CSSS 594 / POLS 559:
Time Series and Panel Data for the Social Sciences

Basic Temporal Concepts:
Trends, Stochastic Processes, and Seasonality

Christopher Adolph

Department of Political Science
and
Center for Statistics and the Social Sciences
University of Washington, Seattle

Data in temporal context

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

Observations are identically 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, as we all know
(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

Deterministic Trends. Each period entails another β increase in $\mathrm{E}(y_{it})$

Dynamic Stochastic Processes. y_{it} is a function of $y_{i,t-p}$ and/or $\varepsilon_{i,t-q}$, for p and $q \in 0, 1, \ldots$

Seasonality (Cycles). y_{it} is an additiive or multiplicative function of $y_{i,t-c}$ for some fixed cycle c (e.g., $c = 12$ for months)

All three have their uses, but the middle category is the most powerful/flexible set of tools for modeling dynamics

Notation

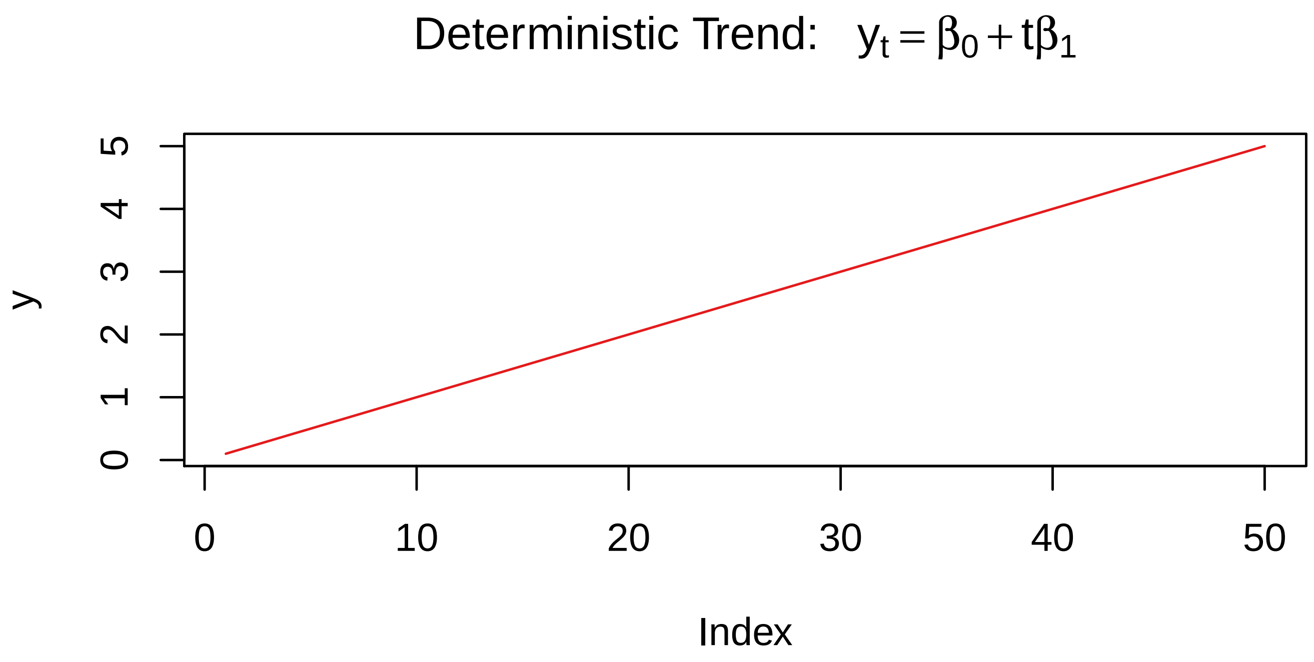
We assume time is discrete

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 (i.e., non-panel) time series y_t, we have T observations total

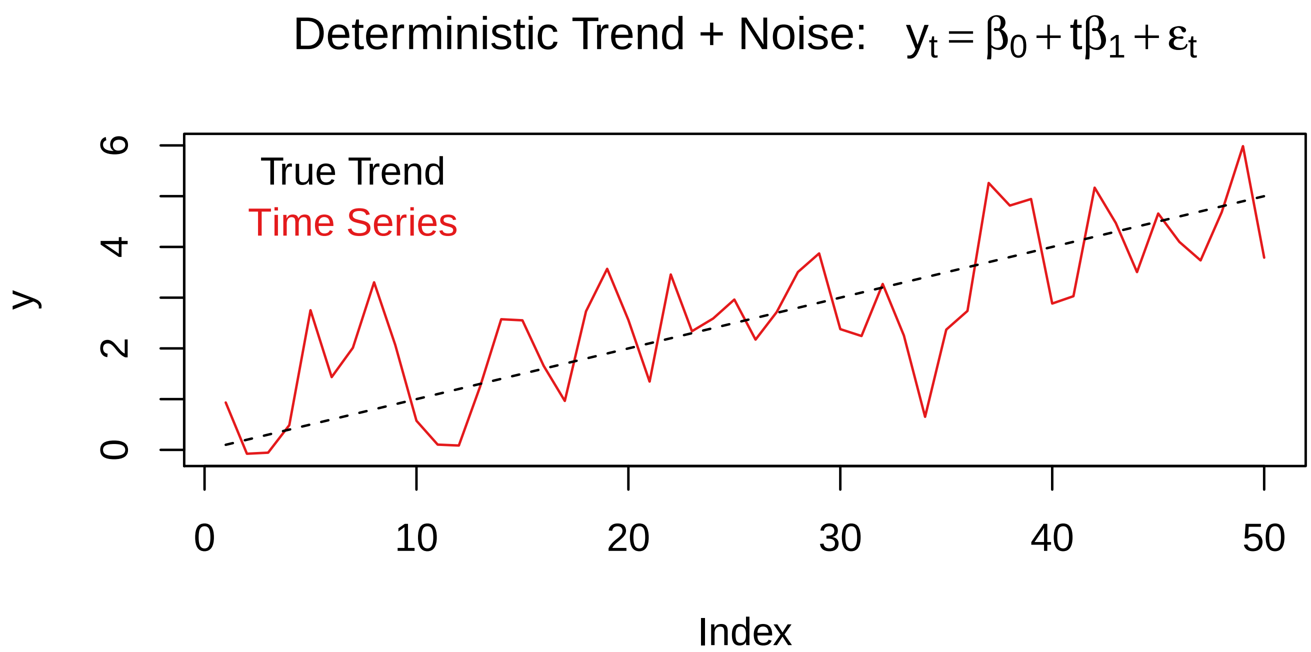


The simplest way that time might matter is through a *deterministic trend*

$$y_t = \beta_0 + t\beta_1$$

In this model, y_t starts at β_0 and increases by *exactly* β_1 units each period

(Notice there is no error term above – this y_t is purely systematic)

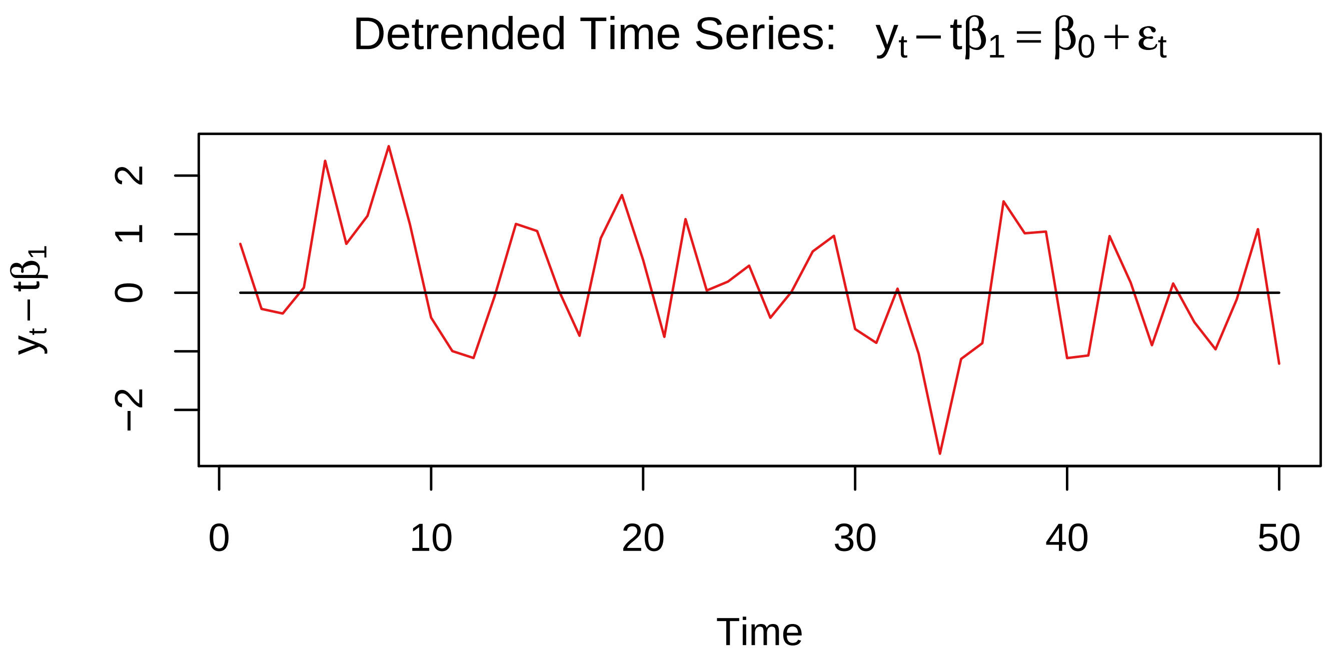


We could easily add a stochastic component to this model

$$y_t = \beta_0 + t\beta_1 + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Now the data are "noisy,"
but time only affects the data through the deterministic trend

That is, time has a simple, systematic effect on y:
every period, just add β_1 more units to the expected value of y

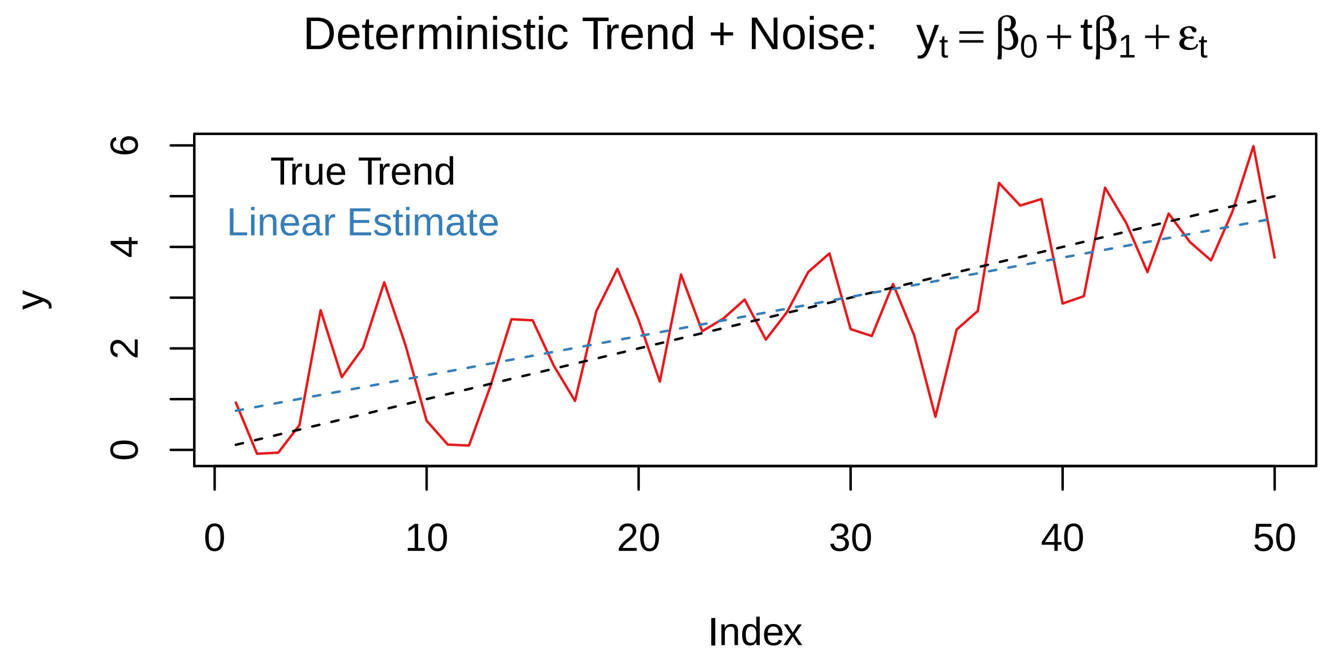


We can remove the trend by substracting $t\hat{\beta}_1$ from the time series

$$y_t - t\hat{\beta}_1 = \gamma_0 + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2)$$

Now the *detrended* time series is simply *white noise*

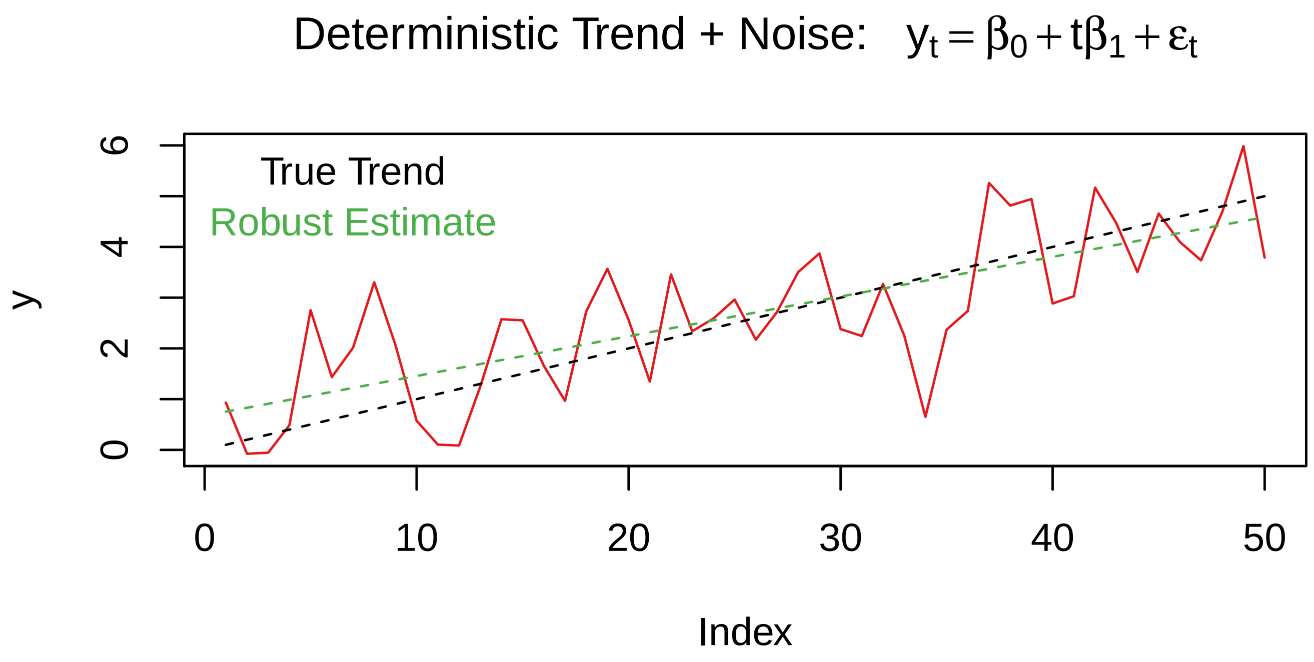
White noise has no correlation from observation to observation,
so in this example, there is no longer any correlation across time



$$y_t = \beta_0 + t\beta_1 + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

A warning: Estimating β_1 without bias can be tricky.

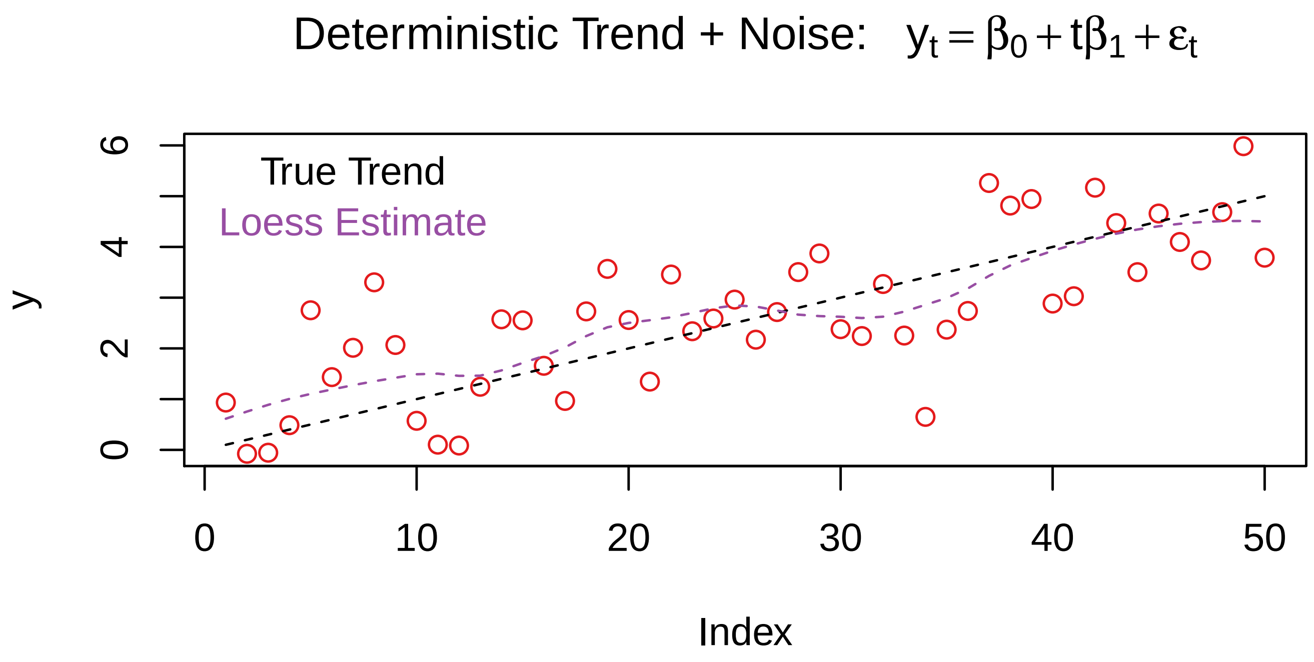
1. The first (few) and last (few) observations are high leverage cases – they have extreme values on the covariate t – so they can lead to bias if they have large error terms.



$$y_t = \beta_0 + t\beta_1 + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Even though the problem is caused by influential outliers, robust and/or resistant regression doesn't always help

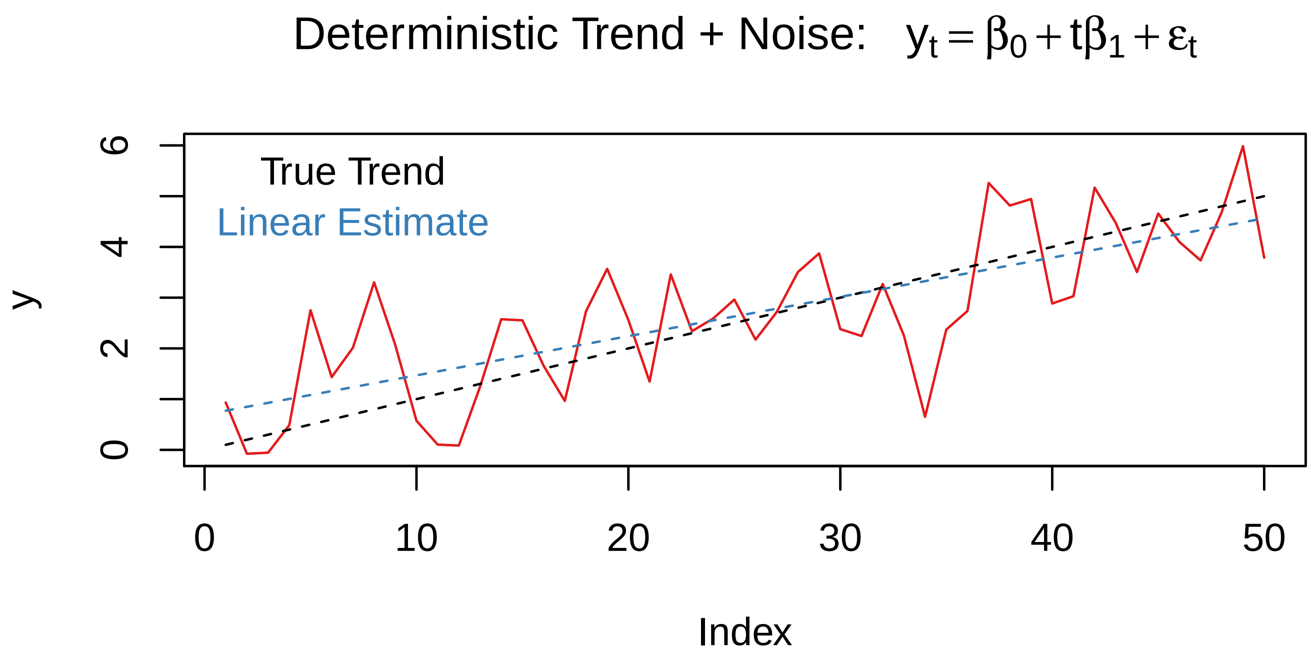
It didn't reduce the bias at all in my example



$$y_t = \beta_0 + t\beta_1 + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$

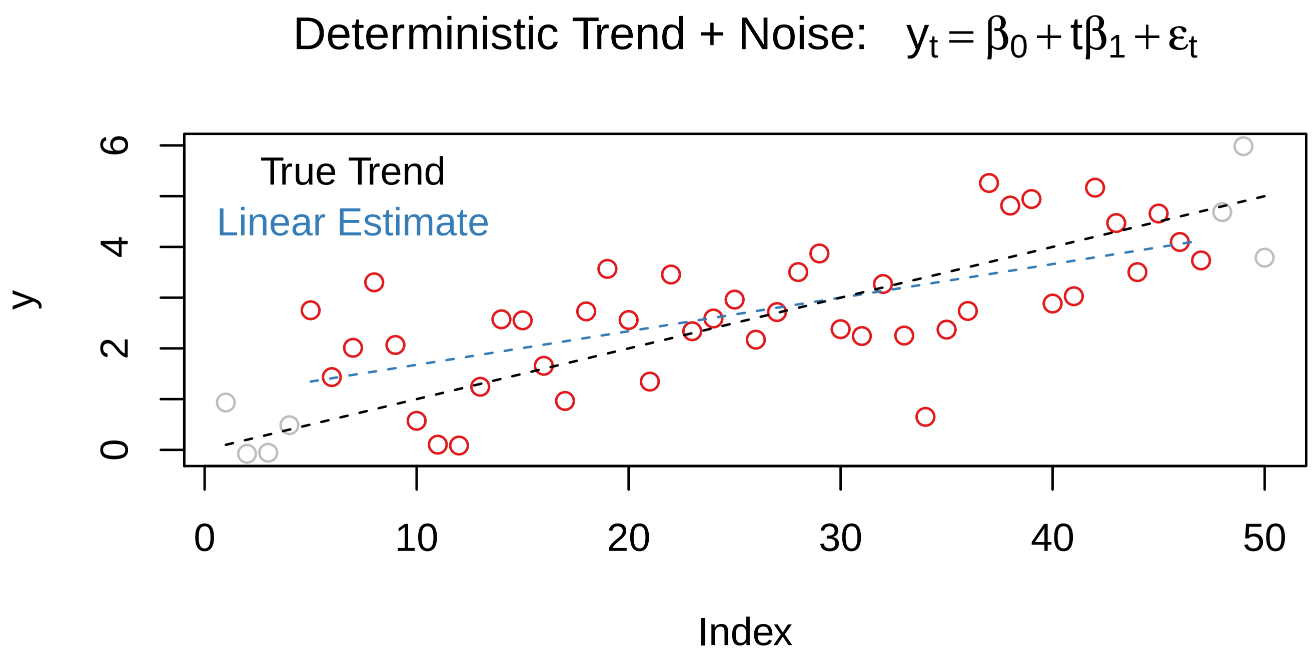
Nor does a local, robust regression necessarily show the true trend

Here is loess with a bandwidth of 0.5 – *still biased*



2. The choice of the starting and ending periods can thus dramatically shift the estimate of $\hat{\beta}_1$, even changing its sign

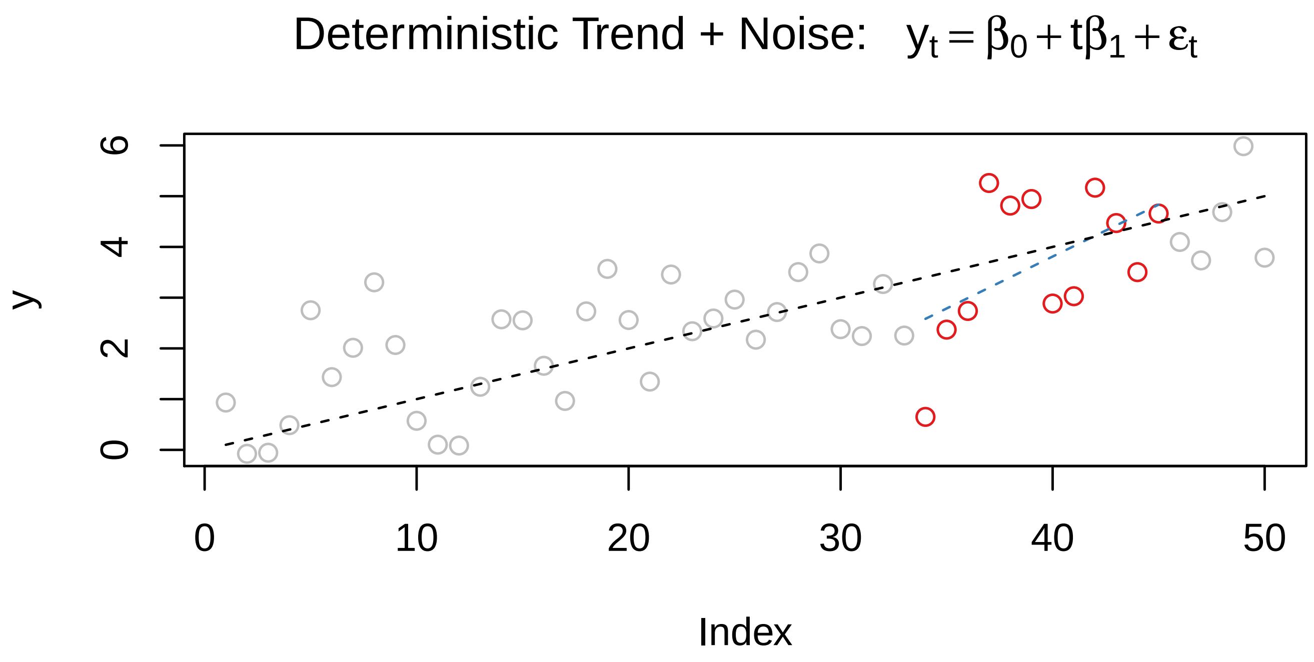
"Choice" doesn't just mean your decision to truncate the series, but the happenstance of data generation and collection



Suppose we'd collected only the middle periods of data (in red)–
then, in this case, we'd get even more attenuation bias

If these were a cross-sectional sample,
it would be obvious that systematic choice of a subset for analysis
could induce selection bias

Truncation in time is simply a form of selection bias,
but we might not realize it because choosing ranges of years may seem "neutral"

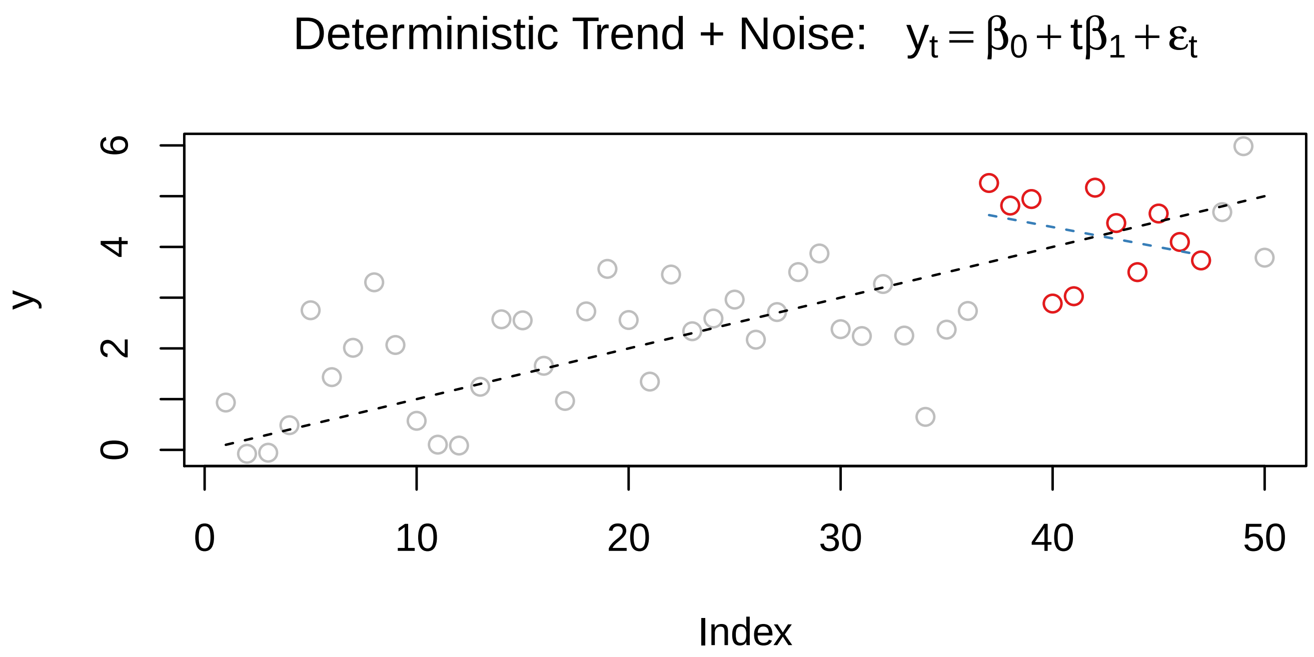


It's not. We can even chose a window of years to get a particular "trend"

Choosing this window of years maximizes upward bias

Honesty and willingness to perform sensitivity analyses are critical here

For example, a large fraction of dishonest studies
on climate change choose time series starting points to hide trends



Potential bias in trend estimation from cherrypicking ranges of time is large

Choosing this window of years flips the trend sign to negative,
even though the true trend is positive!

What would help? Lots of periods, strong trends, low variance

Deterministic Trends

If the only way time could affect our data were through deterministic trends, time series would be a much easier topic in statistics

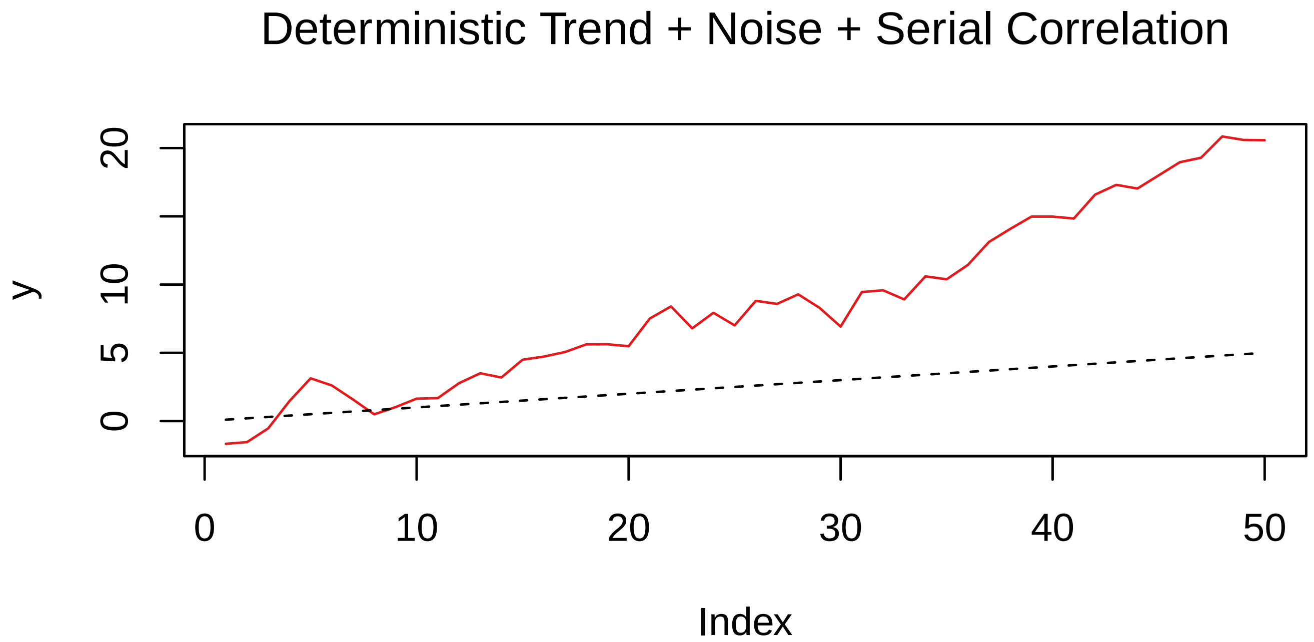
Setting aside the problem of bias,
we could just use least squares estimate the model:

$$y_t = \beta_0 + t\beta_1 + x_{2t}\beta_2 + x_{3t}\beta_3 + \cdots_{kt}\beta_k + \varepsilon_t$$

But time can matter in other ways; in particular, through the stochastic component

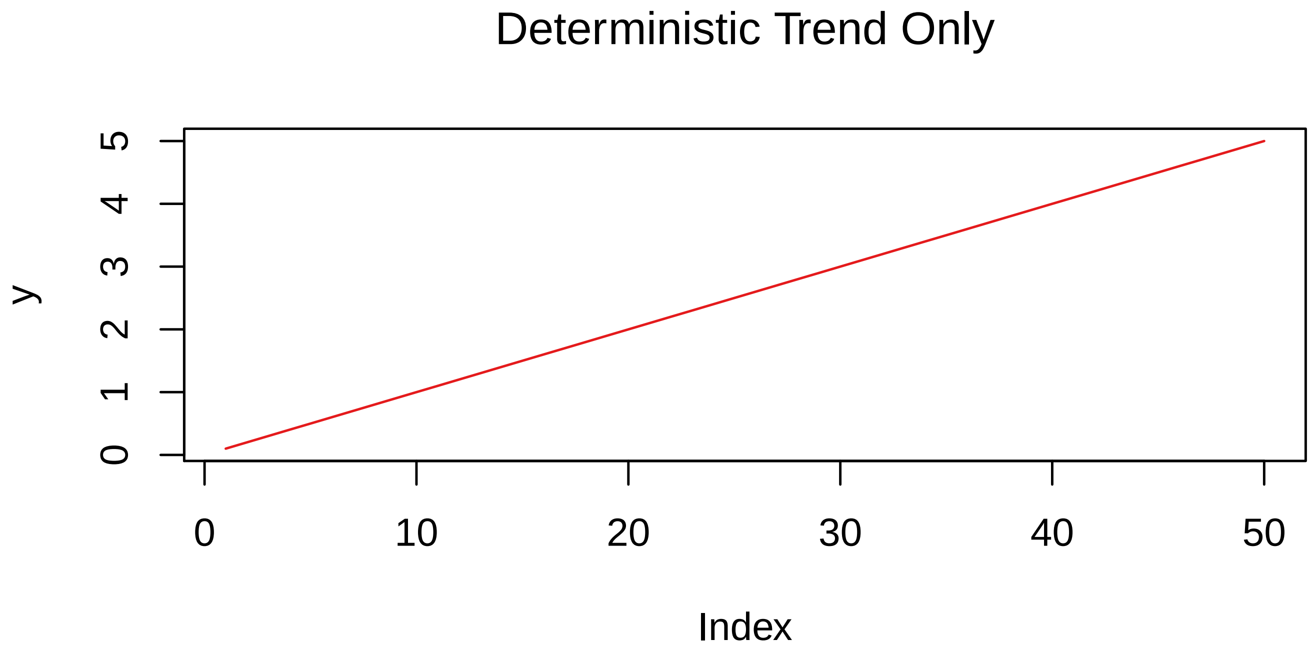
In this way, the recent past can have a strong *probabilistic* effect on the future

Some of these stochastic processes can even mimic deterministic trends,
at least for some period of time



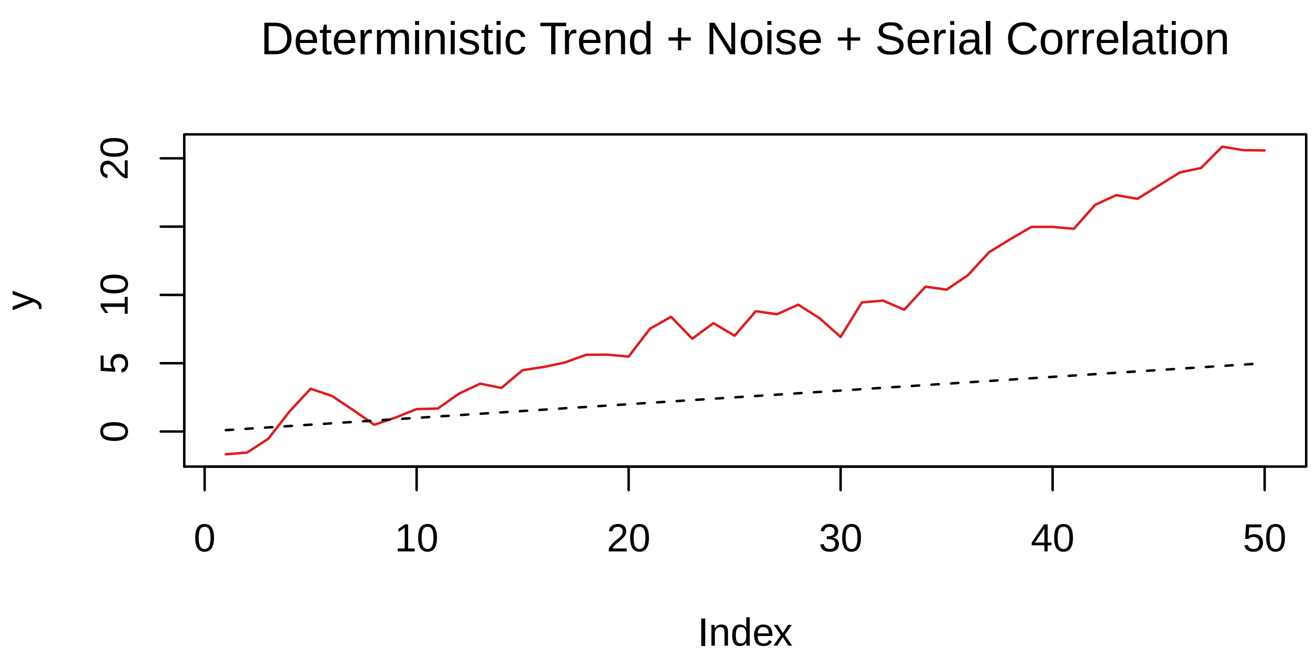
This time series includes two kinds of temporal dependence

There is a deterministic trend. . .



