Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS

2007 Pavel Vaněček

Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS

RNDr. Pavel Vaněček

Estimators of Random Coefficient Autoregressive Models

Department of Probability and Mathematical Statistics
Supervised by: doc. RNDr. Zuzana Prášková, CSc.
Study program: Mathematics
Study branch: m5 Econometrics and operational research

Karlova Univerzita v Praze
Matematicko-fyzikální fakulta

DIZERTAČNÍ PRÁCE

RNDr. Pavel Vaněček

Odhady v autoregresních modelech s náhodnými koeficienty

Katedra pravděpodobnosti a matematické statistiky
Školitelka: doc. RNDr. Zuzana Prášková, CSc.
Studijní program: Matematika
Studijní obor: m5 Ekonometrie a operační výzkum

I would like to thank to my supervisor doc. Prášková for her precious experience and patient guidance. I also thank to my parents for their continuous support during my whole study.

This work was supported by grant GAČR 201/05/H007 and partially supported by grant GAČR 201/03/0945.

I state that I wrote this thesis by myself and I agree with its lending.

In Prague, October 13th, 2007 Pavel Vaněček

Název práce: Odhady v autoregresních modelech s náhodnými koeficienty
Autor: RNDr. Pavel Vaněček
Katedra: Katedra pravděpodobnosti a matematické statistiky
Školitelka: doc. RNDr. Zuzana Prášková, CSc.
e-mail školitelky: praskova@karlin.mff.cuni.cz

Abstrakt: Práce se zabývá autoregresními modely s náhodnými koeficienty. Zaměřuje se na třídu odhadů parametrů těchto modelů, která byla původně navržena pouze pro modely prvního řádu. Rozšiřuje tyto odhady jednak pro obecnější modely prvního řádu a také pro modely vyšších řádů a vícerozměrné modely. Odhady navzájem porovnává, dokazuje jejich striktní konzistenci, asymptotickou normalitu a pro jednorozměrný model prvního řádu i rychlost konvergence rozdělení odhadu k normálnímu rozdělení. Vlastnosti odhadů demonstruje pomocí množství simulačních studií.

Klíčová slova: RCA, martingalové diference, odhady závislé na volbě generující funkce, Berry-Esséenova hranice

Title: Estimators of Random Coefficient Autoregressive Models
Author: RNDr. Pavel Vaněček
Department: Department of Probability and Mathematical Statistics
Supervisor: doc. RNDr. Zuzana Prášková, CSc.
Supervisor's e-mail address: praskova@karlin.mff.cuni.cz

Abstract: This work investigates Random Coefficient Autoregressive time series models. It focuses on estimation of their parameters using a general class of estimators initially proposed for the first-order models only. It relaxes assumptions of the models on which the estimators were studied originally and extends the concept to higher-order and multivariate models. It compares the estimators, proves their strong consistency and asymptotical normality. For the univariate first-order models, it also establishes the rate of convergence for the distribution of the estimators to normal distribution. Statistical properties of the estimators are illustrated through numerous simulations.

Keywords: RCA, martingale differences, functional estimators, Berry-Esséen bound

Contents

Notation

$\mathcal{F}_{-\infty}^{t} = \mathcal{F}_t = \sigma(X_s, s \leq t)$	σ-field generated by history of random process $\{X_t\}$ up to time t
$\mathcal{F}_{t+m}^{+\infty} = \sigma(X_s, s \geq t+m)$	σ-field generated by future of random process $\{X_t\}$ from time $t+m$
$\mathrm{E}[X_t \mid \mathcal{F}_{t-1}]$	conditional mean value of random variable X_t given σ-field $\mathcal{F}_{t-1}$
$\mathrm{var}[X_t \mid \mathcal{F}_{t-1}]$	conditional variance of random variable X_t given σ-field $\mathcal{F}_{t-1}$
$X \xrightarrow{D} N\big(0, V\big)$	convergence in distribution of random variable X to a random variable having normal distribution with mean 0 and variance V
$X_n \xrightarrow{a.s.} X$ as $n \to +\infty$	convergence almost surely, equivalent to $\mathrm{P}(X_n \to X \text{ as } n \to +\infty) = 1$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{I}$	identity matrix of particular dimension

Chapter 1

Introduction

Random coefficient autoregressive models, abbreviated in this work as RCA, have been studied for a long time which indicates their usefulness in econometric practice. One of the first pieces of work concerning these models were paper [2] by Anděl published in 1976 and publication [25] by Nicholls and Quinn from 1982.

These models belong to a broad class of conditional heteroscedastic time series models because of their varying conditional variance. Well-known autoregressive conditional heteroscedastic models (ARCH) introduced by Engle in [11] are in fact just special cases of RCA models, see [31] or [32]. In the latter article, Tsay states that "RCA models were widely investigated by time series analysts and ARCH models were investigated by econometricians" which together with second-order equivalency between RCA and ARCH models (meaning that they have the same conditional expectation and variance) justify the necessity of studying RCA models. It is nicely expressed by the author a few pages later: "Recognition of this connection provides useful insight into these two types of models. First, a combination of research in time series and econometrics provides sufficient theory for understanding and using these models: to econometricians, properties of the ARCH models, conditional and unconditional, can be derived directly form those of the RCA models available in the time series literature; to time series analysts, applications of the ARCH models in econometrics justify the practical value of the RCA models." Tsay then introduced extension of RCA concept called conditional heteroscedastic autoregressive and moving averages models (CHARMA) and the rest of the article is devoted to estimation of such models. CHARMA models are close to ARMA-ARCH models that can be briefly described as standard autoregressive moving average models with ARCH type error process, see for instance [38].

The relation between RCA and ARCH processes is investigated in [16] as well, where the author found a connection between RCA process and AR-ARCH process. There is also investigated an extension of RCA model into a model with heteroscedastic error process variance of which depends on time. Estimation of AR-ARCH process and tests for stationarity under weaker conditions are also described in the sequence of articles [21] and [22] written by Ling where the author uses double-autoregressive model as a special case of ARMA-ARCH models.

RCA models represent a natural extension of the well-known autoregressive processes. Specifically, process $\{X_t, t \in \mathbb{Z}\}$ is called RCA process of order p, in short RCA(p) process, if X_t for each $t \in \mathbb{Z}$ satisfies

$$X_t = \sum_{i=1}^{p} (\beta_i + B_{t,i}) X_{t-i} + Y_t, \tag{1.1}$$

where $\beta_1, \ldots, \beta_p$ are unknown constant parameters, $\{Y_t, t \in \mathbb{Z}\}$ is an error process with zero expectation and $\{\mathbf{B}_t = (B_{t,1}, \ldots, B_{t,p})', t \in \mathbb{Z}\}$ denotes a p-dimensional random coefficient process with zero expectation.

Natural discussion about necessity of such an extension of AR processes might appear, meaning whether the coefficients might not be treated as non-random without significant loss of accuracy. This leads to testing of randomness of the coefficients which is mentioned in [25] (Chapter 6) and summarized in [1] and [13], for instance.

Let us come back to RCA models given by equation (1.1). Their stochastic setup was initially defined by the following set of assumptions:

A1: Random coefficient process $\{\mathbf{B}_t, t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed random vectors with zero mean and constant finite variance matrix $\mathbf{\Sigma}$ (in case that $\mathbf{B}_t$ are univariate random variables, their variance is denoted as σ_B^2).

A2: $\{\mathbf{B}_t, t \in \mathbb{Z}\}$ and $\{Y_t, t \in \mathbb{Z}\}$ are mutually independent.

A3: Error process $\{Y_t, t \in \mathbb{Z}\}$ is a sequence of independent random variables with zero mean and finite variance σ^2.

Anděl in [2] derived conditions for existence of the RCA process as a weakly stationary solution of the singly-infinite stochastic difference equation (SDE) (1.1) starting at $t = 0$. In our doubly-infinite setup, given by equation (1.1) and assumtions A1–A3, the stationary solution is granted by the conditions derived in [25] which will be stated later. The authors of [25] also studied various estimators of unknown parameters $\boldsymbol{\beta}$, σ^2 and $\mathbf{\Sigma}$, in which case they needed more strict modification of assumption A3.

A4: Error process $\{Y_t, t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed random variables with zero mean and finite variance σ^2.

This classical stochastic setup was relaxed during past years in a lot of ways. For instance, Hwang and Basawa in [14] studied estimators of RCA models where the random coefficients and the error process are correlated through a time-independent covariance matrix. They called such models Generalized RCA process and investigated both estimation of the covariance structure and limit distribution of the least-squares estimators. Alternatively, the error process can be heteroscedastic with time-dependent variance σ_t^2, see [16], [17], and even then the assumption of independence could be weakened into martingale difference property, see [15].

In this work we will employ another type of generalization related to the choice of the error process in RCA model, previously used for instance in [5]. Namely, during particular sections we will use the following assumption:

A5: Error process $\{Y_t, t \in \mathbb{Z}\}$ is an ergodic and strictly stationary martingale difference sequence with respect to $\mathcal{F}_t = \sigma\left(B_s, Y_s; s \leq t\right)$ with finite variance σ^2.

Concerning parameter estimation of RCA models, we go down the non-Bayesian time domain path following the work of Nicholls and Quinn (see [25]). However, Bayesian approach seems to be also favorable and gives useful results, see for instance [37]. In that paper, Wang and Ghosh worked with RCA process as a three-level hierarchical model (they called it RCAR there) and did parameter estimation and unit-root testing in this Bayesian setup. Using simulation they concluded that this approach is suitable for smaller sample sizes with presence of volatility in the data and it is robust against model misspecification.

Estimation of parameters in RCA models, especially of the first-order, has been done in great detail by many authors. Basic least-squares estimators and their weighted versions are commonly used with various relaxations of model assumptions. Maximum likelihood estimator is also very popular (for proper derivation of the estimator see for instance [25]), although it does not have explicit expression in general. Quasi-maximum likelihood estimator of parameters in the first-order RCA models under optimal conditions is studied in [4], more elaborate estimators, namely generalized M-estimators, are introduced in [19] where asymptotic normality in a semiparametric model setup is discussed. Authors in [39] deal with an integer-valued RCA processes and they utilize estimator proposed by Schick in [28].

The aim of this work is to develop theory of the estimators of RCA models originally introduced by Schick in [28] for the first-order RCA process only. He proposed a whole class of functional $\sqrt{n}$-consistent estimators and we extent this concept for generalized RCA processes for higher-order and multivariate first-order cases. The following chapters are structured similarly, beginning with model definition and defining the function estimator. The model definition parts include theorems about existence of the generalized RCA process. After that we prove strong consistency and asymptotical normality of the functional estimator for all mentioned RCA models. We estimate the asymptotical variances, or variance matrices, and prove strong consistency of proposed estimators. We also discuss the optimal choice of the functional estimator in the context of efficiency of the estimators.

We begin with the first-order RCA models. Chapter 2 properly defines RCA(1) process and describes its basic properties. Then we mention commonly used least-squares estimators and maximum likelihood estimator. Section 2.4 presents functional estimator and findings done by Schick in [28]. In the subsequent paragraphs, we investigate statistical properties of such estimators, devoting the extra attention to rate of convergence to normal distribution.

Chapter 3 introduces the concept of functional estimators into higher-order RCA models. We derive strong consistency and asymptotical normality of such estimators. The asymptotical variance matrix is derived both in general and for three specific choices of the estimator. We also suggest a consistent estimator of the variance matrix there.

In a similar manner, Chapter 4 deals with multivariate first-order models. Such models have been briefly studied in [25]. Estimation of special cases of non-random RCA models, traditional vector autoregressive processes, is thoroughly discussed for instance in [23] (Section 3.2 for the least-squares estimation, Section 3.4 for the maximum likelihood estimation). There is also a remark about the random coefficient models in Section 18.2.1 and about an extension of multivariate AR models into stable ARMA models with time dependent coefficients. The regularity conditions for such models could also be found in [26].

Some conclusions and possible further topis are mentioned in Chapter 5. Since the statistical inference presented in the proofs requires more complex techniques, special theorems about convergency and more elaborate matrix operators are needed. Therefore, Chapter 6 consists of non-trivial auxiliary lemmas used throughout the previous chapters. For instance, the proofs concerning the rate of convergence in Section 2.4.4 require concepts and properties concerning the dependency structure (mixing, L_p-mixingale and near-epoch dependency in L_p-norm) together with the Berry-Esséen theorem. Some proofs, especially in Section 3.2.1, are based on matrix handling, so the Kronecker product, *vec* and *vech* operators and their relations are stated in Chapter 6.

We present illustrative simulations and figures throughout the whole text. All of them were done using software package *R*. Since we have not found any native estimation procedure in *R* for RCA processes, we created a few functions by ourselves, especially for simulation and estimation of RCA models. Commented source codes of the main procedures are attached in Chapter 7.

Chapter 2

Estimators of the first-order models

2.1 Model specification

We will study the first-order RCA models in this chapter. We will present various estimators of unknown parameters and compare them both theoretically and via simulation studies.

The first-order RCA process, abbreviated as RCA(1), is a special case of equation (1.1) for $p = 1$, namely

$$X_t = (\beta + B_t)\, X_{t-1} + Y_t, \tag{2.1}$$

where the previous notation remains unchanged except that we omit unnecessary indices and variance of univariate random coefficient process $\{B_t, t \in \mathbb{Z}\}$ will be denoted as σ_B^2. For further statistical inference, we will need the following stationarity assumption:

A6: $\beta^2 + \sigma_B^2 < 1$

Definition 2.1. ***RCA(1) process***
Real-valued random process $\{X_t, t \in \mathbb{Z}\}$ is called RCA(1) process if X_t satisfies $\forall t \in \mathbb{Z}$ equation (2.1) where B_t and Y_t fulfill assumptions A1, A2, A4, and A6.
Remark that (2.1) can be written in the form $X_t = \beta X_{t-1} + u_t$, where $u_t = B_t X_{t-1} + Y_t$.

If we denote $\mathcal{F}_t = \sigma\left(B_s, Y_s; s \le t\right)$ and notice that both B_t and Y_t are independent of $\mathcal{F}_{t-1}$, we can derive conditional moments of RCA(1) process.

$$\mathrm{E}[X_t|\mathcal{F}_{t-1}] = \mathrm{E}[(\beta + B_t)X_{t-1} + Y_t|\mathcal{F}_{t-1}] = \beta X_{t-1}, \tag{2.2}$$

$$\begin{aligned}
\operatorname{var}\left[X_t|\mathcal{F}_{t-1}\right] &= \mathrm{E}\big[\left(X_t - \mathrm{E}[X_t|\mathcal{F}_{t-1}]\right)^2|\mathcal{F}_{t-1}\big] = \mathrm{E}\big[\left(B_t X_{t-1} + Y_t\right)^2|\mathcal{F}_{t-1}\big] = \\
&= \mathrm{E}\big[B_t^2 X_{t-1}^2|\mathcal{F}_{t-1}\big] + 2\mathrm{E}[B_t X_{t-1} Y_t|\mathcal{F}_{t-1}] + \mathrm{E}\big[Y_t^2|\mathcal{F}_{t-1}\big] = \\
&= \sigma_B^2 X_{t-1}^2 + \sigma^2.
\end{aligned} \tag{2.3}$$

In case $\sigma_B^2 > 0$, the model has non-constant conditional variance which is sometimes referred to as conditional heteroscedasticity. When both error and coefficient processes are mutually independent *iid* sequences and when the stationarity assumption is met, there exists strictly stationary and ergodic solution of equation (2.1) (see [25], Corollary 2.2.1 and Theorem 2.7). In other words, there exists RCA(1) process according to Definition 2.1. From Section 2.4.2 to the end of this chapter, we will study a generalization of the previously defined RCA(1) process.

Definition 2.2. ***Generalized RCA(1) process***
Real-valued random process $\{X_t, t \in \mathbb{Z}\}$ is called Generalized RCA(1) process (GRCA(1) for short) if X_t satisfies $\forall t \in \mathbb{Z}$ equation (2.1) where B_t and Y_t fulfill assumptions A1, A2, A5 and stationarity condition A6.

There exist other generalizations in literature (see for instance [14]) which are also abbreviated as GRCA, so please do not get confused. Our extension covers for instance cases of RCA models with ARCH or GARCH error process.

Theorem 2.1.
There exists a Generalized RCA(1) process according to Definition 2.2 which is $\mathcal{F}_t$-measurable, ergodic and strictly stationary.

Proof: We will basically follow similar proofs from [25]. We will show that there exists an ergodic and strictly stationary $\mathcal{F}_t$-measurable process W_t, defined as $W_t \stackrel{\text{def}}{=} \sum_{j=0}^{+\infty} \prod_{k=0}^{j-1} (\beta + B_{t-k}) Y_{t-j}$, that satisfies stochastic difference equation (2.1).
We can formally iterate process $\{X_t\}$ given by equation (2.1)

$$\begin{aligned} X_t &= (\beta + B_t)X_{t-1} + Y_t = Y_t + (\beta + B_t)\left((\beta + B_{t-1})X_{t-2} + Y_{t-1}\right) = \\ &= Y_t + (\beta + B_t)Y_{t-1} + (\beta + B_t)(\beta + B_{t-1})X_{t-2} = \cdots = \\ &= Y_t + (\beta + B_t)Y_{t-1} + (\beta + B_t)(\beta + B_{t-1})Y_{t-2} + \\ &\quad + (\beta + B_t)(\beta + B_{t-1})(\beta + B_{t-2})Y_{t-3} + \ldots . \end{aligned}$$

Let us define $W_{t,r} \stackrel{\text{def}}{=} \sum_{j=0}^{r} \prod_{k=0}^{j-1} (\beta + B_{t-k}) Y_{t-j}$, for $t \in \mathbb{Z}$ and $r = 0, 1, 2, \ldots$ (where the product is defined as 1 for $j = 0$). Using independence of $\{B_t\}$, its mutual independence of $\{Y_t\}$, and assumption $\beta^2 + \sigma_B^2 < 1$ we have that variables $\{W_{t,r}, r \in \mathbb{N}\}$ possess Cauchy property for fixed t, because for $s < r$ it holds that

$$\begin{aligned} &\mathrm{E}|W_{t,r} - W_{t,s}|^2 = \\ &= \mathrm{E}\left|(\beta + B_t) \cdot \ldots \cdot (\beta + B_{t-s})Y_{t-s-1} + \ldots + (\beta + B_t) \cdot \ldots \cdot (\beta + B_{t-r+1})Y_{t-r}\right|^2 = \\ &= \sum_{i=s+1}^{r} \sum_{j=s+1}^{r} \mathrm{E}\left((\beta + B_t) \cdot \ldots \cdot (\beta + B_{t+1-i})Y_{t-i} \cdot (\beta + B_t) \cdot \ldots \cdot (\beta + B_{t+1-j})Y_{t-j}\right) = \\ &= \sum_{i=s+1}^{r} \mathrm{E}\left((\beta + B_t)^2 \cdot \ldots \cdot (\beta + B_{t+1-i})^2 Y_{t-i}^2\right) = \sum_{i=s+1}^{r} (\beta^2 + \sigma_B^2)^i \cdot \sigma^2 \le \\ &\le \sum_{i=s+1}^{+\infty} (\beta^2 + \sigma_B^2)^i \cdot \sigma^2 \to 0 \quad \text{as} \quad s \to +\infty. \end{aligned}$$

For $s > r$ the inference is similar. Thus, for each $t \in \mathbb{Z}$ there exists a limit in quadratic mean of $\{W_{t,r}, r \in \mathbb{N}\}$ for $r \to +\infty$ which is equal to W_t. W_t is a function of $Y_t, Y_{t-1}, \ldots$ and $B_t, B_{t-1}, \ldots$ which does not depend on time t, so $W_t \in \mathcal{F}_t$. Since both $\{B_t\}$ and $\{Y_t\}$ are strictly stationary, ergodic, and mutually independent, $\{W_t\}$ is also strictly stationary and ergodic (see Lemma 6.1). Moreover,

$$(\beta + B_t)W_{t-1} = (\beta + B_t)\sum_{j=0}^{+\infty}\prod_{k=0}^{j-1}(\beta + B_{t-1-k})Y_{t-1-j} = \sum_{j=0}^{+\infty}\prod_{k=-1}^{j-1}(\beta + B_{t-k-1})Y_{t-1-j} =$$
$$= \sum_{j=1}^{+\infty}\prod_{k=0}^{j-1}(\beta + B_{t-k})Y_{t-j} = W_t - Y_t,$$

so process $\{W_t\}$ satisfies equation (2.1) and we can refer to this process as Generalized RCA process. □

The previous theorem gives us the explicit solution of equation (2.1) in form

$$X_t = \sum_{j=0}^{+\infty}\prod_{k=0}^{j-1}(\beta + B_{t-k})Y_{t-j}, \tag{2.4}$$

where the product is defined as 1 for $j = 0$. Condition $\beta^2 + \sigma_B^2 < 1$ ensures that the infinite sum (2.4) exists in the L_2 sense. Recently published condition $-\infty \leq \mathrm{E}\ln|\beta + B_0| < 0$ derived in [4] guarantees that the sum converges almost surely (the latter condition means that $\beta + B_0$ is "less than one" on average, see [3]). We can conclude from (2.4), using independence of B_t and Y_t, that

$$\mathrm{E}X_t = 0, \tag{2.5}$$
$$\operatorname{var} X_t = \frac{\sigma^2}{1 - \beta - \sigma_B^2}. \tag{2.6}$$

The essential task is how to estimate the unknown parameters – in our case parameters β, σ_B^2, and σ^2. The last two are usually treated as auxiliary requisites for the first one to be estimated. Equation (2.1) defines the meaning of unknown parameters: Parameter β may be seen as magnitude of the dependence of present value X_t on past value X_{t-1}, σ_B^2 expresses variability of this dependence and σ^2 simply stands for volatility of the error (noise) process that blurs the dependence. Figure 2.1 indicates the influence of β and σ_B^2 using simulated RCA(1) models. Two upper plots differ in variance σ_B^2 and we can see that larger variance means more extremal values, more clustering of volatility. Two lower plots illustrate situation $\sigma_B^2 = 0$, when random coefficient process RCA(1) simplifies to standard autoregressive process AR(1), and we can see impact of the sign of parameter β – for negative β the values more frequently change their sign.

Suppose for a while, that we have a sample of size n drawn from some general model with unknown parameter θ to be estimated. We are interested whether estimator an $\widehat{\theta}_n$ is

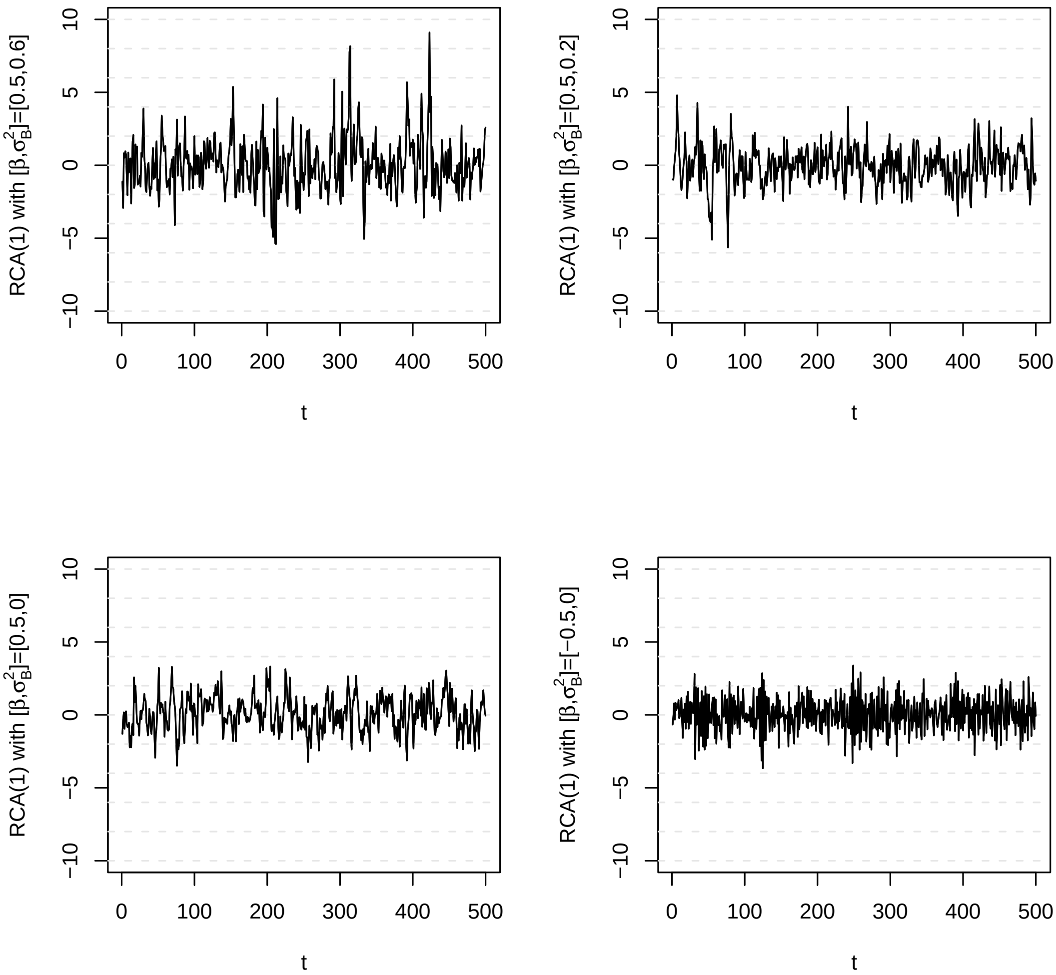

Figure 2.1: Simulated RCA(1) processes with $N(0,1)$ distributed error process Y_t independent of normally distributed random coefficients B_t where $[\beta, \sigma_B^2] = [0.5, 0.6]$ (upper left panel), $[0.5, 0.2]$ (upper right panel), $[0.5, 0]$ (lower left panel) and $[-0.5, 0]$ (lower right panel).

strongly consistent, in which case we know that the estimator is close to its true value when the sample size is large enough (without any information about rate of the convergence). Providing that the estimator is asymptotically normal, we can also estimate the variance of difference $\widehat{\theta}_n - \theta$. We can numerically compare estimators by either their asymptotic variances or their mean square errors defined as $MSE_\theta = \mathrm{E}(\widehat{\theta}_n - \theta)^2$. Whenever it is possible we omit in notation the explicit dependence on sample size n, so we put $\widehat{\theta} = \widehat{\theta}_n$ for instance.

2.2 Least-squares estimators

The similarity of RCA and AR models evokes that we could estimate the parameters of RCA process using the conditional least-squares method (LS estimator). The estimator which minimizes quantity

$$\sum_{t=1}^{n}(X_t - \mathrm{E}[X_t|\mathcal{F}_{t-1}])^2 = \sum_{t=1}^{n}(X_t - \beta X_{t-1})^2$$

is equal to

$$\widehat{\beta}_{LS} = \frac{\sum\limits_{t=1}^{n} X_t \cdot X_{t-1}}{\sum\limits_{t=1}^{n} X_{t-1}^2}. \tag{2.7}$$

Parameters σ^2 and σ_B^2 can be estimated as follows: Analogously to equation (2.3), we can derive that $\mathrm{E}[u_t^2|\mathcal{F}_{t-1}] = \sigma^2 + \sigma_B^2 X_{t-1}^2$. Then the linear regression equation

$$\widehat{u}_t^2 = \sigma^2 + \sigma_B^2 \cdot X_{t-1}^2 + \varepsilon_t$$

is solved where $\widehat{u}_t = X_t - \widehat{\beta}_{LS} X_{t-1}$ are the estimated residuals and ε_t is a noise process. The standard Gauss-Markov solution of the regression equation, provided that $n > 0$ and that $\{X_t^2, t = 0, 1, 2, \ldots, n\}$ is not a constant sequence, is equal to

$$\begin{aligned}
\widehat{\sigma}_B^2 &= \frac{\sum_{t=1}^{n}(X_{t-1}^2 - \overline{X})(X_t - \widehat{\beta}_{LS}X_{t-1})^2}{\sum_{t=1}^{n}(X_{t-1}^2 - \overline{X})^2}, \\
\widehat{\sigma}^2 &= \frac{1}{n}\sum_{t=1}^{n}(X_t - \widehat{\beta}_{LS}X_{t-1})^2 - \widehat{\sigma}_B^2\overline{X},
\end{aligned} \tag{2.8}$$

where $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2$.

Obviously $\mathrm{E}u_t = 0$. Although the variance of errors u_t is constant, which can easily be seen using variance of X_t given by equation (2.6) and noticing that

$$\mathrm{var}\, u_t = \mathrm{E}\left(\mathrm{E}\left[u_t^2|\mathcal{F}_{t-1}\right]\right) = \mathrm{E}\left(\sigma_B^2 X_{t-1}^2 + \sigma^2\right) = \sigma_B^2 \frac{\sigma^2}{1 - \beta - \sigma_B^2} + \sigma^2 = \sigma^2 \frac{1 - \beta}{1 - \beta - \sigma_B^2},$$

conditional variance of u_t quadratically depends on X_{t-1}. This us true as long as $\sigma_B^2 > 0$, in case of $\sigma_B^2 = 0$ both conditional and non-conditional variance of u_t reduces to σ^2. Taking into account this general conditional heteroscedasticity of errors, we can improve the estimator by minimizing

$$\sum_{t=1}^{n} \frac{(X_t - \mathrm{E}[X_t|\mathcal{F}_{t-1}])^2}{\mathrm{var}\left[X_t|\mathcal{F}_{t-1}\right]} = \sum_{t=1}^{n} \frac{u_t^2}{\mathrm{E}[u_t^2|\mathcal{F}_{t-1}]} = \sum_{t=1}^{n} \frac{(X_t - \beta X_{t-1})^2}{\sigma_B^2 X_{t-1}^2 + \sigma^2}.$$

We gain the weighted conditional least-squares estimator (WLS estimator)

$$\widehat{\beta}_{WLS} = \frac{\sum\limits_{t=1}^{n} \frac{X_t \cdot X_{t-1}}{\sigma^2 + \sigma_B^2 \cdot X_{t-1}^2}}{\sum\limits_{t=1}^{n} \frac{X_{t-1}^2}{\sigma^2 + \sigma_B^2 \cdot X_{t-1}^2}}. \tag{2.9}$$

However, to evaluate the latter formula we also need to estimate parameters σ^2 and σ_B^2. We usually use two-step estimation procedure: Calculate $\widehat{\beta}_{LS}$ first, find $\widehat{\sigma}^2$ and $\widehat{\sigma}_B^2$ according to equations (2.8) and then compute $\widehat{\beta}_{WLS}$ by (2.9), where the true values of unknown variances are replaced by their estimates $\widehat{\sigma}^2$ and $\widehat{\sigma}_B^2$, as the final estimate.

