STAT 248: Removal of Trend & Seasonality
Handout 4

GSI: Gido van de Ven

September 24th, 2010

1 Introduction

Today's section we will start with some basic models for time series that take an important place in the analysis of time series. Then we will turn to ways to remove a trend and/or seasonal component from your data in order to make a series stationary.

2 Examples of Basic Time Series Models

2.1 White Noise

A sequence of uncorrelated random variables, $\{Z_t; t \in \{1, ..., n\}\}$, with mean 0 and finite variance σ_Z^2 is called a **white noise** process. It is for example used as a model for noise in some engineering applications. The term white noise comes from the fact that, as we will see in a later lab, a frequency analysis of this model shows that all frequencies enter equally, something that is also the case if you investigate the frequencies contained in white light. A white noise process will be denoted as $Z_t \sim \mathrm{WN}(0, \sigma_Z^2)$.

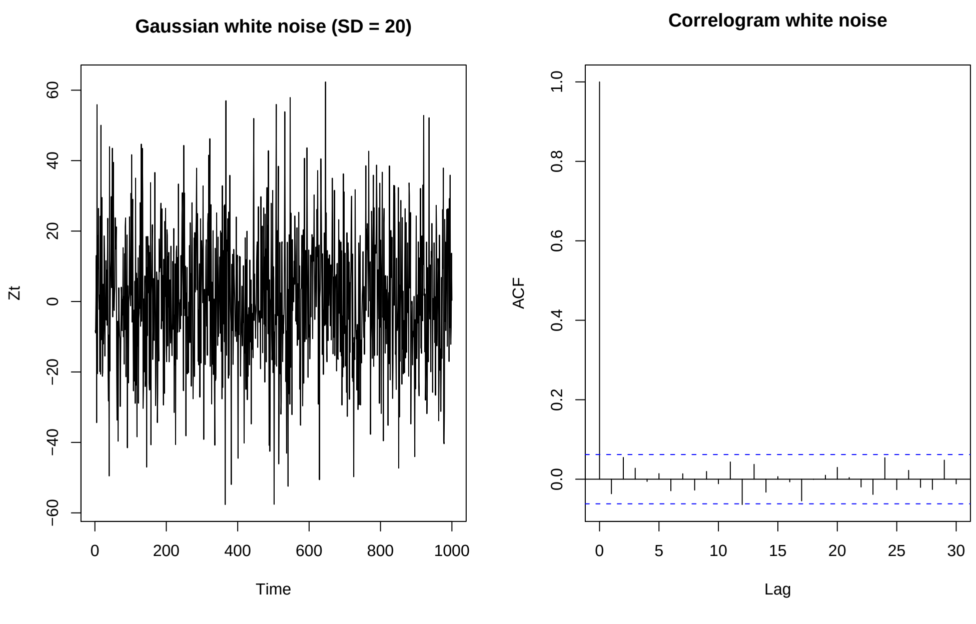

Figure 1: *Simulation (N=1000) of Gaussian white noise with* $\sigma_Z = 20$.

```
> Zt = as.ts(rnorm(1000, sd = 20))
> par(mfrow = c(1,2))
> plot(Zt, main = "Gaussian white noise (SD = 20)")
> acf(Zt, main = "Correlogram")
```

Note that above definition fixes the first two moments of the random variables Z_t, the rest of their distributions are left unspecified. It is even possible that Z_i and Z_j have different higher moments (and hence different distributions). However, sometimes we will require the random variables Z_t, besides independent, also to be identically distributed with mean 0 and variance σ_Z^2. In that case, we will denote the series as $Z_t \sim \text{IDD}(0, \sigma_Z^2)$.

A wide class of stationary processes can be generated by using white noise as the forcing term in a set of linear difference equations. These processes are called autoregressive-moving average (ARMA), and will be discussed next week.