This time series includes two kinds of temporal dependence

There is a deterministic trend. . . (in particular, the same one as before)



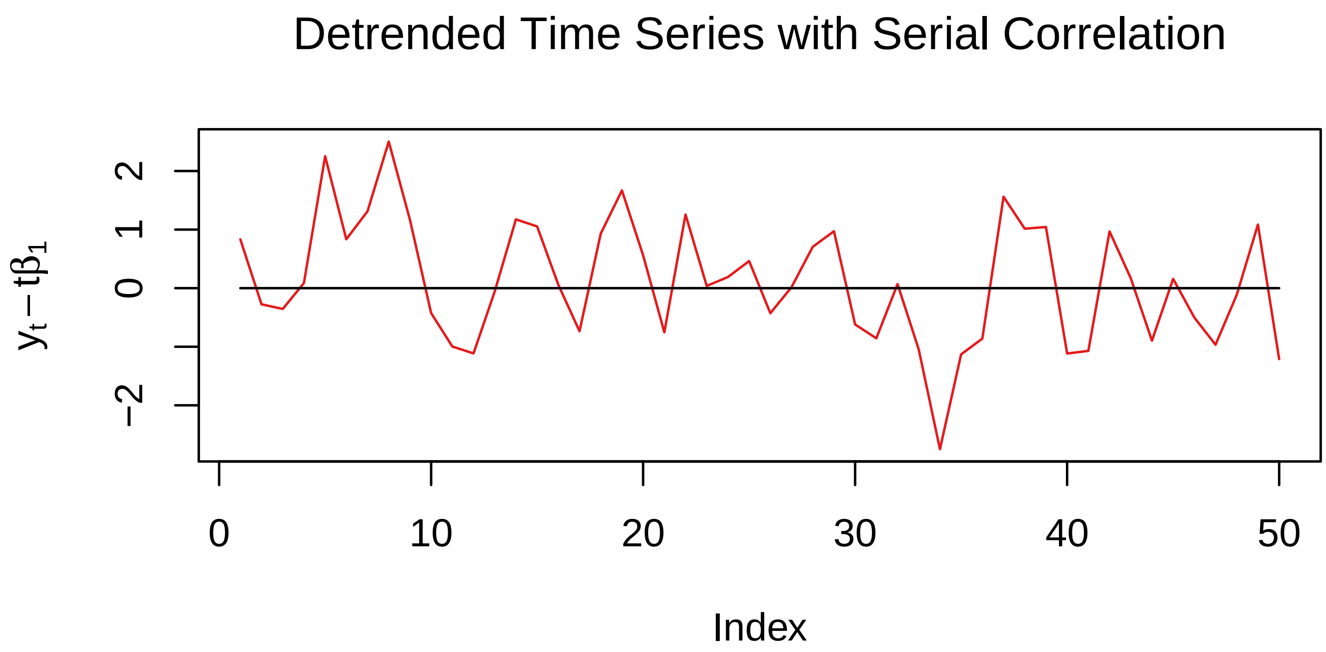
This time series includes two kinds of temporal dependence

There is a deterministic trend. . . (in particular, the same one as before)

But notice that the apparent “trend” in the data is much larger than the deterministic component of the trend

The gap is produced by stochastic variation that can mimic or magnify trends

Put another way, history plays a more complex role in this example than simply increasing y by β units each period



If we remove the deterministic trend – that is, subtract $t\beta_1$ – we find the detrended time series

We still see greater similarity between adjacent time periods than we do among more distant periods

That is, there is some kind of stochastic dependence among the errors in historically proximate periods

We call this *serial correlation*

Deterministic Trends

So when is it okay to model your time series using a trend variable?

1. Look at the data as a time series plot (always the most important step!) and see if it looks like a very clear trend

2. Check to see if your results are strongly sensitive to small changes in observation period

3. If you have panel data, check whether each cross-sectional unit displays the same trend

If you meet these conditions, controlling for a trend may be okay
(or if the time series is very short, it may simply be all you can do)

But if you don't meet these conditions,
or still have serial correlation after detrending,
you will need to do more

Dynamic stochastic processes

In addition to deterministic trends, history can have stochastic effects, inducing serial correlation in our data

Several conceptually different ways to think of the stochastic effect of history:

1. Past realizations – like y_{t-1} – influence current levels of y

2. Past shocks to y – like ε_{t-1} – influence current levels of y

3. Past expectations of y – like $\mathrm{E}(y_{t-1})$ – influence current levels of y

To focus on these more subtle dynamics, we will do two things for the rest of this topic:

- We will work in time series which we can assume have been *detrended*; that is, assume we have removed $t\beta$ if necessary

- We will ignore any other covariates that might influence y

Examples of dynamic stochastic processes

Past realizations of y influence current levels of y.

Examples: Unemployment; Welfare state spending uses last year's budget as baseline

Past shocks to Y influence current levels of Y

Examples: Some forms of financial volatility? Voting in U.S. Congress?

Past expectations of Y influence current levels of Y

Examples: 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, or an autoregressive process of degree 1.

However, the distant 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.

For AR(1), recursion reveals the following:

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

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:
$\varepsilon_{t-1}, \ldots \varepsilon_{t-\infty}$

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

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)
```

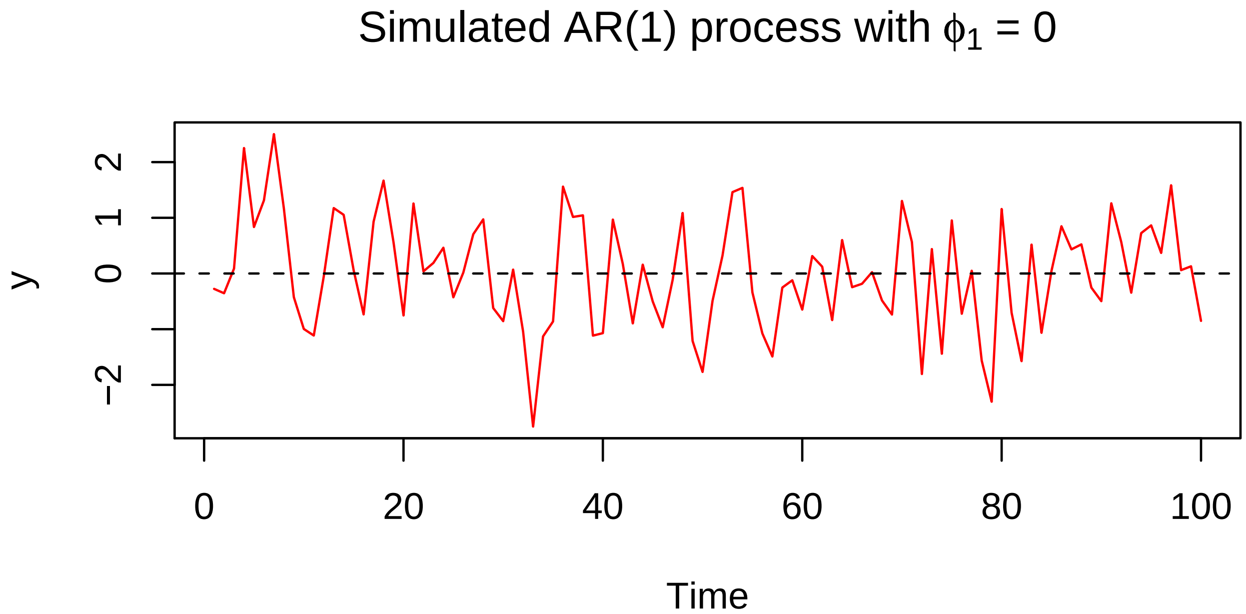
Simulating time series

A more flexible simulator I wrote to make Problem Set 1
(available at the course website)

To make the same series as above. . .

