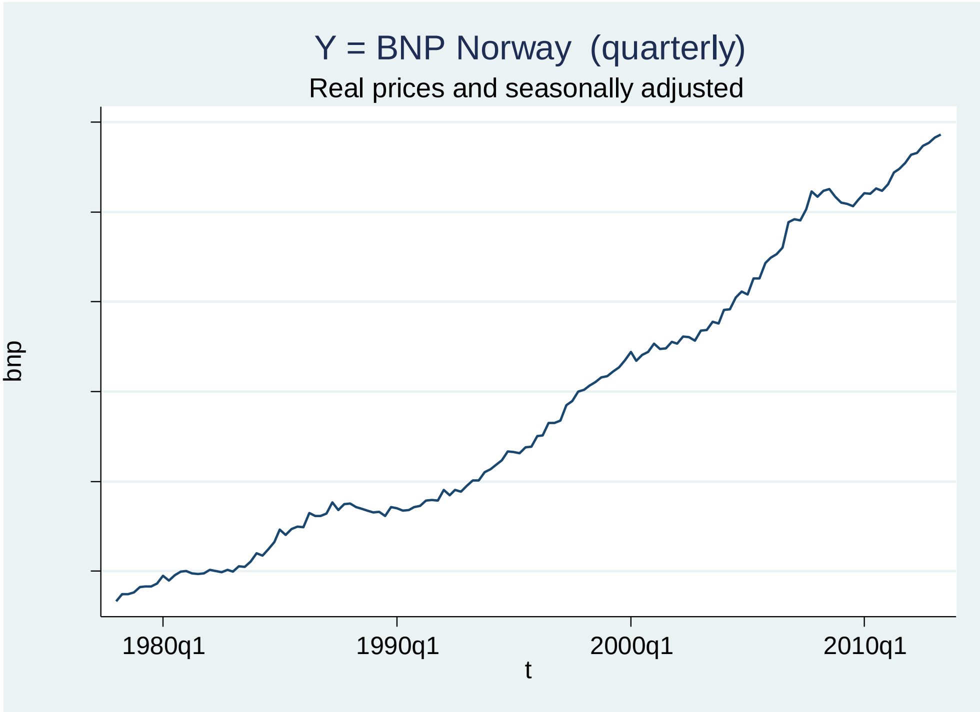
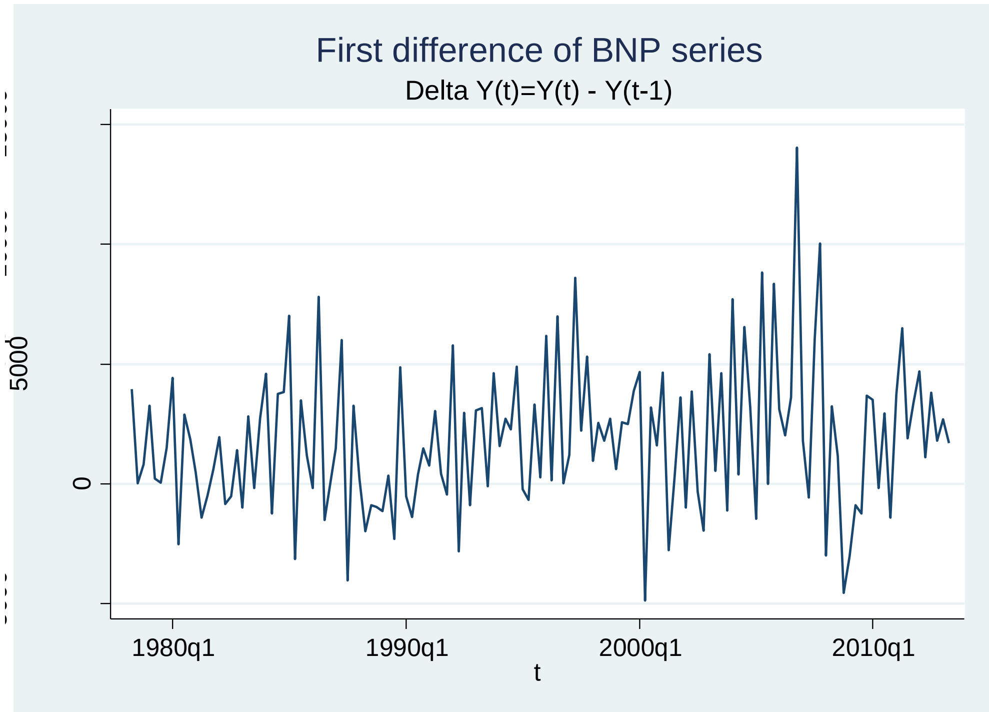
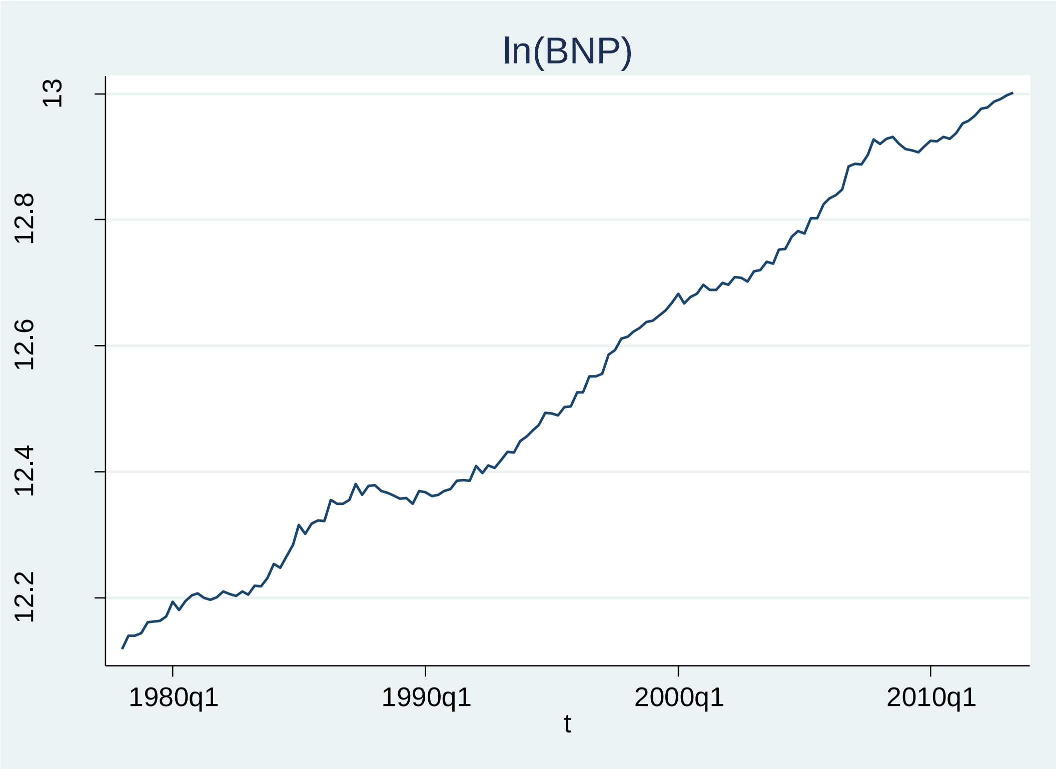
HG
First lecture - 15. Jan. 2014