2.3 Maximum likelihood estimator

Other widely used method for parameter estimation is maximum likelihood procedure. It is based upon additional assumption that joint distribution of the errors and the random coefficients is normal but it works even if this assumption is violated (then it is called quasi-maximum likelihood method). Conditional likelihood function L_n is defined as the conditional density of vector $(X_1, \ldots, X_n)$ given $\mathcal{F}_0$. Its maximum is equivalent to the maximum of function

$$l_n(\beta, \sigma^2, \sigma_B^2) = -\frac{1}{2} \sum_{t=1}^{n} \left(\ln(\sigma^2 + \sigma_B^2 \cdot X_{t-1}^2) + \frac{(X_t - \beta X_{t-1})^2}{\sigma^2 + \sigma_B^2 \cdot X_{t-1}^2} \right). \tag{2.10}$$

Function (2.10) is non-linear in σ^2 and σ_B^2 and it is impossible to find its maximum explicitly. Usual approach to this optimalization task (see [25]) is to introduce auxiliary parameter $r = \frac{\sigma_B^2}{\sigma^2}$, rewrite (2.10) into

$$l_n(\beta, r, \sigma^2) = -\frac{1}{2} n \ln \sigma^2 - \frac{1}{2} \sum_{t=1}^{n} \ln(1 + rX_{t-1}^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{n} \frac{(X_t - \beta X_{t-1})^2}{1 + rX_{t-1}^2}, \tag{2.11}$$

differentiate it with respect to β and σ^2, respectively

$$\frac{\partial l_n}{\partial \beta}(\beta, r, \sigma^2) = \frac{1}{\sigma^2} \sum_{t=1}^{n} \frac{(X_t - \beta X_{t-1})X_{t-1}}{1 + rX_{t-1}^2},$$

$$\frac{\partial l_n}{\partial \sigma^2}(\beta, r, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^{n} \frac{(X_t - \beta X_{t-1})^2}{1 + rX_{t-1}^2}$$

and set the partial derivatives equal to zero to find the extreme. The results depend on parameter r and have form

$$\beta(r) = \frac{\sum_{t=1}^{n} \frac{X_t X_{t-1}}{1 + rX_{t-1}^2}}{\sum_{t=1}^{n} \frac{X_{t-1}^2}{1 + rX_{t-1}^2}}, \tag{2.12}$$

$$\sigma^2(r) = \frac{1}{n}\sum_{t=1}^{n}\frac{(X_t - \beta(r)X_{t-1})^2}{1 + rX_{t-1}^2}. \tag{2.13}$$

Finally, terms (2.12) and (2.13) are put back into equation (2.11), where the last term is omitted because it has no impact on the extreme, and expression

$$l_n(r) = -\frac{1}{2}n\ln\sigma^2(r) - \frac{1}{2}\sum_{t=1}^{n}\ln(1 + rX_{t-1}^2) \tag{2.14}$$

is maximized with respect to r. Optimal $\widehat{r}$ defines maximum likelihood estimators (2.12) and (2.13), estimator $\widehat{\sigma}_B^2$ equals $\widehat{r}\cdot\widehat{\sigma}^2(\widehat{r})$.

We could also differentiate directly function (2.10) with respect to β:

$$\frac{\partial l_n}{\partial\beta}(\beta, \sigma^2, \sigma_B^2) = \sum_{t=1}^{n}\frac{(X_t - \beta X_{t-1})X_{t-1}}{\sigma^2 + \sigma_B^2 X_{t-1}^2} \tag{2.15}.$$

Then solution of equation $\frac{\partial l_n}{\partial\beta}(\beta, \sigma^2, \sigma_B^2) = 0$ is exactly equal to the weighted least-squares estimator (2.9), so in our case of RCA(1) models the maximum likelihood estimator coincides with the weighted least-squares estimator. Notice, that if the random coefficient has zero variance ($\sigma_B^2 = 0$) then those estimators equal the ordinary least-squares estimator (2.7).

Figure 2.2 illustrates the estimation procedure using simulated RCA(1) process (500 observations) with true parameter $\beta = 0.3$, $\sigma_B^2 = 0.4$ and $\sigma^2 = 1$ in which stationarity assumption A6 is met ($\beta^2 + \sigma_B^2 = 0.49 < 1$). We successively computed expressions (2.12), (2.13) and (2.14) for values $r = 0, 0.05, 0.10, \ldots, 2$. Left panel of 2.2 shows curve $l_n(r)$, right panel displays $\beta(r)$. The colored dots highlight the values for specific choices of parameter $r = \frac{\sigma_B^2}{\sigma^2}$: Red color stands for the least-squares estimator ($r = 0$), green color for the theoretical weighted least-squares estimator ($r = 0.4$ because we know that $\sigma_B^2 = 0.4$ and $\sigma^2 = 1$) and blue color for choice $r = 1$ will play a special role in the following chapters. The maximum likelihood estimator of β is defined as the maximum of function $l_n(r)$, which is achieved for $r = 0.29$ with $l_n(0.29) = -103.92$ and corresponds to $\widehat{\beta}_{ML} = 0.275$ (brown lines in the figure). Finally, the green horizontal line in the right panel indicates the weighted least-squares estimator (2.9) provided that we pretend not to know true values of σ_B^2 and σ^2.

Firstly, least squares estimator $\widehat{\beta}_{LS} = 0.2568$ heavily underestimates true value $\beta = 0.3$. Secondly, notice that the maximum likelihood estimator ($\widehat{\beta}_{ML} = 0.2749$) is close to the weighted least-squares estimator (both theoretical $\widehat{\beta}_{WLS}^{TH} = 0.2783$ and "properly" estimated $\widehat{\beta}_{WLS} = 0.2779$). The reason, why the maximum likelihood estimator is not exactly equal to the weighted least-squares estimator as it should be according to the theory, is that the likelihood curve is plotted using estimated parameter $\sigma^2(r)$ instead of its theoretical value 1. Finally, although estimator $\beta(r)$ given by (2.12) with $r = 1$ equals 0.2908 and still underestimates the true value, it is the closest estimator of all mentioned. Since the curve in the right panel of Figure 2.2 increases as r increases, in this case we could obtain the best estimator if we set $r = 1.74$.

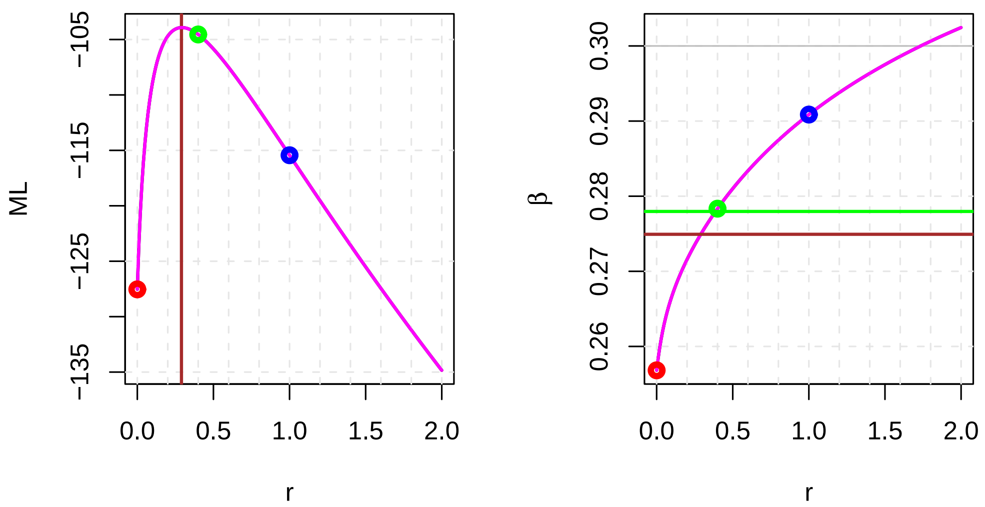

Figure 2.2: Left panel: Log-likelihood function $l_n(r)$ given by (2.14). Right panel: Estimator of parameter β given by (2.12). Based on simulated RCA(1) model with $\beta = 0.3$, $\sigma_B^2 = 0.4$ and $\sigma^2 = 1$.

2.4 Functional estimator

More general approach to RCA estimation was proposed by Schick in his paper [28]. He introduced a whole class of estimators that are asymptotically equivalent to the quasi-maximum likelihood estimator but require weaker moment conditions and can be computed more easily. Since we will explore his ideas further, let us describe the basic concept in more detail.

Consider the RCA(1) model according to Definition 2.1 and put $w(x) = \sigma^2 + \sigma_B^2 x^2$. Let $\phi(x)$ be a measurable function that satisfies $x\phi(x) > 0$ for every $x \neq 0$. Define

$$\widehat{\beta}(\phi) = \frac{\sum_{t=1}^{n} \phi(X_{t-1}) \cdot X_t}{\sum_{t=1}^{n} \phi(X_{t-1}) \cdot X_{t-1}} \tag{2.16}$$

and

$$V(\phi) = \frac{\mathrm{E}\left(\phi^2(X_0) w(X_0)\right)}{\left(\mathrm{E}(\phi(X_0) X_0)\right)^2}. \tag{2.17}$$

The motivation for the functional estimator might be the following: Consider the weighted least-squares estimator given by equation (2.9) which could be rewritten into form

$$\widehat{\beta}_{WLS} = \frac{\sum\limits_{t=1}^{n} \frac{X_{t-1}}{\sigma^2+\sigma_B^2 \cdot X_{t-1}^2} \cdot X_t}{\sum\limits_{t=1}^{n} \frac{X_{t-1}}{\sigma^2+\sigma_B^2 \cdot X_{t-1}^2} \cdot X_{t-1}}$$

and notice that it is exactly equal to equation (2.16) with $\phi(x) = \frac{x}{\sigma^2+\sigma_B^2 x^2} = \frac{x}{w(x)}$. Similarly, the ordinary least-squares estimator $\widehat{\beta}_{LS}$ given by equation (2.7) corresponds to the choice $\phi(x) = x$. Thus, one could allow also other functions ϕ to determine the contribution of the past values X_{t-1} to the estimator.

2.4.1 Choice of generating function

Schick proved in [28], Theorem 1, that for bounded functions ϕ the estimator $\widehat{\beta}(\phi)$ is strongly consistent estimator of parameter β and that it is asymptotically normal with zero mean and variance $V(\phi)$, namely

$$\sqrt{n}(\widehat{\beta}(\phi) - \beta) \xrightarrow{D} N\left(0, V(\phi)\right) \text{ as } n \to \infty.$$

He also proved that the smallest asymptotic variance $V(\phi)$ is achieved for the weighted least-squares estimator. This can be seen from equation (2.17) using Cauchy-Schwarz inequality and noticing that for function $\psi(x) = \frac{x}{w(x)}$ and any function ϕ

$$\begin{aligned} V(\psi) &= \frac{\mathrm{E}\left(\frac{X_0^2}{w^2(X_0)} w(X_0)\right)}{\left(\mathrm{E}\left(\frac{X_0}{w(X_0)} X_0\right)\right)^2} = \frac{1}{\mathrm{E}\left(\frac{X_0^2}{w(X_0)}\right)} = \\ &= \frac{1}{\mathrm{E}\left(\frac{X_0^2}{w(X_0)}\right)} \cdot \frac{\mathrm{E}\left(\phi^2(X_0) w(X_0)\right)}{\mathrm{E}\left(\phi^2(X_0) w(X_0)\right)} \leq \frac{\mathrm{E}\left(\phi^2(X_0) w(X_0)\right)}{\left(\mathrm{E}\left(\phi(X_0) X_0\right)\right)^2} = V(\phi). \end{aligned}$$

Since $V(\phi) = V(c \cdot \phi)$ for any non-zero constant c, the optimal estimator is determined uniquely as a consequence of the previous Cauchy-Schwarz inequality.

The same conclusion about the optimal choice of the generating function can be proved using similar techniques as Rao did in [27] for inference of the lower bound of variance for a general estimator. Since we will need such techniques for RCA(p) and multivariate RCA(1) models later on, let us describe it in detail and prove the previous result once again. The motivation for such approach stems from maximum likelihood function derived in Section 2.3. Consider equation (2.15), where the first derivative of log-likelihood function l_n with respect to parameter β is derived, and assume that parameters σ^2 and σ_B^2 are known. Now we can compute the Fisher information

$$I(\beta) = \mathrm{E}\left(\frac{\partial l_n}{\partial \beta}\right)^2 = \mathrm{E}\left(\sum_{t=1}^{n} \frac{X_{t-1} u_t}{\sigma^2 + \sigma_B^2 X_{t-1}^2}\right)^2 =$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} \mathrm{E}\left(\mathrm{E}\left[\left(\frac{X_{t-1}u_t}{\sigma^2+\sigma_B^2X_{t-1}^2}\right)^2 |\mathcal{F}_{t-1}\right]\right) = \\
&= \sum_{t=1}^{n} \mathrm{E}\left(\frac{X_{t-1}^2}{(\sigma^2+\sigma_B^2X_{t-1}^2)^2}\cdot \mathrm{E}\big[u_t^2|\mathcal{F}_{t-1}\big]\right) = \sum_{t=1}^{n} \mathrm{E}\left(\frac{X_{t-1}^2}{\sigma^2+\sigma_B^2X_{t-1}^2}\right) = \\
&= n\cdot \mathrm{E}\frac{X_0^2}{\sigma^2+\sigma_B^2X_0^2} = n\cdot\frac{1}{V(\psi)}
\end{aligned}
$$

using strict stationarity of process $\{X_t\}$ and the fact that sequence $\{\frac{X_{t-1}u_t}{\sigma^2+\sigma_B^2X_{t-1}^2}\}$ is martingale difference w.r. to $\mathcal{F}_t$ and thus it is a sequence of non-correlated variables. So the inverse of the Fisher information corresponds to the variance of the weighted least-squares estimator.

Now, let us define random variables

$$
\begin{aligned}
T_1 &= \sum_{t=1}^{n}\phi(X_{t-1})(X_t-\beta X_{t-1}) = \sum_{t=1}^{n}\phi(X_{t-1})u_t, \\
T_2 &= \sum_{t=1}^{n}\frac{X_{t-1}(X_t-\beta X_{t-1})}{\sigma_B^2X_{t-1}^2+\sigma^2} = \sum_{t=1}^{n}\frac{X_{t-1}u_t}{w(X_{t-1})}.
\end{aligned} \tag{2.18}
$$

We have

$$
\begin{aligned}
\mathrm{E}T_1 &= \sum_{t=1}^{n}\mathrm{E}\big(\phi(X_{t-1})u_t\big) = \sum_{t=1}^{n}\mathrm{E}\big(\mathrm{E}[\phi(X_{t-1})(B_tX_{t-1}+Y_t)|\mathcal{F}_{t-1}]\big) = \\
&= \sum_{t=1}^{n}\mathrm{E}\big(\phi(X_{t-1})X_{t-1}\cdot\mathrm{E}[B_t|\mathcal{F}_{t-1}]+\phi(X_{t-1})\cdot\mathrm{E}[Y_1|\mathcal{F}_{t-1}]\big) = 0.
\end{aligned}
$$

Since $\mathrm{E}[\phi(X_{t-1})u_t|\mathcal{F}_{t-1}] = \phi(X_{t-1})\cdot\mathrm{E}[u_t|\mathcal{F}_{t-1}] = 0$, sequence $\{\phi(X_{t-1})u_t\}$ is a non-correlated martingale difference which means that

$$
\begin{aligned}
\mathrm{E}T_1^2 &= \sum_{t=1}^{n}\mathrm{E}\big(\phi(X_{t-1})^2u_t^2\big) = \sum_{t=1}^{n}\mathrm{E}\big(\mathrm{E}\big[\phi(X_{t-1})^2u_t^2|\mathcal{F}_{t-1}\big]\big) = \\
&= \sum_{t=1}^{n}\mathrm{E}\big(\phi(X_{t-1})^2\cdot\mathrm{E}\big[u_t^2|\mathcal{F}_{t-1}\big]\big) = \sum_{t=1}^{n}\mathrm{E}\big(\phi(X_{t-1})^2\cdot(\sigma_B^2X_{t-1}^2+\sigma^2)\big) = \\
&= \sum_{t=1}^{n}\mathrm{E}\big(\phi(X_0)^2\cdot(\sigma_B^2X_0^2+\sigma^2)\big) = n\cdot\mathrm{E}\big(\phi(X_0)^2\cdot w(X_0)\big).
\end{aligned} \tag{2.19}
$$

Similarly, we could prove that $\mathrm{E}T_2 = 0$ and $\{\frac{X_{t-1}(X_t-\beta X_{t-1})}{\sigma_B^2X_{t-1}^2+\sigma^2}\}$ is a martingale difference sequence. Thus

$$
\mathrm{E}T_2^2 = \sum_{t=1}^{n}\mathrm{E}\left(\frac{X_{t-1}^2u_t^2}{w(X_{t-1})^2}\right) = \sum_{t=1}^{n}\mathrm{E}\left(\frac{X_{t-1}^2}{w(X_{t-1})^2}\cdot\mathrm{E}\big[u_t^2|\mathcal{F}_{t-1}\big]\right) =
$$

$$= \sum_{t=1}^{n} \mathrm{E}\left(\frac{X_{t-1}^2}{w(X_{t-1})^2} \cdot w(X_{t-1})\right) = n \cdot \mathrm{E}\left(\frac{X_0^2}{w(X_0)}\right). \tag{2.20}$$

Finally, covariance between variables T_1 and T_2 equals

$$\begin{aligned}
\mathrm{E}T_1T_2 &= \mathrm{E}\left(\sum_{t=1}^{n} \phi(X_{t-1})u_t \cdot \sum_{s=1}^{n} \frac{X_{s-1}u_s}{w(X_{s-1})}\right) = \\
&= \mathrm{E}\left(\sum_{t=1}^{n} \frac{\phi(X_{t-1})X_{t-1}u_t^2}{w(X_{t-1})} + \sum_{\substack{t=1 \\ t\neq s}}^{n}\sum_{s=1}^{n} \frac{\phi(X_{t-1})X_{s-1}u_tu_s}{w(X_{s-1})}\right) = \\
&= \sum_{t=1}^{n} \mathrm{E}\left(\frac{\phi(X_{t-1})X_{t-1}}{w(X_{t-1})} \cdot w(X_{t-1})\right) + \sum_{\substack{t=1 \\ t\neq s}}^{n}\sum_{s=1}^{n} \mathrm{E}\left(\frac{\phi(X_{t-1})X_{s-1}u_tu_s}{w(X_{s-1})}\right) = \\
&= n \cdot \mathrm{E}\big(\phi(X_0) \cdot X_0\big)
\end{aligned} \tag{2.21}$$

using the fact that for $t > s$ (and analogously for $t < s$)

$$\begin{aligned}
\mathrm{E}\left(\frac{\phi(X_{t-1})X_{s-1}u_tu_s}{w(X_{s-1})}\right) &= \mathrm{E}\left(\mathrm{E}\left[\frac{\phi(X_{t-1})X_{s-1}u_tu_s}{w(X_{s-1})}|\mathcal{F}_{t-1}\right]\right) = \\
&= \mathrm{E}\left(\frac{\phi(X_{t-1})X_{s-1}u_s}{w(X_{s-1})} \cdot \mathrm{E}[u_t|\mathcal{F}_{t-1}]\right) = 0.
\end{aligned}$$

Now we define random vector $(T_1, T_2)'$ whose variance matrix equals

$$\begin{pmatrix} \mathrm{E}T_1^2, & \mathrm{E}T_1T_2 \\ \mathrm{E}T_1T_2, & \mathrm{E}T_2^2 \end{pmatrix} = n \cdot \begin{pmatrix} \mathrm{E}\phi(X_0)^2w(X_0), & \mathrm{E}\phi(X_0)X_0 \\ \mathrm{E}\phi(X_0)X_0, & \mathrm{E}\frac{X_0^2}{w(X_0)} \end{pmatrix}, \tag{2.22}$$

where the elements of the matrix were computed in (2.19) - (2.21). Variance matrix (2.22) is positively semi-definite (or non-negatively definite) which means that its determinant has to be non-negative. Thus

$$\begin{aligned}
&\mathrm{E}\phi(X_0)^2w(X_0) \cdot \mathrm{E}\frac{X_0^2}{w(X_0)} - \big(\mathrm{E}\phi(X_0)X_0\big)^2 \geq 0 \\
\iff &\frac{\mathrm{E}\phi(X_0)^2w(X_0)}{\big(\mathrm{E}\phi(X_0)X_0\big)^2} \geq \frac{1}{\mathrm{E}\frac{X_0^2}{w(X_0)}} \\
\iff &V(\phi) \geq V(\psi)
\end{aligned}$$

which agrees to the previous proved result using Cauchy-Schwarz inequality.

The choice of function ϕ has tremendous impact on estimator $\widehat{\beta}(\phi)$. Generally, estimator $\widehat{\beta}(\phi)$ is invariant w.r. to multiplication of function $\phi(x)$ by non-zero constant c,

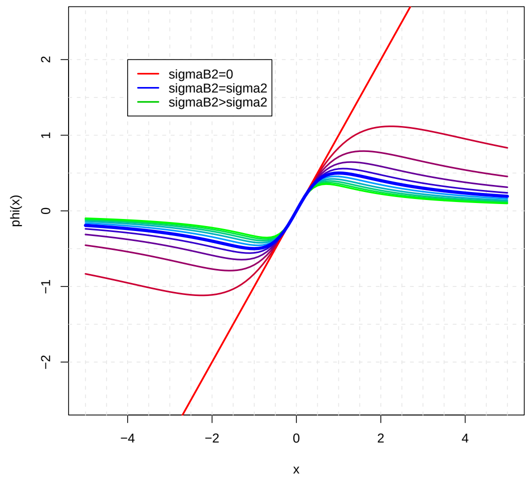

Figure 2.3: Functions $\phi(x) = \frac{x}{1+\frac{\sigma_B^2}{\sigma^2}x^2}$ for $\sigma^2 = 0.5$ and σ_B^2 varying from 0 to 1. Red straight line $\phi(x) = x$ corresponds to $\sigma_B^2 = 0$, blue curve to $\sigma^2 = \sigma_B^2 = 0.5$.

thus $\widehat{\beta}(\phi) = \widehat{\beta}(c \cdot \phi)$. We have already seen that $\phi(x) = x$ corresponds to the least-squares estimator (such function ϕ is not bounded but the statistical properties for the estimator could be proved separately, see for instance [25]). The optimal function ϕ, in the sense of the minimal asymptotical variance, leads to the weighted least-squares estimator that is also equal to maximum likelihood estimator in the normal case. Notice that the optimal estimator requires the knowledge of σ^2 and σ_B^2 (or their ratio at least). Schick solved this problem by introducing a class of consistent estimators of σ^2 and σ_B^2 depending on arbitrarily chosen bounded function χ. Then he proved (see Theorem 3 in [28]) that both strong consistency and asymptotic normality remain valid when we plug those estimates $\widehat{\sigma}^2$ and $\widehat{\sigma}_B^2$ into the optimal estimator.

Figure 2.3 displays functions $\phi(x) = \frac{x}{1+\frac{\sigma_B^2}{\sigma^2}x^2}$ with fixed $\sigma^2 = 0.5$ and various σ_B^2. Red straight line $\phi(x) = x$ corresponds to the choice of function ϕ for the least-squares estimator (2.7), blue curve corresponds to the choice $\sigma^2 = \sigma_B^2$. Generally, the larger parameter σ_B^2 is the more bounded function ϕ. This means that functional estimator (2.16) with such function ϕ down-weights larger observed values – they are "suspected" to be caused by more volatile random coefficients.

Generally, a simple stress-test of choice of function ϕ was performed and Figure 2.4 displays the results. We simulated RCA(1) process of 100 observations with parameters $\beta = 0.3$, $\sigma_B^2 = 0.4$ and $\sigma^2 = 1$. The true value of parameter β was then estimated using estimator $\widehat{\beta}(\phi)$ with $\phi(x) = \frac{x}{1+cx^2}$ for $c = 0, 0.05, 0.10, \ldots, 2$. We repeated such simulation and estimation procedure 1000 times, computed average values of estimator for each choice of constant c and plotted the results as a function of c, see Figure 2.4.

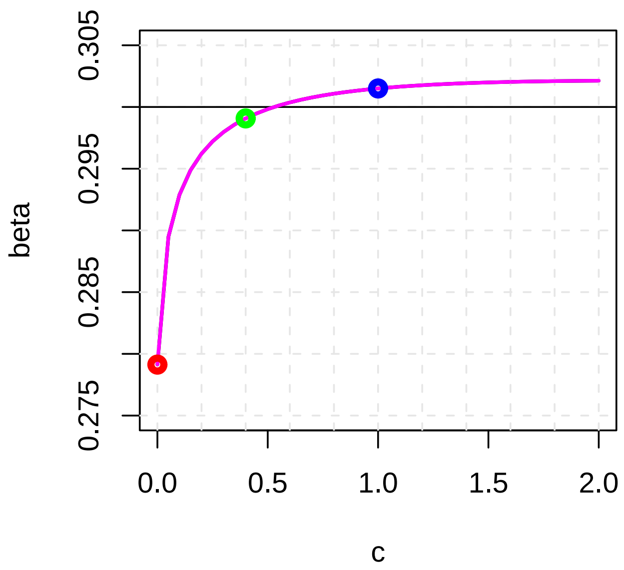

Figure 2.4: Average estimator $\widehat{\beta}(\phi)$ (true value $\beta = 0.3$) for various choices of function $\phi(x) = \frac{x}{1+cx^2}$ for $c = 0, 0.05, 0.10, \ldots, 2$. Red dot ($c = 0$) corresponds to the least-squares estimator, green dot ($c = 0.4$) to the weighted least-squares estimator and blue dot ($c = 1$) to functional estimator used later on.

We highlighted important choices of c there – red dot with $c = 0$ corresponds to the least-squares estimator, green dot with $c = 0.4$ corresponds to the weighted least-squares estimator and blue dot with $c = 1$ corresponds to the functional estimator that will be used later on mainly.

According to results of the previous simulation, we can conclude that the least-squares estimator underestimates true value of parameter (in average of about 7%) whereas the weighted least-squares estimator and the functional estimator with $c = 1$ are closer to the true value (relative differences are about -0.3% and $+0.5\%$ in average, respectively).

2.4.2 Consistency and asymptotical normality

In the previous paragraph we described a class of estimators of RCA(1) process where the estimators depend on chosen function ϕ. The mentioned results concerning consistency and asymptotical normality were proved by Schick in [28] under assumptions that both error process and coefficient process are mutually independent sequences of *iid* random variables. In this section we relax the assumption about independence of error process and replace it by martingale difference sequence property.

We managed to prove consistency and asymptotical normality for the functional estimator in our GRCA(1) setup.

Theorem 2.2.
Let $\{X_t, t \in \mathbb{Z}\}$ be a Generalized RCA(1) according to Definition 2.2.
Let $\mathrm{E}[Y_t^2|\mathcal{F}_{t-1}] = \sigma^2 < +\infty$ a.s. for all t and let $\phi(x)$ be a measurable function such that $0 < \mathrm{E}\left(\phi(X_0) \cdot X_0\right) < +\infty$.
Then the estimator $\widehat{\beta}(\phi)$ defined by (2.16) is strongly consistent estimator of β and $\sqrt{n}(\widehat{\beta}(\phi) - \beta)$ converges in distribution to normal distribution with zero mean and variance $V(\phi)$ defined by (2.17).

Proof: The key expression for derivation of the properties of the estimator is

$$\widehat{\beta}(\phi) - \beta = \frac{\frac{1}{n}\sum\limits_{t=1}^{n} \phi(X_{t-1}) \cdot (X_t - \beta X_{t-1})}{\frac{1}{n}\sum\limits_{t=1}^{n} \phi(X_{t-1}) \cdot X_{t-1}}. \tag{2.23}$$

According to Theorem 2.1, process $\{X_t\}$ is strictly stationary and ergodic, so both $\phi(X_{t-1})(X_t - \beta X_{t-1})$ and $\phi(X_{t-1})X_{t-1}$ are also strictly stationary and ergodic. The ergodic theorem (see e.g. [10], Theorem 13.12) gives us convergence of their sample averages to their expected values. Moreover,

$$\mathrm{E}[\phi(X_{t-1}) \cdot (X_t - \beta X_{t-1})|\mathcal{F}_{t-1}] = \phi(X_{t-1}) \cdot \mathrm{E}[Y_t|\mathcal{F}_{t-1}] = 0,$$

so $\phi(X_{t-1})\cdot(X_t - \beta X_{t-1})$ is a martingale difference sequence with respect to filtration $\{\mathcal{F}_t\}$ (thus, its expectation is zero). The denominator of (2.23) converges to non-zero and finite value $\mathrm{E}\left(\phi(X_0) \cdot X_0\right)$, and we have $\widehat{\beta}(\phi) - \beta \to 0$ a.s. as $n \to \infty$ that is equivalent to strong consistency of estimator $\widehat{\beta}(\phi)$.
Similarly, the Lindeberg-Levy theorem for martingales (see [7]) applied to martingale difference sequence $\phi(X_{t-1}) \cdot (X_t - \beta X_{t-1})$ with variance

$$\begin{aligned}
\operatorname{var}\left(\phi(X_{t-1})(X_t - \beta X_{t-1})\right) &= \mathrm{E}\left(\mathrm{E}\left[\phi^2(X_{t-1})(X_t - \beta X_{t-1})^2|\mathcal{F}_{t-1}\right]\right) = \\
&= \mathrm{E}\left(\phi^2(X_{t-1}) \cdot \mathrm{E}\left[(B_t X_{t-1} + Y_t)^2|\mathcal{F}_{t-1}\right]\right) = \\
&= \mathrm{E}\left(\phi^2(X_{t-1}) \cdot (X_{t-1}^2\sigma_B^2 + \sigma^2)\right) = \mathrm{E}\left(\phi^2(X_0) \cdot w(X_0)\right)
\end{aligned}$$

yields

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \phi(X_{t-1}) \cdot (X_t - \beta X_{t-1}) \xrightarrow{D} N\left(0, \mathrm{E}\left(\phi^2(X_0) \cdot w(X_0)\right)\right) \quad \text{as } n \to \infty. \tag{2.24}$$

Equation (2.23) immediately gives that

$$\sqrt{n}(\widehat{\beta}(\phi) - \beta) = \frac{\frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n} \phi(X_{t-1}) \cdot (X_t - \beta X_{t-1})}{\frac{1}{n}\sum\limits_{t=1}^{n} \phi(X_{t-1}) \cdot X_{t-1}},$$

where the numerator converges according to (2.24) and the denominator converges almost surely (and consequently in distribution) to $\mathrm{E}\phi(X_0)X_0$ from the ergodic theorem discussed earlier. Therefore, $\sqrt{n}(\widehat{\beta}(\phi) - \beta) \xrightarrow{D} N(0, V(\phi))$ for $n \to \infty$ where $V(\phi)$ is defined by (2.17). $\square$

Similarly to the extension of the least-squares estimator of parameter β to the functional estimator depending on chosen function $\phi(x)$, Schick also generalized the estimators of variances σ_B^2 and σ^2 into a broader class of the estimators depending on function $\psi(x)$. Namely, let $\psi(x)$ be a measurable function, denote $\overline{\psi} = \frac{1}{n}\sum_{t=1}^{n}\psi(X_{t-1})$, $\overline{X} = \frac{1}{n}\sum_{t=1}^{n}X_{t-1}^2$ and let $\gamma = \mathrm{cov}(\psi(X_0), X_0^2) \neq 0$. Then the estimators are defined as

$$\begin{aligned}
\widehat{\sigma}_B^2(\psi) &= \frac{\sum_{t=1}^{n}(\psi(X_{t-1}) - \overline{\psi})(X_t - \widehat{\beta}(\phi)X_{t-1})^2}{\sum_{t=1}^{n}(\psi(X_{t-1}) - \overline{\psi})X_{t-1}^2}, \\
\widehat{\sigma}^2(\psi) &= \frac{1}{n}\sum_{t=1}^{n}(X_t - \widehat{\beta}(\phi)X_{t-1})^2 - \widehat{\sigma}_B^2(\psi)\overline{X},
\end{aligned} \tag{2.25}$$

where $\widehat{\beta}(\phi)$ is an arbitrary estimator of parameter β. Since the proof of the consistency of such estimators given by Schick is based solely on the ergodicity of RCA(1) process, it is valid also for our Generalized RCA(1) process. More interesting is the connection of these functional estimators to estimators derived by (2.8) – estimators $\widehat{\sigma}_B^2$ and $\widehat{\sigma}^2$ stated there are the special cases of $\widehat{\sigma}_B^2(\psi)$ and $\widehat{\sigma}^2(\psi)$ for $\psi(x) = x^2$ and $\phi(x) = x$, which could be easily derived noticing that $\sum_{t=1}^{n}(X_{t-1}^2 - \overline{X})X_{t-1}^2 = \sum_{t=1}^{n}(X_{t-1}^2 - \overline{X})^2$.