```
y <- makeTS(n = 1000,      # Simulate 1000 periods
            ar = 0.67,  # AR(1); phi = 0.67
            ma = NULL,      # MA(0)
            trend = 0.01,  # Determ. trend (+0.01 units/period)
            seasons = c(0,0,0, 0,0,0, 0,0,0, 0,0,0),
                                    # 12 period cycle
            adjust = c("level"), # Additive cycles; alternative
                      # is "factor" which multiplies last round
            varY = 1,       # Error term variance
            initY = 0,      # Prior level of time series, for lags
            initE = 0,      # Prior error, for lags
            burnin = 10     # Discard this many iterations at start
            )
```

but as you can see, we could add deterministic trends or seasonal cycles


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

Autocorrelation functions

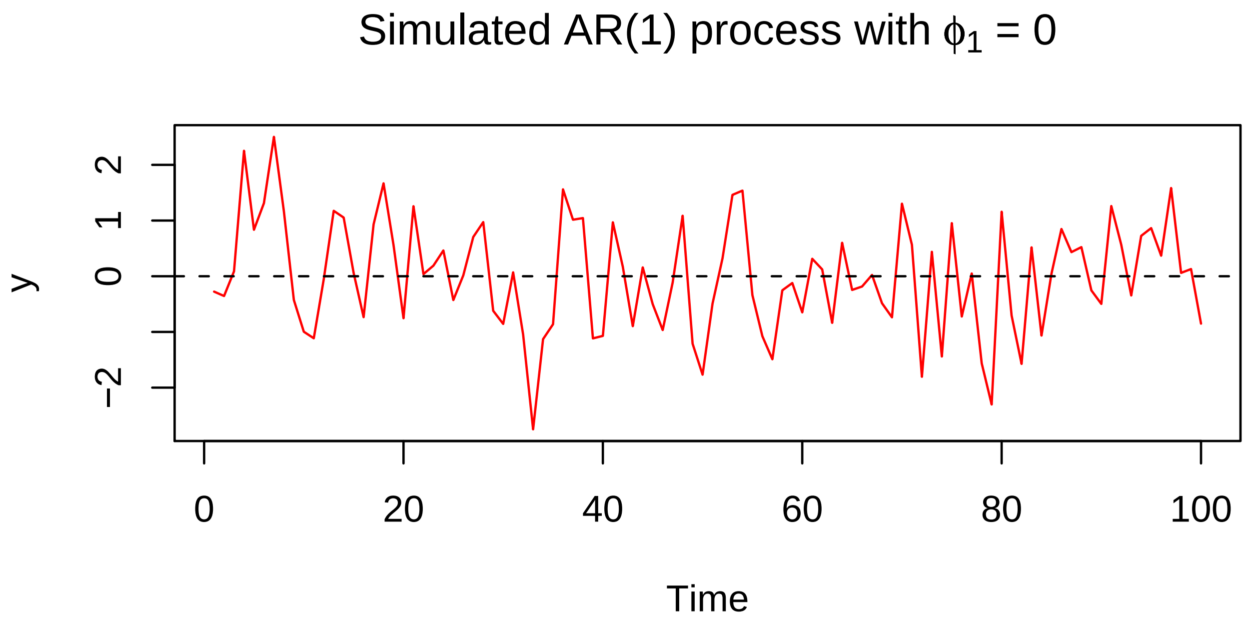
Simplest way to "see" autocorrelation is to calculate and plot the correlation between observations separated by a given distance k for $k = 1, 2, 3, \ldots$.

Define the autocorrelation function as

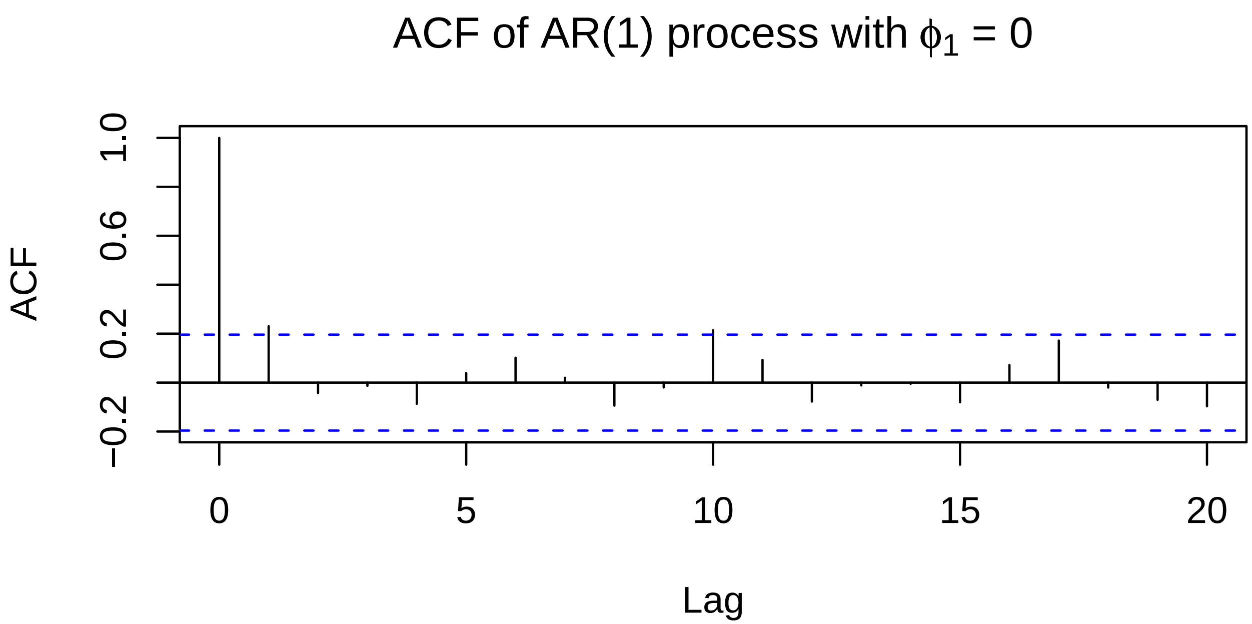
$$\text{ACF}_j = \frac{\text{cov}(y_t, y_{t+j})}{\text{var}(y_t)}$$

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

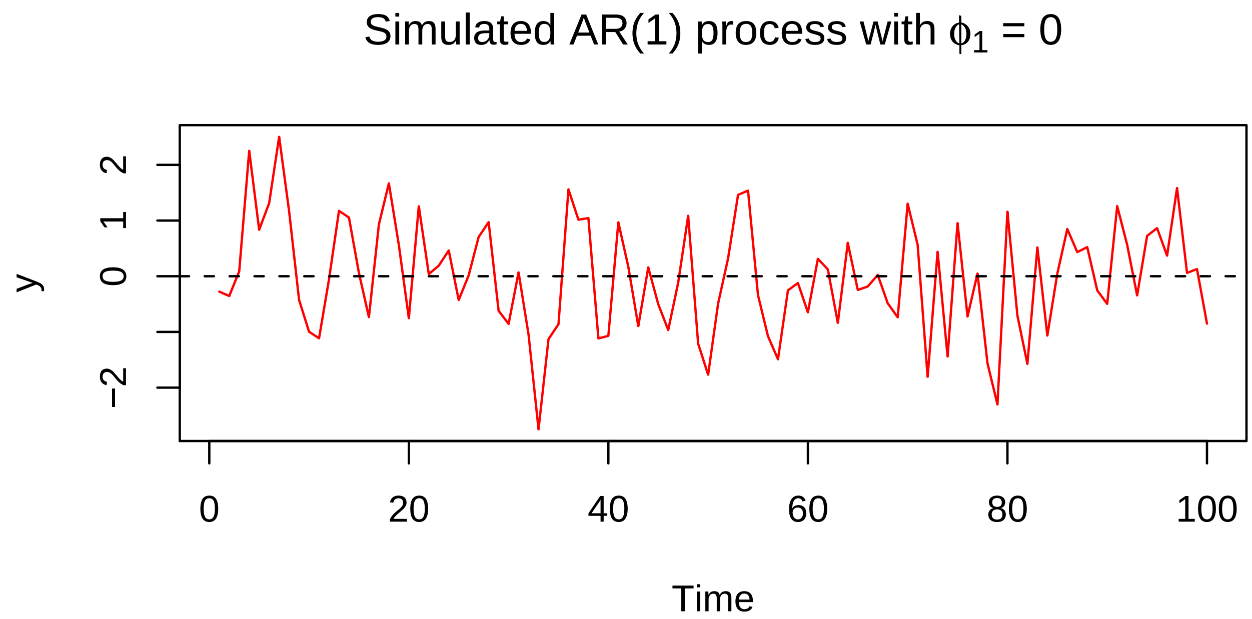
Autocorrelation functions

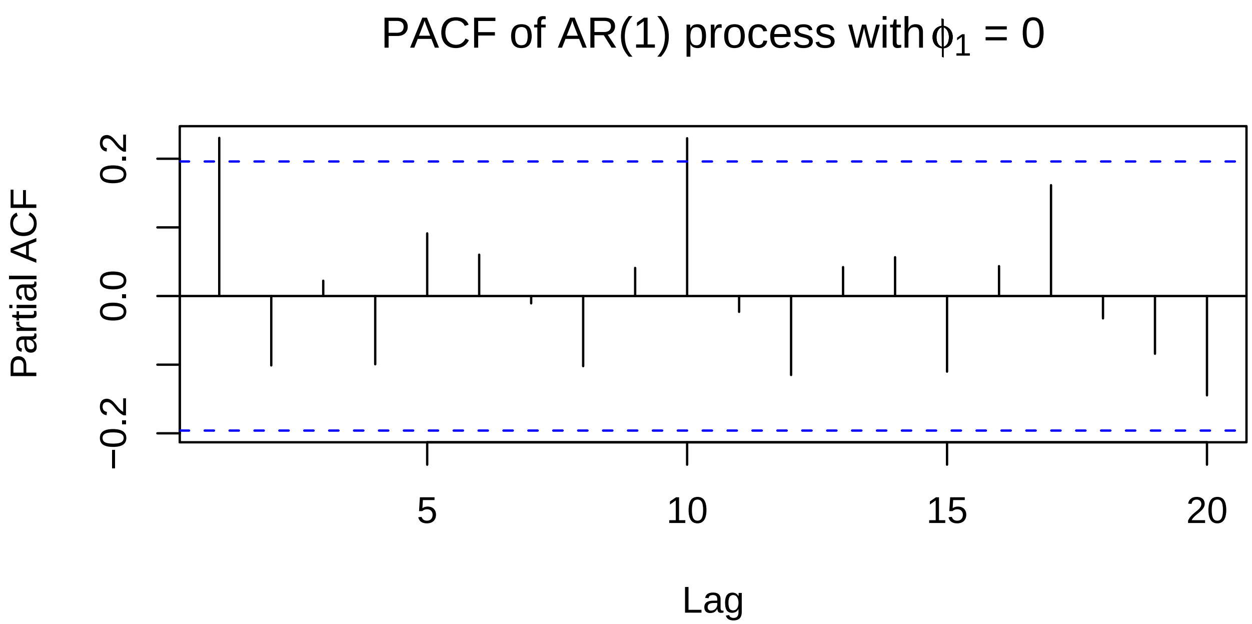
A useful refinement of the ACF is to "partial out" the effects of intervening lags

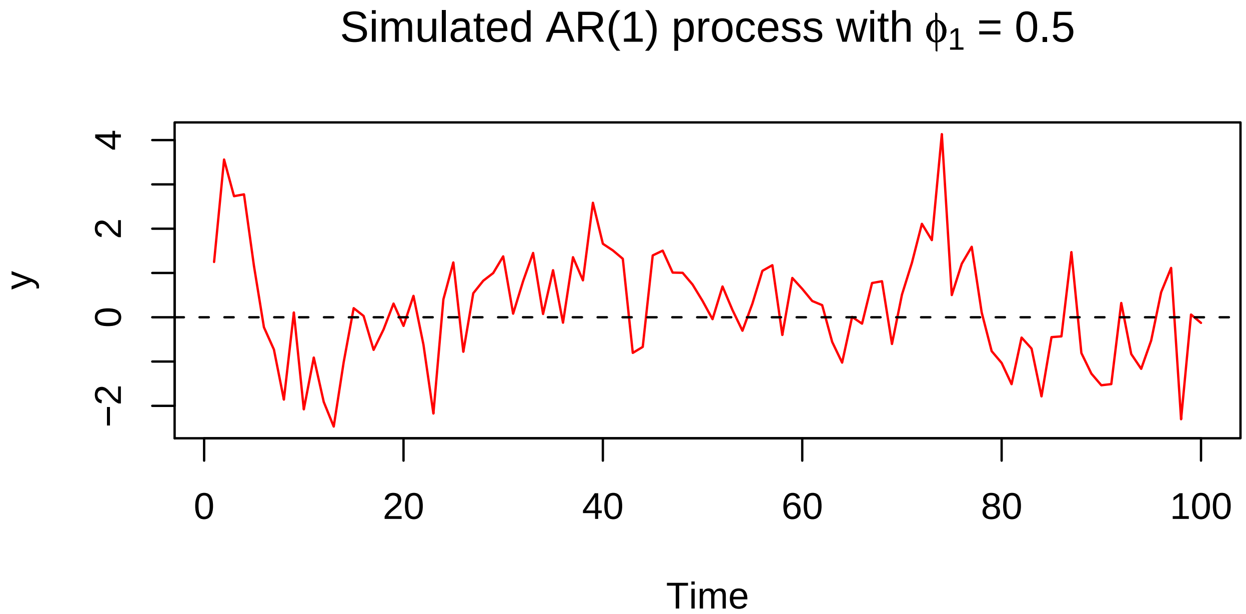
That is, we want to isolate the conditional correlation of y_t and y_{t-k} controlling for the values y_{t-1} to y_{t-k+1}.

We call this the partial autocorrelation function, or PACF.

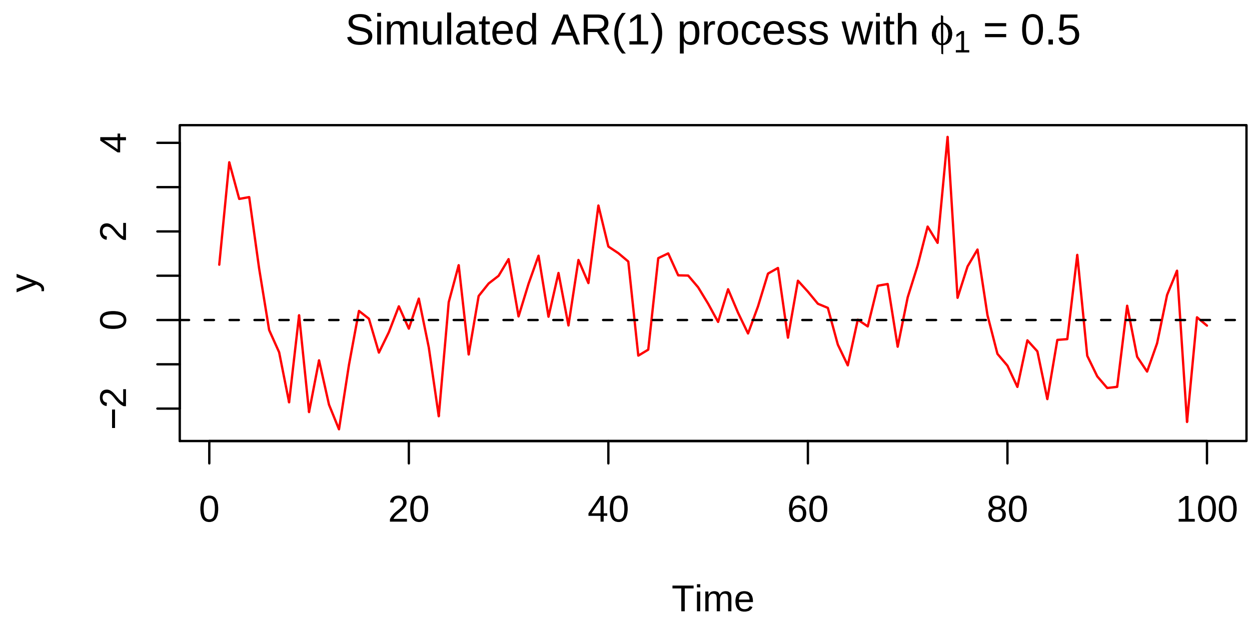
In `R`, just do `pacf(y)`


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

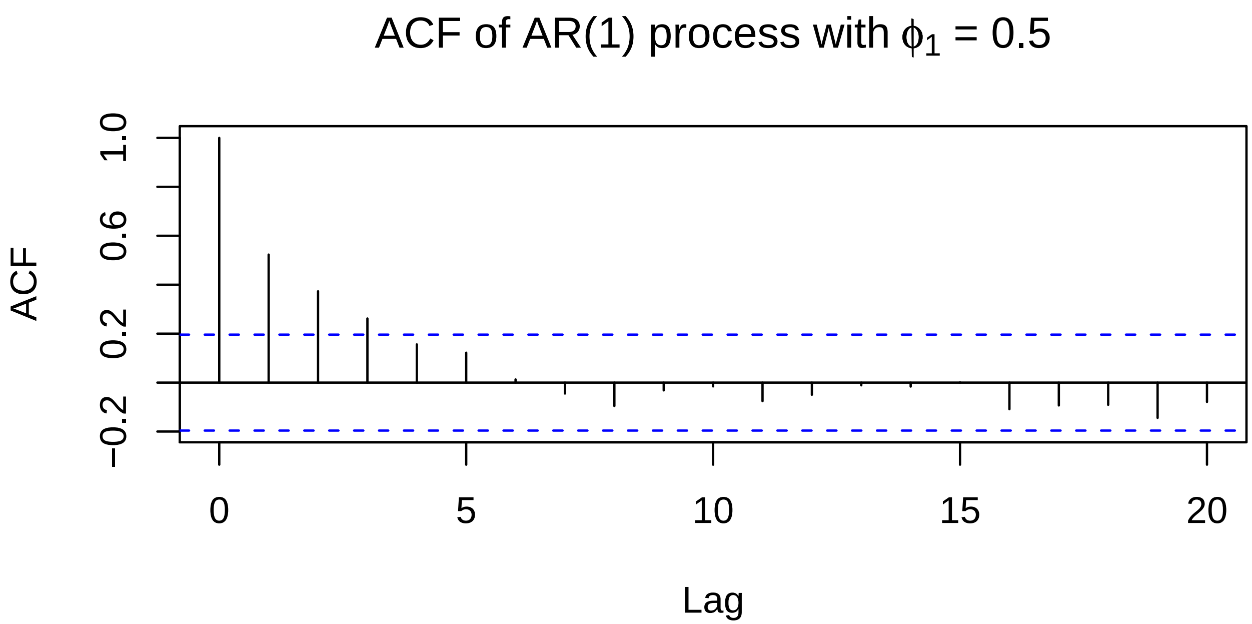

PACF of AR(1) process with $\phi_1 = 0$
Partial ACF
−0.2
0.0
0.2
5
10
15
20
Lag



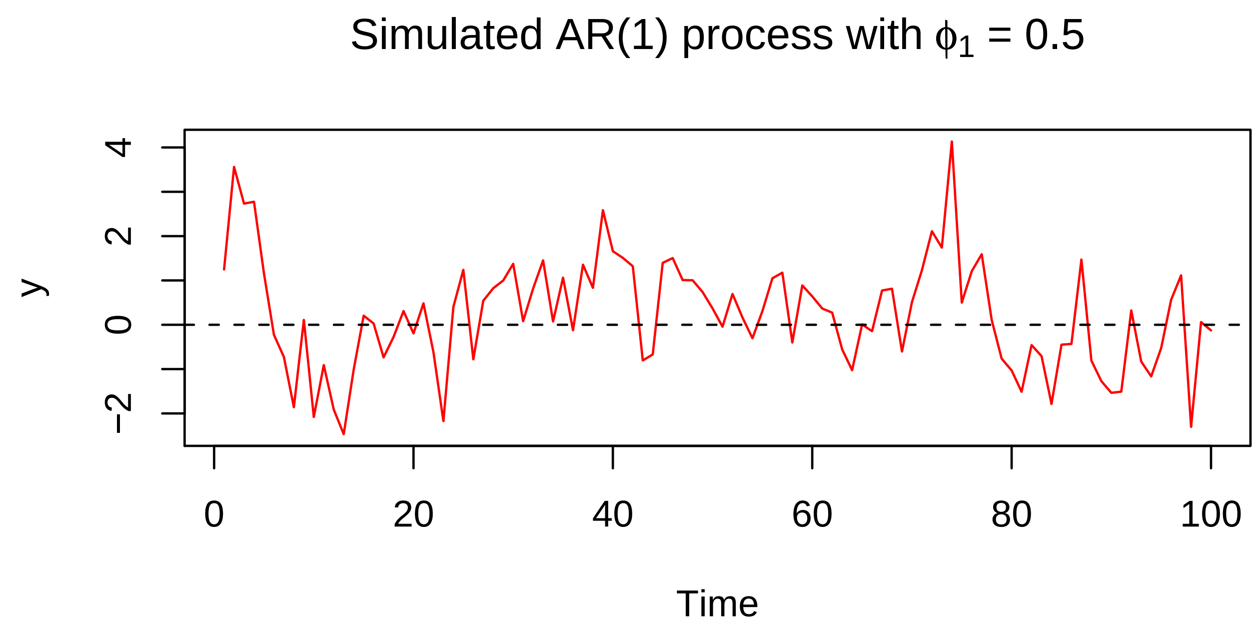
Simulated AR(1) process with ϕ_1 = 0.5
y
4
2
0
−2
0
20
40
60
80
100
Time

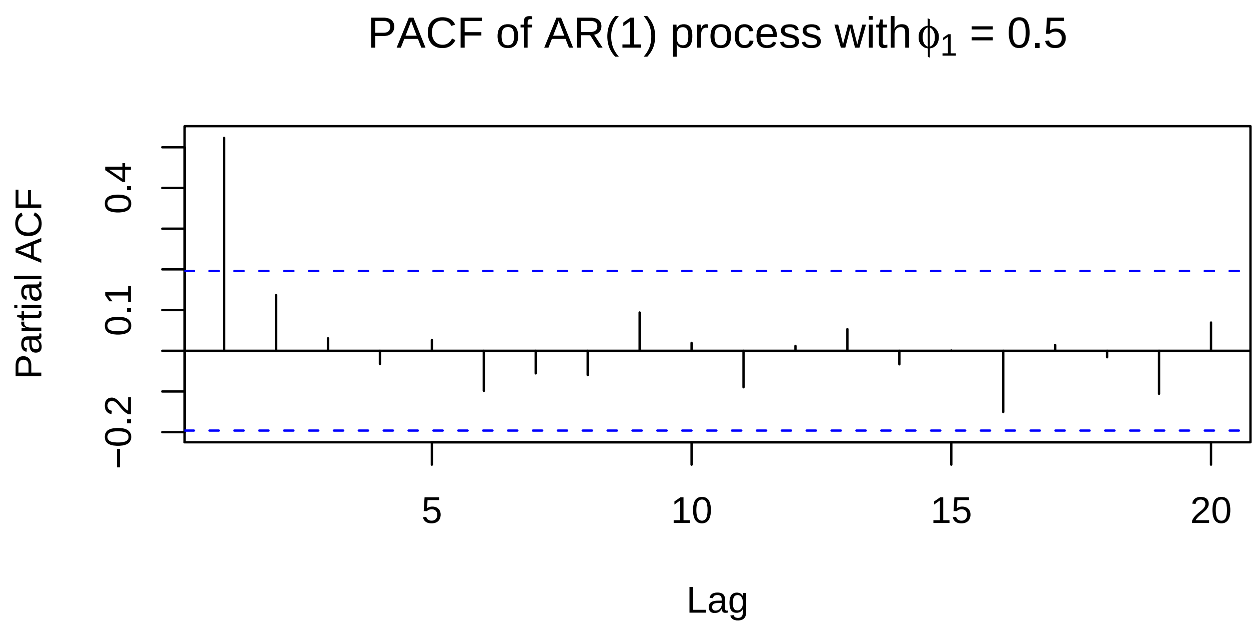


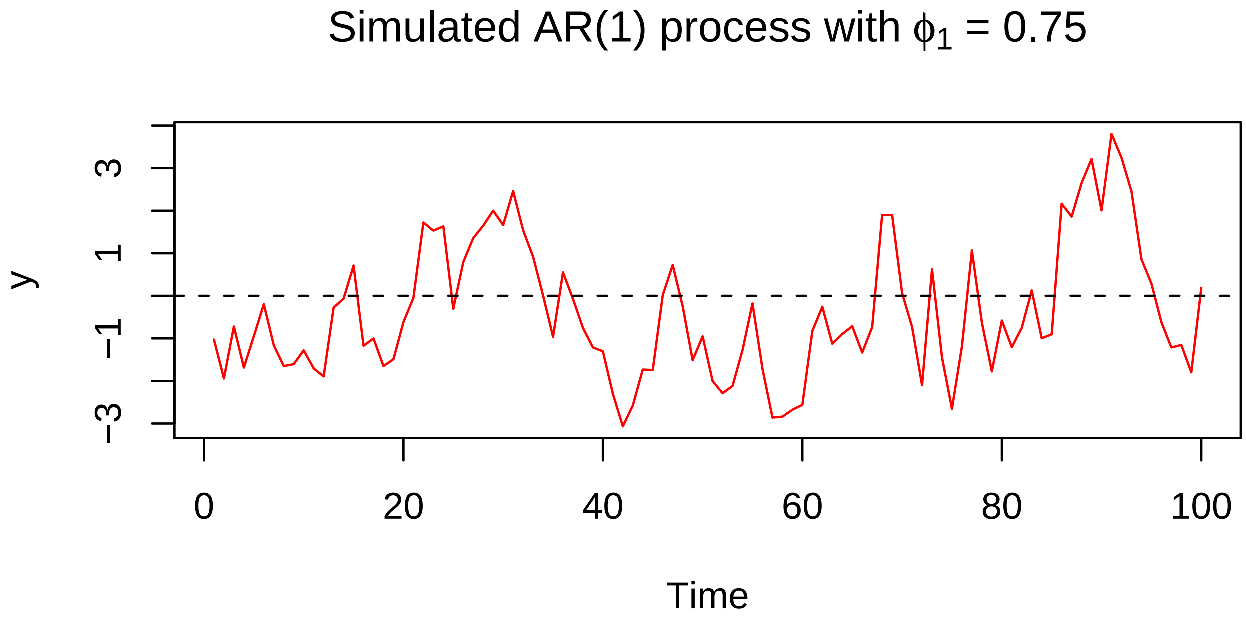
Simulated AR(1) process with $\phi_1 = 0.5$
y
4
2
0
−2
0
20
40
60
80
100
Time



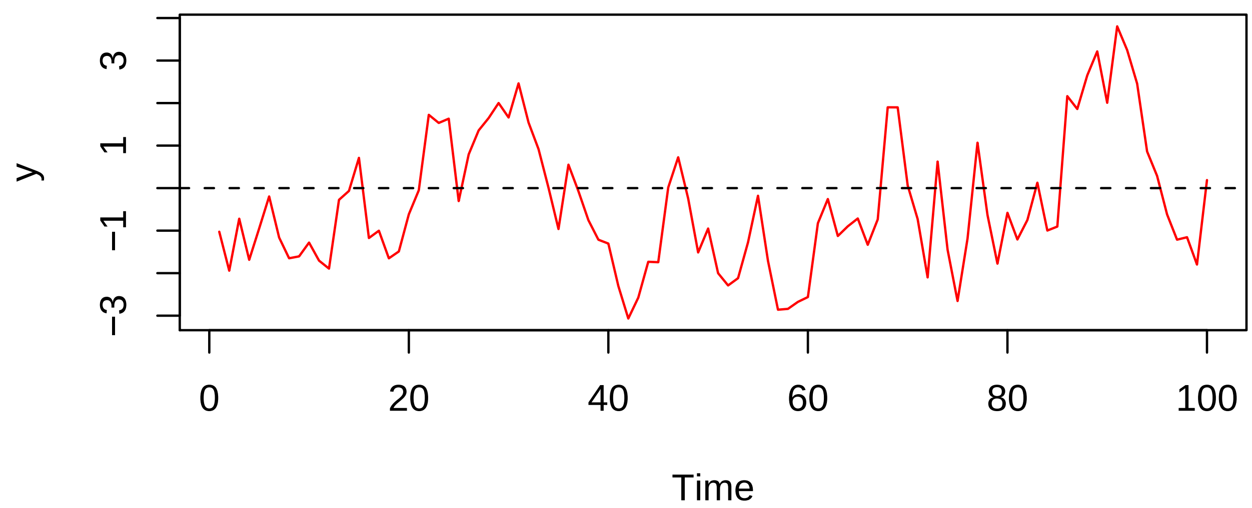
ACF of AR(1) process with $\phi_1 = 0.5$
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag


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

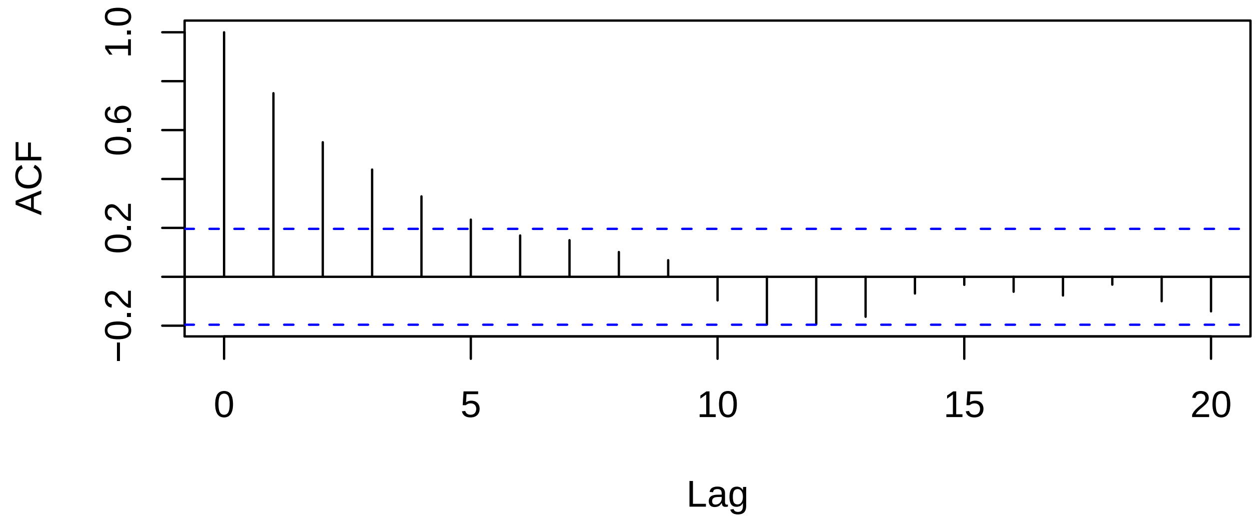

PACF of AR(1) process with $\phi_1 = 0.5$
Partial ACF
0.4
0.1
−0.2
5
10
15
20
Lag

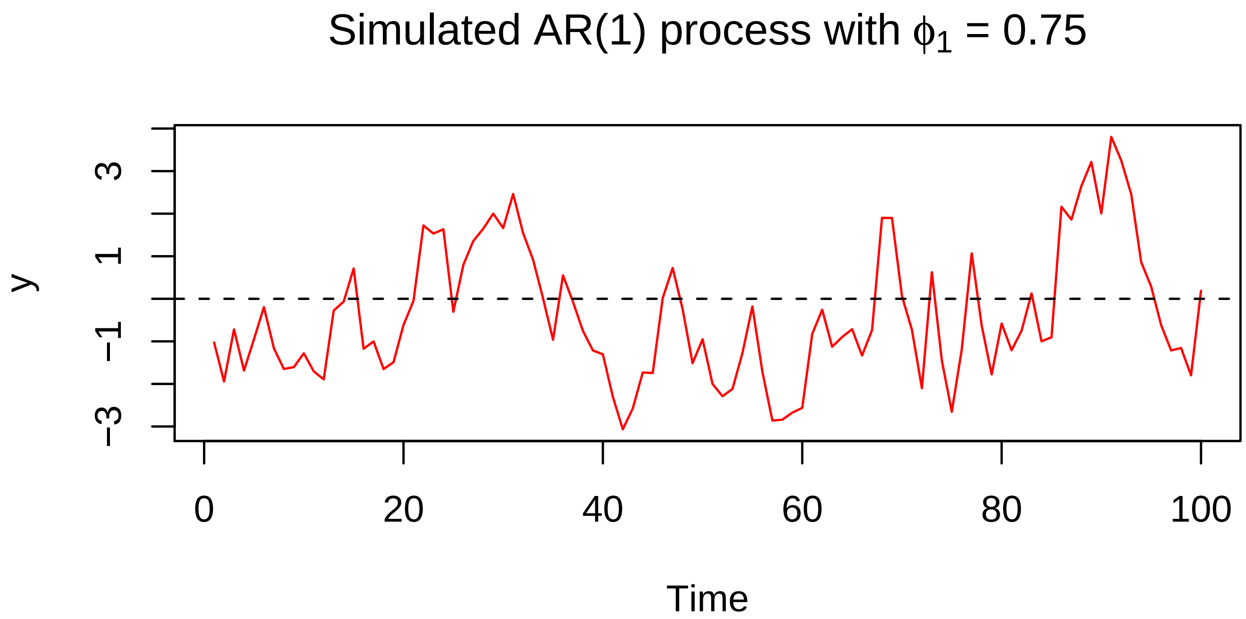

Simulated AR(1) process with $\phi_1 = 0.75$
y
−3
−1
1
3
0
20
40
60
80
100
Time

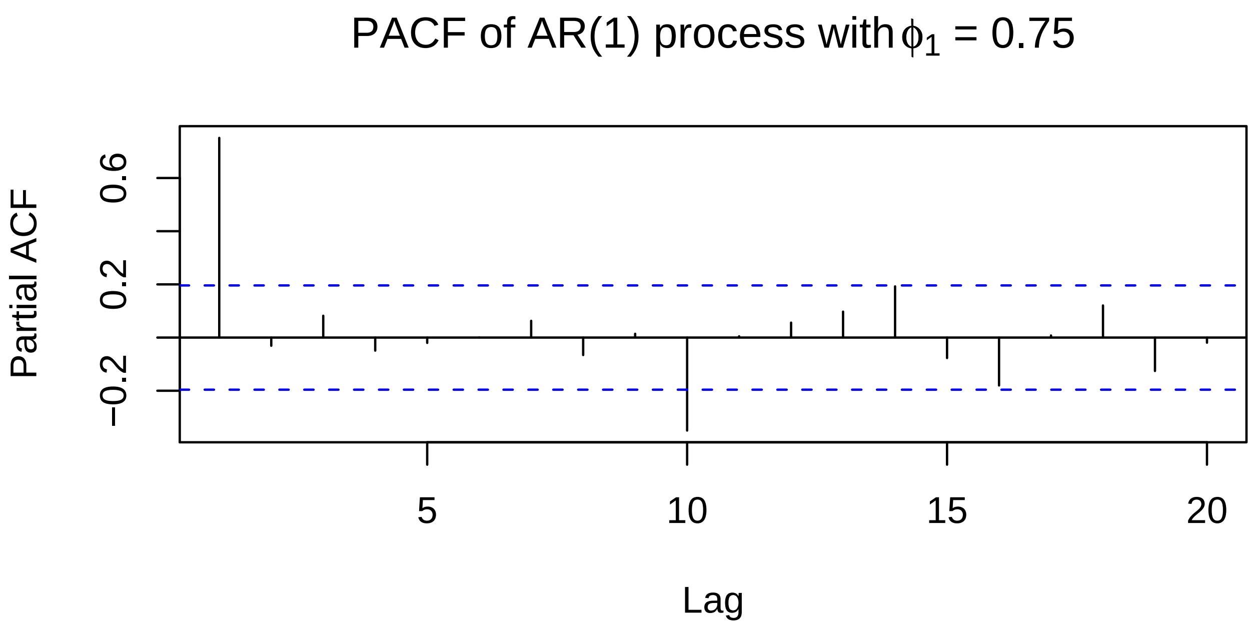
Simulated AR(1) process with $\phi_1 = 0.75$

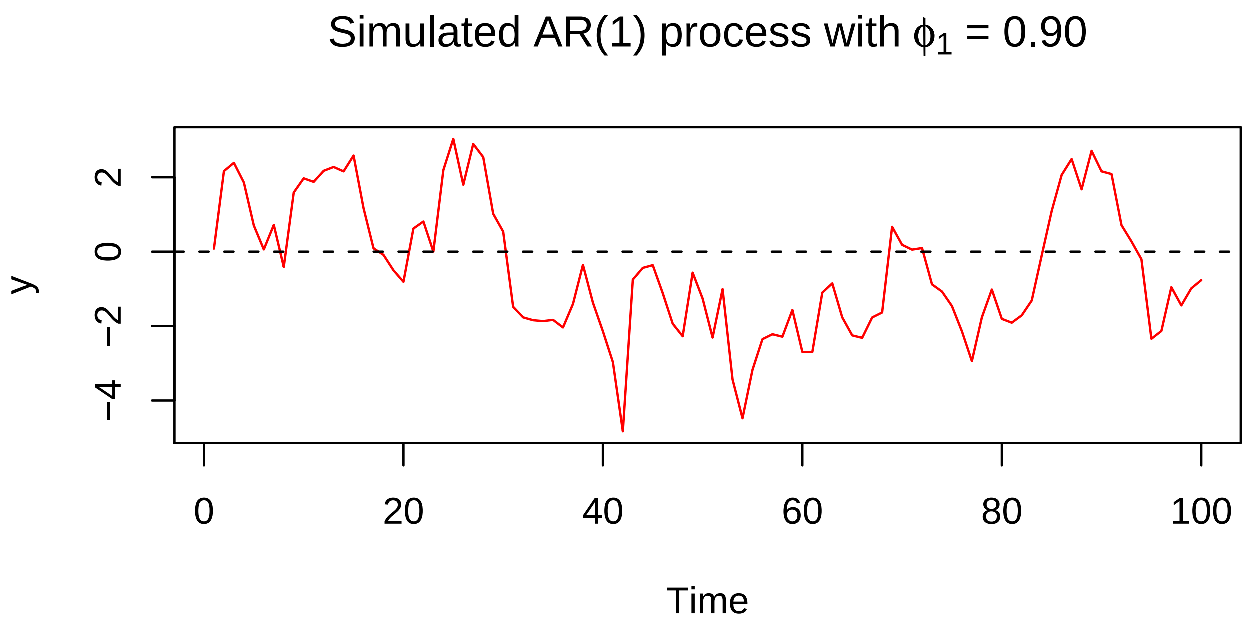


ACF of AR(1) process with $\phi_1 = 0.75$

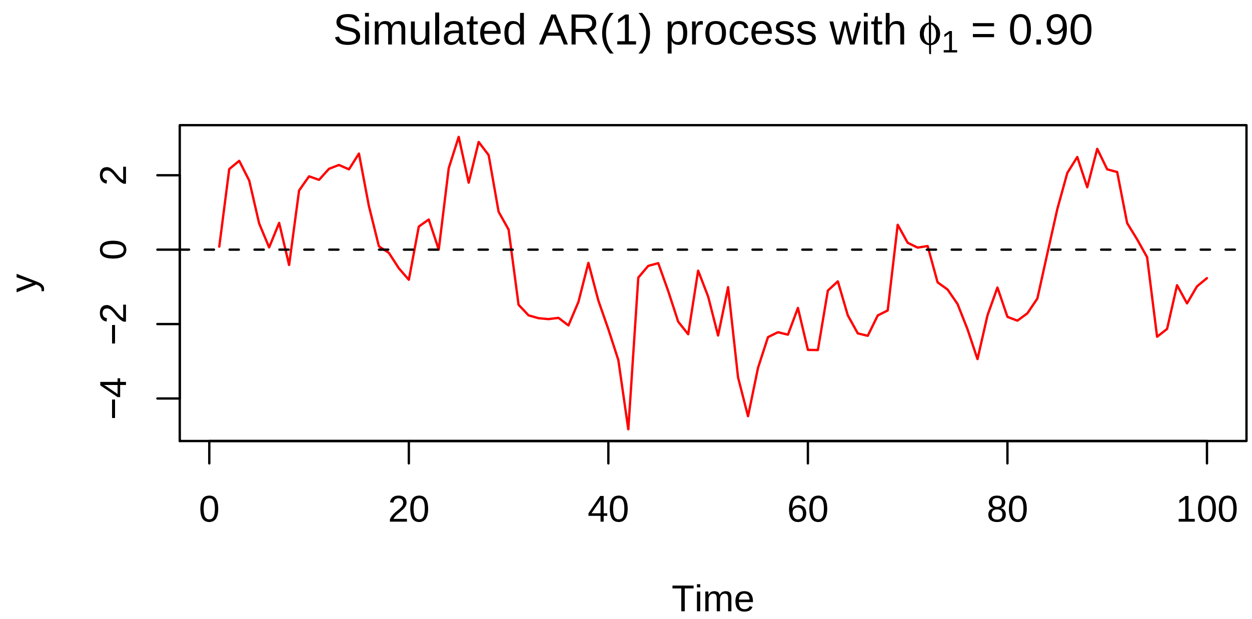


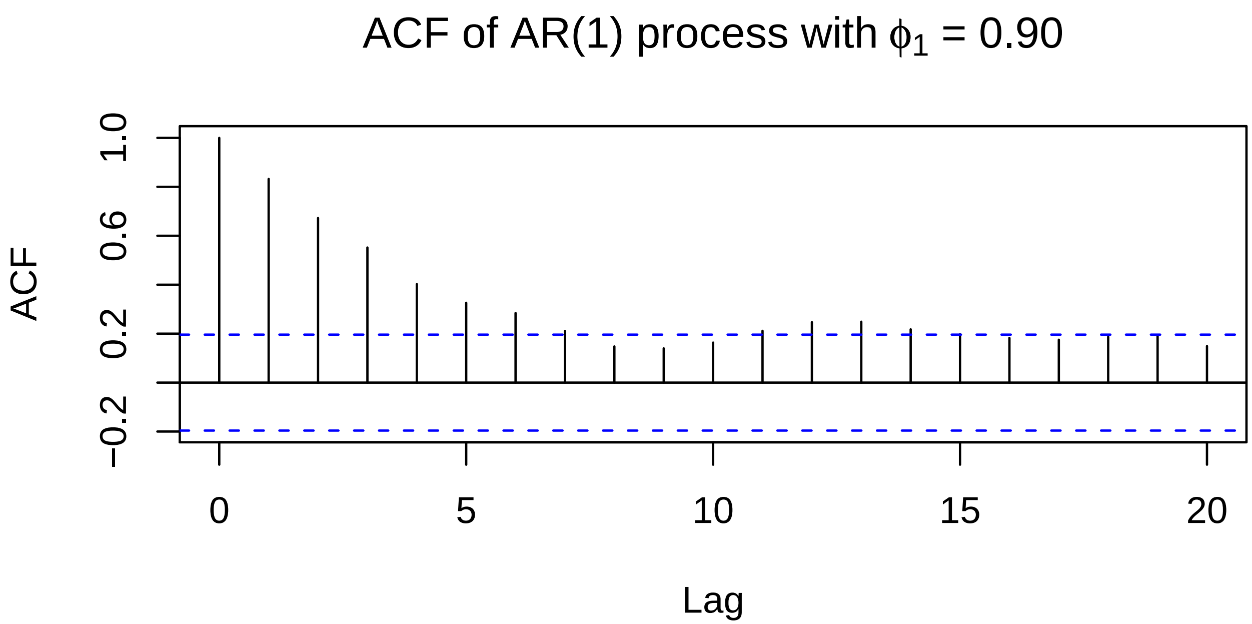

Simulated AR(1) process with ϕ_1 = 0.75
y
−3
−1
1
3
0
20
40
60
80
100
Time

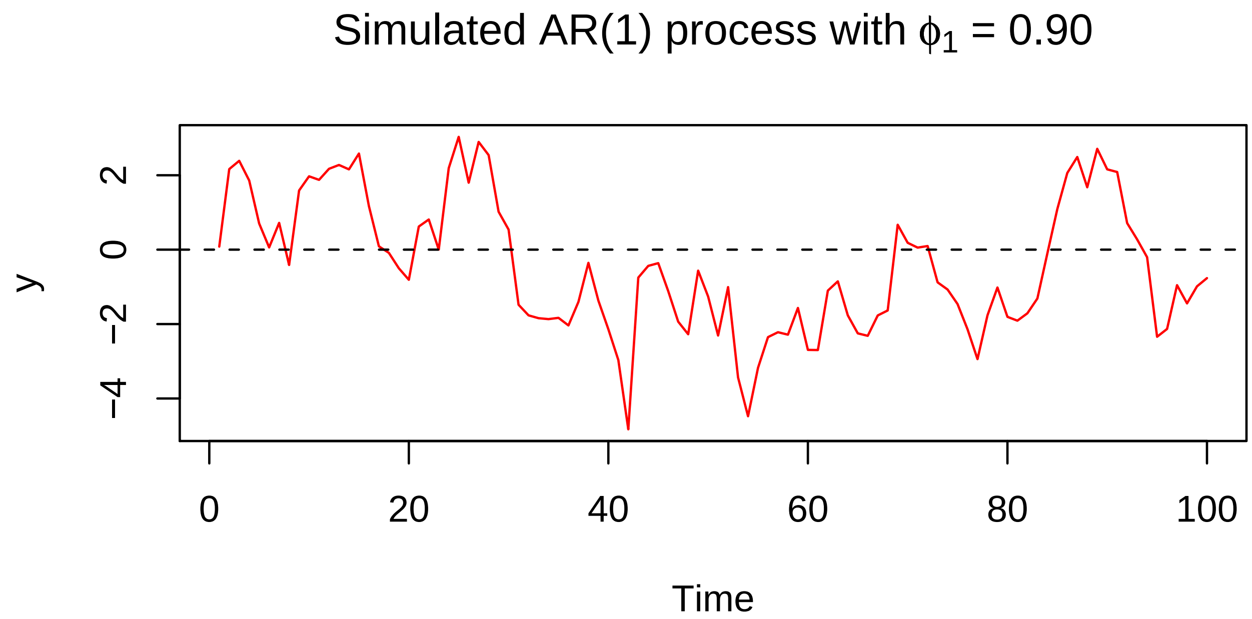

PACF of AR(1) process with ϕ_1 = 0.75
Partial ACF
−0.2
0.2
0.6
5
10
15
20
Lag



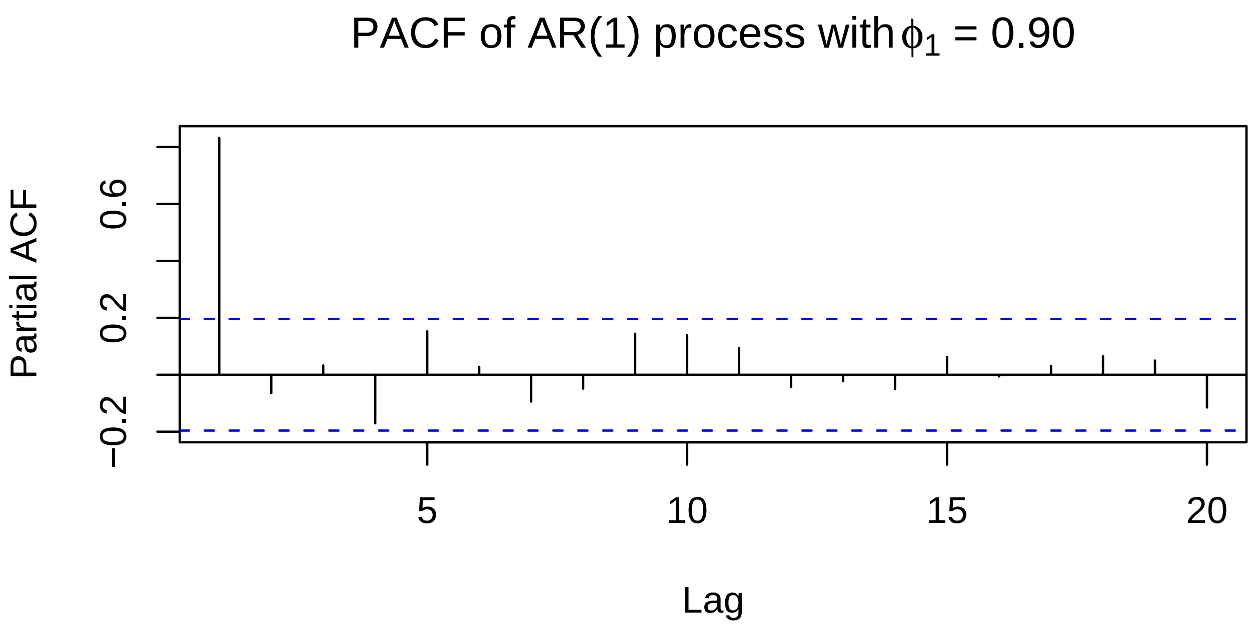
Simulated AR(1) process with $\phi_1 = 0.90$
y
2
0
−2
−4
0
20
40
60
80
100
Time


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

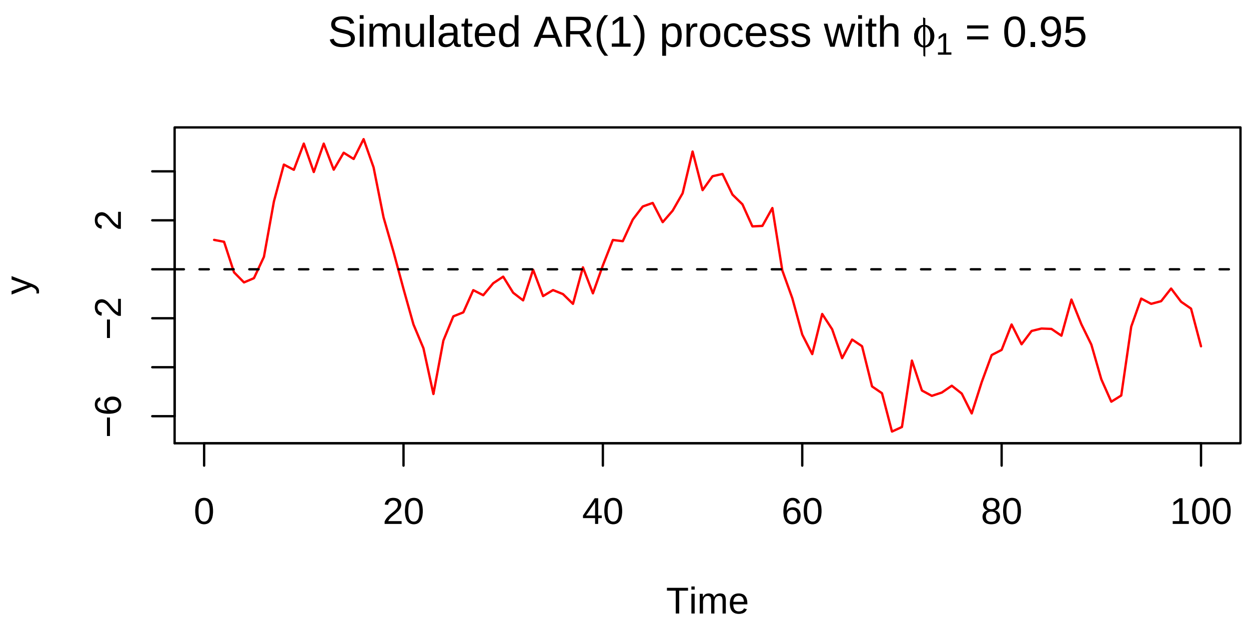

ACF of AR(1) process with ϕ_1 = 0.90
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag

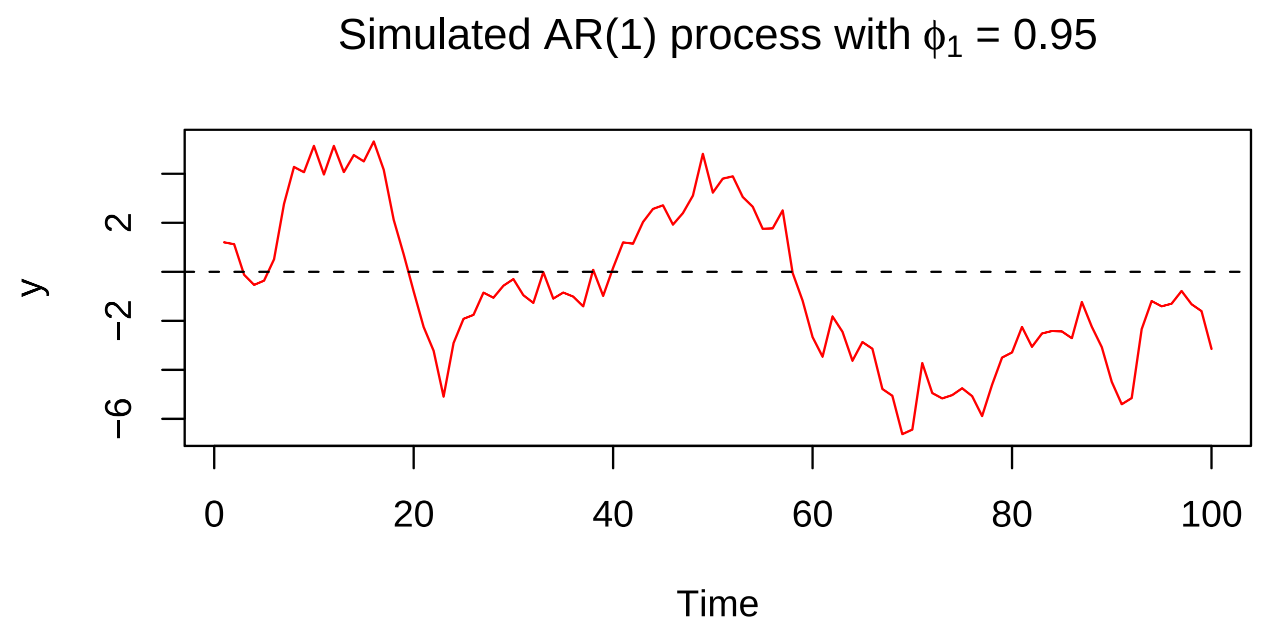


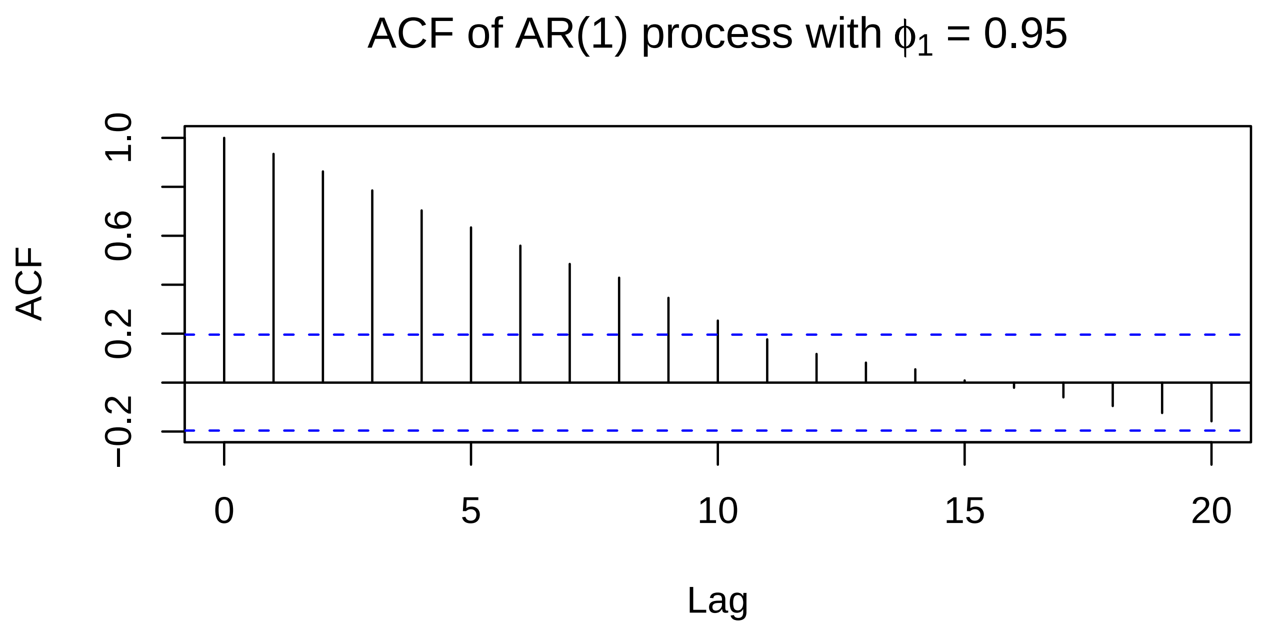
Simulated AR(1) process with ϕ_1 = 0.90
y
−4
−2
0
2
0
20
40
60
80
100
Time



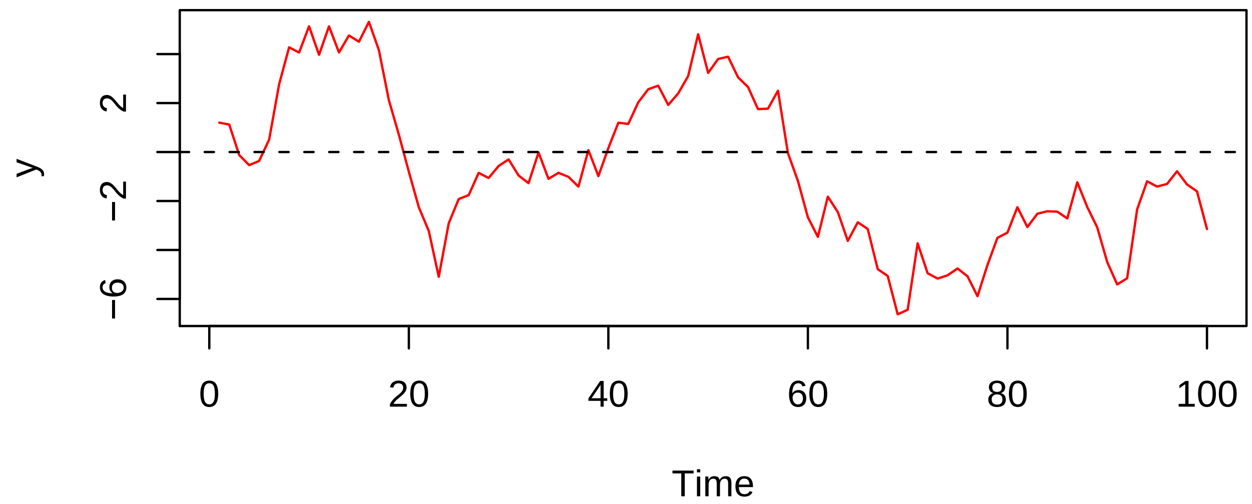
PACF of AR(1) process with ϕ_1 = 0.90
Partial ACF
−0.2
0.2
0.6
5
10
15
20
Lag


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

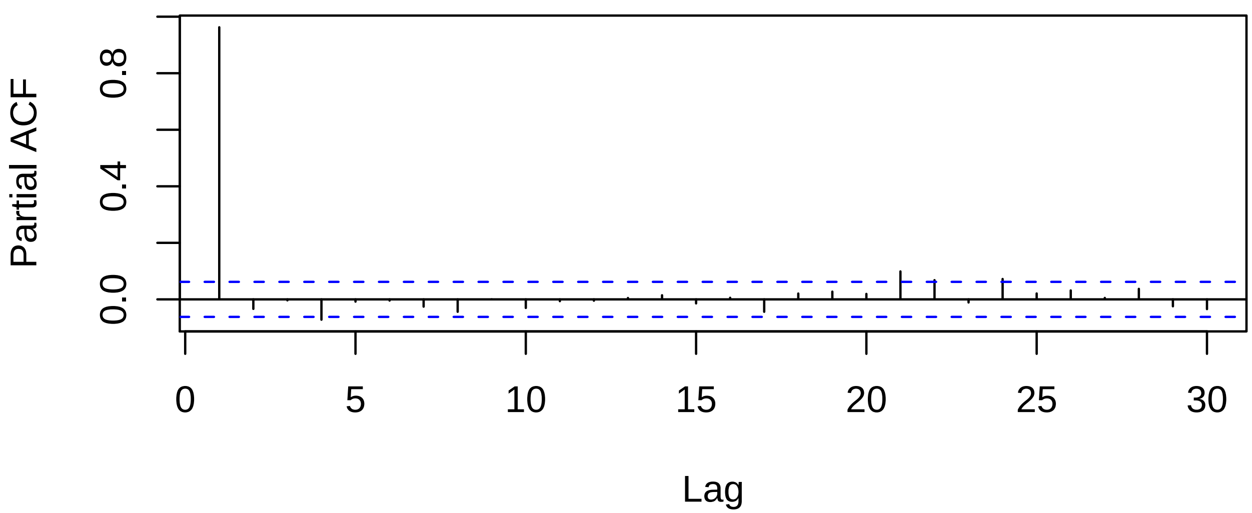

Simulated AR(1) process with $\phi_1 = 0.95$
y
−6
−2
2
0
20
40
60
80
100
Time

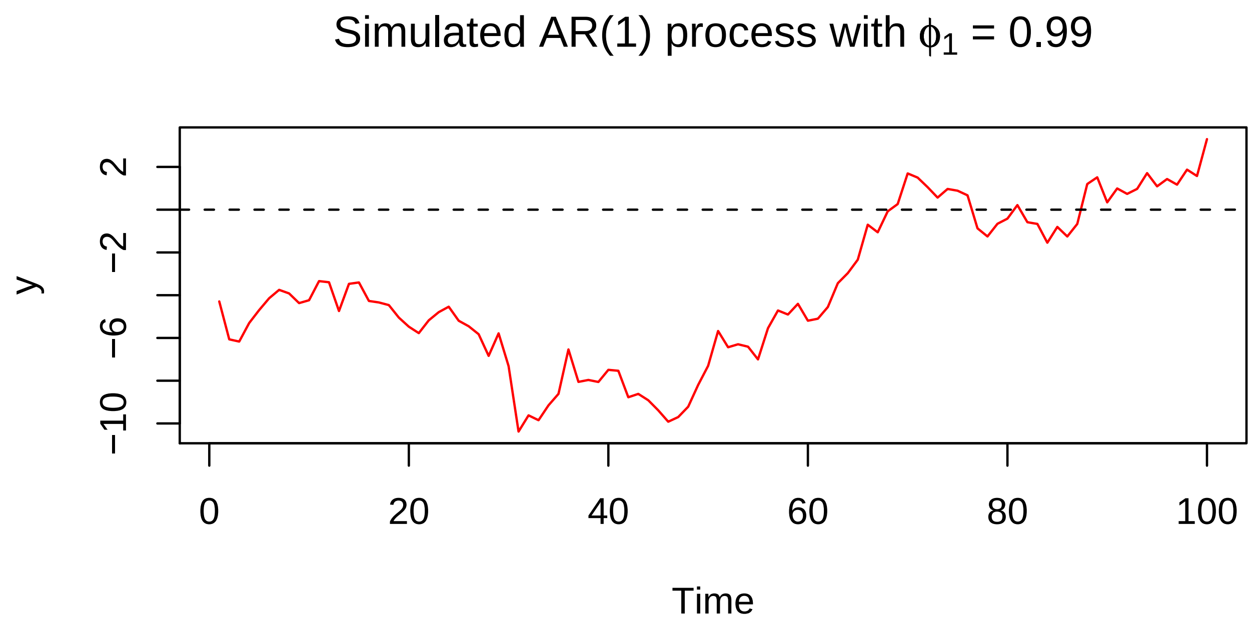

