Intro to Time Series Analysis

Michael Winton

June 16, 2018

Introduction to Time Series Analysis

In classical linear regression, we focus on using independent variables to characterize a dependent variable. In other words, we would be characterizing the conditional mean function of the dependent variable, conditional on the independent variables, relying on iid assumption.

In univariate TSA, we focus on characterizing one series, with observations that are typically dependent on each other. It's that dependency that we want to characterize. We're analyzing T observations of a single variable, so we have a $Tx1$ vector (contrast to the nxk matrix for CLR).

Terminology

In theory a time series begins in infinite past and continues into infinite future (important for deriving properties of TS models).

A *Stochastic Process* is a statistical phenomenon that evolves according to probability laws. A *time series* is a discrete realization of this "data generating process," a set of observations generated sequentially in time. Time indices are $t_1, t_2, ...t_N$ and observations are $z(t_1)...z(t_N)$. If they're equidistant, we simplify to $z_1, ...z_N$.

A *Discrete-Time Stochastic Process* is a sequence of random variables (e.g. $\{...z_{-1}, z_0, z_1, ...z_N\}$). A finite subset of this realization (e.g. $\{z_1, ...z_N\}$) is called a *sample path*, which can be considered a collection of observations.

IMPORTANT: in CLR, a cross-section of observations is considered *many realizations or draws*. In time series analysis, a collection of observations is considered only *one draw* or *one realization*.

General Approach to TSA

1. Based on theory/subject-matter knowledge, consider a useful class of models
2. Collect and clean data
3. Perform ETSDA (graphical and tabular methods)
4. Examine and statistically test whether the series is *stationary*.
5. If it's not, transform it (e.g. detrending, seasonality removal, log, difference transforms)
6. Model using a *stationary* or *integrated* TS model.
7. Examine validity of model's assumptions.
8. Among valid models, choose the best one.
9. Conduct forecasting!

Common Empirical TS Patterns

Pattern 1: Trend with flutuation around the trend

- Systematic upward trend
- Seasonality around trend (also visible in box plots - e.g. by month)
- Fluctuation increases over time

Pattern 2: Change in Structure

- Long period of range-bounded fluctuation, followed by a consistent upward trend w/ variation

Pattern 3: Variation around Stable Mean

- Fluctuation generally stable around a fixed mean
- Spikes in variation from time to time ("volatlity clustering")

Pattern 4: Periodicity

- Periodicity - cycles appear with fixed frequency

Simple TS Models

Model 1: white noise

- collection of uncorrelated random variables with mean=0 and some variance.
- "Gaussian WN": these RVs are iid and have standard normal distribution (popular model)
- the most fundamental component in TS modeling and assumption testing
- a deterministic dynamic model can be transformed into a deterministic stochastic model using the addition of white noise.
- histogram is a useful tool, but shouldn't be used alone since it loses the time element!

Model 2: symmetric, equal-weight, moving average model

- common way to smooth / remove volatility from a TS where white noise was introduced
- can be used to generate dependency between observations
- centered moving average can smooth out variation in order to make trend visible
- length of moving average can be chosen to smooth out seasonality

Model 3: autoregressive models (specifically, order p=2)

- establishes relationship between observed values and past values and white noise component
- we attach a coefficient to past values that dampens over time
- model appears fairly symmetric (histogram)

Model 4: random walk and deterministic trend

- model includes a deterministic trend, along with some random walk along the way
- random walk can happen with or without drift (e.g. upward drift of historical prices for a stock)

- these models are generally very persistent

Statistical Principles and Measures of Dependency

A complete description of a TS as a collection of k random variables requires a joint distribution function:

$$F(c_1, c_2, ...c_n) = P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, ...x_{t_n} \leq c_n)$$

Characterizing this in its general form is impossible. A single realization of a TS doesn't offer enough information to characterize the underlying joint distribution function.

One of the most important probabilitistic features is the dependency structure embedded in joint distributions.

The *mean function* of a stochastic process is:

$$\mu_x(t) = E(x_t) = \int_{-\infty}^{\infty} x_t f_t(x_t) dt$$

If the function is constant over time, then the underlying stochastic process is said to be *stationary in the mean.* The *variance* for a TS that is stationary in the mean is:

$$\sigma_x^2(t) = E(x_t - \mu)^t = \int_{\infty}^{\infty} (x_t - \mu)^2 f_t(x_t) dx_t$$

Note that both mean and variance are functions of the index (t_i). However, it's impossible to estimate the different variances at different points in time with only a single TS (ie. single realization of the SP). As a result, another popular assumption is *stationarity of variance.*

In the context of TSA, we speak of covariance and correlation as between multiple RVs in the same series, so we refer to them as *autocovariance* and *autocorrelation.* The autocovariance function (avcf) is:

$$\gamma_x(s,t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)] \forall s,t$$

By properties of covariance, we know that $\gamma_x(s,t) = \gamma_x(t,s)$ and $\gamma_x(s,s) = cov(x_s, x_s) = var(x_s)$.

