الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY
Faculté des sciences
Département de mathématiques

جامعة باجي مختار – عنابـــة

Course handout

For use: third year licence students in Applied Mathematics

Time series

Author : Djebar ahlem

Academic year : 2023/2024

Contents

Preface

In this course, I study some elements of time series and some linear models, in accordance with the specialty LICENCE programs "Applied Mathematics". I present time series modeling, discuss how to model and estimate discrete time random processes of a classical time series and the decomposition model with forecasting by exponential smoothing. In the second chapter, I am interested in modeling linear processes and two special cases of such processes: the AR processes, MA processes. I also study in the third chapter, linear processes ARMA with ARIMA and SARIMA processes. I give in a very detailed way the definitions and the characteristic functions of each process as well as a probabilistic study is made for each model, giving the necessary and sufficient conditions ensuring the stationarity, the causality and inversibility of these processes with simulations and correlograms, include the different estimation methods, in particular, the maximum likelihood method, Yule-Walker method and the least squares method. These estimation methods are used to estimate the parameters of the three models studied. As for the last chapter, it is devoted to the nonlinear models ARCH and its natural gen eralization : the class of processes GARCH. So in this chapter I first presented the definition and representation of GARCH models. Next, I establish the strict and second-order stationarity condition, from the first-order GARCH model, because the proofs are easier and the results are more explicit. Then, I study the problem of the estimation and the adjustment of a model of type (G)ARCH, where the estimation by maximum likelihood will play a preponderant role.

Chapter 1

Introduction

A time series is a series of observations ordered chronologically. They meet naturally in a wide variety of fields.

Examples include: the economy (unemployment rate, GNP, etc.), finance (action rate, interest rate, etc.), ecology (ozone pollution, CO, etc.), transport (with the famous example of international air traffic), demography, etc.

There are multiple study objectives. Forecasting is probably the most common goal. It is a question of predicting the future values of a variable thanks to the values observed in the present and the past of this same variable; the problem is therefore not the same as in regression where one seeks to predict the level of a variable (the answer) depending on the level of other variables (predictors). Other stated objectives of the time series study include the problem of estimating a trend. The systematic study of time series dates back to the end of the 2nd World War and has continued to intensify since (the computer revolution has even given a serious boost to everything that concerns practical applications).

The descriptive study of these series was not sufficient to predict their future behaviour. The need to prediction this type of series has led scientists to build mathematical models to meet this need.

This brief is divided into five chapters:

In the first chapter, we present general information on time series with presentation of the different methods of exponential smoothing namely simple exponential smoothing, double exponential smoothing and Holt-Winters exponential smoothing with an application of these R predictioning methods.

The second chapter will present the moving average process and autoregressive process.

ARMA process will be defined and discussed in the next chapter. Finally, the last chapeter : The ARCH model explains the future volatility based on current observables. The ARCH process describes the variability as a weighted average of previously estimated squared errors from historical data, and these weights provide more influence to the recent information and less to that of

the distant past. and the generalized autoregressive conditional heteroscedastic GARCH models.

Chapter 2

Descriptive analysis and discrete time random processes

2.1 Introduction

The theory of time series or chronoligical series discussed in this chapter is applied nowadays in fields as varied as econometrics, medicine or demography .... etc. We are interested in the evolution over time of a phenomenon, in order to write, explain and then predict this phenomenon in the future. We thus have observations at different dates, that is, a series of numerical values indicated by time.

2.2 Definition

A time series, or chronological series, is a finite sequence of numerical values representing the evolution of a specific quantity over time. Such sequences of random variables can be expressed mathematically in order to analyze their behavior, usually to understand their past evolution and to predict their future behavior.

A time series can be seen as a series of repeated observations of the same phenomenon on different dates (for example the average daily temperature in a given place, the average electricity consumption each month, the price of a barrel of oil each day...). The dates are often equidistant (daily, monthly, quarterly or annual series) except in a few cases (for example, daily economic data are not always available on non-working days).

A time series $(x_t)_{1<t<n}$ is usually represented by means of a graph with the dates on the x-axis and the observed values ordered.

Example

number of passengers on airlines between the years 1994 and 1960 is present in the R database : plot(AirPassengers) Figure1.

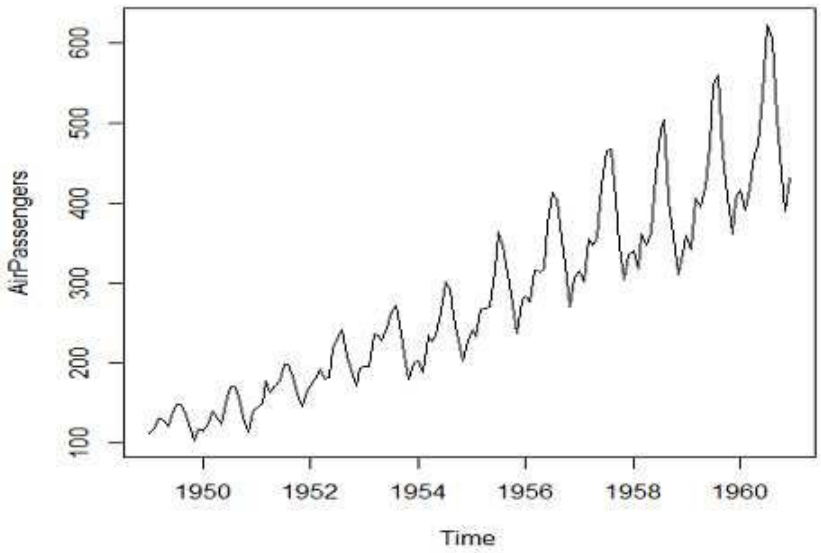


*Figure*1. *AIR* PASSEGERS

2.2.1 The Goal

The study of time series is used to make short, medium and long-term predictions. Quantitative and qualitative predictioning methods are available. Quantitative methods are divided into two categories. There are extrapolation methods that produce predictions on the principle of a correlation of the studied variable with time and explanatory methods that rely on correlations between the studied variable and different explanatory variables. The most widely used methods of the first category are the subject of the current predictioning method and are for the most part illustrated with examples taken from different sources.

2.2.2 Areas of application

Examples of univariate time series can be found in several areas:

1) Finance and econometrics: changes in stock market indices, prices, economic data, agricultural or industrial production.

2) Insurance: claims analysis.

3) Medicine / Biology: monitoring changes in pathologies, analysis of electroencephalograms and electrocardiograms.

4)Earth and Space Sciences: tidal indices, variations in physical phenomena (meteorology), evolution of sunspots, avalanche phenomena.

5) Signal processing: communication signals, radar, sonar, speech analysis.

6) Data processing: successive measurements of position or direction of a moving object (trajectography).

7) Metrology: phase or oscillator frequency variation.

8) Demography: population change.

2.3 Descriptive indices of a time series

It is useful to have some numerical measures that summarize a time series:

2.3.1 Central Trend Indices

We use as an indicator of the central trend the average:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2.3.2 Dispersion indices

We use as an indicator of dispersion the empirical variance :

$$\sigma^2 = \frac{1}{n}\sum_{t=1}^{n}(x_t - \overline{x})^2$$

2.3.3 Dependency indices

These notions, more specific to the study of time series, inform on the dependence between the data.

Auto-covariance: Empirical auto-covariance of order h provides information on the dependence between data that are excluded from h no time.

$$\sigma_h^2 = \frac{1}{n-h}\sum_{t=1}^{n-h}(x_t - \overline{x})(x_{t+h} - \overline{x})$$

For reasons of statistical common sense, we will consider empirical covariances only up to an order h not too large.

The function that at h associates σ_h^2 is called the auto-covariance function (empirical).

2.3.4 Auto-correlation

Empirical auto-correlations are the quotients of empirical covariances by empirical variance :

$$\rho_h = \frac{\sigma_h^2}{\sigma^2}$$

We note x_t the value observed at time t, the process $(x_t)_{1<t<n}$ is then called time series. The goal is to find a model that best represents the data, this is the problem of modeling. The objective is to detect the presence of a trend, a seasonality and to extract the signal structure in the data but on the main mission remains the prediction of future values of the series.

2.3.5 The 2 data locations over time

The data in a quarterly, monthly and daily time series are:

- identified using a single t-index that represents the number of months between the first observation and the observation of the data in question + 1, (the first observation being x_1)
- identified using two indices i and j, where i represents the year of observation, j its "month". "Month" can be a quarter, a month, a day.

Note indifferently the series (x_t) and $(x_{i,j})$.

2.3.6 The link between x_t and $x_{i,j}$

Example :

Let be a quarterly series. $x_{4,2} = 4370$ *so* $4370 = x_t$ *with* $t = (4-1)4+2 = 14$. Let be a monthly series. $x_{26} = 542$ *so* $542 = x_{i,j}$ *or* $26 = 2*12+2$ *so*, $i = 3, j = 2$.

If the data flow over n years, and each year contains p month then the observation of the month of the first year $x_{i,j}$ is also noted x_t with $t = (i-1)p + j$.

i takes the values 1, 2, 3, ... ,n.

j takes the values 1, 2, 3, ... ,p.

t takes the values 1, 2, 3, ... ,np

Note:

p = 4 for quarterly series, p = 12 for monthly series.

These 2 tracking modes give two types of data presentation and two different graphs.

2.4 Components of a Time Series

2.4.1 Trend and seasonality of a time series.

A time series can be broken down into three terms:

$$Y_t = T_t + S_t + \epsilon_t$$

where:

- T_t is a deterministic function that represents long-term variations called trend.

- S_t:is a periodic deterministic function called seasonality period T such as :

$$\sum_{i=1}^{T} S_{t+i} = 0, \forall t \in \mathbb{N}$$

It represents periodic period effects such as p that reproduce more or less identically from one period to another, it is noted by: $S_t, t = 1, \ldots\ldots\ldots, T$. It is generally assumed to be periodic $S_{t+p} = S_t$ of a period p .

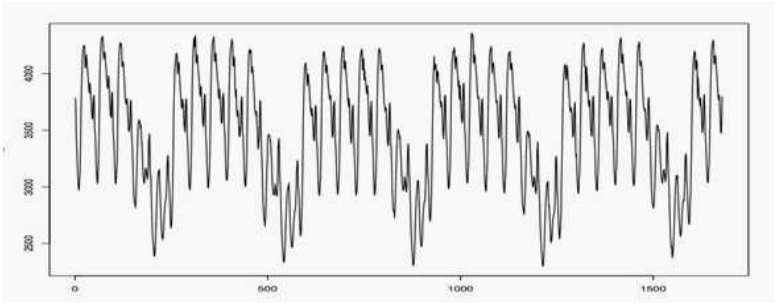


figure 2

- ϵ_t is a stationary random noise called residue.

Residues noted ϵ_t is the unstructured part of the phenomenon. It is modeled by a series of random variables ϵ_t , centered ,uncorrelated and the same variance . We speak of White Noise.

The goal is to learn how to model and estimate the different components and to make predictions on the future values of the initial time series. (the previous model is called additive, replacing + by * we will have the multiplicative model).

2.5 The composition models of these 3 components.

2.5.1 The additive model

In an additive model, it is assumed that the three components: trend, seasonal variations and accidental variations are independent of each other.

We consider that the x_t series is written as the sum of these 3 components: $x_t = T_t + S_t + \epsilon_t$

Graphically, the amplitude of variations is constant around the trend

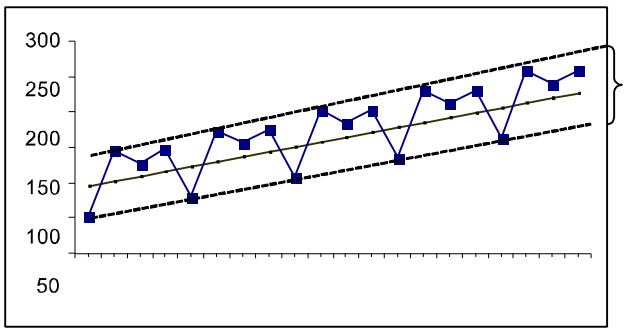


figure 3

2.5.2 The multiplicative model

1° multiplicative model form

Seasonal variations are assumed to depend on the trend. x_t is written as:

$$x_t = T_t * S_t + \epsilon_t$$

Graphically, the amplitude of (seasonal) variations varies.

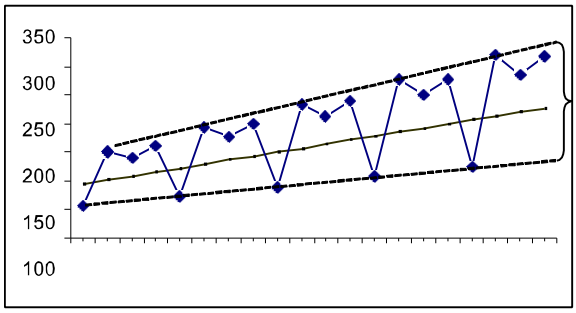


figure 4

2° multiplicative model form

It is assumed that seasonal and accidental variations depend on the trend. And we consider that x_t is written as follows:

$$x_t = T_t * S_t * \epsilon_t$$

Notes:

1- In the case of a series x_t with positive values, this second multiplicative model is reduced to an additive model considering the series $(ln(x_t))$:

$$ln(x_t) = ln(T_t) + ln(S_t) + ln(\epsilon_t).$$

2- The only difference between the 2 multiplicative models is in the estimation of t, which is not very important.

2.6 Trend estimate

2.6.1 Adjustment

Mayer Method: Straight Fit

We adjust the scatterplot $(t; x_t)$ by a line passing through two calculated points: We cut the series into 2 subsets of the same effective.

For each of the 2 subsets, the mean of t and the mean of x_t are calculated. This results in 2 points $(\overline{t_1}; \overline{x_1})$,$(\overline{t_2}; \overline{x_2})$, called mean points.

It remains to draw the right passing through these 2 points.

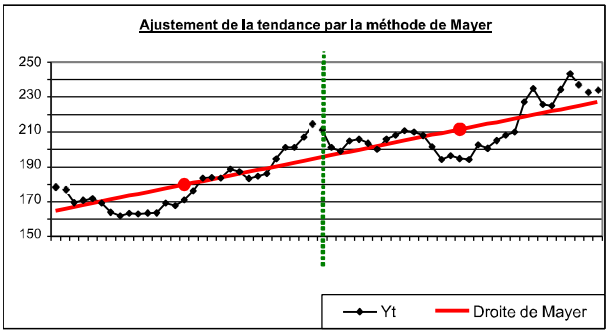


figure 5.Mayer Method

Note:

Median points can be calculated instead of mean points. This limits the influence of outliers.

Linear trend

The least squares method is used to adjust the time series x_t , with the function

$$T_t = at + b.$$

We determine the right of the least squares $(y = at + b)$ of the scatterplot $(t; x_t)$.

(That is, the right that minimizes the distance $\sum_t (x_t - (at + b))^2$)

So,

$$a = \frac{cov(t, x)}{v(t)} \quad \text{and} \quad b = \overline{x} - a\overline{t}$$

with

$$\begin{aligned} cov(t,x) &= \frac{1}{np}\sum_{t=1}^{np}(t-\overline{t})(x_t-\overline{x}) \\ v(t) &= \frac{1}{np}\sum_{t=1}^{np}(t-\overline{t})^2 \\ \overline{t} &= \frac{1}{np}\sum_{t=1}^{np}t_i \end{aligned}$$

Remark 1 *The least squares line adjusts best in the sense of least squares (this is the one that passes closest to the set of points), but it does not always model the trend well, this is the case for the 2nd series with an «aberrant» value.*

Polynomial trend

The least squares method can be used to adjust a trend in the form of a selected degree polynomial.

The observation of the graph of the series gives an idea of the degree of the polynomial (according to the shape of the curve)

A compromise must be made between :

- obtain residues that fluctuate around 0 with the lowest possible amplitude (this requires a high degree)
- use the lowest possible degree polynomial.

We choose the minimum degree of the polynomial that gives a correct adjustment: there is a degree from which we do not earn much by continuing to increase the degree.

Other trend: variable change

For some other trends, one can reduce to a linear or polynomial trend using a variable change.

Examples:

- Adjust $T_t = \frac{1}{at+b}$ is to adjust a linear trend on the series $Z_t = \frac{1}{x_t}$
- Adjusting $T_t = e^{at+b}$ is like adjusting a linear trend on the $Z_t = ln(x_t)$ series
- Adjusting $T_t = ln(at^2+bt+c)$ is like adjusting a trend at^2+bt+c on the series $Z_t = \exp(x_t)$

2.6.2 Nonparametric estimation

In some situations, it is not easy to find the degree of polynomail of adjustment for T_t. For example, in the following figure,

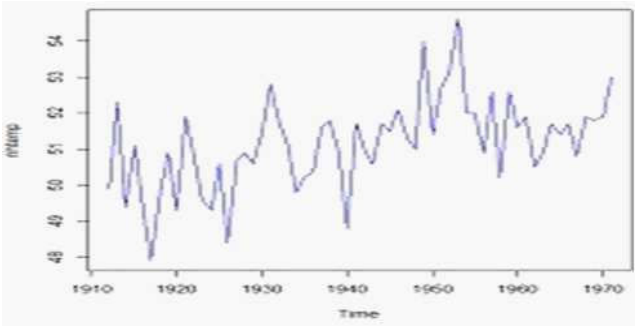


figure 6, verage Year ly Temperature in New Haven

it is not possible to use the least squares method because the polynomial used initially for mt is neither linear nor quadratic. One could use a polynomial with a high degree, but the number of parameters to be estimated would be important and would make calculations tedious. In this situation, the nonparametric theory of trend estimation is used, which does not assume that the trend is polynomial a priori. To understand this technique, suppose mt is linear in an interval [t - q, t + q]. In this case, a good trend estimator is given by

$$T_t = \frac{1}{2q+1} \sum_{k=-q}^{q} x_{t+k}$$

We can also see the interval [t - q, t + q] as a "window" on the observations, which we move when t varies. At each value of t, the estimator T_t calculates the average of observations falling into this sliding window (it is said that an estimate is made by moving average).

2.7 Calculation of seasonal variations

The trend T_t is estimated by adjustment or smoothing (moving average or median). We will now estimate seasonal variations S_t.

2.7.1 Calculation of seasonal coefficients S_j

Since it has been assumed that seasonal variations are repeated identically each year, a seasonal coefficient is estimated for each month, the seasonal variation for all months j will be the seasonal coefficient for month j. We consider the data without trend, we rank them by year (in line) and by month (in column). The

average of the data without trend for the month j of the n years is calculated, giving an initial estimate of the seasonal coefficient S_j. This is done for each month $j(j = 1, 2, \ldots, p)$.

Additive model case :

$$S_j = \frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - T_{i,j})$$

Case of the multiplicative model

$$S_j = \frac{1}{n}\sum_{i=1}^{n}\frac{x_{i,j}}{T_{i,j}}$$

Example 2 *1*

a) Decomposition with additive model and trend adjustment (least squares or Mayer method).

	Trimester 1	Trimester 2	Trimester 3	Trimester 4	
1990	-1685,15	502,28	1988,7	-656,87	
1991	-1473,45	855,98	1490,4	-730,17	
1992	-1335,75	700,68	1652,1	-756,47	
1993	-1537,04	830,38	1712,81	-954,77	
1994	-1507,34	713,08	1758,51	-879,07	
1995	-1625,64	828,78	1731,21	-953,37	**MEAN**
S_j	-1527,4	738,5	1722,3	-821,8	27,9
S_j'	**-1555,3**	**710,6**	**1694,4**	**-849,7**	**0,0**

where -1555,3= -1527,4-27,9

2.7.2 Seasonal coefficient correction

Because of the principle of conservation of areas: «the influence of seasonal variations on a year is zero», seasonal coefficients must be corrected.

For this we start by calculating the average of the seasonal coefficients :

$$\overline{S} = \frac{1}{p}\sum_{j=1}^{p} S_j$$

1. Case of additive model

The principle of conservation of areas means that the average of seasonal coefficients S must be zero.

So if S is not zero, we calculate the adjusted seasonal coefficients by subtracting from each of the S_j the mean $\overline{S}$:

$$S'_j = S_j - \overline{S}$$

2. Case of the multiplicative model

The principle of conservation of areas means that the average of the seasonal coefficients S must be equal to 1.

So if S is different from 1, we calculate the adjusted seasonal coefficients by dividing each of the Sj by the average $\overline{S}$:

$$S'_j = S_j / \overline{S}$$

2.8 Decomposition of a time series

2.8.1 Seasonally adjusted series or CVS series

Definition 3 *The seasonally adjusted or seasonally adjusted series referred to as the CVS series is the x_t time series from which seasonal variations have been removed.*

1) In the case of the additive model:
The seasonally adjusted series is

$$D_t = x_t - S_t$$

2) In the case of the multiplicative model:
The seasonally adjusted series is

$$D_t = x_t / S_t$$

Remark 4 *1) The peculiarity of the CVS series is that the data of D_t are directly comparable,*

2) From the CVS series, the trend can be re-evaluated by adjusting or smoothing (least squares or Mayer on D_t, or moving averages on $D_t \ldots$), in order to have a better estimate of the trend.

2.8.2 Adjusted series and accidental variations

Adjusted Series

The adjusted series, noted $\widehat{x_t}$, is obtained by recomposing the two estimated components: the trend and seasonal variations depending on the model chosen.

-In the case of the additive model:

$$\widehat{x_t} = T_t + S_t$$

-In the case of the multiplicative model :

$$\widehat{x_t} = T_t * S_t$$

Accidental or residual variations

- In the case of the additive model, the difference between the x_t series and its adjusted $\widehat{x_t}$ series represents accidental or residual variations : $\epsilon_t = x_t - \widehat{x}_t$.

- In the case of the multiplicative model (1st form), the difference between the x_t series and its series adjusted $\widehat{x_t}$ represents accidental or residual variations: $\epsilon_t = x_t - \widehat{x}_t$.

- In the case of the multiplicative model (2nd form) the relationship between the x_t series and its series adjusted $\widehat{x_t}$ represents accidental or residual variations:$\epsilon_t = x_t/\widehat{x}_t$

Decomposition of a time series

For a time series (x_t).

1. Plot the graph of (x_t) and the graph of the superimposed curves.
2. C_t trend is estimated and traced.
3. The composition model is chosen as additive or multiplicative.
4. Seasonal variations are estimated S_t.
5. Calculate and trace the CVS Dt series.
6. The adjusted x_t series is calculated.
7. Accidental variations are estimated t .

2.9 Forecasting by exponential smoothing

2.9.1 Introduction

Exponential smoothing is an empirical method of smoothing and predicting random time-series data. As in the moving average method, each data is smoothed successively starting from the initial value. Exponential smoothing gives past observations an exponentially decreasing weight with their seniority.

Three types of exponential smoothing are presented:

1. Simple exponential smoothing
2. Double Exponential Smoothing
3. Holt-Winters method

2.9.2 Simple exponential smoothing

We have $x_1; ...; x_n$ and we want to estimate x_{n+h} ($h \in \mathbb{N}^*$). For $\alpha \in]0; 1[$, the prediction is determined by simple exponential smoothing

$$\widehat{x}_{n,h} = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}$$

(a weighted average of past observations is done).

Remark 5 — *The smaller the α, the more importance is given to ancient observations. Note that $\widehat{x}_{n,h}$ does not depend on h.*

— *The sum of the weights does not make 1.*

— *The prediction does not depend on h.*

You can calculate by recurrence using the update formula:

$$\widehat{x}_{n,h} = \alpha x_n + (1-\alpha)\widehat{x}_{n-1,h}.$$

These formulas allow to calculate quickly (limiting complexity).

Lemma 6 *The prediction $\widehat{x}_{n,h}$ is (asymptotically) the solution of*

$$\widehat{x}_{n,h} = \underset{x}{argmin} \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - x)^2.$$

It's a least squares problem.

Proof. Either

$$f : x \mapsto \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - x)^2$$

We start by looking for critical points. We have:

$$f'(x) = -2 \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - x)$$

So f' is canceled in

$$x_0 = \frac{\sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}}{\sum_{j=0}^{n-1} (1-\alpha)^j} = \frac{\alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}}{1-(1-\alpha)^n} \sim \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j}.$$

A quick study of the variation table of f tells us that x_0 is the absolute minimum. It is because of the equivalent that we say that $\widehat{x}_{n,h}$ is the "asymptotic" solution to the square moindes problem. ■

Remark 7 *To choose α, we can use remark 2.1 or calculate, for any α, the estimators calculated with parameter $\widehat{x}_{n,h}(\alpha)$. We then look at the quadratic error*

$$E_2(\alpha) = \sum_{t=1}^{n-h} (x_{t+h} - \widehat{x}_{n,h}(\alpha))^2.$$

If this error is small, it is because the parameter α produces powerful predictions, given the data $x_1, .., x_n$. One can choose between several parameters $\alpha_1, .., \alpha_p$ by taking

$$\alpha = \underset{\alpha_1,..,\alpha_p}{\arg\min} E_2(\alpha_i)$$

Remark 8 *Note $x_1, .., x_n$ the temperature at noon on the walk of the English public (measured under cover) for n day. The series $x_1, .., x_n$ is supposed to take its values in integers. For each t in$\{1, 2, ..., n\}$, x_{t+1} can be compared with a prediction based on $x_1, .., x_t$. We are interested in two predictions: $\widehat{x}_{t+1}^{(1)}$ is the prediction of Météo France, $\widehat{x}_{t+1}^{(2)} = x_t$*

For i = 1, 2, we consider two errors:

$$\begin{aligned} E_1(i) &= \sum_{t=1}^{n-1} \mathbf{1}_{\widehat{x}_{t,1}^{(i)} = x_{t+1}}, \\ E_2(i) &= \sum_{t=1}^{n-1} (\widehat{x}_{t,1}^{(i)} - x_{t+1})^2. \end{aligned}$$

In general, $E_1(i) < E_1(i)$, so if you look at the E_1 criterion, you don't need a weather prediction.

Example 9 *Let's look at the example in Table 1. We have*

	SA	*SU*	*MO*	*TU*	*WE*	*TH*	*FR*
Weather Forecast FRANCE	*26*	*24*	*24*	*25*	*25*	*24*	*23*
measured temperature	*24*	*25*	*25*	*25*	*23*	*23*	*22*
city temperature	*21*	*24*	*25*	*25*	*25*	*23*	*23*

Table 1. Weather Forecast

$$E_1(1) = 6, E_1(2) = 5.$$

So for the E_1 error, the naive method is better than the Weather France prediction. Moreover

$$\begin{aligned} E_2(1) &= 4+1+1+0+4+1+1 = 12, \\ E_2(2) &= 9+1+0+0+4+0+1 = 15. \end{aligned}$$

So, for the E_2 error, the naive method is less good than the Weather France prediction.

2.9.3 Double Exponential Smoothing

We try to adjust, at the moment t, a line of equation :$y_t = a_1 + a_2(t-n)$. The double exponential smoothing prediction is as follows (where $\alpha \in]0;1[$)

$$\widehat{x}_{n,h} = \widehat{a}_1 + \widehat{a}_2 h,$$

where $(\widehat{a}_1, \widehat{a}_2)$ solution of

$$(\widehat{a}_1, \widehat{a}_2) = \arg\min_{(a_1,a_2)} \sum_{j=1}^{n-1} (1-\alpha)^j (x_{n-j} - (a_1 + a_2 j))^2.$$

Lemma 10 *The solutions to the above minimization problem are (asymptotically)*

$$\begin{cases} \widehat{a}_1 = -L_2(n) + 2L_1(n) \\ \widehat{a}_2 = \frac{\alpha}{\alpha-1}(-L_2(n) + L_1(n)), \end{cases}$$

with

$$\begin{cases} L_1(n) = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} \\ L_2(n) = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j L_1(n-j). \end{cases}$$

(We notice that we have here two successive exponential smoothings.)