ACF of AR(1) process with $\phi_1 = 0.95$
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
20
Lag

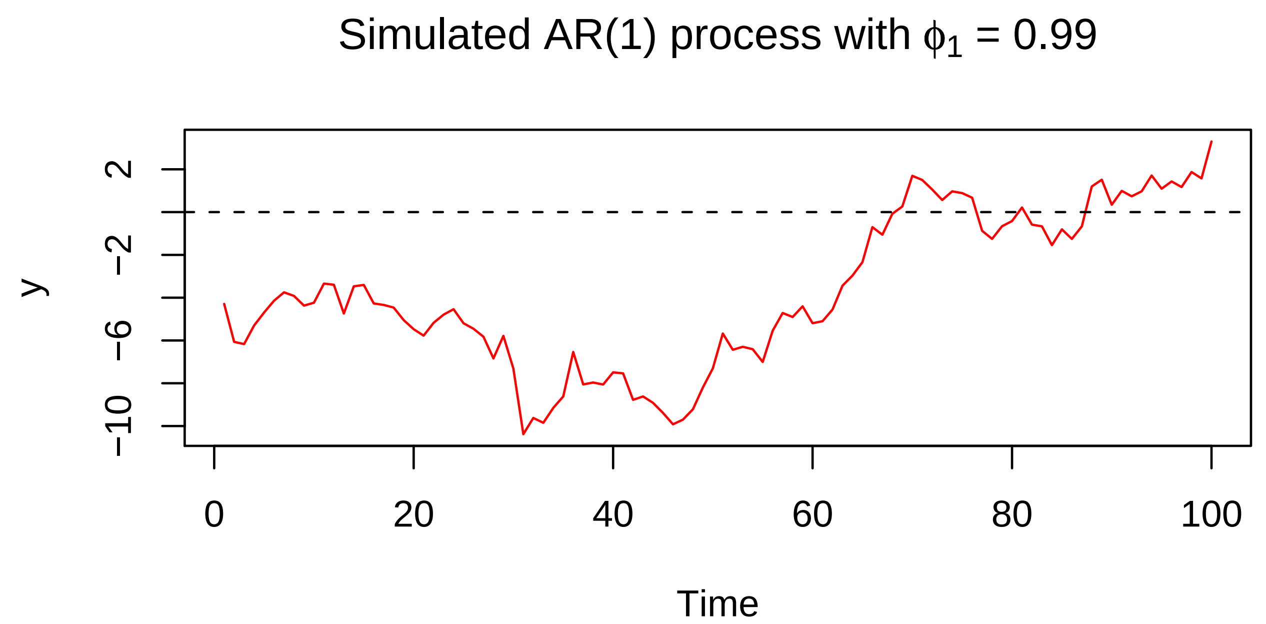
Simulated AR(1) process with $\phi_1 = 0.95$

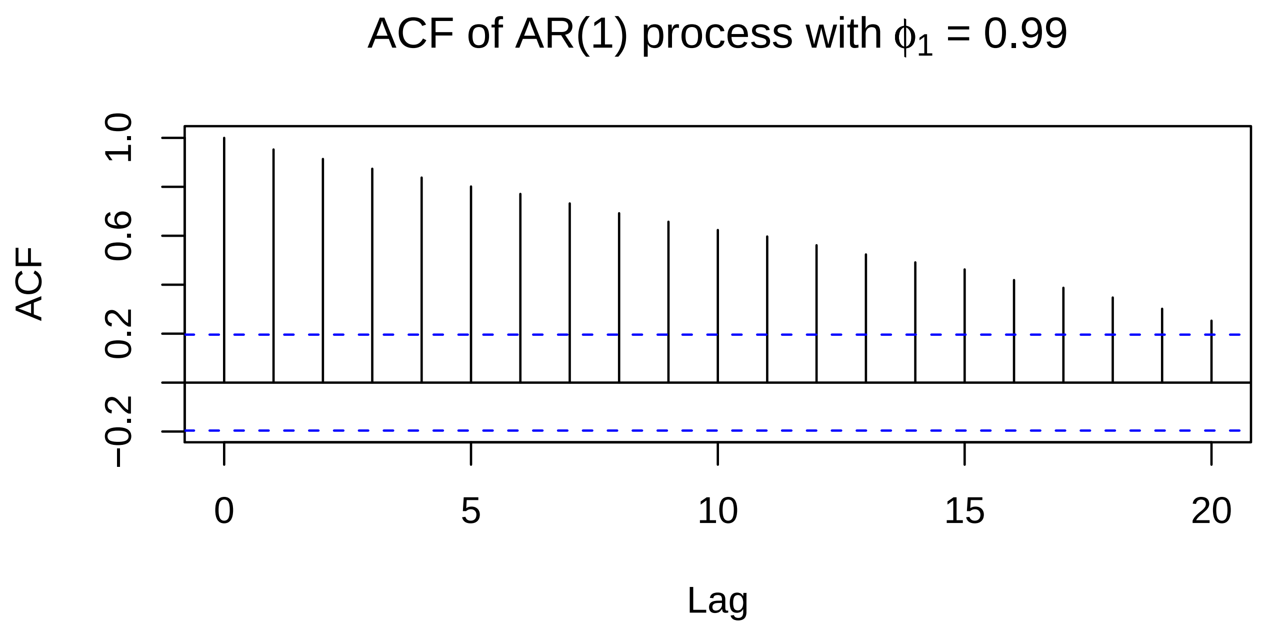


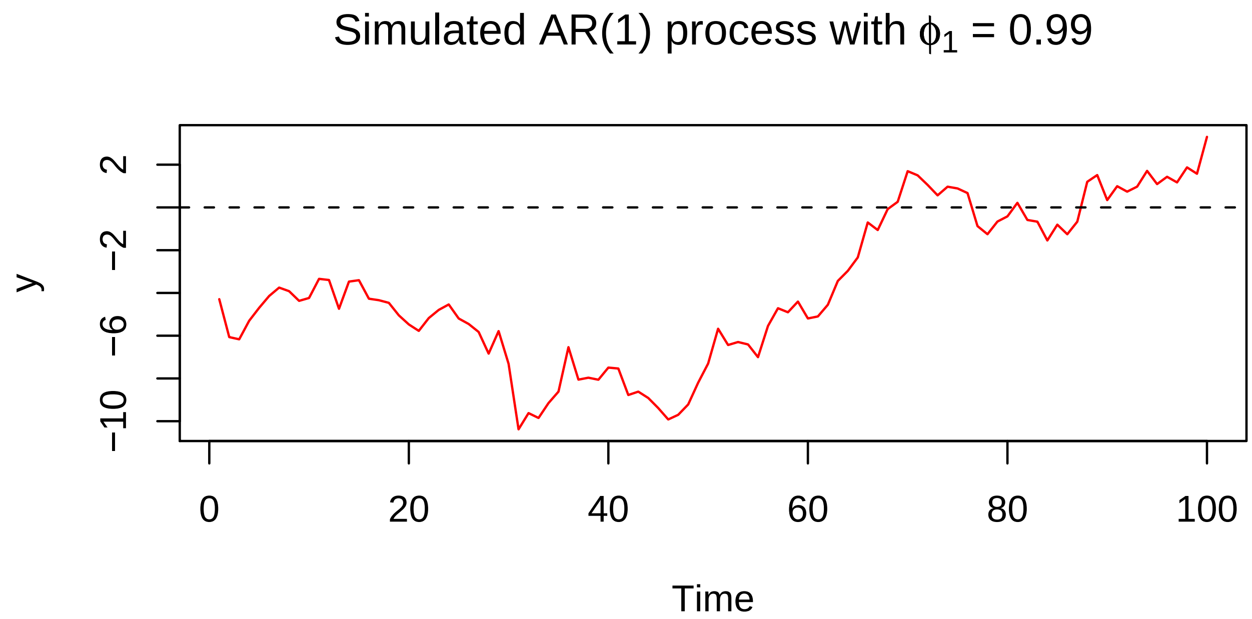
PACF of AR(1) process with $\phi_1 = 0.95$

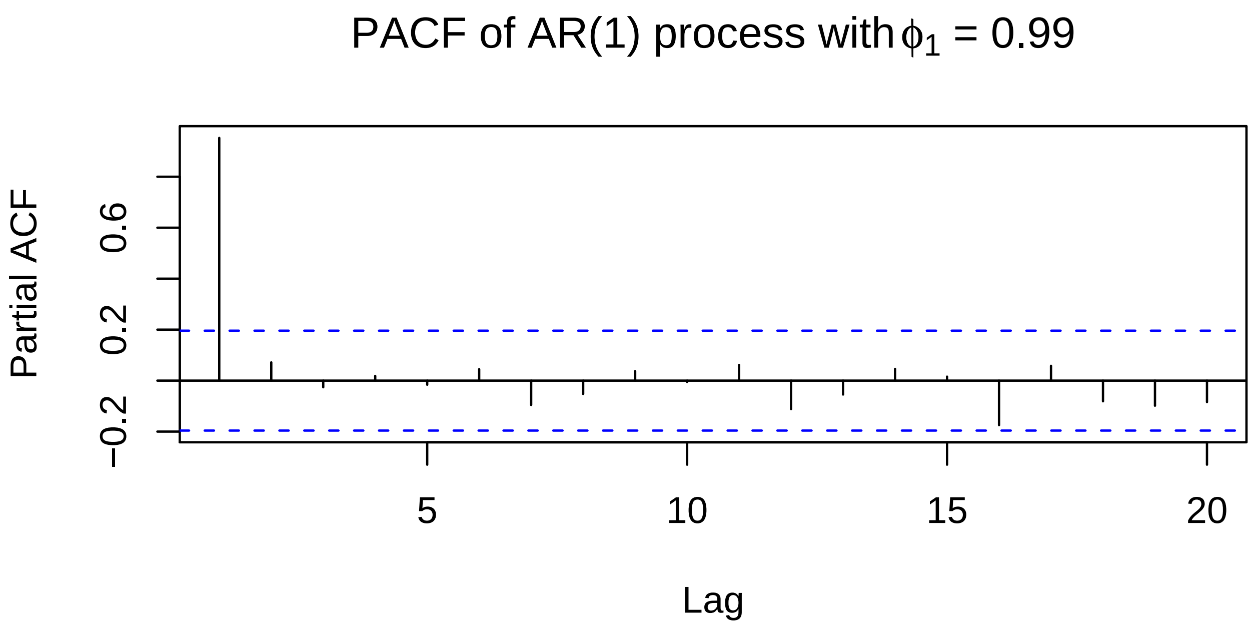



Simulated AR(1) process with ϕ_1 = 0.99
y
2
−2
−6
−10
0
20
40
60
80
100
Time


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

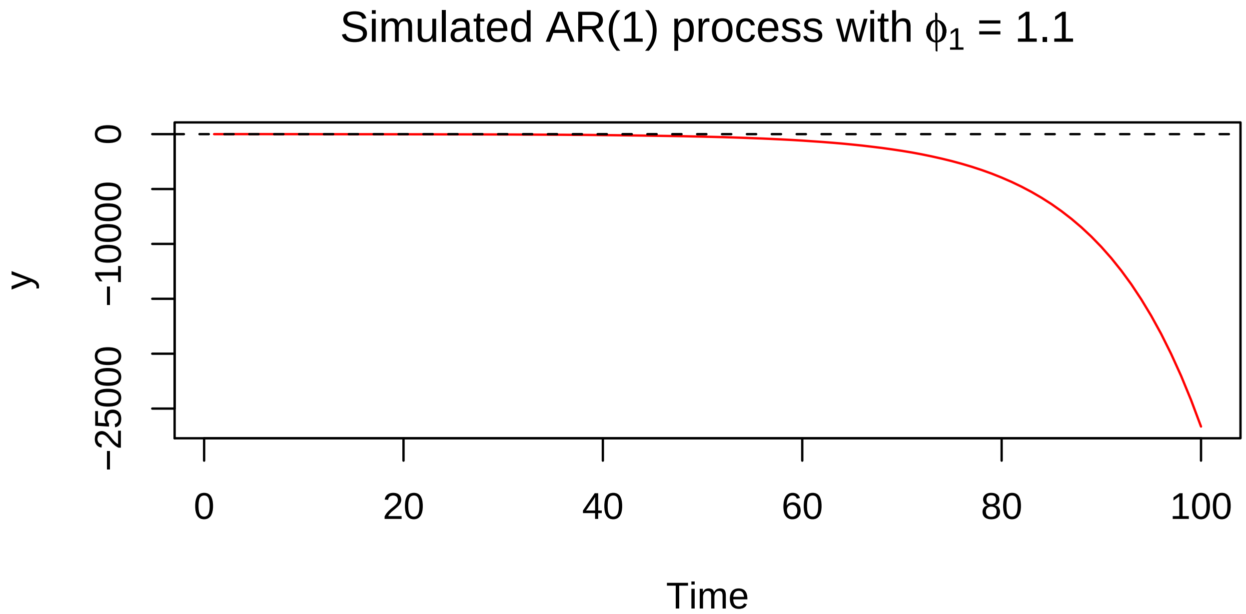

ACF of AR(1) process with ϕ_1 = 0.99
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag


Simulated AR(1) process with ϕ1 = 0.99
y
2
−2
−6
−10
0
20
40
60
80
100
Time

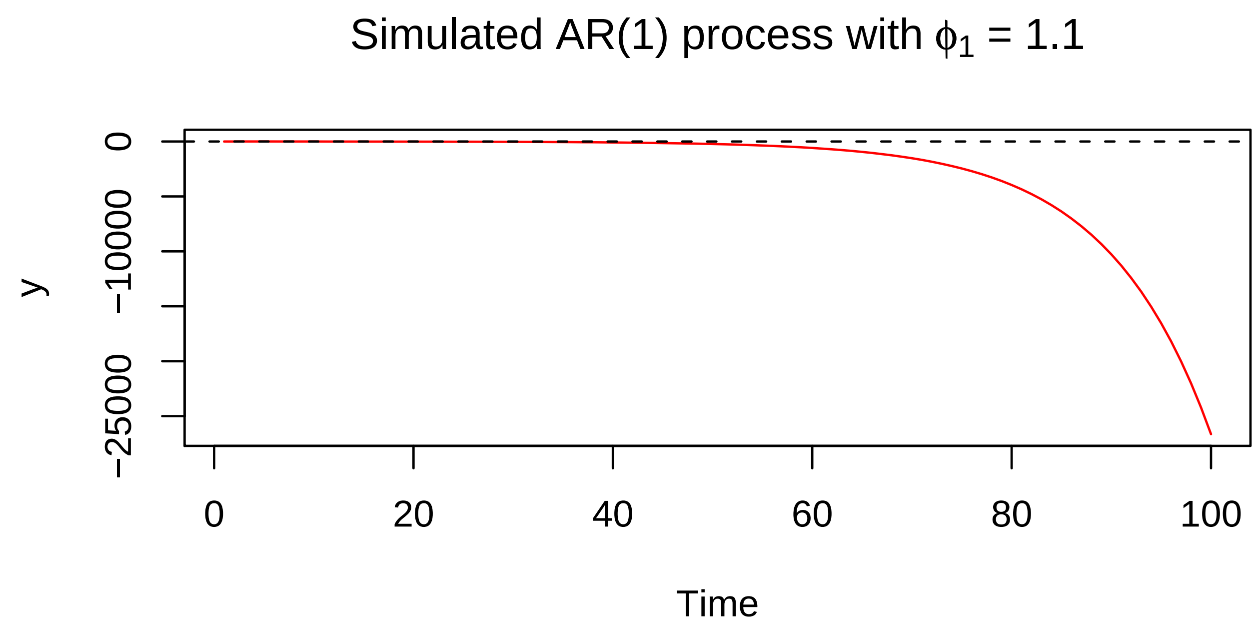

PACF of AR(1) process with ϕ1 = 0.99
Partial ACF
0.6
0.2
−0.2
5
10
15
20
Lag

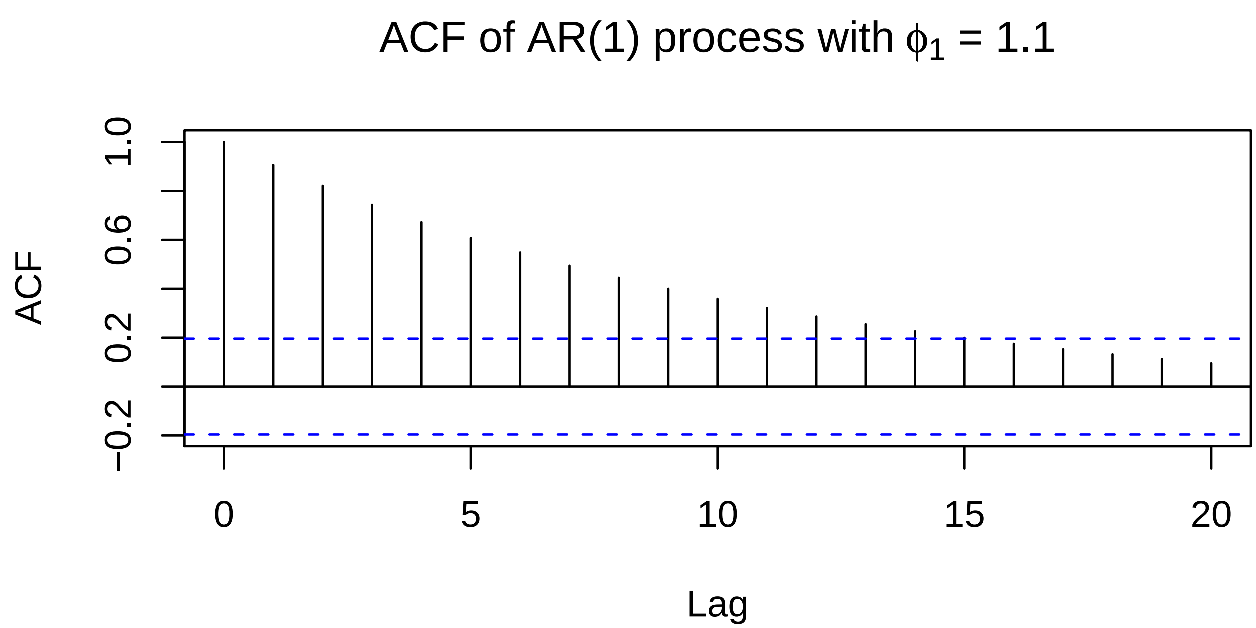
Observations about AR(1) when $|\phi_1| < 1$

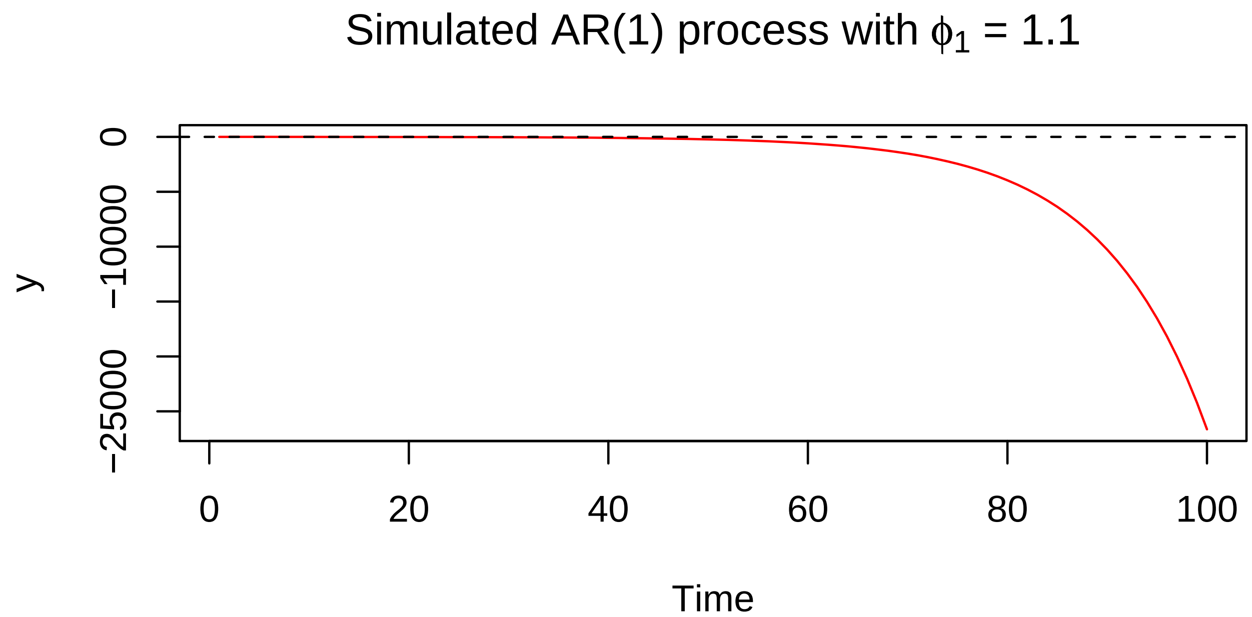
- As $|\phi_1| \to 0$, series reverts to its mean at 0 quickly.
- As $|\phi_1| \to 1$, series takes longer to revert to mean.
- Still gets there eventually. (Even for $\phi_1 = 0.99$?)
- ACF appears to gradually decline in the lag, as expected
- ACF decays more slowly as $|\phi_1| \to 1$
- But for all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

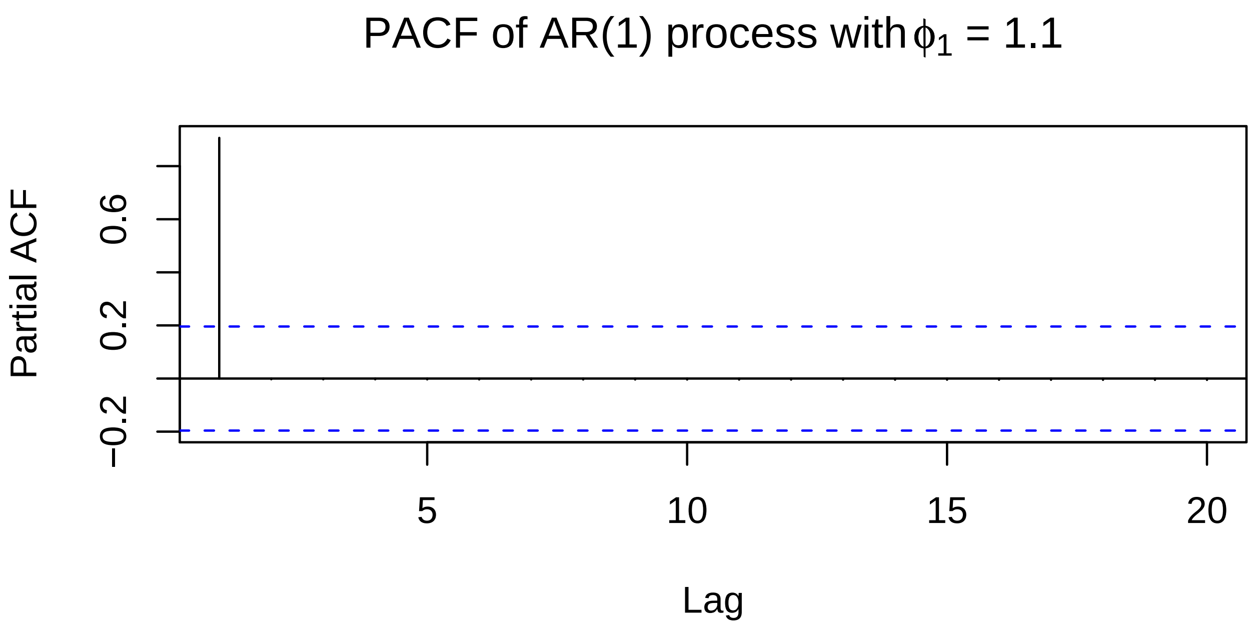


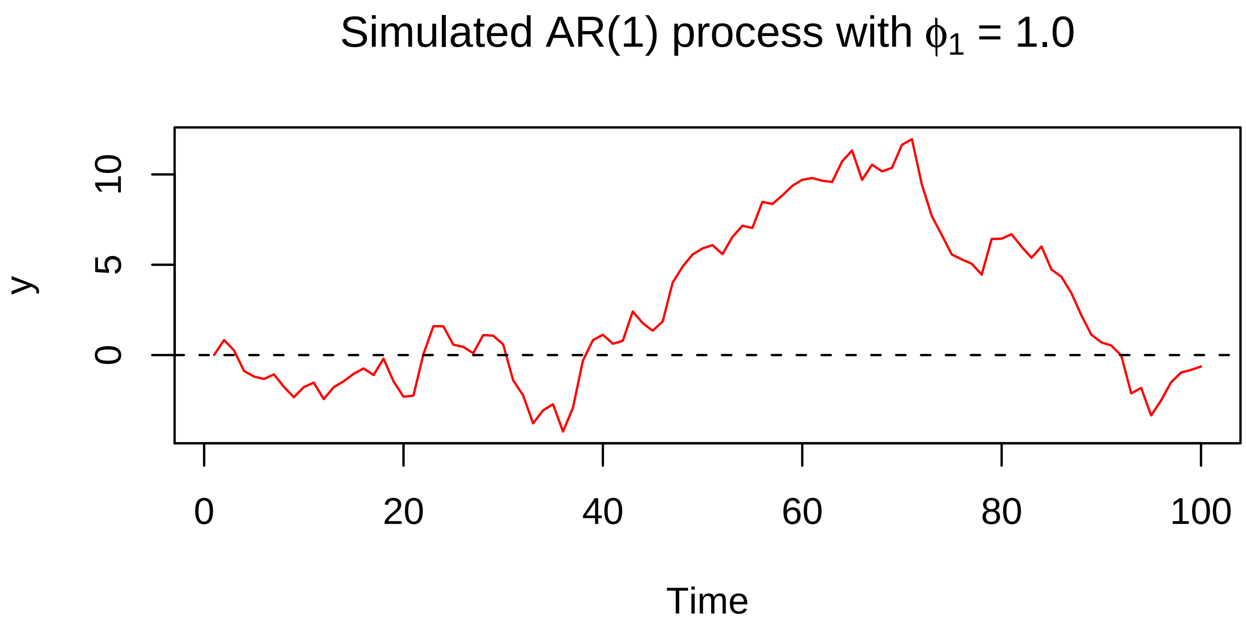
Simulated AR(1) process with $\phi_1 = 1.1$
y
0
−10000
−25000
0
20
40
60
80
100
Time

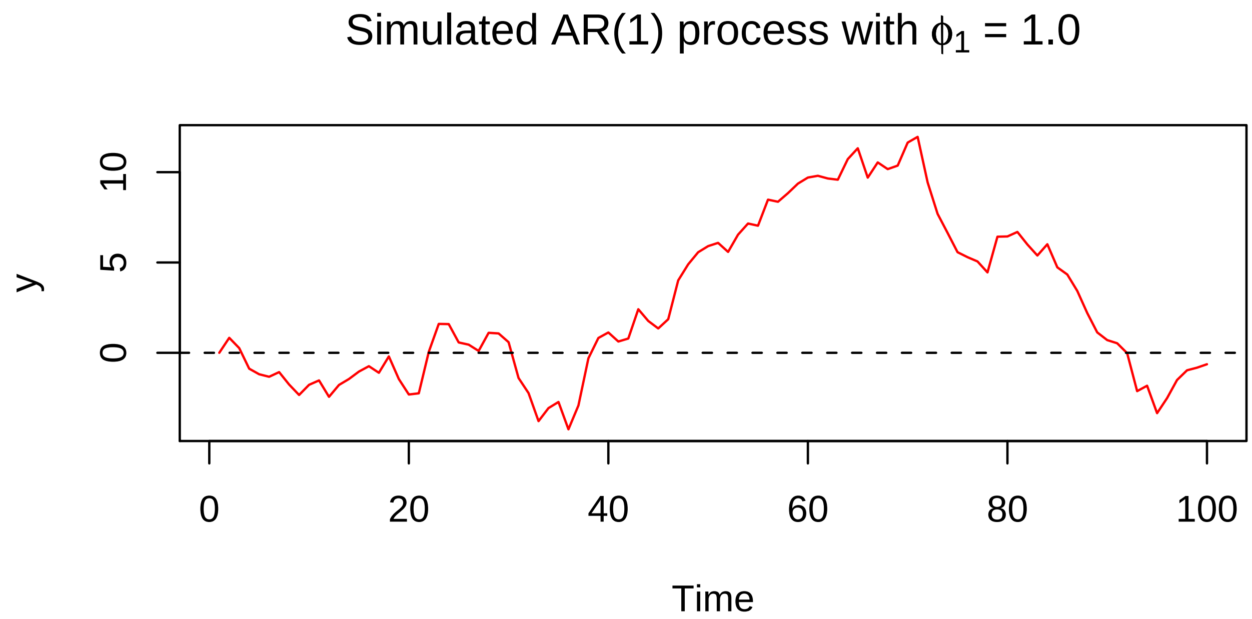

Simulated AR(1) process with $\phi_1 = 1.1$
0
−10000
−25000
y
0
20
40
60
80
100
Time

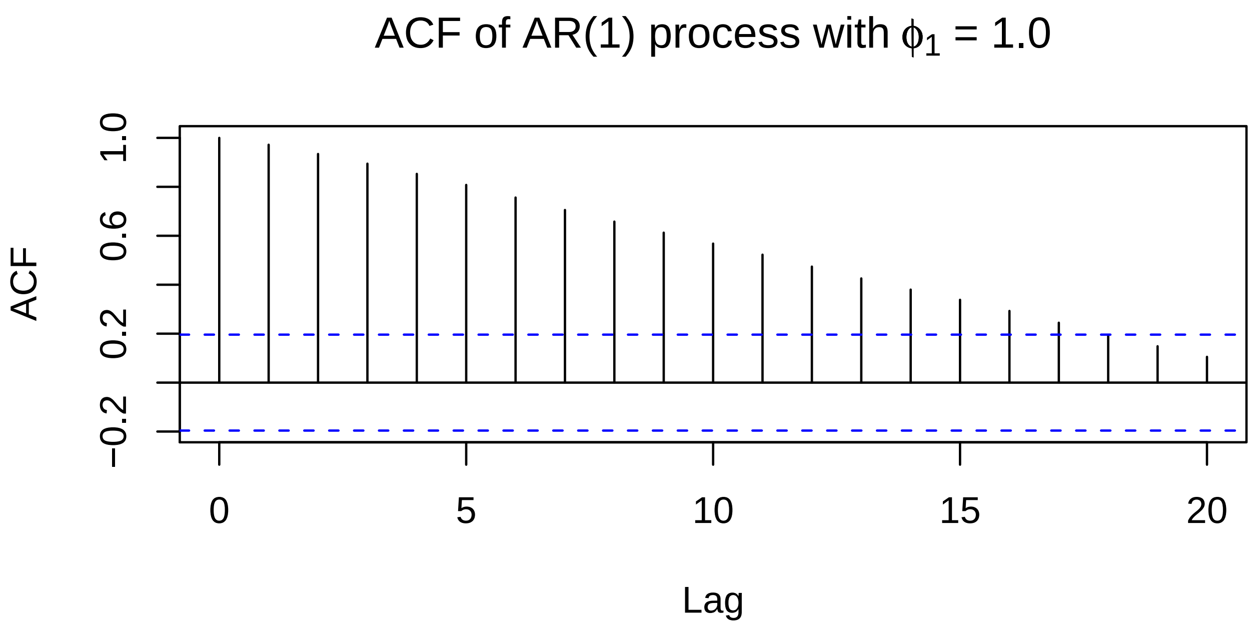

ACF of AR(1) process with $\phi_1 = 1.1$
1.0
0.6
0.2
−0.2
ACF
0
5
10
15
20
Lag


Simulated AR(1) process with ϕ_1 = 1.1
y
0
−10000
−25000
0
20
40
60
80
100
Time

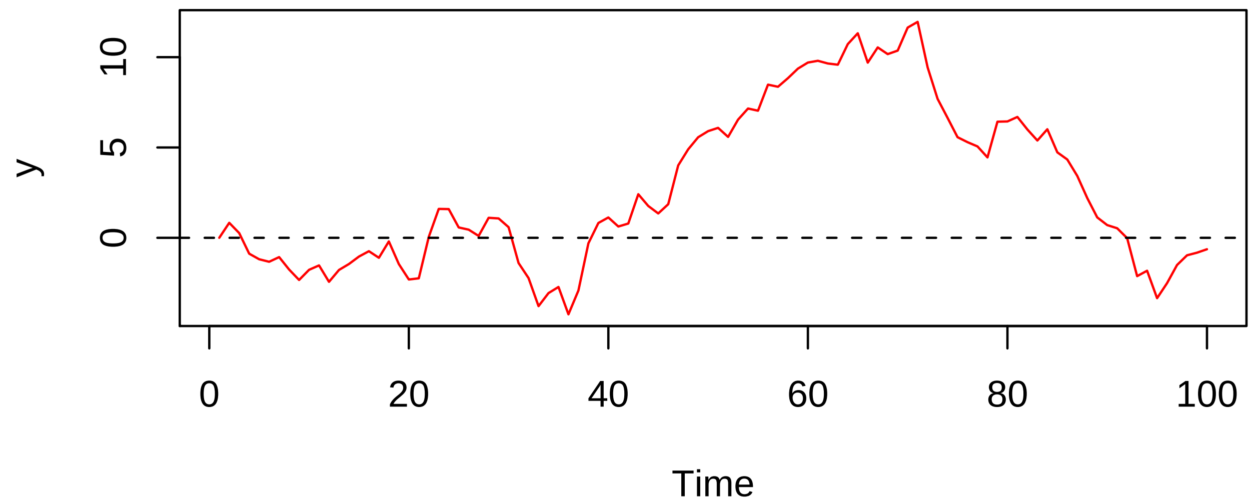

PACF of AR(1) process with ϕ_1 = 1.1
Partial ACF
0.6
0.2
−0.2
5
10
15
20
Lag


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

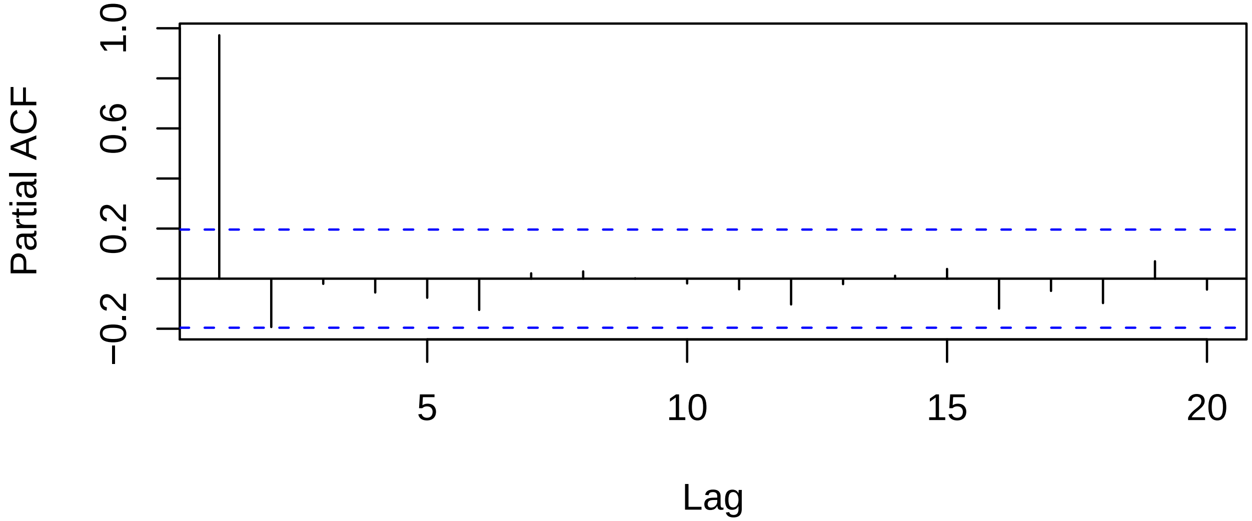

Simulated AR(1) process with ϕ1 = 1.0
y
0
5
10
0
20
40
60
80
100
Time

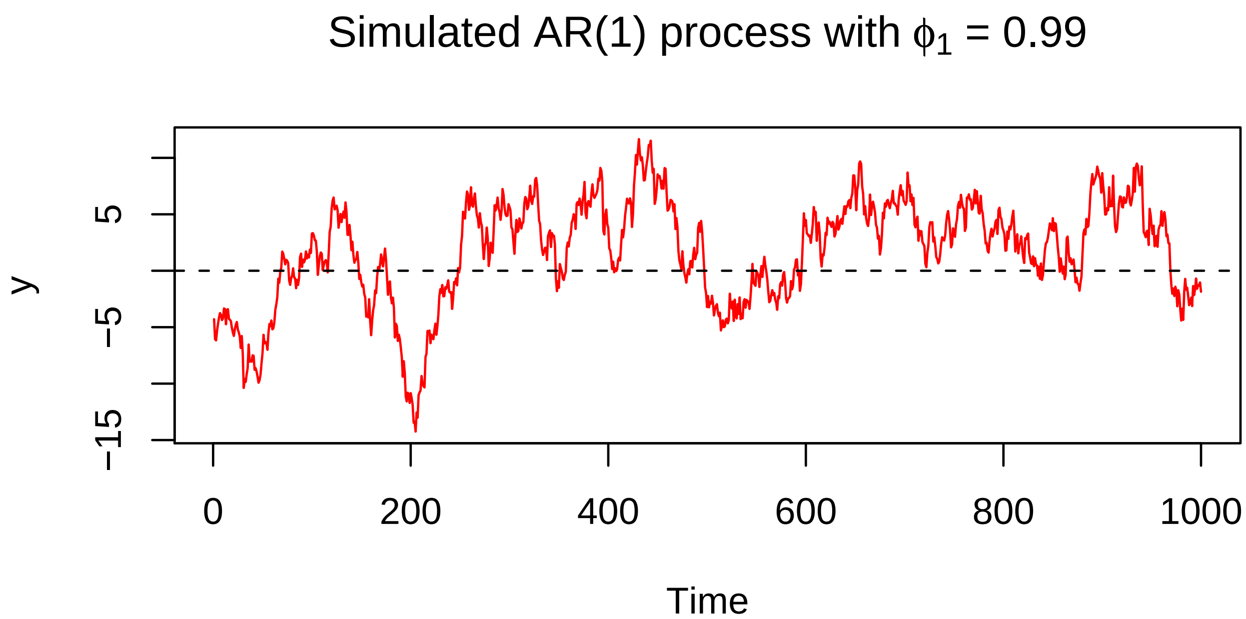

ACF of AR(1) process with ϕ1 = 1.0
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
20
Lag

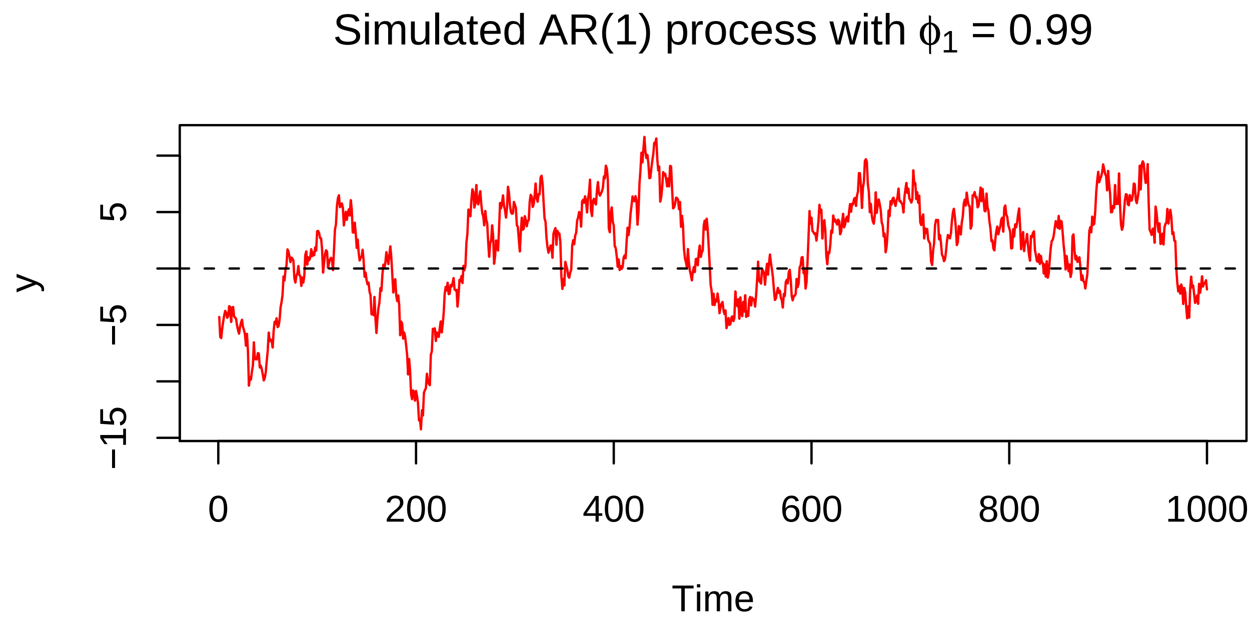
Simulated AR(1) process with $\phi_1 = 1.0$



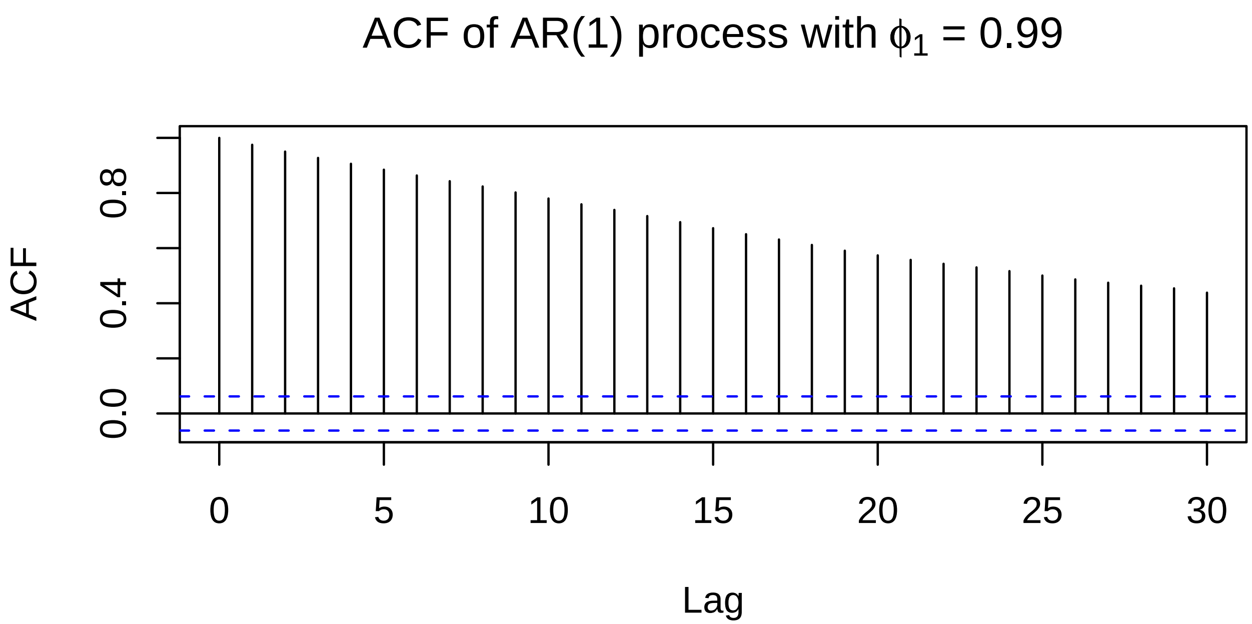
PACF of AR(1) process with $\phi_1 = 1.0$



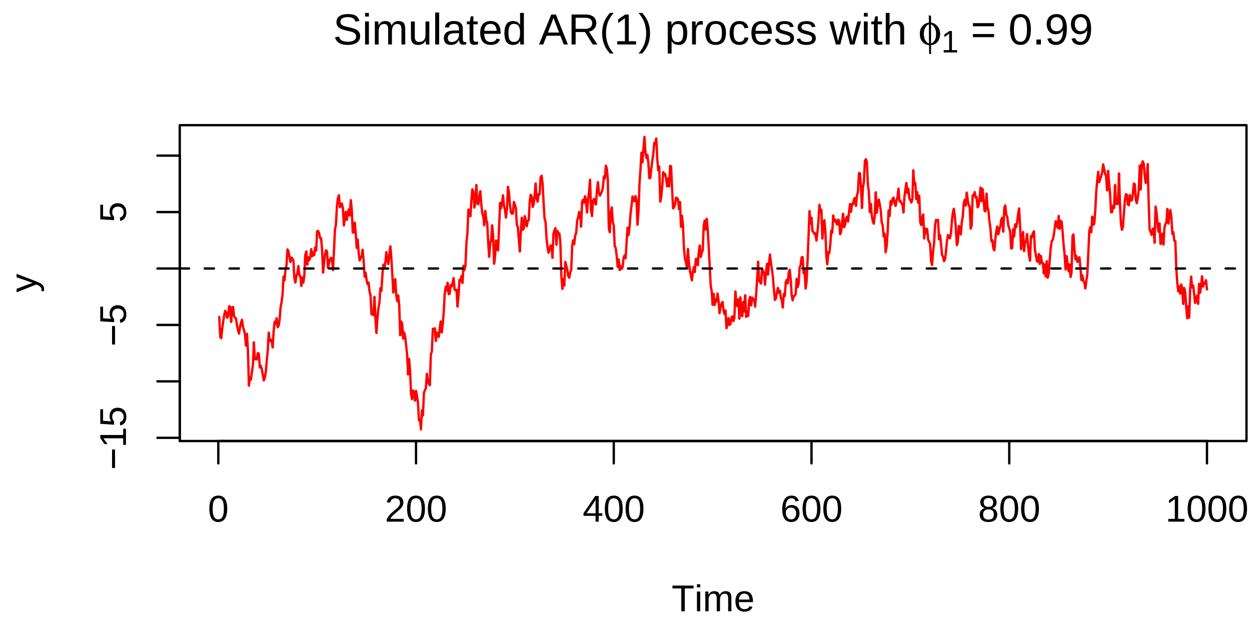

Simulated AR(1) process with ϕ1 = 0.99
y
5
−5
−15
0
200
400
600
800
1000
Time



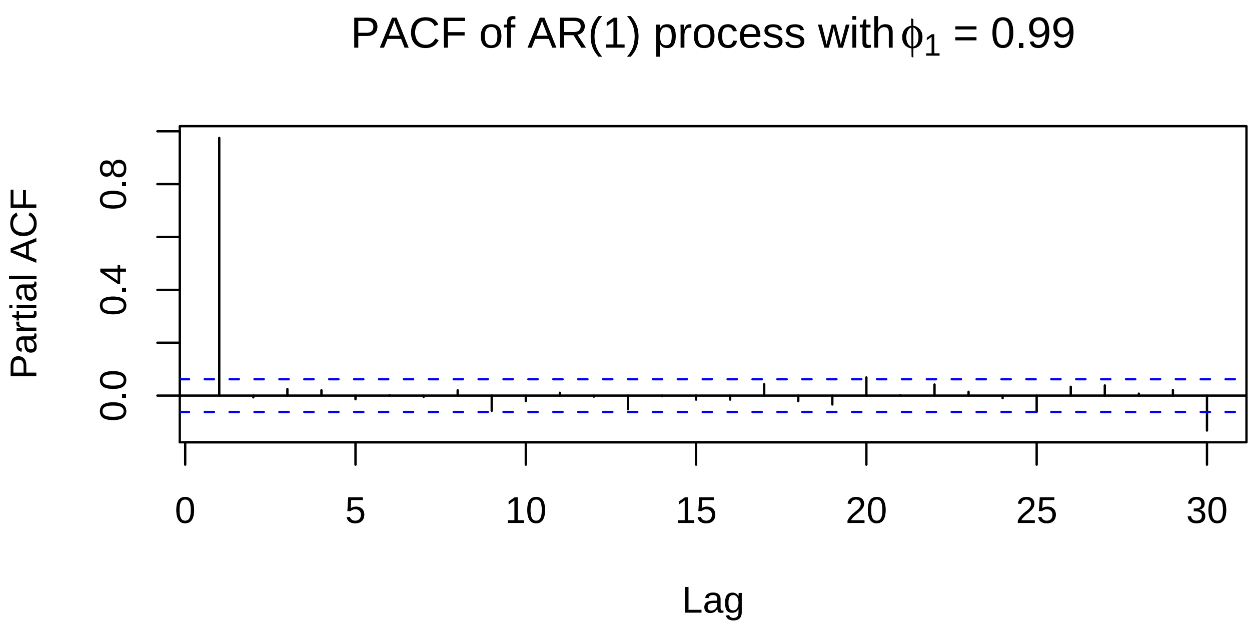
Simulated AR(1) process with ϕ_1 = 0.99
y
5
−5
−15
0
200
400
600
800
1000
Time



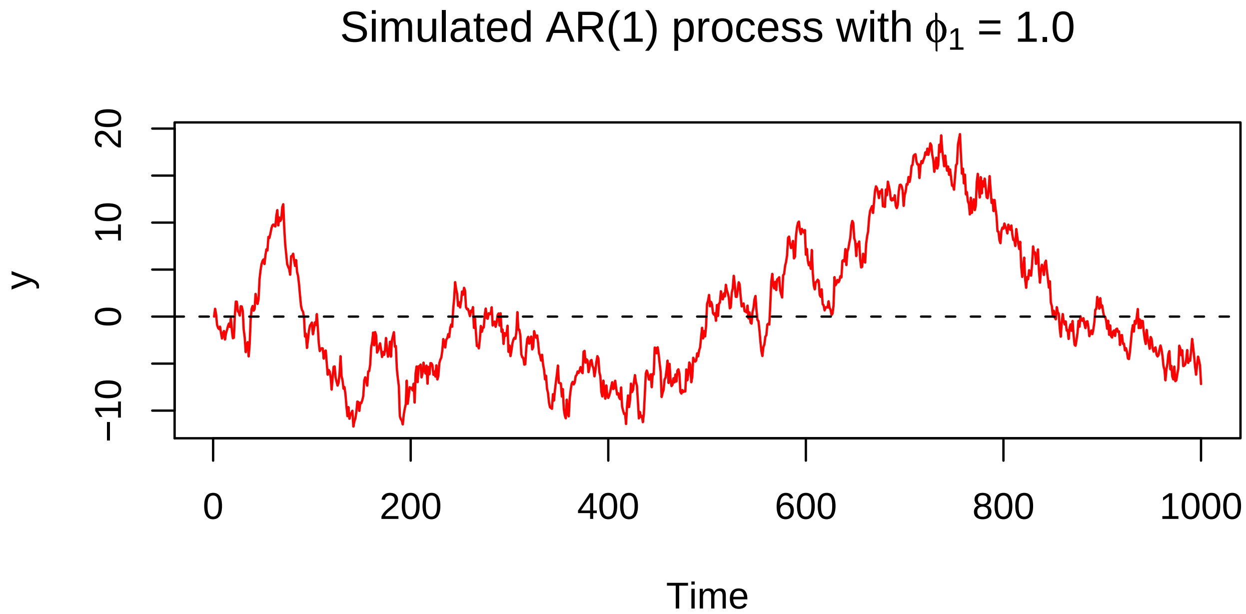
ACF of AR(1) process with ϕ_1 = 0.99
ACF
0.8
0.4
0.0
0
5
10
15
20
25
30
Lag



Simulated AR(1) process with ϕ_1 = 0.99
y
5
−5
−15
0
200
400
600
800
1000
Time

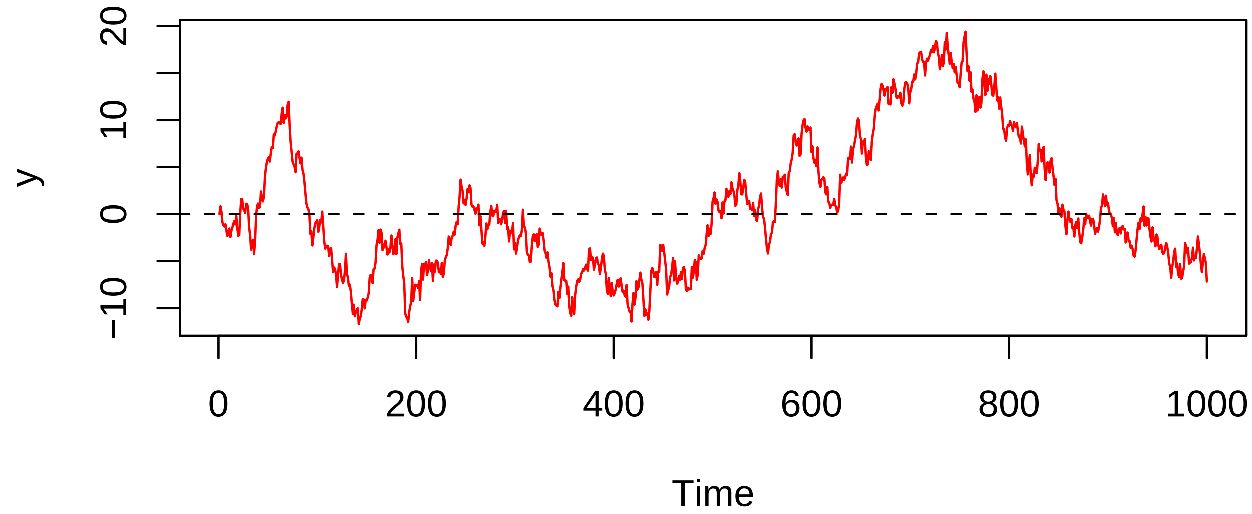


PACF of AR(1) process with ϕ_1 = 0.99
Partial ACF
0.0
0.4
0.8
0
5
10
15
20
25
30
Lag

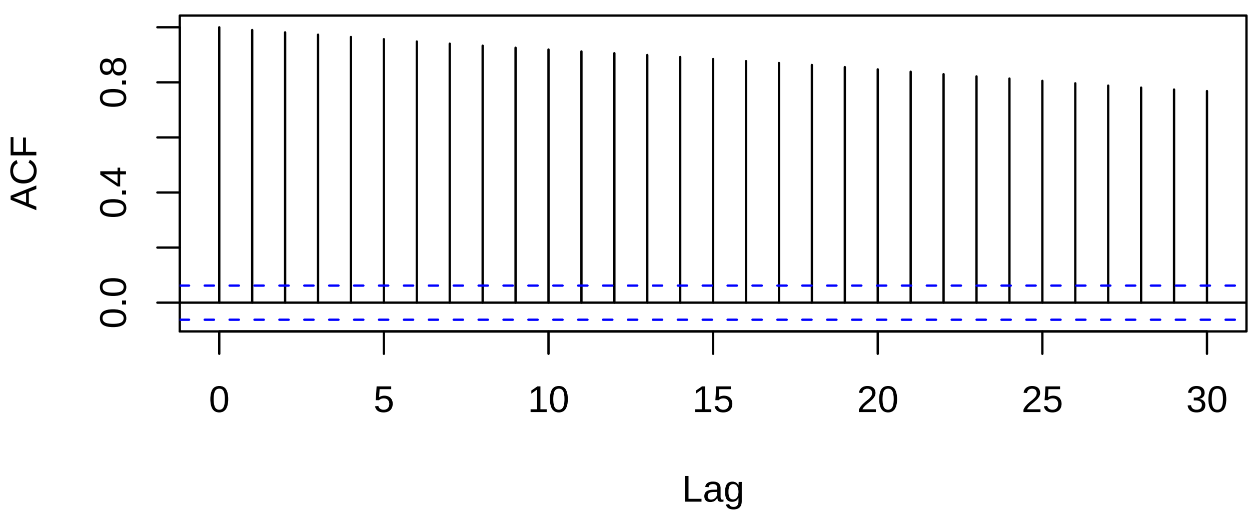


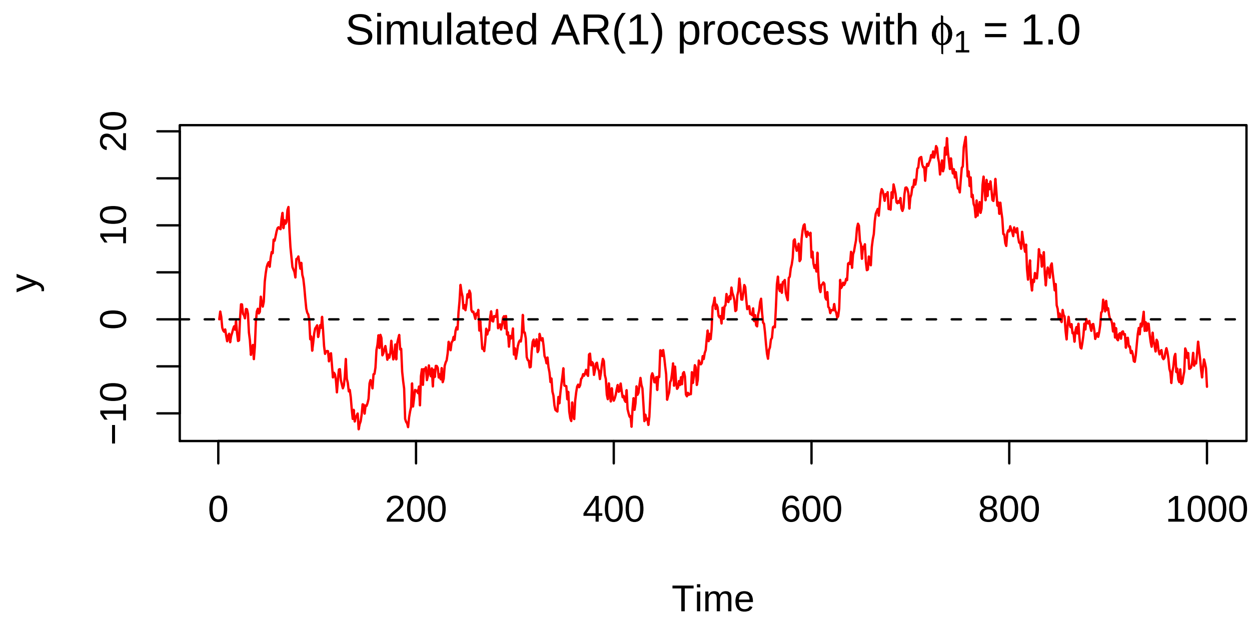
Simulated AR(1) process with $\phi_1 = 1.0$
y
20
10
0
−10
0
200
400
600
800
1000
Time

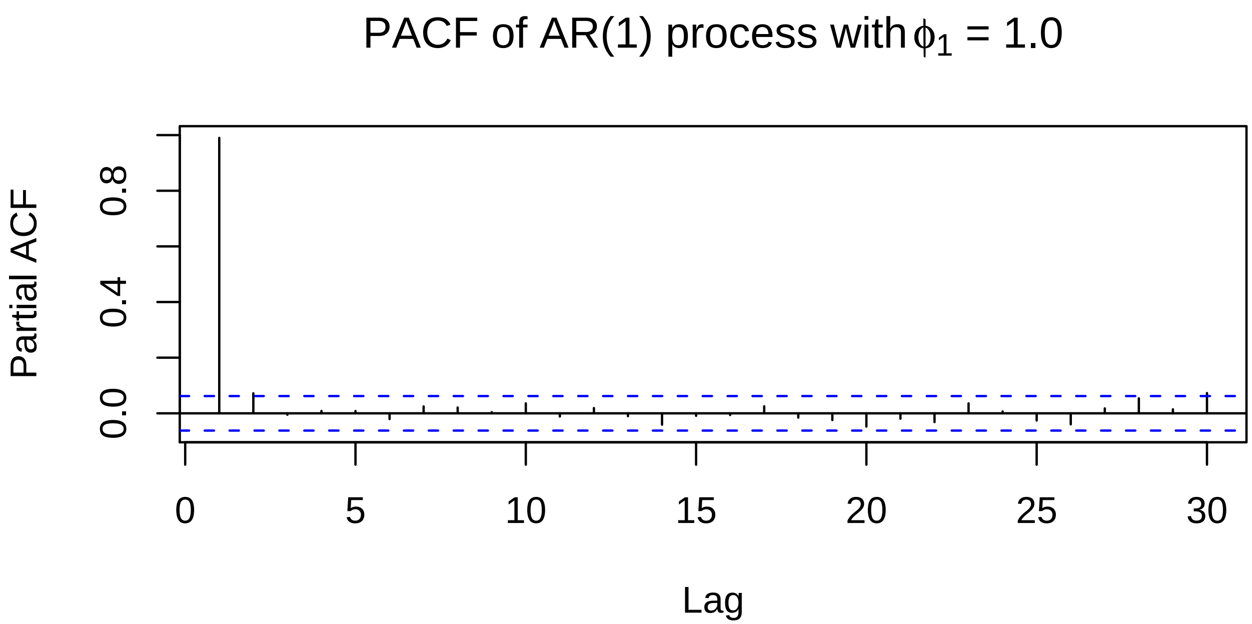
Simulated AR(1) process with $\phi_1 = 1.0$



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




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


PACF of AR(1) process with ϕ_1 = 1.0
Partial ACF
0.0
0.4
0.8
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

Recall we are abstracting from deterministic trends and other covariates

As such, we might be working with a detrended series like

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

Although y_t trends upward deterministically before detrending,
the rest of the components of $y_t - t\beta$ are mean-stationary

Time series that are stationary after detrending are *trend stationary*

Covariance stationarity

