PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA.
Ministry of Higher Education and Scientific Research.

MOULOUD MAMMERI UNIVERSITY, TIZI-OUZOU
Faculty of Sciences
Department of Mathematics

MASTER'S DISSERTATION
in
MATHEMATICS

Option
Probability and Statistics

Theme

Consistency and asymptotic normality of Maximum Likelihood estimators of Autoregressive Random Coefficient Models

Presented by
YACEF Cylia

In front of the examiners board :

HAMADOUCHE Djamal	Professor	UMMTO	President
HAMAZ Abdelghani	Senior Lecturer A	UMMTO	Advisor
ACHEMINE Farida	Senior Lecturer A	UMMTO	Examiner

Graduated on 07 / 07 / 2020

Acknowledgements

First of all, I would like to express my most sincere thanks and also my deepest appreciationto Mr **HAMAZ Abdelghani**, for having done me the honour of accepting to be my advisor, for his support, his patience and his availability, while closely following the elaboration of my dissertation.

I would also like to thank Mr **Hamadouche Djamel**, professor at UMMTO, who has honoured me by accepting the presidency of my dissertation jury.

I am very grateful to Mrs **ACHEMINE Farida** for her interest in my work and for agreeing to be a member of the jury.

The quality teaching provided by the training team of the Master's Degree in Probability and Statistics of the University Mouloud Mammeri of Tizi-Ouzou and all the teachers has nourished my reflections and represented a deep intellectual satisfaction, so thank you to the teacher-researchers.

I thank my dear parents, who have always been there for me, "You have sacrificed everything for my brother and me sparing neither health nor effort. You have given me a magnificent model of hard work and perseverance. I am indebted to an education that I'm proud".
A big thank you to my little brother Idir for your unlimited support.

Finally, I thank all my friends that i love so much, Yasmine, Tinhinane, Assia, Ferodja, Rezika, Lynda. For their sincere friendship and trust, to whom i owe my gratitude.
To all of you, I present my thanks, respect and gratitude.

Abstract

This dissertation is devoted to the study of some properties of the random coefficient autoregressive process. This process is commonly referred to as a sequence $\{X_t, t \in \mathbb{Z}\}$ *fully described by its past values multiplied by random coefficients and disturbed by white noise. We treat some problems of the study of such processes: stationarity, conditions of existence of a stationary solution and unknown parameters estimation. We end this work with simulations carried out by the R language.*

Keywords Autoregressive models, Random Coefficients autoregressive, estimation, Maximum likelihood estimation.

Contents

Introduction

Time series is a branch which is widely used in economics, its purpose is to study variables over time. Among its main objectives are the determination of trends within these series as well as the stability of the values (and their variation) over the time. A distinction is made between linear models (mainly AR and MA, for Auto-Regressive and Moving Average), conditional models (in particular ARCH, for Auto Regressive Conditional Heteroscedasticity) and in our case the random coefficient autoregressive model (RCA).
The analysis of these series touches a lot of fields of professional life, and more precisely that of decisional computing, medicine, astronomy, physics, biology, meteorology...

Classical formulations are not always appropriate for modeling this kind of models as the models considered are generally linear with constant coefficients.
In most situations, it is not expected to these models are the best class to fit a real data set. The advantage of this theory is that it is complete under classical assumptions. For this class of models, there are efficient algorithms to determine the order of the model and to obtain estimators of unknown parameters that have good statistical properties. But even when time series modeling by a linear model was at its peak, the demands of the field have shown that this modeling has become limiting. Indeed, many data series reflect asymmetry, regime changes, the presence of rupture, periodic dynamics...
For all these series, a mathematical framework must be found that is an accurate representation of the process that generated the observational data, i.e. an adequate mathematical framework. Hence the interest to consider random coefficient models which constitute a generalization of constant coefficient models.
These models emerged in the early 1970s, the RCA(1) model has been introduced and studied by several authors such as Andel(1976), Quinn and Nicholls(1983), Brandt(1986), Bougerol and Picard(1992).

Nicholls and Quinn (1982) studied systematically the probabilistic and statistical properties of the RCA(p) model and gave the stationarity conditions and process stability. They derived estimators of unknown parameters and established the asymptotic properties of these estimators.
Their method requires the compactness of the parameter space. Tjøstheim (1986) relaxed this assumption but retained the same moment conditions.

Subba Rao (2005) showed that a certain class of non-stationary random processes can be locally approximated by stationary processes. This class includes time-dependent coefficient RCA processes.

The non-linear models analysis has had a growing interest in the literature. Recent studies have been conducted by Sue et al (2006),Aue and al (2006) who adopted the maximum likelihood method to estimate the parameters of an RCA(1) process and showed that these estimators are highly consistent and asymptotically Gaussian.

Berkes et al. (2009) studied the quasi-maximum-likelihood estimation of the parameters of a non-stationary RCA(1) under second-order conditions with respect to innovation and perturbation processes.

This dissertation is organized as follows:

Chapter 1: Generalities about time series

In this chapter we introduce stochastic processes by developing elementary tools such as strong and weak stationarity. Also, we study the moving average autoregressive Models (ARMA) by giving their stationarity, causality and invertibility conditions. We focused on estimating the parameters of these models by studying their properties and developing various recursive methods of their computations.

Chapter 2: Random coefficient autoregressive processes of order 1

We are interested in studying a random coefficient autoregressive process of order one

RCA(1). We begin by giving some important concepts, in particular the conditions for the existence of a stationary solution, then we'll estimate the process unknown parameters by the quasi-maximum likelihood method.

Chapter 3: Random coefficient autoregressive processes of order p

We will generalize on a n-order random coefficient autoregressive process RCA(n). We begin by studying the stationarity, giving the conditions for the existence of a stationary solution, then we estimate the process unknown parameters by the maximum likelihood method.

Chapter 1

Generalities about time series

1.1 Introduction

So far, we have no analytical form for the autoregressive moving average (ARMA) parameters estimates, particularly those obtained by maximizing the likelihood function which is nonlinear. Therefore, researchers use the iterative optimization. However iterative method may fail to find the maximum because of the highly nonlinear likelihood function in general and direct methods of research require a large number of function evaluations.
several works were devoted to develop the likelihood function.Whittle and Durbin developed approximations to univariate MA models. Godolphin extends the method of Whittle's to the univariate ARMA models and shows how to calculate the parameter estimates directly from the sample autocorrelations of the observed data. Durbin's approximation generates an algorithm in which the MA coefficients are calculated from the coefficients of a long-order autoregressive. Hannan and Rissanen and and Koreisha and Pukkila extended the Durbin method to univariate and multivariate ARMA models, respectively. Several similar methods for approximating maximum likelihood have been proposed. Due to the non-linearity of the likelihood function, there is no analytical solution for estimating the maximum. This is also the reason why many algorithms were provided independently by Solo, Dugre, Shea, Mittnik, Mauricio and others. However, approaches used in previous work to maximize the likelihood function are iterative methods. Y. Boularouk, K. Djeddour (2015) proposed a new approximation of the likelihood function by a quadratic form, and calculate its maximum.

1.2 Stochastic processes and time series

1.2.1 Stochastic processes

A stochastic process represents a discrete or continuous-time evolution of several random variables. It makes it possible to model systems which their behavior is only partially predictable. The theory is based on the probability calculations and statistics.
The fields of application are very diverse: modeling and traffic management in transport networks and electrical systems, signal processing and filtering, quantum mechanics, chemistry, meteorology and even music, and more generally the complex technical systems management subject to random disturbances.

Definition 1.2.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $T \subseteq \mathbb{R}_+$ the parameters space, and $E \subset \mathbb{R}$. The process is in continuous time if T is continuous, (e.g $T = [0, \infty)$), and in discrete time if T is discrete (e.g., $T \subset \mathbb{Z}$).

A stochastic process is a measurable function $X: \begin{array}{l} T \times \Omega \longrightarrow E, \\ (t, w) \longmapsto X(t, w) = X_t(w) \end{array}$ noted as $\{X_t, t \in \mathbb{T}\}$.

It can be :

- A function of time, produced in the context of a random experiment. for example: music produced by the "shuffle" function on IPOD. The process represents music played in random order which is decided for all tracks at the times, in a continuous time.

- A family of random variables indexed by time. For example: position of a butterfly at times 0, 1, 2, 3...

  Example 1.2.1. We want to study the weather's evolution from one day to the next. Often the weather given day may depend on the previous day's weather. We have observations on days 0,1,.., and the weather conditions on day t: (status 0: sun, status 1: rain).
  Denote:

$$X_t = \begin{cases} 0 & \textit{sun on day } t \\ 1 & \textit{rain on day } t. \end{cases}$$

The stochastic process is given by: $\{X_t\} = \{X_0, X_1, ...\}$.

1.2.2 Time series

By language abuse, time series is meant a discrete process. A time series is a sequence of values that evolve on the time. To do the link with the discrete processes, we consider it as a realization of a process. So, the use of the theoretical results on the processes apply on the time series.

We consider that a time series $\{X_t\}$ is the result of different fundamental components:

- trend: represents the long-term evolution of the studied series. It translates the average behavior of the series.
- The seasonal component: corresponds to a phenomenon which is repeated at regular time (periodicals) intervals. In general, it's a seasonal phenomenon from where we got the seasonal variations term.
- The residual component: corresponds to irregular fluctuations, in general of low intensity but of random nature.
- Cyclical phenomenon: it's a repeating phenomenon but unlike the seasonality over periods which are not fixed and generally longer.

The time series study is very useful when looking to analyze, understand or even predict an evolving phenomenon over time. The goal is therefore to draw conclusions from the observed series. The steps to follow are:

1. Propose a probabilistic model in order to represent data.
2. Estimate the parameters of the chosen model and check the quality of adjustment to data (validation of the model).
3. Forecast using the model and evaluation of errors.

Time series are applied in several domains, such as:

- Finance (stock market, sale ...),
- Epidemiology (Monthly deaths due to pulmonary diseases...),
- Industry (production, consumption),
- Sociology (unemployment, strikes).

The following section is devoted to the concept of stationarity. This plays a fundamental role in the time series study.

1.3 Stationarity

One of the big questions in time series study is whether these follow a stationary process. The structure of the underlying process should be studied, whether or not it evolves over time. If the structure remains the same, the time series is then stationary, otherwise it is considered as non-stationary. The concept of stationarity is important in the time series modeling, the fallacious regression problem show that a linear regression with non-stationary variables is not valid.

Stationarity also plays an important role in predicting time series, the prediction interval is being different depending on whether the series is stationary or not.

In practice, weak stationarity is used more than strict stationarity.

Definition 1.3.1. Auto-covariance function

Let $\{X_t, t \in \mathbb{Z}\}$ be a process such that $V(X_t) < \infty$ for all $t \in \mathbb{Z}$, the auto-covariance function $\gamma_x(,)$ is defined by

$$\gamma_X(r,s) = E\left[(X_r - E(X_r))(X_s - E(X_s))\right] \quad r,s \in \mathbb{Z}.$$

1.3.1 Second order stationarity

A process $\{X_t, t \in \mathbb{Z}\}$ is weakly stationary or second order stationary if:

i) $E(X_t) = \mu$ (constant doesn't depend on t).

ii) $E(X_t^2) < +\infty,\ \forall\, t \in \mathbb{Z}$.

iii) $\gamma_X(r,s) = \gamma_X(r+t, s+t) \quad \forall h, t \in \mathbb{Z}$.

Remark 1.3.1. If $\{X_t, t \in \mathbb{Z}\}$ is stationary, then $\gamma_X(r,s) = \gamma_X(r-s,0) \quad \forall h, t \in \mathbb{Z}$. It is therefore more convenient to redefine the auto-covariance function of a stationary process as:

$$\gamma_X(h) = \gamma_X(h,0) = cov(X_{t+h}, X_t), \quad \forall r, s, t \in \mathbb{Z}.$$

1.3.2 Strong stationarity

A time series $\{X_t, t \in \mathbb{Z}\}$ is strictly stationary if its distribution does not change over time: $\forall \ t_1 \ < ... \ < t_k$

$$(X_{t_1}, X_{t_2}, ... \ X_{t_k}) \underset{\forall k,h}{\overset{\mathcal{D}}{\sim}} (X_{t_{1+h}}, X_{t_{2+h}}, ... \ X_{t_{k+h}}).$$

Remark 1.3.2. A strictly stationary process having its moments of order 2 finished is weakly stationary. The converse is not true in general.

- Counter-example

  Let $\{X_t\}$ be a sequence of independent random variables such that :

  $$\begin{cases} X_t \sim \varepsilon(1) & \textit{when } t \textit{ is peer} \\ X_t \sim N(1,1) & \textit{when } t \textit{ is odd.} \end{cases}$$

  $\gamma_X(0) = 1$ and $\gamma_X(h) = 0$ when $h \neq 0$ then, $\{X_t\}$ is stationary. However X_1 and X_2 doesn't have the same distribution, so, $\{X_t\}$ isn't strictly stationary.

There is an important class of processes for which the converse is verified.

Indeed, if a process $\{X_t, t \in \mathbb{Z}\}$ is second order stationary, then $E(X_t) = \mu$. Furthermore, we consider the indexes $t_1, ..., t_k$ so that $X = (X_{t_1}, ..., X_{t_k})^T$. So, the covariance matrix satisfies $\Gamma = (\gamma(t_i, t_j))_{i,j}$, and $\gamma(t_i, t_j)$ depends only on $|t_i - t_j|$.
So, $f_X(x)$ is only a function of deadlines $|t_i - t_j|$.

Then, under the stationarity assumption in the second order, the distribution of $X = (X_{t_1}, ..., X_{t_k})^T$ is only a function of deadlines.

So, a weakly stationary Gaussian process is stationary in the strict sense.

1.4 Second order linear process filtering and ARMA process

1.4.1 Second order stationary process filtering

We are interested in the properties of the obtained process $\{Y_t, t \in \mathbb{Z}\}$ as an image of the process $\{X_t, t \in \mathbb{Z}\}$ by the linear filter:

$$Y_t = \sum_{k=-\infty}^{+\infty} \psi_k \; X_{t-k}.$$

Theorem 1.4.1. *Let $(\psi_k)_k$ be an absolutely summable sequence $\sum_{k\geq 0} |\psi_k| < +\infty$ and $\{X_t, t \in \mathbb{Z}\}$ a second order stationary process with mean μ_X and autocovariance function $\gamma_X(h)$. So, the process $Y_t = \sum_{k=-\infty}^{+\infty} \psi_k \; X_{t-k}$ is second order stationary, with:*

$$\begin{cases} \mu_Y &= \mu_X \sum_{j=-\infty}^{+\infty} \psi_j \\ \gamma_Y(h) &= \sum_{k=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \psi_j \; X_k \; \gamma_X(h+k-j). \end{cases} \tag{1.1}$$

Before defining the moving average autoregressive model (ARMA), we will introduce the white noise, MA(q) and AR(p) process.

1.4.2 White noise

White noise is a realization of a random process in which the spectral density of power is the same for all frequencies in the bandwidth. It is a sequence of uncorrelated random variables with zero mean and constant variance.
It has an important role in modeling disturbances to signal measurements.

Definition 1.4.1. A process $\{\epsilon_t, t \in \mathbb{Z}\}$ is called white noise (pure random process) if it verifies

i) $E(\epsilon_t) = 0.$

ii) $V(\epsilon_t) = \sigma^2.$

iii) $cov(\epsilon_t, \epsilon_{t+h}) = 0 \; \forall \; |h| \; \neq \; 0.$

If moreover the t variables are independent, then it is a strong white noise.

1.4.3 MA(q) process

A moving average in t is a finite linear combination of values in the series corresponding to dates around t, so it performs a smoothing of the series.
These tools were the first to be used in time series study by the physicist Poynting in 1884, to eliminate accidental or periodic variations in a series.
Then, from 1914 onwards, the great figures of statistics such as Student, Pearson and Yule, for example, became interested in this type of problem.

Definition 1.4.2. A process $\{X_t; t \in \mathbb{Z}\}$ is a moving average model of order q if X_t is written as:

$$X_t = \epsilon_t + \sum_{j=1}^{q} \theta_j \; \epsilon_{t-j}, \tag{1.2}$$

where $\{X_t\} \sim \mathcal{BB}(0, \sigma^2)$ and $\theta_j \in \mathbb{R}$
with

- $\mu_X = \sum_{i=0}^{q} \theta_i \; E(\epsilon_{t-i}) = 0.$

- $\gamma_X(h) = \begin{cases} \sigma_\epsilon^2 \sum_{i=0}^{q-|h|} \theta_i \; \theta_{i+h} & if \; 0 \leq |h| \leq q \\ \sigma^2 & If \; not. \end{cases}$

1.4.4 AR(p) process

The autoregressive process is a regression model for a time series, in which the series is explained by its past values rather than by other variables.

A process $\{X_t, t \in \mathbb{Z}\}$ is a p-order autoregressive if $\{X_t\}$ is second order stationary, and if it is the solution to the recurrence equation:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + \epsilon_t, \tag{1.3}$$

where $\{\epsilon_t\} \sim \mathcal{BB}(0, \sigma^2)$, and $\phi_i \in \mathbb{R}$.

▶ **Stationary conditions**

A p-order autoregressive process is said to be second order stationary, if and only if it is the solution of the recurrence equation (1.3).

Let $\Phi(z)$ be the characteristic function of the process $\{X_t\}$ defined by:

$$\Phi(z) = z^p - \phi_1 z^{p-1} - ... - \phi_p,$$

and $\lambda_1, \lambda_2, ..., \lambda_p$ are the roots of this equation.
If $|\lambda_i| > 1$, then $\{X_t\}$ is stationary.

1.4.5 ARMA(p,q) process

It is a model that has been used extensively both theoretically and practically, which was introduced by Box and Jenkins (1976). It is a stochastic process describing time-varying random phenomena, which is written as:

$$X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}, \tag{1.4}$$

where $\theta_0 = 1$, $\{\epsilon_t\} \sim \mathcal{BB}(0, \sigma^2)$, and ϕ_i, $\theta_j \in \mathbb{R}$. Its characteristic functions are :

$$\begin{cases} \Phi(\epsilon) = 1 - \sum\limits_{i=1}^{p} \phi_i \ \epsilon^i \\ \\ \Theta(\epsilon) = 1 + \sum\limits_{j=1}^{q} \theta_j \ \epsilon^j. \end{cases} \tag{1.5}$$

▶ **Existence of a stationary solution**

To look for the existence of a stationary solution, we solve the system (1.5), if $\Phi(.)$ and $\Theta(.)$ have no common roots, then the equation admits a second order stationary solution if and only if $\Phi(z) \neq 0$ et $|z| = 1$.
This solution is unique and can be written as: $X_t = \sum_{j\in\mathbb{Z}} \psi_i \, \epsilon_{t-j}$,
where the ψ_j are coefficients to be developed from $\frac{\Theta(z)}{\Phi(z)} = \sum_{k\in\mathbb{Z}} \psi_k \, z^k$ converging in the crown $\{z \in \mathbb{C} | \delta_1 \leq |z| \leq \delta_2\}$, with

$$\begin{cases} \delta_1 = & max\,\{z \in \mathbb{C} ||z| < 1, \Phi(z) = 0\} \\ \delta_2 = & min\,\{z \in \mathbb{C} ||z| > 1, \Phi(z) = 0\}\,. \end{cases}$$

1.4.6 Causality

The notion of causality plays an important role in the time series study. In a time series system, knowing such causal relations allows us to know where to intervene in order to influence the system in the desired direction.

Theorem 1.4.2. *Let $\{X_t, t \in \mathbb{Z}\}$ be an ARMA(p,q), we assume that $\Phi(z)$ and $\Theta(z)$ have no common roots.*
If $\Phi(z) \neq 0$ for $|z| \leq 1$, then $\{X_t, t \in \mathbb{Z}\}$ admits a weakly stationary causal solution given by

$$X_t = \sum_{j=0}^{+\infty} \psi_j \epsilon_{t-j}. \tag{1.6}$$

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

$$(1 - \phi_z - ... - \phi_p z^p)(\psi_0 + \psi_1 z + ...) = 1 + \theta_1 z + ... + \theta_q z^q,$$

equation of $z^j,\ j = 0, 1...$, we find that

$$\begin{aligned} 1 &= \psi_0, \\ \theta_1 &= \psi_1 - \psi_0\phi_1, \\ \theta_2 &= \psi_1 - \psi_0\phi_1 - \psi_1\phi_2, \\ &\vdots \end{aligned}$$

or equivalently

$$\psi_j - \sum_{k=1}^{p} \phi_k \psi_{j-k} = \theta_j, \quad j = 0, 1, ... \tag{1.7}$$

where $\psi_0 = 1,\ \theta_j = 0\ for\ j > q$, and $\psi_j = 0\ for\ j < 0$.

The causality of the autoregressive polynomial is thus a sufficient condition to ensure the stationarity of the ARMA process satisfying the condition of uniqueness of the solution X_t.

1.4.7 Invertibility

The series is said to be invertible if it has an autoregressive writing, i.e if it is explained by its past values rather than by other variables.

Theorem 1.4.3. *Let $\{X_t\}_t$ be an ARMA(p,q), it's invertible and there is a sequence of numbers $(\Pi_j)_{j\in\mathbb{Z}}$ absolutely summable such that:*

$$\epsilon_t = \sum_{j=0}^{+\infty} \Pi_j X_{t-j}, \tag{1.8}$$

if and only if:

i) $\Phi(z)$ and $\Theta(z)$ have no common roots.

ii) $\Theta(z) \neq 0\ for\ |z| < 1$.

Interchanging the roles of (1.7) AR and MA polynomials, we find from that the sequence $\{\pi_j\}$ is determined by the equations

$$\Pi_j + \sum_{k=1}^{q} \theta_q \Pi_{j-k} = -\phi_j, \quad j = 0, 1, ... \tag{1.9}$$

where $\phi_0 = -1,\ \phi_j = 0,\ for\ j > p$, and $\Pi_j = 0\ for\ j < 0$.

Remark 1.4.1. An AR(p) process is always invertible, and any MA(q) process is causal.

1.5 Model identification and validation

For modelling the ARMA(p,q) model, we use the Box-Jenkins methodology. It suggests a three-step procedure: identification, estimation and model validation.

- In this first step, the objective is to determine from the observation of the simple and partial autocorrelation functions in the ARMA(p,q) models family the adequate model.

  Definition 1.5.1. The autocorrelation function (ACF) quantifies the linear influence of the time lag between two process observations by a classical correlation calculation. It is given by:

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}, \tag{1.10}$$

  with $\rho_X(h) \in [-1, 1]$.

