ARIMA estimation: theory and applications

Matteo Barigozzi

1 General features of ARMA models

Let's consider a generic time series x_t[1] and let's define:

- the autocovariance function: $\gamma(h) = Cov(y_{t+h}, y_t)$
- the autocorrelation function (ACF): $\rho(h) = Corr(y_{t+h}, y_t) = \frac{\gamma(h)}{\gamma(0)}$
- the partial autocorrelation function (PACF):

$$\begin{aligned}\alpha(0) &= 1\\ \alpha(h) &= \phi_{hh} \quad \forall h > 0\end{aligned}$$

  where ϕ_{hh} is the coefficient of y_1 in predicting y_{h+1} in terms of the intermediate observations $y_2 \dots y_h$ i.e. $P_h y_{h+1} = \phi_{h1} y_h + \dots + \phi_{hh} y_1$.

The PACF in practice eliminates the effects of the intervening values between y_1 and y_h. We say that the process is weakly stationary if

$$E[y_t] = \mu \qquad\qquad \forall t$$

$$Cov(y_{t+h}, y_t) = Cov(y_{t+h+j}, y_{t+j}) \quad \forall j \text{ and } \forall h$$

Now let's assume to have a stationary time series y_t; this will be modeled in the most general case as an autoregressive of order p, moving average of order q process i.e. an ARMA(p,q). The model is the following [2]:

$$A(L) y_t = B(L) u_t \quad \text{with } \; u_t \sim \text{white noise}(0, \sigma^2)$$

where A(L) and B(L) are polynomials in the lag operator L:

$$\begin{aligned}A(z) &= 1 - a_1 z - \dots - a_p z^p\\ B(z) &= 1 + b_1 z + \dots + a_q z^q\end{aligned}$$

We assume that they don't have common zeros and we have the definitions:

[1] This can be viewed as a specification of the joint distribution of a sequence of random variables $\{X_t\}$ whose realizations are x_t.

[2] A white noise u_t is a stationary process such that:

$$E[u_t] = 0 \quad \text{and} \quad \gamma(h) = \begin{cases} \sigma^2 & \text{if} \quad h = 0 \\ 0 & \text{if} \quad h \neq 0 \end{cases}$$

.

- y_t is causal and stationary iff $A(z) \neq 0 \; \forall z \in \mathbb{C}$ s.t. $|z| \leq 1$ i.e it has roots only outside the unit circle; [3]

- y_t is invertible iff $B(z) \neq 0 \; \forall z \in \mathbb{C}$ s.t. $|z| \leq 1$ i.e. it has roots only outside the unit circle and the representation is called fundamental.

A causal AR process can be rewritten as an infinite MA: $y_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$ with $\psi(z) = B(z)/A(z)$.
An invertible MA can be rewritten as an infinite AR: $u_t = \sum_{j=0}^{\infty} \phi_j y_{t-j}$ with $\phi(z) = A(z)/B(z)$.

2 Deterministic and stochastic trend

Actually the real data that we have are never stationary thus it is useful to represent the general time series x_t as consisting of three parts: a trend m_t , a seasonal component s_t and a random noise component y_t which is the stationary part. Indeed most of economic time series appear to increase over time and the study of time trends is the main concern of growth theorist. We're instead interested in the irregular part since it is believed to contain the information about the business cycle. Hence detrending and deseasonalizing a time series will make it stationary thus allowing us to use the ARMA estimation techinques [4]. Many macroeconomists believe that technological advancements have permanent effect on the trend of macroeconomy. Given that technological innovations are stochastic the trend should reflect this underlying randomness. Thus it is useful to consider models with stochastic trend. Let's compare the two possible methods to eliminate the trend af a time series.

- Differencing the process until it becomes stationary; suppose we need to do it d times, we then say that the process is integrated of order d or I(d) and it has d unit roots. It will be modelled as an ARIMA(p,d,q) and that's the subject of the following work.

- Detrending by computing the follwing regression with least-squares:

  $$x_t = \alpha + \beta t + y_t$$

  and keeping y_t which should be the stationary part.

In the first case we have a difference stationary model while in the second we have a trend stationary model. The belief that trend is not changing over time will lead to eliminate it by using a linear deterministic regression. However the trend may not be deterministic as argued by real-business cycle theorists. If a variable is trend stationary then any economic shock will not have long run effects, while if it is difference stationary there might be persistent effects of the shocks (e.g. technological shocks). Note that detrending a difference stationary process

[3] Note that if it has roots inside the unit circle the process is not causal but it's still stationary while if it has roots on the unit circle it is non-stationary.

[4] For the moment we assume that the hypothesis of constant variance holds; later on we'll consider also the problem of errors' heteroshdasticity.

will lead to a non stationary process. This and other significant results were found by Nelson and Plosser (1982) in analyzing many macroeconomic variables. Their work challanged the traditional view of trend stationary processes by showing that economic variables are in fact difference stationary. Their results are strongly indicative of unit root processes. Unit root tests are designed exactly in a way such that a linear trend is included and once it's removed the resulting process is tested for stationarity, if the result is not stationary then we say that the process has a unit root and thus it's difference stationary. Finally let's note that saying that macroeconomic variables are difference stationary doesn't imply that real-business cycle theory is for sure the correct way to interpret economy.

3 The tests for unit roots

A slowly decaying ACF is indicative of a large characteristic root, true unit root (i.e. stochastic trend) or trend stationary process. Formal tests can be done to determine which is the case. As already discussed a general (deseasonalized) process x_t can be decomposed into:

- a deterministic trend that for simplicity is assumed linear: $m_t = \alpha + \beta t$;
- a noise term that for simplicity is assumed to be an ARMA(p,q): $A(L)y_t = B(L)u_t$ with u_t white noise [5].

