UNIVERSITY OF TARTU
FACULTY OF MATHEMATICS AND COMPUTER SCIENCES

PREDICTIONS BY NON-INVERTIBLE ARMA MODELS

Master Thesis

Author: **Agris Vaselāns**
Student ID: B25700
Supervisor: PhD Raul Kangro

TARTU 2013

Table of Contents

Introduction **2**

1. $ARMA$ model and the concept of non-invertibility **4**

1.1. Non-invertibility 4

1.2. The Autocovariance Generating Function 8

2. Forecasting **13**

2.1. Multivariate Gaussian 13

2.2. Forecasting invertible Gaussian model 14

2.2.1. One step ahead prediction 17

2.3. Non-invertible Gaussian processes 18

2.3.1. Strictly non-invertible $MA(1)$ 19

2.4. Non-invertible non-Gaussian $MA(1)$ process 22

2.4.1. Predictions by non-invertible non-Gaussian MA(1) 22

2.4.2. Simulation studies 29

3. Application **36**

3.1. The Overshorts data 36

3.2. Predictions by non-invertible $MA(1)$ 38

Summary **40**

Kokkuvõte **42**

Bibliography **44**

Appendix **45**

1. Graphs - Section1 45
2. Graphs - section 2.4.1 45
3. Coefficient test of the Section 2.4.1 46
4. Efficiency code of the Section 2.4.1 46
5. Graphs and code of the Chapter 3 47
6. Sage computations 1 48
7. Sage computations 2 53
8. Sage computations 3 58
9. Sage computations 4 63

Introduction

In the time series analysis one of the most used predicting models are of so called auto regressive moving average ($ARMA$) type. These models are well studied in numerous monographies and research papers. One of the basic assumptions used in the derivation process of the prediction equations is the invertibility of the underlying process. Usually invertibility is assumed as a prerequisite and very little attention is paid to the forecasting of non-invertible processes.

Recent papers [1, pp. 227-229.] shows that nowadays more and more researchers consider and examine the case when the underlying process does not satisfy invertibility condition. Non-invertible processes have been studied also quite a long time ago, but they have become the object of interest due to new applications in sciences (signal detection, financial analysis) also the rapid development of computer sciences and computation possibilities take part in the growing interest of such processes.

In basic time series course non-invertible processes usually are discussed very briefly, but globally the interest in such processes is increasing, therefore the aims of this thesis are:

- to investigate theoretically the questions related to predicting further values of non-invertible $ARMA$ processes;
- to do the computer simulations and compare different methods.

To cover these aims both theoretical and simulation studies are provided. Therefore in the beginning we give a very short introduction and necessary background of stationary $ARMA$ processes needed to give the definition of the non-invertibility of $ARMA$ process. We proceed with another natural assumption used in the derivation process of the prediction equations. The assumption of Gaussian distributed random variables (innovations) gives some prerogatives and simplifies the derivation of the prediction equations also in case of non-invertible process. We briefly discuss the gains which are represented as a useful collection of consecutive theorems that leads to the minimum mean square error predictor in case of non-invertible process with Gaussian distributed data. To extend our studies of non-invertible processes we continue with studies of non-Gaussian, non-invertible process. This situation requires more specific analysis which is provided by a case study of a non-invertible moving average $MA(1)$ process with uniformly distributed innovations (error process).

The thesis consists of 3 main sections with suitable subsections. In the first section the basic concept of a non-invertibility is given. The second section is dedicated to the forecasting of an $ARMA$ process. In this section the derivation of prediction equations in case of invertible process is given, then the derivation of the forecast of a non-invertible process with Gaussian distributed data is described and the sections concludes with the derivation of the minimum mean square error predictor in case of the non-invertible process with uniformly distributed innovation series. Results are illustrated with computer simulations and corresponding graphs. In the last section a real world application is considered and corresponding results are given. Since all sections require some computational work and appropriate programming, the collection of suitable codes written in the R language [2] and scripts of the open-source mathematical software system `Sage` [3], which provides the symbolic calculations needed for this thesis, can be found in the appendix.

1. *ARMA* model and the concept of non-invertibility

In this section we give the basic definitions of an autoregressive moving average *ARMA* process. Here we state also the basic results and theorems that will be used as basis of further investigation of the non-invertible *ARMA* process, which is the main object of this thesis. Note that in this case we just consider discrete processes, where time $t \in \mathbb{Z}$ and also the time horizon $h \in \mathbb{Z}$.

1.1. Non-invertibility

We start with some definitions in order to recall the basic terms and also to agree on notation. We begin with definitions of autocovariance and stationarity. Mostly, in this section we follow the definitions given in monograph by Stoffer and Shumway [4].

Definition 1. *The mean function of a stochastic process x_t is*

$$\mu_t = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx,$$

provided it exists, where E denotes the usual expected value operator and $f_t(x)$ denotes process distribution density function.

Definition 2. *The autocovariance function of a finite variance process x_t with mean value function μ_t is defined as*

$$\gamma_x(s,t) = E[(x_s - \mu_s)(x_t - \mu_t)],$$

for all s and t.

When no possible confusion exists about which time series we are referring to, we will drop the subscript and write $\gamma_x(s,t)$ as $\gamma(s,t)$.

Definition 3. *A weakly stationary time series, $\{x_t\}$, is a finite variance process such that*

- *the mean value function, μ_t is constant and does not depend on time t, and*
- *the covariance function, $\gamma(s,t)$, depends on s and t only through their difference $|s-t|$.*

Henceforth, we will use the term stationarity to mean weak stationarity.

Definition 4. *The autocovariance function of a stationary time series will be written as*

$$\gamma(h) = E[(x_{t+h} - \mu)(x_t - \mu)].$$

Note that

$$\begin{aligned} \gamma(h) &= \gamma(t+h-t) \\ &= E[(x_{t+h} - \mu)(x_t - \mu)] \\ &= E[(x_t - \mu)(x_{t+h} - \mu)] \\ &= \gamma(t-(t+h)) \\ &= \gamma(-h). \end{aligned}$$

When we have defined the autocovariance function, we can give the definitions of the white noise process and Gaussian white noise, which plays an important role in the analysis of $ARMA$ processes and also in the derivation of the prediction equations.

Definition 5. *The process w_t (with mean 0 and variance σ^2) is said to be the white noise process, if and only if w_t has zero mean and covariance function*

$$\gamma(h) = \begin{cases} 0, h \neq 0 \\ \sigma^2, h = 0. \end{cases}$$

Definition 6. *We say that stationary process x_t is* autoregressive process *of order p, abbreviated $AR(p)$, if*

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t,$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$) and w_t is the white noise process.

Unless stated otherwise, we assume that w_t is a Gaussian white noise series with variance σ_w^2.

Definition 7. *The process x_t is said to be an $AR(p)$ process with mean μ if $x_t - \mu$ is an $AR(p)$ process.*

Note, that if the process is said to be with mean μ, then instead of writing

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \ldots + \phi_p(x_{t-p} - \mu) + w_t,$$

we can also write

$$x_t = \alpha + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t,$$

where $\alpha = \mu(1 - \phi_1 - \ldots - \phi_p)$.

A compact way of defining $ARMA$ processes is by defining the back-shift operator and introducing the autoregressive and moving average equations and operators.

Definition 8. *We define the back-shift operator by*

$$Bx_t = x_{t-1},\ t \in \mathbb{Z}$$

and extend it to powers $B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2}$*, and so on. Thus,*

$$B^k x_t = x_{t-k},\ t \in \mathbb{Z}.$$

Definition 9. *The autoregressive operator is defined to be*

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

Definition 10. *We say that stationary process* x_t *is* moving average process *of order* q*, or* $MA(q)$ *process, if*

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}$$

where $\theta_1, \theta_2, ..., \theta_q\ (\theta_q \neq 0)$ *are constants (parameters) and* w_t *is the white noise process.*

Definition 11. *The process* x_t *is said to be an* $MA(q)$ *process with mean* μ *if* $x_t - \mu$ *is an* $MA(q)$ *process.*

Unless stated otherwise, we assume w_t to be the Gaussian white noise process.

Definition 12. *The moving average operator is*

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q.$$

We may also write the $MA(q)$ process in the equivalent form

$$x_t = \theta(B) w_t.$$

Finally, we give the formal definition of the $ARMA(p, q)$ process.

Definition 13. *A stationary process* $\{x_t; t \in \mathbb{Z}\}$ *is said to be* $ARMA(p, q)$ *process if*

$$x_t = \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q},$$

where $\phi_p \neq 0$, $\theta_q \neq 0$*, and* w_t *is the white noise process.*

The parameters p and q are called the autoregressive and the moving average orders, respectively. Unless stated otherwise, w_t is the Gaussian white noise sequence.

Definition 14. *The process* x_t *is said to be an* $ARMA(p, q)$ *process with mean* μ *if* $x_t - \mu$ *is an* $ARMA(p, q)$ *process.*

So, the compact form given by the autoregressive and the moving average operators is $\phi(B)x_t = \theta(B)w_t$. Similarly as in case of AR, also we consider the process with non zero mean [4, p. 93.]. If x_t has a non-zero mean μ, we set $\alpha = \mu(1-\phi_1-\ldots-\phi_p)$ and write the model as

$$x_t = \alpha + \phi_1 x_{t-1} + \ldots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \ldots + \theta_q w_{t-q}.$$

A convenient way of defining invertibility (non-invertibility) of the $ARMA$ processes is given by AR and MA polynomials we are going to define next.

Definition 15. *The AR and MA polynomials are defined as*

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p,\ \phi_p \neq 0,$$

and

$$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q,\ \theta_q \neq 0,$$

respectively, where z is a complex number.

Let us define causality and invertibility of an $ARMA$ process.

Definition 16. *An $ARMA(p,q)$ model, $\phi(B)x_t = \theta(B)w_t$, is said to be* causal, *if the time series $\{x_t;\ t \in \mathbb{Z}\}$ can be written as a one-sided linear process:*

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} = \psi(B)w_t,$$

where $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$, and $\sum_{j=0}^{\infty} |\psi_j| < \infty$; we set $\psi_0 = 1$.

Now we are able to continue with one of the most important terms in this thesis- the invertibility of an $ARMA$ process.

Definition 17. *An $ARMA(p,q)$ model, $\phi(B)x_t = \theta(B)w_t$, is said to be* invertible, *if the time series $\{x_t; t \in \mathbb{Z}\}$ can be written as*

$$\pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} = w_t,$$

where $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$, and $\sum_{j=0}^{\infty} |\pi_j| < \infty$; we set $\pi_0 = 1$.

When we have defined the invertibility of the $ARMA$ process, it is natural to give the criteria of invertibility, so that we can use the criteria to classify the processes into invertible and non-invertible classes.

Lemma 1. *Invertibility of an $ARMA(p,q)$ process . An $ARMA(p,q)$ model is invertible if and only if $\theta(z) \neq 0$ for $|z| \leq 1$. The coefficients π_j of $\pi(B)$ can be* *determined by solving*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)},\ |z| \leq 1.$$

Note that the discussion about the invertibility of $ARMA$ process reduces to the investigation of the invertibility of MA part of the process because of the invertibility of every AR process.

1.2. The Autocovariance Generating Function

In well known time series literature by Brockwell & Davis [5] we can find an useful tool called autocovariance generating function, that will help us to deal with the situation of a non-invertible Gaussian process later, when we will try to derive the minimum mean square error predictor and compare it with the best linear predictor.

But now we start with the definition of the autocovariance generating function.

Definition 18. *If x_t is a stationary process with autocovariance function $\gamma(\cdot)$, then it's autocovariance generating function is defined by*

$$G(z) = \sum_{k=-\infty}^{\infty} \gamma(k) z^k,\ z \in \mathbb{C},$$

provided the series converges for all z in some annulus $r^{-1} < |z| < r$ with $r > 1$.

It's said there [5, p. 103.] that, frequently the generating function is easy to calculate, in which case the autocovariance at lag k may be determined by identifying the coefficient of either z^k orz^{-k}. Clearly $\{x_t\}$ is white noise if and only if the autocovariance generating function $G(z)$ is constant for all z. If

$$x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},\ w \sim N(0, \sigma^2) \tag{1.2.1}$$

and there exists $r > 1$ such that

$$\sum_{j=-\infty}^{\infty} |\psi_j||z|^j < \infty,\ , r^{-1} < |z| < r$$

the generating function G($\cdot$) takes a very simple form. It is easy to see that

$$\gamma(k) = Cov(x_{t+k}, x_t) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k},$$

and hence that

$$
\begin{aligned}
G(z) &= \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} z^k \\
&= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_{j+k} z^k \right] \\
&\quad \textit{Let us define } l := j+k \\
&= \sigma^2 \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \psi_l z^{l-j} \\
&= \sigma^2 \sum_{l=-\infty}^{\infty} \psi_l z^l \sum_{j=-\infty}^{\infty} \psi_j z^{-j}.
\end{aligned}
$$

Defining

$$\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j,\ r^{-1} < |z| < r, \tag{1.2.2}$$

we can write this result more neatly in the form

$$G(z) = \sigma^2 \psi(z)\psi(z^{-1}),\ \ r^{-1} < |z| < r.$$

Lemma 2. *[5, pp. 103-104.] The Autocovariance Generating Function of an $ARMA(p,q)$ process $\phi(B)x_t = \theta(B)w_t$ for which $\phi(z) \neq 0$ when $|z| = 1$, we can express the process $ARMA(p,q)$ as in* Equation *1.2.1 with coefficients defined in* Equation *1.2.2 of the form*

$$\psi(z) = \phi^{-1}(z)\theta(z),\ r^{-1} < |z| < r,$$

for some $r > 1$. Hence

$$G(z) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})},\ r^{-1} < |z| < r. \tag{1.2.3}$$

In the same literature there is proposed such theorem.

Theorem 1. *[5, p. 105.] Let $\{x_t\}$ be the $ARMA(p,q)$ process satisfying the equations*

$$\phi(B)x_t = \theta(B)w_t,$$

where $\phi(z) \neq 0$ and $\theta(z) \neq 0$ for all $z \in \mathcal{C}$ such that $|z| = 1$. Then there exist polynomials, $\tilde{\phi}(z)$ and $\tilde{\theta}(z)$, non-zero for $|z| \leq 1$, of degree p and q respectively, and a white noise sequence $\{\tilde{w}_t\}$ such that $\{x_t\}$ satisfies the causal invertible equations.

Proof [5, p. 105.] Define

$$\tilde{\phi}(z) = \phi(z) \prod_{r<j\leq p} \frac{(1 - a_j z)}{(1 - a_j^{-1} z)},$$

$$\tilde{\theta}(z) = \theta(z) \prod_{s<k\leq q} \frac{(1-b_k z)}{(1-b_k^{-1} z)},$$

where $a_{r+1}, \ldots, a_p$ and $b_{s+t}, \ldots, b_q$ are the zeroes of $\phi(z)$ and $\theta(z)$ which lie inside the unit circle. Since $\tilde{\phi}(z) \neq 0$ and $\tilde{\theta}(z) \neq 0$ for all $|z| \leq 1$, it suffices to show that the process defined by

$$\begin{aligned}
\tilde{w}_t &= \tilde{\phi}(B)\tilde{\theta}(B)^{-1} x_t \\
&= \left(\prod_{r<j\leq p} \frac{(1-a_j B)}{(1-a_j^{-1} B)} \right) \left(\prod_{s<k\leq q} \frac{(1-b_k^{-1} B)}{(1-b_k B)} \right) w_t
\end{aligned}$$

is white noise. We can show (we use *Equation 1.2.3*) that the autocovariance generating function for $\{\tilde{w}_t\}$ is given by

$$\begin{aligned}
G(z) &= \sigma^2 \left(\prod_{r<j\leq p} \frac{(1-a_j z)}{(1-a_j^{-1} z)} \frac{(1-a_j z^{-1})}{(1-a_j^{-1} z^{-1})} \right) \left(\prod_{s<k\leq q} \frac{(1-b_k^{-1} z)}{(1-b_k z)} \frac{(1-b_k^{-1} z^{-1})}{(1-b_k z^{-1})} \right) \\
&= \sigma^2 \left(\prod_{r<j\leq p} \frac{(1-a_j(z+z^{-1})+a_j^2)}{(1-a_j^{-1}(z+z^{-1})+a_j^{-2})} \right) \left(\prod_{s<k\leq q} \frac{(1-b_k^{-1}(z+z^{-1})+b_k^{-2})}{(1-b_k(z+z^{-1})+b_k^2)} \right) \\
&= \sigma^2 \prod_{r<j\leq p} |a_j|^2 \prod_{s<k\leq q} |b_k|^{-2}.
\end{aligned}$$

Since $G(z)$ is constant, we conclude that $\{\tilde{w}_t\}$ is white noise as asserted.

