POLS/CSSS 503:
Advanced Quantitative Political Methodology

Time Series: Stochastic Processes

Christopher Adolph*
University of Washington, Seattle

May 12, 2011

*Assistant Professor, Department of Political Science and Center for Statistics and the Social Sciences.

Data in temporal context

All the models we've looked at have been based on an assumption:

Observations are idenitcally and *independently* distributed, conditional on covariates

Often this is an unrealistic assumption:

- Clustering in physical space (geography)
- Clustering in latent space (networks)
- Temporal dependence (time series)

Inter-dependent observations are intrinsic to social phenomena. Ask any historian or sociologist

Can we rescue iid somehow?

Data in temporal context

Time's arrow: The past shapes the future.

Perhaps if we condition on the past (control for it), then we'll have iid error terms

But it turns out there are many ways to think about the effect of the past

Notation

We assume time is discrete.

Observations take place in periods. No two observations happen at the same time.

Periods could be years, months, quarters, days, etc.

Index our observations as y_t, $t = 1, \ldots, T$

For a single (ie, non-panel) time series y_t, we have T observations total.

Today we won't consider any covariates.

Dynamic processes

Several conceptually different ways to think of the effect of history

- Past realizations of y influence current levels of y
- Past shocks to y influence current levels of y
- Past expectations of y influence current levels of y

Examples of dynamic processes

Past realizations of y influence current levels of y.

Example: Unemployment; Welfare state spending uses last years budget as baseline

Past shocks to Y influence current levels of Y

Example: Some forms of financial volatility? Voting in Congress?

Past expectations of Y influence current levels of Y

Example: Polling time series (shocks are partly measurement error); anything determined by modelers?

Let's incorporate these dynamics into our baseline model.

Past realizations of y

$$y_t = y_{t-1}\phi_1 + \varepsilon_t$$

Known as an *autoregressive process*

Each new realization of y_t incorporates the last period's realization, y_{t-1}

Note that only one lag of y_t appears in our model.

This is an AR(1) process, of an autoregressive process of degree 1.

However, the distance past still has an effect. Implied by above:

$$y_{t-1} = y_{t-2}\phi_1 + \varepsilon_{t-1}$$

and so

$$y_{t-2} = y_{t-3}\phi_1 + \varepsilon_{t-2}$$

. . . and on and on back to the "original" period

Autoregressive Processes

Recursive reparameterization: Iterating through all past periods and substituting back into the first formula.

Autoregressive Processes

For AR(1), recursion reveals the following:

$$
\begin{aligned}
y_t &= y_{t-1}\phi_1 + \varepsilon_t \\
&= \left(y_{t-2}\phi_1 + \varepsilon_{t-1}\right)\phi_1 + \varepsilon_t \\
&= y_{t-2}\phi_1^2 + \varepsilon_{t-1}\phi_1 + \varepsilon_t \\
&= \left(y_{t-3}\phi_1 + \varepsilon_{t-2}\right)\phi_1^2 + \varepsilon_{t-1}\phi_1 + \varepsilon_t \\
&= y_{t-3}\phi_1^3 + \varepsilon_{t-2}\phi_1^2 + \varepsilon_{t-1}\phi_1 + \varepsilon_t \\
&\dots \quad \text{substitute through } y_{t-k} \\
&= y_{t-k}\phi_1^k + \sum_{j=0}^{k-1} \varepsilon_{t-j}\phi^j \\
&\dots \quad \text{substitute through } y_{t-\infty} \\
&= \sum_{j=0}^{\infty} \varepsilon_{t-j}\phi^j
\end{aligned}
$$

Autoregressive Processes

$$y_t = \sum_{j=0}^{\infty} \varepsilon_{t-j} \phi^j$$

In the limit, if y_t is AR(1), then
y_t includes the effects of every disturbance back to the beginning of time:
$\varepsilon_{t-1}, \dots \varepsilon_{t-\infty}$

Effect of history lasts forever in autoregressive processes

So what would happen if $|\phi_1| < 1$?

And if $|\phi_1| > 1$?

And if $|\phi_1| = 1$ exactly?

Autoregressive Processes

But if $-1 < \phi_1 < 1$, the effect of the past approaches zero as time passes, but never completely fades

If $|\phi_1| > 1$, the process is *explosive*, tending quickly to infinity. Not a reasonable model of any natural or social process (at least anything that lasts very long!)

If $\phi_1 = 1$ exactly, we have a *random walk* or *unit root*. Very persistent effects of history, and many unusual properties.

Simulating time series

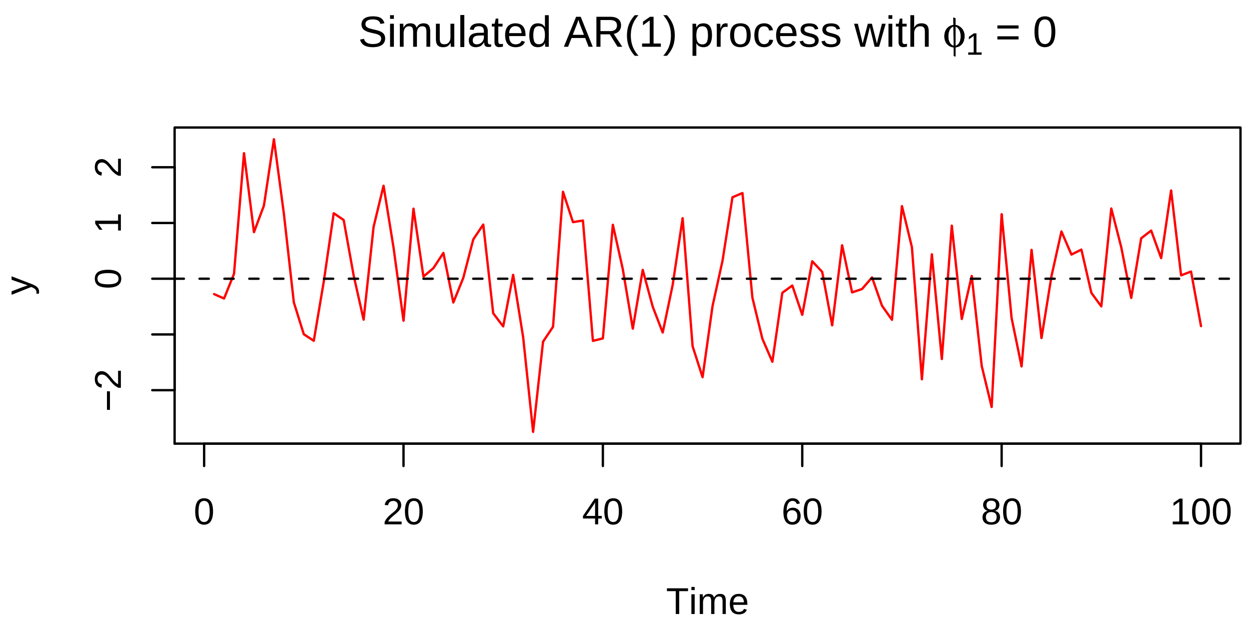
Basic idea:

1. Start at some y_0, e.g., $y_0 = 0$, and choose a ϕ_1 to study
2. Draw $y_1 = y_0\phi_1 + \varepsilon_1$, where $\varepsilon_1 \sim N(0, \sigma^2)$
3. Iterate on step 2 to generate y_2, y_3, etc.

In `R`, just use `arima.sim()`

E.g., to simulate 1000 iterations of an AR(1) with $\phi_1 = 0.67$:

```
y <- arima.sim(list(order = c(1,0,0),
                    ar = 0.67,
                    ma = NULL),
               n=1000)
```



Here are 100 draws from an AR(1) with $\phi_1 = 0$

Nice, but can we "see" that these iterations are not serially correlated (ie, a "white noise" process, or just $N(0, \sigma^2)$)?

Autocorrelation functions

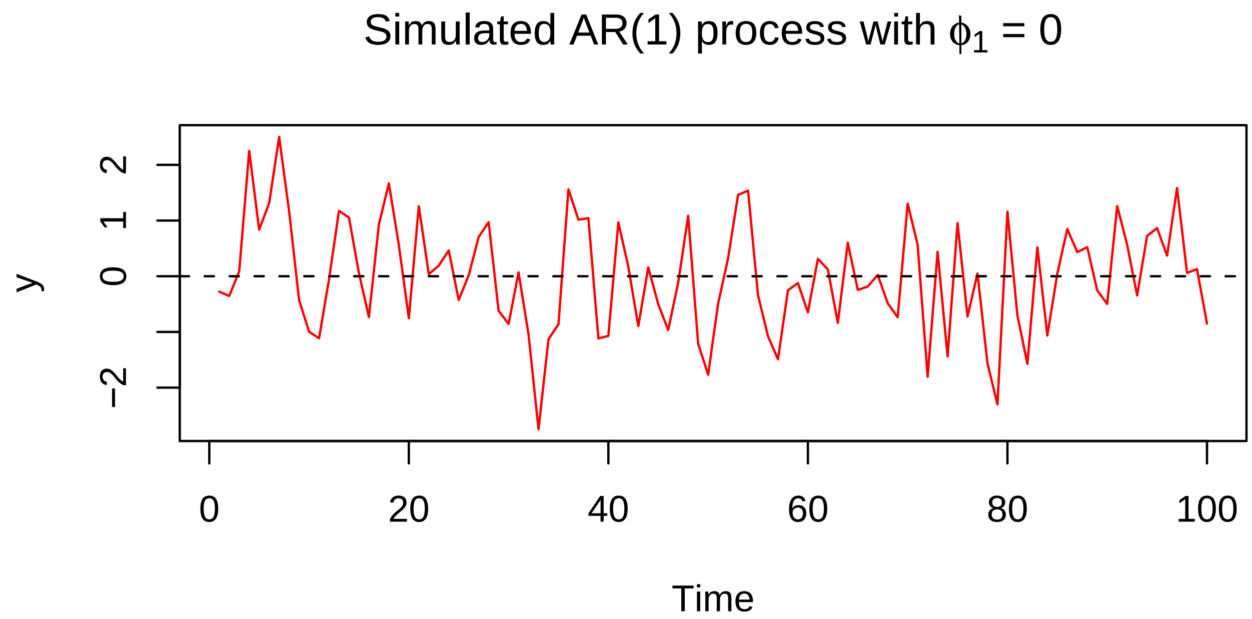
Simplest way to "see" autocorrelation is to calculate and plot the correlation between observations separated by a given distance k for $k = 1, 2, 3, \ldots$.

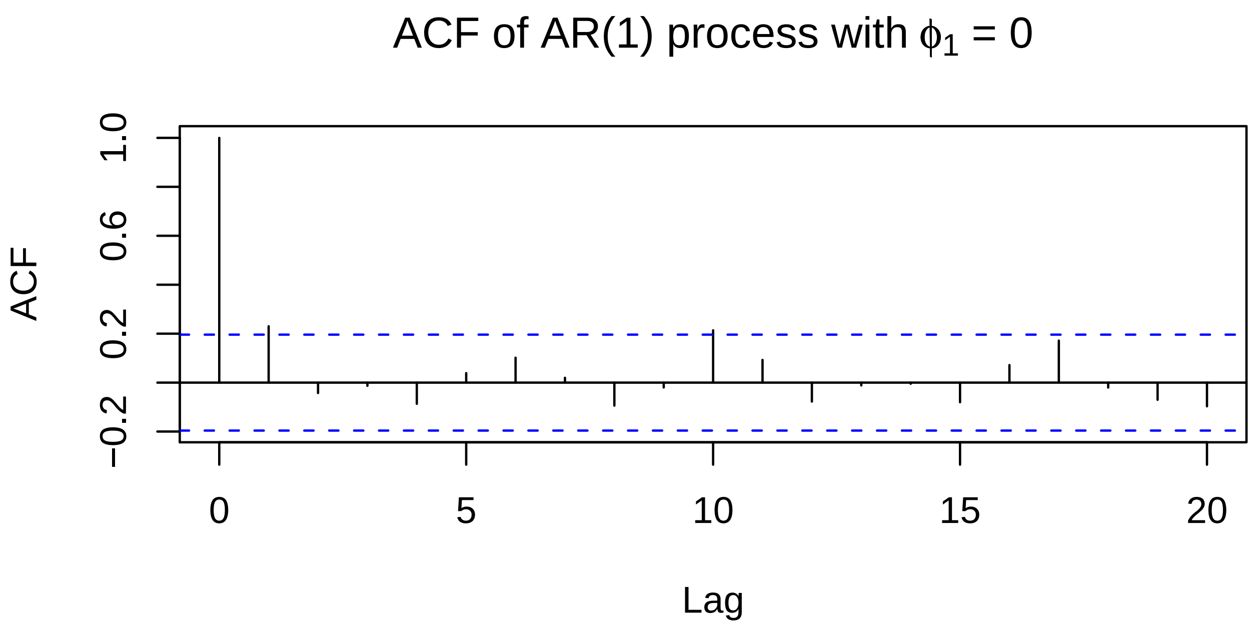
Define the autocorrelation function as

$$\mathrm{ACF}_j = \frac{\mathrm{cov}(x_t, x_{t+j})}{\mathrm{var(x_t)}}$$

Note that for an AR(1) with ϕ_1, the ACF_j is just ϕ_1^j

To have `R` estimate the ACF, just use `acf(y)`


Simulated AR(1) process with $\phi_1 = 0$
y
−2
0
1
2
0
20
40
60
80
100
Time


ACF of AR(1) process with $\phi_1 = 0$
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
20
Lag

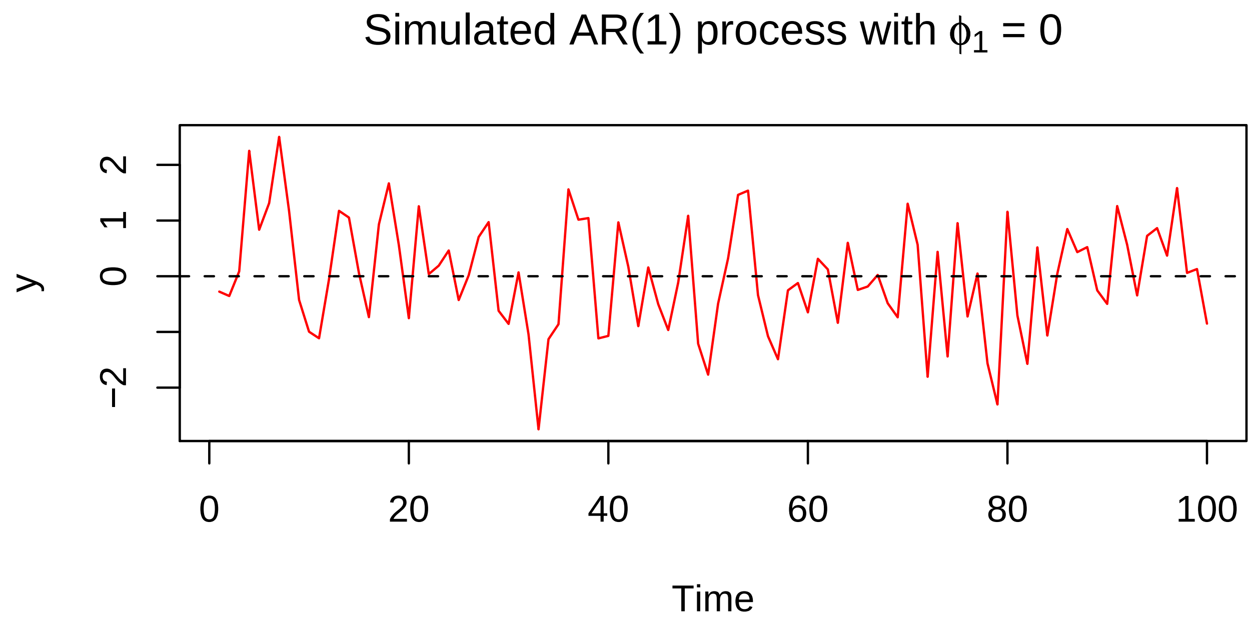
Autocorrelation functions

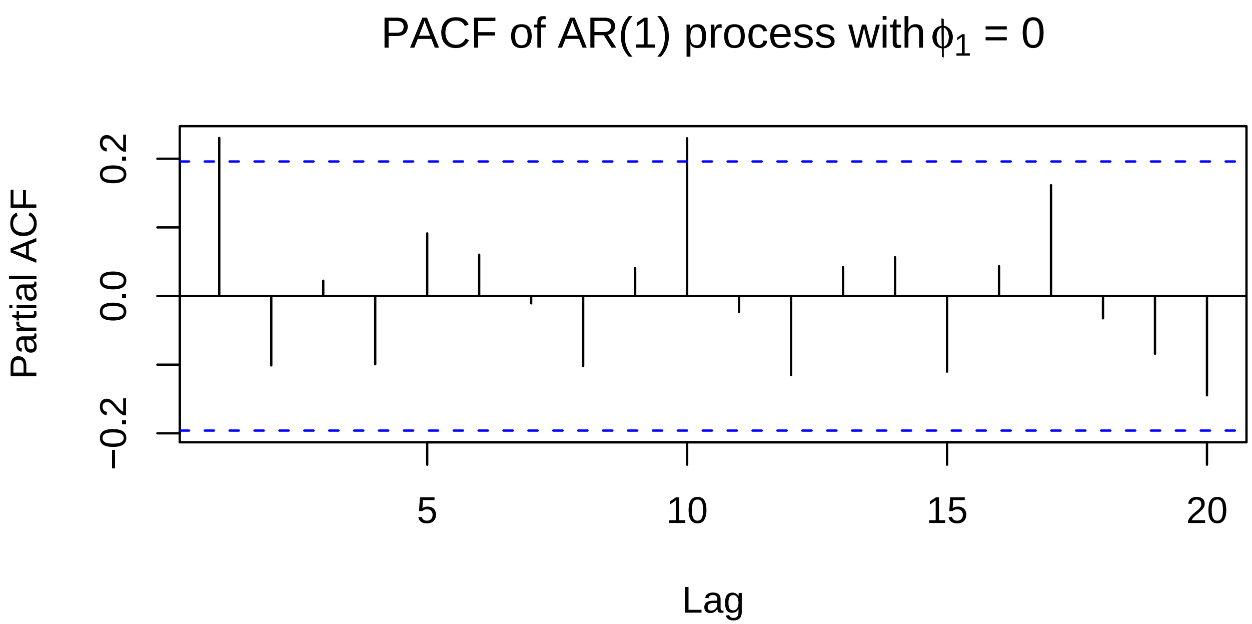
A useful refinement of the ACF is to "partial out" the effects of intervening lags

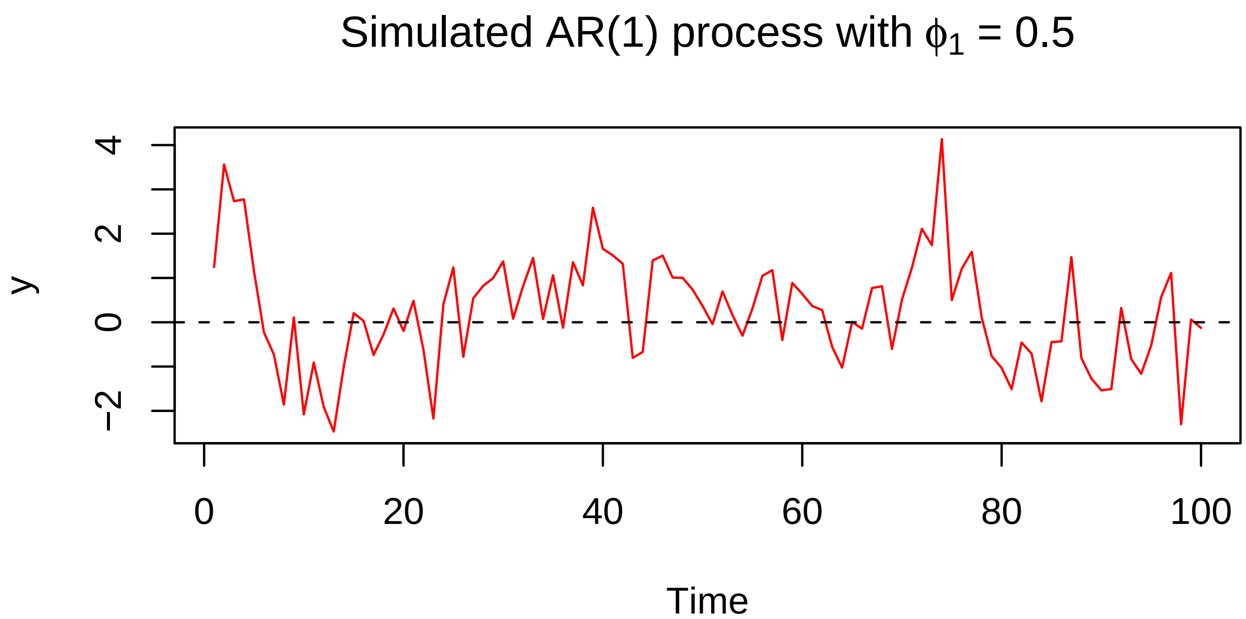
That is, we want to isolate the conditional correlation of y_t and y_{t-k} controlling for the values y_{t-1} to y_{t-k+1}.

We call this the partial autocorrelation function, or PACF.

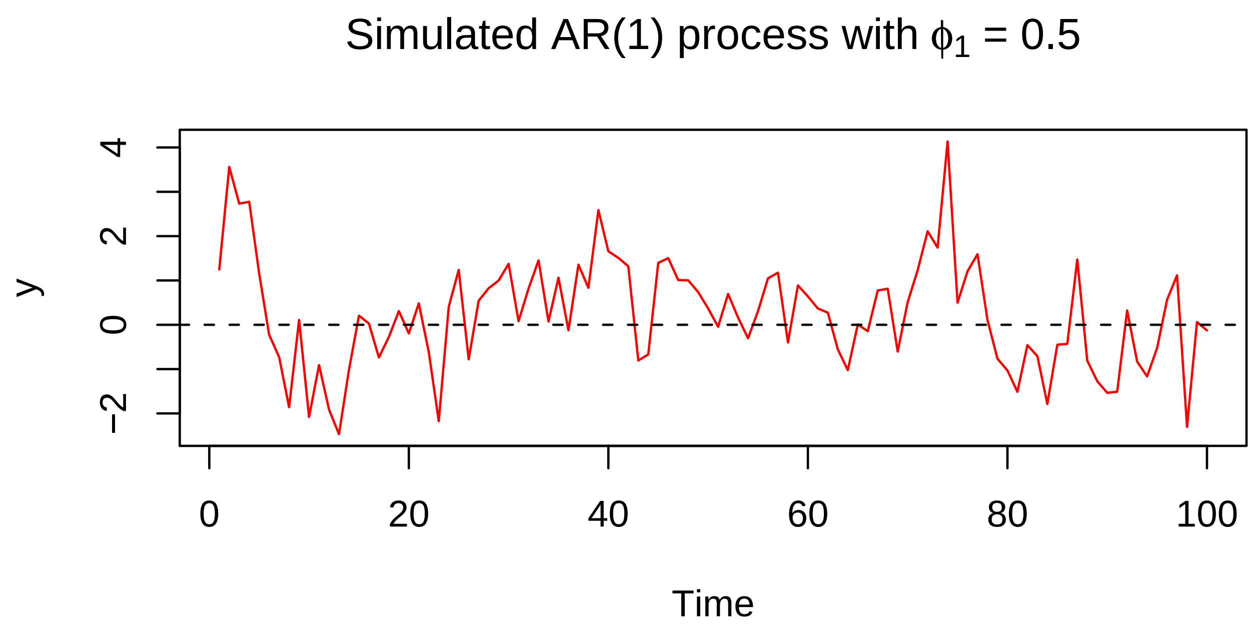
In `R`, just do `pacf(y)`


Simulated AR(1) process with $\phi_1 = 0$
y
−2
0
1
2
0
20
40
60
80
100
Time

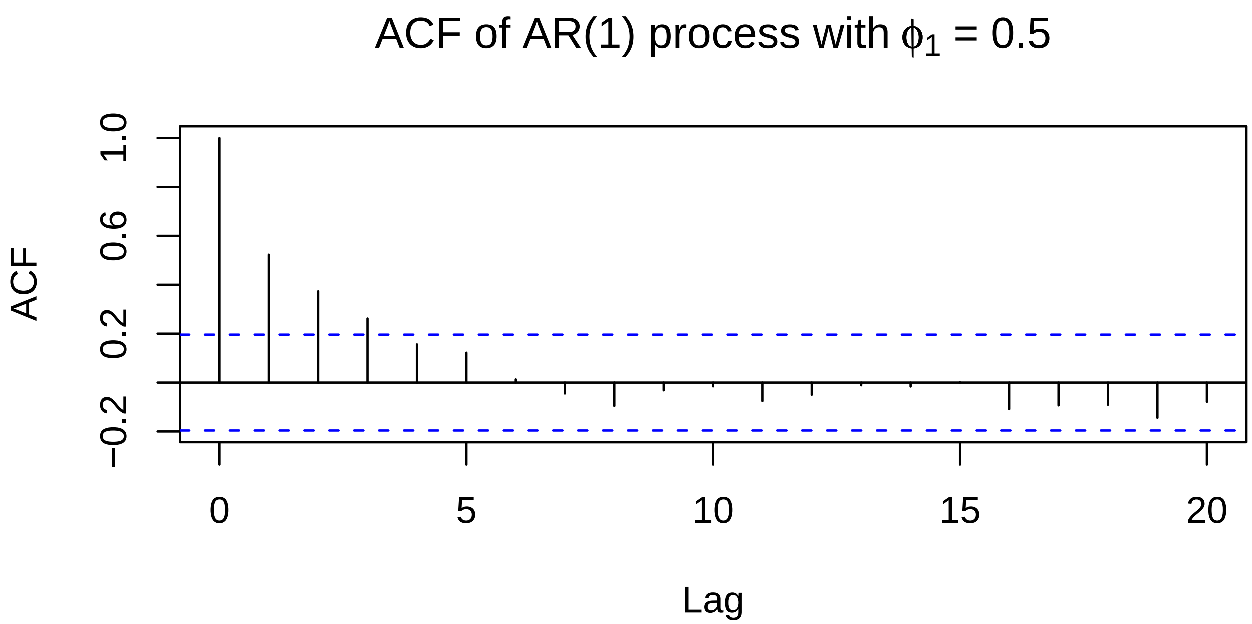

PACF of AR(1) process with $\phi_1 = 0$
Partial ACF
−0.2
0.0
0.2
5
10
15
20
Lag



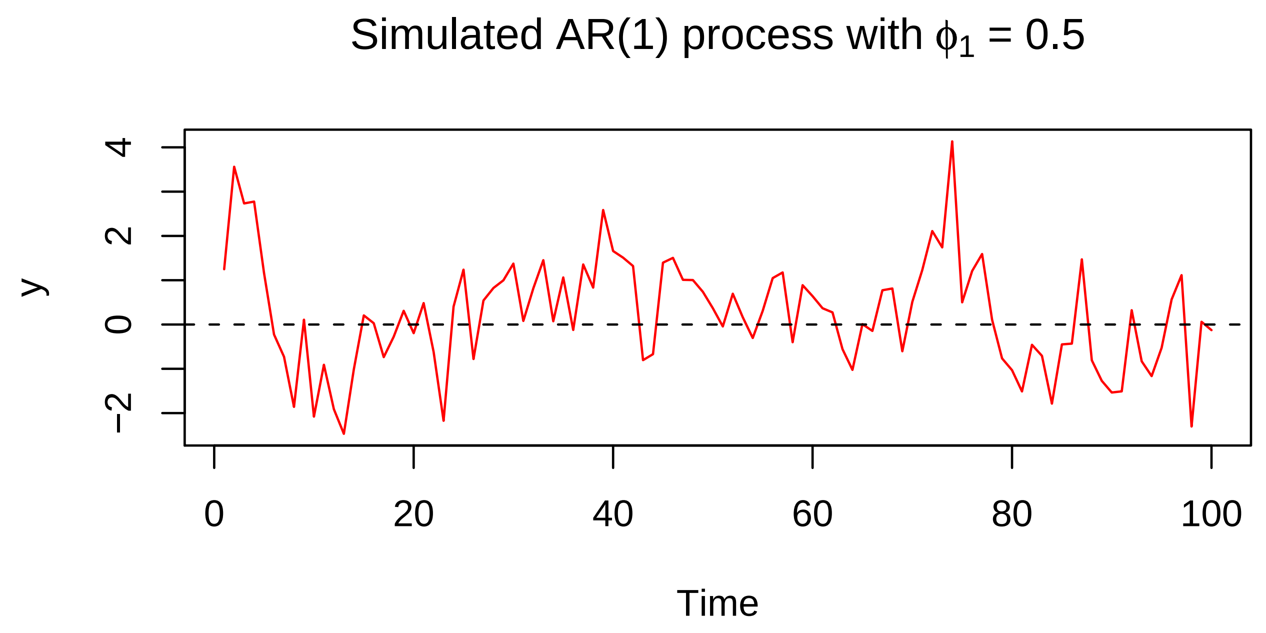
Simulated AR(1) process with ϕ_1 = 0.5
y
−2
0
2
4
0
20
40
60
80
100
Time

