Univariate time-series analysis

1. Basic concepts

A time series contains observations of the random variable Y at certain points of time. This is often denoted as : $\{y_{-\infty},...,y_{\infty}\}=\{y_t\}_{-\infty}^{\infty}$. We have observations only for a finite number of periods (1 to T): $\{y_1,y_2,...,y_T\}=\{y_t\}_{t=1}^{T}$.

A special feature of time-series is that we can observe only a single realization of the underlying process. In other words, we cannot just redraw samples from the population, as our reality itself is the only sample available to us. The use of standard statistical methods needs a fundamental assumption referred to as "fictional population". Please do not make the mistake of not taking this concept seriously.

Let us assume you have data on all OECD countries for the period 1960-2010, and you would like to estimate a model explaining their economic performance. The question is if what you have is the population itself, or just a sample. Some would say that it is the population, since they are interested in the OECD countries itself. The consequence is that no statistical inference is needed, because if we observe the population, we do not need any statistical tests. Yet, even in time-series analysis we rely heavily on statistical inference theory. The explanation is that what we observe is just one of the infinite possible scenarios. Under different circumstances, the effect of, say, investments could have been different on per capita income. In other words, we cannot consider the data we observe as the population itself. Reality is a realization (a draw from) an unknown process (Data Generating Process - DGP) that we wish to understand, uncover based on the observed realization.

Another important feature of time series is that the observations have a fixed order, you cannot reorder the observations as you like such as in cross-sectional data. As a result, serial correlation must be given proper attention.

We can characterize a time series (y_t) by three fundamental statistics:

Expected value ($E(y_t)$), variance ($\gamma_0=Var(y_t)=E(y_t-E(y_t))^2$) and auto-covariance:
$\gamma_s=Cov(y_t,y_{t-s})=E\left((y_t-E(y_t))(y_{t-s}-E(y_{t-s}))\right)$

The information on autocorrelation is summarized in the **Correlogram**, which is a useful graphical diagnostic tool. It consists of two measures: **autocorrelation function** (ACF) and **partial autocorrelation function** (PACF).

Autocorrelation at lag k , or the value of the Autocorrelation function at lag k: ACF_k, is defined as:

$$ACF_k=\rho_k=\frac{\gamma_k}{\gamma_0}=\frac{Cov(y_t,y_{t-k})}{Var(y_t)}$$

It is worthwhile to compare this with the standard formula for linear correlation:

$$Cor(y_t,y_{t-k})=\frac{Cov(y_t,y_{t-k})}{\sqrt{Var(y_t)\cdot Var(y_{t-k})}}$$

They are not the same, but they should be reasonably close provided y is homoscedastic.

Useful to understand ACF(k) is to see it as a regression coefficient. Let us take the following regression:

$$y_{t-k} = \beta_0 + \beta_1 y_t + \varepsilon_t \text{ with } \varepsilon_t \sim IID(0, \sigma_\varepsilon^2)$$

The OLS estimator for the slope coefficient is:

$$\hat{\beta}_1 = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} = ACF(k)$$

How to test is autocorrelation is present? We use two types of portmanteau tests (portmanteau test are hypothesis tests, where the H0 is exactly specified while the alternative hypothesis is just loosely defined.) These are referred to as Q tests.

Box-Pierce: $Q_{BP}(k) = T\sum_{i=1}^{k} q_i^2$ and Ljung-Box: $Q_{LB}(k) = T(T+2)\sum_{i=1}^{k} \frac{q_i^2}{T-k}$

In case of no autocorrelation (H0: up to lag k all ACFs are null), both statistics should follow a chi-squared distribution with k degrees-of-freedom. The Ljung-Box statistics should be preferred in small samples.

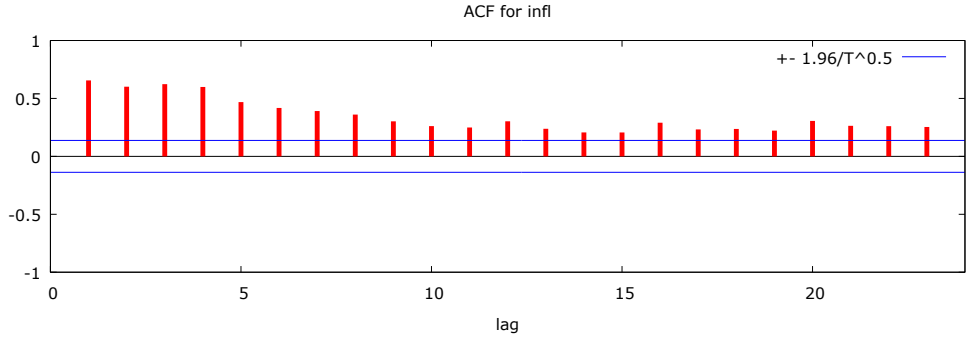
Partial Autocorrelation Function (PACF): the autocorrelation between y_t and y_{t-k} with the correlation between y_t and all lags $y_{t-1}…y_{t-k-1}$ removed.

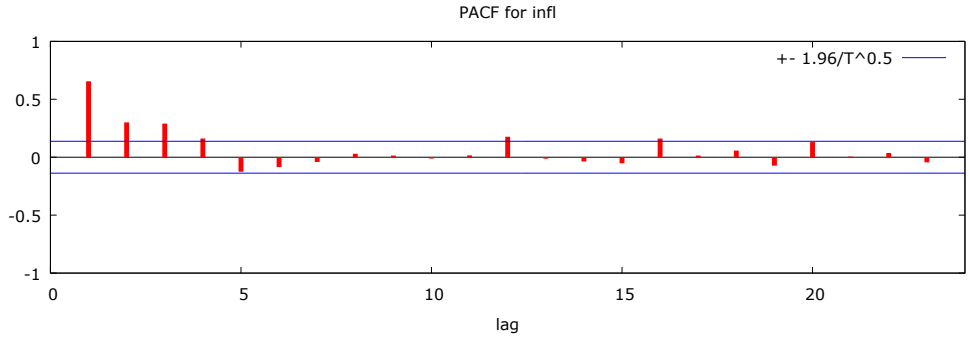
This is the easiest to see as a coefficient of a multivariate regression:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_k y_{t-k} + \varepsilon_t$$

Then β_k is the PACF(k). That is, the relationship between y_t and y_{t-k} with all lower order autocorrelation captured.

Let us see an empirical example from Greene 5-1.gdt in Gretl. We estimate the correlogram of the inflation rate:





The ACF indicates that values as old as 20 quarters (5 years) still have statistically significant effect on the inflation observed today. Such series often referred to as having long-memory. This behaviour of the inflation rate may be a hint that it is not stationary (see next section on this).