The unit root tests are concerned with the behavior of the second term but the specification of a deterministic trend is crucial in the hypotesis testing. Invertibility is also required thus B(L) has roots only outside the unit circle. We know that if x_t is trend stationary then y_t is stationary therefore the roots of A(L) are all outside the unit circle. In a difference stationary process instead y_t has one unit root in the autoregressive part. In this case $\Delta y_t = (1 - L)y_t$ is stationary. The unit root hypothesis is that y_t is difference stationary i.e. it is I(1) [6]. We can further decompose y_t into:

- a cyclical component (the business cycle) C_t stationary with zero mean so that has no lung run impact on the level of x_t;
- a stochastic trend T_t that incorporates all shocks that have long run effects.

In trend stationary models there is no stochastic trend and y_t is already the cyclical part, while in difference stationary models we need the Beveridge and Nelson (1981) decomposition to isolate the two components: y_t is divided into a random walk component built as a moving average of u_t hence it represents the long-run effect of a shock u_t and into a cyclical part (or noise) that has no long run effect.

[5] In general it is possible to allow for some degree of heterogeneity in the innovations which permits richer correlations. This is the subject of the Philips-Perron test.

[6] We don't consider the case of multiple unit roots.

3.1 The simplest case - Dickey Fuller test

Let's take y_t as an AR(1) process:

$$y_t = a y_{t-1} + u_t \quad \text{or} \quad \Delta y_t = \gamma y_{t-1} + u_t$$

We could limit ourselves to test the previous regression but we want to include also the possibility of a deterministic trend. Considering only the linear case we can estimate the two regressions:

$$\begin{aligned} x_t &= \alpha + \beta t + y_t \\ \Delta y_t &= \gamma y_{t-1} + \epsilon_t \end{aligned}$$

This is equivalent to estimate the single regression:

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \epsilon_t$$

with the following system of hypothesis:

$$\begin{cases} H_0: \ \gamma = 0 & \text{i.e. the process has a unit root thus it's I(1)} \\ H_1: \ \gamma < 0 & \text{i.e. the process is trend stationary} \end{cases}$$

The asymptotic distribution of t-values for γ is asymmetric and the critical values can be found in tables for different sets of regressors used. Such distributions were estimated by Dickey and Fuller (1979 and 1981) using Monte Carlo simulations. They reported three tables for these critical values for three different models:

$$\begin{aligned} \Delta x_t &= \gamma x_{t-1} + \epsilon_t & &\text{random walk} \\ \Delta x_t &= \alpha + \gamma x_{t-1} + \epsilon_t & &\text{random walk with drift} \\ \Delta x_t &= \alpha + \beta t + \gamma x_{t-1} + \epsilon_t & &\text{random walk with drift and linear trend} \end{aligned}$$

3.2 The general case - Augmented Dickey Fuller test

For a general AR(p) process it's enough to test the hypothesis $\gamma = 0$ by considering the following regression:

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{p-1} a_i \Delta x_{t-i} + \epsilon_t \tag{1}$$

The critical values for the test are the same of the AR(1) case and they depend on the presence or not of the linear trend and constant terms.
The test assumes that the errors ϵ_t are independent and have constant variance, but we actually don't know which is the real data generating process. We have four problems:

1. the true data generating process may contain also moving average components then we must extend the test to this case;
2. the true orders either of the autoregressive part and of the moving average part are unknown;
3. multiple roots may be present;

4. we don't know if deterministic trend and intercept are present or not.

Since an invertible MA can be transformed into an autoregressive model we can easily generalize the test for an ARMA(p,q) process by taking the following regression:

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{\infty} a_i \Delta x_{t-i} + \epsilon_t$$

obviously this regression cannot be estimated unless we set an upper bound for the autoregressive order. In practice following Said and Dickey (1984) we can approximate an unknown ARIMA(p,1,q) with an ARIMA(n,1,0) with $n \sim T^{1/3}$ where T is the length of the series.
Concerning the second problem we note that including too many lags will reduce the power of rejecting the null hypothesis, and too few will not capture the actual error process. One possibile solution is to start estimating with a relatively long lag n^* and check the t-statistic for the coefficient of lag n^* in the previous regression, if the value is insignficant we move to $n^* - 1$ and repeat the procedure until we reach a significant lag. Finally a test for the whiteness of residuals must be done in order to be sure not to have correlations among them [7].
We can repeat the Dickey Fuller test on the first differenced process and if a second unit root is present we differentiate the process again. This can be repeated until the result is stationary i.e. when the unit root test rejects the null hypothesis.
Concerning the last problem the presence of additional estimated parameters reduces degrees of freedom and the power of the test. Reduced power implies that one might find a unit root when in fact is not present. The issue is to decide which deterministic regressors we must include in the test. Dickey and Fuller provided tables to test simultaneously the significance of the coefficients in such model; by means of these we can try to find a procedure to find the best model for the deterministic part. This is represented graphically in figure 1 and we see that starting from the most general case of equation (1) we move to the simplest one with only a random walk component. This procedure is possible because the t-statistics for the cases with drift and linear trend converge to the standard normal (Campbell and Perron (1991)). However this method must not be applied mechanically and plotting the data plus theoretical considerations are usually important to determine the presence of deterministic regressors.

Finally it's important to recognize that these tests have low power to distinguish between a real unit root and a near unit root process, and they strongly depend on the set of regressors used. Moreover using the spectral theory of time series i.e. considering the process in the frequency domain, seems to be a better tool to decompose the process into a cycle and a trend. Many filtering procedures are available to accomplish the task of isolating the business cycle.

4 Estimation of the ARMA

We now deal with a stationary ARMA process y_t with zero mean. The theoretical behavior of ACF and PACF for an ARMA(p,q) process is the following:
AR(p)

$$\begin{cases} \rho(h) \to 0 & \text{for} \quad h \to \infty \\ \alpha(h) = 0 & \text{for} \quad h > p \end{cases}$$

[7] See below for the description of the Ljung-Box test that accomplishes this task.