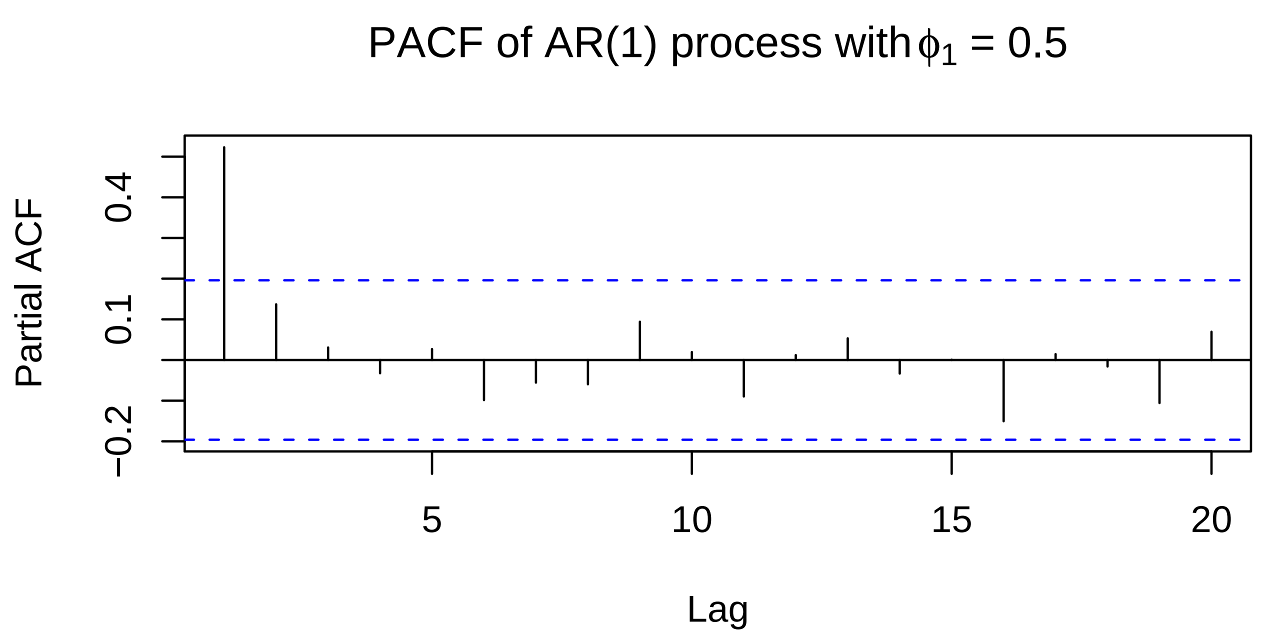


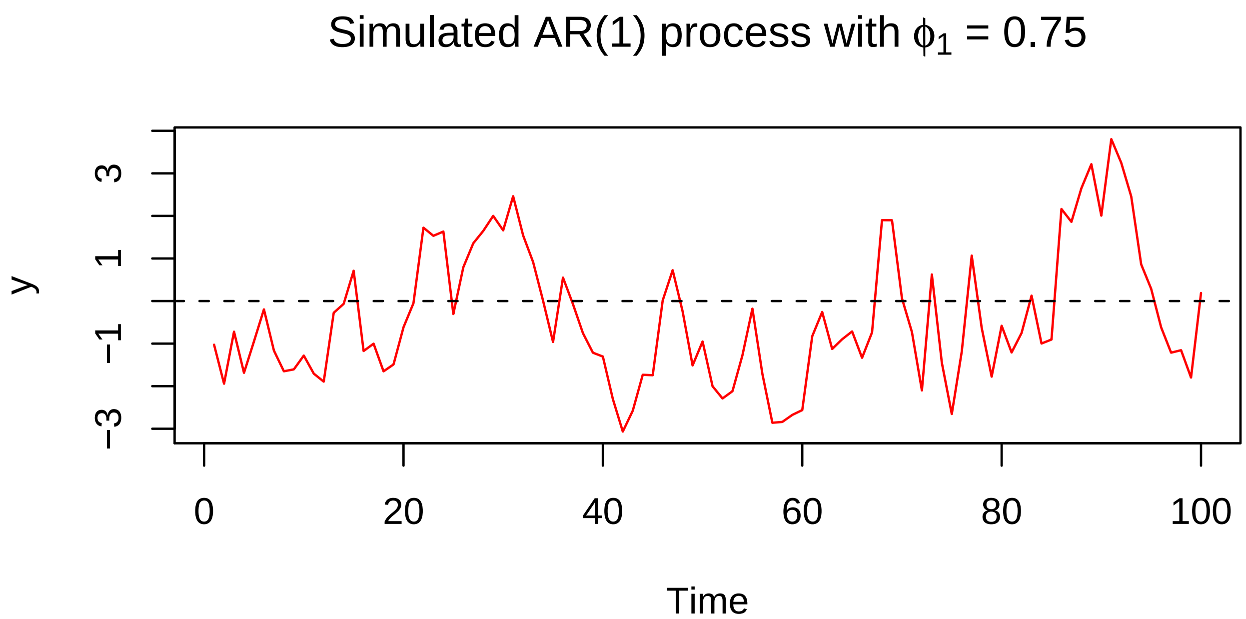
Simulated AR(1) process with $\phi_1 = 0.5$
y
4
2
0
−2
0
20
40
60
80
100
Time



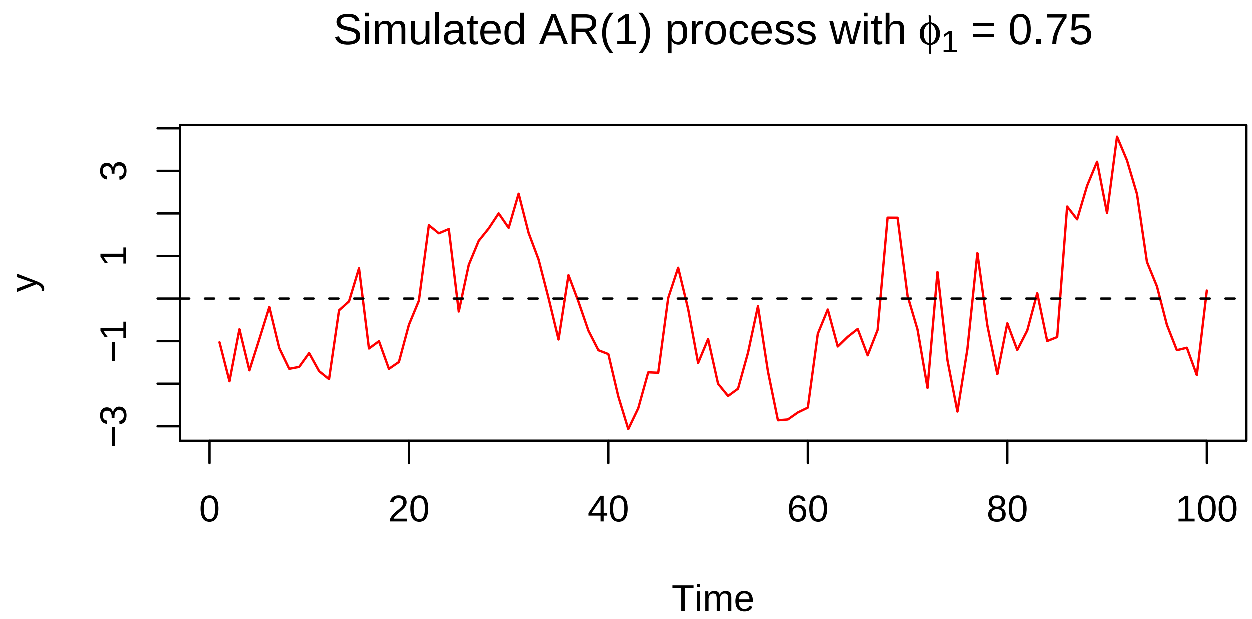
ACF of AR(1) process with $\phi_1 = 0.5$
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag

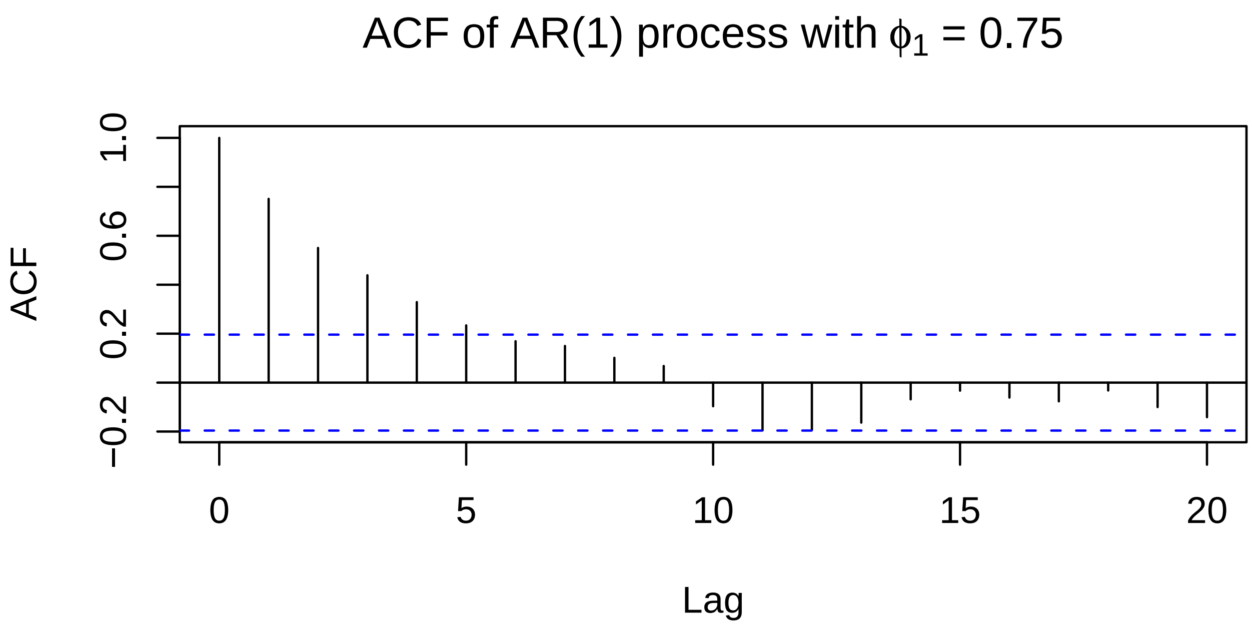

Simulated AR(1) process with ϕ1 = 0.5
y
4
2
0
−2
0
20
40
60
80
100
Time

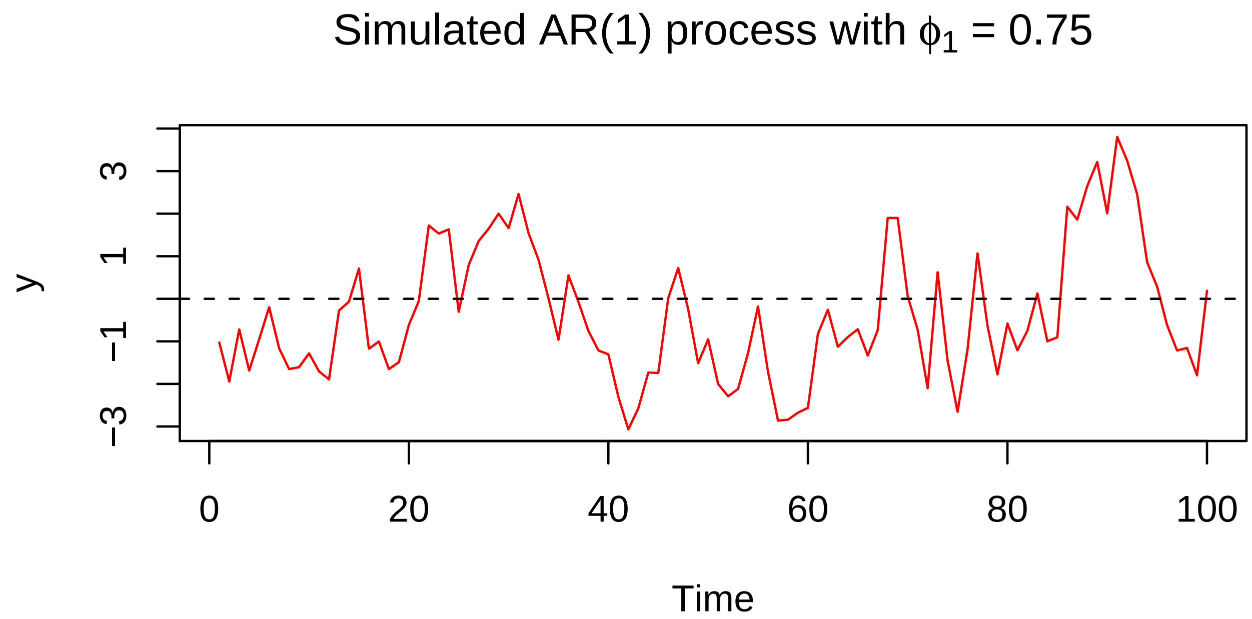

PACF of AR(1) process with ϕ1 = 0.5
Partial ACF
0.4
0.1
−0.2
5
10
15
20
Lag



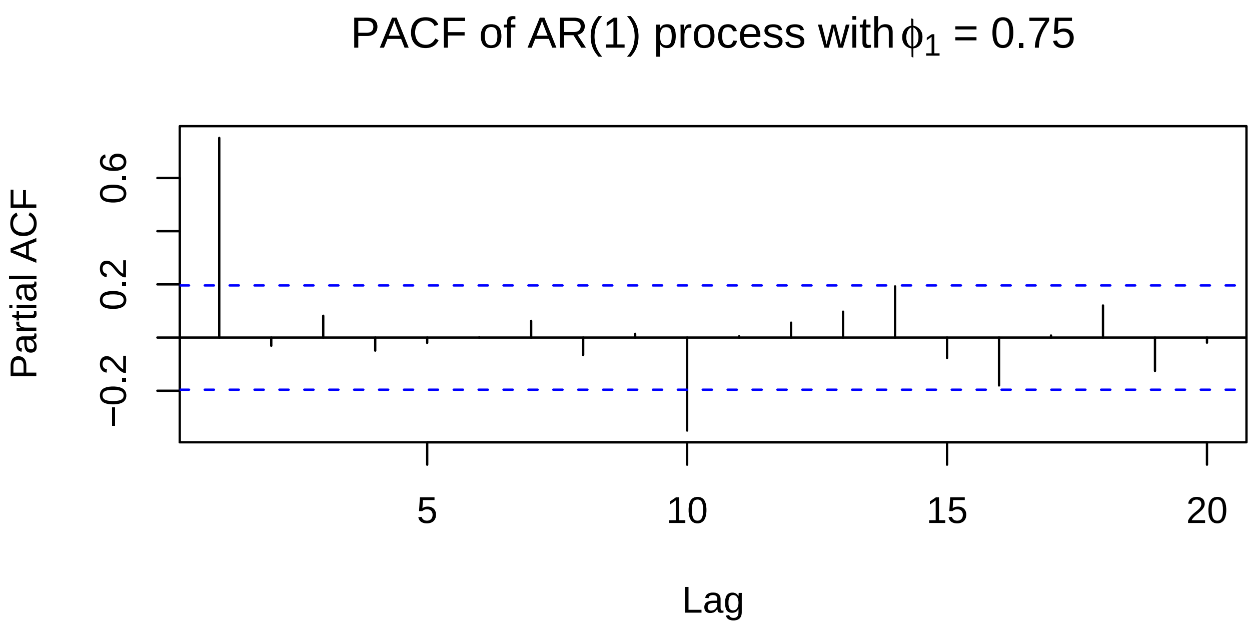
Simulated AR(1) process with $\phi_1 = 0.75$
y
−3
−1
1
3
0
20
40
60
80
100
Time


Simulated AR(1) process with ϕ1 = 0.75
y
−3
−1
1
3
0
20
40
60
80
100
Time

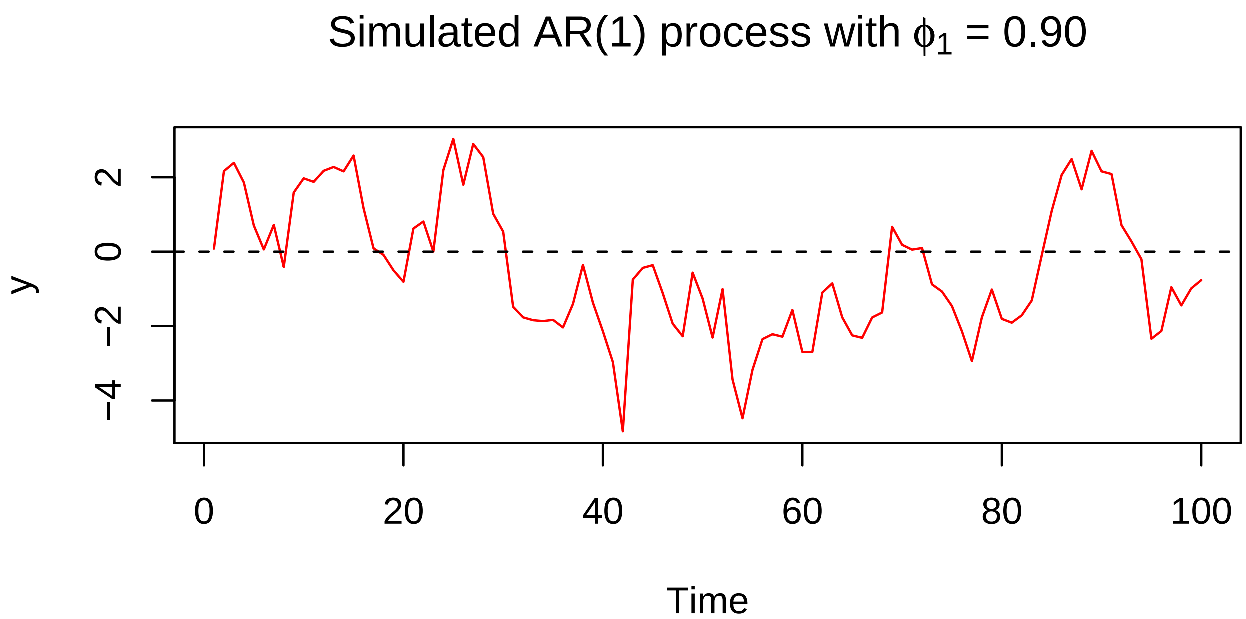

ACF of AR(1) process with ϕ1 = 0.75
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
20
Lag

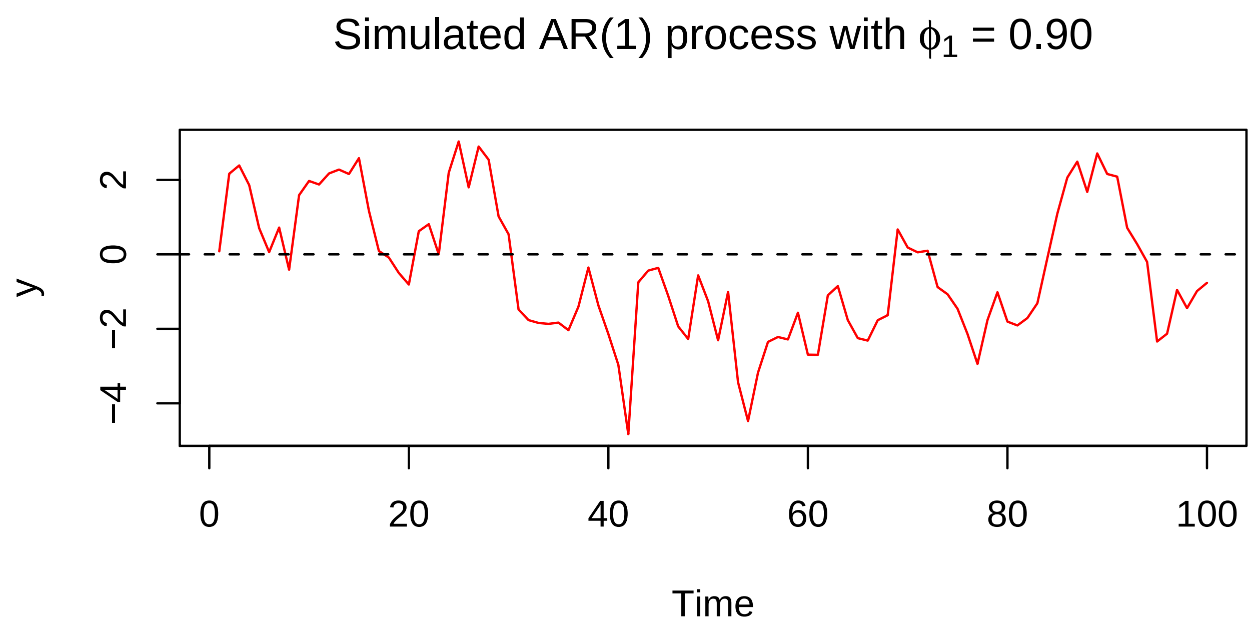


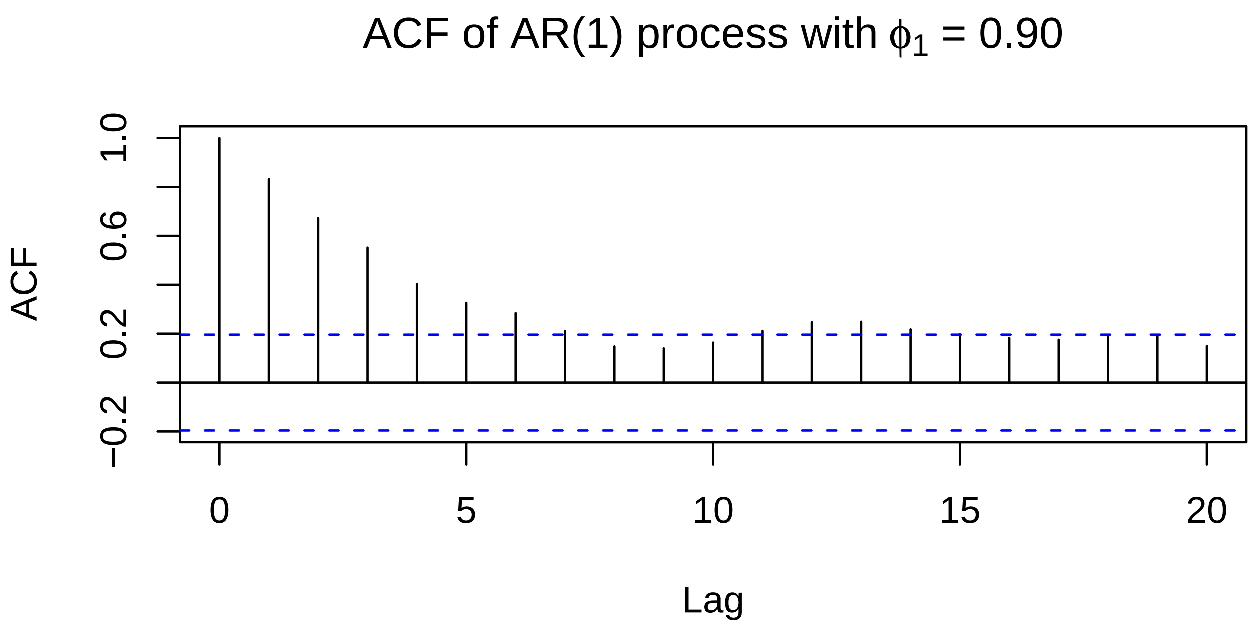
Simulated AR(1) process with ϕ1 = 0.75
y
−3
−1
1
3
0
20
40
60
80
100
Time

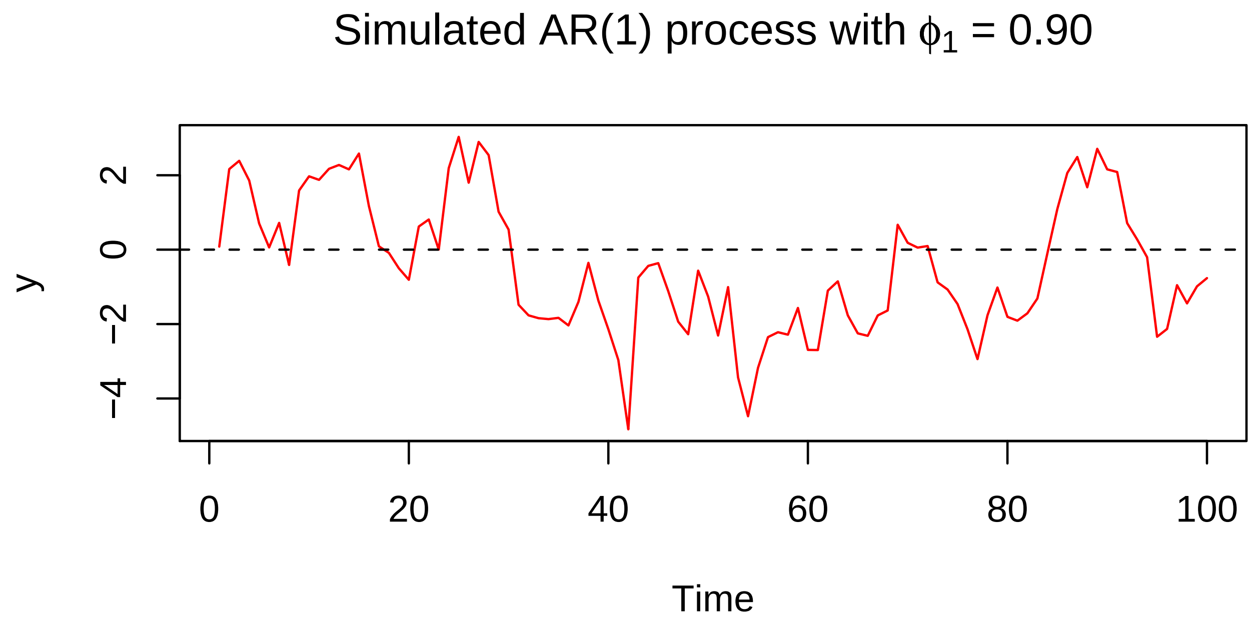


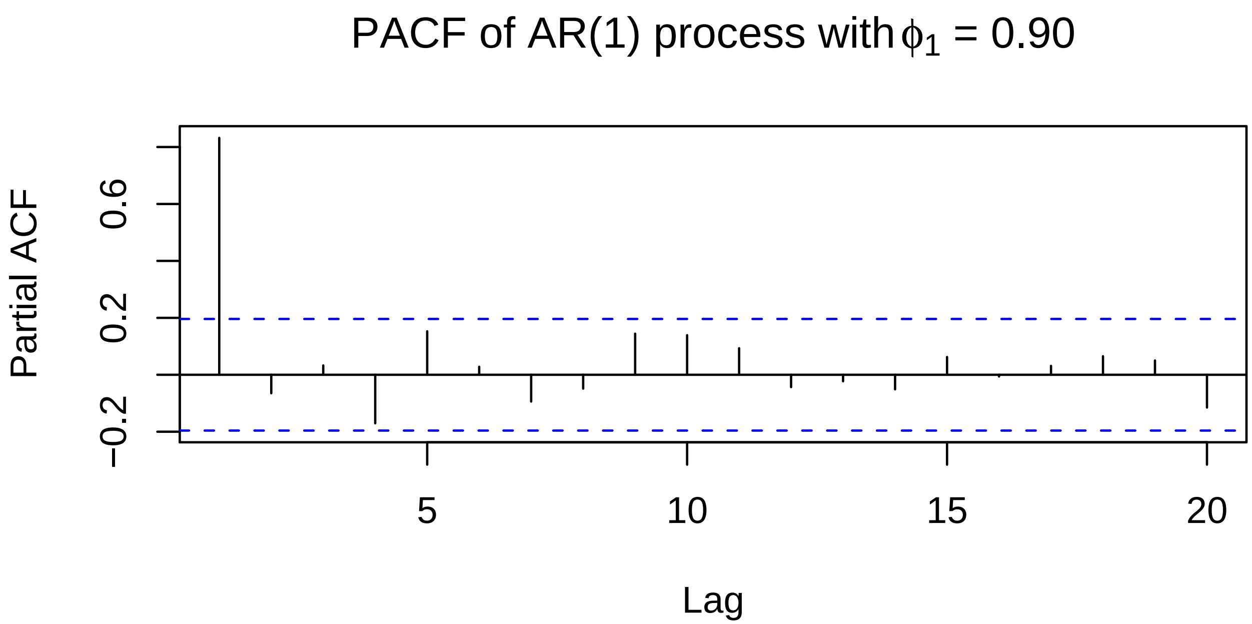
PACF of AR(1) process with ϕ1 = 0.75
Partial ACF
−0.2
0.2
0.6
5
10
15
20
Lag