Note that the definition of *ARMA* processes is quite formal and includes processes x_t which are based on the future innovations $\{w_\tau, \tau > t\}$ (future values). One can show that this approach is not applicable in case of real world situation, when we can express the process value using only the past values of the process. In order to show that we construct an example.

Example 1. *We choose a simple $AR(1)$ process*

$$x_t = 2x_{t-1} + w_t,$$

where w_t are independent of x_i, $i < t$.

According to proposed procedure the process defined by

$$w_t^* = (1 - 0.5B)x_t$$

should be a white noise process, but one can show that it's not true

$$\begin{aligned} w_t^* &= (1-0.5B)x_t \\ &= (1-0.5B)(2x_{t-1}+w_t) \\ &= 2x_{t-1}-x_{t-2}+w_t-0.5w_{t-1} \\ &= 2(2x_{t-2}+w_{t-1})-x_{t-2}+w_t-0.5w_{t-1} \\ &= 3x_{t-2}+1.5w_{t-1}+w_t \end{aligned}$$

We can also show that

$$x_t = 2^t x_0 + 2^{t-1}w_1 + 2^{t-2}w_2 + \ldots + w_t$$

and compute

$$Ex_t = 2^t Ex_0 = 0,$$

$$\begin{aligned} EX_t^2 &= E(2^t x_0 + 2^{t-1}w_1 + 2^{t-2}w_2 + \ldots + w_t)^2 \\ &= 2^{2t}Ex_0^2 + \sigma_w^2(2^{2(t-1)} + 2^{2(t-2)} + \ldots + 1) \\ &= 2^{2t}\sigma_w^2 + \sigma_w^2(2^{2(t-1)} + 2^{2(t-2)} + \ldots + 1) \\ &= \sigma_w^2(2^{2t} + 2^{2(t-1)} + 2^{2(t-2)} + \ldots + 1) \end{aligned}$$

If the process w_t^* is a white noise process then $Var(w_t^*) = const.$, but

$$Ew_t^* = E(3x_{t-2}+1.5w_{t-1}+w_t) = 3\cdot 2^{t-2}Ex_0 = 0.$$

$$\begin{aligned} E(w_t^*)^2 &= E(3x_{t-2}+1.5w_{t-1}+w_t)^2 \\ &= 9Ex_{t-2}^2 + 2.25\sigma_w^2 + \sigma_w^2. \\ &= 9(\sigma_w^2(2^{2(t-2)} + 2^{2(t-3)} + 2^{2(t-4)} + \ldots + 1)) + 3.25\sigma_w^2 \end{aligned}$$

$$Var(w_t^*) = E(w_t^*)^2 - (Ew_t^*)^2 = 9(\sigma_w^2(2^{2(t-2)}+2^{2(t-3)}+2^{2(t-4)}+\ldots+1))+3.25\sigma_w^2 \to \infty,\ t\to\infty.$$

So, one should be careful, when applying the procedure proposed in the proof. We conclude that the part of theorem that accords to MA process is applicable and we use this scheme later in the following way: since this proof is constructive, we can apply polynomial $\tilde{\theta}(z)$ in order to obtain corresponding invertible process for each non-invertible process with MA polynomial $\theta(z)$ and vice versa. So, we can construct an invertible MA process for each non-invertible MA process with the same covariance structure as that of the non-invertible process.

We can do some simulations in R, to verify this computationally. After writing down appropriate program and executing this code we get some nice graphs of processes (see **Figure 1**), autocovariance functions, partial autocovariance functions. The first two rows show a non- invertible process and the corresponding invertible

MA process. The last rows of graphs verifies that w^* obtained is a white noise process.

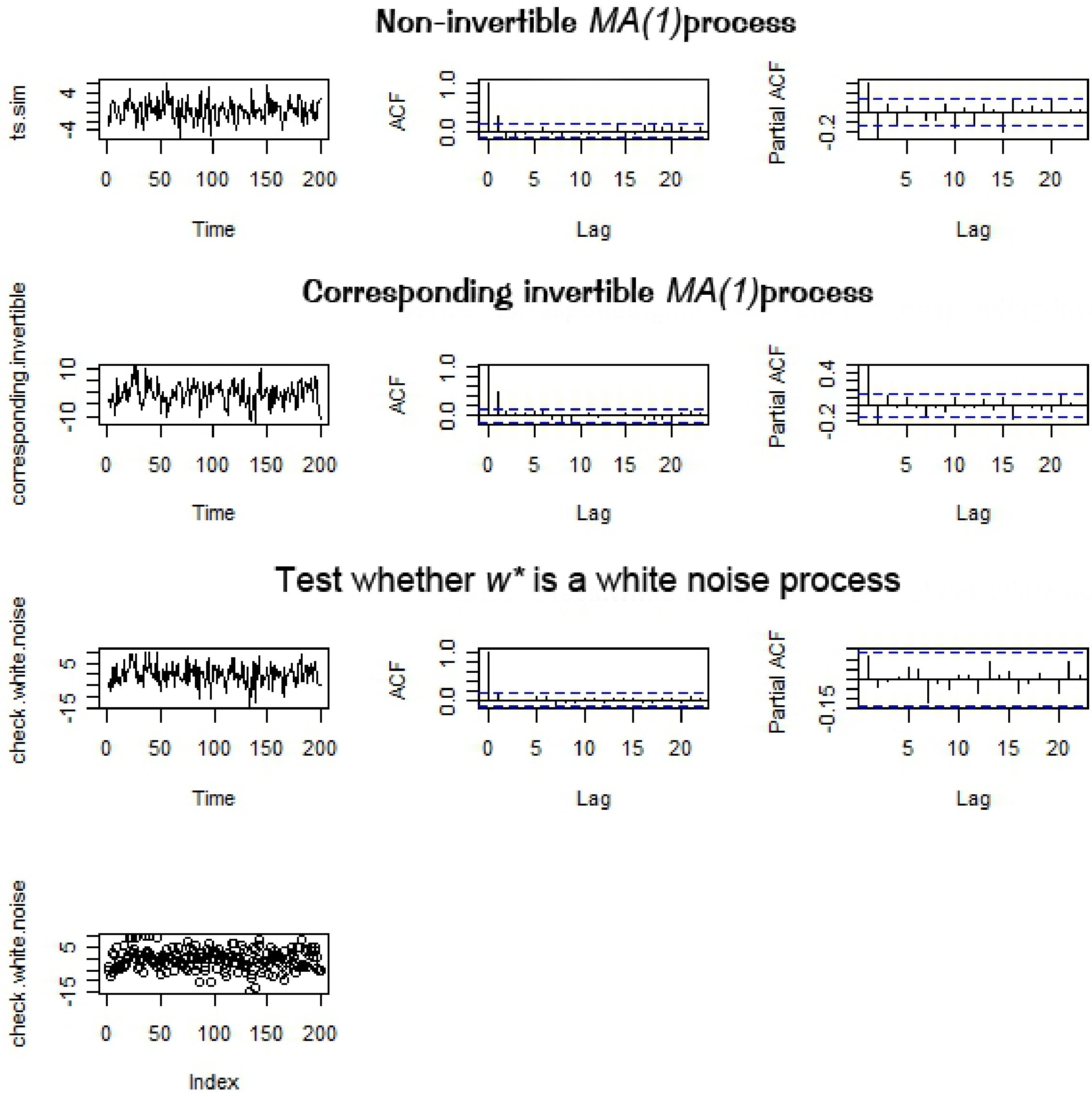

Figure 1.1 : Non-invertible MA, corresponding invertible MA and computational verification of theorem.

In the last graph we can see that it looks like a white noise.

2. Forecasting

This chapter describes appropriate approaches of forecasting $ARMA$ processes for both the invertible case and the non-invertible one, besides we differentiate the non-invertible case into two subcases: a case with Gaussian innovations and the non-Gaussian case of $MA(1)$ process with uniformly distributed random innovation series. We start this chapter with a small subsection about multivariate Gaussian distribution, where we recall the definition to agree on the notation and some basic properties needed for out further purposes.

2.1. Multivariate Gaussian

The assumption about Gaussian distributed process values is one of the basic assumptions when predicting time series. In this section we give basic definitions and characteristics of multivariate Gaussian distributed random variables that are used in further sections. We start with the definition of standard Gaussian vector, which helps to define multivariate Gaussian vector.

Definition 19. *[6, p. 2.] A random vector $X = (X_j)_{j=1}^n \in R^n$ is called standard Gaussian, if its components are independent and have a standard normal distribution. The distribution of X has a density*

$$p(x) = \frac{1}{(2\pi)^{n/2}} e^{\frac{-(x,x)}{2}},\ x \in R^n,$$

where (x, x) denotes the scalar product.

There exist two equivalent definitions of a general Gaussian vector in R^n.

Definition 20. *[6, p. 2.] A random vector $Y \in R^n$ is called Gaussian, if it can be represented as $Y = a + LX$, where X is a standard Gaussian vector, $a \in R^n$, and $L : R^n \to R^n$ is a linear mapping.*

Definition 21. *[6, p. 2.] A random vector $Y \in R^n$ is called Gaussian, if the scalar product (ν, Y) is a normal random variable for each $\nu \in R^n$.*

In this source [6, pp. 3-4.] we can find an useful explanation and basic properties of multivariate Gaussian. Similarly to the univariate notation $N(a, \sigma^2)$, the family of n-dimensional Gaussian distributions also admits a reasonable parametrization. Recall that for any random vector $Z = (Z_j) \in R^n$ one understands the expectation component-wise, i.e. $EZ = (EZ_j)$, while its covariance operator $K_Z : R^n \to R^n$ is defined by

$$cov((\nu_1, Z), (\nu_2, Z)) = (\nu_1, K_Z \nu_2).$$

If all components of a vector Z have finite second moments, then the expectation EZ and covariance operator KZ exists. There are no restrictions on the expectation value, while the covariance operator is necessarily non-negative definite and symmetric. In other words, there exists an orthonormal base (e_j) such that K has a diagonal form $K_Z e_j = \lambda_j e_j$ with $\lambda_j \geq 0$. We write $Y \sim N(a, K)$ if Y is a Gaussian vector with expectation a and covariance operator K. In particular, for a standard Gaussian vector we have $X \sim N(0, E_n)$, where $E_n : R^n \rightarrow R^n$ is the identity operator.

We continue with some properties given in [6, pp. 4.-5.] and start with the uniqueness of $N(a, K)$ that follows from the fact that a pair (a, K) determines the distribution of (ν, Y) as $N((\nu, a), (\nu, K\nu))$, hence by classical Cramer-Wold theorem (cf. Pranab et al [7]) the entire distribution is determined uniquely. So, if $X_1, X_2, \ldots X_p$ are multivariate Gaussian, then conditioning on $X_1, \ldots X_q$ gives the remaining variables $X_{q+1}, \ldots X_p$ a Gaussian distribution as well. This is a very useful property we are going to use in further sections.

Further we try to show, how this property of multivariate Gaussian and some results presented in the next section together with the transformation from non-invertible to invertible MA process can be used in order to obtain the one (or k) step ahead forecast of a non-invertible MA (and $ARMA$). The scheme is quite simple: we are going to find the invertible representation of the non-invertible process with the same covariance structure and then apply the property of the Gaussian distribution, by which the mean and covariance structure fully determines the distribution. Therefore we can apply usual forecast procedure for the obtained invertible representation of non-invertible process. This means, that in practice, if we have the prior information about Gaussian distributed random variables, we fit invertible model to the data and obtain the desired one (k) step ahead prediction. But let us do it step by step in the next sections.

2.2. Forecasting invertible Gaussian model

In this section we describe the usual approach of forecasting of an $ARMA$ process as well as describe the differences in forecasting a non-invertible MA process.

We follow the scheme given in [4, pp. 110-121.]. In the beginning we recall the goal, which is to predict future values of a time series, $x_{n+m}, m = 1, 2, ...$, where m denotes the process value m steps ahead, based on the data collected to the present, $x = \{x_n, x_{n-1}, ..., x_1\}$. Throughout this section, we will assume x_t is stationary and the model parameters are known. First, we define the measure, which gives us a possibility to compare two different predictions and gives us possibility to define the best one in this sense. In the following definition we declare that in this case we are interested in the minimum mean square error predictor (MMSEP).

Lemma 3. *The* minimum mean square error predictor (MMSEP) *of x_{n+m} is*

$$x_{n+m}^n = E(x_{n+m}|x_n, x_{n-1}, ..., x_1)$$

because the conditional expectation minimizes the mean square error

$$E[x_{n+m} - g(x)]^2,$$

where $g(x)$ is a function of the observations x.

The proof of this lemma is quite straightforward and can be found in [8, pp.121-122.].

First, we will restrict attention to predictors that are linear functions of the data, that is, predictors of the form

$$x_{n+m}^n = a_0 + \sum_{k=1}^{n} a_k x_k,$$

where $a_0, a_1, ..., a_n$ are real numbers. Linear predictors of the form that minimize the mean square prediction error are called best linear predictors (BLPs). As we shall see, linear prediction depends only on the second-order moments of the process, which are easy to estimate from the data.

Before that, let us step back and recall some results from algebra, we are going to use to prove the equivalence of the MMSEPs and BLPs. We start with the basic projection theorem in the Hilbert space.

Theorem 2. *(Projection theorem) [4, pp. 523.] Let M be a closed subspace of the Hilbert space H and let y be an element in H. Then, y can be uniquely represented as*

$$y = \hat{y} + z,$$

where $\hat{y}$ belongs to M and z is orthogonal to M; that is, $(z, w) = 0 \, \forall w \in M$. Furthermore, the point $\hat{y}$ is closest to y in the sense that, for any w in M, $||y - w|| \geq ||y - \hat{y}||$, where equality holds if and only if $w = \hat{y}$.

Using the notation of the theorem, we call the mapping $P_M y = \hat{y}$, for $y \in H$, the projection mapping of H onto M. In addition, the closed span of a finite set $\{x_1, ..., x_n\}$ of elements in a Hilbert space, H, is defined to be the set of all linear combinations $w = a_1 x_1 + \ldots + a_n x_n$, where $a_1, ..., a_n$ are scalars. This subspace of H is denoted by $M = \overline{sp}\{x_1, ..., x_n\}$.

Now we are ready to prove the equivalence of the MMSEPs and BLPs in case of Gaussian distributed process values. the result is formulated in the following theorem.

