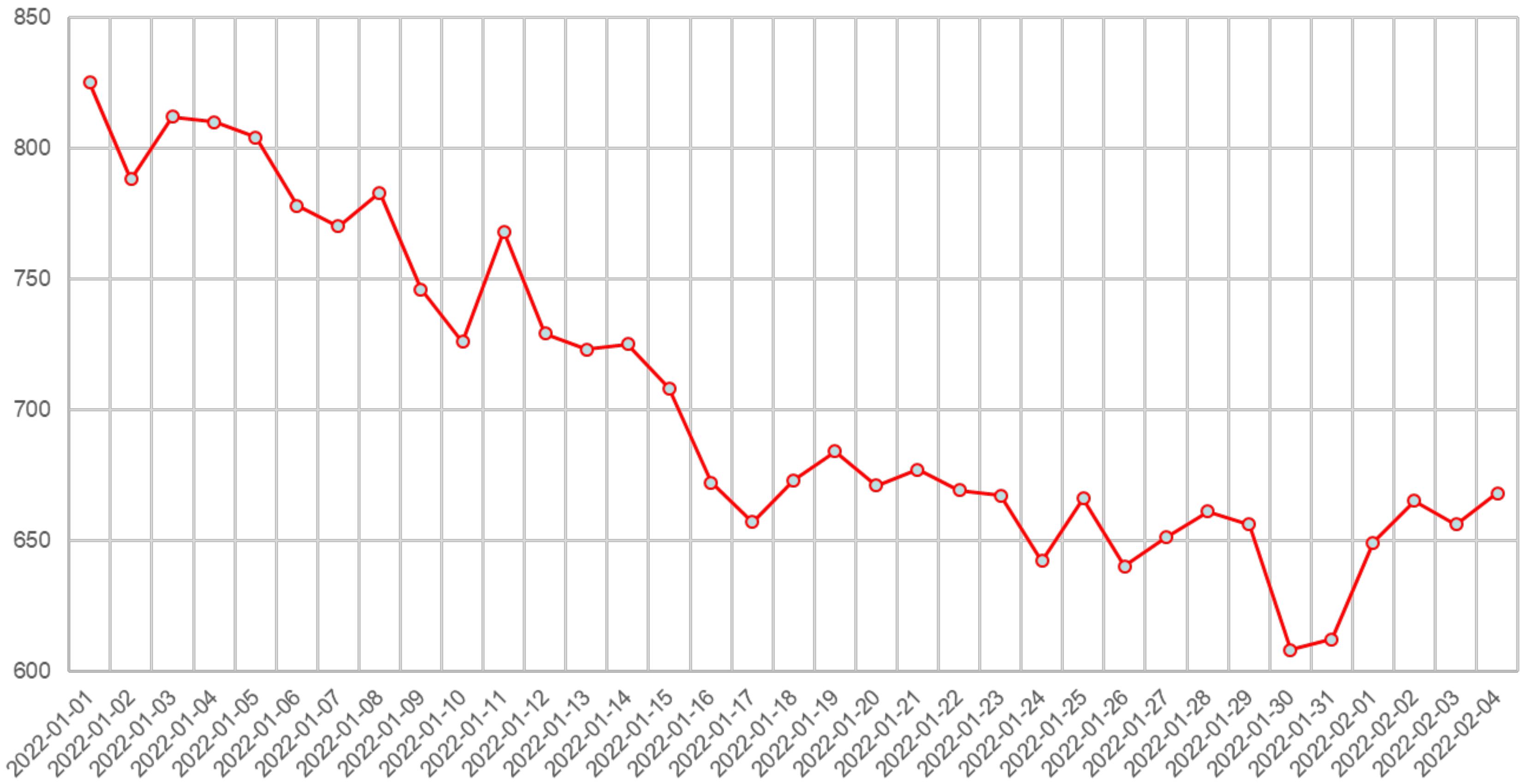
If for a weakly stationary process $Cov(X_t,X_{t+s}) \to 0$ as $s \to \infty$, the process is called weakly dependent (or weakly persistent)

Example: Monthly Inflation Rate in Russia, 2016-2022,%.
Does it look stationary? Why or why not?





Example: Number of COVID-19 deaths in Russia, 2022, Daily. Does it look stationary? Why or why not?



Some examples of time series processes to be tested for stationarity:

$X_t = \beta_2 X_{t-1} + \varepsilon_t$ – autoregressive process of the order 1, AR(1)

$X_t = \beta_1 + \beta_2 X_{t-1} + \varepsilon_t$ – AR(1) with a constant

$X_t = X_{t-1} + \varepsilon_t$ - random walk

$X_t = \beta_1 + X_{t-1} + \varepsilon_t$ - random walk with drift (β_1 is called a drift)