Remark 11 *It is because L_2 is an exponential smoothing of L1 that we speak of double exponential smoothing.*

Proof. note that

$$C(a_1, a_2) = \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - (a_1 + a_2 j))^2.$$

We start by looking for the critical points of C. We have

$$\begin{aligned} \frac{\partial C}{\partial a_1}(a_1, a_2) &= -2 \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - (a_1 + a_2 j)), \\ \frac{\partial C}{\partial a_2}(a_1, a_2) &= -2 \sum_{j=0}^{n-1} (1-\alpha)^j j (x_{n-j} - (a_1 + a_2 j)). \end{aligned}$$

So we look for (a_1, a_2) system solutions

$$\begin{cases} -\sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} + \sum_{j=0}^{n-1} (1-\alpha)^j a_1 + \sum_{j=0}^{n-1} (1-\alpha)^j j a_2 = 0 \\ -\sum_{j=0}^{n-1} (1-\alpha)^j j x_{n-j} + \sum_{j=0}^{n-1} (1-\alpha)^j j a_1 + \sum_{j=0}^{n-1} (1-\alpha)^j j^2 a_2 = 0. \end{cases}$$

Let's recall the formulas

$$\begin{aligned} \sum_{j=0}^{+\infty} (1-\alpha)^j &= \frac{1}{\alpha}, \\ \sum_{j=0}^{+\infty} j(1-\alpha)^j &= \frac{1-\alpha}{\alpha^2}, \\ \sum_{j=0}^{+\infty} j^2 (1-\alpha)^j &= \frac{(1-\alpha)(2-\alpha)}{\alpha^3} \end{aligned}$$

(which can be easily found by manipulating whole series). Thus, by replacing certain partial sums of series by their limits,

$$\begin{cases} -\alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} + a_1 + \frac{1-\alpha}{\alpha} a_2 = 0 \\ -\alpha^2 \sum_{j=0}^{n-1} (1-\alpha)^j j x_{n-j} + (1-\alpha) a_1 + \frac{(1-\alpha)(2-\alpha)}{\alpha} a_2 = 0. \end{cases}$$

Beware, the previous systems are not equivalent. The larger the n, the more they appear. We will continue the calculation from the previous system and that is why we speak of «asymptotic» solution in the statement of the lemma. Dosage (for all n). we pose (for all n)

$$\begin{cases} L_1(n) = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j} \\ L_2(n) = \alpha \sum_{j=0}^{n-1} (1-\alpha)^j L_1(n-j). \end{cases}$$

we note that

$$\begin{aligned} L_2(n) &= \alpha \sum_{j=0}^{n-1} (1-\alpha)^j L_1(n-j) \\ &= \alpha^2 \sum_{j=0}^{n-1} (1-\alpha)^j \sum_{j=0}^{n-1} (1-\alpha)^j x_{n-j-i} \\ (k &= i+j) = \alpha^2 \sum_{k=0}^{n-1} x_{n-k}(1-\alpha)^k (k+1). \end{aligned}$$

The system then becomes

$$\left\{ \begin{array}{c} -\alpha L_1(n) + a_1 + \frac{1-\alpha}{\alpha} a_2 = 0 \\ -L_2(n) + \alpha L_1(n) + (1-\alpha)a_1 + \frac{(1-\alpha)(2-\alpha)}{\alpha} a_2 = 0. \end{array} \right.$$

and

$$\begin{aligned} -L_1(n)(2-\alpha) + a_1(2-\alpha) + L_2(n) - \alpha L_1(n) - (1-\alpha)a_1 &= 0 \\ -L_2(n) + 2L_1(n) &= a_1 \end{aligned}$$

$$\begin{aligned} a_2 &= \frac{\alpha}{1-\alpha}(L_1(n) - a_1) \\ &= \frac{\alpha}{1-\alpha}(L_2(n) - L_1(n)). \end{aligned}$$

Let us show that the function C is convex. For (a_1, a_2) and (b_1, b_2) in $\mathbb{R}^2$ and $\lambda \in [0; 1]$, we have

$$\begin{aligned} C(\lambda(a_1,a_2) + (1-\lambda)(b_1,b_2)) &= \sum_{j=0}^{n-1} (1-\alpha)^j (x_{n-j} - (\lambda a_1 + (1-\lambda)b_1 + (\lambda a_2 + (1-\lambda)b_2)j))^2 \\ &= \sum_{j=0}^{n-1} (1-\alpha)^j (\lambda(x_{n-j} - (a_1 + ja_2)) + (1-\lambda)(x_{n-j} - (b_1 + b_2 j))^2 \end{aligned}$$

(convexity of square function)

$$\begin{aligned} &\leq \sum_{j=0}^{n-1} (1-\alpha)^j (\lambda(x_{n-j} - (a_1 + ja_2))^2 + (1-\lambda)(x_{n-j} - (b_1 + b_2 j))^2) \\ &= \lambda C(a_1, a_2) + (1-\lambda) C(b_1, b_2). \end{aligned}$$

This shows that the C function is convex. The only critical point is therefore the absolute minimum. ■

The update forms are

$$\left\{ \begin{array}{c} \widehat{a}_1(n) = \widehat{a}_1(n-1) + \widehat{a}_2(n-1) + (2\alpha - \alpha^2)(x_n - x_{n-1,1}) \\ \widehat{a}_2(n) = \widehat{a}_2(n-1) + \alpha^2 (x_n - x_{n-1,1}), \end{array} \right. \tag{2.1}$$

with initialization

$$\begin{cases} \widehat{a}_1(0) = x_1 \\ \widehat{a}_2(0) = x_2 - x_1 \end{cases}$$

These formulas make it possible to understand how to calculate the coefficients recursively (which limits the complexity of the calculation).

2.9.4 Holt-Winters Method

Non-seasonal method

In the vicinity of moment n, we try to adjust a line of equation

$$y_t = a_1 + a_2(t - n).$$

The prediction in $n + h$ will be

$$\widehat{x}_{n,h} = \widehat{a}_1(n) + \widehat{a}_2(n)h.$$

Two smoothing constants β and α are chosen in $]0;1[$. The $\widehat{a}_1$, $\widehat{a}_2$ are calculated recursively by

$$\begin{cases} \widehat{a}_1(n) = \alpha x_n + (1-\alpha)(\widehat{a}_1(n-1) + \widehat{a}_2(n-1)) \\ \widehat{a}_2(n) = \beta(\widehat{a}_1(n) - \widehat{a}_1(n-1)) + (1-\beta)\widehat{a}_2(n-1). \end{cases}$$

Remark 12 *This method is more flexible than the previous method. The α plays a role in the n ordinate estimate and theβ plays a role in the slope estimate. The smaller α and β are, the more the distant past is taken into account.*

Lemma 13 *To obtain a double exponential smoothing with $1 - \alpha'$ parameter (which has no dedicated instruction in R), simply do a Holt-Winters smoothing without seasonal component with*

$$\begin{cases} \alpha = 1 - (\alpha')^2 \\ \beta = \frac{1-\alpha'}{1+\alpha'}. \end{cases}$$

We find the simple exponential smoothing of parameter α if β=0 (and $\alpha \neq 0$).

Proof. For a double exponential smoothing of parameter 1-α' in $]0;\ 1[$, we can write the update formulas (2.1) in the form (we note beforehand that $(1-\alpha')(2-(1-\alpha')) = 1-(\alpha')^2$,

$$\begin{aligned}
\widehat{a}_1(n) &= (\alpha')^2(\widehat{a}_1(n-1) + \widehat{a}_2(n-1)) + (1-(\alpha')^2)(x_n - \widehat{x}_{n-1,1} + \widehat{a}_1(n-1) + \widehat{a}_2(n-1)) \\
&= (\alpha')^2(\widehat{a}_1(n-1) + \widehat{a}_2(n-1)) + (1-(\alpha')^2)x_n \\
\widehat{a}_2(n) &= \widehat{a}_2(n-1) + \frac{(1-\alpha')^2}{1-(\alpha')^2}(\widehat{a}_1(n) - (\widehat{a}_1(n-1) + \widehat{a}_2(n-1)) \\
&= \widehat{a}_2(n-1) + \frac{1-\alpha'}{1+\alpha'}\widehat{a}_1(n) + (\widehat{a}_1(n-1) + \widehat{a}_2(n-1))\frac{(1-\alpha')}{1+\alpha'} \\
&= (1-\beta)\widehat{a}_2(n-1) + \beta(\widehat{a}_1(n) - \widehat{a}_1(n-1))
\end{aligned}$$

We find the update formula (2.4) with $\alpha = 1-(\alpha')^2, \beta = \frac{1-\alpha'}{1+\alpha'}$ ■

Additive seasonal method

We search, in the vicinity of n, to adjust an equation graph

$$y_t = a_1 + a_2(t-n) + s_t$$

(with s. periodic, period T). Choose α, β, δ in]0; 1[(and also T). Recursion formulas are :

$$\begin{cases} \widehat{a}_1(n) = \alpha(x_n - \widehat{s}_{n-T}) + (1-\alpha)(\widehat{a}_1(n-1) + \widehat{a}_2(n-1)) \\ \widehat{a}_2(n) = \beta(\widehat{a}_1(n) - \widehat{a}_1(n-1)) + (1-\beta)\widehat{a}_2(n-1)) \\ \widehat{s}_n = \delta(x_n - \widehat{a}_1(n)) + (1-\delta)\widehat{s}_{n-T}. \end{cases}$$

The prediction takes the form

$$\begin{cases} \widehat{x}_{n,h} = \widehat{a}_1(n) + h\widehat{a}_2(n) + \widehat{s}_{n+h-T}, & 1 \leq h \leq T \\ \widehat{x}_{n,h} = \widehat{a}_1(n) + h\widehat{a}_2(n) + \widehat{s}_{n+h-2T}, & \text{T+1} \leq h \leq 2T \\ .... \end{cases}$$

The following initial values allow to start the calculation at T + 2

$$\begin{cases} \widehat{a}_1(T+1) = x_{T+1} \\ \widehat{a}_2(T+1) = \frac{x_{T+1} - x_1}{T} \\ \widehat{s}_j = x_j - (x_1 + (T-1)\widehat{a}_2(T+1)), & 1 \leq j \leq T \end{cases}$$

Seasonal multiplicative method

In the vicinity of n, we try to adjust an equation curve

$$y_t = (a_1 + a_2(t-n)) * s_t$$

(with s. periodic, period T). Choose α, β, δ in]0; 1[(and also T). Recursion formulas are :

$$\begin{aligned} \widehat{a}_1(n) &= (1-\alpha)\frac{x_n}{\widehat{s}_{n-T}} + \alpha(\widehat{a}_1(n-1) + \widehat{a}_2(n-1)) \\ \widehat{a}_2(n) &= (1-\beta)(\widehat{a}_1(n) - \widehat{a}_1(n-1)) + \beta\widehat{a}_2(n-1) \\ \widehat{s}_n &= (1-\delta)\frac{x_n}{\widehat{a}_1(n)} + \delta\widehat{s}_{n-T}. \end{aligned}$$

$$\begin{cases} \widehat{x}_{n,h} = (\widehat{a}_1(n) + h\widehat{a}_1(n)) * \widehat{s}_{n+h-T}, & 1 \leq h \leq T \\ \widehat{x}_{n,h} = (\widehat{a}_1(n) + h\widehat{a}_2(n)) * \widehat{s}_{n+h-2T}, & 1\text{+T} \leq h \leq 2T \\ ... \end{cases}$$

Initialization: it's very hard but we can just put the first $\widehat{a}_2$to 0 and the first $\widehat{a}_1$ to 1 (and the system fixes later).

Remark 14 *The two methods above (additive seasonal and multiplicative seasonal) can only be implemented if we have a number of periods to estimate the periodic part. If the number of data is insufficient, R returns an error message. The seasonal additive or seasonal multiplicative method is chosen according to the graphic description of the data. Basically: constant amplitudes additive seasonal method, non constant amplitudes nultiplicative seasonal method.*

Exercise 15 *Either the turnover achieved by the company Mail*

t1	*54*
t2	*65*
t3	*55*
t4	*20*
t5	*48*
t6	*85*
t7	*95*
t8	*74*
t9	*58*
t10	*62*
t11	*31*
t12	*74*

1. calculate the moving averages : $MM_3(X_t)$, $MM_4(X_t)$and $MM_5(X_t)$.

2. Plot the original series and the two series generated by the moving averages.

Exercise 16 *Let the time series of data recorded semi-annually*

X_t
5
6
2
4
8
5
2
6

1.Plot the series graph.

2.Assuming a linear trend additive model, estimate the trend and seasonal coeffcients.

3.Forecast the series values for the following year.

4.Make a forecast by LES with α =0:7:

Chapter 3

Random processes with discrete times

3.1 Introduction

In statistics, by time series the modeling of a sequence of random and sequentially observed events, usually on a temporal scale. Many real data streams admit such a representation, we can think for example of electricity consumption, the evolution of a stock exchange, the gross domestic product, the tidal cycle or the progression of a chemical process. Although astronomers and meteorologists may claim authorship, the main advances in the formalization of time series seem to lie with statisticians and econometers. Today we find its legacy in the fields of economics and finance of course, but also in engineering, signal processing or the natural and social sciences. It was from the publication of the pioneering works of Box and Jenkins in 1970 that the theory of time series as we know it today was born.

The latter are therefore all the more apt to reflect real flows and natural events within which chronological evolution is manifest. Among the panel of applications from the theory, it is mainly modeling and prediction that focus all attention.

While we seek on the one hand to build the model best suited to a dataset without prejudging events occurring outside the study interval, we focus on the other on the optimal learning of observations to infer future behavior.

Prediction of future values of stochastic processes by models: autoregressive, moving average, autoregressive moving average. These models may include certain stationarity, inversibility, estimation, etc. properties that are analysed and modelled by R software. R is a programming language for data analysis and modeling. R can be used as an object oriented language as well as a statistical environment in which lists of instructions can be executed in sequence without user intervention.

Then the objective of this chapter is the modeling study of stationary series and R. The organization of the chapter is as follows:

In the first section, devoted to the presentation of linear stochastic processes, namely, stationary linear models in which the autoregressive model AR is found, the MA moving average model and a general presentation on statistical programming software or language R.

In the second section represents the study of ARMA moving average autoregressive process.

The third section is devoted to the presentation of AR autoregressive model, MA moving average, ARMA moving average autoregressive by R software.

3.2 Stationary processes

3.2.1 Definitions

Definition 17 *A sequence $\{X_1, ..., X_T\}$ of random variables is called A discrete time process.*

Definition 18 *Let $X = \{X_t\}_{t\in\mathbb{Z}}$ be a process such as $E[X^2] < \infty$ for $t \in \mathbb{Z}$ It will be said that this process is (weakly) stationary (or stationary of order 2) if the following 2 conditions are checked for any t :*

- $E(X_t) = m$ *for all* $t \in \{1,...,T\}$
- $E(X_t X_{t+h}) = \gamma(h)$ *is independent of* t.

3.2.2 Some classic examples of processes

White noise

Definition 19 *The ϵ_t process is called a "white noise" if it forms a sequence of independent and identically distributed variables. It will be said centered if $E[\epsilon_t] = 0$ and reduced if $Var[\epsilon_t] = 1$.*

Proposition 20 *Let X_t be a white noise checking $E(X_t) = m$ and $Var[X_t] = \sigma^2$. Then X_t is a stationary process.*

The first condition simply means that the expectation of the process is independent of time. The second condition involves of course the independence of the autocovariance function from time (stationarity). But it also implies that the autocovariance terms (for h $\geq$0) are all zero. Only the variance is nonzero. That is, it means that white noises are particular stationary processes without "memory". The level of the series considered today has no impact on its level of tomorrow, just as the level of yesterday has no impact on the level of today.

Gaussian Processes

In statistics, Gaussian vectors and processes play a particular role because they are easy to manipulate. In particular, they are completely characterized by their two-order structure. It is therefore easy to estimate and model their statistical properties.

Definition 21 *The $(X_n)_{n\in\mathbb{Z}}$ process is a Gaussian process if all its marginal laws are Gaussian. That is to say if whatever is k and $j_1, ..., j_k$, the vector $(X_{j1}, ..., X_{jk})$ is a Gaussian vector.*

It is clear that a stationary Gaussian porcessus is strictly stationary.

Exercise 22 *Let $(X_n)_{n\in\mathbb{Z}}$ be a sequence of independent, centered, and all Gaussian random variables of variance 1. Which of the following processes are Gaussian? Which ones are stationary? Then give the expression of their covariance function.*

1. $Y_n = X_n X_{n+1}$
2. $Yn = X_n X_{n+1} ... X_{n+k}$
3. $Yn = X_n^2 X_{n+1}^2$
4. $Yn = X_n^2 + X_{n+1}^2$
5. $Yn = n + X_n$

3.2.3 Moving average process

Definition 23 *It will be said that the $\{Z_t\}$ process follows a "moving average of order q" (MA(q)) model if there is a centered white noise $\{\epsilon_t\}$ and constants β_k, $k = 0,...,q$ such as*

$$Z_t = \sum_{k=0}^{q} \beta_k \epsilon_{t-k}$$

Proposition 24 *Let X_t be a process following a moving average model, with ϵ_t a centered white noise checking $E[\epsilon_t] = \sigma^2 < \infty$, then the X_t process is stationary. Its autocovariance function is given by*

$$\sigma_h^2 = \sum_{k=0}^{q-|k|} \beta_k^2 \sigma^2 \qquad si \ \ |h| \in 0, ..., q \ \ \ and \ \ \ 0 \ otherwise$$

Theorem 25 *Wold decomposition*

Any stationary process of order two $(X_n)_{n\in\mathbb{Z}}$ can be represented as:

$$X_t = \sum_{k=0}^{\infty} \beta_k \epsilon_k + \kappa_t$$

where the β_k parameters satisfy $\beta_0 = 1$, $\beta_k \in \mathbb{R}$ or $k > 0$, $\sum_{k=1}^{\infty} \beta_k^2 < \infty$ and where ϵ_k is a white noise. It is said that the sum of past shocks corresponds to the stochastic linear component of X_t. The term κ_t refers to the deterministic linear component such as $Cov(\kappa_t, \epsilon_{t-k}) = 0$ for all $k \in \mathbb{Z}$.

3.2.4 Autoregressive process

Definition 26 *It will be said that the X_t process follows an autoregressive model of order p (AR(p)) if there is a reduced centered white noise $\{\epsilon_t\}$, such that ϵ_t is independent of X_0,... , X_{t-1} and constants $\alpha_1, ..., \alpha_p$ and σ such as for t $\in \{1, ..., T\}$ one has*

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + ... + \alpha_p(X_{t-p} - \mu) + \sigma\epsilon_t$$

3.2.5 Partial auto-correlation

Partial auto-correlation is an important notion for time series, but it is a bit tricky and requires "reminders" of geometry in a random variable space.

Reminders

The real random variables of integrable square defined on the same probability space (Ω, A, P) form a vector space that can be provided with the scalar product $\langle X, Y\rangle = E[XY]$. We note this space $L^2 = L^2(\Omega, A, P)$.

- Orthogonality. Two variables X and Y of L^2 have orthogonal when

$$E[XY] = 0.$$

- Distance. The distance between two variables is defined as

$$d(X, Y) = \sqrt{E[(X - Y)^2]}$$

- Generated space. We consider N random variables $X_1, ..., X_N$. The subspace generated by these variables, which will be noted with $vec(X_1, ..., X_N)$ is the subset of all linear conbinations of these variables.
- Projection : The projection of a random variable Y on the $vec(X_1, ..., X_N)$ will be noted

$$P_{vec(X_1,...,X_N)}(Y).$$

It is the random variable of this space that is closest to Y in the sense of the distance L^2 seen above. It is therefore the linear combination $P_{vec(X_1,...,X_N)}(Y) = \gamma_1 X_1 + ... + \gamma_N X_N$ such that

$$E[(\gamma_1 X_1 + ... + \gamma_N X_N)^2] \leq E[(Y - \lambda_1 X_1 - ... - \lambda_N X_N)^2] \quad \forall \lambda_1, ..., \lambda_N.$$

The best linear explanation of the variable Y in terms of $X_1, ..., X_N$.
To calculate the projected, the following property is used

Proposition 27 *The projected* $P_{vec(X_1,...,X_N)}(Y) = \gamma_1 X_1 + ... + \gamma_N X_N$ *is characterized by* Y *-*$P_{vec(X_1,...,X_N)}(Y)$*being orthogonal to variables* $X_1, ..., X_N$.

this means that

$$\forall k = 1, ..., N, \quad E[(Y - \lambda_1 X_1 - ... - \lambda_N X_N) X_k] = 0$$

is still

$$\gamma_1 E(X_1 X_k) + ... + \gamma_N E(X_N X_k) = E(Y X_k)$$

What is also written in the form of a system of equations

$$\begin{pmatrix} E(X_1^2) & ... & E(X_1 X_N) \\ ... & \ddots & .. \\ E(X_1 X_N) & ... & E(X_N^2) \end{pmatrix} \begin{pmatrix} \gamma_1 \\ ... \\ \gamma_N \end{pmatrix} = \begin{pmatrix} E(X_1 Y) \\ ... \\ E(X_N Y) \end{pmatrix}$$

If the matrix is invertible, the γ is easily obtained.

Partial correlation coefficient between two variables

We give ourselves N $\geq$3 random variables $X_1, ..., X_N$ and we will define a correlation coefficient between X_1 and X_N that ignores the influence of $X_2, ..., X_{N-1}$.

This notion is an answer to the following question: it happens that 2 phenomena are strongly correlated, but that this correlation is due to the influence of an external factor and not to a strong link between the two phenomena.

Definition 28 *The partial correlation coefficient in* X_1 *and* X_N *apart from the influence of* $X_2, ..., X_{N-1}$ *is the correlation coefficient between the two variables to which their best explanations in terms of* $X_2, ..., X_{N-1}$ *have been decided, either*

$$r_{X_2,...,X_{N-1}}(X_1, X_N) = \rho(X_1 - P_{X_2,...,X_{N-1}}(X_1), X_N - P_{X_2,...,X_{N-1}}(X_N))$$

Partial autocorrelation of a stationary process

We now consider a centered stationary process (..., X_{-1}, X_0, X_1,...).

Definition 29 *Partial auto-correlation* $r(h)$ *for* $h \neq 0$ *is defined as follows.*

$$\begin{aligned} r(1) &= \rho(1) \\ r(h) &= r_{X_2,...,X_{N-1}}(X_1, X_{N+1}) \qquad \forall h \geq 2 \\ r(h) &= r(-h) \qquad \forall h \neq 0 \end{aligned}$$

3.3 Autoregressive process

Definition 30 *The model defined by the equation*

Theorem 31 ***Definition 32***

$$\begin{aligned} X_t &= \varsigma + \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \sigma^2 \epsilon_t \\ X_0 &= 0 \end{aligned}$$

for $t \geq p$ has a stationary solution if and only if the (complex) roots of its characteristic polynomial

$$\Phi(z) = 1 - (\alpha_1 z + ... + \alpha_p z^p)$$

are of module strictly greater than 1.

Remark 33 *order model $p = 1$: we find the condition of the previous theorem.*

It is easy to show that the process is stationary by writing its mean and covariance.

Exercise 34 *Give the conditions of stationarity of an AR(2) process.*

Solution 35 *A process AR(2) :*

$$X_t = \varsigma + \alpha_1 X_{t-1} + .\alpha_2 X_{t-2} + \epsilon_t$$

the caracteristic polynom is given by :

$$(1 - \alpha_1 L - \alpha_2 L^2 2) X_t = \epsilon_t$$

the resolution of the equation leads to the following conditions :

$$\begin{aligned} 1)\alpha_1 + \alpha_2 &< 1 \\ 2)\alpha_2 - \alpha_1 &< 1 \\ 3\ ) \quad |\alpha_2| &< 1 \end{aligned}$$

Proposition 36 *The average of the AR(p) process is :*

$$E[X_t] = \frac{\varsigma}{1 - (\alpha_1 + ... + \alpha_p)}$$

The variance is :

$$\gamma_0 = \alpha_1 \gamma_1 + ... + \alpha_p \gamma_p + \sigma^2$$

and the autocovariance is $\gamma_h = \alpha_1 \gamma_{h-1} + ... + \alpha_p \gamma_{h-p}$.

The characteristic polynomial of this recurrence relationship is $z^p - \alpha_1 z^{p-1} - ... - \alpha_p = z^p \Phi(z)$. The λ_i roots of this polynomial are the inverses of the roots of the polynomial Φ, and are therefore of module strictly less than 1. Since $|\lambda_i| < 1$, autocovariance decreases exponentially with h. These results extend directly to the autocorrelation of an $AR(p)$ process.

Causality and invertibility

Definition 37 *An AR(p) process is said to be causal when there is a sequence of numbers ψ_k such as:$\sum_k |\psi_k| < \infty$ and*

$$X_t = \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}$$

Proposition 38 *The process AR(p) is causal and stationary if and only if the polynomial in z: $\Phi(z) = 1 - \sum_i^{\infty} \alpha_i z_i$ is such that*

$$\Phi(z) \neq 0 \quad for\ all\ z \in \mathbb{C} \quad such \quad as \ |z| \leq 1.$$

Inversibility By definition, we have directly that any autoregressive process is automatically inversible. The equation

$$\epsilon_t = \sum_{i=0}^{\infty} \alpha_i X_{t-i} \qquad \text{for all t} \in \mathbb{Z}$$

is sometimes referred to as "AR(∞)" representation of the process. This is one of the main motivations to study it: A linear process that is inversible allows such an AR(∞) representation with rapidly increasing α_i coefficients. So, there is a good approximation by an $AR(p)$ process, with a rather small p-order. This will make it possible to treat the given linear process in the simple and efficient framework of estimation and predication of autoregressive processes of order p.

3.3.1 Statistical inference for autoregressive models

Estimation of parameters

An observation $x_0, ..., x_T$ of length $T+1$ of a stationary process X_t is available supposed to follow an AR(p) model, ie checking

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + ... + \alpha_p X_{t-p} + \sigma \epsilon_t$$

with $t \in \mathbb{Z}$ and $X_0 = x_0$ and $\alpha_1, ..., \alpha_p$ and σ unknown parameters. We then try to estimate these parameters using the available observations.

For this different methods are possible

- Least squares
- Maximum likelihood. The estimation of an $AR(P)$ model by the maximum likelihood method is delicate because the likelihood function is very complex and has no analytical derivative. This difficulty stems from the interdependence of values, as well as the fact that not all previous observations are available for the first p values.
- Method of moments: use of Yule-Walker equations.

Maximum Conditional Likelihood One way to simplify the complexity of the likelihood function is to condition this function to the first p observations. The log-likelihood function becomes:

$$\log\ \mathcal{L}(x_1,...,x_T) = \frac{-(T-p)}{2}\log(2\pi) - \frac{-(T-p)}{2}\log(\sigma^2) - \sum_{t=p+1}^{T}\frac{1}{2\sigma^2}(x_t - \mu - \alpha_1 x_{t-1} - ... - \alpha_p x_{t-p})$$

Maximizing this function in relation to the 3b1 parameters corresponds to minimizing model errors. The maximum conditional likelihood estimator thus corresponds to the least squares estimator. The resulting estimator will be equivalent to the unconditional estimator in large samples and both have the samc asymptotic distribution (Hamilton 1994, p. 126). It can bc biascd.

3.3.2 Prediction in autoregressive models

In this paragraph, it is assumed that $\{X_t\}$ is a stationary process that follows a model $AR(p)$, ie check

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + ... + \alpha_p(X_{t-p} - \mu) + \sigma\epsilon_t$$

with ϵ_t a reduced centered white noise.

Objective:

we seek to predict the value taken by the process at the moments t + 1, t + 2, from the knowledge of the values taken by this process up to the moment t, ie $x_0, ..., x_t$

The model is rewritten in the form

$$X_t = \alpha_1(X_{t-1} - \mu) + \alpha_2(X_{t-2} - \mu) + ... + \alpha_p(X_{t-p} - \mu) + \mu + \sigma\epsilon_t$$

In general, the following quantities are used :

$$\widehat{X}_{t+1|t} = \alpha_1(X_t - \mu) + ... + \alpha_p(X_{t-p+1} - \mu) + \mu$$

to predict X_{t+1} from X_1,..., X_t.

$$\widehat{X}_{t+2|t} = \alpha_1(\widehat{X}_{t+1|t} - \mu) + ... + \alpha_p(X_{t-p+2} - \mu) + \mu$$

And in general

$$\widehat{X}_{t+k|t} = \alpha_1(\widehat{X}_{t+k-1|t} - \mu) + ... + \alpha_p(X_{t+k-p+2} - \mu) + \mu$$

Remark 39 *In the case of Order 1 models, we have*

$$\widehat{X}_{t+1|t} - \mu = \alpha_1(X_t - \mu), \widehat{X}_{t+2|t} - \mu = \alpha_1^2(X_t - \mu), ...$$

It is easy to check by recurrence that

$$\widehat{X}_{t+k|t} - \mu = \alpha_1^k(X_t - \mu)$$

so in particular that$\widehat{X}_{t+k|t} \to \mu$ *when k tends to infinity.*

Proposition 40 *(Accepted) - If $\{X_t\}$ is a stationary process that follows an AR(p) model, then the best prediction, in the sense of least squares, of X_{t+1} knowing $x_t, ..., x_{t-p+1}$ is given by $\widehat{x}_t$ defined by*

$$\widehat{x}_{t+1|t} = \alpha_1 x_1 + ... + \alpha_p x_{t-p+1} + \mu$$

Quality of the prediction

Prediction at a time step We have by definition $\widehat{X}_{t+1|t} - X_{t+1} = \sigma\epsilon_{t+1}$. ϵ_{t+1} represents the prediction error at a time step. In particular, the estimate is unbiased and the variance of the estimation error is σ^2. In general, in order to build intervals of prediction, one is led to suppose that ϵ follows a centered and reduced Gauss law. We deduce from this that $\widehat{X}_{t+1|t} - X_{t+1}$ also follows a law of centered Gauss law and variance σ^2. The interval

$$[\widehat{X}_{t+1|t} - \sigma\Phi^{-1}(1 - 0,95/2),\ \widehat{X}_{t+1|t} + \sigma\Phi^{-1}(1 - 0,95/2)]$$

is called the 95% prediction interval. In practice, σ and $\widehat{X}_{t+1|t}$ are unknown and are estimated.

Prediction at two steps of time We have by definition $\widehat{X}_{t+2|t} - X_{t+2} = \sigma\epsilon_{t+2} + \alpha_1\sigma\epsilon_{t+1}$. In particular, the estimate is non-biased and the variance of the estimation error is $(1 + \alpha_1^2)\sigma^2$.You can also build prediction interval.

3.3.3 Simulations and Correlograms

AR(1)

Let's begin with an AR(1) process. This is similar to a random walk, except that does not have to equal unity. Our model is going to have. The R code for creating this simulation is given as follows:

```
> set.seed(1)
> x <- w <- rnorm(100)
> for (t in 2:100) x[t] <- 0.6*x[t-1] + w[t]
```

Notice that our for loop is carried out from 2 to 100, not 1 to 100, as x[t-1] when is not indexable. Similarly for higher order AR(p) processes, must range from to 100 in this loop.

We can plot the realisation of this model and its associated correlogram using the layout function:

```
> layout(1:2)
> plot(x, type="l")
> acf(x)
```

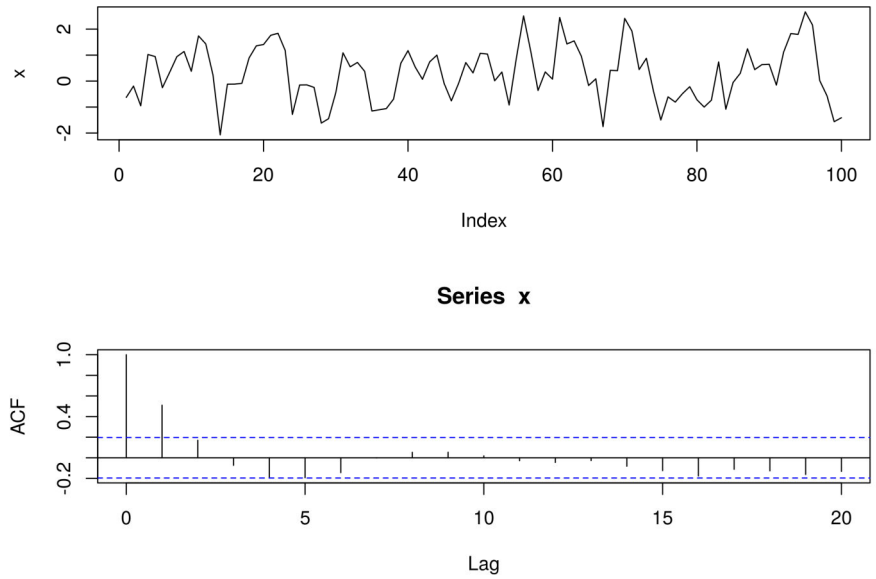


figure 6.1

Let's now try fitting an AR(p) process to the simulated data we've just generated, to see if we can recover the underlying parameters. You may recall that we carried out a similar procedure in the article on white noise and random walks.