2.2 Moving Average

If we replace the white noise series Z_t by a moving average that smoothes the series, the resulting process is called a **moving average**. For example, let us consider a random variable V_t that will average the current value with its immediate neighbors in the past and the future. That is,

$$V_t = \frac{1}{3}(Z_{t-1} + Z_t + Z_{t+1}) \qquad \text{for } t = 2, 3, ..., n-1$$

This is a 3-point moving average of the series Z_t. If we look at the moving average it shows a smoother version of the first series, reflecting that the slower oscillations are more apparent and some of the faster oscillations are taken out.
Note that in this example all involved points have equal weights. To use different weights is possible as well, an example of that is the general MA(q)-model which will be discussed next week.

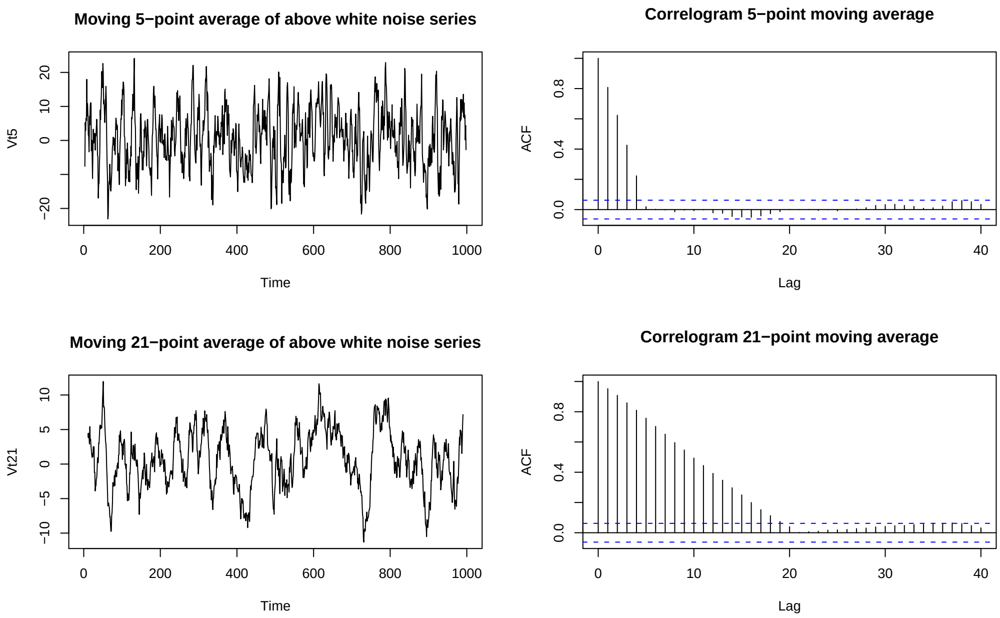

Figure 2: *Moving averages of the white noise process of figure 1.* Above: *5-points moving average.* Below: *21-points moving average.*

```
> par(mfrow = c(2,2))
> Vt = filter(Zt, rep(1/5, 5), sides = 2)
> plot(Vt, main = "Moving 5-point average of above white noise series")
> acf(Vt, main="Correlogram 5-point moving average", lag.max=40, na.action=na.pass)
> Vt = filter(Zt, rep(1/21, 21), sides = 2)
> plot(Vt, main = "Moving 21-point average of above white noise series")
> acf(Vt, main="Correlogram 21-point moving average", lag.max=40, na.action=na.pass)
```

2.3 Auto regressions

Now, let us consider the process generated by the following equation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t \qquad \text{for } t = 1, 2, ..., n$$

Such a process is called a (second order) **auto regression**. Instead of directly on the shocks at different time points, the current value of this series depends on the actual values at different time points. Actually, in above example, the value X_t depends on *all* previous shocks. (Why?)
This equation represents a regression or prediction of the current value X_t of a time series as a function of the past two values of the series. Note that along with above equation we should also provide startup values X_0 and X_{-1}.