Stationarity of mean, variance, and autocovariance produce a large class of *stationary TS models.*

Second-Order Stationarity Assumption

If a TS is "second-order stationary", that means it's stationary in both mean and variance, so $\mu_t = \mu$ and $\sigma_t^2 = \sigma^2$ for all t. Then we write the *autocovariance* function as:

$$\gamma_k = Cov(x_t, x_{t+k}) = E[(x_t - \mu)(x_{t+k} - \mu)]$$

Then the *autocorrelation* function (acf) is:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma^2}$$

If $k = 0$, then it follows that $\gamma_k = \sigma^2$ and $rho_k = 1$. **The dependence between values of a TS is important, so it's important to estimate autocorrelation with precision.**

We use *moment principles* to estimate values of acvf and acf from their sample equivalents. Note the division by T, the total number of measurements (not t).

$$\text{sample acvf: } \hat{\gamma_k} = \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \hat{x})(x_{t+k} - \hat{x})$$

$$\text{sample acf: } \hat{\rho_k} = \frac{\hat{\gamma_k}}{\hat{\gamma_0}} = \frac{\frac{1}{T}\sum_{t=1}^{T-k}(x_t - \hat{x})(x_{t+k} - \hat{x})}{\frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x})^2}$$

Partial Autocorrelation

This is a conditional correlation, conditional on other explanatory variables being accounted for in a TS model. The partial autocorrelation of a process z_t at lag k (ϕ_{kk}) is the autocorrelation between z_t and z_{t-k}, adjusting for effects of variables $z_{t-1}, z_{t-2}, ...z_{t-k+1}$. In other words, if you did a linear regression of z_t on all of the other z variables, ϕ_{kk} would be the coefficient for z_{t-k}. (Note: this kind of regression would be called an *autoregression.*)

The partial autocorrelation summarizes the dynamics of a process, and as such it's a powerful tool for identifying the *order* p of an $AR(p)$ model.

Strict vs. Weak Stationarity

A time series x_t is *strictly* stationary if the distribution is **unchanged** for any time shift. This isn't practical for any modeling. Mathematically:

$$F(x_{t_1}...x_{t_n}) = F(x_{t_1+m}...x_{t_n+m}) \forall t_1..t_n \text{ } and \text{ } m$$

A *weak stationarity* (aka. *second-order stationarity*) is more practical. A time series x_t is *weakly* stationary if its mean and variance are stationary, and its $cov(x_t, x_{t+k})$ only depends on k (can be written as $\gamma(k)$).

White Noise Model

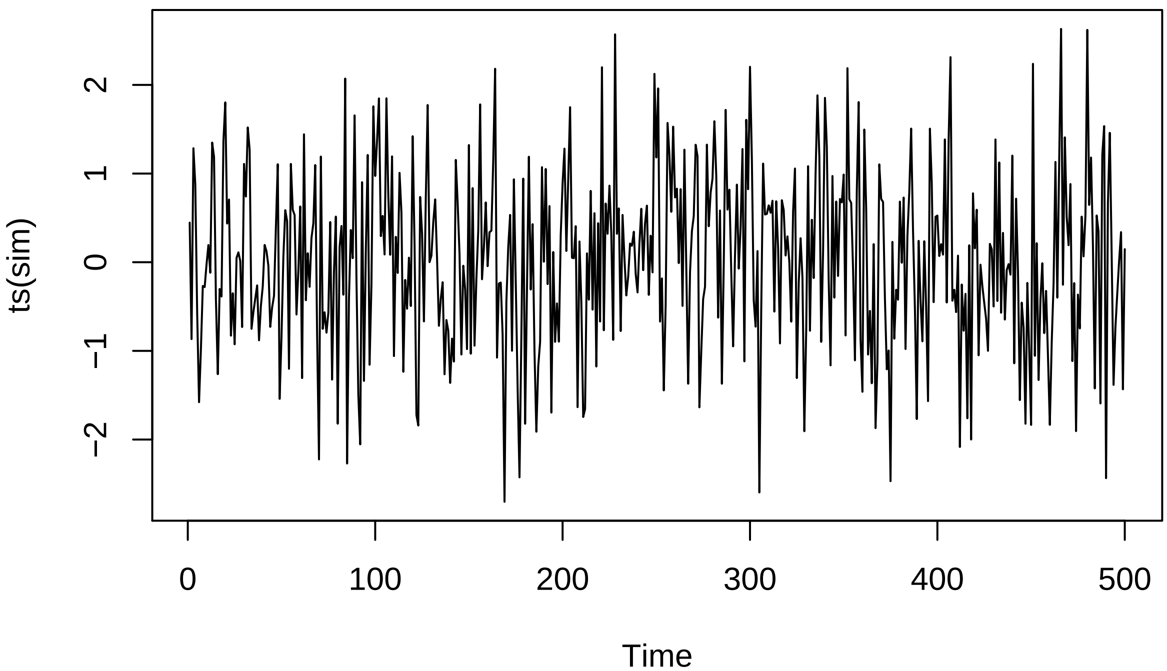
Recall that the discrete white noise process (w_t) is a sequence of RVs indexed by t that are iid and have:

$$E(w_t) = \mu_w = 0$$

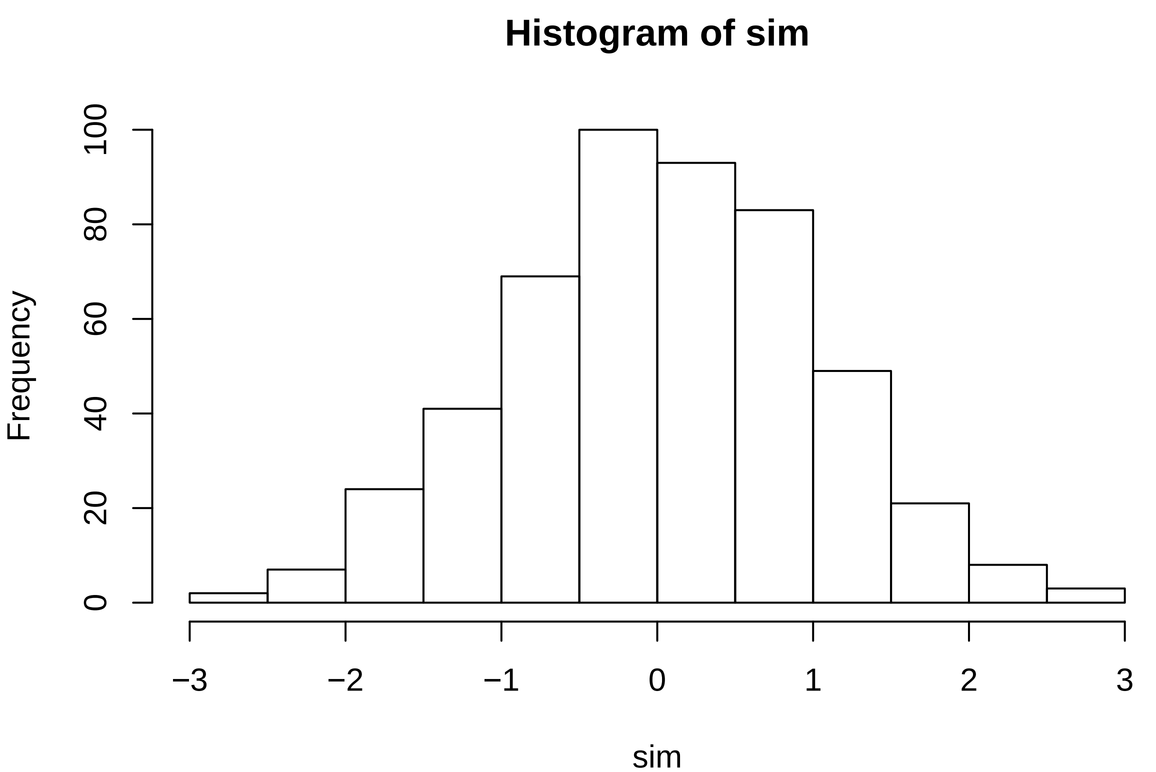
$$\gamma_k = cov(w_t, w_{t+k}) = \begin{cases} \sigma_w^2, k = 0 \\ 0, k \neq 0 \end{cases}$$

$$\rho_k = \begin{cases} 1, k = 0 \\ 0, k \neq 0 \end{cases}$$

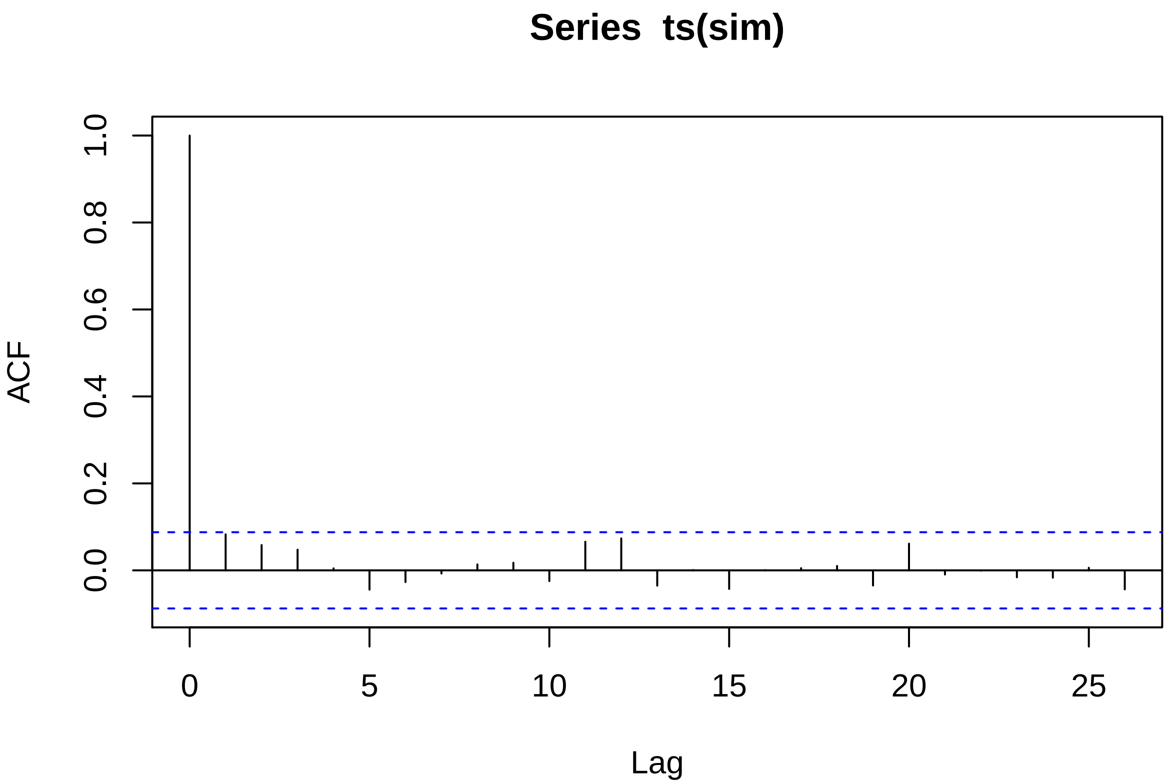
```
r   # White Noise simulation   sim <- rnorm(500, mean=0, sd=1)   layout(1:1)
plot(ts(sim))
```




```
r   hist(sim)
```



```r
acf(ts(sim))
```



As expected, the series appears random, and the ACF shows no significant correlation with any lags.

Discrete White Noise is important because we use it as a model for residuals. When we try to fit other time series models to our observed data, we use Discrete White Noise to confirm that we have eliminated any remaining serial correlation from the residuals, resulting in a good model fit.