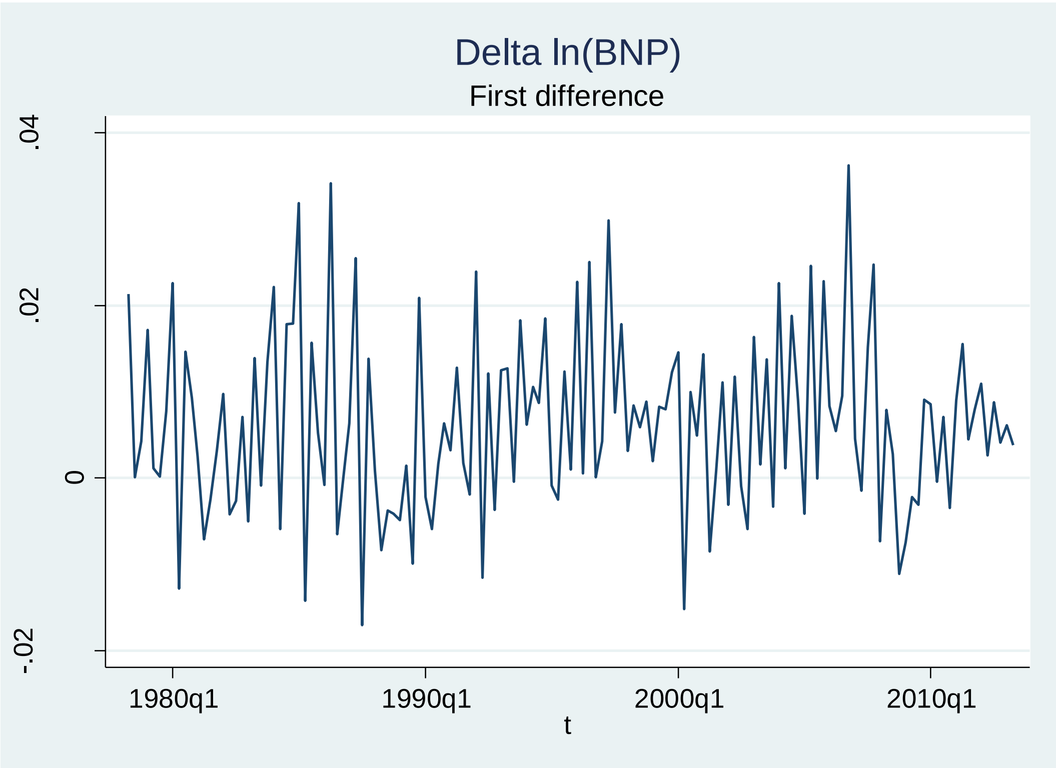
1. **Introduction.** A *time series* is a series of random variables (rv's), $\{Y_t;\ t=1,2,\ldots,T\}$ where the index, t, represents time in fixed units, eg., days, quarters, years etc.