Theorem 3. *[4, pp. 526.] Under the established notation and conditions, if $(y, x_1, ..., x_n)$ is multivariate normal, then*

$$E(y|x_1, ..., x_n) = P_{\overline{sp}}\{1, x_1, ..., x_n\}y.$$

Before the proof of this theorem, let us recall two useful properties of conditional expectation $E(Y|Z)$.

Property 1. *Let Y be a random variable with $E(|Y|) < \infty$ and $\mathcal{G}$ be a sigma algebra, then*

- *(Taking out what is known.) If Y is $\mathcal{G}$ measurable and bounded, then $E(YZ|\mathcal{G}) = YE(Z|\mathcal{G})$, $P - a.s.$*
- *(Independence rule) If Y is independent of $\mathcal{G}$, then $E(Y|\mathcal{G}) = E(Y)$, $P - a.s.$*

Proofs of the properties of the conditional expectation can be found in [9, pp. 48-50.]

Proof [4, pp. 526.] First, by the projection theorem, the conditional expectation of y given $x = \{x_1, \ldots, x_n\}$ is the unique element $E(y|x_1, \ldots, x_n)$ that satisfies the orthogonality principle. We will show that $\hat{y} = P_{sp}\{1, x_1, ..., x_n\}y$ is that element. Consider

$$E(y|x_1, \ldots, x_n) = E(y - \hat{y} + \hat{y}||x_1, \ldots, x_n).$$

In fact, by the projection theorem, $\hat{y}$ satisfies

$$(y - \hat{y}, x_i) = 0, \ \ i = 0, 1, \ldots, n,$$

where we have set $x_0 = 1$. But $(y - \hat{y}, x_i) = cov(y - \hat{y}, x_i) = 0$, implying that $y - \hat{y}$ and $(x_1, ..., x_n)$ are independent because the vector $(y - \hat{y}, x_1, ..., x_n)^T$ is multivariate normal. Thus, if $y - \hat{y}$ are independent we continue with

$$E(y|x_1, \ldots, x_n) = E(y - \hat{y} + \hat{y}|x_1, \ldots, x_n) = E(y - \hat{y}) + E(\hat{y}|x_1, \ldots, x_n),$$

and now we have applied the property of independence. We can do more and apply the property, that says that we can take out what is known. Here $\hat{y}$ is a linear combination of $x_1, \ldots, x_n$, therefore it is $x_1, \ldots, x_n$ measurable. Recall, that $0 = (y - \hat{y}, 1) = E(y - \hat{y})$, hence,

$$E(y|x_1, \ldots, x_n) = E(y - \hat{y}) + E(\hat{y}|x_1, \ldots, x_n) = 0 + \hat{y} = \hat{y},$$

which completes the proof.

This theorem states that if the process is Gaussian, minimum mean square error predictors and best linear predictors are the same.

Hence we can continue with the following property, which is based on the projection theorem and gives us the system of equations, where the solution of this system is a long awaited predictions.

Property 2. *[4, pp. 111.] Given data $x_1, ..., x_n$, the best linear predictor,*

$$x_{n+m}^n = a_0 + \sum_{k=1}^{n} a_k x_k,$$

of x_{n+m}, for $m \geq 1$, is found by solving

$$E[(x_{n+m} - x_{n+m}^n)x_k] = 0, \ \ k = 0, 1, ..., n,$$

where $x_0 = 1$.

These equations are called the prediction equations, and they are used to solve for the coefficients $\{a_0, a_1, ..., a_n\}$. If $E(x_t) = \mu$, the first equation ($k = 0$) implies

$$E(x_{n+m}^n) = E(x_{n+m}) = \mu.$$

Thus, taking expectation, we have

$$\mu = a_0 + \sum_{k=1}^{n} a_k \mu \ \ or \ \ a_0 = \mu \left(1 - \sum_{k=1}^{n} a_k\right).$$

Hence, the form of the BLP is

$$x_{n+m}^n = \mu + \sum_{k=1}^{n} a_k (x_k - \mu).$$

Thus, until we discuss estimation, there is no loss of generality in considering the case that $\mu = 0$, in which case, $a_0 = 0$.

2.2.1. One step ahead prediction

Again we follow the scheme given in [4, pp. 112-115.]. Consider, first, one-step-ahead prediction. That is, given $\{x_1, ..., x_n\}$, we wish to forecast the value of the time series at the next time point, x_{n+1}. The BLP of x_{n+1} is

$$x_{n+1}^n = \phi_{n1} x_n + \phi_{n2} x_{n-1} + \ldots + \phi_{nn} x_1,$$

where, for purposes that will become clear shortly, we have written a_k as $\phi_{n,n+1-k}$, for $k = 1, ..., n$. Using property described above, the coefficients $\{\phi_{n1}, \phi_{n2}, \ldots, \phi_{nn}\}$ satisfy

$$E\left[\left(x_{n+1} - \sum_{j=1}^{n} \phi_{nj} x_{n+1-j}\right) x_{n+1-k}\right] = 0,\ k = 1, ..., n,$$

or

$$\sum_{j=1}^{n} \phi_{nj} \gamma(k-j) = \gamma(k), k = 1, ..., n.$$

The prediction equations can be written in matrix notation as

$$\Gamma_n \phi_n = \gamma_n, \tag{2.2.1}$$

where $\Gamma_n = \{\gamma(k-j)\}_{j,k=1}^{n}$ is an $n \times n$ matrix, $\phi_n = (\phi_{n1}, ..., \phi_{nn})^T$ is an $n \times 1$ vector, and $\gamma_n = (\gamma(1), ..., \gamma(n))^T$ is an $n \times 1$ vector. The matrix Γ_n is non-negative definite. If Γ_n is singular, there are many solutions for these equations, but, by the projection theorem, x_{n+1}^n is unique. If Γ_n is non-singular, the elements of ϕ_n are unique, and are given by

$$\phi_n = \Gamma_n^{-1} \gamma_n.$$

For $ARMA$ models, the fact that $\sigma_w^2 > 0$ and $\gamma(h) \to 0$ as $h \to \infty$ is enough to ensure that Γ_n is positive definite (an additional information about non-singularity can be found in [10, pp. 74-75.]). It is sometimes convenient to write the one-step-ahead forecast in vector notation

$$x_{n+1}^n = \phi_n^T x,$$

where $x = (x_n, x_{n-1}, ..., x_1)^T$.

2.3. Non-invertible Gaussian processes

So far we have described all necessary parts to describe the prediction scheme of a non-invertible Gaussian $ARMA$ process. Here we describe briefly the scheme of reaching one step ahead prediction, but without loss of generality the scheme remains the same also for m steps ahead.

First, we recall that according to *Theorem 1* and conclusions for each non-invertible process there exists an invertible process with the same covariance structure. So, if we fit the invertible model to given data, we get this invertible representation even when the underlying process is non-invertible.

Secondly, we recall that in *Section 2.1.* we agreed that Gaussian distribution is unique and it is fully determined by the covariance matrix and vector of mean values. Therefore the distribution is the same for bot the fitted invertible model and the non-invertible one.

As the third step in this argumentation we point out that for Gaussian processes the minimum mean square error predictor and the best linear predictor is the same. This result is given in *Theorem 3*. When we have presented this scheme, we can find the minimum mean square error predictions by *Equation 2.2.1*.

But there is one special case, which is usually called “strictly” non-invertible case, which takes some more attention and particular consideration.

2.3.1. Strictly non-invertible $MA(1)$

Here we consider the case of a strictly non-invertible $MA(1)$ process. The term “strictly non-invertible process” refers to the situation when MA polynomial has one or more unit roots. Some more information and different approach of analysing strictly non-invertible processes can be found in Plosser and Schwert publication [11].

But in this subsection we are going to look at one specific example to illustrate the situation.

Predictions by strictly non-invertible $MA(1)$

This corresponds to the forecasting of the process with unit root in MA polynomial. Predicting MA processes with unit root in the MA polynomial do not cause a lot of problems in the simplest cases but there are some conclusions that clarify the situation. Let us consider the $MA(1)$ process

$$x_t = w_t - w_{t-1}.$$

Recall (compute), that for this process

$$\gamma(h) = \begin{cases} 2\sigma_w^2,\ h = 0, \\ -\sigma_w^2,\ h = \pm 1, \\ 0,\ |h| \geq 2. \end{cases}$$

Therefore,

$$\Gamma_n = \sigma_w^2 \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ \dots & & \dots & & \dots & \\ 0 & 0 & \dots & 0 & -1 & 2 \end{pmatrix}.$$

Let's consider case $n = 1$ when we have observations (process values) x_1 and x_2, then

$$x_2^1 = \phi_{11}x_1 = \gamma(1)\gamma^{-1}(0)x_1 = -\frac{1}{2}x_1.$$

And let us consider also case $n = 2$ when we have observations x_1 and x_2, then

$$x_3^2 = \phi_{21}x_2 + \phi_{22}x_1 = (\gamma(1), \gamma(2))\Gamma^{-1}(x_2, x_1)^T.$$

This time ($n = 2$)

$$\Gamma^{-1} = \sigma_w^{-2}\frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Thus,

$$\begin{aligned} x_3^2 &= (\gamma(1), \gamma(2))\Gamma^{-1}(x_2, x_1)^T = (-\sigma_w^2, 0)\sigma_w^{-2}\frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}(x_2, x_1)^T \\ &= \frac{-2}{3}x_2 - \frac{1}{3}x_1 \end{aligned}$$

Hence, for finite dimensional case we can continue this procedure.

At first let us state a lemma that will help us to get required inversion of $k \times k$ symmetric tridiagonal matrices. We are looking to apply this result in the case of our covariance matrix.

Lemma 4. *[12, pp. 1511-1513] Let M_k be a tridiagonal matrix of the form*

$$M_k = \begin{pmatrix} D & 1 & 0 & 0 & 0 & \dots & & & & 0 \\ 1 & D & 1 & 0 & 0 & \dots & & & & 0 \\ 0 & 1 & D & 1 & 0 & \dots & & & & 0 \\ \dots & & & & & \dots & & & & \dots \\ & & & & & & & & & \\ \dots & & & & & \dots & & & & \dots \\ 0 & 0 & 0 & & & \dots & 0 & 1 & D & 1 & 0 \\ 0 & 0 & 0 & & & \dots & 0 & 0 & 1 & D & 1 \\ 0 & 0 & 0 & & & \dots & 0 & 0 & 0 & 1 & D \end{pmatrix},$$

let also $\mathcal{M}_k = det(M_k)$, then

- *(i)$\mathcal{M}_{i+1} = D\mathcal{M}_i - \mathcal{M}_{i-1}$ with boundary conditions $\mathcal{M}_0 = 1$, $\mathcal{M}_1 = D$*
- *(ii)and the inversion of M_k in case of $D = -2$ is given by $M^{-1} = R = (r_{ij})$, where*

$$r_{ij} = -\frac{(i + j - |j - i|)(2k + 2 - |j - i| - i - j)}{4(k + 1)}.$$

Let us now show that in case of $D = -2$ we get $\mathcal{M}_k = (-1)^k(k + 1)$. We apply here the principle of mathematical induction: **Proof**

- Base: $\mathcal{M}_0 = 1$, $\mathcal{M}_1 = -2$ follows directly from the lemma.

- Inductive step:

  (i) Assume that $\mathcal{M}_n = (-1)^n(n+1)$.

  (ii) Show that $\mathcal{M}_{n+1} = (-1)^{n+1}(n+2)$. We apply lemma and get

$$\begin{aligned}\mathcal{M}_{n+1} &= -2\mathcal{M}_n - \mathcal{M}_{n-1}\\ &= -2(-1)^n(n+1) - (-1)^{n-1}n\\ &= (-1)^{n+1}(n+2).\end{aligned}$$

Now we continue our example of predicting MA in case of $(n = N)$, $N \in \mathbb{R}$. Note that $\Gamma_N = -\sigma_w^2 M_N$ with D=-2. Hence $det(\Gamma_N) = (-1)^{N-1}(N+1)\sigma_w^2$ and the inversion of Γ_N can be expressed $\Gamma_N^{-1} = R = (r_{ij})$, where

$$r_{ij} = \frac{(i+j-|j-i|)(2N+2-|j-i|-i-j)}{4(N+1)\sigma_w^2}.$$

Furthermore we can rewrite the formula of r_{ij} because $det(\Gamma_N) = (-1)^{N-1}(N+1)$. Therefore,

$$r_{ij} = \frac{(i+j-|j-i|)(2N+2-|j-i|-i-j)}{4|det(\Gamma_N)|}.$$

Hence

$$x_N^{N+1} = (\gamma(1), \gamma(2), \ldots, \gamma(N))\Gamma_N^{-1}(x_N, x_{N-1}, \ldots, x_1)^T,$$

and we have $(\gamma(1), \gamma(2), \ldots, \gamma(N)) = \sigma_w^2(-1, 0, \ldots, 0)$, we only need the first line of Γ_N^{-1} in order to get prediction equation. Let us denote $\Gamma_N^{-1} = (g_1, g_2, \ldots, g_N)^T$, where $g_i = (r_{ij})$, $j = 1, 2 \ldots, N$ denotes rows of matrix Γ_N^{-1}. This means that we need only g_1 and now it is easy to show that one can use formulas of r_{ij} and obtain

$$g_1 = \frac{1}{|det\Gamma_N|}(N, N-1, \ldots, 1).$$

Finally, we get the prediction of form

$$\begin{aligned}x_N^{N-1} &= (\gamma(1), \gamma(2), \ldots, \gamma(N))\Gamma_N^{-1}(x_N, x_{N-1}, \ldots, x_1)^T\\ &= \frac{-1}{(N+1)}\left((N, N-1, \ldots, 1)(x_N, x_{N-1}, \ldots, x_1)^T\right)\\ &= \frac{-1}{(N+1)}\left(Nx_N + (N-1)x_{N-1} + \ldots + x_1\right).\end{aligned}$$

In conclusion we can say that

- the plus is that the form of prediction is clear and can be expressed as exact formula,
- the main drawback is that the prediction is linearly dependent of all history, which is not a good property because we have to have all the history to get a reasonable prediction (Recall, that in case of invertible process the weights are exponential and therefore we get better estimates, because the convergence is fast enough, but in this case, as it is in case of the harmonic series $\sum_{n=1}^{\infty} 1/n$, the sum of weights is divergent. Therefore we cannot fix a time point, when the rest part of the process do not influence the final result, even if we set the the length of used process values very large.).

On the other hand in real world examples at such situation, when we have process with a unit root in MA polynomial are quite unlikely, so in conclusion we can say, that in case of Gaussian processes the derivation of the prediction equations is well defined and in the most of the cases gives us reasonable results.

2.4. Non-invertible non-Gaussian $MA(1)$ process

To extend studies of predictions by non-invertible $ARMA$ processes we continue with the analysis of non-Gaussian non-invertible $ARMA$ processes. As expected such analysis is not common in the literature, but there are some research papers dedicated to this kind of problems, e.g., Breidt and Hsu [13].

We concentrate on the most simple case, that is described in the following example. This example will illustrate some difficulties faced during the derivation process and the gains of derived formula can be found in subsection, where the computer simulation results are presented. Actually, this example can be considered as a case study of such processes.