Now we can state and prove a theorem about the consistent estimator of the asymptotical variance of estimator $\widehat{\beta}(\phi)$.

Theorem 2.3.
Let $\{X_t, t \in \mathbb{Z}\}$ be Generalized RCA(1) according to Definition 2.2. Let $\sigma^2 < +\infty$ and let $\phi(x)$ be measurable function such that $0 < \mathrm{E}\left(\phi(X_0) \cdot X_0\right) < +\infty$. Let $\widehat{\sigma}_B^2$ and $\widehat{\sigma}^2$ be strongly consistent estimators of σ_B^2 and σ^2, respectively.
Define

$$\widehat{V}(\phi) = n \cdot \frac{\sum\limits_{t=1}^{n}\phi^2(X_t)(\widehat{\sigma}^2 + \widehat{\sigma}_B^2X_t^2)}{\left(\sum\limits_{t=1}^{n}\phi(X_t)X_t\right)^2}. \tag{2.26}$$

Then estimator $\widehat{V}(\phi)$ is strongly consistent estimator of asymptotical variance $V(\phi)$ defined by (2.17).

Proof: Estimator $\widehat{V}(\phi)$ could be rewritten into the form

$$\widehat{V}(\phi) = \frac{\widehat{\sigma}^2 \cdot \frac{1}{n}\sum\limits_{t=1}^{n}\phi^2(X_t) + \widehat{\sigma}_B^2 \cdot \frac{1}{n}\sum\limits_{t=1}^{n}\phi^2(X_t)X_t^2}{\left(\frac{1}{n}\sum\limits_{t=1}^{n}\phi(X_t)X_t\right)^2}.$$

Similarly to the proof of Theorem 2.2, we know that $\phi^2(X_t)$, $\phi^2(X_t)X_t^2$ and $\phi(X_t)X_t$ are strictly stationary and ergodic which together with the consistency of $\widehat{\sigma}_B^2$ and $\widehat{\sigma}^2$ gives

$$\widehat{\sigma}^2 \cdot \frac{1}{n}\sum_{t=1}^{n} \phi^2(X_t) \to \sigma^2 \cdot \mathrm{E}\left(\phi^2(X_0)\right),$$
$$\widehat{\sigma}_B^2 \cdot \frac{1}{n}\sum_{t=1}^{n} \phi^2(X_t)X_t^2 \to \sigma_B^2 \cdot \mathrm{E}\left(\phi^2(X_0)X_0^2\right),$$
$$\frac{1}{n}\sum_{t=1}^{n} \phi(X_t)X_t \to \mathrm{E}\left(\phi(X_0)X_0\right).$$

The combination of the limiting values equals the variance matrix (2.17) which completes the proof. □

2.4.3 Simulation study

In the previous section we described some estimators of RCA and Generalized RCA models. Since those estimators differ only in the choice of generating function ϕ, they posses similar asymptotical properties. In the subsequent simulation, we will see that there can be substantial differences for final sample size among the estimators.

The aim of the study is to vary length of simulated RCA(1) time series from 50 to 500 observations and compare the least-squares estimator $\widehat{\beta}_{LS}$, the weighted least-squares estimator $\widehat{\beta}_{WLS}$ and the functional estimator $\widehat{\beta}(\phi)$ defined by (2.16) with $\phi(x) = \frac{x}{1+x^2}$. Moreover, we are curious to know whether the assumption about independence of errors Y_t is crucial.

Setup of the simulations is the following: We simulate a sequence of observations from RCA(1) model given by equation

$$X_t = (0.3 + B_t)X_{t-1} + Y_t, \tag{2.27}$$

where random coefficients $\{B_t, t \in \mathbb{Z}\}$ are independent $N(0, 0.4)$ random variables. Error process $\{Y_t, t \in \mathbb{Z}\}$ is a sequence of identically distributed variables with zero mean and unit variance independent of $\{B_t\}$. Notice that since true parameter $\beta = 0.3$ and $\sigma_B^2 = 0.4$, stationarity condition $\beta^2 + \sigma_B^2 < 1$ is fulfilled. Constellation $[\beta, \sigma_B^2] = [0.3, 0.4]$ has been chosen to illustrate the behavior of the estimators. We also performed more exhaustive simulation for general $[\beta, \sigma_B^2] \in \mathbb{R}_+^2$ such that $\beta^2 + \sigma_B^2 < 1$ which results will be described later.

We consider two cases:

1. $\{Y_t\}$ are independent and normally distributed random variables so X_t forms classical RCA(1) model.

2. $\{Y_t\}$ stands for ARCH(1) process of the form

$$Y_t = \sigma_t Z_t \quad a.s.,$$
$$\sigma_t^2 = 0.5 + 0.5 Y_{t-1}^2 \quad a.s.,$$

   where $\{Z_t, t \in \mathbb{Z}\}$ is a sequence of independent and identically $N(0,1)$ distributed variables. Thus, $\{X_t\}$ is Generalized RCA(1) model according to Definition 2.2 because ARCH process is obviously a martingale difference sequence of no longer independent random variables. The particular choice of ARCH process does not seem to be crucial because we achieved similar results using other ARCH processes.

Simulation procedure:

- set sample size $n \in \{50, 100, 150, \ldots, 500\}$
- do for each choice n of sample size:
  - repeat 1000 times
    1. simulate RCA(1) time series according to (2.27) (n observations)
    2. estimate true parameter $\beta = 0.3$ using $\widehat{\beta}_{LS}$, $\widehat{\beta}_{WLS}$ and $\widehat{\beta}(\phi)$ with $\phi(x) = \frac{x}{1+x^2}$, respectively and estimate asymptotical variance $V(\phi)$ for each estimator
  - for each estimator compute $\overline{\widehat{\beta}} = \frac{1}{1000}\sum_{i=1}^{1000}\widehat{\beta}^i$, $MSE = \frac{1}{1000}\sum_{i=1}^{1000}\left(\widehat{\beta}^i - \overline{\widehat{\beta}}\right)^2$ and $\overline{\widehat{V}}(\phi) = \frac{1}{1000}\sum_{i=1}^{1000}\widehat{V}^i(\phi)$
- for both types of errors Y_t plot $\overline{\widehat{\beta}}$, MSE and $\overline{\widehat{V}}(\phi)$ for each estimator as function of sample size n.

The weighted least-squares estimator requires the knowledge of variances σ^2 and σ_B^2 (see expression (2.9)). We estimate those parameters as we have discussed previously using linear regression estimates (2.8). Figures 2.5 to 2.7 display achieved results.

In brief, the weighted least-squares estimator and the functional estimator behave similarly whereas the ordinary least-squares estimator tends to underestimate the true value of parameter, especially in case of ARCH errors.

More precisely, the functional estimator seems to perform even little bit better than WLS estimator in the sense that average estimate of parameter β is usually closer to its true value 0.3. Variance of the functional estimator also decreases to zero most quickly, its asymptotical variance is stable. Figure 2.7 also reveals known fact that the smallest asymptotical variance is theoretically achieved by the weighted least-squares estimator. The worst results are achieved by LS estimator in Generalized RCA(1) model with ARCH errors. It not only underestimates the true value of parameter about 10% for small sample sizes but also its MSE is roughly twice as large as for the other estimators and converges very slowly. The estimate of asymptotical variance is not stable and even increases as sample size rises.

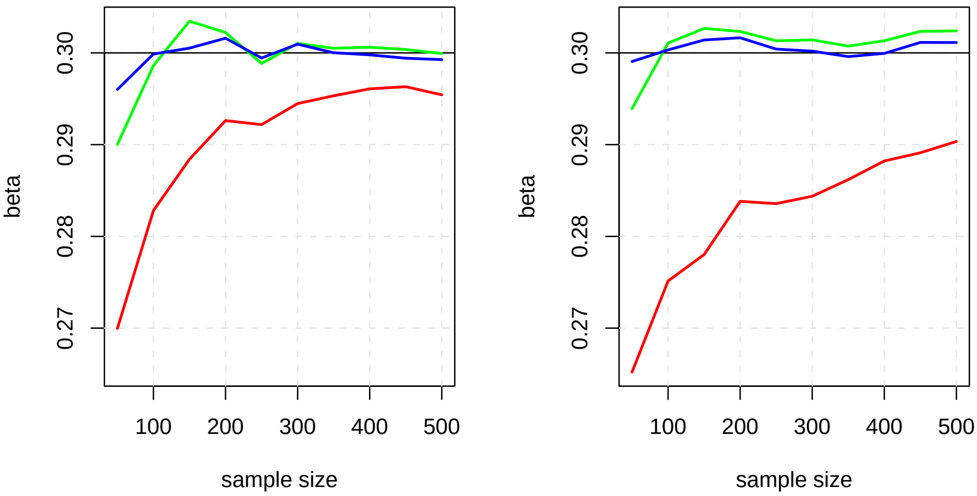

Figure 2.5: Average estimate $\overline{\widehat{\beta}}_{LS}$ (red curve), $\overline{\widehat{\beta}}_{WLS}$ (green curve) and $\overline{\widehat{\beta}}(\phi)$ with $\phi(x) = \frac{x}{1+x^2}$ (blue curve) for parameter β (true value 0.3) in simulated Generalized RCA(1) process with *iid* error process Y_t (left panel) and ARCH errors (right panel).

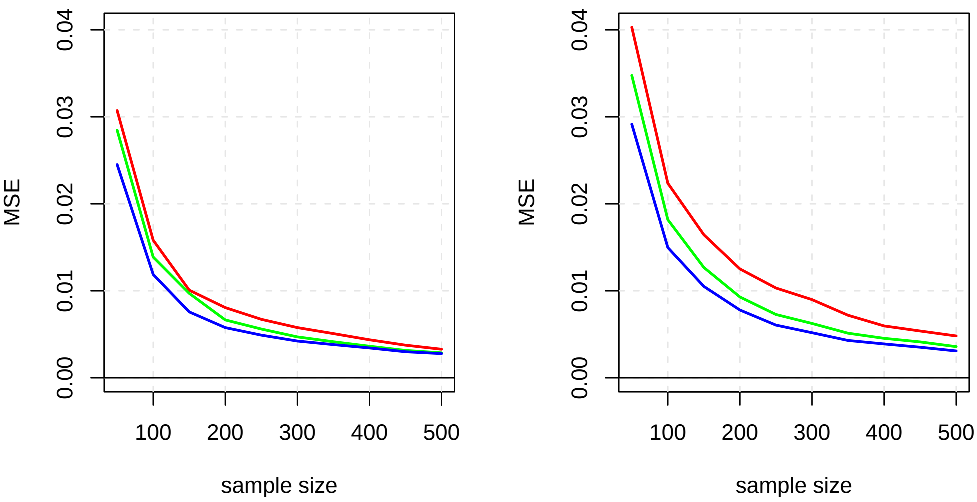

Figure 2.6: MSE of estimator $\widehat{\beta}_{LS}$ (red curve), $\widehat{\beta}_{WLS}$ (green curve) and $\widehat{\beta}(\phi)$ with $\phi(x) = \frac{x}{1+x^2}$ (blue curve) in simulated Generalized RCA(1) process with *iid* error process Y_t (left panel) and ARCH errors (right panel).

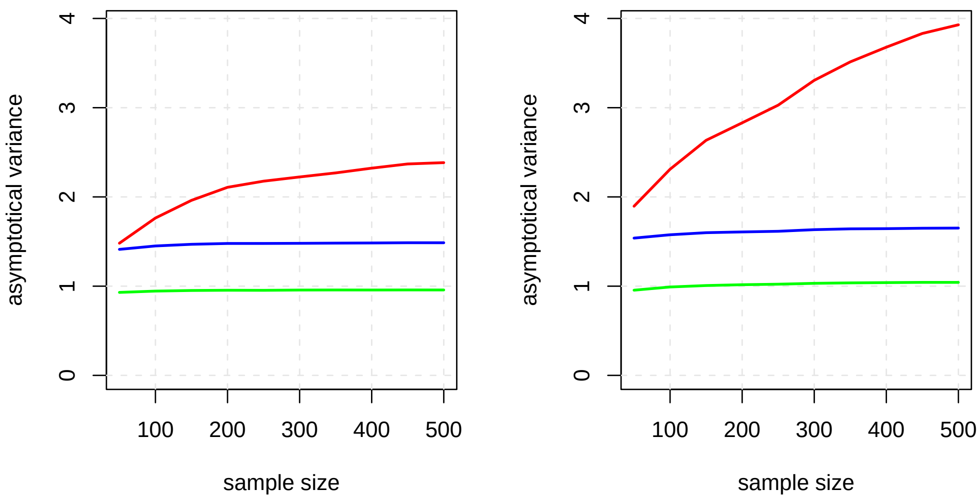

Figure 2.7: Average asymptotical variance of estimator $\widehat{\beta}_{LS}$ (red curve), $\widehat{\beta}_{WLS}$ (green curve) and $\widehat{\beta}(\phi)$ with $\phi(x) = \frac{x}{1+x^2}$ (blue curve) in simulated Generalized RCA(1) process with *iid* error process Y_t (left panel) and ARCH errors (right panel).

The difference between the weighted least-squares estimator and the functional estimator lies in estimation of auxiliary parameters σ^2 and σ_B^2 – WLS needs them to be estimated formerly while the function estimator does not. We can observe that in practice there is no substantive difference it the estimate, even the functional estimator performs better without their knowledge.

This simulation study might be compared to the simulations we published earlier (see [34]). The conclusions are similar to the previous ones. The possibly larger smoothness of current curves is caused by larger number of simulations performed (1000 instead of 100) and by slightly different approach in generation of the time series due to computational requirements. Technically, here we generated sequences of the maximum length 500 observations and truncated them for smaller sample sizes whereas previously we generated each sequence independently.

The previous results emphasized the effect of the increasing sample size. Let us add another factor into the simulations, namely freedom in $[\beta, \sigma_B^2]$. Specifically, we varied both β and σ_B^2 in $0, 0.1, 0.2, \ldots, 0.9$ (maintaining that $\beta^2 + \sigma_B^2 < 1$) and for each choice of $[\beta, \sigma_B^2]$ accomplished the previous simulation procedure. The results are displayed on Figures 2.8 and 2.9, precise numbers are included in Table 2.1.

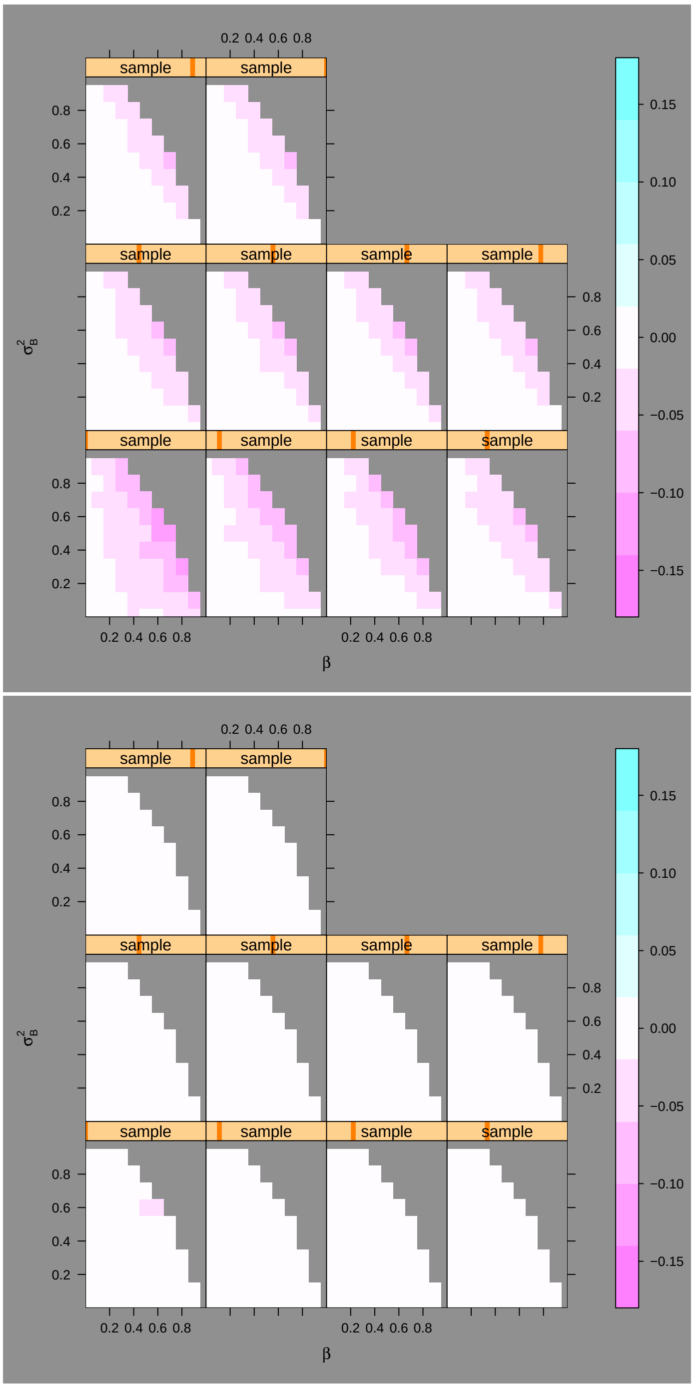

Figure 2.8: Average differences $\overline{\widehat{\beta}}_{LS} - \beta$ and $\overline{\widehat{\beta}}(\phi) - \beta$ with $\phi(x) = \frac{x}{1+x^2}$ (upper figure and lower figure, respectively) for various parameters β and σ_B^2 (horizontal and vertical axis in each panel). Sample size varies across the panels from 50 (left bottom panel) to 500 (right upper panel).

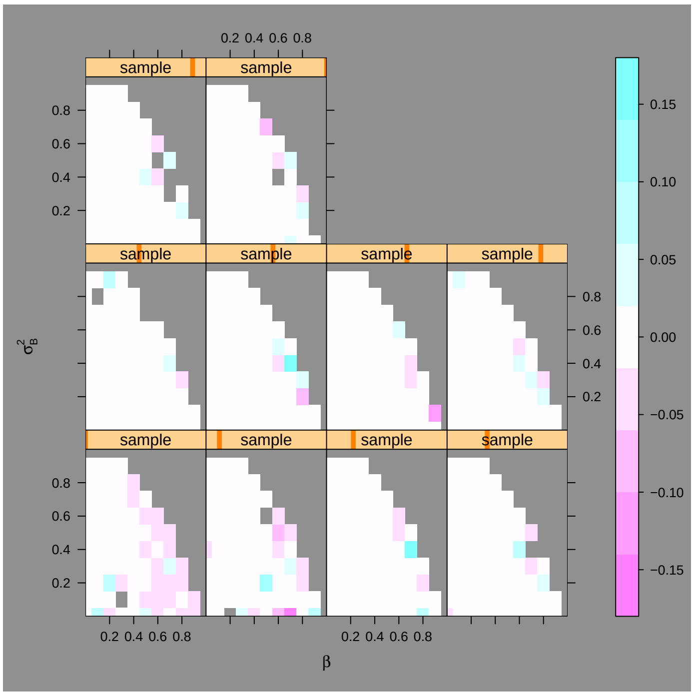

Figure 2.9: Average difference $\overline{\widehat{\beta}}_{WLS} - \beta$ for various parameters β and σ_B^2 (horizontal and vertical axis in each panel). Sample size varies across the panels from 50 (left bottom panel) to 500 (right upper panel).

Figure 2.8 consists of two pictures – upper figure shows the results of the simulation for the least-squares estimator $\overline{\widehat{\beta}}_{LS}$, lower figure shows the same for the functional estimator $\widehat{\beta}(\phi)$ with $\phi(x) = \frac{x}{1+x^2}$. The results for the weighted least-squares estimator are separately displayed on Figure 2.9. Each figure consists of 10 panels each of which stands for the particular sample size of the simulations (ordered by rows from bottom to top and from left right, sample sizes are $n = 50, 100, 150, \ldots, 500$). Each panel displays the average difference of the particular estimator $\overline{\widehat{\beta}} - \beta$ as a function of β and σ_B^2 (the numbers which the plots are based on are reported in Table 2.1). In other words, the color of each small square defined by coordinates $[\beta, \sigma_B^2]$ visualizes the result of 1000 simulations of RCA(1) model with given the parameters β and σ_B^2. If we traced the small square $[\beta, \sigma_B^2] = [0.3, 0.4]$ across the panels, we would obtain the line plot 2.5 for instance.

General rule how to read these figures says that the more pink color means the larger under fit of the estimator. We can see that the least-squares estimator underfits the models especially for smaller sample sizes which has already been revealed by the previous simulation. In addition, these figures also illustrate the effect of the magnitude of the random coefficients and their variance. The pink color is more concentrated to the boundary $\beta + \sigma_B^2 = 1$ which width is much broader for small sample sizes. On the other hand, low values of the true parameters β and σ_B^2 lead to significantly better estimators.

	β									
σ_B^2	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
	-0.017	-0.029	-0.044	-0.068						
0.9	-0.002	-0.005	-0.007	-0.006						
	-0.010	-0.014	-0.014	-0.017						
	-0.008	-0.007	-0.029	-0.067	-0.087					
0.8	-0.006	0.006	-0.003	-0.018	-0.015					
	-0.009	0.002	0.006	-0.019	-0.028					
	-0.001	-0.022	-0.036	-0.050	-0.088	-0.100				
0.7	-0.001	0.001	-0.004	-0.003	-0.018	-0.017				
	0.000	-0.005	-0.008	-0.008	-0.031	-0.015				
	-0.010	-0.019	-0.031	-0.043	-0.059	-0.085	-0.114			
0.6	-0.008	-0.009	-0.012	-0.003	0.002	-0.021	-0.020			
	-0.010	-0.016	-0.013	-0.011	-0.008	-0.026	-0.029			
	-0.005	-0.005	-0.023	-0.035	-0.054	-0.060	-0.101	-0.128		
0.5	-0.014	0.002	-0.008	-0.008	-0.013	-0.007	-0.019	-0.016		
	-0.008	0.002	-0.015	-0.010	-0.010	-0.015	-0.052	-0.040		
	-0.005	-0.005	-0.025	-0.023	-0.053	-0.064	-0.077	-0.091		
0.4	-0.007	-0.001	-0.018	-0.005	-0.015	-0.019	-0.004	-0.005		
	-0.005	0.000	0.013	-0.008	-0.008	-0.025	-0.009	-0.024		
	-0.003	-0.002	-0.020	-0.024	-0.033	-0.046	-0.059	-0.074	-0.104	
0.3	0.000	0.005	-0.007	-0.003	-0.010	-0.010	-0.012	-0.008	-0.016	
	0.001	-0.002	-0.008	-0.005	-0.011	-0.016	-0.024	0.057	-0.031	
	-0.007	-0.012	-0.015	-0.021	-0.023	-0.035	-0.043	-0.061	-0.075	
0.2	-0.003	-0.012	-0.005	-0.006	-0.008	-0.012	-0.014	-0.013	-0.009	
	-0.009	-0.009	0.066	-0.023	-0.014	-0.015	-0.021	-0.040	-0.038	
	0.000	-0.012	-0.014	-0.012	-0.024	-0.028	-0.037	-0.035	-0.047	-0.063
0.1	0.002	-0.004	-0.011	-0.006	-0.013	-0.012	-0.015	-0.007	-0.010	-0.017
	0.011	0.002	-0.011	0.197	-0.017	-0.025	-0.029	-0.023	-0.018	-0.038
	0.000	0.000	-0.011	-0.011	-0.021	-0.019	-0.014	-0.030	-0.029	-0.035
0.0	0.003	0.003	-0.009	-0.008	-0.011	-0.011	-0.003	-0.015	-0.011	-0.016
	0.003	0.071	-0.035	-0.013	-0.011	0.031	-0.035	-0.014	-0.039	-0.044

Table 2.1: Average differences between the true and estimated value of estimated parameter ($\overline{\widehat{\beta}} - \beta$) for various choices of parameters β and σ_B^2. Sample size of the simulation equals 50 observations. Presented estimators are LS (top rows), functional estimator (middle rows), WLS (bottom rows).

The functional estimator did better job than the least-squares estimator and it depends neither on the sample size nor on magnitudes of the true parameters. The weighted least-squares estimator behaves similarly to the functional one. However, there were certain constellations of the parameters when this estimator was not stable. We also performed the same simulations for a few different variances σ^2 and for various choices of the ARCH process as the noise process. Since they corresponded to the previous results, we does not present them here.

The conclusions from this simulations might be that the least-squares estimator is a reasonable and computationally simple estimator when there is low randomness in the coefficients and if there is no dependency structure in the errors of RCA(1) process. If there is a dependence present, in our case of the form of ARCH process, in tends to underestimate the parameter whereas the weighted least-squares estimator or the functional estimator behaves nicely. Also the least-squares estimator requires larger sample sizes not to underfit the true value of the parameter.

2.4.4 Rate of convergence

We proved consistency and asymptotic normality of the estimator of type (2.16) in Generalized RCA(1) model specified in Definition 2.2. We also performed a simulation study which revealed that the functional estimators seem to possess better statistical properties than the conventional estimators. The aim of this section is to derive the rate of convergence to normal distribution of such estimators for sufficiently large class of functions.

Special case $\phi(x) = x$ has been studied for instance by Basu and Roy in [5] or [6]. In the first mentioned article, the authors established the rate of convergence of LS estimator in both univariate and multivariate RCA models. In the second paper, they proved similar results but they used general autoregressive model with fixed coefficients instead of RCA process. Similar work was done in [8] for the first-order autoregressive process.

Unfortunately, such approach cannot be used for a general function $\phi(x)$, so we have to define class of all admissible functions $\phi(x)$. Denote $h(x) = x\phi(x)$ and presume that $h(x) > 0$ for every $x \neq 0$ and that $h(x)$ fulfills Lipschitz condition for given RCA process X_t:

Definition 2.3. *Lipschitz function*
Let $\{X_t, t \in \mathbb{Z}\}$ be real random process and $h(x)$ be real measurable function.
Then $h(x)$ fulfills the Lipschitz condition for given process X_t if

$$|h(X_s) - h(X_t)| \leq c_h |X_s - X_t| \quad \text{a.s.} \tag{2.28}$$

for all $s, t \in \mathbb{Z}$ and some constant $c_h > 0$.

Remark:

1. If process $\{X_t\}$ is uniformly bounded by some positive constant c_X then function $h(x) = x^2$ satisfies Definition 2.3. This can be seen by noticing that $|X_s^2 - X_t^2| = |X_s + X_t||X_s - X_t| \leq (|X_s| + |X_t|)|X_s - X_t| \leq 2c_X|X_s - X_t|$.

2. Generally, if process $\{X_t\}$ in Definition 2.3 satisfies $\mathrm{P}(X_t \in I, \forall t \in \mathbb{Z}) = 1$ for some bounded interval $I \in \mathbb{R}$, then any function $h(x)$ with bounded first derivation on I suits.
   This follows from the Lagrange Mean Value Theorem that states that for such function $h(x)$ and for each $x, y \in I$ there exists $z \in (x, y)$ such that $h(y) - h(x) = h'(z)(y-x)$. So $|h(X_s)-h(X_t)| = |h'(Z)|\cdot|(X_s-X_t)| \leq \max_{z\in I}|h'(z)|\cdot|(X_s-X_t)| = c_h|X_s - X_t|$ a.s..
   Thus, function $h(x) = \frac{x^2}{1+x^2}$, which corresponds to $\phi(x) = \frac{x}{1+x^2}$, fits for instance.

We employ similar techniques as Basu and Roy did in [5] where they explored the rate of convergence of LS estimator in RCA models. To prove the main result of this section we need a few auxiliary lemmas. The commonly known ones can be found in Chapter 6 consisting of these auxiliaries exclusively (Lemmas 6.2–6.5), the more specific propositions will be proved in this section.

It also turned out that we need a Hoeffding-type exponential inequality for a function of RCA(1) process $\{X_t\}$. Specifically we have to ensure that

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1}\big(h(X_t) - \mathrm{E}h(X_t)\big)\right| > \varepsilon\right) \leq c \cdot \mathrm{e}^{-dn\varepsilon^2}, \tag{2.29}$$

where $h(x)$ is given measurable function and c, d are some constants. This inequality is well known when $h(x) = x$ and either $\{X_t\}$ are independent or they form a martingale difference sequence. For general function $h(x)$ there were similar inequalities proved for uniformly ergodic Markov chains $\{X_t\}$ (see for instance [12]) or under some assumptions for ergodic time series (working paper [30]). Unfortunately, none of these generalizations of the Hoeffding inequality was applicable in our case so we proved inequality (2.29) using similar techniques as the authors of [17] did for heteroscedastic RCA(1) model.

For that purpose, we employed mixing concept (for further details see [10]):

Definition 2.4. ***Strong mixing***
Let $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ be a sequence of random vectors. Let $\mathcal{F}_{-\infty}^t = \sigma(\boldsymbol{V}_s, s \leq t)$, and $\mathcal{F}_{t+m}^{+\infty} = \sigma(\boldsymbol{V}_s, s \geq t+m)$ for each $t, m \in \mathbb{Z}$.
The sequence $\{\boldsymbol{V}_t\}$ is said to be α-mixing (or strong mixing) if $\lim_{m\to\infty}\alpha_m = 0$ where

$$\alpha_m = \sup_{t\in\mathbb{Z}}\left(\sup_{G\in\mathcal{F}_{-\infty}^t, H\in\mathcal{F}_{t+m}^{+\infty}}\big|\mathrm{P}(G\cap H) - \mathrm{P}(G)\mathrm{P}(H)\big|\right).$$

Precisely, a sequence is called α-mixing of size $-a_0$ if $\alpha_m = O(m^{-a})$ for some $a > a_0$.

Definition 2.5. ***Mixingale***
Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of integrable random variables. Let $\mathcal{F}_t = \sigma(X_s, s \leq t)$ for each $t \in \mathbb{Z}$.

The sequence $\{X_t, \mathcal{F}_t\}$ is called an L_p-mixingale if, for $p \geq 1$, there exist sequences of non-negative constants $\{c_t, t \in \mathbb{Z}\}$ and $\{\zeta_m, m \in \mathbb{N}_0\}$ such that $\zeta_m \to 0$ as $m \to +\infty$, and

$$\|\mathrm{E}[X_t|\mathcal{F}_{t-m}]\|_p \leq c_t\zeta_m,$$
$$\|X_t - \mathrm{E}[X_t|\mathcal{F}_{t+m}]\|_p \leq c_t\zeta_{m+1}$$

hold for all $t \in \mathbb{Z}$ and $m \in \mathbb{N}_0$.
The L_p-mixingale is of size $-a_0$ if $\zeta_m = O(m^{-a})$ for some $a > a_0 > 0$.