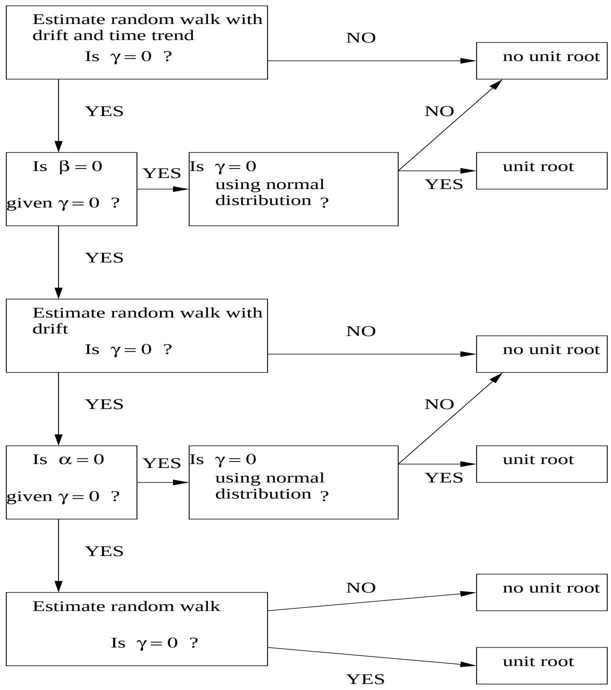

Figure 1: Dickey Fuller test

MA(q)

$$\begin{cases} \rho(h) = 0 & \text{for} \quad h > q \\ \alpha(h) \to 0 & \text{for} \quad h \to \infty \end{cases}$$

Hence from the plot of the sample ACF and PACF we can infer at least roughly the orders p and q.

4.1 Identification

The first step is to compute the sample ACF and PACF and to test their significance. Given a sample of T observations the sample ACF is given by:

$$\hat{\rho}(s) = \frac{\sum_{t=s+1}^{T}(y_t)(y_{t-s})}{\sum_{t=1}^{T}(y_t)^2}$$

Under the null hypothesis of an i.i.d. process this is normally distributed with mean zero and variance $Var(\hat{\rho}(s)) = Tw^2(s)$ where

$$w^2(1) = 1$$
$$w^2(s) = (1 + 2\sum_{j=1}^{s-1}\hat{\rho}^2(j)) \qquad \forall\ s > 1$$

Thus if we find values of ACF that satisfy $|\hat{\rho(s)}| < 1.96T^{-1/2}w(s)$ then we can reject the null hypothesis with a confidence level of 0.95 and we say that the autocorrelation at lag s is statistically not different from zero. We will therefore have at most an MA(s-1) process. For large samples however we have $w(s) = 1$, $\forall s$. A similar reasoning can be made in the case of an AR(p) model we can test the null hypothesis on the PACFs $\alpha(s) = 0 \; \forall s > p$ at confidence level of 0.95 if $|\alpha(s)| < 1.96T^{-1/2}$.

4.2 Estimation

Once p and q are specified we need to estimate the model:

$$A(L)y_t = B(L)u_t \quad \text{with} \;\; u_t \sim \text{white noise}(0, \sigma^2)$$

The parameters to be estimated are the (p+q)-vector of the ARMA coefficients ($\vec{\beta}$) and the variance of the white noise (σ^2). The algorithm the is used is maximum likelihood estimation (ML). First we define the transformed process $z_t = y_t/\sigma^2$ and its mean squared errors $r_t = E(z_{t+1} - \hat{z}_{t+1})^2$ where $\hat{z}_t$ is the forecast of the process at time t. Now computing the one step forecast errors within the time series we compute the gaussian likelihood function:

$$L(\vec{\beta}, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2) r_0 \ldots r_T}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{j=1}^{T} \frac{(y_j - \hat{y}_j)^2}{r_{j-1}} \right\}$$

the estimated parameters ($\hat{\vec{\beta}}$,$\hat{\sigma}^2$) will therefore maximize this function [8]. Many algorithms exist to maximize a function but we won't enter in any further detail. Once the values for q and p are given for different model specifications we can try to refine the analysis by introducing some criteria in order to have a parsimonious model as suggested originally by Box and Jenkins. We must avoid overfitting (i.e. the introduction of too many parameters) or we will loose power in prediction. Indeed the prediction errors (mean squared errors) depend not only on the variance of residuals (that decreases if more parameters are estimated), but also on the estimation errors of the parameters. We have two criteria to select among different models. The correct model will minimize [9] :

$$\begin{aligned} AIC(\hat{\vec{\beta}}) &= -2\ln L(\hat{\vec{\beta}}, S(\hat{\vec{\beta}})/T) + 2(p+q+1) \\ BIC(\hat{\sigma}^2) &= -2\ln L(\hat{\vec{\beta}}, S(\hat{\vec{\beta}})/T) + (p+q+1)\ln T \end{aligned}$$

respectively called Akaike Information Criterion and Bayesian Information Criterion.

[8] Although our time series is not gaussian it is possible to prove that the estimators are asymptotically the same.

[9] Define:

$$S(\hat{\vec{\beta}})) = \sum_{j=1}^{T} \frac{(y_j - \hat{y}_j)^2}{r_{j-1}}$$

4.3 Diagnostic checking

A final test can be done on the residuals of the ARMA estimation. If the specified model is correct they should be completely uncorrelated so that they don't contain any information about the process; hence they must be a white noise. Given the residuals ϵ_t we compute easily their sample ACF ($\hat{\rho}(k)$) and from that we can compute:

$$Q = T(T+2)\sum_{k=1}^{s}\frac{\hat{\rho}(k)^2}{(T-k)}$$

Under the null hypothesis of an i.i.d. process we have $Q \sim \chi^2_{s-p-q}$ hence rejecting this hypothesis implies that at least one value of ACF of the residuals from lag 1 to lag s is statistically different from zero; this is the Ljung Box Q-test [10]. Note that now we have $s-p-q$ degrees of freedom since their number is reduced by the estimated parameters.

5 Modeling the variance

Many economic time series exhibit periods of unusual volatility followed by relative tranquillity. Hence in such cases it's inappropriate to model the time series assuming that its variance is constant through time. This heteroschedasticity of a process is tipical of financial time series which are naturally characterized by high volatility. In such a context it is also clear that we are intereseted in conditional forecasts instead of long term forecasts; indeed the conditional forecast error is always less than the unconditional case. Engle (1982) showed how it's possible to model simultaneously the mean and the variance of a process.