2.4.1. Predictions by non-invertible non-Gaussian MA(1)

What happens when we violate the restriction about Gaussian distributed innovations in MA process? Let's look at an example

$$x_t = w_t - \theta w_{t-1}.$$

In this example we consider case $w_t \sim_{i.i.d.} U[0,1]$, $\theta > 1$. Note, that this corresponds to the $MA(1)$ process

$$x_t = 0.5(1-\theta) + (w_t - 0.5) - \theta(w_{t-1} - 0.5)$$

with zero mean as defined in *Section 1.* At the beginning we try to find joint distribution density function and then conditional distribution density function

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_2,X_1}(x_2,x_1)}{f_{X_1}(x_1)}.$$

In this case joint distribution density function f_{x_1} is quite easy to find (recall that $x_1 = w_1 - \theta w_0$). Let us rewrite this process x_1 as following $x_1 = Z = X + Y$, where $X \sim U[0,1]$ and $Y \sim U[-\theta, 0]$, then

$$\begin{aligned}
F_Z(z) &= \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy \\
& \quad Since\, f_Y(y) = \frac{1}{\theta},\ -\theta \le y \le 0,\ and\, 0\, otherwise \\
&= \frac{1}{\theta}\int_{-\theta}^{0} F_X(z-y)dy \\
& \quad Apply\, change\, of\, variables\, t := z - y \\
&= -\frac{1}{\theta}\int_{z+\theta}^{z} F_X(t)dt \\
&= \frac{1}{\theta}\int_{z}^{z+\theta} F_X(t)dt.
\end{aligned}$$

Now the integrand is 0, unless $-\theta \le z \le 1$, therefore we split integral by domains:

- if $z \le -\theta$:

$$\begin{aligned}
F_Z(z) &= \frac{1}{\theta}\int_{z}^{z+\theta} F_X(t)dt \\
&= \frac{1}{\theta}\int_{z}^{z+\theta} 0dt = 0
\end{aligned}$$

- if $-\theta \le z \le -\theta+1$:

$$\begin{aligned} F_Z(z) &= \frac{1}{\theta}\int_z^{z+\theta} F_X(t)dt \\ &= \frac{1}{\theta}\int_z^0 0dt + \frac{1}{\theta}\int_0^{z+\theta} tdt \\ &= \frac{(z+\theta)^2}{2\theta} \end{aligned}$$

- if $-\theta+1 \le z \le 0$:

$$\begin{aligned} F_Z(z) &= \frac{1}{\theta}\int_z^{z+\theta} F_X(t)dt \\ &= \frac{1}{\theta}\int_z^0 0dt + \frac{1}{\theta}\int_0^1 tdt + \frac{1}{\theta}\int_1^{z+\theta} 1dt \\ &= \frac{1}{\theta}(z+\theta-0.5) \end{aligned}$$

- if $0 \le z \le 1$:

$$\begin{aligned} F_Z(z) &= \frac{1}{\theta}\int_z^{z+\theta} F_X(t)dt \\ &= \frac{1}{\theta}\int_z^1 tdt + \frac{1}{\theta}\int_1^{z+\theta} 1dt \\ &= \frac{2\theta-(1-z)^2}{2\theta} \end{aligned}$$

- and also if $z \ge 1$:

$$\begin{aligned} F_Z(z) &= \frac{1}{\theta}\int_z^{z+\theta} F_X(t)dt \\ &= \frac{1}{\theta}\int_z^{z+\theta} 1dt = 1. \end{aligned}$$

Thus,

$$F_{X_1}(z) = F_{X+Y}(z) = \begin{cases} 0, & z \le -\theta \\ \frac{(z+\theta)^2}{2\theta}, & -\theta \le z \le -\theta+1 \\ \frac{1}{\theta}(z+\theta-0.5), & -\theta+1 \le z \le 0 \\ \frac{2\theta-(1-z)^2}{2\theta}, & 0 \le z \le 1 \\ 1, & z \ge 1 \end{cases}$$

Now we have obtained $F_{X_1}(x_1)$, and have to continue with joint distribution $F_{X_1,X_2}(x_1,x_2)$. We recall that x_1 and x_2 depend on w_0, w_1, w_2, where $w_i \sim U[0,1]$

and try to find

$$F_{X_1,X_2}(x_1,x_2) = P(w_1 - \theta w_0 \leq x_1, w_2 - \theta w_1 \leq x_2) = \iiint\limits_D dw_0 dw_1 dw_2,$$

where D is defined by the intersection of the unit cube $[0,1]^3$ and hyperplanes defined by x_1 and x_2.

To illustrate this situation, we can look at the geometrical interpretation of the joint distribution function. Turns out, that joint distribution function describes the volume that is defined by two hyperplanes and the unit cube. These two hyperplanes are defined by the equations of x_1 and x_2, where x_1 depends on uniformly distributed $U[0,1]$ random variables w_0 and w_1, and x_2, respectively, depends on w_1 and w_2. If we construct unit cube with the origin $(w_0 = 0, w_1 = 0, w_2 = 0)$ and plot the hyperplanes defined by the equations of x_1 and x_2, we can determine the volume of interest for the pairs of x_1 and x_2 and set up the triple integrals to find the value of them. Unfortunately we have to deal with different cases defined by different regions of x_1.

The three main different cases are show in the ***Figure 2.1***. The first cube corresponds to the following derivation process, where we put a restriction on x_1 such that $\{-\theta \leq x_1 \leq -\theta + \frac{1}{\theta}\}$. The other two cubes describes the situation, when $x_1 \in [1-\theta, 0)$ (the cube with blue and red hyperplanes and quadrilateral basis) and $x_1 \in [0,1)$ (the cube with violet and red hyperplanes and quadrilateral basis (and small triangular part left outside)).

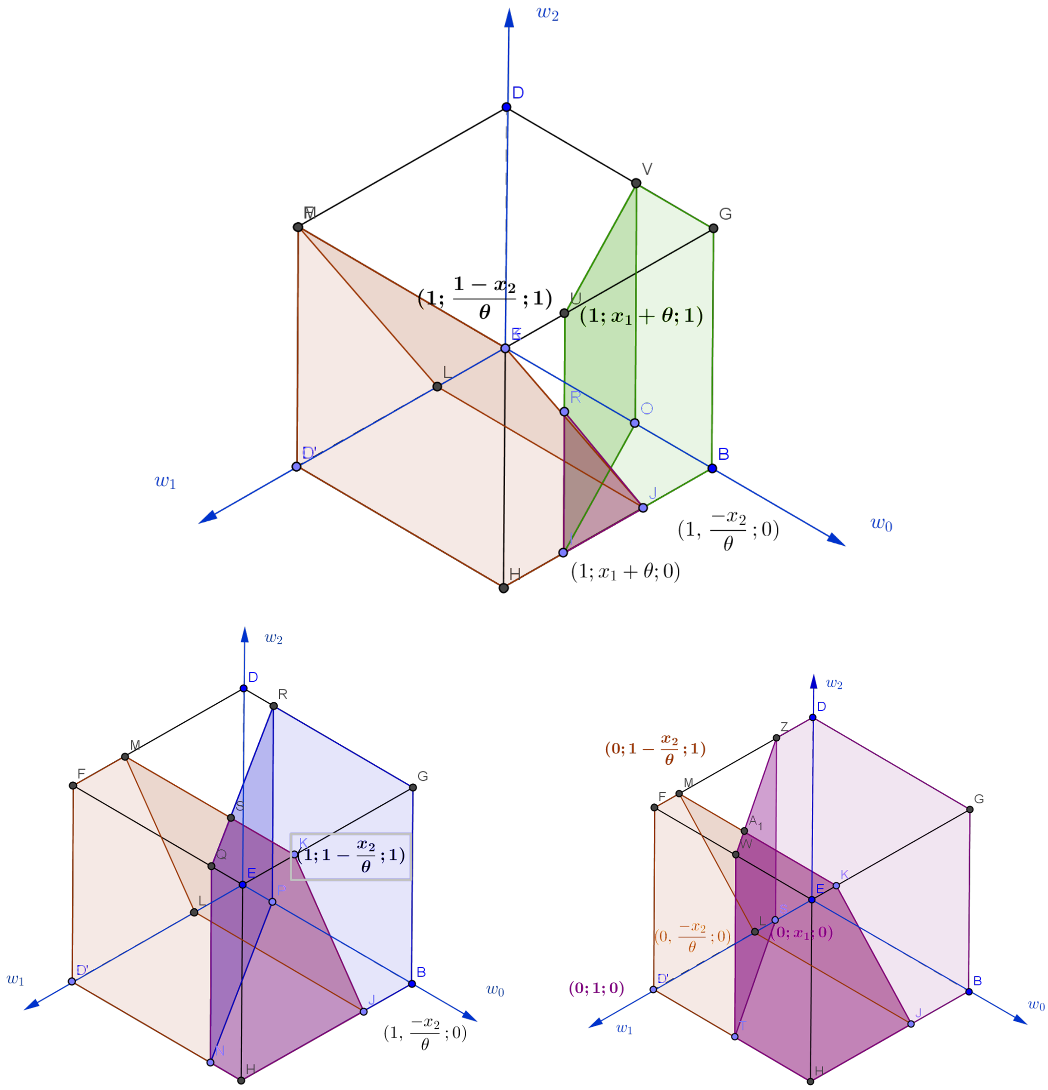

Figure 2.1 Triangular base; Quadrilateral base; Quadrilateral base

We are going to fully present one case, but all the necessary derivations and brief clarifications and commentaries can be found in the appendix, where this derivation part is given as several pieces of code of `Sage` package [3] (for each of the regions).

One can show that $-\theta \le X_1 \le 1$. Here we consider only the band $\{-\theta \le x_1 \le -\theta + \frac{1}{\theta}\}$. From the model we know that also $-\theta \le X_2 \le 1$, but we split this interval into smaller pieces in order to get the distribution function over $[-\theta \le x_1 \le -\theta + \frac{1}{\theta}] \times [-\theta \le x_2 \le 1]$. Thus we continue with case analysis (this describes different placement of the red hyperplane in ***Figure 2.1***, when the green one is defined):

1. $\{-\theta \le x_1 \le -\theta + \frac{1}{\theta}\}$ and $\{-\theta \le x_2 < -\theta(X_1 + \theta)\}$, for this intersection

$$F_{X_2,X_1}(x_2, x_1) = 0$$

2. $\{-\theta \leq x_1 \leq -\theta + \frac{1}{\theta}\}$ and $\{-\theta(x_1+\theta) \leq x_2 < 0\}$, for this intersection we set integral

$$F_{X_2,X_1}(x_2,x_1) = \int_{-\frac{x_2}{\theta}}^{x_1+\theta} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{x_2+\theta w_1} dw_2 dw_0 dw_1$$
$$= \frac{(\theta^3 x_1^3 + 3\theta^4 x_1^2 + 3\theta^5 x_1 + \theta^6 + x_2^3 + 3x_2\theta^2 x_1^2 + 6x_2\theta^3 x_1 + 3x_2\theta^4 + 3x_1\theta x_2^2 + 3\theta^2 x_2^2)}{6\theta^3}$$

3. $\{-\theta \leq x_1 \leq -\theta + \frac{1}{\theta}\}$ and $\{0 \leq x_2 < 1 - \theta(x_1+\theta)\}$, for this intersection we set integral

$$\begin{aligned} F_{X_2,X_1}(x_2,x_1) &= \int_0^{x_1+\theta} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{x_2+\theta w_1} dw_2 dw_0 dw_1 \\ &= \frac{(x_1+\theta)^2(x_1\theta + \theta^2 + 3x_2)}{6\theta} \end{aligned}$$

4. $\{-\theta \leq x_1 \leq -\theta + \frac{1}{\theta}\}$ and $\{1 - \theta(x_1+\theta) \leq x_2 < 1\}$, for this intersection we set integral

$$F_{X_2,X_1}(x_2,x_1) = \int_0^{\frac{1-x_2}{\theta}} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{x_2+\theta w_1} dw_2 dw_0 dw_1 + \int_{\frac{1-x_2}{\theta}}^{x_1+\theta} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{1} dw_2 dw_0 dw_1$$
$$= \frac{(-3x_1\theta x_2^2 + 6\theta^2 x_2 + 3\theta^4 - x_2^3 - 3x_2)}{6\theta^3}$$
$$+ \frac{(-3x_1\theta - 3\theta^2 + 6x_1\theta x_2 + 3\theta^2 x_1^2 + 6\theta^3 x_1 + 1 - 3\theta^2 x_2^2 + 3x_2^2)}{6\theta^3}$$

5. $\{-\theta \leq x_1 \leq -\theta + \frac{1}{\theta}\}$ and $\{x_2 \geq 1\}$, for this intersection we set integral

$$\begin{aligned} F_{X_2,X_1}(x_2,x_1) &= \int_{-\frac{x_1}{\theta}}^{1} \int_0^{x_1+\theta w_0} \int_0^{1} dw_2 dw_1 dw_0 \\ &= \frac{(x_1^2 + 2\theta x_1 + \theta^2)}{2\theta}. \end{aligned}$$

Thus, for $-\theta \leq x_1 \leq -\theta + \frac{1}{\theta}$ we have obtained:

$$F_{X_2,X_1}(x_2,x_1) =$$
$$\begin{cases} 0, & -\theta \le x_2 < -\theta(x_1+\theta) \\ \frac{(\theta^3x_1^3+3\theta^4x_1^2+3\theta^5x_1+\theta^6+x_2^3+3x_2\theta^2x_1^2+6x_2\theta^3x_1+3x_2\theta^4+3x_1\theta x_2^2+3\theta^2x_2^2)}{6\theta^3}, & -\theta(x_1+\theta) \le x_2 < 0 \\ \frac{(x_1+\theta)^2(x_1\theta+\theta^2+3x_2)}{6\theta}, & 0 \le x_2 < 1-\theta(x_1+\theta) \\ \frac{(-3x_1\theta x_2^2+6\theta^2x_2+3\theta^4-x_2^3-3x_2)}{6\theta^3} & \\ +\frac{(-3x_1\theta-3\theta^2+6x_1\theta x_2+3\theta^2x_1^2+6\theta^3x_1+1-3\theta^2x_2^2+3x_2^2)}{6\theta^3}, & 1-\theta(x_1+\theta) \le x_2 < 1 \\ \frac{(x_1^2+2\theta x_1+\theta^2)}{2\theta}, & x_2 \ge 1 \end{cases}$$

Now we have obtained joint distribution $F_{X_1,X_2}(x_1,x_2)$.
We can obtain conditional density function:

$$f_{X_2}(x_2|X_1=c) = \frac{f_{X_2,X_1}(x_2,c)}{f_{X_1}(c)}.$$

From previous calculations we know that

$$F_{X_1}(x_1) = \begin{cases} 0, & x_1 \le -\theta \\ \frac{(x_1+\theta)^2}{2\theta}, & -\theta \le x_1 \le -\theta+1 \\ \frac{1}{\theta}(x_1+\theta-0.5), & -\theta+1 \le x_1 \le 0 \\ \frac{2\theta-(1-x_1)^2}{2\theta}, & 0 \le x_1 \le 1 \\ 1, & x_1 \ge 1 \end{cases}$$

In our case we are interested only in region $-\theta \le x_1 \le -\theta+\frac{1}{\theta}$, so, in the function

$$F_{X_1}(x_1) = \frac{(x_1+\theta)^2}{2\theta}.$$

and therefore in this case we have

$$f_{X_1}(x_1) = \frac{(x_1+\theta)}{\theta}.$$

and

$$f_{X_2,X_1}(x_2,x_1) = \begin{cases} 0, & -\theta \le x_2 < -\theta(x_1+\theta) \\ \frac{x_2+\theta x_1+\theta^2}{\theta^2}, & -\theta(x_1+\theta) \le x_2 < 0 \\ \frac{x_1+\theta}{\theta}, & 0 \le x_2 < 1-\theta(x_1+\theta) \\ -\frac{x_2-1}{\theta^2}, & 1-\theta(x_1+\theta) \le x_2 < 1 \\ 0, & x_2 \ge 1 \end{cases}$$

We can choose $x_1 = c \in [-\theta, \frac{1}{\theta} - \theta]$ and then obtain a conditional density function of x_2 given $x_1 = c$:

$$f_{X_2}(x_2|X_1 = c) = \frac{f_{X_2,X_1}(x_2, c)}{f_{X_1}(c)} = \begin{cases} 0, & -\theta \leq x_2 < -\theta(c+\theta) \\ \frac{(c\theta + x_2 + \theta^2)}{(\theta(c+\theta))}, & -\theta(c+\theta) \leq x_2 < 0 \\ 1, & 0 \leq x_2 < 1 - \theta(c+\theta) \\ \frac{(1-x_2)}{\theta(c+\theta)}, & 1 - \theta(c+\theta) \leq x_2 < 1 \\ 0, & x_2 \geq 1 \end{cases}.$$

And, finally,

$$E(X_2|X_1 = c) = \frac{1}{2}(1 - c\theta - \theta^2),\ c \in [-\theta, \frac{1}{\theta} - \theta].$$

Similarly we continue for each domain of x_1 (full derivation can be found in appendix). But here we present the final result. We have obtained the least mean square error one step ahead estimate. Therefore

$$E(X_2|X_1 = c) = \begin{cases} 0, & c \in (-\infty - \theta) \\ 0.5(1 - c\theta - \theta^2), & c \in [-\theta, 1 - \theta) \\ 0.5(1 - \theta), & c \in [1 - \theta, 0) \\ 0.5(-c\theta - \theta + 1,), & c \in [0, 1) \\ 0, & c \in (1, \infty). \end{cases} \tag{2.4.1}$$

Actually, we can see that the minimum mean square error one step ahead prediction is not the same as the best linear prediction in all cases. In further computational case studies we are going to show that this difference can be rather large in case of the non-invertible $MA(1)$ model. We are also going to show that the one step ahead prediction does not improve if we base our prediction on all previous values of the process $\{x_t\}$.