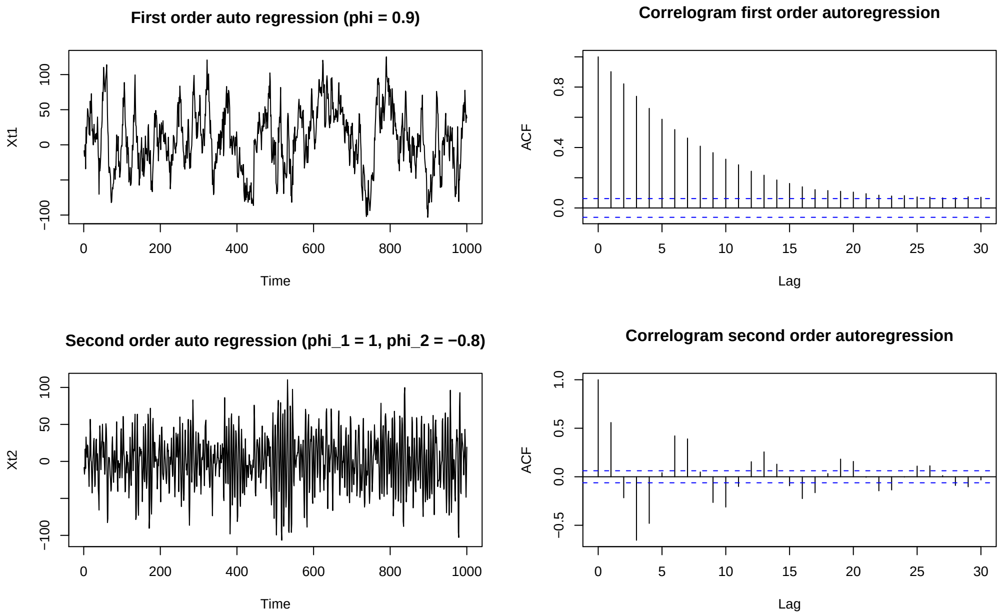

Figure 3: *Auto regressions with the white noise process of figure 1 as input,* X_0 *and* X_{-1} *are set to* 0. Above: $\phi_1 = 0.8$ *and* $\phi_2 = 0$. Below: $\phi_1 = 0.8$ *and* $\phi_2 = 0$.

```
> Xt1 = filter(Zt, 0.9, method = "recursive")
> Xt2 = filter(Zt, c(1, -0.8), method = "recursive")
> par(mfrow = c(2,2))
> plot(Xt1, main = "First order auto regression (phi = 0.9)")
> acf(Xt1, main = "Correlogram first order autoregression")
> plot(Xt2, main = "Second order auto regression (phi1 = 1, phi2 = −0.8)”)
> acf(Xt2, main = ”Correlogramsecondorderautoregression”)
```

2.4 Random Walk

If we take a first order auto regression with $\phi_1 = 1$, the resulting process is called a **random walk**:

$$Y_t = Y_{t-1} + Z_t \qquad \text{for } t = 1, 2, ..., n, \text{with initial condition } Y_0 = 0$$

The term random walk comes from the fact that the value of the time series at time t is the value of the series at time $t-1$ plus a completely random movement determined by Z_t. We may rewrite it as a cumulative sum of white noise variates:

$$Y_t = \sum_{j=1}^{t} Z_j \qquad \text{for } t = 1, 2, ..., n \tag{1}$$

A variant of the random walk model is the **random walk with drift**, given by:

$$Y_t = \delta + Y_{t-1} + Z_t \qquad \text{for } t = 1, 2, ..., n, \text{with initial condition } Y_0 = 0$$

The constant δ is called the drift. This model can also be rewritten as a cummulative sum:

$$Y_t = \delta t + \sum_{j=1}^{t} Z_j \qquad \text{for } t = 1, 2, ..., n$$

The random walk with drift is clearly not stationary (why?), but what about the pure random walk process? This one is also not stationarity, what can be seen by calculating its variance, which turns out to be a function of the time t:

$$Var(Y_t) = Var(\sum_{j=1}^{t} Z_j) = t\sigma_Z^2 \qquad \text{for } t = 1, 2, ..., n$$