Stochastic Model with Deterministic Linear Trend

Consider a model with a deterministic linear trend: $x_t = a + bt + w_t$, where w_t is a white noise process with mean=0 and variance = σ_w^2. The expected value of x_t is:

$$E(x_t) = E(\beta_0 + \beta_1 t + w_t) = \beta_0 + \beta_1 t + E(w_t) = \beta_0 + \beta_1 t$$
$$Var(x_t) = Var(\beta_0 + \beta_1 t + w_t) = Var(w_t) = \sigma_w^2$$

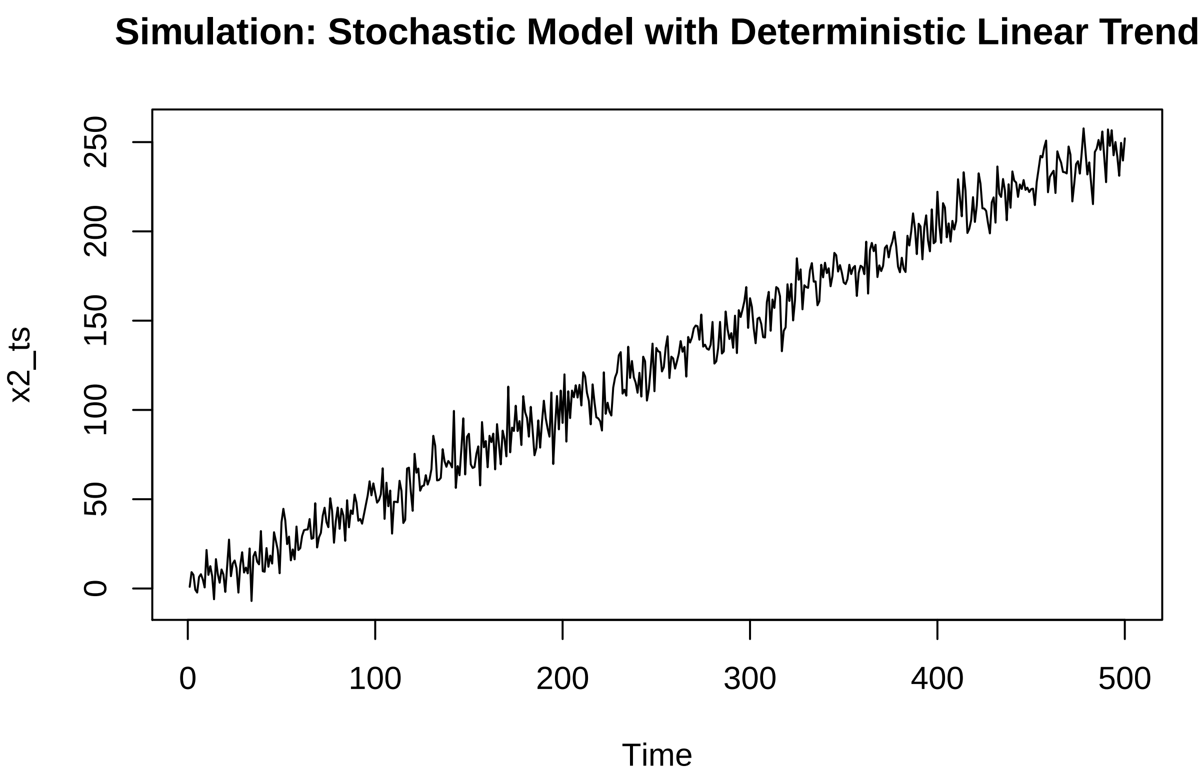
This means that the expected value (mean) is changing with time, but the variance is **not**.

To look at correlation:

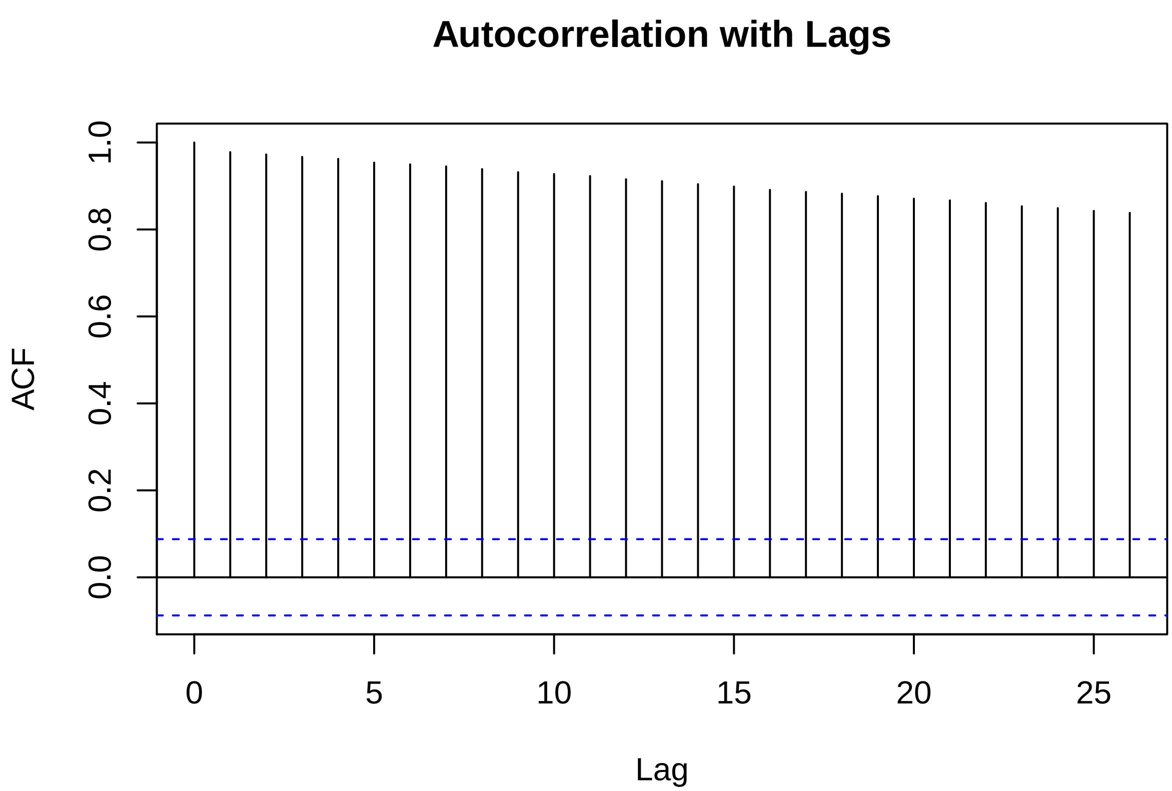
$$Cov(x_t, x_{t-1}) = Cov(\beta_0 + \beta_1 t + w_t, \beta_0 + \beta_1(t-1) + w_{t-1}) = Cov(w_t, w_{t-1}) = 0$$