2.4.2. Simulation studies

In this section we are going to compare the best linear prediction and the minimum mean square error predictor by simulation studies of predicting the MA process (described in last *Subection 2.4.1.*) values one step ahead.

Let us start just with the graph of the conditional expected value represented in *Equation 2.4.1* and the graph of the best linear predictor (BLP), which is based on the previous process value and obtained by the linear regression of x_t on x_{t-1}. Recall, that the coefficient estimates of the linear regression of the simple linear model $x_t = \beta x_{t-1} + \alpha$ are

$$\hat{\beta} = \frac{Cov(x_{t-1}, x_t)}{Var(x_{t-1})}, \quad \hat{\alpha} = Ex_t - \hat{\beta}Ex_{t-1}.$$

The derivation of coefficients and interpretation in terms of the covariance and mean value can be found in [8, pp. 4-7] and the derivation of coefficients and its generalization in [14, pp. 17-19., 130-135.] In this case, when we have $x_t = w_t - \theta w_{t-1}$, where $w_i \sim i.i.d.U[0,1], i = 0, 1, \ldots, t$, we can get coefficient estimates $\hat{\beta}$ and $\hat{\alpha}$. To simplify the calculations, we denote $\tilde{w}_i = w_i - 0.5, i = 0, 1, 2, \ldots, t$. Note, that in this case $E\tilde{w}_i = 0, i = 0, 1, 2, \ldots, t$ and $Var(\tilde{w}_i) = E\tilde{w}_i^2 = 1/12, i = 0, 1, 2, \ldots, t$. Therefore

$$\hat{\beta} = \frac{Cov(x_{t-1}, x_t)}{Var(x_{t-1})} = \frac{E[(\tilde{w}_{t-1} - \theta\tilde{w}_{t-2})(\tilde{w}_t - \theta\tilde{w}_{t-1})]}{E(\tilde{w}_{t-1} - \theta\tilde{w}_{t-2})^2} = \frac{-\frac{1}{12}\theta}{\frac{1}{12}(1+\theta^2)} = \frac{-\theta}{(1+\theta^2)}$$

and

$$\hat{\alpha} = Ex_t - \hat{\beta}Ex_{t-1} = 0.5(1-\theta) - \hat{\beta}0.5(1-\theta) = 0.5(1-\theta)(1 + \frac{\theta}{1+\theta^2}).$$

Finally, we get the best linear estimate of x_t, when the value of x_{t-1} is given

$$\hat{x_t} = \frac{-\theta}{1+\theta^2}x_{t-1} + \frac{1-\theta}{2}\left(1 + \frac{\theta}{1+\theta^2}\right). \tag{2.4.2}$$

This will help to clarify the obtained results because this function has such a specific form. In the **Figure 2.2** the yellow line represents MMSEP given in *Equation 2.4.1*, but the red line corresponds to the best linear predictor (BLP), where we predict the value of x_2 by *Equation 2.4.2*, when the value of x_1 is given. In the limiting case, when $\theta = 1$, then we can see that the overall best linear prediction is the same as the minimum mean square error predictor, but for larger values of θ the piecewise linearity of the MMSEP is noticeable. Also the region, where the MMSEP takes constant value (actually, it is the mean value of the process) widens by increasing the value of coefficient θ, but the region, where the MMSEP is non-constant becomes relatively small, but the slope becomes even sharper.

At first let us do the simulations and compare the usual approach, when we fit invertible model and use the prediction equations defined by *Equation 2.2.1* and the obtained MMSEP in this special case as in *Equation 2.4.1*. Let the innovation series be as required in last example, where $w_t \sim U[0,1]$. We generate the process

$$x_t = w_t - \theta w_{t-1}$$

M times and each time for the obtained realization calculate the one step ahead

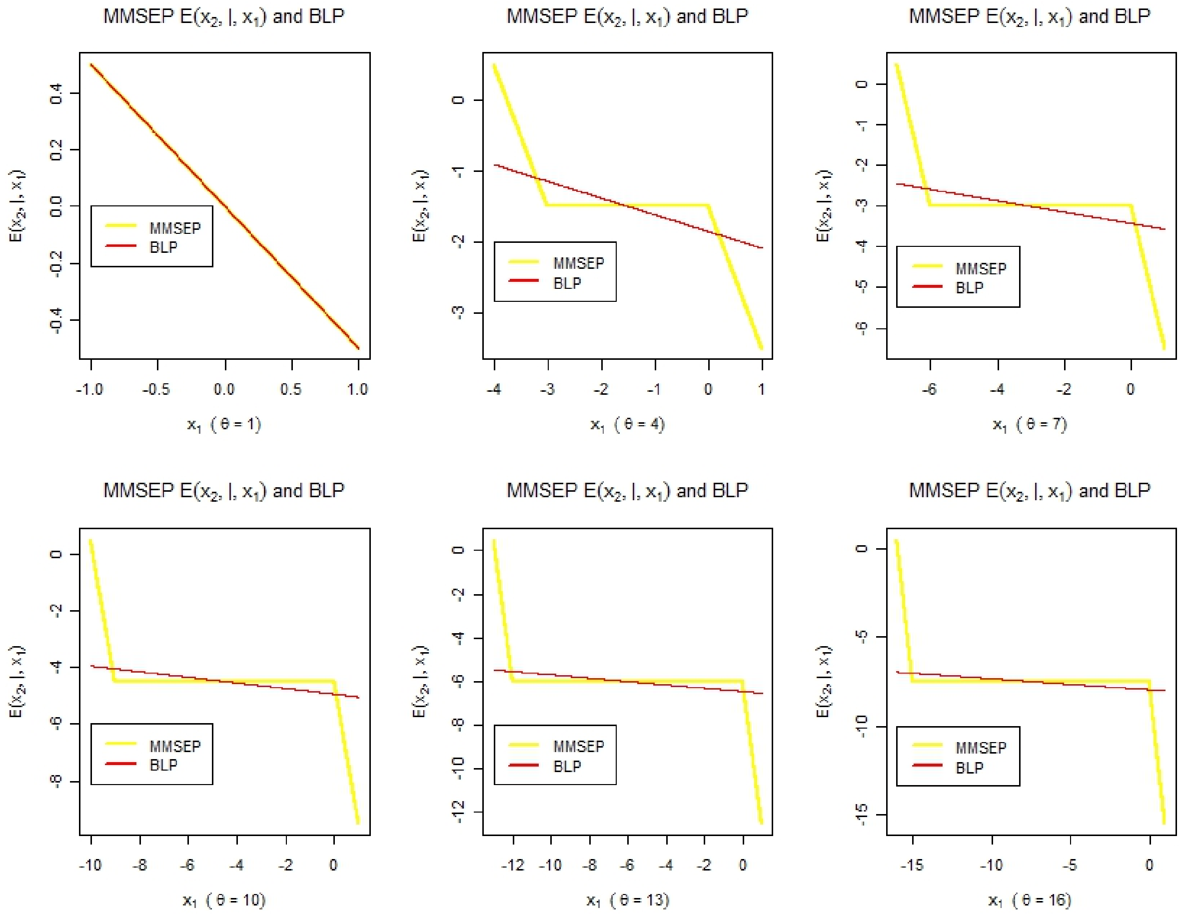

Figure 2.2 : The plots of MSSEP (conditional expected value function $E(X_2|X_1 = c)$ given in *Equation 2.4.1*) and BLP given in *Equation 2.4.2* for different value of θ.

predictions for the last 20% of data. Then we calculate the difference between the predicted value and the real value. Then we calculate the mean square error of predictions for each series. After all M realizations we count the number of times, when the mean square error was smaller for the estimates obtained by MMSEP formula compared with the mean square error in case of typical approach. This percentage can be found in *Table 2.4.2.* . Here we recall, that in case of typical approach we use all previous process values, when calculating the next process value, but when using obtained piecewise linear MMSEP defined by *Equation 2.4.1* we use just the previous process value. We can do all this also for Gaussian distributed data and make sure that derived *Equation 2.4.1* is valid only for the purpose it was derived. All these results are collected and presented in *Table 2.4.2.* . The mean value of mean square errors for each situation is also given.

From the results we can see that in the case of uniformly distributed innovations

	Gaussian			Uniform $U[0,1]$		
θ	%	$\widehat{MSE}_{stand}$	$\widehat{MSE}_{method}$	%	$\widehat{MSE}_{stand}$	$\widehat{MSE}_{method}$
2.0	0.1	4.1898	6.0653	87.1	0.34897	0.33379
5.0	0	24.972	127.2321	85.4	2.1187	1.9676
15	0	223.50	8051.035	79.5	19.078	18.286
25	0	618.97	60133	78.3	52.997	51.281

Table 2.4.1 Simulation results $M = 1000$, $N = 200$

w_t we get more precise results with the derived formula despite the fact that we used only one previous value to predict the next one, but in the usual approach we need all previous values. The second noticeable thing is that there is noticeable reduction of the efficiency, when value of coefficient θ increases. This can be explained by the form of the exact MMSEP, which is equal to the mean value for a wide range of previous value of the process and is slightly different only, when the previous value is very close to the boundary.

One could claim that MMSEP converges to one overall linear prediction expression, when we increase the number of previous values used in the derivation of MMSEP as it can be shown in case of invertible model. Let us prove the fact, that in case if the model is invertible (in this case $|\theta| < 1$), then also in case of non-Gaussian distributed innovations the MMSEP converges to one overall linear model. Recall, that we are interested in the model

$$x_t = w_t - \theta w_{t-1}, \tag{2.4.3}$$

where $w_t \sim U[0,1]$. If we denote $\tilde{w}_i = w_i - 0.5, i = 0, 1, \ldots, t$ and $\tilde{x}_i = x_i - 0.5(1 - \theta), i = 0, 1, \ldots, t$ substitute these variables in *Equation 2.4.3*, then we obtain $MA(1)$ process with mean zero

$$\tilde{x}_t = \tilde{w}_t - \theta \tilde{w}_{t-1}, \tag{2.4.4}$$

Assume, that we want to get MMSEP of x_t, when we have all previous process values.

Recall the form of the minimum mean square error predictor (MMSEP) stated in *Lemma 3*. Therefore we are interested in

$$\begin{aligned}
E(\tilde{x}_t|\tilde{x}_{t-1},\ldots,\tilde{x}_1) &= E(\tilde{w}_t + \theta\tilde{w}_{t-1}|\tilde{x}_{t-1},\tilde{x}_{t-2},\ldots,\tilde{x}_0) \\
&= E(\tilde{w}_t + \theta(\tilde{x}_{t-1} + \theta\tilde{w}_{t-2})|\tilde{x}_{t-1},\tilde{x}_{t-2},\ldots,\tilde{x}_0) \\
&\ldots \\
&= E(\tilde{w}_t - \theta\tilde{x}_{t-1} - \theta^2\tilde{x}_{t-2} - \ldots - \theta^{t-1}\tilde{x}_1 - \theta^t\tilde{w}_0)|\tilde{x}_{t-1}\ldots,\tilde{x}_0) \\
&= E(\tilde{w}_t) + E(-\theta\tilde{x}_{t-1} - \theta^2\tilde{x}_{t-2} - \ldots - \theta^{t-1}\tilde{x}_1 - \theta^t\tilde{w}_0)|\tilde{x}_{t-1}\ldots,\tilde{x}_0) \\
&= E(\tilde{w}_t) + E(-\theta\tilde{x}_{t-1} - \theta^2\tilde{x}_{t-2} - \ldots - \theta^{t-1}\tilde{x}_1 - \theta^t\tilde{w}_0)|\tilde{x}_{t-1}\ldots,\tilde{x}_0) \\
&= -\theta\tilde{x}_{t-1} - \theta^2\tilde{x}_{t-2} - \ldots - \theta^{t-1}\tilde{x}_1 - \theta^t\tilde{w}_0
\end{aligned}$$

And if we consider the limiting case in this invertible model ($|\theta| < 1$), when $t \to \infty$ then we get, that the last term $\theta^t\tilde{w}_0 \to 0$, $t \to \infty$. Thus, if we increase the number of predictors (use a larger history of the observed process values), then the MMSEP converges to the linear expression we just obtained. This is true for invertible model, but in the next paragraph we are going to look at the non-invertible case.

Let us investigate this non-invertible non-Gaussian case by computer simulations. Let us simulate M time series of length N with the innovation series be $w_t \sim U[0,1]$ and the data generating process

$$x_t = w_t - \theta w_{t-1}.$$

Then we fit linear model to the data: we take the set $\{x_1,\ldots x_{N-1}\}$ as predictors and try to fit the model for dependent variable x_N. This way we obtain one overall model. We can define also different sets of the same data based on the value of x_{N-1} and then fit the model for each subset.

If there exists one best overall linear model, then for each set we should get approximately the same model and it should be close to the overall model. The result of such simulations is given in **Figure 2.3**. The first three rows correspond to the model coefficients for each previous process value, but the last row represents the bar chart of the coefficient at x_{N-1} for each submodel (subset).