As it turns out R provides a useful command ar to fit autoregressive models. We can use this method to firstly tell us the best order of the model (as determined by the AIC above) and provide us with parameter estimates for the, which we can then use to form confidence intervals.

For completeness, let's recreate the series:

```
> set.seed(1)
> x <- w <- rnorm(100)
> for (t in 2:100) x[t] <- 0.6*x[t-1] + w[t]
```

Now we use the ar command to fit an autoregressive model to our simulated AR(1) process, using maximum likelihood estimation (MLE) as the fitting procedure.

We will firstly extract the best obtained order:

```
> x.ar <- ar(x, method = "mle")
> x.ar$order
1
```

The ar command has successfully determined that our underlying time series model is an AR(1) process.

We can then obtain the parameter(α) estimates:

> x.ar$ar

0.5231187

Finally, we can use the standard error (with the asymptotic variance) to construct 95% confidence intervals around the underlying parameter(s). To achieve this, we simply create a vector c(-1.96, 1.96) and then multiply it by the standard error:

x.ar$ar + c(-1.96, 1.96)*sqrt(x.ar$asy.var)

0.3556050 0.6906324

The true parameter does fall within the 95% confidence interval, as we'd expect from the fact we've generated the realisation from the model specifically.

3.4 Moving Average Processes

In this section we consider moving average processes. These processes form a flexible class of models for many observed phenomena. They are constructed from the idea that the observation at time t is explained linearly by the observations of a white noise. To study these processes, we will start with characteristic functions by giving definitions with the representation of their trajectories.

Definition 41 *A process $(X_t, t \in \mathbb{Z})$, is said to be a moving average process of order q, denoted $MA(q)$ if it is defined by :*

$$X_t = \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$$

where θ_i are real and $\epsilon_t \rightsquigarrow$white noise(0, σ_ϵ^2) .

The delay operator L is introduced (this is simply defined as: $(B^i X_t = X_{t-i})$ to simplify the formula of an $MA(q)$:

$$\begin{aligned} X_t &= (1 + \theta_1 L + ... + \theta_q L^q)\epsilon_t \\ X_t &= \Theta(L)\epsilon_t \end{aligned}$$

where $\Theta(L)$ is a delay polynomial of order q associated with the X_t process.

Lemma 42 *By a direct calculation, we have*

$$\begin{aligned} E(X_t) &= E[\sum_{i=1}^{q} \theta_i \epsilon_{t-i}] = 0, \\ var(X_t) &= \gamma(0) = E(X_t^2) = \sum_{i=1}^{q}\sum_{j=1}^{q} \theta_i \theta_j E[\epsilon_{t-i}\epsilon_{t-j}] \end{aligned}$$

we have

$$E[\epsilon_{t-i}\epsilon_{t-j}] = \begin{cases} 0 & if \quad j \neq i \\ \sigma_\epsilon^2 & if \quad j=i \end{cases} .$$

So,

$$var(X_t) = \sigma_\epsilon^2 \sum_{i=0}^{q} \theta_i^2$$

3.4.1 autocovariance function and autocorrelation function

Lemma 43 *The autocovariance function is given by:*

$$\gamma(h) = E(X_t X_{t+h}) = \begin{cases} \sigma^2 \sum_{i=0}^{q} \theta_{|h|+i}\theta_i & if \quad |h| \leq q \ , \ h \in \mathbb{Z} \\ 0 & else \end{cases} .$$

Proof. (td) ■

Lemma 44 *The autocorrelation function is given by:*

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} \frac{\sum_{i=0}^{q-|h|} \theta_{|h|+i}\theta_i}{\sum_{i=0}^{q} \theta_i^2} & if \quad |h| \leq q \ , \ h \in \mathbb{Z} \\ 0 & else \end{cases}$$

Stationarity and causality

The stationarity of a time series is a fundamental property. It indicates whether the characteristics of it change over time or not. In this section, we move to the probabilistic study of $MA(q)$ processes. In particular, stationarity, causality and invcrsibility.

Let $(X_t, t \in \mathbb{Z})$ a $MA(q)$ process, is stationary by definition because :

1. $E(X_t) = E(\sum_{i=0}^{q} \theta_i \epsilon_{t-i}) = \sum_{i=0}^{q} \theta_i E(\epsilon_{t-i}) = 0$ does not depend on t.

2. $var(X_t) = \sigma_\epsilon^2 \sum_{i=0}^{q} \theta_i^2$ does not depend on t.

3. $\gamma(h) = \begin{cases} \sigma^2 \sum_{i=0}^{q} \theta_{|h|+i}\theta_i & \text{if} \quad |h| \leq q \ , \ \text{h} \in \mathbb{Z} \\ 0 & \text{else} \end{cases}$ does not depend on t.

Proposition 45 *Sometimes when we speak of a causal process, we say that it has an $MA(\infty)$ representation, that is to say,*

$$X_t = \sum_{k=0}^{\infty} a_k \epsilon_{t-k}.$$

So, $MA(q)$ is an $MA(\infty)$ where

$$a_k = \begin{cases} \theta_k & if \quad 0 \leq \quad k \leq q \\ 0 & otherwise \end{cases}$$

so a moving average process of order q is always causal.

Definition 46 *A process is said to be invertible, if there is a sequence of real constants (b_k) such as :*

$$\sum_{k=0}^{\infty} |b_k| < \infty \quad \text{and} \quad \epsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}.$$

Proposition 47 *It is said that a process is invertible, if it admits an $AR(\infty)$ representation , that is,*

$$\epsilon_t = \sum_{k=0}^{\infty} b_k X_{t-k}$$

Now, we determine the inversibility conditions of a process $(X_t, t \in \mathbb{Z})$ moving average of order q defined by

$$X_t = \Theta(L)\, \epsilon_t.$$

Indeed, be $\omega_1, ..., \omega_q$ the roots of the equation

$$1 - \theta_1 L - ... - \theta_q L^q = 0$$

.The absence of the double root is assumed, hence

$$X_t = \left(1 - \omega_1^{-1} L\right) \left(1 - \omega_2^{-1} L\right) ... \left(1 - \omega_q^{-1} L\right) \epsilon_t$$

with

$$\begin{aligned} \epsilon_t &= \left(1 - \omega_1^{-1} L\right)^{-1} \left(1 - \omega_2^{-1} L\right)^{-1} ... \left(1 - \omega_q^{-1} L\right)^{-1} X_t \\ &= (\sum_{i=0}^{\infty} \omega_1^{-i} L^i) ... (\sum_{i=0}^{\infty} \omega_q^{-i} L^i) X_t \\ &= \sum_{s=0}^{\infty} (\sum_{s=i_1+...+i_q} \omega_1^{-i_1} ... \omega_q^{-i_q}) X_{t-s}. \end{aligned}$$

For this last writing to have a meaning, it is necessary that the sums converge which implies that the roots $\omega_1, ..., \omega_q$, possibly complex, have a module greater than the unit or, in other words, that they are located outside the unitary circle: $\|\omega_i\| > 1$, $\forall i \in \overline{0,q}$.

It is also possible to express the conditions of inversibility not on the roots of polynomial $1 - \theta_1 L - ... - \theta_q L^q = 0$ but directly on the coefficients $\theta_1, ..., \theta_q$ of the process.

3.4.2 Statistical inference for Moving Average Processes

As for AR processes, we can build an iterative algorithm to estimate the coefficients of an MA process based on empirical covariance. However, maximum likelihood estimators have better statistical properties. They are used by default in R and SAS.

Example 48 *inferance for MA(1) (TD)*

Prediction of MA processes

For these processes, it is easier to project not on the space generated by $X_1, ..., X_t$ but on $X_{-\infty}, ..., X_t$. This is where the AR(∞) representation of MA processes is useful. Indeed this representation implies that spaces $X_{-\infty}, ..., X_t$ and $\epsilon_{-\infty}, ..., \epsilon_t$ are equal.

Proposition 49 *For all t, the predictor at horizon 1 is*

$$\widehat{X}_{n+1} = P_{[X_{-\infty},...,X_t]}(X_{t+1}) = P_{[\epsilon_{-\infty},...,\epsilon_t]}(X_{t+1}) = b_1\epsilon_t + ... + b_q\epsilon_{t-q}$$

The forecast error at horizon 1 is

$$X_{t+1} - \widehat{X}_{n+1} = X_{t+1} - (b_1\epsilon_t + ... + b_q\epsilon_{t-q}) = \epsilon_{t+1}$$

You can easily see from the representation

$$X_{t+h} = \epsilon_{t+h} + b_1\epsilon_{t+h-1} + ... + b_q\epsilon_{t+h-q}$$

that the predictor is null as soon as the horizon h exceeds q.

$$X_{t+h} - \widehat{X}_{n+h} = X_{n+1}$$

3.4.3 Simulations and Correlograms

Let's start with a MA(1) process with $\theta = 0.6$, As with the AR(p) models in the previous article we can use R to simulate such a series and then plot the correlogram. Since we've had a lot of practice in the previous Time Series Analysis article series of carrying out plots, I will write the R code in full, rather than splitting it up:

```
> set.seed(1)
> x <- w <- rnorm(100)
> for (t in 2:100) x[t] <- w[t] + 0.6*w[t-1]
> layout(1:2)
> plot(x, type="l")
> acf(x)
```

The output is as follows:

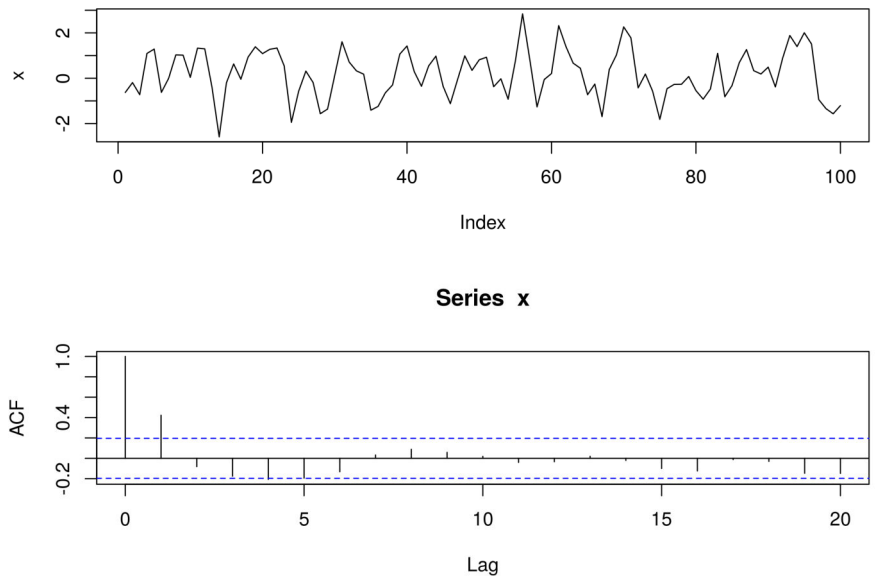


figure 7

Instead, we must use the more general arima command and set the autoregressive and integrated components to zero. We do this by creating a 3-vector and setting the first two components (the autogressive and integrated parameters, respectively) to zero:

> x.ma <- arima(x, order=c(0, 0, 1))
> x.ma

Call:
arima(x = x, order = c(0, 0, 1))
Coefficients:
ma1 intercept
0.6023 0.1681
s.e. 0.0827 0.1424
sigma^2 estimated as 0.7958: log likelihood = -130.7, aic = 267.39

The major difference between arima and ar is that arima estimates an intercept term because it does not subtract the mean value of the series. Hence we need to be careful when carrying out predictions using the arima command.

Exercise 50 *Let the process AR(2) define by :*

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \epsilon_t \tag{3.1}$$

with ϵ_t is an white noise and variance σ^2.
1. Calculate autocovaries of order 1 and 2 based on variance.
2. Taking the values: $\varphi_1 = 1.5$ and $\varphi_2 = 0.56$. show that this process is stationary.
3. Calculate the variance.
4. Deduce autocovaries and autocorrelations.

Exercise 51 *Let the process MA(1) define by :*

$$X_t = \epsilon_t - \theta_1 \epsilon_{t-1}$$

with ϵ_t is an white noise N(0,1)
1. Is this process stationary?
2.Calculate the variance and autocovariance in order 1.
3.Give the forecast on horizon 1.

Chapter 4

ARMA Processes

4.1 Introduction to ARMA processes

4.1.1 ARMA (1,1) processes

Introduction

In this chapter we introduce, through an example, some of the key properties of an important class of linear processes known as ARMA (autoregressive moving average) processes. These are defined by linear difference equations with constant coefficients. As our example we shall consider the ARMA(1,1) process. Higher-order ARMA processes will be discussed in this chapter.

Definition 52 *The time series $\{X_t\}$ is an ARMA(1, 1) process if it is stationary and satisfies (for every t)*

$$X_t - \alpha X_{t-1} = \epsilon_t + \theta \epsilon_{t-1} \tag{4.1}$$

where $\epsilon_t \rightsquigarrow WN(0, \sigma^2)$ and $\alpha + \theta \neq 0$.

Using the backward shift operator B, (4.1) can be written more concisely as

$$\alpha(L) X_t = \theta(L) \epsilon_t \tag{4.2}$$

where $\alpha(L)$ and $\theta(L)$ are the linear filters
$\alpha(L) = 1 - \alpha L$ and $\theta(L) = 1 + \theta L$,respectively.

We first investigate the range of values of α and θ for which a stationary solution of (4.1) exists. If $|\alpha| < 1$, let $\chi(z)$ denote the power series expansion of $1/\alpha(z)$,

i.e., $\sum_{j=0}^{\infty} \alpha^j z^j$, which has absolutely summable coefficients. Then we conclude that $\chi(L)\alpha(L) = 1$. Applying $\chi(L)$ to each side of (4.2) therefore gives

$$X_t = \chi(L)\theta(L)\epsilon_t = \psi(L)\epsilon_t$$

where $\psi(L) = \sum_{j=0}^{\infty} \psi_j L^j = (1 + \alpha L + \alpha L^2 + ...)(1 + \theta L)$.

By multiplying out the right-hand side, we find that$\psi_0 = 1$ and $\psi_j = (\alpha + \theta)\alpha^{j-1}$ for $j \geq 1$.

we conclude that the MA(∞) process

$$X_t = \epsilon_t + (\alpha + \theta) \sum_{j=1}^{\infty} \alpha^{j-1} \epsilon_{t-j} \tag{4.3}$$

is the unique stationary solution of (4.1).

Now suppose that $|\alpha| > 1$. We first represent $1/\alpha(z)$ as a series of powers of z with absolutely summable coefficients by expanding in powers of z^{-1} , we have

$$\frac{1}{\alpha(z)} = -\sum_{j=1}^{\infty} \alpha^{-j} z^{-j}.$$

.

Then we can apply the same argument as in the case where $|\alpha| < 1$ to obtain the unique stationary solution of (4.1). We let $\chi(L) = -\sum_{j=1}^{\infty} \alpha^{-j} L^{-j}$ and apply $\chi(L)$ to each side of (4.2) to obtain

$$X_t = \chi(L)\theta(L)\epsilon_t = -\theta\alpha^{-1}\epsilon_t - (\alpha + \theta) \sum_{j=1}^{\infty} \alpha^{-j-1} \epsilon_{t+j}. \tag{4.4}$$

If $\alpha = \pm 1$, there is no stationary solution of (4.1). Consequently, there is no such thing as an ARMA(1,1) process with $\alpha = \pm 1$ according to our definition.

We can now summarize our findings about the existence and nature of the stationary solutions of the ARMA(1,1) recursions (4.2) as follows:

• A stationary solution of the ARMA(1,1) equations exists if and only if $\alpha = \pm 1$.

• If $|\alpha| < 1$, then the unique stationary solution is given by (4.3). In this case we say that $\{X_t\}$ is causal or a causal function of $\{\epsilon_t\}$, since X_t can be expressed in terms of the current and past values $\epsilon_s, s \leq t$.

• If $|\alpha| > 1$, then the unique stationary solution is given by (4.4). The solution is noncausal, since X_t is then a function of $\epsilon_s, s \geq t$.

Just as causality means that X_t is expressible in terms of $\epsilon_s, s \leq t.$, the dual concept of invertibility means that ϵ_t is expressible in terms of $X_s, s \leq t$. We show now that the ARMA(1,1) process defined by (4.1) is invertible if$|\theta| < 1$. To demonstrate this, let $\xi(z)$ denote the power series expansion of $1/\theta$(z), i.e., $\sum_{j=0}^{\infty}(-\theta)^j z^j$, which has absolutely summable coefficients. we have $\xi(L)\theta(L) = 1$, and applying $\xi(L)$ to each side of (4.2) gives

$$\epsilon_t = \xi(L)\alpha(L)X_t = \pi(L)X_t,$$

where

$$\pi(L) = \sum_{j=0}^{\infty} \pi_j L^j = (1 - \theta L + (-\theta)^2 L^2 + ...)(1 - \alpha L).$$

By multiplying out the right-hand side, we find that

$$\epsilon_t = X_t - (\alpha + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} X_{t-j}. \tag{4.5}$$

Thus the ARMA(1,1) process is invertible, since ϵ_t can be expressed in terms of the present and past values of the process $X_s, s \leq t$. An argument like the one used to

show noncausality when $|\alpha| > 1$ shows that the ARMA(1,1) process is non-invertible when $|\theta| > 1$, since then

$$\epsilon_t = -\alpha\theta^{-1} X_t + (\alpha + \theta) \sum_{j=1}^{\infty} (-\theta)^{-j-1} X_{t+j}. \tag{4.6}$$

We summarize these results as follows:

• If $|\theta| < 1$, then the ARMA(1,1) process is invertible, and ϵ_t is expressed in terms of $X_s, s \leq t.$, by (4.5).

• If $|\theta|>1, then\ the ARMA(1,1) process is noninvertible, and \epsilon_t$ is expressed in terms of $X_s, s \geq\leq t$ by (4.6).

4.2 ARMA(p, q) Processes

In previous section we presented an ARMA(1,1) process and discussed some of its key properties. These included existence and uniqueness of stationary solutions of the

defining equations and the concepts of causality and invertibility. In this section we extend these notions to the general ARMA(p, q) process.

Definition 53 *$\{X_t\}$ is an ARMA(p, q) process if $\{X_t\}$ is stationary and if for every t,*

$$X_t - \alpha_1 X_{t-1} - ... - \alpha_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \tag{4.7}$$

,

where $\{\epsilon_t\} \rightsquigarrow WN(0, \sigma^2)$ and the polynomials $(1 - \alpha_1 z - ... - \alpha_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors.

The process $\{X_t\}$ is said to be an ARMA(p, q) process with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process. It is convenient to use the more concise form of (4.7)

$$\alpha(L) X_t = \theta(L) \epsilon_t, \tag{4.8}$$

where $\alpha(.)$ and $\theta(.)$ are the pth and qth-degree polynomials

$$\alpha(z) = 1 - \alpha_1 z - ... - \alpha_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$

and B is the backward shift operator . The time series $\{X_t\}$ is said to be an autoregressive process of order p (or AR(p)) if $\theta(z) = 1$, and a moving-average process of order q (or MA(q)) if $\alpha(z) = 1$. An important part of Definition is the requirement that $\{X_t\}$ be stationary. For the ARMA(1,1) equations (4.1), that a stationary solution exists (and is unique) if and only if $\alpha_1 \neq \pm 1$. The latter is equivalent to the condition that the autoregressive polynomial $\alpha(z) = 1 - \alpha_1 z \neq 0$ for $z \neq \pm 1$. The analogous condition for the general ARMA(p, q) process is $\alpha(z) = 1 - \alpha_1 z - ... - \alpha_p z^p \neq 0$ for all complex z with $|z| = 1$. (Complex z is used here, since the zeros of a polynomial of degree $p > 1$ may be either real or complex. The region defined by

the set of complex z such that $|z| = 1$ is referred to as the unit circle.) If $\alpha(z) \pm 0$ for all z on the unit circle, then there exists $\delta > 0$ such that

$$\frac{1}{\alpha(z)} = \sum_{j=-\infty}^{\infty} \chi_j z^j \qquad \text{for all } 1\text{-}\delta < |z| < 1 + \delta,$$

and $\sum_{j=-\infty}^{\infty} |\chi_j| < \infty$. We can then define $1/\alpha(L)$ as the linear filter with absolutely summable coefficients

$$\frac{1}{\alpha(L)} = \sum_{j=-\infty}^{\infty} \chi_j L^j.$$

Applying the operator$\chi(L); = \frac{1}{\alpha(L)}$ to both sides of (4.8), we obtain

$$X_t = \chi(L)\,\alpha(L) X_t = \chi(L)\,\theta(L)\,\epsilon_t = \psi(L)\,\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_{j\epsilon_{t-j}}, \tag{4.9}$$

where $\psi_j(z) = \chi(z)\,\theta(z) = \sum_{-\infty}^{+\infty} \psi_j z^j$. for the ARMA(1,1) process, it follows that $\psi(B)\,\epsilon_t$ is the unique stationary solution of (4.7).

Existence and Uniqueness: A stationary solution $\{X_t\}$ of equations (4.7) exists (and is also the unique stationary solution) if and only if

$$\alpha(z) = 1 - \alpha_1 z - ... - \alpha_p z^p \neq 0 \quad for \quad all \quad |z| = 1.$$

Causality: An ARMA(p, q) process $\{X_t\}$ is causal, or a causal function of $\{\epsilon_t\}$, if there exist constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{4.10}$$

Causality is equivalent to the condition

$$\alpha(z) = 1 - \alpha_1 z - ... - \alpha_p z^p \neq 0 \quad \text{for all} \quad |z| \leq 1 \tag{4.11}$$

The proof of the equivalence between causality and (4.11) follows from elementary properties of power series. From (4.10) we see that $\{X_t\}$ is causal if and only if $\chi(z) = 1/\alpha(z) = \sum_{j=0}^{\infty} \chi_j z^j$ (assuming that α(z) and θ(z) have no common factors). But this, in turn, is equivalent to (4.11).

The sequence $\{\psi_j\}$ in (4.10) is determined by the relation $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\alpha(z)}$, or equivalently by the identity

$$(1 - \alpha_1 z - ... - \alpha_p z^p)(\psi_0 + \psi_1 z + ...) = 1 + \theta_1 z + ... + \theta_q z^q.$$

Equating coefficients of z^j , j = 0, 1, . . ., we find that
$1 = \psi_0$,
$\theta_1 = \psi_1$ - $\psi_1 \alpha_1$,
$\theta_2 = \psi_2$ - $\psi_1 \alpha_1 - \psi_0 \alpha_2$,
...
or equivalently,

$$\psi_j - \sum_{k=1}^{p} \alpha_k \psi_{j-k} = \theta_j, \quad \text{j=0,1,...,} \tag{4.12}$$

where $\theta_0 := 1$, $\theta_j := 0$ for$j > q$, and $\psi_j := 0$ for $j < 0$.

Invertibility: An ARMA(p, q) process $\{X_t\}$ is invertible if there exist constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and

$$\epsilon_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}$$

Invertibility is equivalent to the condition

$$\theta(z) = 1 + \theta_1 z + ... + \theta_p z^p \neq 0 \quad \text{for all} \quad |z| \leq 1. \tag{4.13}$$

Example 54 *An ARMA(1,1) process*
Consider the ARMA(1,1) process$\{X_t\}$ *satisfying the equations*

$$X_t - 0.5X_{t-1} = \epsilon_t + 0.4\epsilon_{t-1}, \{\epsilon_t\} \rightsquigarrow WN(0, \sigma^2). \tag{4.14}$$

(3.1.9)
Since the autoregressive polynomial $\alpha(z) = 1 - 0.5z$ *has a zero at* $z = 2$, *which is located outside the unit circle, we conclude from (4.10) and (4.11) that there exists a unique ARMA process satisfying (4.14) that is also causal.*
The coefficients $\{\psi_j\}$ *in the MA*(∞) *representation of* $\{X_t\}$ *are found directly from (4.12):*

- ψ_0 =1,
- ψ_1 =0.4 +0 .5,
- $\psi_2 = 0$ .5(0.4 +0 .5),
- .
- .
- .
- $\psi_j = 0$ $.5^{j-1}$(0.4 +0 .5), j =1, 2, ...

The MA polynomial $\theta(z) = 1 + 0.4z$ *has a zero at z =-1/0.4= -2.5, which is also located outside the unit circle. This implies that* $\{X_t\}$ *is invertible with coefficients* $\{\pi_j\}$ *given by(4.13)*

- $\pi_0 = 1$,
- $\pi_1 = -(0.4 + 0.5)$,
- $\pi_2 = -(0.4 + 0.5)(-0.4)$,
- .
- .
- .
- $\pi_j = -(0.4 + 0.5)(-0.4)^{j-1}, \quad j = 1, 2, ...,$

The ACF and PACF of an ARMA(p, q) Process

In this section we discuss the methods for computing the autocovariance function $\gamma(\cdot)$ of a causal ARMA process $\{X_t\}$. The autocorrelation function is readily found from the ACVF on dividing by $\gamma(0)$. The partial autocorrelation function (PACF) is also found from the function $\gamma(\cdot)$.

Calculation of the ACVF First we determine the ACVF $\gamma(\cdot)$ of the causal ARMA(p, q) process defined by

$$\alpha(L) X_t = \theta(L) \epsilon_t, \{\epsilon_t\} \rightsquigarrow WN(0, \sigma^2) \tag{4.15}$$

where $\alpha(z) = 1 - \alpha_1 z - ... - \alpha_p z^p \neq 0$ and $\theta(z) = 1 + \theta_1 z + ... + \theta_p z^p$. The causality assumption implies that

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \tag{4.16}$$

From this representation, we obtain

$$\gamma(h) = E(X_{t+h} X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}.$$

The Autocorrelation Function Recall that the ACF of an ARMA process $\{X_t\}$. is the function $\rho(\cdot)$ found immediately from the ACVF $\gamma(\cdot)$ as

$$\rho(\cdot) = \frac{\gamma(\cdot)}{\gamma(0)}$$

The Partial Autocorrelation Function The partial autocorrelation function (PACF) of an ARMA process $\{X_t\}$ is the η function $\eta(\cdot)$ defined by the equations :

$$\eta(0) = 1$$

and

$$\eta(h) = \alpha_{hh}, h \geq 1,$$

where α_{hh} is the last component of

$$\eta_h = \Gamma_h^{-1}\gamma_h,$$

and

$$\Gamma_h^{-1} = [\gamma(i-j)]_{i,j=1}^{h}$$

and

$$\gamma_h = [\gamma(1), \gamma(2), ..., \gamma(h)]'.$$

4.3 Identification and estimation of ARMA models

In practice, when an AR, MA or ARMA model has to be adjusted to actual data, the first question that arises is the choice of orders p and q of the ARMA model (it is considered that the AR and MA are a special case of ARMA with respectively q = 0 and p = 0). In the case of an ARMA there is a way to determine the p and q orders using the wedge method based on some determinant of correlation matrices that we do not develop here. Another approach is to consider a credible set of ARMA(p,q) models and then select a candidate by a model selection method based on BIC or AIC information criteria.

Once the order and type of model are chosen, to estimate the coefficients of the models several approaches are possible.

1. **Yule-Walker equations.** We have previously obtained that in the case of an AR process, the coefficients verify: $\begin{pmatrix} -\alpha_1 \\ \vdots \\ -\alpha_p \end{pmatrix} = [\rho(i-j)]^{-1} \begin{pmatrix} \rho(1) \\ \vdots \\ \rho(p) \end{pmatrix}$

by exploiting the fact that empirical autocorrelations are a convergent estimator of autocorrelations, we can obtain an estimator of the coefficients α_j.

To obtain these equations, we used orthogonality relationships. Reminder:

$$X_t + \sum_{j=1}^{p} X_{t-j} = \epsilon_t$$

by mulitfolding by X_{t-h}, taking the mean and noting that for $h > 0$, $\mathrm{E}(\epsilon_t X_{t-h}) = 0$ one obtains the relation at the origin of the Yule-Walker equations:

$$\gamma(h) + \sum_{j=1}^{p} \gamma(h-j) = 0$$

This is equivalent to calculating the linear projection of X_t on its past thus minimizing in α_j the quadratic risk:

$$E\left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j}\right)^2$$

Another approach is to directly minimize the empirical risk:

$$\sum_{t=p+1}^{n} \left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j}\right)^2$$

We can show that the asymptotic behavior of this estimator is the same as that obtained by Yule-Walker.