  Definition 1.5.2. The Partial Autocorrelation Function (PACF) quantifies the exact influence of a past observation on the current value of the process by removing all intermediate observations from the study. It is given by $\alpha(h) = \phi_{h,h}$, where

$$\phi_{h,h} = \left(1 - \sum_{k=1}^{h-1} \phi_{h-1,k}\rho(k)\right)^{-1} \left(\rho(h) - \sum_{k=1}^{h-1} \phi_{h-1,k}\rho(h-k)\right), \quad \forall h \geq 2,$$

  with $\phi_{1,1} = \rho(1)$ and $\alpha(0) = 1$ pour tout $h \in \mathbb{N}$

  Partial auto-correlations are identically null beyond order p and auto-correlations are identically null beyond order q.

- Estimation: This step consists of estimating the parameters of the appropriate model selected.

- Model validation: This refers to various statistical tests of specification to check if the model is congruent, i.e. it cannot be put in default. These statistical tests consist in testing that the residuals of the estimated model do not exactly follow the white noise but are close to it, in other words the residuals must be autocorrelated and do not show heteroscedasticity.

1.6 Estimation of ARMA(p,q) parameters

Now, We'll proceed to the estimation of the unknown parameters of the ARMA(p,q).

1.6.1 Estimation of μ

An unbiased estimator of the mean μ of a stationary process $\{X_t, t \in \mathbb{Z}\}$ is the empirical mean:

$$\overline{X} = \frac{1}{n} \sum_{j=0}^{n} X_j. \tag{1.11}$$

Proposition 1.6.1. *Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary process with mean μ and autocovariance function γ, then*

$$E((\overline{X}_n - \mu)^2) = V(\overline{X}_n) \xrightarrow[n\to+\infty]{} 0 \quad if\ \gamma(n) \to 0.$$

$$n\ E((\overline{X}_n - \mu)^2) \xrightarrow[n\to+\infty]{} \sum_{h\in z} \gamma(h) \quad if\ \sum_{h\in z} |\gamma(h)| < +\infty.$$

Proof.

$$\begin{aligned} n\ V(\overline{X}_n) &= n\ E((\overline{X}_n - \mu)^2) \\ &= E((\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n} n\mu)^2) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E((X_i - \mu)(X_j - \mu)) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_X(i-j) \\ &= \sum_{k=-(n-1)}^{n-1} (1 - \frac{|h|}{n})\gamma_X(h). \end{aligned}$$

- If $\gamma(n) \to 0$, then according to Caesaro's theorem $\sum_{|h|<n} \frac{1}{n}|\gamma(h)| \to 0$, so $V(\overline{X}_n) \xrightarrow[n\to+\infty]{} 0$.
- If $\sum_{h\in\mathbb{Z}} |\gamma(h)| < +\infty$, then the dominated convergence theorem gives us the second result.

► **Confidence interval**

The following theorem is used to construct a confidence interval:

Theorem 1.6.1. *If* $\{X_t, t \in \mathbb{Z}\}$ *is a stationary process such that:*

$$X_t = \mu + \sum_{j\in\mathbb{Z}} \psi_j \epsilon_{t-j},$$

where ϵ_t *iid* $(0, \sigma_\epsilon^2)$, $\quad \sum_{j\in\mathbb{Z}} |\psi_j| < +\infty \quad$ *and* $\quad \sum_{j\in\mathbb{Z}} |\psi_j| \neq 0.$

Then $\sqrt{n}(\overline{X}_n - \mu) \sim N(0, \nu)$, *with* $\nu = \sum_{h\in\mathbb{Z}} \gamma_X(h)$. *So, we can construct the confidence interval of* μ.

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sqrt{\nu}} \sim N(0,1),$$

$$\mu \in \left[\frac{-1.96\sqrt{\nu}}{\sqrt{n}} + \overline{X}_n,\ \frac{1.96\sqrt{\nu}}{\sqrt{n}} + \overline{X}_n\right]. \tag{1.12}$$

1.6.2 Innovations algorithm for the ARMA process

To simplify the innovations algorithm, we apply it to the transformed process X_t such that

$$\begin{cases} w_t = \sigma_\epsilon^{-1}\ X_{t-1} & for\ \ t = 1,\ ...\ max(p,q) \\ w_t = \sigma_\epsilon^{-1}\ (X_t + \phi_1 X_{t-1} +\ \ ...\ \ + \phi_p X_{t-p}) & t > max(p,q). \end{cases} \tag{1.13}$$

Once the auto-covariances are known, we calculate the innovations recursively using the following algorithm:

$$\begin{cases} r_0 & = \frac{\gamma_X(0)}{\sigma_\epsilon^2} \\ \theta_{m,m-k} & = r_K^{-1}\left(\frac{\gamma_X(m-k)}{\sigma_\epsilon^2} - \sum_{j=0}^{j-1} \theta_{m,m-j}\theta_{n,n-j}\ r_j\right) \\ r_m & = \frac{\gamma_X(0)}{\sigma_\epsilon^2} - \sum_{j=0}^{j-1} \theta_m \theta_{m,m-j}\ r_j. \end{cases} \tag{1.14}$$

► **Application of the algorithm for MA(1) and ARMA(1,1)**

- Let $\{X_t, t \in \mathbb{Z}\}$ be an MA(1)defined by $X_t = \epsilon_t + \theta\epsilon_{t-1}$.
  We start by calculating r_0

$$\begin{aligned} r_0 &= \frac{\gamma_X(0)}{\sigma_\epsilon^2} \\ &= \frac{\sigma_\epsilon^2 \left(\sum_{j=0}^{1} \theta_j^2 \right)}{\sigma_\epsilon^2} \\ &= \frac{\sigma_\epsilon^2 \left(1+\theta^2\right)}{\sigma_\epsilon^2} \\ &= 1+\theta^2. \end{aligned}$$

So, by the innovations algorithm, the calculation of the r_i yields

$$\begin{cases} r_0 = & 1+\theta^2 \\ r_m = & 1+\theta^2 - \frac{\theta^2}{r_{m-1}}. \end{cases} \tag{1.15}$$

- If $\{X_t, t \in \mathbb{Z}\}$ is an autoregressive moving average ARMA(1,1) process, then the calculation of the r_i by the innovations algorithm yields

$$\begin{cases} r_0 = & (1+2\theta\phi+\theta^2)/(1-\theta^2) \\ r_m = & 1+\theta^2 - \frac{\theta^2}{r_{m-1}}. \end{cases} \tag{1.16}$$

1.6.3 Maximum likelihood estimation

Likelihood function maximization is the most widely used method for estimating parameters.

Let $\{X_t, t \in \mathbb{Z}\}$ be an ARMA(p,q):

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = \epsilon_t + \theta_1\epsilon_{t-1} + \ldots + \theta_q\epsilon_{t-q}, \tag{1.17}$$

X_t is assumed to be centered, otherwise it will be replaced by $\widehat{X}_t = X_t - \overline{X}$, and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ are independent and identically distributed.
Since ϵ_t are Gaussian, the likelihood function for a sample of size n is:

$$f(\Phi, \Theta, \sigma_\epsilon^2) = \frac{1}{\sqrt{(2\Pi\sigma^2)^n r_0 \ldots r_{n-1}}} \exp -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \frac{\epsilon_i^2(\Phi, \Theta)}{r_{i-1}}, \tag{1.18}$$

where
$\Phi = (\phi_1, \ \ldots, \phi_p)'$, $\Theta = (\theta_1, \ldots, \theta_q)'$, $\sigma_\epsilon^2 = \sigma^2 r_i$, σ^2 is the variance of the process.

To calculate the r_i we use the innovations algorithm. Maximizing f with respect to $\beta = (\Phi, \Theta)$ is equivalent to the minimization of

$$l_n(\Phi, \Theta) = \frac{1}{n} \sum_{i=1}^{n} \ln(r_{i-1}) + \ln \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\epsilon_i^2(\Phi, \Theta)}{r_{i-1}} \right). \tag{1.19}$$

This function has a non-linear form which implies the non-existence of an analytical solution.

To calculate this explicit form, we use :

$$\sum_{i=1}^{n} \frac{\epsilon_i^2(\Phi, \Theta)}{r_{i-1}} = \left[\left(L_\Theta^{-1} \ (L_\Theta X - F e_*) \right) \ A^{\frac{1}{2}} \right]^2, \tag{1.20}$$

where

$$L_\Phi = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & . & -\phi_p & 0 & . & . & 0 \\ 0 & 1 & -\phi_2 & . & -\phi_p & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 & . & . & 0 \\ 0 & 0 & 0 & . & -\phi_p & 0 & . & . & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & . & . & 1 \end{bmatrix}, \qquad L_\Theta = \begin{bmatrix} 1 & \theta_1 & \theta_2 & . & \theta_q & 0 & . & . & 0 \\ 0 & 1 & \theta_2 & . & \theta_q & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 & . & . & 0 \\ 0 & 0 & 0 & . & \theta_q & 0 & . & . & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & . & . & 1 \end{bmatrix},$$

$$F = \begin{bmatrix} H_p & M_q \\ 0 & 0 \end{bmatrix}, \qquad H_p = \begin{bmatrix} \phi_p & \phi_{p-1} & . & . & \phi_1 \\ 0 & \phi_p & . & . & \phi_2 \\ . & 0 & . & . & . \\ . & . & . & . & \phi_{p-1} \\ . & . & . & . & \phi_p \end{bmatrix}, \qquad M_q = \begin{bmatrix} \theta_p & \theta_{p-1} & . & . & \theta_1 \\ 0 & \theta_p & . & . & \theta_2 \\ . & 0 & . & . & . \\ . & . & . & . & \theta_{p-1} \\ . & . & . & . & \theta_p \end{bmatrix},$$

$$A = \sigma_\epsilon^2 \begin{bmatrix} r_0 & 0 & . & 0 \\ 0 & r_1 & . & 0 \\ . & 0 & . & . \\ 0 & 0 & . & r_n \end{bmatrix}, \qquad e_* = (X_0, ..., \ X_{1-p}, \epsilon_0, ..., \epsilon_{1-q})'.$$

In practice, the vector e_* is assumed to be zero.

► **Some properties of ARMA's likelihood function**

The likelihood function of an ARMA process has certain characteristics that allow us to calculate the maximum differently.
Through the proposal (1.6.1) and lemma (1.6.1), we'll prove the essence of these characteristics.

Proposition 1.6.2. *Let $\{X_t, t \in \mathbb{Z}\}$ be an ARMA(p,q), and $\beta = (\Phi, \Theta)$ the vector of unknown parameters. So, the $(r_i)_{i>0}$ verify the following properties:*

1) $r_i > 1; \forall i = 0, ..., n.$

2) $\frac{1}{n}\sum\limits_{i=1}^{n} \ln r_{i-1} \xrightarrow[n \to +\infty]{} 0.$

Proof. Let $\{X_t, t \in \mathbb{Z}\}$ be an invertible ARMA(p,q) given by $\epsilon_t = \sum\limits_{j=0}^{+\infty} \Pi_j X_{t-j}$.
We know that the random variables ϵ_t are independent and identically distributed, then:

$$
\begin{aligned}
\sigma_\epsilon^2 = V(\epsilon_{t+1}) \;\; &= E\left(\sum_{j=0}^{\infty} \Pi_j \; X_{t+1-j}\right)^2 \\
&\le \sigma_\epsilon^2 r_t \le \sigma_\epsilon^2 E\left(\sum_{j=0}^{t} \Pi_j \; X_{t+1-j}\right)^2 \\
&= \sigma_\epsilon^2 E\left(\epsilon_{t+1} - \sum_{j=0}^{\infty} \Pi_j \; X_{t+1-j} + \sum_{j=1}^{t} \Pi_j \; X_{t+1-j}\right)^2 \\
&= \sigma_\epsilon^2 E\left(\epsilon_{t+1} - \sum_{j=t+1}^{\infty} \Pi_j \; X_{t+1-j}\right)^2 \\
&= \sigma_\epsilon^2 \left(1 + E\left(\sum_{j=t+1}^{\infty} \Pi_j \; X_{t+1-j}\right)^2\right) \\
&\le \sigma_\epsilon^2 \left(1 + \left(\sum_{j=t+1}^{\infty} |\Pi|_j\right) \gamma_X(0)\right).
\end{aligned}
\tag{1.21}
$$

From the first inequality we deduce that $r_{i-1} \ge 1$ and $r_0 \ge 1$, hence $\prod_{i=1}^{n} r_i > 1$. Moreover, since $\left(\sum_{j=t+1}^{\infty} |\Pi|_j\right) \gamma_X(0) \to 0$ and $r_t \underset{t\to\infty}{\longrightarrow} 1$, then by the Césaro convergence

$$
0 \;\le\; \ln\,(\prod_{i=1}^{n} r_{i-1})^{\frac{1}{n}} \;=\; (\frac{1}{n}\sum_{j=1}^{n} \ln\; r_{i-1}) \underset{n\to\infty}{\longrightarrow} \; 0. \tag{1.22}
$$

Lemma 1.6.1. *For $\beta \in \mathbb{R}^{p+q}$, the likelihood function $l_n(\beta)$ is equal to a second order polynomial almost surely.*

Proof. From the property (1) of the proposition (1.5.1), it follows that $\sum_{j=1}^{n} \ln\; r_i$ is bounded by a real A positive.

i. Using relations (1), (2) and inequality $\ln(x) \le 1 + x$, it follows:

$$
l(\beta) \;\le A + \sum_{i=0}^{n} \epsilon^2 \;= b + \left[\left(L_\Theta^{-1}\;(L_\Theta X - F\widehat{e_*})\right)\right]^2,
$$

which is an integrable second-order polynomial.

ii. Since the sequence $(r_n)_n$ converges to 1, and $\frac{1}{n}\sum_{i=1}^{n}\ln(r_i) \xrightarrow[n\to+\infty]{} 0$, we conclude that $l_n(\beta)$ converges to a second order polynomial.

So, by the monotonous convergence theorem, the results (i) and (ii) imply that the log-likelihood function l_n is equal to a second order polynomial almost surely.
We will approach it by a second order polynomial and his minimum will be an approximation for the minimum of l_n.

1.6.4 Calculation procedure

To calculate the maximum likelihood, we follow these steps:

1- We start by calculating the initial estimation of the parameters using Yule-Walker's equations and other algorithms such as the Hannan-Rissanen algorithm.

2- We estimate the confidence intervals in the univariate context for each parameter using the initial values, then we take the product of these intervals $\prod_{i=1}^{p+q}]a_i, b_i[$.

3- Generation of p + q samples of m elements $[(\theta_{ij}), l_n(\theta_{ij})]_{j=1,m}$ from the confidence intervals of the parameters.

4- We make a numerical approximation by a second order polynomial using the least squares method.

5- We calculate the maximum.

1.6.5 Initial values

In this section, we will examine some initial estimation techniques for the parameters $(\phi_1, ..., \phi_p)$, $(\theta_1, ..., \theta_p)$ from the observations $x_1, ..., x_n$ of the causal process ARMA(p,q).

- Yule-Walker's procedure for pure autoregressive models, (we can adapt it to the models $q > 0$ except that they will be less effective).

- Burg's algorithm gives a higher value of likelihood than that obtained by Yule-Walker's equations.

- the Hannan-Rissanen algorithm for mixed models (($p > 0,\ q > 0$): is usually the best for finding causal models that are required to initialize the likelihood maximization.
- The innovations algorithm frequently gives a slightly higher value than the Hannan-Rissanen algorithm for the MA(q) process.

▶ **The Yule-Walker Equations**

Let $\{X_t\}_t$ be an AR(p) which admits a causal solution

$$X_t = \sum_{j=0}^{+\infty} \psi_j \, \epsilon_{t-j},$$

where : $\psi(\epsilon) = \sum_{j=0}^{+\infty} \psi_j \, \epsilon^t = \frac{1}{\phi(\epsilon)}$ pour $|\epsilon| \leq 1.$

We multiply each side by $X_{t-l},\ l = 0, \dots, p,$

$$X_{t-l}X_t - \phi_1 X_{t-l}X_{t-1} - \ \dots \ - \phi_p X_{t-l}X_{t-p} = X_{t-l}\epsilon_t.$$

We take the expectation, and we get:
$E(X_{t-l}X_t) - E(\phi_1 X_{t-l}X_{t-1}) - \ \dots \ - E(\phi_p X_{t-l}X_{t-p}) = E(X_{t-l}\epsilon_t).$

$\Rightarrow \gamma_X(l) - \phi_1\gamma_X(l-1) - \ \dots \ - \phi_p\gamma_X(l-p) = 0.$

$\Rightarrow \gamma_X(l) = \phi_1\gamma_X(l-1) + \ \dots \ + \phi_p\gamma_X(l-p).$

We get the following system:

$$\begin{cases} \phi_1\gamma_X(0) + \phi_2\gamma_X(1) + \ \dots \ + \phi_p\gamma_X(p-1) & = \gamma_X(1), \\ \phi_1\gamma_X(1) + \phi_2\gamma_X(2) + \ \dots \ + \phi_p\gamma_X(p-2) & = \gamma_X(2), \\ \phi_1\gamma_X(p-1) + \phi_2\gamma_X(p-2) + \ \dots \ + \ \phi_p\gamma_X(0) & = \gamma_X(p-1). \end{cases} \quad (1.23)$$

▶ **Matrix writing:**

$$\Gamma_p \Phi = \gamma_p. \quad (1.24)$$

With: $\Phi = \begin{pmatrix} \phi_1 \\ . \\ . \\ . \\ \phi_p \end{pmatrix}$, $\gamma_p = \begin{pmatrix} \gamma_1 \\ . \\ . \\ . \\ \gamma_p \end{pmatrix}$, and $\Gamma_p = \begin{bmatrix} \gamma_0 & . & . & . & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & . & . & \gamma_{p-2} \\ . & . & . & . & . \\ . & . & . & . & . \\ \gamma_{p-1} & . & . & . & \gamma_0 \end{bmatrix}$.

⋇ **Estimation of σ^2 and ϕ_j:**

- To estimate σ^2, we have just to multiply each side of the process by X_t (case where $l = 0$), and proceed to expectations to get the equation of $\widehat{\sigma}^2$:

$$X_t X_t - \phi_1 X_t X_{t-1} - \ ... \ - \phi_p X_t X_{t-p} = X_t X_t.$$

$$\gamma_X(0) - \phi_1 \gamma_X(1) - \ ... \ - \phi \gamma_X(p) = \widehat{\sigma}^2. \tag{1.25}$$

- So the estimator $\widehat{\phi}$ of ϕ, is given by:

$$\widehat{\phi} = \widehat{\Gamma}_p^{-1}\, \widehat{\gamma}_p. \tag{1.26}$$

▶ **Hannan-Rissanen's Algorithm**

The equations defined for a causal ARMA(p,q) model has the form of a linear regression model with a coefficients vector $(\phi_1,\ ..., \phi_p)'$. This suggests the use of simple least squares regression to obtain an initial estimate.

The application of this technique is complicated by the fact that in the general ARMA equations, X is generated not only on x but also on z.

The application of this technique is complicated when $(q > 0)$ by the fact that in the equations of an ARMA(p,q), X_t is written not only according to its past but also according to $(\epsilon_{t-1},\ ..., \epsilon_{t-q})$.

Nevertheless, it is still possible to apply a least squares regression to the estimate of ϕ and θ by replacing the unobserved quantities $(\epsilon_{t-1},\ ...\ , \epsilon_{t-q})$ by their estimated values

$(\widehat{\epsilon}_{t-1},\ \dots\ ,\widehat{\epsilon}_{t-q})$.
The parameters ϕ and θ will then be estimated by rewriting X_t as a function of $X_{t-1},\ \dots, X_{t-p}$, $\widehat{\epsilon}_{t-1},\ \dots, \widehat{\epsilon}_{t-q}$

Now, we give the main steps of the Hannan-Rissanen estimation procedure.