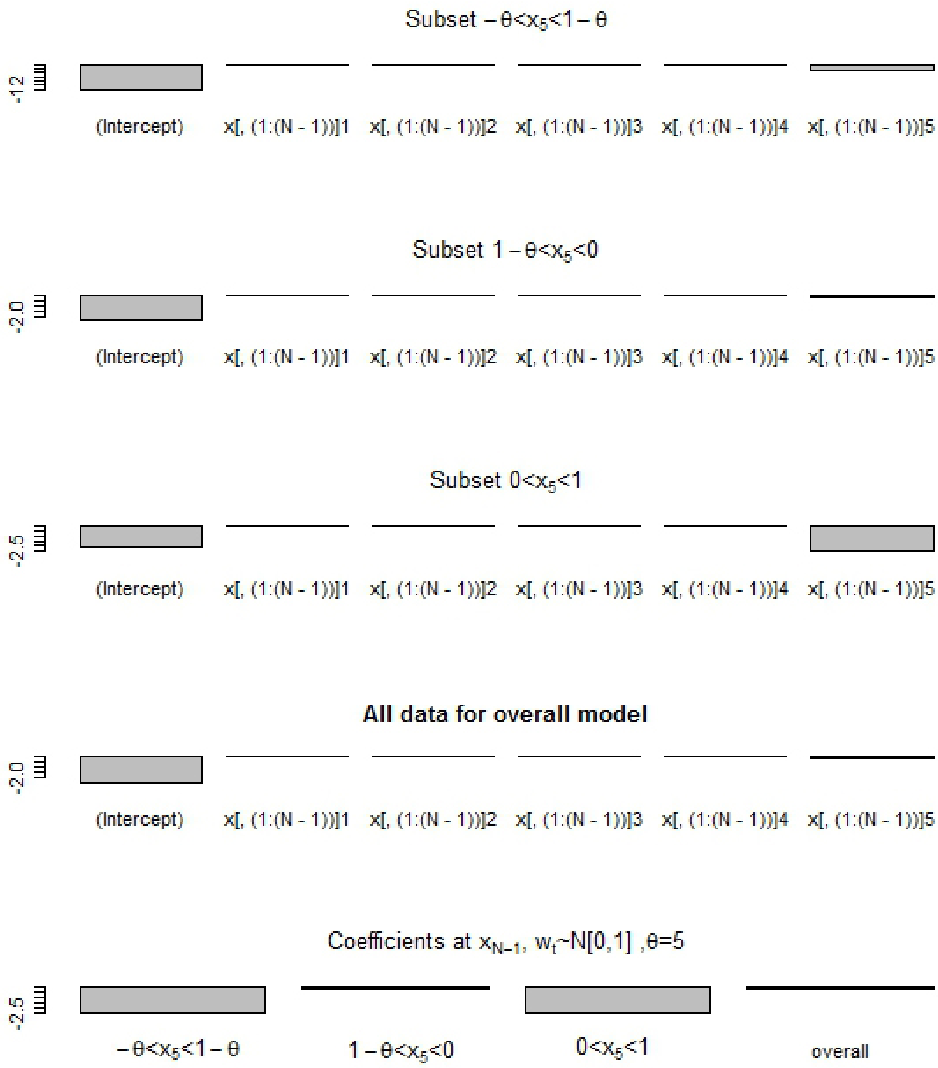

Figure 2.3 : The coefficients of fitted model for 3 subsets and the overall model, $M = 10000,\ N = 6$ and $\theta = 5$.

When we look at the coefficient at x_{N-1} for each submodel (subset) at the last row of in **Figure 2.3**, we can clearly see that the models are different for each of the subsets, therefore with this computational example we demonstrate, that it is quite unlikely, that there exists one overall best minimum mean square error predictor, which can be expressed as a linear function of previous process values.

As we noticed the efficiency of the method, when the coefficient θ increases can be analysed. This is done in the next step. As shown in the **Figure 2.4**, the decrease of efficiency of the MMSEPs, that are based on assumption about uniformly distributed data against the predictions based on Gaussian distributed random innovations is noticeable, but at the same time the decrease , although it is steady, remains quite

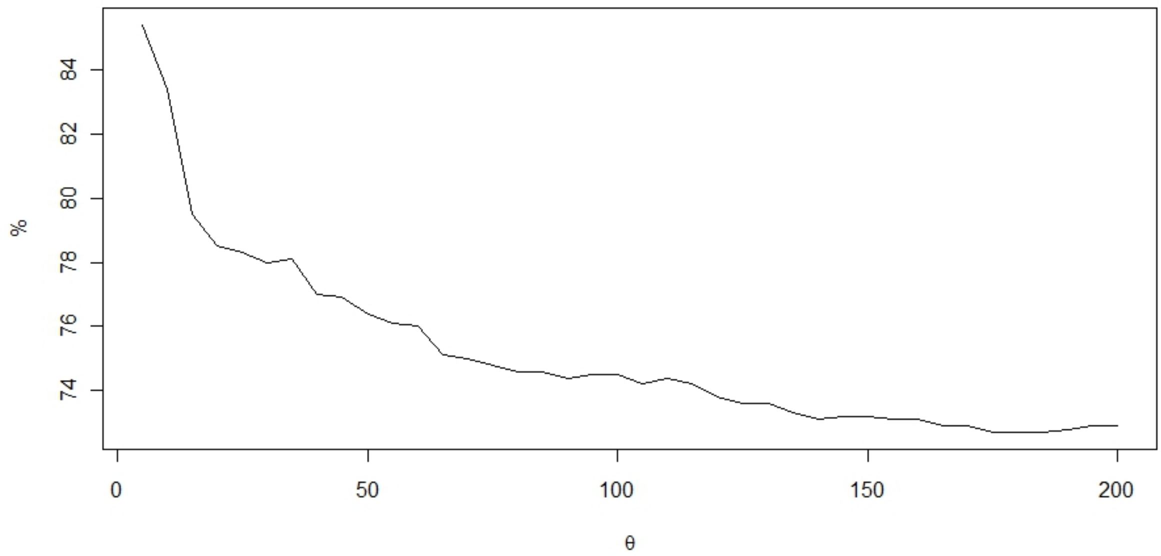

Figure 2.4 Efficiency of MMSEP (as a function of θ), $M = 1000$, $N = 200$

slow. Especially if we take into account that very large values of *theta* are quite unlikely, because then the coefficient θ in $MA(1)$ in the corresponding invertible model converges to zero. So, we can conclude, that for every reasonable choice of θ, the MMSEPs based on appropriate information about the distribution of random innovations gives reasonably better prediction estimates. Therefore it is suggested to use maximal information we can get out of the data.

3. Application

In real world situations data, which could be modelled by non-invertible $ARMA$ models are not very common. Although, we can find some interesting datasets. In publication [13] Breidt and Hsu analyse US unemployment rates from January 1948 to October 1997. They fit $SARIMA(0;1;5) \times (0;0;2)_{12}$ model to data and analyse non-invertibility in the seasonal part. They propose a different approach, how to deal with non-invertibility in this dataset. Results are promising, but as we have derived formula for non-Gaussian non-invertible $ARMA$ model ($MA(1)$), we will consider another example and look at the results.

Another dataset, where non-invertible model can be applied is described in the monograph by Brockwell and Davis [15, pp. 97-98.]. There they only fit the model and argue, that model could be non-invertible, but we are going to investigate this dataset more in details and try to do some forecasts.

3.1. The Overshorts data

We start by a small data description given in the monograph [15, p. 97].The dataset consists of 57 consecutive daily overshorts from an underground gasoline tank at a filling station in Colorado. If y_t is the measured amount of fuel in the tank at the end of the tth day and a_t is the measured amount sold minus the amount delivered during the course of the tth day, then the overshort at the end of day t is defined as $x_t = y_t - y_{t-1} + a_t$. Due to the error in measuring the current amount of fuel in the tank, the amount sold, and the amount delivered to the station, we view y_t, a_t, and x_t as observed values from some set of random variables Y_t, A_t , and X_t for $t = 1, \ldots, 57$. (In the absence of any measurement error and any leak in the tank, each x_t would be zero.)

So, we are interested in the *overshorts* dataset $x_t, t = 1, \ldots, 57$. In the beginning we can look at data (**Figure 3.1**). To give some more information, why we fit $MA(1)$ model to these data, we also should look at the graphs of autocorrelation function (ACF) (**Figure 3.2**) and also partial autocorrelation function ($PACF$) (**Figure 3.2**). As we can see, the graph of ACF shows, that only the first (obviously also zero correlation) correlation drops outside the confidence interval, but if we look at graph of $PACF$ (**Figure 3.3**), we can see, that it decreases almost linearly. Therefore choice of the moving average 1 ($MA(1)$) model is reasonable.

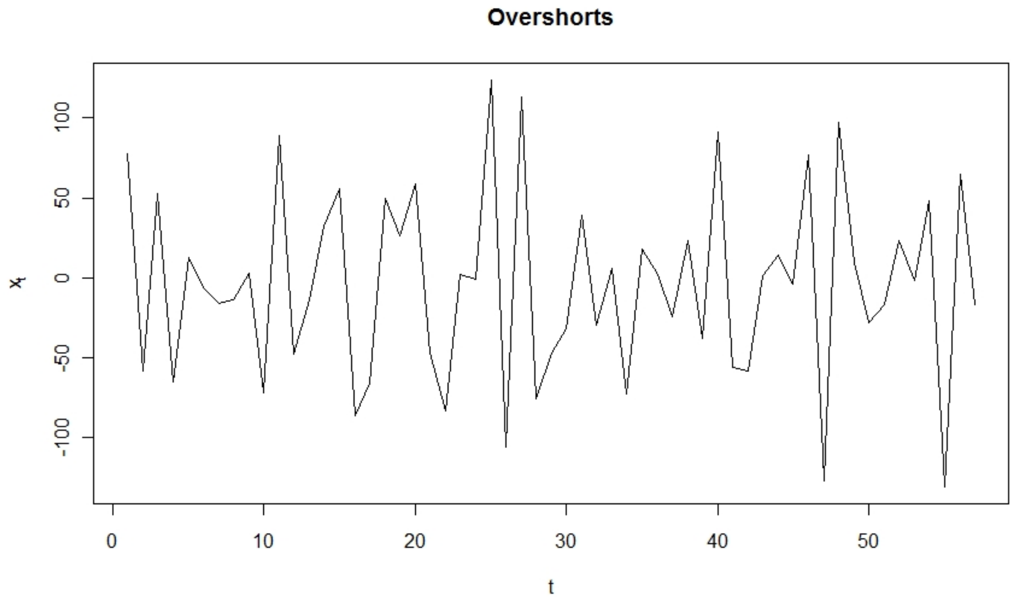

Figure 3.1 The overshorts data x_t

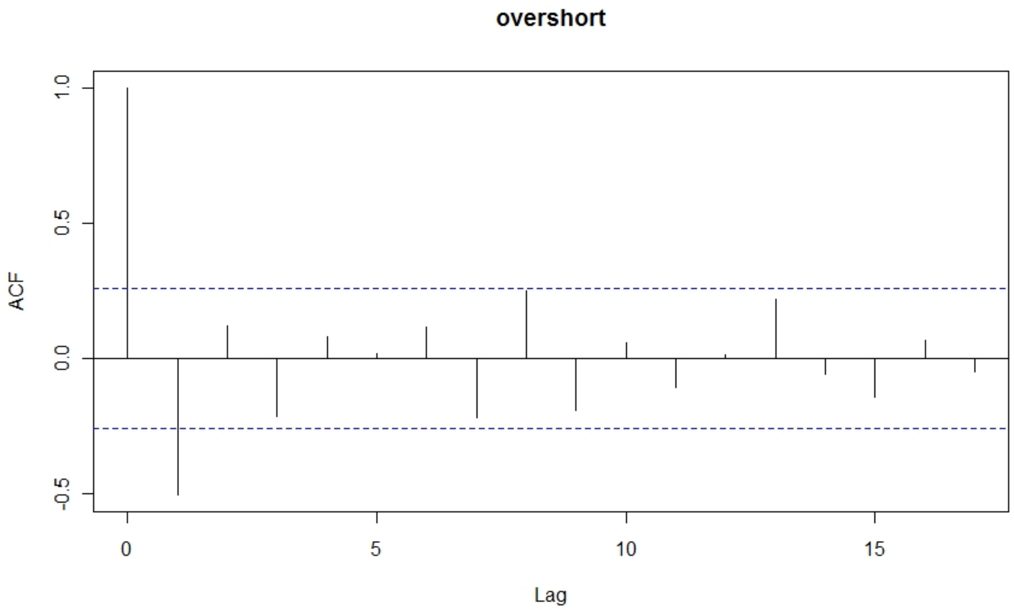

Figure 3.2 The ACF graph of overshorts data x_t

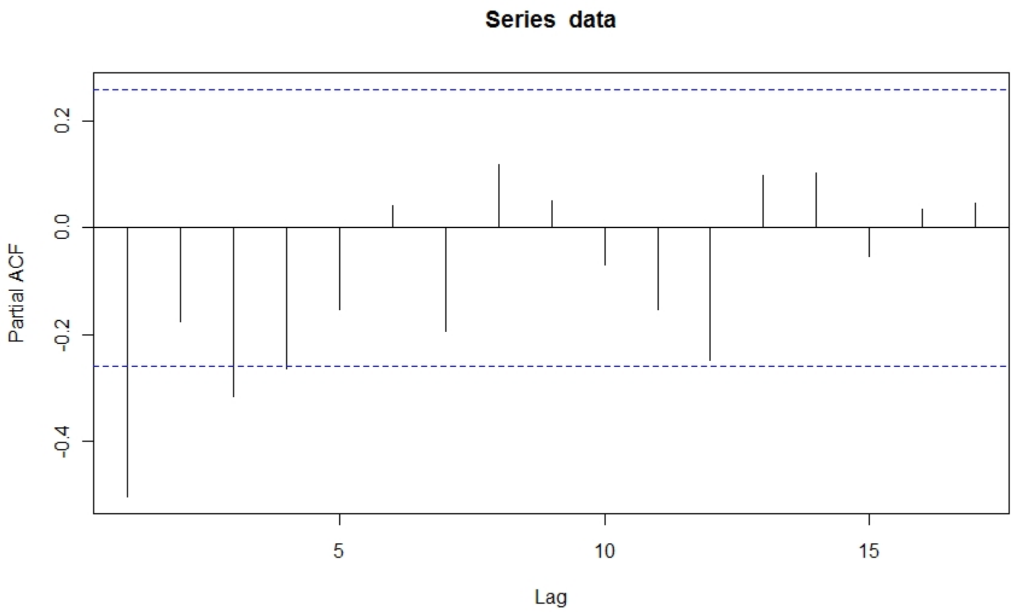

Figure 3.3 The $PACF$ graph of overshorts data x_t

3.2. Predictions by non-invertible $MA(1)$

When we have argued about the choice of the particular model, we can continue with the most interesting part, where we are going to do forecasts by prediction equations: both the classical approach, which assumes the Gaussian distribution, and the approach, where we assume uniformly distributed innovations, however, in this case we cannot argue, that the uniform distribution is more appropriate. For both, we fit the $MA(1)$ model $x_t = \mu_t + w_t - \theta w_{t-1}$, where the w_t terms denote innovations and μ_t denote the intercept, each time, when we want to do the forecast (we fit the model based on n previous values.) After fitting the model we also obtain the coefficient θ for our derived formula 2.4.1

$$E(X_2|X_1 = c) = \begin{cases} 0, & c \in (-\infty - \theta) \\ 0.5(1 - c\theta - \theta^2), & c \in [-\theta, 1 - \theta) \\ 0.5(1 - \theta), & c \in [1 - \theta, 0) \\ 0.5(-c\theta - \theta + 1,), & c \in [0, 1) \\ 0, & c \in (1, \infty). \end{cases} .$$

Here we want to point out, that the usual approach (defined by *Equation 2.2.1*) uses all available previous values, while our derived formula uses only the last process value. So, we are going to fit the model on a part of the data and look forward to predict the next process value, after that we compute the error and take the error squared. Then we perform the next step similarly, we fit the model on given data and try to obtain the next value. To compare these methods we are going to start this process at different time moments and look at the results. The results are collected in *Table 3.2.1*.