Therefore, the stochastic model with a deterministic linear trend is not (strictly or weakly) stationary, but it can be transformed into a stationary model.

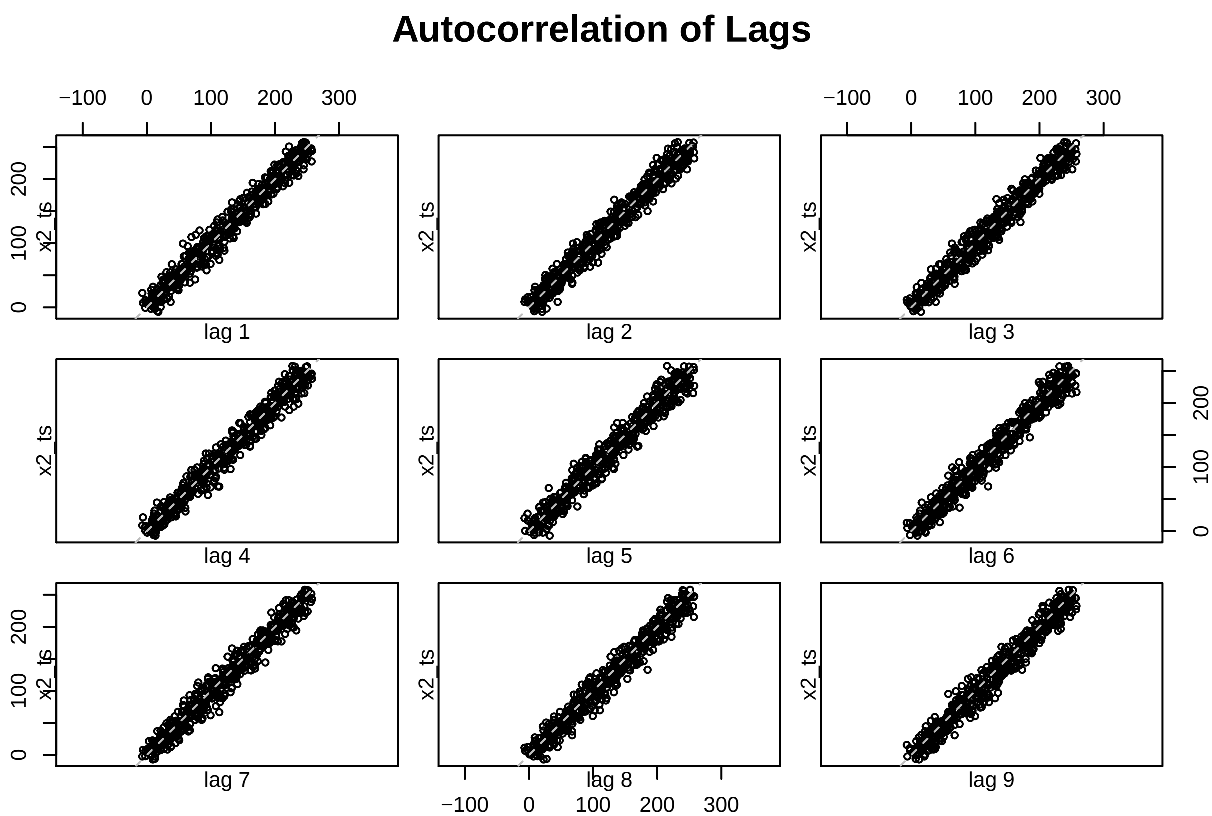
```r
# run a simulation
beta_0 <- 1
beta_1 <- 0.5
sigma_w <- 10
t <- seq(1,500)
w_t <- rnorm(500, 0, sigma_w)
x2 <- beta_0 + beta_1 * t + w_t
x2_ts <- ts(x2)

# plot
layout(1:1)
plot(x2_ts, main="Simulation: Stochastic Model with Deterministic Linear Trend")
```




```r
acf(x2_ts, main="Autocorrelation with Lags")
```




```
lag.plot(x2_ts, lags=9, main="Autocorrelation of Lags")
```



As expected, we see the mean changes with time, the variance doesn't, and the ACF plot shows that the series has *very high persistence* (ie. highly correlated with lags). The scatterplot also show clear correlation.