1) Fitting a higher order AR(m) model ($m > max(p,q)$) using Yule-Walker's equations. If $(\widehat{\phi}_{m1},\ \dots, \widehat{\phi}_{mm})$ is the estimated coefficients vector, the residuals are then estimated from the equations $\widehat{\epsilon}_t = X_t - \widehat{\phi}_{m1}X_{t-1} -\ \dots\ - \widehat{\phi}_{mm}X_{t-m}$ for $t = m+1, n$

2) Once the residue $(\widehat{\epsilon}_t)_t$ estimates are calculated, the parameter vector $\beta = (\Phi, \Theta)$ is estimated using the least squares method through a linear regression of X_t as a function of de $X_{t-1},\ \dots\ , X_{t-p}, \epsilon_{t-1},\ \dots, \epsilon_{t-q}$ for $t = \overline{m+q+1, n}$ by minimizing the sum of squares

$$S(\beta) = \sum_{t=m+q+1}^{n} \left(X_t - \phi_1 X_{t-1} -\ \dots\ - \phi_p X_{t-p} - \theta_1 \widehat{\epsilon}_{t-1} -\ \dots\ - \theta_q \widehat{\epsilon}_{t-q})^2 \right), \quad (1.27)$$

with respect to β.
Then, the hannan-rissanen estimator is given by:

$$\widehat{\beta} = (\epsilon' \epsilon)^{-1}\ \epsilon' Y_n, \quad (1.28)$$

with:
$Y_n = (X_{m+1+q},\ \dots\ X_n)'$ and

$$\epsilon = \begin{bmatrix} X_{m+q} & X_{m+q-1} & \cdots & X_{m+q+1-p} & \widehat{\epsilon}_{m+q} & \widehat{\epsilon}_{m+q-1} & \cdots & \widehat{\epsilon}_{m+1} \\ X_{m+q+1} & X_{m+q} & \cdots & X_{m+q+2-p} & \widehat{\epsilon}_{m+q+1} & \widehat{\epsilon}_{m+q} & \cdots & \widehat{\epsilon}_{m+2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ X_{n-1} & X_{n-2} & \cdots & X_{n-p} & \widehat{\epsilon}_{n-1} & \widehat{\epsilon}_{n-2} & \cdots & \widehat{\epsilon}_{n-q} \end{bmatrix}.$$

The Hannan-Nissan estimator of the white noise variance is

$$\widehat{\sigma}^2_{HR} = \frac{S(\beta)}{n-m-q}. \quad (1.29)$$

3) Hannan and Rissenan included a third step to improve the estimate.
Using the estimator $\widehat{\beta} = (\widehat{\phi}_1,\ ...,\widehat{\phi}_p,\widehat{\theta}_1,\ ...,\widehat{\theta}_q)'$ from the 2nd step, we have:

$$\widetilde{\epsilon}_t = \begin{cases} X_t - \sum_{i=1}^{p} \widehat{\phi}_j X_{t-j} - \sum_{i=1}^{q} \widehat{\theta}_j \widehat{\epsilon}_{t-j} & if\ t > max(p,q) \\ 0 & if\ t \leq max(p,q). \end{cases} \tag{1.30}$$

For $t = 1, ..., n$, Hannan and Rissenan posed:

$$V_t = \begin{cases} \sum_{i=1}^{p} \widehat{\phi}_j V_{t-j} + \widetilde{\epsilon}_t & if t > max(p,q) \\ 0 & if t \leq max(p,q), \end{cases}$$

and

$$W_t = \begin{cases} -\sum_{i=1}^{p} \widehat{\theta}_j W_{t-j} + \widetilde{\epsilon}_t & if t > max(p,q) \\ 0 & if t \leq max(p,q). \end{cases}$$

We note that V_t and W_t are autoregressive. If $\widehat{\beta}^*$ is the estimator of β found by writing $\widetilde{\epsilon}$ as a function of $(V_{t-1},\ ...,V_{t-p},W_{t-1},\ ...,W_{t-q})$, i.e $\widehat{\beta}^*$ minimizes

$$S^*(\beta) = \sum_{t=max(p,q)+1}^{n} \left(\widetilde{\epsilon}_t - \sum_{j=1}^{p} \beta_j V_{t-j} - \sum_{k=1}^{q} \beta_{k+p} W_{t-k} \right)^2 . \tag{1.31}$$

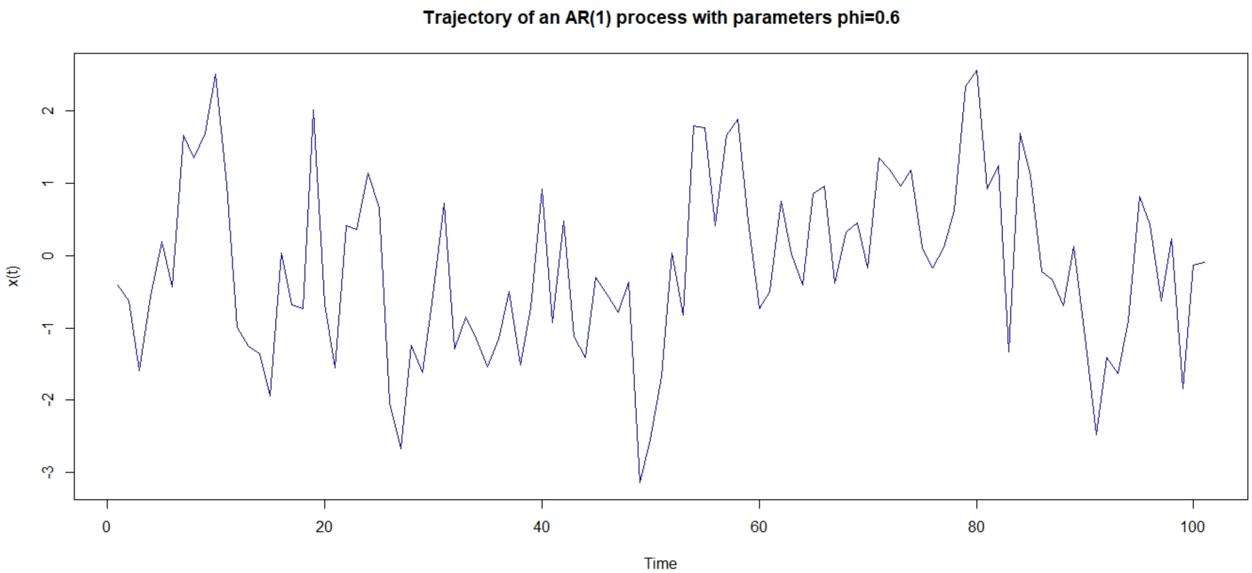
So the estimator of β is $\widetilde{\beta} = \widehat{\beta}^* + \widehat{\beta}$. The new estimator $\widetilde{\beta}$ has the same asymptotic efficiency as the maximum likelihood estimator.

In interactive time series modelling, using the model produced by step 2 as the initial model for the calculation of the maximum likelihood estimator itself, the third step is eliminated.

1.6.6 Application

We generated a first-order autoregressive process $\{X_t\}_{t=1,100}$ with parameter $\phi = 0.6$.

We plot this time series as a function of time



Maximization of the log likelihood function of an AR(1) is equivalent to the minimization of

$$l_n(\Phi,\Theta) = \frac{1}{n}\sum_{i=1}^{n}\ln\ r_{i-1} + \ln\sum_{i=1}^{n}(\tfrac{1}{n}[(L_\Theta^{-1}(L_\Phi(X-\overline{X})-Fe_*))A^{1/2}]^2) \quad (1.32)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\ln\ r_{i-1} + \ln\left(\sum_{i=1}^{n}\frac{(x_i-\phi x_{i-1})^2}{r_{i-1}}\right),$$

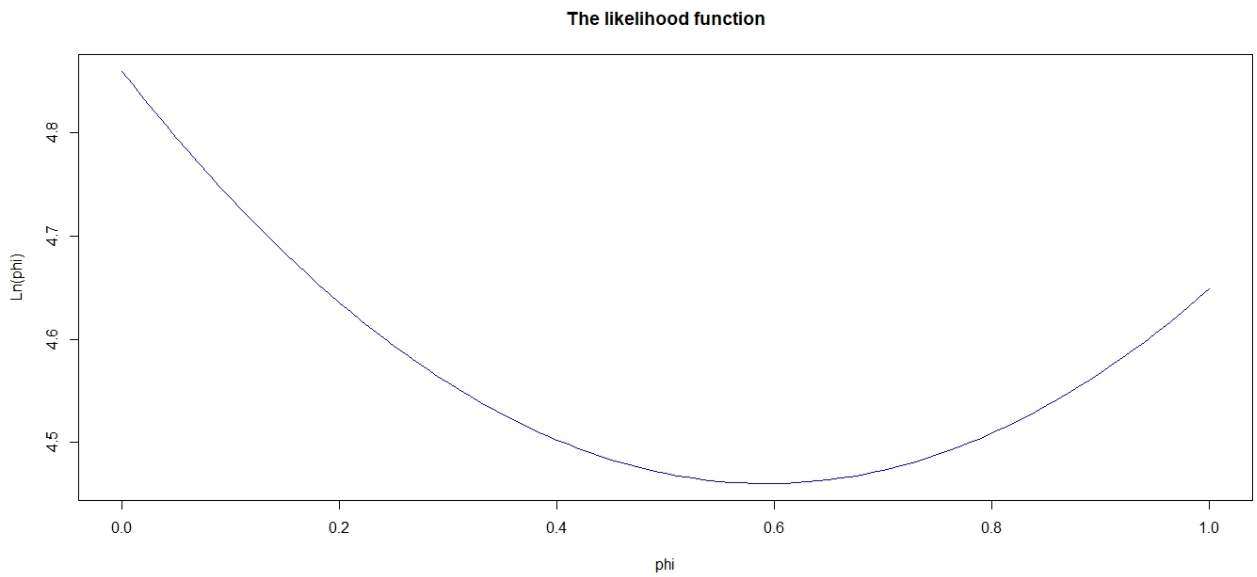
where $\overline{X}$ is the empirical mean,$L_\Theta = I_n$, e_* is estimated by $\widehat{e_*} = (0,...,0)$, is the matrix defined in section (1.6.3) with $\phi_1 = \phi$, $\phi_i = 0$ for all $i > 1$ and $r_i = 1$ for all $i \geq 1$.

We used Yule-Walker's equations to get $\phi_{initial} = 0.6011303$.
Then we used the asymptotic auto-covariance matrix to calculate the confidence region
We got $]0.3844883, 0.8177723[$.

From this interval we generate a sample of $m = 100$ values of ϕ_i, for which we calculate the values of our function l_n and so obtain our sample $(\phi_i, y_i = l_n(\phi_i))_{i=1,100}$.

ϕ	0.5809	0.4809	0.5469	0.7269	0.5609
$l_n(\phi)$	4.64905	4.63428	4.64114	4.71454	4.64404
ϕ	0.7259	0.6019	0.4649	0.7499	0.6309
$l_n(\phi)$	4.71393	4.76578	4.63433	4.72911	4.71700
ϕ	0.5539	0.5849	0.5019	0.8119	0.6659
$l_n(\phi)$	4.64253	4.65017	4.63523	4.78116	4.68125
ϕ	0.5169	0.4079	0.6419	0.6779	0.5179
$l_n(\phi)$	4.77725	4.63991	4.72393	4.64622	4.77803
ϕ	0.4199	0.4789	0.5139	0.4089	0.5989
$l_n(\phi)$	4.63804	4.63425	4.63630	4.63974	4.65441
ϕ	0.4559	0.7229	0.5699	0.6889	0.4259
$l_n(\phi)$	4.63465	4.71212	4.64617	4.69286	4.78431
ϕ	0.3889	0.4529	0.6079	0.6989	0.7049
$l_n(\phi)$	4.64525	4.63480	4.65739	4.76353	4.70162
ϕ	0.4009	0.5369	0.8009	0.4429	0.5409
$l_n(\phi)$	4.64118	4.63938	4.76503	4.63548	4.64005
ϕ	0.5549	0.4959	0.6249	0.5289	0.5359
$l_n(\phi)$	4.64274	4.63484	4.78352	4.63815	4.63921
ϕ	0.4799	0.5939	0.7679	0.5389	0.8019
$l_n(\phi)$	4.63426	4.64257	4.74124	4.63971	4.78909
ϕ	0.6759	0.6089	0.5059	0.5899	0.7209
$l_n(\phi)$	4.68615	4.70389	4.63555	4.65163	4.71092
ϕ	0.3899	0.4189	6 0.4049	0.6939	0.6539
$l_n(\phi)$	4.64829	4.63818	4.64044	4.75982	4.67567
ϕ	0.6289	0.5299	0.4829	0.4029	0.4059
$l_n(\phi)$	4.78669	4.63829	4.63432	4.64081	4.64026
ϕ	0.5039	0.4769	0.6399	0.6689	0.5159
$l_n(\phi)$	4.63539	4.63423	4.66958	4.68270	4.63651
ϕ	0.6899	0.6139	0.6899	0.6049	0.4849
$l_n(\phi)$	4.75688	4.65949	4.69339	4.65638	4.63437
ϕ	0.7179	0.5259	0.8049	0.7359	0.4099
$l_n(\phi)$	4.70913	4.63773	4.76804	4.72012	4.63958
ϕ	0.5479	0.5379	0.5029	0.6709	0.4459
$l_n(\phi)$	4.64133	4.63954	4.76653	4.68367	4.63525
ϕ	0.6629	0.6599	0.7709	0.4629	0.6339
$l_n(\phi)$	4.65023	4.73577	4.74332	4.63438	4.66711
ϕ	0.8069	0.6449	0.6809	0.5239	6 0.6119
$l_n(\phi)$	4.76956	4.67170	4.68869	4.63747	4.77338
ϕ	0.6209	0.6019	0.4639	0.5089	0.4879
$l_n(\phi)$	4.66205	4.65538	4.64995	4.63581	4.63447

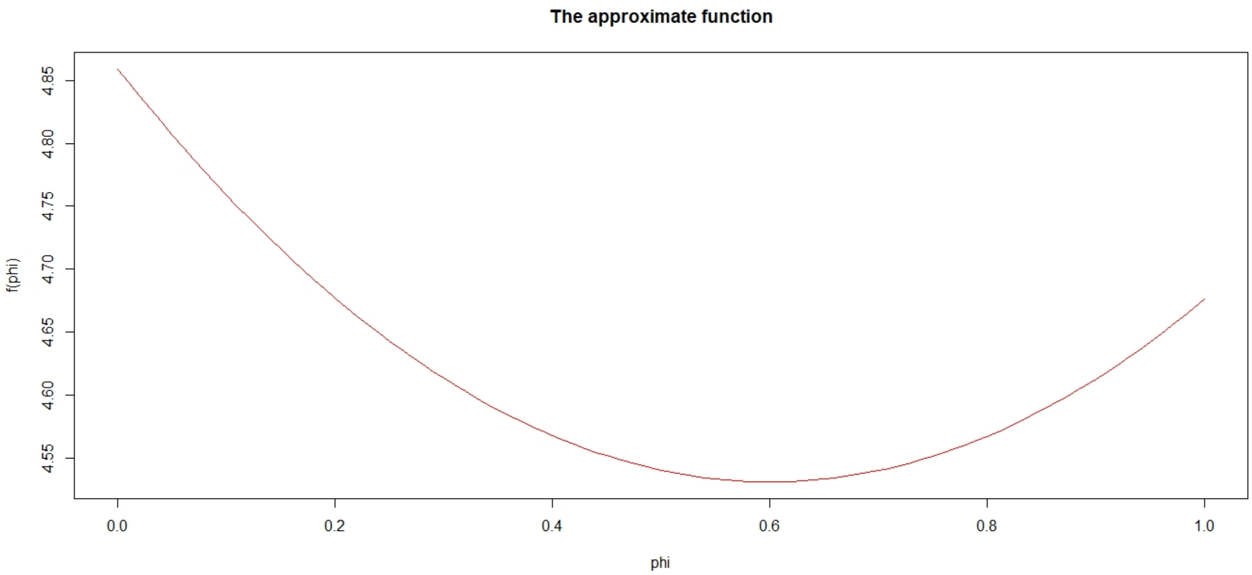
The likelihood function representation is



Using the method of least squares we obtain the approximation of the likelihood function:

$$f(\phi) = 0.9094810757\phi^2 - 1.092486010\phi + 4.8590633527.$$

Therefore the approximation of ϕ is $\widehat{\phi} = \frac{-b}{2a} = \frac{1.092}{2\times 0.909} = 0.6006601$. To illustrate the quality of the approximation of the log-likelihood function, we give its graphical representation below:


The approximate function
f(phi)
4.55
4.60
4.65
4.70
4.75
4.80
4.85
0.0
0.2
0.4
0.6
0.8
1.0
phi

Chapter 2

Random coefficient autoregressive processes of order 1

2.1 Introduction

Over the past three decades, there has been an interest increasing for non-linear time series models. One of the first examples is the random coefficient model introduced and studied by Nicholls and Quinn (1982). The first-order random coefficient autoregressive model, abbreviated as RCA(1), is given by the equations

$$X_k = (\varphi + b_k)X_{k-1} + e_k, \qquad k \in \mathbb{Z}. \tag{2.1}$$

where φ is a real parameter. We assume also, throughout this chapter, that

$$\begin{array}{l}\{(b_k, e_k); k \in \mathbb{Z}\} \text{ are pairs of random variables independent and identical} \\ \text{distributed defined on the same probabilistic space } (\Omega, \mathcal{F}, P).\end{array} \tag{2.2}$$

The time series $\{X_k; k \in \mathbb{Z}\}$ satisfying (2.1) have been used in the context of random perturbations of dynamical systems and have found a variety of applications, for example in finance and biology (cf. Nicholls and Quinn, 1982; and Tong, 1990). In this chapter, we are interested in estimation by the quasi-maximal likelihood method of unknown parameters of an RCA(1) process, based on the work of Aue *et al* (2006). The procedure was initially introduced by Quinn and Nicholls (1981) for more general RCA processes (see Chapter 4 of the Nicholls and Quinn, 1982), who established a strong consistency and central limit theorem. However, their results require additional restrictive assumptions, such as the existence of second order moments of random variables in b_k and e_k. In contrast, Aue, A., Horvath, L., Steinebach, J. (2006) imposed conditions natural (and minimal) sequences only at $\{e_k\}$ and $\{b_k\}$. Other studies in this direction have been conducted by Schick

(1996), which dealt with estimating the deterministic coefficient φ in (2.1). His method could, in principle, be extended to obtain similar results for the joint estimation of all parameters.

This chapter is organized as follows. In section 2, we discuss some basic properties of RCA(1) time series such as the necessary and sufficient conditions for the existence and uniqueness of a strictly stationary solution to (2.1) as well as the existence of moments. The main results are given in section 3, where the high consistency (almost certain convergence) and the asymptotic normality of the quasi-maximum likelihood estimator are obtained under optimal assumptions.

2.2 RCA(1) Model Structure

We assume that the process $\{X_k; k \in \mathbb{Z}\}$ satisfies the equation (2.1), so the necessary and sufficient condition for the existence and uniqueness of a solution can be derived from Bougerol and Picard (1992). Pioneering studies related to the bilinear equations, which can also be applied to RCA processes (1), were conducted by Quinn (1980, 1982) who studied the conditions of low stationarity of $\{X_k; k \in \mathbb{Z}\}$ without assuming (2.2). First, we study the properties of

$$X = \sum_{i=0}^{\infty} e_{-i} \prod_{j=0}^{i-1} (\varphi + b_{-j}). \tag{2.3}$$

Let $\log^+(x) = \max\{\log(x), 0\}$ be the positive part of the logarithm.

Lemma 2.1. *We assume that* (2.2) *is verified,*

$$\mathbf{E}\left[\log^+ |e_0|\right] < \infty \text{ and } \mathbf{E}\left[\log^+ |\varphi + b_0|\right] < \infty. \tag{2.4}$$

i) *If*

$$-\infty \leq \mathbf{E}\left[\log |\varphi + b_0|\right] < 0, \tag{2.5}$$

then X defined in (2.3) *converges absolutely (a.s.).*

ii) *Conversely, if $P(\{e_0 = 0\} < 1$ and $P\{|X| < \infty\} = 1$, the* (2.5) *is verified.*

Proof. .

i) We first assume that (2.5) holds. By the strong law of large numbers, there is a random variable i_0 such that

$$\log|\varphi+b_{-1}|+\log|\varphi+b_{-2}|+...+\log|\varphi+b_{-i}|\leq\frac{1}{2}\gamma,\qquad if\ i\geq i_0,\tag{2.6}$$

when $-\infty<\gamma=E\log|\varphi+b_0|<0$. Using (2.6), we'll have

$$|X|\leq\sum_{i=0}^{i_0}|e_{-i}|\prod_{j=0}^{i-1}\log|\varphi+b_{-j}|+\prod_{k=0}^{i_0}\log|\varphi+b_{-k}|\sum_{i=i_0+1}^{\infty}|e_{-i}|e^{i\gamma/2},\tag{2.7}$$

The lemma 2.2 of Berkes and al (2003) shows that

$$p\left\{\sum_{i=0}^{\infty}|e_{-i}|e^{i\gamma/2}<\infty\right\}=1,$$

which completes the lemma 1.1 when $-\infty<E\log|\varphi+b_0|<0$.

if $E\log|\varphi+b_0|=-\infty$, then (2.7) Checked for all $\gamma<0$ (cf. Chow and Teicher, 1997, p. 125-126), then the argument of (2.9) can be used again.

ii) Since $P\{|e_0|=0\}<1$ then, there's a constant $a>0$ such that $P\{|e_0|\geq a\rangle 0$. We define the event A_k as

$$A_k=\left\{\omega:\left(|e_{-k}|,\prod_{i=0}^{k-1}|\varphi+b_{-i}|\right)\in[a,\infty)\times[1,\infty)\right\},\quad k\in\mathbb{N}.$$

Next, we introduce a growing string of tribes defined by $\mathcal{F}_0=\{\emptyset,\Omega\}$
and $\mathcal{F}_k=\sigma((e_i,b_i),-k\leq i\leq 0)$. On notes that$A_k\in\mathcal{F}_k$.
Applying (2.2), we get

$$P\{A_k|\mathcal{F}_k\}=P\{|e_0|\geq a\}I\left\{\sum_{i=0}^{k-1}\log|\varphi+b_{-i}|\geq 0\right\}.$$

Hence

$$\sum_{k=1}^{\infty}P\{A_k|\mathcal{F}_k\}=P\{|e_0|\geq a\}\sum_{k=1}^{\infty}I\left\{\sum_{i=0}^{k-1}\log|\varphi+b_{-i}|\geq 0\right\}.\tag{2.8}$$

We're going to show that

$$\sum_{k=1}^{\infty}I\left\{\sum_{i=0}^{k-1}\log|\varphi+b_{-i}|\geq 0\right\}=\infty\ a.s.\tag{2.9}$$

If $E\log|\varphi + b_0| > 0$, then according to the strong law of large numbers, (2.9) is verified.

If $E\log|\varphi + b_0| = 0$, then Chung-Fuchs' law (cf. Chow and Teicher, 1997, p. 154) yields (2.9).

Corollary 3.2 of Durrett (1996, p. 240), (2.8) and (2.9) gives
$P\{A_k\ infinitely\ often\} = 1$, and

$$P\left\{\lim_{k\to\infty} e_k \prod_{j=0}^{k-1}(\varphi + b_{-j}) = 0\right\} = 0.$$

This contradicts the existence of X, and completes the proof of Lemma 1.1.

■

Lemma 2.2. *We assume that* (2.2) *and* (2.5) *are verified, and that* $E|e_0|^\epsilon < \infty$ *and* $E|b_0|^\epsilon < \infty$ *with* ϵ *>0. So, there's a* δ *>0 such that* $E|X|^\delta < \infty$.

Proof. Let $M(t) = E|\varphi + b_0|^t$, we note that $0 \le t \le \epsilon$. $M(0) = 1$ and $M(0^+)$ =1 (based on (2.5)), so M is decreasing towards a neighborhood of 0. Then there is $\delta > 0$ such that $M < 1$.
. We can assume, without loss of generality, that $0 < \delta \le 1$.
We know that $(a+b)^\delta \le a^\delta + b^\delta\ for\ tout\ a, b \ge 0$ by concavity,

$$|X|^\delta \le \left(\sum_{i=0}^{\infty} |e_{-i}| \prod_{j=0}^{i-1} |\varphi + b_{-j}|\right)^\delta \le \sum_{i=0}^{\infty} |e_{-i}|^\delta \prod_{j=0}^{i-1} |\varphi + b_{-j}|^\delta.$$

Based on (2.2) and $M(\delta)$ <1, we get

$$E|X|^\delta \le E|e_0|^\delta \sum_{i=0}^{\infty} E\left[\prod_{j=0}^{i-1} |\varphi + b_{-j}|^\delta\right] = E|e_0|^\delta \sum_{i=0}^{\infty} M^i(\delta) < \infty.$$

■

The following lemma is a generalization of lemma 1.2, it provides a condition for the existence $E|X|^\nu$, where $\nu \ge 1$.

Lemma 2.3. *Assume that* (2.2) *and* (2.5) *are verified for* $\nu \ge 1$,
and that $E|e_0|^\nu < \infty$, $E|b_0|^\nu < \infty$ *and* $E|\varphi + b_0|^\nu < \infty$. *so,* $E|X|^\nu < \infty$.

Proof. Using the fact that$\{(b_k, e_k); k \in \mathbb{Z}\}$ are pairs of independent random variables and identically distributed defined on the same probabilistic space $(\Omega, \mathcal{F}, P)$. And by applying Minkowski's inequality (cf. Chow and Teicher, 1997, p. 110), we get

$$(E|X|^{\nu})^{1/\nu} =\leq (E|e_0|^{\nu})^{1/\nu} \sum_{i=0}^{\infty} (E|\varphi + b_0|^{\nu})^{1/\nu} < \infty.$$

■

Remark 2.2.1. If e_0 and b_0 are independent, $E(b_0) = E(e_0) = 0$, $E(b_0^2) = \omega^2$, and $E(e_0^2) < \infty$, then $E(X^2) < \infty$ if and only if $\varphi^2 + \omega^2 < 1$, according to (Andél (1976) of Nicholls and Quinn (1982), p. 31).)