$X_t = \beta_1 + \beta_2 t + \varepsilon_t$ – deterministic trend

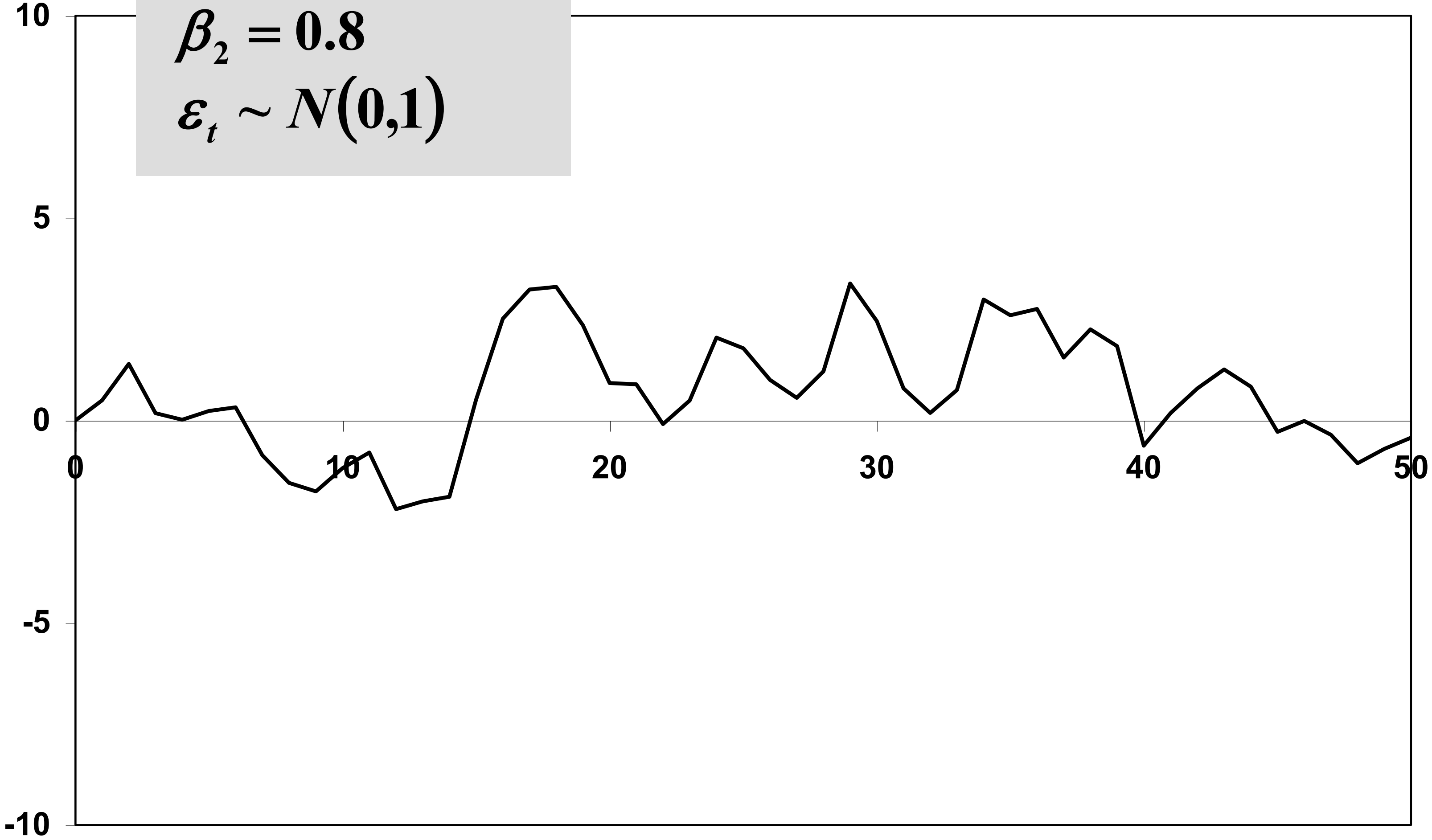
$X_t = \beta_1 + \beta_2 X_{t-1} + \ldots + \beta_{m+1} X_{t-m} + \varepsilon_t$ - *autoregressive process of the order m, AR(m)*

$X_t = \beta_1 + \varepsilon_t + \beta_2 \varepsilon_{t-1} + \ldots + \beta_{k+1} \varepsilon_{t-k}$ – moving average of the order k, MA(k)

$X_t = \beta_1 + \alpha_1 X_{t-1} + \ldots + \alpha_m X_{t-m} + \varepsilon_t + \beta_2 \varepsilon_{t-1} + \ldots + \beta_{k+1} \varepsilon_{t-k}$ - ARMA(m,k)

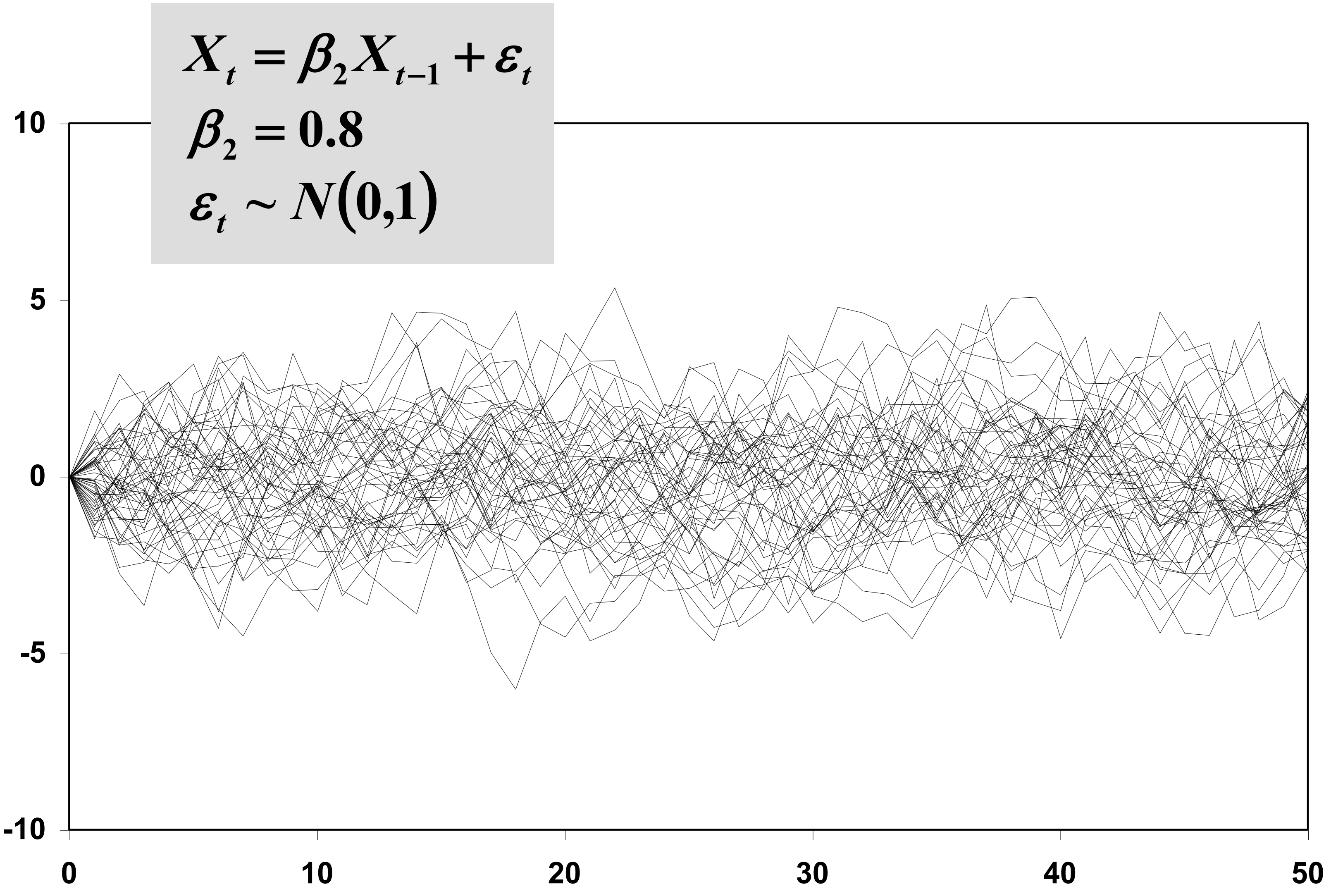
AR(1) process with $|\beta_2|$<1

$$X_t = \beta_2 X_{t-1} + \varepsilon_t$$
$$\beta_2 = 0.8$$
$$\varepsilon_t \sim N(0,1)$$