The PACF is significant at the first three lags, which may indicate an AR(3) structure behind the observed series (that is DGP may be like this: $\pi_t = \gamma_0 + \gamma_1 \pi_{t-1} + \gamma_2 \pi_{t-2} + \gamma_3 \pi_{t-3} + \varepsilon_t$)

We also obtain an alphanumerical version of the correlogram with test statistics:

```
Autocorrelation function for infl

  LAG      ACF           PACF          Q-stat. [p-value

    1   0.6555  ***    0.6555 ***     88.5176  [0.000]
    2   0.6013  ***    0.3009 ***    163.3691  [0.000]
    3   0.6233  ***    0.2902 ***    244.2048  [0.000]
    4   0.5991  ***    0.1615 **     319.2622  [0.000]
    5   0.4683  ***   -0.1252 *      365.3572  [0.000]
    6   0.4182  ***   -0.0849        402.2948  [0.000]
    7   0.3912  ***   -0.0408        434.7951  [0.000]
    8   0.3607  ***    0.0297        462.5652  [0.000]
    9   0.3024  ***    0.0145        482.1843  [0.000]
   10   0.2610  ***   -0.0115        496.8774  [0.000]
   11   0.2492  ***    0.0160        510.3402  [0.000]
   12   0.3028  ***    0.1765 **     530.3155  [0.000]
   13   0.2382  ***   -0.0125        542.7383  [0.000]
   14   0.2066  ***   -0.0365        552.1365  [0.000]
   15   0.2065  ***   -0.0521        561.5753  [0.000]
   16   0.2903  ***    0.1614 **     580.3285  [0.000]
   17   0.2322  ***    0.0145        592.3934  [0.000]
   18   0.2369  ***    0.0571        605.0166  [0.000]
   19   0.2223  ***   -0.0725        616.1939  [0.000]
   20   0.3058  ***    0.1310 *      637.4571  [0.000]
   21   0.2637  ***    0.0053        653.3557  [0.000]
   22   0.2605  ***    0.0365        668.9621  [0.000]
   23   0.2536  ***   -0.0442        683.8254  [0.000]
```

It is visible that the Q-stat rejects the H0 at all lags.

2. **Stationarity and non-stationarity** (pay extra attention – fundamental concepts will be introduced)

When dealing with time-series, unlike with cross-sectional data, we cannot assume that all characteristics (statistical moments) of the underlying DGP are unique, in the sense that they do not depend on time. Yet, standard methods such as correlation or linear regression were all based on the assumption that the DGPs probability distribution remain the same for the whole sample. Let us take two practical examples.

Cross-sectional example: We wish to estimate the effect of childhood nutrition on the height of young adults. For this reason we draw a sample of 100. We have all reason to assume that if we were to draw a sample of 1000, the expected value of the heights would not change, the only thing that would change is the accuracy of our estimates. The reason is that the underlying process, the DGP for heights remains the same for all individuals: its mean, and variance will be the same independently of how large sample we choose. Length is hence stationary. The same applies to the explanatory variable, nutrition. Standard statistical techniques can be used.

Time-series example: We wish to estimate the effect of education on the per capita income of Hungary for the period 1970-2006. Obviously, since per capita income and education have trends, the expected value would depend on the sample length chosen. The average education has a mean of 9.48 for the 1970-2006 period, while 10.29 for the 1990-2006 period. This is because it has a trend. The same applies to GDP per capita: its expected value depends on the length of the sample, hence its statistical means are not unique. We cannot argue hence that all observations are drawn from the same probability distribution, and standard statistical techniques would lead to different estimates depending on the sample size or which period we take. Obviously, these series are non-stationary.

Let us define stationarity now.

Stationarity in the strict sense: A process is called **strictly stationary** if its probability distribution is independent of time. In other words: if we have a time series observed between time t and T, then the joint probability distribution of these observations should be the same as those observed between any t+k and T+k.

What we actually say here is that the main characteristics of the time series should remain the same whenever we observe it. This is however a theoretical concept because it can be tested through the statistical moments. It is more useful to formulate the definition in terms of moments, which leads to the weak version of stationarity.

Stationarity in the weak sense or **covariance stationarity** defines stationary processes in term of their moments. This makes this definition easier to work with and testable.

Time series y is called covariance stationary if:

1. It has a finite mean: $E(y_t) = \mu < \infty$. (This means that the expected value should be independent of time – and finite)

2. It has a finite variance: $Var(y_t) = \sigma_y^2 < \infty$

3. Its autocovariance depends on the difference of the observations (k) only and is independent of time: $Cov(y_t, y_{t-s}) = Cov(y_{t+k}, y_{t-s+k})$

But how can we find out if a series is stationary? First, there is some arithmetical way to find this out, provided we know the DGP. If this is not the case (usually) we can rely on different statistical testing methods, called unit-root tests.

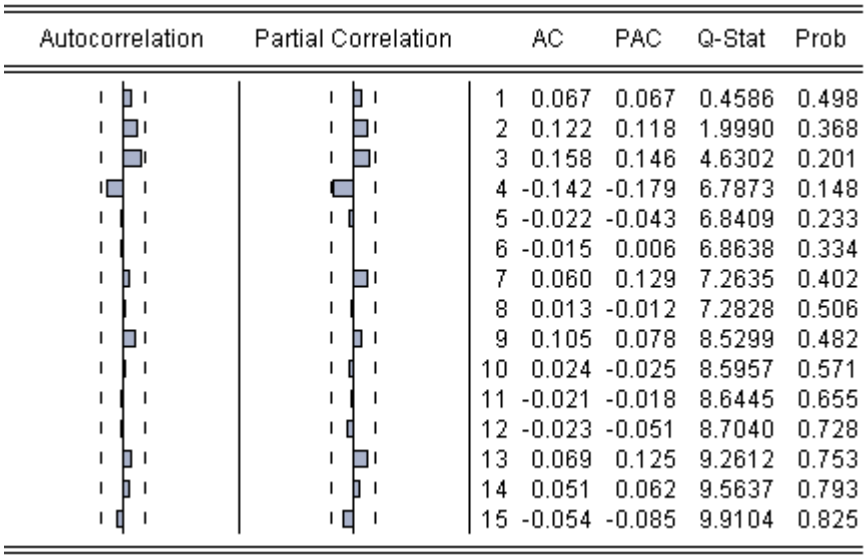
Box and Jenkins identifies three basic type of stationary time-series.

White-noise (WN): White noise has the properties of the ideal disturbance (or error) term. It has zero mean, it is homoscedastic and exhibits no serial correlation whatsoever. Time series ε is a white-noise if:

$E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2_{\varepsilon_t} = \sigma^2_{\varepsilon_{t-k}} = \sigma^2_{\varepsilon}$ for all k, and $Cov(\varepsilon_t, \varepsilon_{t-k}) = 0$ if k≠0.

The white noise is purely random, it contains no useful information whatsoever. By above definition, a white noise is stationary.

The correlogram of a random standard normal variable (a white-noise)



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.067	0.067	0.4586	0.498
		2	0.122	0.118	1.9990	0.368
		3	0.158	0.146	4.6302	0.201
		4	-0.142	-0.179	6.7873	0.148
		5	-0.022	-0.043	6.8409	0.233
		6	-0.015	0.006	6.8638	0.334
		7	0.060	0.129	7.2635	0.402
		8	0.013	-0.012	7.2828	0.506
		9	0.105	0.078	8.5299	0.482
		10	0.024	-0.025	8.5957	0.571
		11	-0.021	-0.018	8.6445	0.655
		12	-0.023	-0.051	8.7040	0.728
		13	0.069	0.125	9.2612	0.753
		14	0.051	0.062	9.5637	0.793
		15	-0.054	-0.085	9.9104	0.825

Moving average (MA): A moving average process of order q equals a (an optional) constant, white noise term, and the linear combination of the lagged errors from t-1 to t-q.