Theorem 2.1 gives us the necessary and sufficient condition of the existence of a unique solution of (2.1). The solution of (2.1) has the same distribution as X, and we see that in theorem 2.1 that if X exists, then it is unique and strictly stationary. We say that $\{X_k; k \in \mathbb{Z}\}$ is a non-anticipative solution of (2.1), if X_j is independent of $\{(b_k, e_k); k \in \mathbb{Z}\}$ for all$j \in \mathbb{Z}$.

Theorem 2.1. *We suppose that* (2.1) *and* (2.4) *are verified.*

i) *If* (2.5) *is satisfied, then*

$$X_k = \sum_{i=0}^{\infty} e_{k-i} \prod_{i=0}^{i-1} (\varphi + b_{k-j}),$$

converges absolutely a.s, and the process $\{X_k; k \in \mathbb{Z}\}$ *is the unique stationary non anticipative solution of* (2.1).

ii)

$$\text{If } P\{c_1 b_0 + c_2 e_0 = c_3\} < 1 \text{ for all real } c_1, c_2, c_3 \text{ such that } (c_1, c_2) \neq (0,0), \text{ and (2.1) has a non-anticipative solution, so (2.5) is verified.} \quad (2.10)$$

Proof. The first part of the theory is an immediate consequence from lemma 2.1. The condition (2.10) indicates that X_k is irreducible in the sense of Bougerol and Picard (1992), so their Theorem 2.5 therefore implies the second part.

2.3 Parameter estimation

We assume that

$$\mathbf{E}[b_0, e_0] = (0, 0) \text{ and } \mathrm{Cov}(b_0, e_0) = \begin{bmatrix} \omega^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}, \ \omega^2, \sigma^2 > 0. \tag{2.11}$$

In this section, we are interested in estimating the parameters vector $\theta = (\varphi, \omega^2, \sigma^2)$ based on the quasi-maximum likelihood method. We note that for $k \in \mathbb{Z}$,

$$\mathbf{E}\left(X_t \backslash \mathcal{G}_{k-1}\right) = \varphi X_{k-1} \text{ and } \ \mathrm{var}(X_k \backslash \mathcal{G}_{k-1}) = \omega^2 X_{k-1}^2 + \sigma^2,$$

where $\mathcal{G}_{k-1} = \sigma((b_i, e_i) : i \leq k-1$.

Thus, if the vector (b_k, e_k) is Gaussian, then the conditional likelihood can be given by

$$L_n(\mathbf{u}) = \prod_{i=1}^{n} \left(\frac{1}{2\pi(xX_{i-1}^2 + y)} \right)^{\frac{1}{2}} \exp\left(-\frac{(X_i - sX_{i-1})^2}{2(xX_{i-1}^2 + y)} \right),$$

where $\mathbf{u} = (s, x, y)$. The estimator of the quasi-maximum likelihood $\widehat{\theta}$ is defined as

$$\sup_{u \in \Gamma} L_n(\mathbf{u}) = L_n(\widehat{\theta}), \tag{2.12}$$

with a certain choice of $\boldsymbol{\Gamma} \subset \mathbb{R}^3$. It is more convenient to work with the log-likelihood function

$$l_n(\mathbf{u}) = \frac{1}{n} \sum_{n=1}^{n} g_i(\mathbf{u}), \quad g_i(\mathbf{u}) = \frac{(X_i - sX_{i-1})^2}{xX_{i-1}^2 + y} + \log(xX_{i-1}^2 + y).$$

So, we can write (2.12) as

$$\sup_{u \in \Gamma} l_n(\mathbf{u}) = l_n(\widehat{\theta}).$$

We assume that

$$\Gamma = \left\{ (s, x, y); \ -s_0 \leq s \leq s_0, \ \frac{1}{x_0} \leq x \leq x_0, \ \frac{1}{y_0} \leq y \leq y_0 \right\},$$

with

$$s_0 > 0, \ x_0 > 1, \ \textit{and} \ y_0 > 1. \tag{2.13}$$

We'll now calculate these estimators, it is more convenient to minimise $\widetilde{l}_n(\theta)$ than to maximize $l_n(\theta)$. We assume $\tau = \frac{\omega^2}{\sigma^2}$, then the function $\widetilde{l}_n$ which depends on τ is given by

$$\begin{aligned}\widetilde{l}_n(\varphi,\tau,\boldsymbol{\sigma^2}) &= l_n(\varphi,\omega^2,\sigma^2)\\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{(X_i-\varphi X_{i-1})^2}{\tau\sigma^2X_{i-1}^2+\sigma^2}+\frac{1}{n}\sum_{i=1}^{n}\log(\sigma^2+\tau\sigma^2X_{i-1}^2)\\ &= \frac{1}{\sigma^2}n^{-1}\sum_{i=1}^{n}\frac{(X_i-\varphi X_{i-1})^2}{\tau X_{i-1}^2+1}+n^{-1}\sum_{i=1}^{n}\log(1+\tau X_{i-1}^2)+\log\sigma^2.\end{aligned}$$

Deriving with respect to the parameters s and y, we obtain

$$\frac{\partial}{\partial\varphi}\widetilde{l}_n(\varphi,\tau,\boldsymbol{\sigma^2}) = -2\frac{1}{\sigma^2}n^{-1}\sum_{i=1}^{n}\frac{X_{i-1}(X_i-\varphi X_{i-1})}{1+\tau X_{i-1}^2},$$

$$\frac{\partial}{\partial\sigma^2}\widetilde{l}_n(\varphi,\tau,\boldsymbol{\sigma^2}) = \frac{1}{\sigma^2}-n^{-1}\sigma^{-4}\sum_{i=1}^{n}\frac{(X_i-\varphi X_{i-1})^2}{1+\tau X_{i-1}^2},$$

We get the estimators s and y based on τ

$$\varphi(\tau) = \left\{\sum_{i=1}^{n}\frac{X_iX_{i-1}}{1+\tau X_{i-1}^2}\right\}\left\{\sum_{i=1}^{n}\frac{X_{i-1}^2}{1+\tau X_{i-1}^2}\right\}^{-1}$$

$$\sigma^2(\tau) = n^{-1}\left\{\sum_{i=1}^{n}\frac{(X_i-\varphi(\tau)X_{i-1})^2}{1+\tau X_{i-1}^2}\right\}.$$

We can obtain the maximum likelihood estimator of s, x and y by calculating $\widehat{\tau}$. We can consider the minimum of the following function:

$$l_n(\tau) = \inf_{\sigma^2}\widetilde{l}_n(\varphi,\tau,\boldsymbol{\sigma^2})-1.$$

By calculating $\inf_{\sigma^2}\widetilde{l}_n(\varphi,\tau,\boldsymbol{\sigma^2})$, we get

$$\inf_{y}\widetilde{l}_n(\varphi,\tau,\sigma^2) = \log\left(n^{-1}\sum_{i=1}^{n}\frac{(X_i-\varphi(\tau)X_{i-1})^2}{1+\tau X_{i-1}^2}\right)+n^{-1}\sum_{i=1}^{n}\log(1+\tau X_{i-1}^2)+1,$$

so,

$$l_n(\tau) = \log(\sigma^2(\tau))+n^{-1}\sum_{i=1}^{n}\log(1+\tau X_{i-1}^2).$$

By minimizing the above function, we obtain $\widehat{\tau}$, so

$$\begin{aligned} s &= \varphi(\widehat{\tau}), \\ y &= \sigma^2(\widehat{\tau}), \\ x &= \widehat{\tau}\sigma^2(\widehat{\tau}). \end{aligned}$$

Even if the likelihood function was derived under the assumption of normality, consistency and normality asymptotic of $\widehat{\theta}_n$ will be derived without this distribution hypothesis.

The first main result of this section is the strong consistency of the quasi-maximum likelihood estimator $\widehat{\theta}_n$.

Theorem 2.2. *If* (2.2), (2.4), (2.5), (2.10), (2.11) *and* (2.13) *hold, then*

$$P\left\{(\varphi + b_0)e_0 = 0\right\} < 1, \tag{2.14}$$

and,

$$\theta \in \Gamma, \tag{2.15}$$

so,

$$\widehat{\theta}_n \to \theta \ a.s \ (n \to \infty). \tag{2.16}$$

Remark 2.3.1. If e_0 and b_0 are independent, then (2.14) is satisfied, or equivalently,

$$P\left\{(\varphi + b_0)e_0 \neq 0\right\} > 0,$$

We can easily verify from the facts that

$$P\{b_0 > 0\} > 0,\ P\{b_0 < 0\} > 0,\ P\{e_0 > 0\} > 0,\ P\{e_0 < 0\} > 0,$$

which implies, for example, in the case of $\varphi \geq 0$, that

$$0 < P\{\varphi + b_0 > 0\}P\{e_0 > 0\} \leq P\{(\varphi + b_0)e_0 > 0\}.$$

A similar argument applies to the case $\varphi < 0$.

Remark 2.3.2. We've imposed (2.5) only on parameters that is the necessary and sufficient condition for the existence of a unique strictly stationary version of the RCA(1) process.

Nicholls and Quinn (1982) proved that $\varphi^2 + \omega^2 < 1$ is equivalent to the existence of a weakly stationary solution, if $\{b_k; k \in \mathbb{Z}\}$and $\{e_k; k \in \mathbb{Z}\}$ are independent, and (2.11) is satisfied.

If (2.2) holds, then this solution is also strictly stationary.

Remark 2.3.3. We assume $\omega^2 > 0$ and $\sigma^2 > 0$. One of these assumptions can be abandoned. For example Assume that $\omega^2 \leq 0$ and by choosing Γ such that

$$\Gamma = \left\{(s,x,y);\ -s_0 \leq s \leq s_0,\ 0 \leq x \leq x_0,\ \frac{1}{y_0} \leq y \leq y_0\right\},$$

Theorem 2.2 remains true if:

i) $\omega^2 = 0$, so $\{X_k; k \in \mathbb{Z}\}$ is an autoregressive process of order 1.

ii) We also assume that $E(X_0^2) < \infty$.

iii) If $\sigma^2 < \infty$, then we can choose Γ such that

$$\Gamma = \left\{(s,x,y);\ -s_0 \leq s \leq s_0,\ \frac{1}{x_0} \leq x \leq x_0,\ 0 \leq y \leq y_0\right\}.$$

We now consider the asymptotic normality of $n^{\frac{1}{2}}(\widehat{\theta}_n - \theta)$. Let

$$Y_{i-1}(\mathbf{u},k,\gamma) = \frac{X_{i-1}^k}{(xX_{i-1}^2 + y)^\gamma}, \quad k = 0,1,...,2\gamma, \quad \gamma \in \mathbb{N}. \tag{2.17}$$

The proof will depend on the properties of

$$g_i'(\mathbf{u}) = (\frac{\partial}{\partial s}g_i(\mathbf{u}), \frac{\partial}{\partial x}g_i(\mathbf{u}), \frac{\partial}{\partial y}g_i(\mathbf{u})),$$

where

$$\begin{aligned}
g_i(\mathbf{u}) &= \frac{(X_i - sX_{i-1})^2}{xX_{i-1}^2 + y} + \log(xX_{i-1}^2 + y) \\
&= \frac{[(\varphi + b_i)X_{i-1} + e_i - sX_{i-1}]^2}{xX_{i-1}^2 + y} + \log(xX_{i-1}^2 + y) \\
&= \frac{(\varphi + bi - s)^2 X_{i-1}^2 + e_i^2 + 2e_i(\varphi + bi - s)X_{i-1}}{xX_{i-1}^2 + y} + \log(xX_{i-1}^2 + y) \\
&= (\varphi + bi - s)Y_{i-1}(\mathbf{u},2,1) + 2e_i(\varphi + bi - s)Y_{i-1}(\mathbf{u},1,1) + \frac{e_i^2}{xX_{i-1}^2 + y} + \log(xX_{i-1}^2 + y),
\end{aligned}$$

then the partial derivatives are given by:

$$\frac{\partial}{\partial s}g_i(\mathbf{u}) = -2(\varphi + bi - s)Y_{i-1}(\mathbf{u},2,1) - 2e_iY_{i-1}(\mathbf{u},1,1), \tag{2.18}$$

$$\begin{aligned}\frac{\partial}{\partial x}g_i(\mathbf{u}) = \;& -(\varphi + bi - s)^2Y_{i-1}(\mathbf{u},4,2) - 2e_i(\varphi + bi - s)Y_{i-1}(\mathbf{u},3,2) \\ & -e_i^2Y_{i-1}(\mathbf{u},2,2) + Y_{i-1}(\mathbf{u},2,1),\end{aligned} \tag{2.19}$$

$$\begin{aligned}\frac{\partial}{\partial y}g_i(\mathbf{u}) = \;& -(\varphi + bi - s)^2Y_{i-1}(\mathbf{u},2,2) - 2e_i(\varphi + bi - s)Y_{i-1}(\mathbf{u},1,2) \\ & -e_i^2Y_{i-1}(\mathbf{u},0,2) + Y_{i-1}(\mathbf{u},0,1).\end{aligned} \tag{2.20}$$

Let

$$A = E(g_1'(\theta)^T g_1'(\theta)), \tag{2.21}$$

The next result shows that the quasi-maximum likelihood estimator is asymptotically normal. Let $N(\theta, A)$ be a normal random vector with mean θ and covariance matrix A, and int(Γ) inside of Γ. Let

$$H = H(\theta) = \begin{bmatrix} 2\alpha(2,1) & 0 & 0 \\ 0 & \alpha(4,2) & \alpha(2,2) \\ 0 & \alpha(2,2) & \alpha(0,2) \end{bmatrix}, \tag{2.22}$$

with

$$\alpha(k,\gamma) = E\left\{\frac{X_0^2}{(\omega^2X_0^2 + \sigma^2)^\gamma}\right\} < \infty, \quad k = 0,1,...,2\gamma, \quad \gamma \in \mathbb{N}. \tag{2.23}$$

Theorem 2.3.1. *If* (2.2), (2.4), (2.5), (2.11), (2.13), (2.14), $\theta \in int(\Gamma)$ *hold,*

$$Eb_0^4 < \infty, \;\; Ee_0^4 < \infty, \tag{2.24}$$

and

$$P\{c_1b_0 + e_0 \in \{c_2, c_3\}\} < 1 \;\; \textit{for any real} \;\; c_1, c_2, c_3 \; \textit{with} \; c_1 \neq 0, \tag{2.25}$$

then

$$n^{1/2}(\widehat{\theta} - \theta) \xrightarrow{\mathcal{D}} N(0, H^{-1}AH),$$

where A and B are non-singular matrices defined in (2.21) *and* (2.22) *respectively.*

Remark 2.3.4. If b_0 and e_0 are independent, the condition (2.25) is satisfied.

Otherwise, we assume that

$$P\{c_1 b_0 + e_0 \in \{c_2, c_3\}\} = 1 \quad \text{for any real } \; c_1, c_2, c_3 \text{ with } c_1 \neq 0, \tag{2.26}$$

We note that $c_2 \neq c_3$, for example, $c_2 < c_3$, then $c_2 < 0 < c_3$, since $E(c_1 b_0 + e_0) = 0$, but $E(c_1 b_0 + e_0)^2 = c_1^2 \omega^2 + \sigma^2 > 0$.

Now, from (2.26),

$$\int P\{c_1 b_0 + e_0 \in \{c_2, c_3\} | e_0 = v\} dP_{e_0}(v) = 1,$$

i.e,

$$P\left\{ b_0 \in \left\{ \frac{c_2 - v}{c_1}, \frac{c_3 - v}{c_1} \right\} \right\} = 1 \quad \text{for } P_{e_0} \text{ for some } v.$$

Therefore, b_0 has a two-point distribution, for example

$$P\{b_0 = u_1\} > 0, P\{b_0 = u_2\} > 0, \quad \text{for some } \; u_1 < 0 < u_2.$$

Similarly,

$$P\{e_0 = v_1\} > 0, P\{e_0 = v_2\} > 0, \quad \text{for some } v_1 < 0 < v_2.$$

This implies,

$$P\{c_1 b_0 + e_0 = c_1 u_1 + v_1\} > 0,$$

$$P\{c_1 b_0 + e_0 = c_1 u_1 + v_2\} > 0,$$

$$P\{c_1 b_0 + e_0 = c_1 u_2 + v_2\} > 0,$$

with $c_1 u_1 + v_1 < c_1 u_1 + v_2 < c_1 u_2 + v_2$, if $c_1 > 0$ which contradicts (2.26). A similar argument applies in case of $c_1 < 0$.

2.3.1 Proof of Main Results

To demonstrate theorem 2.2, we start with two technical lemmas.

Lemma 2.4. *If* (2.2), (2.4), (2.5), (2.11), *and* (2.13) *are verified, so*

$$E\left\{\inf_{\mathbf{u}\in\Gamma} g_1(\mathbf{u})\right\} > -\infty,$$

and $E\left\{g_1(\mathbf{u})\right\}$ *is continuous on* Γ.

Proof. From lemma 2.2, we can see that

$$E\left[\sup_{\mathbf{u}\in\Gamma}\left\{|\log(xX_0^2+y)|\right\}\right] \leq E\left|\log\left(\frac{X_0^2}{x_0}+\frac{1}{y_0}\right)\right| + E|\log(xX_0^2+y)| + |\log y_0| < \infty.$$

So, we have

$$E\left[\inf_{\mathbf{u}\in\Gamma}\left\{g_1(\mathbf{u})\right\}\right] \geq E\left[\inf_{\mathbf{u}\in\Gamma}\left\{\frac{(X_1-sX_0)^2}{xX_0^2+y}\right\}\right] - E\left[\sup_{\mathbf{u}\in\Gamma}\left\{\log(xX_0^2+y)\right\}\right] > -\infty,$$

since $(X_1-sX_0)^2(xX_0^2+y)^{-1}\geq 0$ for all $\mathbf{u}\in\Gamma$.
Using (2.17), we can rewrite $g_i(\mathbf{u})$ as

$$\begin{aligned} g_i(\mathbf{u}) = \; & (\varphi-s+b_i)^2Y_{i-1}(\mathbf{u},2,2)+e_i^2Y_{i-1}(\mathbf{u},0,1)+2(\varphi-s+b_i)e_iY_{i-1}(\mathbf{u},1,1) \\ & +\log(xX_{i-1}^2+y). \end{aligned}$$

It's clear that for all $\mathbf{u}\in\Gamma$, all $k=0,1,2$,

$$E\left\{\frac{|X_{i-1}|^k}{xX_{i-1}^2+y}\right\} < \infty. \tag{2.27}$$

Therefore, from (2.11), and (2.23),we have

$$E\left\{g_1(\mathbf{u})\right\} = \left((\varphi-s)^2+\omega^2\right)\alpha(2,1)+\sigma^2\alpha(0,1)-E\left\{\log Y_0(\mathbf{u},0,1)\right\}. \tag{2.28}$$

So $E(g_1(\mathbf{u}))$ is continous on $\mathbf{u}\in\Gamma$.

Lemma 2.5. *If* (2.2), (2.4), (2.5), (2.10), (2.11), (2.13), *and* (2.14) *are verified, so*

$$E[g_1(\mathbf{u})] > E[g_1(\theta)] \;\; \textit{for all}\; \mathbf{u}\neq\theta, \mathbf{u}\in\Gamma.$$

Proof. From (2.28), we have

$$E[g_1(\mathbf{u})] = (\varphi - s)^2 E\left[\frac{X_0^2}{xX_0^2 + y}\right] + E\left[h\left(\frac{\omega^2 X_0^2 + \sigma^2}{xX_0^2 + y}\right)\right] + E\left[\log(\omega^2 X_0^2 + \sigma^2)\right],$$

where $h(x) = x - \log x$, $x > 0$, which is an increasing function, so, $h(x) > h(1) = 1$ for every real $x \neq 1$. Therefor $E(g_1(\mathbf{u})) \geq E(g_1(\theta))$, and $E(g_1(\mathbf{u})) = E(g_1(\theta))$ if and only if

$$s = \varphi \quad \text{and} \quad P\left\{\frac{\omega^2 X_0^2 + \sigma^2}{xX_0^2 + y} = 1\right\} = 1. \tag{2.29}$$

If $\omega^2 X_0^2 + \sigma^2 = xX_0^2 + y$ a.s, then $(\omega^2 - x)X_0^2 = y - \sigma^2$ a.s.

Now, if $\omega^2 \neq x$ or $\sigma^2 \neq y$, then, $P\{X_0^2 = 1\} = 1$, où $c \in \mathbb{R}$. By replacing in (2.1), we get

$$\begin{aligned} X_1^2 &= [(\varphi + b_1)X_0 + e_1]^2 \\ &= (\varphi + b_1)^2 X_0^2 + e_1^2 + 2e_1 X_0(\varphi + b_1) \\ &= (\varphi + b_1)^2 c + e_1^2 + 2e_1 X_0(\varphi + b_1). \end{aligned} \tag{2.30}$$

Using the fact that the process $\{X_k; k \in \mathbb{Z}\}$ is stationary, we also have $P\{X_1^2 = 1\} = 1$. We replace X_1^2 in (2.30), and we get

$$c = (\varphi + b_1)^2 c + e_1^2 + 2e_1 X_0(\varphi + b_1). \tag{2.31}$$

We note that, according to (2.1), (2.2) and (2.11), we have $E(X_1) = E(\varphi + b1)E(X_0) + E(e_1)$, with $|\varphi| < 1$, and so, by stationarity of $\{X_k; k \in \mathbb{Z}\}$, $E(X_k) = 0$ for all $k \in \mathbb{Z}$. By applying the conditional expectation for (2.31) with respect to (b_1, e_1), we have
$c = (\varphi + b_1)^2 c + e_1^2$ almost surely. Plugging the latter relation back into (2.31), and noting that $|X_0| = \sqrt{c} \neq 0$, we should have $P\{(\varphi + b_1)e_1 = 0\} = 1$, which contradicts the assumption (2.14). So, from (2.29),we deduce that $s = \varphi, x = \omega^2$ and $y = \sigma^2$.