5.1 ARCH

Let's consider an AR(1) process: $y_t = a_1 y_{t-1} + u_t$. Its conditional variance at time $t+1$ is just the forecast error variance:

$$Var(y_{t+1}|y_t) = E_t[(y_{t+1} - a_1 y_t)^2] = E_t(u^2_{t+1})$$

Until now we assumed that the residuals have a constant variance σ^2, but if this is not the case we can model it as something like an AR(r) process [11] where this time the white noise term $\nu_t \sim (0,1)$ is multiplicative instead of being additive and it's independent of u_{t-1}:

$$u_t = \nu_t\sqrt{\alpha_0 + \alpha_1 u^2_{t-1} + \ldots + \alpha_r u^2_{t-r}}$$

If there's homoschedasticity then $\alpha_i = 0, \forall\, i > 0$ and we're back to the usual case with $\alpha_0^2 = \sigma^2$. Otherwise we have an Autoregressive Conditional Heteroschedasticity model of order r for y_t

[10] Remember that i.i.d. $\Rightarrow$ white noise but not the converse.

[11] Practically we will use the estimated residuals $\hat{u}_t$ of the AR estimation of y_t.

(ARCH(r)). The residuals are uncorrelated and their unconditional mean and variance are unaffected. But they are not independent since their conditional variance is affected by ν_t:

$$E_{t-1}(u_t^2) = \alpha_0 + \sum_{i=1}^{r} \alpha_i u_{t-i}^2$$

It's clear that large realizations of u_t^2 imply large conditional variance at time $t+1$. To ensure stability of the AR process and to avoid negative conditional variance we need $\alpha_0 > 0$ and $0 < \alpha_i < 1$, $\forall i > 0$. Finally note that the error structure outlined will interact with the autocorrelation structure of y_t. Any shock in ν_t will create persistently large variance of u_t and the larger are the ACF values of y_t the longer will be the persistence on the variance of y_t.

5.2 GARCH

We can allow the conditional variance to be an ARMA process:

$$\begin{aligned} u_t &= \nu_t \sqrt{h_t} \\ h_t &= \alpha_0 + \sum_{i=1}^{r} \alpha_i u_{t-i}^2 + \sum_{i=1}^{s} \beta_i h_{t-i} \end{aligned}$$

The conditional variance of y_t is now given by h_t which is an ARMA(s,r) process; we speak in this case of Generalized Autoregressive Conditional Heteroschdasticity model of orders s and r (GARCH(s,r)). This representation is useful because for an high order ARCH exists a more parsimonious but equivalent GARCH representation that is easier to identify and estimate. In practice to identify such models we first fit y_t as an ARMA as explained above and we take the estimated residuals $\hat{u}_t$. The ACF of such residuals should indicate a white noise process but the ACF of $\hat{u}_t^2$ should indicate the GARCH order, indeed under this hypothesis they are an ARMA(s,r) process. Hence we can test heteroschedasticity qualitatively just by looking at ACF and PACF of the squared residuals and quantitatively by applying the Engle test. In the latter case we estimate the linear regression:

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 \hat{u}_{t-1}^2 + \ldots + \alpha_r \hat{u}_{t-r}^2$$

and we test the null hypothesis that $\alpha_i = 0 \ \forall i > 0$, this means no explantory power of the regression so that the residual sum of squares R^2 is very low. Given that $TR^2 \sim \chi_r^2$ we test the null hypothesis and its rejection means the existence of ARCH effects of order r [12]. Once a GARCH structure is specified it's possible to estimate its parameters jointly with the original ARMA just using ML algorithm.

[12] Given that an ARMA is always equivalent to a simple AR of higher order the presence of ARCH effects is equivalent to GARCH effects.

6 Application to real data

We apply all the previous procedures to real data. We considered two italian quarterly time series (both seasonally adjusted):

1. real GDP from 1970 to 1999 (figure 2);
2. long term government bond yields from 1957 to 1999 (figure 3);

The source of these data is International Monetary Fund. To reduce the variance we applied the logarithmic transformation to both series.

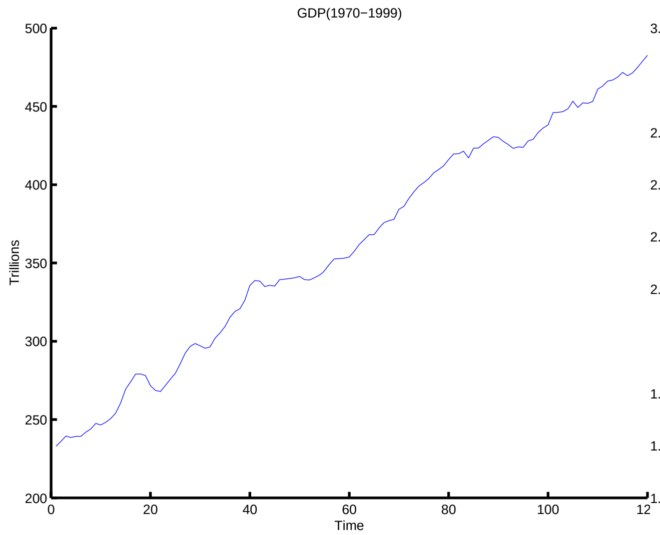

Figure 2: Italian real GDP

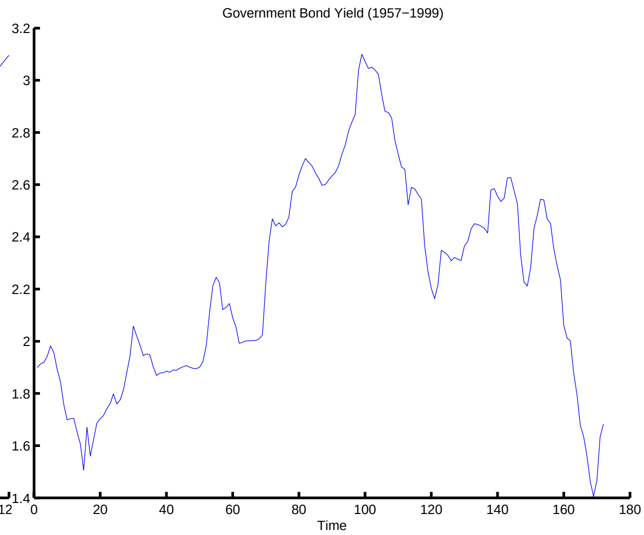

Figure 3: Government bond yields

6.1 Results for the GDP

6.1.1 Augmented Dickey Fuller tests

Following Nelson and Plosser we tested the following equation with 4 lags [13]:

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \sum_{i=1}^{4} a_i \Delta x_{t-i} + \epsilon_t$$

the coefficient of the fourth lag turns out to be non significative thus we reduce the number of lags and test again. This time all the coefficients of lags are significative and moreover we have a unit root, thus we can test for $\beta = 0$ given $\gamma = 0$. The result gives a non significant time trend in agreement with the idea that GDP is difference stationary. Following the procedure

[13] All the estimations were performed using MatLab functions either written on purpose or taken from the Garch Toolbox.

outlined in figure 1 we find that also the constant term is not significant. At the end we have the simple regression:

$$\Delta x_t = \gamma x_{t-1} + \sum_{i=1}^{3} a_i \Delta x_{t-i} + \epsilon_t$$

and we find that indeed $\gamma = 0$ (results in table 1). For this regression we computed also the Q test of whiteness of residuals at 10 lags and we obtain $Q = 9.75$ with a critical value of 18.31; hence the residuals are white. To obtain a stationary series we calculate the first differences $\Delta x_t \equiv y_t$. On y_t we test again for the presence of a unit root (after eliminating non significant lags) and we find that y_t is really stationary.

Regressors	t-Statistic	Critical Value
Test the significance of lag 4		
$x_{t-1}(\gamma)$	-1.95	
Δx_{t-1}	4.19	
Δx_{t-2}	2.14	
Δx_{t-3}	-2.24	
Δx_{t-4}*	-0.84	1.66
α	2.01	
βt	1.62	
Test the significance of lag 3 and the trend given a unit root		
$x_{t-1}(\gamma)$*	-2.08	-3.45
Δx_{t-1}	4.82	
Δx_{t-2}	2.02	
Δx_{t-3}	-2.78	1.66
α	2.13	
βt*	1.77	2.79
Test the significance of constant given a unit root		
$x_{t-1}(\gamma)$*	-2.02	-2.89
α*	2.22	2.54
Test for a unit root		
$x_{t-1}(\gamma)$*	4.21	-1.94
Test of stationarity of first differences		
$y_{t-1}(\gamma)$	-5.21	-1.94

Table 1: Augmented Dickey Fuller tests for GDP. * Means acceptance of the null hypothesis of zero coefficient. After the first two regressions we have 3 significative lags thus we don't report anymore their t-values, we just look for unit roots.

6.1.2 Box Jenkins estimation

First of all we subtract the mean from y_t; next we take a look at ACF and PACF with confidence intervals given by Bartlett's formula or given by the hypothesis of normal distribution of sample ACF that holds for large samples (figure 4). From this inspection we decide to try

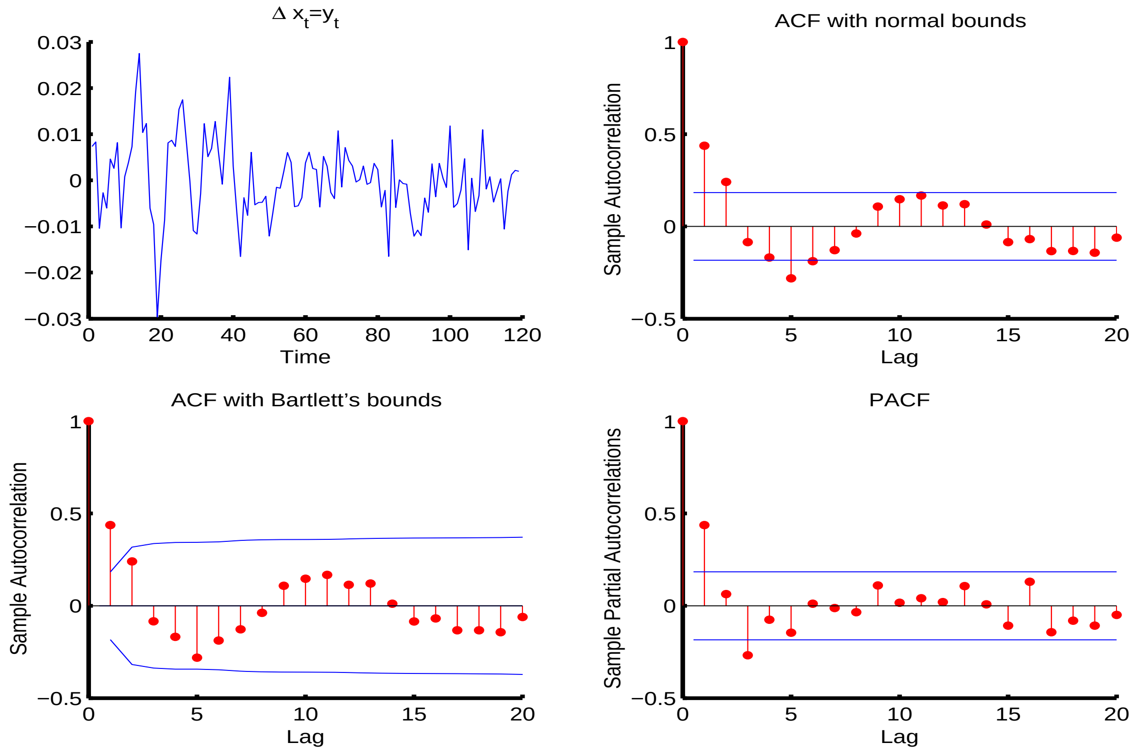

Figure 4: ACF and PACF

four different models estimated through ML algorithm. We report the stimated coefficients, Q values at 10 and 20 lags (critical values in parentheses) and Information Criteria, for different ARMA models. * Means residuals are not white i.e. Q test rejects the null hypothesis.