$y_t = \alpha + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t$, $\varepsilon_t \sim WN(0, \sigma^2_{\varepsilon})$ an MA(q) process.

The expected value of an MA(q) process is:

$E(y_t) = \alpha_0 + \sum_{j=1}^{q} \beta_j E(\varepsilon_{t-j}) + E(\varepsilon_t) = \alpha_0$, since this does not depend on t, this condition fulfils the requirements of stationarity.

The variance of an MA(q) process is:

$\sigma^2_{y_t} = \sum_{j=1}^{q} \beta_j^2 \sigma^2_{\varepsilon} + \sigma^2_{\varepsilon} = \left(1 + \sum_{j=1}^{q} \beta_j^2\right) \sigma^2_{\varepsilon}$, that is, as long as q is finite or $\sum_{j=1}^{\infty} \beta_j^2 < \infty$ for all j, this process has a finite variance.

The first-order autocovariance of an MA(1) process is

$$Cov(y_t, y_{t-1}) = E\left(\left(\alpha + \beta_1\varepsilon_{t-1} + \varepsilon_t\right)\left(\alpha + \beta_1\varepsilon_{t-2} + \varepsilon_{t-1}\right)\right) - E\left(y_{t-1}\right)E\left(y_{t-1}\right) =$$

$$= \beta_1 E\left(\varepsilon_{t-1}\varepsilon_{t-1}\right) = \beta_1\sigma_\varepsilon^2$$

and zero for higher orders.

The ACF is:

$$\gamma_k = \begin{cases} \dfrac{\beta_1}{\left(1+\beta_1^2\right)}, k = 1 \\ 0 \quad , k > 1 \end{cases}$$

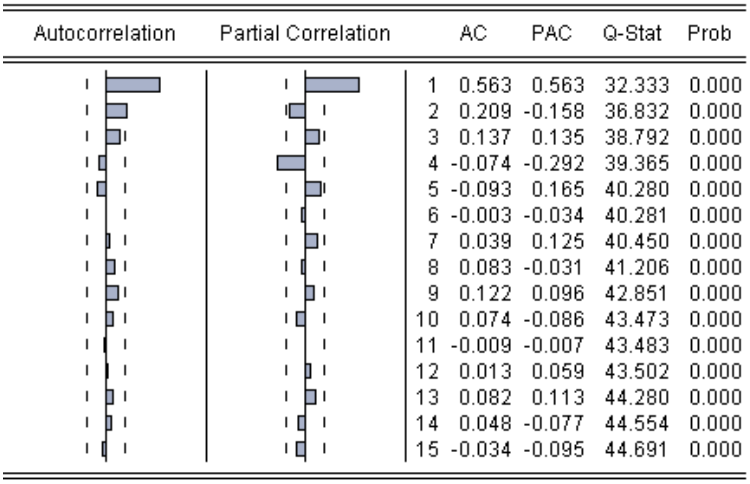
The PACF can be obtained by inverting the MA(1) to an AR(∞) process:

$$y_t = (1+\beta L)\varepsilon_t \rightarrow (1+\beta L)^{-1} y_t = \varepsilon_t \rightarrow \sum_{k=o}^{\infty}\left(-\beta_1 L\right)^k y_t = \varepsilon_t \text{ or}$$

$$y_t - \beta_1 y_{t-1} + \beta_1^2 y_{t-2} - \beta_1^3 y_{t-3} + \ldots = \varepsilon_t \text{ or } y_t = \beta_1 y_{t-1} - \beta_1^2 y_{t-2} + \beta_1^3 y_{t-3} - \ldots + \varepsilon_t$$

Where the coefficients are the respective PACF coefficients.

Correlogram of the MA(1) process: $y_t = \varepsilon_t + .8\varepsilon_{t-1}$



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.563	0.563	32.333	0.000
		2	0.209	-0.158	36.832	0.000
		3	0.137	0.135	38.792	0.000
		4	-0.074	-0.292	39.365	0.000
		5	-0.093	0.165	40.280	0.000
		6	-0.003	-0.034	40.281	0.000
		7	0.039	0.125	40.450	0.000
		8	0.083	-0.031	41.206	0.000
		9	0.122	0.096	42.851	0.000
		10	0.074	-0.086	43.473	0.000
		11	-0.009	-0.007	43.483	0.000
		12	0.013	0.059	43.502	0.000
		13	0.082	0.113	44.280	0.000
		14	0.048	-0.077	44.554	0.000
		15	-0.034	-0.095	44.691	0.000

Autoregressive (AR) model: An autoregressive process of order p equals a (an optional) constant, white noise term, and the linear combination of the lagged dependent variable from t-1 to t-p.

$y_t = \alpha_0 + \sum_{i=1}^{p}\alpha_i y_{t-i} + \varepsilon_t \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ an AR(p) process.

The expected value of an AR(p) process is:

$E(y_t) = \alpha_0 + \sum_{i=1}^{p}\alpha_i E(y_{t-i}) = \dfrac{\alpha_0}{1 - \sum_{i=1}^{p}\alpha_i}$, provided $\sum_{i=1}^{p}\alpha_i < 1$, the expected value of an AR(p) process is finite.

The variance of an AR(p) process is:

$\sigma_{y_t}^2 = \sum_{i=1}^{p} \alpha_i^2 \sigma_{y_{t-i}}^2 + \sigma_\varepsilon^2 \rightarrow \sigma_{y_t}^2 = \frac{\sigma_\varepsilon^2}{1 - \sum_{i=1}^{p} \alpha_i^2}$, that is, as long as $\sum_{i=1}^{p} \alpha_i^2 < 1$, this process has a finite positive variance.

The autocovariance of an AR(1) process is easily obtained by inverting it to an infinite MA process:

$$(1-\alpha_1 L) y_t = \alpha_0 + \varepsilon_t \rightarrow$$

$$y_t = (1-\alpha_1 L)^{-1}(\alpha_0 + \varepsilon_t) = (1 + \alpha_1 L + \alpha_1^2 L^2 + \alpha_1^3 L^3 + \ldots)(\alpha_0 + \varepsilon_t) = y_0 + \sum_{i=0}^{t-1} \alpha_0 \alpha_1^i + \sum_{i=0}^{t-1} \varepsilon_{t-i} \alpha_1^i$$

