

Stochastic processes III

The ASTA team

Contents

1 Models with exogenous variables	1
1.1 Exogenous variables	1
1.2 Regression models with exogenous variables	1
1.3 Example	2
1.4 Simulation of the example	2
1.5 Estimation and model checking	3
1.6 Prediction	5
1.7 An example with delay	6
1.8 ARMAX models	9
2 Discrete vs. continuous time	10
2.1 Continuous time stochastic processes	10
2.2 Wiener process	10
2.3 Stochastic differential equations	11
2.4 Stochastic differential equations	12

1 Models with exogenous variables

1.1 Exogenous variables

- The ARMA processes are flexible models for x_t for $t = 1 \dots, n$ evolving randomly over time, but it does not include the possibility that anything is influencing x_t
- An exogeneous variable is another variable, say y_t , that influences the behaviour of x_t
- Here y_t may be another stochastic process, which we do not model, but only consider as given, or it might be something we can control.



1.2 Regression models with exogenous variables

- We can combine regression models with ARMA models to obtain a stochastic process which is influenced by exogenous variables.
- Consider a linear regression, but where the noise term is an ARMA process:

$$y_t = \gamma x_t + \epsilon_t, \quad \alpha(B)\epsilon_t = \beta(B)w_t$$

- If we isolate y_t and insert into the ARMA expression, we get something that looks more like an ARMA process but with y_t adjusted by the exogenous variable:

$$\alpha(B)(y_t - \gamma x_t) = \beta(B)w_t$$

- Modelling a dataset with this model will allow us to simultaneously include influence by earlier times of the process itself, but also from external sources.
- The purpose of fitting such a model is both to obtain a good model for the evolution of the data and to obtain an understanding of the relation between y_t and x_t .
- Note that here γ is a single number, and x_t is a single stochastic process, but we can also include multiple stochastic processes, and let γ be a vector instead.

1.3 Example

- As an example consider a simple linear regression combined with an AR(1) process for noise terms:

$$y_t = \gamma x_t + \epsilon_t, \quad \epsilon_t = \alpha_1 \epsilon_{t-1} + w_t$$

or

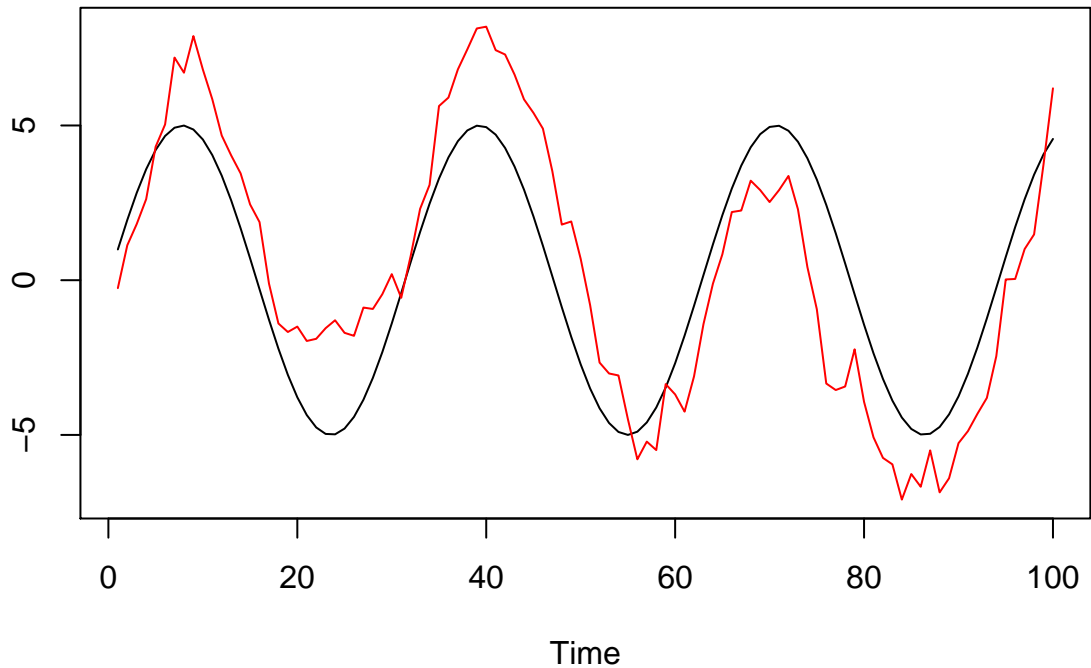
$$y_t = \alpha_1 y_{t-1} + \gamma(x_t - \alpha_1 x_{t-1}) + w_t$$

- Notice that the model behaves like an AR(1) process, but instead of having a constant mean of 0, its mean is constantly adjusted by the exogenous variable.