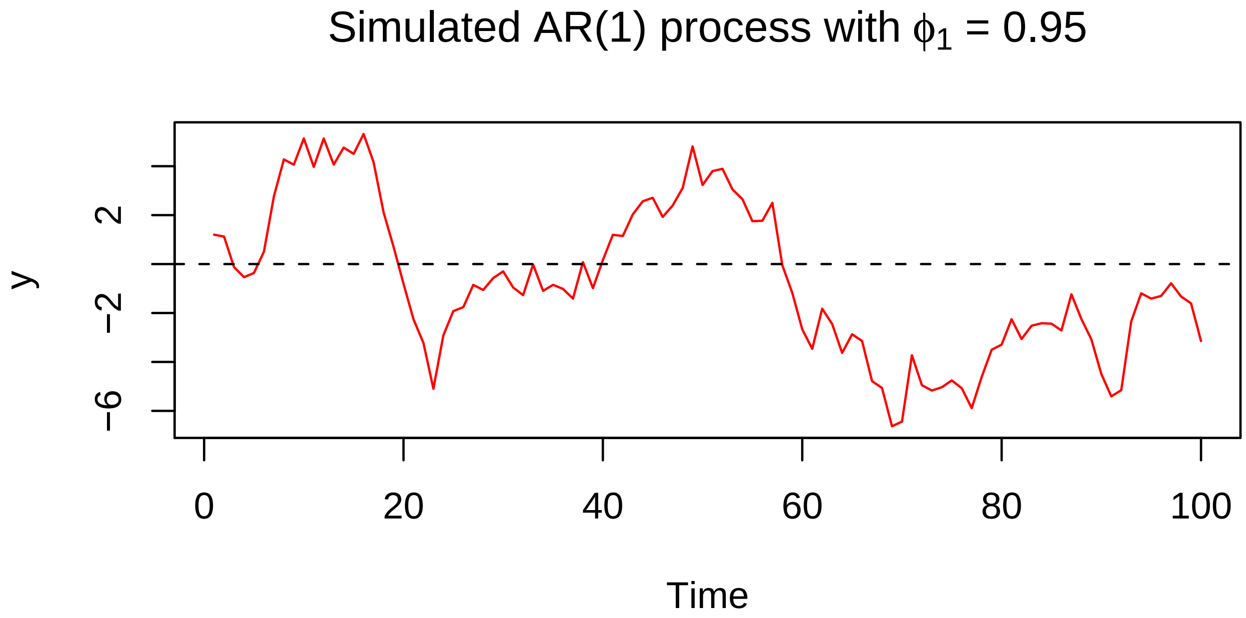

Simulated AR(1) process with ϕ_1 = 0.90
y
2
0
−2
−4
0
20
40
60
80
100
Time

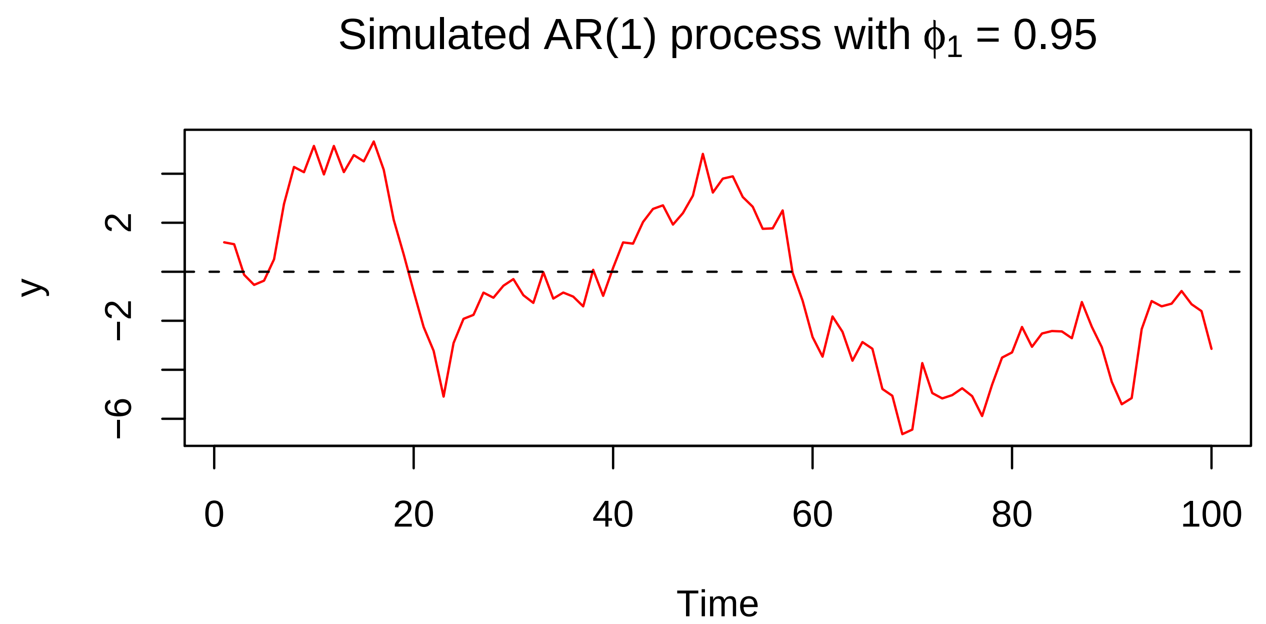

Simulated AR(1) process with ϕ_1 = 0.90
y
2
0
−2
−4
0
20
40
60
80
100
Time

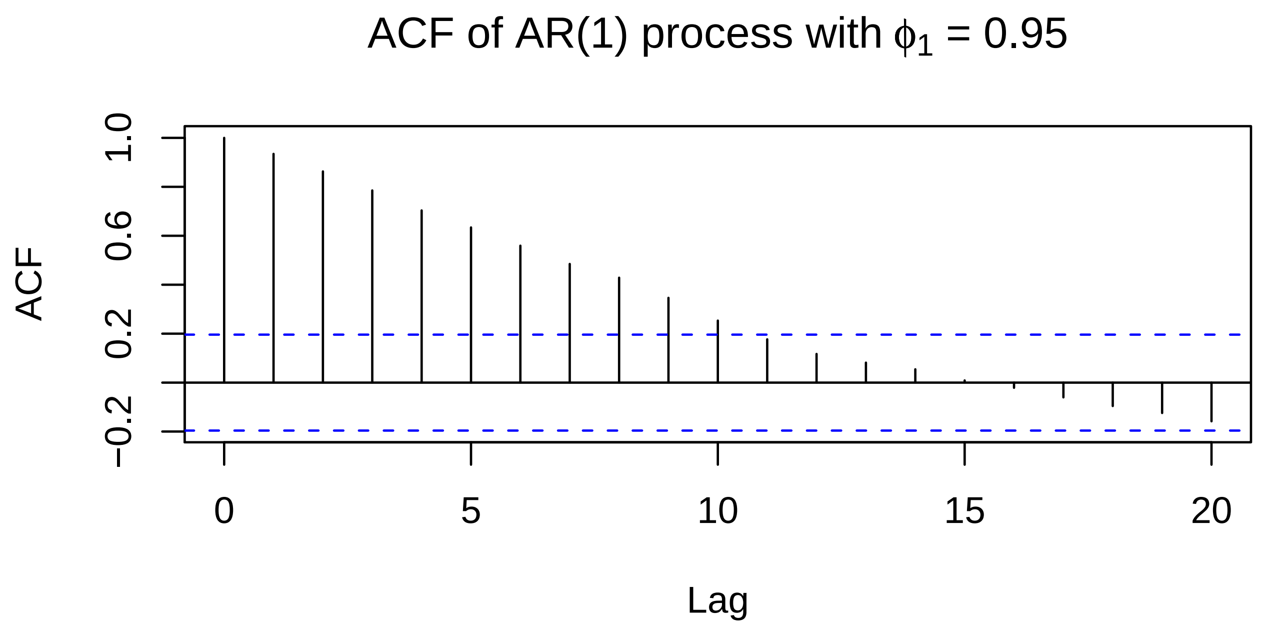

ACF of AR(1) process with ϕ_1 = 0.90
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag

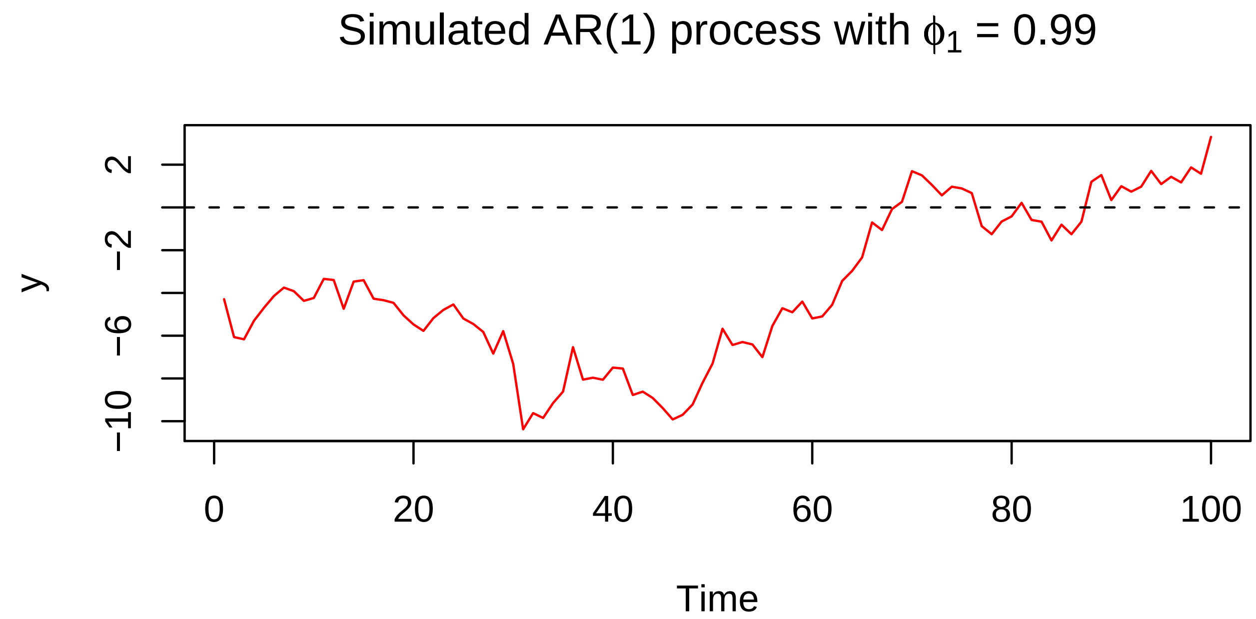

Simulated AR(1) process with ϕ_1 = 0.90
y
−4
−2
0
2
0
20
40
60
80
100
Time

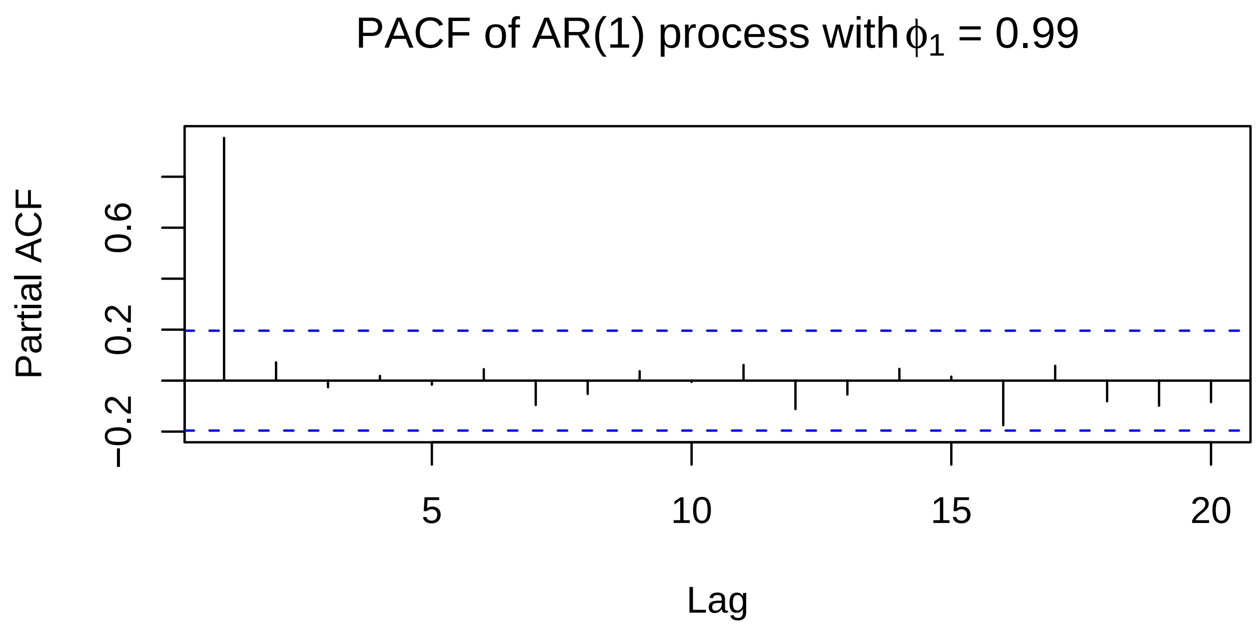

PACF of AR(1) process with ϕ_1 = 0.90
Partial ACF
−0.2
0.2
0.6
5
10
15
20
Lag


Simulated AR(1) process with ϕ_1 = 0.95
y
−6
−2
2
0
20
40
60
80
100
Time


Simulated AR(1) process with ϕ_1 = 0.95
y
2
−2
−6
0
20
40
60
80
100
Time

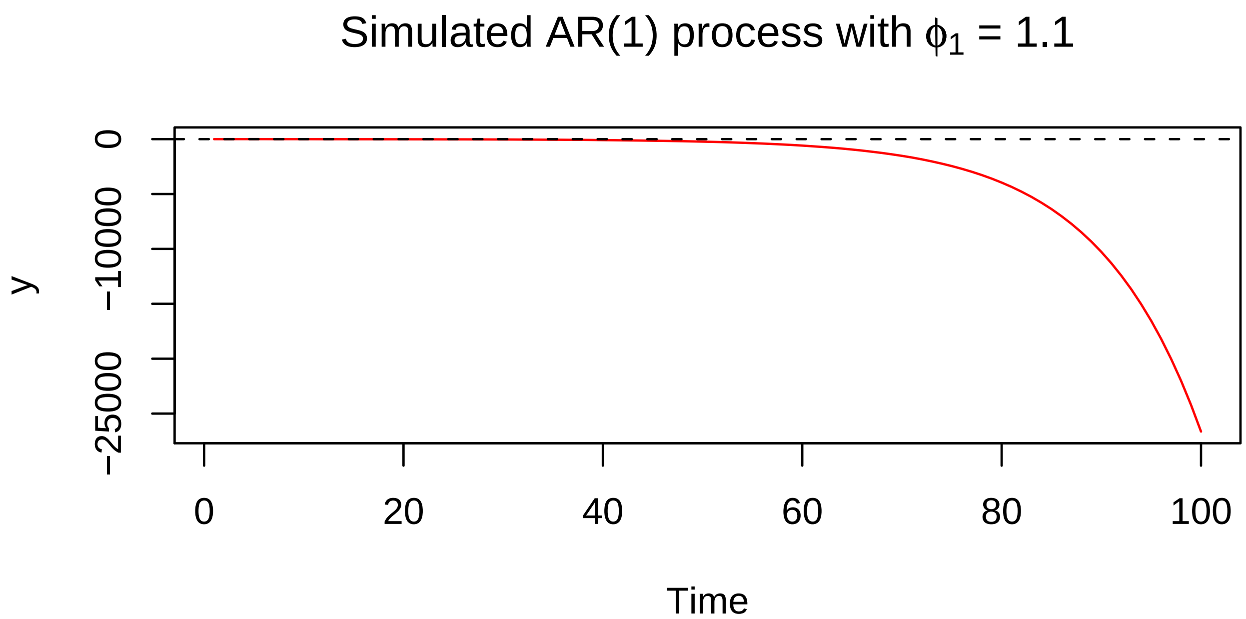

ACF of AR(1) process with ϕ_1 = 0.95
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag


Simulated AR(1) process with ϕ_1 = 0.99
y
2
−2
−6
−10
0
20
40
60
80
100
Time


PACF of AR(1) process with ϕ_1 = 0.99
Partial ACF
0.6
0.2
−0.2
5
10
15
20
Lag

Observations about AR(1) when $|\phi_1| < 1$

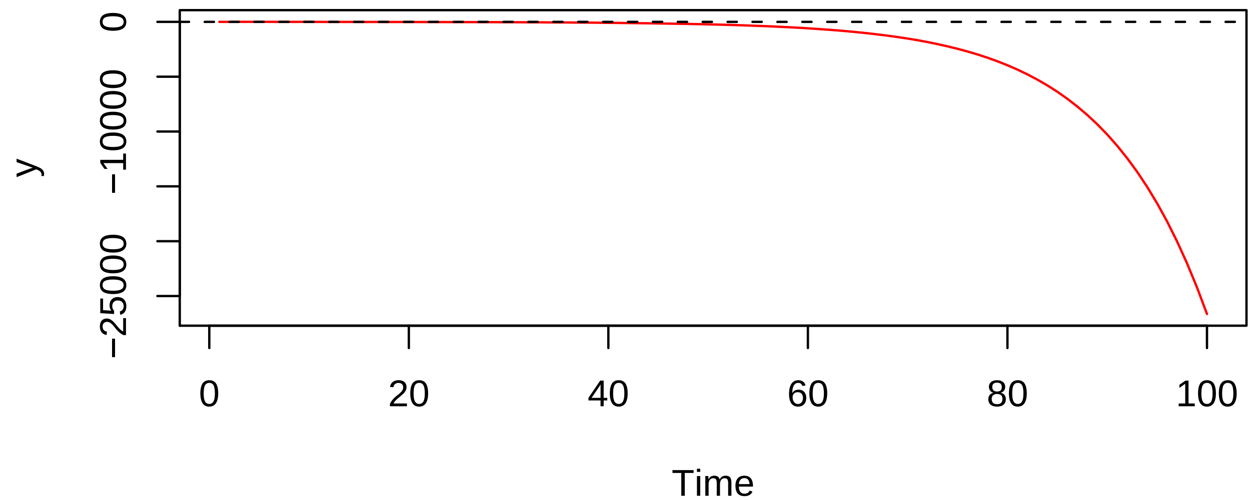
- As $|\phi_1|$ approaches 0, series reverts to its mean at 0 quickly.
- As $|\phi_1|$ approaches 1, series takes longer to revert to mean.
- Still gets there eventually. (Even for $\phi_1 = 0.99$?)
- ACF appears to gradually decline in the lag, as expected
- ACF decays more slowly as $|\phi_1|$ approaches 1
- But for all process, PACF is large only for lag of 1, because all series are AR(1)



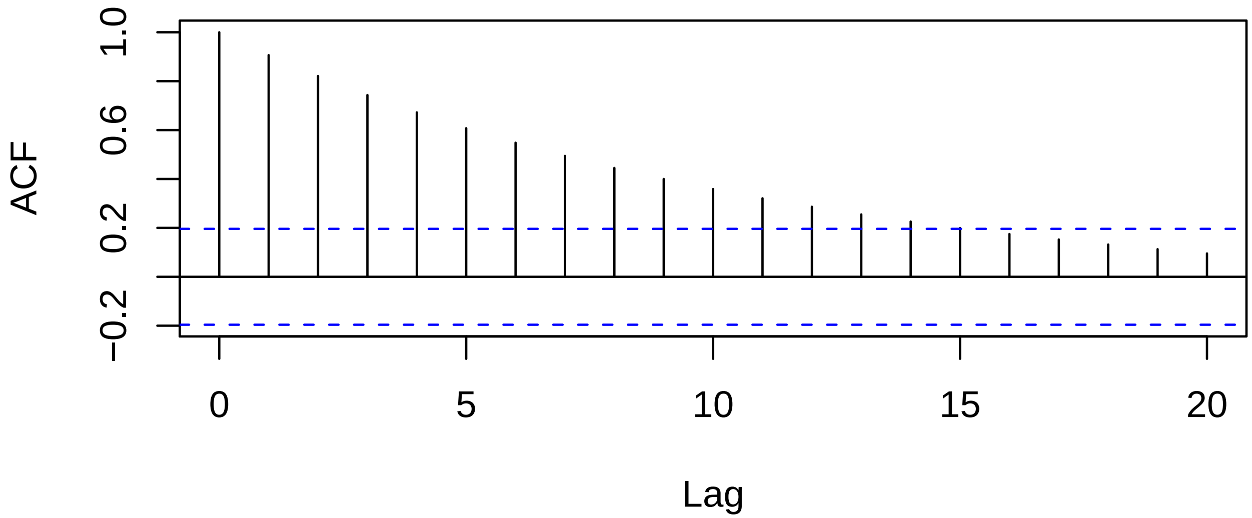
What's happening here?

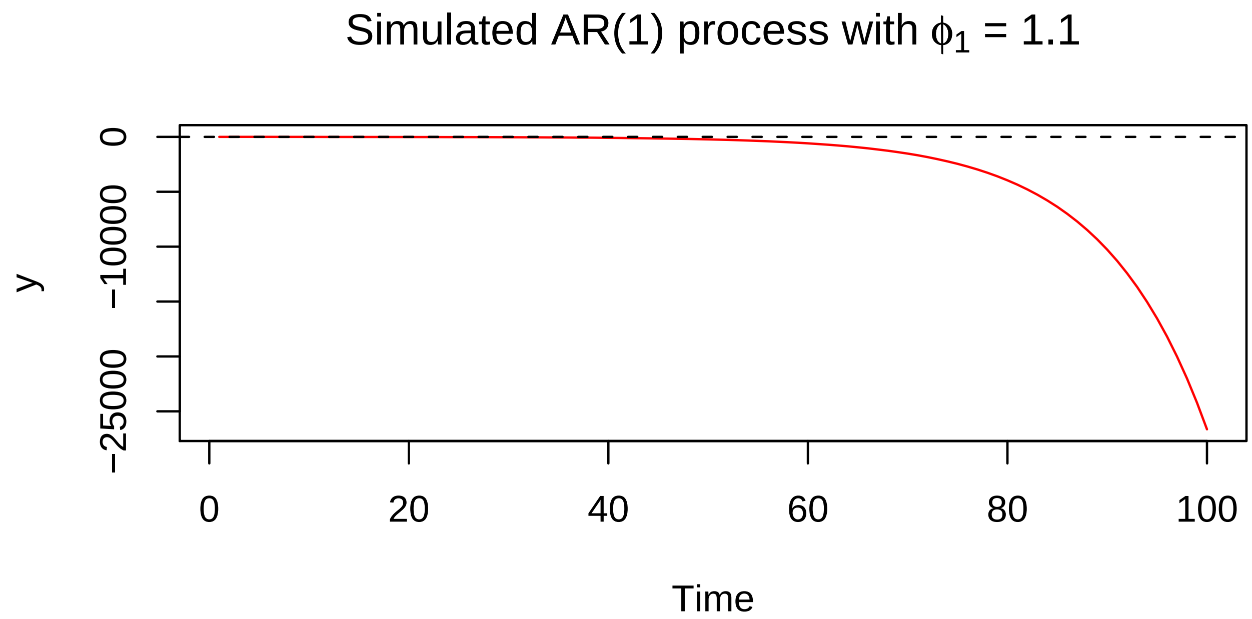
Explosive process. Shouldn't ever see this in real data

Simulated AR(1) process with $\phi_1 = 1.1$

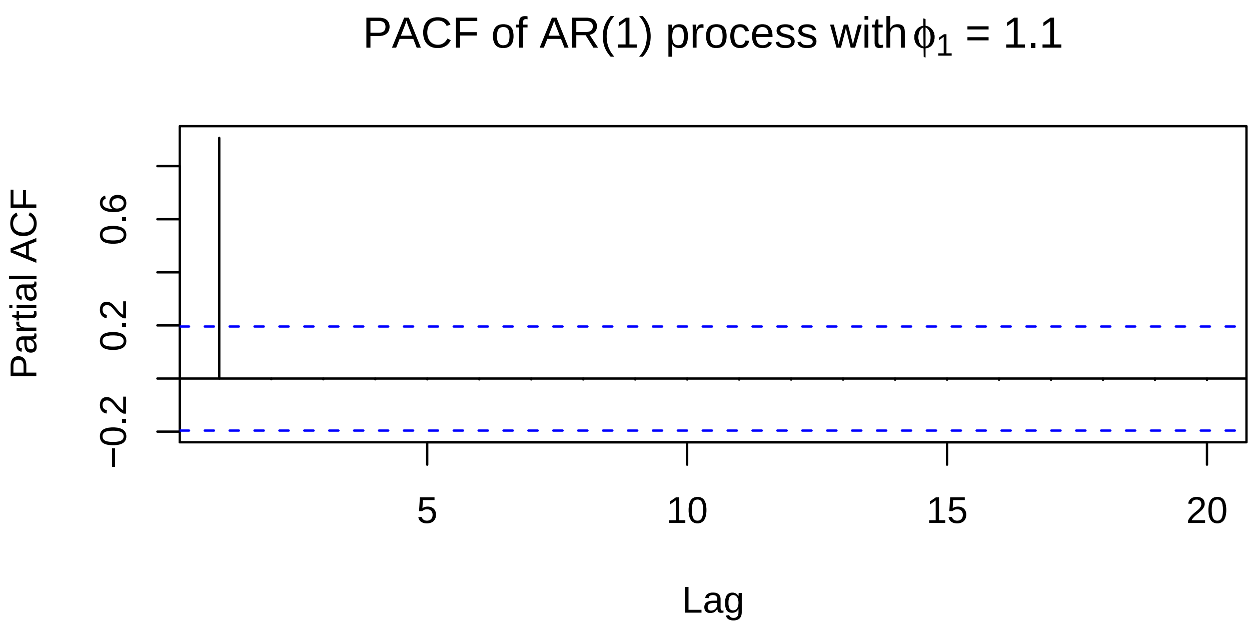


ACF of AR(1) process with $\phi_1 = 1.1$

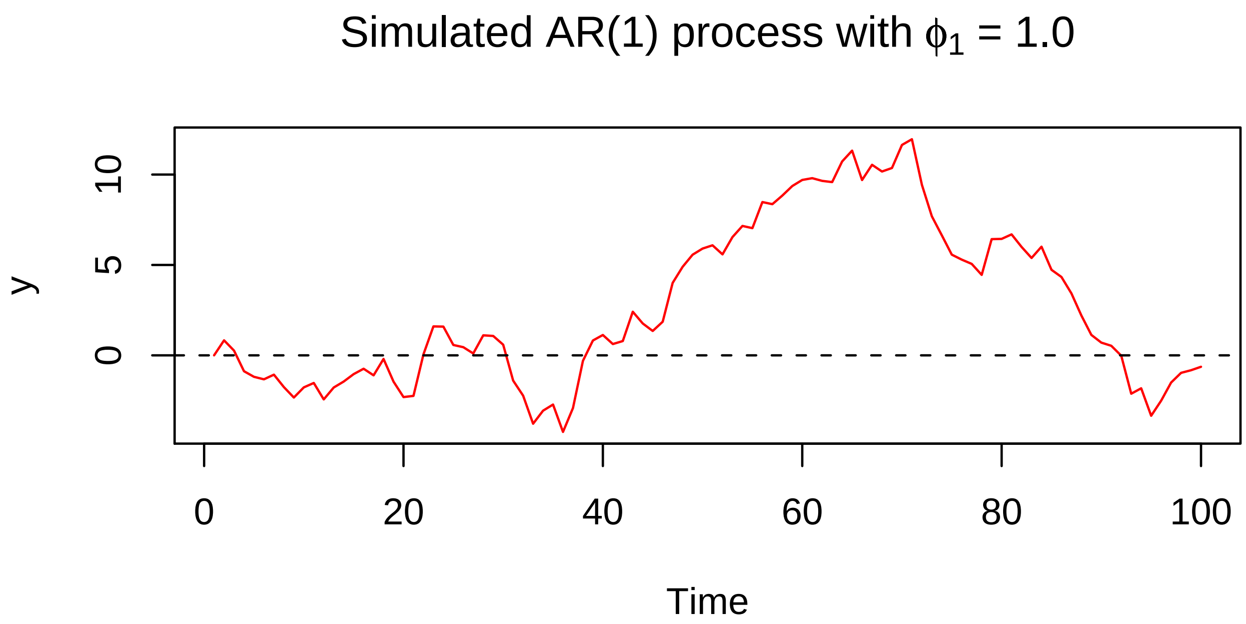




Simulated AR(1) process with $\phi_1 = 1.1$
y
0
−10000
−25000
0
20
40
60
80
100
Time



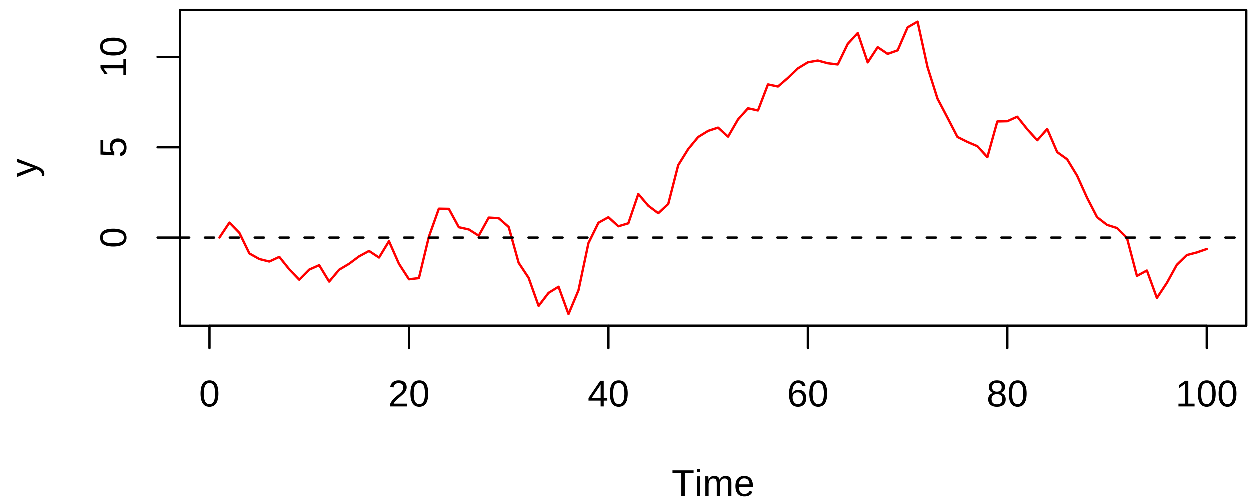
PACF of AR(1) process with $\phi_1 = 1.1$
Partial ACF
0.6
0.2
−0.2
5
10
15
20
Lag



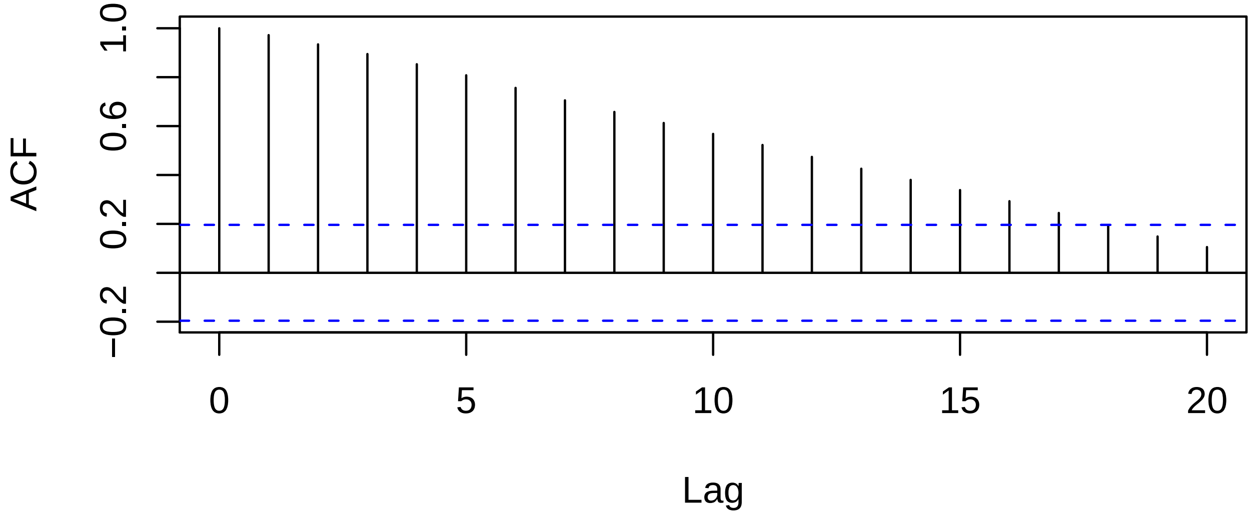
Is $\phi_1 = 1$ *really* that different from $\phi_1 = 0.99$?