Proof. of theorem 2.2

This proof follows the general method developed by Pfanzagl (1969). First, from (2.15), we notice that,

$$\inf_{\mathbf{u}\in\Gamma} l_n(\mathbf{u}) \leq l_n(\theta),$$

and therefore

$$\limsup_{n\to\infty} \inf_{\mathbf{u}\in\Gamma} l_n(\mathbf{u}) \leq \limsup_{n\to\infty} l_n(\theta, \quad a.s,$$

The sequence $\{g_i(\theta); i \in \mathbb{Z}\}$, is stationary and ergodic, with $E|g1(\theta)| < \infty$. So, by the ergodic theorem, we have

$$\lim_{n\to\infty} l_n(\theta) = E(g_1(\theta)), \quad a.s,$$

which results in

$$\limsup_{n\to\infty} \inf_{\mathbf{u}\in\Gamma} l_n(\mathbf{u}) \leq E(g_1(\theta)), \quad a.s. \tag{2.32}$$

For each positive integer n, $l_n(\mathbf{u})$is continuous on the compact Γ set, and therefore, $\widehat{\theta} \in \Gamma$. From (2.32), we have

$$\limsup_{n\to\infty} l_n(\mathbf{u}) \leq E(g_1(\theta)), \quad a.s. \tag{2.33}$$

Let $C \subset \Gamma$ be a compact such that the distance between C and θ is positive.

We can build the set C easily, for example, Let U be the open center ball θ, with a small enough radius. It's clear that the $C = \Gamma\ U$ is satisfying the assumptions required on C.

Since $g_1(\mathbf{u})$ is continuous on Γ (and so on C), for any $r < E[g_1(\mathbf{u})], \mathbf{u} \in C$, there is an open ball $U(\mathbf{u})$ around $\mathbf{u}$ such that

$$r < E[\inf_{\mathbf{t}\in U(\mathbf{u})} \{g_1(\mathbf{t})\}],$$

(cf. Pfanzagl, 1969, p. 271). The sets $U(\mathbf{u}), \mathbf{u} \in C$, are an open covering of C, and so, they contain finite covering, $U(\mathbf{u_1}), U(\mathbf{u_2}), ..., U(\mathbf{u_k})$ of C. By the ergodic theorem, we have, for any $1 \leq j \leq k$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \inf_{\mathbf{u}\in U(\mathbf{u_j})} g_i(\mathbf{u}) = E\left[\inf_{\mathbf{u}\in U(\mathbf{u_j})} g_1(\mathbf{u})\right] > r \quad a.s. \tag{2.34}$$

noting that

$$\inf_{\mathbf{u}\in C} l_n(\mathbf{u}) \geq \min_{1\leq j\leq k} \inf_{\mathbf{u}\in U(\mathbf{u_j})} l_n(\mathbf{u}) \geq \min_{1\leq j\leq k} \frac{1}{n} \sum_{i=1}^{n} \inf_{\mathbf{u}\in U(\mathbf{u_j})} g_i(\mathbf{u}),$$

from (2.34), we have

$$\liminf_{n\to\infty} \inf_{\mathbf{u}\in C} l_n(\mathbf{u}) \geq \min_{1\leq j\leq k} \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \inf_{\mathbf{u}\in U(\mathbf{u_j})} g_i(\mathbf{u}) > r,$$

except for the event B_r with $P(B_r) = 0$ for any rational r satisfying

$$r < \inf_{\mathbf{u} \in C} E[g_1(\mathbf{u})]. \tag{2.35}$$

Let

$$B = \bigcup \{B_r; r \text{ is rational and satisfies (2.35)}\},$$

then, $P(B_r) = 0$, and, on the complement of B, we have

$$\liminf_{n\to\infty} \inf_{\mathbf{u} \in C} l_n(\mathbf{u}) > r,$$

for any rational r satisfied (2.35). So

$$\liminf_{n\to\infty} \inf_{\mathbf{u} \in C} l_n(\mathbf{u}) \geq \inf_{\mathbf{u} \in C} E[g_1(\mathbf{u})] \quad a.s. \tag{2.36}$$

According to lemma 2.4, $E[g_1(\mathbf{u})]$ is continuous, and from the lemma 2.5, $\theta \neq C$ is the only minimum of $E[g_1(\mathbf{u})]$, so (2.36) gives us

$$\liminf_{n\to\infty} \inf_{\mathbf{u} \in C} l_n(\mathbf{u}) \geq E[g_1(\mathbf{u})] \quad a.s. \tag{2.37}$$

Let U be an open ball with a center θ and a small enough radius, and $U^* = U \cap \Gamma$. If $\widehat{\theta}_n \notin U^*$ is infinitely often there is a random subsequence n_k, where $C^* = \Gamma\ U^*$, we have

$$\liminf_{n\to\infty} \inf_{\mathbf{u} \in C^*} l_n(\mathbf{u}) \leq \liminf_{k\to\infty} l_{n_k}(\widehat{\theta}_{n_k}) \leq \limsup_{n\to\infty} l_n(\widehat{\theta}_n) \quad a.s. \tag{2.38}$$

However, (2.38) contradicts (2.33) and (2.37). So, for any U^*, as above, there is a random variable n_0 such that $\widehat{\theta}_n \in U^*$ for any $n \leq n_0$.

The proof of theorem 2.3 is also divided into several parts lemmas.

Lemma 2.6. *Suppose* (2.2), (2.4), (2.5), (2.11), (2.13), (2.15), (2.24), *and* (2.25) *hold. Then* $E[g_1(\theta)]$ *=0, and the A matrix of* (2.21) *exists and it is nonsingular.*

Proof. From (2.2), (2.11), (2.18)-(2.20) and (2.23), so

$$\frac{\partial}{\partial s} g_i(\mathbf{u}) = \frac{\partial}{\partial x} g_i(\mathbf{u}) = \frac{\partial}{\partial y} g_i(\mathbf{u}) = 0.$$

It is clear that A is finite and non-negative, it is enough to show that it's nonsingular. Suppose that A is singular, then

$$c_1 \frac{\partial}{\partial s} g_1(\mathbf{u}) + c_2 \frac{\partial}{\partial x} g_1(\mathbf{u}) + c_3 \frac{\partial}{\partial y} g_1(\mathbf{u}) = 0 \quad a.s, \tag{2.39}$$

with c_1, c_2 and c_3 constants not all zero, using the formulas of the partial derivatives, we get

$$(c_3 + c_2 X_0^2)\frac{(b_1 X_0 + e_1)^2}{(\omega^2 X_0^2 +_\sigma^2)^2} - 2c_1 X_0 \frac{b_1 X_0 + e_1}{\omega^2 X_0^2 +_\sigma^2} + \frac{c_3 + c_2 X_0^2}{\omega^2 X_0^2 +_\sigma^2} = 0 \quad a.s. \tag{2.40}$$

We notice that $c_2 \neq 0$ or $c_3 \neq 0$. Otherwise, $c_1 \neq 0$ and $X_0(b_1 X_0 + e_1) = 0$ a.s, which implies

$$(b_1 X_0 + e_1) I_{\{X_0 \neq 0\}} = 0 \qquad a.s.$$

By taking the conditional expectation with respect to (b_1, e_1), we get

$$b_1 E(b_1 X_0 + e_1) I_{\{X_0 \neq 0\}} + e_1 P\{X_0 \neq 0\} = e_1 P\{X_0 \neq 0\} = 0.$$

Since $P\{X_0 \neq 0\} > 0$, $e_1 = 0$ a.s, which contradicts the assumptions. So, $c_2 \neq 0$ or $c_3 \neq 0$, which results in $P\{c_3 + c_2 X_0^2 \neq 0\} = 1$,because otherwise$P\{X_0^2 = c\} = 1$for some c, which is a contradiction. From (2.40), we have

$$P\{b_1 X_0 + e_1 \in \{c_2, c_3\}\} = 1,$$

where $c_2 = c_2(X_0)$, $c_3 = c_3(X_0)$, which implies

$$E[P\{b_1 X_0 + e_1 \in \{c_2(X_0), c_3(X_0)\}|X_0\} = 1,$$

i.e, since X_0 and (b_1, e_1) are independent,

$$\int P\{b_1 X_0 + e_1 \in \{c_2(X_0), c_3(X_0)\}\} dP_{X_0} = 1,$$

resulting in

$$P\{b_1 X_0 + e_1 \in \{c_2(X_0), c_3(X_0)\}\} = 1, \text{ for } P_{X_0} \text{ for all } x.$$

In view of $P\{X_0 \neq 0\} > 0$, the latter equality contradicts the assumption (2.25).

The next result shows that $n^{1/2} l_n'(\theta)$ converges in distribution to a non-degenerated random vector of dimension three.

Lemma 2.7. *We suppose that* (2.2), (2.4), (2.5), (2.11), (2.13), (2.15),(2.24) *and* (2.25) *are verified, so*

$$n^{1/2} l'_n(\theta) \xrightarrow{\mathcal{D}} N(\theta, A),$$

where A is defined in (2.21).

Proof. We note that

$$n^{1/2} l'_n(\theta) = n^{1/2} \sum_{i=1}^{n} g'_i(\theta),$$

with $E[g'_i(\theta)] = 0$, $cov(g'_i(\theta)) = A$, and $g'_i(\theta)$, $g'_j(\theta)$ are uncorrelated if $i \neq j$.

In addition, from (2.2) and (2.11), we have

$$E\left(\sum_{i=1}^{k} g'_i(\theta) | \mathcal{G}_{k-1}\right) = \sum_{i=1}^{k-1} g'_i(\theta).$$

Thus, Billingsley's Theorem 23.1 (1968, p. 206) and an application of the Cramér-Wold device (Billingsley, 1968, pp. 48-49) yield the lemma 2.7.

Set $H(\mathbf{u}) = E[g''_1(\mathbf{u})]$. By (2.11), similarly to (2.27), it's obvious that $H(\mathbf{u})$ is finite for any $\mathbf{u} \in \Gamma$.

Let $|\cdot|$ be the maximum norm of a matrix.

Lemma 2.8. *We suppose that* (2.2), (2.4), (2.5), (2.11), (2.13), (2.14), (2.15) *et* (2.24) *are verified, so*

$$\sup_{\mathbf{u} \in \Gamma} |l''_n(\mathbf{u}) - H(\mathbf{u})| \to 0 \quad a.s. \tag{2.41}$$

Also, $H(\mathbf{u})$ is continuous on Γ, and $H = H(\mathbf{u})$ as in (2.22) *is nonsingular.*

Proof. The ergodic theorem yields

$$\limsup_{n \to \infty} \sup_{\mathbf{u} \in \Gamma} |l_n^{(3)}(\mathbf{u})| \leq \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{\mathbf{u} \in \Gamma} |g_i^{(3)}(\mathbf{u})| = E\left[\sup_{\mathbf{u} \in \Gamma} |g_1^{(3)}(\mathbf{u})|\right] < \infty \quad a.s.$$

Therefore, the mean value theorem applied to the coordinates of $l''_n(\mathbf{u})$, yields (2.41).

It's immediate that $H(\mathbf{u})$ is continuous on Γ.

If we show that

$$H^* = \begin{bmatrix} \alpha(4,2) & \alpha(2,2) \\ \alpha(2,2) & \alpha(0,2) \end{bmatrix}$$

is nonsingular, the proof of lemma 2.8 will be complete.

We note that $H^* = E[\eta^T \eta]$, where

$$\eta = \left(\frac{X_0^2}{\omega^2 X_0^2 + \sigma^2}, \frac{1}{\omega^2 X_0^2 + \sigma^2} \right).$$

If $H^*$is singular, then there is c_1 and c_2, with $(c_1, c_2) \neq (0,0)$, such that

$$P\left\{ c_1 \frac{X_0^2}{\omega^2 X_0^2 + \sigma^2} + c_2 \frac{1}{\omega^2 X_0^2 + \sigma^2} = 0 \right\} = 1, \tag{2.42}$$

since (2.42) implies $c_1 \neq 0$, we have

$$P\left\{ X_0^2 = -\frac{c_2}{c_1} \right\} = 1, \tag{2.43}$$

which is the same contradiction as the proof of lemma 2.6.

Proof. Of theorem 2.3

In lemma 2.8, we proved that H^{-1} exists. Since $l_n(\mathbf{u})$ is is continuously differentiable on $\text{int}(\Gamma)$, by theorem 2.2, we have $l'_n(\widehat{\theta}_n) = 0$, and therefore $l'_n(\widehat{\theta}_n) - l'_n(\theta_n) = -l'(\theta)$. Applying the mean value theorem to the coordinates of $l'_n(\mathbf{u})$, we get by the lemma 2.8 and the theorem 2.2

$$(\widehat{\theta}_n - \theta)(H + o(1)) = -l'(\theta) \quad a.s.$$

The result is derived from lemma 2.7.

2.4 Simulation

We will simulate two random coefficients autoregressive models of order one, respectively a stationary model and a model close to the non-stationary zone, for the samples sizes 50, 200 and 500 using the R software. We will estimate their parameters by the maximum likelihood method and calculate their RMSE. Then we will study their asymptotic normality.

For each process we have minimized the following function

$$l_n(\tau) = \log(\sigma^2(\tau)) + n^{-1} \sum_{i=1}^{n} \log(1 + \tau X_{i-1}^2),$$

then we replaced the value of τ in $\varphi(\tau)$ and $\sigma^2(\tau)$.

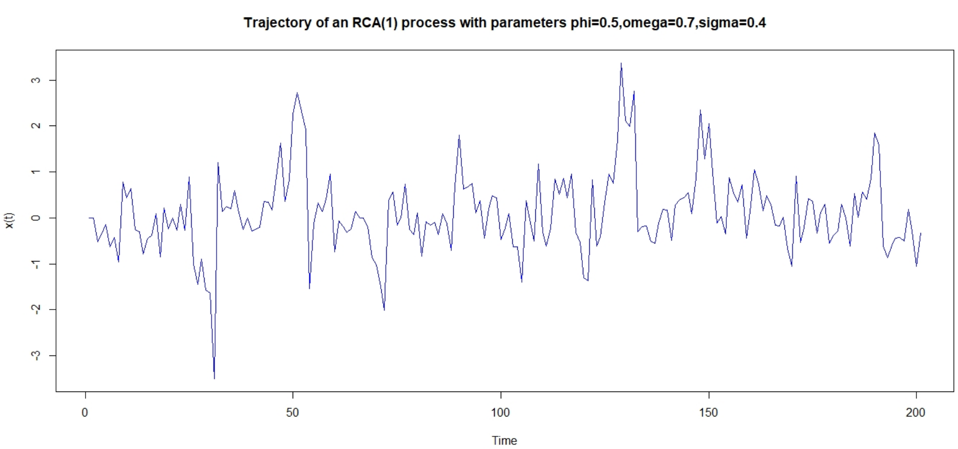
Let $\{X_t, t \in \mathbb{Z}\}$ be an RCA(1) process given by $X_t = (\varphi + b_t)X_{t-1} + \epsilon_t$, and $\theta = (\varphi, w^2, \sigma^2)$.

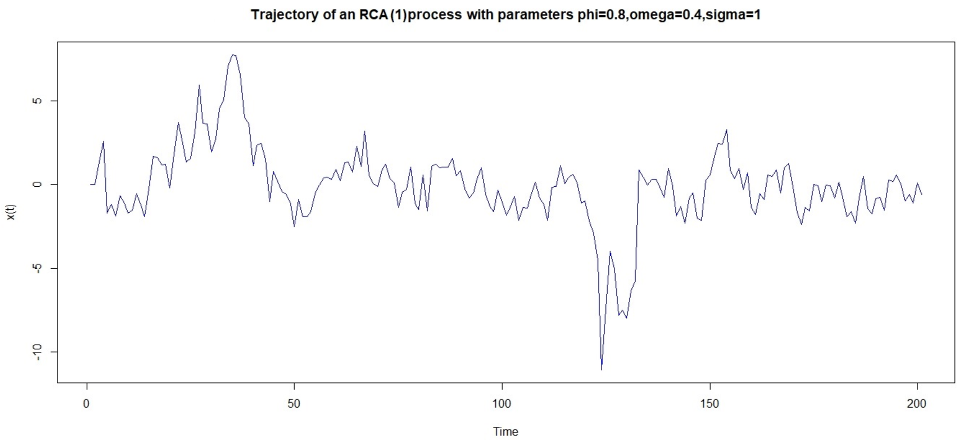
```
RCA1 <- function(n=100,phi=0.7, omega=0.5,sigma=0.6){
  X <- rep(0,n+50)

  e <- rnorm(n+50,mean=0,sd=sigma)
  b <- rnorm(n+50,mean=0,sd=omega)
  if (phi^2+omega^2>1){print("The process is non-stationary.")}
  else
  if (phi^2+omega^2<1){ for (i in 2:n+50){ X[i]<-(phi+b[i])*X[i-1] +
     e[i] }
   X <- X[50:(n+50)]
   X
  }


plot(X,type="l",col="blue",xlab="Time",
ylab="x(t)",main="Trajectory of an RCA(1)  process with parameters
phi=0.8,omega=0.4,sigma=1")
```

We plot the trajectories of 2 processes with $\theta_1 = (0.5, 0.49, 0.16)$ and $\theta_2 = (0.8, 0.16, 1)$ respectively.






```
L <- function(tau){
  X<-RCA1(n=50,phi=0.8,omega=0.4,sigma=1)
  Xd<-X[1:n]     # vector of X_{i-1}
  X<-X[2:(n+1)]#vector of X_{i}
  c1= (X*Xd)/(1+ tau*(Xd^2))
  c2=sum(c1)
  c3= (Xd^2)/(1+ tau*(Xd^2))
  c4= 1/(sum(c3))
  c5=((X -(c2*c4)*Xd)^2)/(1+ tau*(Xd^2))
  c6=log(1+ tau*(Xd^2))
  log((1/n)*sum(c5)) + (1/n)*sum(c6)
}
tau = optimize(L, interval=c(0,5))$minimum
phiT=(sum((X*Xd)/(1+tau*(Xd^2))))*(1/(sum((Xd^2)/(1+tau*(Xd^2)))))
sigmaT=(1/n)*sum(((X-((sum((X*Xd)/(1+tau*(Xd^2))))*(1/(sum((Xd^2)/(1+
```

```
tau*(Xd^2)))))))*Xd)^2)/(1+ tau*(Xd^2)))

s=phiT       #estimator of varphi

y=sigmaT #estimator of sigma^2

x=tau*y #estimator de omega^2
```

The estimators and the root mean square error (RMSE) are given in the following tables

	N	φ	w^2	σ^2
The true values		0.5	0.49	0.16
The estimated values	50	0.5279	0.3066	0.2250
RMSE		0.0279	0.1833	0.0650
The estimated values	200	0.4843	0.3473	0.1971
RMSE		0.0156	0.1426	0.0371
The estimated values	500	0.4865	0.4992	0.1635
RMSE		0.0134	0.0092	0.0035

	N	φ	w^2	σ^2
The true values		0.8	0.16	1
The estimated values	50	0.8287	0.3939	0.5431
RMSE		0.0287	0.2339	0.4568
The estimated values	200	0.8329	0.3096	0.7435
RMSE		0.0329	0.1496	0.2564
The estimated values	500	0.8139	0.1830	0.9636
RMSE		0.01396	0.0230	0.0363

We notice when n increases, the root-mean-square error decreases, i.e. we get more accurate estimators.
We also note that the RMSE for the model where the non-random parameter close to the non-stationarity zone is larger than the other model.

To examine asymptotic normality of these estimators, we calculate their kurtosis and their skewness.
We have obtained the following results.

Parameters estimated by the ML method	Kurtosis	Skewness
φ	3.1139	-0.37698
w^2	2.3486	0.2606
σ^2	2.5748	0.16061

Parameters estimated by the ML method	Kurtosis	Skewness
φ	2.4230	0.22080
w^2	2.3971	0.00369
σ^2	2.6401	0.56195

We note that the model's parameter estimators have a distribution that approaches a normal distribution since their kurtosis is about 3 and the absolute value of their skewness is less than 0.5.

We also note that the estimators of the first model are closer to normality than those of the second model.

Conclusion:

The higher the value of n, the better the quality of the estimator, the RMSE proves it. Same for normality.
Conversely, the closer we get to the non-stationarity, the poorer the quality of the estimator. The RMSE values highlight this. Also, we deviate from the normality.

Chapter 3

Random coefficient autoregressive processes of order p

3.1 Introduction

More recently, Ledolter (1980) has extended Garbade's (1977) procedure to include autoregressive models, while Conlisk (1974), (1976) has derived conditions for the stability of RCA models. Andel (1976) has argued that when modelling time series data in such fields as hydrology, meteorology and biology, the coefficients of the model under consideration arise "as a result of complicated processes and actions which usually have many random features". This has led him to consider scalar RCA models and to derive conditions for their second order stationarity.

Nicholls and Quinn (1982) extended the work of Andel (1976) to the multivariate case, and also obtained the conditions for the existence of a stationary solution. They also estimated the unknown parameters of the model and established a strong consistency and central limit theorem.

This chapter is organized as follows. In section 1 and 2 we discuss some basic properties of RCA(p) time series such as the necessary and sufficient conditions for the existence and uniqueness of a strictly stationary solution to $\{X_t\}$. The quasi-maximum likelihood estimators are obtained in section 3, their high consistency (almost certain convergence) and asymptotic normality are obtained under optimal assumptions.

3.2 Definition

A p-variate $\{X_t, t \in \mathbb{Z}\}$ is said to follow a random coefficient autoregressive model of order n, i.e RCA(n), if it satisfies an equation of the form

$$X_t = \sum_{i=1}^{n} (\psi_i + b_{it}) X_{t-1} + \epsilon_t. \tag{3.1}$$

The following assumptions hold for this model

i) The $(p \times p)$ matrices $\psi_i, i = 1, ..., n$ are constant .

ii) Let $b_t = (b_{n,t}, ..., b_{1,t})'$, then $\{b_t, t = 0, \pm 1, \pm 2, ...\}$ is an independent sequence of $(p \times np)$ matrices with mean zero and covariance matrix C.

iii) $\{\epsilon_t, t = 0, \pm 1, \pm 2, ...\}$ is an independent sequence of p-variate random variables with mean zero and covariance matrix G, $\{b_t\}$ and $\{\epsilon_t\}$ are independent.

3.3 Stationarity

In the case of the scalar RCA model. that is the model with $p = 1$, Andel (1976) has obtained conditions for the existence of a singly infinite process $\{X_t, t = 1 - n, ..., 0, 1, ...\}$ satisfaing (3.1). Nicholls and Quinn (1982) extended the work of Andel (1970) to the case of a multivariate RCA model and also obtained the conditions of existence of a stationary solution of (3.1)

In deriving necessary and sufficient conditions for stationarity, we assume conditions (i)-(iii) along with the condition

iv) there is no non-zero $(p \times 1)$ constant vector such that $y' X_t$ is purely linearly deterministic, i.e, it's determined exactly as a linear function of $\{X_t\}$.

Condition (iv) is imposed in order that X(t) have no linearly deterministic component. In fact given (i)-(iii). conditions will be found for (iv) to hold.

3.3.1 Singly-Infinite Stationarity

Definition 3.3.1. A process $\{X_t, t \in \mathbb{Z}\}$ is second order stationary if and only if $E[X_i] = \mu_i, i \leq 0$ is constant and $E[(X_i - \mu_i)(X_i - \mu_i)'] = V_{ij}, i, j \leq 0$ depends only on the value of (j-i).