2. **Lesser conditional squares.**

if ϵ_t is Gaussian, the vector of the observations $(X_1, ..., X_n)$ is Gaussian and we can then calculate its likelihood. A simplified approach is to calculate the density of $(X_{p+1}, ..., X_n)$conditionally to $(X_1, ..., X_p)$ which is Gaussian. Knowing that $X_t/X_{t-1}, ..., X_{t-p}$ is a Gaussian of expected $\sum_{j=1}^{p} \alpha_j X_{t-j}$ and variance σ^2, using the formula of successive conditioning:

$$\begin{aligned} f(X_{p+1}, ..., X_n/X_1, ..., X_p) &= \prod_{t=p+1}^{n} f(X_t/X_{t-1}, ..., X_{t-p}) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n-p}{2}}} \exp\left(-\frac{\sum_{t=p+1}^{n}\left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j}\right)^2}{2\sigma^2}\right) \end{aligned}$$

maximizing this conditional likelihood in α means, by ln, minimizing:

$$(\alpha_1, ..., \alpha_p, \sigma^2) \to \frac{\sum_{t=p+1}^{n}\left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j}\right)^2}{\sigma^2} + (n-p)\ln(\sigma^2)$$

The approach is therefore to obtain $\widehat{\alpha}_{css}$ (for conditional sum of square) by minimizing:

$$\widehat{\alpha}_{css} = \arg\min_{\alpha} \sum_{t=p+1}^{n} \left(X_t - \sum_{j=1}^{p} \alpha_j X_{t-j} \right)^2$$

then

$$\widehat{\sigma}^2_{css} = \frac{1}{n-p} \sum_{t=p+1}^{n} \left(X_t - \sum_{j=1}^{p} \widehat{\alpha}_{css} X_{t-j} \right)^2$$

This approach is generalized to ARMA, as is the calculation of exact likelihood, which we shall not discuss here.

Let us return to the choice of the order of an ARMA(p,q). We have seen previously that the study of partial autocor- and autocorrelations can make it possible to pre-select a number of plausible models. Once the parameters of these models have been estimated, we can then select the one that minimizes the following criteria:

• **AIC Akaike Information Criterion,** adapted to the forecasting problem, defined by :

$$AIC(\alpha, \theta, \sigma^2) = -2log(L(\theta, \alpha, \sigma^2)) + 2k$$

where L is the likelihood, k is the number of parameters in the model so p + q or p + q + 1 if the constant is in the model.

• **BIC Bayesian Information Criterion**, adapted to the forecasting problem, defined by :

$$BIC(\alpha, \theta, \sigma^2) = -2log(L(\theta, \alpha, \sigma^2)) + log(n)k$$

the principle is therefore to select a model that fits well with the data (high likelihood) while penalizing models with too many parameters.

4.3.1 Simulations and Correlograms

As with the autoregressive and moving average models we will now simulate various ARMA series and then attempt to fit ARMA models to these realisations. We carry this out because we want to ensure that we understand the fitting procedure, including how to calculate confidence intervals for the models, as well as ensure that the procedure does actually recover reasonable estimates for the original ARMA parameters.

In Part 1 and Part 2 we manually constructed the AR and MA series by drawing N samples from a normal distribution and then crafting the specific time series model using lags of these samples. However, there is a more straightforward way to simulate AR, MA, ARMA and even ARIMA data, simply by using the arima.sim method in R.

Let's start with the simplest possible non-trivial ARMA model, namely the ARMA(1,1) model. That is, an autoregressive model of order one combined with a moving average model of order one. Such a model has only two coefficients α, and β, which represent the first lags of the time series itself and the "shock" white noise terms.

We need to specify the coefficients prior to simulation. Let's take $\alpha = 0.5$and $\beta = -0.5$

```
> set.seed(1)
> x <- arima.sim(n=1000, model=list(ar=0.5, ma=-0.5))
> plot(x)
```

The output of the realisation of an ARMA(1,1) Model,is as follows:

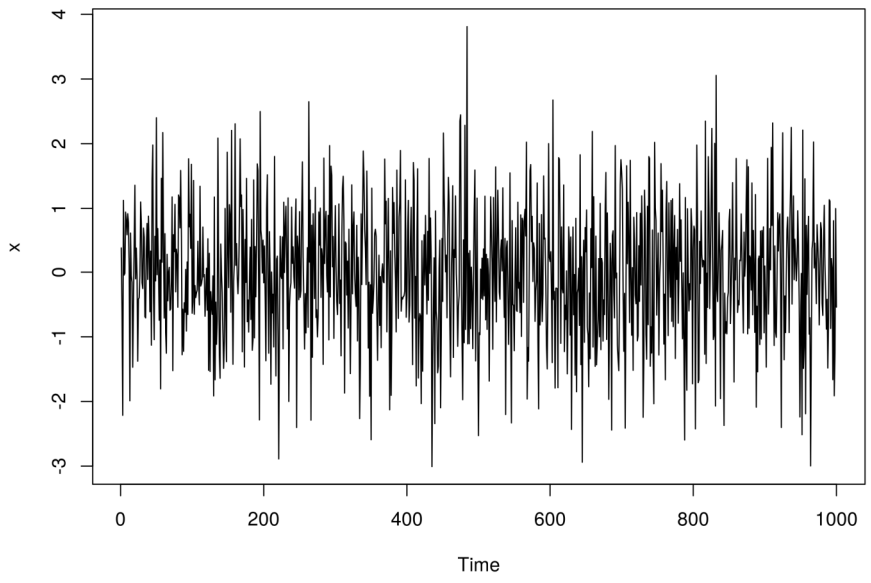


figure 8

Let's also plot the correlogram:

> acf(x)

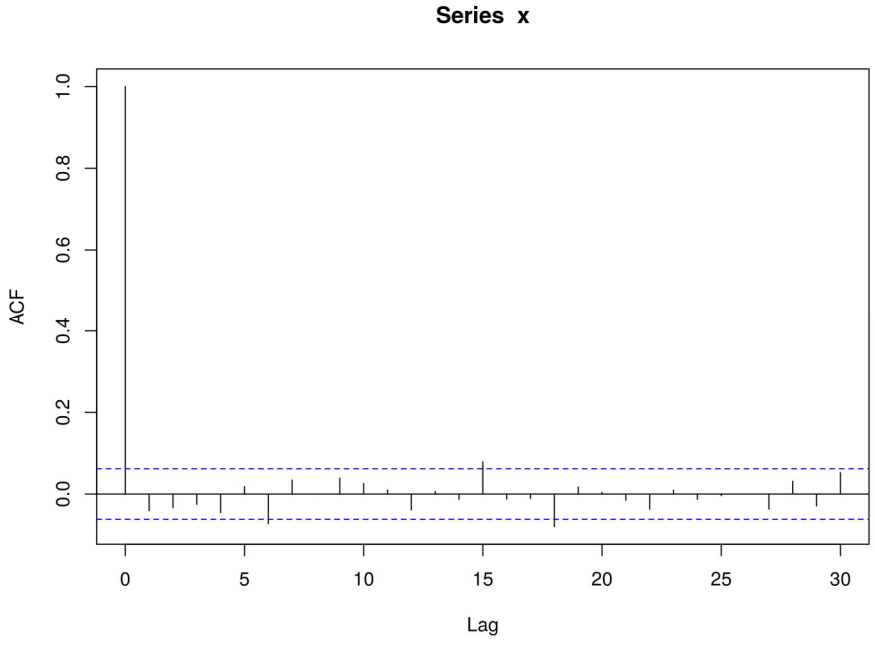


figure 9

We can see that there is no significant autocorrelation, which is to be expected from an ARMA(1,1) model.

Finally, let's try and determine the coefficients and their standard errors using the arima function:

> arima(x, order=c(1, 0, 1))

output

Call:
arima(x = x, order = c(1, 0, 1))
Coefficients:

	ar1	ma1	intercept
	-0.3957	0.4503	0.0538
s.e.	0.3727	0.3617	0.0337

sigma^2 estimated as 1.053: log likelihood = -1444.79, aic = 2897.58

We can calculate the confidence intervals for each parameter using the standard errors:

> -0.396 + c(-1.96, 1.96)*0.373
-1.12708 0.33508
> 0.450 + c(-1.96, 1.96)*0.362

-0.25952 1.15952

The confidence intervals do contain the true parameter values for both cases, however we should note that the 95% confidence intervals are very wide (a consequence of the reasonably large standard errors).

4.4 Non-stationary Process: ARIMA and SARIMA

4.4.1 ARIMA process

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. To better comprehend the data or to forecast upcoming series points, both of these models are fitted to time series data. ARIMA models are applied in some cases where data show evidence of non-stationarity in the sense of mean (but not variance/autocovariance), where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity of the mean function (i.e., the trend). When the seasonality shows in a time series, the seasonal-differencing could be applied to eliminate the seasonal component. Since the ARMA model, according to the Wold's decomposition theorem, is theoretically sufficient to describe a regular (a.k.a. purely nondeterministic) wide-sense stationary time series, we are motivated to make stationary a non-stationary time series, e.g., by using differencing, before we can use the ARMA model. Note that if the time series contains a predictable sub-process (a.k.a. pure sine or complex-valued exponential process), the predictable component is treated as a non-zero-mean but periodic (i.e., seasonal) component in the ARIMA framework so that it is eliminated by the seasonal differencing.

The autoregressive (AR) part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The moving average (MA) part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model. Seasonal ARIMA models are usually denoted ARIMA(p,d,q)(P,D,Q)m, where m refers to the number of periods in each season, and the uppercase P,D,Q refer to the autoregressive, differencing, and mov-

ing average terms for the seasonal part of the ARIMA model. ARIMA models can be estimated following the Box–Jenkins approach.

Definition ARIMA process

Given time series data X_t where t is an integer index and the X_t are real numbers, an ARMA (p', q) model is given by

$$X_t - \alpha_1 X_{t-1} - ... - \alpha_{p'} X_{t-p'} = \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q},$$

Definition 55 *or equivalently by*

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t$$

where L is the lag operator, the $\alpha_i$are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the ϵ_t are error terms. The error terms ϵ_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Assume now that the polynomial $\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right)$ has a unit root (a factor (1 - L)) of multiplicity d. Then it can be rewritten as:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \varphi_i L^i\right) (1 - L)^d$$

An ARIMA(p,d,q) process expresses this polynomial factorisation property with p=p'-d, and is given by:

$$\left(1 - \sum_{i=1}^{p'-d} \varphi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t$$

and thus can be thought as a particular case of an ARMA(p+d,q) process having the autoregressive polynomial with d unit roots. (For this reason, no process that is accurately described by an ARIMA model with $d > 0$ is wide-sense stationary.)

The above can be generalized as follows.

$$\left(1 - \sum_{i=1}^{p'-d} \varphi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t$$

This defines an ARIMA(p,d,q) process with drift $\frac{\delta}{1 - \sum_i \varphi_i}$.

we can reformulate the ARIMA model as follows :

$$\Phi(L)\triangle^d X_t = \Theta(L)\epsilon_t$$

where $(1-L)^d = \triangle^d$ *and* $\Phi(L)$ *is a polynomial with p degree and* $\Theta(L)$ *a polynomial with q degree.*

4.4.2 Differencing

A stationary time series's properties do not depend on the time at which the series is observed. Specifically, for a wide-sense stationary time series, the mean and the variance/autocovariance keep constant over time. Differencing in statistics is a transformation applied to a non-stationary time-series in order to make it stationary in the mean sense (viz., to remove the non-constant trend), but having nothing to do with the non-stationarity of the variance or autocovariance. Likewise, the seasonal differencing is applied to a seasonal time-series to remove the seasonal component. From the perspective of signal processing, especially the Fourier spectral analysis theory, the trend is the low-frequency part in the spectrum of a non-stationary time series, while the season is the periodic-frequency part in the spectrum of it. Therefore, the differencing works as a high-pass (i.e., low-stop) filter and the seasonal-differencing as a comb filter to suppress the low-frequency trend and the periodic-frequency season in the spectrum domain (rather than directly in the time domain), respectively.

To difference the data, the difference between consecutive observations is computed. Mathematically, this is shown as

$$\triangle y_t = y_t - y_{t-1}$$

Differencing removes the changes in the level of a time series, eliminating trend and seasonality and consequently stabilizing the mean of the time series. Sometimes it may be necessary to difference the data a second time to obtain a stationary time series, which is referred to as second-order differencing:

$$\begin{aligned} y_t^* &= \triangle y_t - \triangle y_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

Another method of differencing data is seasonal differencing, which involves computing the difference between an observation and the corresponding observation in the previous season e.g a year. This is shown as:

$$\triangle y_t = y_t - y_{t-m}$$

Where m=duration of season and the differenced data are then used for the estimation of an ARMA model.

Examples

Some well-known special cases arise naturally or are mathematically equivalent to other popular forecasting models. For example:

- An ARIMA(0, 1, 0) model is given by $X_t = X_{t-1} + \epsilon_t$ which is simply a random walk.
- An ARIMA(0, 1, 0) with a constant, given by $X_t = c + X_{t-1} + \epsilon_t$ which is a random walk with drift.
- An ARIMA(0, 0, 0) model is a white noise model.
- An ARIMA(0, 1, 2) model is a Damped Holt's model.
- An ARIMA(0, 1, 1) model without constant is a basic exponential smoothing model.
- An ARIMA(0, 2, 2) model is given by $X_t = 2X_{t-1} - X_{t-2} + (\alpha + \beta - 2)\,\epsilon_{t-1} + (1 - \alpha)\,\epsilon_{t-2} + \epsilon_t$ which is equivalent to Holt's linear method with additive errors, or double exponential smoothing.

Choosing the order

The order p and q can be determined using the sample autocorrelation function (ACF), partial autocorrelation function (PACF), and/or extended autocorrelation function (EACF) method.

Other alternative methods include AIC, BIC, etc. To determine the order of a non-seasonal ARIMA model, a useful criterion is the Akaike information criterion (AIC). It is written as :

$$AIC = -2\log(L) + 2(p + q + k)$$

where L is the likelihood of the data, p is the order of the autoregressive part and q is the order of the moving average part. The k represents the intercept of the ARIMA model. For AIC, if k = 1 then there is an intercept in the ARIMA model ($c \neq 0$) and if k = 0 then there is no intercept in the ARIMA model (c = 0). The corrected AIC for ARIMA models can be written as :

$$AICc = AIC + \frac{2\,(p + q + k)\,(p + q + k + 1)}{T - p - q - k - 1}$$

The Bayesian Information Criterion (BIC) can be written as

$$BIC = AIC + ((\log T) - 2)\,(p + k + q)$$

The objective is to minimize the AIC, AICc or BIC values for a good model. The lower the value of one of these criteria for a range of models being investigated, the better the model will suit the data. The AIC and the BIC are

used for two completely different purposes. While the AIC tries to approximate models towards the reality of the situation, the BIC attempts to find the perfect fit. The BIC approach is often criticized as there never is a perfect fit to real-life complex data; however, it is still a useful method for selection as it penalizes models more heavily for having more parameters than the AIC would.

AICc can only be used to compare ARIMA models with the same orders of differencing. For ARIMAs with different orders of differencing, RMSE can be used for model comparison.

4.4.3 Estimation of coefficients

Forecasts using ARIMA models

The ARIMA model can be viewed as a "cascade" of two models. The first is non-stationary:

$$Y_t = (1-L)^d X_t$$

while the second is wide-sense stationary:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \epsilon_t$$

Now forecasts can be made for the process Y_t, using a generalization of the method of autoregressive forecasting.

Forecast intervals The forecast intervals (confidence intervals for forecasts) for ARIMA models are based on assumptions that the residuals are uncorrelated and normally distributed. If either of these assumptions does not hold, then the forecast intervals may be incorrect. For this reason, researchers plot the ACF and histogram of the residuals to check the assumptions before producing forecast intervals.

95% forecast interval:

$$\widehat{y}_{T+h|T} \pm 1.96\sqrt{V_{T+h|T}}$$

where $V_{T+h|T}$ is the variance of $y_{T+h}\,|y_1, ..., y_T$ ..

For h = 1 ,

$$V_{T+h|T} = \widehat{\sigma}^2$$

For all ARIMA models regardless of parameters and orders. For ARIMA(0,0,q), $y_t = e_t + \sum_{i=1}^{q} \theta_i e_{t-i}$.

$$V_{T+h|T} = \widehat{\sigma}^2 \left[1 + \sum_{i=1}^{h-1} \theta_i e_{t-i}\right], \quad \text{for h=2,3.....}$$

in general, forecast intervals from ARIMA models will increase as the forecast horizon increases.

Variations and extensions

A number of variations on the ARIMA model are commonly employed. If multiple time series are used then the X_t can be thought of as vectors and a VARIMA model may be appropriate. Sometimes a seasonal effect is suspected in the model; in that case, it is generally considered better to use a SARIMA (seasonal ARIMA) model than to increase the order of the AR or MA parts of the model. If the time-series is suspected to exhibit long-range dependence, then the d parameter may be allowed to have non-integer values in an autoregressive fractionally integrated moving average model, which is also called a Fractional ARIMA (FARIMA or ARFIMA) model.

Exemlpe of ARIMA

We'll attempt to predict stock returns by using ARIMA. We'll be using some financial data from tidyquant :

```
# get historical data for single stock. e.g. google
library(tidyquant)
jnj = tq_get("JNJ", get="stock.prices", from="1997-01-01") %>%
tq_transmute(mutate_fun=to.period,period="months")
```

Let's say we are primarily interested in the closing prices of our stock :

```
library(ggplot2)
# showing monthly return for single stock
ggplot(jnj, aes(date, close)) + geom_line()
```

figure 10

We are interested in predicting the returns of JNJ. We compute the log difference of the closing price to stationarize the time series.

```
plot(diff(log(jnj$close)),type='l', main='log returns plot')
```

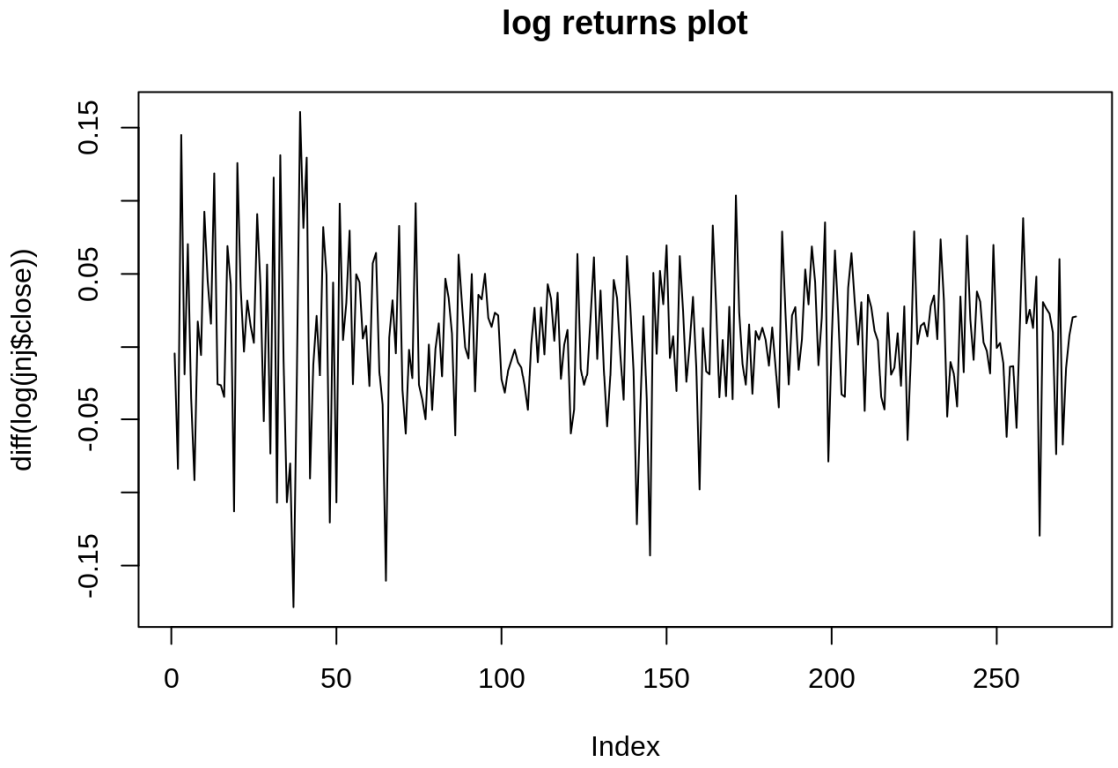


figure 11

Returns look stationary in the plot above. Let's double check with the Dickey Fuller Test of Stationarity:

```
library(tseries)
adf.test(diff(log(jnj$close)), alternative="stationary", k=0)
##
## Augmented Dickey-Fuller Test
##
## data: diff(log(jnj$close))
## Dickey-Fuller = -17.966, Lag order = 0, p-value = 0.01
## alternative hypothesis: stationary
```

The Dickey-Fuller test returns a p-value of 0.01, resulting in the rejection of the null hypothesis and accepting the alternate, that the data is stationary.

It is quite common in financial analysis to predict stock returns. By taking the difference between stocks, we are essentially stationarizing the time series. Though not all stock returns are stationary, in many experiments regarding financial analysis, many assume it is.

In short, ACF and PACF will allow us to determine the order of our parameters for our ARIMA model.

```
acf(diff(log(jnj$close)))
```

```
pacf(diff(log(jnj$close)))
```

Our findings in the ACF/PACF section suggest that model ARIMA(1, 0, 1) might be the best fit. Building an ARIMA model is easy with the forecast package; we just call the function 'arima', and specify our parameters.

```
library(forecast)
(fit <- arima(diff(log(jnj$close)), c(3, 0, 1)))
##
## Call:
## arima(x = diff(log(jnj$close)), order = c(3, 0, 1))
##
## Coefficients:
##          ar1     ar2      ar3      ma1  intercept
##       0.6295  0.0387  -0.0954  -0.7394     0.0056
## s.e.  0.2644  0.0759   0.0775   0.2636     0.0019
##
## sigma^2 estimated as 0.002493:  log likelihood = 432.35,  aic = -852.69
```

Before we make predictions, let's see how our model fitted with our training data.

```
plot(as.ts(diff(log(jnj$close))) )
lines(fitted(fit), col="red")
```

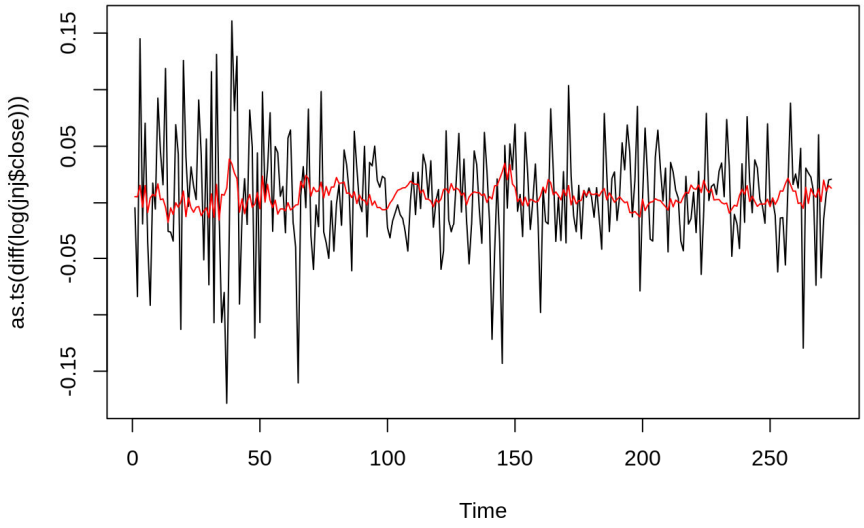


figure 12

Moving on, we can make a prediction of future stock returns with the forecast.Arima function. We predict the returns on the next 5 months:

futurVal <- forecast(fit,h=5, level=c(99)) #confidence level 99%

plot(forecast(futurVal))

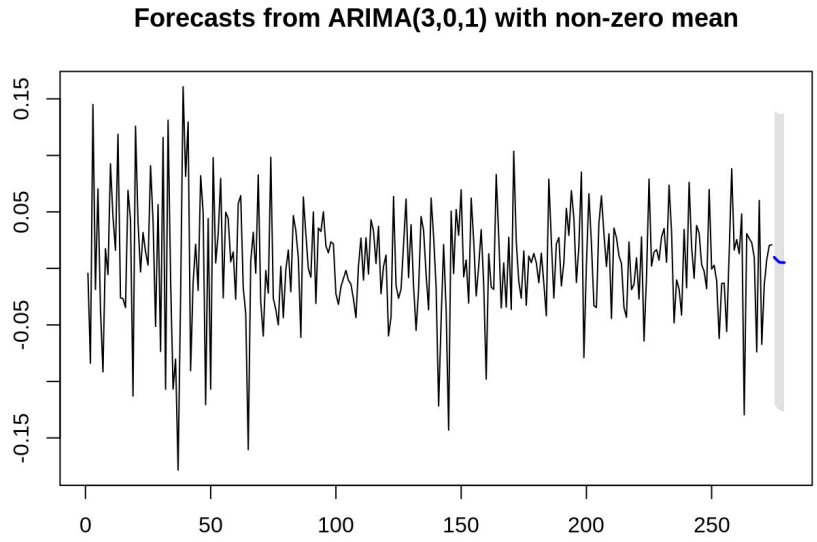


figure 13

```
# 5 predicted values
futurVal$mean
## Time Series:
## Start = 275
## End = 279
## Frequency = 1
## [1] 0.009774685 0.007420760 0.005455234
0.005188938 0.005169821
```

4.4.4 SARIMA process

Introduction

Autoregressive Integrated Moving Average, or ARIMA, is a forecasting method for univariate time series data.

As its name suggests, it supports both an autoregressive and moving average elements. The integrated element refers to differencing allowing the method to support time series data with a trend.

A problem with ARIMA is that it does not support seasonal data. That is a time series with a repeating cycle. ARIMA expects data that is either not seasonal or has the seasonal component removed, e.g. seasonally adjusted via methods such as seasonal differencing. Seasonal Autoregressive Integrated

Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component. It adds three new hyperparameters to specify the autoregression (AR), differencing (I) and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality.

Enter SARIMA (Seasonal ARIMA). This model is very similar to the ARIMA model, except that there is an additional set of autoregressive and moving average components.The additional lags are offset by the frequency of seasonality (ex. 12 — monthly, 24 — hourly). SARIMA models allow for differencing data by seasonal frequency, yet also by non-seasonal differencing. Knowing which parameters are best can be made easier through automatic parameter search frameworks such as pmdarina.

Definition 56 *The SARIMA model defined in this section constitutes a straightforward extension of the nonseasonal ARMA and ARIMA models presented in last two sections, respectively. In their book, Box and Jenkins (1976, Ch. 9) defme this model and justify why it is useful for describing certain kinds of seasonal series.*

Let X_1,X_2,. . , ,X_η represent a sequence of seasonal observations. If, for example, there were n years of data for which each year contains s seasons. this would mean that η is equal to ns. For the case of 15 years of monthly data, then would be a total of 15×12 = 180 observations. If the seasonal time series were not normally distributed and/or the variance of the series changes over time (k., the series is heteroscedastic), one could alleviate this problem by invoking a Box-Cox tramformation (Box and Cox, 1964) defined as

$$X_t^{(\lambda)} = \begin{cases} \lambda^{-1}\left[(X_t + c)^\lambda - 1\right], & \lambda \neq 0 \\ \ln(X_t + c) & \lambda = 0 \end{cases} \tag{4.17}$$

The parameter λ is the Box-Cox power transformation and c is a positive number which is chosen to be just large enough to cause all the enmes in the time series to be positive. If non- normality and heteroscedasticity in the given series were not detected prior to fitting a SARIMA model to the data, these problems would show up in the residuals of the fitted model. At that time, an appropriate Box-Cox transformation could be selected and the parameters of the SARIMA model could then be estimated for the transformed series.