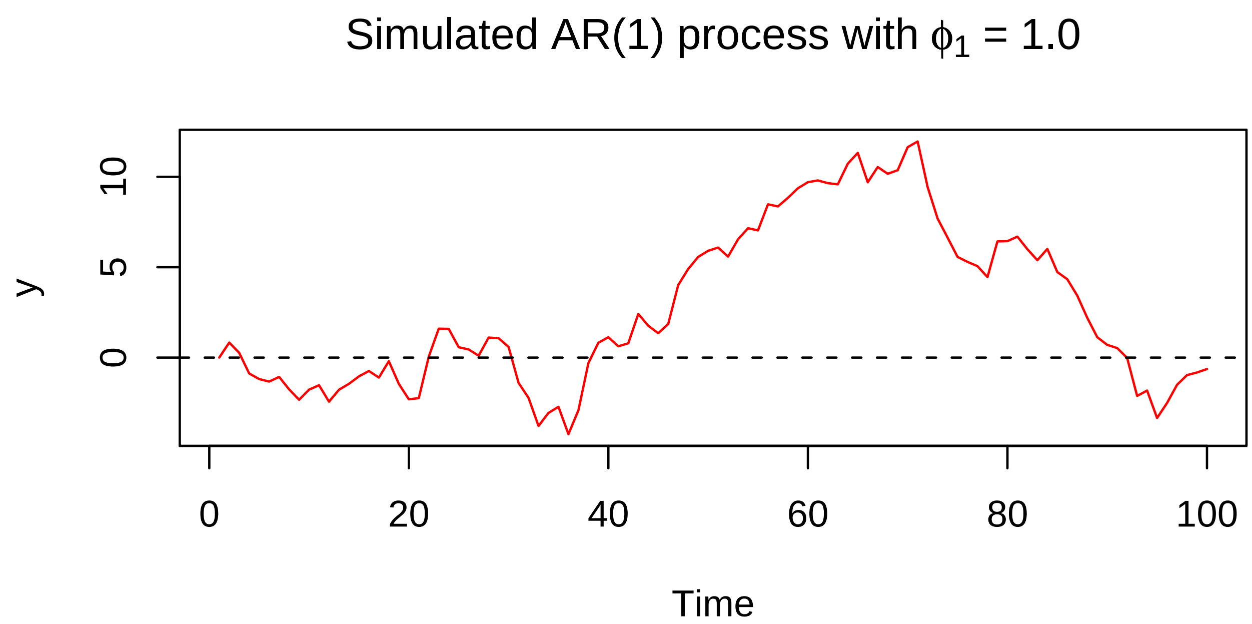
Will both eventually bounce around the mean?

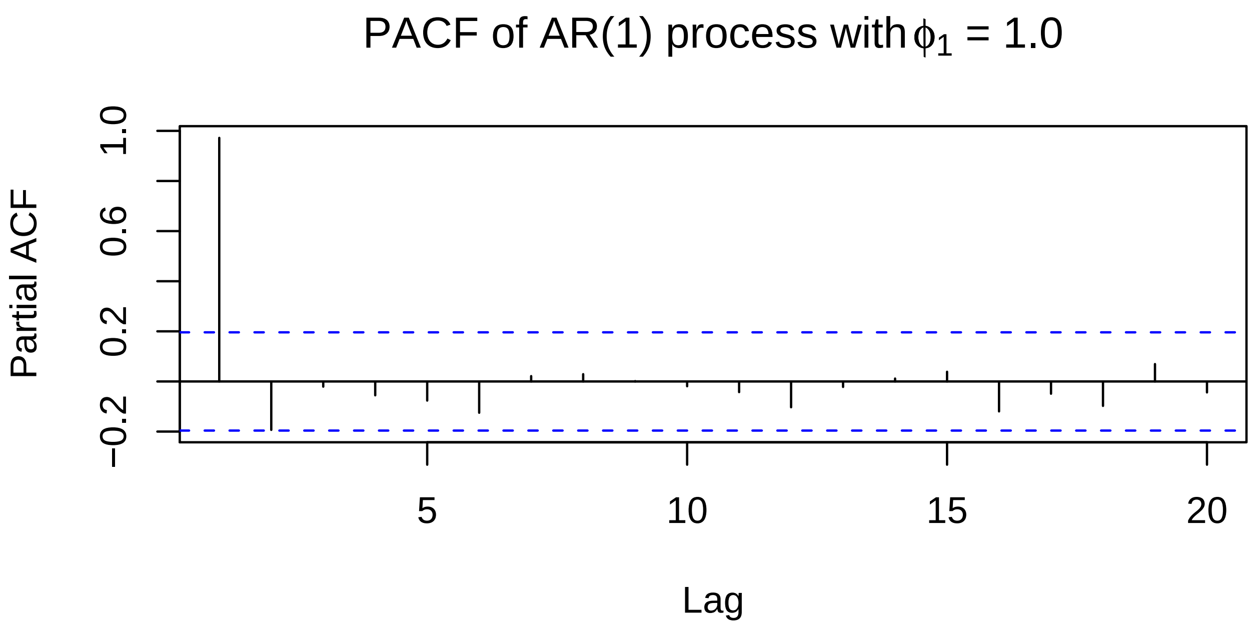
Simulated AR(1) process with $\phi_1 = 1.0$

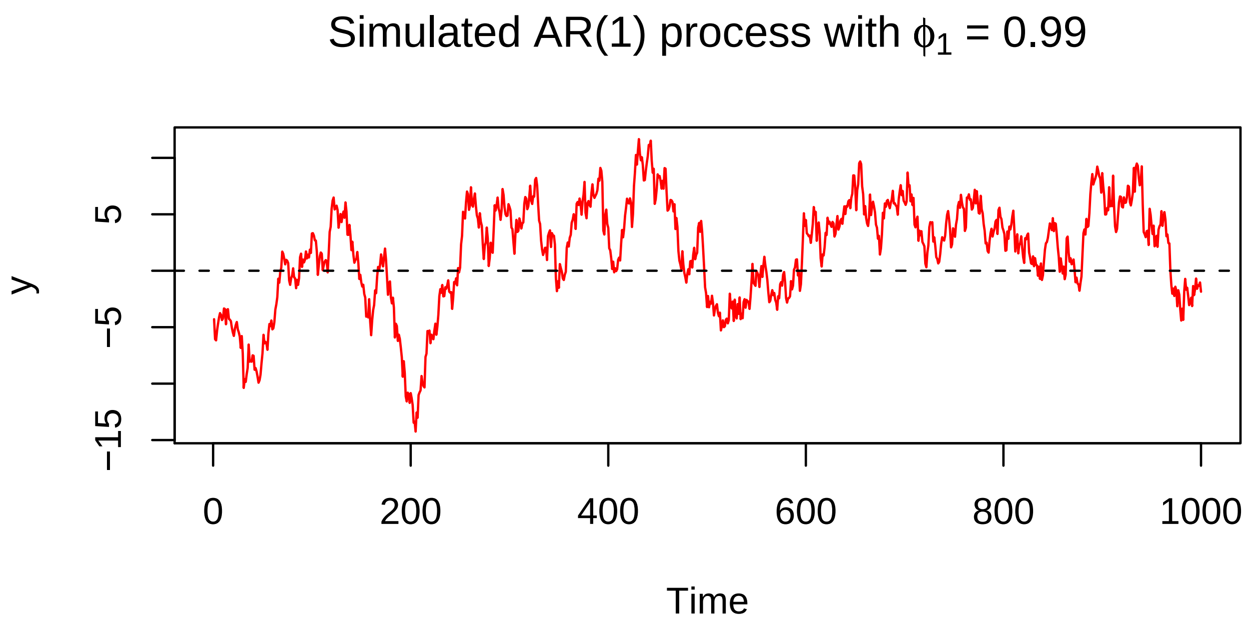


ACF of AR(1) process with $\phi_1 = 1.0$




Simulated AR(1) process with ϕ_1 = 1.0
y
10
5
0
0
20
40
60
80
100
Time


PACF of AR(1) process with ϕ_1 = 1.0
Partial ACF
1.0
0.6
0.2
−0.2
5
10
15
20
Lag

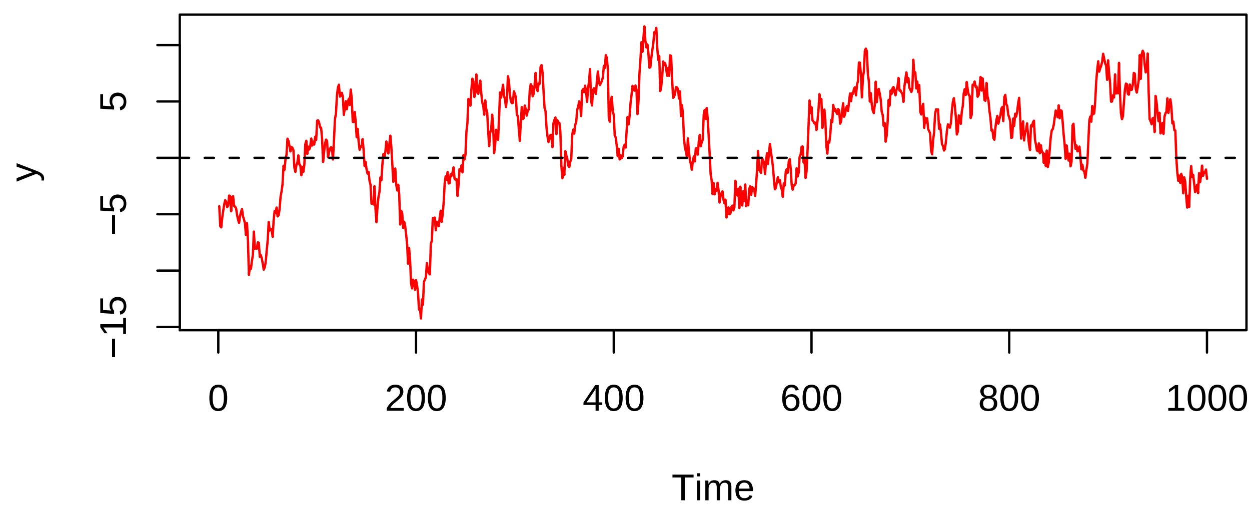


Let's extend the $\phi_1 = 0.99$ sequence

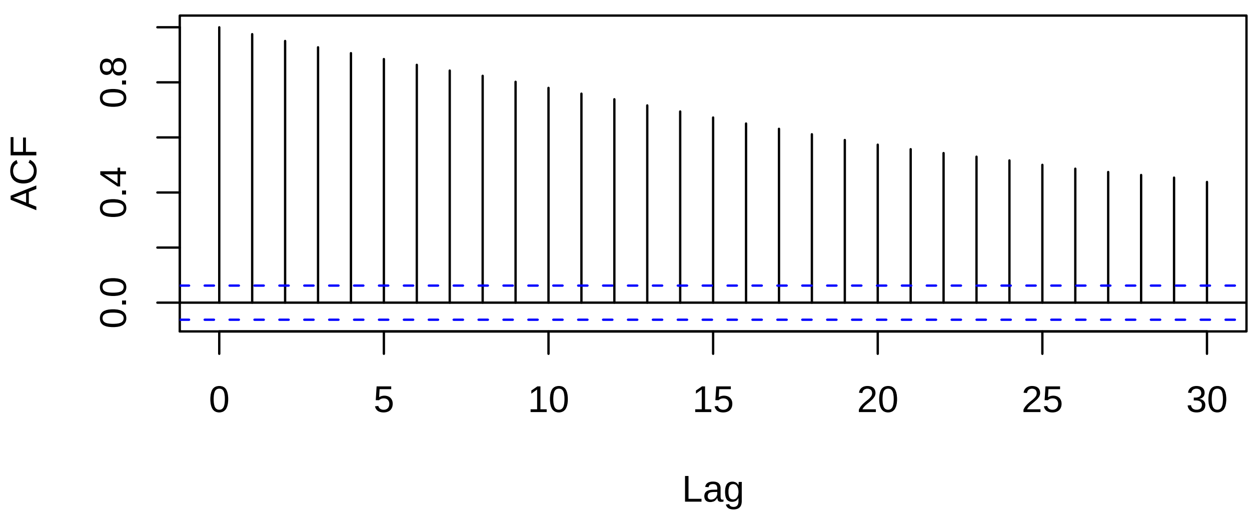
Note that it stays away from the mean for a while, but always comes back

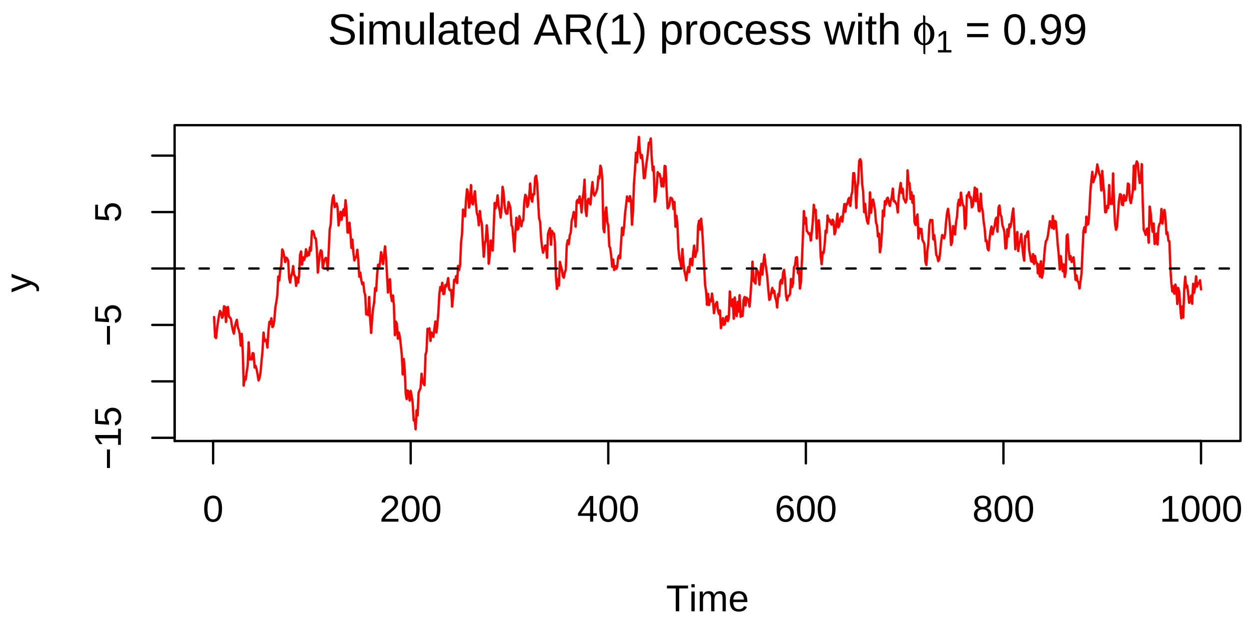
So does the $\phi_1 = 0.9999$ sequence, and the $\phi_1 = 0.999999$ sequence

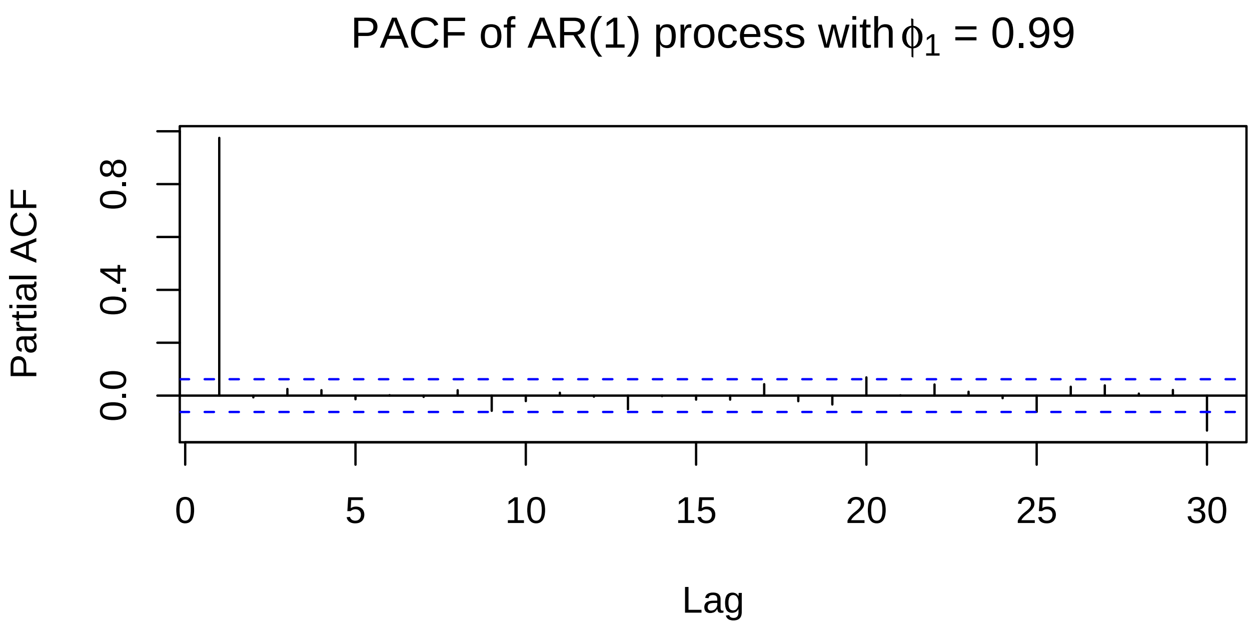
Simulated AR(1) process with $\phi_1 = 0.99$

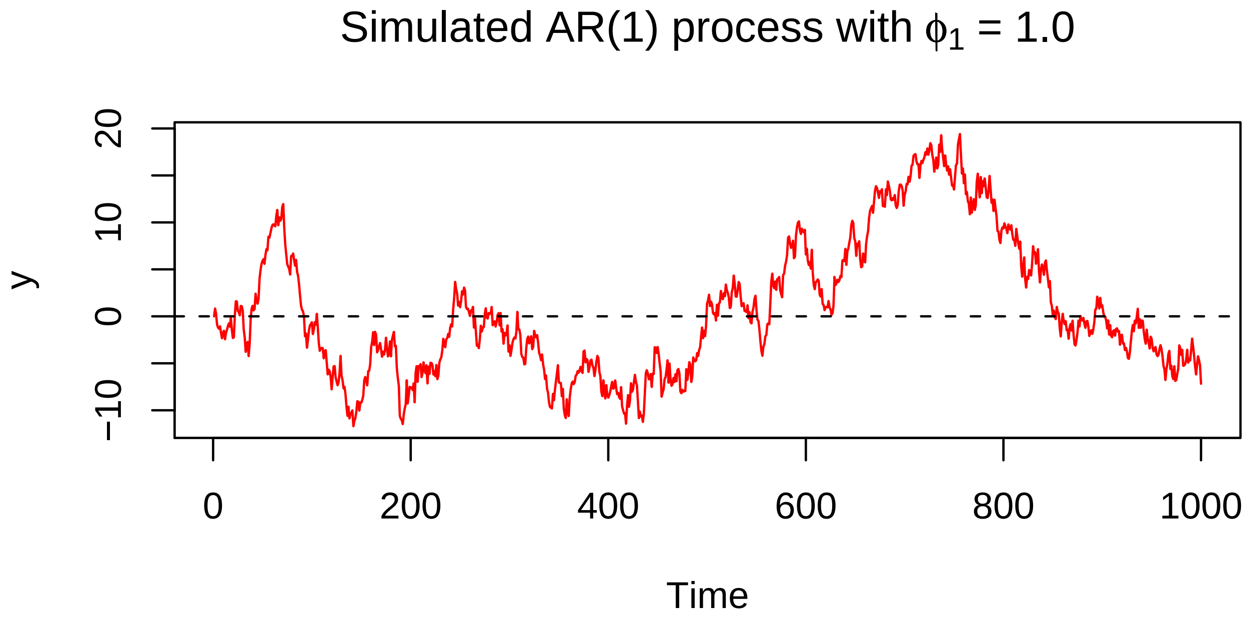


ACF of AR(1) process with $\phi_1 = 0.99$




Simulated AR(1) process with $\phi_1 = 0.99$
y
−15
−5
5
0
200
400
600
800
1000
Time


PACF of AR(1) process with $\phi_1 = 0.99$
Partial ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag

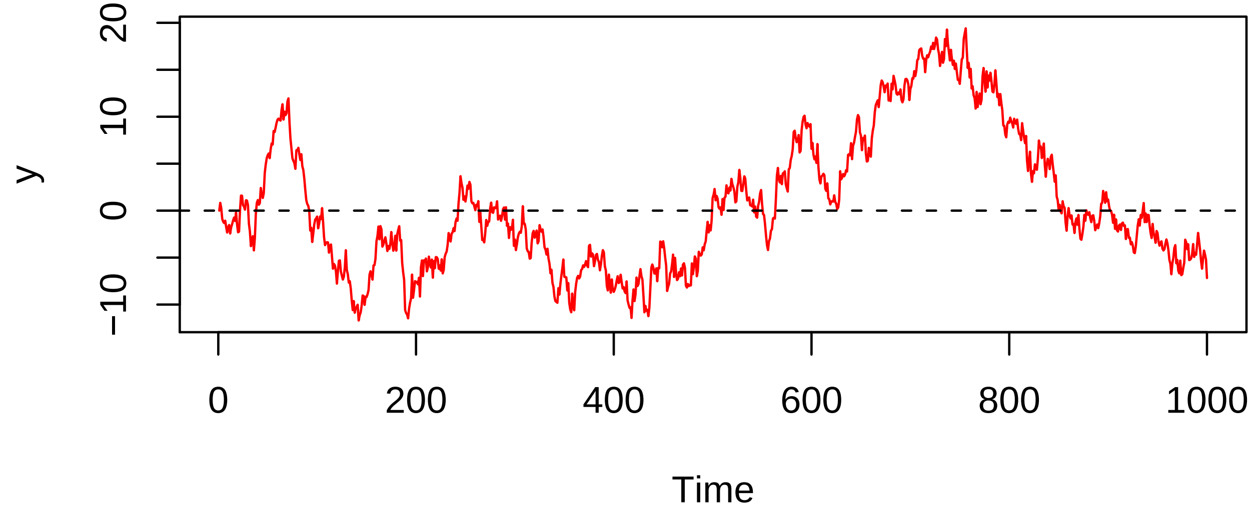


$\phi_1 = 1$ doesn't mean revert ever.

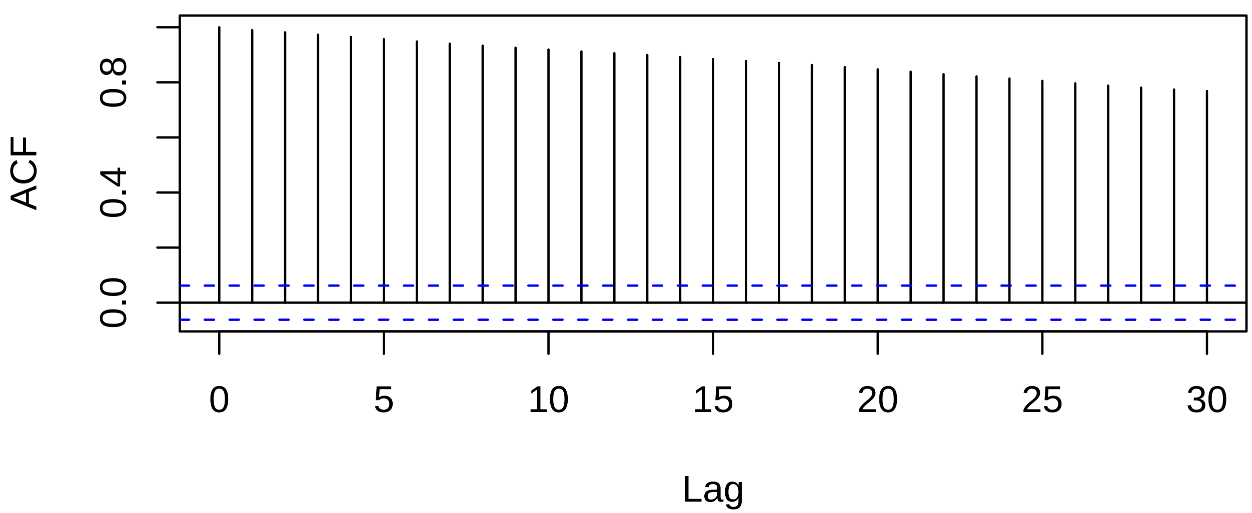
Only passes back through 0 by chance.