Definition 2.6. ***Near-epoch dependency***
Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of integrable random variables. Let $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ be a sequence, possibly vector-valued, let us define for each $t \in \mathbb{Z}$ the filtration $\{\mathcal{F}_{t-m}^{t+m}, m \in \mathbb{N}_0\}$ such that $\mathcal{F}_{t-m}^{t+m} = \sigma(\boldsymbol{V}_{t-m}, \ldots, \boldsymbol{V}_{t+m})$.
The sequence $\{X_t\}$ is said to be near-epoch dependent in L_p-norm (L_p-NED) on $\{\boldsymbol{V}_t\}$ if, for $p > 0$, there exist sequences of non-negative constants $\{d_t, t \in \mathbb{Z}\}$ and $\{\nu_m, m \in \mathbb{N}_0\}$ such that $\nu_m \to 0$ as $m \to +\infty$, and

$$\left\|X_t - \mathrm{E}\big[X_t|\mathcal{F}_{t-m}^{t+m}\big]\right\|_p \leq d_t\nu_m$$

hold for all $t \in \mathbb{Z}$ and $m \in \mathbb{N}_0$.
The L_p-NED is of size $-a_0$ if $\nu_m = O(m^{-a})$ for some $a > a_0 > 0$.

To prove the main theorem, we have to make an assumption about mutual structure between the random coefficients and the error process.

A7: $\{(B_t, Y_t), t \in \mathbb{Z}\}$ is α-mixing of size $-a$ for some $a > 2$.

The idea behind using just introduced concepts for proving exponential inequality (2.29) is briefly the following: A Lipschitz function maintains L_p-NED property, L_p-NED process is L_p-mixingale under certain assumptions, and a sum of L_p-mixingale items can be divided into a sum of martingale difference items and some residuum.

Theorem 2.4.
Let $\{X_t, \mathcal{F}_t\}$ be a uniformly bounded stationary L_1-mixingale of size -1 with bounded constant sequence $\{c_t, t \in \mathbb{Z}\}$ from Definition 2.5.
Then there exists decomposition

$$X_t = R_t - R_{t+1} + W_t, \tag{2.30}$$

where $\mathrm{E}|R_t| < +\infty$ and $\{W_t, \mathcal{F}_t\}$ is a stationary martingale difference sequence. Moreover, both sequences $\{R_t\}$ and $\{W_t\}$ are also uniformly bounded. Equation (2.30) immediately implies that

$$\sum_{t=0}^{n-1} X_t = R_0 - R_n + \sum_{t=0}^{n-1} W_t. \tag{2.31}$$

Proof: This theorem extends Theorem 16.6 from [10] and we will make use of the proof stated there. Decompositions (2.30) and (2.31) are established in the cited theorem, what remains is to prove the boundedness. Let us for each $m \in \mathbb{N}$ and $t \in \mathbb{Z}$ define a random variable

$$R_{m,t} = \sum_{s=0}^{m} (\mathrm{E}[X_{t+s}|\mathcal{F}_{t-1}] - X_{t-s-1} + \mathrm{E}[X_{t-s-1}|\mathcal{F}_{t-1}]).$$

Using triangular inequality and L_1-mixingale property we have that

$$|R_{m,t}| \leq \sum_{s=0}^{m} \left(\left|\mathrm{E}[X_{t+s}|\mathcal{F}_{t-1}]\right| + \left|X_{t-s-1} - \mathrm{E}[X_{t-s-1}|\mathcal{F}_{t-1}]\right|\right) \leq$$
$$\leq \sum_{s=0}^{m} (c_{t+s} \cdot \zeta_{s+1} + c_{t-s-1} \cdot \zeta_{s+1}) \leq 2c_1 \cdot \sum_{s=0}^{m} \zeta_{s+1} \leq 2c_1 \cdot \sum_{s=0}^{m} c_2 \frac{1}{(s+1)^a} \leq d,$$

where $c_t \leq c_1$ and $\zeta_{s+1} \leq c_2 \frac{1}{(s+1)^a}$ for some $c_1, c_2 > 0$. Constant $d > 0$ exists because of $a > 1$. Thus, $\{R_{m,t}\}$ are bounded uniformly in both m and t. In the proof mentioned above there is shown that for each $t \in \mathbb{Z}$, $R_{m,t}$ converges to R_t a.s. as $m \to +\infty$. Thus, R_t are uniformly bounded in t because of the boundedness of $R_{m,t}$.
Triangular inequality applied to rearranged equation (2.30) yields $|W_t| = |X_t - R_t + R_{t+1}| \leq c_X + 2d$, which ensures uniform boundedness of W_t. $\square$

We have to assume some additional conditions concerning the Generalized RCA(1) process $\{X_t\}$ to employ the previous techniques.

A8: $\mathrm{P}(|X_0| \leq c_X) = 1$ for some constant $c_X > 0$.

A9: $\mathrm{P}(|B_0| \leq c_B) = 1$ for some constant $c_B > 0$.

Since the Generalized RCA(1) process $\{X_t\}$ is strictly stationary, the boundedness of X_0 according to A8 ensures uniform boundedness of the whole sequence $\{X_t\}$, i.e. there exists constant $c_X > 0$ such that $\mathrm{P}(|X_t| \leq c_X, \forall t \in \mathbb{Z}) = 1$.

Theorem 2.5.
Under Assumptions A8 and A9 both $\{B_t\}$ and $\{Y_t\}$ in the definition of the Generalized RCA(1) process are uniformly bounded.

Proof:
Firstly, sequence $\{B_t\}$ is uniformly bounded according to Assumptions A3 and A9 using its strictly stationarity property.
Secondly, equation (2.1) gives us $|X_t| = |(\beta + B_t)X_{t-1} + Y_t| \geq |B_t X_{t-1} + Y_t| - |\beta X_{t-1}|$, consequently $|B_t X_{t-1} + Y_t| \leq |X_t| + |\beta X_{t-1}| \leq (1 + |\beta|)c_X$ a.s. due to boundedness of $\{X_t\}$, so we know that $\{B_t X_{t-1} + Y_t\}$ is uniformly bounded.
Finally, $|B_t X_{t-1} + Y_t| \geq |Y_t| - |B_t||X_{t-1}|$, thus $|Y_t| \leq |B_t X_{t-1} + Y_t| + |B_t||X_{t-1}| \leq c$ for some constant $c > 0$ and all $t \in \mathbb{Z}$ because all processes on the right hand side of the inequality are uniformly bounded. $\square$

Theorem 2.6.
Let $\{X_t, t \in \mathbb{Z}\}$ be a Generalized RCA(1) process that satisfies Assumptions A7, A8 and A9. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuous measurable function such that $h(x) = x\phi(x)$ satisfies Definition 2.3. Denote $Z_t = h(X_t) - \mathrm{E}h(X_t)$ for each $t \in \mathbb{Z}$.
Then there exists the same decomposition of process $\{Z_t, t \in \mathbb{Z}\}$ as in Theorem 2.4 for $\mathcal{F}_t = \sigma(B_s, Y_s; s \le t)$ where both sequences $\{R_t\}$ and $\{W_t\}$ are uniformly bounded.

Proof:
Firstly, we will show that $\{X_t\}$ is L_2-NED on $\{B_t, Y_t\}$ of arbitrary size. Let us denote $\mathcal{F}_{t-m}^{t+m} = \sigma(B_s, Y_s; s = t-m, \dots, t+m)$ and verify Definition 2.6:
Definition of RCA(1) model and $\mathcal{F}_{t-m}^{t+m}$-measurability of B_s and Y_s for $s = t, t-1, \dots, t-m$ yields

$$
\begin{aligned}
&\big\|X_t - \mathrm{E}\big[X_t|\mathcal{F}_{t-m}^{t+m}\big]\big\|_2 = \big\|(\beta + B_t)X_{t-1} + Y_t - \mathrm{E}\big[(\beta + B_t)X_{t-1} + Y_t|\mathcal{F}_{t-m}^{t+m}\big]\big\|_2 = \\
&= \big\|(\beta + B_t)X_{t-1} - (\beta + B_t)\mathrm{E}\big[X_{t-1}|\mathcal{F}_{t-m}^{t+m}\big]\big\|_2 = \\
&= \big\|(\beta + B_t)\left(X_{t-1} - \mathrm{E}\big[X_{t-1}|\mathcal{F}_{t-m}^{t+m}\big]\right)\big\|_2 = \cdots = \\
&= \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) \cdot \left(X_{t-m-1} - \mathrm{E}\big[X_{t-m-1}|\mathcal{F}_{t-m}^{t+m}\big]\right)\right\|_2 \le \\
&\le \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) X_{t-m-1}\right\|_2 + \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) \mathrm{E}\big[X_{t-m-1}|\mathcal{F}_{t-m}^{t+m}\big]\right\|_2 \le \\
&\le \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) c_X\right\|_2 + \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) c_X\right\|_2 .
\end{aligned}
$$

The last inequality holds due to the uniform boundedness of sequence $\{X_t\}$ by a positive constant c_X. Notice that

$$
\left\|\prod_{i=0}^{m}(\beta + B_{t-i})\right\|_2 = \left(\mathrm{E}\prod_{i=0}^{m}(\beta + B_{t-i})^2\right)^{\frac{1}{2}} = \left(\mathrm{E}(\beta + B_t)^2\right)^{\frac{m+1}{2}} = \left(\beta^2 + \sigma_B^2\right)^{\frac{m+1}{2}} .
$$

So we have

$$
\big\|X_t - \mathrm{E}\big[X_t|\mathcal{F}_{t-m}^{t+m}\big]\big\|_2 \le 2 \cdot c_X \cdot \left(\beta^2 + \sigma_B^2\right)^{\frac{m+1}{2}} .
$$

Since it is assumed that $\beta^2 + \sigma_B^2 < 1$, term $(\beta^2 + \sigma_B^2)^{\frac{m+1}{2}}$ is of order $O(m^{-a})$ for any $a > 0$ and thus $\{X_t\}$ is L_2-NED with constants $d_t = 2 \cdot c_X$ in Definition 2.6.
Secondly, $\{h(X_t)\}$ is L_2-NED on $\{B_t, Y_t\}$ of arbitrary size with constants $d_t = 2 \cdot c \cdot c_h$ (due to Lemma 6.4 and the Lipschitz property of the function $h(x)$ with some constant c_h) and so is $\{Z_t\}$ (adding a constant does not violate the NED condition).
Finally, if we knew that $\{Z_t\}$ is L_1-mixingale of size -1 we would have the desired decomposition of Z_t according to Theorem 2.4. We will show that $\{Z_t\}$ is even L_2-mixingale (and the inequality $\|\cdot\|_1 \le \|\cdot\|_2$ will ensure its L_1-mixingale property). This can be seen by applying Lemma 6.5 to sequence $\{Z_t\}$ (which is L_r-bounded for any $r > 1$ because the sequence $\{X_t\}$ is bounded and the function $h(x)$ is continuous) and to α-mixing

sequence $\{B_t, Y_t\}$ of size $-a$ for some $a > 2$ (Assumption A7). $\{Z_t\}$ is L_2-NED of arbitrary size $-b$, so the size of L_2-mixingale Z_t is $-\min(b, a(1/2 - 1/r))$. Notice, that $\min(b, a(1/2 - 1/r)) = a/2 - a/r > 1$ because the term a/r can be arbitrary small by increasing r. Constants c_t in Definition 2.5 are equal to $O(\max(\|X_t\|_r, 2 \cdot c \cdot c_h))$, they are bounded due to the boundedness of $\{X_t\}$, thus both sequences $\{R_t\}$ and $\{W_t\}$ in the decomposition are uniformly bounded according to Theorem 2.4. □

Theorem 2.7.
Let $\{X_t, t \in \mathbb{Z}\}$ be a Generalized RCA(1) process that satisfies Assumptions A7, A8 and A9. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuous measurable function such that $h(x) = x\phi(x)$ satisfies Lipschitz condition (2.28).
Then there exist constants $c > 0$ and $d > 0$ such that exponential inequality (2.29) *holds.*

Proof:
Denote $Z_t = h(X_t) - \mathrm{E}h(X_t)$ for each $t \in \mathbb{Z}$. Then Theorem 2.6 gives us decomposition (2.31) of $\{Z_t\}$, namely $\sum_{t=0}^{n-1} Z_t = R_0 - R_n + \sum_{t=0}^{n-1} W_t$, where $\{W_t\}$ is a uniformly bounded martingale difference sequence and $\{R_t\}$ is a uniformly bounded random sequence. Now we get

$$\begin{aligned}
\mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1} Z_t\right| > \varepsilon\right) &= \mathrm{P}\left(\left|\frac{1}{n}\left(R_0 - R_n + \sum_{t=0}^{n-1} W_t\right)\right| > \varepsilon\right) \le \\
&\le \mathrm{P}\left(\frac{1}{n}\left|R_0 - R_n\right| + \left|\frac{1}{n}\sum_{t=0}^{n-1} W_t\right| > \varepsilon\right) \le \\
&\le \mathrm{P}\left(\left|R_0 - R_n\right| > n \cdot \frac{\varepsilon}{2}\right) + \mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1} W_t\right| > \frac{\varepsilon}{2}\right) \le \\
&\le \mathrm{P}\left(\left|R_0\right| + \left|R_n\right| > n \cdot \frac{\varepsilon}{2}\right) + c_1 \cdot \mathrm{e}^{-d_1 n \left(\frac{\varepsilon}{2}\right)^2}
\end{aligned}$$

using the triangular inequality and the Hoeffding inequality for bounded martingale differences (see for instance [10], Theorem 15.20). An upper bound for the first summand might be easily obtained as follows:

$$\begin{aligned}
\mathrm{P}\left(|R_0| + |R_n| > n \cdot \frac{\varepsilon}{2}\right) &= \mathrm{P}\left(\mathrm{e}^{\varepsilon \cdot (|R_0| + |R_n|)} > \mathrm{e}^{n \cdot \frac{\varepsilon^2}{2}}\right) \le \mathrm{E}\left(\mathrm{e}^{\varepsilon \cdot (|R_0| + |R_n|)}\right) \cdot \mathrm{e}^{-n \cdot \frac{\varepsilon^2}{2}} \le \\
&\le c_2 \cdot \mathrm{e}^{-2n\left(\frac{\varepsilon}{2}\right)^2},
\end{aligned}$$

where finite constant c_2 can be found thanks to uniform boundedness of $\{R_t\}$. If we combine this result with the previous estimate, we get the desired result

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1} Z_t\right| > \varepsilon\right) \le \max(c_1, c_2) \cdot \mathrm{e}^{-\min(d_1, 2) \cdot n \cdot \left(\frac{\varepsilon}{2}\right)^2} = c \cdot \mathrm{e}^{-dn\varepsilon^2}$$ □

Now, let us formulate and prove the main theorem of this section.

Theorem 2.8.
Let $\{X_t, t \in \mathbb{Z}\}$ be a Generalized RCA(1) process according to Definition 2.2 with $\mathrm{E}[Y_t^2|\mathcal{F}_{t-1}] = \sigma^2 < +\infty$ *a.s. for all t. Let Assumptions A7, A8 and A9 be satisfied.*
Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuous measurable function such that $h(x) = x\phi(x)$ satisfies $h(x) > 0$ for $x \neq 0$ and the Lipschitz condition (2.28).
Then there exists $c > 0$ such that for each $n \in \mathbb{N}$

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\sqrt{n}\frac{\left(\widehat{\beta}(\phi) - \beta\right)}{\sqrt{V(\phi)}} \leq x \right) - \Phi(x) \right| \leq c \cdot \frac{(\ln n)^3}{\sqrt{n}}, \tag{2.32}$$

where estimator $\widehat{\beta}(\phi)$ is defined by (2.16) *and asymptotic variance $V(\phi)$ is defined by* (2.17).

Remark: This theorem covers both the LS estimator (choice $\phi(x) = x$) and the estimator with the smallest asymptotic variance in model with $\sigma_B^2 = \sigma_Y^2$ (choice $\phi(x) = \frac{x}{1+x^2}$), for further discussion see the beginning of Paragraph 2.4.1.

Proof: We basically follow similar proof from [5] for scalar RCA(1) model.
For each $n \in \mathbb{N}$ let us define

$$f_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \phi(X_{t-1})(X_t - \beta X_{t-1}),$$
$$g_n = \frac{1}{n} \sum_{t=1}^{n} \phi(X_{t-1})X_{t-1} = \frac{1}{n} \sum_{t=1}^{n} h(X_{t-1}).$$

Then $\sqrt{n}(\widehat{\beta}(\phi) - \beta) = \frac{f_n}{g_n}$. Denote $U(\phi) = \mathrm{E}\left(\phi^2(X_1)w(X_1)\right)$ where $w(x) = \sigma^2 + \sigma_B^2 x^2$ and notice that $V(\phi) = \frac{U(\phi)}{(\mathrm{E}h(X_1))^2} > 0$, so we have

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\sqrt{n}\frac{\left(\widehat{\beta}(\phi) - \beta\right)}{\sqrt{V(\phi)}} \leq x \right) - \Phi(x) \right| = \sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{\mathrm{E}h(X_1)}{\sqrt{U(\phi)}} \frac{f_n}{g_n} \leq x \right) - \Phi(x) \right|.$$

Lemma 6.2 gives us for any $\varepsilon > 0$

$$\begin{aligned} \sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{\mathrm{E}h(X_1)}{\sqrt{U(\phi)}} \frac{f_n}{g_n} \leq x \right) - \Phi(x) \right| \leq & \sup_{y\in\mathbb{R}} \left| \mathrm{P}\left(\frac{f_n}{\sqrt{U(\phi)}} \leq y \right) - \Phi(y) \right| + \\ & + \mathrm{P}\left(\left| \frac{g_n}{\mathrm{E}h(X_1)} - 1 \right| > \varepsilon \right) + \varepsilon. \end{aligned} \tag{2.33}$$

It can be easily derived that $\{\phi(X_{t-1})(X_t - \beta X_{t-1})\}$ is a martingale difference sequence with variance $U(\phi)$ defined previously. Function ϕ is continuous, process $\{X_t\}$ is uniformly bounded, thus $\{\phi(X_{t-1})(X_t - \beta X_{t-1})\}$ is also uniformly bounded. So Lemma 6.3 can be applied to the first term on the right hand side of inequality (2.33) and we have

$$\sup_{y\in\mathbb{R}} \left| \mathrm{P}\left(\frac{f_n}{\sqrt{U(\phi)}} \le y \right) - \Phi(y) \right| \le d \cdot \frac{(\ln n)^3}{\sqrt{n}},$$

where $d > 0$ is some constant.
The second term on the right hand side of inequality (2.33) can be arranged into

$$\begin{aligned} \mathrm{P}\left(\left| \frac{g_n}{\mathrm{E}h(X_1)} - 1 \right| > \varepsilon \right) &= \mathrm{P}\big(\left| g_n - \mathrm{E}h(X_1) \right| > \varepsilon \cdot \mathrm{E}h(X_1) \big) = \\ &= \mathrm{P}\left(\left| \frac{1}{n} \sum_{t=1}^{n} \big(h(X_{t-1}) - \mathrm{E}h(X_{t-1}) \big) \right| > \varepsilon \cdot \mathrm{E}h(X_1) \right), \end{aligned}$$

using the definition of g_n and the strict stationarity of $\{X_t\}$. All conditions of Theorem 2.7 are met and we have

$$\mathrm{P}\left(\left| \frac{1}{n} \sum_{t=1}^{n} \big(h(X_{t-1}) - \mathrm{E}h(X_{t-1}) \big) \right| > \varepsilon \cdot \mathrm{E}h(X_1) \right) \le p \cdot \mathrm{e}^{-qn\varepsilon^2},$$

where $p, q > 0$ are some constants.
If we sum up all derived results we gain that for any $\varepsilon > 0$ there exist positive constants d, p, q such that

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\sqrt{n} \left(\widehat{\beta}(\phi) - \beta \right) \le x \right) - G(x) \right| \le d \cdot \frac{(\ln n)^3}{\sqrt{n}} + p \cdot \mathrm{e}^{-qn\varepsilon^2} + \varepsilon.$$

Setting $\varepsilon = \frac{(\ln n)^3}{\sqrt{n}}$ we obtain the desired upper bound $c \cdot \frac{(\ln n)^3}{\sqrt{n}}$ for some positive constant c. □

Chapter 3

Estimators of higher-order models

3.1 Model specification

In this chapter we will return our attention to RCA(p) models of general order p which special case for $p = 1$ has been studied earlier. Similarly to the previous chapter, we will be interested in estimation of parameter $\boldsymbol{\beta}$, this time a p-dimensional vector. We will describe commonly used estimators (see for instance [25]) and introduce an extension of the functional estimator originally proposed by Schick in [28] for the first-order models and which we studied in Chapter 2. We will also derive statistical properties of the functional estimator and compare it to the least-squares estimator.

Equation (1.1) can be rewritten into the form closer to standard autoregressive expression

$$X_t = \sum_{i=1}^{p} \beta_i X_{t-i} + u_t = \boldsymbol{\beta}'\mathbf{Z}_{t-1} + u_t \tag{3.1}$$

with a new error process $u_t = \sum_{i=1}^{p} B_{t,i}X_{t-i} + Y_t = \mathbf{B}_t'\mathbf{Z}_{t-1} + Y_t$, where

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)', \quad \mathbf{B}_t = (B_{t,1}, B_{t,2}, \dots, B_{t,p})' \quad \text{and} \quad \mathbf{Z}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})'.$$

Another useful representation could be via multivariate RCA(1) model of the form

$$\mathbf{Z}_t = (\mathbf{B} + \mathbf{C}_t)\mathbf{Z}_{t-1} + \mathbf{U}_t, \tag{3.2}$$

where $\mathbf{U}_t = (Y_t, 0, \dots, 0)$ and

$$\mathbf{B} = \begin{pmatrix} \beta_1 & \beta_2 & \dots & \beta_{p-1} & \beta_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \mathbf{C}_t = \begin{pmatrix} B_{t,1} & B_{t,2} & \dots & B_{t,p} \\ 0 & 0 & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Similarly to the first-order case, we will use the stochastic setup given by assumptions A1, A2 and A5. The representation (3.2) of process $\{\mathbf{X}_t\}$ as a special multivariate RCA(1) process enables us to benefit from the theorems that will be proved in Chapter 4. Namely, we can adopt the stationarity assumption

A10: All the eigenvalues of matrix $\mathrm{E}(\mathbf{C}_t \otimes \mathbf{C}_t) + (\mathbf{B} \otimes \mathbf{B})$ are less than unity in modulus.

Definition 3.1. ***Generalized RCA(p) process***
Real-valued random process $\{X_t, t \in \mathbb{Z}\}$ is called Generalized RCA(p) process if X_t satisfies $\forall t \in \mathbb{Z}$ equation (1.1) and fulfills assumptions A1, A2, A5 and stationarity condition A10.

Theorem 4.1 in Chapter 4 proves that there exists strictly stationary and ergodic solution of equation (3.2) (or equivalent equation (1.1)), which is measurable with respect to $\mathcal{F}_t = \sigma\left((\mathbf{B}_s', Y_s)'; s \le t\right)$, and the solution is of the form

$$\mathbf{Z}_t = \sum_{j=0}^{+\infty} \left[\prod_{i=0}^{j-1} (\mathbf{B} + \mathbf{C}_{t-i})\right] \cdot \mathbf{U}_{t-j}, \tag{3.3}$$

where the product is defined as 1 for $j = 0$. This expression can be directly compared to its special case (2.4) for RCA(1) models.

Assuming that the conditional variance of error process $\{Y_t\}$ is constant and time independent, the conditional moments of process $\{X_t\}$ can be computed as follows:

$$\begin{aligned} \mathrm{E}[X_t|\mathcal{F}_{t-1}] &= \mathrm{E}\big[\boldsymbol{\beta}'\mathbf{Z}_{t-1} + u_t|\mathcal{F}_{t-1}\big] = \boldsymbol{\beta}'\mathbf{Z}_{t-1} + \mathrm{E}\big[\mathbf{B}_t'\mathbf{Z}_{t-1} + Y_t|\mathcal{F}_{t-1}\big] = \\ &= \boldsymbol{\beta}'\mathbf{Z}_{t-1} + \mathbf{Z}_{t-1}' \cdot \mathrm{E}\mathbf{B}_t + \mathrm{E}[Y_t|\mathcal{F}_{t-1}] = \boldsymbol{\beta}'\mathbf{Z}_{t-1}, \end{aligned} \tag{3.4}$$

$$\begin{aligned} \mathrm{var}\,[X_t|\mathcal{F}_{t-1}] &= \mathrm{E}\big[\,(X_t - \mathrm{E}[X_t|\mathcal{F}_{t-1}])^2\,|\mathcal{F}_{t-1}\big] = \mathrm{E}\big[u_t^2|\mathcal{F}_{t-1}\big] = \\ &= \mathrm{E}\Big[(\mathbf{B}_t'\mathbf{Z}_{t-1} + Y_t)^2\,|\mathcal{F}_{t-1}\Big] = \\ &= \mathrm{E}\big[Y_t^2 + 2\cdot\mathbf{B}_t'\mathbf{Z}_{t-1}Y_t + (\mathbf{B}_t'\mathbf{Z}_{t-1})^2|\mathcal{F}_{t-1}\big] = \\ &= \sigma^2 + 2\cdot\mathbf{B}_t'\mathbf{Z}_{t-1}\cdot\mathrm{E}[Y_t|\mathcal{F}_{t-1}] + \mathrm{E}\big[\mathbf{Z}_{t-1}'\mathbf{B}_t\mathbf{B}_t'\mathbf{Z}_{t-1}|\mathcal{F}_{t-1}\big] = \\ &= \sigma^2 + \mathbf{Z}_{t-1}'\cdot\mathrm{E}\left(\mathbf{B}_t\mathbf{B}_t'\right)\cdot\mathbf{Z}_{t-1} = \sigma^2 + \mathbf{Z}_{t-1}'\cdot\boldsymbol{\Sigma}\cdot\mathbf{Z}_{t-1} = \\ &= \sigma^2 + \left(\mathbf{Z}_{t-1}'\otimes\mathbf{Z}_{t-1}'\right)\cdot\mathrm{vec}(\boldsymbol{\Sigma}) = \\ &= \sigma^2 + \mathrm{vec}\left(\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'\right)'\cdot\mathbf{K}_n'\mathrm{vech}(\boldsymbol{\Sigma}), \end{aligned} \tag{3.5}$$

where the last equalities are obtained using Lemma 6.8 (identities b, c) and Lemma 6.9 from Chapter 6 where the duplication matrix $\mathbf{K}_n$ is also defined.

3.2 Functional estimator

The well-known estimator of parameter $\boldsymbol{\beta}$ (see for instance [25]) is the least-squares estimator defined by

$$\widehat{\boldsymbol{\beta}}_{LS} = \left(\sum_{t=1}^{n} \mathbf{Z}_{t-1}\mathbf{Z}'_{t-1}\right)^{-1} \cdot \left(\sum_{t=1}^{n} \mathbf{Z}_{t-1}X_t\right) \tag{3.6}$$

and it possesses essential statistical properties, namely strong consistency and asymptotical normality

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}\right) \longrightarrow N\left(0, \sigma^2\mathbf{V}^{-1} + \mathbf{V}^{-1}\mathrm{E}\left(\mathbf{Z}_0\mathbf{Z}'_0\mathbf{Z}'_0\boldsymbol{\Sigma}\mathbf{Z}_0\right)\mathbf{V}^{-1}\right),$$

where $\mathbf{V} = \mathrm{E}\mathbf{Z}_0\mathbf{Z}'_0$.

Similarly to the first-order case, the unknown parameters σ^2 and $\boldsymbol{\Sigma}$ can be estimated using a linear regression model: The conditional variance of the theoretical residuals u_t is derived in equation (3.5). If we substitute u_t by the estimated residuals $\widehat{u}_t = X_t - \widehat{\boldsymbol{\beta}}'_{LS}\mathbf{Z}_{t-1}$ into that expression, we obtain the regression equation

$$\widehat{u}_t^2 = \sigma^2 + \mathbf{A}'_{t-1}\gamma + \varepsilon_t$$

with regressors $(\mathbf{1}|\mathbf{A}'_{t-1}) = (\mathbf{1}|\mathrm{vec}(\mathbf{Z}_{t-1}\mathbf{Z}'_{t-1})' \cdot \mathbf{K}'_n)$, parameters $(\sigma^2, \gamma')' = (\sigma^2, \mathrm{vech}(\boldsymbol{\Sigma})')'$ to estimate and a noise process ε_t. The solution of the regression equation is

$$\begin{aligned}\widehat{\gamma} &= \left(\sum_{t=1}^{n}(\mathbf{A}_{t-1} - \overline{\mathbf{A}})(\mathbf{A}_{t-1} - \overline{\mathbf{A}})'\right)^{-1} \sum_{t=1}^{n}(X_t - \widehat{\boldsymbol{\beta}}'_{LS}\mathbf{Z}_{t-1})^2(\mathbf{A}_{t-1} - \overline{\mathbf{A}}), \\ \widehat{\sigma}^2 &= \frac{1}{n}\sum_{t=1}^{n}(X_t - \widehat{\boldsymbol{\beta}}'_{LS}\mathbf{Z}_{t-1})^2 - \widehat{\gamma}'\overline{\mathbf{A}},\end{aligned} \tag{3.7}$$

where $\overline{\mathbf{A}} = \frac{1}{n}\sum_{t=1}^{n}\mathbf{A}_{t-1}$. Nicholls and Quinn proved in [25], Theorem 3.2, that under higher moment conditions of process $\{X_t\}$ such estimators are strongly consistent and asymptotically normal. We could generalize estimators (3.7) in the same manner as Schick did for the first-order RCA(1) models (see formulas (2.25)). Namely, we could impose a measurable function $\psi : \mathbb{R}^p \to \mathbb{R}^p$, define $\mathbf{B}_{t-1} = \mathbf{K}_n \cdot \mathrm{vec}(\psi(\mathbf{Z}_{t-1})\mathbf{Z}'_{t-1})$ instead of the quantities $\mathbf{A}_{t-1}$ and reformulate expressions (3.7).

Other commonly used estimators are the weighted least-squares or the maximum likelihood estimators. However, the exact derivation of such estimators is not straightforward and precise statistical inference requires stronger assumptions. Inspired by the extension of the least-squares estimator of the first-order RCA models into a class of the functional estimators, we propose an analogous extension for the higher-order models. Consider measurable function $\phi : \mathbb{R}^p \to \mathbb{R}^p$ and define

$$\widehat{\boldsymbol{\beta}}(\phi) = \left(\sum_{t=1}^{n} \phi(\mathbf{Z}_{t-1})\mathbf{Z}'_{t-1}\right)^{-1} \cdot \left(\sum_{t=1}^{n} \phi(\mathbf{Z}_{t-1})X_t\right). \tag{3.8}$$

The following theorem states basic properties of estimator (3.8):

Theorem 3.1.
Consider RCA(p) model according to Definition 3.1.
Let $\mathrm{E}[Y_t^2|\mathcal{F}_{t-1}] = \sigma^2 < +\infty$ *a.s. for all* t *and let* $\phi : \mathbb{R}^p \to \mathbb{R}^p$ *be measurable function such that* $\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'$ *is a finite positive definite matrix and* $\mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)'\left(\mathbf{Z}_0'\boldsymbol{\Sigma}\mathbf{Z}_0\right)\right)$ *is a finite matrix. Denote*

$$\mathbf{V}(\phi) = \left(\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right)^{-1} \cdot \mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)' \cdot w(\mathbf{Z}_0)\right) \cdot \left(\mathrm{E}\mathbf{Z}_0\phi(\mathbf{Z}_0)'\right)^{-1}, \tag{3.9}$$

where $w(\mathbf{z}) = \sigma^2 + \mathbf{z}'\boldsymbol{\Sigma}\mathbf{z}$ *for* $\mathbf{z} \in \mathbb{R}^p$.
Then estimator (3.8) *is a strongly consistent and asymptotically normal estimator of parameter* $\boldsymbol{\beta}$ *with asymptotical variance matrix defined by* (3.9).