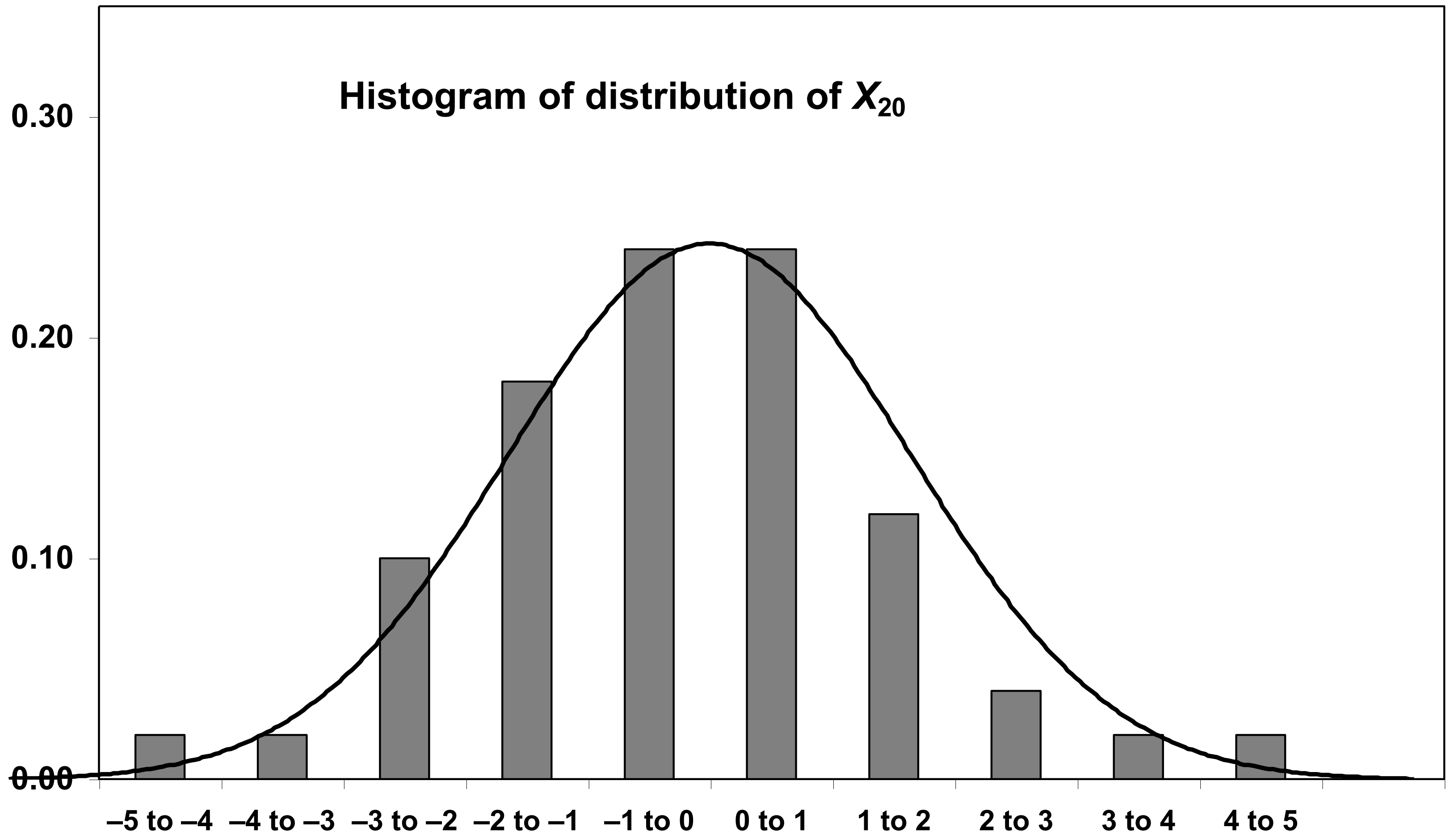
The AR(1) process $X_t = \beta_2 X_{t-1} + \varepsilon_t$ where $|\beta_2| < 1$ and innovation term ε_t is iid—independently and identically distributed—with zero mean and finite variance.

AR(1) process with $|\beta_2|<1$



The figure shows 50 realizations of the process.

AR(1) process with $|\beta_2|<1$



If the number of realizations were increased, each histogram would converge to the normal distribution

AR(1) process with $|\beta_2|$<1

Conditions for weak stationarity:

1. The population mean of the distribution is independent of time.

$$X_t = \beta_2 X_{t-1} + \varepsilon_t \qquad\qquad -1 < \beta_2 < 1$$

$$X_{t-1} = \beta_2 X_{t-2} + \varepsilon_{t-1}$$

$$X_t = \beta_2^2 X_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t$$

$$X_t = \beta_2^t X_0 + \beta_2^{t-1} \varepsilon_1 + \ldots + \beta_2^2 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t$$

$$\mathrm{E}(X_t) = \beta_2^t X_0$$

$E(X_t) = \beta_2^t X_0$ since E(ε)=0. $E(X_t) \to 0$ since $|\beta_2|$ <1. In the special case $X_0 = 0$, we have $E(X_t) = 0$. Since the expectation is not a function of time (asymptotically), the first condition is satisfied.

AR(1) process with $|\beta_2|<1$

Conditions for weak stationarity:

2. The variance of the distribution is independent of time.

$$\boldsymbol{X_t = \beta_2^t X_0 + \beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2\varepsilon_{t-2} + \beta_2\varepsilon_{t-1} + \varepsilon_t}$$

$$\begin{aligned}
\mathrm{Var}(X_t) &= \mathrm{Var}(\beta_2^t X_0 + \beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2\varepsilon_{t-2} + \beta_2\varepsilon_{t-1} + \varepsilon_t) \\
&= \mathrm{Var}(\beta_2^{t-1}\varepsilon_1) + \ldots + \mathrm{Var}(\beta_2^2\varepsilon_{t-2}) + \mathrm{Var}(\beta_2\varepsilon_{t-1}) + \mathrm{Var}(\varepsilon_t) \\
&= \beta_2^{2(t-1)}\sigma_\varepsilon^2 + \ldots + \beta_2^4\sigma_\varepsilon^2 + \beta_2^2\sigma_\varepsilon^2 + \sigma_\varepsilon^2 \\
&= \left(\beta_2^{2(t-1)} + \ldots + \beta_2^4 + \beta_2^2 + 1\right)\sigma_\varepsilon^2 = \left(\frac{1-\beta_2^{2t}}{1-\beta_2^2}\right)\sigma_\varepsilon^2
\end{aligned}$$

Since $|\beta_2| < 1$, β_2^{2t} tends to zero as *t* becomes large.
Thus, Var(X_t) tends to a limit independent of time.
The Condition 2 holds (asymptotically).