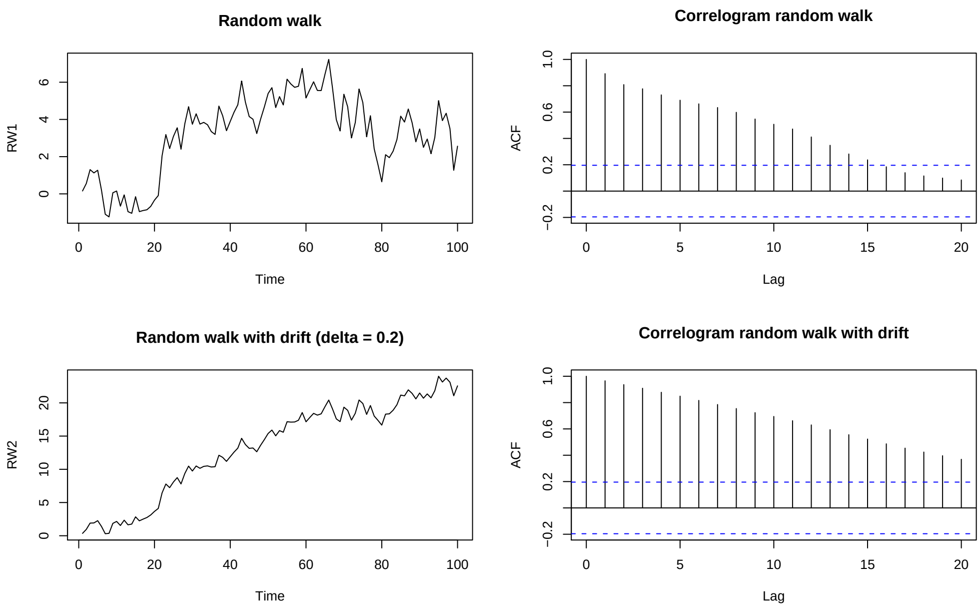

Figure 4: Above: *Simulated random walk.* Below: *Simulated random walk with drift* $\delta = 0.2$.

```
> RW1 = as.ts(cumsum(Zt))
> RW2 = 0.2*(1:length(Zt)) + as.ts(cumsum(Zt))
> par(mfrow = c(2,2))
> plot(RW1, main = "Random walk")
> acf(RW1, main = "Correlogram random walk")
> plot(RW2, main = "Random walk with drift (delta = 0.2)")
> acf(RW2, main = "Correlogram random walk with drift")
```

3 Non-stationarity: Trend and seasonality

Last week we discussed the importance of stationarity in analyzing time series. There exist a lot of methods for the succesfull analysis of stationary processes. However, by far not all time series are stationary.

As discussed in the second lab, as a first step in the analysis of a time series, it is often wise to make some plots of the data. This gives a first indication whether or not your observed series is stationary. For example, it could attent you on the presence of (far) outliers or apperent discontinuities in the data. For outliers you should carefully inspect whether there is any evidence that they are errroneous (and thus should be disregarded), or whether they may be correct (in which case they contain valuable information about the data generation process). In case of disountinuities it might be better to split the data up into homogenous (and perhaps stationary) segments before further analyzing the data. Inspection of the graph may also suggest the possibility of representing the data as a realization of the following process (the ***classical decomposition model***):

$$X_t = \mu_t + S_t + Z_t \qquad \text{for } t = 1, 2, ..., n \tag{2}$$

where μ_t is a slowly changing function known as a "trend" component, S_t is a function with known period d referred to as a "seasonal" component and Z_t is a random noise component (not necessarily white noise) which is stationary. Our aim will be to identify and extract the deterministic components μ_t and S_t in the hope that the residual or noise component Z_t turns out to be a stationary random process. Then we will use the theory of stationary processes to find a satisfactory probabilistic model for the process $\{Z_t\}$, to analyze its properties and to use it in conjunction with μ_t and S_t for the purpose of prediction and control of $\{X_t\}$.