Remark: Function $\phi(\mathbf{x}) = \mathbf{x}$, that corresponds to the least-squares estimator, fulfills the assumptions provided that $\mathrm{E}X_t^4 < +\infty$ (see [25], Theorem 3.1). If function ϕ is bounded then the second assumption reduces to $\mathrm{E}X_t^2 < +\infty$.

Proof: Using definition of functional estimator (3.8) we can infer that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta} &= \left(\frac{1}{n}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'\right)^{-1} \cdot \left(\frac{1}{n}\sum_{t=1}^{n}\left[\phi(\mathbf{Z}_{t-1})X_t - \phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'\boldsymbol{\beta}\right]\right) = \\ &= \left(\frac{1}{n}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'\right)^{-1} \cdot \left(\frac{1}{n}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})u_t\right). \end{aligned} \tag{3.10}$$

Likewise in the proof of Theorem 2.2, the components of sequences $\{\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}', t \in \mathbb{Z}\}$ and $\{\phi(\mathbf{Z}_{t-1})u_t, t \in \mathbb{Z}\}$ are strictly stationary and ergodic. Moreover, the components of the latter sequence form a martingale difference sequence because

$$\mathrm{E}[\phi(\mathbf{Z}_{t-1})u_t|\mathcal{F}_{t-1}] = \phi(\mathbf{Z}_{t-1})\mathrm{E}[u_t|\mathcal{F}_{t-1}] = \mathbf{0}.$$

The ergodic theorem (see e.g. [10], Theorem 13.12) tells us that the first term in (3.10) almost surely converges to $\left(\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right)^{-1}$ whereas the second term converges to $\mathrm{E}(\phi(\mathbf{Z}_0)u_1) = \mathbf{0}$, which completes the proof of consistency of the estimator.

Similarly to equation (3.10) notice that

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta}\right) = \left(\frac{1}{n}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'\right)^{-1} \cdot \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})u_t\right).$$

Choose arbitrary $\alpha \in \mathbb{R}^p$ and note that also $\{\alpha'\phi(\mathbf{Z}_{t-1})u_t\}$ forms an univariate martingale difference sequence with variance

$$\begin{aligned}\mathrm{var}(\alpha'\phi(\mathbf{Z}_{t-1})u_t) &= \mathrm{E}\left(\mathrm{E}\Big[\left(\alpha'\phi(\mathbf{Z}_{t-1})\right)^2 u_t^2|\mathcal{F}_{t-1}\Big]\right) = \\ &= \mathrm{E}\left(\left(\alpha'\phi(\mathbf{Z}_{t-1})\right)^2\left(\sigma^2 + \mathbf{Z}_{t-1}'\boldsymbol{\Sigma}\mathbf{Z}_{t-1}\right)\right) = \end{aligned}$$

$$= \alpha' \mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)'\left(\sigma^2 + \mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0\right)\right)\alpha < +\infty.$$

Due to Lindeberg-Levy theorem for martingales (see [7]) this means that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\alpha'\phi(\mathbf{Z}_{t-1})u_t = \alpha'\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})u_t\right)$$

has asymptotically one-dimensional normal distribution with zero mean and the variance computed above. Thus, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})u_t$ has asymptotically p-dimensional normal distribution and so does $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta}\right)$ with the asymptotical variance matrix equal to (3.9). □

3.2.1 Asymptotical variance matrix

Asymptotical variance matrix of the functional estimator is relatively complex and strongly depends on the choice of function ϕ. In this section we will compute the exact form of the matrix for a few particular choices of the function and then suggest a consistent estimator of the variance matrix.

Let us consider asymptotical variance matrix $\mathbf{V}(\phi)$ given by (3.9).

1. Choice $\phi(\mathbf{z}) = \mathbf{z}$ corresponds to the least-squares estimator (3.6) and the variance matrix directly simplifies into

$$\mathbf{V}(\phi) = \sigma^2\mathbf{V}^{-1} + \mathbf{V}^{-1}\cdot \mathrm{E}\left(\mathbf{Z}_0\mathbf{Z}_0'\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0\right)\cdot\mathbf{V}^{-1},$$

where $\mathbf{V} = \mathrm{E}\mathbf{Z}_0\mathbf{Z}_0' = \mathrm{var}\,\mathbf{Z}_0$.

2. Choice $\phi(\mathbf{z}) = \frac{\mathbf{z}}{\sigma^2+\mathbf{z}'\mathbf{\Sigma}\mathbf{z}}$ results in the weighted least-squares estimator with the variance matrix

$$\begin{aligned}\mathbf{V}(\phi) &= \left(\mathrm{E}\frac{\mathbf{Z}_0\mathbf{Z}_0'}{\sigma^2+\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0}\right)^{-1}\cdot \mathrm{E}\left(\frac{\mathbf{Z}_0\mathbf{Z}_0'}{(\sigma^2+\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0)^2}\left(\sigma^2+\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0\right)\right)\cdot \\ &\cdot\left(\mathrm{E}\frac{\mathbf{Z}_0\mathbf{Z}_0'}{\sigma^2+\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0}\right)^{-1} = \left(\mathrm{E}\frac{\mathbf{Z}_0\mathbf{Z}_0'}{\sigma^2+\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0}\right)^{-1} = \\ &= \left(\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right)^{-1}.\end{aligned}$$

3. Choice $\phi(\mathbf{z}) = \frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}}$ corresponds to the special case of the weighted least-squares estimator for $\mathbf{\Sigma} = \sigma^2\mathrm{I}$ and its variance matrix equals

$$\mathbf{V}(\phi) = \sigma^2\mathbf{F}^{-1} + \mathbf{F}^{-1}\mathrm{E}\left(\frac{\mathbf{Z}_0\mathbf{Z}_0'\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0}{(1+\mathbf{Z}_0'\mathbf{Z}_0)^2}\right)\mathbf{F}^{-1},$$

where $\mathbf{F} = \mathrm{E}\frac{\mathbf{Z}_0\mathbf{Z}_0'}{1+\mathbf{Z}_0'\mathbf{Z}_0}$.

Theorem 3.1 gives the exact formula for the asymptotical variance matrix of estimator $\boldsymbol{\beta}(\phi)$ for general function ϕ. The three specific choices of ϕ given above indicate that evaluation of the matrix is rather complicated. Naturally, we are interested in a consistent estimator of this matrix.

Theorem 3.2.
Consider RCA(p) model according to Definition 3.1. Let $\phi : \mathbb{R}^p \to \mathbb{R}^p$ be measurable function such that matrix $\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'$ is finite and positive definite.
Let $\widehat{\sigma}_n^2$ and $\widehat{\boldsymbol{\Sigma}}_n$ be strongly consistent estimators of σ^2 and $\boldsymbol{\Sigma}$, respectively. Denote

$$\mathbf{P}_n = \frac{1}{n}\sum_{t=1}^{n} \phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}',$$
$$\mathbf{Q}_n = \frac{1}{n}\sum_{t=1}^{n} \left[\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})'(\widehat{\sigma}_n^2 + \mathbf{Z}_{t-1}'\widehat{\boldsymbol{\Sigma}}_n\mathbf{Z}_{t-1})\right].$$

Then

$$\widehat{\mathbf{V}}_n(\phi) = \mathbf{P}_n^{-1} \cdot \mathbf{Q}_n \cdot \mathbf{P}_n'^{-1} \tag{3.11}$$

is a strongly consistent estimator of the asymptotical variance matrix $\mathbf{V}(\phi)$ given by (3.9).

Proof: Denote $\mathbf{Q} = \phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)'\,(\sigma^2 + \mathbf{Z}_0'\boldsymbol{\Sigma}\mathbf{Z}_0)$ and notice that

$$\mathbf{V}(\phi) = (\mathrm{E}\mathbf{P}_1)^{-1} \cdot \mathrm{E}\mathbf{Q} \cdot (\mathrm{E}\mathbf{P}_1')^{-1}.$$

Since $\{X_t, t \in \mathbf{Z}\}$ is strictly stationary and ergodic so is $\{\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}', t \in \mathbf{Z}\}$ and the ergodic theorem tells us that $\mathbf{P}_n \xrightarrow{a.s.} \mathrm{E}\mathbf{P}_1 = \mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'$ as $n \to +\infty$. Term $\mathbf{Q}_n$ might be rewritten as

$$\mathbf{Q}_n = \widehat{\sigma}_n^2 \cdot \frac{1}{n}\sum_{t=1}^{n} \phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})' + \frac{1}{n}\sum_{t=1}^{n} \left[\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})'(\mathbf{Z}_{t-1}'\widehat{\boldsymbol{\Sigma}}_n\mathbf{Z}_{t-1})\right]. \tag{3.12}$$

The first term of (3.12) converges to $\sigma^2 \cdot \mathrm{E}\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)'$ due to consistency of $\widehat{\sigma}_n^2$ and ergodicity of $\{\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})', t \in \mathbf{Z}\}$. The convergence of the second term in (3.12) has to be proved more technically (using the Kronecker product, the *vec* operator and the related identities from Lemma 6.6 and Lemma 6.8 stated in Chapter 6):

$$\mathrm{vec}\left(\frac{1}{n}\sum_{t=1}^{n} \phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})'(\mathbf{Z}_{t-1}'\widehat{\boldsymbol{\Sigma}}_n\mathbf{Z}_{t-1})\right) =$$
$$= \frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\left(\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})'\right)_{p\times p} \cdot \left(\mathbf{Z}_{t-1}'\widehat{\boldsymbol{\Sigma}}_n\mathbf{Z}_{t-1}\right)_{1\times 1} = \frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\left(\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})'\right) \cdot$$
$$\cdot\, \mathrm{vec}\left(\mathbf{Z}_{t-1}'\widehat{\boldsymbol{\Sigma}}_n\mathbf{Z}_{t-1}\right) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\left(\phi(\mathbf{Z}_{t-1})_{p\times 1}\phi(\mathbf{Z}_{t-1})_{1\times p}'\right) \cdot \left((\mathbf{Z}_{t-1}' \otimes \mathbf{Z}_{t-1}')\mathrm{vec}(\widehat{\boldsymbol{\Sigma}}_n)\right) =$$

$$=\frac{1}{n}\sum_{t=1}^{n}\left(\phi(\mathbf{Z}_{t-1})\otimes\phi(\mathbf{Z}_{t-1})\right)_{p^2\times 1}\cdot(\mathbf{Z}_{t-1}'\otimes\mathbf{Z}_{t-1}')_{1\times p^2}\cdot\mathrm{vec}(\widehat{\mathbf{\Sigma}}_n)_{p^2\times 1}=$$
$$=\left(\frac{1}{n}\sum_{t=1}^{n}\left(\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'\right)\otimes\left(\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'\right)\right)\cdot\mathrm{vec}(\widehat{\mathbf{\Sigma}}_n).$$

The latter term thanks to the ergodicity of $\{\mathbf{Z}_t\}$ and consistency of $\widehat{\mathbf{\Sigma}}_n$ converges to

$$\mathrm{E}\left(\left(\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right)\otimes\left(\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right)\right)\cdot\mathrm{vec}(\mathbf{\Sigma})=\mathrm{E}\left(\left(\phi(\mathbf{Z}_0)\otimes\phi(\mathbf{Z}_0)\right)\cdot\left(\mathbf{Z}_0'\otimes\mathbf{Z}_0'\right)\cdot\mathrm{vec}(\mathbf{\Sigma})\right)=$$
$$=\mathrm{E}\left(\mathrm{vec}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)'\right)\cdot\mathrm{vec}\left(\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0'\right)\right)=\mathrm{vec}\left(\mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)'\mathbf{Z}_0'\mathbf{\Sigma}\mathbf{Z}_0'\right)\right).$$

Thus, $\mathbf{Q}_n \xrightarrow{a.s.} \mathrm{E}\mathbf{Q}$ as $n\to+\infty$ which completes the proof that $\widehat{\mathbf{V}}_n(\phi)\xrightarrow{a.s.}\mathbf{V}(\phi)$. □

3.2.2 Optimal estimator

Similarly to Section 2.4.1 where we studied the impact of the generating function of the functional estimators for the first-order RCA models, we will find an optimal choice of the function for the higher-order models in this section. Optimality of an estimator will be defined using its asymptotical variance matrix. Namely, estimator $\widehat{\boldsymbol{\beta}}(\psi)$ defined by equation (3.8) is optimal if its asymptotical variance matrix $\mathbf{V}(\psi)$ defined by (3.9) satisfies $\mathbf{V}(\phi)-\mathbf{V}(\psi)\geq\mathbf{0}$ for any estimator $\widehat{\boldsymbol{\beta}}(\phi)$ with variance matrix $\mathbf{V}(\phi)$ (the difference of the variance matrices is a positively semi-definite matrix). This property ensures, for instance, that the variances of all elements of vector estimator $\widehat{\boldsymbol{\beta}}(\psi)$ are the smallest possible.

Now let us assume that the conditional variance of error process $\{Y_t\}$ equals σ^2. We will show that the optimal estimator is the weighted-least squares estimator which corresponds to the choice $\psi(\mathbf{z})=\frac{\mathbf{z}}{\sigma^2+\mathbf{z}'\mathbf{\Sigma}\mathbf{z}}=\frac{\mathbf{z}}{w(\mathbf{z})}$ and which asymptotical variance matrix has been derived in Section 3.2.1, point 2. The derivation mimics the similar inference for RCA(1) models performed in Section 2.4.1. Let us define two p-dimensional random vectors

$$\mathbf{T}_1=\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})(X_t-\boldsymbol{\beta}'\mathbf{Z}_{t-1})=\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})u_t,$$
$$\mathbf{T}_2=\sum_{t=1}^{n}\frac{\mathbf{Z}_{t-1}(X_t-\boldsymbol{\beta}'\mathbf{Z}_{t-1})}{\sigma^2+\mathbf{Z}_{t-1}'\mathbf{\Sigma}\mathbf{Z}_{t-1}}=\sum_{t=1}^{n}\frac{\mathbf{Z}_{t-1}u_t}{w(\mathbf{Z}_{t-1})}.$$

Since both sequences $\{\phi(\mathbf{Z}_{t-1})u_t\}$ and $\{\frac{\mathbf{Z}_{t-1}u_t}{w(\mathbf{Z}_{t-1})}\}$ are martingale differences w.r. to $\mathcal{F}_t$, we know that both $\mathrm{E}\mathbf{T}_1=0$ and $\mathrm{E}\mathbf{T}_2=0$. Variance matrix of vector $\mathbf{T}_1$ then equals

$$\mathrm{E}\mathbf{T}_1\mathbf{T}_1'=\mathrm{E}\left(\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})u_t\cdot\sum_{s=1}^{n}\phi(\mathbf{Z}_{s-1})'u_s\right)=$$
$$=\mathrm{E}\left(\sum_{t=1}^{n}\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})'\cdot u_t^2+\sum_{\substack{t=1\\t\neq s}}^{n}\sum_{s=1}^{n}\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{s-1})'\cdot u_tu_s\right)=$$

$$= \sum_{t=1}^{n} \mathrm{E}\left(\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})' \cdot \mathrm{E}\big[u_t^2|\mathcal{F}_{t-1}\big]\right) = \sum_{t=1}^{n} \mathrm{E}\left(\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{t-1})' \cdot w(\mathbf{Z}_{t-1})\right)$$
$$= n \cdot \mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)' \cdot w(\mathbf{Z}_0)\right) \tag{3.13}$$

using strict stationarity of the process $\{\mathbf{Z}_t\}$ and the fact that for $t > s$ (for $t < s$ the same conclusion holds)

$$\mathrm{E}\left(\sum_{\substack{t=1\\t>s}}^{n}\sum_{s=1}^{n} \phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{s-1})' \cdot u_t u_s\right) = \mathrm{E}\left(\sum_{\substack{t=1\\t>s}}^{n}\sum_{s=1}^{n} \mathrm{E}[\phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{s-1})' \cdot u_t u_s|\mathcal{F}_{t-1}]\right) =$$
$$= \mathrm{E}\left(\sum_{\substack{t=1\\t>s}}^{n}\sum_{s=1}^{n} \phi(\mathbf{Z}_{t-1})\phi(\mathbf{Z}_{s-1})' \cdot u_s \cdot \mathrm{E}[u_t|\mathcal{F}_{t-1}]\right) = 0.$$

Analogously

$$\mathrm{E}\mathbf{T}_2\mathbf{T}_2' = \mathrm{E}\left(\sum_{t=1}^{n} \frac{\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}' \cdot u_t^2}{w(\mathbf{Z}_{t-1})^2} + \sum_{\substack{t=1\\t\neq s}}^{n}\sum_{s=1}^{n} \frac{\mathbf{Z}_{t-1}\mathbf{Z}_{s-1}' \cdot u_t u_s}{w(\mathbf{Z}_{t-1})w(\mathbf{Z}_{s-1})}\right) =$$
$$= \sum_{t=1}^{n} \mathrm{E}\left(\frac{\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}'}{w(\mathbf{Z}_{t-1})^2} \cdot \mathrm{E}\big[u_t^2|\mathcal{F}_{t-1}\big]\right) = n \cdot \mathrm{E}\left(\frac{\mathbf{Z}_0\mathbf{Z}_0'}{w(\mathbf{Z}_0)}\right). \tag{3.14}$$

The cross-covariance matrix of $\mathbf{T}_1$ and $\mathbf{T}_2$ could be rewritten into

$$\mathrm{E}\mathbf{T}_1\mathbf{T}_2' = \mathrm{E}\left(\sum_{t=1}^{n} \phi(\mathbf{Z}_{t-1})u_t \cdot \sum_{s=1}^{n} \frac{\mathbf{Z}_{s-1}' u_s}{w(\mathbf{Z}_{s-1})}\right) =$$
$$= \mathrm{E}\left(\sum_{t=1}^{n} \frac{\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}' \cdot u_t^2}{w(\mathbf{Z}_{t-1})} + \sum_{\substack{t=1\\t\neq s}}^{n}\sum_{s=1}^{n} \frac{\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{s-1}' \cdot u_t u_s}{w(\mathbf{Z}_{s-1})}\right) =$$
$$= \sum_{t=1}^{n} \mathrm{E}\left(\frac{\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{t-1}'}{w(\mathbf{Z}_{t-1})} \cdot w(\mathbf{Z}_{t-1})\right) + \sum_{\substack{t=1\\t\neq s}}^{n}\sum_{s=1}^{n} \mathrm{E}\left(\frac{\phi(\mathbf{Z}_{t-1})\mathbf{Z}_{s-1}' \cdot u_t u_s}{w(\mathbf{Z}_{s-1})}\right) =$$
$$= n \cdot \mathrm{E}\left(\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right) \tag{3.15}$$

Variance matrix of $2p$-dimensional random vector $(\mathbf{T}_1', \mathbf{T}_2')'$ is equal to the block-matrix

$$\begin{pmatrix} \mathrm{E}\mathbf{T}_1\mathbf{T}_1', & \mathrm{E}\mathbf{T}_1\mathbf{T}_2' \\ \mathrm{E}\mathbf{T}_2\mathbf{T}_1', & \mathrm{E}\mathbf{T}_2\mathbf{T}_2' \end{pmatrix} = n \cdot \begin{pmatrix} \mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)' \cdot w(\mathbf{Z}_0)\right), & \mathrm{E}\left(\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right) \\ \mathrm{E}\left(\mathbf{Z}_0\phi(\mathbf{Z}_0)'\right), & \mathrm{E}\left(\frac{\mathbf{Z}_0\mathbf{Z}_0'}{w(\mathbf{Z}_0)}\right) \end{pmatrix}, \tag{3.16}$$

where the elements were computed in (3.13) - (3.15). Case $p = 1$ has been studied previously and it is based on the non-negativity of the determinant of matrix (3.16). For general $p \geq 1$ we will employ the following theorem which is a modification of the theorem concerning computation of the determinant for a block matrix:

Theorem 3.3.
Consider block matrix

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{pmatrix},$$

where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *are* $p \times p$*-dimensional matrices. Let* $\mathbf{M} \geq \mathbf{0}$, $\mathbf{B}$ *be a regular matrix, and* $\mathbf{C}$ *be a symmetric regular matrix.*
Then

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{B}'^{-1} - \mathbf{C}^{-1} \geq \mathbf{0}.$$

Proof: We will prove that $\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}' \geq \mathbf{0}$ from which the result could be obtained by multiplication of matrices $\mathbf{B}^{-1}$ and $\mathbf{B}'^{-1}$, respectively. Choose arbitrary $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^p$. Matrix $\mathbf{M}$ is positively semi-definite which means that

$$(\mathbf{x}', \mathbf{y}')\mathbf{M}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{y}'\mathbf{B}'\mathbf{x} + \mathbf{x}'\mathbf{B}\mathbf{y} + \mathbf{y}'\mathbf{C}\mathbf{y} = \mathbf{x}'\mathbf{A}\mathbf{x} + \mathbf{y}'\mathbf{C}\mathbf{y} + 2\mathbf{x}'\mathbf{B}\mathbf{y} \geq 0.$$

We want to know that for any $\mathbf{x} \in \mathbb{R}^p$ holds

$$\mathbf{x}'\mathbf{A}\mathbf{x} - \mathbf{x}'\mathbf{B}\mathbf{C}^{-1}\mathbf{B}'\mathbf{x} \geq 0.$$

Comparing the previous two inequalities we conclude that it suffices to find $\mathbf{y} \in \mathbf{R}^p$ such that

$$\mathbf{y}'\mathbf{C}\mathbf{y} + 2\mathbf{x}'\mathbf{B}\mathbf{y} = -\mathbf{x}'\mathbf{B}\mathbf{C}^{-1}\mathbf{B}'\mathbf{x}.$$

Denote $\mathbf{z} = \mathbf{B}'\mathbf{x}$. Then $\mathbf{y}$ has to satisfy elliptical equation

$$\mathbf{y}'\mathbf{C}\mathbf{y} + 2\mathbf{z}'\mathbf{y} + \mathbf{z}'\mathbf{C}^{-1}\mathbf{z} = 0. \tag{3.17}$$

Assuming for the moment that $p = 1$, the equation simplifies into $cy^2 + 2zy + \frac{z^2}{c} = 0$ which is solved by $y = -\frac{z}{c}$. And really the choice $\mathbf{y} = -\mathbf{C}^{-1}\mathbf{z}$ fulfils equation (3.17) which completes the proof. □

Let us come back to block matrix (3.16). If the block elements $\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'$ and $\mathrm{E}\frac{\mathbf{Z}_0\mathbf{Z}_0'}{w(\mathbf{Z}_0)}$ are regular matrices, we could employ theorem 3.3 and conclude that

$$\begin{aligned} &\left(\mathrm{E}\phi(\mathbf{Z}_0)\mathbf{Z}_0'\right)^{-1} \mathrm{E}\left(\phi(\mathbf{Z}_0)\phi(\mathbf{Z}_0)' \cdot w(\mathbf{Z}_0)\right) \left(\mathrm{E}\mathbf{Z}_0\phi(\mathbf{Z}_0)'\right)^{-1} - \left(\mathrm{E}\frac{\mathbf{Z}_0\mathbf{Z}_0'}{w(\mathbf{Z}_0)}\right)^{-1} \geq \mathbf{0} \\ &\Longleftrightarrow \mathbf{V}(\phi) - \mathbf{V}(\psi) \geq \mathbf{0}, \end{aligned}$$

where $\mathbf{V}(\phi)$ is the asymptotical variance matrix of general estimator $\widehat{\boldsymbol{\beta}}(\phi)$ and $\psi(\mathbf{z}) = \frac{\mathbf{z}}{w(\mathbf{z})}$ corresponds to the asymptotical variance matrix of the weighted least-squares estimator.

3.3 Simulation study

In the previous section we extended the concept of functional estimator to the higher-order RCA processes. The aim of this study is to compare two functional estimators – one corresponding to the least-squares estimator, the other to the special case of the weighted least-squares estimator. We are interested in the impact of variances of random coefficients to the estimators.

Setup of the simulations is the following: We simulate a sequence of observations from RCA(2) model given by equation

$$X_t = (0.4 + B_{t,1})X_{t-1} + (0.2 + B_{t,2})X_{t-2} + Y_t, \tag{3.18}$$

where random coefficients $\{\mathbf{B}_t = (B_{t,1}, B_{t,2})', t \in \mathbb{Z}\}$ are independent normally distributed $N_2(\mathbf{0}, \mathbf{\Sigma})$ random variables and error process $\{Y_t, t \in \mathbb{Z}\}$ is a sequence of independent and normally distributed $N(0,1)$ variables independent of $\{B_t\}$.
Simulation procedure involves the following steps:

- set $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ where $\sigma_1^2, \sigma_2^2 \in \{0; 0.05; 0.10; \ldots; 0.35\}$
  - repeat 100 times
    1. simulate RCA(2) time series (500 observations)
    2. estimate parameter $\boldsymbol{\beta}$ using $\widehat{\boldsymbol{\beta}}_{LS}$ and $\widehat{\boldsymbol{\beta}}(\phi)$ with $\phi(\mathbf{z}) = \frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}}$, respectively
  - for each estimator compute $\overline{\widehat{\beta}} = \frac{1}{100}\sum_{i=1}^{100} \widehat{\beta}^i$ and $MSE = \frac{1}{100}\sum_{i=1}^{100} \left(\widehat{\beta}^i - \overline{\widehat{\beta}}\right)^2$
- plot $\overline{\widehat{\beta}}$ and MSE as functions of σ_1^2 for the fixed level of σ_2^2 for each estimator

Before the actual simulation, let us examine stationarity assumption A10 closely. This assumption states that all eigenvalues of matrix $\mathrm{E}(\mathbf{C}_t \otimes \mathbf{C}_t) + (\mathbf{B} \otimes \mathbf{B})$ should be less than unity in order there exists a stationary RCA process. In our case

$$\mathbf{C}_t = \begin{pmatrix} B_{t,1} & B_{t,2} \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.4 & 0.2 \\ 1 & 0 \end{pmatrix}.$$

Using independence of $B_{t,1}$ and $B_{t,2}$, that follows from their non-correlated jointly normal distribution, we have

$$\mathrm{E}(\mathbf{C}_t \otimes \mathbf{C}_t) + (\mathbf{B} \otimes \mathbf{B}) =$$

$$= \mathrm{E}\begin{pmatrix} B_{t,1}^2 & B_{t,1}B_{t,2} & B_{t,1}B_{t,2} & B_{t,2}^2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0.16 & 0.08 & 0.08 & 0.04 \\ 0.4 & 0 & 0.2 & 0 \\ 0.4 & 0.2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} =$$

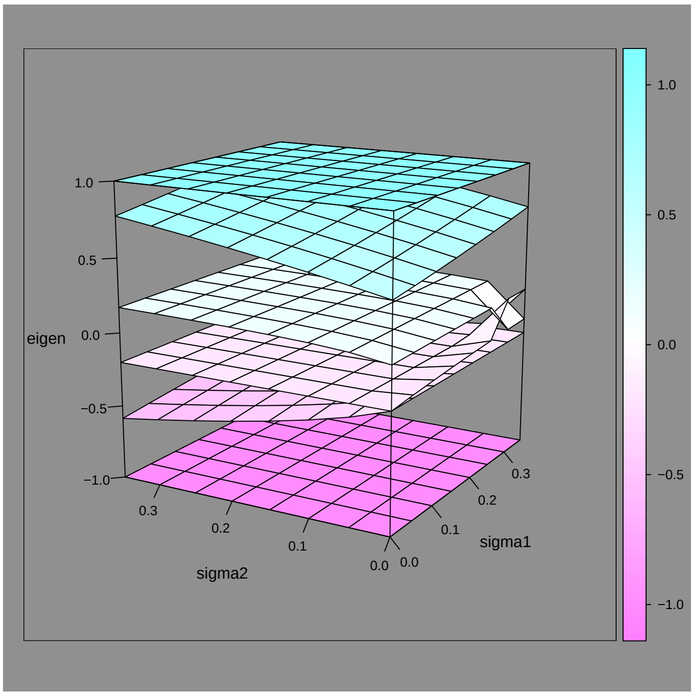

Figure 3.1: Eigenvalues from staionarity assumption as functions of variances σ_1^2 and σ_2^2. Each surface stands for a single eigenvalue.

$$= \begin{pmatrix} \sigma_1^2 + 0.16 & 0.08 & 0.08 & \sigma_2^2 + 0.04 \\ 0.4 & 0 & 0.2 & 0 \\ 0.4 & 0.2 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

The latter matrix has four eigenvalues for each choice of σ_1^2 and σ_2^2. Figure 3.1 displays those eigenvalues as functions of the variances. Each surface stands for one eigenvalue and we can see that all of them are bounded by values of z axis -1 and 1, respectively (pink bottom plane and light blue upper plane). Thus, each choice of variance matrix $\mathbf{\Sigma}$ leads to a stationary RCA(2) process.

Results of the simulation study are displayed on Figures 3.2 and 3.3. Compared estimators $\widehat{\boldsymbol{\beta}}_{LS}$ and $\widehat{\boldsymbol{\beta}}(\frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}})$ behave similarly when random coefficients have low variances. However, $\widehat{\boldsymbol{\beta}}_{LS}$ on average underestimates the true value of parameter when variances increase and the estimated values vary depending on simulations. On the other hand, $\widehat{\boldsymbol{\beta}}(\frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}})$ is stable no matter how large variances of random coefficients are and the estimated values do not vary so much.

We also performed the same simulations with the only difference, namely we set true parameter $\boldsymbol{\beta}$ in equation (3.18) (where it equals $(0.4, 0.2)'$) to $(0.2, 0.4)'$ or $(0, 0)'$, respectively. Figures 3.4 and 3.5 display results for $\boldsymbol{\beta} = (0.2, 0.4)'$ which do not differ from the previous ones and the least-squares estimator still underestimates the true parameter when variances increase. Figures 3.6 and 3.7 show results for $\boldsymbol{\beta} = (0, 0)'$. There both estimators behave similarly no matter how large the variances are, with a little bigger MSE in case of the least-squares estimator.