Model	Coefficient (std.error)	Q(20)	Q(10)	AIC	BIC
ARMA(1,1)					
AR(1) MA(1)	0.49 (0.19) -0.07 (0.22)	23.55 (31.41)	17.13* (15.51)	-818.65	-810.31
ARMA(1,2)					
AR(1) MA(1) MA(2)	0.22 (0.25) 0.20 (0.24) 0.27 (0.13)	15.56 (31.41)	9.36 (14.07)	-822.09	-810.97
AR(1)					
AR(1)	0.44 (0.08)	24.24 (30.14)	17.61* (16.92)	-820.38	-814.82
MA(2)					
MA(1) MA(2)	0.41 (0.08) 0.30 (0.10)	17.95 (28.87)	10.57 (15.51)	-823.43	-815.09

Given that ARMA(1,1) and AR(1) don't give white residuals we discard them and then according to the significance of the coefficients we choose MA(2) as a model for GDP. Moreover

for this model AIC and BIC are lower than for all the others.
We can now verify the forecast power of the model by computing the one step forecast error for the conditional mean in all the models. This is defined as:

$$MSE = E(y_{t+1} - y_t)^2$$

were the expectation is computed according to the estimated parameters of the model; we have:

MSE(ARMA(1,2))	$5.91\ 10^{-4}$
MSE(ARMA(1,1))	$8.30\ 10^{-4}$
MSE(AR(1))	$7.95\ 10^{-4}$
MSE(MA(2))	$4.04\ 10^{-4}$

We definitely prefer MA(2) and on this model we also perform an Engle test to check for homoschedasticity. The result gives $TR^2 = 1.34$ with a critical value of 3.84 thus we accept the null hypothesis of no ARCH effect. This is clear looking at figure (5). In this case the estimated unconditional variance of the residuals is $\hat{\sigma}^2 = 0.55\,10^{-4}$.

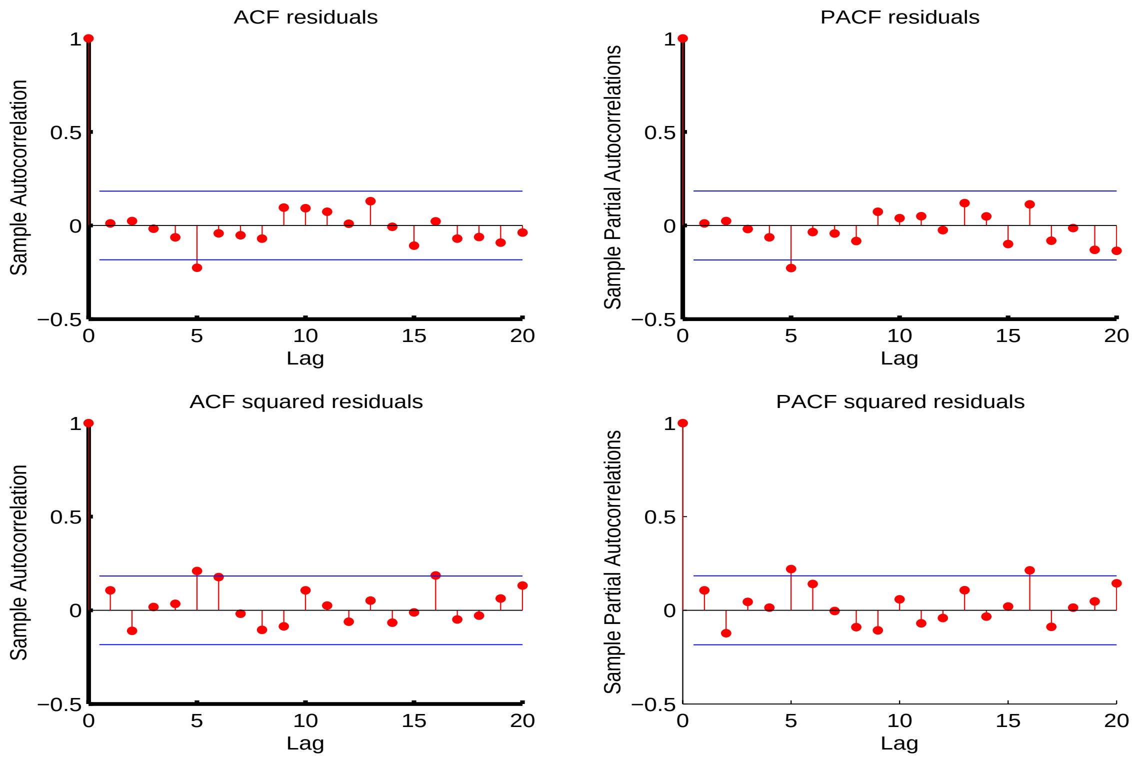

Figure 5: ACF and PACF for residuals and squared residuals - GARCH test

6.2 Results for the government bond yields

6.2.1 Augmented Dickey Fuller tests

By looking at figure (3) we don't consider any linear trend and for the rest we proceed as in the previous case. We first look for the significant number of lags then we check for significance of the constant term given the presence of a unit root. At the end we have the simple regression:

$$\Delta x_t = \gamma x_{t-1} + a_i \Delta x_{t-1} + \epsilon_t$$

and we find that indeed $\gamma = 0$ (results in brief in table 2). For this regression we computed also the Q test of whiteness of residuals at 10 lags and we obtain $Q = 9.17$ with a critical value of 18.31; hence the residuals are white. To obtain a stationary series we calculate the first differences $\Delta x_t \equiv y_t$. On y_t we test again for the presence of a unit root (after eliminating non significant lags) and we find that y_t is really stationary.

Regressors	t-Statistic	Critical Value
Test for a unit root		
$x_{t-1}(\gamma)^*$	-0.41	-1.94
Δx_{t-1}	6.04	
Test of stationarity of first differences		
$y_{t-1}(\gamma)$	-8.26	-1.94

Table 2: Augmented Dickey Fuller tests for government bond yields. * Means acceptance of the null hypothesis of zero coefficient i.e. of a unit root.