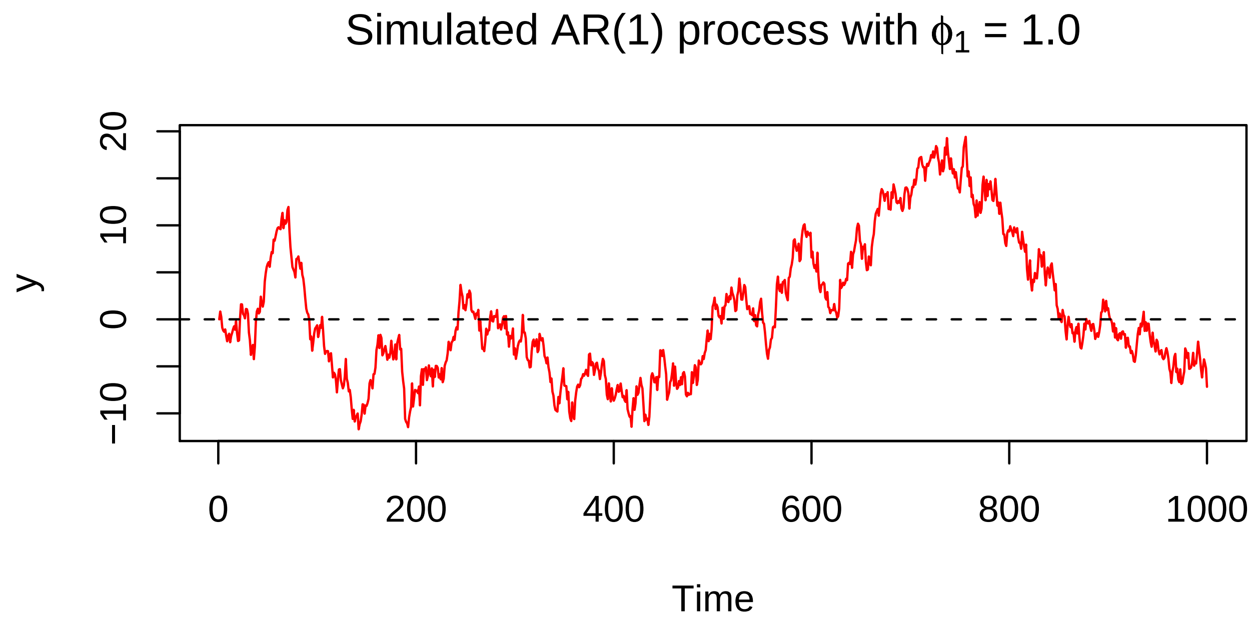
Major difference from $\phi_1 = 0.99$, but almost identical ACF, PACF

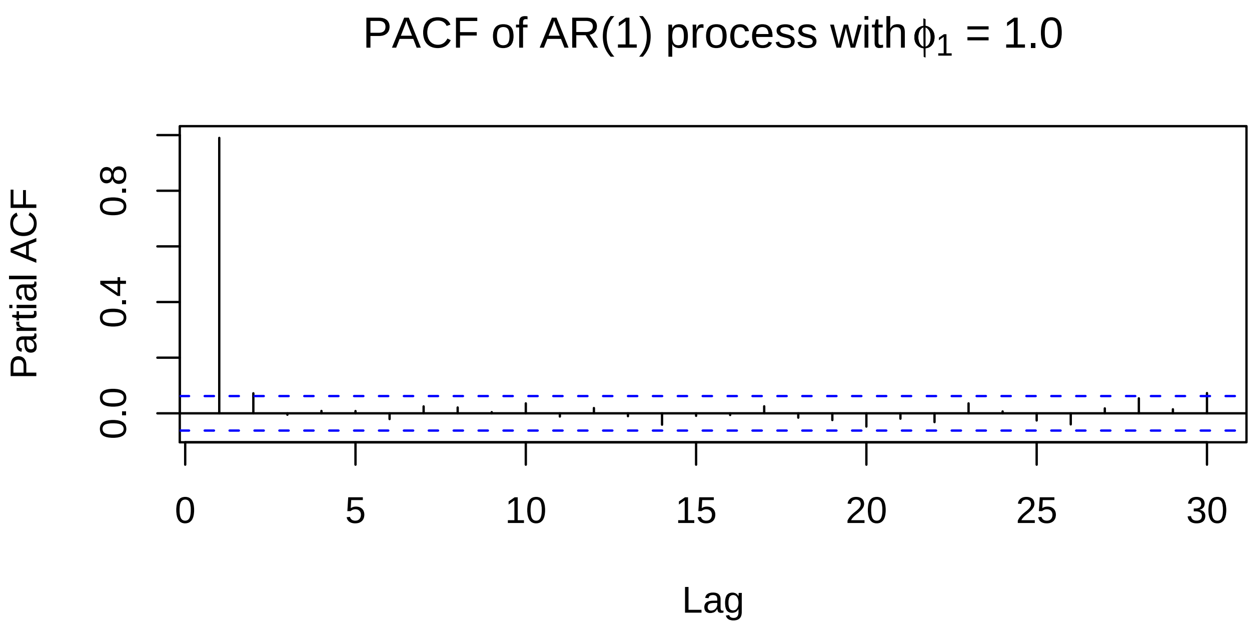
Simulated AR(1) process with $\phi_1 = 1.0$



ACF of AR(1) process with $\phi_1 = 1.0$




Simulated AR(1) process with ϕ_1 = 1.0
y
20
10
0
−10
0
200
400
600
800
1000
Time


PACF of AR(1) process with ϕ_1 = 1.0
Partial ACF
0.8
0.4
0.0
0
5
10
15
20
25
30
Lag

Stationary Series

AR(1) processes with $|\phi_1| \geq 1$ lack three related properties which all AR(1) processes with smaller ϕ_1 possess:

1. Mean stationarity

2. Covariance stationarity

3. Ergodicity

Mean stationarity

A time series is *mean stationary* if its mean does not depend on t

Formally, $\mathrm{E}(y_t) = \mu$ for all t

Note that we are abstracting from trends and other covariates

Suppose we add t itself as a covariate

$$y_t = 0.5y_{t-1} + t\beta + \varepsilon_t$$

Although this series trends upward, the time series component remain mean-stationary

Covariance stationarity

A time series is *covariance stationary* if neither the mean of x_t nor the covariance of x_t and x_{t+k} depend on t

The covariance may still depend on the length of time between two observations, k

Formally,

$$\mathrm{cov}(y_t, y_{t+k}) = \mathrm{E}\left((y_t - \mu)(y_{t+k} - \mu)\right) = \gamma_k$$

for all t and k

Also known as "weak stationarity"

Ergodicity

A time series is *ergodic* if it converges in probability to its mean

Formally, ergodicity implies $\mathrm{E}(y_t) \rightarrow \mu$ as $t \rightarrow \infty$

Usually the same in practice as stationarity

Unless μ itself is a random function of time

Non-stationary time series

Non-stationary series are "random walks"

Non-stationarity creates several problems

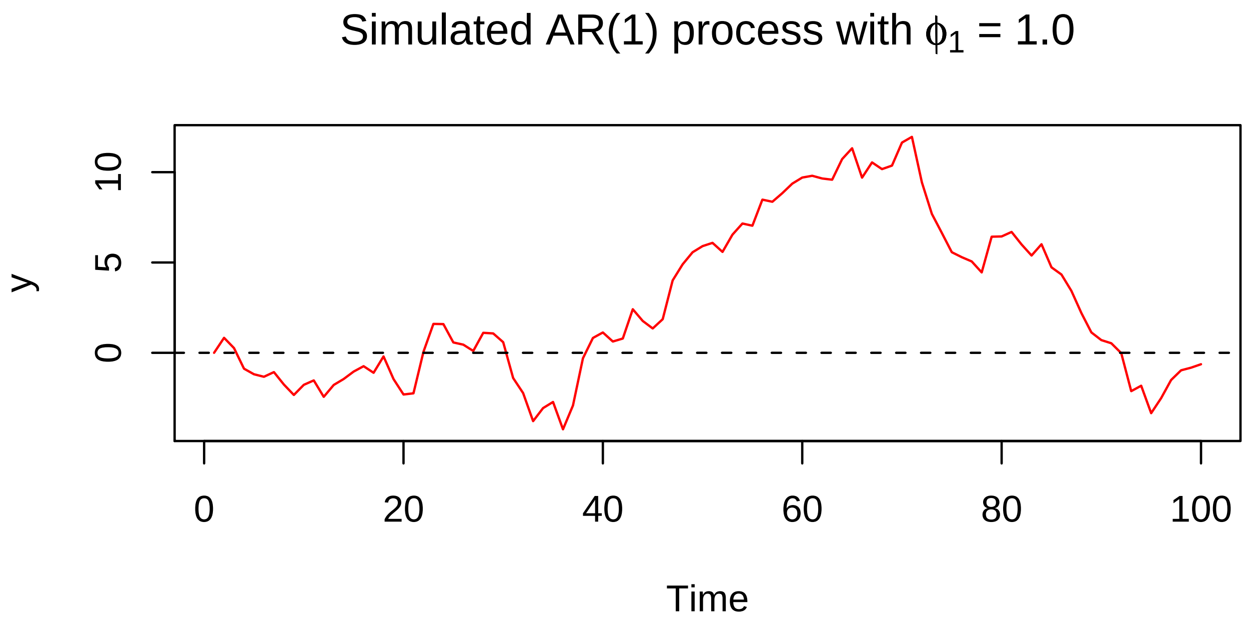
- The ACF and PACF are not defined (since covariances depend on t),
  so hard to distinguish a random walk ($\phi = 1$)
  from a stationary process with large lags ($\phi = 0.99$)

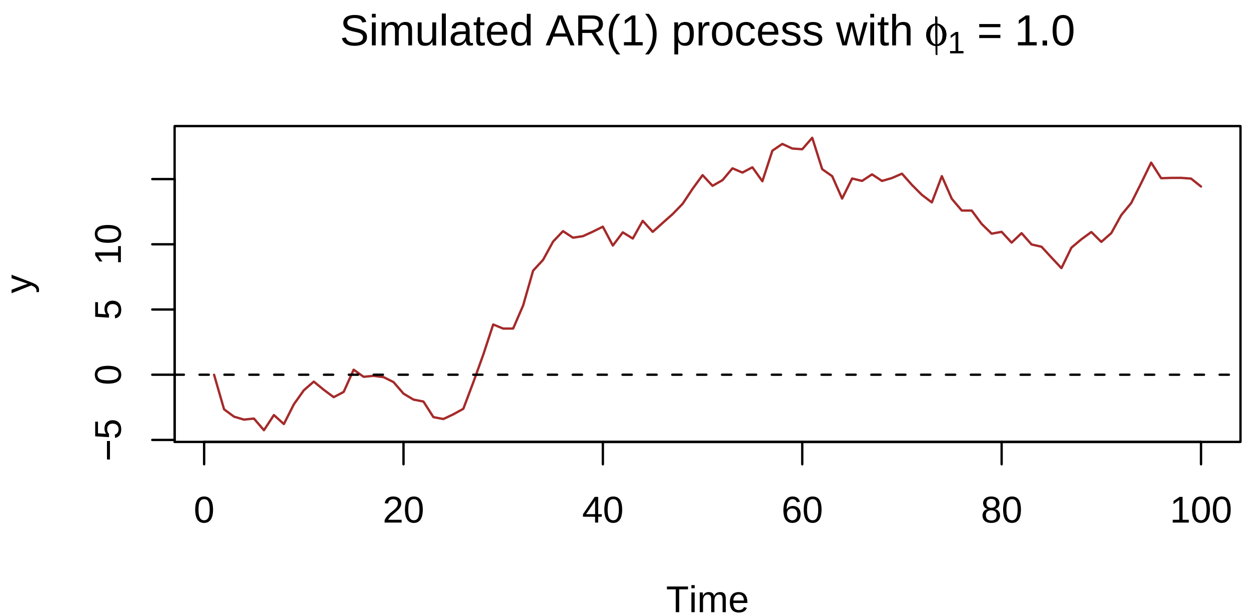
- Long-run forecasts are hard—don't tend towards any particular mean

- Spurious regression: Regressing one random walk on another
  tends to find large correlations even when the series are really independent

Spurious regression a major problem.

Identified by Granger & Newbold 1974;
called into question a vast amount of past (and future) econometric work.


Simulated AR(1) process with ϕ_1 = 1.0
y
0
5
10
0
20
40
60
80
100
Time


Simulated AR(1) process with ϕ_1 = 1.0
y
−5
0
5
10
0
20
40
60
80
100
Time

Spurious correlation

These were the first two random walks I generated. They are correlated ≈ 0.6 over the first 100 observations, and ≈ 0.3 over the first 1000

Few social science relationships are this strong. . .
and these are totally unrelated variables!

Many time series look like random walks over the period we can observe them

Grave danger of spurious "significant" findings

Techniques to mitigate this problem later in the course

Techniques to analyze stationary time series next time

Autoregression with p lags: AR(p) process

An autoregressive process may have many lags, e.g.,

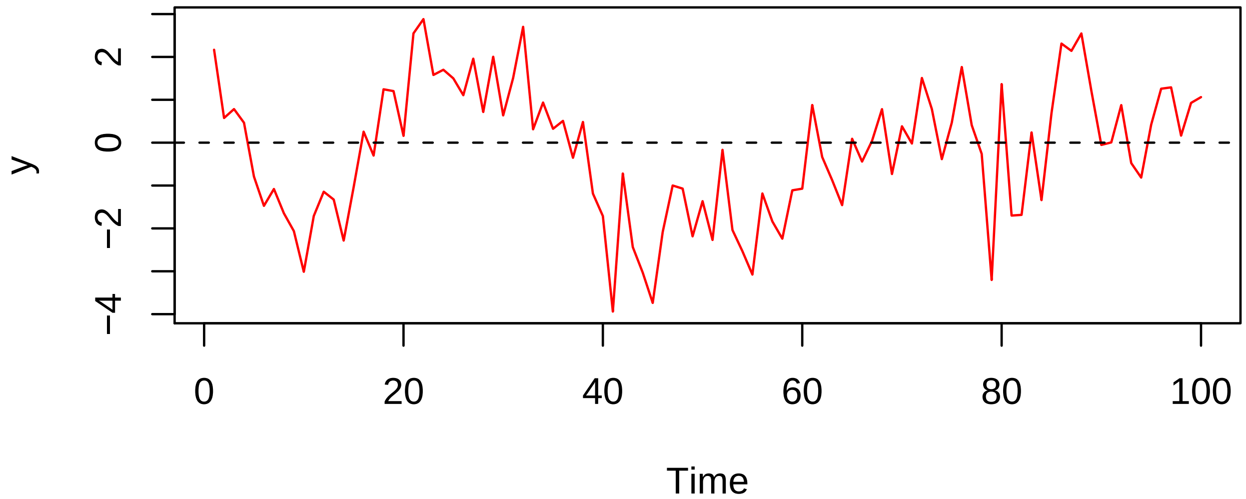
$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + \ldots + y_{t-p}\phi_p + \varepsilon_t$$

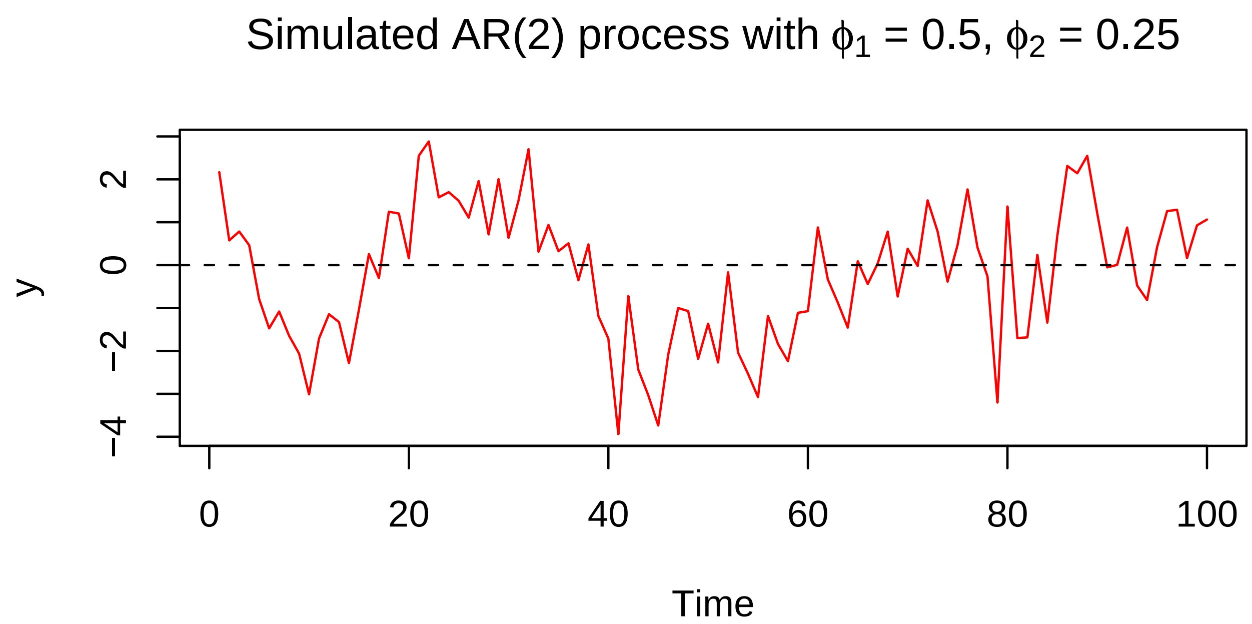
This general case is known as AR(p).

The distant past has a *direct* effect on the present

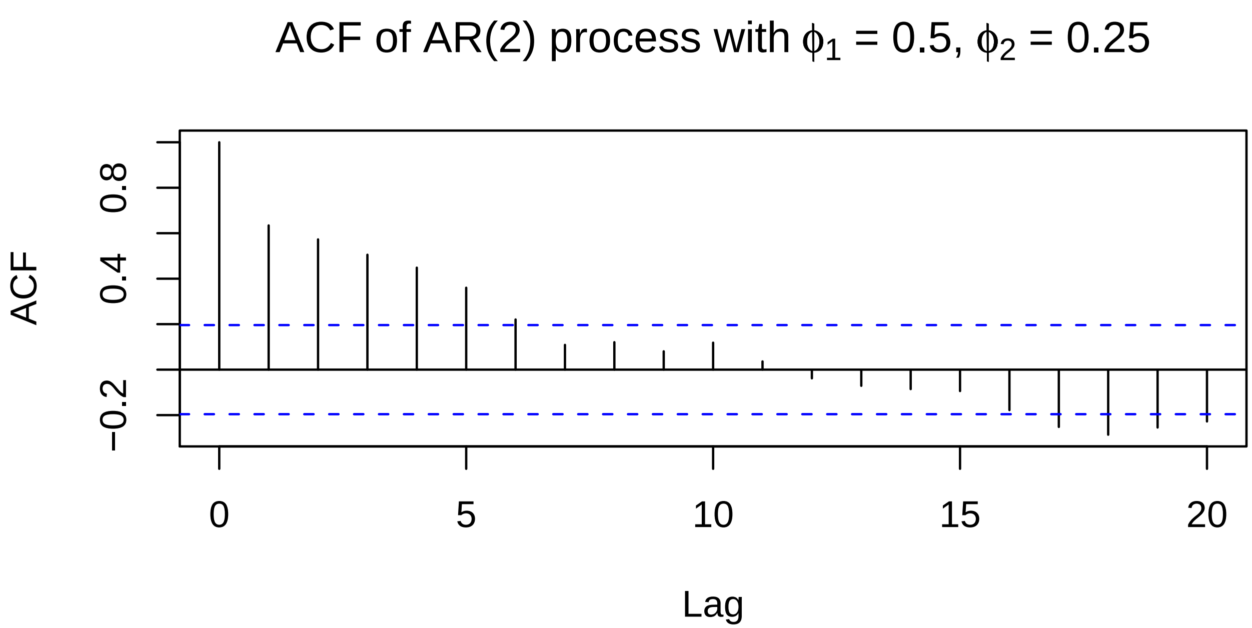
Distant past should show up in PACF, not just ACF

Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$

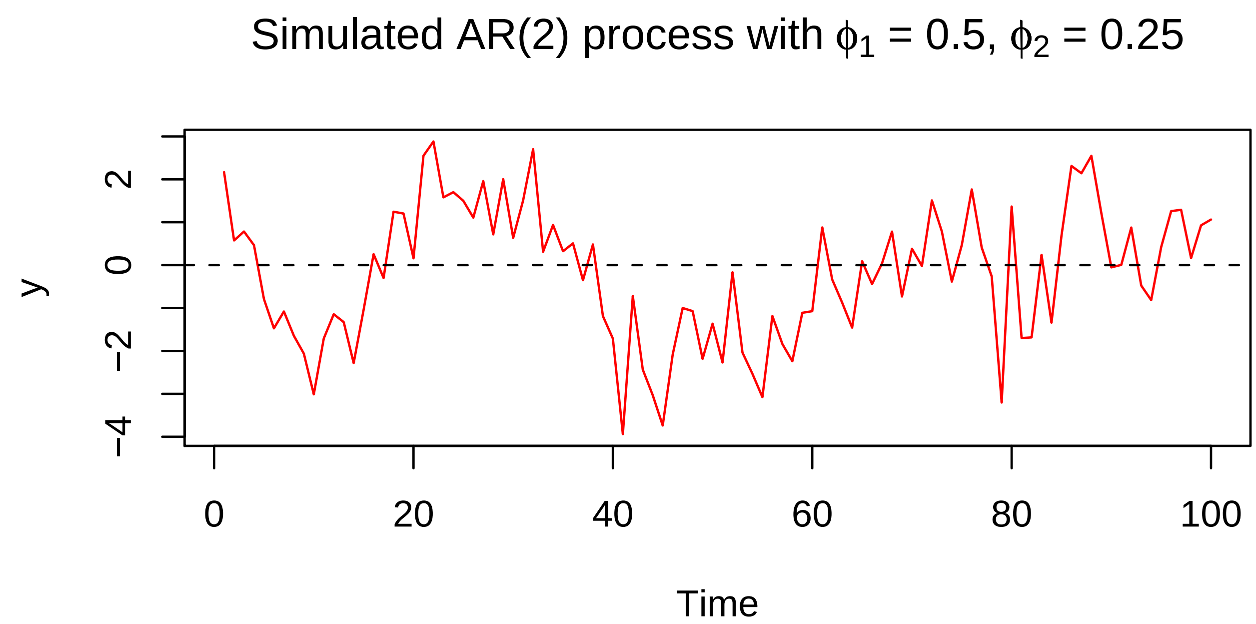


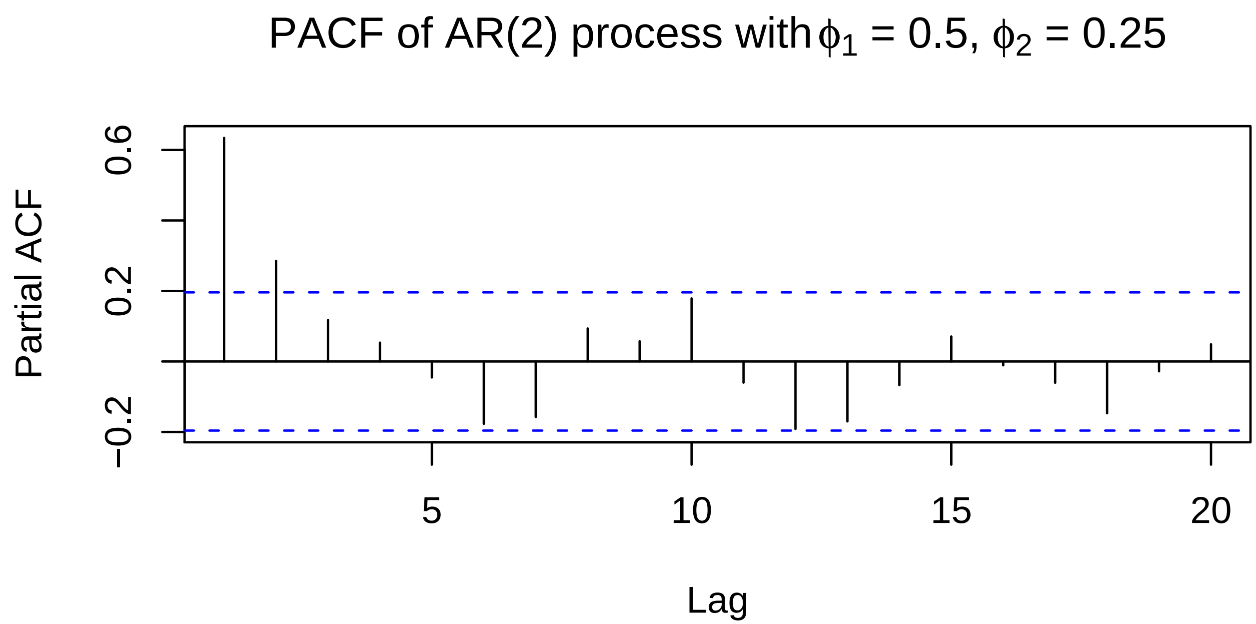


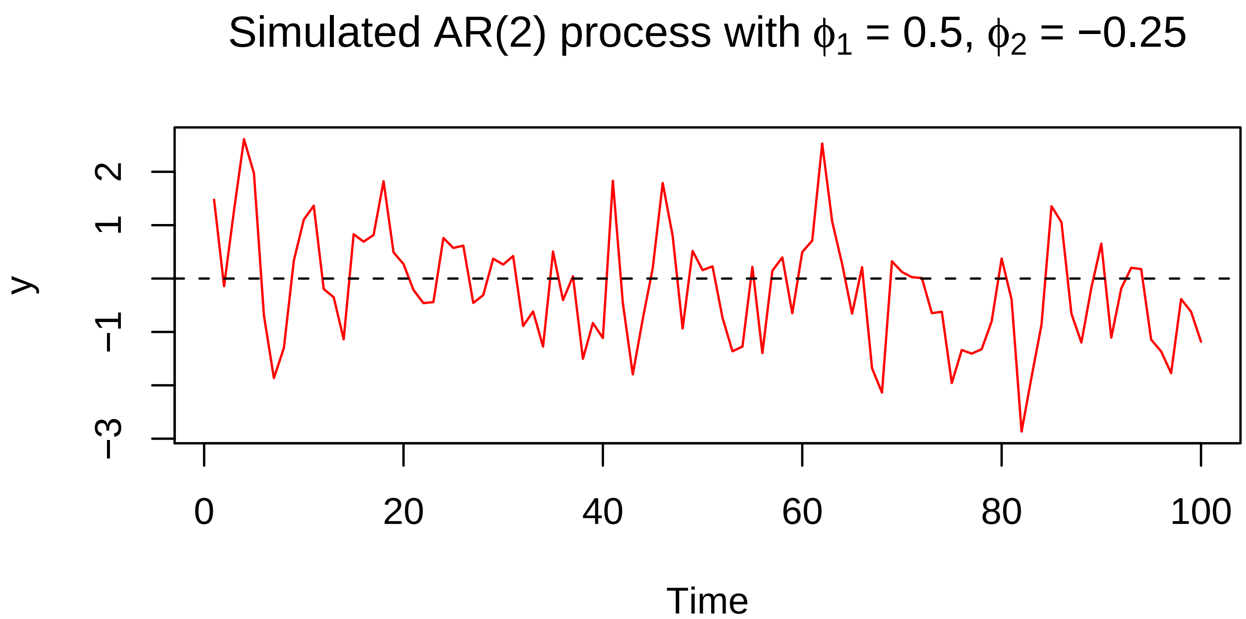
Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$
y
−4
−2
0
2
0
20
40
60
80
100
Time



ACF of AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$
ACF
−0.2
0.4
0.8
0
5
10
15
20
Lag

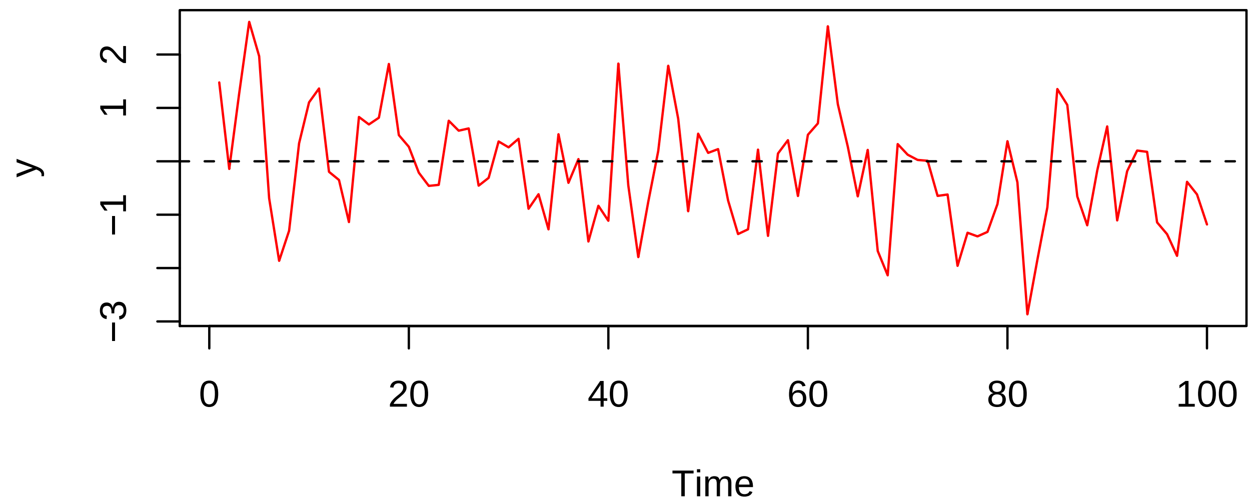

Simulated AR(2) process with ϕ1 = 0.5, ϕ2 = 0.25
y
−4
−2
0
2
0
20
40
60
80
100
Time


PACF of AR(2) process with ϕ1 = 0.5, ϕ2 = 0.25
Partial ACF
−0.2
0.2
0.6
5
10
15
20
Lag

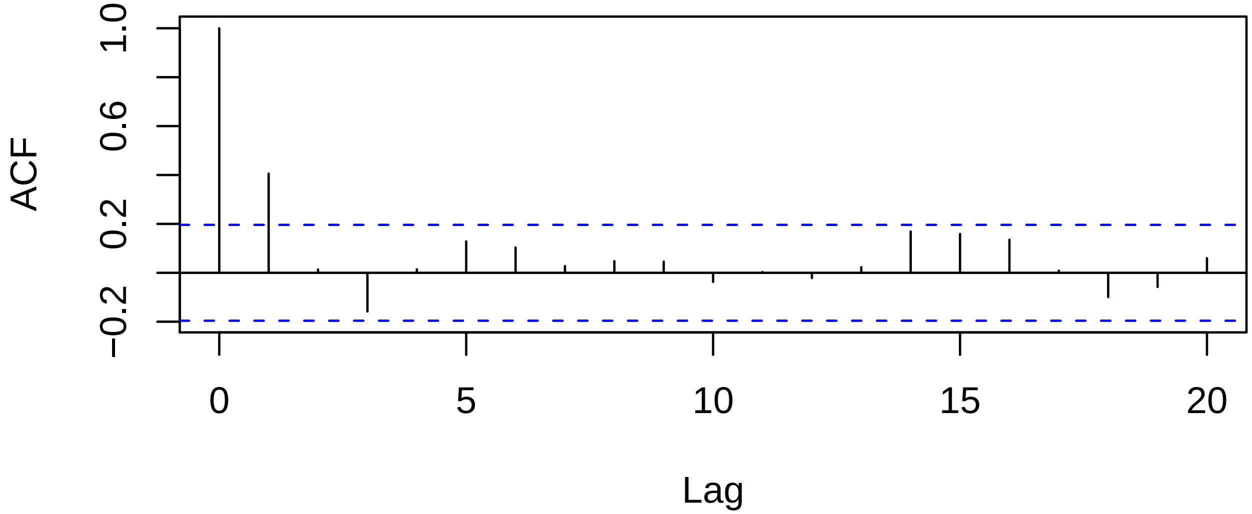


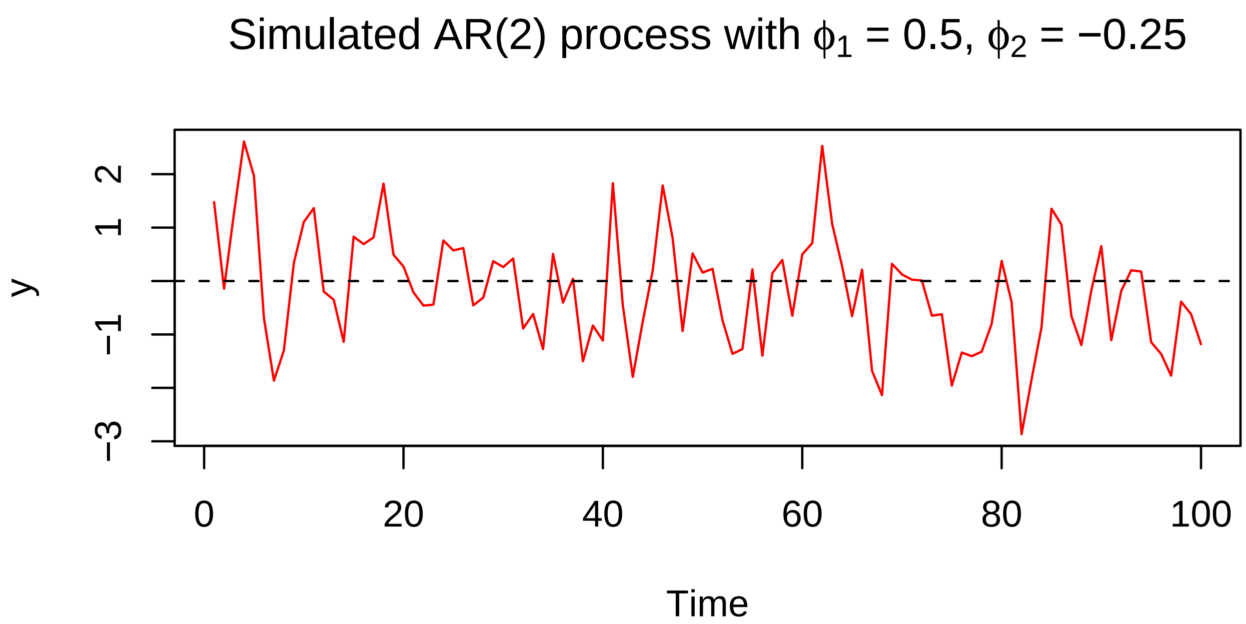
Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = -0.25$
y
−3
−1
1
2
0
20
40
60
80
100
Time

Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = -0.25$

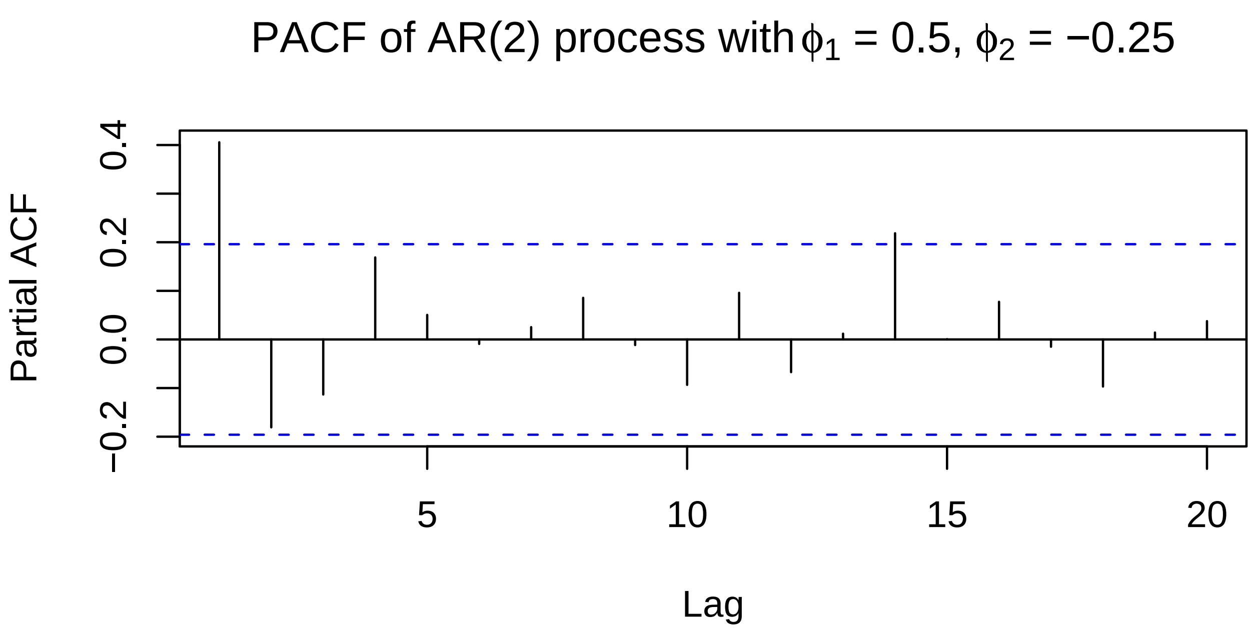


ACF of AR(2) process with $\phi_1 = 0.5$, $\phi_2 = -0.25$





Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = -0.25$
y
−3
−1
1
2
0
20
40
60
80
100
Time



PACF of AR(2) process with $\phi_1 = 0.5$, $\phi_2 = -0.25$
Partial ACF
−0.2
0.0
0.2
0.4
5
10
15
20
Lag

Alternative representations of lag structure

The most obvious way to write out an AR(p) process is the equation used above:

$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + \ldots + y_{t-p}\phi_p + \varepsilon_t$$

But there are alternatives using the *lag operator*, L
(sometimes called the backshift operator)

Define $\mathrm{L}y_t = y_{t-1}$.

L is an operation that shifts y_t back one period

Repeated applications of L create more distant lags: $\mathrm{L}^k y_{t-1} = y_{t-k}$

Somewhat unusual notation: makes an operation look like a variable

Will turn out to be handy

Example of lag operator: random walk

Recall the equation for a random walk (note that $\phi_1 = 1$):

$$\begin{aligned}
y_t &= \mathrm{L}y_t + \varepsilon_t \\
y_t - \mathrm{L}y_t &= \varepsilon_t \\
(1 - \mathrm{L})y_t &= \varepsilon_t \\
y_t &= \frac{\varepsilon_t}{1 - \mathrm{L}} \\
y_t &= (1 + \mathrm{L} + \mathrm{L}^2 + \mathrm{L}^3 + \ldots + \mathrm{L}^\infty)\varepsilon_t \\
y_t &= \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \ldots + \varepsilon_{t-\infty}
\end{aligned}$$

So the Lag operator shows us that the random walk consists of all past disturbances with equal weight.

Unit roots

Now suppose we have an AR(2) process:

$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + \varepsilon_t$$

Using the lag operator, this is

$$y_t = \phi_1 \mathrm{L} y_t + \phi_2 \mathrm{L}^2 y_t + \varepsilon_t$$

Rearranging, we find

$$\left(1 - \phi_1 \mathrm{L} + \phi_2 \mathrm{L}^2\right) y_t = \varepsilon_t$$

Isolate the polynomial:

$$1 - \phi_1 \mathrm{L} + \phi_2 \mathrm{L}^2$$

Unit roots

$$1 - \phi_1 \mathrm{L} + \phi_2 \mathrm{L}^2 = 0$$

Setting this equal to 0, and solving for the roots of L yields 2 numbers

If the absolute value of both roots > 1, then y_t is stationary

If either root $= 1$ or -1, or is a *unit root*, then y_t is non-stationary

For AR(2) this is conceptually easy; if the sum of ϕ_1, ϕ_2 is 1 or -1, you have a non-stationary series

Unit roots

This generalizes to AR(p):

$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + y_{t-3}\phi_3 + \ldots + y_{t-p}\phi_p\varepsilon_t$$

Using the lag operator, this is

$$y_t = \phi_1 \mathrm{L} y_t + \phi_2 \mathrm{L}^2 y_t + \phi_3 \mathrm{L}^3 y_t + \ldots + \phi_p \mathrm{L}^p y_t + \varepsilon_t$$

Rearranging, we find

$$\left(1 - \phi_1 \mathrm{L} + \phi_2 \mathrm{L}^2 + \phi_3 \mathrm{L}^3 + \cdots + \phi_p \mathrm{L}^p\right) y_t = \varepsilon_t$$

Unit roots

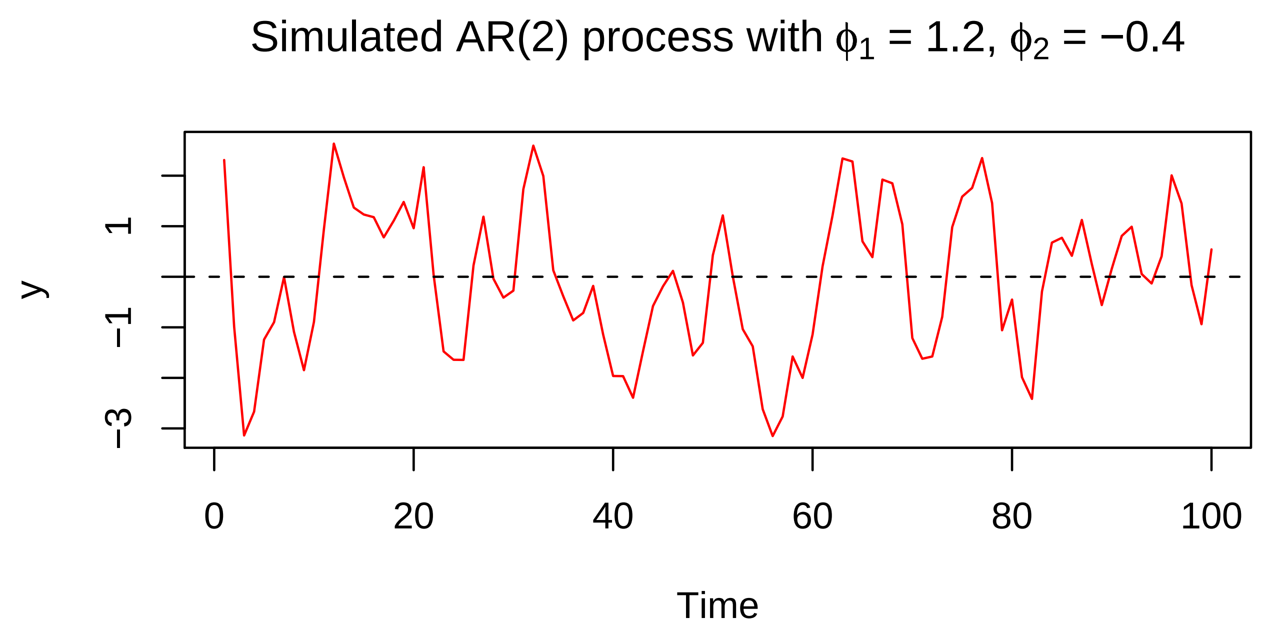
$$\left(1 - \phi_1 \mathrm{L} + \phi_2 \mathrm{L}^2 + \phi_3 \mathrm{L}^3 + \cdots + \phi_p \mathrm{L}^p\right) y_t = \varepsilon_t$$

You can solve for the roots for a given polynomial using `polyroot` in `R`

Should worry if any (empirical) roots “close” to 1.

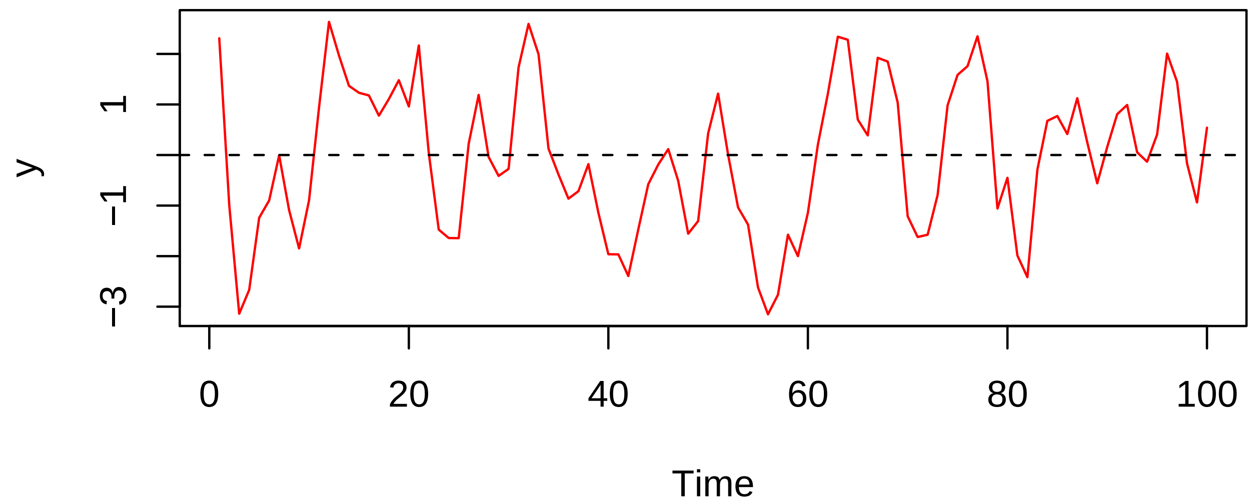
More on finding unit roots next time

Very hard to do well, unless t is very large (not likely in political science)

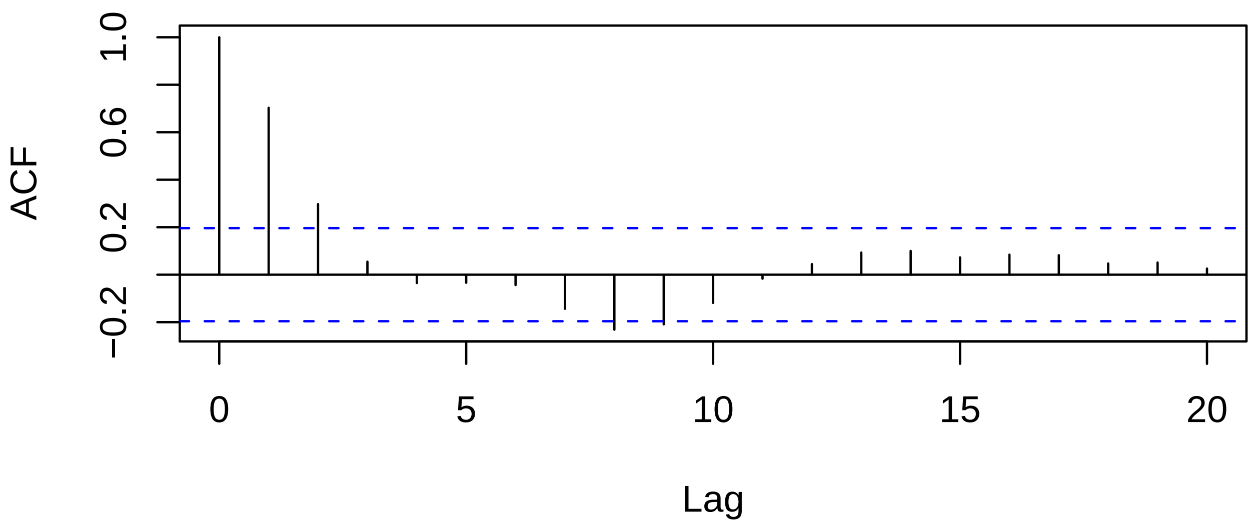


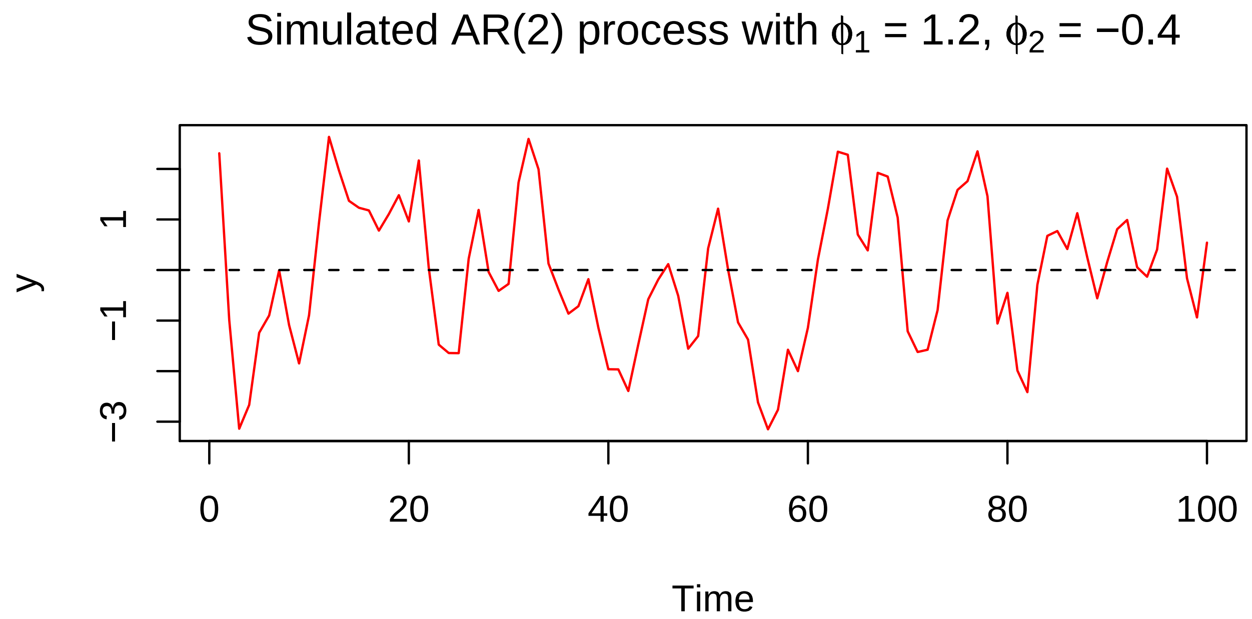
Simulated AR(2) process with ϕ1 = 1.2, ϕ2 = −0.4
y
1
−1
−3
0
20
40
60
80
100
Time

Simulated AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.4$

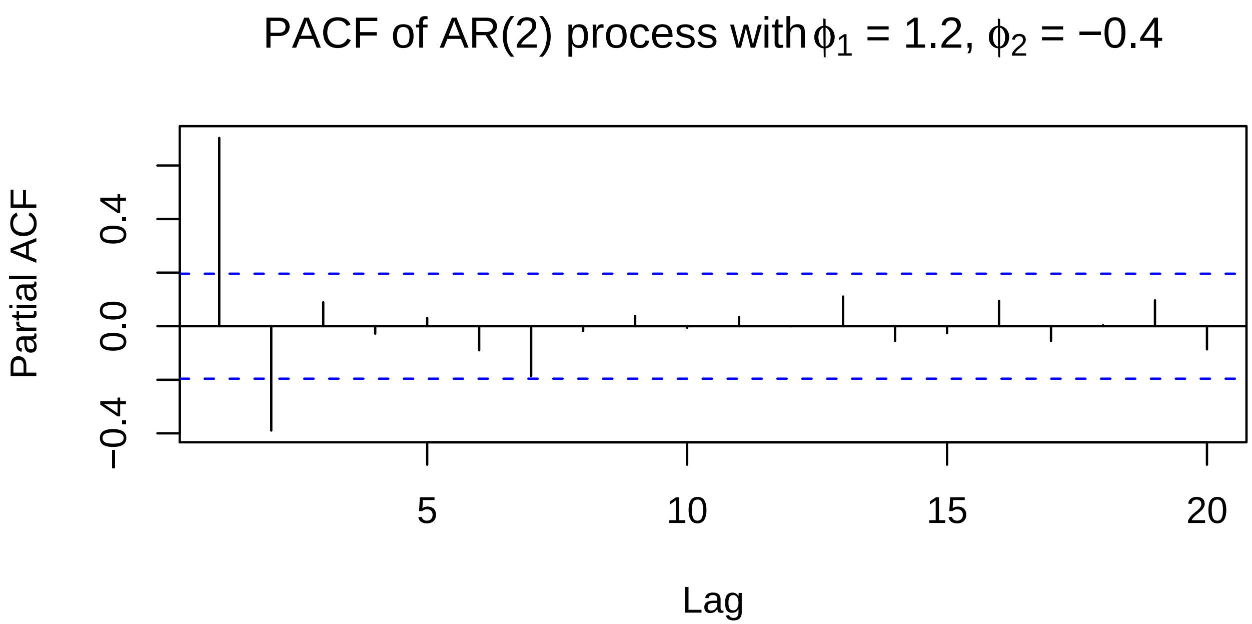


ACF of AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.4$

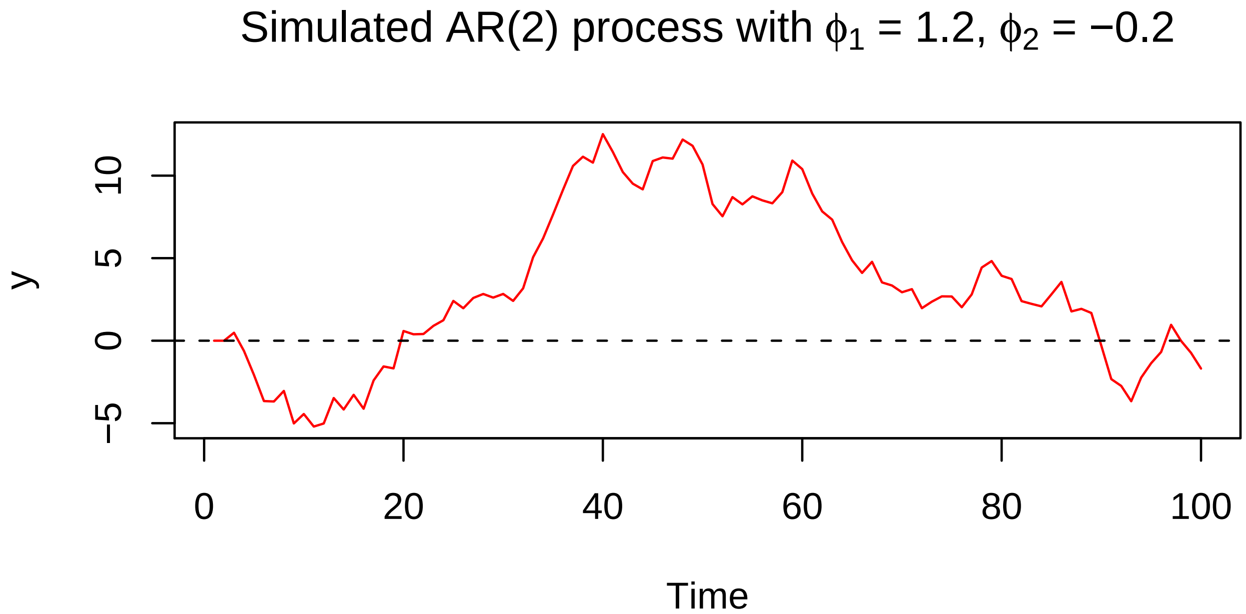




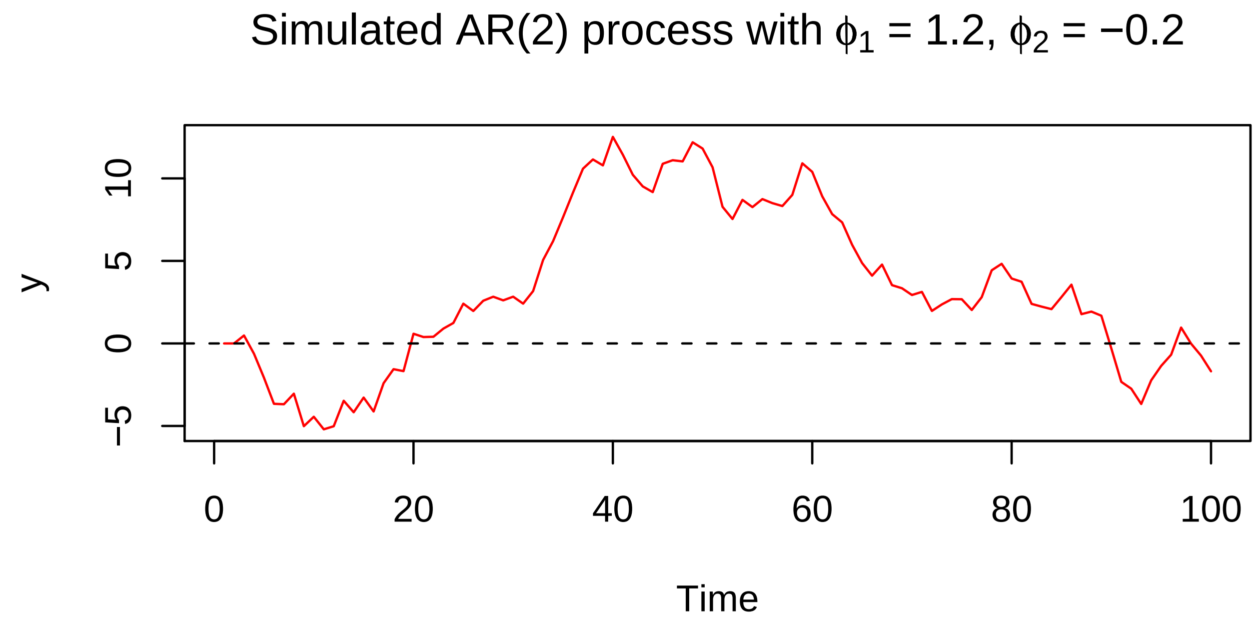
Simulated AR(2) process with ϕ1 = 1.2, ϕ2 = −0.4
y
−3
−1
1
0
20
40
60
80
100
Time



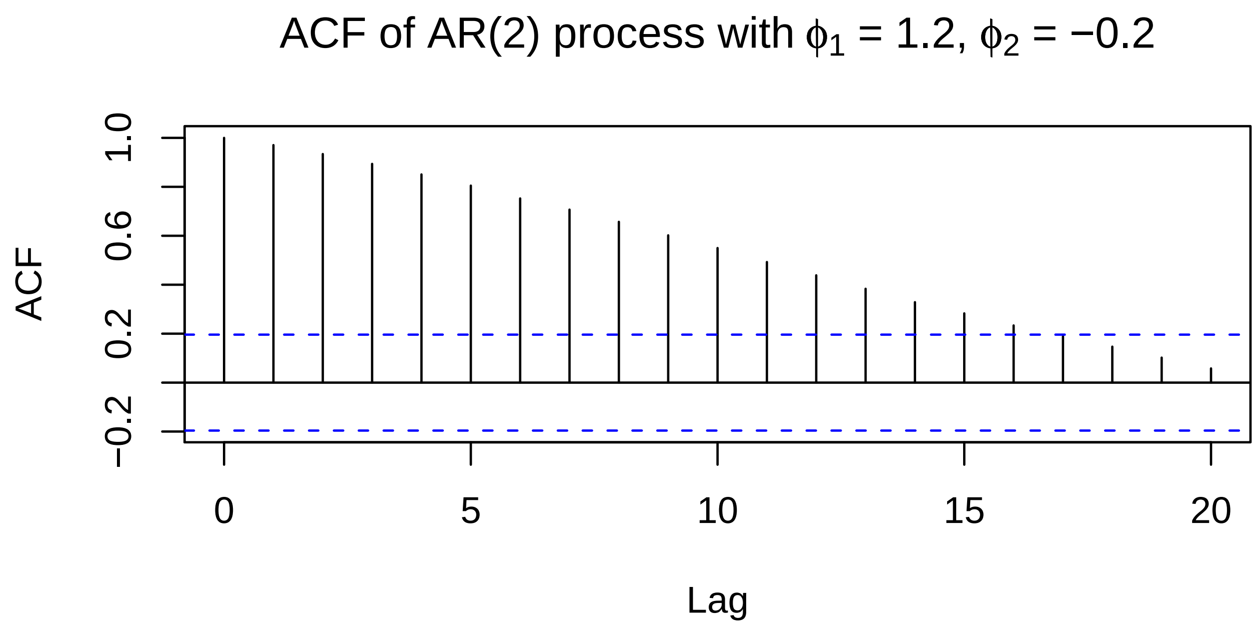
PACF of AR(2) process with ϕ1 = 1.2, ϕ2 = −0.4
Partial ACF
−0.4
0.0
0.4
5
10
15
20
Lag



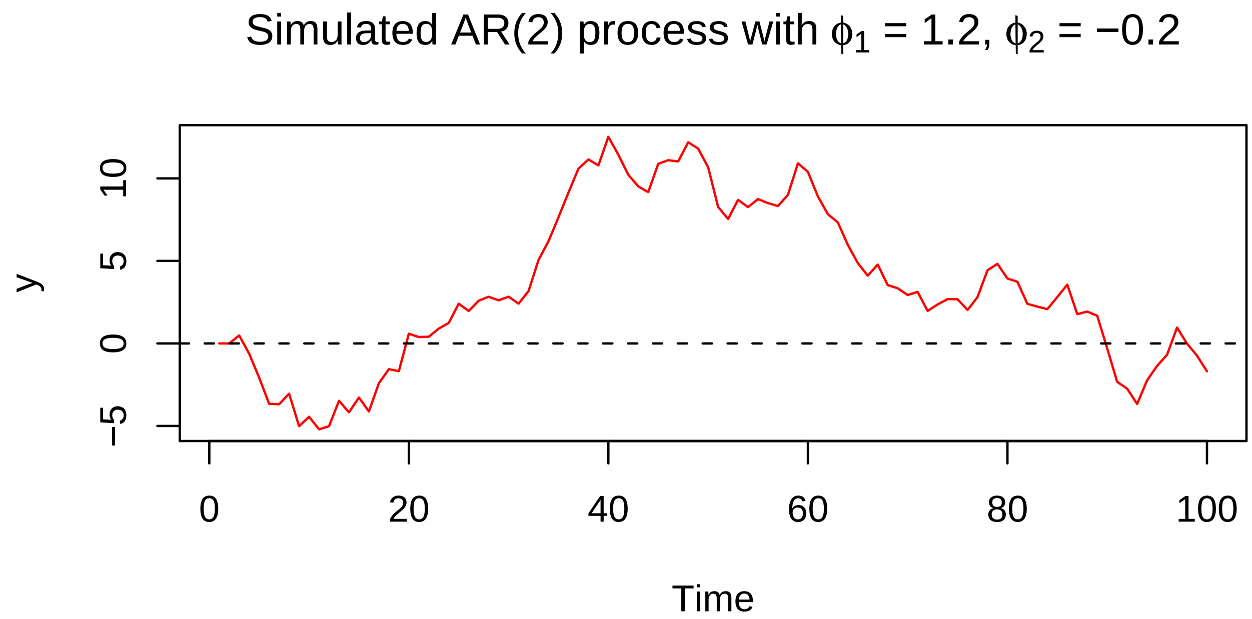
Does this picture remind you of another time series process?