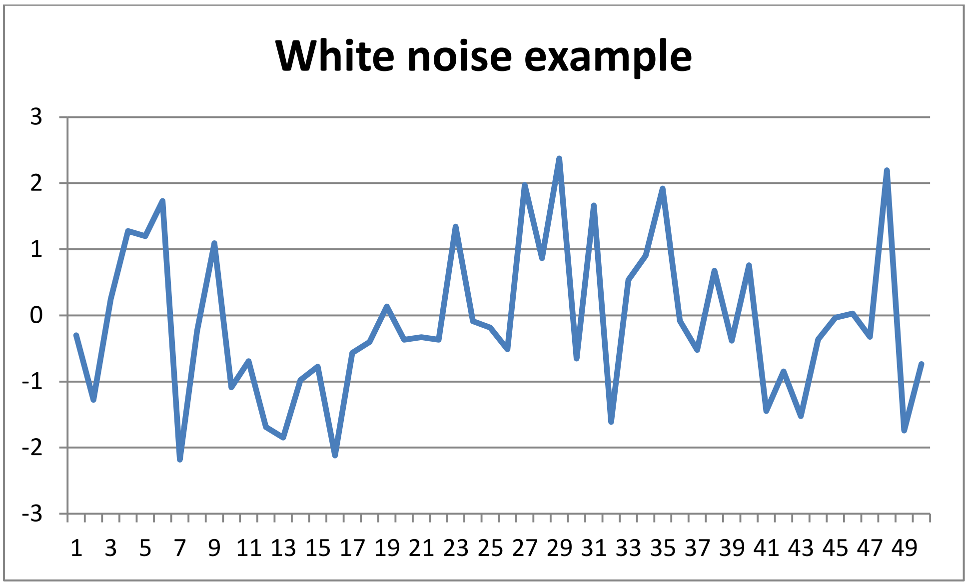


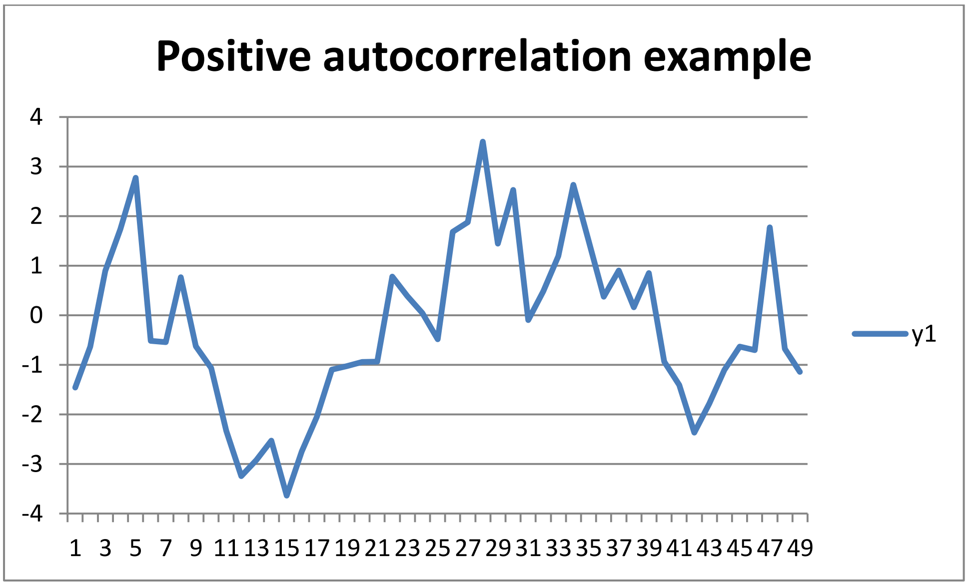


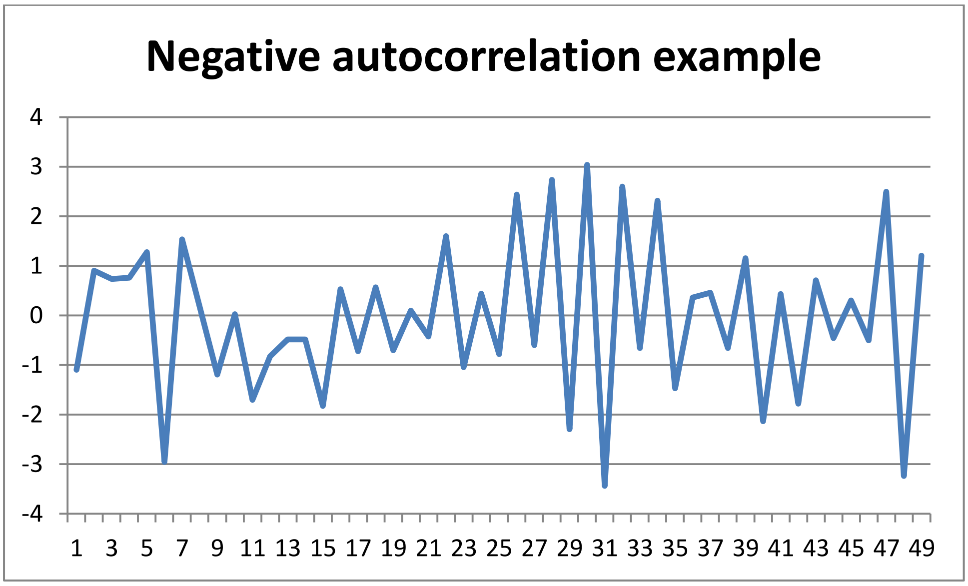
- In time series analysis we try to separate the *stationary* part from the time series, $Y_t = X_t + U_t$, where, e.g., X_t may represent a trend part, and U_t the stationary part.
- Sometimes stationarity can be achieved by differencing. $\Delta Y_t = Y_t - Y_{t-1}$ (first difference), or $\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}$ (second difference), or higher, $\Delta^k Y_t$.
- The stationary part is important for establishing a basis for inference (estimation, forecasting, random variation, uncertainty, etc) for the series.
- **DEF:** The series $\{Y_t\}$ is called *strictly stationary* if for any $t_1 < t_2 < \cdots < t_r$, the joint distribution of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$ is *the same as* the distribution of $(Y_{t_1+h}, Y_{t_2+h}, \ldots, Y_{t_r+h})$ for any *h*. This implies, in particular, that $E(Y_t) = \mu$ and $\text{var}(\text{Y}_t) = \sigma^2$ must be constant (do not depend on *t*), and also that $\text{cov}(Y_t, Y_{t+j}) = \gamma_j$ does not depend on *t*.
- **DEF:** The series $\{Y_t\}$ is called *covariance stationary* (or *weakly stationary*) if it fulfills the last property in the previous bullet point, i.e., if $E(Y_t) = \mu,\ \text{var}(\text{Y}_t) = \sigma^2,$ and $\text{cov}(Y_t, Y_{t+j}) = \gamma_j$ do not depend on *t*.
- The simplest case of a stationary series is *white noise*, $\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t, \ldots$, where the ε_t's are *iid* (independent and identically) distributed with $E(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) = \sigma^2$. The iid assumption implies strict stationarity. Sometimes we also use the weaker concept, *2nd order white noise* which is only covariance stationary with $E(\varepsilon_t) = 0$ and $\text{var}(\varepsilon_t) = \sigma^2$.
- Note (!) that in the gaussian case (i.e., when the joint distributions of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_r})$ are multivariate normal), covariance stationarity implies strict stationarity. This follows since the joint distribution in the Gaussian case is completely determined by the expectations and the covariance matrix.
- The delta BNP series shows increasing variance (volatility) – i.e., lack of stationarity.