From this it is obvious that the dynamic multipliers are α_1^k for all lags k. The covariance is:

$$\gamma_k = \alpha_1 \sigma_y^2$$

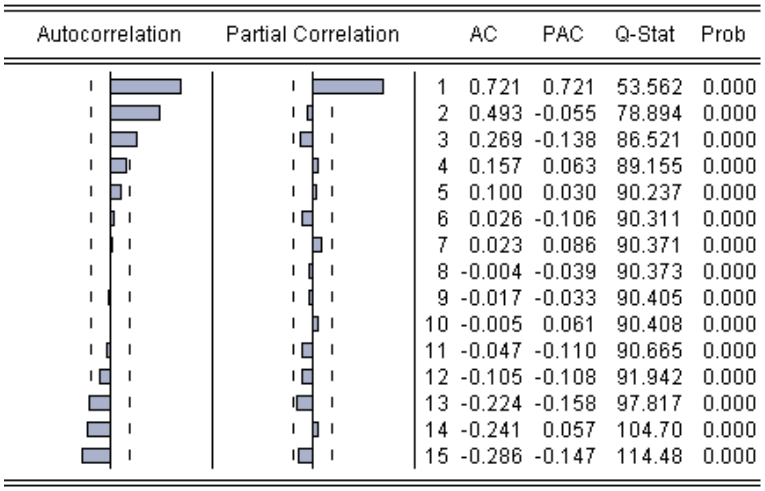
$$\gamma_0 = \sigma_y^2 = \sigma_\varepsilon^2 \sum_{i=0}^{t-1} \alpha_1^{2i} = \frac{\sigma_\varepsilon^2}{1-\alpha_1^2}$$

The ACF at lag k is: $ACF_k = \alpha_1^k$

The PACF at lag k is obvious:

$$PACF(k) = \begin{cases} q, & \text{if } k = 1 \\ 0, & \text{if } k > 1 \end{cases}$$

Correlogram of the AR(1) process: $y_t = 0.7 y_{t-1} + \varepsilon_t$



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.721	0.721	53.562	0.000
		2	0.493	-0.055	78.894	0.000
		3	0.269	-0.138	86.521	0.000
		4	0.157	0.063	89.155	0.000
		5	0.100	0.030	90.237	0.000
		6	0.026	-0.106	90.311	0.000
		7	0.023	0.086	90.371	0.000
		8	-0.004	-0.039	90.373	0.000
		9	-0.017	-0.033	90.405	0.000
		10	-0.005	0.061	90.408	0.000
		11	-0.047	-0.110	90.665	0.000
		12	-0.105	-0.108	91.942	0.000
		13	-0.224	-0.158	97.817	0.000
		14	-0.241	0.057	104.70	0.000
		15	-0.286	-0.147	114.48	0.000

Be careful that AR processes can also be non-stationary.

3. A general solution of difference equations (and the characterization of AR models)

You should recognize that autoregressive time-series models are stochastic difference equations. As a result they can be treated and solved so. Let us define an AR(p) model, where w_t is a stochastic forcing function. It may or may not include a constant term, but it does always include a random, IID disturbance term (ε_t) or innovation which is assumed to be a white noise process.

$y_t = \sum_{i=1}^{p} \gamma_i y_{t-1} + w_t$ where $w_t = \alpha + \varepsilon_t$ (0.1) $\varepsilon_t \sim\sim iid(0, \sigma_\varepsilon^2)$

Even though the behaviour of the above model can be described with a recursive approach, hence by expressing y_t as a function of the starting value y_0 and the past innovations, this may become quite difficult as p is increasing. There is a better way to arrive at a general solution.

All difference equation as in (1.1) can be written in a matrix form:

$\xi_t = F\xi_{t-1} + \nu_t$ (0.2) where $\xi_t = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$, $\xi_{t-1} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{p-1} \end{bmatrix}$, $\nu_t = \begin{bmatrix} w_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ and

$F = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} & \gamma_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$, where F is called the companion matrix and has dimension pxp.

In case of p=2 we have:

$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} \gamma_1 & \gamma_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}$ (0.3) which is the same as a a system of 2 equations:

$$\begin{aligned} y_t &= \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + w_t \\ y_{t-1} &= y_{t-1} \end{aligned} \quad (0.4)$$

So basically what we do here is to transform any p-order differential equation into a system of p equations.

The great advantage of doing so is that we actually transformed a p order problem into a first order problem. For example a a P order equation can be rewritten recursively as follows:

$\xi_t = F\xi_{t-1} + \nu_t = F^2\xi_{t-2} + \nu_t + F\nu_{t-1} = F^t\xi_0 + \sum_{j=0}^{t} F^j \nu_{t-j}$ (0.5)

Or you can start out from period 0 as well:

$\xi_0 = \xi_0$

$\xi_1 = F\xi_0 + \nu_1$

$\xi_2 = F^2\xi_0 + \nu_2 + F\nu_1$

$\xi_3 = F^3\xi_0 + \nu_3 + F^2\nu_2 + F\nu_1$

$\xi_t = F^t \xi_0 + \sum_{i=1}^{t} F^{t-i} v_i$ (0.6) which is equivalent with (1.5), just we changed the numbering of the forcing process from backward to forward.

So when we wish to estimate the effect of an innovation in period i on y in period t (i≤t), then can simply express it from (1.6) as follows:

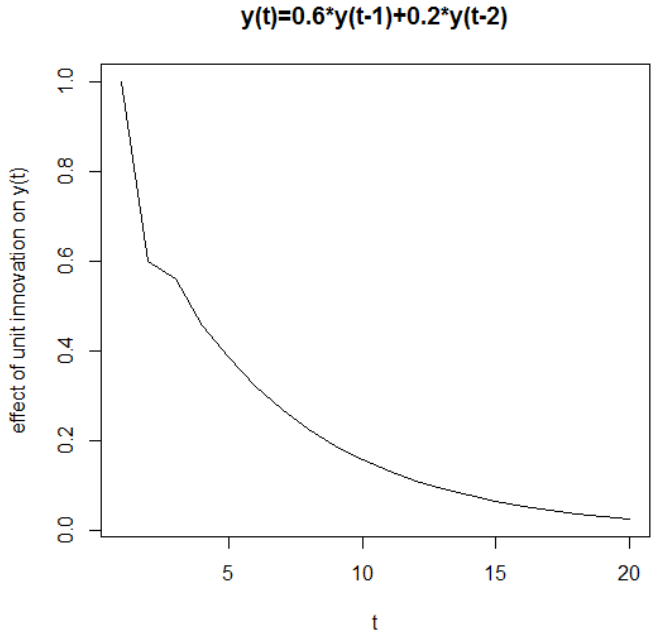
$\frac{\partial \xi_t}{\partial \varepsilon_i} = F^{t-i}$ (0.7) The question of course, what is F^{t-i}? Fortunately, since F is a quadratic matrix, we can make use of the diagonalization. Hence $F^t = E\Lambda^t E^{-1}$ (0.8), where E is the matrix of eigenvectors, and Λ is a diagonal matrix of eigenvalues. Since Λ is diagonal:

$\Lambda^t = \begin{bmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p^t \end{bmatrix}$. Hence, $\lim_{t\to\infty} F^t = 0$ only if $\lim_{t\to\infty} \Lambda^t = 0$ which is only possible if all eigenvalues of F are less than one. This is the condition for stationarity. If this is not the case, and if only one of the roots (eigenvalues) equals one in absolute terms, the effect of an innovation will never decay. What is more, it can even be explosive (increasing as we get farther away from the innovation), if at least one of the eigenvalues exceeds one.