1.4 Simulation of the example

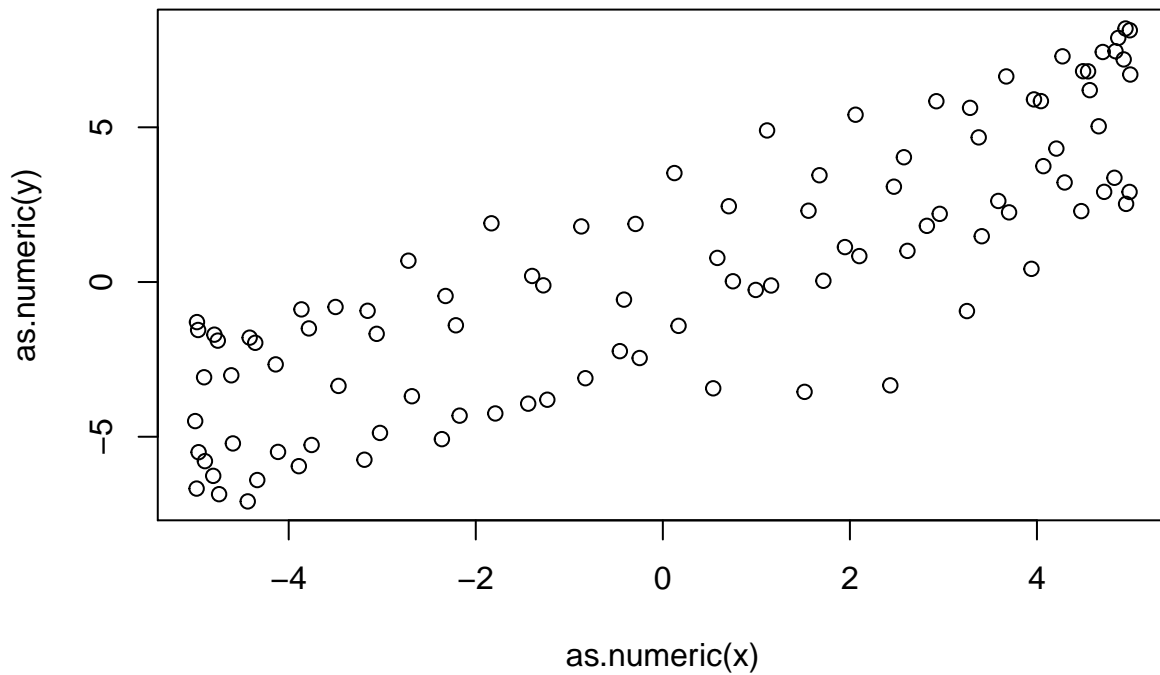
- We simulate some data resembling the example, where we just let x_t follow a sine curve:

```
alpha = 0.9; gamma = 1; n = 100
x = as.ts(5*sin(1:n/5))
eps = arima.sim(model=list(ar=alpha),n=n)
y = gamma*x+eps
ts.plot(x,y,col=1:2)
```



- We should think of the red curve as some data we want to model, and the black curve as another variable which we believe may influence the data. - We can also plot x_t against y_t to get a view of the relation between the two variables rather than the time evolution.

```
plot(as.numeric(x),as.numeric(y))
```



1.5 Estimation and model checking

- We can estimate the parameters using the arima function in R.

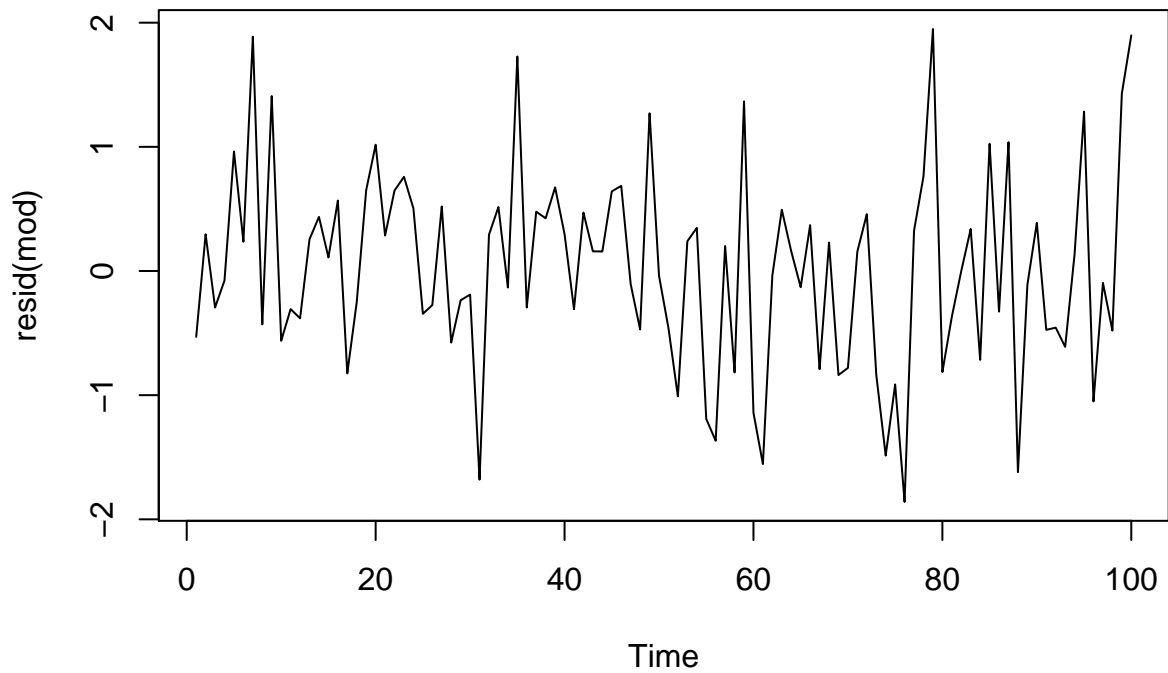
- We fit a linear regression model with AR(1) noise to the simulated data (i.e. the true model used for simulation):

```
mod=arima(y,order=c(1,0,0),xreg=x); mod
```

```
##  
## Call:  
## arima(x = y, order = c(1, 0, 0), xreg = x)  
##  
## Coefficients:  
##      ar1  intercept      x  
##    0.9312    0.1709  1.0364  
## s.e. 0.0331    1.0367  0.1098  
##  
## sigma^2 estimated as 0.6428:  log likelihood = -120.81,  aic = 249.62
```