Starting point of the prediction, t	$\widehat{MSE}_{stand}$	$\widehat{MSE}_{uniform}$
45	2803.986	2501.685
46	3016.124	2709.409
47	2786.663	2445.599
48	2719.33	1906.934
49	2757.434	1994.122
50	2643.249	1815.562
51	2861.123	1999.356
52	3275.432	2172.416
53	3500.631	2564.849
54	3839.837	3183.499
55	2034.201	3508.332
56	670.1151	226.4287

Table 3.2.1 Simulation results

This time we can say that almost every time, except the case $t = 55$, new approach was more effective, than the usual fit. This is remarkable result, because, in the last case, we use just the previous value and even then the estimated mean square error is smaller than in usual approach, where we use all the information. There is a nice explanation for the case $t = 55$. First we recall that the total number of points is 57, so there are just three predictions, which determine MSE for both approaches. And, if we look at the time series **Figure 3.1**, we can notice, that time point 55 is slightly different, actually, at $t = 55$ process reaches its absolute minimum, which was not predicted in this case. So, in general, new approach is reasonable and could be suggested for predicting such time series.

Summary

In the beginning we declared, that in this thesis we are going to investigate predictions by non-invertible $ARMA$ models. We started with the definition of invertibility and declared non-invertibility to be the opposite of invertibility, briefly discussed transformation to invertibility. Then we moved further to the forecasting of $ARMA$ processes and especially to the predicting of non-invertible processes, which was provided also with detailed investigation of $MA(1)$ in each of the cases: the Gaussian case (invertible and non-invertible) and non-Gaussian (invertible and extended derivation of non-invertible one in case of uniformly distributed innovations).

The first part is more general than the second one, where the predictions by non-invertible $ARMA$ processes are described because it involves more specific analysis in each particular situation. However, the case of predictions by non-invertible $ARMA$ draws the greatest interest. Therefore we have chosen $MA(1)$ model as the most simplest way, how to deliver the results and give the reader a general intuition of the particular case. We deal with "strictly" non-invertible case, which gives some interesting result.Then we argue, why in case of Gaussian distributed data we can fit invertible model to the data generated by non-invertible Gaussian process and use the forecasts of the invertible model, which is remarkable result to point out due to the frequent occurrence of Gaussian distributed innovations. But one of the largest efforts was put to the non-invertible non-Gaussian case, where the Uniform distribution was chosen for modelling innovations. The derivation takes some effort and some technical parts can also be found in the appendix. When the resulting prediction equation is obtained, then some comparison is needed. Therefore we have included simulation studies, which show that, first, the prediction equation is piecewise linear and there is no one overall limiting linear model (fully continuous, non-piecewise linear function). Second, the exact formula (minimum mean square error predictor) gives better results even when using one previous process value compared with linear predictions using all previous process values but the third important thing is that it can be applied only in case, when we have strong arguments of applying such model. As a small drawback from the last section in *Chapter* 2, we can also mention, that the efficiency of this method slowly declines, when the value of the moving average coefficient θ increases. Finally, there is also a real-world application provided. We fit the $MA(1)$ model to the data and compare the predictions by non-invertible Gaussian innovations with the ones by non-invertible case with uniformly distributed innovations. It's quite surprising and interesting, that in this case the results obtained by specially derived formula of the case, when innovations are uniformly distributed, are far more better. It's especially amazing, because data describes daily overshorts from an underground gasoline tank at a filling station in Colorado, therefore the choice of uniformly distributed innovations can be considered as non-typical at the first look.

In conclusion I have to say that the thesis covers an interesting but a limited insight of predicting non-invertible $ARMA$ processes. These ideas (especially in case of non-invertible non-Gaussian processes) can be extended and more described in some further discussion. As well as the idea of applying non-Gaussian predictions in practice can be considered more frequently, when the possibility of non-invertible model is not rejected, because then there is reasonable possibility of better forecasts.

Mittepööratavate ARMA mudelite abil prognoosimine

Agris Vaselāns

Kokkuvõte

Aegridade modelleerimisel ja prognoosimisel kasutatakse laialdaselt ARMA tüüpi mudeleid, mida on põhjalikult uuritud paljudes teadusartiklites ning käsitletud enamikus aegridadele pühendatud õpikutes. Selliste mudelite korral eeldatakse, et vaadeldava aegrea hetkeväärtus avaldub lineaarselt minevikuväärtuste ning mingi juhusliku häirituse hetke- ja minevikuväärtuste kaudu. Enamasti tehakse nii teoreetilistes käsitluste kui ka praktilistes rakendustes eeldus, et vaadeldava rea hetke- ja minevikuväärtuste põhjal on võimalik leida juhusliku häirituse hetkeväärtus; sellise omaduse olemasolul nimetatakse vastavat mudelit pööratavaks. Selline lähenemine ei pruugi aga anda kuigi häid tulemusi, kui vaadeldava rea andmed on tegelikult tekitatud mittepööratavale mudelile vastava juhusliku protsessi poolt.

Käesoleva bakalaureusetöö eesmärgiks on uurida nii teoreetiliselt kui ka arvutisimulatsioonide teel mittepööratavate ARMA tüüpi protsesside tulevikuväärtuste prognoosimisega seotud küsimusi.

Töö on jaotatud kolmeks peatükiks. Esimese peatükis tuuakse aegridade käsitlemiseks vajalikud põhimõisted ja tulemused. Teises peatükis uuritakse põhjalikumalt mittepööratavate ARMA protsesside prognoosimisega seotud küsimusi. Kõigepealt vaadeldakse protsesse juhul, kui häiritused on normaaljaotusega. Tõestatakse (kirjandusele tuginedes), et peaaegu alati (välja arvatud nn range mittepööratavuse erijuht) saadakse sellistel eeldustel optimaalsed prognoosid aegreale vastava pööratava mudeli sobitamise ning selle põhjal ennustamise teel. Samuti vaadeldakse ühte rangelt mittepööratava protsessi juhtu, mille jaoks tuletakse optimaalse prognoosi valemid. Kõige olulisemad autori originaaltulemused on saadud mittepöörava MA(1) protsessi uurimisel. Sellisel juhul on tuletatud ühte minevikuvaatlust kasutava optimaalse prognoosi funktsioon ning näidatud simulatsioonide teel, et seda kasutades on võimalik vastava protsessi tulevikuväärtuseid ennustada paremini, kui andmetele sobitatud pööratava mudeli abil. Kolmandas peatükis on toodud näide ühest praktilisest andmestikust, kus eelmises peatükis tuletatud ennustusfunktsioon annab samuti paremaid tulemusi, kui standardse ARMA tüüpi mudeliga on võimalik saada.

Töö ei ole kindlasti ammendav ülevaade mittepööratavate protsesside prognoosimisest, kuid annab aimu selle valdkonna probleemidest ja võimalikest lähenemistest. Saadud tulemuste põhjal võib järeldada, et juhul, kui aegrea puhul ei ole õigustatud

normaaljaotustega häirituste eeldus, ei pruugi standardsete vahenditega sobitatud aegrea mudelid sugugi parimaid prognoose anda ning et siis tasuks tõsiselt kaaluda ka häirituste jaotusele vastavate mittepööratavate mudelite sobitamist ja nende põhjal ennustamist.

Bibliography

[1] Meitz M. Saikkonen P. Maximum likelihood estimation of a non-invertible *arma* model with autoregressive conditional expectation. *Journal of Multivariate analysis*, 114, 2012.

[2] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[3] W. A. Stein et al. *Sage Mathematics Software (Version 5.9)*. The Sage Development Team, 2013. `http://www.sagemath.org`.

[4] Stoffer D.S Shumway R.H. *Time Series Analysis and Its Applications with R Examples 2nd edition*. Springer Science+Business Media, LLC, New York, 2006.

[5] Davis A. Richard Brockwell J. Peter. *Time Series: Theory and Methods. Second Edition*. Springer Science +Business Media, LLC, New York, 2006.

[6] Lifshits M. *Lectures on Gaussian Processes*. SpringerBriefs in Mathematics, Heidelberg, 2012.

[7] De Lima Antonio C. Pedroso Sen K. Pranab, Singer M. Julio. *From Finite Sample to Asymptotic Methods in Statistics*. Cambridge University Press, Cambridge, 2009.

[8] Fry M.J Bingham N.H. *Regression: Linear Models in Statistics*. Springer-Verlag London Limited 2010, London, 2010.

[9] N.H. Bingham and R. Kiesel. *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer Finance. Springer, 2004.

[10] Hamilton J.D. *Time series analysis*. Princeton University Press, New Jersey, 1994.

[11] Schwert G.W Plosser I.C. Estimation of a non-invertible moving average process: The case of overdifferencing. *Journal of Econometrics*, 6:pp. 199–224., 1977.

[12] O'Connell R. F. Hu G. Y. Analytical inversion of symmetric tridiagonal matrices. *J. Phys*, 29, 1996.

[13] Hsu N-J. Breidt J.F. Best mean square prediction for moving averages. *Statistica Sinica*, 15:pp. 427–446., 2005.

[14] Sheather S.J. *A Modern Approach to Regression with R*. Springer Science +Business Media, LLC, New York, 2009.

[15] Davis A. Richard Brockwell J. Peter. *Introduction to Time Series and Forecasting, Second Edition*. Springer-Verlag New York, Inc., New York, 2002.

Appendix

1. Graphs - Section1

```
# An ARIMA simulation
par(mfrow=c(4,3))
xx<-rnorm(200,0,1)
ts.sim <- arima.sim(list(order = c(0,0,1), ma = c(2)), n=200, innov=xx)
ts.plot(ts.sim)
acf(ts.sim)
pacf(ts.sim)
sd(ts.sim)
xx2<-rnorm(200,0,4)
corresponding.invertible<-arima.sim(list(order = c(0,0,1), ma = c(0.5)), n = 200,innov=xx2)
ts.plot(corresponding.invertible)
acf(corresponding.invertible)
pacf(corresponding.invertible)
sd(corresponding.invertible)


#check
check.white.noise<-ts.sim[-200]+xx2[-1]
ts.plot(check.white.noise)
acf(check.white.noise)
pacf(check.white.noise)
plot(check.white.noise) #looks like a white noise - at least for this case the
#theorem seems to be working
```

2. Graphs - section 2.4.1

```
####

#Conditional expected value (f-n)

Ex2.x1<-function(x,theta){
return((x<(-theta))*0+(x>=(-theta))*(x<(1-theta))*
(0.5*(1-x*theta-theta^2))+
(x>=(1-theta))*(x<(0))*(0.5*(1-theta))+(x>=0)*(x<=(1))*
(0.5*(1-x*theta-theta))+
(x>1)*0)
}

#Best linear predictor - BLP
BLP<-function(x,theta){
return((-theta/(1+theta^2))*x+((1-theta)/2)*(1+(theta/(1+theta^2))))
}


par(mfrow=c(2,3))
for (i in 1:6){
theta<-1+(i-1)*3
x<-seq(-theta,1,by=0.001)
z<-Ex2.x1(x,theta)
plot(x,z,col="yellow",type="l",lwd="2.5",
main=expression(paste("MMSEP ",E(x[2],"|",x[1])," and BLP")),
ylab=expression(paste(E(x[2],"|",x[1]))),
xlab=bquote(x[1]~~( ~   theta ~ "=" ~ .(theta) )))

legend(-theta,-theta+i, # places a legend at the appropriate place
c(expression(paste("MMSEP")),expression(paste("BLP"))), # puts text in the legend
lty=c(1,1), # gives the legend appropriate symbols (lines)
lwd=c(2.5,2.5),col=c("yellow","red"))
```

```
lines(x,BLP(x,theta), lty=1, col="red",lwd="0.5")

}
```

3. Coefficient test of the Section 2.4.1

```
#Parameters:
N<-6 #number of x values
M<-10000 # number of simulations
theta<-5 #coef. of theta in MA model (x_t+w_t-theta*w_{t-1})
w<-matrix(0,ncol=N+1,nrow=M)
for (i in 1:M){w[i,]<-runif(N+1)} #inovations
#w
x<-w[,2:(N+1)]-theta*w[,1:N]


lm(x[,N]~x[,(1:(N-1))],subset=(x[,N-1]<=1-theta))
lm(x[,N]~x[,(1:(N-1))],subset=((x[,N-1]>1-theta)&&(x[,N-1]<=0)))
lm(x[,N]~x[,(1:(N-1))],subset=(x[,N-1]>=0))

par(mfrow=c(5,1))

barplot(coefficients(lm(x[,N]~x[,(1:(N-1))],subset=(x[,N-1]<=1-theta))),
main=expression(paste("Subset ",-theta,"<",x[5],"<",1-theta)))
barplot(coefficients(lm(x[,N]~x[,(1:(N-1))],subset=((x[,N-1]>1-theta)&&(x[,N-1]<=0)))),
main=expression(paste("Subset ",1-theta,"<",x[5],"<",0)))
barplot(coefficients(lm(x[,N]~x[,(1:(N-1))],subset=(x[,N-1]>=0))),
main=expression(paste("Subset ",0,"<",x[5],"<",1)))

#overall
barplot(coefficients(lm(x[,N]~x[,(1:(N-1))])),main="All data for overall model")

barplot(c(coefficients(lm(x[,N]~x[,(1:(N-1))],subset=(x[,N-1]<=1-theta)))[N],
coefficients(lm(x[,N]~x[,(1:(N-1))],subset=((x[,N-1]>1-theta)&&(x[,N-1]<=0))))[N],
coefficients(lm(x[,N]~x[,(1:(N-1))],subset=(x[,N-1]>=0)))[N],
coefficients(lm(x[,N]~x[,(1:(N-1))]))[N]),
main=expression(paste("Coefficients at ", x[N-1],", ",w[t],"~", N,"[0,1] ,",theta,"=5")),
names.arg=c(expression(1-theta<x[N-1]),expression(x[N-1]<0) , expression(x[N-1]<1), "overall"))
#
```

4. Efficiency code of the Section 2.4.1

```
###########################efficiency ######################
standard.andcond.together.efficiency3<-function(x,n.start,theta){
N<-length(x)
predictions<-rep(NA,N);predictions2<-rep(NA,N)
j<-n.start
repeat{
predictions[j]<-predict(arima(x[(1:(j-1))],order=c(0,0,1)), n.ahead = 1)$pred
predictions2[j]<-(0.5*(1-theta*x[j-1]-theta^2)*(x[j-1]<=1-theta)+
     0.5*(1-theta)*(x[j-1]>1-theta)*(x[j-1]<=0)+
     0.5*(1-theta*x[j-1]-theta)*(x[j-1]>0))
j<-j+1
if(j>N) break()
}
error.squared<-(predictions-x)^2
error.squared2<-(predictions2-x)^2
return(cbind(predictions,error.squared,predictions2,error.squared2))
}
```

```
standard.MSE<-function(x){
return(mean(standart.aproach.MA1.for.efficiency3(x,((8*N)/10+1))[((8*N)/10+1):N,2]))
}

cond.exp.MSE<-function(x){
return(mean(cond.expectation.aproach.MA1.for.efficiency2(x,(8*N)/10+1,theta)[((8*N)/10+1):N,2]))
}

#Parameters:
N<-200 #number of x values
j<-1 #initial theta
MM<-40 #max theta
rratio<-rep(NA,MM)
avg.MSE.stand<-rep(NA,MM)
avg.MSE.cexp<-rep(NA,MM)
M<-1000 #number of replications
w<-matrix(runif((N+1)*M),ncol=(N+1))

both.MSE<-function(x){
res<-standard.andcond.together.efficiency3(x,(8*N)/10+1,theta)
return(cbind(mean(res[((8*N)/10+1):N,2]),mean(res[((8*N)/10+1):N,4])))
}
#Get result for each 5th value of theta
repeat{
theta<-j*5 #coef. theta in MA model (x_t=w_t-theta*w_{t-1})
results<-rep(NA,M);MSE.stand<-rep(NA,M);MSE.cexp<-rep(NA,M)
x<-matrix(w[,2:(N+1)]-theta*w[,1:N],ncol=N)
MSE<-apply(x,1,both.MSE) #compute MSE stand. for each serie
###print(MSE[1,]);print(MSE[2,])
results<- MSE[1,]>MSE[2,]
avg.MSE.stand[j]<-mean(MSE[1,])
avg.MSE.cexp[j]<-mean(MSE[2,])
ratio<-(sum(results)/length(results))
print(j)
print(ratio)
rratio[j]<-ratio*100
##rratio<-vector of ratios in non-invertibe/uniform
j<-j+1
if(j>MM) break()
}
par(mfrow=c(1,1))
plot(5*(1:MM),rratio,type="l",xlab=expression(paste(theta)),ylab="%")
```