6.2.2 Box Jenkins estimation

Again we subtract the mean from y_t; next we take a look at ACF and PACF with confidence intervals given by Bartlett's formula or given by the hypothesis of normal distribution of sample ACF that holds for large samples (figure 6). From this inspection we decide to try three different models estimated through ML algorithm. We report the estimated coefficients, Q values at 10 lags and we performed an Engle test (for which we report the ARCHstat with critical values in parentheses) and Information Criteria, for different ARMA models. * Means residuals are not homoschedastic i.e. Engle test rejects the null hypothesis. ** Means rejection of Q test for whiteness of residuals.

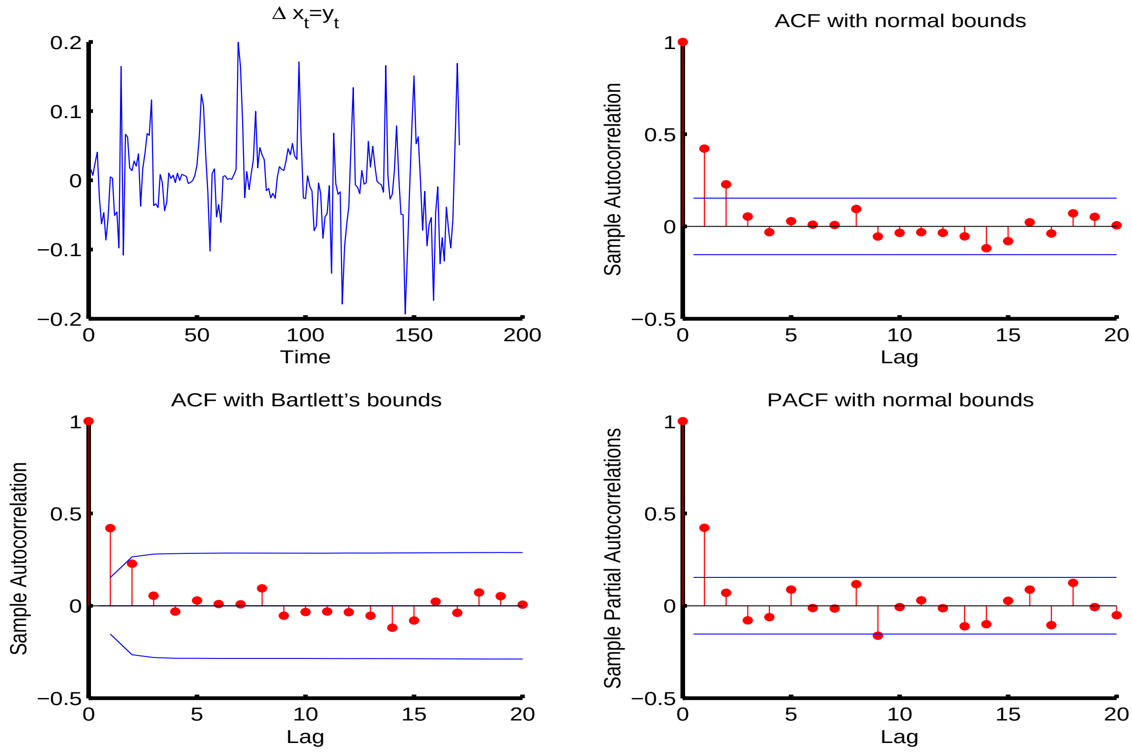

Figure 6: ACF and PACF

Model	Coefficient (std.error)	ARCHstat	Q(10)	AIC	BIC
ARMA(1,1)					
AR(1) MA(1)	0.52 (0.17) -0.12 (0.17)	4.29* (3.84)	9.52 (15.51)	-483.66	-474.24
ARMA(1,2)					
AR(1) MA(1) MA(2)	0.24 (0.31) 0.15 (0.31) 0.19 (0.13)	3.07 (3.84)	16.92** (14.07)	-483.31	-470.74
AR(1)					
AR(1)	0.42 (0.06)	5.08* (3.84)	10.39 (16.92)	-485.09	-478.80

According to significance of coefficients, AIC and BIC and Q test the chosen model is an AR(1). Government bond yields are a financial time series thus we expect ARCH effects and indeed this is the case. From figure 7 we can draw the same conclusion. Therefore we tested the joint AR(1) ARCH(1) model to have an estimate also of the conditional variance of residuals:

$$\begin{aligned} y_t &= a_1 y_{t-1} + u_t \\ u_t &= \nu_t \sqrt{\alpha_0 + \alpha_1 u_{t-1}^2} \end{aligned}$$