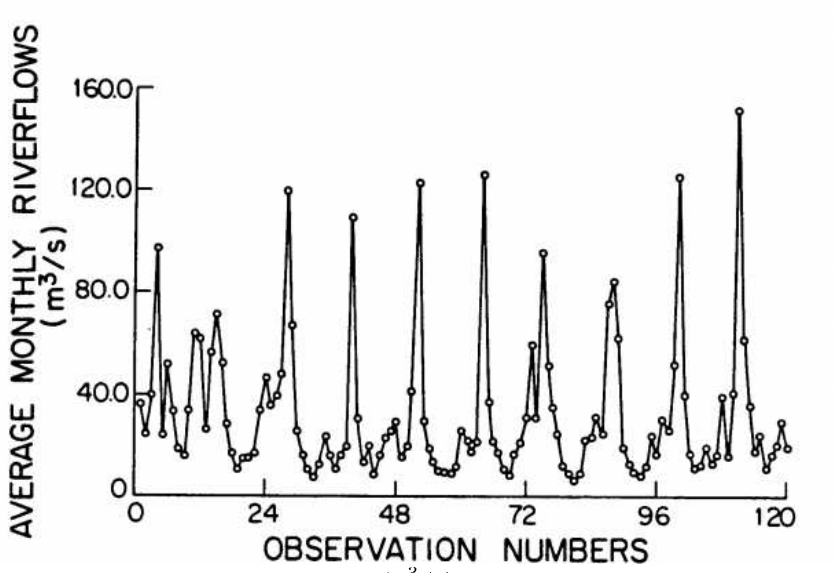


figure 13.1. avrage monthly flows (m^3/s) of the saugeen river at walkerton. ontario. canada from january 1967 until december 1976

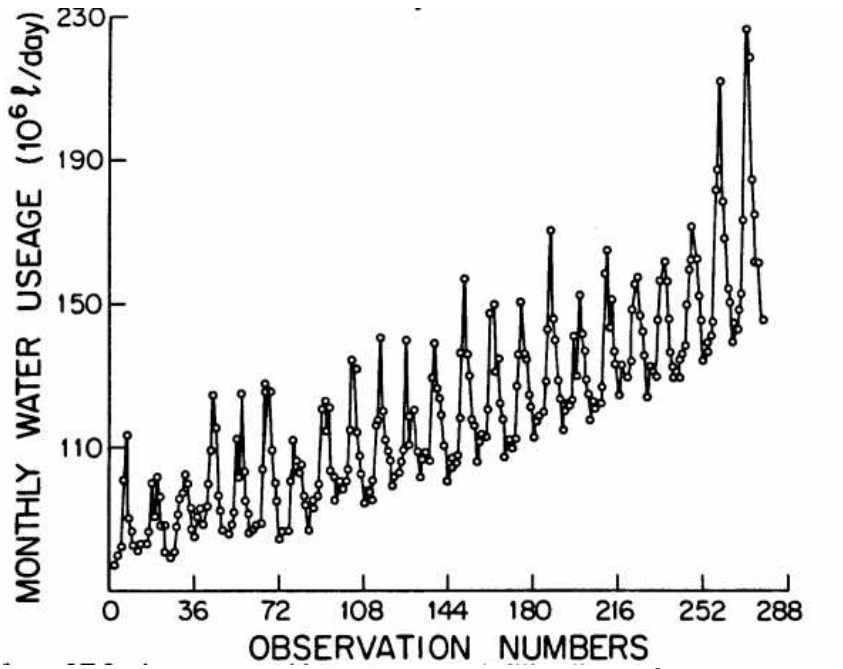


figure 14. avrage monthly water useage(million litres per day) (m^3/s) for the city of London. ontario. canada from january 1966 until december 1988

Figures (13.1 *and 14) graphically depicts how the magnitudes of observations can change across the seasons in a cyclic manner and also from year to year within a given season. To eliminate nonstationarity within each season, one can employ the seasonal differencing operator defined by*

$$\nabla_s X_t^{(\lambda)} = (1 - L^s) X_t^{(\lambda)} = X_t^{(\lambda)} - X_{t-s}^{(\lambda)} \qquad for \quad t = s+1, s+2, ..., \eta \tag{4.18}$$

where s is the number of seasons per year and L^s is the backward shift operator defined by $L^s X_t^{(\lambda)} = X_{t-s}^{(\lambda)}$. When dealing with monthly data, notice that the relationship in (4.18) *only connects observations within the same season. Hence, when using seasonal differencing with monthly data, an observation in March is only subtracted from the observation in March of the previous year. If the $X_t^{(\lambda)}$ series is of length $\eta = sn$. the number of observations in the differenced series is η- s. The differencing operator in* (4.18) *is applied just enough times to remove the seasonal nonstationarity. If it were necessary to apply the seasonal differencing operator in* (4.18) *D times to produce a series of length η- sD, the resulting series would be given by*

$$\nabla_s^D X_t^{(\lambda)} = (1 - L^s)^D X_t^{(\lambda)} \tag{4.19}$$

For purposes of explanation, consider once again a time series consisting of monthly observations. To model correlation among, say, March observations in the differenced series, one may wish to introduce appropriate model parameters. More specifically, to accomplish this task of linking March observations together one can use a model of the form

$$\Phi(L^s) \nabla_s^D X_t^{(\lambda)} = \Theta(L^s) \alpha_t \tag{4.20}$$

where $\Phi(L^s)$ and $\Theta(L^s)$ art the seasonal autoregressive (AR) and seasonal moving average (MA) operators, respectively, and α_t is a residual series which may conrain nonseasonal correlation both the seasonal AR and MA operators arc defined in order to describe relationships within the same season. In particular, the seasonal AR operator is defined as

$$\Phi(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - ... - \Phi_p L^{p_s}$$

where Φ_i is the ith AR parameter and P is the order of the AR operator. Because the power of each differencing operator is always an integer multiple of s. only the observations within each season an related to one another when using this operator. Hence, for the case of March observations in a monthly series, only the March observations are connected together using $\Phi(L^s)$. To describe the relationship of the residuals, a„ within a given season. the seasonal MA operafor is defined using

$$\Theta(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - ... - \Theta_q L^{q_s}$$

where Θ_i is the ith MA parameter and q is the order of the MA operator. Since the exponents of L in $\Theta(L^s)$ are always integer multiples of s, the residuals in the same season are linked with another when using $\Theta(L^s)$.

Theoretically one could define a separate model as in (4.20) for each season of the year. However, to keep the model as parsimonious as possible, one can assume that (4.20) *can be used for all of the seasons. Therefore, one is making the assumption that the correlation within all of the seasons is the same. For the case of monthly data this means that the relationship among all of the March observations is exactly the same as each of the other months.*

The error components or residuals, α_t, may contain nonseasonal nonstationarity which can be removed by using the nonseasonal diferencing operator defined as

$$\nabla^d \alpha_t = (1 - L^s)^d \alpha_t \tag{4.21}$$

where d is the order of the nonseasonal differencing operator which is selected just large enough to remove all of the nonseasonal nonstationarity. The sequence produced using (4.21) is theoretically a stationary nonseasonal series. The nonseasonal correlation can then be captured by writing the ARMA model as

$$\phi(L) \nabla^d \alpha_t = \theta(L) \alpha_t \tag{4.22}$$

where $\phi(L)$ is the nonseasonal AR operator of order p defined as

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$$

and $\theta(L)$ is the nonseasonal MA operator of order q written as

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q$$

The α_t's an the innovations which art identically independently distributed (IID) with a mean of zero and variance of σ_α^2. Hence $\alpha_t \sim$IID(0,σ_α^2). In order to obtain maximum likelihood estimates for the model parameters of a SARIMA model, in the next section the restriction of normality is also placed upon the α_t,'s so that they are assumed to be distributed as NID(0, σ_α^2).

Because of the form of (4.22), *the correlation among seasons is the same no matter what season one is dealing with. Hence. when entertaining monthly data, the correlation between. say, the March and February observations is defined to be the same as that between any other adjacent months such as October and September. For the periodic models, this restriction is dropped by allowing for a separate correlation smcture for each Season of the year.*

To define the overall seasonal model, one simply combines equations (4.22) and (4.20). This can be accomplished by solving for α_t, in (4.22) and substituting this result into (4.20) to obtain the SARIMA (seasonal autoregressive integrated moving average) model

$$\phi(L) \Phi(L^s) \nabla^d \nabla_s^D x_t^{(\lambda)} = \theta(L) \Theta(L^s) \alpha_t \tag{4.23}$$

Because the operators in (4.23) are multiplied together rather than summed, this model is often called a multiplicative SARIMA model.

When fitting the SARIMA model to a given time series of length η, one first transforms the data, if necessary, using the Box-Cox transformation in (4.17). Following this, the data can be differenced both seasonally and nonseasonally. It does not matter which differencing operation is carried out first. One then obtains the stationary series given by

$$\omega_t = \nabla^d \nabla_s^D x_t^{(\lambda)} \tag{4.24}$$

where the length of the ω_t, series is $\eta' = \eta - d - sD$. The seasonal and nonseasonal correlation in the ω_t, sequence is modelled by using the seasonal and nonseasonal AR and MA operators, respectively. Hence, ω_t is modelled by employing

$$\phi(L)\,\Phi(L^s)\,\omega_t = \theta(L)\,\Theta(L^s)\,\alpha_t \tag{4.25}$$

In some applications, ω_t may be a stationary seasonal series which is not obtained by differencing the original series. The model in (4.25) is called a seasonal ARMA or SARMA model of the ω_t series.

Notation

For a given application, one may first wish to indicate the key parameters included in a SARIMA model, without writing down all of the parameter estimates either in a table or else using the difference equation in (4.23). An economical notation for summarizing the strucm of the SARIMA model in (4.23) is (p,d,q)x(P,D,Q)$_s$. The first set of brackets contains the orders of the nonseasonal operators while the orders of the seasonal operators are listed inside the second set of brackets. More specifically, p, d and q stand for the orders of the nonseasonal AR. differencing and MA operators. respectively. In the second set of brackets, P, D and Q give the orders of the seasonal AR, differencing and MA operators, respectively. The subscript s appearing to the right of the second set of brackets points out the number of seasons per year.

For the case of monthly data for which s = 12, a specific example of a SARIMA model is $(2,1,1)\times(1,1,2)_{12}$ Suppose that the original series were transformed using natural logarithms. By utilizing (4.23), this model is written using a finite difference equation as

$$(1-\phi_1 L-\phi_2 L)\left(1-\Phi_1 L^{12}\right)(1-L)\left(1-L^{12}\right)\ln(x_t) = (1-\theta_1 L)\left(1-\Theta_1 L^{12}-\Theta_2 L^{24}\right)\alpha_t$$

If the data an stationary, nonseasonal or seasonal differencing is not required. A stationary model is indicated as (p,0,q)x(P,0,Q)$_s$.

The summary notation for a pun MA model is (0,d,q)x(0,D,Q)$_s$. When a model contains no MA parameters, the SARIMA model is written as (p,d,0)x(P,D,0)$_s$. Likewise, a nonstationary nonseasonal ARIMA model is denoted as ARIMA(p,d,q) rather than the more cumbersome notation given by SARIMA(p,d,q)x(O,O,O)$_s$.

4.4.5 Stationarity and Invertibility

For a nonseasonal model, the conditions of stationarity and invertibility are discussed in past Sections. Recall that for an ARMA model to be stationary the roots of the characteristics equation $\phi(L) = 0$ must lie outside the unit circle. Likewise, for invertibility the roots of $\theta(L) = 0$ must fall outside the unit circle. In addition to the aforesaid conditions, the properties of the seasonal AR and MA operators must be specified for the SARIMA model if it is to be fitted to the stationary ω_t, series in (4.25). For seasonal stationarity the roots of the characteristic equation $\Phi(L^s) = 0$ must lie outside the unit circle. Similarly, for seasonal invertibility, the roots of the characteristic equation $\Theta(L^s) = 0$ must fall outside the unit circle.

4.4.6 Autocorrelation function

The derivations of the theoretical ACF for nonseasonal AR, MA, and ARMA processes are presented in this cours, respectively. One could follow a similar approach to that used for the nonseasonal case to develop the formula for the theoretical ACF of the stationary seasonal process, ω_t, given in (4.25). Subsequent to this, the procedure developed for the nonseasonal cast can be used to obtain the theoretical ACF for the seasonal model. To use the theoretical results for the ACF developed for the nonseasonal ARMA model, simply replace the nonseasonal AR and MA operators by their combined counterparts for the seasonal model. The algorithm of McLeod (1975) can then be used to determine the theoretical ACF for the SARIMA process.

4.4.7 Model construction

Introduction

The most appropriate SARIMA model to fit to a given seasonal time series can be ascertained by following the identification, estimation, and diagnostic check stages of model construction. In the previous section, it is shown how the design of the ARIMA family of models is a straightforward extension of the nonseasonal models presented in past section. Likewise. as is explained in this section, the tools used for SARIMA model building are either the same or else closely related versions of the nonseasonal model construction.

Identification

Introduction The purpose of the identification stage is to determine the nonseasonal and seasonal differencing required to produce stationarity and also the orders of both the nonseasonal and the seasonal AR and MA operators for the ω_t, series in (4.25). Although each identification technique is discussed separately, in practical applications the output from all the techniques is interpreted and compared together in order to design the type of model to be estimated.

For some types of seasonal time series, it is known in advance whether or not the data sets should be transformed using the Box-Cox transformation in (4.17). For instance, average monthly riverflow series often require a natural logarithmic transformation to cause the residuals of the fitted models to be approximately normally distributed and homoscedastic. In many applications, analysts may not realize that Box-Cox transformations are needed until after the model parameters have been estimated and the statistical properties of the residuals are examined. The analysts should keep in mind that usually a Box-Cox transformation does not change the design of the AR and MA operators needed in the model or models to fit to the transformed time series. However, this is not true in general, and as is pointed out by Granger and Newbold [1976], certain transformations can change the type of model to estimate for a given time series.

Therefore, even though it is usually not necessary to perform the identification stage for the transformed data if it has already been done for the corresponding untransformed series, a practitioner should be aware that in some instances this may not be the case. When a transformation does change the type of model to be used, diagnostic checks would detect this fact and then the design of the model to fit to the transformed data can be properly identified.

For a SARIMA model application there should be at least seven years of seasonal data and also at least 50 data points overall in order to get reasonable MLE's (maximum likelihood estimates) for the model parameters. If one were analyzing a monthly series, one would require at least 12x7 = 84 observations. Therefore. one should proceed with the identification stage only if the minimum requid amount of information is present.

Tools When examining a specified time series analysis for the first time, one may wish to utilize the exploratory data analysis tools. The purpose of exploratory data analysis is to discover the basic statistical characteristics of a data set by examining simple graphical and numerical output. Subsequent to obtaining a general understanding of the statistical properties of the time series, one may wish to design a specific SARIMA model to fit to the series by studying the graphical output from the following techniques.

1. **Plot of the Original Series** - A graph of the observations in the series against time is an important exploratory data analysis method that should always be used in model identification. Characteristics of the data which are usually easily uncovered from a perusal of a time series plot include seasonality, nonstationarity due to ands in the mean levels of the seasons or years, changing variance, extreme values, comlation or dependence among observations, and long term cycles. The nonstationarity present in the data can often be removed using seasonal and/or nonseasonal differencing. The seasonal and nonseasonal comlation in the time series can be modelled by appropriately deciding upon which AR and MA operators should be included in the SARIMA model. Graphs of the next four functions can

be used for specifically designing the components nded in the SARIMA modeL

2. **ACF (autocorrelation function)** - The theoretical ACF measures the amount of linear dependence between observations in a time series that are separated by k time lags. The sample estimate, r_k for ρ_k, To use the sample ACF in model identification, calculate and then plot r_k, up to a maximum lag of roughly $\frac{\eta}{4}$ along with the approximate 95% confidence limits. The graph of the sample ACF and the other three graphs described below should include at least 2s or 3s lags, where s is the number of seasons per year. In this way, the cyclic behaviour caused by seasonality and any decaying or truncation properties of r_k over k, can be visually detected.

   The first step is to examine a plot of the ACF to detect the presence of nonstationarity in the given series. For seasonal data with the seasonal length equal to s. the ACF often follows a wave pattern with peaks at s, 2s, 3s, and other integer multiples of s. As is shown by Box and Jenkins (1976. pp. 174175). if the estimated ACF at lags that are integer multiples of the seasonal length s do not die out rapidly, this may indicate that seasonal differencing is needed to produce stationarity. Failure of other ACF estimates to damp out may imply that nonseasonal differencing is also required. If the length of the original series is η, the number of data points in the differenced series would be η' = η - d + sD. Li (1991) develops some statistical tests for determining the orders of differencing required for a seasonal time series.

   If the stationary ω_t, series is not white noise, one can use the sample ACF to help decide upon which AR and MA parameters are needed in the SARIMA model. When the process is a pure MA(0,d,q)x(0,D,Q)$_s$, model, the sample ACF truncates and is not significantly different from zero after lag q + sQ. For this case, the variance of r_k after lag q + sQ is (Bartlett, 1946)

$$Var\left[r_k\right] = \frac{1}{\eta'}\left(1 + 2\sum_{i=1}^{q+sD} r_i^2\right), \qquad k > q + sQ \tag{4.26}$$

   where η' stands for the length of the ω_t series after differencing.

   If r_k attenuates at lags that are multiples of s, this implies the presence of a seasonal AR component. The failure of the ACF to truncate at other lags may imply that a nonseasonal AR term is required.

   As defined in (4.24). the stationary ω_t, series created by differencing either the original or transformed series is given as

$$\omega_t = \nabla^d \nabla_s^D x_t^{(\lambda)}$$

   when the exponent λ indicates that the original x_t series may be transformed using the Box-Cox transformations in (4.17). After the data have

been differenced just enough times to produce both seasonal and nonseasonal stationarity, then check the ACF of the ω_t series to determine the number of AR and MA parameters required in the model. The ω_t series is also used at the other steps of the identification procedure. Of course, if no differencing is required. the ω_t series is simply the $x_t^{(\lambda)}$ series. As noted earlier, the graph of the sample ACF for w, should include at least 2s or 3s lags.

If a series is white noise, then r_k is approximately NID$(0, \frac{1}{\eta'})$. This result allows one to test whether a given series is white noise by checking to see if the ACF estimates are significantly different from zero, Simply plot confidence limits on the ACF diagram and see if a significant number of r_k values fall outside the chosen confidence interval.

3. **PACF (partial autocorrelation function)** - After writing the SARIMA in(4.25), the theoretical PACF is defined for the ω_t series by using the Yule-Walker equations. the sample PACF can be estimated. For model identification, simply calculate and plot the sample PACF to at least lag 2s along with the 95% confidence limits. the sample PACF can be utilized for deciding upon which AR and MA parameters are needed for properly representing the data.

   When the process is a pure AR(p,d,0)x(P,D,0), model, the sample PACF cuts off and is not significantly different from zero after lag p + sP. After lag p + sP, the sample PACF is approximately NID$(0, \frac{1}{\eta'})$.

   If the sample PACF damps out at lags that are multiples of s, this suggests the incorporation of a seasonal MA component into the model. The failure of the sample PACF to truncate at other lags may imply that a nonseasonal MA term is required,

4. **IACF (inverse autocorrelation function)** Theoretically, the IACF of the ω_t series is defined to be the ACF of the (q,d,p)x(Q,D, P), process that is written as

$$\theta(L)\,\Theta(L^s)\,\omega_t = \phi(L)\,\Phi(L^s)\,\alpha_t \tag{4.27}$$

   The model in (4.27) is called the dual model while the SARIMA(p,d,q)x(P,D,Q), model in (4.23) or (4.25) is refed to as the primal model (Mcleod, 1984). the sample IACF is plotted up to a lag of at least 2s or 3s or not more than $\frac{\eta'}{4}$. If the ω_t, series is white noise, the sample IACF is approximately NID$(0, \frac{1}{\eta'})$. For the case of white noise, the values of the sample IACF should not fall outside the 95% confidence limits off $\pm\frac{1.96}{\sqrt{\eta'}}$ more than once in twenty lags

5. **LPACF (inverse partial autocorrelation tunction)** - The theoretical IPACF originally defined by Hipel et al. (1977). For model identification, the sample IPACF and its 95% confidence limits are plotted up to a lag

of at least 2s or 3s. If the ω_t series is white noise, then the values of the sample IPACF should not be significantly different from zero and should fall within the 95% confidence limits.

6. **Cumulative periodogram white noise test** - As was mentioned previously. the sample ACF plot is an accepted means of checking whether the given data are white noise. The sample PACF, IACF, and IPACF can also be employed in this capacity. However, the cumulative periodogram provides another means of checking for white noise.

   In addition to verifying whether a series is uncorrelated, the cumulative periodogram can also detects certain types of correlation. In particular, it is an effective procedure for finding hidden periodicities.

4.4.8 Estimation

Introduction

Often, identification methods cannot clearly determine which is the single best SARIMA model to fit to the time series under study. Rather. anywhere from one to four model designs may be tentatively identified At the estimation stage, MLE's can then be obtained for the parameters in each of the models. Subsequently, discrimination methods can be used for selecting the best model from the set of calibrated models. The techniques for choosing the best model include the AIC discussed in this section. If none of the fitted models adequately describes the data, appropriate design modifications can be made before estimating the parameters of the most recent iteration and repeating the above procedure until a suitable model is found.

Algorithms

Two algorithms are discussed in this section for obtaining approximate MLE's for the parameters of the SARMA model fitted to the w, series in (4.25). Besides using the basic definition of the SARMA model, both methods are based on the assumption that the innovations are normally independently distributed with a mean of zero and variance of σ_α^2.

The first approach is to use the algorithm developed for the nonseasonal ARMA model while the second one is to employ a more computationally efficient procedure which takes into account the specific mathematical structure of the SARMA model. Because the orders of nonseasonal and seasonal differencing operators required to produce the stationary ω_t, series in (4.25) are not estimated but selected based upon identification results, only the parameters included in the SARMA model for fitting to ω_t, have to be estimated.

the first estimation method works well when the number of seasons per year is not more than 12. However, for bimonthly data and weekly series for which s = 24 and 52. nspectively, the estimation becomes computationally inefficient. To overcome this problem, the maximum likelihood approach of McLeod and

Salas (1983) can be used. This estimation method, which is based upon the modified sum of squares method of McLeod (1977). is designed according to the multiplicative structure of the AR and MA operators in the SARIMA model. It works for yearly (s = 1). monthly (s = 12). weekly (s = 52), daily (s = 365) as well as any other types of seasonal series.

Subsequent to estimating the model parameters for the models separately fitted to the time series under study, one cm calculate the value of the AIC for each model in order to select the model which has the minimum AIC value.

One approach is to carry out an exhaustive AIC study by fitting a large range of SARIMA models to the time Series and then picking the one having the minimum AIC. In the second main approach, the identification techniques of ARIMA section are used to select a handful of models for which the parameters and AIC values am estimated. Once again, one selects the model having the minimum AIC value.

The general formula for the AIC is given by

$$AIC = -2\ln(ML) + 2k$$

where ML denotes the maximized value of the likelihood function and k is the number of independently adjusted parameters in the model. Approximate formulae can be devised for determining the AIC for a SARIMA model which contains differencing operators. Because the amount of data has been reduced from a total of η to $\eta' = \eta - d - sD$ points when there is both nonseasonal and seasonal differencing. Hence, the AIC for a SARIMA model can be roughly calculated as

$$AIC = \frac{\eta}{\eta'}(-2\ln(ML)) + 2k \tag{4.28}$$

The total number of model parameters is $k = p + q + P + Q + 1$, where the unity term allows for the estimate of the variance of the model residuals. Usually, the mean of the differenced series can be assumed to be zero. However, if the mean of the differenced series is also estimated, k must be increased by unity. Also, k is increased by one if $\lambda \neq 1$.

Tests for Whiteness

If a calibrated SARIMA model adequately describes a time series, the estimated innovations, $\widehat{\alpha}_t$. or residuals should be white, due to the independence assumption of the $\widehat{\alpha}_t$. To determine whether the residuals are white noise, the best procedure is to examine the residual autocorrelation function (RACF).

Portmanteau tests In order to test that the estimated residues follow a white noise, the hypothalate is tested. In practice, two tests are used

1- Box-Pierce test : The purpose of this test is to test the non autocorrelated character of the residues.

2- Ljung and Box test : This test is to be applied, preferably to the Box-Pierce test, when the sample is small, the Ljung-Box statistic, given by:

$$LB = \eta'(\eta' + 2) \sum_{k=1}^{L} \tau_k^2 \left(\widehat{\alpha}\right) / \left(\eta' - k\right)$$

This statistic is χ^2 disbibuted on (L - P -p - Q - q) degrees of freedom.

Test for Periodic Correlation To test whether or not periodic correlation is contained in the residuals of a fitted SARIMA model, one can employ a statistical test presented by McLeod (1993). The periodic autocorrelation at lag k for season m may be written as

$$r_k^{(m)}\left(\widehat{\alpha}_{r,m}\right) = \frac{\sum_{r=1}^{n} \widehat{\alpha}_{r,m}\widehat{\alpha}_{r,m-k}}{\sqrt{\sum_{r=1}^{n} \widehat{\alpha}_{r,m}^2 \sum_{r=1}^{n} \widehat{\alpha}_{r,m+k}^2}}$$

where $\widehat{\alpha}_{r,m}$is the estimated innovation or residual for the rth year and rnth season. n is the number of years of seasonal data, and s is the number of seasons per year. Over the s seasons, the residual autocomlations at lag one given by $r_1^{(m)}\left(\widehat{\alpha}_{r,m}\right)$, m = 1.2, . . . , s, are approximately jointly normally distributed with mean zero, diagonal covariance matrix, and $var(r_1^{(m)}\left(\widehat{\alpha}_{r,m}\right)) = n^{-I}$. A diagnostic check for detecting periodic autocorrelation in the residuals of a fitted SARIMA model is given by

$$S = n \sum_{m=1}^{s} \left(r_1^{(m)}\left(\widehat{\alpha}_{r,m}\right)\right)^2$$

which should be approximately χ^2 distributed on s degrees of freedom. if the model is adequate. When the calculated value for S is larger than that found in the tables for a given sigruficance level, the calibrated model does not capture the periodic correlation.

Normality Tests Many standard tests are available to check whether data are normally distributed. Additionally, the graph of the cumulative distribution of the residua should appear as a straight line when plotted on normality paper if the residuals are normally distributed. For instance, the residuals should not be sigruficantly skewed or prosess a significantly large kurtosis coefficient under the assumption that the residuals are normally distributed.

Homoscedasticity Checks Heterodsticity or changes in variance can arise in a number of different ways including:

1- the variance changes over time,

2- the magnitude of the variance is a function of the current level of the series.

By following the identification, estimation and diagnostic checks stages of model construction, one can conveniently determine a reasonable SARIMA model for describing a the series. These three construction stages are summarized in next Figure

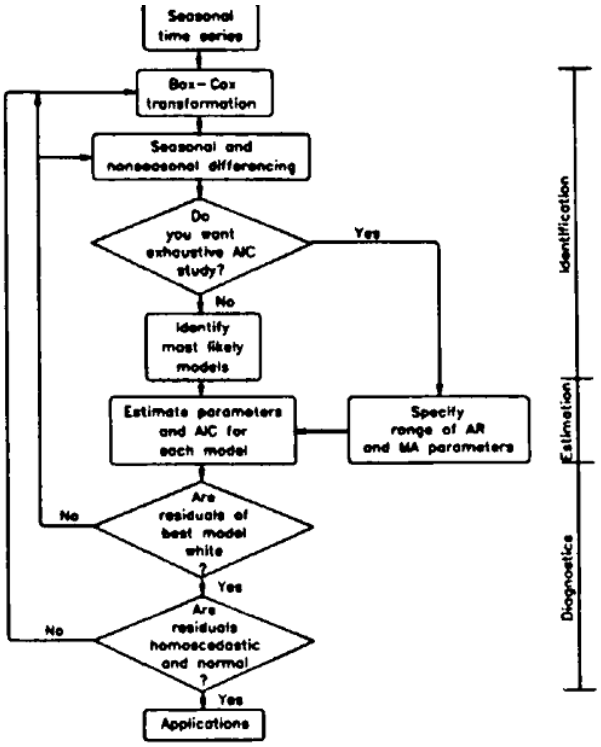


figure 15. *constructing a SARIMA* model

4.4.9 Exemple of SARIMA

AirPassengers example
fit the classic airline model using arima()
ap.arima <- arima(log(AirPassengers), order = c(0,1,1), seasonal = c(0,1,1))
same model using two equivalent ways to specify it
ap.baseA <- sarima(log(AirPassengers) ~0 $|ma(1, c(-0.3)) + sma(12, 1, c(-0.1)) + i(1) + si(12, 1), ss.method = "base"$)
ap.baseB <- sarima(log(AirPassengers) ~0 $|ma(1, c(-0.3)) + sma(12, 1, c(-0.1)) + i(2) + s(12)$,
ss.method = "base")
ap.baseA
summary(ap.baseA)

```
ap.baseB
summary(ap.baseB)
## as above, but drop 1-B from the model:
ap2.arima <- arima(log(AirPassengers), order = c(0,0,1), seasonal = c(0,1,1))
ap2.baseA <- sarima(log(AirPassengers) ~0 |ma(1, c(−0.3))+sma(12, 1, c(−0.1))+
si(12, 1), ss.method = "base")
ap2.baseB <- sarima(log(AirPassengers) ~0 |ma(1, c(−0.3))+sma(12, 1, c(−0.1))+
i(1) + s(12), ss.method = "base")
## for illustration, here the non-stationary part is
## (1-B)^2(1+B+...+B^5) = (1-B)(1-B^6)
## ( compare to (1-B)(1-B^{12}) = (1-B)(1-B^6)(1+B^6) )
ap3.base <- sarima(log(AirPassengers) ~0 |ma(1, c(−0.3))+sma(12, 1, c(−0.1))+
i(2) + s(6), ss.method = "base")
## further unit roots, equivalent specifications for the airline model
tmp.su <- sarima(log(AirPassengers) ~0 |ma(1, c(−0.3))+sma(12, 1, c(−0.1))+
i(2) + s(2) + su(12, 1 : 5), ss.method = "base")
tmp.su$interna$delta_poly
prod(tmp.su$interna$delta_poly)
zapsmall(coef(prod(tmp.su$interna$delta_poly)))
tmp.su
tmp.u <- sarima(log(AirPassengers) ~0 |ma(1, c(−0.3))+sma(12, 1, c(−0.1))+
i(2) + s(2) + u((1 : 5)/12),
ss.method = "base")
tmp.u
```

Exercise 57 *Let the process ARMA(1,1) defined by :*

$$X_t = \varphi_1 X_{t-1} + \epsilon_t - \theta_1 \epsilon_{t-1}.$$

with $\epsilon_t \sim \quad WN(0, \sigma^2)$

1. Is this process stationary?

2. Calculate the variance and autocovariance in order 1.

3. Give the forecast on horizon 1.

Exercise 58 *Let* (Y_t) *be the ARMA plus noise time series defined by*

$$Y_t = X_t + W_t,$$

where $(W_t)_t \sim WN(0, \sigma_w^2)$ *,* (X_t) *is the ARMA(p,q) process satisfying*

$$\phi(L) X_t = \theta(L) Z_t \quad , \quad Z_t \sim WN(0, \sigma_z^2)$$

and $E(W_s Z_t) = 0$ *for all s and t .*

1.Show that (Y_t) *is stationary and find its autocovariance in terms of* σ_w^2 *and the ACVF of* (X_t)*.*

2.Show that the process $U_t := \phi(L) Y_t$ *is r-correlated, where r max(p,q) and hence,is an MA(r) process. Conclude that* (Y_t) *is an ARMA(p,r) process.*

Exercise 59 *1.Find the seasonal coefficients for a quarterly series that follows the model* $z_t = 10 + cos(\pi t/2 + \pi/8) + \epsilon_t$

2. Prove that the series can be modelled using $\triangle^4 z_t = (1 - L^4)\epsilon_t$.

Exercise 60 *Find the theoretical autocorrelation function of the process* $(1 - 0.4L)w_t = (1 + 0.5L^{12})$.

Exercise 61 *Find the theoretical autocorrelation function of the process* $(1 - 0.4L)(1 - 0.8L^{12})w_t = \epsilon_t$.

Exercise 62 *Find the theoretical autocorrelation function of the process* $w_t = (1 - \theta L)(1 - \Theta L^{12})\epsilon_t$ *and compare it with that of the non-multiplicative process* $w_t = (1 - \theta L - \Theta L^{12})\epsilon_t$.