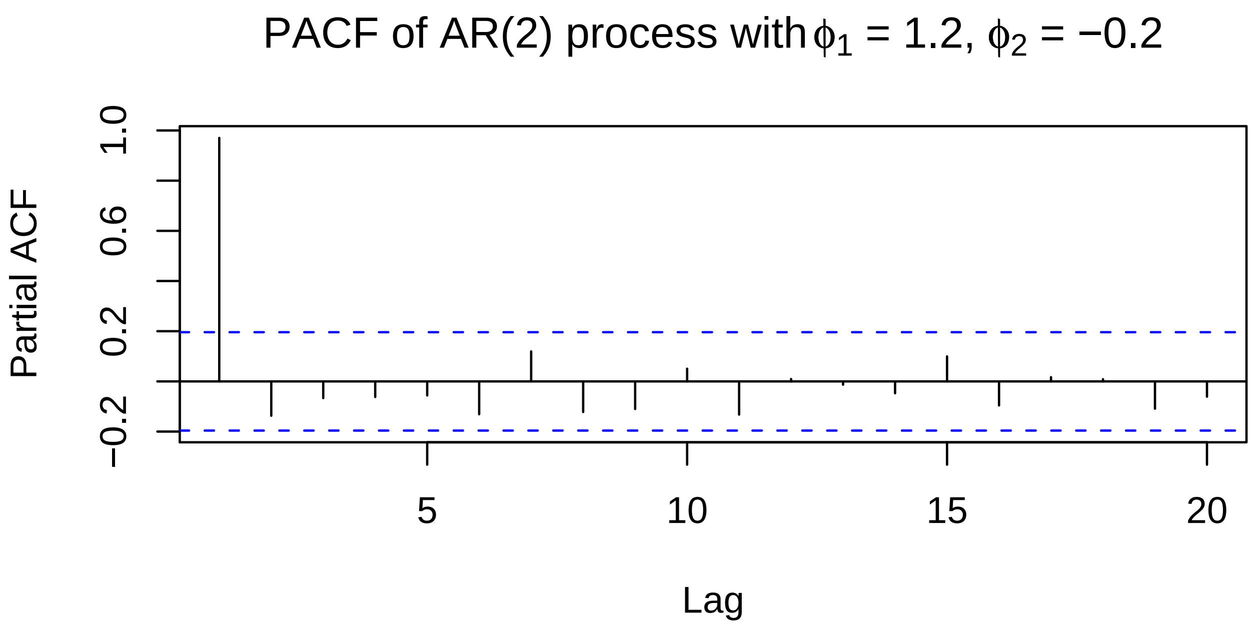
Simulated AR(2) process with ϕ1 = 1.2, ϕ2 = −0.2
y
10
5
0
−5
0
20
40
60
80
100
Time



ACF of AR(2) process with ϕ1 = 1.2, ϕ2 = −0.2
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag



Simulated AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.2$
y
−5
0
5
10
0
20
40
60
80
100
Time



PACF of AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.2$
Partial ACF
−0.2
0.2
0.6
1.0
5
10
15
20
Lag

Past shocks

Suppose we think y_t responds to past shocks ε_{t-q} only, not past values of y_t

Many financial examples (day-trading)

A political example: Voting after a major roll call in Congress

A model that response to last period's disturbance:

$$y_t = \varepsilon_{t-1}\rho_1 + \varepsilon_t$$

This is known as a *moving average* process of order 1

So called because the stochastic component is a weighted average of the current and previous error

MA(1) Processes

Notice something interesting when we calculate the autocorrelations for lags 1 and 2.

Remember that because ε_t is white noise, $\mathrm{cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ for all $k \geq 1$

$$
\begin{aligned}
\mathrm{E}(y_t - \mu)(y_{t-1} - \mu) &= \mathrm{E}(\varepsilon_t + \rho_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \rho_1 \varepsilon_{t-2}) \\
&= \mathrm{E}(\varepsilon_t \varepsilon_{t-1} + \rho_1 \varepsilon_{t-1}^2 + \rho_1 \varepsilon_t \varepsilon_{t-2} + \rho_1^2 \varepsilon_{t-1} \varepsilon_{t-2}) \\
&= 0 + \rho \sigma^2
\end{aligned}
$$

MA(1) Processes

Notice something interesting when we calculate the autocorrelations for lags 1 and 2.

Remember that because ε_t is white noise, $\mathrm{cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ for all $k \geq 1$

$$
\begin{aligned}
\mathrm{E}(y_t - \mu)(y_{t-1} - \mu) &= \mathrm{E}(\varepsilon_t + \rho_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \rho_1 \varepsilon_{t-2}) \\
&= \mathrm{E}(\varepsilon_t \varepsilon_{t-1} + \rho_1 \varepsilon_{t-1}^2 + \rho_1 \varepsilon_t \varepsilon_{t-2} + \rho_1^2 \varepsilon_{t-1} \varepsilon_{t-2}) \\
&= 0 + \rho \sigma^2 + 0
\end{aligned}
$$

MA(1) Processes

Notice something interesting when we calculate the autocorrelations for lags 1 and 2.

Remember that because ε_t is white noise, $\mathrm{cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ for all $k \geq 1$

$$
\begin{aligned}
\mathrm{E}(y_t - \mu)(y_{t-1} - \mu) &= \mathrm{E}(\varepsilon_t + \rho_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \rho_1 \varepsilon_{t-2}) \\
&= \mathrm{E}(\varepsilon_t \varepsilon_{t-1} + \rho_1 \varepsilon_{t-1}^2 + \rho_1 \varepsilon_t \varepsilon_{t-2} + \rho_1^2 \varepsilon_{t-1} \varepsilon_{t-2}) \\
&= 0 + \rho \sigma^2 + 0 + 0
\end{aligned}
$$

MA(1) Processes

Notice something interesting when we calculate the autocorrelations for lags 1 and 2.

Remember that because ε_t is white noise, $\mathrm{cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ for all $k \geq 1$

$$\begin{aligned}
\mathrm{E}(y_t - \mu)(y_{t-1} - \mu) &= \mathrm{E}(\varepsilon_t + \rho_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \rho_1 \varepsilon_{t-2}) \\
&= \mathrm{E}(\varepsilon_t \varepsilon_{t-1} + \rho_1 \varepsilon_{t-1}^2 + \rho_1 \varepsilon_t \varepsilon_{t-2} + \rho_1^2 \varepsilon_{t-1} \varepsilon_{t-2}) \\
&= 0 + \rho \sigma^2 + 0 + 0
\end{aligned}$$

In MA(1), y_t and y_{t+1} are correlated

MA(1) Processes

However, for any larger lags . . .

$$\begin{aligned}
\mathrm{E}(y_t-\mu)(y_{t-2}-\mu) &= \mathrm{E}(\varepsilon_t+\rho_1\varepsilon_{t-1})(\varepsilon_{t-2}+\rho_1\varepsilon_{t-3}) \\
&= \mathrm{E}(\varepsilon_t\varepsilon_{t-2}+\rho_1\varepsilon_{t-1}\varepsilon_{t-2}+\rho_1\varepsilon_t\varepsilon_{t-3}+\rho_1^2\varepsilon_{t-1}\varepsilon_{t-3}) \\
&= 0+0
\end{aligned}$$

MA(1) Processes

However, for any larger lags . . .

$$
\begin{aligned}
\mathrm{E}(y_t - \mu)(y_{t-2} - \mu) &= \mathrm{E}(\varepsilon_t + \rho_1 \varepsilon_{t-1})(\varepsilon_{t-2} + \rho_1 \varepsilon_{t-3}) \\
&= \mathrm{E}(\varepsilon_t \varepsilon_{t-2} + \rho_1 \varepsilon_{t-1} \varepsilon_{t-2} + \rho_1 \varepsilon_t \varepsilon_{t-3} + \rho_1^2 \varepsilon_{t-1} \varepsilon_{t-3}) \\
&= 0 + 0 + 0
\end{aligned}
$$

MA(1) Processes

However, for any larger lags . . .

$$\begin{aligned}
\mathrm{E}(y_t-\mu)(y_{t-2}-\mu) &= \mathrm{E}(\varepsilon_t+\rho_1\varepsilon_{t-1})(\varepsilon_{t-2}+\rho_1\varepsilon_{t-3}) \\
&= \mathrm{E}(\varepsilon_t\varepsilon_{t-2}+\rho_1\varepsilon_{t-1}\varepsilon_{t-2}+\rho_1\varepsilon_t\varepsilon_{t-3}+\rho_1^2\varepsilon_{t-1}\varepsilon_{t-3}) \\
&= 0+0+0+0
\end{aligned}$$

MA(1) Processes

However, for any larger lags . . .

$$\begin{aligned} \mathrm{E}(y_t - \mu)(y_{t-2} - \mu) &= \mathrm{E}(\varepsilon_t + \rho_1 \varepsilon_{t-1})(\varepsilon_{t-2} + \rho_1 \varepsilon_{t-3}) \\ &= \mathrm{E}(\varepsilon_t \varepsilon_{t-2} + \rho_1 \varepsilon_{t-1} \varepsilon_{t-2} + \rho_1 \varepsilon_t \varepsilon_{t-3} + \rho_1^2 \varepsilon_{t-1} \varepsilon_{t-3}) \\ &= 0 + 0 + 0 + 0 \end{aligned}$$

In MA(1), y_t and y_{t+k} are *uncorrelated* if $k > 1$

Shocks die out completely after 1 period. PACF will be 0 after 1 period.

So MA(1) processes are always stationary and ergodic.

The MA(q) process

We can add any number of moving average terms to our equation

$$y_t = \varepsilon_{t-1}\rho_1 + \varepsilon_{t-2}\rho_2 + \ldots + \varepsilon_{t-q}\rho_q + \varepsilon_t$$

This is known as a moving average process of order q, or an MA(q) process

Note that as in the AR(1), the effect of past shocks dies out after q periods

So MA(q) processes are always stationary and ergodic for finite q.

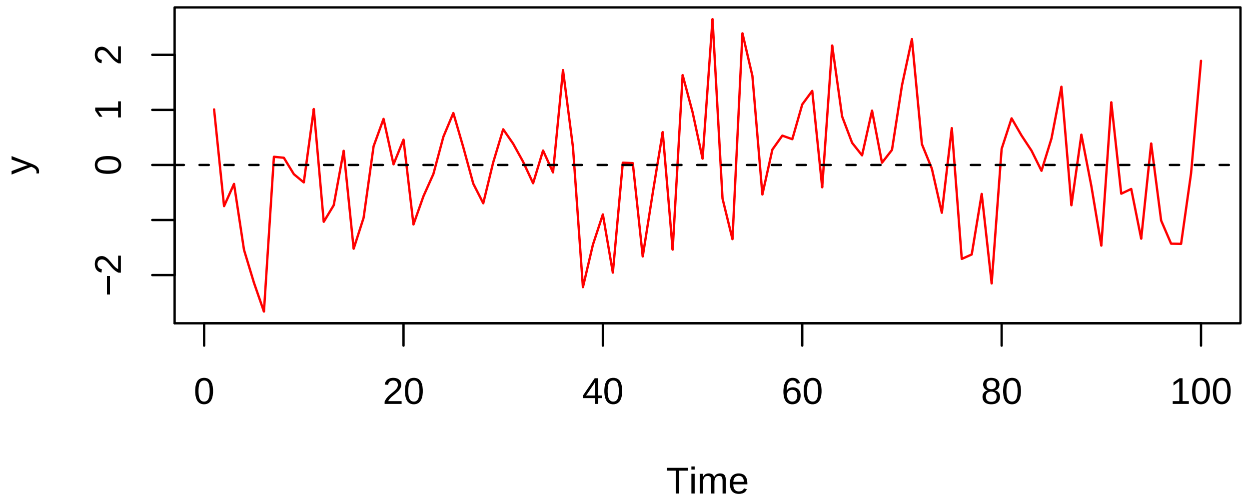
Contrast to the AR(1) or AR(p), in which shocks never (quite) die out, and non-stationarity can occur

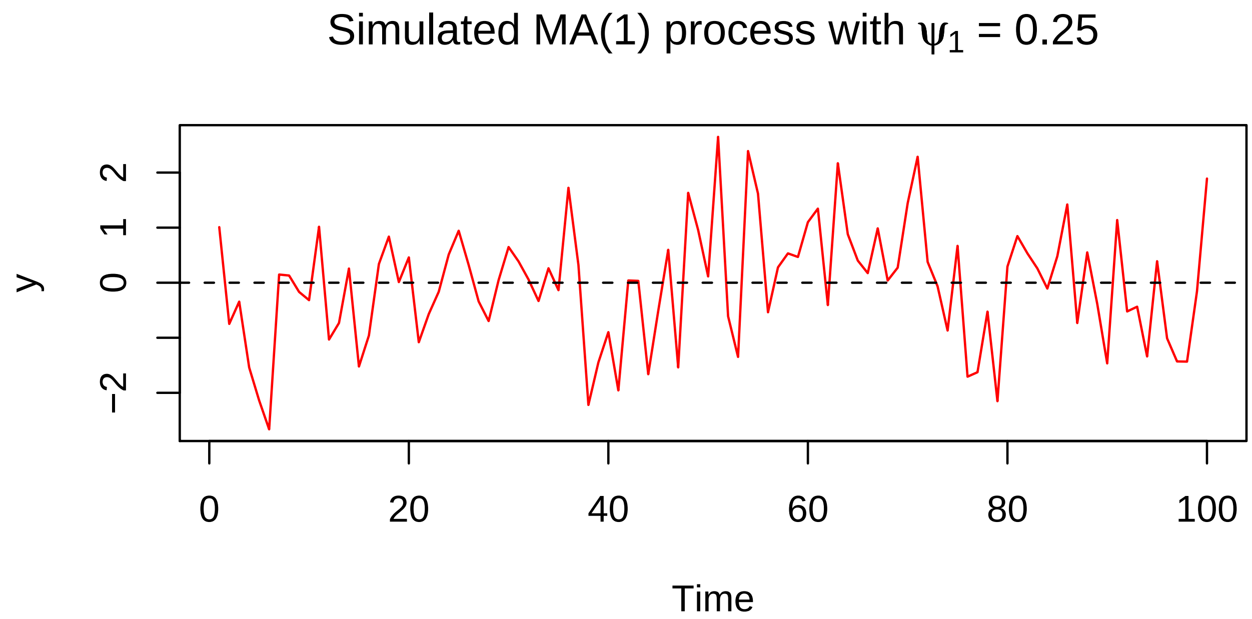
Simulating MA(q)

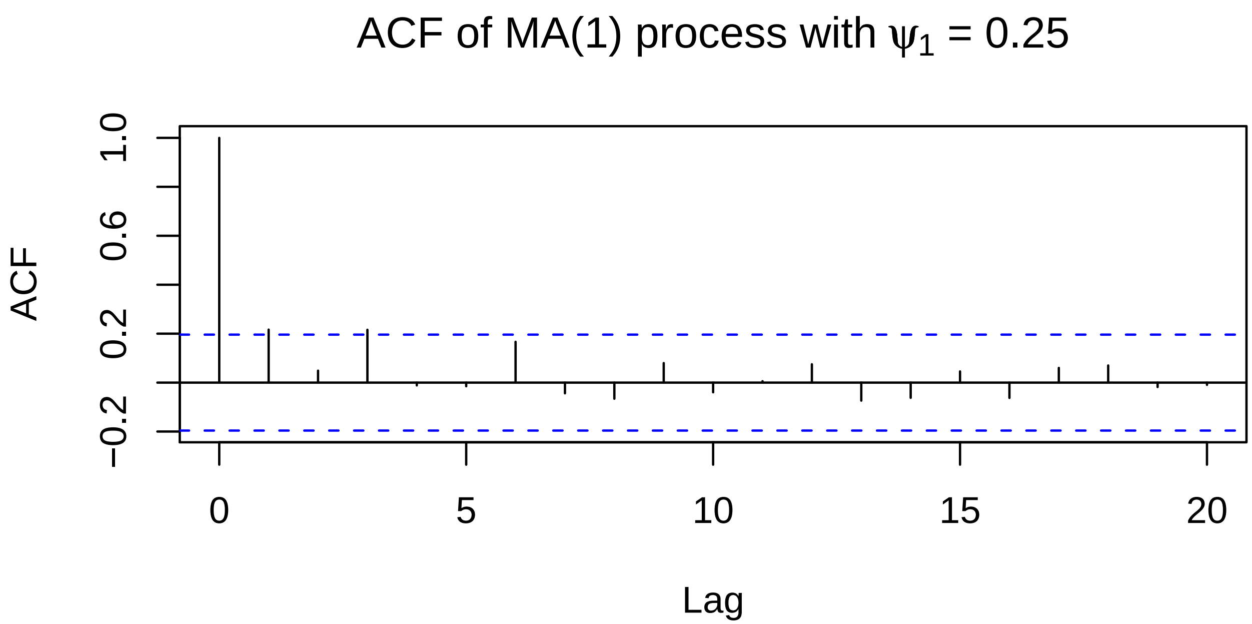
To simulate 1000 iterations of an MA(2) with $\rho_1 = 0.67$ and $\rho_2 = 0.5$:

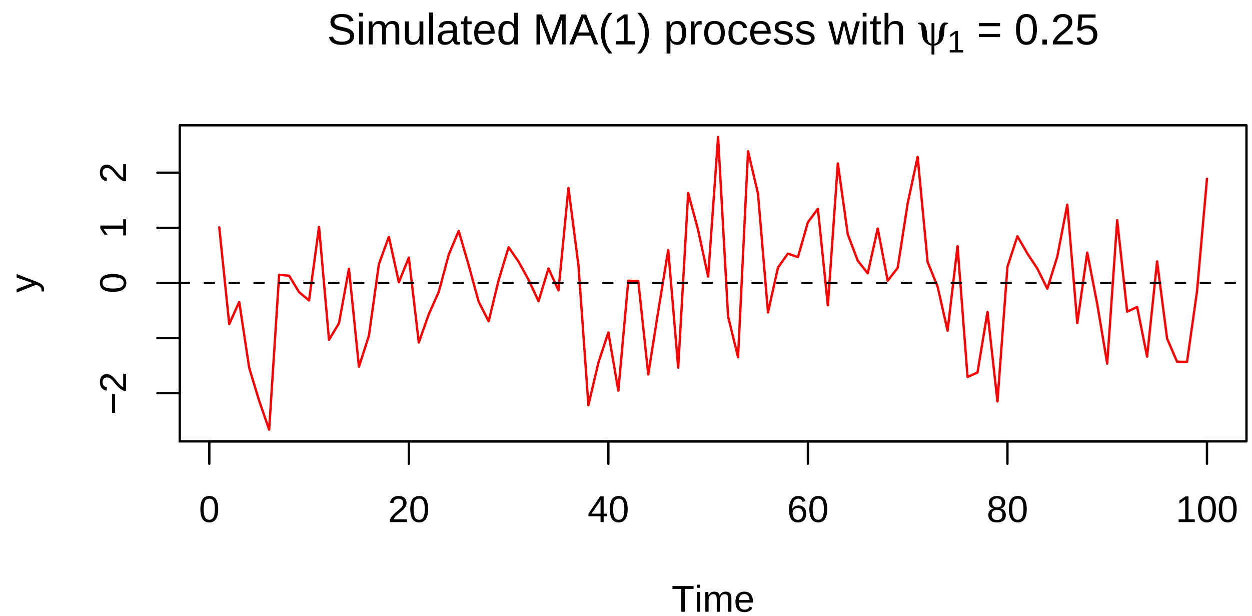
```
y <- arima.sim(list(order = c(0,0,2),
                    ar = NULL,
                    ma = c(0.67,0.5)),
               n=1000)
```

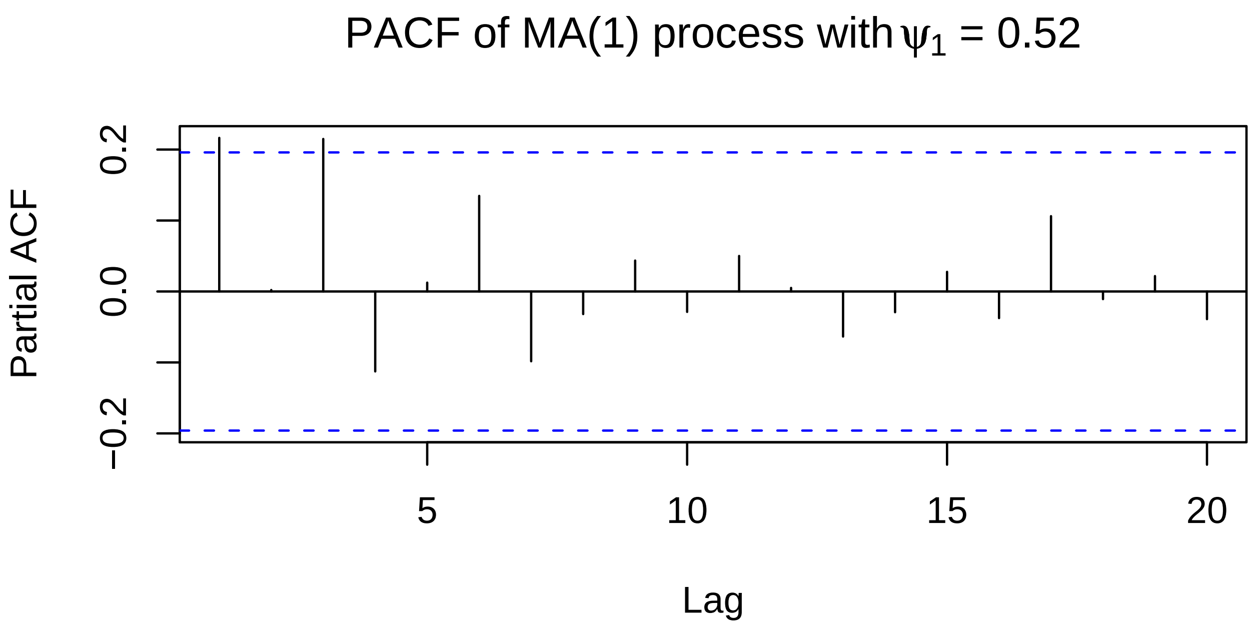
Simulated MA(1) process with $\psi_1 = 0.25$

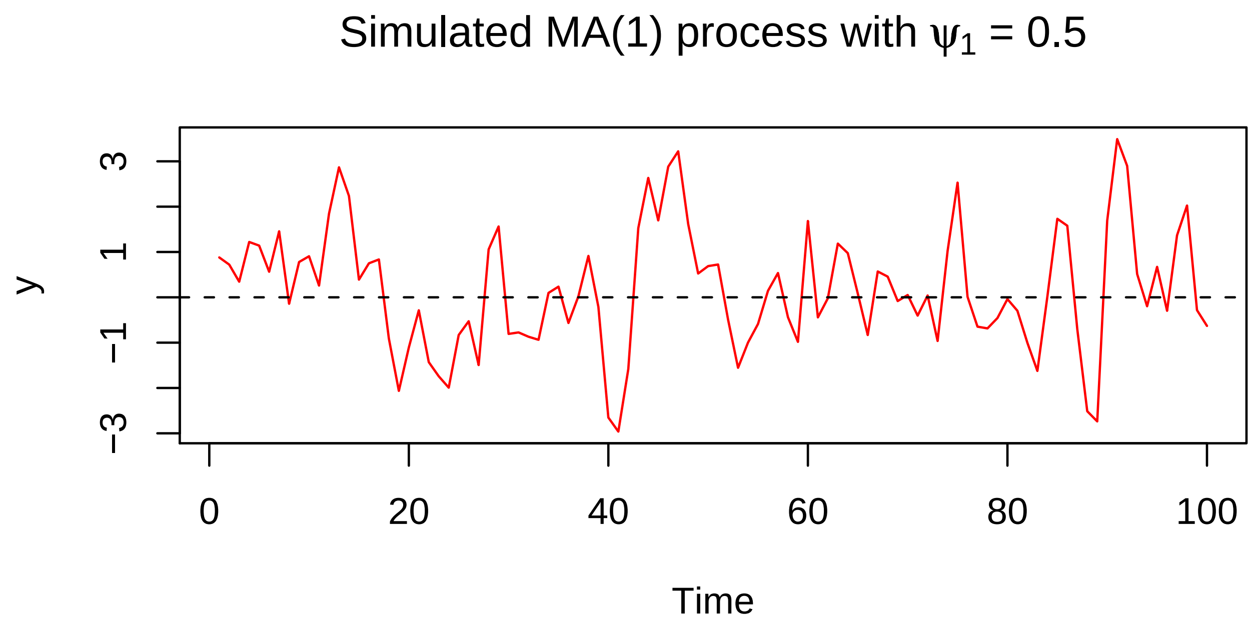



Simulated MA(1) process with ψ1 = 0.25
y
−2
0
1
2
0
20
40
60
80
100
Time


ACF of MA(1) process with ψ1 = 0.25
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
20
Lag

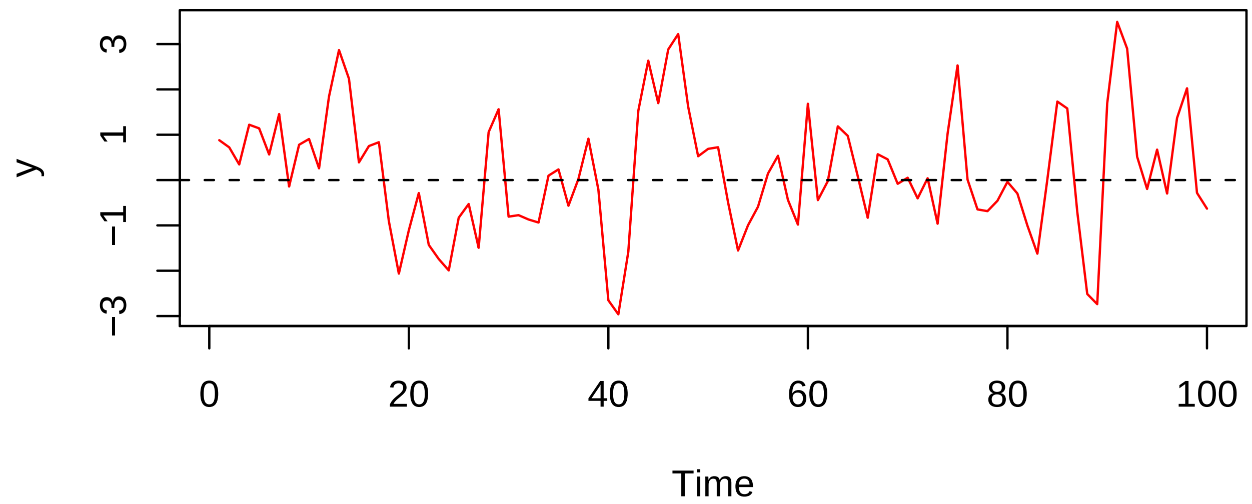

Simulated MA(1) process with ψ1 = 0.25
y
−2
0
1
2
0
20
40
60
80
100
Time


PACF of MA(1) process with ψ1 = 0.52
Partial ACF
−0.2
0.0
0.2
5
10
15
20
Lag

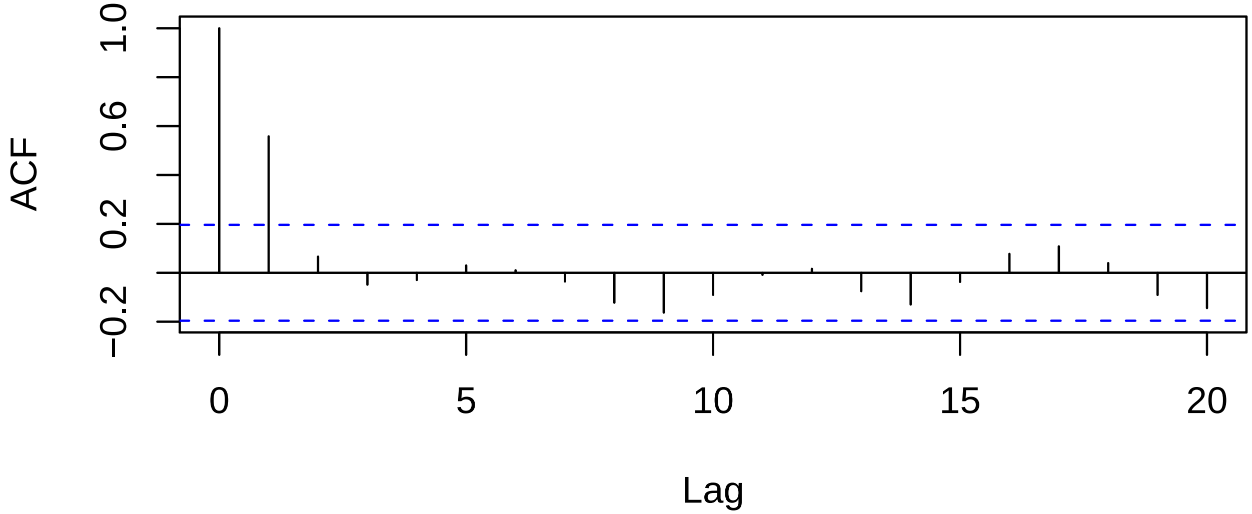


Simulated MA(1) process with ψ1 = 0.5
y
3
1
−1
−3
0
20
40
60
80
100
Time

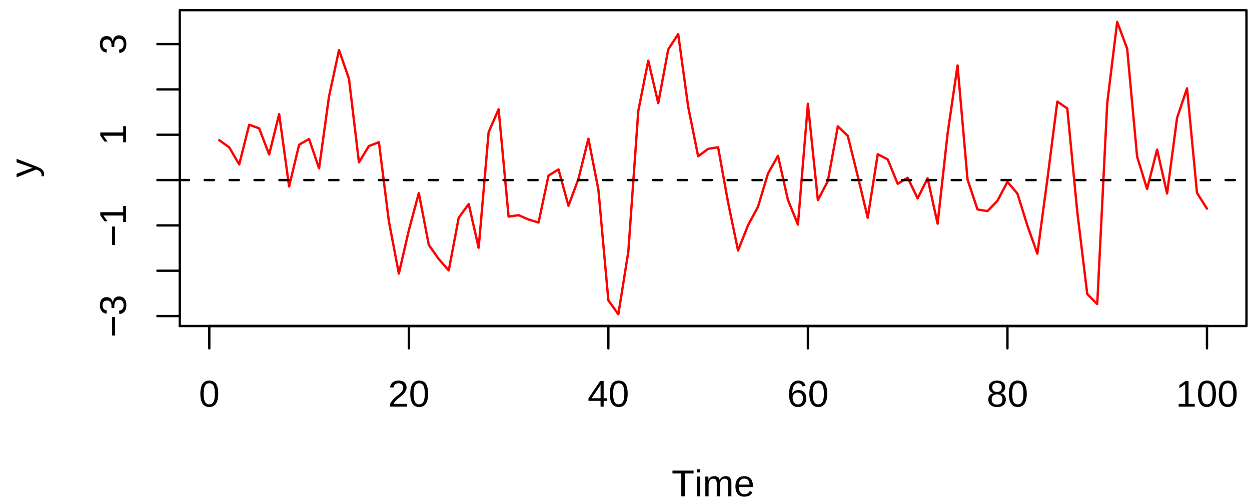
Simulated MA(1) process with $\psi_1 = 0.5$