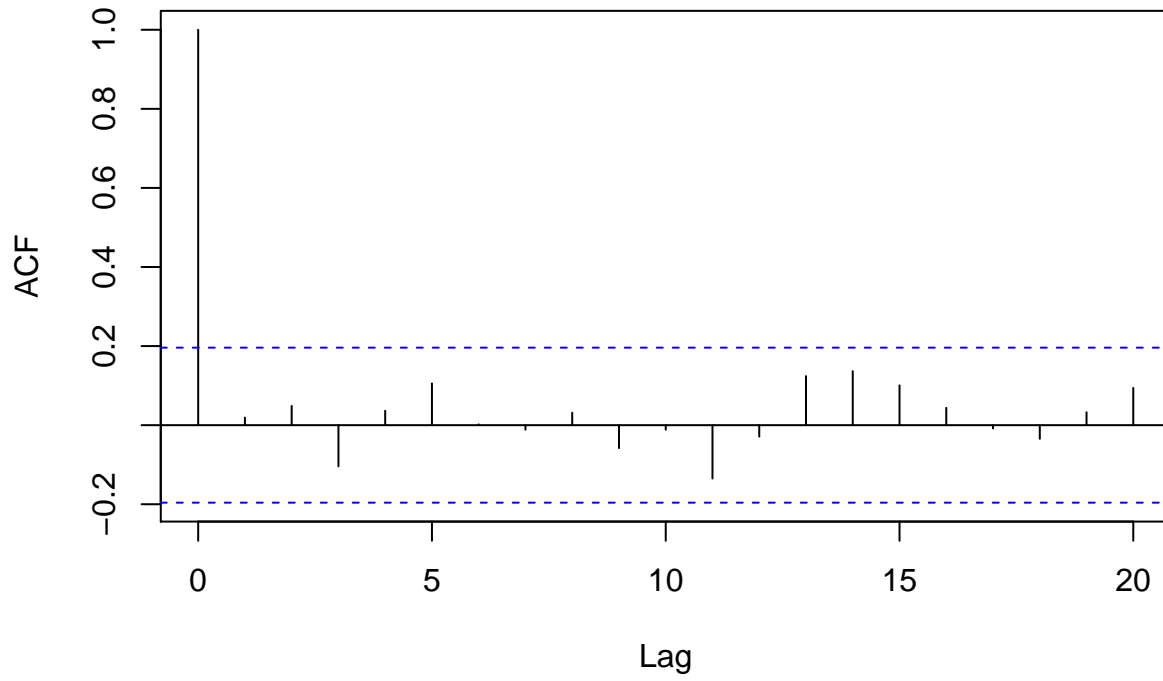
- The output looks as usual, but we have an additional term describing the relation between y_t and x_t .
- Model checking can be performed as usual.

```
plot(resid(mod))
```



```
acf(resid(mod))
```

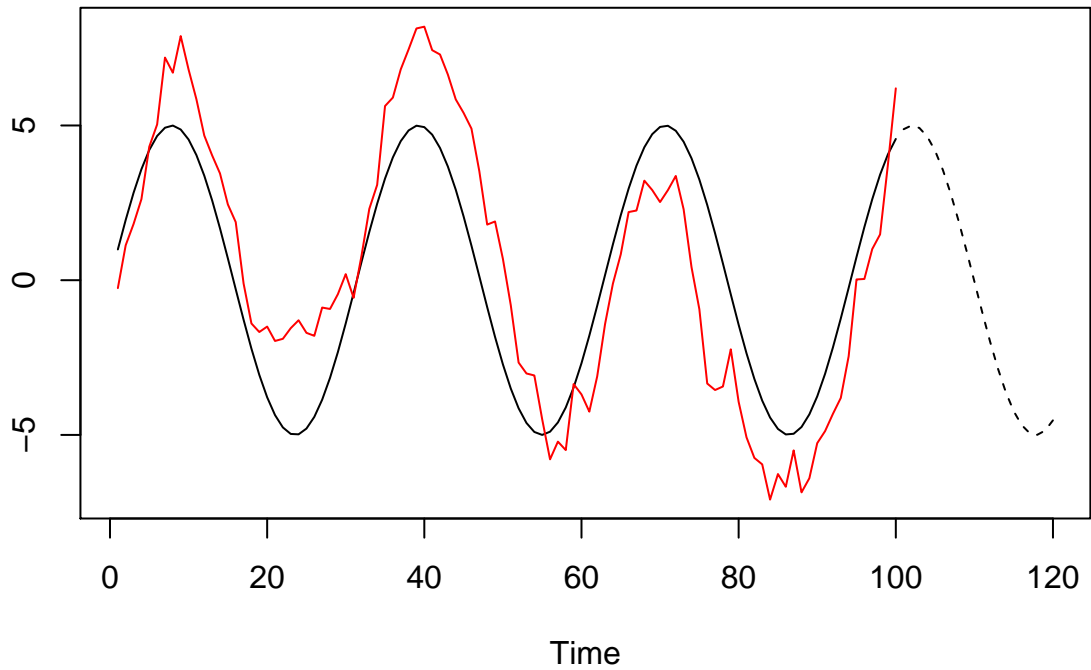
Series resid(mod)