An alternative approach to make a series stationary, which was developed by Box and Jenkins, is to apply difference operators repeatedly to the data $\{X_t\}$ until the differenced observations resemble a realization of a stationary process $\{W_t\}$. We will go over several approaches to trend and seasonality removal by a) estimation of μ_t and S_t and b) by differencing the data $\{X_t\}$.

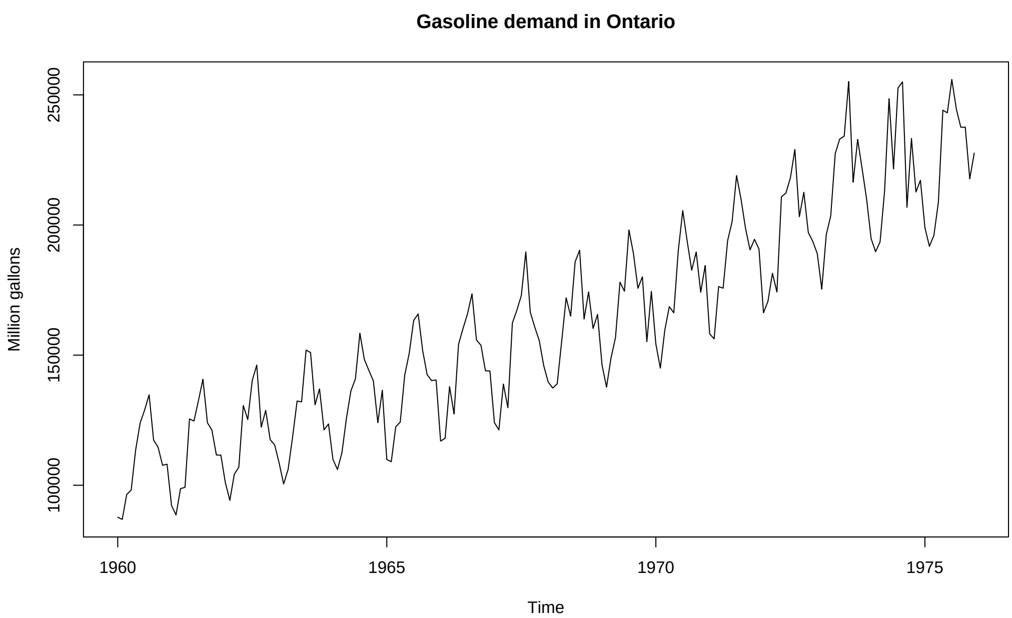

Figure 5:

The various methods will be illustrated using a dataset from Abraham and Ledolter (1983) on the monthly gasoline demand in Ontario over the period 1960 - 1975. The data can be downloaded from the section website (`http://www.stat.berkeley.edu/~gido/`) as **gas.dat**. Figure 5 shows a time plot of the data. This series seems to have both a linear trend and a seasonal term.

3.1 Least squares estimation of μ_t

In this procedure we attempt, for the trend component m_t, to fit a parametric family of functions, e.g.

$$\mu_t = \beta_0 + \beta_1 t + \beta_2 t^2 \qquad \text{for } t = 1, 2, ..., n$$

to the data $X_t, t = 1, ..., n$ by choosing the parameters (in this illustration β_0, β_1 and β_2) to minimize the sum of squares $\sum_{t=1}^{n}(X_t - \mu_t)^2$. The estimated values of the noise process Z_t, are the residuals obtained by the subtractions of $\hat{\mu}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{\beta}_2 t^2$ from the data X_t. The goal is for them to be stationary.