Chapter 5

ARCH and GARCH models

5.1 Introduction

Traditional time series models assume that the variance is constant, meaning the statistical dispersion remains unchanged across different time periods. However, it is not uncommon for empirical time series to exhibit volatility clusters. In other words, when the volatility is high it is likely to remain high and when it is low and it is likely to remain lowClearly, the homogeneity of variance assumption is violated in such situations. Therefore, it is fundamentally inaccurate to fit the statistical models that assume constant variance over time. To address this issue, especially in the case of modeling financial volatility, Engle (1982) introduced the Autoregressive Conditional Heteroscedastic (ARCH), model. The ARCH model explains the future volatility based on current observables. The ARCH process describes the variability as a weighted average of previously estimated squared errors from historical data, and these weights provide more influence to the recent information and less to that of the distant past. Furthermore, an ARCH process can handle a higher number of extreme values than what is expected from a standard normal distribution, hence it is more applicable during high volatility periods.Tim Bollerslev (1986) expanded on the ARCH model and proposed a class of heteroscedastic models called the Generalized Autoregressive Conditional Heteroscedastic (GARCH) models. Subsequently, the GARCH formulation was rapidly expanded to include what is known as the GARCH family of models. The GARCH model constitutes of three variance components: a constant variance portraying the long-run average, the variance forecast from the previous period, and the variance arising from the new information. The weights of the last two forecasts govern the fluctuations to the long-run average returns due to the arrival of new information and the volatility observed in the immediate past.

The ARCH and the GARCH family models became popular among researchers due to their versatility in modeling financial data. Motivated by the fact that the ARCH and the GARCH models are useful tools to explain the real-world phenomena and successfully forecast future volatiles, researchers added variations to the standard ARCH and GARCH models.

5.2 The autoregressive conditional heteroscedastic (ARCH)

The Autoregressive Conditional Heteroscedastic (ARCH) formulation was first proposed by Engle (1982) to model the time dependent variance of a time series. The ARCH process is used to model the conditional variances which is also referred to as conditional volatility, and it is expressed as a linear function of the squared errors. The large squared returns may signal a relatively high volatile period while series of small squared returns may signal a relatively low volatile period.

5.2.1 The ARCH (p) Model presentation

Definition 63 *Let ϵ_t be the real valued discrete-time stochastic process, and F_{t-1} denotes the sigma field generated from the information set up to time t, Then the ARCH model of orderp is formulated as:*

$$\begin{aligned} \epsilon_t &= \sigma_t z_t; \qquad z_t \rightsquigarrow i.i.d.\ N(0,1) \qquad (5.1)\\ \epsilon_t \,|F_{t-1} &\rightsquigarrow N\left(0, \sigma_t^2\right), \\ Var\left(\epsilon_t \,|F_{t-1}\right) &= E\left(\epsilon_t^2 \,|F_{t-1}\right) - \left[E\left(\epsilon_t \,|F_{t-1}\right)\right]^2 = E\left(\epsilon_t^2 \,|F_{t-1}\right) = \sigma_t^2, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2, \\ \alpha_0 &> 0, \alpha_i > 0, i = 1, ..., p. \end{aligned}$$

When p = 1, the ARCH (p) model can be rewritten as the ARCH (1) process. Sometimes it is important to rearrange the ARCH (1) model as an AR (1) process. To do so, define the serially uncorrelated zero-mean stochastic process $\{\eta_t\}$such that: $\eta_t = \epsilon_t^2 - \sigma_t^2$. After replacing σ_t^2 in equation (5.1) by $\epsilon_t^2 - \eta_t$, the ARCH process of order 1 can also be specified as the AR (1) model: $\epsilon_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \eta_t$. Here,

$$E\left[E\left(\epsilon_t^2 \,|F_{t-1}\right)\right] = E\left(\sigma_t^2\right) = \frac{\alpha_0}{1-\alpha_1} > 0 \quad and \quad 0 \leq \alpha_1 < 1.$$

In a separate study done by Ling and McAleer (2002), the condition $0 \leq \alpha_1 < 1$ is shown to be the necessary and sufficient condition for the weak stationarity

of an ARCH (1) process. A main application of the ARCH model is to forecast future conditional variances. Assume σ_{t+1}^2 is the one step ahead conditional variance in the ARCH (1) model. Then one can write:

$$\sigma_{t+1}^2 = E\left(\epsilon_t^2 \left| F_{t-1} \right.\right) = \alpha_0 + \alpha_1 \sigma_t^2.$$

5.2.2 Parameter Estimation of ARCH Models.

Engle (1982) stated that the parameters can be estimated by the Maximum Likelihood Estimation (MLE) method. Let Θ be the parameter vector such that $\Theta = (\alpha_0,\alpha_1...,\alpha_p)$ and p is the order of the ARCH model. Then under the normality assumption, the likelihood function for the ARCH (p) is formulated as follows:

$$f\left(\epsilon_t, ..., \epsilon_n \left| \Theta \right.\right) = f\left(\epsilon_t, ..., \epsilon_p \left| \Theta \right.\right) * \prod_{t=p+1}^{n} f\left(\epsilon_t \left| \epsilon_1, ..., \epsilon_{t-1}, \Theta \right.\right).$$

Here, $f\left(\epsilon_t, ..., \epsilon_n \left| \Theta \right.\right)$ is the joint pdf function of $\{\epsilon_t\}_{t=1}^{p}$, where conditional pdf of ϵ_t given sigma field generated by all the information set up to time t-1, F_{t-1} is:

$$\begin{aligned} f\left(\epsilon_t \left| F_{t-1} \right.\right) &= f\left(\epsilon_t \left| \epsilon_1, ..., \epsilon_{t-1}, \Theta \right.\right) \\ &= \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right), \end{aligned}$$

where, σ_t^2 is defined as given in the equation (5.1). As pointed out by Engle (1982), the exact form of the joint pdf of f $f\left(\epsilon_t, ..., \epsilon_p \left| \Theta \right.\right)$ is complicated, and therefore it is replaced

by a joint distribution obtained by conditioning on the firstp observations. The resulting conditional likelihood function of ARCH (p) can be written as:

$$\prod_{t=p+1}^{n} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\epsilon_t^2}{2\sigma_t^2}\right)$$

Finally, the conditional log likelihood function can be presented as:

$$l\left(\Theta \left| \{\epsilon_t\}_{t=p+1}^{n} \right.\right) = -\sum_{t=p+1}^{n} \left[\frac{1}{2}\log\left(2\pi\right) + \frac{1}{2}\log\left(\sigma_t^2\right) + \frac{\epsilon_t^2}{\sigma_t^2}\right].$$

Engle (1982) also proved that the parameters are asymptotically independent hence, they can be maximized separately using available numerical optimization methods.

figure 16, shows the simulated ARCH (1) time series plot of size n =500, with the parameter vector 0 = (0.10,0.90). The simulated data exhibited volatility clustering, which means larger (smaller) changes in volatility are followed by

larger (smaller) changes in volatility and they group together. Since the order of the ARCH process is 1, it has the ability to adjust quickly after a large fluctuation in volatility.

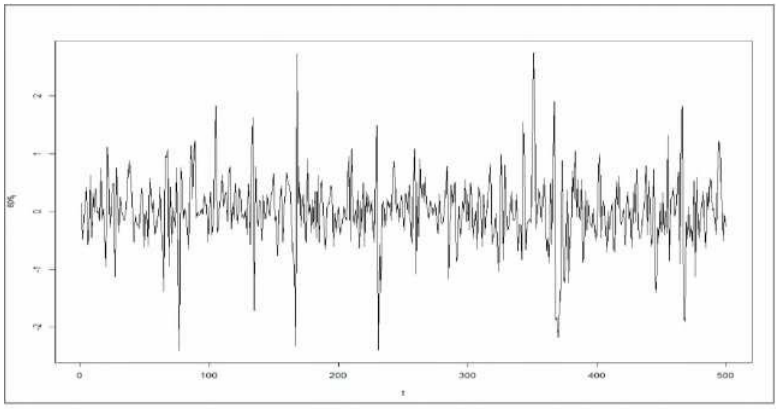


figure16 simulated ARCH(1) time serieswith Θ=(0,10. 0,90)

Figure 17 exhibits the Autocorrelation (ACF) and the Partial ACF (PACF) of the simulated ARCH (1) time series, mentioned above. According to the ACF and the PACF plots, only lag 8 had a mildly significant result. Except for the lag 8 all the other lags in the ACF and the PACF plots, failed to show significant correlations. Therefore, the simulated data are serially uncorrelated for all practical purposes.

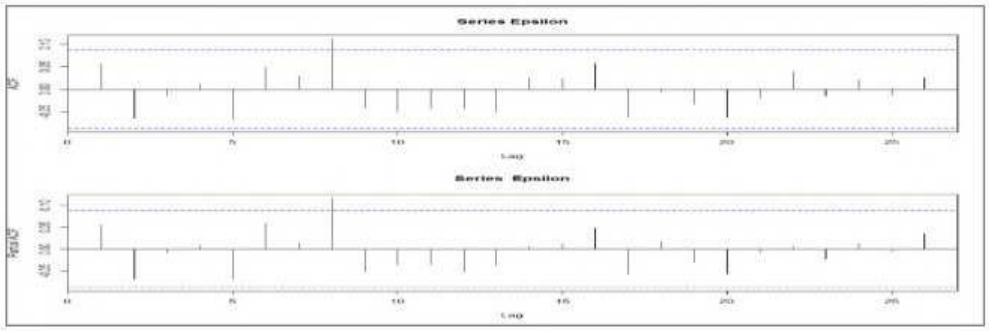


Figure17.Sample ACF and PACF plots of simulatedARCH(1)process data withΘ=(0,10. 0,90)

5.2.3 Leptokurtosis and ARCH models

Kurtosis is the scaled fourth moment of a probability distribution and is defined (for a random variable x_t with mean μ) by

$$\kappa(x_t) = \frac{E(x_t - \mu)^4}{\left[E(x_t - \mu)^2\right]^2} = \frac{\mu_4}{\mu_2^2}$$

Kurtosis measures the fatness of the tails of a distribution, which is the probability of 'outliers'. For the normal distribution, $\kappa(x_t) = 3$. Distributions with fatter tails than the normal, like the t-distribution, have $\kappa(x_t) > 3$ and are said to be leptokurtic. Distributions with thinner tails than the normal have $\kappa(x_t) < 3$ and are said to be platykurtic. In empirical investigation of financial data series, many series show evidence of leptokurtosis. ARCH model disturbances have the property of being leptokurtic. For the case of the ARCH(1) model, the kurtosis of the unconditional distribution of ut is given by

$$\kappa(x_t) = 3\frac{(1-\alpha_1^2)}{1-3\alpha_1^2} > 3$$

for $\alpha_1 > 0$ and $3\alpha_1^2 < 1$ (otherwise the kurtosis is not finite). Thus ARCH models exhibit fatter tails than the normal distribution although this may not be enough fully to account for the kurtosis observed in real financial data series.

5.2.4 Testing an ARCH model

After a model is selected and estimated, it is generally desirable to test whether it adequately represents the data. A useful array of tests can readily be constructed from calculating Lagrange multiplier tests against particular parametric alternatives. Since almost any moment condition can be formulated as the score against some alternative, these tests may also be interpreted as conditional moment tests; see Newey (1985) and Tauchen (1985). Whenever one computes a collection of test statistics, the question of the appropriate size of the full procedure arises. It is generally impossible to control precisely the size of a procedure when there are many correlated test statistics and conventional econometric practice does not require this. When these tests are viewed as diagnostic tests, they are simply aids in the model building process and may well be part of a sequential testing procedure anyway. In this section, we will show how to develop tests against a variety of interesting alternatives to any particular model.

Let an ARCH model such that $\epsilon_t = z_t.\sigma_t$ with $z_t \rightsquigarrow N(0,1)$ and $\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 = \alpha_0 + \alpha(L)\epsilon_t^2$. Let the nested hypothesis H_0, H_0: $\alpha_1 = ... = \alpha_p = 0$; and against the alternative hypothesis H1: α_i not all null.

• If H_0 is accepted, the variance of the error is constant $\sigma_t^2 = \alpha_0$.Otherwise the terms of the error follow an ARCH whose order p is to be determined.

The test is based either on a classic Fisher test or on the Lagrange multiplier test LM=n x R^2 with:n=number of observations used to calculate the regression,

R^2=coefficient of determination

• If LM>χ^2(p) at p degrees of freedom read from the table at a set threshold (typically 0.05), H_0 is rejected; the process is considered justifiable from an

ARCH(p) model. Another approach is to calculate the residuals correogram to the squares of the initial model. If terms from this correlogram are significantly different from 0, so one can conclude to an ARCH type specification.

5.3 The Generalized autoregessive conditional heteroscedastic model (GARCH)

While conventional time series and econometric models operate under an assumption of constant variance, the ARCH (Autoregressive Conditional Heteroskedastic) process introduced in Engle (1982) allows the conditional variance to change over time as a function bf past errors leaving the unconditional variance constant. This type of model behavior has already proven useful in modelling several different economic phenomena. In Engle (1982), Engle (1983) and Engle and Kraft (1983), models for the inflation rate are constructed recognizing that the uncertainty of inflation tends to change over time. In Coulson and Robins (1985) the estimated inflation volatility is related to some key macroeconomic variables.

Models for the term structure using an estimate of the conditional variance as a proxy for the risk premium are given in Engle, Lilien and Robins (1985).

The same idea is applied to the foreign exchange market in Domowitz and Hakkio (1985). In Weiss (1984) ARMA models with ARCH errors are found to be successful in modelling thirteen different U.S. macroeconomic time series. Common to most of the above applications however, is the introduction of a rather arbitrary linear declining lag structure in the conditional variance equation to take account of the long memory typically found in empirical work, since estimating a totally free lag distribution often will lead to violation of the non-negativity constraints. In this section a new, more general class of processes, GARCH (Generalized Autoregressive Conditional Heteroskedastic), is introduced, allowing for a much more flexible lag structure.

The extension of the ARCH process to the GARCH process bears much resemblance to the extension of the standard time series AR process to the general ARMA process and, as is argued below, permits a more parsimonious description in many situations.

5.3.1 The GARCH(p,q) process

The ARCH process introduced by Engle (1982) explicitly recognizes the difference between the unconditional and the conditional variance allowing the latter to change over time as a function of past errors. The statistical properties of this new parametric class of models has been studied further in Weiss (1982) and Milhoj (1984).

In empirical applications of the ARCH model a relatively long lag in the conditional variance equation is often called for, and to avoid problems with negative variance parameter estimates a fixed lag structure is typically imposed, cf. Engle (1982), Engle (1983) and Engle and Kraft (1983). In this light it seems

of immediate practical interest to extend the ARCH class of model to allow for both a longer memory and a more flexible lag structure.

Let ϵ_t denote a real-valued discrete-time stochastic process, and ψ_t, the information set (σ-field) of all information through time t. The GARCH(p, q) process (Generalized Autoregressive Conditional Heteroskedasticity) is then given by

$$\begin{aligned} \epsilon_t \left| \psi_{t-1} \right. &\sim N\left(0, h_t\right), \qquad (5.2) \\ h_t &= \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \\ &= \alpha_0 + A\left(L\right) \epsilon_t^2 + B\left(L\right) h_i, \end{aligned}$$

where

$$\begin{array}{rcl} p \geq 0, & & q > 0 \\ \alpha_0 > 0, & \alpha_i \geq 0, & i = 1, ..., q, \\ \beta_i \geq 0, & & i = 1, ..., p. \end{array}$$

For p = 0 the process reduces to the ARCH(q) process, and for p = q = 0 ϵ_t is simply white noise. In the ARCH(q) process the conditional variance is specified as a linear function of past sample variances only, whereas the GARCH(p, q) process allows lagged conditional variances to enter as well.

This corresponds to some sort of adaptive learning mechanism. The GARCH(p,q) regression model is obtained by letting the ϵ_t be innovations in a linear regression,

$$\epsilon_t = y_t - x_t' b,$$

where y_t is the dependent variable, x, a vector of explanatory variables, and b a vector of unknown parameters.

If all the roots of $1 - B(z) = 0$ lie outside the unit circle, (5.2) can be rewritten as a distributed lag of past ϵ_t^2's,

$$\begin{aligned} h_t &= \alpha_0 \left(1 - B\left(1\right)\right)^{-1} + A\left(L\right) \left(1 - B\left(L\right)\right)^{-1} \epsilon_t^2 \\ &= \alpha_0 \left(1 - \sum_{i=1}^{p} \beta_i\right)^{-1} + \sum_{i=1}^{\infty} \delta_i \epsilon_{t-i}^2, \end{aligned}$$

which together with (5.2) may be seen as an infinite-dimensional ARCH(∞) process. The δ_i's are found from the power series expansion of

$$\begin{aligned} D(L) &= A(L)\left(1 - B(L)\right)^{-1}, \qquad (5.3) \\ \delta_i &= \alpha_i + \sum_{j=1}^{n} \beta_j \delta_{i-j}, \qquad i = 1, ..., q, \\ &= \sum_{j=1}^{n} \beta_j \delta_{i-j}, \qquad i = q+1, ... \end{aligned}$$

where $n = min(p, i-1\}$. It follows, that if $B(1) < 1$, δ_i will be decreasing for i greater than $m = max\{p, q\}$. Thus if $D(1) < 1$, the GARCH(p, q) process can be approximated to any degree of accuracy by a stationary ARCH(q) for a sufficiently large value of q. But as in the ARMA analogue, the GARCH process might possibly be justified through a Wald's decomposition type of arguments as a more parsimonious description.

From the theory on finite-dimensional ARCH(q) processes it is to be expected that D(1) $<$ 1, or equivalentlyA(1) + B(1) $<$ 1, suffices for wide-sense stationarity; cf. Milhoj (1984). This is indeed the case.

Theorem 64 *The GARCH(p, q) process as defined in (5.2) is wide-sense stationary with :*

$$E(\epsilon_t) = 0, var(\epsilon_t) = \alpha_0(1 - A(1) - B(1))^{-1}$$

and

$$cov(\epsilon_t, \epsilon_s) = 0$$

for $t \neq s$ if and only /f

$$A(1) + B(1) < 1.$$

As pointed out by Sastry Pantula and an anonymous referee, an equivalent representation of the GARCH(p, q) process is given by

$$\epsilon_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i^2 \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \epsilon_{t-j}^2 - \sum_{j=1}^{p} \beta_j \nu_{t-j} + \nu_t, \qquad (5.4)$$

and

$$\nu_t = \epsilon_t^2 - h_t = \left(\eta_t^2 - 1\right) h_t, \qquad (5.5)$$

where

$$\eta_t \overset{i.i.d}{\sim} N(0,1).$$

Note, by definition ν_t is serially uncorrelated with mean zero. Therefore, the GARCH(p, q) process can be interpreted as an autoregressive moving average process in ϵ_t^2 of orders $m = max(p,q\}$ and p, respectively.

Although a parametcrization along the lines of (5.4) might be more meaningful from a theoretical time series" point of view, (5.2) is easier to work with in practice.

5.3.2 The GARCH(1, 1) process

The simplest but often very useful GARCH process is of course the GARCH(1,1) process given by (5.2) and

$$h_t = \alpha_0 \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1}, \quad \alpha_0 > 0, \qquad \alpha_1 \geq 0, \quad \beta_1 \geq 0. \tag{5.6}$$

From Theorem 1, $\alpha_1 + \beta_1 < 1$ suffices for wide-sense stationarity, and in general we have:

Theorem 65 *For the GARCH(1, 1) process given by (5.2) and (5.4) a necessary and sufficient condition for existence of the 2ruth moment is*

$$\mu(\alpha_1, \beta_1, m) = \sum_{j=0}^{m} \binom{m}{j} a_j \alpha_1^j \beta_1^{m-j} < 1, \tag{5.7}$$

where

$$a_0 = 1 \quad \text{and } a_j = \prod_{i=1}^{j} (2j-1), \qquad j = 1, ... \tag{5.8}$$

The 2mth moment can be expressed by the recursive formula

$$E\left[\epsilon_t^{2m}\right] = a_m \left[\sum_{n=0}^{m-1} a_n^{-1} E\left(\epsilon_t^{2n}\right) \alpha_0^{m-n} \binom{m}{m-n} \mu(\alpha, \beta, n)\right] * [1 - \mu(\alpha_1, \beta_1, m)]^{-1}. \tag{5.9}$$

Proof. By normality

$$E\left[\epsilon_t^{2m}\right] = a_m E\left[h_t^m\right], \tag{5.10}$$

Where a_m is defined in (5.8). The binomial formula yields

$$h_t^m = \left(\alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1}\right)^m$$

$$= \sum_{n=0}^{m} \left(\begin{array}{c} m \\ n \end{array} \right) \alpha_0^{m-n} \sum_{j=0}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) \alpha_1^j \beta_1^{n-j} \epsilon_{t-1}^{2j} h_{t-1}^{n-j}.$$

Because

$$E\left[\epsilon_{t-1}^{2j} h_{t-1}^{n-j} \left|\psi_{t-2}\right.\right] = a_j h_{t-1}^n,$$

We have

$$E\left[h_t^m \left|\psi_{t-2}\right.\right] = \sum_{n=0}^{m} h_{t-1}^n \left(\begin{array}{c} m \\ n \end{array} \right) \alpha_0^{m-n} \sum_{j=0}^{n} \left(\begin{array}{c} n \\ j \end{array} \right) a_j \alpha_1^j \beta_1^{n-j}. \tag{5.11}$$

Let

$$\omega_t = \left(h_t^m, h_t^{m-1}, ..., h_t\right)'$$

then by (5.11)

$$E\left[\omega_t \left|\omega_{t-2}\right.\right] = d + C\omega_{t-1}, \tag{5.12}$$

here C is an m $\times$ $muppertriangular matrix with diagonal elements.$

$$\mu\left(\alpha_1, \beta_1, i\right) = \sum_{j=0}^{i} \left(\begin{array}{c} i \\ j \end{array} \right) a_j \alpha_1^j \beta_1^{i-j}, \qquad i = 1, ..., m. \tag{5.13}$$

Substituting in (5.12) yields

$$E\left[\omega_t \left|\omega_{t-k-1}\right.\right] = \left(I + C + C^2 + C^{k-1}\right) d + C^k \omega_{t-k}.$$

Since the process is assumed to start indefinitely far in the past with finite 2m moments, the limit as k goes to infinity exists and does not depend on t if and only if all the eigenvalues of C lie inside the unit circle,

$$\lim_{k\to\infty} E\left[\omega_t \left|\omega_{t-k-1}\right.\right] = \left(I - C\right) d = E\left[\omega_t\right]$$

Because C is upper triangular, the eigenvalues are equal to the diagonal elements as given in (5.13). Tedious, but rather straightforward calculations show that $\mu(\alpha_1, \beta_1, i) < 1$ implies $\mu(\alpha_1, \beta_1, i-1) < 1$ for $\alpha_1 + \beta_1 \leq 1$, and $\mu(\alpha_1, \beta_1, m) < 1$ suffices for the 2mth moment to exist; cf. fig. 1.

Finally (5.9) follows from (5.10) and (5.11) by rearranging terms. ■

The conditions for existence of the first twelve moments are illustrated in fig. 1.

It follows by symmetry that if the 2ruth moment exists, $E(\epsilon_t^{2m-1}) = 0$.

For $\beta_1 = 0$, (5.7) reduces to the well-known condition for the ARCH(1) process, Engle (1982). Thus if

$$\alpha_1 > (a_m)^{-1/m}$$

in the ARCH(1) process, the 2mth moment does not exist, whereas even if

$$\sum_{i=1}^{\infty} \delta_i = \alpha_1 (1 - \beta_1)^{-1} > (a_m)^{-1/m}$$

in the GARCH(1,1) process, the 2mth moment might very well exist because of the longer memory in this process.

In the GARCH(1,1) process the mean lag in the conditional variance equation is given by :

$$\varsigma = \sum_{i=1}^{\infty} i\delta_i / \sum_{i=1}^{\infty} \delta_i = (1 - \beta_1)^{-1},$$

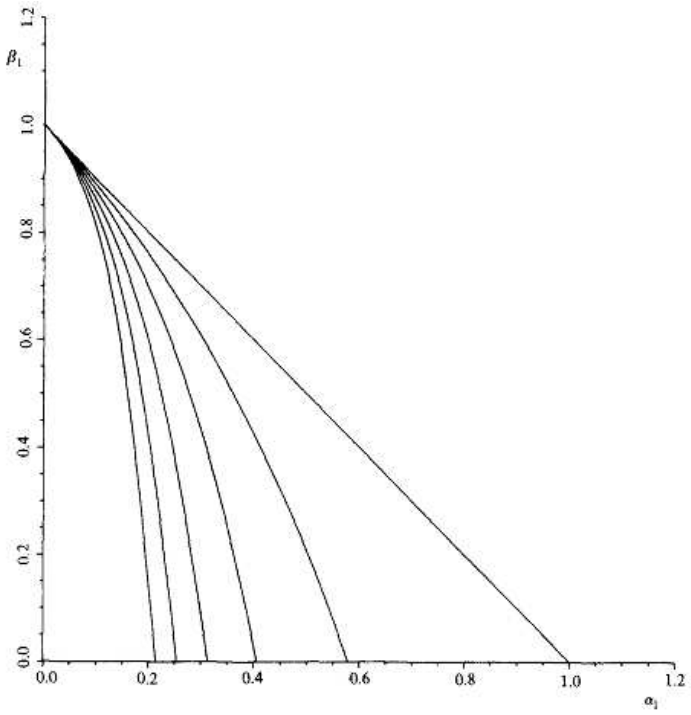


Fig. 1. Moment conditions for GARCH(1,1).

figure 18

and the median lag is found to be

$$\nu = -\log 2/\log \beta_1,$$

where

$$\sum_{i=1}^{\nu} \delta_i / \sum_{i=1}^{\infty} \delta_i = 1/2$$

and the δ_i's are defined in (5.3); cf. Harvey (1982).

If $3\alpha_1^2+2\alpha_1\beta_1+\beta_1^2 < 1$, the fourth-order moment exists and by past Theorem

$$E\left[\epsilon_t^2\right] = \alpha_0(1-\alpha_1-\beta_1)^{-1},$$

and

$$E\left[\epsilon_t^4\right] = 3\alpha_0^2(1+\alpha_1+\beta_1)\left[(1-\alpha_1-\beta_1)\left(1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2\right)\right]^{--1},$$

The coefficient of kurtosis is therefore

$$\begin{aligned}\kappa &= \left(E\left[\epsilon_t^4\right] - 3E\left[\epsilon_t^2\right]^2\right) E\left[\epsilon_t^2\right]^{-2} \\ &= 6\alpha_1^2\left(1-\beta_1^2-2\alpha_1\beta_1-3\alpha_1^2\right)^{-1}\end{aligned}$$

which is greater than zero by assumption. Hence the GARCH(1,1) process is leptokurtic (heavily tailed), a property the process shares with the ARCH(q) process; cf. Milhoj (1984).

5.3.3 Autocorrelation and partial autocorrelation structure

The use of autocorrelation and partial autocorrelation functions to identify and check time series behaviour of the ARMA form in the conditional mean is well established; cf. Box and Jenkins (1976). In this section, the autocorrelation and partial autocorrelation functions for the squared process are shown to be useful in identifying and checking time series behaviour in the conditional variance equation of the GARCH form. The idea of using the squared process to check for model adequacy is not new; cf. Granger and Anderson (1978) where it is found that some of the series modelled in Box and Jenkins (1976) exhibit autocorrelated squared residuabs even though the residuals themselves do not seem to be correlated over time.

Consider the general GARCH(p, q) process as specified in (2.1) and (5.2), and let us assume the process has finite fourth-order moment. 3 Let the covariance function for ϵ_t^2 be denoted

$$\gamma_n = \gamma_{-n} = cov(\epsilon_t^2, \epsilon_{t-n}^2).$$

The general conditions for the existence of finite fourth-order moment are unknown. However, given a specific order of the model the conditions may be derived following the same line of arguments as lead to lastTheorem for the GARCH(1,1) process. For instance the necessary and sufficient condition for the GARCH(1,2) process is found to be

$$\alpha_2 + 3\alpha_1^2 + 3\alpha_2^2 + \beta_1^2 + 2\alpha_1\beta_1 - 3\alpha_2^3 + 3\alpha_1^2\alpha_2 + 6\alpha_1\alpha_2\beta_1 + \alpha_2\beta_1^2 < 1,$$

and for the GARCH(2,1) the condition is

$$\beta_2 + 3\alpha_1^2 + \beta_1^2 + \beta_2^2 + 2\alpha_1\beta_1 - \beta_2^3 - \alpha_1^2\beta_2 + 2\alpha_1\beta_1\beta_2 + \beta_1^2\beta_2 < 1.$$

In Milhoj (1984) the condition for the ARCH(q) process is derived and expressed in terms of the inverse of a qxq matrix, $3\varphi'\left(I-\phi\right)^{-1}\varphi < 1$, where $\varphi' = (\alpha_1.....\alpha_q), \phi_{ij} = \phi_{i+j} + \phi_{i-j}, i, j =$
1.....q,and $\phi_k = 0\ for\ \ k \leq 0\ and\ k > q$.

It follows then immediately from (5.4) and (5.5) that

$$\begin{aligned}\gamma_n &= \sum_{i=1}^{q} \alpha_i\gamma_{n-1} + \sum_{i=1}^{p} \beta_i\gamma_{n-i} \\ &= \sum_{i=1}^{m} \varphi_i\gamma_{n-i}, \qquad n \geq p+1 \end{aligned} \tag{5.14}$$

where m = max(p, q),

$$\varphi_i = \alpha_i + \beta_i, \qquad i = 1, ..., q,$$

$$\alpha_i \equiv 0 \quad \text{for} \quad i < q \quad \text{and} \quad \beta_i \equiv 0 \quad \text{for} \quad i > p.$$

From (5.14) we get the following analogue to the Yule-Walker equations :

$$\rho_n = \gamma_n \gamma_0^{-1} = \sum_{i=1}^{m} \varphi_i \rho_{n-i}, \qquad n \geq p+1 \tag{5.15}$$

Thus, the first p autocorrelations for ϵ_t^2 depend 'directly' on the parameters $\alpha_i, \beta_i \ \forall i$but given $\rho_p, ..., \rho_{p+1-m}$ the above difference equation uniquely determines the autocorrelations at higher lags. This is similar to the result for the autocorrelations for an ARMA(m, p) process; cf, Box and o3Jenkins (1976). Note also, that (5.15) depends on the parameters α_i, β_i only through $\varphi_i \ \ \forall i$

Let ϕ_{kk} denote the kth partial autocorrelation for ϵ_t^2 found by solving the set of k equations in the k unknown $\phi_{k1}, ..., \phi_{kk}$.

$$\rho_n = \sum_{i=1}^{k} \phi_{ki} \rho_{n-i}, \qquad n = 1, ..., k. \tag{5.16}$$

By (5.15) ϕ_{kk} cuts off after lag q for an ARCH(q) process

$$\begin{aligned} \phi_{kk} &\neq 0, \qquad k \leq q, \\ &= 0, \qquad k > q. \end{aligned} \tag{5.17}$$

This is identical to the behaviour of the partial autocorrelation function for an AR(q) process. Also from (5.15) and well-known results in the time series literature, the partial autocorrelation function for ϵ_t^2 for a GARCH(p, q) process is in general non-zero but dies out; see Granger and Newbold (1977).