1.6 Prediction

- Prediction can only be performed if we know the behavior of x_t for future time points, for example if we are able to control it.
- For the previous example we assume that the sine curve continues:

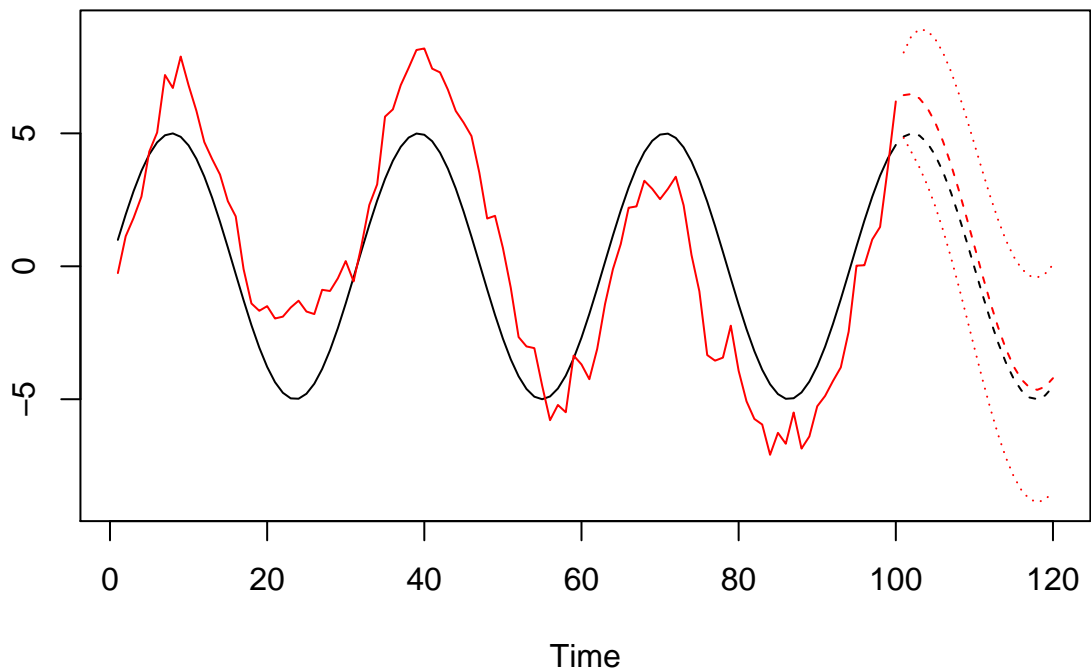
```
mnew = 20  
xnew = lag(as.ts(5*sin(((n+1):(n+mnew))/5)),-n)  
ts.plot(x,y,xnew,col=c(1,2,1),lty=c(1,1,2))
```



predict function.

- we use the

```
p = predict(mod, n.ahead=nnew, newxreg=xnew)
ts.plot(x, y, xnew, p$pred, p$pred+2*p$se, p$pred-2*p$se, col=c(1,2,1,2,2,2), lty=c(1,1,2,2,3,3))
```