A time series is *covariance stationary* if neither the mean of y_t nor the covariance of y_t and y_{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 its sample moments converge in probability to the population moments

Put another way, in an ergodic time series,
sample mean and sample autocorrelation functions will tend to be the same as the mean and autocorrelation of the entire (conceivably infinite) series

Put still another way, for an ergodic process,
subsamples tends to have the same behavior (mean, variance, and autocorrelation) regardless of which section of the series is sampled

Formally, ergodicity implies that for a sample $\mathbf{y}$ taken from a time series with population mean μ and population autocorrelation γ_k,

$$\bar{\mathbf{y}} \xrightarrow{p} \mu \qquad \text{and} \qquad \text{cov}(y_t, y_{t+k}) \xrightarrow{p} \gamma_k$$

where $\xrightarrow{p}$ indicates convergence in probability

Stationarity time series are usually also ergodic,
unless μ or γ_k are random functions 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 ($\phi = 0.99$)

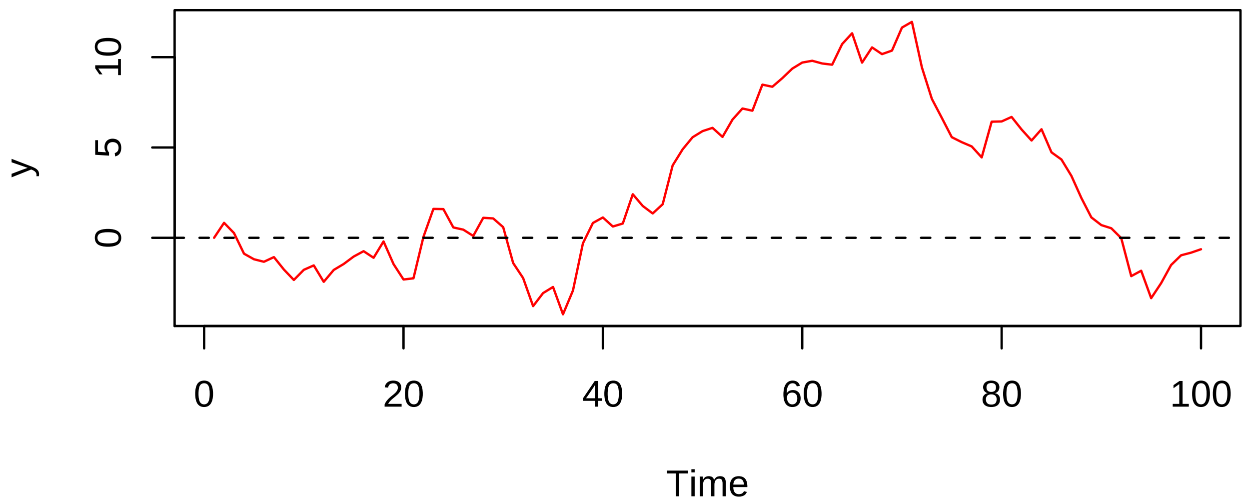
- Long-run forecasts are hard—don't tend towards any particular mean

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

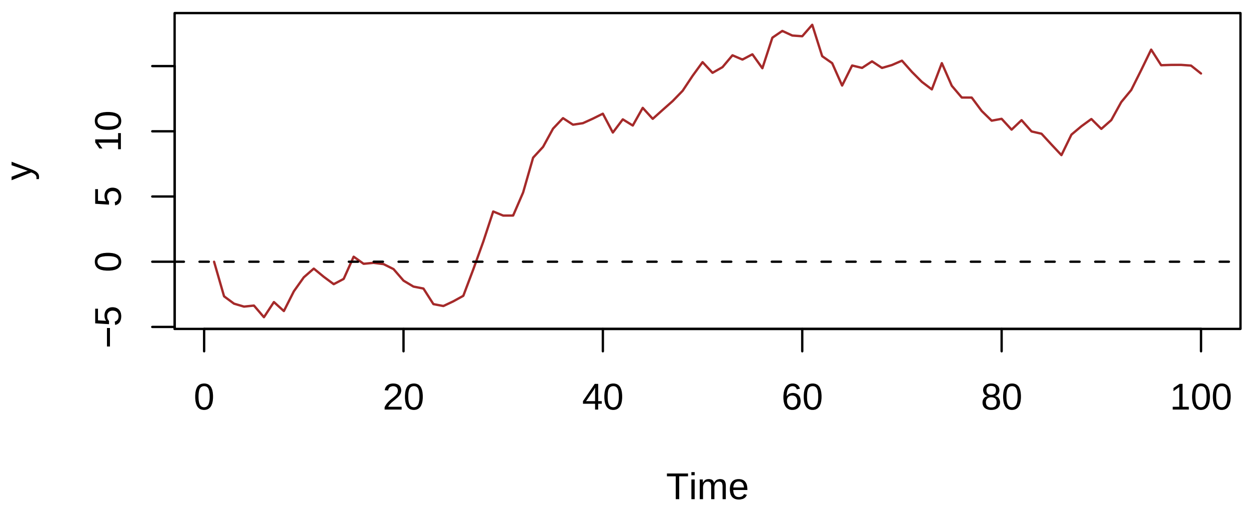
Spurious regression a major problem.

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

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



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



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.,

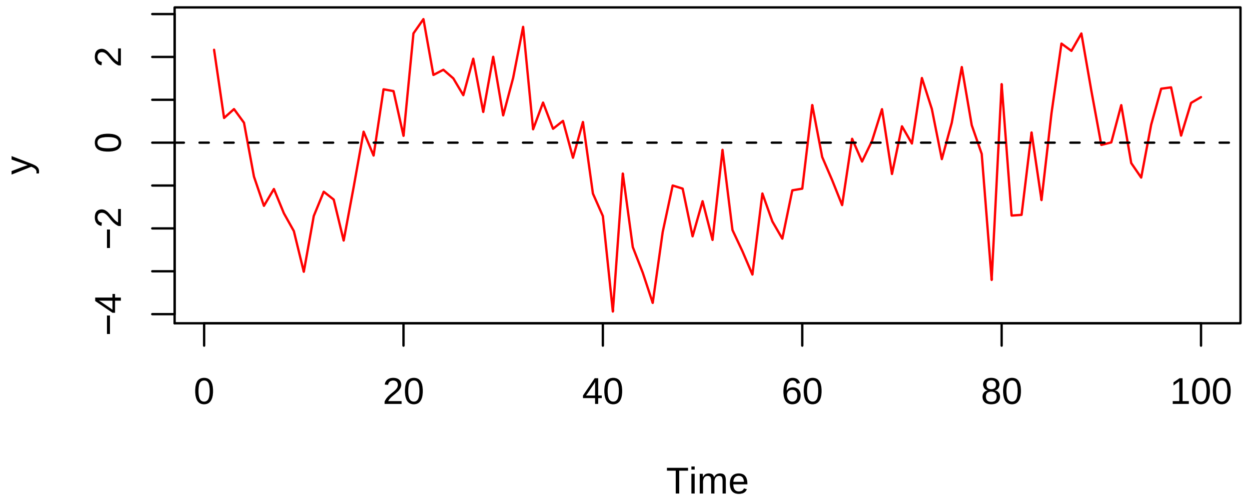
$$y_t = y_{t-1}\phi_1 + y_{t-2}\phi_2 + \ldots + y_{t-p}\phi_p + \varepsilon_t$$

This general case is known as AR(p).

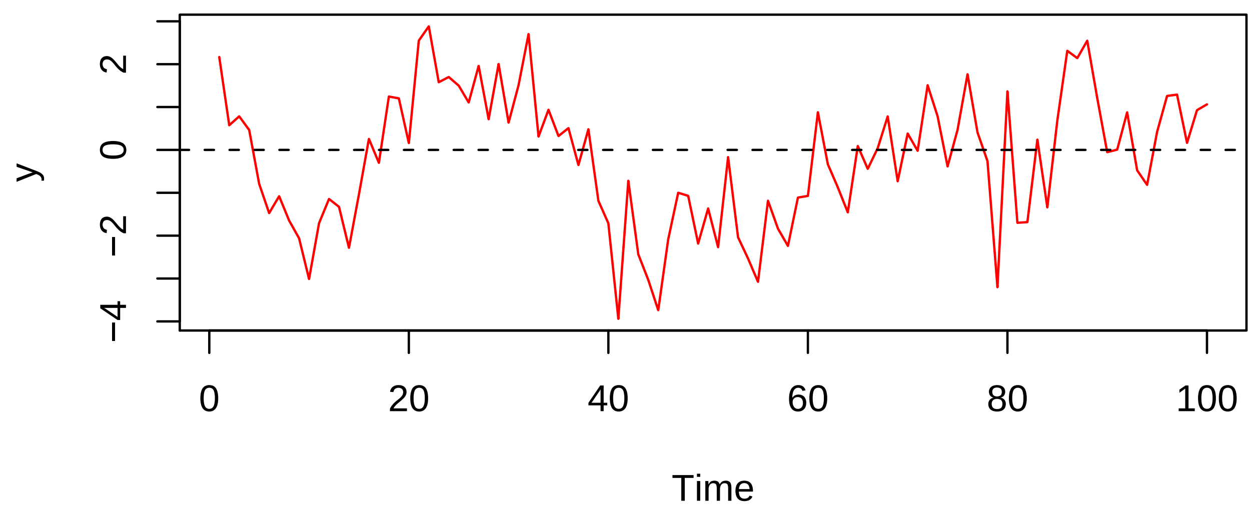
The distant past has a *direct* effect on the present

Distant past should show up in PACF, not just ACF

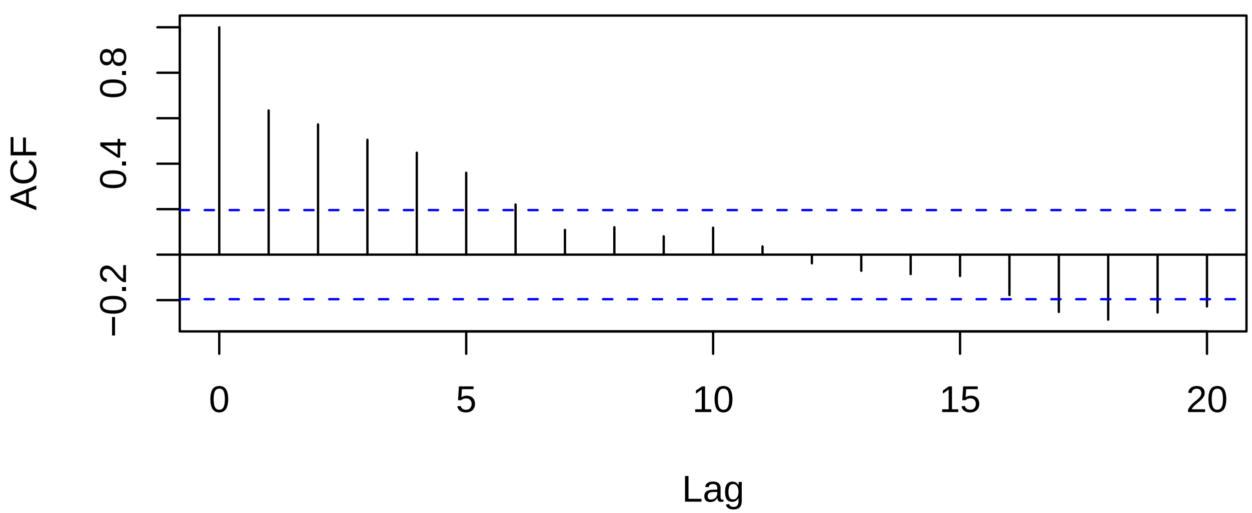
Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$



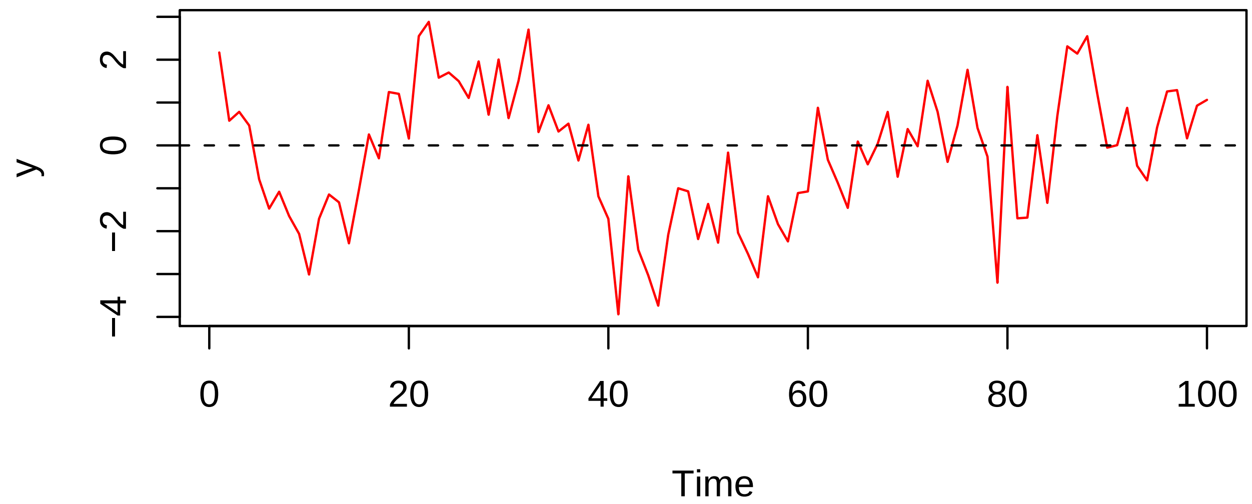
Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$



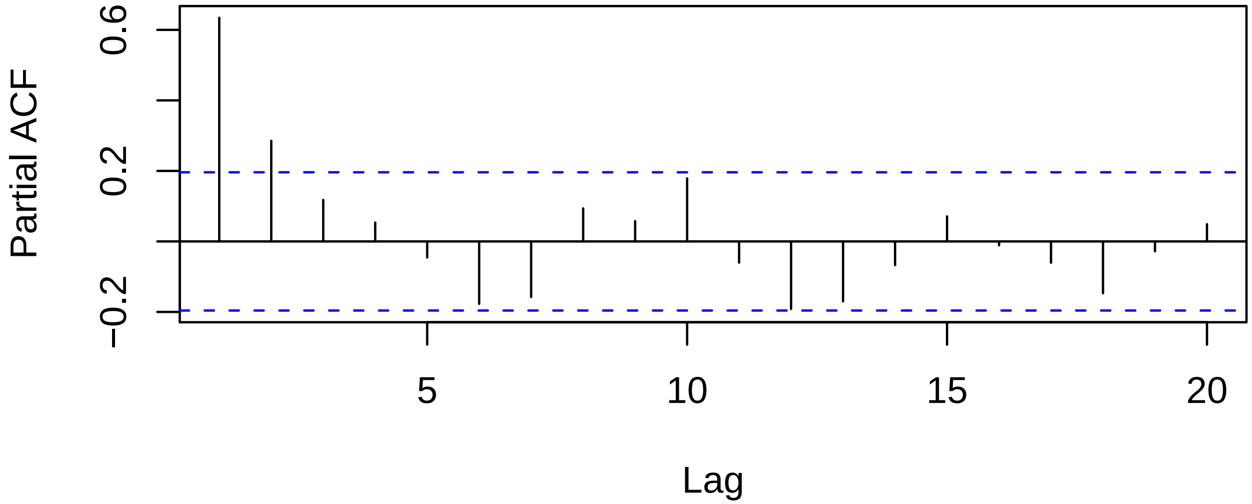
ACF of AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$

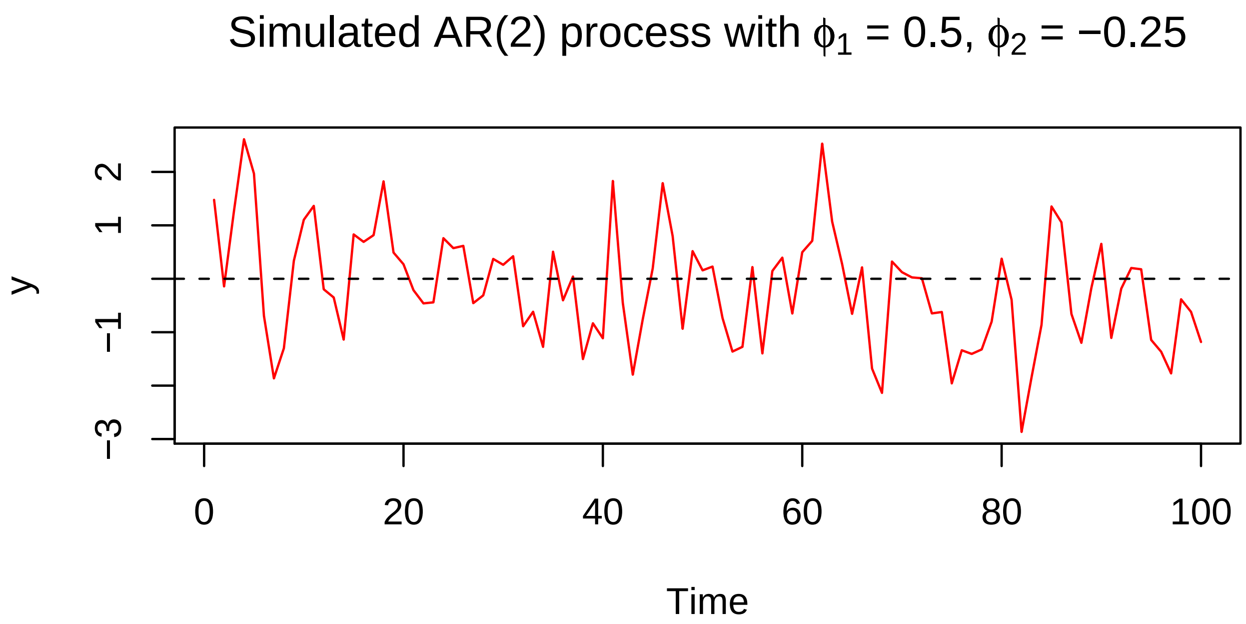


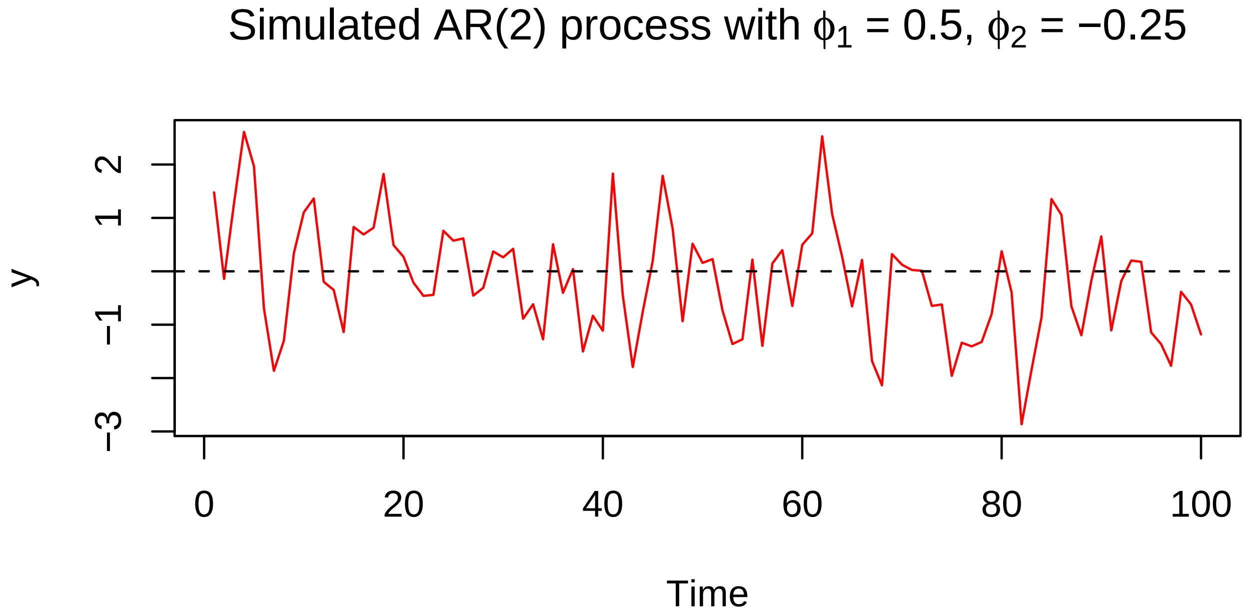
Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$

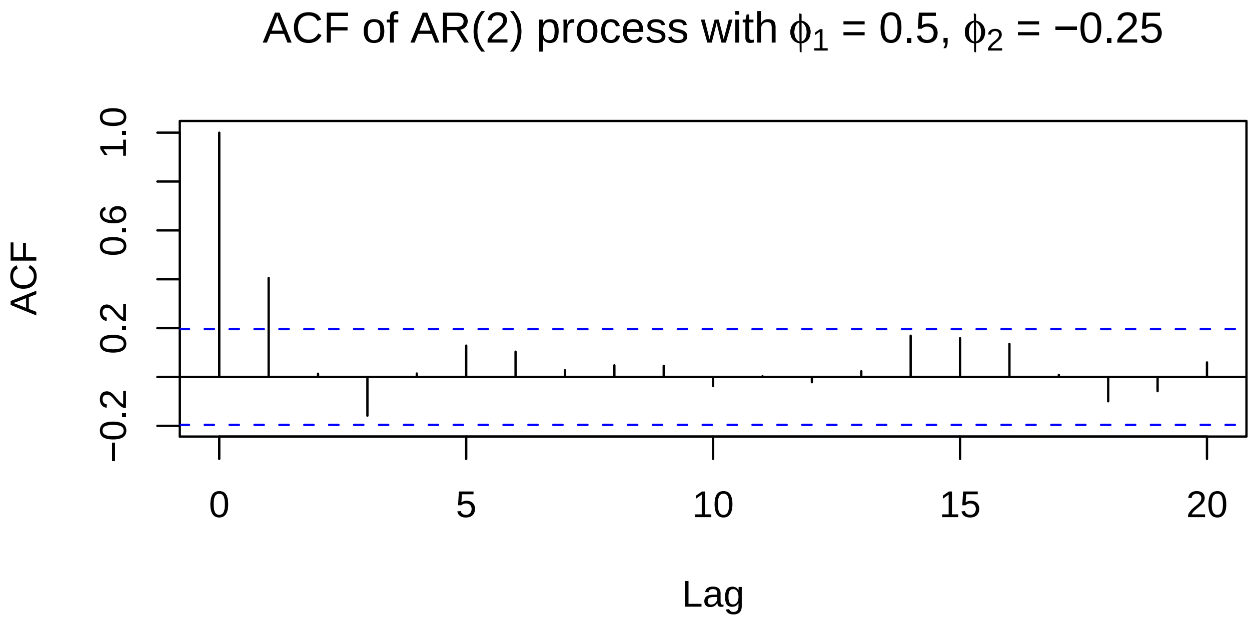


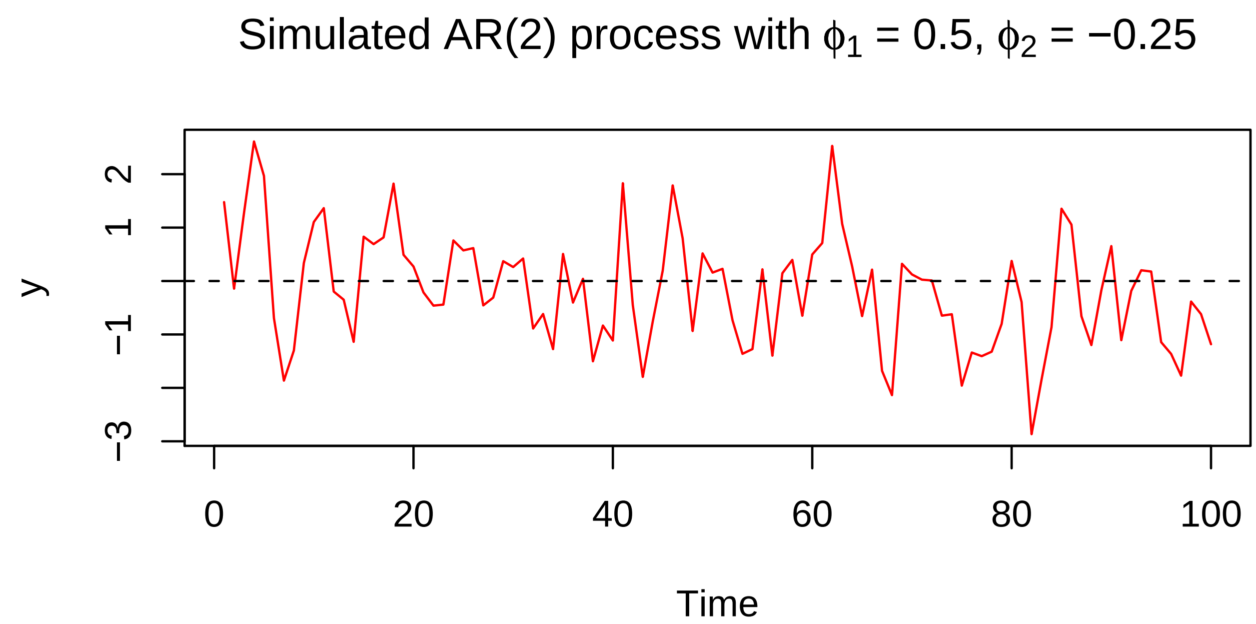
PACF of AR(2) process with $\phi_1 = 0.5$, $\phi_2 = 0.25$

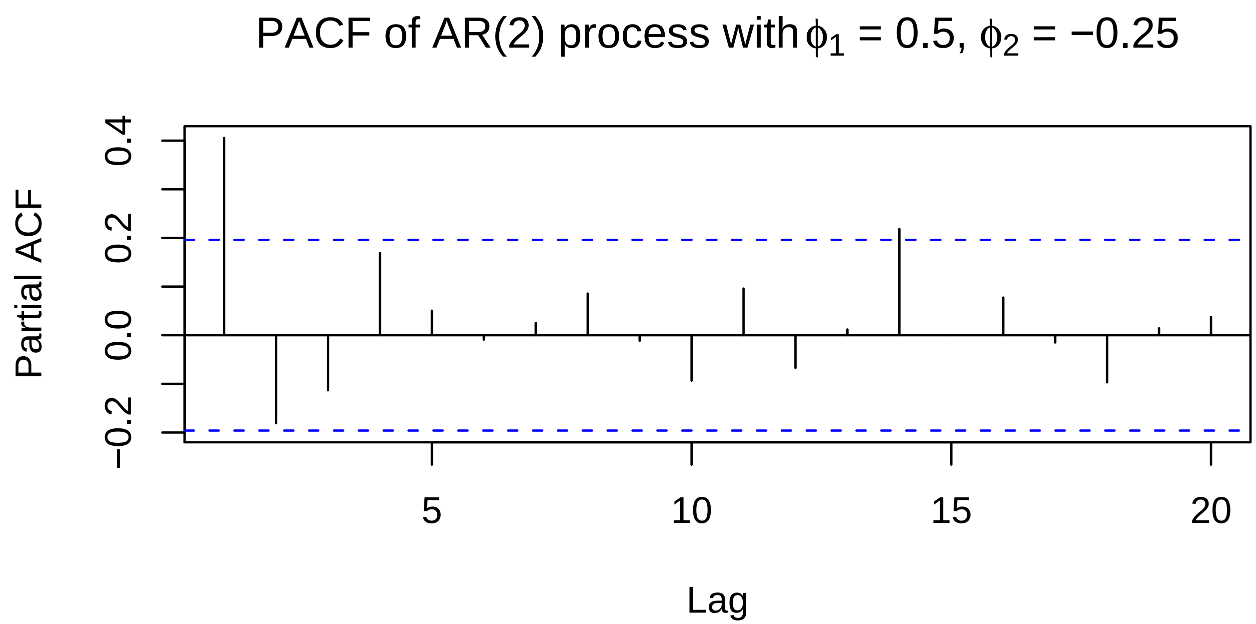



Simulated AR(2) process with $\phi_1 = 0.5$, $\phi_2 = -0.25$
y
2
1
−1
−3
0
20
40
60
80
100
Time


Simulated AR(2) process with ϕ1 = 0.5, ϕ2 = −0.25
y
−3
−1
1
2
0
20
40
60
80
100
Time


ACF of AR(2) process with ϕ1 = 0.5, ϕ2 = −0.25
ACF
−0.2
0.2
0.6
1.0
0
5
10
15
20
Lag


Simulated AR(2) process with ϕ1 = 0.5, ϕ2 = −0.25
y
2
1
−1
−3
0
20
40
60
80
100
Time


PACF of AR(2) process with ϕ1 = 0.5, ϕ2 = −0.25
Partial ACF
0.4
0.2
0.0
−0.2
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 a 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

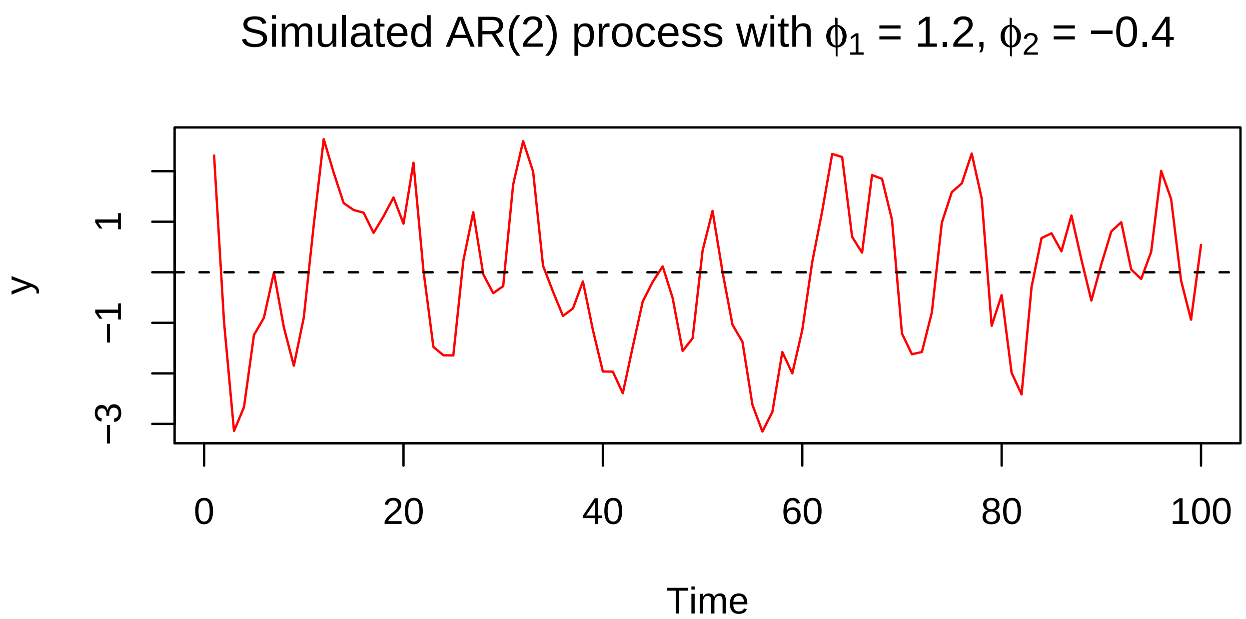
$$\left(1 - \phi_1 \mathrm{L} - \phi_2 \mathrm{L}^2 - \phi_3 \mathrm{L}^3 - \cdots - \phi_p \mathrm{L}^p\right) y_t = \varepsilon_t$$

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

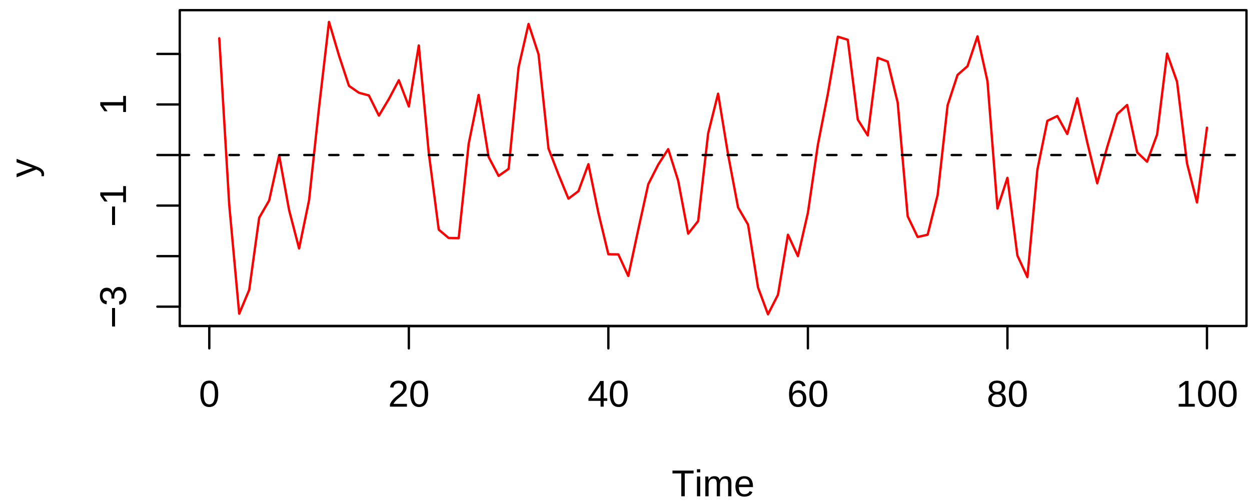
Should worry if any (empirical) roots "close" to 1.

More on finding unit roots in *real* data next time

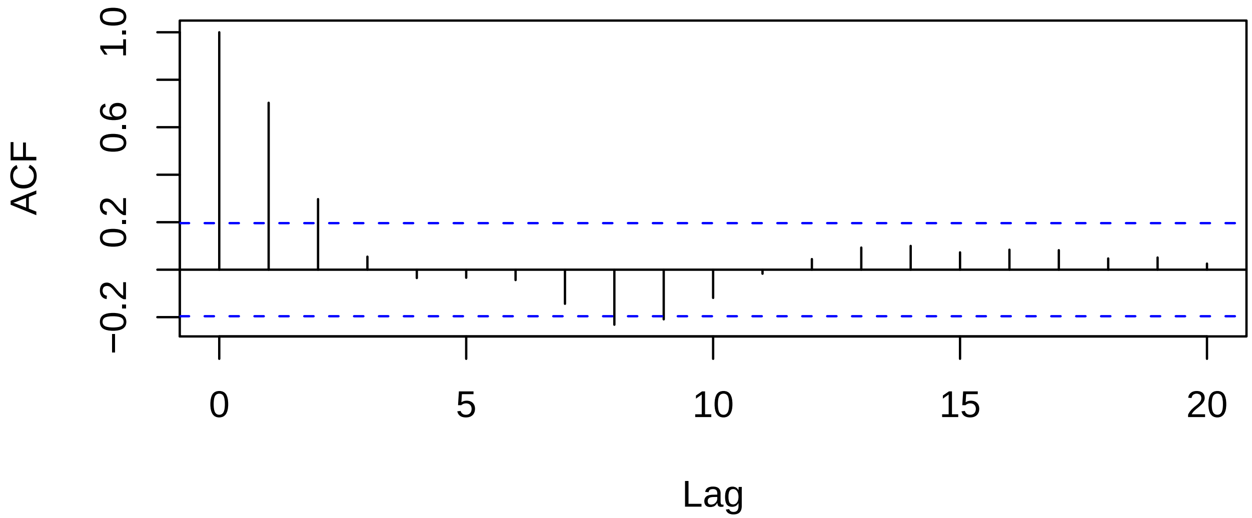
Very hard to do well, unless t is very large
(larger than we ever see in social science)


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

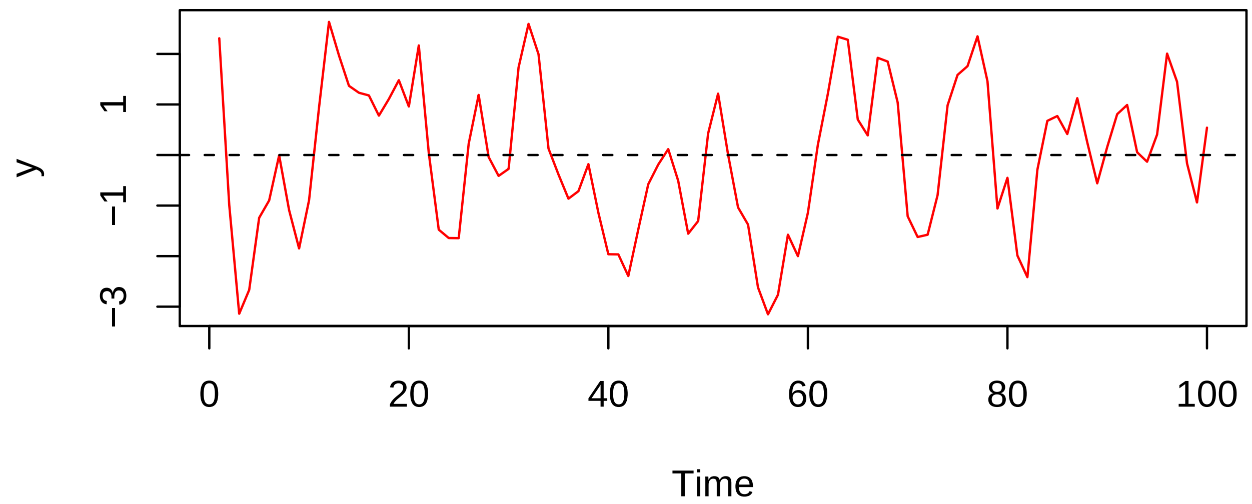
Simulated AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.4$



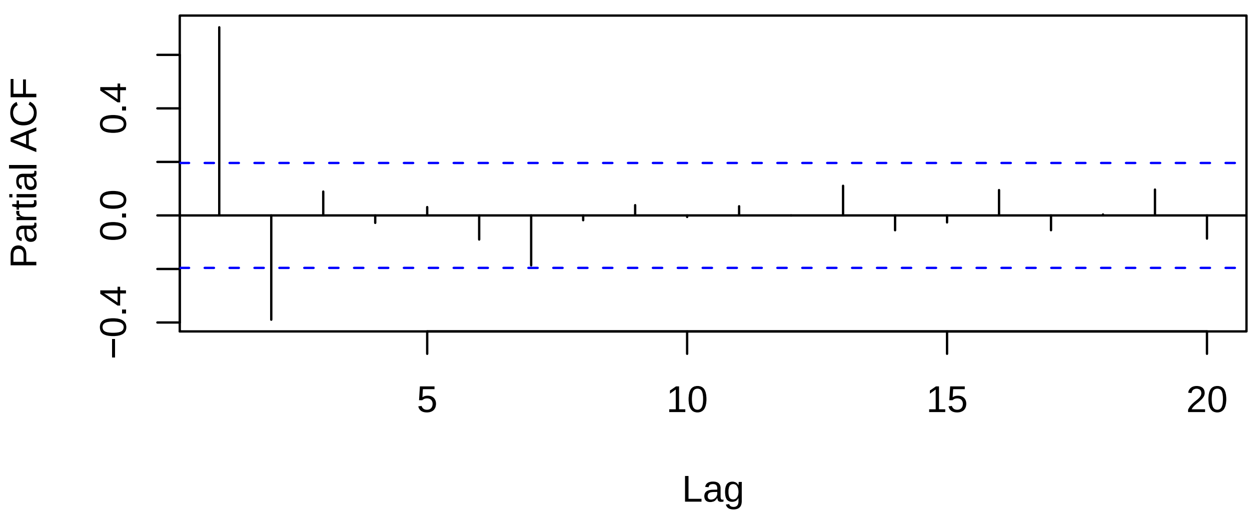
ACF of AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.4$

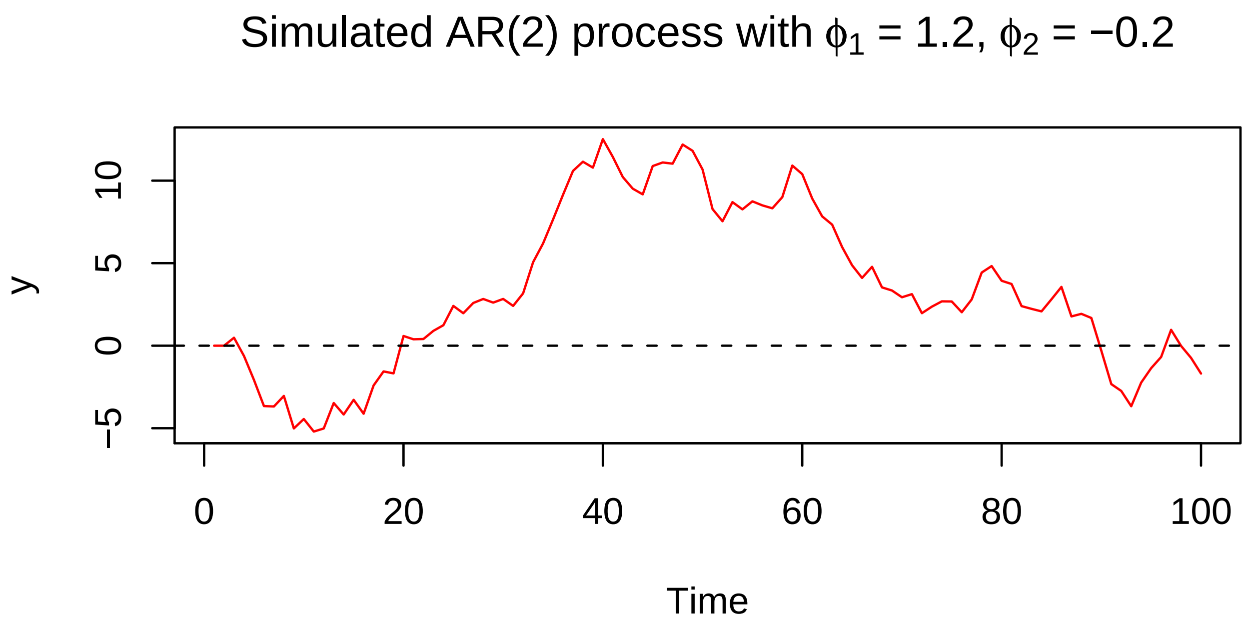


Simulated AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.4$