In order to derive the stationarity conditions for the process $\{X_t\}$ satisfying (3.1), it is more convenient to give it a Markovian representation, i.e, to represent the p-order RCA model as a RCA(1) model.

Let $\{Z_t\}$ be a $(np \times 1)$ vector defined by $Z_t = [(X'_{t+1-n}, ..., X'_t)']$.
We can rewrite the equation (3.1) in terms of Z_t by:

$$\begin{aligned} Z_t = & \; M\ Z_{t-1} \ + \ L \otimes \ (b'_t\ Z_{t-1} \ + \ \epsilon_t), \\ = & \; (M + L \otimes b'_t) Z_{t-1} + L \otimes \epsilon_t \\ = & \; (M + K_t) Z_{t-1} + \eta_t, \end{aligned} \tag{3.2}$$

where L is the $n \times 1$ vector with only nonzero entry the n^{th}, which is 1, $K_t = L \otimes b'_t$, $\eta_t = L \otimes \epsilon_t$, and M is the $(np \times np)$ matrix defined by

- $M = \begin{bmatrix} 0 & I \\ \psi_n \ldots & \psi_1 \end{bmatrix}$,

  · M(1,1) is the $(n-1)p \times p$ null matrix

  · M(1,2) is the $(n-1)p \times (n-1)p$ identity

- $K_t = \begin{bmatrix} & 0 & \\ & & \\ b_{p,t} & \cdots & b_{1,t} \end{bmatrix}$,

- $\eta_t = L \ \otimes \ \epsilon_t \ = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \otimes \ \epsilon_t \ = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \epsilon_t \end{bmatrix}$,

- $$\begin{aligned} E[\eta_t\eta_t'] &= E[(L \otimes \epsilon_t)(L \otimes \epsilon_t)'] \\ &= E[(L \otimes \epsilon_t)(L' \otimes \epsilon_t')] \\ &= (LL') \otimes E[\epsilon_t\epsilon_t'] \\ &= J \otimes G, \end{aligned}$$
  where $J = LL'$.

$$Z_t = \begin{bmatrix} X_{t+1-n} \\ X_{t+2-n} \\ \vdots \\ X_t \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ \psi_n + b_{n,t} & \psi_{n-1} + b_{(n-1),t} & & & \cdots & \psi_1 + b_{1,t} \end{bmatrix} \times \begin{bmatrix} X_{t-n} \\ X_{t+1-n} \\ \vdots \\ X_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \epsilon_t \end{bmatrix}.$$

Let $\widetilde{C} = E[K_t \otimes K_t]$.

Theorem. ***(Nicholls and Quinn (1982))*** *The process* $\{X_t,\ t = 1-n, ..., 0, 1, ...\}$ *generated by* (3.1) *from* $t = 1$*, is second order stationary if and only if:*

- $\mu_1 = \mu_0$,
- $V_{1,1} = V_{0,0}$,

where $\mu_i = E[Z_i]$*, and* $V_{i,j} = E[Z_i Z_j']$.

Proof. We assumed that Y_0 is independent of $\{\epsilon_t\}$ and $\{b_t\}$. We also assume that:

$$\begin{cases} \mu_t = \mu, & t = 0, 1, ..., h, \\ V_{t,t-s} = V_{t-u,t-s-u} = \Sigma_s, & t = s+1, ..., h, u = 1, ..., t-s, s = 0, ..., h. \end{cases}$$

Where $h = 1$, this condition is reduced to

$$\begin{cases} \mu_1 = \mu_0 = \mu, \\ V_{0,0} = V_{1,1} = \Sigma_0. \end{cases}$$

According to (3.2), under the above assumptions, we have:

$$\begin{aligned} \mu_{h+1} &= E[Z_{h-1}] \\ &= E[(M + K_{h+1})\ Z_h + \eta_{h+1}] \\ &= E[M\ Z_h\ +\ K_{h+1}\ Z_h + \eta_{h+1}] \\ &= M\ E[Z_h]\ +\ E[K_{h+1}\ Z_h] + E[\eta_{h+1}] \\ &= M\ E[Z_h]\ +\ E[K_{h+1}]\ E[Z_h] + E[\eta_{h+1}]. \end{aligned}$$

Since K_{h+1} and η_{h+1} are independent of $K_h, ..., K_1$, $\eta_h, ..., \eta_1$ and Z_0, so:

$$E[K_{h+1}] = E[L \otimes b_{h+1}] = L \otimes E[b_{h+1}] = 0, \forall h,$$

$$E[\eta_{h+1}] = E[L \otimes \epsilon_{h+1}] = L \otimes E[\epsilon_{h+1}] = 0, \forall h.$$

We'll get:

$$\begin{aligned} \mu_{h+1} &= ME[Z_h] \\ &= M\mu_h \\ &= M\mu_{h-1}. \end{aligned}$$

Under the recurrence hypothesis:

$$\begin{aligned} \mu_{h+1} &= ME[Z_{h-1}] \\ &= \mu_h. \end{aligned}$$

Now, we're looking to find V_{h+s}, V_{h+s-1} for:

► Case where $1 \leq s \leq h$: By hypothesis we have:

$$V_{h,h} = V_{h-1,h-1} = ... = V_{0,0},$$

$$V_{t,t-s} = V_{t-u,t-s-u} = \Sigma_s, t = s+1, ..., h, u = 1.$$

We also have K_{h+1} and η_{h+1} are independent of $K_h, ..., K_1$, $\eta_h, ..., \eta_1$ and Z_0, so:

$$\begin{aligned}
V_{h+1,h+1-s} &= E[Z_{h+1}\ Z'_{h+1-s}] \\
&= E\left\{[(M\ +\ K_{h+1})\ Z_h\ +\ \eta_{h+1}]\ Z'_{h+1-s}\right\} \\
&= E\left\{[(M\ +\ K_{h+1})\ Z_h\ Z'_{h+1-s}]\ +\ [\eta_{h+1}\ Z'_{h+1-s}]\right\} \\
&= M\ E[Z_h\ Z'_{h+1-s}]\ +\ E[K_{h+1}\ Z_h\ Z'_{h+1-s}]\ +\ E[\eta_{h+1}\ Z'_{h+1-s}] \\
&= M\ E[Z_h\ Z'_{h+1-s}]\ +\ \underbrace{E[K_{h+1}]}_{=0}\ E[Z_h\ Z'_{h+1-s}]\ +\ \underbrace{E[\eta_{h+1}]}_{=0}\ E[Z'_{h+1-s}] \\
&= M\ V_{h,h+1-s} \\
&= M\ V_{h-1,h+1-1-s} \\
&= M\ V_{h-1,h-s} \\
&= M\ E[Z_{h+1}\ Z'_{h-s}] \\
&= V_{h,h-s}.
\end{aligned}$$

► Case where $S = 0$

$$\begin{aligned}
V_{h+1,h+1} &= E[Z_{h+1}\ Z'_{h+1}] \\
&= E\left\{[(M\ +\ K_{h+1})\ Z_h + \eta_{h+1}]\ [(M\ +\ K_{h+1})\ Z_h + \eta_{h+1}]'\right\} \\
&= E[(M\ +\ K_{h+1})\ Z_h\ Z'_h(M\ +\ K_{h+1})'] + E[\eta_{h+1} Z'_h(M\ +\ K_{h+1})'] \\
&+\ E[(M\ +\ K_{h+1})\ Z_h \eta'_{h+1}]\ +\ E[\eta_{h+1}\eta'_{h+1}] \\
&= E[(M\ +\ K_{h+1})\ Z_h\ Z'_h(M\ +\ K_{h+1})']\ +\ E[\eta_{h+1}\eta'_{h+1}].
\end{aligned}$$

$$
\begin{aligned}
vec\, V_{h+1,h+1} &= vec\, E[(M + K_{h+1})\, Z_h\, Z_h'(M + K_{h+1})'] + vec\, E[\eta_{h+1}\eta_{h+1}'] \\
&= E[vec\,((M + K_{h+1})\, Z_h\, Z_h'(M + K_{h+1})')] + vec\, E[\eta_{h+1}\eta_{h+1}'] \\
&= E[(M + K_{h+1}) \otimes (M + K_{h+1}) vec\,(Z_h\, Z_h')] + vec\, E[\eta_{h+1}\eta_{h+1}'] \\
&= E[(M \otimes M) + (K_{h+1} \otimes K_{h+1})]\, vec\,(Z_h\, Z_h') + vec\, E[\eta_{h+1}\eta_{h+1}'] \\
&= [(M \otimes M) + E[(K_{h+1} \otimes K_{h+1})]\, vec\,(Z_h\, Z_h') + vec\,(J \otimes G) \\
&= (M \otimes M + \widetilde{C})\, vec\, V_{h,h} + vec\,(J \otimes G) \\
&= (M \otimes M + \widetilde{C})\, vec\, V_{h-1,h-1} + vec\,(J \otimes G) \\
&= vec\, V_{h,h}.
\end{aligned}
$$

Corollary 3.3.1. ***(Nicholls and Quinn (1982))***

The process $\{X_t, t = 1-n, ..., 0, 1, ...\}$ *generated by* (3.1) *is said to be stationary if and only if*
$E[Z_0] = \mu$ *satisfies* $M\mu = \mu$,
and $E[Z_0\, Z_0'] = V$ *satisfies*

$$
vec\, V = (M \otimes M + \widetilde{C})\, vec\, V_{h-1,h-1} + vec\,(J \otimes G). \tag{3.3}
$$

Proof. Let $\{X_t, t = 1-n, ..., 0, 1, ...\}$ be the process generated by (3.1).

► **The Necessary Condition**

If the process is second order stationary, then:

- $E[Z_t] = E[Z_{t-1}] = ... = E[Z_0]$,
  $\mu_h = E[Z_h] = ME[Z_{h-1}] = M\mu_{h-1}$,
  so, $\mu_h = M\, \mu_{h-1}$, where $\mu_h = \mu, \forall h$, so $\mu = M\, \mu$.

- Let $V = E[Z_0\, Z_0']$ and $V_1 = E[Z_1\, Z_1']$, we assume that $V = V_1$.
  We're trying to find $vec\, V_1$.

Following the same procedure as in the proof of the previous theorem, and we get

$$\begin{aligned} vec\, V_1 &= vec\, [E[Z_1\, Z_1']] \\ &= vec\, [E[((M\, +\, K_1)\, Z_0\, +\, \eta_1)(((M\, +\, K_1)\, Z_0\, +\, \eta_1))']] \\ &= vec\, E[M\, Z_0\, Z_0'\, M']\, +\, vec\, E[M\, Z_0\, Z_0'\, K_1']\, +\, vec\, E[K_1\, Z_0\, Z_0'\, M'] \\ &+\, vec\, E[K_1\, Z_0\, Z_0'\, K_1']\, +\, vec\, (J \otimes G). \end{aligned}$$

Since $vec\, E[M\, Z_0\, Z_0'\, K_1'] = 0$ and $vec\, E[K_1\, Z_0\, Z_0'\, M'] = 0$, so:

$$\begin{aligned} vec\, V_1 &= vec\, E[M\, Z_0\, Z_0'\, M']\, +\, vec\, E[K_1\, Z_0\, Z_0'\, K_1']\, +\, vec\, (J \otimes G) \\ &= (M \otimes M)\, vec\, E[Z_0 Z_0']\, +\, E[K_1 \otimes K_1]\, vec\, E[Z_0\, Z_0']\, +\, vec\, (J \otimes G) \\ &= (M \otimes M) vec\, V\, +\, \underbrace{E[K_1 \otimes K_1]}_{\widetilde{C}}\, vec\, V\, +\, vec\, (J \otimes G) \\ &= (M \otimes M + \widetilde{C})\, vec\, V + vec\, (J \otimes G) \\ &= vec\, V. \end{aligned}$$

► **The sufficient condition**

- is $M\, \mu = \mu$, where $\mu = E[Z_0]$.
  $\mu_1 = E[Z_1]\, =\, M\, E[Z_0]\, =\, M\, \mu_0\, =\, \mu_0$, so the mean is constant.

- If $vec\, V\, =\, (M \otimes M + \widetilde{C})\, vec\, V + vec\, (J \otimes G)$,
  so, $V_1 = V\, \Leftrightarrow\, E[Z_1\, Z_1']\, =\, E[Z_0\, Z_0']$.
  We can show by recurrence that $E[Z_t\, Z_t'] = E[Z_0\, Z_0']\ \forall t$.
  So the process $\{X_t, t = 1 - n, ..., 0, 1, ...\}$ is stationary, a solution to the equation (3.3) is given by:

$$\begin{aligned} vec\, V &= (M \otimes M + \widetilde{C})\, vec\, V + vec\, (J \otimes G) \\ &\Rightarrow\, vec\, V\, -\, (M \otimes M + \widetilde{C})\, vec\, V\, =\, vec\, (J \otimes G) \\ &\Rightarrow\, (I\, -\, (M \otimes M + \widetilde{C}))\, vec\, V\, =\, vec\, (J \otimes G), \end{aligned}$$

where

$$vec\,(J \otimes G) \;=\; (I \;-\; (M \otimes M + \widetilde{C}))\; vec\, V,$$

with $(I \;-\; (M \otimes M + \widetilde{C}))$ be invertible.

3.3.2 Existence of a Stationary Solution

We are now trying to give the conditions for a stationary solution for (3.1).

Let F_t be the tribe generated by $\{(\epsilon_s, b_s), s \leq t\}$.

In order to try to find the $\mathcal{F}$-measurable solution to the equation (3.1), we define the product of matrices $\prod_{l=1}^{j} \; A_l$ such that:

$$\prod_{l=1}^{j} A_l = \begin{cases} A_i A_{i+1} ... A_j, & i \leq j \\ I & j = i-1, \end{cases}$$

then, if $S_{t,r} = \prod\limits_{l=0}^{r} (M + K_{t-l})$; $R_{t,r} = S_{t,r} Z_{t-r-1}$, we have, iterating (3.2),

$$\begin{aligned} Z_t &= (M \;+\; K_t)\,[(M \;+\; K_{t-1})\; Z_{t-2} \;+\; \eta_{t-1}] \;+\; \eta_t \\ &= (M \;+\; K_t)\,(M \;+\; K_{t-1})\; Z_{t-2} \;+\; (M \;+\; K_t)\,\eta_{t-1} \;+\; \eta_t \\ &= \sum_{j=0}^{r} \; S_{t,j-1}\; \eta_{t-j} \;+\; R_{t,r}, \end{aligned}$$

Besides, if $\Sigma_{t,r} = Z_t - R_{t,r}$ knowing that $E[\eta_{t-1}\eta'_{t-1}] = 0$ if $i \neq j$, then:

$$\begin{aligned} vec\; E(\Sigma_{t,r}\; \Sigma'_{t,r}) \;&=\; vec\; E\left[(\sum_{j=0}^{r} \; S_{t,j-1}\; \eta_{t-j})\; (\sum_{j=0}^{r} \; S_{t,i-1}\; \eta_{t-i})' \right] \\ &=\; \sum_{j=0}^{r} \sum_{i=0}^{r} \; E(S_{t,j-1} \otimes S_{t,i-1}\; vec\; (\eta_{t-j}\; \eta'_{t-i})) \\ &=\; vec\; E\left[\sum_{j=0}^{r} \; S_{t,j-1}\; \eta_{t-j}\; \eta'_{t-j} S'_{t,j-1} \right] \\ &=\; E\left[\sum_{j=0}^{r} \; S_{t,j-1} \otimes S_{t,j-1}\; vec\; (\eta_{t-j}\; \eta'_{t-j}) \right] \\ &=\; E\left[\sum_{j=0}^{r} \; \left\{ \prod_{l=0}^{j-1} [(M \;+\; K_{t-l}) \otimes (M \;+\; K_{t-l})] \right\} \; vec\; (\eta_{t-j}\; \eta'_{t-j}) \right]. \end{aligned}$$

Since $\{b_t, t \in \mathbb{Z}\}$ and $\{\eta_t\}$ are independent, then:

$$
\begin{aligned}
vec\ E(\Sigma_{t,r}\ \Sigma'_{t,r}) &= E\left[\sum_{j=0}^{r}\left\{\prod_{l=0}^{j-1}\left[(M\ +\ K_{t-l})\otimes(M\ +\ K_{t-l})\right]\ vec\ E(\eta_{t-j}\ \eta'_{t-j})\right\}\right] \\
&= \sum_{j=0}^{r}\left\{\prod_{l=0}^{j-1} E\left[(M\otimes M)+(M\otimes K_{t-l})+(K_{t-l}\otimes M)+(K_{t-1}\otimes K_{t-l})\right]\right. \\
\times &\quad \left. vec\ E(\eta_{t-j}\ \eta'_{t-j})\right\}.
\end{aligned}
$$

Since $b_t, t\in\mathbb{Z}$ is centered, and M is non-random, so:

$E[M\otimes K_{t-l}] = E[K_{t-l}\otimes M] = 0.$

$$
\begin{aligned}
vec\ E(\Sigma_{t,r}\ \Sigma'_{t,r}) &= \sum_{j=0}^{r}\left\{\prod_{l=0}^{j-1}\left[(M\otimes M)\ E(K_{t-l}\otimes K_{t-l})\right]\ vec\ E(\eta_{t-j}\ \eta'_{t-j})\right\} \\
&= \sum_{j=0}^{r}\left\{\prod_{l=0}^{j-1}\left[(M\otimes M)\ +\ \widetilde{C}\right]\ vec\ (J\otimes G)\right\} \\
&= \sum_{j=0}^{r}(M\otimes M\ +\ \widetilde{C})^{j}\ vec\ (J\otimes G).
\end{aligned}
$$

Theorem. ***(Nicholls and Quinn (1982))***

In order to X_t be a stationary solution $\mathcal{F}_t$-measurable, it's necessary that $\sum_{j=0}^{r}(M\otimes M\ +\ \widetilde{C})^{j}\ vec\ (J\otimes G)$, converges when $r\to\infty$ and sufficient that this occurs with the matrix H defined such that

$$vec\ H = vec\ G + C\ \sum_{j=0}^{\infty}(M\otimes M+\widetilde{C})^{j}\ vec\ (J\otimes G).$$

If the matrix$(M\otimes M+\widetilde{C})$ has no unitary eigenvalue, then this last condition is both necessary and sufficient, so the unique X_t solution is obtained from:

$$Z_t = \sum_{j=1}^{\infty}\left(\prod_{l=0}^{j-1}(M+K_{t-l})\ \eta_{t-j}\right)+\eta_t. \tag{3.4}$$

Proof. : F. Nicholls Barry G. Quinn, Random Coefficient Autoregressive Models: An Introduction, pages 21-32.

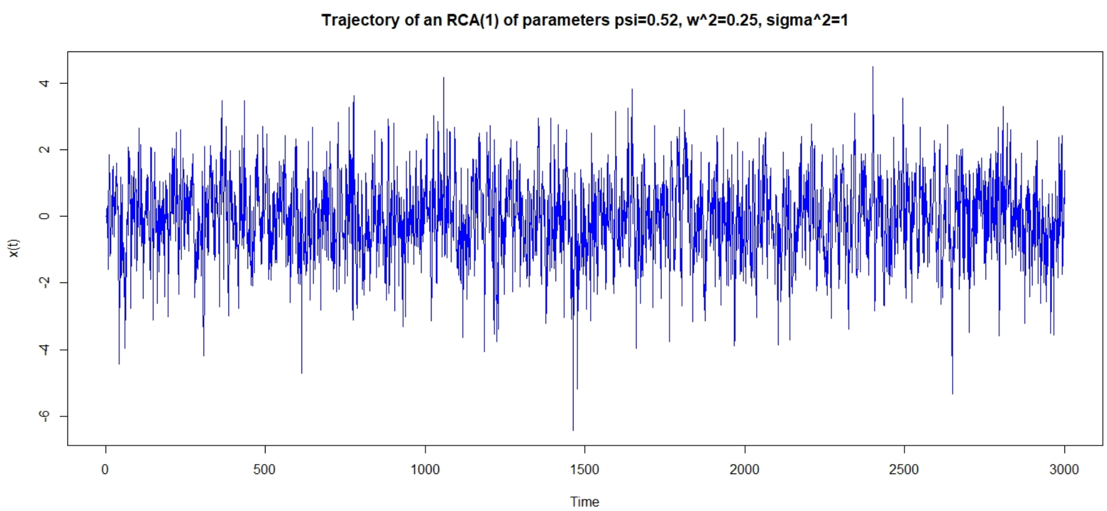
3.3.3 Strong Stationarity

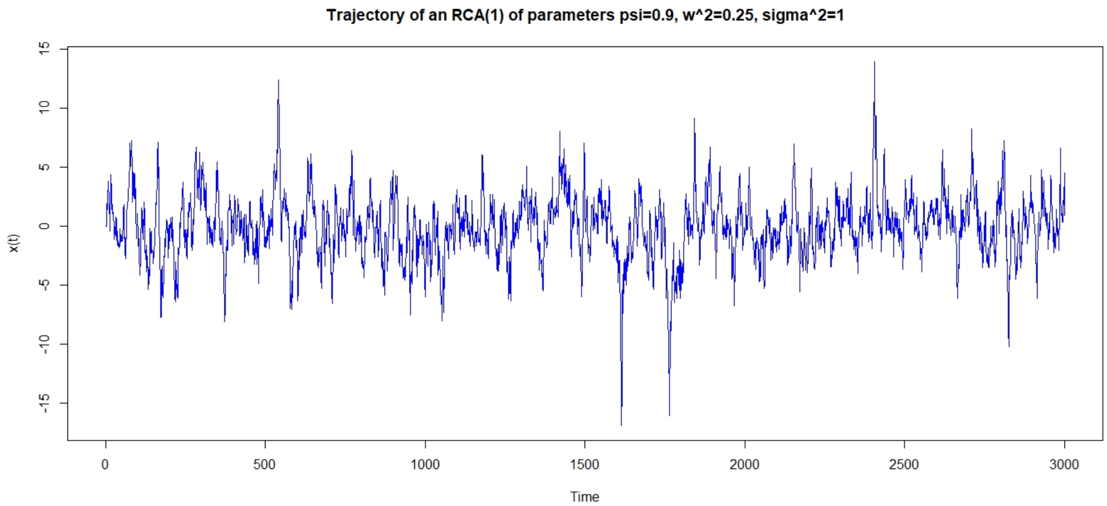
The previous sections have made no assumptions about $\{\epsilon_t\}$ and $\{b_t\}$ except they're independent second-order stationary processes that are mutually Independents.
$\{\epsilon_t\}$ and $\{b_t\}$ are sequences of identically distributed random variables, in this case, they are also strictly stationary and ergodic, it is possible to deduce stronger properties for the X_t $\mathcal{F}_t$-measurable solution of equation (3.1).