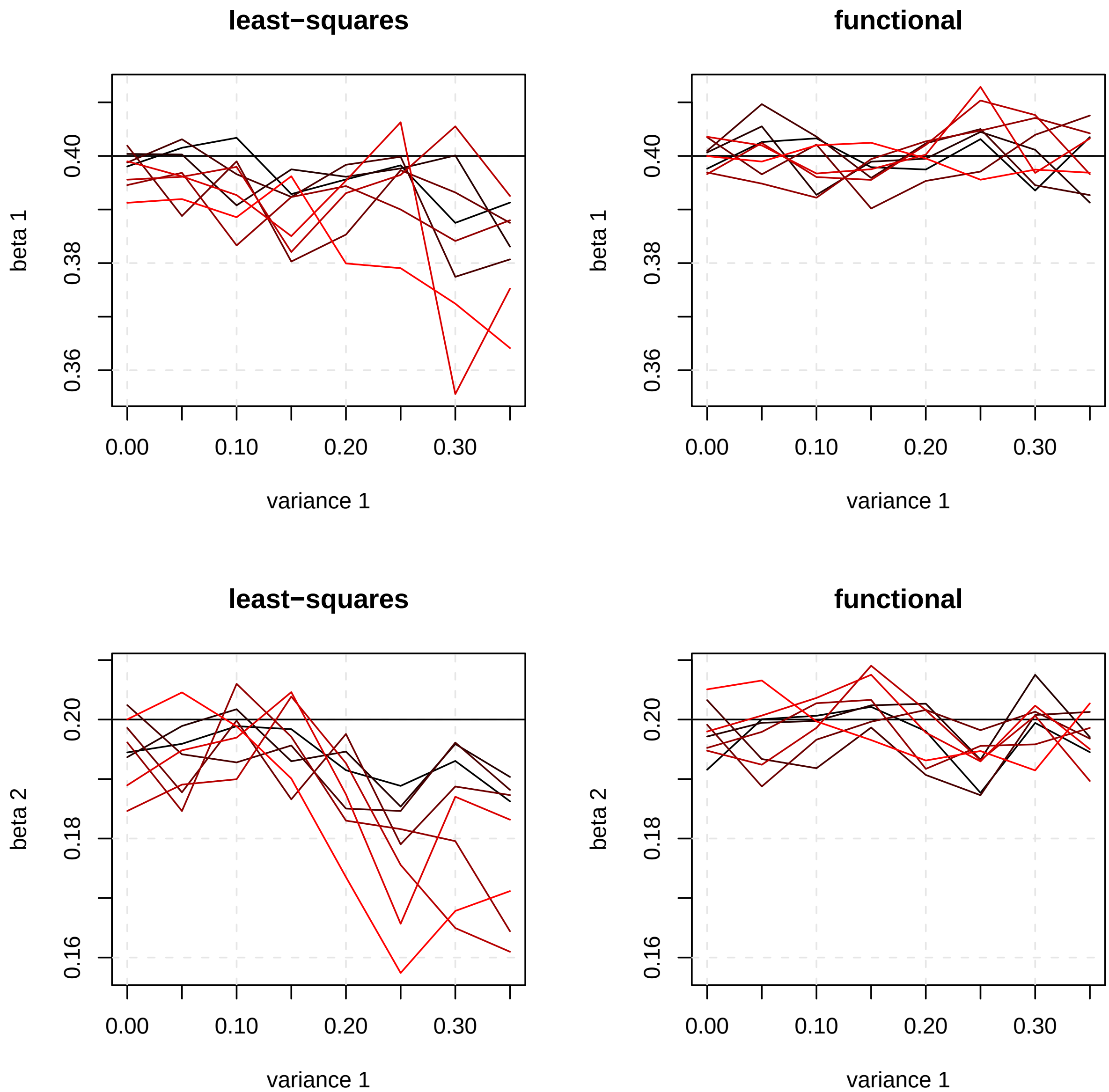

Figure 3.2: Average estimate of true parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0.4, 0.2)'$ in simulated RCA(2) model. Upper panels stand for estimation of β_1, lower for β_2. Left panels display results for the least-squares estimator, right panels for the functional estimator. Horizontal axis stands for variance σ_1^2, curves corresponds to various choices of variance σ_2^2 (color of the curves ranges from black to red as σ_2^2 increases).

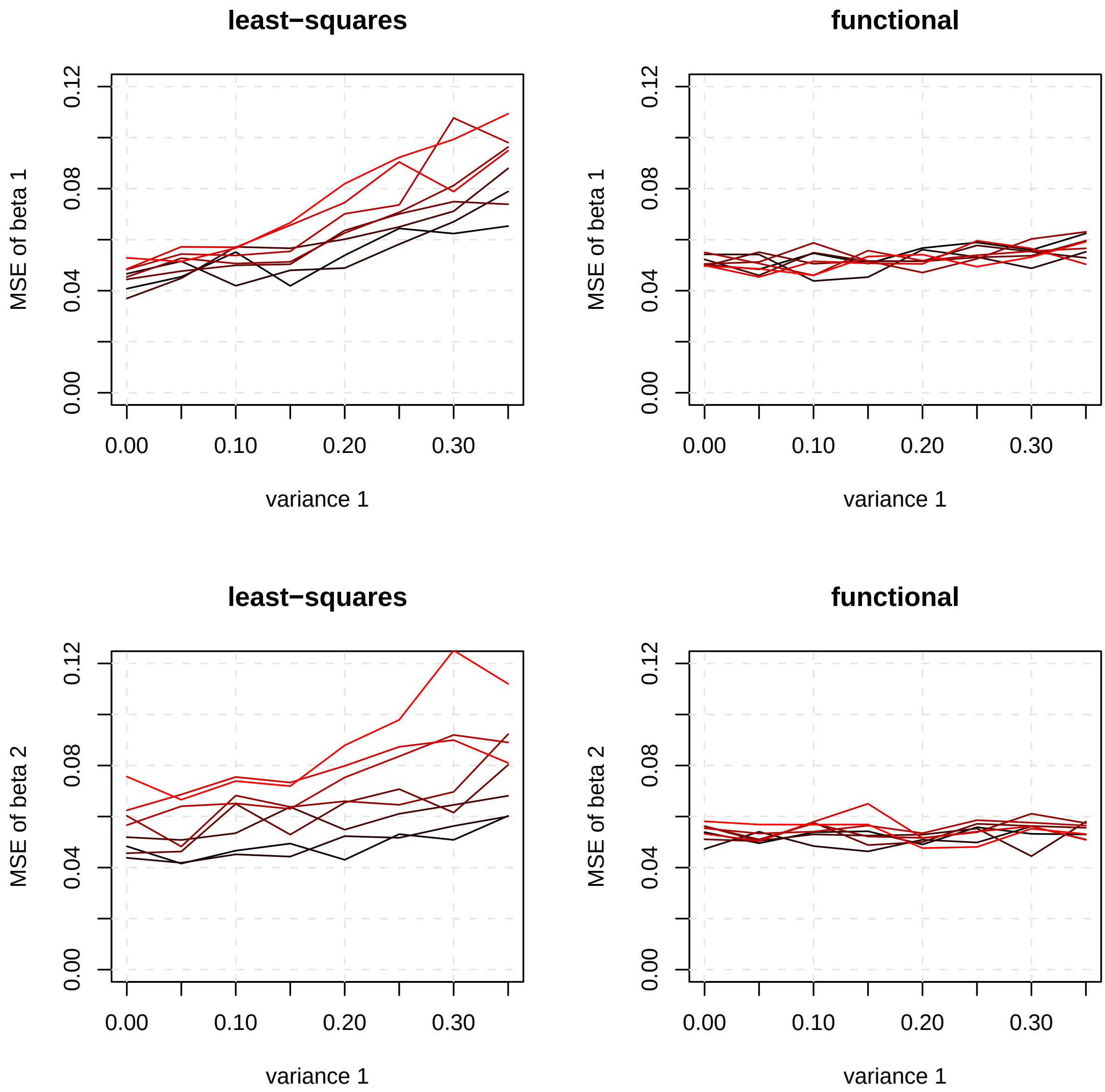

Figure 3.3: MSE of estimators of parameter $\boldsymbol{\beta}$ in simulated RCA(2) model (true parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0.4, 0.2)'$). Upper panels stand for MSE concerning to β_1, lower for β_2. Left panels display results for the least-squares estimator, right panels for the functional estimator. Horizontal axis stands for variance σ_1^2, curves corresponds to various choices of variance σ_2^2 (color of the curves ranges from black to red as σ_2^2 increases).

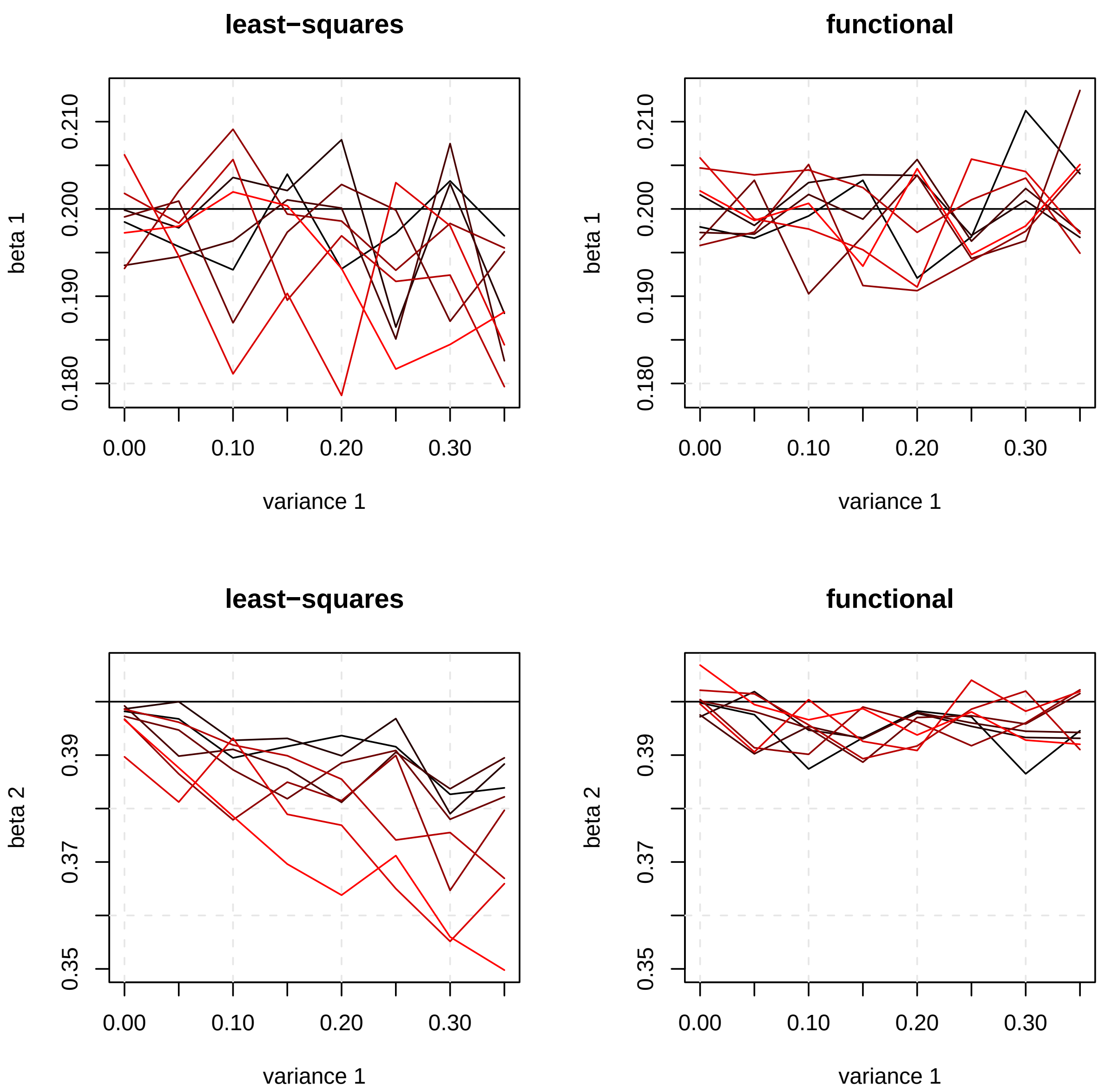

Figure 3.4: Average estimate of true parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0.2, 0.4)'$ in simulated RCA(2) model. Upper panels stand for estimation of β_1, lower for β_2. Left panels display results for the least-squares estimator, right panels for the functional estimator. Horizontal axis stands for variance σ_1^2, curves corresponds to various choices of variance σ_2^2 (color of the curves ranges from black to red as σ_2^2 increases).

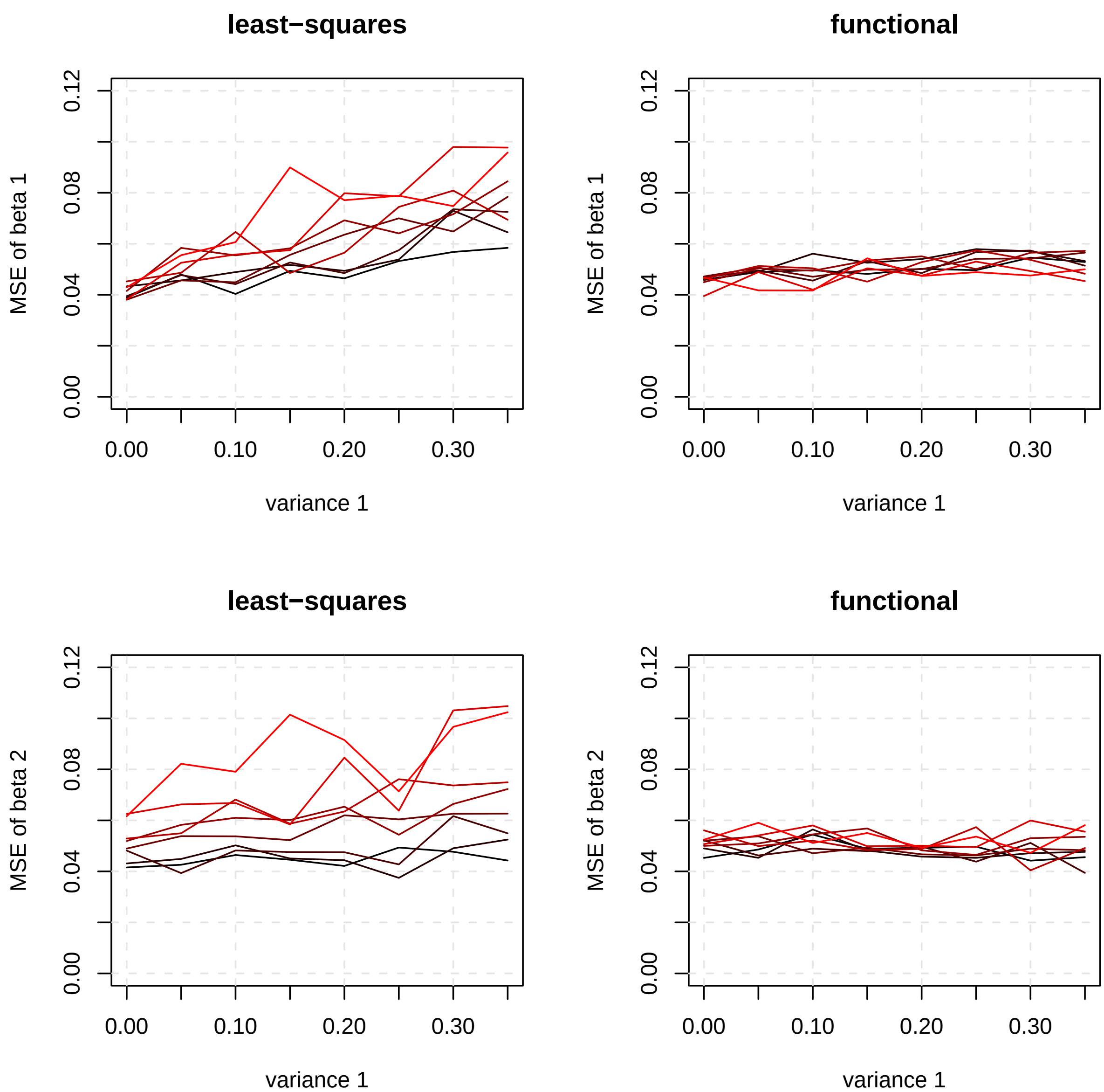

Figure 3.5: MSE of estimators of parameter $\boldsymbol{\beta}$ in simulated RCA(2) model (true parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0.2, 0.4)'$). Upper panels stand for MSE concerning to β_1, lower for β_2. Left panels display results for the least-squares estimator, right panels for the functional estimator. Horizontal axis stands for variance σ_1^2, curves corresponds to various choices of variance σ_2^2 (color of the curves ranges from black to red as σ_2^2 increases).

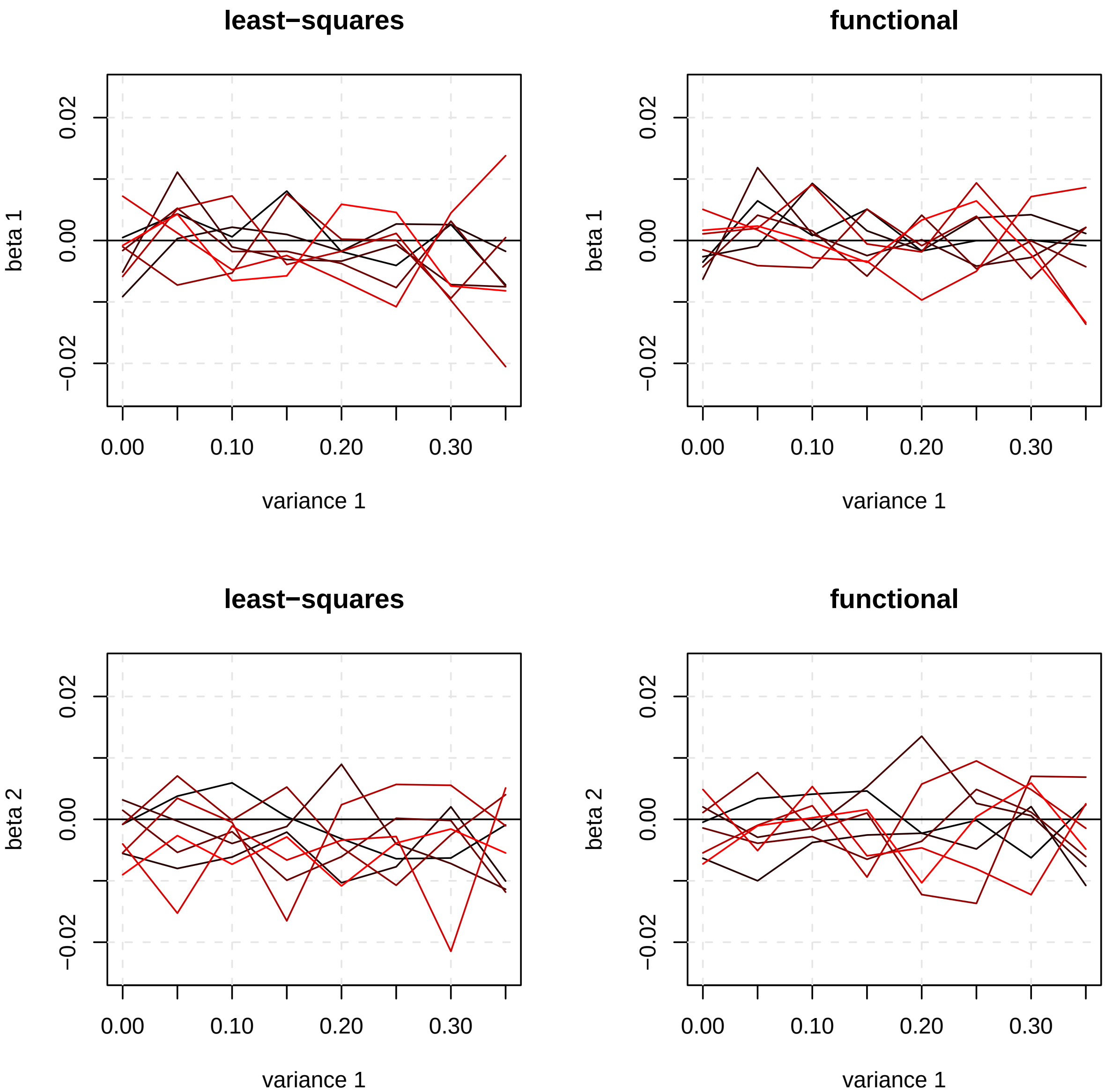

Figure 3.6: Average estimate of true parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0,0)'$ in simulated RCA(2) model. Upper panels stand for estimation of β_1, lower for β_2. Left panels display results for the least-squares estimator, right panels for the functional estimator. Horizontal axis stands for variance σ_1^2, curves corresponds to various choices of variance σ_2^2 (color of the curves ranges from black to red as σ_2^2 increases).

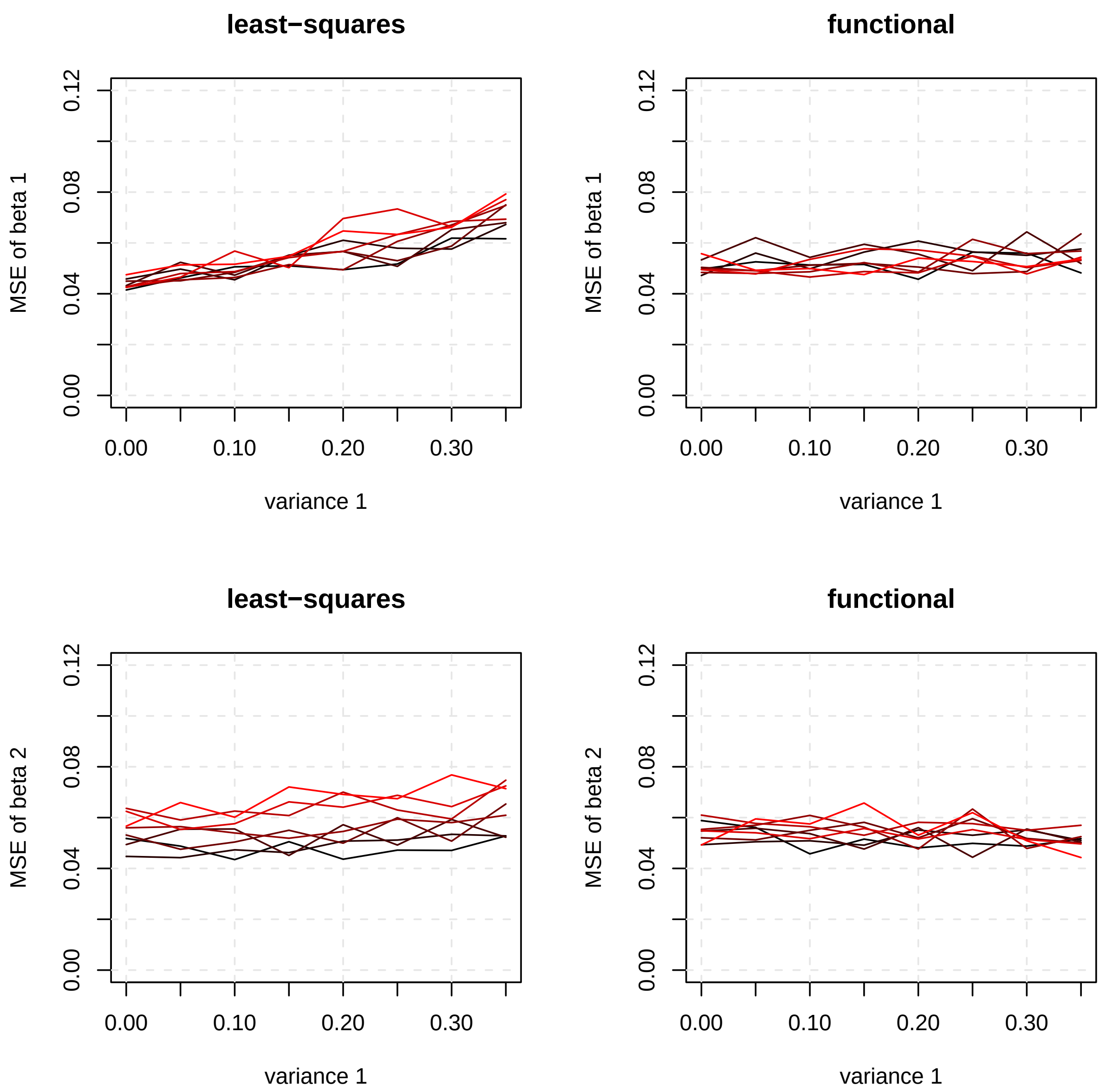

Figure 3.7: MSE of estimators of parameter $\boldsymbol{\beta}$ in simulated RCA(2) model (true parameter $\boldsymbol{\beta} = (\beta_1, \beta_2)' = (0, 0)'$). Upper panels stand for MSE concerning to β_1, lower for β_2. Left panels display results for the least-squares estimator, right panels for the functional estimator. Horizontal axis stands for variance σ_1^2, curves corresponds to various choices of variance σ_2^2 (color of the curves ranges from black to red as σ_2^2 increases).

Chapter 4

Estimators of multivariate models

4.1 Model specification

We will extend the concept of functional estimator to the multivariate RCA(1) models in this chapter, assuming that values of the process belong to $\mathbb{R}^m$ for general $m \in \mathbb{N}$. We will restrict ourselves to the first-order model only, so the results might be compared to univariate RCA(1) models studied in Chapter 2. Similarly to Chapter 3, we will specify the model, briefly describe the usual least-squares estimator and define a class of the functional estimators. We will proof strong consistency and asymptotical normality of such estimators and we will demonstrate their properties via a simple simulation study.

Process $\mathbf{X}_t = (X_t^1, \dots, X_t^m)' \in \mathbb{R}^m$ is called multivariate RCA(1) model of the first order if $\mathbf{X}_t$ for each $t \in \mathbb{Z}$ satisfies

$$\mathbf{X}_t = (\boldsymbol{\beta} + \mathbf{B}_t)\mathbf{X}_{t-1} + \mathbf{Y}_t, \tag{4.1}$$

where $\boldsymbol{\beta}$ is a matrix of unknown parameters, $\{\mathbf{B}_t, t \in \mathbb{Z}\}$ is a sequence of random matrices and $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ is a random vector error process. Let us denote

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \dots & \beta_{1m} \\ \vdots & & \\ \beta_{m1} & \dots & \beta_{mm} \end{pmatrix}, \quad \mathbf{B}_t = \begin{pmatrix} B_t^{11} & \dots & B_t^{1m} \\ \vdots & & \\ B_t^{m1} & \dots & B_t^{mm} \end{pmatrix}, \quad \mathbf{Y}_t = \begin{pmatrix} Y_t^1 \\ \vdots \\ Y_t^m \end{pmatrix}.$$

Equation (4.1) might be rewritten into

$$\mathbf{X}_t = \boldsymbol{\beta}\mathbf{X}_{t-1} + \mathbf{u}_t \tag{4.2}$$

using new vector error process $\mathbf{u}_t = \mathbf{B}_t\mathbf{X}_{t-1} + \mathbf{Y}_t = (\mathbf{X}'_{t-1} \otimes \mathbf{I}) \cdot \mathrm{vec}(\mathbf{B}_t) + \mathbf{Y}_t$ (for the second equation see Lemma 6.8, property *d*) that will play an important role later on.

We assume that error process $\mathbf{Y}_t$ is independent of the *iid* distributed random coefficients $\mathbf{B}_t$. Similarly to the univariate case, we allow the error process to be a stationary martingale difference sequence.

A11: All eigenvalues of matrix $\mathrm{E}(\mathbf{B}_0 \otimes \mathbf{B}_0) + (\boldsymbol{\beta} \otimes \boldsymbol{\beta})$ are less than unity in modulus.

Definition 4.1. ***Generalized multivariate RCA(1) process***
Random vector process $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ is called Generalized multivariate RCA(1) process if $\mathbf{X}_t$ satisfies $\forall t \in \mathbb{Z}$ equation (4.1) and if stationarity condition A11 is met. Matrix process $\{\mathbf{B}_t\}$ is assumed to be a centered *iid* sequence with finite positive definite matrix $\mathbf{\Sigma} = \mathrm{E}(\mathrm{vec}\mathbf{B}_t \cdot \mathrm{vec}'\mathbf{B}_t)$ independent of error process $\{\mathbf{Y}_t\}$ which is an ergodic and strictly stationary martingale difference sequence with respect to $\mathcal{F}_t = \sigma\left(\mathbf{B}_s, \mathbf{Y}_s; s \leq t\right)$ with finite positive definite variance matrix $\mathbf{G}$.

If we denote $\mathbf{\Sigma} = \mathrm{E}(\mathrm{vec}\mathbf{B}_t \cdot \mathrm{vec}'\mathbf{B}_t)$ and assume that $\mathrm{E}[\mathbf{Y}_t\mathbf{Y}_t'|\mathcal{F}_{t-1}] = \mathbf{G}$, we can derive the conditional moments of process $\{\mathbf{X}_t\}$

$$\mathrm{E}[\mathbf{X}_t|\mathcal{F}_{t-1}] = \boldsymbol{\beta}\mathbf{X}_{t-1},$$

$$\begin{aligned}
\mathrm{var}\,[\mathbf{X}_t|\mathcal{F}_{t-1}] &= \mathrm{E}\big[\,(\mathbf{X}_t - \mathrm{E}[\mathbf{X}_t|\mathcal{F}_{t-1}])\,(\mathbf{X}_t - \mathrm{E}[\mathbf{X}_t|\mathcal{F}_{t-1}])'\,|\mathcal{F}_{t-1}\big] = \mathrm{E}[\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}] = \\
&= \mathrm{E}\Big[\big((\mathbf{X}_{t-1}' \otimes \mathbf{I})\mathrm{vec}(\mathbf{B}_t) + \mathbf{Y}_t\big)\,\big((\mathbf{X}_{t-1}' \otimes \mathbf{I})\mathrm{vec}(\mathbf{B}_t) + \mathbf{Y}_t\big)'\,|\mathcal{F}_{t-1}\Big] = \\
&= \mathrm{E}\big[(\mathbf{X}_{t-1}' \otimes \mathbf{I})\mathrm{vec}(\mathbf{B}_t)\mathrm{vec}'(\mathbf{B}_t)(\mathbf{X}_{t-1} \otimes \mathbf{I})|\mathcal{F}_{t-1}\big] + \mathrm{E}[\mathbf{Y}_t\mathbf{Y}_t'|\mathcal{F}_{t-1}] = \\
&= (\mathbf{X}_{t-1}' \otimes \mathbf{I}) \cdot \mathbf{\Sigma} \cdot (\mathbf{X}_{t-1} \otimes \mathbf{I}) + \mathbf{G}.
\end{aligned}$$

The conditional variance is a $m \times m$ matrix and sometimes it is useful to express it as a vector. Using Lemma 6.8 from Chapter 6, matrix identity c, we have

$$\mathrm{vec}\,(\,\mathrm{var}\,[\mathbf{X}_t|\mathcal{F}_{t-1}]) = \big[(\mathbf{X}_{t-1}' \otimes \mathbf{I}) \otimes (\mathbf{X}_{t-1}' \otimes \mathbf{I})\big] \cdot \mathrm{vec}(\mathbf{\Sigma}) + \mathrm{vec}(\mathbf{G})$$

which could be transformed into

$$\begin{aligned}
\mathrm{vech}\,(\,\mathrm{var}\,[\mathbf{X}_t|\mathcal{F}_{t-1}]) &= \mathrm{vech}\,(\mathrm{E}[\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}]) = \mathbf{H}_m\mathrm{vec}\,(\,\mathrm{var}\,[\mathbf{X}_t|\mathcal{F}_{t-1}]) = \\
&= \mathbf{H}_m\,\big[(\mathbf{X}_{t-1}' \otimes \mathbf{I}) \otimes (\mathbf{X}_{t-1}' \otimes \mathbf{I}) \cdot \mathbf{K}_{m^2}'\mathrm{vech}(\mathbf{\Sigma}) + \mathbf{K}_m'\mathrm{vech}(\mathbf{G})\big] = \\
&= \mathbf{H}_m(\mathbf{X}_{t-1}' \otimes \mathbf{I}) \otimes (\mathbf{X}_{t-1}' \otimes \mathbf{I})\mathbf{K}_{m^2}' \cdot \mathrm{vech}(\mathbf{\Sigma}) + \mathrm{vech}(\mathbf{G}), \qquad (4.3)
\end{aligned}$$

where $\mathbf{H}_m$ and $\mathbf{K}_{m^2}$ are the matrices from Lemma 6.9 in Chapter 6.

The following theorem proves that there exists a unique strictly stationary solution of the stochastic difference equation (4.1). The solution is measurable with respect to $\mathcal{F}_t = \sigma\left((\mathbf{B}_s', \mathbf{Y}_s)'; s \leq t\right)$, and is of the form

$$\mathbf{X}_t = \sum_{j=0}^{+\infty} \left[\prod_{i=0}^{j-1}(\boldsymbol{\beta} + \mathbf{B}_{t-i})\right] \cdot \mathbf{Y}_{t-j}, \qquad (4.4)$$

where the product is defined as 1 for $j = 0$.