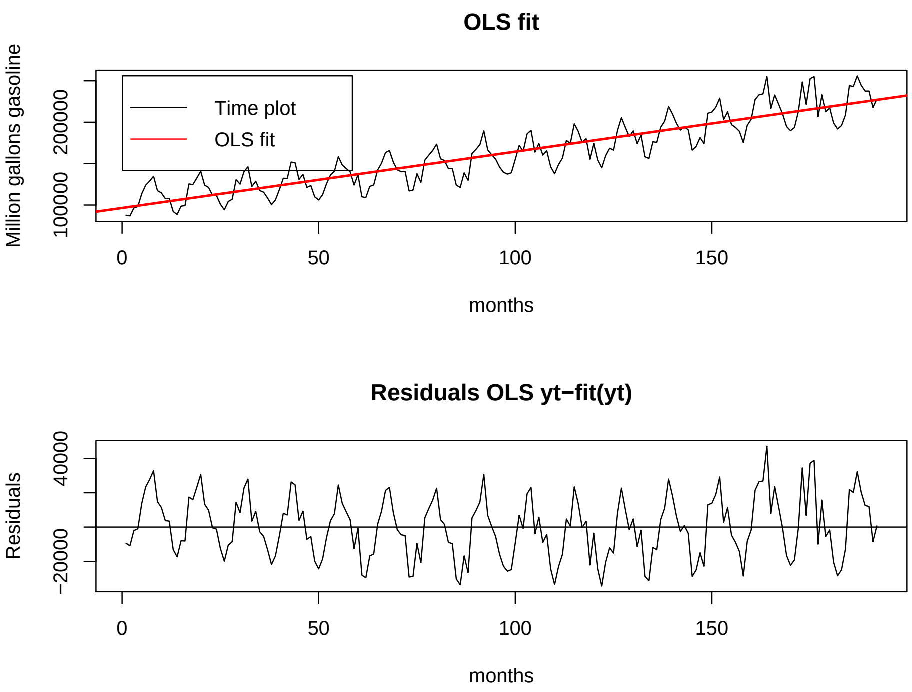

Figure 6: *Illustration of least squares estimation of the trend component in the gasoline data using the trend model* $\mu_t = \beta_0 + \beta_1 t$.

```
> time = 1:length(gas.ts)
> fit = lm(gas.ts time)
> par(mfrow=c(2,1))
> plot(time, gas.ts, type='l', xlab='months', ylab='Million gallons gasoline',
main="OLS fit")
> legend(.1, 256000, c("Time plot", "OLS fit"), col = c(1,2), text.col = "black",
merge = TRUE, lty=c(1,1))
> abline(fit,col=2,lwd=2)
> plot(t.gas, fit$resid, type='l', xlab="months", ylab="Residuals",
main="Residuals OLS [yt-fit(yt)]")
> abline(a=0, b=0)
```

3.2 Smoothing by means of moving average

Let q be a non-negative integer and consider the two sided moving average,

$$V_t = \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t-j} \qquad \text{for } t = q+1, q+2, ..., n-q$$

Assuming that X_t was generated by the process $X_t = \mu_t + Z_t$ and that μ_t is approximately linear over the interval $[t-q, t+q]$ and that the average of the error terms over this interval is zero, then V_t will be close to μ_t.

It is useful to think of $\{V_t\}$ as a process obtained from X_t by application of a linear operator or **linear filter**, $V_t = \sum_{j=-\infty}^{\infty} a_j X_{t+j}$ with weights $a_j = \frac{1}{2q+1}$, $-q \leq j \leq q$, and $a_j = 0$, $|j| > q$. This particular filter is a **"low-pass"** filter since it takes the data $\{X_t\}$ and removes from it the rapidly fluctuating (or high frequency) component $\{\hat{Z}_t\}$, and leaves the relatively slowly varying estimated trend term $\{\hat{\mu}_t = V_t\}$.

Remark 1: If $\frac{1}{2q+1}\sum_{j=-q}^{q} Z_{t-j} \sim 0$ (what usually happens if we choose q "large enough"), this filter will allow a linear trend function $\mu_t = \beta_0 + \beta_1 t$ to pass without distortion. However, we must beware of choosing q to be too large since if μ_t is not linear, the filtered process, although smooth, will not be a a good estimate of μ_t.