Moving Average Model (Order = 1)

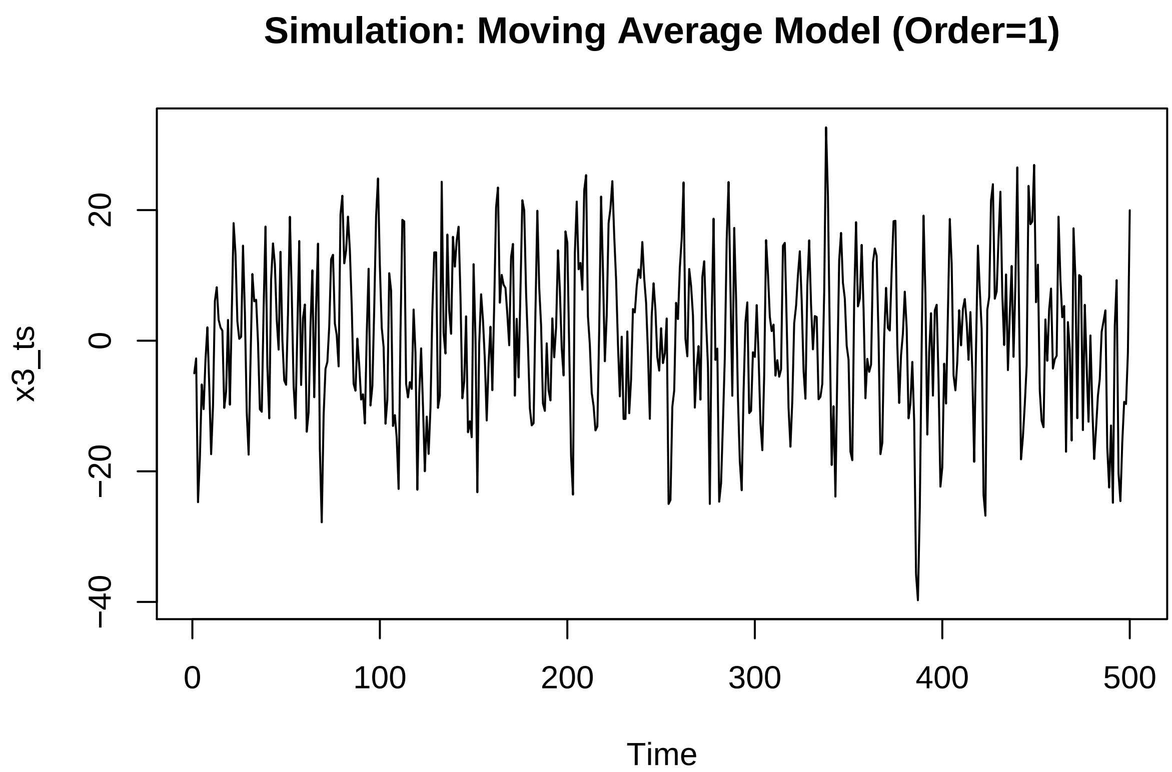
A MA(1) model takes the form: $x_t = \alpha w_t + \beta w_{t-1}$ where w_t is a discrete white noise series with mean=0 and variance σ_w^2. Sometimes the w_{t-1} term is referred to as a "shock." The expected value of x_t is:

$$E(x_t) = E(\alpha w_t + \beta w_{t-1}) = \alpha E(w_t) + \beta E(w_{t-1}) = 0$$

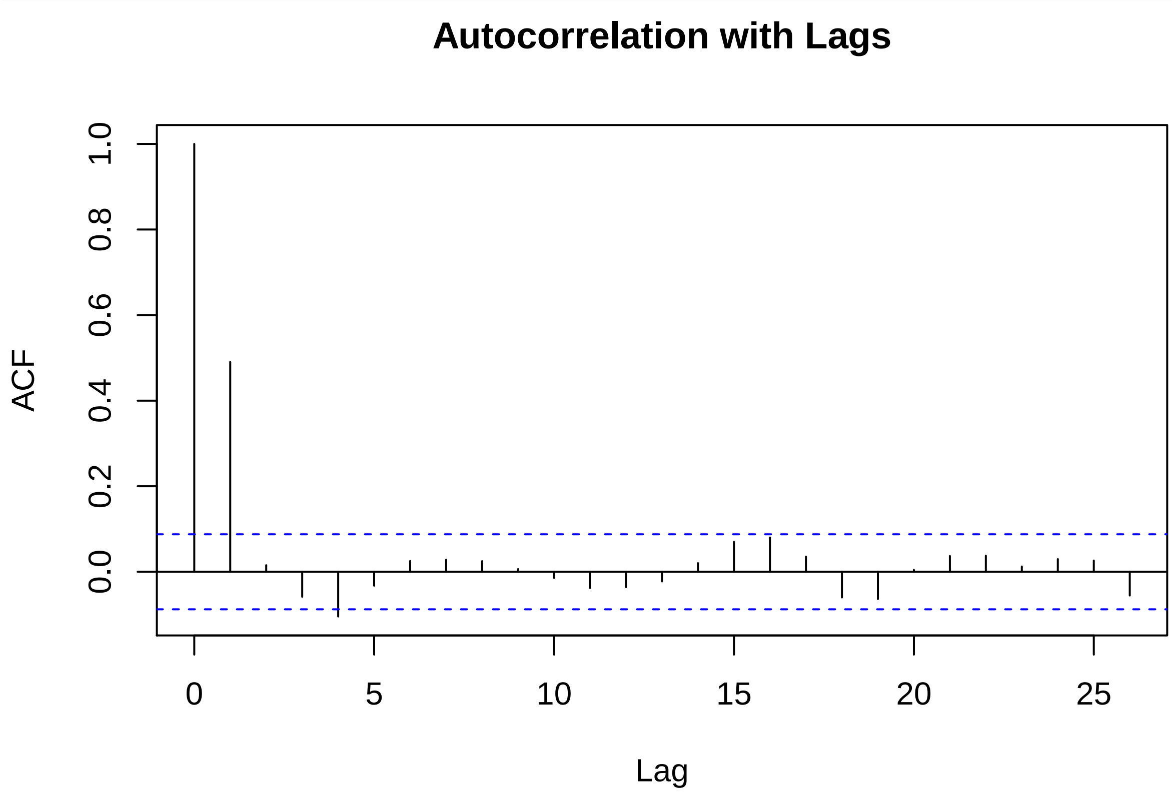
The autocorrelation function is:

$$\rho_k = \begin{cases} 1, k = 0 \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} \beta_i^2}, k = 1...q \\ 0, k > q \end{cases}$$