We can use the above form to simulate the behaviour of the process $y_t = 0.6y_{t-1} + 0.2y_{t-2} + \varepsilon_t$ (0.9). Copy this small code into R to find out how:

```
F<-matrix(NA,2,2)

F<-cbind( c(.6,1),c(0.2,0))

Y<-matrix(NA,20,2)

Y[1,]<-c(1,0)

eF<-eigen(F)

for (i in 2:20) {

Y[i,]<-F%*%Y[i-1,]

t[i]=i}

plot(t,Y[,1],type="l",main="y(t)=0.6*y(t-1)+0.2*y(t-2)",ylab="effect of unit innovation on y(t)")
```

You should obtain the following graph:



What we did was simply to assign some starting value in the vector Y (y_1=1,y_0=0), and we simply used (1.6) to estimate the value of y at any period t.

Alternatively we can solve the equation (1.9) so that $y_0 = 0$ and $y_1 = 1$.The general solution to any p-order homogenous difference equation is $y_t = \sum_{i=1}^{p} A_i \lambda_i^t$ (0.10). So if $y_t = 0.6y_{t-1} + 0.2y_{t-2} + \varepsilon_t$ then the roots of the companion matrix are 0.84 and -0.24 (rounded to 2 digits).

The eigenvalues can be obtained from the companion matrix as follows:

$F = \begin{bmatrix} 0.6 & 0.2 \\ 1 & 0 \end{bmatrix}$ since $Fv = \lambda v \rightarrow (F - \lambda I)v = 0$

$F - \lambda I = \begin{bmatrix} 0.6-\lambda & 0.2 \\ 1 & -\lambda \end{bmatrix} \rightarrow |F - \lambda I| = \lambda^2 - 0.6\lambda - 0.2 = 0$ which is the characteristic polynomial.

the roots (eigenvalues) are: 0.838 and -0.2385

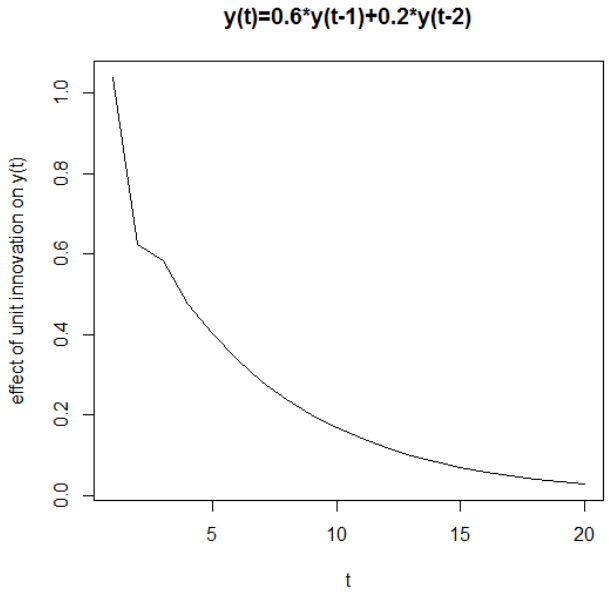
Hence $y_t = A_1 0.84^t + A_2(-0.24)^t$ is a general solution. No we can go for a particular solution.

if $y_0 = 0$ then $A_1 = -A_2$ $y_t = A(0.84^t - (-0.24)^t)$

If $y_1 = 1$, then $1 = A(0.84 - (-0.24)) \rightarrow A = \frac{1}{1.04}$ the particular solution is then:

$y_t = \frac{1}{1.04}(0.84^t - (-0.24)^t)$ which can be also plotted in R:

```
x<-(1/1.04)*(.84^t-(-.24)^t)

plot(t,x,type="l",main="y(t)=0.6*y(t-1)+0.2*y(t-2)",ylab="effect of unit innovation on y(t)")
```



Not surprisingly, this is the same as what we simulated from the matrix form.

Let us look at some other cases.

Example 1: non-stationary series with real roots

$$y_t = 0.8y_{t-1} + 0.2y_{t-2} + \varepsilon_t$$

$$F = \begin{bmatrix} 0.8 & 0.2 \\ 1 & 0 \end{bmatrix}$$

The eigenvalues are 1 and -0.2, the general solution is hence:

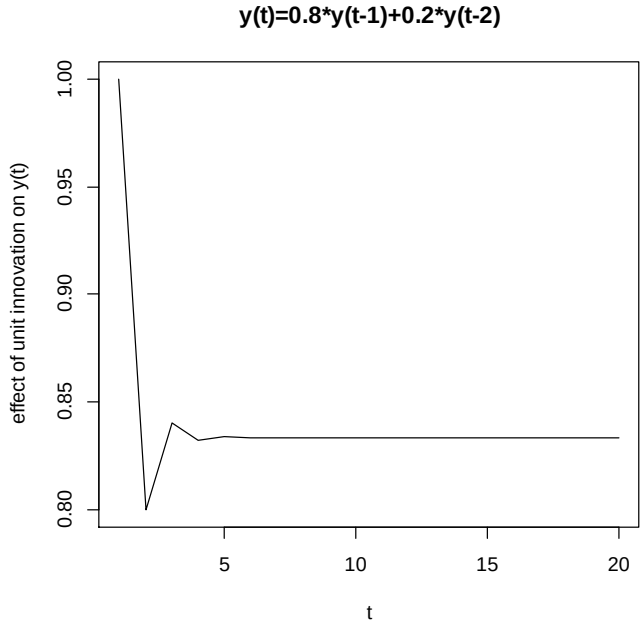
$y_t = A_1 1^t + A_2(-0.2)^t$ Now we look for the particular solution so that $y_0 = 0$ and $y_1 = 1$.

$0 = A_1 + A_2 \rightarrow A_1 = -A_2$ $y_t = A(1^t - (-0.2)^t)$

and $1 = A(1-(-0.2)) \rightarrow A = \frac{1}{1.2}$ so the particular solution is:

$$y_t = \frac{1}{1.2}(1^t - (-0.2)^t)$$

You can observe that one of the eigenvalues (or characteristic roots) of the companion matrix was one (unit-root). and this assures that the effect of the unit impulse in t=1 will not fade away.



Example 2: Explosive series

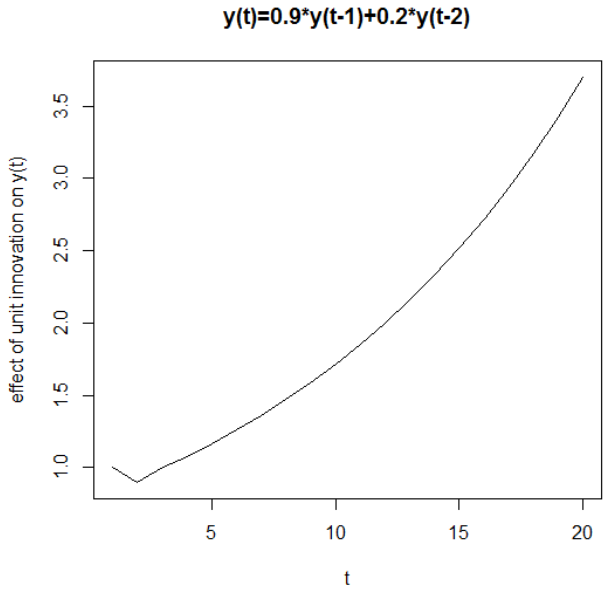
This is by no means usual in economics or social sciences. Yet, for the example's sake we can try this one:

$$y_t = 0.9y_{t-1} + 0.2y_{t-2} + \varepsilon_t$$

The eigenvalues of F are 1.08, -0.18, hence this process is explosive.