Remark 2: By clever choosing the weights $\{a_j\}$ it is possible to design a filter which allows a larger class of trend functions (e.g. kth degree polynomials) to pass undistorted through the filter. Take for example the Spencer 15-point moving average. This filter has the following weights:

$$\begin{aligned} a_i &= 0 \qquad |i| > 7 \\ a_i &= a_{-i} \qquad |i| \leq 7 \\ \{a_0, a_1, ..., a_7\} &= \frac{1}{320}\{74, 67, 46, 21, 3, -5, -6, 3\} \end{aligned}$$

As you will show in today's theoretical problem, the Spencer 15-point moving average filter $\{a_j\}$ does not distort a cubic trend.

Remark 3: Note that by clever choosing the weights $\{a_j\}$ it is also possible to let seasonal components pass the filter undistorted. For example, if you have a series with seasonal componemt with known period d, then you could create a linear filter where only the weights a_{jd} for $j = ..., -1, 0, 1, ...$ are allowed to be non-zero.

Remark 4: Many other techniques are available for smoothing time series data based on methods from scatterplot smoothers. The general setup is:

$$X_t = f_t + Z_t$$

where f_t is some smooth function of time, and Z_t is a stationary process. Examples of other kind of smoothers are: kernel smoothers, locally weighted regression (lowess), splines.

3.3 Differencing to de-trend a series

Instead of attempting to remove the noise by smoothing as in Method 2, we now attempt to eliminate the trend term by differencing. We define the first difference operator ∇ by:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

where B is called the backward shift operator,

$$BX_t = X_{t-1}$$

Powers of the operators B and ∇ are defined in the obvious way, i.e. $B^j(X_t) = X_{t-j}$ and $\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$, $j \geq 1$ with $\nabla^0(X_t) = X_t$

If the operator ∇ is applied to a linear trend function $\mu_t = \beta_0 + \beta_1 t$ then we obtain the constant function $\nabla X_t = \beta_1$. In the same way any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k.

Remark: These considerations suggest the possibility, given a sequence $\{X_t\}$ of data, of applying the operator ∇ repeatedly until we find a sequence $\{\nabla^d X_t\}$ which can plausibly be modeled as a realization of a stationary process. It is often found in practice that the order k of differencing required is quite small, frequently one or low. This depends on the fact that many functions can be well approximated, on an interval of finite length, by a polynomial of reasonably low degree.

3.4 Differencing to de-seasonalized a series

An example of a model for seasonal data is:

$$x_t = S_t + Z_t$$

where S_t is a seasonal component that varies slowly from one year to the next, according to a random walk:

$$S_t = S_{t-12} + W_t$$

In this model Z_t and W_t are uncorrelated white noise processes. The tendency of data to follow this type of model will be exhibited in a sample ACF that is large and decays very slowly at lags $h = 12k$, for $k = 1, 2, ...$

This kind of seasonality is known as **additive seasonality**. Show that the seasonal difference operator ∇_{12} of order 1 acts on X_t, to produce a stationary series. A seasonal difference operator of order D is defined as

$$\nabla_s^D x_t = (1 - B^s)^D x_t$$

where $D = 1, 2, ...$ takes integer values. Typically, $D = 1$ is sufficient to obtain seasonal stationarity. Note that this seasonal difference operation is not only able to remove additive seasonality, but also other forms of seasonality (compare with how the general difference operator is able to remove linear trends).

4 Bibliography

This handout is based on handouts prepared by Irma Hernandez-Magallanes, previous GSI for this course. Additional sources that are used, and that could be useful for you:

- "Time Series: Data Analysis and Theory" by David R. Brillinger
- "Time Series: Theory and Methods" by Peter Brockwell & Richard Davis
- "Time Series Analysis and Its Applications: With R Examples" by Robert Schumway & David Stoffer