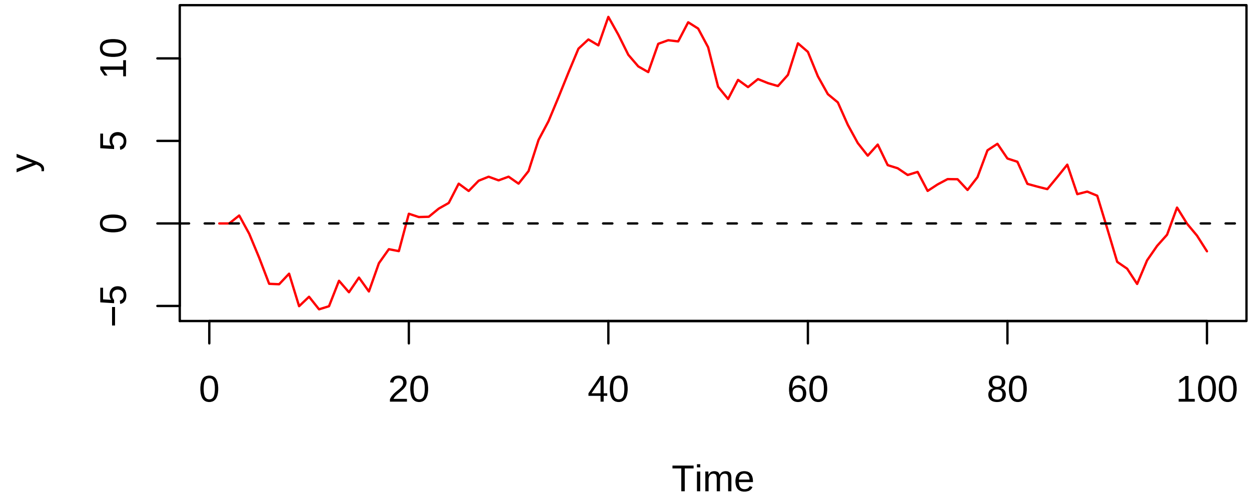
PACF of AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.4$



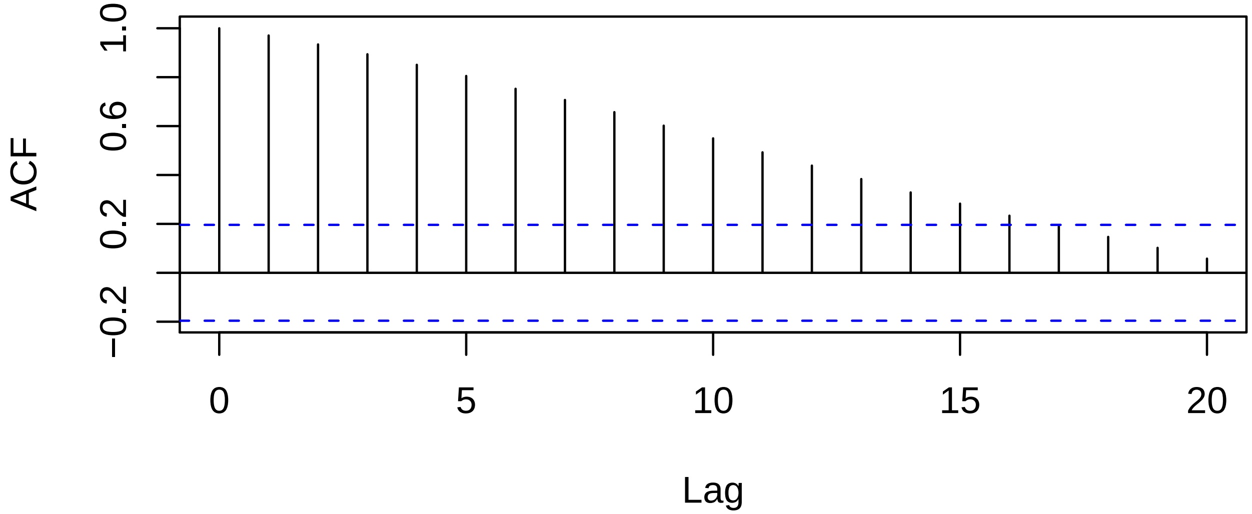


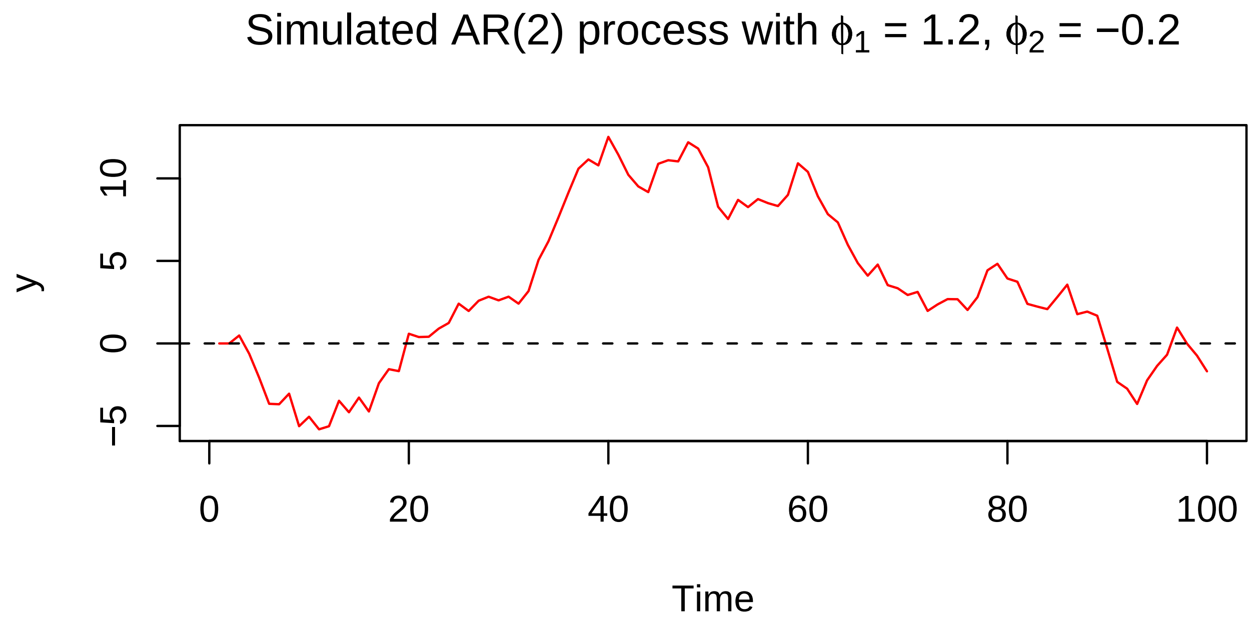
Simulated AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.2$
y
−5
0
5
10
0
20
40
60
80
100
Time

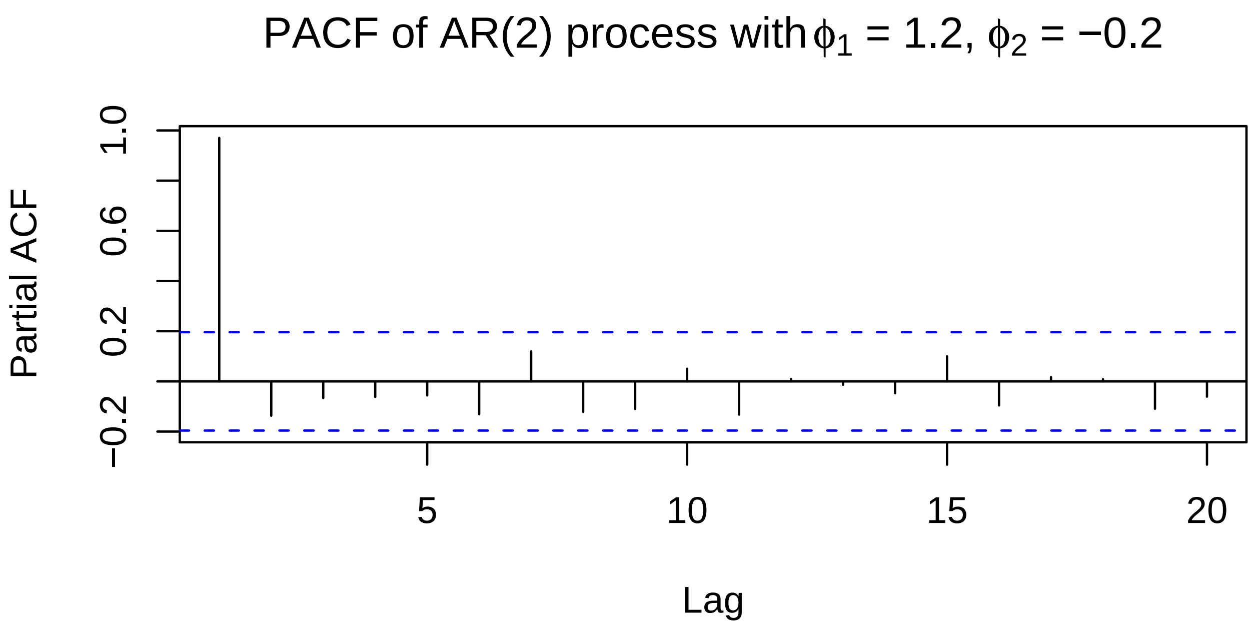
Simulated AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.2$



ACF of AR(2) process with $\phi_1 = 1.2$, $\phi_2 = -0.2$




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


PACF of AR(2) process with ϕ1 = 1.2, ϕ2 = −0.2
Partial ACF
1.0
0.6
0.2
−0.2
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, $\operatorname{cov}(\varepsilon_t, \varepsilon_{t+k}) = 0$ 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 MA(1), the effect of past shocks dies out after q periods

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

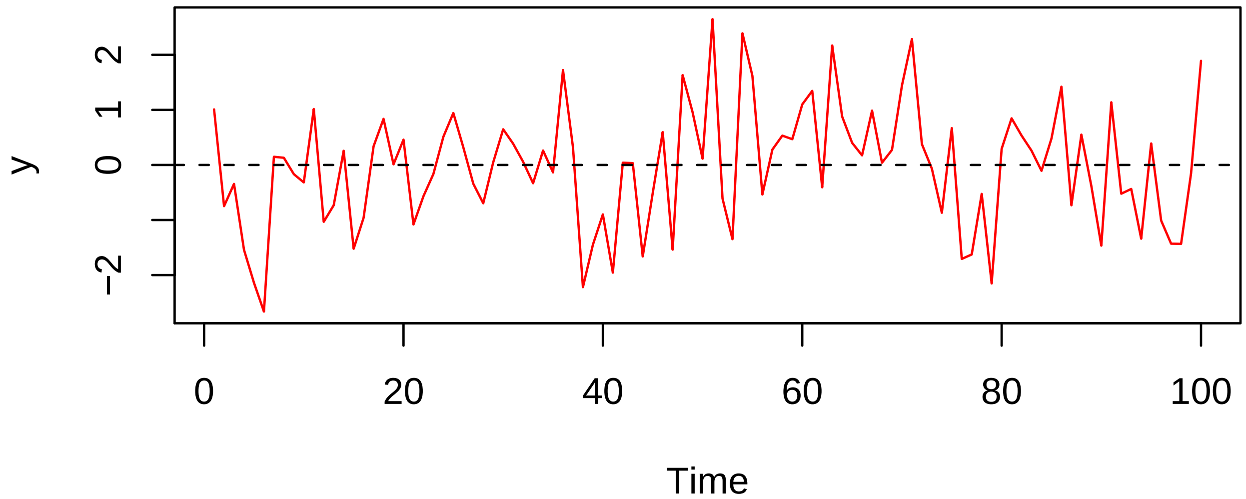
Contrast to the AR(1) or AR(p), in which shocks never (quite) die out, and non-stationarity can occur

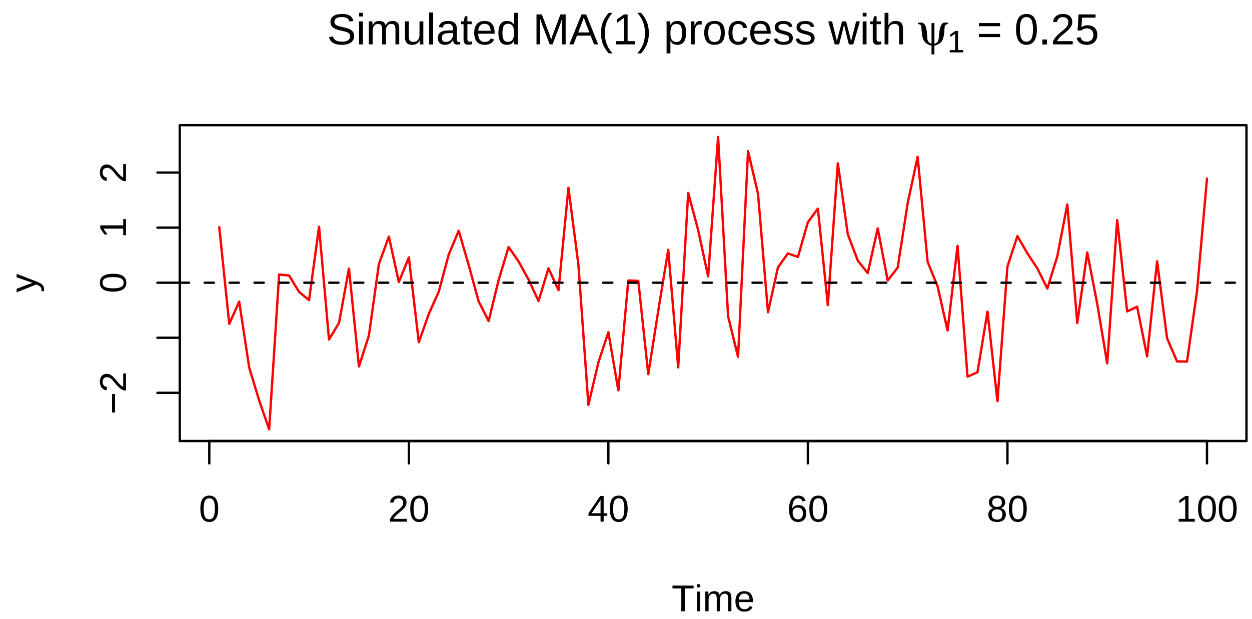
Simulating MA(q)

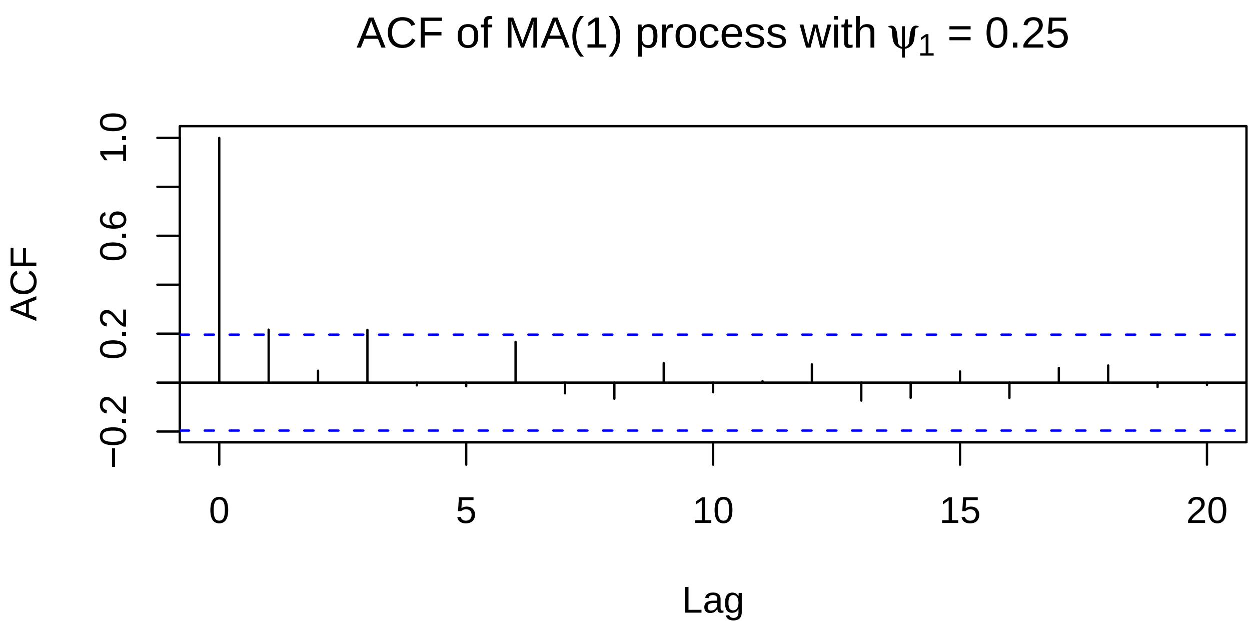
To simulate 1000 iterations of an MA(2) with $\rho_1 = 0.67$ and $\rho_2 = 0.5$:

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

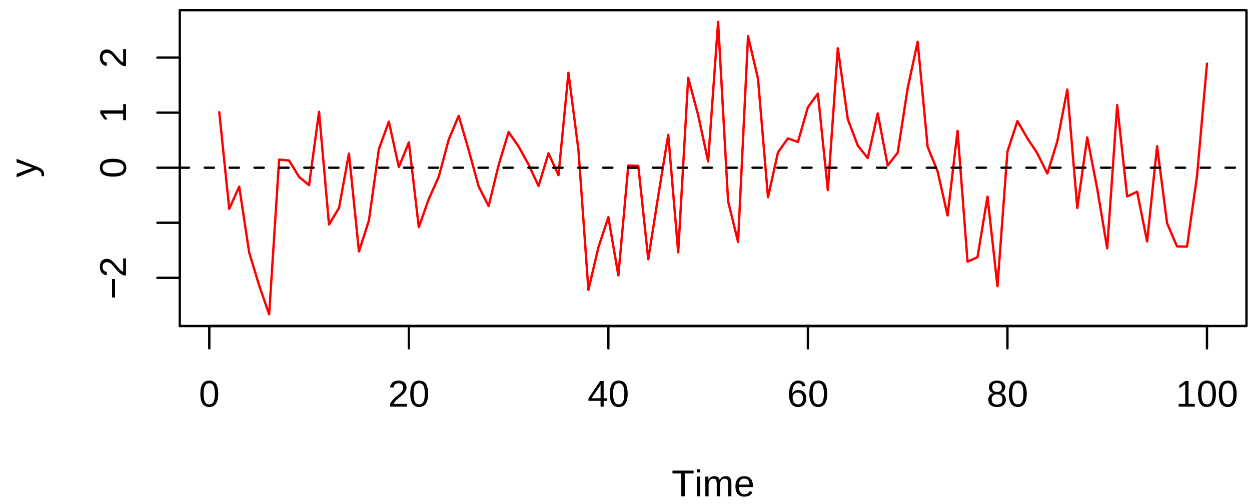
Simulated MA(1) process with $\psi_1 = 0.25$



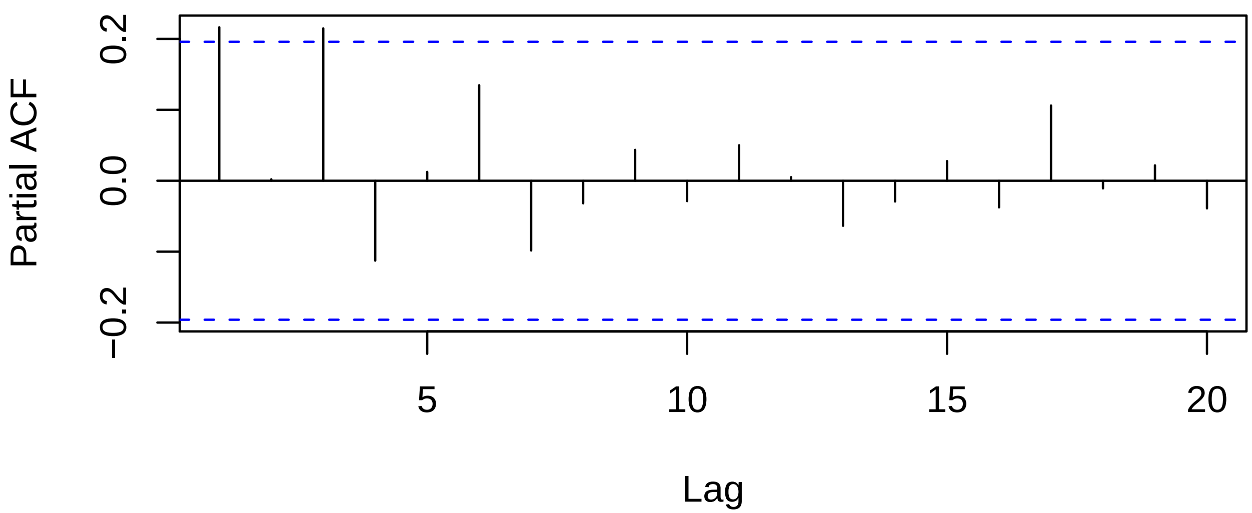

Simulated MA(1) process with ψ1 = 0.25
y
−2
0
1
2
0
20
40
60
80
100
Time


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

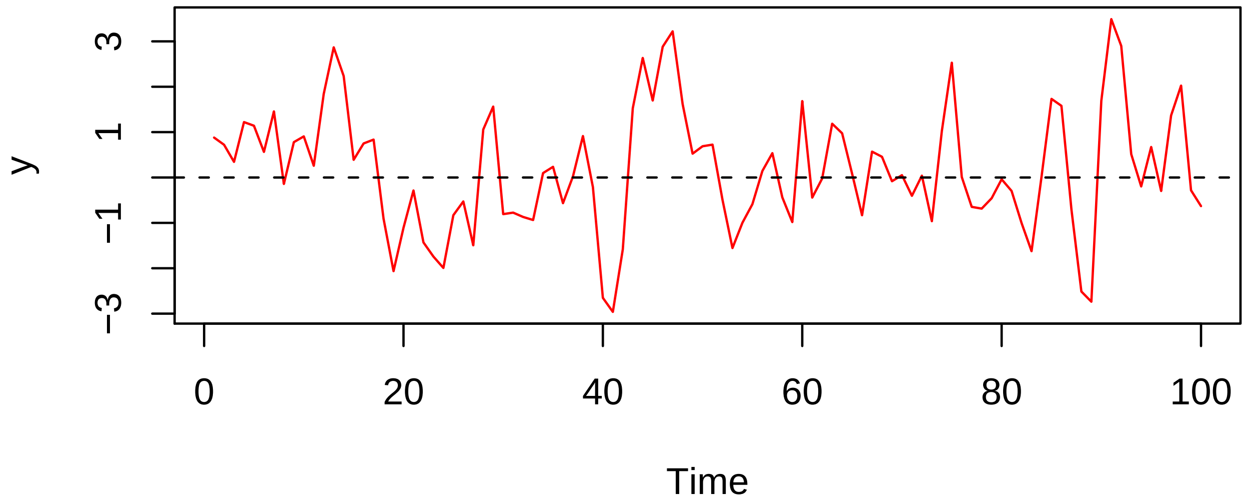
Simulated MA(1) process with $\psi_1 = 0.25$

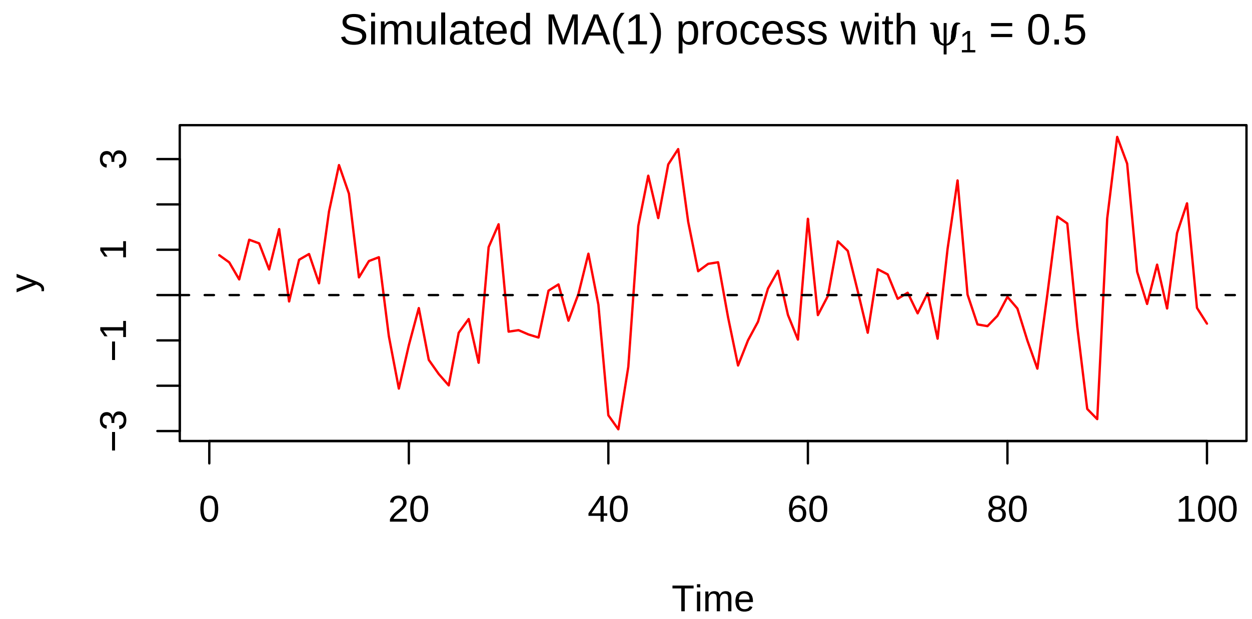


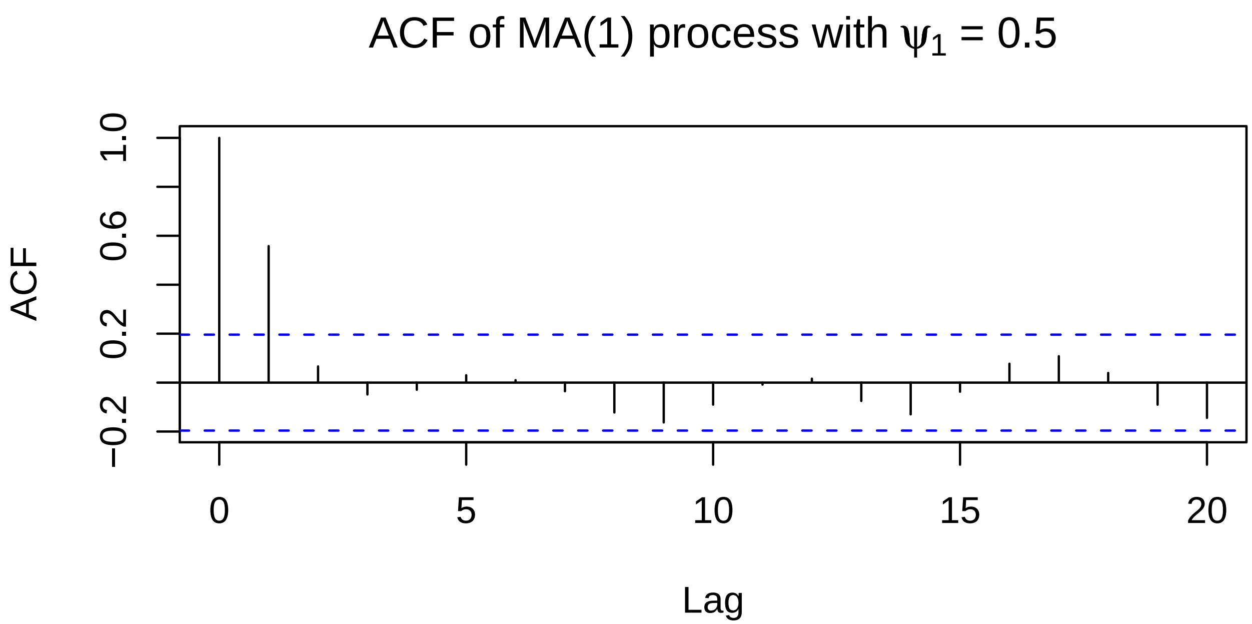
PACF of MA(1) process with $\psi_1 = 0.52$

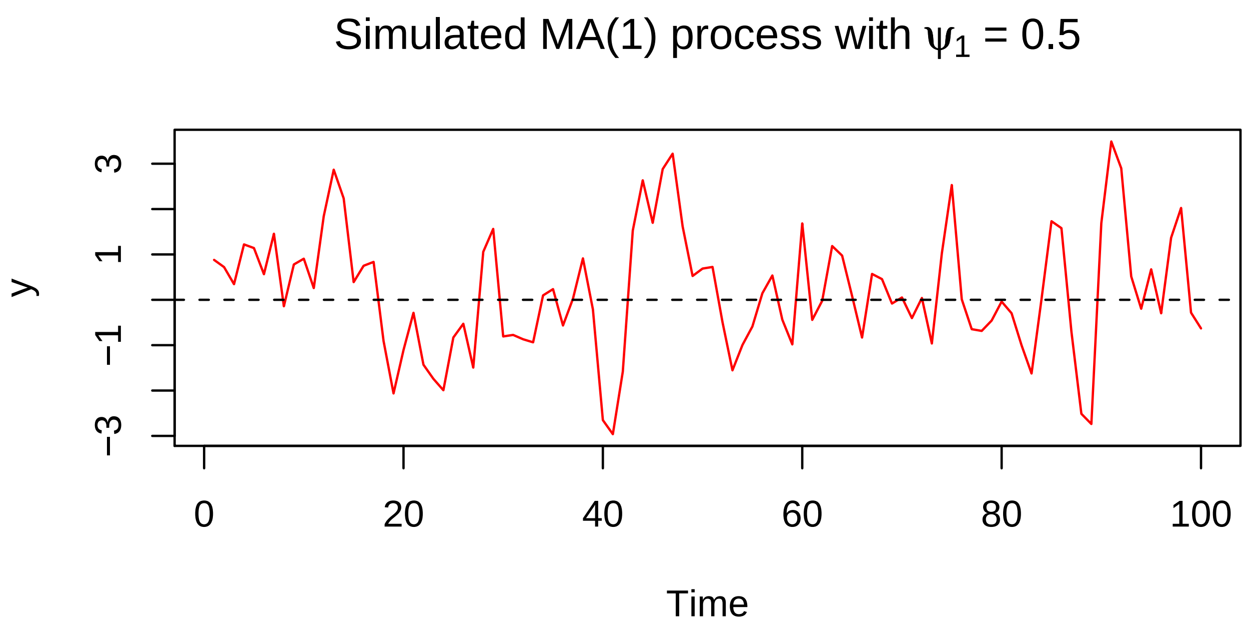


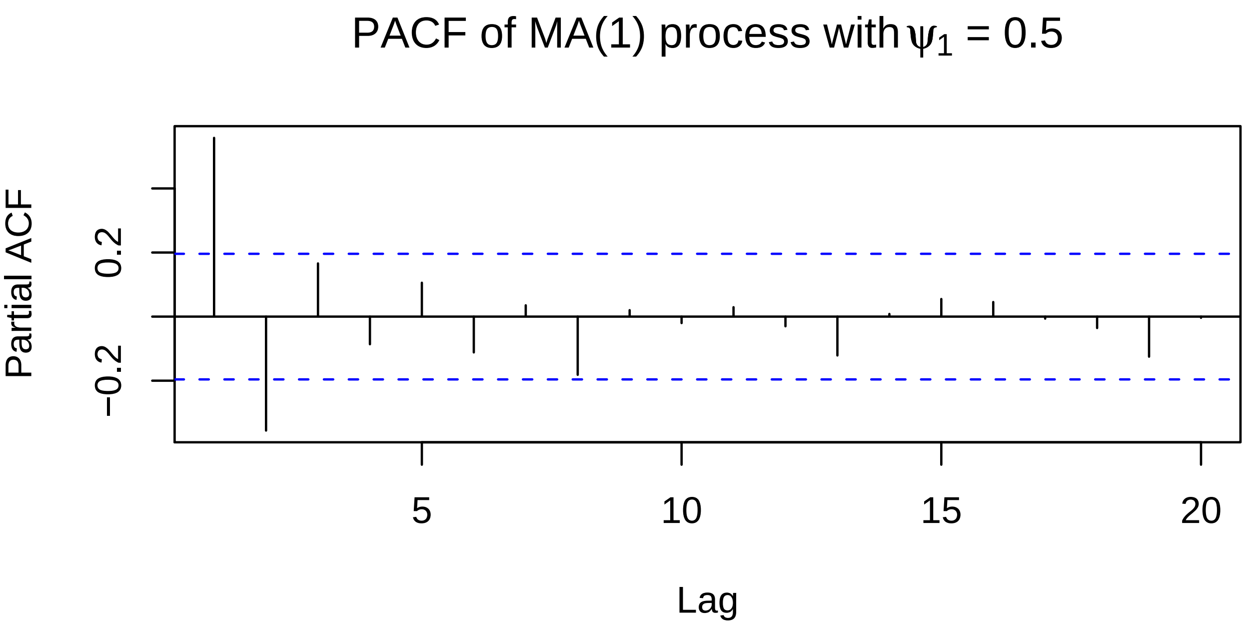
Simulated MA(1) process with $\psi_1 = 0.5$

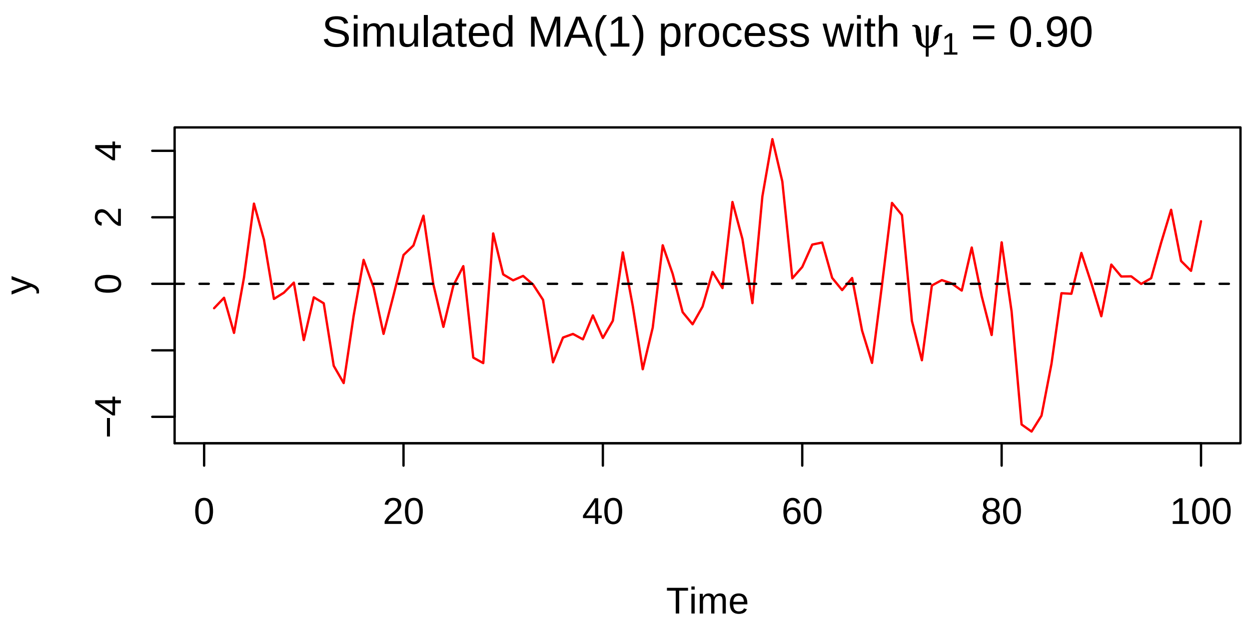



Simulated MA(1) process with $\psi_1 = 0.5$
y
3
1
−1
−3
0
20
40
60
80
100
Time

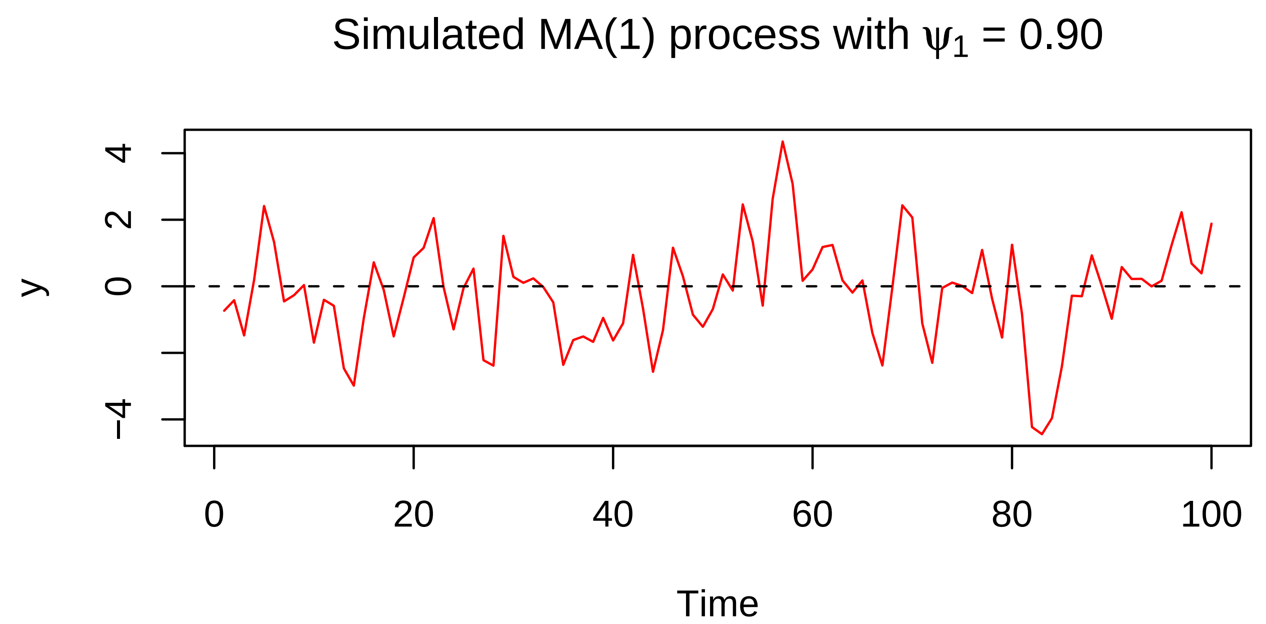

ACF of MA(1) process with $\psi_1 = 0.5$
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag


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

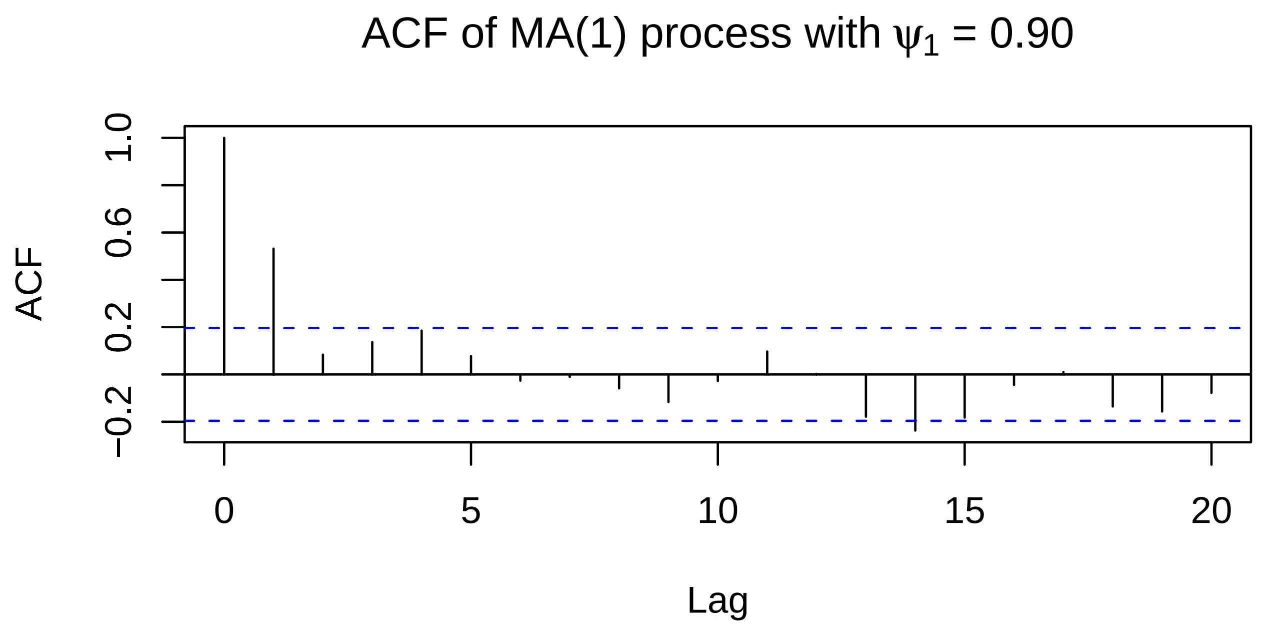

PACF of MA(1) process with ψ1 = 0.5
Partial ACF
0.2
−0.2
5
10
15
20
Lag



Simulated MA(1) process with ψ_1 = 0.90
y
4
2
0
−4
0
20
40
60
80
100
Time

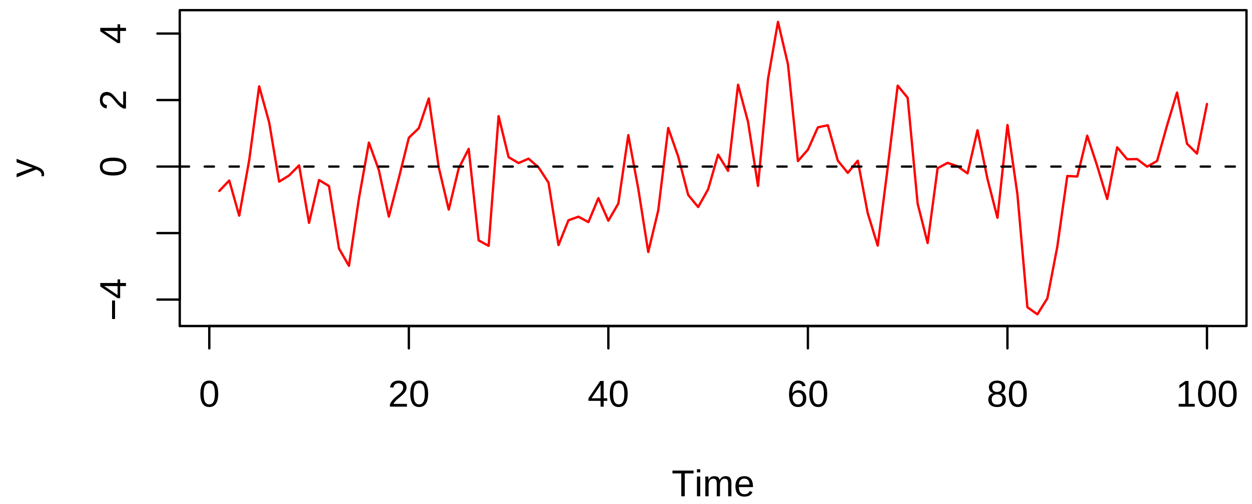


Simulated MA(1) process with ψ1 = 0.90
y
4
2
0
−4
0
20
40
60
80
100
Time

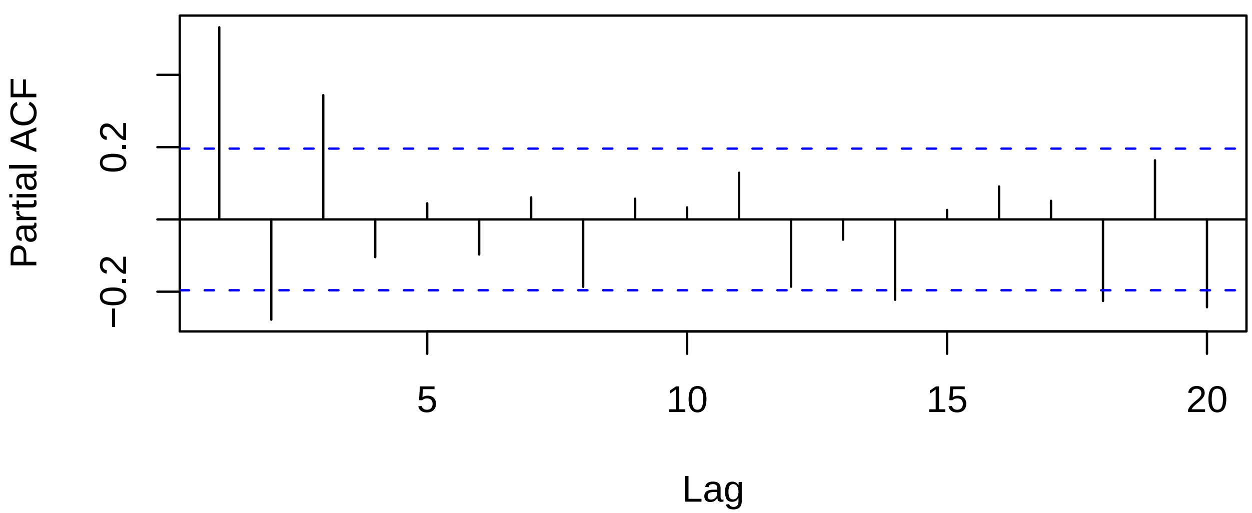


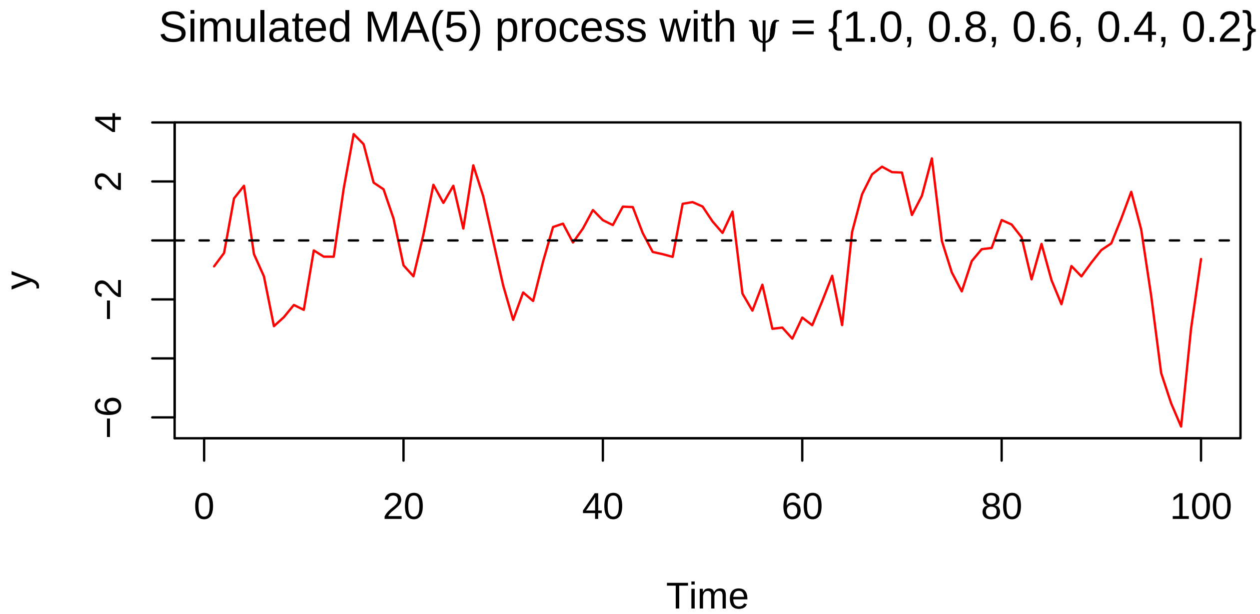
ACF of MA(1) process with ψ1 = 0.90
ACF
1.0
0.6
0.2
−0.2
0
5
10
15
20
Lag

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

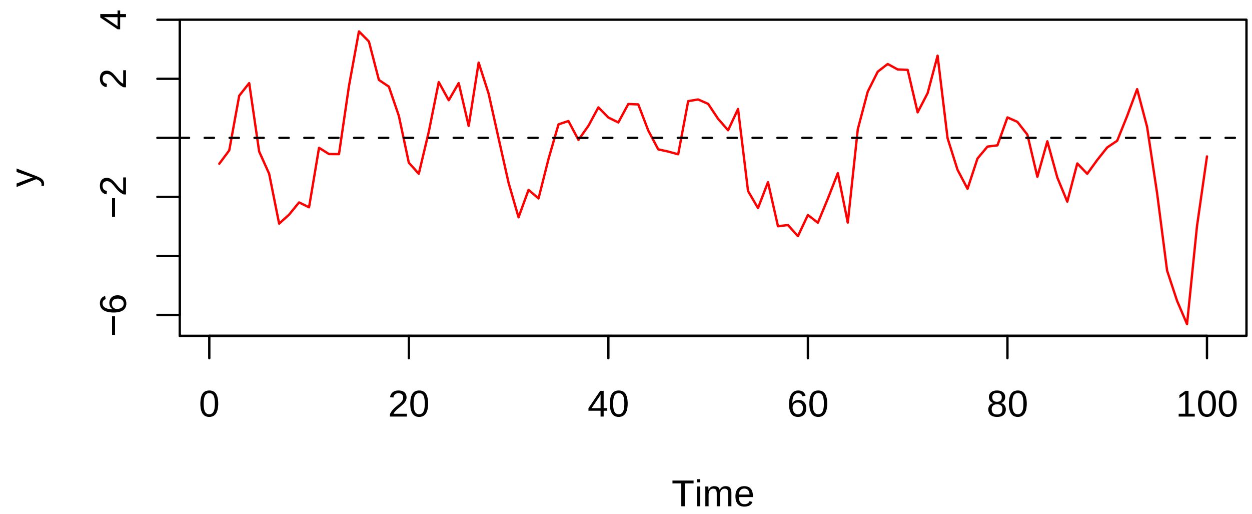


PACF of MA(1) process with $\psi_1 = 0.90$

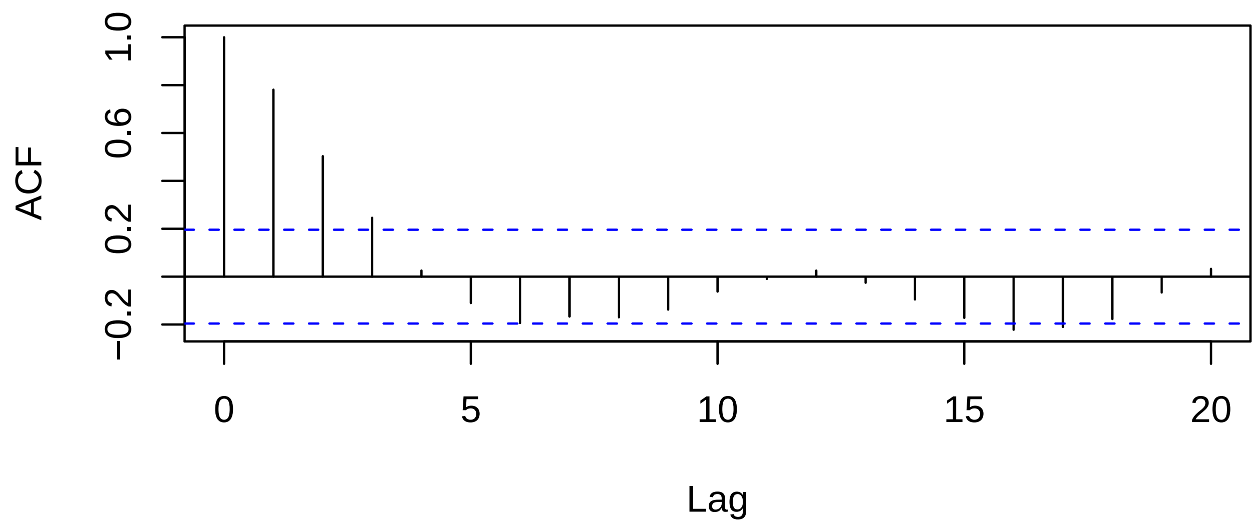


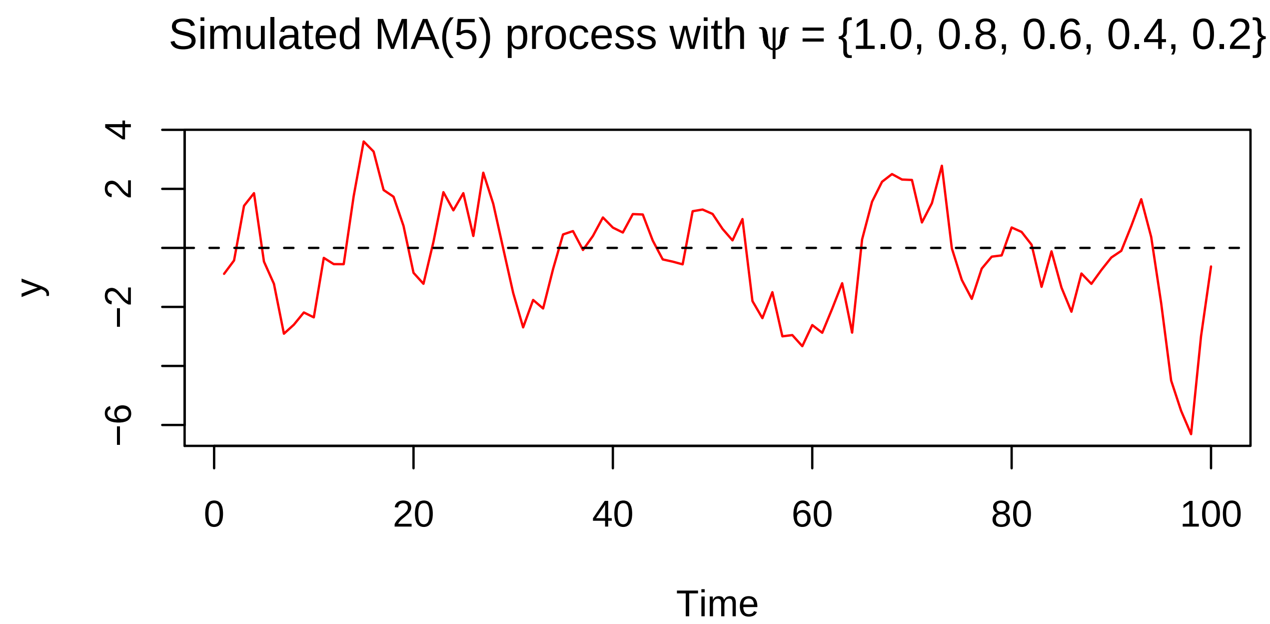

Simulated MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}
y
4
2
−2
−6
0
20
40
60
80
100
Time

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

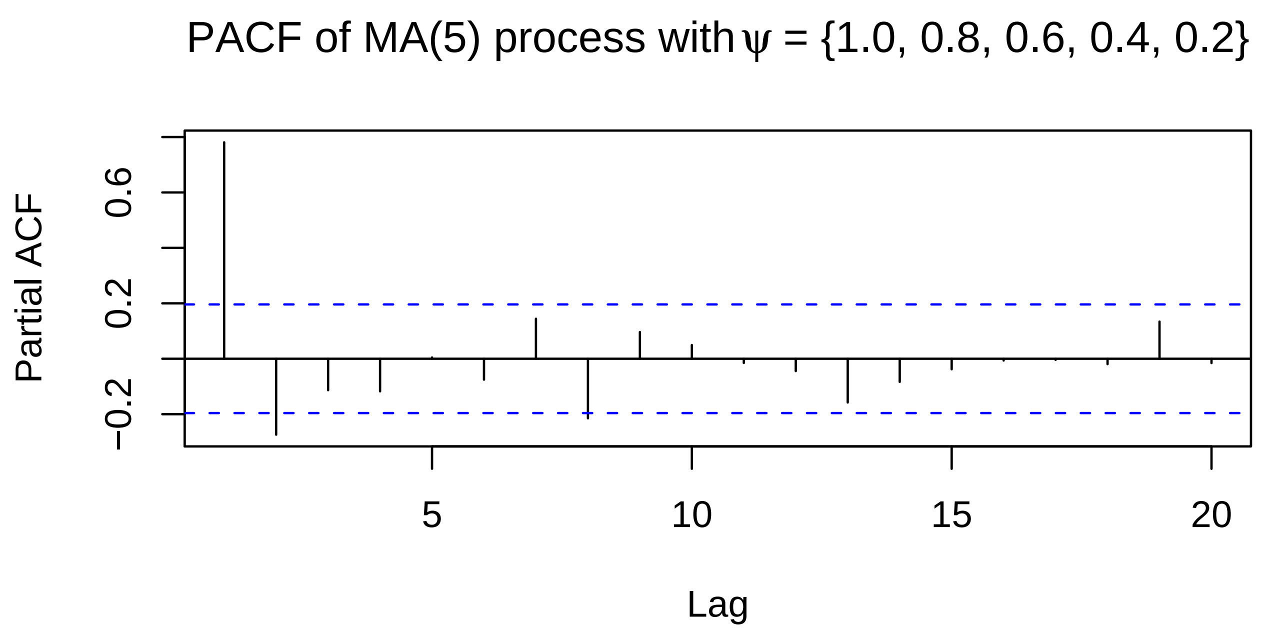


ACF of MA(5) process with $\psi = \{1.0, 0.8, 0.6, 0.4, 0.2\}$





Simulated MA(5) process with ψ = {1.0, 0.8, 0.6, 0.4, 0.2}
y
−6
−2
2
4
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
−0.2
0.2
0.6
5
10
15
20
Lag

Equivalence of AR and MA processes

For any stationary 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 *invertible* 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

Invertible here refers to moving averages in which shocks have declining weight as they pass into the past

Invertibility also requires that each $|\rho| < 1$

A sort of mirror case to stationarity for autoregressive processes:
if we had an MA(∞) with $\rho_1 = \rho_2 = \cdots = \rho_\infty = 1$, the process would be nonstationary

Note that non-invertible MA(q) processes can still be stationary for $q < \infty$, but can't be represented using AR(∞) processes