The solution if $y_0 = 0$ and $y_1 = 1$ is $y_t = \frac{1}{1.26}(1.08^t - (-0.18)^t)$

The graph reflects the explosive nature of this model:



Example 3: Stationary process with complex root

Let us have the following AR(2) model:

$$y_t = 0.8y_{t-1} - 0.6y_{t-2} + \varepsilon_t$$

The eigenvalues of F are 0.4+0.663i and 0.4-0.663i. The process is stationary since the real parts of the roots are all smaller the one. Yet, complex roots has an additional meaning: the process oscillates. Fortunately, all complex numbers z can be rewritten in a trigonometric form:

$z = a + bi = r(\cos\varphi + i\sin\varphi)$ where $\varphi = \tan^{-1}\left(\frac{b}{a}\right)$ and $r = |z| = \sqrt{(a+bi)^2}$

Hence $r = 0.774$ and $\varphi = 1.0279$

The general solution when the roots are complex is:

$$y_t = A_1 r^t \cos(\varphi t) + A_2 r^t \sin(\varphi t)$$

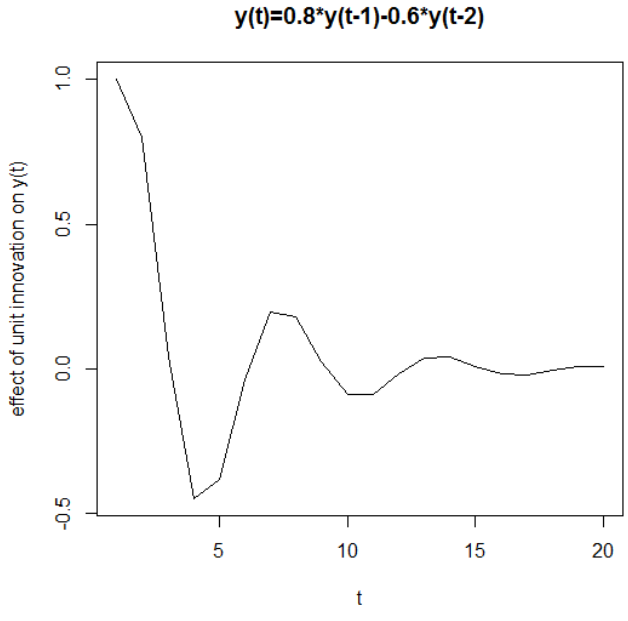
Hence: $y_t = A_1 (0.774)^t \cos(1.0279t) + A_2 (0.774)^t \sin(1.0279t)$

If $y_0 = 0$ and $y_1 = 1$ then the particular solution: $0 = A_1$

$$1 = 0.774A\sin(1.0279) \rightarrow A = \frac{1}{0.774\sin(1.0279)}$$

the particular solution is then:

$$y_t = \frac{(0.774)^t \sin(1.0279t)}{0.774\sin(1.0279)}$$



4. Lag operators

Lag operators can make our life easier, when used properly. Let L be an operator with the following characteristics:

$Ly_t = y_{t-1}$ that is multiplying by L creates a time series lagged by one period. Since $L(Ly_t) = L^2 y_t = y_{t-2}$. Generally: $L^k y_t = y_{t-k}$

The lag operator is commutative, distributive and associative:

For example:

$(L^1 + L^2)y_t = y_{t-1} + y_{t-2}$, $L^1 L^2 y_t = L^3 y_t = y_{t-3}$, $L^{-1} y_t = y_{t+1}$

The operation: $(1-L)y_t = y_t - y_{t-1}$ is called first-differencing.

An important operation is inversion:

$(1-\theta L)^{-1} y_t = (1+\theta L + \theta^2 L^2 + \theta^3 L^3 + \ldots) y_t = \theta(L) y_t$ provided $|\theta| < 1$ (the series at the right-hand side is absolute summable).

Proof: multiply both sides by $(1-\theta L)$:

$$y_t = (1-\theta L)(1+\theta L + \theta^2 L^2 + \theta^3 L^3 + \ldots) y_t = (1+\theta L + \theta^2 L^2 + \theta^3 L^3 + \ldots) - (\theta L + \theta^2 L^2 + \theta^3 L^3 + \theta^4 L^4 + \ldots)$$
$$y_t = (1-\theta^n L^n) y_t$$

If $\lim_{n\to\infty}(\theta^n L^n y_t) = 0$ if $|\theta| < 1$, the equality holds.

Bear in mind that the inversion was true only if the effect of an observation in the distant past on current value was negligible. This is what we expect from stationary series. **Only stationary series are invertible.**

The above finding is known as **Wold's theorem**. Any stationary time-series can be expressed as an infinite MA series. This is a fundamental result in statistics. A practical version is the inversion of a stationary AR process to an infinite MA process.

Let $y_t = \gamma_0 + \gamma_1 y_{t-1} + \varepsilon_t$ with lag operators this is equivalent with: $(1-\gamma_1 L)y_t = \gamma_0 + \varepsilon_t$. Multiplying both sides by the inverse $(1-\gamma_1 L)^{-1}$ yields:

$$y_t = (1-\gamma_1 L)^{-1}(\gamma_0 + \varepsilon_t) = (1+\gamma_1 L + {\gamma_1}^2 L^2 + {\gamma_1}^3 L^3 + \ldots)(\gamma_0 + \varepsilon_t) =$$
$$= \gamma_0 \sum_{i=0}^{\infty} {\gamma_1}^i + \sum_{i=0}^{\infty} {\gamma_1}^i \varepsilon_{t-i} = \frac{\gamma_0}{1-\gamma_1} + \sum_{i=0}^{\infty} {\gamma_1}^i \varepsilon_{t-i} \quad \text{, if } 0 < |\gamma_1| < 1$$

We can use the lag operator form as well to find the characteristic roots of a difference equation.

A p-order autoregressive process is stationary if all the roots (solutions) of the following polynomial:

$1-\theta_1 z-\theta_2 z^2-\theta_p z^p=0$ lie outside the unit circle (if the roots are real (not complex) it means that all roots should be larger than one in absolute value).

Some examples:

Ex 1.1: $y_t=0.5y_{t-1}+\varepsilon_t$

the polynomial form is:

$(1-0.5L)y_t=\varepsilon_t$

The characteristic equation is:

$1-0.5z=0$, where the root is: $z=2$ so this AR(1) model is stationary.