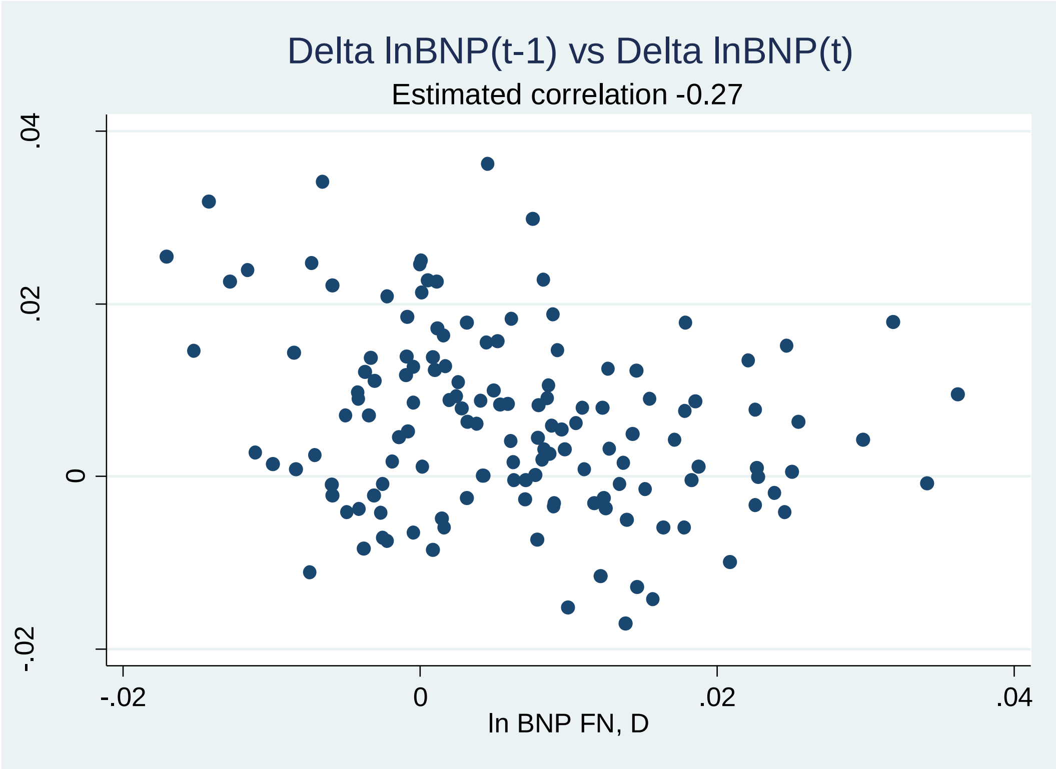

ln(BNP)
13
12.8
12.6
12.4
12.2
1980q1
1990q1
2000q1
2010q1
t


Delta ln(BNP)
First difference
.04
.02
0
-.02
1980q1
1990q1
2000q1
2010q1
t


White noise example
3
2
1
0
-1
-2
-3
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49


Positive autocorrelation example
4
3
2
1
0
-1
-2
-3
-4
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
y1


Negative autocorrelation example
4
3
2
1
0
-1
-2
-3
-4
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49



2. Autocovariance and autocorrelation function

For any time series $\{Y_t\}$ we can define auto-covariance function,

$$\gamma(s,t) = \text{cov}(Y_s, Y_t)$$

If the series is stationary, the covariance will only depend on the distance in time between *s* and *t*, i.e.,

(2) $\gamma(s,t) = \gamma(h)$ where $h = |t-s|$

The function, $\gamma_h = \gamma(h) = \text{cov}(Y_t, Y_{t+h}),\ h = 0, \pm 1, \pm 2, \ldots$, is called the autocovariance function for a covariance stationary time series.