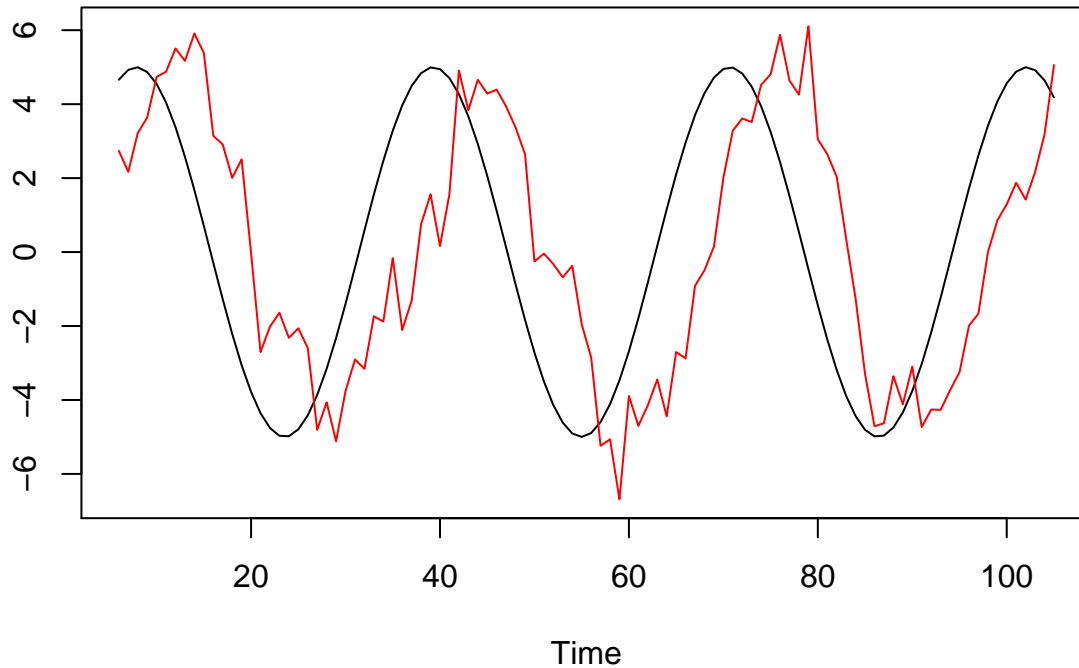
1.7 An example with delay

- When we observe two stochastic processes, and want to model one conditional on the other one with a regression model with ARMA noise, we may encounter the problem that there is a delay before the

influence of x_t has an effect on y_t , and we may not know what this delay is.

- That is, we want to model y_t based on x_{t-k} for some unknown value of k .
- First we simulate a dataset with a built-in delay, and then we model this afterwards.

```
alpha = 0.5; gamma = 1; n = 100; delay = 5
x = as.ts(5*sin(1:(n+delay)/5))
eps = arima.sim(model=list(ar=alpha),n=n+delay)
y = gamma*lag(x,-delay)+eps
dat = ts.intersect(x,y)
ts.plot(dat[,1],dat[,2],col=1:2)
```



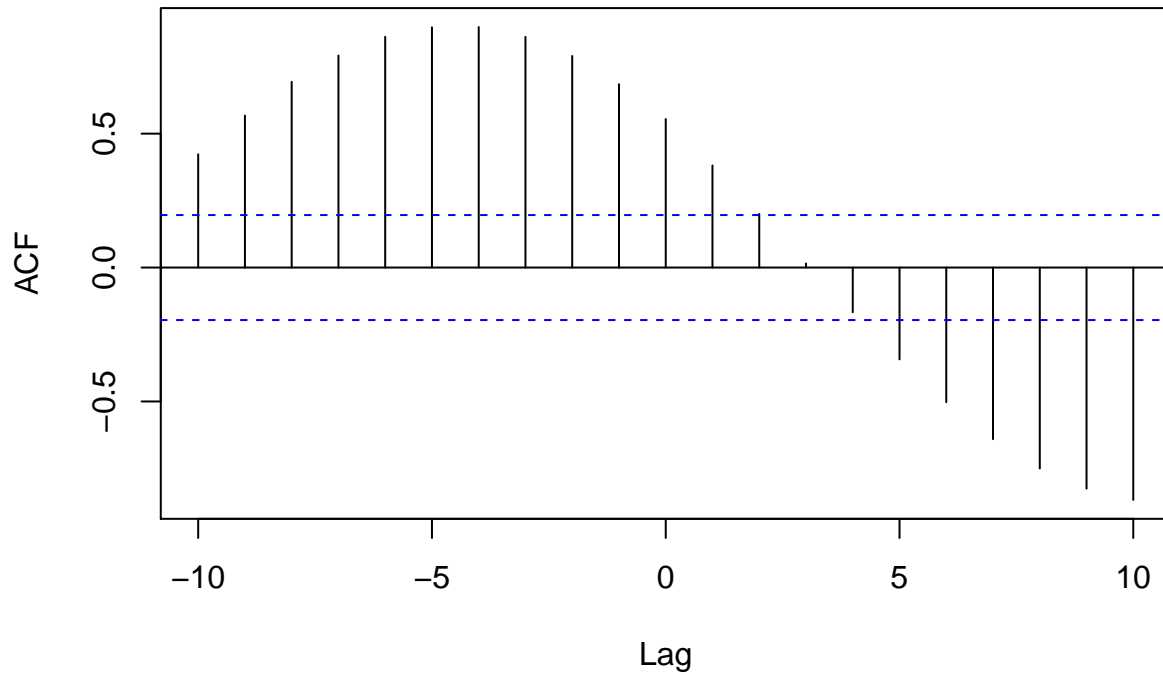
- The cross correlation function is used for checking the relation between two time series at different time points:

$$\rho_{xy}(t, t+k) = \frac{E[(x_t - \mu_t^x)(y_{t+k} - \mu_{t+k}^y)]}{\sigma_t^x \sigma_{t+k}^y}$$

Values that are close to 1 or -1 indicate that the time series are closely related if one is delayed by k time steps.

```
cc = ccf(dat[,1],dat[,2],lag.max=10)
```

dat[, 1] & dat[, 2]