AR(1) process with $|\beta_2|$<1

Conditions for weak stationarity:

3. **The covariance between its values at any two time points depends only on the distance between those points, and not on time.**

$$X_t = \beta_2 X_{t-1} + \varepsilon_t$$

$$X_{t+s} = \beta_2 X_{t+s-1} + \varepsilon_{t+s} = \beta_2^2 X_{t+s-2} + \beta_2 \varepsilon_{t+s-1} + \varepsilon_{t+s} =$$
$$= \beta_2^s X_t + \beta_2^{s-1} \varepsilon_{t+1} + \ldots + \beta_2^2 \varepsilon_{t+s-2} + \beta_2 \varepsilon_{t+s-1} + \varepsilon_{t+s}$$

$$\mathrm{Cov}(X_t, X_{t+s}) = \mathrm{Cov}(X_t, \beta_2^s X_t) +$$
$$+ \mathrm{Cov}(X_t, [\beta_2^{s-1} \varepsilon_{t+1} + \ldots + \beta_2^2 \varepsilon_{t+s-2} + \beta_2 \varepsilon_{t+s-1} + \varepsilon_{t+s}]) = \beta_2^s \,\mathrm{Var}(X_t)$$

$$-does\ not\ asymptotically\ depend\ on\ t.$$

Lagging and substituting *s* times, we can express X_{t+s} in terms of X_t and the innovations ε_{t+1}, ..., ε_{t+s}. X_t is independent of the innovations after time *t*. As we have seen, Var(X_t) is independent of *t*, apart from a transitory initial effect.
Hence the condition 3 holds (asymptotically).
As $\beta_2^s \to 0$ if $s \to \infty$, the process is weakly persistent.

AR(1) process with $|\beta_2|<1$ and $\beta_1 \neq 0$.

Add intercept to the process: does it influence stationarity?
Conditions for weak stationarity:

1. **The mean of the distribution is independent of time.**

$$X_t = \beta_1 + \beta_2 X_{t-1} + \varepsilon_t$$

$$X_t = \beta_1(\beta_2 + 1) + \beta_2^2 X_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t$$

$$X_t = \beta_2^t X_0 + \beta_1\left(\beta_2^{t-1} + \ldots + \beta_2 + 1\right) + \beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t$$

$$= \beta_2^t X_0 + \beta_1 \frac{1-\beta_2^t}{1-\beta_2} + \beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t$$

$$E(X_t) = \beta_2^t X_0 + \beta_1 \frac{1-\beta_2^t}{1-\beta_2} \rightarrow \frac{\beta_1}{1-\beta_2}$$

Taking expectations, E(X_t) tends to $\beta_1/(1 - \beta_2)$ since the term β_2^t tends to zero. It is non-zero, but asymptotically independent of time.

AR(1) process with $|\beta_2|<1$ and $\beta_1 \neq 0$.

Conditions for weak stationarity:

2. The variance of the distribution is independent of time.

$$X_t = \beta_2^t X_0 + \beta_1 \frac{1-\beta_2^t}{1-\beta_2} + \beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t$$

$$\mathrm{Var}(X_t) = \mathrm{Var}\left(\beta_2^t X_0 + \beta_1 \frac{1-\beta_2^t}{1-\beta_2} + \beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t\right) =$$

$$= \mathrm{Var}(\beta_2^{t-1}\varepsilon_1 + \ldots + \beta_2^2 \varepsilon_{t-2} + \beta_2 \varepsilon_{t-1} + \varepsilon_t) = \left(\frac{1-\beta_2^{2t}}{1-\beta_2^2}\right)\sigma_\varepsilon^2 \rightarrow \frac{\sigma_\varepsilon^2}{1-\beta_2^2}$$

The variance is unaffected by the addition of intercept.

STATIONARY PROCESSES: AR(1) process with $|\beta_2|<1$ and $\beta_1 \neq 0$.

Conditions for weak stationarity:

3. **The covariance between its values at any two time points depends only on the distance between those points, and not on time.**

$$X_t = \beta_1 + \beta_2 X_{t-1} + \varepsilon_t$$

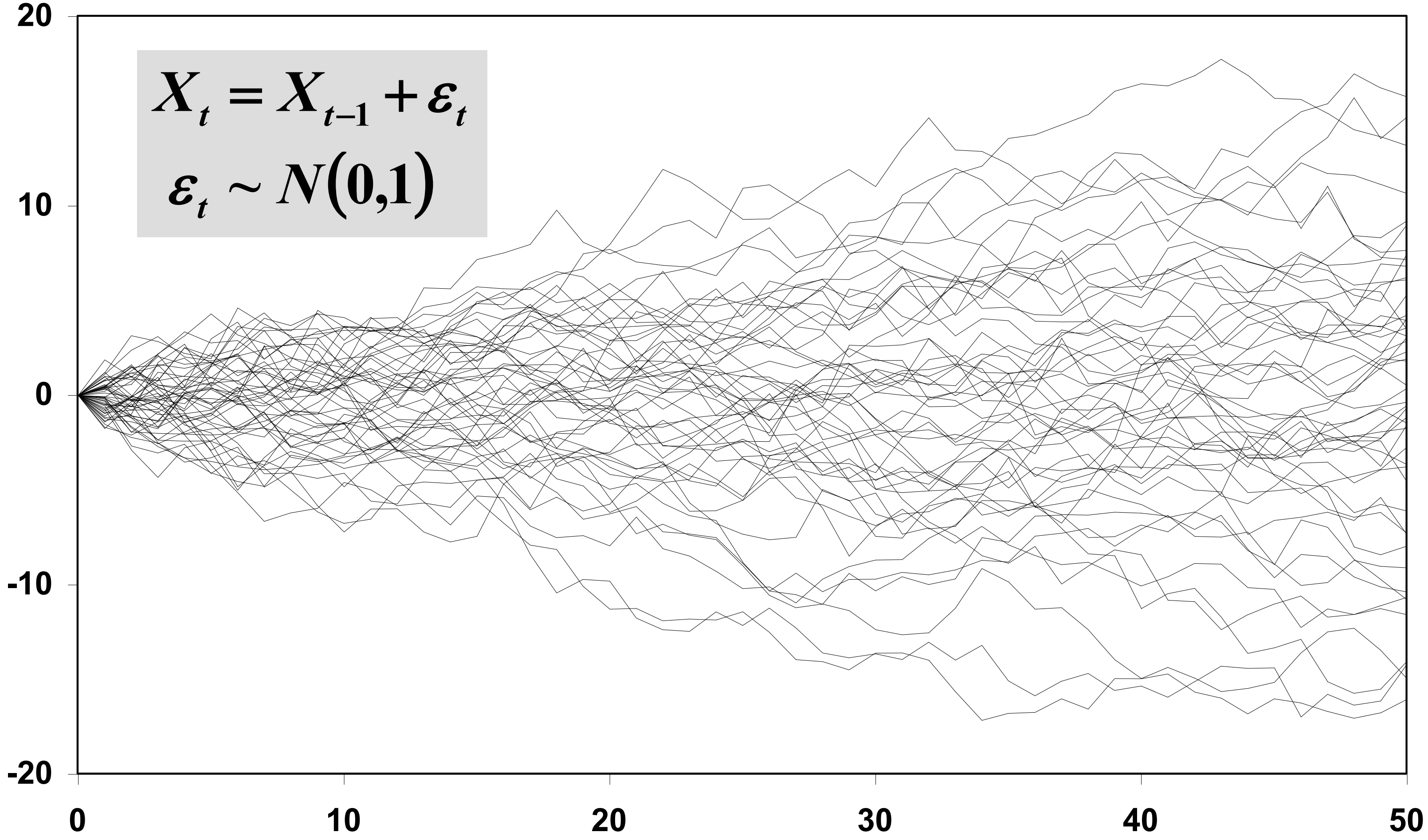
$$X_{t+s} = \beta_1 + \beta_2 X_{t+s-1} + \varepsilon_{t+s}$$