- Notice that $\gamma(h) = \gamma(-h)$ for all *h*, which follows from the stationarity.
- Notice that the (constant) variance is, $\text{var}(Y_t) = \text{cov}(Y_t, Y_t) = \gamma(0) = \gamma_0$

Having the auto covariance at hand, we can define *the autocorrelation function,*

(3) $$\rho_h = \rho(h) = \text{corre}(Y_t, Y_{t+h}) = \frac{\text{cov}(Y_t, Y_{t+h})}{\sqrt{\text{var}(Y_t)\text{var}(Y_{t+h})}} = \frac{\gamma(h)}{\sqrt{\gamma(0)\gamma(0)}} = \frac{\gamma(h)}{\gamma(0)}, \quad h = 0, \pm 1, \pm 2, \dots$$

Having observed the time series, $Y_1, Y_2, \dots, Y_T$, these functions can be estimated, consistently by (see H(4)):

The sample autocovariance/correlation function

The sample autocovariance function

$$\hat{\gamma}_{-h} = \hat{\gamma}_h = \frac{1}{T}\sum_{t=h+1}^{T}(Y_t - \bar{Y})(Y_{t-h} - \bar{Y}) \quad \text{for } h = 0,1,\dots,T-1, \text{ where } \bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$$

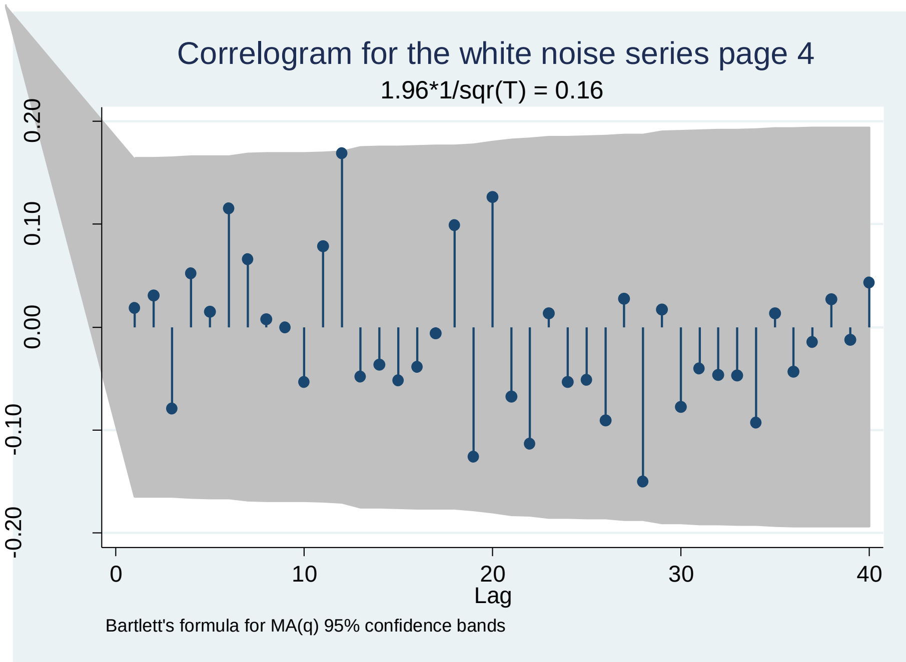
The sample autocorrelation function

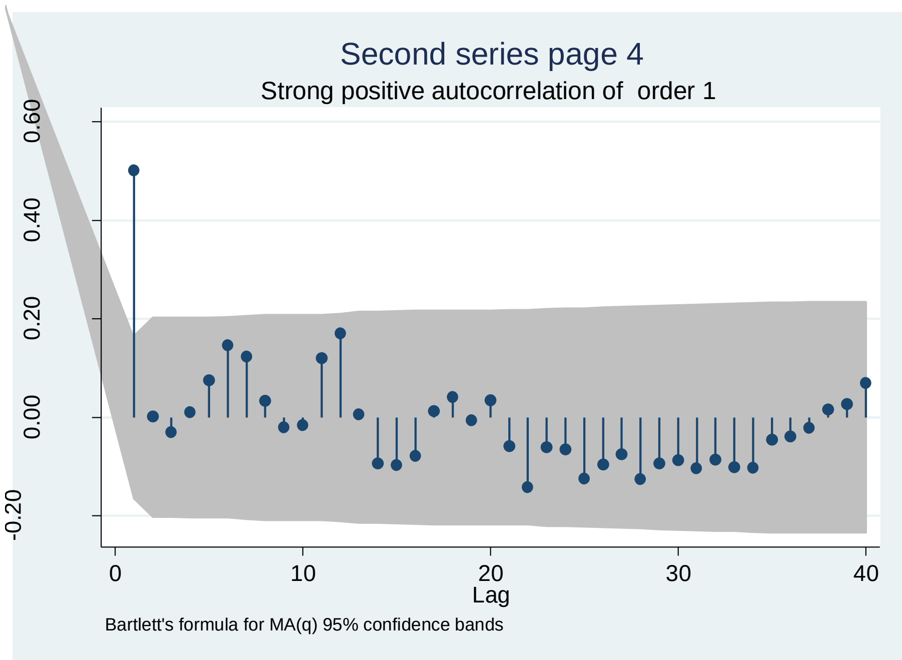
$$\hat{\rho}_{-h} = \hat{\rho}_h = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad h = 0,1,2,\dots$$

- Notice that it is common to use the same T in the denominator of all covariances which make the sample covariances slightly downward biased. The reason for this is to keep the covariance matrices based on $\{\hat{\gamma}_h\}$ as proper covariance matrices (i.e., non-negative definite).
- However, it can be proven that $\{\hat{\gamma}_h, \hat{\rho}_h\}$ are all consistent, and (if the 4$^{\text{th}}$ order moments of Y_t is finite), asymptotically normally distributed when T is large, (the formula for the asymptotical variances is slightly complicated, but can be found in many time series text books).
- In the special case that $\{Y_t\}$ is white noise (written shortly, $Y_t \sim WN(0, \ \sigma^2 = \gamma(0))$, we have that $\rho_h = 0$ for all $h \neq 0$. Then it turns out that (if the 4$^{\text{th}}$ order moments exist) that, if T is large, then $\hat{\rho}_h$ is approximately normally distributed with expectation 0 and variance, $1/T$.