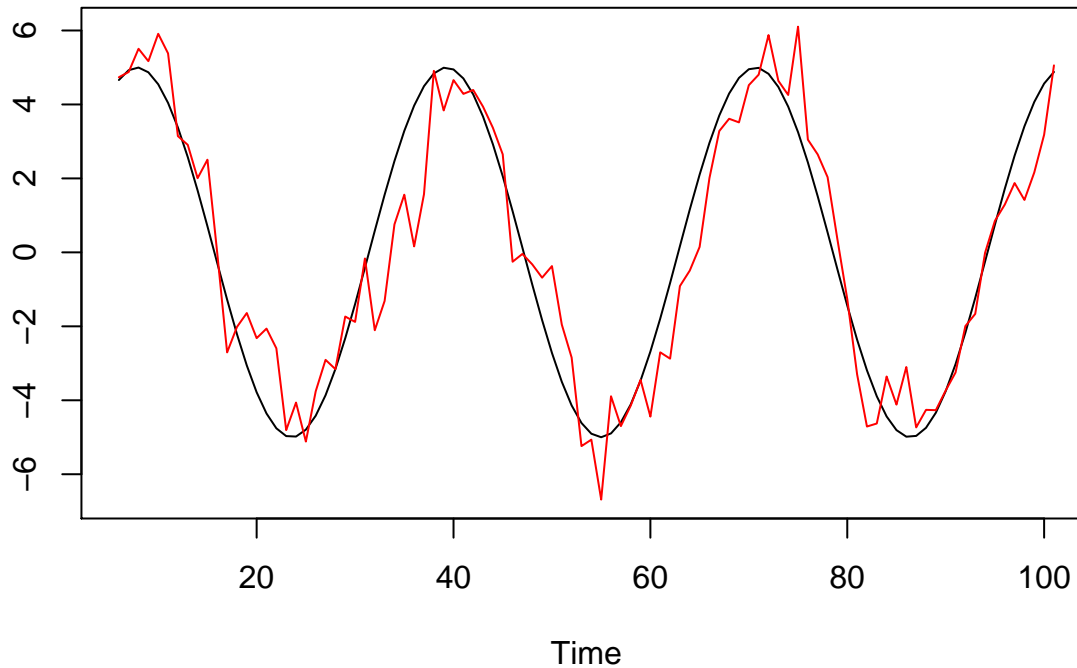


```
estlag = cc$lag[which(cc$acf==max(cc$acf))];estlag
```

```
## [1] -4
```

- Plotting the data with this lags can be useful to check the choice:

```
dat2 = ts.intersect(as.ts(dat[,1]),lag(dat[,2],k=-estlag))  
ts.plot(dat2[,1],dat2[,2],col=1:2)
```

- We can now fit a model based on the lag:

```
mod=arima(dat2[,2],order=c(1,0,0),xreg=dat2[,1]); mod
```

```
##
## Call:
## arima(x = dat2[, 2], order = c(1, 0, 0), xreg = dat2[, 1])
##
## Coefficients:
##      ar1  intercept  dat2[, 1]
##  0.5595   -0.0524   0.9128
## s.e.  0.0837    0.2225   0.0588
##
## sigma^2 estimated as 0.9432:  log likelihood = -133.6,  aic = 275.2
```

1.8 ARMAX models

- An alternative way of including exogenous variables into an ARMA model is an ARMAX model.
- The $ARMAX(p, q, b)$ model is an $ARMA(p, q)$ model including b terms of an exogenous variable, i.e. it is defined by

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{i=1}^b \gamma_i x_{t-i} + w_t + \sum_{i=1}^q \beta_i w_{t-i}$$

- Using the backshift operator, this can be written as

$$\alpha(B)y_t = \gamma(B)x_t + \beta(B)w_t$$

with $\alpha(B) = 1 - \sum_{i=1}^p \alpha_i B^i$, $\beta(B) = 1 + \sum_{i=1}^q \beta_i B^i$, and $\gamma(B) = \sum_{i=1}^b \gamma_i B^i$.

- Compare with the regression with ARMA noise:

$$\alpha(B)(y_t - \gamma x_t) = \beta(B)w_t \quad \Rightarrow \quad \alpha(B)y_t = \alpha(B)\gamma x_t + \beta(B)w_t$$

Notice that the difference is only how the model includes the exogenous variable. It is mostly a matter of taste which kind of model you should choose. Only the regression with ARMA noise is included into R as standard.

2 Discrete vs. continuous time

There are two fundamentally different model classes for time series data.

- Discrete time stochastic processes
- Continuous time stochastic processes

So far we have only looked at the discrete time case. We will finish today's lecture by looking a bit at the continuous time case, just to give you an idea of this topic.

2.1 Continuous time stochastic processes

In this setup we see the underlying X_t as a continuous function of t for t in some interval $[0, T]$.

In principle we imagine that there are infinitely many data points, simply because there are infinitely many time points between 0 and T . Of course this is never true: In practice we will always only have finitely many data points.