Theorem 4.1.
There exists the Generalized multivariate RCA(1) process according to Definition 4.1 which is $\mathcal{F}_t$-measurable, ergodic and strictly stationary.

Proof: The existence of RCA process and its properties are profoundly discussed in [25], Section 2.3. However, the definition of RCA process stated there differs in the assumption that $\mathbf{Y}_t$ are supposed to be independent. We will benefit from the proofs given there and we will show that the crucial steps remain valid also for the RCA process according to our definition.

Iterating equation (4.1) we get

$$\begin{aligned}\mathbf{X}_t &= \mathbf{Y}_t + (\boldsymbol{\beta} + \mathbf{B}_t)\left[\mathbf{Y}_{t-1} + (\boldsymbol{\beta} + \mathbf{B}_{t-1})\mathbf{X}_{t-2}\right] = \\ &= \mathbf{Y}_t + (\boldsymbol{\beta} + \mathbf{B}_t)\mathbf{Y}_{t-1} + (\boldsymbol{\beta} + \mathbf{B}_t)(\boldsymbol{\beta} + \mathbf{B}_{t-1})\mathbf{X}_{t-2} = \ldots = \\ &= \sum_{j=0}^{r} \mathbf{S}_{t,j-1} \cdot \mathbf{Y}_{t-j} + \mathbf{S}_{t,r} \cdot \mathbf{X}_{t-r-1}, \end{aligned} \tag{4.5}$$

where $\mathbf{S}_{t,r} = \prod_{k=0}^{r} (\boldsymbol{\beta} + \mathbf{B}_{t-k})$ for $r = 0, 1, \ldots$ and $\mathbf{S}_{t,-1} = \mathbf{I}$. Denote $\mathbf{W}_{t,r} = \mathbf{X}_t - \mathbf{S}_{t,r} \cdot \mathbf{X}_{t-r-1}$ and if we knew that the sum in (4.5) converges, we could continue with the iteration into infinity when the second term would diminish to zero and we would obtain the solution given by equation (4.4) which is equal to the limit of $\mathbf{W}_{t,r}$ as r goes to infinity.

Now we will compute the variance matrix of $\mathbf{W}_{t,r}$ in the form of *vec* operator. Noticing that

$$\mathrm{E}\left[\mathrm{vec}\left(\mathbf{S}_{t,j-1}\mathbf{Y}_{t-j}\mathbf{Y}'_{t-i}\mathbf{S}'_{t,i-1}\right)\right] = \mathrm{E}\left[(\mathbf{S}_{t,i-1} \otimes \mathbf{S}_{t,j-1})\right] \cdot \mathrm{E}\left[\mathrm{vec}(\mathbf{Y}_{t-j}\mathbf{Y}'_{t-i})\right] = \mathbf{0}$$

for $i \neq j$ where we used matrix identity c from Lemma 6.8 in Chapter 6 and the independence of $\mathbf{S}_{t,r}$ and $\mathbf{Y}_u$ for any time $u \in \mathbb{Z}$, we have

$$\begin{aligned}\mathrm{vec}\left(\mathrm{E}\left[\mathbf{W}_{t,r}\mathbf{W}'_{t,r}\right]\right) &= \mathrm{vec}\left(\mathrm{E}\left[\sum_{j=0}^{r} \mathbf{S}_{t,j-1}\mathbf{Y}_{t-j}\right]\left[\sum_{i=0}^{r} \mathbf{S}_{t,i-1}\mathbf{Y}_{t-i}\right]'\right) = \\ &= \mathrm{vec}\left(\mathrm{E}\left[\sum_{j=0}^{r} \mathbf{S}_{t,j-1}\mathbf{Y}_{t-j}\mathbf{Y}'_{t-j}\mathbf{S}'_{t,j-1}\right]\right) = \\ &= \sum_{j=0}^{r} \mathrm{E}\left(\mathbf{S}_{t,j-1} \otimes \mathbf{S}_{t,j-1}\right) \cdot \mathrm{E}\left[\mathrm{vec}\left(\mathbf{Y}_{t-j}\mathbf{Y}'_{t-j}\right)\right] = \\ &= \sum_{j=0}^{r} \mathrm{E}\left(\prod_{k=0}^{j-1}(\boldsymbol{\beta} + \mathbf{B}_{t-k}) \otimes \prod_{l=0}^{j-1}(\boldsymbol{\beta} + \mathbf{B}_{t-l})\right) \cdot \mathrm{vec}\left(\mathrm{E}\left[\mathbf{Y}_{t-j}\mathbf{Y}'_{t-j}\right]\right) = \\ &= \sum_{j=0}^{r} \mathrm{E}\left(\prod_{k=0}^{j-1}(\boldsymbol{\beta} + \mathbf{B}_{t-k}) \otimes (\boldsymbol{\beta} + \mathbf{B}_{t-k})\right) \cdot \mathrm{vec}(\mathbf{G}) = \\ &= \sum_{j=0}^{r} \prod_{k=0}^{j-1} \left(\boldsymbol{\beta} \otimes \boldsymbol{\beta} + \mathrm{E}\left(\mathbf{B}_t \otimes \mathbf{B}_t\right)\right) \cdot \mathrm{vec}(\mathbf{G}) = \\ &= \sum_{j=0}^{r} \left(\mathrm{E}\left(\mathbf{B}_0 \otimes \mathbf{B}_0\right) + \boldsymbol{\beta} \otimes \boldsymbol{\beta}\right)^j \cdot \mathrm{vec}(\mathbf{G}). \end{aligned} \tag{4.6}$$

The condition under which the sum in equation (4.6) converges is equal to the stationarity assumption A11. That is sequentially proved in Theorem 2.2 and Corollary 2.2.1 in [25], all of which derivations remain valid under our assumption on RCA process. The uniqueness of the process $\{\mathbf{X}_t\}$ is discussed in Corollary 2.2.2, stationarity and ergodicity is proved in Theorem 2.7 in [25]. □

4.2 Functional estimator

The least-squares estimator of multivariate RCA(1) model according to Definition 4.1 is defined (see [25], Section 7.2) as

$$\widehat{\boldsymbol{\beta}}_{LS} = \left(\sum_{t=1}^{n} \mathbf{X}_t \mathbf{X}'_{t-1}\right) \cdot \left(\sum_{t=1}^{n} \mathbf{X}_{t-1} \mathbf{X}'_{t-1}\right)^{-1}. \tag{4.7}$$

There also has been proved its strong consistency, which in this case means that

$$\operatorname{vec}\left(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}\right) \xrightarrow{a.s.} \mathbf{0}$$

and asymptotical normality of $\sqrt{n} \cdot \operatorname{vec}\left(\widehat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\beta}\right)$.

The other parameters of the model, variance matrices Σ and $\mathbf{G}$, can be estimated as follows: Equation (4.3) gives the expression of the conditional mean of $\operatorname{vech}(\mathbf{u}_t\mathbf{u}'_t)$. If we compute the estimated residuals $\widehat{\mathbf{u}}_t = \mathbf{X}_t - \widehat{\boldsymbol{\beta}}_{LS}\mathbf{X}_{t-1}$, Nicholls and Quinn in [25] suggested to estimate parameters $\boldsymbol{\Sigma}$ and $\mathbf{G}$ by regressing $\operatorname{vech}(\mathbf{u}_t\mathbf{u}'_t)$ on

$$(\mathbf{1}|\mathbf{A}'_{t-1}) = (\mathbf{1}|\mathbf{H}_m(\mathbf{X}'_{t-1} \otimes \mathbf{I}) \otimes (\mathbf{X}'_{t-1} \otimes \mathbf{I})\mathbf{K}'_{m^2})$$

with the regression parameters $(\operatorname{vech}(\mathbf{G})', \operatorname{vech}(\boldsymbol{\Sigma}))'$. The solution of the regression equation is equal to

$$\begin{aligned}
\operatorname{vech}(\widehat{\boldsymbol{\Sigma}}) &= \left(\sum_{t=1}^{n} \left(\mathbf{A}_{t-1} - \overline{\mathbf{A}}\right)\left(\mathbf{A}_{t-1} - \overline{\mathbf{A}}\right)'\right)^{-1} \left(\sum_{t=1}^{n} \left(\mathbf{A}_{t-1} - \overline{\mathbf{A}}\right) \cdot \operatorname{vech}\left(\mathbf{u}_t\mathbf{u}'_t\right)\right), \\
\operatorname{vech}(\widehat{\mathbf{G}}) &= \frac{1}{n}\sum_{t=1}^{n} \operatorname{vech}\left(\mathbf{u}_t\mathbf{u}'_t\right) - \overline{\mathbf{A}}' \cdot \operatorname{vech}(\widehat{\boldsymbol{\Sigma}}),
\end{aligned} \tag{4.8}$$

where $\overline{\mathbf{A}} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{A}_{t-1}$. In [25], Theorem 7.2, there is proved that under higher moment conditions of process $\{\mathbf{X}_t\}$ such estimators are strongly consistent (one requires finite 4^{th} moments of the components of vector $\mathbf{X}_t$) and asymptotically normal (finite 8^{th} moments).

Likewise in the extension of the functional estimator to univariate higher-order RCA(p) models, we propose an extension of the least-squares estimator into a broad class of estimators using measurable function $\phi : \mathbb{R}^m \to \mathbb{R}^m$. Thus, let us define estimator

$$\widehat{\boldsymbol{\beta}}(\phi) = \left(\sum_{t=1}^{n} \mathbf{X}_t \phi(\mathbf{X}_{t-1})'\right) \cdot \left(\sum_{t=1}^{n} \mathbf{X}_{t-1} \phi(\mathbf{X}_{t-1})'\right)^{-1}. \tag{4.9}$$

We published the proof of the following theorem for a special multivariate RCA model with *iid* errors (see [36]). Since then we have extended the statement to the generalized RCA models.

Theorem 4.2.
Consider Generalized multivariate RCA(1) process according to Definition 4.1.
Let $\mathrm{E}[\mathbf{Y}_t\mathbf{Y}_t'|\mathcal{F}_{t-1}] = \mathbf{G}$ *a.s for all* t *and let* $\phi : \mathbb{R}^m \to \mathbb{R}^m$ *be measurable function such that matrix* $\mathrm{E}\phi(\mathbf{X}_0)\mathbf{X}_0'$ *is finite and positive definite. Denote*

$$\mathbf{V}(\phi) = (\mathrm{E}\mathbf{P})^{-1}\cdot \mathrm{E}\left(\mathbf{Q}\mathbf{G}\mathbf{Q}'\right)\cdot(\mathrm{E}\mathbf{P}')^{-1} + (\mathrm{E}\mathbf{P})^{-1}\cdot \mathrm{E}\left(\mathbf{P}\boldsymbol{\Sigma}\mathbf{P}'\right)\cdot(\mathrm{E}\mathbf{P}')^{-1}, \qquad (4.10)$$

where $\mathbf{P} = \phi(\mathbf{X}_0)\mathbf{X}_0'\otimes\mathbf{I}$ *and* $\mathbf{Q} = \phi(\mathbf{X}_0)\otimes\mathbf{I}$.
Then estimator $\widehat{\boldsymbol{\beta}}(\phi)$ *defined by* (4.9) *is a strongly consistent and asymptotically normal estimator of parameter* $\boldsymbol{\beta}$ *with asymptotical variance matrix defined by* (4.10).

Proof: According to Definition (4.9) of functional estimator $\widehat{\boldsymbol{\beta}}(\phi)$ we know that

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta} &= \left(\sum_{t=1}^n \mathbf{X}_t\phi(\mathbf{X}_{t-1})' - \boldsymbol{\beta}\sum_{t=1}^n \mathbf{X}_{t-1}\phi(\mathbf{X}_{t-1})'\right)\cdot\left(\sum_{t=1}^n \mathbf{X}_{t-1}\phi(\mathbf{X}_{t-1})'\right)^{-1} = \\ &= \left(\sum_{t=1}^n \mathbf{u}_t\phi(\mathbf{X}_{t-1})'\right)\cdot\left(\sum_{t=1}^n \mathbf{X}_{t-1}\phi(\mathbf{X}_{t-1})'\right)^{-1}.\end{aligned}$$

Now using Lemma 6.8 from Chapter 6, matrix identities f and g, we have

$$\operatorname{vec}(\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta}) = \left[\left(\frac{1}{n}\sum_{t=1}^n \phi(\mathbf{X}_{t-1})\mathbf{X}_{t-1}'\right)^{-1}\otimes\mathbf{I}\right]\cdot\left[\frac{1}{n}\sum_{t=1}^n\left(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I}\right)\mathbf{u}_t\right]. \qquad (4.11)$$

Strict stationarity and ergodicity of $\mathbf{X}_t$ guarantees strict stationarity and ergodicity of both sequences $\{\phi(\mathbf{X}_{t-1})\mathbf{X}_{t-1}', t\in\mathbb{Z}\}$ and $\{(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\mathbf{u}_t, t\in\mathbb{Z}\}$. Moreover, components of the latter sequence form a martingale difference sequence with zero mean value, which can be seen by choosing arbitrary $\alpha\in\mathbb{R}^{m^2}$ and noticing that

$$\mathrm{E}[\alpha'(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\mathbf{u}_t|\mathcal{F}_{t-1}] = \alpha'(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\mathrm{E}[\mathbf{u}_t|\mathcal{F}_{t-1}] = 0.$$

Ergodic theorem (see e.g. [10], Theorem 13.12) tells us that the first term in (4.11) converges almost surely to $(\mathrm{E}\phi(\mathbf{X}_0)\mathbf{X}_0')^{-1}\otimes\mathbf{I}$ and the second term converges to zero, which implies that $\operatorname{vec}(\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta}) \to \mathbf{0}$ as $n\to+\infty$.

Proof of the asymptotical normality requires little bit more computation:

$$\begin{aligned}\operatorname{var}\left(\alpha'(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\mathbf{u}_t\right) &= \mathrm{E}\left(\mathrm{E}\left[\left(\alpha'(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\mathbf{u}_t\right)^2|\mathcal{F}_{t-1}\right]\right) = \\ &= \mathrm{E}\left(\mathrm{E}[\alpha'(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\mathbf{u}_t\mathbf{u}_t'(\phi(\mathbf{X}_{t-1})'\otimes\mathbf{I})\alpha'|\mathcal{F}_{t-1}]\right) = \\ &= \mathrm{E}\left(\alpha'(\phi(\mathbf{X}_{t-1})\otimes\mathbf{I})\cdot\mathrm{E}[\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}]\cdot(\phi(\mathbf{X}_{t-1})'\otimes\mathbf{I})\alpha'\right) =\end{aligned}$$

$$\begin{aligned}
&= \mathrm{E}\left(\alpha'(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}) \cdot \left[(\mathbf{X}'_{t-1} \otimes \mathbf{I})\boldsymbol{\Sigma}(\mathbf{X}_{t-1} \otimes \mathbf{I}) + \mathbf{G}\right] \cdot (\phi(\mathbf{X}_{t-1})' \otimes \mathbf{I})\alpha'\right) = \\
&= \alpha'\mathrm{E}\left((\phi(\mathbf{X}_{t-1})\mathbf{X}'_{t-1} \otimes \mathbf{I})\boldsymbol{\Sigma}(\mathbf{X}_{t-1}\phi(\mathbf{X}_{t-1})' \otimes \mathbf{I})\right)\alpha + \\
&\quad + \alpha'\mathrm{E}\left((\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I})\mathbf{G}(\phi(\mathbf{X}_{t-1})' \otimes \mathbf{I})\right)\alpha. \qquad (4.12)
\end{aligned}$$

Similarly to equation (4.11) notice that

$$\sqrt{n} \cdot \mathrm{vec}(\widehat{\boldsymbol{\beta}}(\phi) - \boldsymbol{\beta}) = \left[\left(\frac{1}{n}\sum_{t=1}^{n} \phi(\mathbf{X}_{t-1})\mathbf{X}'_{t-1}\right)^{-1} \otimes \mathbf{I}\right] \cdot \left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \left(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\right)\mathbf{u}_t\right],$$

where the first term converges to $(\mathrm{E}\phi(\mathbf{X}_0)\mathbf{X}'_0)^{-1} \otimes \mathbf{I}$ and the second term has due to Lindeberg-Levy theorem for martingales (see [7]) asymptotically m-dimensional normal distribution. If we introduce expressions $\mathbf{P}$ and $\mathbf{Q}$ according to Theorem 4.2, the variance matrix $\mathbf{V}(\phi)$ can easily be derived from (4.12). □

4.2.1 Asymptotical variance matrix

Variance matrix $\mathbf{V}(\phi)$ defined by equation (4.10) is very complicated, so we suggest a consistent estimator of the asymptotical variance matrix. Firstly, we need an auxiliary proposition:

Theorem 4.3.
Let $(\Omega, \mathcal{A}, \mathrm{P})$ denote probability space. Let $\{\mathbf{A}_n, n \in \mathbb{N}\}$ and $\{\mathbf{B}_n, n \in \mathbb{N}\}$ be random $r \times r$-dimensional matrix processes. Assume that $\{\mathbf{B}_n\}$ is ergodic with finite second moments and that $\mathbf{A}_n \xrightarrow{a.s.} \mathbf{A}$ as $n \to +\infty$ where $\mathbf{A} \in \mathbb{R}^{r\times r}$ is a finite constant matrix. Then

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{B}_t\mathbf{A}_n\mathbf{B}'_t \xrightarrow{a.s.} \mathrm{E}(\mathbf{B}_1\mathbf{A}\mathbf{B}'_1) \text{ as } n \to +\infty.$$

Proof: We will prove the convergence of the correspondent components using ergodicity and the vec operator (see Chapter 6 for its properties):

$$\begin{aligned}
&\mathrm{vec}\left(\frac{1}{n}\sum_{t=1}^{n} \mathbf{B}_t\mathbf{A}_n\mathbf{B}'_t\right) = \frac{1}{n}\sum_{t=1}^{n} \mathrm{vec}\left(\mathbf{B}_t\mathbf{A}_n\mathbf{B}'_t\right) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{B}_t \otimes \mathbf{B}_t \cdot \mathrm{vec}(\mathbf{A}_n) \xrightarrow{a.s.} \\
&\xrightarrow{a.s.} \mathrm{E}\left(\mathbf{B}_1 \otimes \mathbf{B}_1\right) \cdot \mathrm{vec}(\mathbf{A}) = \mathrm{E}\left(\mathrm{vec}\left(\mathbf{B}_1\mathbf{A}\mathbf{B}'_1\right)\right) = \mathrm{vec}\left(\mathrm{E}\left(\mathbf{B}_1\mathbf{A}\mathbf{B}'_1\right)\right).
\end{aligned}$$

□

Now let us formulate and prove the theorem concerning the consistent estimator:

Theorem 4.4.
Consider multivariate RCA(1) model according to Definition 4.1.

Let $\phi : \mathbb{R}^m \to \mathbb{R}^m$ be measurable function such that matrix $\mathrm{E}\phi(\mathbf{X}_0)\mathbf{X}_0'$ is finite and positive definite. Let $\widehat{\mathbf{G}}_n$ and $\widehat{\mathbf{\Sigma}}_n$ be strongly consistent estimators of $\mathbf{G}$ and $\mathbf{\Sigma}$, respectively. Denote $\mathbf{P}_t = \phi(\mathbf{X}_{t-1})\mathbf{X}_{t-1}' \otimes \mathbf{I}$ and $\mathbf{Q}_t = \phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}$.
Then

$$\widehat{\mathbf{V}}_n(\phi) = n\left(\sum_{t=1}^{n}\mathbf{P}_t\right)^{-1} \cdot \sum_{t=1}^{n}\left(\mathbf{Q}_t\widehat{\mathbf{G}}_n\mathbf{Q}_t'\right) \cdot \left(\sum_{t=1}^{n}\mathbf{P}_t'\right)^{-1} + \\ + n\left(\sum_{t=1}^{n}\mathbf{P}_t\right)^{-1} \cdot \sum_{t=1}^{n}\left(\mathbf{P}_t\widehat{\mathbf{\Sigma}}_n\mathbf{P}_t'\right) \cdot \left(\sum_{t=1}^{n}\mathbf{P}_t'\right)^{-1}$$

is a strongly consistent estimator of asymptotical variance matrix $\mathbf{V}(\phi)$ given by (4.10).

Proof: Process $\{\mathbf{X}_t, t \in \mathbf{Z}\}$ is strictly stationary and ergodic which implies that $\{\mathbf{P}_t, t \in \mathbf{Z}\}$, $\{\mathbf{Q}_t, t \in \mathbf{Z}\}$, $\{\mathbf{P}_t\mathbf{\Sigma}\mathbf{P}_t', t \in \mathbf{Z}\}$, and $\{\mathbf{Q}_t\mathbf{G}\mathbf{Q}_t', t \in \mathbf{Z}\}$ are strictly stationary and ergodic. The ergodic theorem tells us that $\frac{1}{n}\sum_{t=1}^{n}\mathbf{P}_t \xrightarrow{a.s.} \mathrm{E}\mathbf{P}$ and $\frac{1}{n}\sum_{t=1}^{n}\mathbf{Q}_t \xrightarrow{a.s.} \mathrm{E}\mathbf{Q}$ as $n \to +\infty$, where $\mathbf{P}$ and $\mathbf{Q}$ are defined in Theorem 4.2. Consistency of estimators $\widehat{\mathbf{G}}_n$, $\widehat{\mathbf{\Sigma}}_n$ and Theorem 4.3 completes the proof. □

4.2.2 Lower bound for variance matrix

Similarly to the previous chapters, we are interested in the optimal choice of generating function ϕ for estimator $\widehat{\boldsymbol{\beta}}(\phi)$. We will employ the same techniques described in Sections 2.4.1 and 3.2.2, so we will omit unnecessary details. The only difference this time is that we are not able to identify the optimal estimator (optimal in the sense of asymptotical variance matrix). Thus, we will infer the lower bound for asymptotical variance matrix of the estimators.

Let $\mathrm{E}[\mathbf{Y}_t\mathbf{Y}_t'|\mathcal{F}_{t-1}] = \mathbf{G}$ and define two m^2-dimensional random vectors

$$\mathbf{T}_1 = \sum_{t=1}^{n}\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot (\mathbf{X}_t - \boldsymbol{\beta}'\mathbf{X}_{t-1}) = \sum_{t=1}^{n}\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot \mathbf{u}_t,$$

$$\mathbf{T}_2 = \sum_{t=1}^{n}\big(\mathbf{X}_{t-1} \otimes \mathbf{I}\big) \cdot \mathbf{w}(\mathbf{X}_{t-1})^{-1} \cdot (\mathbf{X}_t - \boldsymbol{\beta}'\mathbf{X}_{t-1}) = \sum_{t=1}^{n}\big(\mathbf{X}_{t-1} \otimes \mathbf{I}\big) \cdot \mathbf{w}(\mathbf{X}_{t-1})^{-1} \cdot \mathbf{u}_t,$$

where $\mathbf{w}(\mathbf{z}) = (\mathbf{z}' \otimes \mathbf{I}) \cdot \Sigma \cdot (\mathbf{z} \otimes \mathbf{I}) + \mathbf{G}$. Since both sequences $\{\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot \mathbf{u}_t\}$ and $\{\big(\mathbf{X}_{t-1} \otimes \mathbf{I}\big) \cdot \mathbf{w}(\mathbf{X}_{t-1})^{-1} \cdot \mathbf{u}_t\}$ are martingale differences w.r to $\mathcal{F}_t$, it immediately follows that $\mathrm{E}\mathbf{T}_1 = 0$ and $\mathrm{E}\mathbf{T}_2 = 0$ and the variance matrix of vector $\mathbf{T}_1$ equals

$$\mathrm{E}\mathbf{T}_1\mathbf{T}_1' = \mathrm{E}\left(\sum_{t=1}^{n}\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big)\mathbf{u}_t\right) \cdot \left(\sum_{s=1}^{n}\mathbf{u}_s'\big(\phi(\mathbf{X}_{t-1})' \otimes \mathbf{I}\big)\right) = \\ = \sum_{t=1}^{n}\mathrm{E}\left(\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big)\mathbf{u}_t\mathbf{u}_t'\big(\phi(\mathbf{X}_{t-1})' \otimes \mathbf{I}\big)\right) =$$

$$
\begin{aligned}
&= \sum_{t=1}^{n} \mathrm{E}\left(\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot \mathrm{E}[\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}] \cdot \big(\phi(\mathbf{X}_{t-1})' \otimes \mathbf{I}\big)\right) = \\
&= n \cdot \mathrm{E}\left(\big(\phi(\mathbf{X}_0) \otimes \mathbf{I}\big) \cdot \mathbf{w}(\mathbf{X}_0) \cdot \big(\phi(\mathbf{X}_0)' \otimes \mathbf{I}\big)\right) = \\
&= n \cdot \mathrm{E}\left(\big(\phi(\mathbf{X}_0) \otimes \mathbf{I}\big) \cdot ((\mathbf{X}_0' \otimes \mathbf{I}) \cdot \Sigma \cdot (\mathbf{X}_0 \otimes \mathbf{I}) + \mathbf{G}) \cdot \big(\phi(\mathbf{X}_0)' \otimes \mathbf{I}\big)\right) = \\
&= n \cdot \mathrm{E}\left(\big(\phi(\mathbf{X}_0)\mathbf{X}_0' \otimes \mathbf{I}\big)\, \Sigma\, \big(\mathbf{X}_0\phi(\mathbf{X}_0)' \otimes \mathbf{I}\big) + \big(\phi(\mathbf{X}_0) \otimes \mathbf{I}\big)\, \mathbf{G}\, \big(\phi(\mathbf{X}_0)' \otimes \mathbf{I}\big)\right) = \\
&= n \cdot \mathrm{E}\left(\mathbf{P}\Sigma\mathbf{P}' + \mathbf{Q}\mathbf{G}\mathbf{Q}'\right) \qquad (4.13)
\end{aligned}
$$

using matrix property from Lemma 6.7 and the notation defined in Theorem 4.2. As the direct analogy we can infer that

$$
\mathrm{E}\mathbf{T}_2\mathbf{T}_2' = n \cdot \mathrm{E}\left(\big(\mathbf{X}_0 \otimes \mathbf{I}\big) \cdot \mathbf{w}(\mathbf{X}_0)^{-1} \cdot \big(\mathbf{X}_0' \otimes \mathbf{I}\big)\right). \qquad (4.14)
$$

The cross-covariance matrix of $\mathbf{T}_1$ and $\mathbf{T}_2$ could be computed as follows:

$$
\begin{aligned}
\mathrm{E}\mathbf{T}_1\mathbf{T}_2' &= \mathrm{E}\left(\left(\sum_{t=1}^{n} \big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot \mathbf{u}_t\right) \cdot \left(\sum_{s=1}^{n} \mathbf{u}_s' \cdot \mathbf{w}(\mathbf{X}_{s-1})^{-1} \cdot \big(\mathbf{X}_{s-1}' \otimes \mathbf{I}\big)\right)\right) = \\
&= \sum_{t=1}^{n} \mathrm{E}\left(\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot \mathrm{E}[\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}] \cdot \mathbf{w}(\mathbf{X}_{t-1})^{-1} \cdot \big(\mathbf{X}_{t-1}' \otimes \mathbf{I}\big)\right) = \\
&= \sum_{t=1}^{n} \mathrm{E}\left(\big(\phi(\mathbf{X}_{t-1}) \otimes \mathbf{I}\big) \cdot \mathbf{w}(\mathbf{X}_{t-1}) \cdot \mathbf{w}(\mathbf{X}_{t-1})^{-1} \cdot \big(\mathbf{X}_{t-1}' \otimes \mathbf{I}\big)\right) = \\
&= n \cdot \mathrm{E}\left(\big(\phi(\mathbf{X}_0) \otimes \mathbf{I}\big) \cdot \big(\mathbf{X}_0' \otimes \mathbf{I}\big)\right) = n \cdot \mathrm{E}\left(\phi(\mathbf{X}_0)\mathbf{X}_0' \otimes \mathbf{I}\right) = n \cdot \mathrm{E}\mathbf{P}. \qquad (4.15)
\end{aligned}
$$

Variance matrix of $2m^2$-dimensional random vector $(\mathbf{T}_1', \mathbf{T}_2')'$ is equal to the block-matrix

$$
\begin{pmatrix} \mathrm{E}\mathbf{T}_1\mathbf{T}_1', & \mathrm{E}\mathbf{T}_1\mathbf{T}_2' \\ \mathrm{E}\mathbf{T}_2\mathbf{T}_1', & \mathrm{E}\mathbf{T}_2\mathbf{T}_2' \end{pmatrix}
$$

where the elements were computed in (4.13) - (4.15). Theorem 3.3 tells us that if the block elements $\mathrm{E}\mathbf{T}_1\mathbf{T}_2'$ and $\mathrm{E}\mathbf{T}_2\mathbf{T}_2'$ are regular matrices, we have

$$
\begin{aligned}
&\left(\mathrm{E}\mathbf{T}_1\mathbf{T}_2'\right)^{-1} \cdot (\mathrm{E}\mathbf{T}_1\mathbf{T}_1') \cdot (\mathrm{E}\mathbf{T}_2\mathbf{T}_1')^{-1} - (\mathrm{E}\mathbf{T}_2\mathbf{T}_2')^{-1} \geq \mathbf{0} \\
\Longleftrightarrow\ &(\mathrm{E}\mathbf{P})^{-1} \cdot \mathrm{E}\left(\mathbf{P}\Sigma\mathbf{P}' + \mathbf{Q}\mathbf{G}\mathbf{Q}'\right) \cdot (\mathrm{E}\mathbf{P}')^{-1} - (\mathrm{E}\mathbf{T}_2\mathbf{T}_2')^{-1} \geq \mathbf{0} \\
\Longleftrightarrow\ &\mathbf{V}(\phi) - \left(\mathrm{E}\left((\mathbf{X}_0 \otimes \mathbf{I}) \cdot \mathbf{w}(\mathbf{X}_0)^{-1} \cdot (\mathbf{X}_0' \otimes \mathbf{I})\right)\right)^{-1} \geq \mathbf{0},
\end{aligned}
$$

where $\mathbf{V}(\phi)$ is the asymptotical variance matrix of general estimator $\widehat{\boldsymbol{\beta}}(\phi)$. Thus

$$
\left(\mathrm{E}\left((\mathbf{X}_0 \otimes \mathbf{I}) \cdot \mathbf{w}(\mathbf{X}_0)^{-1} \cdot (\mathbf{X}_0' \otimes \mathbf{I})\right)\right)^{-1}
$$

is the lower bound of the asymptotical variance matrix for all functional estimators $\widehat{\boldsymbol{\beta}}(\phi)$ such that $\mathrm{E}\left(\phi(\mathbf{X}_0)\mathbf{X}_0' \otimes \mathbf{I}\right)$ is regular.

There are a few special cases when we can compute the optimal estimator explicitly. If $\mathbf{\Sigma} = \mathbf{0}$ for instance, which corresponds to the classical AR model, and $\mathbf{G} = \sigma^2\mathbf{I}$ for some $\sigma^2 > 0$, we have $\mathbf{w}(\mathbf{z}) = \sigma^2\mathbf{I}$. Then the lower bound equals

$$\left(\mathrm{E}\left((\mathbf{X}_0 \otimes \mathbf{I}) \cdot \mathbf{w}(\mathbf{X}_0)^{-1} \cdot (\mathbf{X}_0' \otimes \mathbf{I})\right)\right)^{-1} = \sigma^2 \left(\mathrm{E}\left(\mathbf{X}_0\mathbf{X}_0' \otimes \mathbf{I}\right)\right)^{-1}$$

and asymptotical variance matrix of the estimator $\widehat{\boldsymbol{\beta}}(\phi)$ with $\phi(\mathbf{z}) = \mathbf{z}$ achieves this lower bound.