$$X_{t+s} = \beta_1(\beta_2 + 1) + \beta_2^2 X_{t+s-2} + \beta_2 \varepsilon_{t+s-1} + \varepsilon_{t+s}$$

$$X_{t+s} = \beta_1\left(\beta_2^{s-1} + \ldots + \beta_2^2 + \beta_2 + 1\right) + \beta_2^s X_t$$
$$+ \beta_2^{s-1}\varepsilon_{t+1} + \ldots + \beta_2^2 \varepsilon_{t+s-2} + \beta_2 \varepsilon_{t+s-1} + \varepsilon_{t+s}$$

The covariance of X_t and X_{t+s} is not affected by the inclusion of this term because it is a constant. Hence the covariance is the same as before and remains independent of *t*.

NONSTATIONARY PROCESSES: RANDOM WALK



This figure shows the results of 50 simulations. The distribution is not stationary: it changes as *t* increases, becoming increasingly spread out.

NONSTATIONARY PROCESSES: RANDOM WALK

Random walk $$X_t = X_{t-1} + \varepsilon_t$$

$$X_t = X_0 + \varepsilon_1 + \ldots + \varepsilon_{t-1} + \varepsilon_t$$

$$\mathrm{E}(X_t) = X_0 + \mathrm{E}(\varepsilon_1) + \ldots + \mathrm{E}(\varepsilon_n) = X_0$$

$$\sigma^2_{X_t} = \mathrm{Var}(X_0 + \varepsilon_1 + \ldots + \varepsilon_{t-1} + \varepsilon_t) =$$
$$= \mathrm{Var}(\varepsilon_1 + \ldots + \varepsilon_{t-1} + \varepsilon_t) = \sigma^2_\varepsilon + \ldots + \sigma^2_\varepsilon + \sigma^2_\varepsilon = t\sigma^2_\varepsilon$$

The population variance of X_t is directly proportional to *t*, its distribution spreads out as *t* increases. The process is nonstationary.

$$E(y_t) = E(y_0)$$

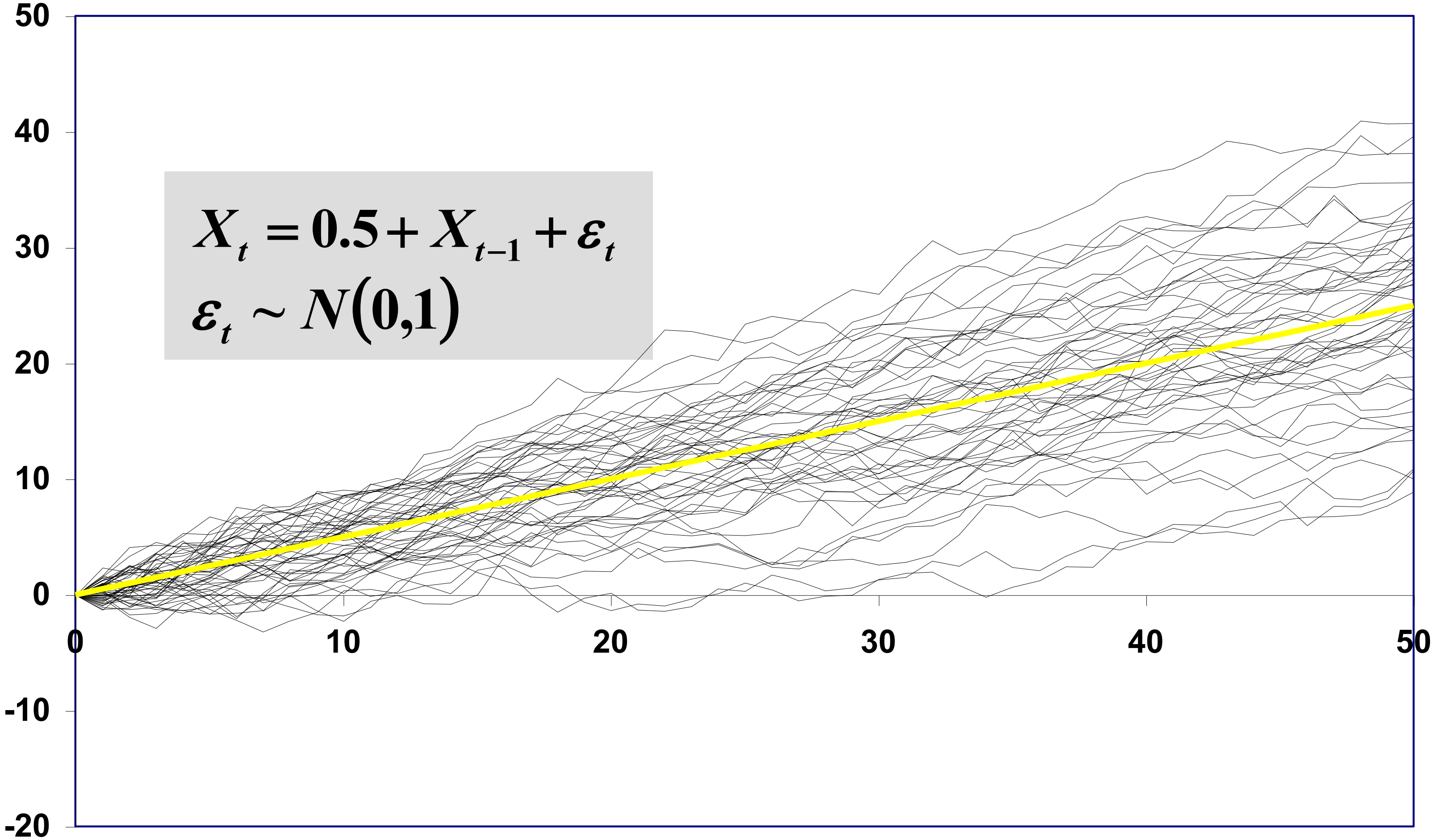
$$Var(y_t) = \sigma_e^2 t$$

$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

The random walk is not covariance stationary because its variance and its covariance depend on time.