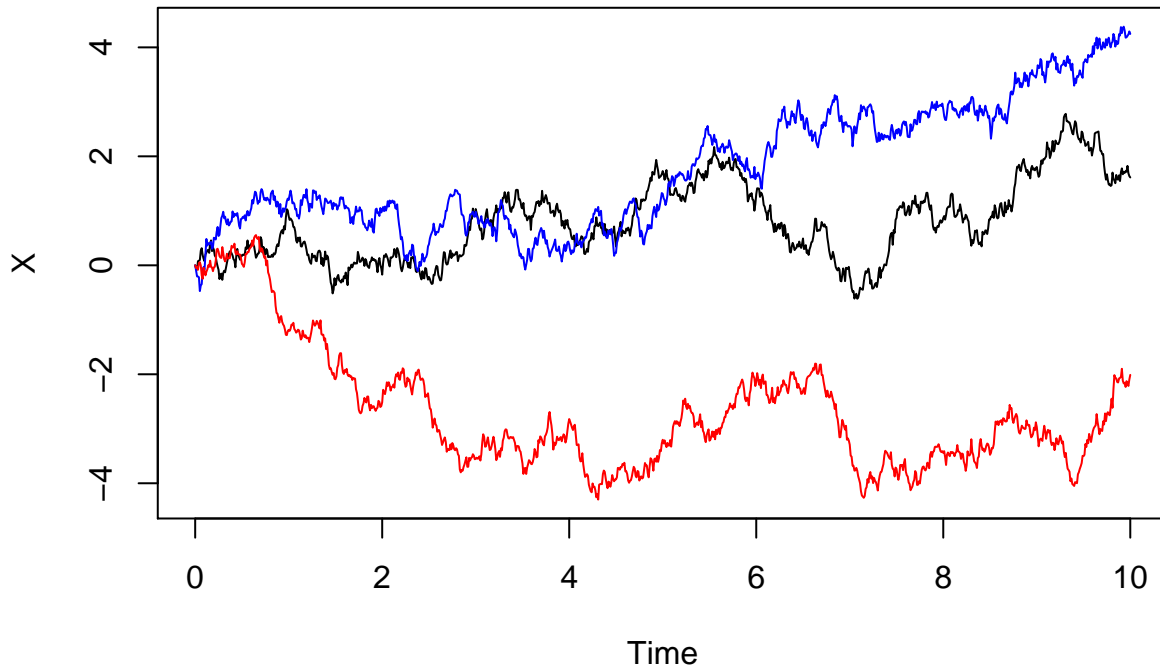
But it makes sense to believe that the real data actually contains all the data points. We are just not able to measure them (and to store them in a computer).

With a model for all datapoints, we are - through simulation - able to describe the behaviour of data. Also between the observations.

2.2 Wiener process

A key example of a process in continuous time will be the so-called Wiener process.

Three simulated realizations (black, blue and red) of this process can be seen here



A Wiener process has the following properties:

- It starts in 0: $W_0 = 0$.
- It has independent increments: For $0 < s < t$ it holds that $W_t - W_s$ is independent of everything that has happened up to time s , that is W_u for all $u \leq s$.
- It has normally distributed increments: For $0 < s < t$ it holds that the increment $W_t - W_s$ is normally distributed with variance $t - s$:

$$W_t - W_s \sim \text{Normal}(\mu = 0, \sigma^2 = t - s).$$

The intuition of this process is that it somehow changes direction all the time: How the process changes after time s will be independent of what has happened before time s . So whether the process should increase or decrease after s will not be affected by how much it was increasing or decreasing before.

This gives the very bumpy behaviour over time.

2.3 Stochastic differential equations

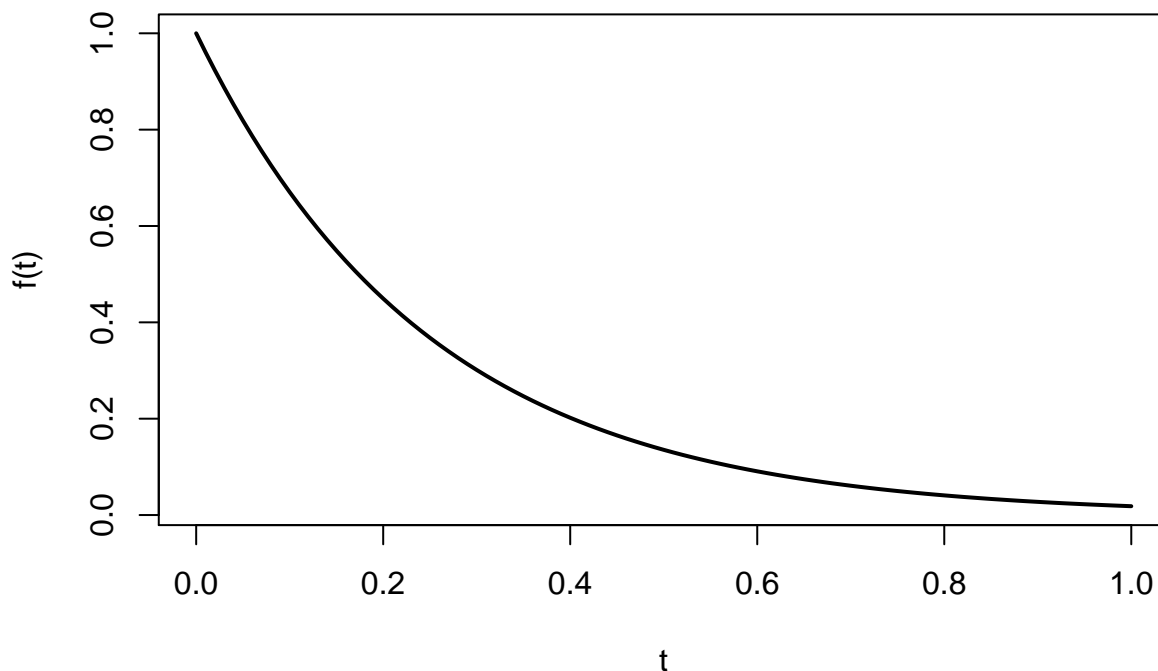
A common way to define a continuous time stochastic process model is through a stochastic differential equation (SDE) which we will turn to shortly, but before doing so we will recall some basic things about ordinary differential equations.