4.3 Simulation study

We introduced and studied a class of functional estimators of parameter β of the multivariate RCA(1) process in the previous sections. This study will compare the least-squares estimator to the particular choice of the functional estimator.

Setup of the simulations is the following: We simulate a sequence of observations from 2-dimensional RCA(1) model given by equation

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \left(\begin{pmatrix} 0.2 & 0.1 \\ 0.3 & 0.4 \end{pmatrix} + \begin{pmatrix} B_t^{11} & B_t^{12} \\ B_t^{21} & B_t^{22} \end{pmatrix}\right) \cdot \begin{pmatrix} X_{t-1}^1 \\ X_{t-1}^2 \end{pmatrix} + \begin{pmatrix} Y_t^1 \\ Y_t^2 \end{pmatrix}, \tag{4.16}$$

where both random coefficients $\mathbf{B}_t$ and error process $\mathbf{Y}_t$ are mutually independent and identically normally distributed. We set $\mathbf{\Sigma} = \mathrm{var}(\mathrm{vec}\mathbf{B}_t) = 0.2 \cdot \mathbf{I}$ and $\mathbf{G} = \mathrm{var}\,\mathbf{Y}_t = \mathbf{I}$. Simulation procedure consists of the following steps:

- repeat 1000 times
    1. simulate 2-dimensional RCA(1) time series (100 observations) given by equation (4.16)
    2. estimate parameter $\boldsymbol{\beta}$ using $\widehat{\boldsymbol{\beta}}_{LS}$ and $\widehat{\boldsymbol{\beta}}(\phi)$ with $\phi(\mathbf{z}) = \frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}}$
- compute $\overline{\widehat{\boldsymbol{\beta}}} = \frac{1}{1000}\sum_{i=1}^{1000} \widehat{\boldsymbol{\beta}}^i$ for each estimator
- plot density estimations of all parameters for both estimators

Let us verify stationarity assumption A11 of the process first. This assumption requires that all eigenvalues of matrix $\mathrm{E}(\mathbf{B}_t \otimes \mathbf{B}_t) + (\boldsymbol{\beta} \otimes \boldsymbol{\beta})$ are less than unity:

$$\begin{aligned} &\mathrm{E}(\mathbf{B}_t \otimes \mathbf{B}_t) + (\boldsymbol{\beta} \otimes \boldsymbol{\beta}) = \\ &= \mathrm{E}\begin{pmatrix} \mathrm{var}\,B_t^{11} & 0 & 0 & \mathrm{var}\,B_t^{12} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathrm{var}\,B_t^{21} & 0 & 0 & \mathrm{var}\,B_t^{22} \end{pmatrix} + \begin{pmatrix} 0.04 & 0.02 & 0.02 & 0.01 \\ 0.06 & 0.08 & 0.03 & 0.04 \\ 0.06 & 0.03 & 0.08 & 0.04 \\ 0.09 & 0.12 & 0.12 & 0.16 \end{pmatrix} = \end{aligned}$$

$$= \begin{pmatrix} 0.24 & 0.02 & 0.02 & 0.21 \\ 0.06 & 0.08 & 0.03 & 0.04 \\ 0.06 & 0.03 & 0.08 & 0.04 \\ 0.29 & 0.12 & 0.12 & 0.36 \end{pmatrix},$$

using mutual independence of random coefficients that follows from their non-correlated jointly normal distribution. The latter matrix has four eigenvalues (0.583, $0.064 + 0.036i$, $0.064 - 0.036i$, 0.050) all of which absolute values (0.583, 0.073, 0.073, 0.050) are less than one. Thus, the simulated process is well-defined stationary 2-dimensional RCA(1) process.

Results of the simulation study are displayed on Figure 4.1 and in Table 4.1. We have 1000 estimated values of true parameters β_{11}, β_{12}, β_{21} and β_{22}. We compare the estimators via both sample means and via density estimations (we use default density estimation in R based on kernel smoothing).

parameter	**true value**	**LS est.**	$\phi(\mathbf{z}) = \frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}}$
β_{11}	0.2	0.163	0.182
β_{12}	0.1	0.088	0.098
β_{21}	0.3	0.257	0.278
β_{22}	0.4	0.391	0.401

Table 4.1: Average values of estimated parameters

The least-squares estimator $\widehat{\boldsymbol{\beta}}_{LS}$ always underestimates the true value, especially for β_{11} and β_{21} whereas functional estimator $\widehat{\boldsymbol{\beta}}(\frac{\mathbf{z}}{1+\mathbf{z}'\mathbf{z}})$ is closer to the true values. The density estimation also reveals bias for the least-squares estimator.

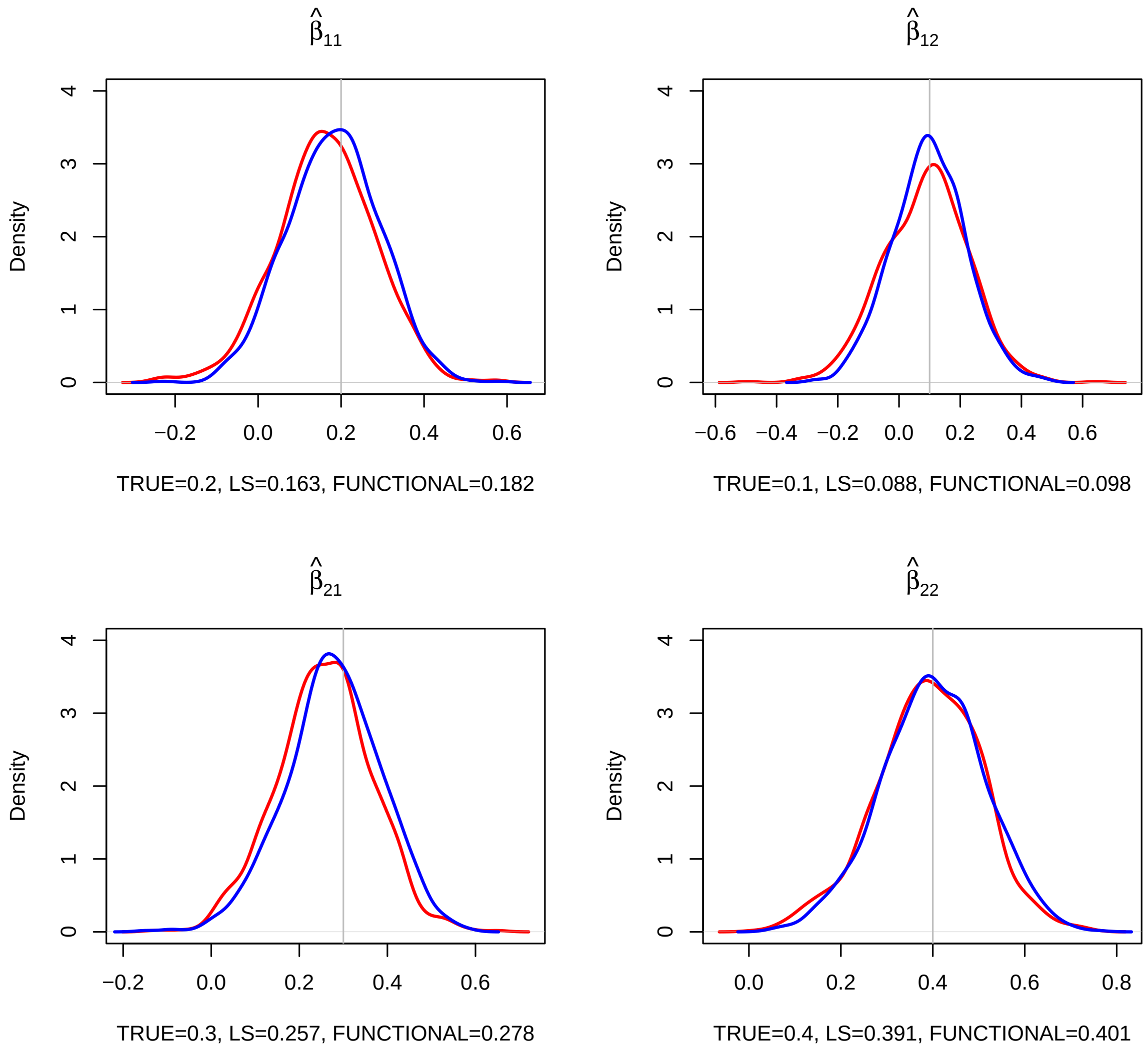

Figure 4.1: Density of estimated values of the parameters (red color for LS estimator, blue color for functional estimator) for simulated 2-dimensional RCA(1) process.

Chapter 5

Conclusions and further topics

This work studied estimation of parameters in Random Coefficient Autoregressive time series models. We introduced the concept of functional estimator originally published by Schick. The idea behind such estimators is based on generalization of the least-squares estimator via imposing chosen measurable function that controls contribution of past observed values.

The functional estimator, which covers both least-squares and weighted least-squares estimators, was thoroughly investigated. We proved its strong consistency and asypmtotical normality (Theorems 2.2, 3.1, and 4.2) and proposed a strongly consistent estimator of its asymptotical variance for the univariate first-order case (Theorem 2.3) and asymptotical variance matrix for the higher-orders and multivariate cases (Theorems 3.2 and 4.4). We deduced that the weighted least-squares estimator has the smallest asymptotical variance among all functional estimators and illustrated impact of the choice of generating function on the estimator for a narrower class of functions where the shape of the functions depends on a single parameter only. We also established the rate of convergence for the distribution of the estimator to normal distribution (Theorem 2.8). For the higher-order RCA processes, we evaluated the asymptotical variance matrices for a few specific choices of the generating function and showed that the optimal estimator (in the sense of asymptotical variance matrix) is the weighted least-squares estimator again. Generally, the techniques of the proofs provided were similar to those for the first-order model, only more attention had to be paid to precise manipulation with matrix terms and a few technical theorems had to be used (for instance Theorem 3.3).

We performed many simulations throughout the previous chapters. Generally speaking, the least-squares estimator usually underestimated the true value of the parameters in contrast to the weighted least-squares estimator. Such underfit was heavier for larger variances of either random coefficients or error process. Since the weighted least-squares estimator requires the knowledge of additional variance parameters, which typically results in an iterative estimation procedure, we proposed a special functional estimator that is equivalent to the weighted least-squares estimator under some special conditions. It turned out that such estimator behaves well even under general conditions, does not require any prior knowledge of the parameters, and preserves the simplicity of the least-squares estimator.

We established the stochastic setup in which we studied behavior of the estimators. In accord with other authors we assumed that the random coefficient process is a sequence of independent and identically distributed random variables (or vectors, matrices, respectively) but we relaxed the error process into a strictly stationary and ergodic martingale difference sequence. Theorems 2.1 and 4.1 proved that there exist such univariate and multivariate RCA processes.

We are aware that such assumptions are rather strong and hard to verify. The assumptions might be further relaxed which would require more complicated limit theorems dealing with weak dependency of random variables, vectors or matrices to prove stated asymptotical results. We could also study the bias of the estimators or eliminate the presumption that conditional variance of the error process is time invariant which simplified the proofs. Another generalization could be based on the demanded mutual independence of the random coefficients and the error process. Some authors have allowed the sequences to be correlated via a constant correlation matrix. Since we widely benefited from the independence through the proofs, we did not relax the model in this manner.

Theorem 2.8 is a general Hoeffding-type exponential inequality that might be of independent interest. It would be possible to extend the result further to the multivariate case and obtain multivariate version of the Berry-Essén inequality.

We have also tried to apply our results to real data. Unfortunately, financial data of logarithmic returns we had at our disposal did not exhibit any autoregressive effect. On the other hand, applications of our results to other theoretical approaches (likewise the authors of [39] have done in a special simple case proposed by Schick in [28]) is a challenge for further research.

Chapter 6

Auxiliary results

Definition 6.1. *Ergodicity*
Let $(\Omega, \mathcal{A}, \mathrm{P})$ denote probability space.
Measurable transformation $T : \Omega \to \Omega$ is said to be measure preserving if $\mathrm{P}(T^{-1}(A)) = \mathrm{P}(A)$ for all $A \in \mathcal{A}$.
Event $A \in \mathcal{A}$ is invariant towards transformation T if $T^{-1}(A) = A$. Let $\mathcal{E}$ denote σ-field of all invariant events under T.
Measure preserving transformation T is ergodic if $\mathrm{P}(E) = 1$ or $\mathrm{P}(E) = 0$ for all $E \in \mathcal{E}$.
Strictly stationary sequence $\{X_t, t \in \mathbb{Z}\}$ defined on $(\Omega, \mathcal{A}, \mathrm{P})$ is called ergodic if

$$X_t(\omega) = X_1(T^{t-1}(\omega))$$

for all $t \in \mathbb{Z}$ and $\omega \in \Omega$, where T is a measure preserving and ergodic transformation.

Theorem 6.1. *Ergodic Theorem*
Let $\{X_t, t \in \mathbb{Z}\}$ be a strictly stationary ergodic integrable sequence.
Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} X_t = \mathrm{E} X_1 \quad a.s..$$

Proof: See [10], Section 13.4, Theorem 13.12. □

Lemma 6.1. *Persistency of ergodicity*
Let $\{X_t, t \in \mathbb{Z}\}$ be a strictly stationary ergodic sequence with values in some measurable space $(\mathcal{X}, \mathcal{B})$.
Let f be a measurable map $f : (\mathcal{X}^\infty, \mathcal{B}^\infty) \to (\mathcal{Y}, \mathcal{T})$, where $\mathcal{X}^\infty$ is the product space, $\mathcal{B}^\infty$ is the product σ-field, and $(\mathcal{Y}, \mathcal{T})$ is a measurable space.
Then $\{Y_t, t \in \mathbb{Z}\}$ where $Y_t = f(\dots, X_{t-1}, X_t, X_{t+1}, \dots)$ is an ergodic sequence.

Proof: See [33], Lemma 4.23. Special case for RCA processes with singly-infinite arguments of function f to the past is proved in [28], Theorem 2.7. □

Lemma 6.2. ***Accuracy of normal approximation***
Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and f and g be $\mathcal{F}$-measurable functions.
Then for any $\varepsilon > 0$

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{f}{g} \leq x\right) - \Phi(x) \right| \leq \sup_{y\in\mathbb{R}} \left|\mathrm{P}\left(f \leq y\right) - \Phi(y)\right| + \mathrm{P}\left(|g-1| > \varepsilon\right) + \varepsilon,$$

where $\Phi(x)$ is the distribution function of the standard normal distribution.

Proof: See [24]. □

Lemma 6.3. ***Berry-Esséen Theorem***
Let $\{U_t, t \in \mathbb{N}\}$ be martingale difference sequence with constant nonzero variance σ^2. Let $\mathrm{P}(|U_t| \leq c, \forall t \in \mathbb{N}) = 1$ for some constant $c > 0$.
Then there exists $d > 0$ such that for each $n \in \mathbb{N}$

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{1}{\sqrt{\sigma^2}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \leq x\right) - \Phi(x) \right| \leq d \cdot \frac{(\ln n)^3}{\sqrt{n}}.$$

Proof: See [18]. □

Lemma 6.4.
Let $\{X_t, t \in \mathbb{Z}\}$ be L_2-NED of size $-a$ on $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ with constants $\{d_t\}$. Let $h(x)$ be function that satisfies Lipschitz assumption 2.3 with Lipschitz constant $c_h > 0$.
Then $\{h(X_t), t \in \mathbb{Z}\}$ is also L_2-NED of size $-a$ on $\{\boldsymbol{V}_t\}$ with constants $\{c_h \cdot d_t\}$.

Proof: It is a special case of Theorem 17.12 in [10]. □

Lemma 6.5.
Let $\{X_t, t \in \mathbb{Z}\}$ be an L_r-bounded zero-mean sequence, for $r > 1$. Let $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ be α-mixing of size $-a$.
If $\{X_t\}$ is an L_p-NED of size $-b$ on $\{\boldsymbol{V}_t\}$, for $1 \leq p < r$ with constants $\{d_t\}$, then $\{X_t, \mathcal{F}_t\}$ is an L_p-mixingale of size $-\min(b, a(1/p - 1/r))$ with constants $\{c_t\}$, such that $c_t = O(\max(\|X_t\|_r, d_t))$.

Proof: See Theorem 17.5 in [10]. □

Definition 6.2. ***Matrix Operators***
Let $\mathbf{A}$ be an $m \times n$ matrix.
Then mn-component vector $vec(\mathbf{A})$ is defined as stacking the columns of $\mathbf{A}$, one on top of the other in order from left to right.

Let $\mathbf{A}$ be an $n \times n$ symmetric matrix.
Then $n(n+1)/2$-component vector $vech(\mathbf{A})$ is defined as stacking those parts of the columns of $\mathbf{A}$ on and below the main diagonal, one on top of the other in order from left to right.

Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{B}$ be a $p \times q$ matrix.
Then the *Kronecker product* $\mathbf{A} \otimes \mathbf{B}$ of $\mathbf{A}$ and $\mathbf{B}$ is the $mp \times nq$ matrix whose (i, j)th block is the $p \times q$ matrix $a_{i,j} \cdot \mathbf{B}$, where $a_{i,j}$ is the (i, j)th element of $\mathbf{B}$.

Lemma 6.6.
Let $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ *and* $\mathbf{D}$ *be any matrices and* $\mathbf{u}$, $\mathbf{v}$ *be any vectors.*
Then

(a) $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$

(b) $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C})$

(c) $(\mathbf{A} + \mathbf{B}) \otimes (\mathbf{C} + \mathbf{D}) = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{D})$

(d) $\mathbf{u}\mathbf{v}' = \mathbf{u} \otimes \mathbf{v}' = \mathbf{v}' \otimes \mathbf{u}$

(e) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$

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

Proof: See [29], Paragraph 3 in Chapter 7, mainly Theorem 7.6. □

Lemma 6.7.
Let $\mathbf{A}_i$ *and* $\mathbf{B}_i$, $i = 1, 2, \ldots, n$, *be any matrices.*
Then

$$\left(\prod_{i=1}^{n} \mathbf{A}_i\right) \otimes \left(\prod_{j=1}^{n} \mathbf{B}_j\right) = \prod_{i=1}^{n} (\mathbf{A}_i \otimes \mathbf{B}_i)\,.$$

Proof: See [29], Paragraph 3 in Chapter 7. □

Lemma 6.8.
Let $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *be any matrices and* $\mathbf{u}$, $\mathbf{v}$, $\mathbf{x}$ *and* $\mathbf{y}$ *be any vectors such that the expressions below are well defined.*
Then

(a) $\mathrm{vec}(\mathbf{u}) = \mathrm{vec}(\mathbf{u}') = \mathbf{u}$

(b) $\mathrm{vec}(\mathbf{u}\mathbf{v}') = \mathbf{v} \otimes \mathbf{u}$

(c) $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A}) \cdot \mathrm{vec}(\mathbf{B})$

(d) $\mathbf{Au} = (\mathbf{u}' \otimes \mathbf{I}) \cdot \mathrm{vec}(\mathbf{A})$

(e) $(\mathbf{A} \otimes \mathbf{I}) \otimes (\mathbf{A} \otimes \mathbf{I}) = (\mathrm{vec}(\mathbf{A}'\mathbf{A}))' \otimes \mathbf{I}$

(f) $\mathrm{vec}(\mathbf{A} \cdot \mathbf{B}) = (\mathbf{B}' \otimes \mathbf{I}) \cdot \mathrm{vec}(\mathbf{A})$

(g) $\mathrm{vec}(\mathbf{u}\mathbf{v}') = (\mathbf{v} \otimes \mathbf{I}) \cdot \mathbf{u}$

(h) $\mathrm{vec}\left((\mathbf{u}\mathbf{v}')(\mathbf{x}\mathbf{y}')\right) = \left((\mathbf{y}\mathbf{x}') \otimes \mathbf{I}\right) \cdot \left(\mathbf{v} \otimes \mathbf{I}\right) \cdot \mathbf{u}$

Proof: Properties a–e are proved in [29], Paragraph 5 in Chapter 7. Properties f and g are direct applications of property c. Property h is a consequence of properties f and g. □

Lemma 6.9.
Let $\mathbf{A}$ *be an* $n \times n$ *symmetric matrix.*
Then there exist constant $(n(n+1)/2) \times n^2$ *matrices* $\mathbf{K}_n$ *and* $\mathbf{H}_n$ *for which*

$$\begin{aligned} \mathrm{vech}(\mathbf{A}) &= \mathbf{H}_n \mathrm{vec}(\mathbf{A}) \\ \mathrm{vec}(\mathbf{A}) &= \mathbf{K}_n' \mathrm{vech}(\mathbf{A}) \end{aligned}$$

such that $\mathbf{H}_n \mathbf{K}_n' = \mathbf{I}_{n(n+1)/2}$. *Matrix* $\mathbf{K}_n$ *is sometimes called the duplication matrix.*

Proof: See Theorem A.1.3 in [25]. □

Chapter 7

Source codes

```
################################################################################
function(beta,sigmaB,sigmaY2,n)
################################################################################
{
################################################################################
# Functionality: Generates RCA(p) model
#
# Requirements: Library MASS
#
# Assumptions:
#  - random coefficients have multi-dimensional normal distribution
#  - error process has normal distribution
#
# Input:
#  - beta = vector of the true parameters beta
#  - sigmaB = variance matrix of random coefficients B_t
#  - sigmaY2 = variance of error process Y_t
#  - n = length of simulated time series
#
# Output: Data frame consisting of 2 columns
#  - t = time index from 1 to length
#  - y = simulated RCA model
#
# Note: First n simulations is used as initial starting values.
#
# Example of usage:
#  - RCA(1) model:
#  gen.rcap(beta=matrix(0.5),sigmaB=matrix(0.3),1,100)
#  - RCA(2) model with correlated coefficients:
#  gen.rcap(matrix(c(1/3,1/3),nrow=2),matrix(c(0.3,0.2,0.2,0.4),nrow=2),1,500)
#
################################################################################
```

```
library(MASS)

p = dim(beta)[1]
N = 2*n
eps <- rnorm(N, 0, sqrt(sigmaY2))
B <- mvrnorm(N, beta, sigmaB)
y <- rep(0, N)

for(i in (p+1):N) {
        y[i] <- B[i,]%*%y[(i-1):(i-p)] + eps[i]
}
data <- data.frame(1:n, y[(n+1):N])
dimnames(data)[[2]]<-list("t","y")
return(data)
}

################################################################################
function(data)
################################################################################
{
################################################################################
# Functionality: Estimates RCA(1) model using LS and weighted LS estimators
#
# Input:
#  - data = one-dimensional sequence of data to be fitted
#
# Output: Row vector consisting of 6 values
#  - beta = estimate of beta parameter using LS estimator
#  - betaW = estimate of beta parameter using weighted LS estimator
#  - varEps = estimate variance of error process
#  - varB = estimate of variance of random coefficient process
#  - as.var = estimate of asymptotical variance of LS beta estimator
#  - as.varW = estimate of asymptotical variance of weighted LS beta estimator
#
################################################################################

data<-c(0,data)
n<-length(data)-1
data0<-data[-length(data)]
data1<-data[-1]

beta<-(data1%*%data0)/(data0%*%data0)
names(beta)<-"beta"
est.eps<-data1-beta*data0
x<-data0^2
```

```
y<-est.eps^2
regrese<-lm(y~x)
c<-coef(regrese)
names(c)<-c("varEps","varB")

var.eps<-fitted.values(regrese)
data1W<-data1/var.eps
data0W<-data0/var.eps
betaW<- (data1W%*%data0)/(data0W%*%data0)
names(betaW)<-"betaW"

as.var<-n*sum(c[1]*(data1^2)+c[2]*(data1^4))/(sum(data1^2)^2)
names(as.var)<-"beta.as.var"
as.varW<-n/sum(data1^2/(c[1]+c[2]*data0^2))
names(as.varW)<-"betaW.as.var"

return(c(beta,betaW,c[1],c[2],as.var,as.varW))
}

################################################################################
function(data,p=1)
################################################################################
{
################################################################################
# Functionality: Estimates RCA(p) model using LS and functional estimator
#
# Assumptions:
#  - function fi is chosen as x/(1+x'x),
#    but it is possible to redefine it in the code
#
# Input:
#  - data = one-dimensional sequence of data to be fitted
#  - p = order of assumed RCA(p) model to be fitted
#
# Output: Data frame consisting of 2 columns
#  - beta.ls = estimate of beta parameter(s) using LS estimator
#  - beta.f = estimate of beta parameter(s) using functional estimator
#
################################################################################

fi.ls<-function(x){return(x)}
fi<-function(x){return(x/(1+t(x)%*%x))}

N<-length(data)

var1 <- 0
```

```
var2 <- 0
var1.ls <- 0
var2.ls <-0

for(i in (p+1):N) {
        var1 <- var1 + fi(data[(i-1):(i-p)])*data[i]
        var2 <- var2 + fi(data[(i-1):(i-p)])%*%t(data[(i-1):(i-p)])

        var1.ls <- var1.ls + fi.ls(data[(i-1):(i-p)])*data[i]
        var2.ls <- var2.ls + fi.ls(data[(i-1):(i-p)])
                          %*%t(data[(i-1):(i-p)])
}

beta.ls <- solve(var2.ls)%*%var1.ls
beta <- solve(var2)%*%var1
solution <- cbind(beta.ls,beta)
dimnames(solution)[[2]]<-list("beta.ls","beta.f")
return(solution)
}
```

Bibliography

[1] Akharif A., Hallin M., Efficient Detection of Random Coefficients in Autoregressive Models, *The Annals of Statistics*, 31, 675–704, 2003.

[2] Anděl J., Autoregressive Series with Random Parameters, *Math. Operationsforsch. u. Statist. 7*, 5, 735–741, 1976.

[3] Aue A., Testing for Parameter Stability in RCA(1) Time Series, *Journal of Statistical Planning and Inference*, 13, 3070–3089, 2006.

[4] Aue A., Horváth L., Steinebach J., Estimation in Random Coefficient Autoregressive Models, *Journal of Time Series Analysis*, 27, 60–67, 2006.

[5] Basu A. K., Sen Roy S., On Rates of Convergence in the Central Limit Theorem for Parameter Estimation in Random Coefficient Autoregressive Models, *Journal of the Indian Statistical Association*, 26, 19–25, 1988.

[6] Basu A. K., Sen Roy S., On Rates of Convergence in the Central Limit Theorem for Parameter Estimation in General Autoregressive Model, *Statistics*, 21, 461–470, 1990.

[7] Billingsley P., The Lindeberg-Levy Theorem for Martingales, *Proc. of the American Mathematical Society 12*, 5, 788–792, 1961.

[8] Bose A., Dasgupta R., Speed of Convergence of the Least-Squares Estimator in Autoregressive Models, *Journal of Statistical Planning and Inference*, 38, 371–381, 1994.

[9] Brockwell P.J, Davis R.A, Time Series: Theory and Methods. Springer, New York, 1991.

[10] Davidson J., Stochastic Limit Theory, *Advanced Texts in Econometrics*, Oxford University Press, New York, Reprinted 2002.

[11] Engle F.R., Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 50, 987–1007, 1982.

[12] Glynn P. W., Ormoneit D., Hoeffding's Inequality for Uniformly Ergodic Markov Chains, *Statistics & Probability Letters*, 56, 143–146, 2002.

[13] Horký T., Testing the Randomness of Autoregressive Coefficients (In Czech), Diploma Thesis, Prague, 2005.

[14] Hwang S.Y., Basawa I.V., Parameter Estimation for Generalized Random Coefficient Autoregressive Processes, *Journal of Statistical Planning and Inference*, 68, 323–337, 1998.

[15] Janečková H., Some Generalizations in a Heteroscedastic RCA(1) Model, *Acta Universitatis Carolinae*, Prague, 2001.

[16] Janečková H., Time Series with Changing Parameters, *Doctoral Thesis*, Prague, 2002.

[17] Janečková H, Prášková Z., CWLS and ML Estimates in a Heteroscedastic RCA(1) Model, *Statistics & Decisions*, 22, 245–259, 2004.

[18] Kato Y., Rates of convergence in central limit theorem for martingale differences, *Bull. Math. Statist.*, 18, 1–8, 1979.

[19] Koul H. L., Schick A., Adaptive Estimation in a Ranodm Coefficient Autoregressive Model, *The Annals of Statistics*, 24, 1025–1052, 1996.

[20] Lachout P., Theory of Probability (In Czech), Karolinum, Praha, 1998.

[21] Ling S., A Double AR(p) Model: Structure and Estimation, *Statistica Sinica*, accepted, 2006.

[22] Ling S., Estimation and Testing Stationarity for Double-Autoregressive Models, *Journal of Royal Statistical Society*, 66, 63–78, 2004.

[23] Lütkepohl H., New Introduction to Multiple Time Series Analysis, Springer, 2005.

[24] Michel R., Pfanzagl J., The Accuracy of the Normal Approximation for Minimum Contrast Estimate, *Z. Wahrsch. Verw. Gebiete*, 18, 73–84, 1971.

[25] Nicholls D.F, Quinn B.G., Random Coefficient Autoregressive Models: An Introduction, *Lecture Notes in Statistics*, 11, Springer, New York, 1982.

[26] Peiris M. S., Thavaneswaran A., Multivariate Stable ARMA Processes with Time Dependent Coefficients, *Metrika*, 54, 131–138, 2001.

[27] Rao C.R., Lineární metody statistické indukce a jejich aplikace, *ACADEMIA*, Praha 1978

[28] Schick A., $\sqrt{n}$-consistent Estimation in a Random Coefficient Autoregressive Model, *The Australian Journal of Statistics*, 38, 155–160, 1996.

[29] Schott J., Matrix Analysis for Statistics, *Wiley Series in Probabality and Statistics*, 1996

[30] Tang Y., A Hoeffding-type Inequality for Ergodic Time Series, *Journal of Theoretical Probability*, 20, 167–176, 2007

[31] Tong H., Nonlinear Time Series: A Dynamic Approach, *Oxford Univ. Press*, 1990.

[32] Tsay R. S., Conditional Heteroscedastic Time Series Models, *Journal of the American Statistical Association*, 82, 590–604, 1987.

[33] Vaart A. W. van der, Time Series, *Lecture Notes*, Vrije Universiteit Amsterdam. http://www.math.vu.nl/sto/onderwijs/timeseries/dictaat02.pdf

[34] Vaněček P., Estimators of Generalized RCA Models, *WDS'04 Proceedings of Contributed Papers*, Part I, MATFYZPRESS, 35–40, 2004.

[35] Vaněček P., Rate of Convergence for a Class of RCA Estimators, *Kybernetika*, 6, 698–709, 2006.

[36] Vaněček P., Estimators of Multivariate RCA Models, *Proceedings of 56th Session of ISI, Section CPM033 - Time Series Analysis I*, 2007.

[37] Wang D., Ghost S. K., Bayesian Analysis of Random Coefficient AutoRegressive Models, submitted to *Bayesian Analysis*.

[38] Weiss A. A., ARMA Models with ARCH Errors, *Journal of Time Series Analysis*, 5, 129–143, 1984.

[39] Zheng H. T., Basawa I. V., Datta S., Inference for p^{th}-order Random Coefficient Integer-valued Autoregressive Processes, *Journal of Time Series Analysis*, 27, 411–440, 2006.