The last property is quite useful when we wish to check if the residuals from a time series analysis seem to behave like white noise. A rough test of this is plotting $\hat{\rho}_h$ (for the residuals) against h, together with the interval

$\left[-(1.96)/\sqrt{T}\,,\ (1.96)/\sqrt{T}\right]$, which represents an approximate 95% variation interval for $\hat{\rho}_h$ under the assumption of white noise. If several $\hat{\rho}_h$ are outside this interval, this is evidence against the white noise hypothesis.





There are also a number of more formal tests of the white noise hypothesis. One common test is, e.g., Portmanteau test (implemented in most packages). If time allows it, I will describe it later:

Stata gave:

```
Portmanteau test for white noise (first graph page 4)
---------------------------------------
Portmanteau (Q) statistic =    32.7239
Prob > chi2(40)           =     0.7861   (p-value)


Portmanteau test for white noise (second graph page 4)
---------------------------------------
Portmanteau (Q) statistic =    79.6478
Prob > chi2(40)           =     0.0002   (p-value)
```

3. White noise series are the main building blocks for constructing stationary time series.

Let $\varepsilon_t \sim WN(0,\sigma^2)$ (i.e., $\{\varepsilon_t\}$ is a white noise series with constant expectation 0, constant variance σ^2, and zero covariances between different ε_t's).
We know that $\{\varepsilon_t\}$ is covariance stationary (strictly stationary if they are jointly normal).

Then a linear combination of different ε_t's is also stationary, for example MA(q) processes (moving average processes of order q).

Def 3.1 $\{Y_t\}$ is a *moving average process of order q* (in short $Y_t \sim \text{MA}(q)$), if Y_t can be written

(4) $$Y_t = \mu + \theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q}, \quad t = \ldots 1,2,3,\ldots$$

where $\mu, \theta_0, \theta_1, \theta_2, \ldots, \theta_q$ are constants.

(In most cases θ_0 is set = 1)

Result 3.2 If $Y_t \sim \text{MA}(q)$ with respect to the white noise series $\{\varepsilon_t\} \sim \text{WN}(0,\sigma^2)$, then

i) $\{Y_t\}$ is covariance stationary with

ii) autocovariance, $$\gamma(-h)=\gamma(h)=\begin{cases}\sigma^2 \sum_{j=0}^{q-h} \theta_j\theta_{j+h} & \text{if } 0 \le h \le q \\ 0 & \text{if } h > q\end{cases}$$

iii) variance, $$\text{var}(Y_t)=\gamma(0)=\sigma^2\left(\theta_0^2+\theta_1^2+\cdots+\theta_q^2\right)$$

iv) autocorrelation, $$\rho(-h)=\rho(h)=\begin{cases}\dfrac{\sum_{j=0}^{q-h}\theta_j\theta_{j+h}}{\theta_0^2+\theta_1^2+\cdots+\theta_q^2} & \text{if } 0 \le h \le q \\ 0 & \text{if } h > q\end{cases}$$

[**"Proof":** Follows from straightforward calculation. First, clearly, $E(Y_t)=\mu$, i.e., a constant.

$$\text{cov}(Y_t, Y_{t+h}) = E(Y_t-\mu)(Y_{t+h}-\mu) =$$
$$= E\left(\theta_0\varepsilon_t+\theta_1\varepsilon_{t-1}+\cdots+\theta_q\varepsilon_{t-q}\right)\left(\theta_0\varepsilon_{t+h}+\theta_1\varepsilon_{t+h-1}+\cdots+\theta_q\varepsilon_{t+h-q}\right)=$$
$$= \cdots \text{some algebra} \cdots = \sigma^2\sum_{j=0}^{q-h}\theta_j\theta_{j+h}$$

which does not depend on *t*. Hence, stationarity.]

For example, for the MA(1) process, $Y_t=\mu+\varepsilon_t+\theta\varepsilon_{t-1}$, we have

$$\gamma(-h)=\gamma(h)=\begin{cases}(1+\theta^2)\sigma^2 & \text{if } h=0 \\ \theta\sigma^2 & \text{if } h=1 \\ 0 & \text{if } h>1\end{cases}$$

$$\rho(-h)=\rho(h)=\begin{cases}\dfrac{\theta}{1+\theta^2} & \text{if } h=1 \\ 0 & \text{if } h>1\end{cases}$$

4. Introduction to dynamic modelling

The simplest case of dynamic modelling of a time series $\{Y_t\}$ is probably the autoregressive model of order 1 (AR(1)), specified as

(5) $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t,\ t = 1,2,3,\ldots$ where φ_0, φ_1 are constants and $\varepsilon_t \sim WN(0,\sigma^2)$

This is a special case of the more general autoregressive model of order *p* (AR(*p*)) for $\{Y_t\}$, specified as

(6) $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t$
where $\varphi_0, \varphi_1, \ldots, \varphi_p$ are constants and $\varepsilon_t \sim WN(0,\sigma^2)$.