Suppose f is an unknown differentiable function and $f(0) = 1$. Recall the mathematical description of a differential equation

$$\frac{df(t)}{dt} = -4f(t)$$

With the condition $f(0) = 1$ this equation has the solution

$$f(t) = \exp(-4t)$$



With a slightly unusual notation we can rewrite this as

$$df(t) = -4 \cdot f(t) dt$$

This equation has the following (hopefully intuitive) interpretation:

- When time increases by a small amount dt the value of f changes (approximately) by $-4f(t) dt$.

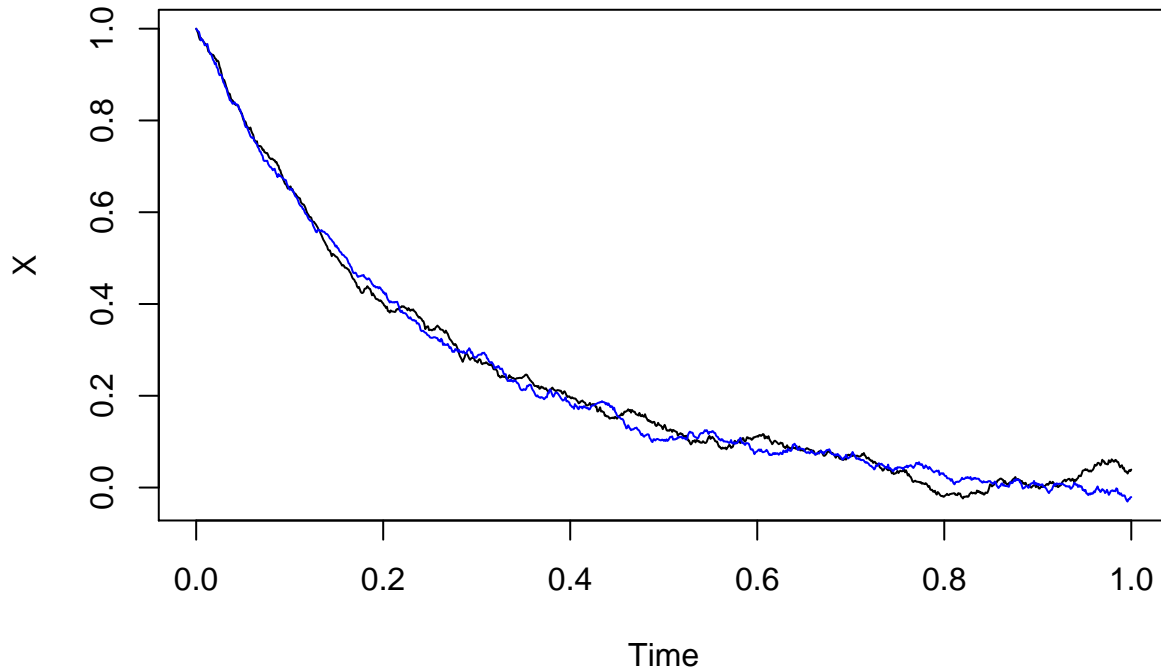
So when t is increased, then f is decreased. And the decrease is determined by the value of $f(t)$. That is why f decreases slower and slower, when t is increased.

We say that the function has a **drift** towards zero, and this drift is determined by the value of the function.

2.4 Stochastic differential equations

It will probably never be true that data behaves exactly like the exponentially decreasing curve on the previous slide.

Instead we will consider a model, where some random noise from a Wiener process has been added. Two different (black/blue) simulated realizations can be seen below



The type of process that is simulated above is described formally by the equation

$$dX_t = -4X_t dt + 0.1dW_t$$

This is called a **Stochastic Differential Equation** (SDE), and the processes simulated above are called solutions of the stochastic differential equation.

The SDE $dX_t = -4X_t dt + 0.1dW_t$ has two terms:

- $-4X_t dt$ is the **drift term**.
- $0.1dW_t$ is the **diffusion term**.

The intuition behind this notation is very similar to the intuition in the equation $df(t) = -4 \cdot f(t) dt$ for an ordinary differential equation. When the time is increased by the small amount dt , then the process X_t is increased by $-4X_t dt$ AND by how much the process $0.1W_t$ has increased on the time interval $[t, t + dt]$.

So this process has a **drift** towards zero, but it is also pushed in a random direction (either up or down) by the Wiener process (more precisely, the process $0.1W_t$)