Intuition check:
The MA(5) series above is an example of a non-invertible but stationary series

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)

Assuming invertibility, for most high order AR(p), MA(q), and ARM(p,q) processes, there is a very similar *low*-order ARMA(p,q) process

For modeling purposes, we would prefer to estimate fewer parameters, so using this low-order ARMA process gives us a lot of flexibility for a small price

This approach is the Box-Jenkins methodology for time series

Tends to outperform more highly parameterized approaches in forecasting

Seasonality

Time series can demonstrate cycles as well as trends and lags

The best known cycles revolve around seasons of the year,
but there are other possibilities:

- with daily data, often weekly cycles
- in politics the cycles may be multiyear (relating to election calendars)
- can have cycles within cycles (days within weeks within months. . .)

For now, assume we just have $k = 12$ seasons, so we will refer to "months" of data

This means that $t = 1$ has something in common with $t = 13$,
and $t = 2$ has something in common with $t = 14$, etc.

There could still be close correlations between $t = 1$ and $t = 2$ (consecutive months), compounded with the correlation between $t = 1$ and $t = 13$ ("same" month in different years)

Additive Seasonality

Two key ways to think about seasonality – additive or multiplicative

Additive seasonal effects: the same month in different years to share a level component

Call this level κ_k. Suppose $\kappa_1 = 15$, $\kappa_2 = -10$, and $\bar{\kappa} = 5$

Then the average month has a κ of 5,
while Januaries tend to be 10 units higher than the average month,
and Februaries 15 units lower

Note that this size of this effect does not, in this setup, depend on the year

If we knew the true κ's, we could remove seasonal differences by subtracting the appropriate κ from each observation

If we didn't know the κ's, we could estimate them (up to a common constant) using a regression controlling for $k - 1$ parameters,
one for each month (less a reference month)

Additive Seasonality

When is an additive approach to seasonality reasonable?

As with any time series method, the key will be what fits the data well (covered in the next topic)

But as a rough guide,
additive seasonality probably best describes time series that fulfill three criteria:

1. the time series shows evidence of cycles,
2. but no deterministic trends (or the series has been detrended),
3. and has stable variance over time

The last two criteria reflect the assumption that an additive seasonal effect is stable over many iterations of the cycle

Multiplicative Seasonality

Multiplicative seasonal effects: A period relates to the previous cycle through a factor change

This factor is usually a positive number between 0 and 1

E.g., if the time series was large last January, it will tend to be so again this January; if the series was small last December, it will also be small this December

This allows the size of the seasonal effects to change over time in various ways

An obvious tool for assessing this kind of seasonality is ARMA itself,
but deployed over longer lags

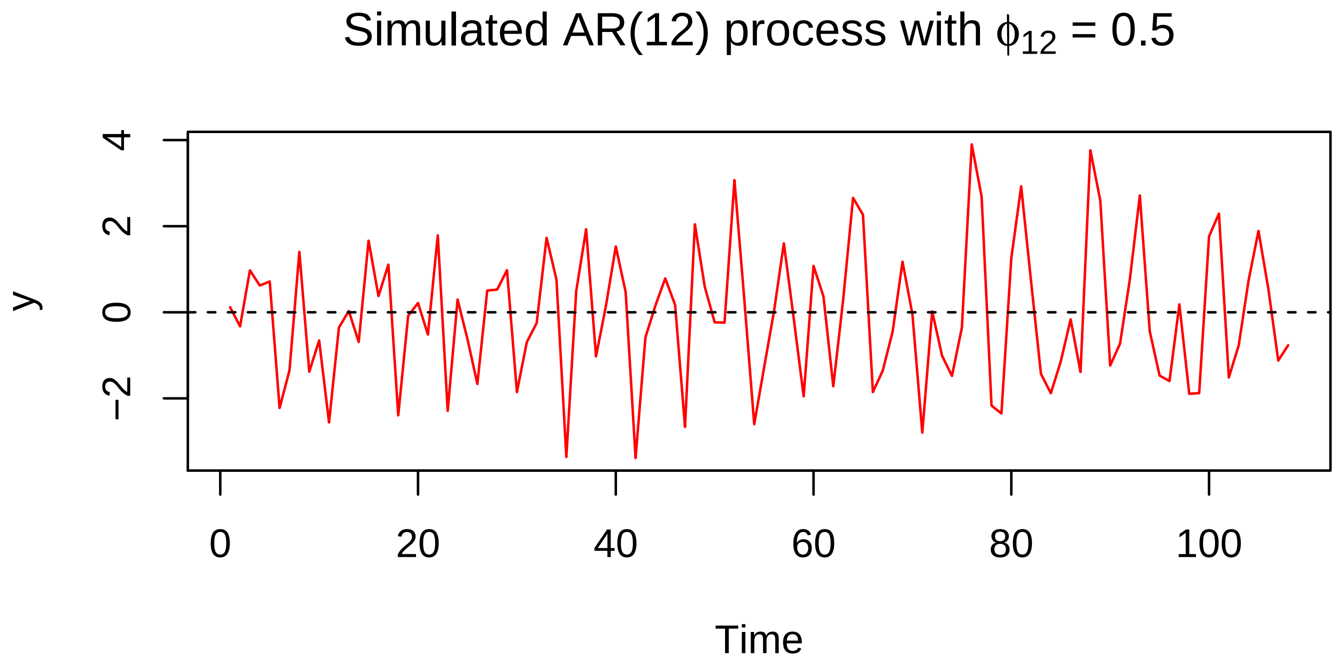
Multiplicative Seasonality with ARMA

Imagine we had an ARMA(12,0) process on monthly data
in which $\phi_1 = \phi_2 = \ldots \phi_1 1 = 0$ and $\phi_{12} = 0.5$

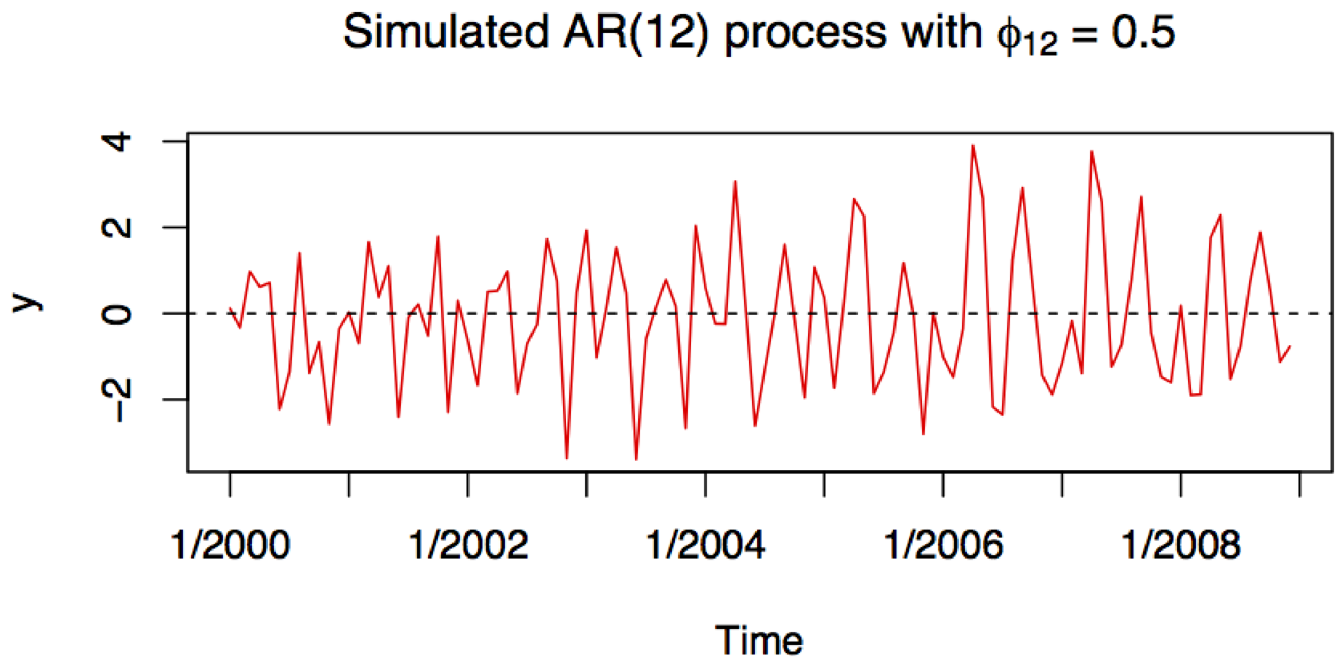
This is a time series in which the present value of y_t depends on the past, but on y_{t-12}, not y_{t-1}

That is, the present y_t depends on the corresponding month of *last year*

Let's look at a simulated time series to fix ideas. . .



Here are $12 \times 9 = 108$ draws from an AR(12) with $\phi_{12} = 0.5$

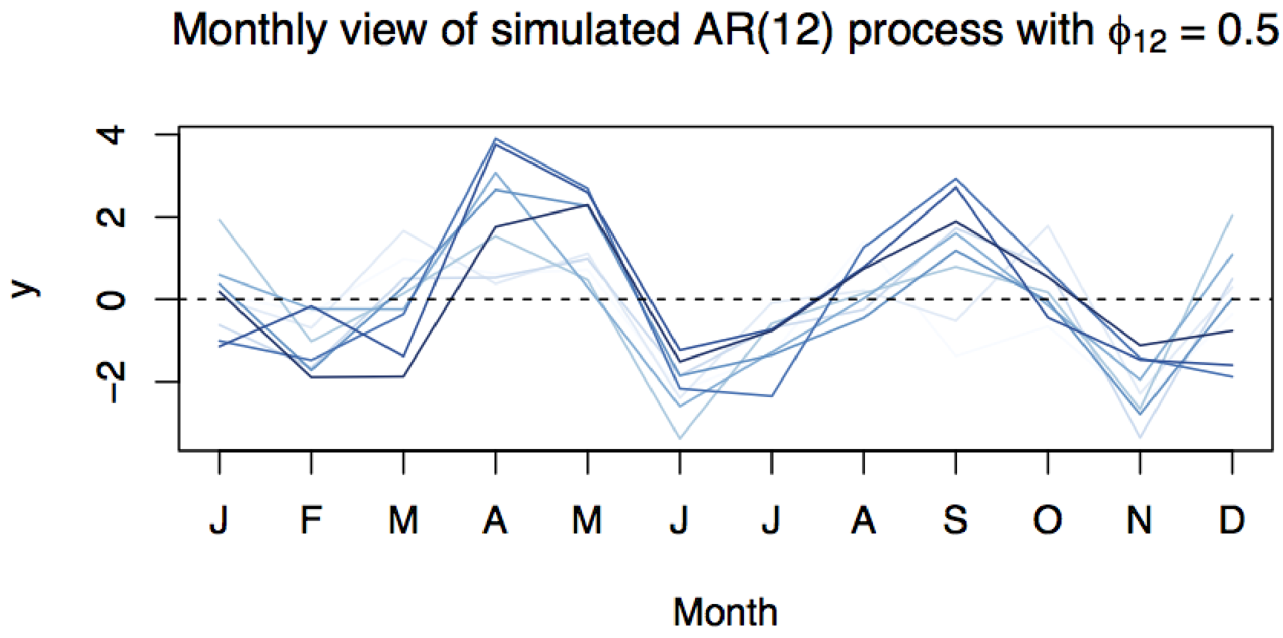


Here are $12 \times 9 = 108$ draws from an AR(12) with $\phi_{12} = 0.5$