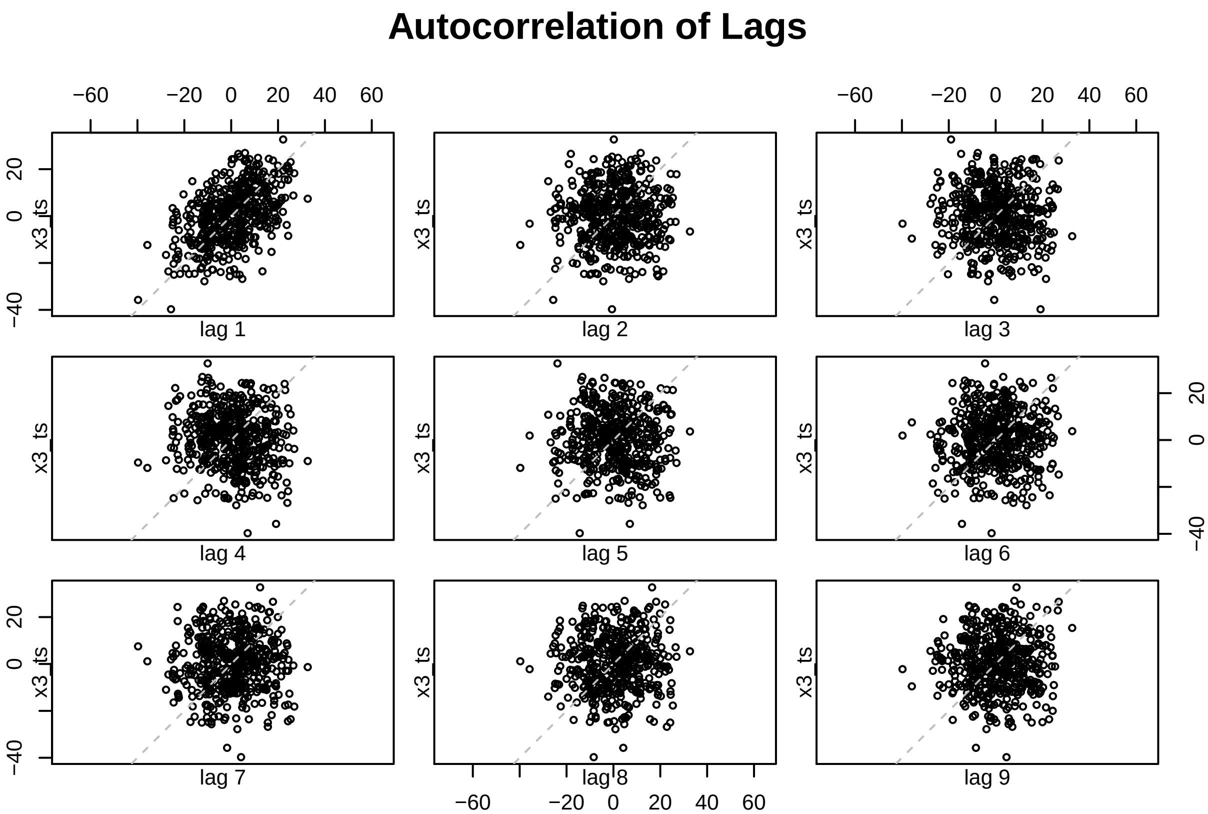
```r
# run a simulation
alpha <- 1
beta <- 0.8
sigma_w <- 10
w_t <- rnorm(500, 0, sigma_w)

x3 <- vector()
x3[1] <- w_t[1]
for (t in 2:500) {
x3[t] <- alpha * w_t[t] + beta * w_t[t-1]
}
x3_ts <- ts(x3)

# plot
layout(1:1)
plot(x3_ts, main="Simulation: Moving Average Model (Order=1)")
```




```
acf(x3_ts, main="Autocorrelation with Lags")
```




```
lag.plot(x3_ts, lags=9, main="Autocorrelation of Lags")
```



A distinguishing feature of this model is that the ACF abruptly drops off after the $order = q$ lags. The scatterplots show some correlation for $lag = 1$, but none after that.

Autoregressive Model (Order = 1)

The AR(1) model takes the form:

$$x_t = \alpha x_{t-1} + w_t$$

where w_t is a discrete white noise series with mean=0 and variance σ_w^2. Because of recursive substitution, an AR(1) process can be written as a linear process, in terms of the sum of infinite white noises:

$$x_t = \alpha x_{t-1} + w_t = \alpha(\alpha x_{t-2} + w_{t-1}) + w_t = \alpha(\alpha(\alpha x_{t-3} + w_{t-2}) + w_{t-1}) + w_t = ...$$

The result can be written more cleanly as:

$$x_t = \sum_{i=0}^{\infty} \alpha^i w_{t-i}$$

Using that form, it's easy to calculate the expected value of x_t:

$$E(x_t) = E(\sum_{i=0}^{\infty} \alpha^i w_{t-i}) = \sum_{i=0}^{\infty} E(\alpha^i w_{t-i}) = \sum_{i=0}^{\infty} \alpha^i E(w_{t-i}) = 0$$