Theorem. ***(Nicholls and Quinn (1982))***

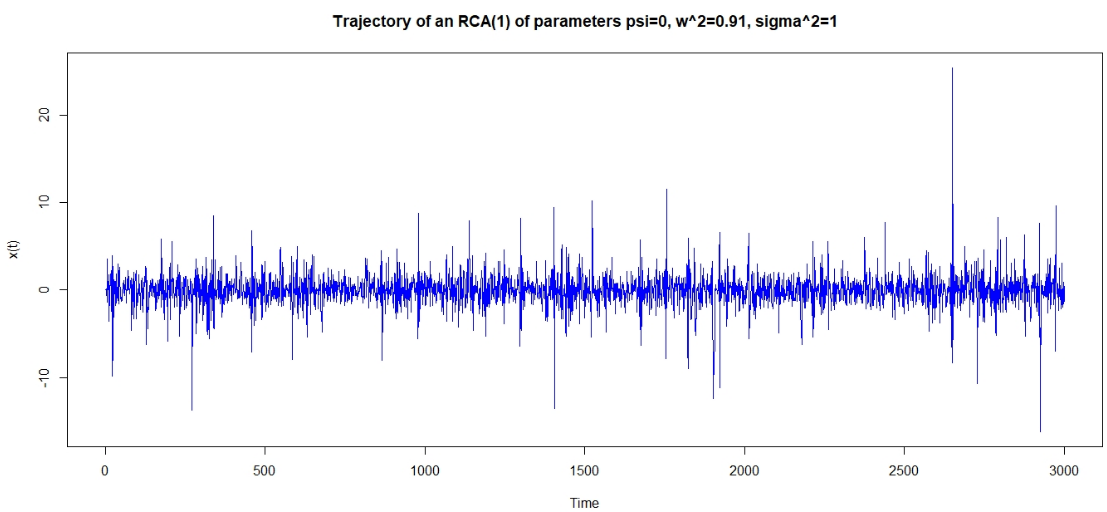
Assume that $\{\epsilon_t\}$ and $\{b_t\}$ are sequences of identically distributed random variables, then, if the second order stationary solution $\mathcal{F}_t$-measurable and unique X_t of (3.1) *exists, then the process$\{X_t\}$ is also strictly stationary and ergodic.*

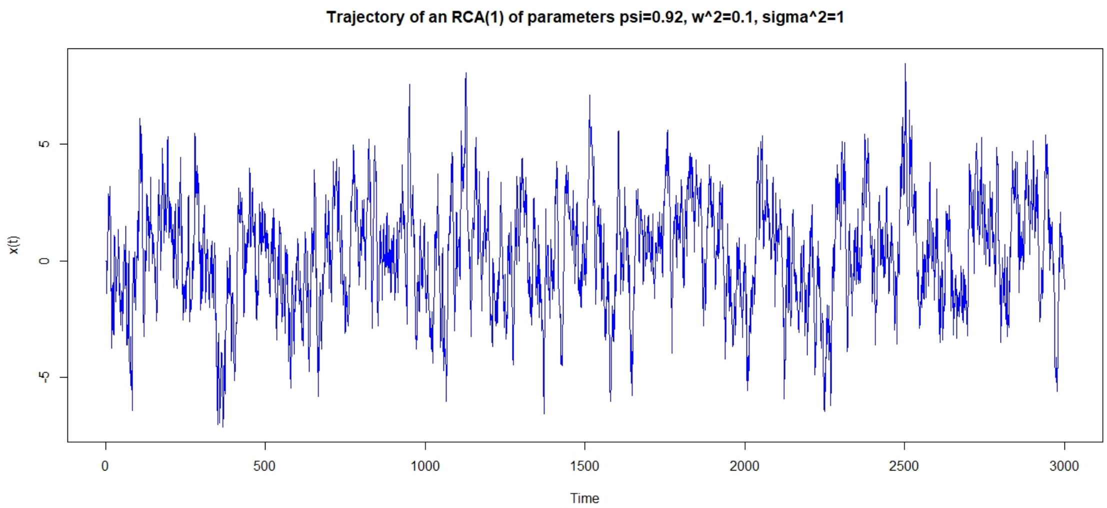
Often a behavior generally associated with non-stationarity, is an indication of the non-linearity of a model. The phenomenon is well illustrated on the following graphs where we simulated a three-mile sample size for different values of ψ, w^2 and σ^2.





These figures represent a stationary RCA(1) process where the non-random parameter is close to the non-stationarity zone. We note that in the second case extreme values are more frequent.





When the variance of the random disturbance increases even for $\psi = 0$, we notice the multiplicity of extreme values. Although the variance of the random disturbance is relatively small, the fact that the non-random parameter ψ is close to the non-stationarity zone makes the extreme values appear.

3.4 Parameter estimation

In the previous section, we have given the conditions for the existence of a stationary solution of the equation (3.1). However, in practice, it is necessary to estimate the unknown parameters of the stationary time series $\{X_t\}$ to provide X_t predictors based on past values of the process.

The procedures for estimating constant coefficient autoregressive, and the asymptotic properties of these estimates are well known.However, the random coefficient autoregressive non-linear in nature and any method of maximum likelihood estimation would be an iterative procedure.

3.4.1 Maximum likelihood approach

Nichols and Quinn (1982) introduced an estimation method based on maximizing the likelihood function constructed as though the processes $\{\epsilon_t\}$ and $\{b_t\}$ were sequences of normally distributed random variables.
this method alleviates the restrictions on the existence of 8^{th}-order moments necessary in

the context of least square estimation.

Let $\{X_t, t \in \mathbb{Z}\}$ a random coefficient autoregressive model of order n, defined by

$$X_t = \sum_{i=1}^{n} (\psi_i + b_{it}) X_{t-1} + \epsilon_t, \tag{3.5}$$

where the following assumptions are made

(i) The $n \times 1$ vector $\psi = (\psi_n, ..., \psi_1)'$ is constant.

(ii) Let $b_t = (b_{nt}, ..., b_{1t})'$, then, $\{b_t, t = 0, \pm 1, \pm 2, ...\}$ is a sequence of i.i.d. random variables with mean zero and covariance matrix w^2.

(iii) $\{\epsilon_t, t = 0, \pm 1, \pm 2, ...\}$ is a sequence of identically and independently distributed (i.i.d.) random variables with zero mean and variance σ^2. Furthermore $\{b_t\}$ is independent of $\{\epsilon_t\}$.

(iv) The variance σ^2 is bounded below by δ_1, while the smallest eigenvalue of w^2 is bounded below by δ_2, where $\delta_1 > 0$ and $\delta_2 > 0$ are both arbitrarily small.

(v)' The parameters $\psi_i, i = 1, ..., n$ and w^2 are such that there is a unique strictly stationary solution to (3.5) which has finite second moments.

i.e. $\{X_t\}$ will be ergodic and strictly stationary, and this also allows for unconstrained optimization to avoid the contentious cases where $\sigma^2 = 0$, or the matrix w^2 has null eigenvalues.

Let $Z_t = (X_{t+1-n}, ..., X_t)'$ be the $n \times 1$ vector, and M the $n \times n$ matrix defined by

$$M = \begin{bmatrix} 0 & I \\ \psi_n \ldots & \psi_1 \end{bmatrix},$$

- M(1,1) is the $(n-1) \times 1$ null matrix,

- M(1,2) is the $(n-1) \times (n-1)$ identity.

So, (3.5) may be rewritten as

$$Z_t = MZ_{t-1} + \xi_t, \tag{3.6}$$

where:
$\xi_t = L \otimes (b'_t\, Z_{t-1} + \epsilon_t)$, where L is the $n \times 1$ vector whose only non-zero entry is the n'th, which is 1.
Then necessary and sufficient conditions for the existence of a unique strictly stationary solution to (3.5) possessing finite second order moments are:

1. The largest eigenvalue modulus of M is less than unity.

2. $(vec\, w^2)'\; vec\, \Sigma < 1$, where $vec\, \Sigma$ is the last column of the matrix $(I - M \otimes M)^{-1}$.

In the derivation of the asymptotic properties of the estimates the boundedness of the second moments of $\{X_t\}$ is required. If the two criteria considered in (1.) and (2.) are bounded away from unity it follows that this moment condition will be satisfied. Then, we replace (v)' with

(v) The largest eigenvalue modulus of M is less than or equal to $(1-\delta_3)$ and $(vec\, w^2)'\; vec\, \Sigma \leq 1 - \delta_4$.
where $\delta_3 > 0$ and $\delta_4 > 0$ are both arbitrarily small.

3.4.2 The maximum likelihood procedure

We consider sample of size N of the process $\{X_t\}$ generated by (3.5). We will give the conditional likelihood function for the values $(X_{1-n}, ..., X_0)$, we also assume that b_t and ϵ_t are Gaussian.
Let $f_s(x_t, ..., x_{t-s+1}|A_{t-s})$ be the density of $x_t, ..., x_{t-s+1}$, given as an event A_{t-s} in the $\sigma-$field $\mathcal{F}_{t-s}$ generated by $x_{t-s}, x_{t-s-1}, ...$

So, from the structure of (3.5), we have

$$f(x_t \mid x_{t-1},\ x_{t-2}, ...) = f(x_t \mid x_{t-1},\ x_{t-2},\ ..., x_{t-n}).$$

Under conditions (i)...(v), ϵ_t and b_t are independent of $X_{t-j}, j > 0$. We start by calculating the conditioned expectation and variance.

$$
\begin{aligned}
E(X_t \mid Z_{t-1}) &= E\left[\left(\sum_{i=1}^{n} (\psi_i + b_{it})X_{t-i} + \epsilon_t\right) \mid Z_{t-1}\right] \\
&= \psi' Z_{t-1}.
\end{aligned}
$$

$$
\begin{aligned}
E(X_t^2 \mid Z_{t-1}) &= E\left[(\epsilon_t + b_t' Z_{t-1})^2 \mid Z_{t-1}\right] \\
&= E\left[\left(\epsilon_t^2 + Z_{t-1}'\, b_t'\, b_t\, Z_{t-1} + 2\epsilon_t\, b_t'\, Z_{t-1}\right) \mid Z_{t-1}\right] \\
&= E\left(Z_{t-1}'\, b_t'\, b_t\, Z_{t-1} \mid Z_{t-1}\right) + E\left(\epsilon_t^2 \mid Z_{t-1}\right) + 2E\left(\epsilon_t\, b_t'\, Z_{t-1} \mid Z_{t-1}\right) \\
&= Z_{t-1}'\, E(b_t'\, b_t) Z_{t-1} + \sigma^2 \\
&= Z_{t-1}'\, w^2\, Z_{t-1} + \sigma^2.
\end{aligned}
$$

Hence, we get the following likelihood function:

$$
\begin{aligned}
f_N(x_1, \ldots, x_N | x_0, \ldots, x_{1-n}) &= \prod_{t=1}^{N} f(x_1 | x_{t-1}, \ldots, x_{t-n}) \\
&= \prod_{t=1}^{N} \frac{1}{\sqrt{2\Pi(Z_{t-1}'\, w^2\, Z_{t-1} + \sigma^2)}} \exp\left\{\frac{-1}{2} \frac{(x_t - \psi' Z_{t-1})^2}{Z_{t-1}'\, w^2\, Z_{t-1} + \sigma^2}\right\} \\
&= (2\Pi)^{-\frac{N}{2}} \prod_{t=1}^{N} \left(w^2\, Z_{t-1} + \sigma^2\right)^{\frac{-1}{2}} \exp\left\{\frac{-1}{2} \frac{(x_t - \psi' Z_{t-1})^2}{Z_{t-1}'\, w^2\, Z_{t-1} + \sigma^2}\right\} \\
&= L_N(\psi, \sigma^2, w^2).
\end{aligned}
$$

It's more convenient to consider the following monotone likelihood function

$$
\begin{aligned}
l_N(\psi, \sigma^2, w^2) &= \tfrac{-2}{N} \log\left[L_N(\psi, \sigma^2, w^2)\right] - \log(2\pi) \\
&= \tfrac{-2}{N} \left[\tfrac{-N}{2} \log(2\pi) - \tfrac{1}{2} \sum_{t=1}^{N} \log\left(\sigma^2 + Z_{t-1}' w^2 Z_{t-1}\right) - \tfrac{1}{2} \left(\sum_{t=1}^{N} \tfrac{(x_t - \psi' Z_{t-1})^2}{Z_{t-1}' w^2 Z_{t-1} + \sigma^2}\right)\right] \\
&\quad - \log(2\pi) \\
&= N^{-1} \sum_{t=1}^{N} \log\left(Z_{t-1}' w^2 Z_{t-1} + \sigma^2\right) + N^{-1} \sum_{t=1}^{N} \tfrac{(x_t - \psi' Z_{t-1})^2}{Z_{t-1}' w^2 Z_{t-1} + \sigma^2}.
\end{aligned}
\tag{3.7}
$$

This function is non-linear in σ^2 and w^2, then there is no closed form expression for the estimates $\widehat{\psi}$, $\widehat{\sigma}^2$ and $\widehat{w}^2$ which minimize l_N.

However, by letting $R = \frac{w^2}{\sigma^2}$, we can reduce the function to be minimized by concentrating out ψ and σ^2. Let $\widetilde{l}_N(\psi, \sigma^2, R) = l_N(\psi, \sigma^2, w^2)$, we'll get:

$$\begin{aligned}\widetilde{l}_N(\psi,\sigma^2,R) &= N^{-1}\sum_{t=1}^{N}\log\left(Z'_{t-1}\sigma^2 R Z_{t-1}+\sigma^2\right)+N^{-1}\sum_{t=1}^{N}\frac{\left(x_t-\psi' Z_{t-1}\right)^2}{Z'_{t-1}\sigma^2 R Z_{t-1}+\sigma^2}\\ &= N^{-1}\sum_{t=1}^{N}\left(\log(\sigma^2)+\log\left(1+Z'_{t-1}RZ_{t-1}\right)\right)+N^{-1}\sigma^{-2}\sum_{t=1}^{N}\frac{\left(x_t-\psi' Z_{t-1}\right)^2}{1+Z'_{t-1}RZ_{t-1}}\\ &= \log(\sigma^2)+N^{-1}\sum_{t=1}^{N}\left(\log\left(1+Z'_{t-1}RZ_{t-1}\right)\right)+N^{-1}\sigma^{-2}\sum_{t=1}^{N}\frac{\left(x_t-\psi' Z_{t-1}\right)^2}{1+Z'_{t-1}RZ_{t-1}}.\end{aligned}$$

Now for fixed R, we can minimize $\widetilde{l}_N$ and get $\widehat{\psi}=\psi_N(R)$ and $\widehat{\sigma}^2=\sigma^2_N(R)$,

$$\frac{\partial\widetilde{l}_N(\psi,\sigma^2,R)}{\partial\psi} = 2N^{-1}\sigma^{-2}\left[\sum_{t-1}^{N}\frac{Z'_{t-1}\psi Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}-\sum_{t-1}^{N}\frac{x_t Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}\right].$$

$$\begin{aligned}\frac{\partial\widetilde{l}_N(\psi,\sigma^2,R)}{\partial\psi}=0 &\Rightarrow 2N^{-1}\sigma^{-2}\left[\sum_{t-1}^{N}\frac{Z'_{t-1}\psi Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}-\sum_{t-1}^{N}\frac{x_t Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}\right]=0\\ &\Rightarrow \sum_{t-1}^{N}\frac{Z'_{t-1}\psi Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}=\sum_{t-1}^{N}\frac{x_t Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}\\ &\Rightarrow \psi(R)_N=\sum_{t-1}^{N}\frac{x_t Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}\left(\sum_{t-1}^{N}\frac{Z'_{t-1}Z_{t-1}}{1+Z'_{t-1}RZ_{t-1}}\right)^{-1}.\end{aligned}$$

$$\frac{\partial\widetilde{l}_N(\psi,\sigma^2,R)}{\partial\sigma^2}=0 \Rightarrow \sigma^2(R)_N=N^{-1}\sum_{t-1}^{N}\frac{\left(x_t-\psi(R)Z_{t-1}\right)^2}{1+Z'_{t-1}rZ_{t-1}}.$$

Having obtained the estimates of ψ and σ^2 in term of R, now we substitute them back into $\widetilde{l}_N$ which is now a function of R only.

Minimizing $\widetilde{l}_N(\psi,\sigma^2,R)$ by using this concentrated maximum likelihood approach, equivalent to minimizing a single function of R.

So, maximum likelihood estimators $\widehat{\psi}_N$, $\widehat{\sigma}^2_N$ and $\widehat{w}^2_N$, are given by :

$$\widehat{\sigma}^2_N=\sigma^2_N(\widehat{R}_N),$$

$$\widehat{\psi}_N^2 = \psi_N^2(\widehat{R}_N),$$
$$\widehat{w}_N^2 = \widehat{\sigma}_N^2\ \widehat{R}_N.$$

This procedure would be useful if one were using an optimization algorithm not requiring first and second derivatives of the function being optimized. If however such derivatives will be required and as in the case of $\widetilde{l}_N$, these are complicated. For this reason it is better to minimize

$$l_N(\psi, r) \quad = \quad \inf_{\sigma^2} \widetilde{l}_N(\psi, \sigma^2, R) - 1,$$

where $r =$ *vech* R is the vector of non redundant elements of R.

Lemma 3.4.1. *If* $y_t = G\ vec\ (Z_{t-1}Z'_{t-1})$*, where* G *is the matrix satisfying*

$$vec\ X = G'\ vech\ X,$$

for all symmetric $(n \times n)$ *matrices* X*, then*

$$\begin{aligned} Z'_{t-1}RZ_{t-1} &= tr(RZ_{t-1}Z'_{t-1}) = [vec\ R]'\ vec\ (Z_{t-1}Z'_{t-1}) \\ &= [vech\ R]'D\ vec\ (Z_{t-1}Z'_{t-1}) = r'y_t. \end{aligned}$$

Hence

$$l_N(\psi, r) = N^{-1}\sum_{t=1}^{N}(\log{(1 + r'y_t)}) + \log\left\{N^{-1}\sum_{t=1}^{N}\frac{(x_t - \psi' Z_{t-1})^2}{1 + r'y_t}\right\}. \tag{3.8}$$

Now, we can easily obtain the first two derivatives of l_N with respect to ψ and r

$$L_N(\widehat{\psi}_N, \widehat{r}_N) = \inf_{(\psi', r')\in\Theta} l_N(\psi, r),$$

$$\widehat{\sigma}_N^2 = N^{-1}\sum_{t=1}^{N}\frac{(X_t - \widehat{\psi}'_N Z_{t-1})^2}{(1 + \widehat{r}'_N y_t)},$$

and

$$vech\ (\widehat{w}_N^2) = \widehat{\sigma}_N^2 \widehat{r}_N.$$

3.4.3 Strong consistency of estimators

The set Θ on which $l_N(\psi, r)$ is minimized depends on two positive numbers δ_3 and δ_5 which are arbitrarily small. Θ is the set of all $n(n+3)/2$ dimensional vectors $[\psi', r']'$ with ψ and r containing n and $n(n+1)/2$ elements respectively, satisfying the following conditions:

(ci) ψ is such that the largest eigenvalue modulus of M is less than or equal to $(1-\delta_3)$.

(cii) By letting R the symmetric square matrix for which $r = vech\ R$, then the smallest eigenvalue of R is greater than or equal to δ_5.

Assuming that $\theta_0 = (\psi_0', r_0')'$ is the vector of the true values of the parameters. And X_t is a strictly stationary $\mathcal{F}_t$-measurable solution of (3.5) satisfying conditions (i)-(v), and for which $\psi = \psi_0$, $w^2 = w_0^2$, $\sigma^2 = \sigma_0^2$. And $r = r_0 = w_0^2/\sigma_0^2$.

The proof of the strong consistency of the maximum likelihood estimators requires that Θ be a subset of the compact $\mathbb{R}^{n(n+3)/2}$. In particular, we know that any continuous function on Θ reaches its minimum and its Maximum.

Theorem. ***(Nicholls and Quinn (1982))***
Let X_t be the strictly stationary solution, $\mathcal{F}_t$-measurable of (3.5) with $\psi = \psi_0$, $r^2 = r_0^2$, $\sigma^2 = \sigma_0^2$ under the conditions (i)-(v). And letting $\theta_0 = (\psi_0', r_0')'$ where $r_0 = w_0^2/\sigma_0^2$. Then, $\lim_{N\to\infty} l_N(\psi, r)$ exist a.s $\forall(\psi', r')' \in \Theta$ and the limit $l_N(\psi, r)$ admits a single minimum in $(\psi', r')' \in \Theta_0$ for $\Theta_0 \in int\Theta$.

Theorem. ***(Nicholls and Quinn (1982))***
If $l_N(\widehat{\psi}_N, \widehat{r}_N) = \min_{\psi,r} l_N(\psi, r)$, and $\theta_0 = (\psi_0, r_0) \in$ int (Θ), then $\widehat{\theta}_0 = (\widehat{\psi}_0, \widehat{r}_0) \in$ int (Θ) converge to θ_0 a.s.

Corollary 3.4.1. ***(Nicholls and Quinn (1982))***
$\widehat{w}_N^2$ and $\widehat{\sigma}_N^2$ converge to w_0^2 and σ_0^2 a.s.

3.4.4 The central limit theorem

The maximum likelihood estimators require only the existence of $E(X_t^2)$ to be strongly consistent.
There is a central limit theorem if the fourth-order moments of $\{\epsilon_t\}$ and $\{b_t\}$ are finite.

To prove the central theorem limit however, it's more convenient to consider log likelihood in the form (3.7) where letting $\gamma = \ vech\ w^2$, (3.7) becomes

$$l_N(\psi, \sigma^2, \gamma) = N^{-1} \sum_{t=1}^{N} \log\left(\sigma^2 + \gamma' y_t\right) + N^{-1} \sum_{t=1}^{N} \frac{\left(x_t - \psi' Z_{t-1}\right)^2}{\sigma^2 + \gamma' y_t}.$$

Of course $l_N(\psi, r)$ defined by (3.8) and $l_N(\psi, \sigma^2, \gamma)$ give the same estimates of the parameters.

Lemma 3.4.2. ***(Nicholls and Quinn (1982))***
The sequence $\left\{\frac{\partial^2 l_N(\theta)}{\partial\theta\partial\theta'}\right\}$ *converges uniformly almost surely on a compact neighborhood of* θ_0 *to* $\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}$.