It is also not weakly dependent because the correlation between observations vanishes very slowly and this depends on how large t is.

NONSTATIONARY PROCESSES: RANDOM WALK WITH DRIFT



The figure shows 50 realizations of such a process. The underlying drift line is highlighted in yellow. It can be seen that the ensemble distribution changes in two ways with time.

NONSTATIONARY PROCESSES: RANDOM WALK WITH DRIFT

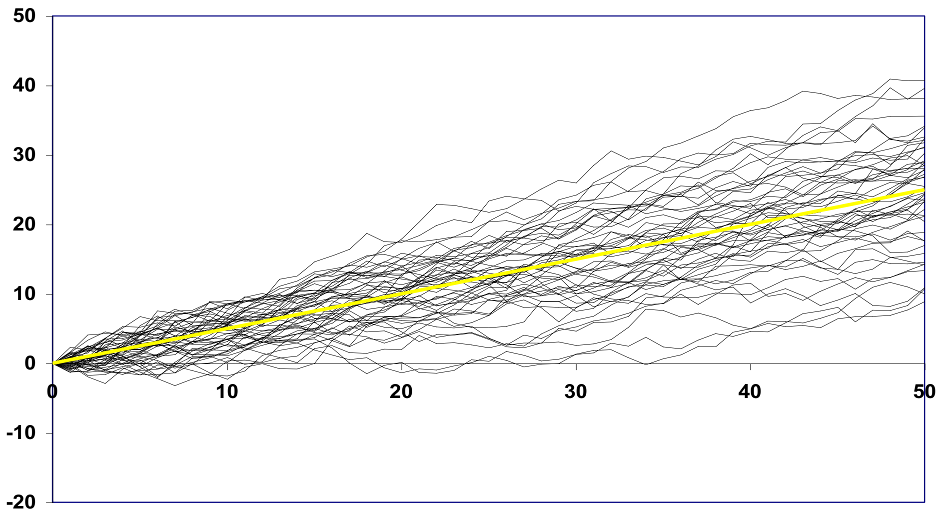
Random walk with drift

$$X_t = \beta_1 + X_{t-1} + \varepsilon_t$$

$$X_t = \beta_1 t + X_0 + \varepsilon_1 + \ldots + \varepsilon_{t-1} + \varepsilon_t$$

$$E(X_t) = X_0 + \beta_1 t$$

$$\sigma_{X_t}^2 = t\sigma_\varepsilon^2$$



Here the expected value changes in time. The population variance grows proportionally to the time. The process is nonstationary.

$$E(y_t) = \alpha_0 t + E(y_0)$$

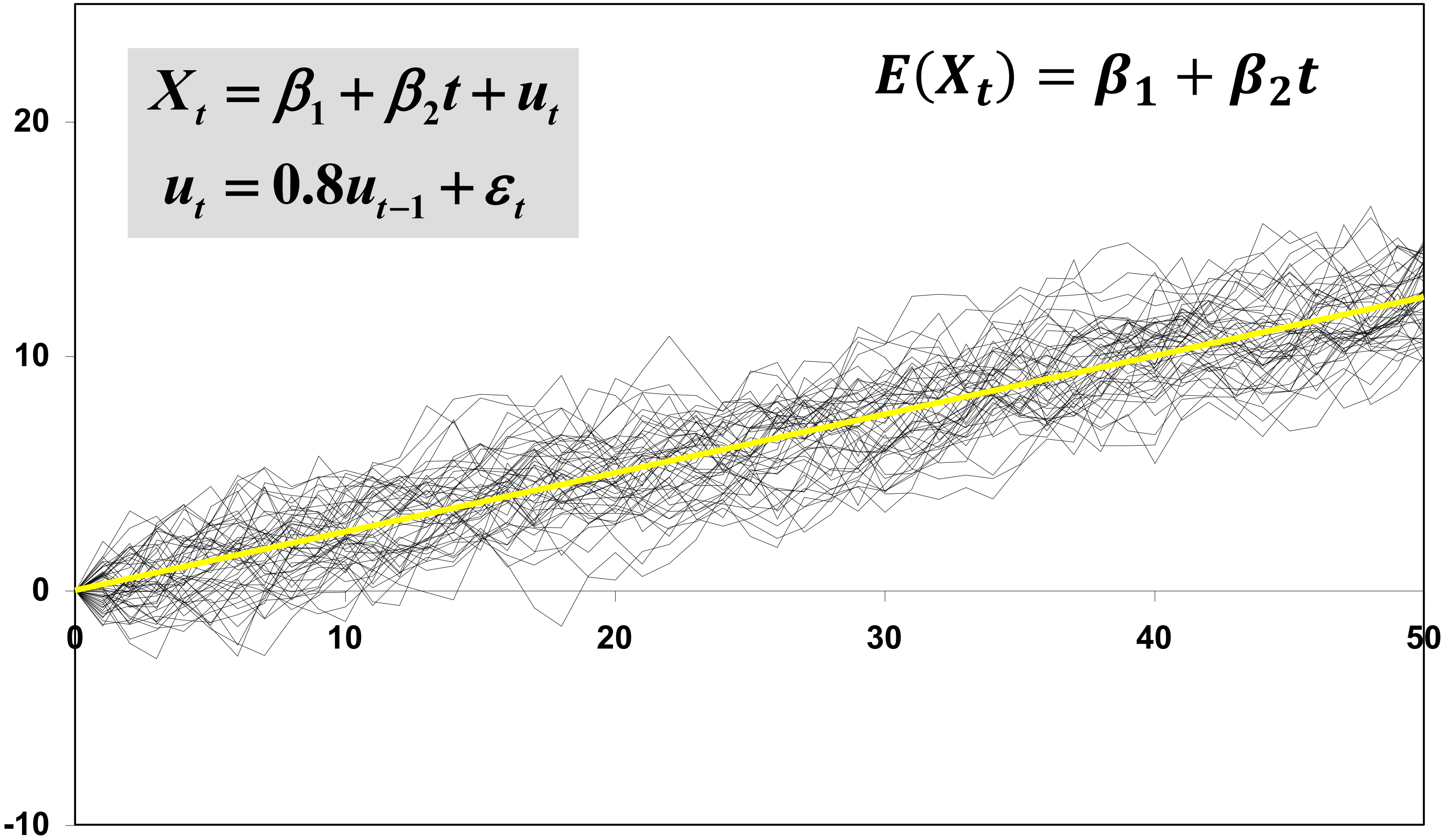
$$Var(y_t) = \sigma_e^2 t$$

$$Corr(y_t, y_{t+h}) = \sqrt{t/(t+h)}$$

Otherwise, the random walk with drift has similar properties as the random walk without drift.