The linear form of x_t also leads to the autocovariance:

$$ \gamma_k = Cov(x_t, x_{t+k}) = Cov (\sum_{i=0}^\infty \alpha^i w_{t-i}, \sum_{j=0}^\infty \a

From this equation, we can get the autocorrelation function:

$$ \rho_k(t) =\frac{\gamma_k}{\gamma_0} = \bigg[\frac{\alpha^k \sigma^2}{(1-\alpha^2)}\bigg]

A theoretical $AR(1)$ model with a _positive correlation_ decays exponentially. An $AR(1)$

```r
# simulate shape of the ACF curve for +/- alphas
rho <- function(k, alpha) {
  alpha^k
}
plot(0:10, rho(0:10, alpha=0.7), type='b', main='alpha=0.7')
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-4-1.pdf)<!-- -->`

```r
plot(0:10, rho(0:10, alpha=-0.7), type='b', main='alpha=-0.7')
abline(h=0, lty=2)
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-4-2.pdf)<!-- -->`

Now run a simulation:

```r
      # run a simulation
      alpha <- 0.8
      sigma_w <- 1
      w_t <- rnorm(500, 0, sigma_w)

      x4 <- vector()
      x4[1] <- w_t[1]
      for (t in 2:500) {
x4[t] <- alpha * x4[t-1] + w_t[t]
      }
      x4_ts <- ts(x4)
```

```
    # plot
    plot(x4_ts, main="Simulation: Autoregressive Model (Order=1)")
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-5-1.pdf)<!-- -->`

```r
    acf(x4_ts, main="Autocorrelation with Lags")
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-5-2.pdf)<!-- -->`

```r
    lag.plot(x4_ts, lags=9, main="Autocorrelation of Lags")
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-5-3.pdf)<!-- -->`

Note the exponential decay on the ACF plot behaves similarly to the simulated plots above.

The dotted lines represent the 95% CI for the ACF. Note that the CI of each of the autocorr

$$ CI: -\frac{1}{n} \pm \frac{2}{\sqrt{n}}$$
While, in theory, points outside the 95% CI would be evidence against the null hypothesis t

Random Walk Model

The Random Walk model is a special case of $AR(1)$, with $\alpha=1$.

As with the general $AR(1)$ case, $E(x_t)=0$. The variance is:

$$ Var(x_t) = t \sigma^2$$

The autocovariance simplifies to:

$$\gamma_k = t \sigma^2$$

Because autocovariance is a function of time, this model is _nonstationary_. Because of th

The autocorrelation is:

$$\rho_k(t) = \frac{Cov(x_t, x_{t+k})}{\sqrt{Var(x_t)Var(x_{t+k})}} =
\frac{t \sigma^2}{\sqrt{t\sigma^2(t+k)\sigma^2}} = \frac{1}{\sqrt{1+k/t}}$$

Example 1: Wave Height

Example comes from Cowpertwait.

```r
url <- 'https://raw.githubusercontent.com/mwinton/Introductory_Time_Series_with_R_datasets/ma
wave_df <- read.table(url, header=TRUE)
wave_ts <- ts(wave_df)

# data is collected at 0.1 sec intervals (396 obs = 39.6 sec)
layout(1:2)
plot(wave_ts)
abline(h=0, lty=2)

# plot first 60
plot(ts(wave_df[1:60,]))
abline(h=0, lty=2)
```

```
![](Time_Series_Intro_files/figure-latex/unnamed-chunk-6-1.pdf)<!-- -->
```

Observations: Data doesn't seem to show any trend or seasonal component, so it should be appr

It's important to plot the "correlogram" of acf and $avcf$ vs. lag using the R function `ac

```r
# look at ACF
acf(wave_ts)
# values are stored in acf(...)$acf
head(acf(wave_ts)$acf, 10)
```

```
![](Time_Series_Intro_files/figure-latex/unnamed-chunk-7-1.pdf)<!-- -->
```

```
##  [1]  1.00000  0.47026 -0.26291 -0.49892 -0.37871 -0.21499 -0.03792
##  [8]  0.17764  0.26932  0.13039
```

Observe the wavelike shape that resembles a shrinking cos function. This is typical of TS

Another visual way to examine dependency structure of a series is a plot of scatterplot matri

```r
lag.plot(wave_ts, lag=9)
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-8-1.pdf)<!-- -->`

Note the alternating direction of correlation in these plots agrees with `acf`.

Example 2: Initial Jobless Claims

Downloaded data from [MacroTrends.net](http://www.macrotrends.net/1365/jobless-claims-histori

```r
#plot ts
unemployment_df <- read.csv('jobless-claims-historical-chart.csv', skip=13, header=TRUE)
unemployment_df <- unemployment_df %>% filter(value>0) %>% mutate(thousands=value/1000)
unemployment_ts <- ts(unemployment_df$thousands, start=c(1967,1), freq=12)
plot(unemployment_ts, xlab="Monthly Series", ylab="First Time Unemployment Claims (in Thousan
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-9-1.pdf)<!-- -->`

Observe that series appears very _persistent_, showing fairly long term upwards and/or downwa

```r
# plot scm against its own lags
lag.plot(unemployment_ts, lags=9, main="Autocorrelation of First Time Jobless Claims (Monthly
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-10-1.pdf)<!-- -->`

```r
# plot acf; manually change axis labels if we don't want it in years
acf(unemployment_ts, type='correlation', main="Autocorrelation - First Time Jobless Claims\n
axis_max <- length(unemployment_ts)
axis(1, at=0:axis_max/12, labels=0:axis_max)
abline(h=0.8, lty=3, col="red")
```

`![](Time_Series_Intro_files/figure-latex/unnamed-chunk-10-2.pdf)<!-- -->`

The scatterplot matrix shows strong correlation for all 9 months displayed. The ACF plot sho