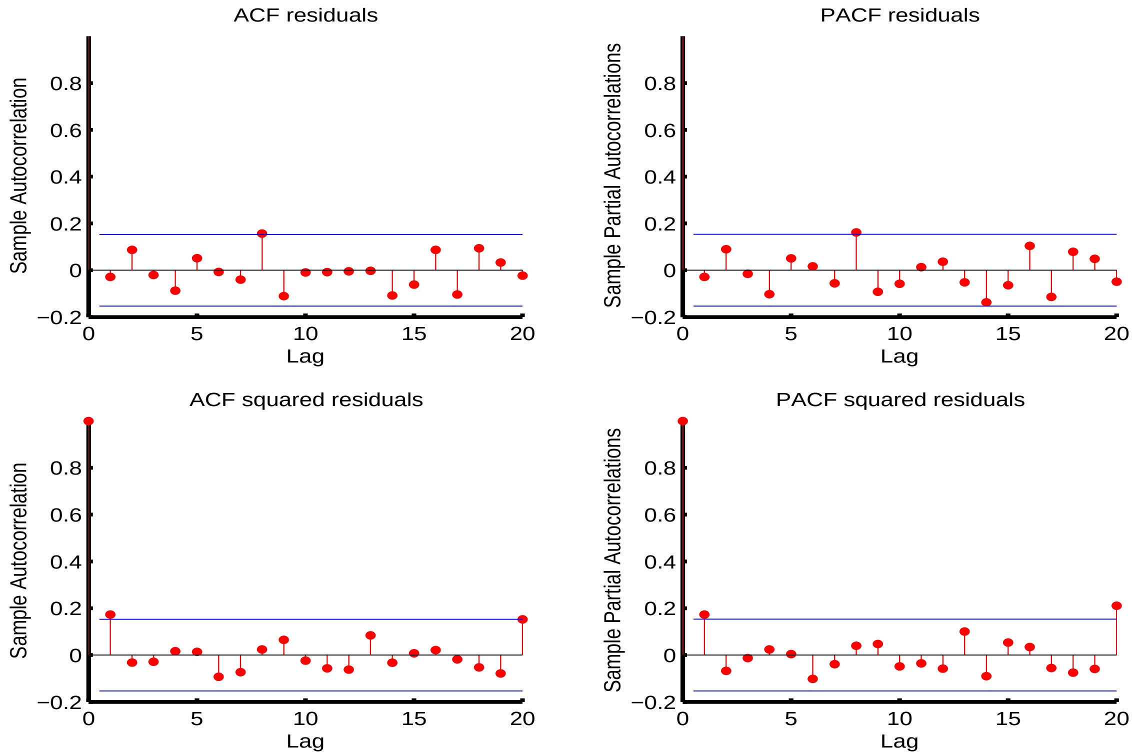

Figure 7: ACF and PACF for residuals and squared residuals - GARCH test

From ML we get:

	Coefficient (Std.Error)
a_1	0.4891 (0.0864)
α_0	0.0028 (0.0002)
α_1	0.1496 (0.0898)

Note that the long term estimated unconditional standard error of the residuals is given by the simple AR(1) estimation or by:

$$\hat{\sigma} = \sqrt{\frac{\alpha_0}{1-\alpha_1}} = 0.0578$$

The contitional standard deviation of the residuals have the following asymptotic behavior:

$$Var_{t+k}(u_t) \xrightarrow{k\to\infty} \hat{\sigma}^2$$

The following picture illustrates this case with present data. Note however the big difference between the two for one step forecasts of the standard error. Since the variance represents roughly the risk we see why in finance we are more interested in conditional rather than unconditional variance. Indeed with conditional variance we commit a smaller error in prediction than the one we would commit in taking in account only unconditional variance.

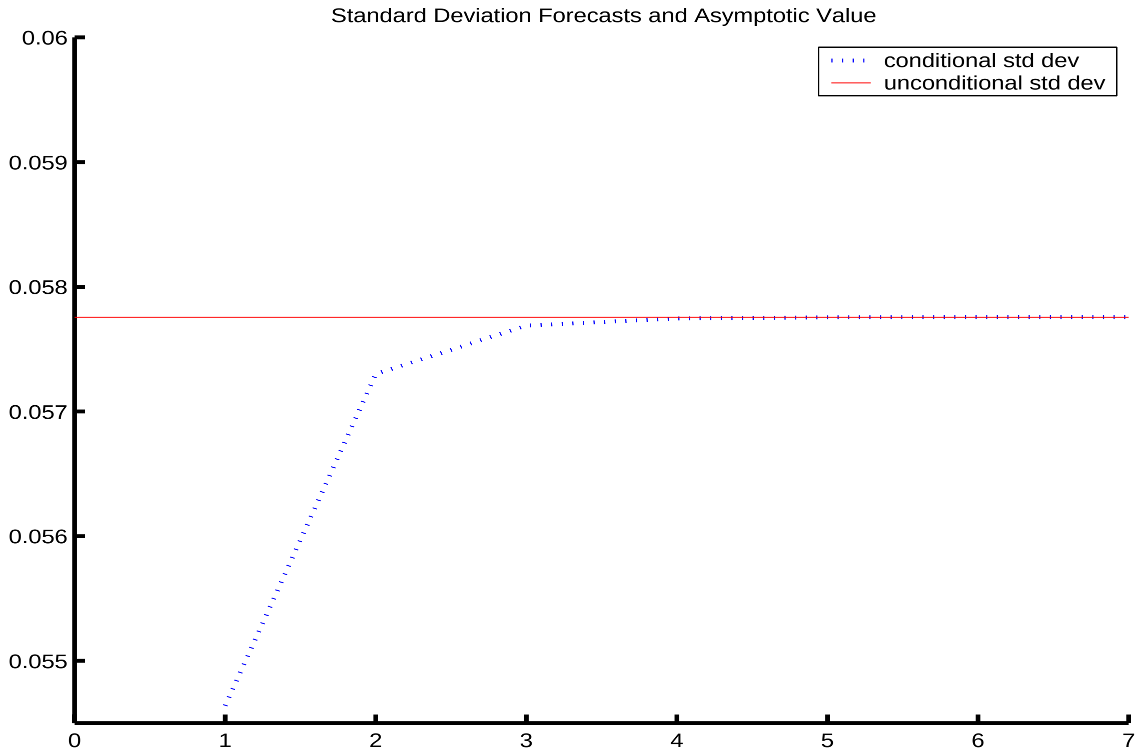

Figure 8: Conditional and unconditional standard error of the residuals

Bibliography

Beveridge, S. and Nelson, C. R. (1981) "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle"', *Journal of Monetary Economics* **7** 151-174.

Campbell, J. Y. and Perron, P. (1991) "Pitfalls and opportunities: what macroeconomists should know about unit roots", NBER Working Paper No. 100.

Dickey, D. A. and Fuller, W. A. (1979) "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association* **74** 427-431.

Dickey, D. A. and Fuller, W. A. (1981) "Likelihood ratio statistics for autoregressive time series with a unit root", *Econometrica* **49** 1057-1072.

Enders, W. (1995) *Applied Econometric Time Series*, New York, Wiley.

Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods*, New York, Springer-Verlag.

Box, G. and Jenkins, G. (1976) *Time Series Analysis, Forecasting and Control*, San Francisco, Holden Day.

Engle, R. F. (1982) "Autoregressive conditional heteroschedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* **50** 987-1007.