We will look at the simplest AR(1) case in (5) in some detail first. Assume that the prosess $\{Y_t\}$ starts at time $t = 0$ with value Y_0. From (5) we see that the dynamic nature of $\{Y_t\}$ is expressed in terms of a stochastic difference equation. To get some further insight in the nature of $\{Y_t\}$, we need to solve this equation with respect to Y_t:

By successive substitution in (5), we get the solution

(7)
$$
\begin{aligned}
Y_t &= \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t = \varphi_0 + \varphi_1(\varphi_0 + \varphi_1 Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \\
&= \varphi_0 + \varphi_1\varphi_0 + \varphi_1^2 Y_{t-2} + \varepsilon_t + \varphi_1\varepsilon_{t-1} = \\
&= \varphi_0 + \varphi_1\varphi_0 + \varphi_1^2(\varphi_0 + \varphi_1 Y_{t-3} + \varepsilon_{t-2}) + \varepsilon_t + \varphi_1\varepsilon_{t-1} \\
&= \cdots\cdots\cdots = \\
&= \varphi_0\left(1 + \varphi_1 + \varphi_1^2 + \cdots + \varphi_1^{t-1}\right) + \varphi_1^t Y_0 + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \cdots + \varphi_1^{t-1}\varepsilon_1
\end{aligned}
$$

We see that the solution is not (!) stationary (e.g., taking expectations on both sides, the ε-terms disappear showing that $E(Y_t)$ is not constant).

However, if $|\varphi_1| < 1$, $\varphi_1^t \underset{t\to\infty}{\to} 0$, and the solution in (7) stabilizes and approaches something stationary when $t \to \infty$. Using the formula for a geometric sum, we get

$$Y_t = \varphi_0 \frac{1-\varphi_1^t}{1-\varphi_1} + \varphi_1^t Y_0 + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \cdots + \varphi_1^{t-1}\varepsilon_1$$

Now, when $t \to \infty$, the first term approaches the constant $\mu = \dfrac{\varphi_0}{1-\varphi_1}$, the second term, $\varphi_1^t Y_0$, approaches 0, and the third term, the sum of ε's approaches the infinite stochastic series,

(8) $$\varepsilon_t + \varphi_1 \varepsilon_{t-1} + \cdots + \varphi_1^{t-1}\varepsilon_1 + \varphi_1^t \varepsilon_0 + \varphi_1^{t+1}\varepsilon_{-1} + \varphi_1^{t+2}\varepsilon_{-2} + \cdots = \sum_{j=0}^{\infty} \varphi_1^j \varepsilon_{t-j}$$

where we imagine that the white noise series $\{\varepsilon_t\}$ has been going on since $-\infty$. The last (imagined) part of (8) will be negligible when *t* is large. It turns out that the infinite series in (8), also is stationary and represents a generalization of the MA-processes above, sometimes called a $\text{MA}(\infty)$ process and sometimes a *linear process*. If we want a stationary solution to (5), an infinite stochastic series like (8) appears to be the answer.

Our final step in building up stationary processes from white noise is provided by the following theorem.

4.1 Theorem. If $\{\varepsilon_t; t = 0, \pm 1, \pm 2, \ldots\}$ is a doubly infinite white noise $(0, \sigma^2)$ time series, and $\{\psi_t; t = 0, \pm 1, \pm 2, \ldots\}$ is a doubly infinite sequence of constants such that $\sum_{t=-\infty}^{\infty} |\psi_t| < \infty$, then

$$Z_t = \sum_{j=-\infty}^{\infty} \psi_{-j}\varepsilon_{t+j} = \cdots + \psi_2 \varepsilon_{t-2} + \psi_1 \varepsilon_{t-1} + \psi_0 \varepsilon_t + \psi_{-1}\varepsilon_{t+1} + \psi_{-2}\varepsilon_{t+2} + \cdots$$

are well defined random variables, and $\{Z_t\}$ is covariance stationary with autocovariance function and variance

$$\gamma(-h) = \gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_{j+h}\psi_j \ , \qquad \text{var}(Z_t) = \gamma(0) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j^2$$

(For a proof of theorem 4.1 see e.g. the appendix to chapter 3 in Hamilton. where it is shown that it suffices with the slightly weaker condition, $\sum_j \psi_j^2 < \infty$.)

4.2 DEF. We say that $\{Z_t\}$ is *causal* if all $\psi_j = 0$ for $j < 0$, i.e., implying

$$Z_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$$

so that Z_t only depends on ε_j's from current time (t) and before – and not on future ε_j's. Otherwise Z_t is called *non-causal*. If we add a constant to Z_t, the same terminology is used (so, e.g., if Z_t is causal, then $U_t = \mu + Z_t$ is causal as well).

For example, $Z_t = 1 + 0.5\varepsilon_{t+1} + 2\varepsilon_t - \varepsilon_{t-2}$ is non-causal, while $U_t = 1 + 2\varepsilon_t - \varepsilon_{t-2}$ is causal.

Returning to the AR(1) example in (5),

(5) $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t,\ t = 1, 2, 3, \ldots$

Case i, $|\varphi_1| < 1$:

Reason to believe there is a causal stationary solution

(9) $Y_t = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots$ where we must have $\mu = E(Y_t)$

implying

(10) $\varphi_1 Y_{t-1} = \varphi_1 \mu + \varphi_1 \psi_0 \varepsilon_{t-1} + \varphi_1 \psi_1 \varepsilon_{t-2} + \varphi_1 \psi_2 \varepsilon_{t-3} + \cdots$

We must then have

$$Y_t - \varphi_1 Y_{t-1} = \mu(1-\varphi_1) + \psi_0\varepsilon_t + (\psi_1 - \varphi_1\psi_0)\varepsilon_{t-1} + (\psi_2 - \varphi_1\psi_1)\varepsilon_{t-2} + \cdots$$
$$\cdots + (\psi_j - \varphi_1\psi_{j-1})\varepsilon_{t-j} + \cdots$$
$$= \varphi_0 + \varepsilon_t$$

Hence, we must have $\varphi_0 = \mu(1-\varphi_1) \quad \Rightarrow \quad \mu = \frac{\varphi_0}{1-\varphi_1}$, and $\psi_0 = 1$.