Random walks with drift are not covariance stationary and not weakly dependent.

NONSTATIONARY PROCESSES: DETERMINISTIC TREND

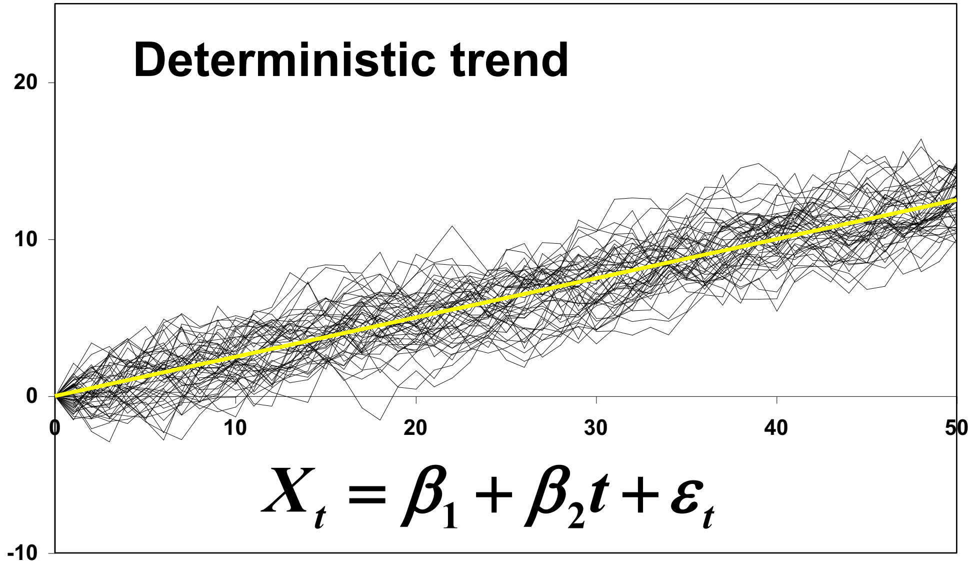


The figure shows 50 realizations where the disturbance term is $u_t = 0.8u_{t-1}+\varepsilon_t$.

The process is nonstationary since $E(X_t)$ depends on t. The variation around the trend line does not grow: $var(X_t)=var(u_t)$, where u_t is stationary.

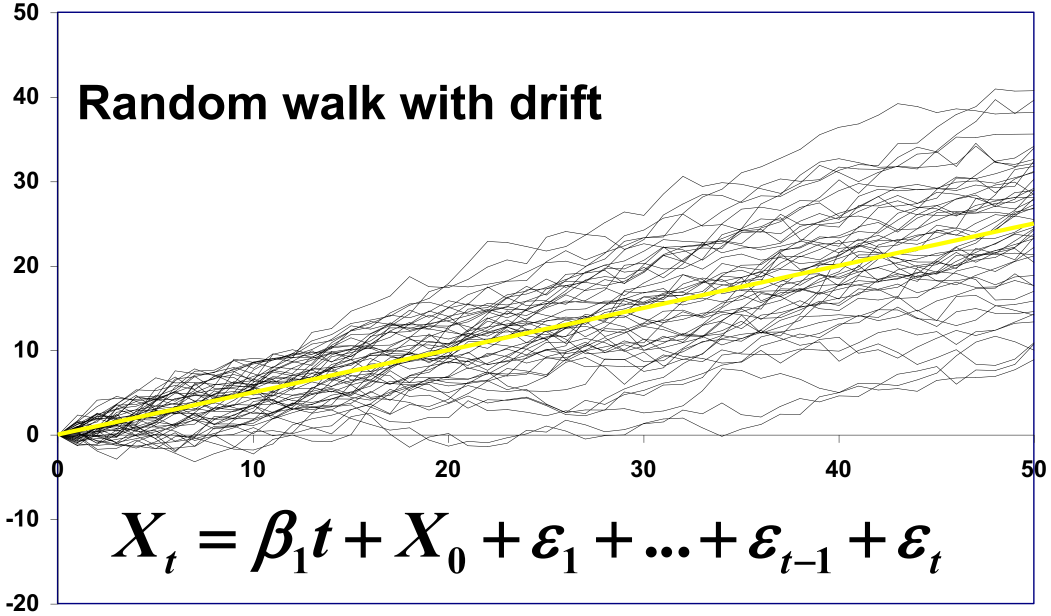
NONSTATIONARY PROCESSES:

DETERMINISTIC TREND AND RANDOM WALK WITH DRIFT



$$X_t = \beta_1 + \beta_2 t + \varepsilon_t$$

$$\mathrm{Var}(X_t) = \sigma^2{}_{\varepsilon}$$



$$X_t = \beta_1 t + X_0 + \varepsilon_1 + \ldots + \varepsilon_{t-1} + \varepsilon_t$$

$$\mathrm{Var}(X_t) = t\sigma^2{}_{\varepsilon}$$

More facts on Stationarity

Necessary condition for stationarity of autoregressive process of the order m (AR(*m*))

$X_t = \beta_1 + \beta_2 X_{t-1} + \ldots + \beta_{m+1} X_{t-m} + \varepsilon_t$ *is*

$$\sum_{i=2}^{m+1} \beta_i < 1.$$

Sufficient condition for stationarity of AR*(m) process is* $$\sum_{i=2}^{m+1} |\beta_i| < 1.$$

Any MA(k) *series* $X_t = \beta_1 + \varepsilon_t + \beta_2 \varepsilon_{t-1} + \ldots + \beta_{k+1} \varepsilon_{t-k}$ (moving averages of the order *k*) is stationary.

Stationarity of the ARMA(m,k) model

$X_t = \beta_1 + \alpha_1 X_{t-1} + \ldots + \alpha_m X_{t-m} + \varepsilon_t + \beta_2 \varepsilon_{t-1} + \ldots + \beta_{k+1} \varepsilon_{t-k}$ depends on its AR part only.

NONSTATIONARY PROCESSES: DIFFERENCING

Difference-stationarity

$$X_t = \beta_1 + X_{t-1} + \varepsilon_t$$

$$\Delta X_t = X_t - X_{t-1} = \beta_1 + \varepsilon_t$$

$$E(\Delta X_t) = \beta_1$$

$$\sigma^2_{\Delta X_t} = \sigma^2_{\varepsilon}$$

$$\mathrm{Cov}(\Delta X_t, \Delta X_{t+s}) = 0$$

If a nonstationary time series can be transformed into a stationary process by differencing p times, it is described as integrated of the order p, or I(p).

Most of nonstationary economic time series are I(1), or at most I(2).