Ex 1.2: $y_t=0.5y_{t-1}+0.3y_{t-2}+\varepsilon_t$

the polynomial form is:

$(1-0.5L-0.3L^2)y_t=\varepsilon_t$

The characteristic equation is:

$1-0.5z-0.3z^2=0$, where the roots are: $z_{1,2}=\dfrac{0.5\pm\sqrt{0.25-4\cdot(-0.3)\cdot 1}}{-0.6}$ $z_1=-2.84, z_2=1.17$

Both roots exceed 1 in absolute terms, the above AR(2) model is stationary.

Ex 1.3: $y_t=0.4y_{t-1}-0.2y_{t-2}+\varepsilon_t$

the inverse polynomial form is:

$(1-0.4L+0.2L^2)y_t=\varepsilon_t$

The characteristic equation is:

$1-0.4z+0.2z^2=0$, where the roots are: $z_{1,2}=\dfrac{0.4\pm\sqrt{0.16-4\cdot 0.2\cdot 1}}{0.4}$.

There is a problem now because the expression under the square root is negative. The roots are complex in this case. No worry:

$\sqrt{0.16-4\cdot 0.2\cdot 1}=\sqrt{-0.64}=\sqrt{0.64}i=0.8i$ ($i=\sqrt{-1}$)

So: $z_{1,2}=\dfrac{0.4\pm 0.8i}{0.4}=1\pm 2i$ The absolute value of these complex numbers is its modulus:

$|1+2i|=|1-2i|=\sqrt{1^2+2^2}=\sqrt{5}=2.236$, which is larger than one. The process is stationary.

This means that the process will return to its mean in an oscillating way.

It is because of potentially complex roots that we mention unit circle in the definition. Simply, the modulus looks like an equation of a circle. What the definition required was that if the complex root was a+bi or a-bi then $a^2+b^2>1$. That is the modulus lies outside of a circle with unit radius.

Do not get confused. In section 3 we were looking for the characteristic roots of the polynomial of an AR form, and we required that all characteristic roots are within the unit circle. What we are looking for in section 4, are the roots of the **inverse polynomial**, hence roots outside the unit-circle are desired for stationary series.

Let us look at the behaviour of the following series:

$$y_t = 1 + 0.5y_{t-1} - 0.3y_{t-2} + u_t$$

By the first method (using the eigenvalues of the companion matrix) we obtain two complex eigenvalues (roots): 0.25±0.487i. The modulus ($0.25^2+0.487^2=0.3$) is less than one, hence the process is stationary.

Let us rewrite the process with lag operators now!

$(1-0.5L+0.3L^2)y_t = 1+u_t$ The characteristic equation is: $1-0.5z+0.3z^2=0$, and the roots of the inverse polynomial are 0.83±1.62i. The modulus ($0.83^2+1.62^2=3.31$) exceeds one, so the process is stationary. Observe that the modulus from this equation is 1 over the modulus of the other (deviations are due to rounding errors).

2.c Testing for the order of integration

By simply looking at a series you will not be able to decide if it is stationary or not or if it needs to be differenced. Let us take the following two series:

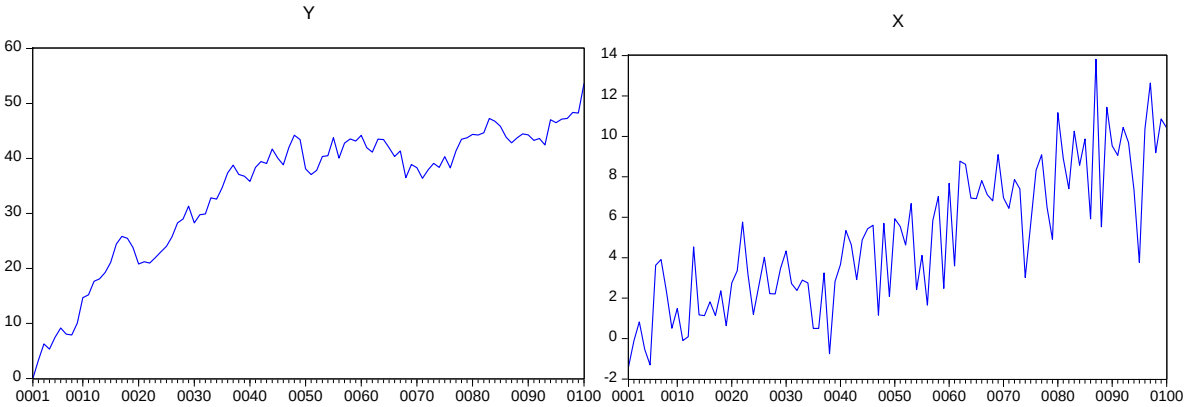
$y_t = \alpha + y_{t-1} + \varepsilon_t$ a random walk with drift (also called stochastic trend model)

and

$y_t = \alpha t + \eta_t$ a deterministic trend model (also called trend stationary)

The expected value of both series is going to be:

$Ey_t = \alpha t$, provided the initial value is zero.



Would you be able to tell which of the above series are produced by a deterministic or a stochastic trend model if you had not been told the truth? The honest is answer is no. You can guess but that is not enough.

Actually the distinction between trend stationary and random walk with drift series came into the focus of debates in macroeconomics during the '70s and remained there as a crucial point of research until the '90s.

The traditional view on business cycles was motivated by Keynesian models, where the long-run output was determined by technology (supply side) leading to some smooth growth of output, and business cycles were attributed to short-run deviations from equilibrium. This idea can be modeled by a trend stationary process:

$\ln y_t = \alpha + \delta t + u_t$, where y denotes per capita income and the parameter δ is the growth rate of per capita income per period. The residual u has all the shocks that would capture the deviations from the long-run growth path. In this model one would expect that the shocks have just an immediate effect so the economy should return to its long-run growth path.

If you do not like this much, you can assume that it takes some time to return to equilibrium, by making the model autoregressive:

$\ln y_t = \alpha_0 + \alpha_1 \ln y_{t-1} + \delta t + u_t$ (0<|α₁|<1), but the model is still trend stationary.

Kydland and Prescott, followed by many, argued that cycles may actually arise even in equilibrium. That is, the equilibrium output of the economy, that had been believed to follow a smooth path, may also exhibit cyclical behaviour. Because this means that there are shocks in real (not nominal) variables, this school is called the Real Business Cycle School or RBC.

This idea is best captured by a stochastic trend model:

$$\ln y_t = \ln y_{t-1} + \delta + u_t$$

The economy will have a tendency to grow at a rate δ, but any shocks (in this context it is called technological or productivity shock) could have a lasting effect. Remember: having a unit-root means exactly the same. So cycles are not always signs of disequilibrium.

This is the reason why the article of Granger and Newbold in 1974 became so influential. People started testing if macroeconomic series playing a role in RBC theory (GDP, employment, TFP) were trend stationary or not.

Dickey Fuller test:

The first and most fundamental unit-root test (1979).

There are three possible null-hypotheses:

1. The DGP is a random-walk: $y_t = y_{t-1} + u_t$ with $u_t \sim IID(0, \sigma_u^2)$.

   In this case the test equation is $\Delta y_t = \rho y_{t-1} + u_t$, where if $\rho = 0$ then the null hypothesis is accepted, if $\rho < 0$ then it is rejected. The test equation is estimated by OLS, but the distribution of the coefficient is not going to converge to normality so you should not use Student's t-distribution to find critical values. Instead critical values are supplied in a different table or built in into some packages.