Also $\psi_1 - \varphi_1\psi_0 = \psi_1 - \varphi_1 = 0 \quad \Rightarrow \quad \psi_1 = \varphi_1$

$\psi_2 - \varphi_1\psi_1 = 0 \quad \Rightarrow \quad \psi_2 = \varphi_1^2$

……

$\psi_j - \varphi_1\psi_{j-1} = 0 \quad \Rightarrow \quad \psi_j = \varphi_1^j$ for all $j = 0,1,2,\ldots$

Hence, if $|\varphi_1| < 1$, (5) has a unique causal stationary solution

(11) $$Y_t = \frac{\varphi_0}{1-\varphi_1} + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \varphi_1^2\varepsilon_{t-2} + \cdots + \varphi_1^j\varepsilon_{t-j} + \cdots$$

Case ii, $|\varphi_1| > 1$:

If we insist on a causal solution, we must have from (7)

(12) $$Y_t = \varphi_0\left(1 + \varphi_1 + \varphi_1^2 + \cdots + \varphi_1^{t-1}\right) + \varphi_1^t Y_0 + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \cdots + \varphi_1^{t-1}\varepsilon_1 =$$
$$= \varphi_0\frac{1-\varphi_1^t}{1-\varphi_1} + \varphi_1^t Y_0 + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \cdots + \varphi_1^{t-1}\varepsilon_1$$

implying

$$E(Y_t) = \varphi_0\frac{1-\varphi_1^t}{1-\varphi_1} + \varphi_1^t E(Y_0) = \frac{\varphi_0}{1-\varphi_1} + \varphi_1^t\left(E(Y_0) - \frac{\varphi_0}{1-\varphi_1}\right)$$

which converges exponentially fast towards $+\infty$ or $-\infty$ (or both), unless $E(Y_0) = \frac{\varphi_0}{1-\varphi_1}$. (In that last case Y_t will still show explosive behavior since it will be dominated by the earliest white noise terms, $\varphi_1^{t-1}\varepsilon_1,\ \varphi_1^{t-2}\varepsilon_2,\ldots$.)

Hence, if $|\varphi_1| > 1$, any causal solution will lead a process with what we call an "explosive" behavior (and, therefore, highly non-stationary).

However, there is a possibility of *non-causal* stationary solution of (5) (!), since (5) can be written

$$Y_{t-1} = -\frac{\varphi_0}{\varphi_1} + \frac{1}{\varphi_1} Y_t - \frac{1}{\varphi_1}\varepsilon_t \quad \text{or}$$

$$Y_t = b_0 + b_1 Y_{t+1} + \eta_{t+1}, \quad \text{where } b_0 = -\frac{\varphi_0}{\varphi_1},\ b_1 = \frac{1}{\varphi_1},\ \ \eta_{t+1} = -\frac{\varepsilon_{t+1}}{\varphi_1} \sim WN\left(0,\ \frac{\sigma^2}{\varphi_1^2}\right)$$

Since, $|b_1| < 1$, we get as above (successive substitution and passing to a limit) a linear stationary, but non-causal, process solution,

$$Y_t = \frac{b_0}{1 - b_1} + \eta_{t+1} + b_1\eta_{t+2} + b_1^2\eta_{t+3} + \cdots + b_1^j \eta_{t+j+1} + \cdots$$

Case iii, $\varphi_1 = \pm 1$ (also called a unit root solution):
We will look at the case $\varphi_1 = 1$ only.

Now the causal solution in (7) becomes

$$Y_t = \varphi_0\left(1 + \varphi_1 + \varphi_1^2 + \cdots + \varphi_1^{t-1}\right) + \varphi_1^t Y_0 + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \cdots + \varphi_1^{t-1}\varepsilon_1 =$$
$$= Y_0 + \varphi_0 t + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$$

So Y_t becomes a deterministic trend plus a process

$Z_t = \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t$, i.e., a "*Random walk*", which are an important type of time series much used in econometrics, among other things, as a tool for modelling (stochastic) trends (for example in cointegration theory). More on this later in the course. Note here that $\{Z_t\}$ is non-stationary! (E.g., $\text{var}(Z_t) = t\sigma^2$ which increases with *t*.)

If $Y_1 \sim AR(1)$ with $|\varphi_1| < 1$, we can find the autocovariance/correlation functions. (We already know $E(Y_t) = \mu = \varphi_0 / (1 - \varphi_1)$).

Write $\sigma_Y^2 = \text{var}(Y_t)$, $\sigma_\varepsilon^2 = \text{var}(\varepsilon_t)$.

Then from (5), $\sigma_Y^2 = \text{var}(Y_t) = \text{var}(\varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t) = \varphi_1^2 \sigma_Y^2 + \sigma_\varepsilon^2$, giving

$$\gamma(0) = \sigma_Y^2 = \frac{\sigma_\varepsilon^2}{1 - \varphi_1^2}$$

, and (see Hamilton section 3.4 or next time)

$$\gamma(-h) = \gamma(h) = \frac{\varphi_1^h \sigma_\varepsilon^2}{1 - \varphi_1^2} \quad \text{for } h = 0, 1, 2, \ldots \quad \Rightarrow \quad \rho(-h) = \rho(h) = \varphi_1^h$$