When working with cyclical data,
it's important to label axes to highlight cycles

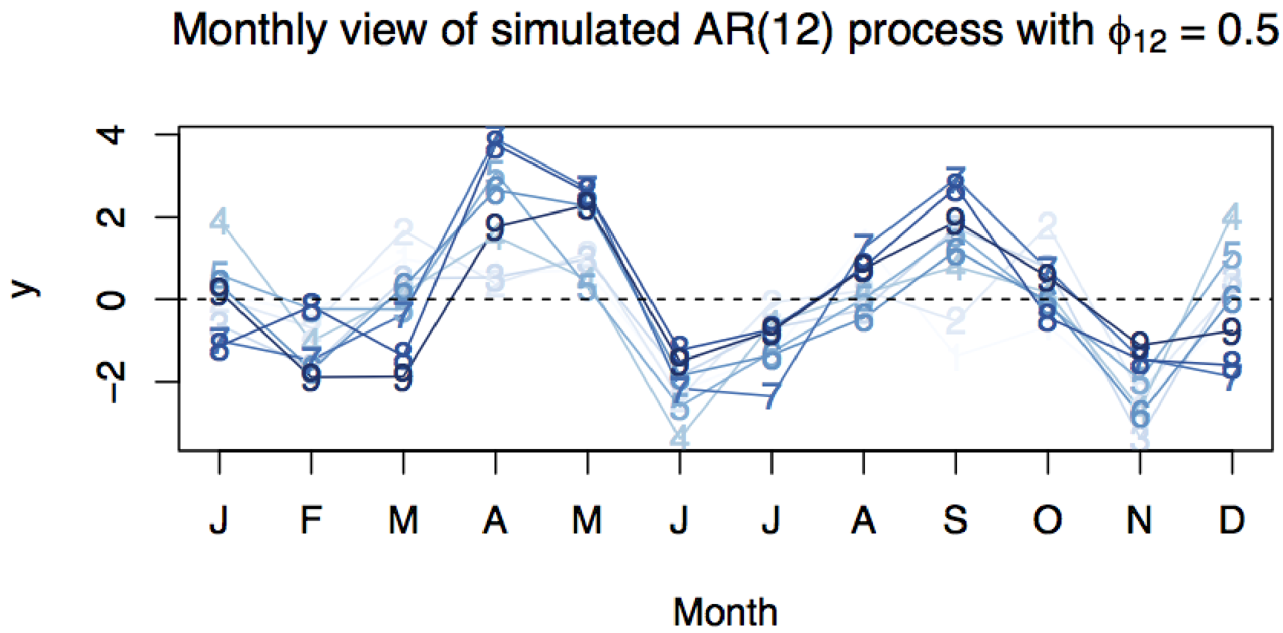
(I'll make up some "years")

But we still can't really see the cyclical pattern. . .



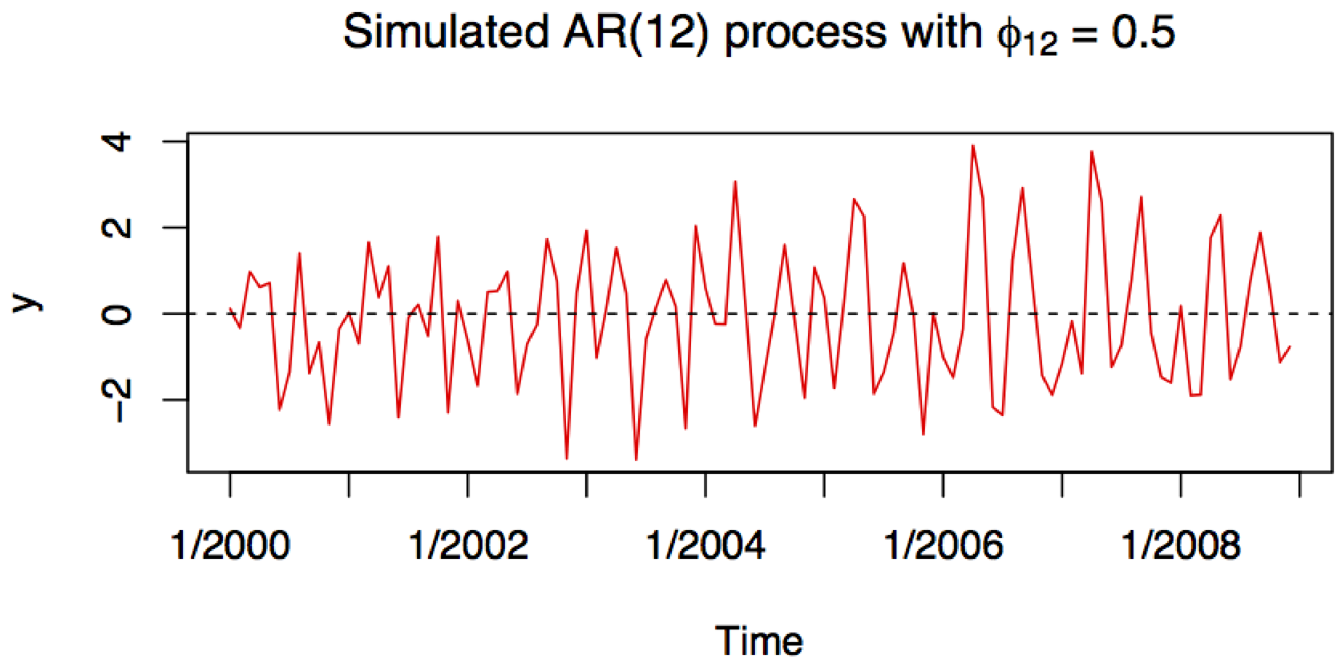
Splicing the time series and overplotting each cycle clarifies the nature of seasonality

This is a key visualization step if you suspect seasonality



Choose colors well to indicate the flow of time – `RColorBrewer` helps

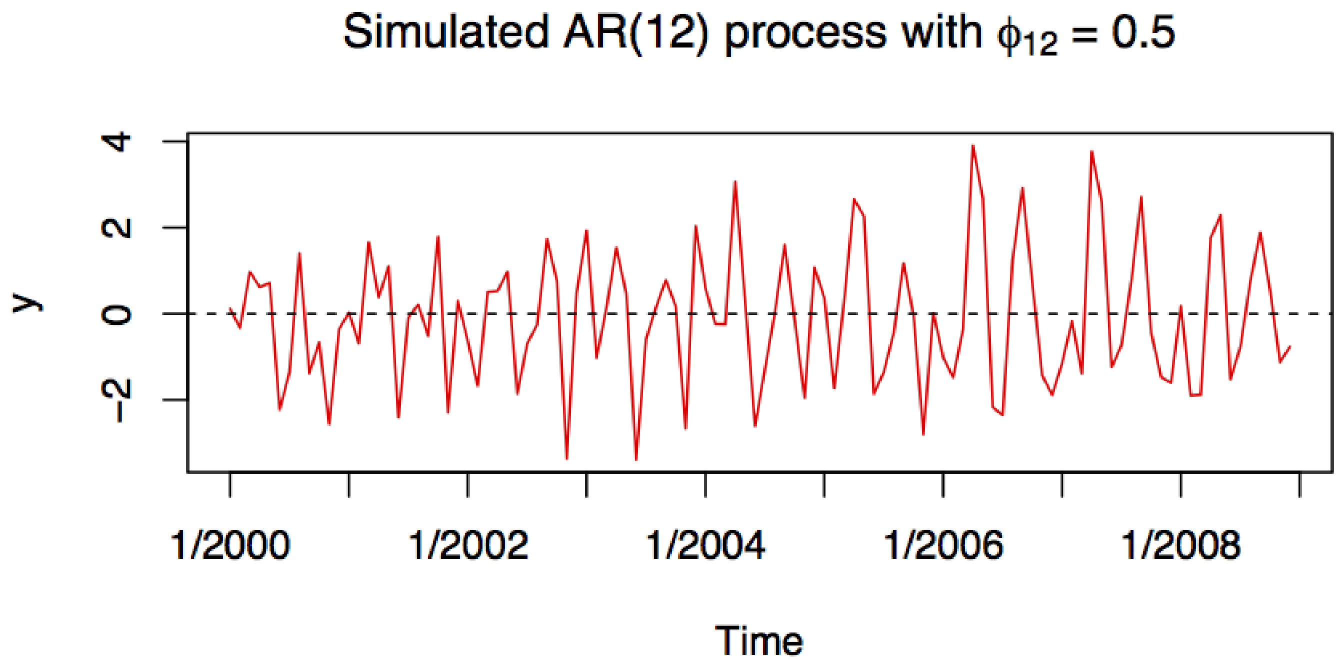
Strong dependence on the prior year's months is now obvious



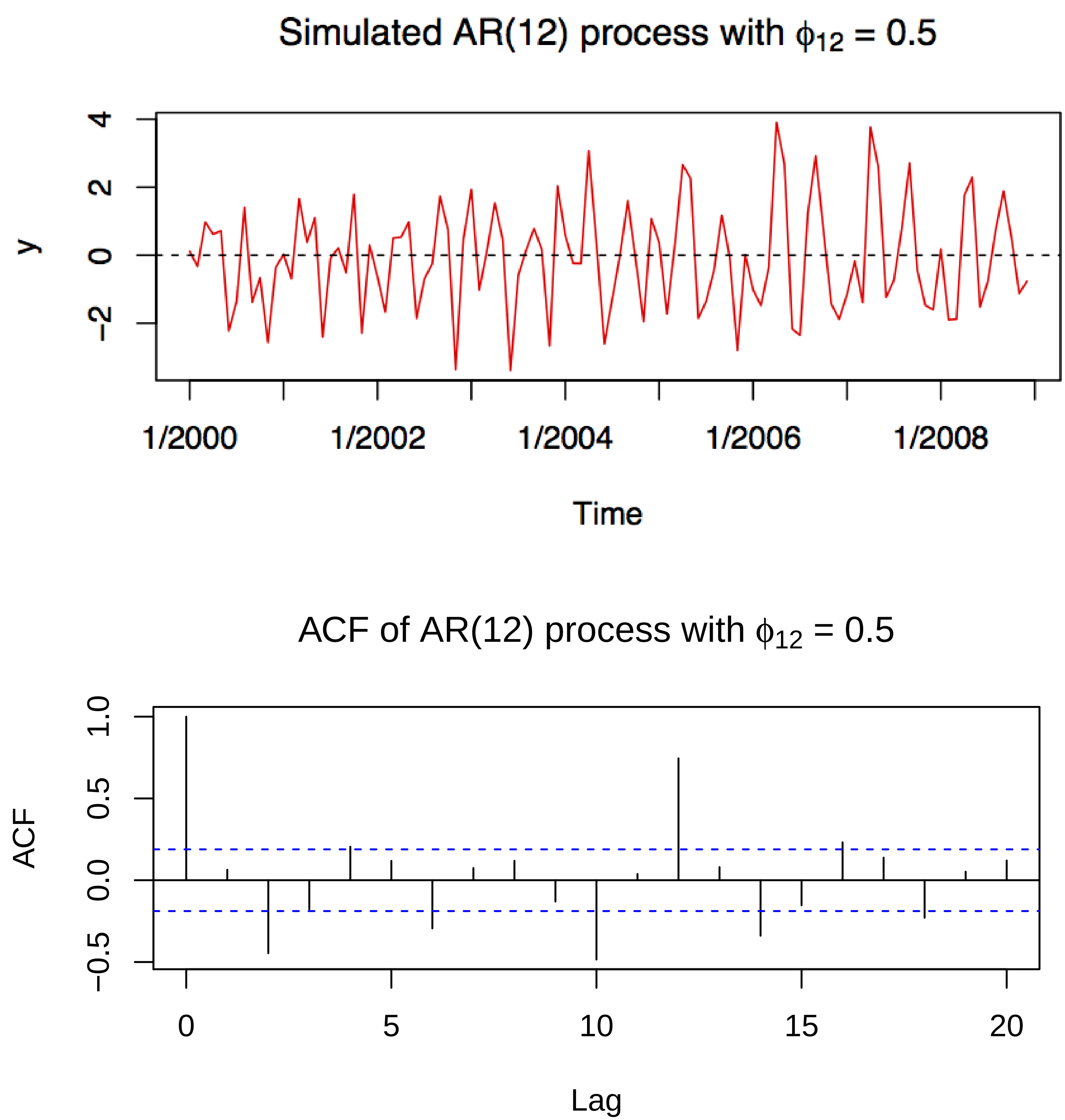
"Imagine we had an ARMA(12,0) process on monthly data
in which $\phi_1 = \phi_2 = \ldots \phi_1 1 = 0$ and $\phi_{12} = 0.5$"

The notation above is cumbersome, so let's call this an ARMA$(1,0)_{12}$ process instead

We will understand the subscript in ARMA$(p,q)_k$
to indicate a further lag applied to the entire ARMA structure

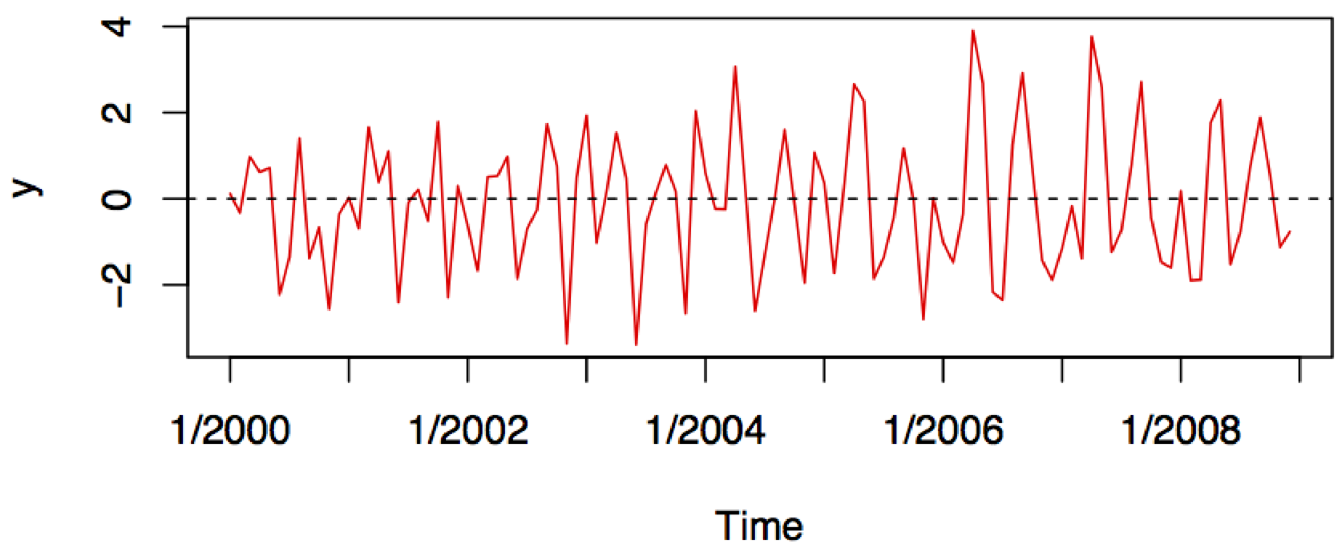


We can apply our ACF and PACF tools as usual, expecting to find lags at 12

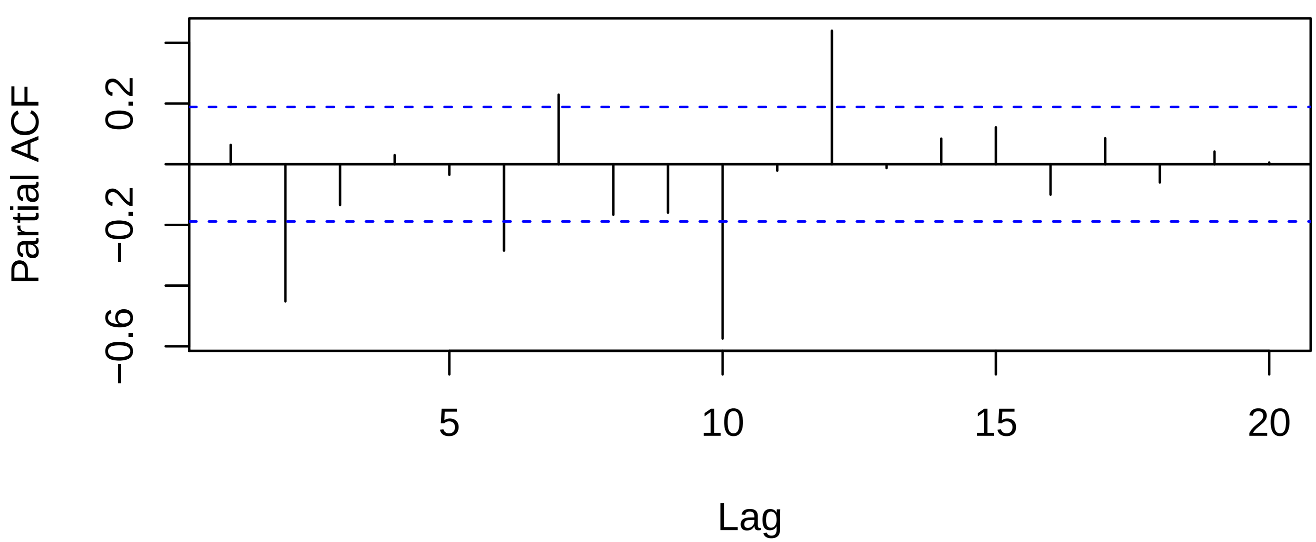


Simulated AR(12) process with $\phi_{12} = 0.5$
y
4
2
0
−2
1/2000
1/2002
1/2004
1/2006
1/2008
Time
ACF of AR(12) process with $\phi_{12} = 0.5$
ACF
1.0
0.5
0.0
−0.5
0
5
10
15
20
Lag

Simulated AR(12) process with $\phi_{12} = 0.5$



PACF of AR(12) process with $\phi_{12} = 0.5$



Seasonality with ARMA

ARMA(1,0)$_k$ processes tend to show stronger correlations at lag k (and $2k$, $3k$, etc.) on the ACF correlogram. . .

. . . and evidence of a direct effect of lag k on the PACF correlogram

Lags this far out might be noise in shorter time series –
important to have a reason to suspect cycles

In this example (the first I made), ACF and PACF are somewhat misleading–
the time series plot by year is *not*

We could combine ordinary ARMA processes with seasonally lagged ones

Thus an ARMA(1,0)(1,0)$_{12}$ process would include a lag for the last period *and* for the corresponding month of last year

That is, if t indicates February 2015,
then y_t will be a function of January 2015 $(t-1)$ *and*
February 2014 $(t-12)$, each with its own ϕ.

Seasonality with ARMA

Suppose you have an ARMA(1,0)$_k$ and the true ϕ_k is 1.0

Seasonality can be non-stationary!

This could cause the same set of errors in inference as before

When we develop models for nonstationary data,
those same models will help with nonstationarity in seasonal lags

Seasonality: Final thoughts

Seasonality is often a nuisance that worsens as our data become more finely grained

As one switches from annual to monthly to weekly to daily data,
more cycles (within cycles) may emerge

Dealing with these cycles is one price for using more information rich data
(cycles are often averaged out in lower frequency time series)

For some, dealing with seasonality is the goal:
e.g., for producing "seasonally adjusted" time series

If this is your goal, be aware that there are many more diagnostic and adjustment techniques for cycles than we can cover here

Seasonality: Final thoughts

Bonus thought:
if you have *very* high frequency data – by the minute, second, or millisecond–
you will encounter new time series behaviors. . .

burstiness: time series that have sudden bursts of activity, which then disappear (in a sense, the real "action" is whether there is an "event" creating a sequence of large y's, rather than the exact value of y_t)

very long lags: what happened 10,000 periods ago may be newly revelant with data measured by the second, but not in data measured by the year

These issues are also mostly beyond the scope of the course,
though the discussion regarding ARFIMA in Box-Steffensmeier et al is one place to start

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 is known as a distributed lag model

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