NONSTATIONARY PROCESSES: DETRENDING

Trend-stationarity

$$X_t = \beta_1 + \beta_2 t + \varepsilon_t$$

$$\hat{X}_t = b_1 + b_2 t$$

$$\tilde{X}_t = X_t - \hat{X}_t = X_t - b_1 - b_2 t$$

If E(b_2)=β_2, then $\tilde{X}_t$ **is stationary.**

NONSTATIONARY PROCESSES EXAMPLE: DETRENDING AND DIFFERENCING

Wages and productivity

$$\widehat{\log}(hrwage) = \underset{(.37)}{-5.33} + \underset{(.09)}{1.64} \log(outphr) - \underset{(.002)}{.018}\, t$$

Include trend because both series display clear trends.

$$n = 41, R^2 = .971, \bar{R}^2 = .970$$

The elasticity of hourly wage with respect to output per hour (productivity) seems implausibly large.

It turns out that even after detrending, both series display sample autocorrelations close to one so that estimating the equation in first differences seems more adequate:

$$\Delta\widehat{\log}(hrwage) = \underset{(.0042)}{-.0036} + \underset{(.173)}{.809}\, \Delta \log(outphr)$$

$$n = 40, R^2 = .364, \bar{R}^2 = .348$$

This estimate of the elasticity of hourly wage with respect to productivity makes much more sense.

Estimation of Regressions with Nonstationary variables

The main problem: getting Spurious Regressions.

The share of Type 1 errors is much higher than the significance level.

Granger and Newbold (1974); Monte-Carlo experiment with 100 estimations of regression of one independent random walk on another:

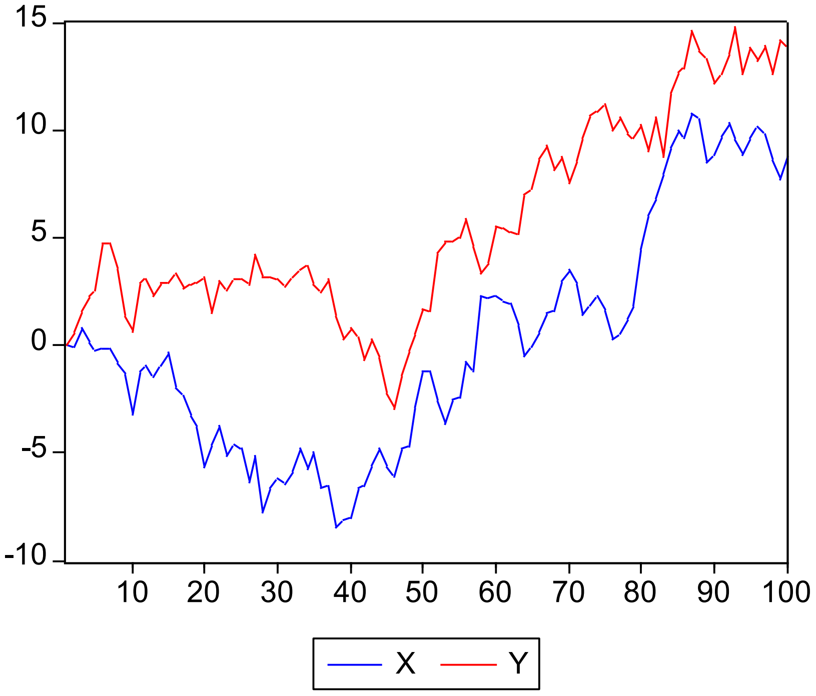
70 significant slope coefficients under 1% significance level,

77 time at the 5% level.

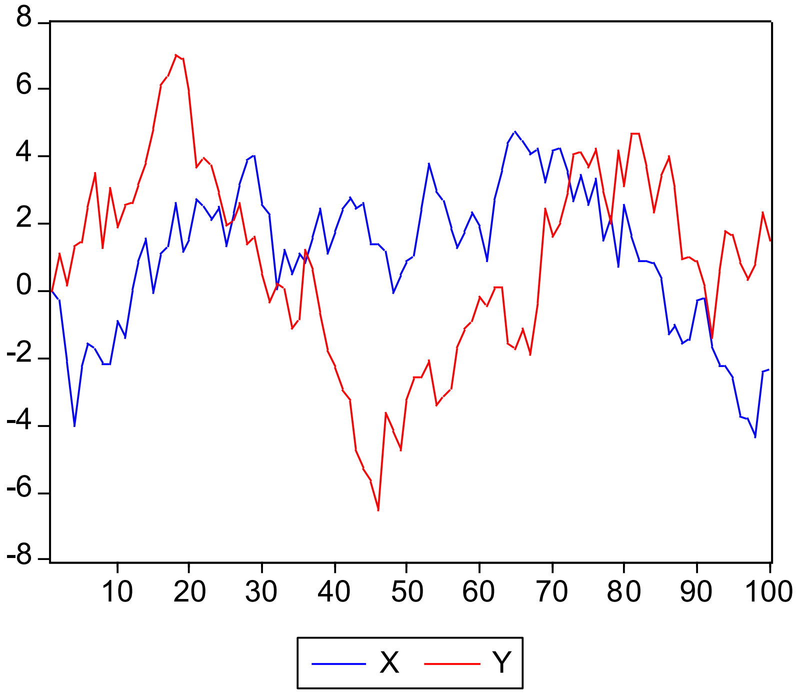
Regressions with Nonstationary Data

Monte Carlo experiment: $X(0)=Y(0)=0$; $X(t)=X(t\text{-}1)+\varepsilon_t$; $Y(t)=Y(t\text{-}1)+\mu_t$ (ε_t and μ_t are white noises with $\sigma^2_{\varepsilon}=\sigma^2_{\mu}=1$), T=100.

Sample 1:



Sample 2:



Spurious Regression Example

Dependent Variable: Y Method: Least Squares

Sample 1:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.634	0.227	24.822	0.0000
X	**0.750**	**0.042**	17.835	0.0000
R-squared	0.764	Mean dependent var	5.702	
S.D. dependent var	4.652	S.E. of regression	2.259	
Sum squared resid	504.71	F-statistic	318.08	
Durbin-Watson stat	**0.284**	Prob(F-statistic)	0.0000	

Sample 2:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.052	0.324	3.248	0.0016
X	**-0.158**	**0.136**	-1.160	0.2488
R-squared	0.014	Mean dependent var	0.884	
S.D. dependent var	2.904	S.E. of regression	2.899	
Sum squared resid	823.6	F-statistic	1.346	
Durbin-Watson stat	**0.136**	Prob(F-statistic)	0.249	

Pay attention to the Durbin-Watson Statistic: it indicates incorrect specification.
Disturbance term is a Random Walk ($u_t=Y_t-\beta_1-0*X_t$).