Lemma 3.4.3. ***(Nicholls and Quinn (1982))***

Let $u_{0,t} = x_t - \psi_0' Z_{t-1}$, $\lambda_{0,t} = \sigma_0^2 + \gamma_0' y_t$, *and* $\eta_t = u_t^2 - \lambda_t$, *then the first-order partial derivatives are given by*

$$\frac{\partial l_N(\theta_0)}{\partial \psi} = -2N^{-1} \sum_{t=1}^{N} \lambda_{0,t}^{-1} u_{0,t} Z_{t-1},$$

$$\frac{\partial l_N(\theta_0)}{\partial \sigma^2} = N^{-1} \sum_{t=1}^{N} \lambda_{0,t}^{-2} (\lambda_{0,t} - u_{0,t}^2),$$

$$\frac{\partial l_N(\theta_0)}{\partial \gamma} = N^{-1} \sum_{t=1}^{N} \lambda_{0,t}^{-2} y_t (\lambda_{0,t} - u_{0,t}^2).$$

And the second derivatives of $l_N(\theta)$ are given by:

$$\frac{\partial^2 l_N(\theta)}{\partial\psi\partial\psi'} = \quad 2N^{-1}\sum_{t=1}^{N} \lambda_t^{-1} Z_{t-1} Z'_{t-1},$$

$$\frac{\partial^2 l_N(\theta)}{\partial\psi\partial\gamma'} = \quad 2N^{-1}\sum_{t=1}^{N} \lambda_t^{-2} u_t Z_{t-1} y'_t,$$

$$\frac{\partial^2 l_N(\theta)}{\partial\psi\partial\sigma^2} = \quad 2N^{-1}\sum_{t=1}^{N} \lambda_t^{-2} u_t Z_{t-1},$$

$$\frac{\partial^2 l_N(\theta)}{\partial\gamma\partial\gamma'} = \quad N^{-1}\sum_{t=1}^{N} 2\lambda_t^{-3} u_t^2 y_t y'_t - N^{-1}\sum_{t=1}^{N} \lambda_t^{-2} y_t y'_t,$$

$$\frac{\partial^2 l_N(\theta)}{\partial\gamma\partial\sigma^2} = \quad N^{-1}\sum_{t=1}^{N} 2\lambda_t^{-3} u_t^2 y_t - N^{-1}\sum_{t=1}^{N} \lambda_t^{-2} y_t,$$

$$\frac{\partial^2 l_N(\theta)}{\partial^2\sigma^4} = \quad N^{-1}\sum_{t=1}^{N} 2\lambda_t^{-3} u_t^2 - N^{-1}\sum_{t=1}^{N} \lambda_t^{-2},$$

where $\lambda_t = \sigma^2 + \gamma' y_t$.
$\left\{\frac{\partial^2 l_N(\theta)}{\partial\theta\partial\theta'}\right\}$ converge a.s to the matrix $\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}$ defined by:

$$\frac{\partial^2 l(\theta)}{\partial\psi\partial\psi'} = \quad 2E\left[\lambda_t^{-1}\ Z_{t-1}\ Z'_{t-1}\right],$$

$$\frac{\partial^2 l(\theta)}{\partial\psi\partial\gamma'} = \quad 2E\left[\lambda_t^{-2}\ u_t\ Z_{t-1}\ y'_t\right],$$

$$\frac{\partial^2 l(\theta)}{\partial\psi\partial\sigma^2} = \quad 2E\left[\lambda_t^{-2}\ u_t\ Z_{t-1}\right],$$

$$\frac{\partial^2 l(\theta)}{\partial\gamma\partial\gamma'} = \quad 2E\left[\lambda_t^{-3}\ u_t^2\ y_t\ y'_t\right]\ -\ E\left[\lambda_t^{-2}\ y_t\ y'_t\right],$$

$$\frac{\partial^2 l(\theta)}{\partial\gamma\partial\sigma^2} \qquad = 2E\left[\lambda_t^{-3}\ u_t^2\ y_t\right]\ -\ E\left[\lambda_t^{-2}\ y_t\right],$$

$$\frac{\partial^2 l(\theta)}{\partial^2\sigma^4} = \quad 2E\left[\lambda_t^{-3}\ u_t^2\right]\ -\ E\left[\lambda_t^{-2}\right].$$

Letting $I = \frac{\partial^2 l(\theta_0)}{\partial\theta\partial\theta'}$, then we get the following submatrix:

$$
\begin{aligned}
I_{11} = \quad & \tfrac{\partial^2 l(\theta_0)}{\partial\psi\partial\psi'} = 2E\left[\lambda_{0,t}^{-1}\ Z_{t-1}\ Z'_{t-1}\right], \\
I_{12} = \quad & \tfrac{\partial^2 l(\theta_0)}{\partial\psi\partial\gamma'} = 2E\left[E[u_{0,t}|F_{t-1}]\ \lambda_{0,t}^{-2}\ Z_{t-1}\ y'_t\right] = 0, \\
I_{13} = \quad & \tfrac{\partial^2 l(\theta_0)}{\partial\psi\partial\sigma^2} = 2E\left[E[u_{0,t}|F_{t-1}]\lambda_{0,t}^{-2}\ Z_{t-1}\right] = 0, \\
I_{22} = \quad & \tfrac{\partial^2 l(\theta_0)}{\partial\gamma\partial\gamma'} = 2E\left[\lambda_{0,t}^{-3}\ E[u_{0,t}^2|F_{t-1}]\ y_t\ y'_t\right] \ - \ E\left[\lambda_{0,t}^{-2}\ y_t\ y'_t\right] \\
= \quad & E[\lambda_{0,t}^{-2}\ y_t\ y'_t], \\
I_{23} = \quad & \tfrac{\partial^2 l(\theta_0)}{\partial\gamma\partial\sigma^2} = E\left[\lambda_{0,t}^{-2}\ y_t\right], \\
I_{33} = \quad & \tfrac{\partial^2 l(\theta_0)}{\partial^2\sigma^4} = E\left[\lambda_{0,t}^{-2}\right].
\end{aligned}
$$

where $E(u_{0,t}|F_{t-1}) = E[(\epsilon_t + \psi' Z_{t-1})|F_{t-1}] = 0$, and $E(u_{0,t}^2|F_{t-1}) = \lambda_{0,t}$.

Theorem. ***(Nicholls and Quinn (1982))***
Let $\{X_t, t \in \mathbb{Z}\}$ be a strictly stationary process $\mathcal{F}_t$-measurable, satisfying (3.5) *under the conditions (i)-(v), then $N^{1/2}(\widehat{\theta}_N - \theta_0) \sim N(0, I^{-1}JI^{-1})$ asymptotically.*

The covariance matrix will be reduced to $2I^{-1}$ if $\{\epsilon_t\}$ and $\{b_t\}$ are jointly normally distributed. And the matrix J is given by

$$
J = \begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J'_{12} & J_{22} & J_{23} \\ J'_{13} & J'_{23} & J_{33} \end{bmatrix},
$$

where

$$
\begin{aligned}
J_{11} = & \; 4\, E[E(u_{0,t}^2|F_{t-1})\, \lambda_{0,t}^{-2}\, Z_{t-1}\, Z_{t-1}'] \\
= & \; 4E[\lambda_{0,t}^{-1}\, Z_{t-1}\, Z_{t-1}'], \\
J_{12} = & \; 2\, E[E(u_{0,t}\, \eta_t|F_{t-1})\, \lambda_{0,t}^{-3}\, Z_{t-1}\, y_t'] \\
= & \; 2\, E[u_{0,t}^3\, \lambda_{0,t}^{-3}\, Z_{t-1}\, y_t'], \\
J_{13} = & \; 2\, E[E(u_{0,t}\, \eta_t|F_{t-1})\, \lambda_{0,t}^{-3}\, Z_{t-1}] \\
= & \; 2\, E[u_{0,t}^3\, \lambda_{0,t}^{-3}\, Z_{t-1}], \\
J_{22} = & \; E[\eta_t^2\, \lambda_{0,t}^{-4}\, y_t\, y_t'], \\
J_{23} = & \; E[\eta_t\, \lambda_{0,t}^{-4}\, y_t], \\
J_{33} = & \; E[\eta_t\, \lambda_{0,t}^{-4}].
\end{aligned}
$$

3.5 Simulation

3.5.1 Simulation of RCA(p)

1) Specify the real and complex eigenvalues $\{\lambda_i; i = 1, ..., p\}$ of the M matrix, all of which must have modules smaller than one.

2) Calculate the parameters $\{b_t; t = 1, ..., p\}$ of

$$\prod_{i=1}^{p}(1 - \lambda_i y) = 1 - \sum_{i=1}^{p} b_i y^i.$$

From there, calculate the square matrixΣ where $vec\ (\Sigma)$ is the last column of $(I - M \otimes M)^{-1}$.

3) Take a positive defined matrix W^* and calculate $tr\ (W^*\Sigma)$.

In order to guarantee second-order stationarity, we must have

$$1 > vec\ (W)'\ vec\ (\Sigma) = tr\ (W\Sigma),$$

so, if we specify the necessary value of $\rho = tr\ (W\Sigma)$, then, $W = W^*(\rho/tr\ (W^*\Sigma)$ checks $\rho = tr\ (W\Sigma)$. It is therefore sufficient to define the matrix W^*.

4) Calculate the lower triangular matrix L which has a positive diagonal and for which $LL' = W$.

5) Generate a vector $(w_1, ..., w_{p+1})'$ of random numbers normally distributed. Specifying $\sigma^2 = E(\epsilon_t^2)$, and we consider $\epsilon_t = \sigma w_1$ and $B_t' = L(w_1, ..., w_{p+1})'$. So, ϵ_t and b_t will be theoretically independent, will have zero averages and will have $\sigma^2 = E[\epsilon_t^2]$ and $E[b_t' b_t] = LL' = W$.

6) Calculate

$$X_t = \sum_{i=1}^{p} (\psi_i + b_{it}) X_{t-1} + \epsilon_t, \ where \ X_t = 0, \ for \ t < 0.$$

7) Repeat steps 5 and 6 $(N + 200)$ times where N is sample size, and ignore the first 200 produced values. This allows X_t to reach a balance.

3.5.2 Simulation of RCA(2)

We will simulate two random coefficients autoregressive models of order two, respectively a stationary model and a model close to the non-stationary zone, for the samples sizes 50, 200 and 500 using the R software. We will estimate their parameters by the maximum likelihood method and calculate their RMSE. Then we will study their asymptotic normality.

We consider the two sequences of random variables b_t and ϵ_t are supposed to be normally distributed. the steps are the following:

```
RCA2 <- function(l1 ,l2 ,W,r ,sge){
phi1 <- l1+l2
phi2 <- l1*l2
   M <- matrix(c(0,phi2 ,1 ,phi1),nrow=2)     #The matrix M
   D <- matrix(c(1,0,0,(-phi2)^2,0,1,(-phi2),(-phi1*phi2),0,
(-phi2),1,(-phi2*phi1),-1,-phi1,-phi1,1-phi1^2),nrow=4) # the
matrix D=((I-M*M)^-1)
  R  <- solve(D)  #the eigenvalues of the
matrix* D
  W  <- matrix(R[,4],nrow=2) # the last line of matrix D
  S  <- W%*%Y # the positive definite matrix W*
  c  <- sum(diag(S)) #the trace of the matrix W*
  K  <- W * (r/c)
  L  <- chol(K)
```

```
  X    <- rep(0,1200)
  w    <- rnorm(3) X[1]<-w[1]*sge
  w    <- rnorm(3) B<-L%*%w[2:3]
X[2]  <- (phi1+B[1])*X[1]+sge*w[1]

       for(i in 3:1200){
       w <- rnorm(3)
       B=L%*%w[2:3]
       X[i]=(phi1+B[1])*X[i-1]+(phi2+B[2])*X[i-2]+sge*w[1]
       }

X <- X[200:(n+200)] X return(X) }
```

For each process we minimized the function (3.8), then we replaced the value of r_N in ψ_N, w_N^2 and σ_N^2.

The estimators and the root mean square error (RMSE) are given in the following tables

	N	ψ_1	ψ_2	$w_{1,1}$	$w_{1,2}$	$w_{2,2}$	σ^2
The true values		0.01	0.3	0.19	0.01	0.19	1
The estimated values	50	0.1964	0.4080	0.2515	0.1207	0.2401	1.2904
RMSE		0.1864	0.108	0.0615	0.1307	0.1107	0.2904
The estimated values	200	0.0501	0.3496	0.2019	0.0346	0.2146	1.1020
RMSE		0.0401	0.0496	0.0119	0.0246	0.0246	0.102
The estimated values	500	0.0118	0.3023	0.1910	0.0110	0.1911	1.001
RMSE		0.0018	0.0023	0.001	0.001	0.0011	0.001

	N	ψ_1	ψ_2	$w_{1,1}$	$w_{1,2}$	$w_{2,2}$	σ^2
The true values		0.87	-0.1	0.12	0.12	0.21	1
The estimated values	50	0.70	-0.213	0.034	0.010	0.076	1.4413
RMSE		0.192	0.113	0.086	0.11	0.134	0.4413
The estimated values	200	0.763	-0.04	0.083	0.076	0.154	1.6607
RMSE		0.107	0.06	0.037	0.044	0.056	0.6607
The estimated values	500	0.784	-0.061	0.091	0.085	0.182	1.421
RMSE		0.086	0.039	0.029	0.035	0.028	0.421

We notice when n increases, the root-mean-square error decreases, i.e. we get more accurate estimators.
We also note that the RMSE for the model where the non-random parameter close to the non-stationarity zone is larger than the other model.

To examine asymptotic normality of these estimators, we calculate their kurtosis and their skewness.
We have obtained the following results.

Parameters estimated by the ML method	Kurtosis	Skewness
ψ_1	3.1034	-1.16547
ψ_1	2.6876	0.4532
$w_{1,1}$	3.0123	0.3446
$w_{1,2}$	2.8764	-0.4365
$w_{2,2}$ σ^2	2.6778	0.1602

Parameters estimated by the ML method	Kurtosis	Skewness
ψ_1	2.5035	-1.1427
ψ_1	2.3699	0.6302
$w_{1,1}$	3.3120	0.4653
$w_{1,2}$	2.5794	-0.5355
$w_{2,2}$ σ^2	2.4973	0.5089

We note that the model's parameter estimators have a distribution that approaches a normal distribution since their kurtosis is about 3 and the absolute value of their skewness is less than 0.5.
We also note that the estimators of the first model are closer to normality than those of the second model.

Conclusion:

The higher the value of n, the better the quality of the estimator, the RMSE proves it. Same for normality.
Conversely, the closer we get to the non-stationarity, the poorer the quality of the estimator. The RMSE values highlight this. Also, we deviate from the normality.

Conclusion

In many disciplines when modeling some phenomenon by means of an autoregressive model, the coefficients arise as a result of a number of complicated factors usually displaying random features. When such situations occur it is more realistic to consider random coefficient autoregressive (RCA) models, in which the coefficients are not constant but subject to random perturbations.

In this dissertation, we have developed such models, which are autoregressive models with random coefficients of order p. Based on the work of Aue and al (2006), we proved the consistency and the asymptotic normality of the maximum likelihood estimators. The validity of the theoretical results was highlighted by simulations.

As future perspectives, it would be interesting to extend these results to stationary random fields.

Annex

※ **Some operators and their properties:**
Let A, B, C and D be finite order matrices.

▶ **Kronecker product**

$$A \otimes B = (a_{ij}B)_{i,j}$$

▶ **vec and vech operators**

- vec(A) is obtained from A by stacking the columns of A, one on top of the other, in order, from left to right.

- If A is a symmetric matrix, vech(A) obtained by vectorizing only the lower triangular part of A.

If matrix products are possible, then

- $vec(A\ B\ C) = (C' \otimes A) vec\ B$

- $trace(A\ B) = (vec\ (B'))'\ vec\ A = (vec\ B)'\ vec\ A$

- $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

- $(A \otimes B)\ (C \otimes D) = (A\ C) \otimes (B\ D)$

- $(A \otimes B)' = A' \otimes B'$

▶ For a p-order symmetric matrix A, we have:

- $vec\ A \ = \ K_p'\ vech\ \ A$

- $vech\ A \ = \ H_p\ vec\ A$

where G and H_p are a p-order symmetric A matrix, $(\frac{p(p+1)}{2} \times p_2)$ such that $H_p G' \ = \ I_{p(p+1)/2)}$

▶ Let M be a $(p \times p)$ matrix such that

$$M = G\ MV' + Q \qquad (*)$$

Where G, V and Q are a $(p \times p)$ order matrices. If $(I - V \ \otimes \ G)^{-1}$ exist, there's a unique solution M of $(*)$ such that :

$$vec\ M \ = \ (I - V \ \otimes \ G)^{-1}\ vec\ Q$$

Proof. (Nicholls and Quinn(1982) page 12).

▶ Let A_i and B_i be a defined matrices, then

$$(\prod_{i=0}^{j} A_i) \otimes (\prod_{k=0}^{j} A_k) = \prod_{i=0}^{j} (A_i \otimes B_i)$$

▶ **Ergodic theorem**

A stochastic process X_t is said to be ergodic

$$\underset{n \to +\infty}{lim} \frac{1}{n} \sum_i x_i = \mu \qquad \Longrightarrow \qquad \underset{n \to +\infty}{lim} E\left[\left(\frac{1}{n} \sum_i x_i - \mu\right)^2\right] = 0$$

An ergodic process is stationary, the reverse is not true.

- Let $T : X \to X$ be a measure-preserving transformation on a measure space (X, σ, μ) and suppose f is a μ-integrable function, i.e. $f \subset L^1(\mu)$. Then we define the following averages: Time average: This is defined as

Time average: This is defined as the average (if it exists) over iterations of T starting from some initial point x
$\widehat{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(\mathrm{T}^k x)$.

► **Irreducible process**

Let $\{X_t\}$ be a generalized autoregressive model with i.i.d. coefficients, such that

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \qquad n \in \mathbb{Z}, \tag{**}$$

where $\{(A_n, B_n); n \in \mathbb{Z}\}$ are pairs of independent and identically distributed random variables defined on a probabilized space $(\Omega, \mathcal{A}, P)$.

An affine subspace H of $\mathbb{R}^d$ is said to be invariant under the model (**).
If $\{\ A_0 x + B_0; x \in H\}$ is contained in H almost surely. The model (**) is called irreducible if $\mathbb{R}^d$ is the only affine invariant subspace.

► **Berkes and al.(2003): Lemma 2.2**

If $\{\xi_k, 0 \le k < \infty\}$ is a sequence of identically distributed random variables, satisfying

$$E \log^+ |\xi_0| < \infty,$$

then, $\sum\limits_{0\le k<\infty} \xi_k Z^k$ converges with probability one for any $|Z| < 1$.

▶ **Dominated convergence theorem**

Let $\{f_n\}$ a measurable sequence of functions, $\forall n \neq 1$,

$$f_n : (\Omega, \mathcal{F}) \to (\overline{\mathbb{R}}, B(\overline{\mathbb{R}})).$$

Assuming that $\forall n \neq 1$, f_n in $\overline{\mathbb{R}}^+$ μ-almost everywhere, and $f_n \leq f_{n+1}; \forall n > 1$. Note then,

$$\lim_{n\to\infty} \int_{\Omega} f_n d\mu = \int_{\Omega} (\lim_{n\to\infty} f_n) d\mu.$$

▶ **Cesàro's theorem**

Let $(a_n)_{n\in\mathbb{N}^*}$ a sequence of real or complex numbers. If it converges to l, then the sequence of its Cesàro means, of general term $c_n = \frac{1}{n}\sum_{k=1}^{n} a_k$, also converges, and its limit is l.

▶ **Skewness**

Skewness refers to distortion or asymmetry in a symmetrical bell curve, or normal distribution, in a set of data.

Let X be a real random variable with mean μ and standard deviation σ, its asymmetry coefficient is defined as the third-order moment of the reduced centrad variable :

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Bibliography

[1]. ANDEL, J. (1976) Autoregressive series with random parameters. Math. Operationsforsh. u. Statist. 7,735-741

[2]. Aue, A., Horvath, L., Steinebach, J. (2006) : Estimation in Random Coefficient Autoregressive Models. Journal of Time Series Analysis. 27 (1) : 61-76.

[3]. Berkes, I., Horvàth, L. and Kokoszka, P. (2003) GARCH processes: structure and estimation. Bernoulli 9, 201-27.

[4]. Billingsley, P. (1968) Convergence of Probability Measures. New York: Wiley.

[5]. Borges, R. and Pfanzagl, J. (1965). One-parameter exponential families generated by transformation groups. Ann. Math. Statist. 36 261–271

[6]. Bougerol Philippe and Nico PicardSource, (Oct., 1992). Strict Stationarity of Generalized Autoregressive Processes. The Annals of Probability, Vol. 20, No. 4, pp.1716-1717.
[67]. Y. Boularouka, K. Djeddourb. (2015), New approximation for ARMA parameters estimate, 116-122.

[8]. Brockwell, Peter J. and Davis, Richard A. (1991). Time series: theory and methods. Springer Verlag., Proposition 3.1.2 page 84.
[9]. Brockwell, Peter J, (2010. Richard A. Davis. Introduction to Time Series and Forecasting, second edition, 45-75.

[10]. Brockwell, Richard A. Davis, Peter J. (2016). Introduction to Time Series and Fore-

casting, third edition, 74-96.

[11]. Conlisk, J. (1974), Stability in a random coefficient model, Int. Eaon. Rev., XV, 529-533.

[12]. Conlisk, J. (1976), A further note on stability in a random coefficient model, Int. Eaon. Rev., XVII, 759-764.

[13].Garbade. Kenneth, (1977). "Two Methods for Examining the Stability of Regression Coefficients", Journal of the American Statistical Association 72. 54-63.

[14]. E.J. Godolphin,(1984). A direct representation for the large sample maximum likelihood estimator of a Gaussian autoregressive moving average process, Biometrika 71. 281–289.

[15]. Hannan, E.J. Rissanen, E.J. (1982), Recursive estimation of mixed autoregressive-moving average order, Biometrika, 69, 81-94.

[16]. S. Koreisha, T. Pukkila, Fast linear estimation methods for vector moving average models, J. Time Series Anal. 10 (1989) 325–339.

[17]. Ledolter, J (1980). Recursive estimation and adaptive forecasting in ARIMA models with time varying coefficients. In Second Applied Series Symposium (D. F. Findley, ed.). Academic Press, New York.

[18]. J.A. Mauricio,(2002). An algorithm for the exact likelihood of a vector autoregressive moving average model, J. Time Series Anal. 23. 473–486.

[19]. S. Mittnik, (1990). Computation of theoretical auto-covariance matrices of multivariate autoregressive moving average time series, J. R. Stat. Soc. Ser. B 52. 151–155.

[20]. D. F. Nicholls and B. G. Quinn, (1982). Random Coefficient Autoregressive Models: An Introduction, Springer-Verlag, New York, NY, USA.

[21]. D. F. Nicholls and B. G. Quinn, (1982) The Estimation of Random Coefficient Autoregressive Models. II.

[22]. B.L. Shea, (1988). A note on the generation of independent realizations of a vector autoregressive moving average process, J. Time Series Anal. 9. 403–410.

[23]. Subba Rao, T (2005), "On some non stationary, non linear random process and their stationary approximations", Wojing papper.

[24]. Tjøstheim, D. (1986). Estimation in nonlinear time series models, Stochastic Processes and Applications 21: 251–273.