In practice, of course, the ϕ_{kk},'s and ϕ_{kk}'S will be unknown. However, the sample analogue, say $\widehat{\rho}_n$ yields a consistent estimate forρ_n and ϕ_{kk} is consistently estimated by the kth coefficient, say $\widehat{\phi}_{kk}$, in a kth-order autoregression for ϵ_t^2; see Granger and Newbold (1977). These estimates together with their asymptotic variance under the null of no GARCH 1/T [cf. Weiss (1984) and McLeod and Li (1983)] can be used in the preliminary identification stage, and are also useful for diagnostic checking.

5.3.4 Estimation of the GARCH regression model

In this section we consider maximum likelihood estimation of the GARCH regression model (5.2). Because the results are quite similar to those for the ARCH regression model, the discussion will be very schematic.

Let $z_t' = (1, \epsilon_{t-1}^2, ..., \epsilon_{t-q}^2, h_{t-1}, ..., h_{t-p})$, $\omega' = (\alpha_0, ..., \alpha_q, \beta_1, ...\beta_p)$ and $\theta \in \Theta$, where $\theta = (b', \omega')$ and Θ is a compact subspace of a Euclidean space such that ϵ_t possesses finite second moments. Denote the true parameters by $\theta_0, \in$ int Θ. We may then rewrite the model as

$$\begin{aligned} \epsilon_t &= \gamma_t - x_t' b, \\ \epsilon_t \,|\psi_{t-1} &\sim N\left(0, h_t\right), \\ h_t &= z_t' \omega. \end{aligned} \tag{5.18}$$

The log likelihood function for a sample of T observations is apart from some constant,

$$\begin{aligned} L_T\left(\theta\right) &= T^{-1} \sum_{i=1}^{T} l_t\left(\theta\right), \\ l_t\left(\theta\right) &= -1/2(\log h_t - 1/2(\epsilon_t^2 h_t^{-1}) \end{aligned} \tag{5.19}$$

Differentiating with respect to the variance parameter yields

$$\begin{aligned} \frac{\partial l_t}{\partial \omega} &= \frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial \omega} \left(\frac{\epsilon_t^2}{h_t} - 1 \right), \\ \frac{\partial^2 l_t}{\partial \omega \partial \omega'} &= \left(\frac{\epsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial \omega'} \left[\frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial \omega} \right] - \frac{1}{2} h_t^{-2} \frac{\partial h_t}{\partial \omega} \frac{\partial h_t}{\partial \omega'} \frac{\epsilon_t^2}{h_t}, \end{aligned} \tag{5.20}$$

Where

$$\frac{\partial h_t}{\partial \omega} = z_t + \sum_{i=1}^{p} \beta_i \frac{\partial h_{t-i}}{\partial \omega} \tag{5.21}$$

The only difference from Engle (1982) is the inclusion of the recursive part in (5.21). Note, $B(1) < 1$ guarantees that (5.21) is stable. Since the conditional

expectation of the first term in (5.20) is zero, the part of Fisher's information matrix corresponding to ω is consistently estimated by the sample analogue of the last term in (5.20) which involves first derivatives only.

Differentiating with respect to the mean parameters yields

$$\frac{\partial l_t}{\partial b} = \epsilon_t x_t h_t^{-1} + \frac{1}{2} h_t \frac{\partial h_t}{\partial b} \left(\frac{\epsilon_t^2}{h_t} - 1 \right), \tag{5.22}$$

$$\begin{aligned} \frac{\partial^2 l_t}{\partial b \partial b'} &= -h_t^{-1} x_t x_t' - \frac{1}{2} h_t^{-2} \frac{\partial h_t}{\partial b} \frac{\partial h_t}{\partial b'} \left(\frac{\epsilon_t^2}{h_t} \right) \\ &\quad -2 h_t^{-2} \epsilon_t x_t \frac{\partial h_t}{\partial b} + \left(\frac{\epsilon_t^2}{h_t} - 1 \right) \frac{\partial}{\partial b'} \left[\frac{1}{2} h_t^{-1} \frac{\partial h_t}{\partial b} \right]. \end{aligned} \tag{5.23}$$

Where

$$\frac{\partial h_t}{\partial b} = -2 \sum_{j=1}^{q} \alpha_j x_{t-j} \epsilon_{t-j} + \sum_{j=1}^{p} \beta_j \frac{\partial h_{t-j}}{\partial b}. \tag{5.24}$$

Again the single difference from the ARCH(q) regression model is the inclusion of the recursive part in (5.24). A consistent estimate of the part of the information matrix corresponding to b is given by the sample analogue of the first two terms in (5.22) but with $\epsilon_t^2 \; h_t^{-1}$ in the second term replaced by its expected value of one. This estimate will also involve first derivatives only.

Finally, the elements in the off-diagonal block in the information matrix may be shown to be zero. Because of this asymptotic independence ω can be estimated without loss of asymptotic efficiency based on a consistent estimate of b, and vice versa.

To obtain maximum likelihood estimates, and second-order efficiency, an iterative procedure is called for. For the ARCH(q) regression model the method of scoring could be expressed in terms of a simple auxiliary regression, but the recursive terms in (5.21) and (5.24) complicate this procedure. Instead the Berndt, Hall, Hall and Hausman (1974) algorithm turns out to be convenient.

Let $\theta^{(i)}$ denote the parameter estimates after the ith iteration. $\theta^{(i+1)}$ is then calculated from

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left(\sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \right)^{-1} \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta},$$

where $\frac{\partial l_t}{\partial \theta}$ is evaluated at $\theta^{(i)}$, and λ_i is a variable step length chosen to maximize the likelihood function in the given direction. Note, the direction vector is easily calculated from a least squares regression of a $\mathrm{T} \times 1 vector of ones on \partial l_t \, \overline{\partial \theta}$.

Also, the iterations for $\omega^{(i)}$ and $b^{(i)}$ may be carried out separately because of the block diagonality in the information matrix.

From Weiss (1982) it follows that the maximum likelihood estimate $\widehat{\theta}_t$ is strongly consistent for θ_0 and asymptotically normal with mean $\theta_0$and covariance matrix $F^{-1} = -E(\frac{\partial^2 l_t}{\partial\theta\partial\theta 4})^{-1}$. However, $F = F$, where $F = E((\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta'}))$, and a consistent estimate of the asymptotic covariance matrix is therefore given by $T^{-l}(\sum_{t=1}^{T}(\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta'}))^{-1}$ from the last BHHH iteration.

Replacing (5.2) with the weaker set of assumptions

$$\begin{aligned} E\left(\epsilon_t \,|\psi_{t-1}\right) &= 0, \\ E\left(\epsilon_t^2 h_t^{-1} \,|\psi_{t-1}\right) &= 1, \\ E\left(\epsilon_t^4 h_t^{-2} \,|\psi_{t-1}\right) &\leq M < \infty \end{aligned}$$

$\widehat{\theta}_T$ is still strongly consistent for θ_0 and asymptotically normal with mean θ_0 but with covariance matrix $F^{-1}FF^{-1}$; see Weiss (1982) and White (1982). Of course, if the true conditional distribution is normal, $F = F$ and therefore $F^{-1}FF^{-1} = F^{-1}$

5.3.5 Testing for GARCH

Because of the complication involved in estimating a GARCH process, it seems of interest to have a formal test for the presence of GARCH instead of just relying on the more informal tools .

Consider the GARCH(p, q) regression model (5.18). As in Engle and Kraft (1983) let us partition the conditional variance equation

$$h_t = z_t'\omega = z_{1t}'\omega_1 + z_{2t}'\omega_2 \tag{5.25}$$

The Lagrange multiplier test statistic for H_0: $\omega_2 = 0$ is then given by

$$\xi_{LM}^{*} = \frac{1}{2} f_0' Z_0 \left(Z_0' Z_0\right)^{-1} Z_0' f_0, \tag{5.26}$$

Where

$$\begin{aligned} f_0 &= \left(\epsilon_1^2 h_1^{-1} - 1, ..., \epsilon_T^2 h_T^{-1} - 1\right)', \\ Z_0 &= \left(h_1 \frac{\partial h_1}{\partial\omega}, ..., h_T \frac{\partial h_T}{\partial\omega}\right)', \end{aligned} \tag{5.27}$$

and both are evaluated under H_0. When H_0 is true,ξ^*_{LM} is asymptotically chi-square with r, the number of elements in ω_2, degrees of freedom. This test differs slightly from the standard results, Breusch and Pagan (1978), in that $\frac{\partial h_t}{\partial \omega}$ does not simplify when the conditional variance equation contains lagged conditional variances; cf. eq. (5.21).

It is well known that by normality an asymptotically equivalent test statistic is

$$\xi_{LM} = T.R^2,$$

where R^2 is the squared multiple correlation coefficient between f_0 and Z_0. From past section this corresponds to $T.R^2$ from the OLS regression in the first BHHH iteration for the general model starting at the maximum likelihood estimates under H_0.

The alternative as represented by z_2 needs some consideration. Straightforward calculations show that under the null of white noise, $Z_0 Z_0'$ is singular if both $p > 0$ and $q > 0$, and therefore a general test for GARCH(p, q) is not feasible. In fact if the null is an ARCH(q) process, $Z_0 Z_0'$ is singular for GARCH($r_1, q + r_2$) alternatives, where $r_1 > 0$ and $r_2 > 0$. It is also interesting to note that for an ARCH(q) null, the LM test for GARCH(r,q) and ARCH(q + r) alternatives coincide. This is similar to the results in Godfrey (1978), where it is shown that the LM tests for AR(p) and MA(q) errors in a linear regression model coincide and that the test procedures break down when a full ARMA(p, q) model is considered. These test results are, of course, not peculiar to the LM test, but concern the Likelihood Ratio and the Wald tests as well. A formal proof of the above statements can be constructed along the same lines as in Godfrey (1978, 1981).

5.4 ARCH and GARCH Models in R

There are many distinct kinds of non-linear time series models. The ARCH or GARCH models, which are used to model and predict volatility, are the most widely used non-linear financial models. The ARCH concept was developed by economist Robert F. Engle III in the 1980s. ARCH immediately improved financial modeling, resulting in Engle winning the 2003 Nobel Memorial Prize in Economic Sciences.

When it comes to capturing the volatility clustering of financial returns, Taylor (1986) and Bollerslev (1986)'s generalised autoregressive conditional heteroscedasticity (GARCH) models dominate. The GARCH model allows the conditional variance to be dependent upon previous own lags and squared error terms.

5.4.1 Steps Involves before estimating Volatility Models

When estimating GARCH type models, the following steps are essential:

i. Check the stationary
-Apply Augmented dicky fuller test
-Time series plot
ii. Check the presence of volatility
-Time series plot
iii. Check normality
-Jarque-Bera Test of Normality
-QQ-plot
-Summary statistics(Kurtosis)
iv. Check for ARCH effect
- Apply LM ARCH test

Remark 66 *If The LM test shows p-value less than 0.05 which indicates that null hypothesis (no arch effect) can be rejected. Therefore, the log of stock returns will have ARCH.or Apply Ljung-Box Q-test on the first m lags of the squared residual series as an alternative to Engle's ARCH test, you can check for serial dependence (ARCH effects) in a residual series by conducting a Ljung-Box Q-test on the first m lags of the squared residual series.*

v. Estimate ARCH and GARCH Models with Normal Innovations using rugarch() package

-If ARCH effect is present in the returns, we will estimate ARCH and GARCH model using rugarch() package.

vi. Estimate ARCH and GARCH Models with Non-Normal Innovations using rugarch() pakage as we can see from the stylized facts, the normal distribution assumption is not true. We will estimate both ARCH and GARCH models with t-distribution innovation.

vii. Model Selection using Information Criterion

5.4.2 Application in R

Install and load required libraries:

```
1  ###installing and loading multiple packages
2  list.packages<-c("fGarch", "PerformanceAnalytics","rugarch","tseries","xts","FinTS")
3  new.packages <- list.packages[!(list.packages %in% installed.packages()[,"Package"])]
4    if(length(new.packages)) install.packages(new.packages)
5     #Loading Packages
6    invisible(lapply(list.packages, require, character.only = TRUE))
```

and set your working directory and upload your data

```
1  setwd("")# in inverted commas, add the path to your working directory.
```

```
2 BMW<-read.csv(BMW.csv)# Upload your data if you are using the R
console or Jupyter notebook.

3 head(BMW)
```

Calculate Financial Returns and remove first NA value

```
1 #set timeseries element

2 BMW$Years <- as.Date(BMW$Years, "%d/%m/%Y")

3 class(BMW$Years)

4 BMW.z = zoo(x=BMW$Prices, order.by=BMW$Years)

5 #calculate log returns and remove first NA value

6 Return.BMW<-Return.calculate(BMW.z, method = "log")[-1]
```

Check the Stationary:

ADF-test either from the "tseries" or "urca" packages can be used.

```
1 #apply ADF test with drift

2 ADF_Returns = ur.df(Return.BMW, type = "drift",selectlags = "AIC"
)

3 #summary of he test

4 summary(ADF_Returns)
```

```
################################################
# Augmented Dickey-Fuller Test Unit Root Test #
################################################

Test regression drift


Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
      Min        1Q    Median        3Q       Max
-0.156291 -0.010786 -0.000023  0.010886  0.138128

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0004118  0.0003170   1.299 0.193978
z.lag.1     -1.0096970  0.0201650 -50.072  < 2e-16 ***
z.diff.lag   0.0552591  0.0145795   3.790 0.000152 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02171 on 4692 degrees of freedom
Multiple R-squared:   0.48,     Adjusted R-squared:  0.4798
F-statistic:  2166 on 2 and 4692 DF,  p-value: < 2.2e-16


Value of test-statistic is: -50.0717 1253.588

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.43 -2.86 -2.57
phi1  6.43  4.59  3.78
```

figure 19

In absolute terms value of test statics(50.0512) is higher than of three critical values. So returns of BMW are stationary.

Check the Presence of Volatility

We check the presence of volatility using time series plot for log returns, square returns and absolute returns.

```
1  # plot returns with squared and absolute returns
2  dataToPlot = cbind(Return.BMW, Return.BMW^2, abs(Return.BMW))
3   colnames(dataToPlot) = c("Returns", "Returns^2", "abs(Returns)")
4   plot.zoo(dataToPlot, main="BMW Daily Returns", col="blue")
```

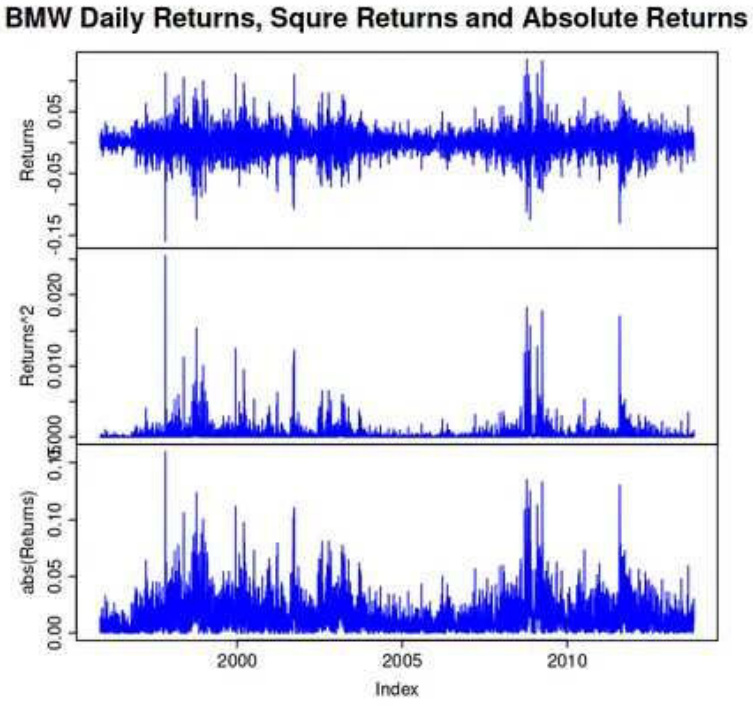


figure 20

Check Normality

QQ-plot to check normality

```
1 ##QQ-plot
2 qqnorm(Return.BMW, main = "BMW Daily Returns -QQ Plot", col =
"blue")
3 qqline(Return.BMW)
```

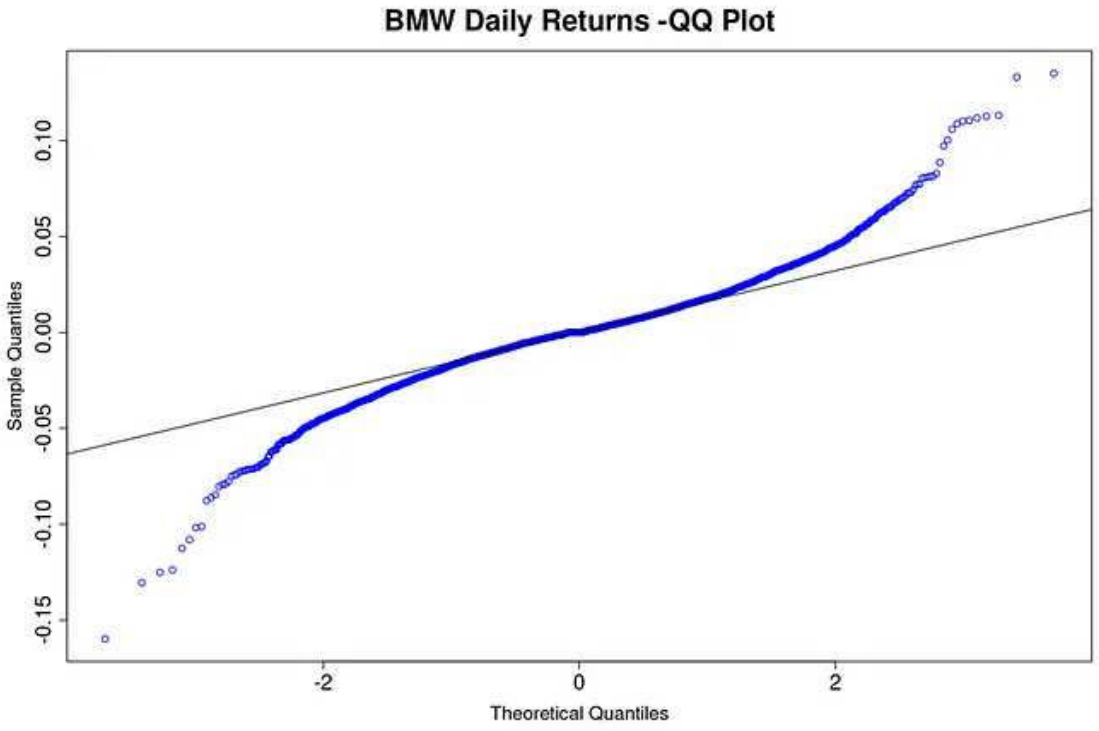


figure 21

Value of Kurtosis to check for normality

```
1  #summary table
2  table.Stats(cbind(Return.BMW))
3  #conduct Jarque-Bera test for normality
4   jarque.bera.test(Return.BMW)
```

	Observations	NAs	Minimum	Quartile 1	Median	Arithmetic Mean	Geometric Mean	Quartile 3	Maximum	SE Mean	LCL Mean (0.95)	UCL Mean (0.95)	Variance	Stdev	Skewness	Kurtosis
Return.BMW	4697	0	-0.1598	-0.0104	0	4e-04	2e-04	0.0112	0.1352	3e-04	-2e-04	0.001	5e-04	0.0218	-0.011	4.3723

figure 22

Jarque Bera test to check normality

```
        Jarque Bera Test

data:  Return.BMW
X-squared = 3741.4, df = 2, p-value < 2.2e-16
```

figure 23

This tells us that the test statistic is 3741.4 and the p-value of the test is

0.000. In this case, we would reject the null hypothesis that the data is not normally distributed.

Check for ARCH effect

I will plot auto-correlation first:

```
1  ## Convert to xts for time series features
2  Return.BMW<-as.xts(Return.BMW)
3  # plot autocorrelations of returns, returns^2 and abs(returns)
4  options(repr.plot.width=15, repr.plot.height=5)
5  par(mfrow=c(1,3))
6   acf(Return.BMW, main="BMW Returns",cex.main=10)
7  acf(Return.BMW^2, main="BMW Returns^2",cex.main=10)
8  acf(abs(Return.BMW), main="BMW abs(Returns)",cex.main=10)
9    par(mfrow=c(1,1))
10   # use Ljung Box.test from stats package to check auto correlation in
squre retruns
11   Box.test(coredata(Return.BMW^2), type="Ljung-Box", lag = 12)
12   #ARCH LM Test
13   ArchTest(Return.BMW)
```

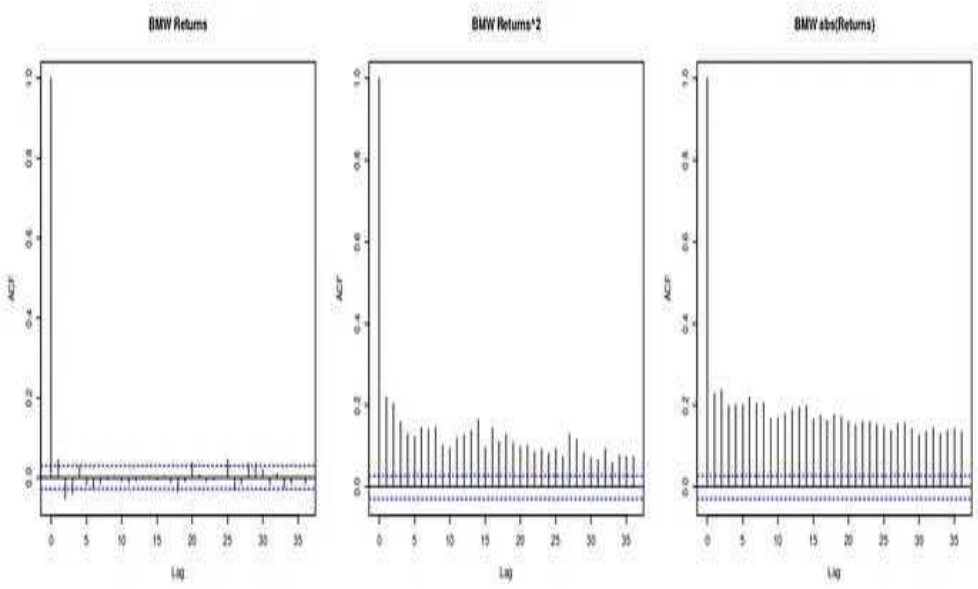


figure 24

Square and absolute returns show high level of auto-correlation.

We can double check present of auto-correlation in square returns by apply Ljung-Box test on square returns.

```
        Box-Ljung test

data:  coredata(Return.BMW^2)
X-squared = 1222.9, df = 12, p-value < 2.2e-16
```

figure 25

p<0.05 so data is not independent. We reject the null hypothesis . It mean auto correlation is present.

ARCH-LM Test

```
        ARCH LM-test; Null hypothesis: no ARCH effects

data:  Return.BMW
Chi-squared = 500.01, df = 12, p-value < 2.2e-16
```

figure 26

The LM test shows p-value less than 0.05 which indicates that null hypothesis (no arch effect) can be rejected. Therefore, the BMW Returns exhibit ARCH effect.

Estimate ARCH and GARCH Models with Normal Innovations using rugarch() Package

ARCH Model estimation with normal innovation

```
1 #Specify the model
2 spec = ugarchspec(variance.model=list(garchOrder=c(1,0)),
3 mean.model=list(armaOrder=c(0,0)),distribution.model="norm")
4 #Fit ARCH Model
5 arch11.fit=ugarchfit(data=Return.BMW,spec=spec)
6 #fitted model outcome
7 arch11.fit
```

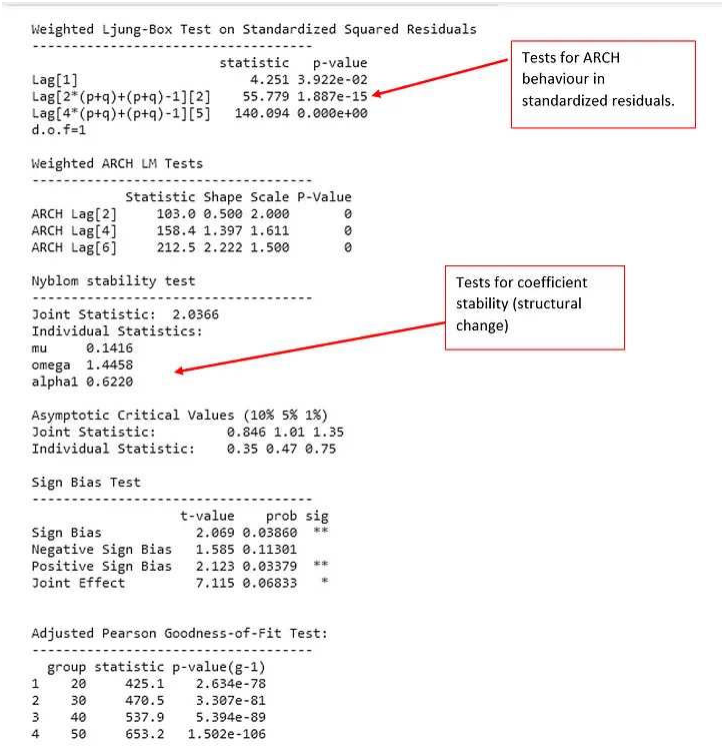
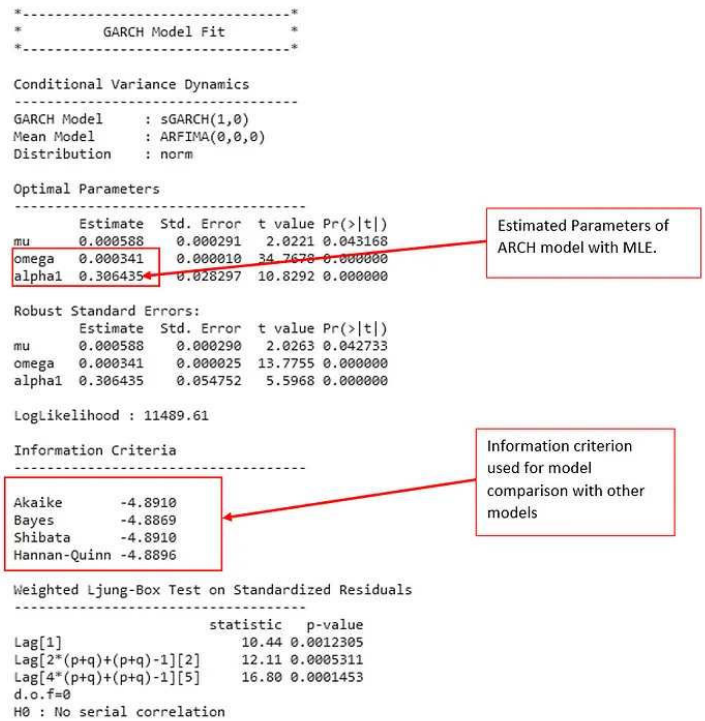

Weighted Ljung-Box Test on Standardized Squared Residuals
statistic p-value
Lag[1] 4.251 3.922e-02
Lag[2*(p+q)+(p+q)-1][2] 55.779 1.887e-15
Lag[4*(p+q)+(p+q)-1][5] 140.094 0.000e+00
d.o.f=1
Tests for ARCH behaviour in standardized residuals.
Weighted ARCH LM Tests
Statistic Shape Scale P-Value
ARCH Lag[2] 103.0 0.500 2.000 0
ARCH Lag[4] 158.4 1.397 1.611 0
ARCH Lag[6] 212.5 2.222 1.500 0
Nyblom stability test
Joint Statistic: 2.0366
Individual Statistics:
mu 0.1416
omega 1.4458
alpha1 0.6220
Tests for coefficient stability (structural change)
Asymptotic Critical Values (10% 5% 1%)
Joint Statistic: 0.846 1.01 1.35
Individual Statistic: 0.35 0.47 0.75
Sign Bias Test
t-value prob sig
Sign Bias 2.069 0.03860 **
Negative Sign Bias 1.585 0.11301
Positive Sign Bias 2.123 0.03379 **
Joint Effect 7.115 0.06833 *
Adjusted Pearson Goodness-of-Fit Test:
group statistic p-value(g-1)
1 20 425.1 2.634e-78
2 30 470.5 3.307e-81
3 40 537.9 5.394e-89
4 50 653.2 1.502e-106

figure 27


----
*   GARCH Model Fit   *

Conditional Variance Dynamics

GARCH Model : sGARCH(1,0)
Mean Model : ARFIMA(0,0,0)
Distribution : norm
Optimal Parameters