2. The DGP is either a random-walk with a constant or a random-walk with drift: $y_t = \alpha + y_{t-1} + u_t$ with $u_t \sim IID(0, \sigma_u^2)$. In this case the test equation is $\Delta y_t = \beta + \rho y_{t-1} + u_t$, where if $\rho = 0$ then the null hypothesis is accepted, if $\rho < 0$ then it is rejected.

3. The DGP is a random-walk with a constant and a deterministic trend: $y_t = \alpha + y_{t-1} + \delta t + u_t$ with $u_t \sim IID(0, \sigma_u^2)$. In this case the test equation is $\Delta y_t = \beta_0 + \rho y_{t-1} + \beta_1 t + u_t$, where if $\rho = 0$ then the null hypothesis is accepted, if $\rho < 0$ then it is rejected. If we reject the null-hypothesis but find that $\beta_1 \neq 0$, then the series are trend-stationary.

So that the DF test works properly the residual should exhibit no serial correlation. But this assumption is too bold. Let us see why:

You can observe that the DF test uses an AR(1) model to test is a series is non-stationary. But what if the real DGP is a higher-order AR model?

$y_t = 0.7y_{t-1} + 0.3y_{t-2} + \varepsilon_t$, this model is also non-stationary since the sum of the two coefficients equals one (alternatively you can estimate the roots that are -1.43 and 1).

After differencing the real DGP becomes:

$$\Delta y_t = -0.3y_{t-1} + 0.3y_{t-2} + \varepsilon_t$$

If you approximate this equation by an AR(1) model you will get a coefficient close to zero but the omitted lags will cause an autocorrelation in u.

There are two solutions:

1. Correcting the test statistics of the DF test for serial correlation and possible heteroscedasticity: this is the **Phillips-Perron** (PP) test.

2. Adding lags of the dependent variable in the test equation of the DF test to capture autocorrelation: **Augmented Dickey-Fuller** test (ADF test).

The augmented Dickey Fuller test has the following test equations:

$$\Delta y_t = \rho y_{t-1} + \sum_{i=1}^{q} \Delta y_{t-i} + u_t, \Delta y_t = \beta + \rho y_{t-1} + \sum_{i=1}^{q} \Delta y_{t-i} + u_t, \Delta y_t = \beta_0 + \rho y_{t-1} + \beta_1 t + \sum_{i=1}^{q} \Delta y_{t-i} + u_t$$

Choosing the q can be done based on model selection statistics, like the AIC or BIC. Smarter softwares select q for you, based on some criteria.

Other modifications for the DF test are also available: the DF-GLS test for example, detrend the series before running a standard ADF procedure.

The main problem of DF type test is the low power. This means that the probability of Type II error is high. For the DF type tests this means that you have a high chance that when you have an autoregressive DGP with a high degree of positive autocorrelation, close to but less than one, the test will not reject the null hypothesis.

Kwiatkowski, Phillips, Schmidt and Shin suggested a test with stationarity as null hypothesis (KPSS-test).

The null hypotheses can be that the series is either stationary or trend stationary. Depending on this the test equations are:

$y_t = \alpha + \hat{u}_t$ or $y_t = \alpha + \delta t + \hat{u}_t$

The statistics is based on the residual:

$S_t = \sum_{s=1}^{t} \hat{u}_s$ which is the recursive sum of the residual.

The test statistics is: $KPSS = \frac{T^{-2} \sum_{t=1}^{T} S_t^2}{\hat{\sigma}^2}$ where $\hat{\sigma}^2$ is the estimated long-run variance of $\hat{u}$.

Using the unit root tests, we can determine the order of integration as follows:

1. We carry out a unit-root test on the level of y: if it is stationary then the process is I(0). If not, we take the difference of the series.

2. We carry out a unit-root test on Δy. If it is stationary then y is I(1), if not, we difference Δy further.

3. The process is followed as long as it needed to achieve stationarity.

Be aware that the unit-root test suffer from a power problem: they tend to accept the null hypothesis of non-stationarity even when it is not true, especially if the autoregressive coefficient is close to one.

2.d Overdifferencing

A word of caution: it is possible to overdifference the series, that is, to take the difference of a stationary series, which will lead to a special autoregressive pattern. Overdifferencing will also increase the noise (measurement errors and sampling error) relative to signal, leading to a decrease in efficiency.

Let us assume that y is stationary:

$y_t = \varepsilon_t$ with $\varepsilon_t \sim IID(0, \sigma_u^2)$

taking first difference yields:

$x_t = \Delta y_t = \varepsilon_t - \varepsilon_{t-1}$

The first consequence is obvious:

$\sigma_x^2 = \sigma_{\varepsilon_t}^2 + \sigma_{\varepsilon_{t-1}}^2 = 2\sigma_\varepsilon^2$ So the differenced series is going to have higher variance.

The second consequence is that x is going to be serially correlated:

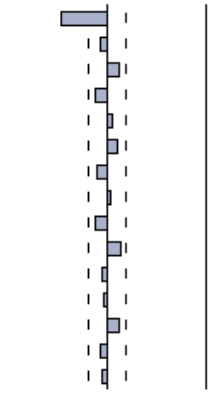
$$Cov(x_t, x_{t-1}) = E\left(\left(\varepsilon_t - \varepsilon_{t-1}\right)\left(\varepsilon_{t-1} - \varepsilon_{t-2}\right)\right) = E\left(\varepsilon_t\varepsilon_{t-1} - \varepsilon_t\varepsilon_{t-2} - \varepsilon_{t-1}^2 + \varepsilon_{t-1}\varepsilon_{t-2}\right) = -\sigma_\varepsilon^2$$

$$\gamma_1 = \frac{Cov(x_t, x_{t-1})}{\sigma_x^2} = -0.5\,,\ \gamma_j = \frac{Cov(x_t, x_{t-j})}{\sigma_x^2} = 0 \text{ for all j>1}$$

That is, if we obtain a variable after differencing that has a first order autocorrelation around -0.5

then we have reason to believe that we overdifferenced it.

ACF and PACF of the first difference of a standard normal variable.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.493	-0.493	24.835	0.000
		2	-0.071	-0.416	25.355	0.000
		3	0.118	-0.218	26.810	0.000
		4	-0.124	-0.299	28.424	0.000
		5	0.054	-0.266	28.738	0.000
		6	0.113	-0.067	30.121	0.000
		7	-0.099	-0.031	31.179	0.000
		8	0.036	0.057	31.325	0.000
		9	-0.139	-0.175	33.462	0.000
		10	0.153	-0.041	36.076	0.000
		11	-0.054	-0.102	36.407	0.000
		12	-0.046	-0.188	36.650	0.000
		13	0.129	-0.066	38.580	0.000
		14	-0.066	0.007	39.096	0.000
		15	-0.064	-0.017	39.589	0.001

Note: observe that the ACF(1) is close to -0.5.