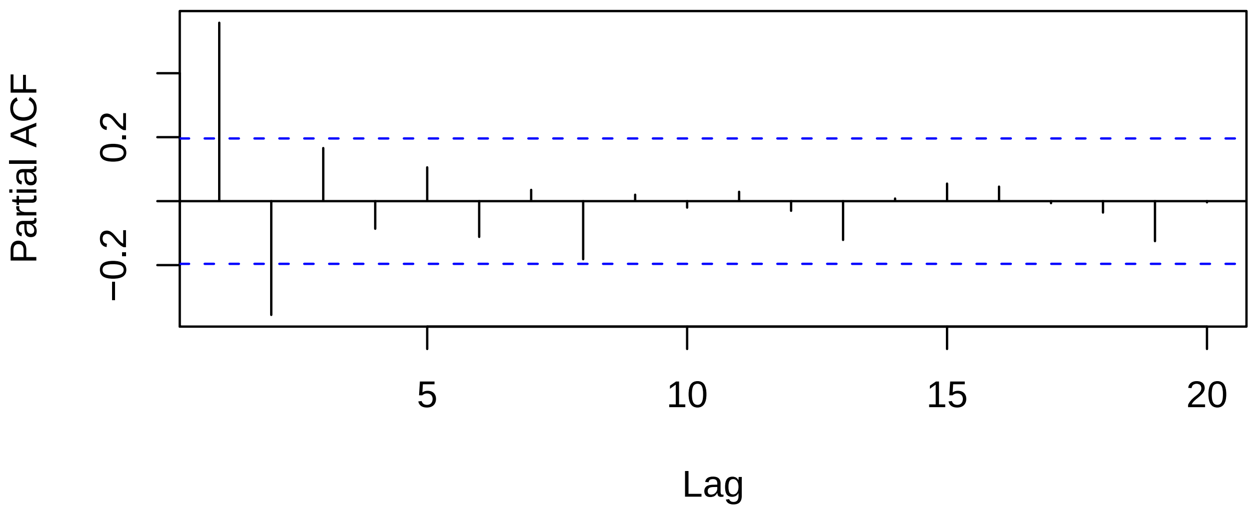
ACF of MA(1) process with $\psi_1 = 0.5$

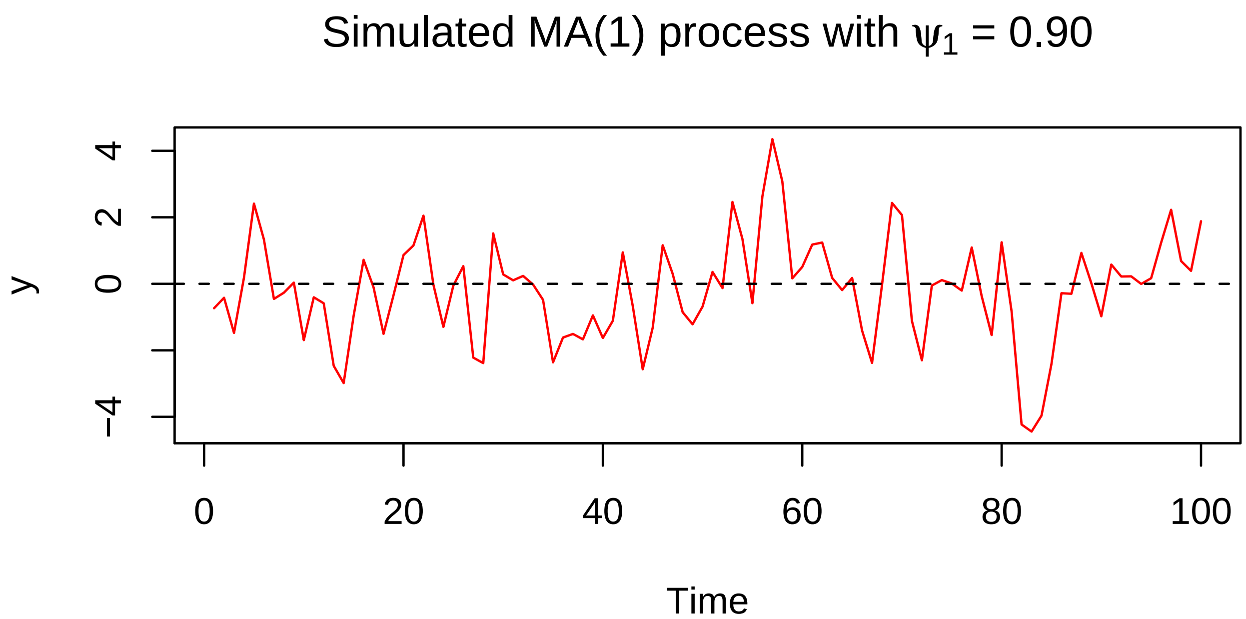


Simulated MA(1) process with $\psi_1 = 0.5$



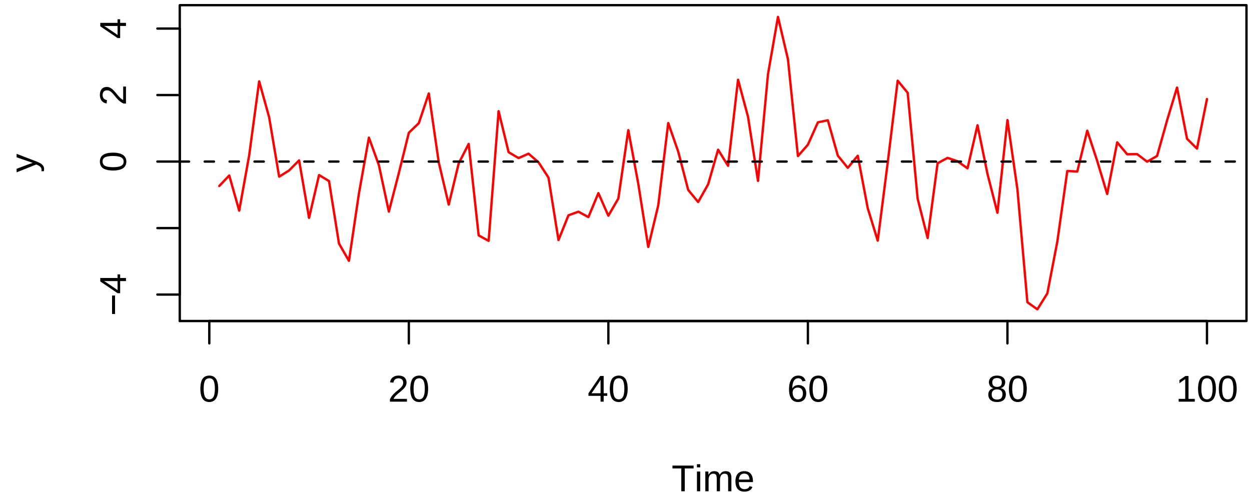
PACF of MA(1) process with $\psi_1 = 0.5$



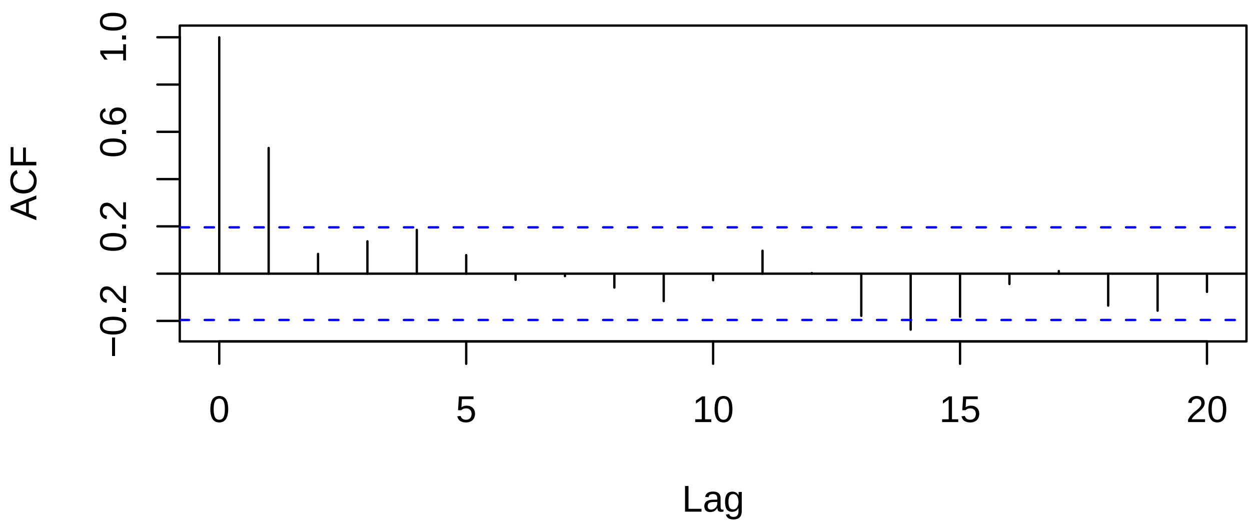


Simulated MA(1) process with ψ1 = 0.90
y
4
2
0
−4
0
20
40
60
80
100
Time

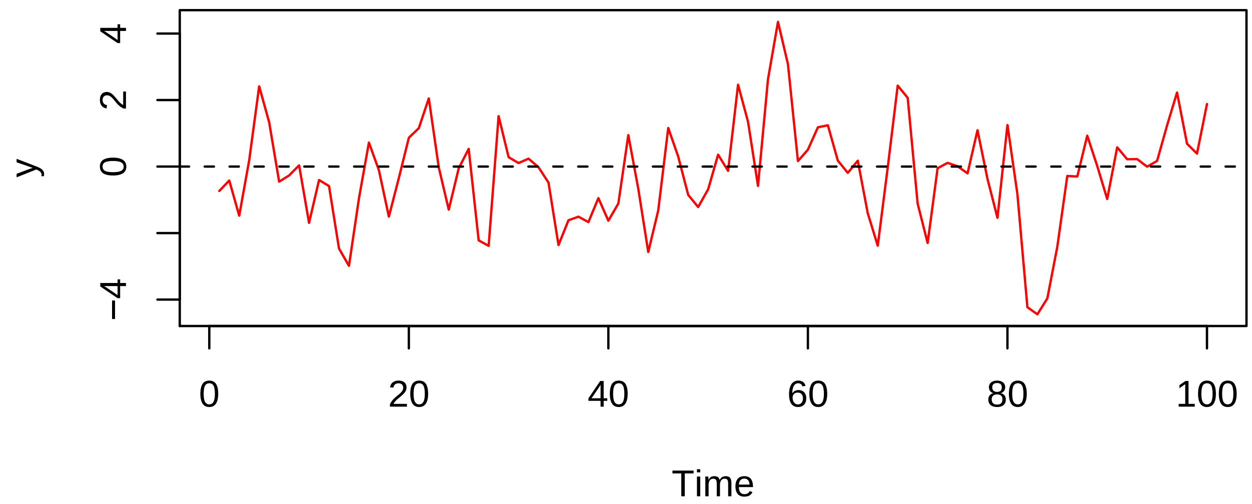
Simulated MA(1) process with $\psi_1 = 0.90$



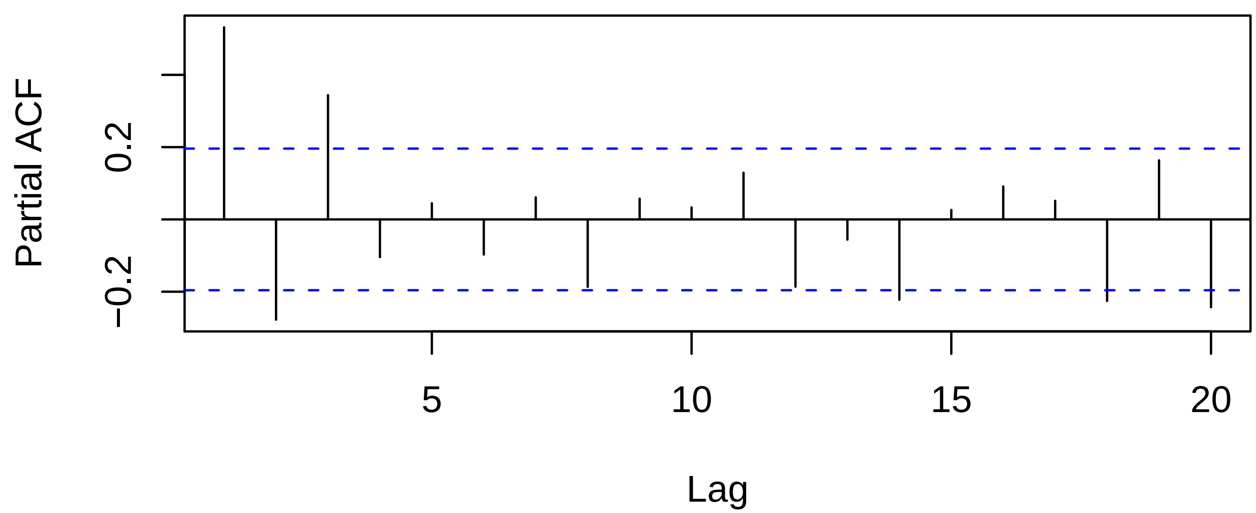
ACF of MA(1) process with $\psi_1 = 0.90$

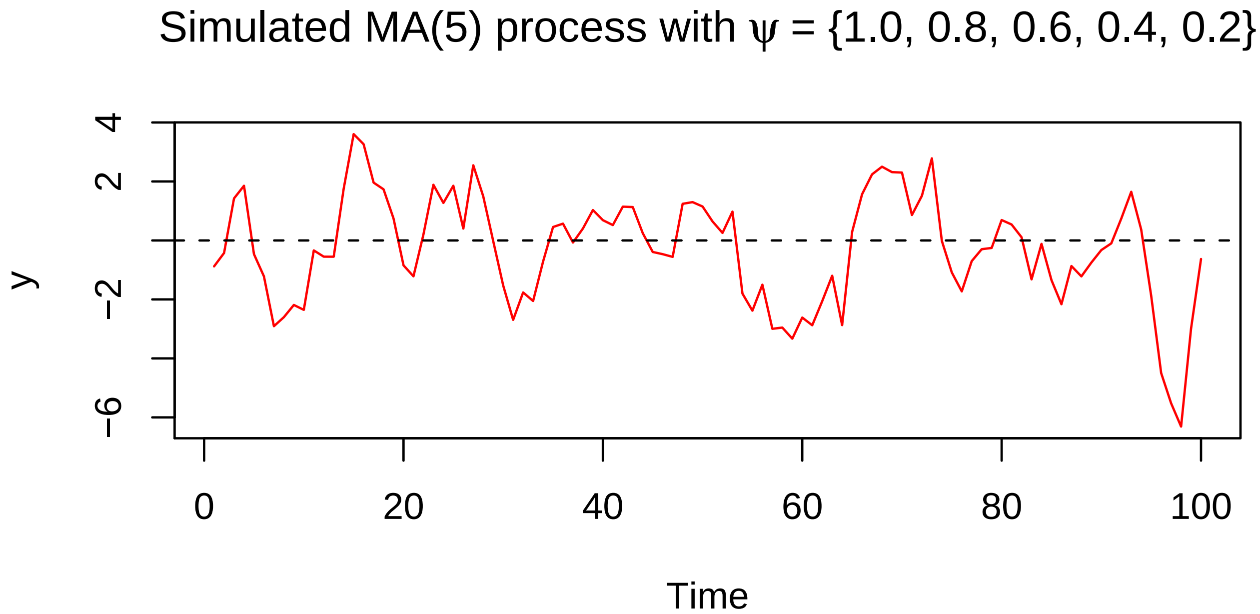


Simulated MA(1) process with $\psi_1 = 0.90$

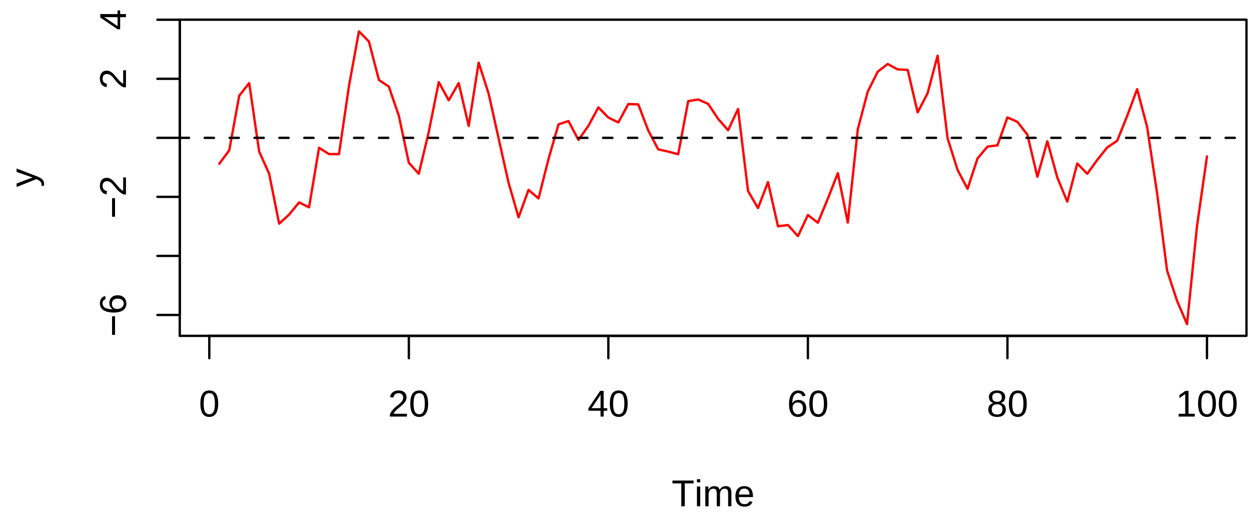


PACF of MA(1) process with $\psi_1 = 0.90$

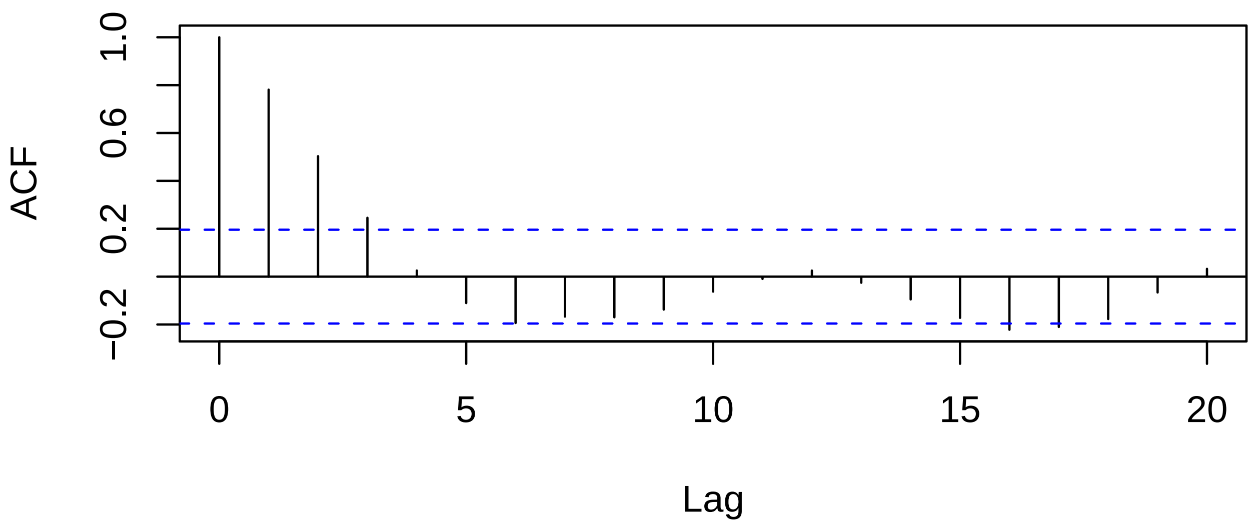


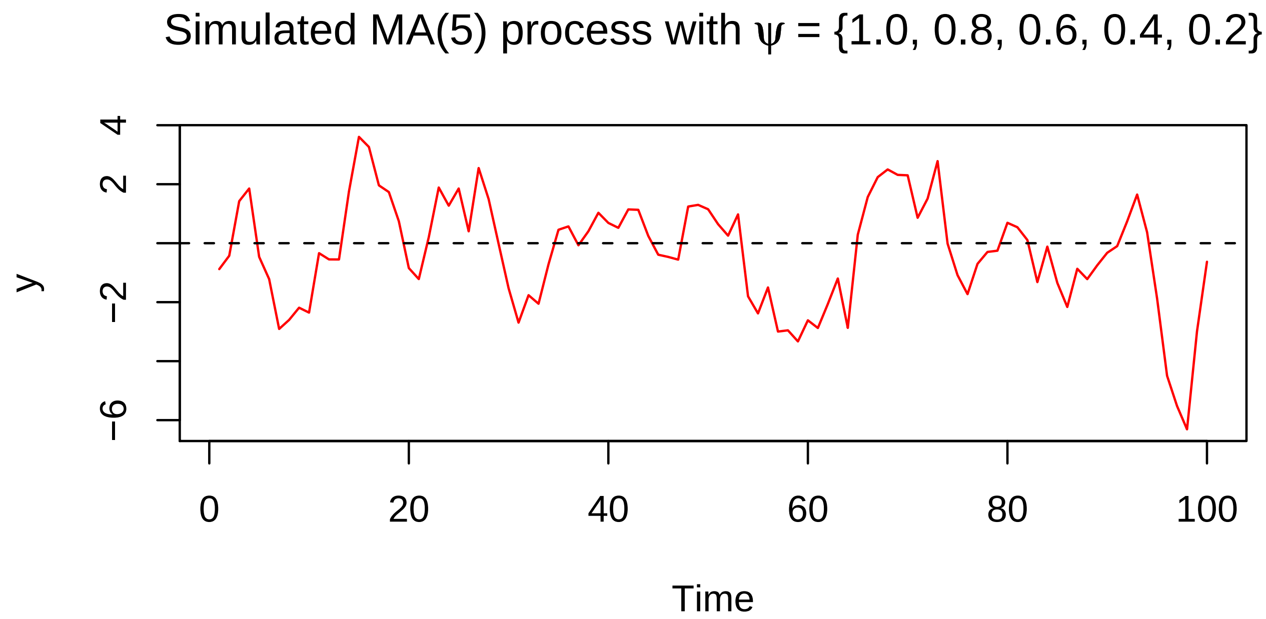

Simulated MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}
y
4
2
−2
−6
0
20
40
60
80
100
Time

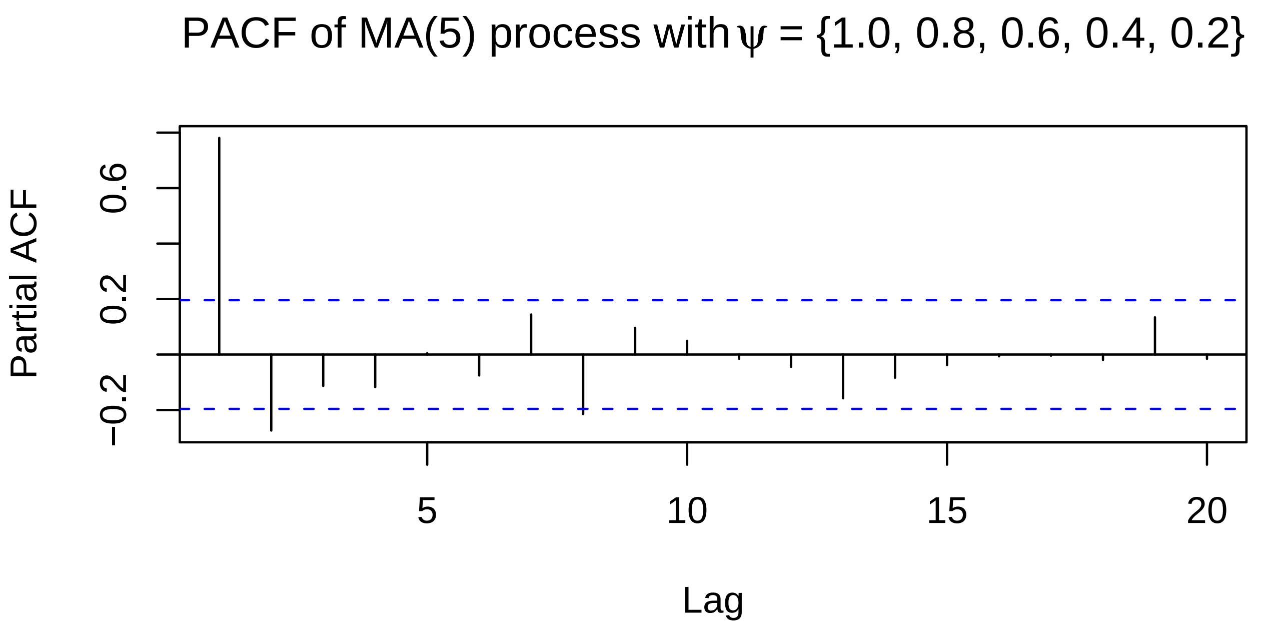
Simulated MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}



ACF of MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}




Simulated MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}
y
4
2
−2
−6
0
20
40
60
80
100
Time


PACF of MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}
Partial ACF
0.6
0.2
−0.2
5
10
15
20
Lag

Equivalence of AR and MA processes

For any AR(1) with parameter ϕ_1, there is some MA(∞) with the right $\rho_1, \rho_2, \ldots, \rho_\infty$ which is equivalent (produces the same time series)

That is, there is an endless pattern of MA terms that exactly replicates the rate of decay of a shock to the time series over time

For any MA(1) with parameter ρ_1, there is some AR(∞) with the right $\phi_1, \phi_2, \ldots, \phi_\infty$ which is equivalent (produces the same time series)

That is, there is an infinite set of AR terms that exact cancel out the long term effect of a shock *except* for a transitory moving average-like effect

Equivalence of AR and MA processes

When looking for the best representation of a time series, and can only choose an AR(p) or an MA(q), one or the other may involve fewer parameters to estimate

But why choose only one? Why not a little of each? Even more efficient: ARMA(p,q)

Past expectations

$$\mu_t = x_t\beta + \mu_{t-1}\phi$$

implies

$$\mu_{t-1} = x_{t-1}\beta + \mu_{t-2}\phi$$

substituting back, we find

$$\begin{aligned}
\mu_t &= x_t\beta + (x_{t-1}\beta + \mu_{t-2}\phi)\phi \\
&= x_t\beta + x_{t-1}\beta\phi + \mu_{t-2}\phi^2 \\
&= x_t\beta + x_{t-1}\beta\phi + (x_{t-2}\beta + \mu_{t-3}\phi)\phi^2 \\
&= x_t\beta + x_{t-1}\beta\phi + x_{t-2}\beta\phi^2 + \mu_{t-3}\phi^3 \\
&= \mu_1\phi^{t-1} + \sum_{j=0}^{t-2} x_{t-j}\beta\phi^j
\end{aligned}$$

The final line can be substituted into a Normal likelihood function.

Past expectations

$$\mu_t = \mu_1 \phi^{t-1} + \sum_{j=0}^{t-2} x_{t-j} \beta \phi^j$$

Notice three things about the final line

1. We still have the first μ_1. We could estimate it.
   Or make some assumption about the first period (e.g., $\mu_1 = y_1$)

2. The present value of y_t turns out to depend on all past values of x

3. But more ancient x_t matter less for smaller $|\phi|$
   $|\phi| > 1$ is again implausible (effects would get bigger and bigger as they aged), eventually becoming infinite

This model is known as a geometric distributed lag.

Ambiguity of different dynamic specifications

We have talked about controlling for

- past realized values (AR processes)
- past expected values (MA process)
- past shocks (distributed lag processes)

But note that these concepts are closely related:

$$y_{t-1} = \mu_{t-1} + \varepsilon_{t-1}$$

Any two are equivalent to the third.

So choosing any two produces identical results to choosing any other two

But with a different interpretation