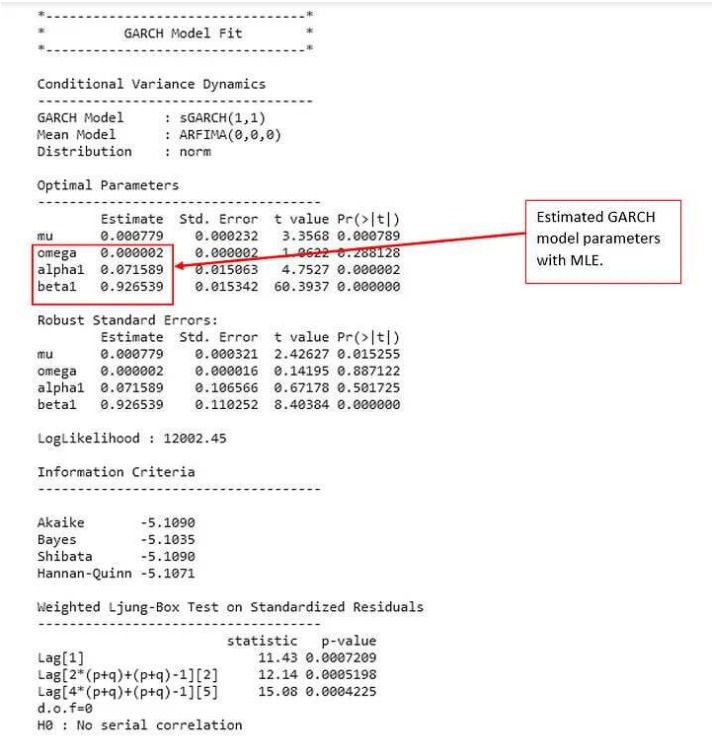
Estimate Std. Error t value Pr(>|t|)
mu 0.000588 0.000291 2.0221 0.043168
omega 0.000341 0.000010 34.7678 0.000000
alpha1 0.306435 0.028297 10.8292 0.000000
Robust Standard Errors:
Estimate Std. Error t value Pr(>|t|)
mu 0.000588 0.000290 2.0263 0.042733
omega 0.000341 0.000025 13.7755 0.000000
alpha1 0.306435 0.054752 5.5968 0.000000
LogLikelihood : 11489.61
Information Criteria

Akaike -4.8910
Bayes -4.8869
Shibata -4.8910
Hannan-Quinn -4.8896
Weighted Ljung-Box Test on Standardized Residuals
----
statistic p-value
Lag[1] 10.44 0.0012305
Lag[2*(p+q)+(p+q)-1][2] 12.11 0.0005311
Lag[4*(p+q)+(p+q)-1][5] 16.80 0.0001453
d.o.f=0
H0 : No serial correlation
Estimated Parameters of ARCH model with MLE.
Information criterion used for model comparison with other models

figure 28

GARCH(1,1) with Normal Innovation

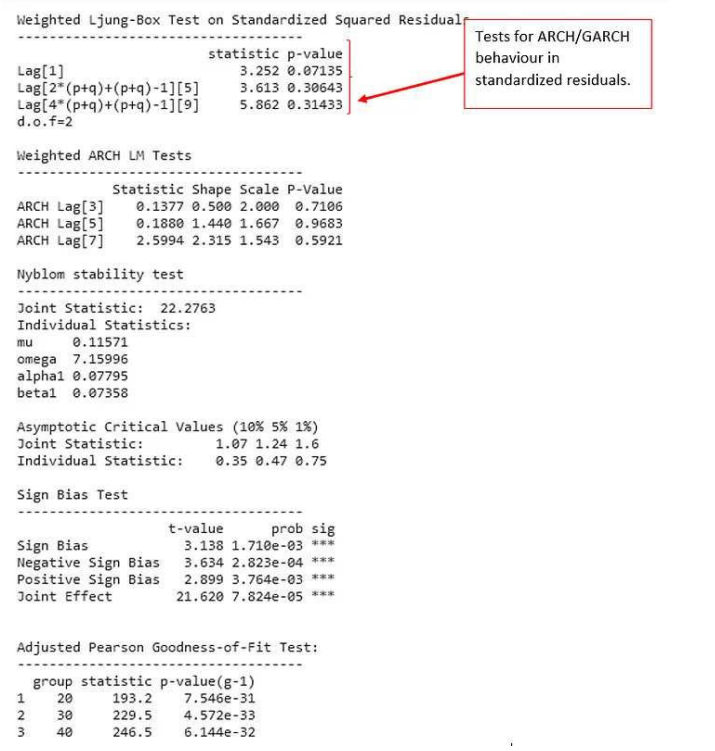
```
#Specify GARCH models:
spec = ugarchspec(variance.model=list(garchOrder=c(1,1)),
mean.model=list(armaOrder=c(0,0)),
distribution.model="norm",)
#fit GARCH model
garch.fit=ugarchfit(data=Return.BMW,spec=spec)
##summary of GARCH fit
garch.fit
```



```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : norm

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000779    0.000232   3.3568 0.000789
omega   0.000002    0.000002   1.0622 0.288128
alpha1  0.071589    0.015063   4.7527 0.000002
beta1   0.926539    0.015342  60.3937 0.000000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000779    0.000321  2.42627 0.015255
omega   0.000002    0.000016  0.14195 0.887122
alpha1  0.071589    0.106566  0.67178 0.501725
beta1   0.926539    0.110252  8.40384 0.000000

LogLikelihood : 12002.45

Information Criteria
------------------------------------

Akaike       -5.1090
Bayes        -5.1035
Shibata      -5.1090
Hannan-Quinn -5.1071

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic   p-value
Lag[1]                      11.43 0.0007209
Lag[2*(p+q)+(p+q)-1][2]     12.14 0.0005198
Lag[4*(p+q)+(p+q)-1][5]     15.08 0.0004225
d.o.f=0
H0 : No serial correlation
```

Estimated GARCH model parameters with MLE.

figure 29



```
Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      3.252 0.07135
Lag[2*(p+q)+(p+q)-1][5]     3.613 0.30643
Lag[4*(p+q)+(p+q)-1][9]     5.862 0.31433
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.1377 0.500 2.000  0.7106
ARCH Lag[5]    0.1880 1.440 1.667  0.9683
ARCH Lag[7]    2.5994 2.315 1.543  0.5921

Nyblom stability test
------------------------------------
Joint Statistic:  22.2763
Individual Statistics:
mu     0.11571
omega  7.15996
alpha1 0.07795
beta1  0.07358

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.07 1.24 1.6
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value      prob sig
Sign Bias            3.138 1.710e-03 ***
Negative Sign Bias   3.634 2.823e-04 ***
Positive Sign Bias   2.899 3.764e-03 ***
Joint Effect        21.620 7.824e-05 ***


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     193.2    7.546e-31
2    30     229.5    4.572e-33
3    40     246.5    6.144e-32
```

Tests for ARCH/GARCH behaviour in standardized residuals.

figure 30

The Ljung-Box test tells us that we find no evidence of serial correlation in the squared residuals of the GARCH model.

```
1 #Conditional Volatility Plot for GARCH Model
2 options(repr.plot.width=10, repr.plot.height=5)
3 plot.ts(sigma(garch.fit), ylab="sigma(t)", col="blue", main = "Conditional Varaince of GARCH Model", cex.main=2)
```

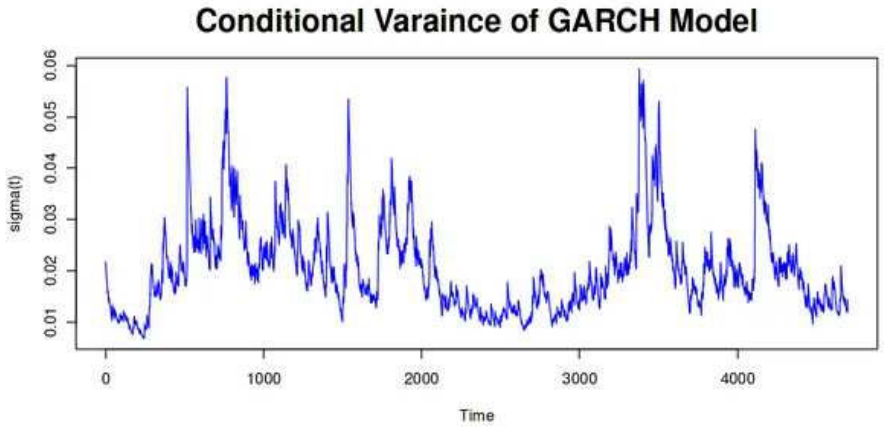


figure 31

ARCH(1) and GARCH(1,1) with normal Innovation are not appropriate Models

QQ-plot of GARCH-normal fitted values

```
1 ##QQ-plot for fitted GARCH Model
2 plot(garch.fit, which=9)
```

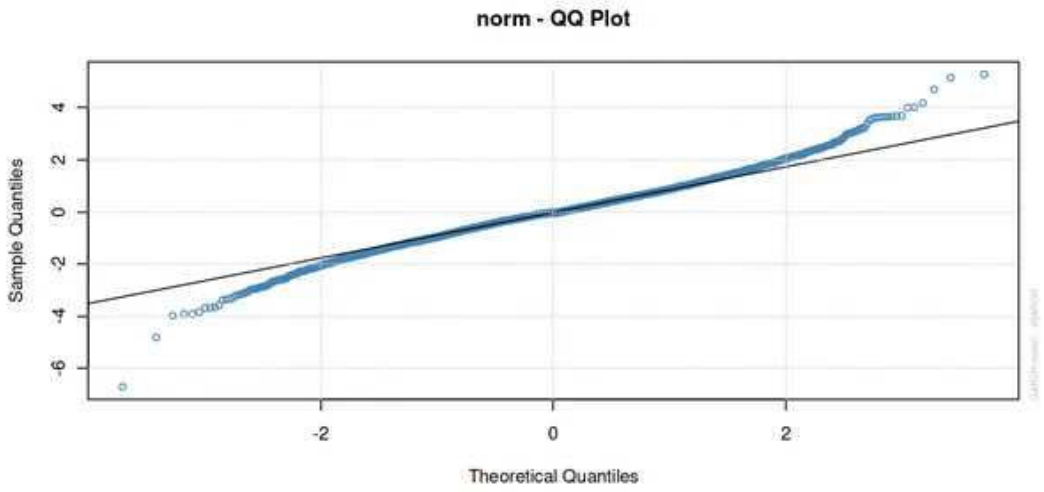


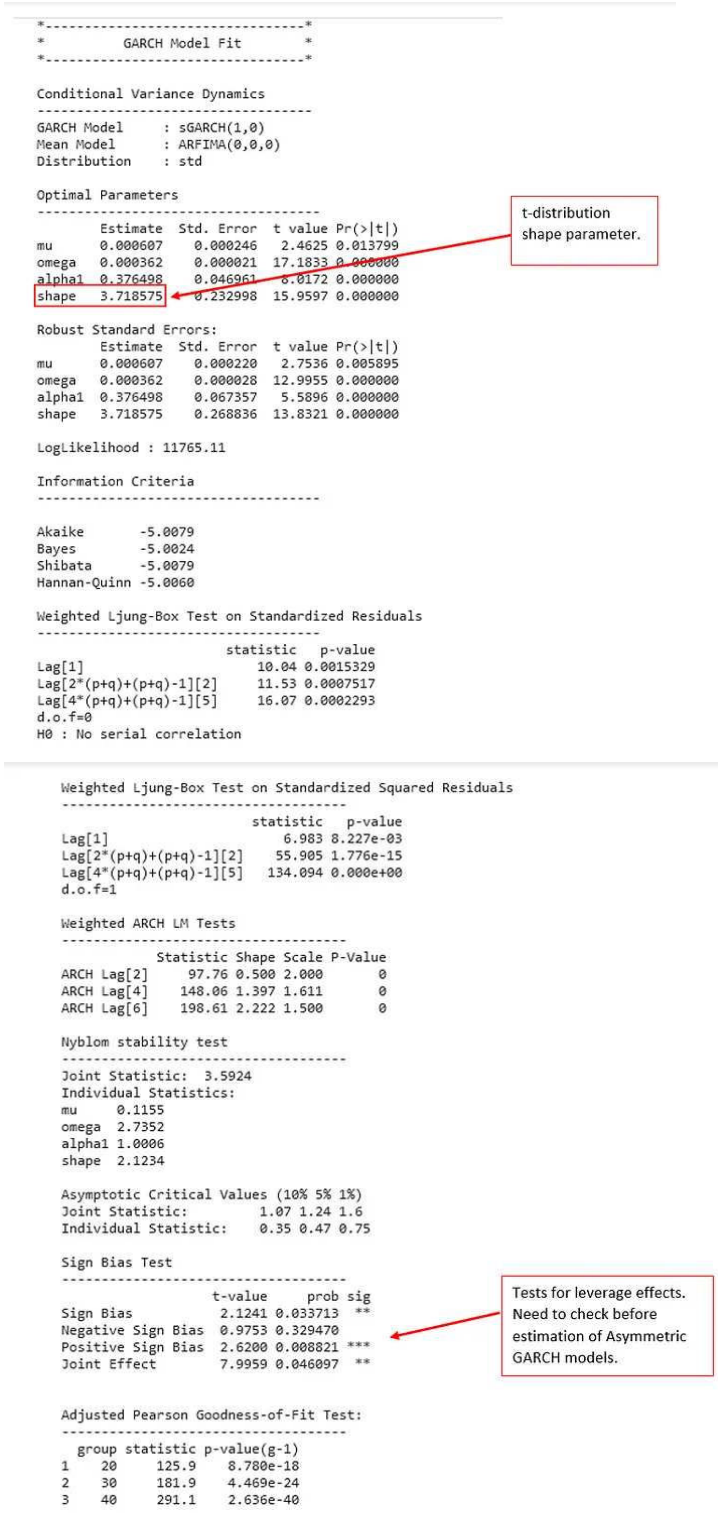
figure 32

Estimation of ARCH and GARCH Models with Non-Normal Innovations using rugarch() Package

ARCH Model with t-distribution innovation

```
#Specify the model
 spec = ugarchspec(variance.model=list(garchOrder=c(1,0)),
 mean.model=list(armaOrder=c(0,0)),distribution.model="std")
  #Fit ARCH Model
 arch11.fit.t=ugarchfit(data=Return.BMW,spec=spec)
```

6 arch11.fit.t


----
*   GARCH Model Fit   *

Conditional Variance Dynamics

GARCH Model   : sGARCH(1,0)
Mean Model   : ARFIMA(0,0,0)
Distribution   : std

Optimal Parameters

        Estimate Std. Error t value Pr(>|t|)
mu   0.000607   0.000246   2.4625 0.013799
omega   0.000362   0.000021 17.1833 0.000000
alpha1 0.376498   0.046961   8.0172 0.000000
shape   3.718575   0.232998 15.9597 0.000000

Robust Standard Errors:
        Estimate Std. Error t value Pr(>|t|)
mu   0.000607   0.000220   2.7536 0.005895
omega   0.000362   0.000028 12.9955 0.000000
alpha1 0.376498   0.067357   5.5896 0.000000
shape   3.718575   0.268836 13.8321 0.000000

LogLikelihood : 11765.11

Information Criteria

Akaike   -5.0079
Bayes   -5.0024
Shibata   -5.0079
Hannan-Quinn -5.0060

Weighted Ljung-Box Test on Standardized Residuals
----
                        statistic   p-value
Lag[1]   10.04 0.0015329
Lag[2*(p+q)+(p+q)-1][2]   11.53 0.0007517
Lag[4*(p+q)+(p+q)-1][5]   16.07 0.0002293
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
----
                        statistic   p-value
Lag[1]   6.983 8.227e-03
Lag[2*(p+q)+(p+q)-1][2]   55.905 1.776e-15
Lag[4*(p+q)+(p+q)-1][5]   134.094 0.000e+00
d.o.f=1

Weighted ARCH LM Tests

            Statistic Shape Scale P-Value
ARCH Lag[2]   97.76 0.500 2.000   0
ARCH Lag[4]   148.06 1.397 1.611   0
ARCH Lag[6]   198.61 2.222 1.500   0

Nyblom stability test

Joint Statistic: 3.5924
Individual Statistics:
mu   0.1155
omega 2.7352
alpha1 1.0006
shape 2.1234

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic: 1.07 1.24 1.6
Individual Statistic: 0.35 0.47 0.75

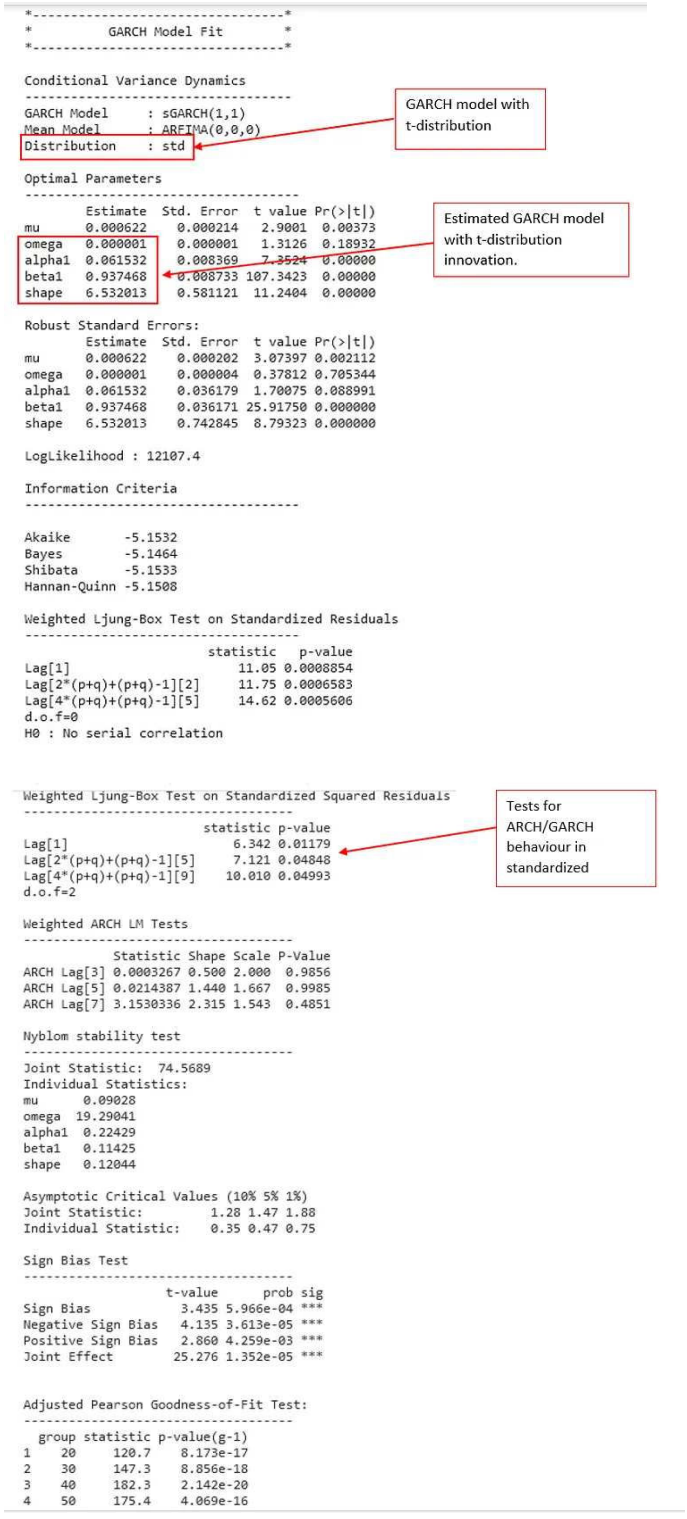
Sign Bias Test
----
                   t-value   prob sig
Sign Bias   2.1241 0.033713 **
Negative Sign Bias 0.9753 0.329470
Positive Sign Bias 2.6200 0.008821 ***
Joint Effect   7.9959 0.046097 **

Adjusted Pearson Goodness-of-Fit Test:
----
  group statistic p-value(g-1)
1   20   125.9   8.780e-18
2   30   181.9   4.469e-24
3   40   291.1   2.636e-40
t-distribution shape parameter.
Tests for leverage effects. Need to check before estimation of Asymmetric GARCH models.

figure 33

GARCH Model with t-distribution Innovation

```
1  #Specify GARCH models:
2  spec = ugarchspec(variance.model=list(garchOrder=c(1,1)),
3  mean.model=list(armaOrder=c(0,0)),
4  distribution.model="std",)
5  #fit GARCH model
6  garch.fit.t=ugarchfit(data=Return.BMW,spec=spec)
7  garch.fit.t
```




```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : sGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000622    0.000214   2.9001  0.00373
omega   0.000001    0.000001   1.3126  0.18932
alpha1  0.061532    0.008369   7.3524  0.00000
beta1   0.937468    0.008733 107.3423  0.00000
shape   6.532013    0.581121  11.2404  0.00000

Robust Standard Errors:
        Estimate  Std. Error  t value Pr(>|t|)
mu      0.000622    0.000202  3.07397 0.002112
omega   0.000001    0.000004  0.37812 0.705344
alpha1  0.061532    0.036179  1.70075 0.088991
beta1   0.937468    0.036171 25.91750 0.000000
shape   6.532013    0.742845  8.79323 0.000000

LogLikelihood : 12107.4

Information Criteria
------------------------------------

Akaike       -5.1532
Bayes        -5.1464
Shibata      -5.1533
Hannan-Quinn -5.1508

Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic   p-value
Lag[1]                     11.05 0.0008854
Lag[2*(p+q)+(p+q)-1][2]    11.75 0.0006583
Lag[4*(p+q)+(p+q)-1][5]    14.62 0.0005606
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                      6.342 0.01179
Lag[2*(p+q)+(p+q)-1][5]     7.121 0.04848
Lag[4*(p+q)+(p+q)-1][9]    10.010 0.04993
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3] 0.0003267 0.500 2.000  0.9856
ARCH Lag[5] 0.0214387 1.440 1.667  0.9985
ARCH Lag[7] 3.1530336 2.315 1.543  0.4851

Nyblom stability test
------------------------------------
Joint Statistic:  74.5689
Individual Statistics:
mu      0.09028
omega  19.29041
alpha1  0.22429
beta1   0.11425
shape   0.12044

Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.28 1.47 1.88
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value      prob sig
Sign Bias            3.435 5.966e-04 ***
Negative Sign Bias   4.135 3.613e-05 ***
Positive Sign Bias   2.860 4.259e-03 ***
Joint Effect        25.276 1.352e-05 ***


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     120.7    8.173e-17
2    30     147.3    8.856e-18
3    40     182.3    2.142e-20
4    50     175.4    4.069e-16
```

figure 34

```
1 ##QQ-plot for fitted GARCH Model with t distribution

2 plot(garch.fit.t, which=9)
```

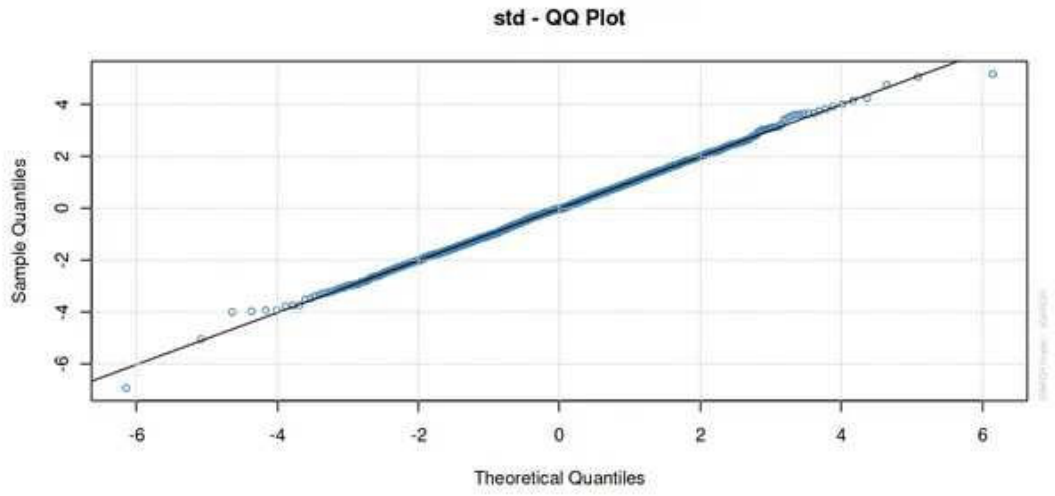


figure 35

We can observe that the QQ-plot for Garch (1,1) does give us better results.

```
1 #various plots for fitted values
2 options(repr.plot.width=15, repr.plot.height=15)
3 plot(garch.fit.t, which= "all")
```

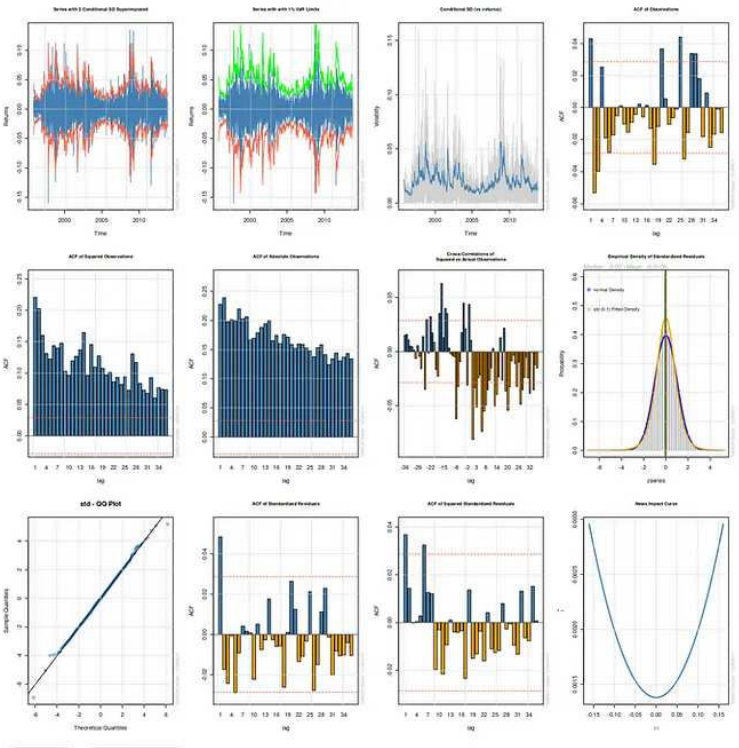


figure 36

Model Selection using Information Criterion

```
#Model selection usinh information critarion
    modcl.list = list("arch(1,1)" = arch11.fit,
   "arch(1,1)-t" = arch11.fit.t,
   "garch(1,1)" = garch.fit,
   "garch(1,1).t" = garch.fit.t)
     info.mat = sapply(model.list, infocriteria)
     rownamcs(info.mat) = rownamcs(infocriteria(garch.fit))
      info.mat
```

A matrix 4 × 4 of type dbl

	arch(1,1)	arch(1,1)-t	garch(1,1)	garch(1,1).t
Akaike	-4.891043	-5.007926	-5.108986	-5.153248
Bayes	-4.886920	-5.002429	-5.103489	-5.146377
Shibata	-4.891044	-5.007928	-5.108988	-5.153251
Hannan-Quinn	-4.889593	-5.005994	-5.107054	-5.150833

GARCH (1,1) with t distribution has the best fit – smallest values of info criterion.

figure 37

Conclusion

A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in the frequency domain. Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. In this handout, i focus on The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving-average or ARMA model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process.

ARCH models are commonly employed in modeling financial time series that exhibit time-varying volatility and volatility clustering, i.e. periods of swings interspersed with periods of relative calm. ARCH-type models are sometimes considered to be in the family of stochastic volatility models, although this is strictly incorrect since at time t the volatility is completely pre-determined (deterministic) given previous values.

Exercise 67 *Let* $(X_t) \sim GARCH(p;q)$ *such that*

$$\begin{aligned} X_t &= \sigma_t Z_t && (5.28)\\ Z_t &\sim \quad iid\ N(0,1), && (5.29)\\ \sigma_t &= \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2. \end{aligned}$$

where the usual conditions on the parameter shold true.

1. Let η_t *denote* $\eta_t = \sigma_t^2\left(Z_t^2 - 1\right)$*.Show that* $E[\eta_t\,|I_{t-1}] = 0$ *where* I_{t-1} *is the information set at time t -1.*

2.Using the previous result and the representation of a GARCH(p;q), show that the stochastic process (X_t^2) *has the following representation :*

$$(X_t^2) = \omega + \sum_{i=1}^{p}(\alpha_i + \beta_i)X_{t-i}^2 + \eta_t - \sum_{i=1}^{q}\beta_i\eta_{t-j}$$

3.Interpret the previous result.

Exercise 68 *Let* $(X_t) \sim GARCH(1;1)$ *such that :*

$$\begin{aligned} X_t &= \sigma_t Z_t && (5.30)\\ Z_t &\sim \ \ iid\ N(0,1), && (5.31)\\ \sigma_t^2 &= \omega + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

$\omega > 0, \alpha$ *and* $\beta > 0 (with\ \alpha + \beta < 1)$. *We further assume*

$$1 - 3\alpha^2 - \beta^2 + 2\beta\alpha > 0$$

1. *Find the autocovariance and autocorrelation function of* X_t^2

References

[1] Bollerslev,T(1986).Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31,307-327.

[2]Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control. Holden-Day, Oakland, California, revised edition.

[3]Box, G. E. P., Pierce, D. A. and Newbold, P. (1987). Estimating and and growth rates in seasonal time series, Journal of the American Statistical Association, 82(397):276-282.

[4] Brockwell.PJ.and.Davis.RA.(1991).Time Series:Theory and Methods,(Seconded.).Springer.

[5] Engle,R.F.(1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation. Econometrica,50,987-1008.

[6]Irvine. K. N. and Eberhardt, A. J. (1992). Multiplicative seasonal AIUMA models for Lake Erie and Lake Ontario water levels, Water Resources Bulletin, 28(2):385-396.

[7] Leemis.L .(2023). Statistical Modeling: Regression, Survival Analysis, and Time Series Analysis (Paperback)

[8]Nelson.D.B(1991).Conditional Heteroskedasticity in Asset Returns: a New Approach. Econometrica,59,347-370.

[9]Sanchez. J. (2023) Time Series for Data Scientists. Cambridge University Press.

[10] William W. S. Wei. (2023)Multivariate Time Series Analysis and Applications (Hardcover).wiley.