5. Graphs and code of the Chapter 3

```
#OSHORTS data
#data<- read.table("C:/Users/user/skola_TARTU/non_invertible/oshorts.dat")
data<-read.table("C:/Users/user/skola_TARTU/non_invertible/oshorts.txt",header=TRUE)
t<-1:length(data[,1])
plot(t,data[,1], type="l", main="Overshorts",ylab=expression(x[t]))
acf(data)
pacf(data)


standard.andcond.together.efficiency3.thetaFromData<-function(x,n.start){
N<-length(x)
predictions<-rep(NA,N);predictions2<-rep(NA,N)
j<-n.start
repeat{
fit<-arima(x[(1:(j-1))],order=c(0,0,1))
#we assume non-inv. model for cond.exp. estimate
theta<- -(1/coefficients(arima(x[(1:(j-1))],order=c(0,0,1)))[1])
predictions[j]<-predict(fit, n.ahead = 1)$pred
```

```
predictions2[j]<-(0.5*(1-theta*x[j-1]-theta^2)*(x[j-1]<=1-theta)+
     0.5*(1-theta)*(x[j-1]>1-theta)*(x[j-1]<=0)+
     0.5*(1-theta*x[j-1]-theta)*(x[j-1]>0))
j<-j+1
if(j>N) break()
}
error.squared<-(predictions-x)^2
error.squared2<-(predictions2-x)^2
return(cbind(predictions,error.squared,predictions2,error.squared2))
}
for (j in 45:56){
res<-standard.andcond.together.efficiency3.thetaFromData(data[,1],j)
a<-mean(res[(j:57),2])
b<-mean(res[(j:57),4])
print(j);print(a);print(b)}
```

6. Sage computations 1

large triangular (basic)

Define variables:

```
w1,w2,w0,x1,x2,theta=var('w1','w2','w0','x1','x2','theta')
```

Define the integration f-n:

```
def myint(a,b,c,d,hold=false):
    if hold==true:
        return sage.calculus.calculus.dummy_integrate(a,b,c,d)
    else:
        return(integral(a,b,c,d))
```

Define diff f-n:

```
def mydiff(a,b,hold):
    if hold==true:
        return(sage.calculus.calculus.dummy_diff(a,b))
    else:
        return(diff(a,b))
```

```
hold=true;print 'F2=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,x1+theta,hold))
hold=false
F2(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,x1+theta,hold)
show(F2(x1,x2,theta))
```

F2=

$$\int_{-\frac{x_2}{\theta}}^{\theta+x_1}\int_{\frac{w_1-x_1}{\theta}}^{1}\int_{0}^{\theta w_1+x_2} 1\,dw_2\,dw_0\,dw_1$$

$$\frac{\theta^4+3\theta^3x_1+3\theta^2x_1^2+\theta x_1^3+3\left(\theta^2+2\theta x_1+x_1^2\right)x_2}{6\theta}+\frac{3\left(\theta^2+\theta x_1\right)x_2^2+x_2^3}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F2 /dx1 dx2';show(mydiff(mydiff(F2,x1,true),x2,true))
hold=false
print('f2=');f2=mydiff(mydiff(F2,x1,true),x2,true);show(f2(x1,x2,theta))
```

Diff: d^2F2 /dx1 dx2

$$(x_1,x_2,\theta)\mapsto\frac{\theta+x_1}{\theta}+\frac{x_2}{\theta^2}$$

f2=

$$\frac{\theta + x_1}{\theta} + \frac{x_2}{\theta^2}$$

```
hold=true;print 'F3=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,x1+theta,hold))
hold=false
F3(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,x1+theta,hold)
show(F3(x1,x2,theta))
```

F3=

$$\int_{-\frac{x_2}{\theta}}^{-\frac{x_2-1}{\theta}} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{\theta w_1+x_2} 1\,dw_2\,dw_0\,dw_1 + \int_{-\frac{x_2-1}{\theta}}^{\theta+x_1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{1} 1\,dw_2\,dw_0\,du$$

$$\frac{\theta^2 + 2\theta x_1 + x_1^2}{2\theta} + \frac{3\left(\theta^2 + \theta x_1\right)x_2^2 + x_2^3}{6\theta^3} + \frac{2\left(\theta^2 + \theta x_1 - 1\right)x_2 - 2\theta^2 - 2\theta x_1 + x_2^2 + 1}{2\theta^3} - \frac{3\left(\theta^2 + \theta x_1\right.}{}$$

```
hold=true
print 'Diff: d^2F3 /dx1 dx2';show(mydiff(mydiff(F3,x1,true),x2,true))
hold=false
print('f3=')
f3=mydiff(mydiff(F3,x1,true),x2,true);
show(f3(x1,x2,theta))
```

Diff: d^2F3 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto \frac{1}{\theta^2}$$

f3=

$$\frac{1}{\theta^2}$$

```
hold=true;print 'F4=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,x1+theta,hold))
hold=false
F4(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,x1+theta,hold)
show(F4(x1,x2,theta))
```

F4=

$$\int_{0}^{-\frac{x_2-1}{\theta}} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_{0}^{\theta w_1+x_2} 1\, dw_2\, dw_0\, dw_1 + \int_{-\frac{x_2-1}{\theta}}^{\theta+x_1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_{0}^{1} 1\, dw_2\, dw_0\, dw_1$$

$$\frac{\theta^2+2\theta x_1+x_1^2}{2\theta} + \frac{2\left(\theta^2+\theta x_1-1\right)x_2-2\theta^2-2\theta x_1+x_2^2+1}{2\theta^3} - \frac{3\left(\theta^2+\theta x_1\right)x_2^2-3\theta^2+x_2^3-3\theta x_1}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F4 /dx1 dx2';show(mydiff(mydiff(F4,x1,true),x2,true))
hold=false
print('f4=')
f4=mydiff(mydiff(F4,x1,true),x2,true);
show(f4(x1,x2,theta))
```

Diff: d^2F4 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto -\frac{x_2}{\theta^2} + \frac{1}{\theta^2}$$

f4=

$$-\frac{x_2}{\theta^2} + \frac{1}{\theta^2}$$

```
hold=true;print 'F5=';
show(myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,0,x1+theta,hold))
hold=false
F5(x1,x2,theta)=myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,0,x1+theta,hold)
show(F5(x1,x2,theta))
```

F5=

$$\int_{0}^{\theta+x_1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_{0}^{1} 1\, dw_2\, dw_0\, dw_1$$

$$\frac{\theta^2+2\theta x_1+x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F5 /dx1 dx2';show(mydiff(mydiff(F5,x1,true),x2,true))
hold=false
print('f5=')
f5=mydiff(mydiff(F5,x1,true),x2,true);
show(f5(x1,x2,theta))
```

Diff: d^2F5 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto 0$$

f5=

0

For cond. density we need fx1

```
print 'fx1:';fx1(x1,x2,theta)=(x1+theta)/theta;show(fx1)
```

fx1:

$$(x_1, x_2, \theta) \mapsto \frac{\theta + x_1}{\theta}$$

Conditional density function (piecewise)

```
f2_cond(x1,x2,theta)=f2/fx1;f2_cond.factor()
```

$$(x_1, x_2, \theta) \mapsto \frac{\theta^2 + \theta x_1 + x_2}{(\theta + x_1)\theta}$$

```
f3_cond(x1,x2,theta)=f3/fx1;f3_cond.factor()
```

$$(x_1, x_2, \theta) \mapsto \frac{1}{(\theta + x_1)\theta}$$

```
f4_cond(x1,x2,theta)=f4/fx1;f4_cond.factor()
```

$$(x_1, x_2, \theta) \mapsto -\frac{x_2 - 1}{(\theta + x_1)\theta}$$

Finally, conditional exp. of x2, x1 given:

```
hold=false
Exp(x1,theta)=myint(f2_cond(x1,x2,theta)*x2,x2,-theta*(x1+theta),1-
theta*(x1+theta))+myint(f3_cond(x1,x2,theta)*x2,x2,1-theta*
(x1+theta),0)+myint(f4_cond(x1,x2,theta)*x2,x2,0,1)
Exp(x1,theta).factor()
```

$$-\frac{1}{2}\theta^2 - \frac{1}{2}\theta x_1 + \frac{1}{2}$$

Test: overall integral of conditional density f-n should be "=1":

```
Exp1(x1,x2,theta)=myint(f2_cond(x1,x2,theta),x2,-theta*(x1+theta),1-
theta*(x1+theta))
Exp2(x1,x2,theta)=myint(f3_cond(x1,x2,theta),x2,1-theta*(x1+theta),0)
Exp3(x1,x2,theta)=myint(f4_cond(x1,x2,theta),x2,0,1)
(Exp1(x1,x2,theta)+Exp2(x1,x2,theta)+Exp3(x1,x2,theta)).factor()
```

1

7. Sage computations 2

middle part - rectangular

Middle part - rectangular prism

Define variables:

```
w1,w2,w0,x1,x2,theta=var('w1','w2','w0','x1','x2','theta')
```

Define the integration f-n:

```
def myint(a,b,c,d,hold=false):
    if hold==true:
        return sage.calculus.calculus.dummy_integrate(a,b,c,d)
    else:
        return(integral(a,b,c,d))
```

Define diff f-n:

```
def mydiff(a,b,hold):
    if hold==true:
        return(sage.calculus.calculus.dummy_diff(a,b))
    else:
        return(diff(a,b))
```

```
hold=true;print 'F2=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,1,hold))
hold=false
F2(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,1,hold)
show(F2(x1,x2,theta))
```

F2=

$$\int_{-\frac{x_2}{\theta}}^{1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_{0}^{\theta w_1+x_2} 1\, dw_2\, dw_0\, dw_1$$

$$\frac{3(2\theta+2x_1-1)x_2+3\theta^2+3\theta x_1-2\theta}{6\theta}+\frac{3\left(\theta^2+\theta x_1\right)x_2^2+x_2^3}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F2 /dx1 dx2';show(mydiff(mydiff(F2,x1,true),x2,true))
hold=false
print('f2=');f2=mydiff(mydiff(F2,x1,true),x2,true);show(f2(x1,x2,theta))
```

Diff: d^2F2 /dx1 dx2

$$(x_1,x_2,\theta) \mapsto \frac{1}{\theta}+\frac{x_2}{\theta^2}$$

f2=

$$\frac{1}{\theta}+\frac{x_2}{\theta^2}$$

```
hold=true;print 'F3=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,1,hold))
hold=false
F3(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,(-x2)/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,1,hold)
show(F3(x1,x2,theta))
```

F3=

$$\int_{-\frac{x_2}{\theta}}^{-\frac{x_2-1}{\theta}}\int_{\frac{w_1-x_1}{\theta}}^{1}\int_{0}^{\theta w_1+x_2} 1\,dw_2\,dw_0\,dw_1+\int_{-\frac{x_2-1}{\theta}}^{1}\int_{\frac{w_1-x_1}{\theta}}^{1}\int_{0}^{1} 1\,dw_2\,dw_0\,dw_1$$

$$\frac{2\theta+2x_1-1}{2\theta}+\frac{3\left(\theta^2+\theta x_1\right)x_2^2+x_2^3}{6\theta^3}+\frac{2\left(\theta^2+\theta x_1-1\right)x_2-2\theta^2-2\theta x_1+x_2^2+1}{2\theta^3}-\frac{3\left(\theta^2+\theta x_1\right)\mathit{x}}{}$$

```
hold=true
print 'Diff: d^2F3 /dx1 dx2';show(mydiff(mydiff(F3,x1,true),x2,true))
hold=false
print('f3=')
f3=mydiff(mydiff(F3,x1,true),x2,true);
show(f3(x1,x2,theta))
```

Diff: d^2F3 /dx1 dx2

$$(x_1,x_2,\theta)\mapsto\frac{1}{\theta^2}$$

f3=

$$\frac{1}{\theta^2}$$

```
hold=true;print 'F4=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,1,hold))
hold=false
F4(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-
x1)/theta,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,(1-x2)/theta,1,hold)
show(F4(x1,x2,theta))
```

F4=

$$\int_0^{-\frac{x_2-1}{\theta}} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{\theta w_1+x_2} 1\, dw_2\, dw_0\, dw_1 + \int_{-\frac{x_2-1}{\theta}}^{1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{1} 1\, dw_2\, dw_0\, dw_1$$

$$\frac{2\theta+2x_1-1}{2\theta} + \frac{2\left(\theta^2+\theta x_1-1\right)x_2 - 2\theta^2 - 2\theta x_1 + x_2^2 + 1}{2\theta^3} - \frac{3\left(\theta^2+\theta x_1\right)x_2^2 - 3\theta^2 + x_2^3 - 3\theta x_1 - \text{:}}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F4 /dx1 dx2';show(mydiff(mydiff(F4,x1,true),x2,true))
hold=false
print('f4=')
f4=mydiff(mydiff(F4,x1,true),x2,true);
show(f4(x1,x2,theta))
```

Diff: d^2F4 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto -\frac{x_2}{\theta^2} + \frac{1}{\theta^2}$$

f4=

$$-\frac{x_2}{\theta^2} + \frac{1}{\theta^2}$$

```
hold=true;print 'F5=';
show(myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,0,1,hold))
hold=false
F5(x1,x2,theta)=myint(myint(myint(1,w2,0,1,hold),w0,(w1-
x1)/theta,1,hold),w1,0,1,hold)
show(F5(x1,x2,theta))
```

F5=

$$\int_0^1 \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^1 1\, dw_2\, dw_0\, dw_1$$

$$\frac{2\theta+2x_1-1}{2\theta}$$

```
hold=true
print 'Diff: d^2F5 /dx1 dx2';show(mydiff(mydiff(F5,x1,true),x2,true))
hold=false
print('f5=')
f5=mydiff(mydiff(F5,x1,true),x2,true);
show(f5(x1,x2,theta))
```

Diff: d^2F5 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto 0$$

```
f5=
```

$$0$$

For cond. density we need fx1 (in region -theta+1..0)

```
print 'fx1:';fx1(x1,x2,theta)=1/theta;show(fx1)
```

```
fx1:
```

$$(x_1,x_2,\theta) \mapsto \frac{1}{\theta}$$

Conditional density function (piecewise)

```
f2_cond(x1,x2,theta)=f2/fx1;f2_cond.factor()
```

$(x_1,x_2,\theta) \mapsto \frac{\theta+x_2}{\theta}$

```
f3_cond(x1,x2,theta)=f3/fx1;f3_cond.factor()
```

$(x_1,x_2,\theta) \mapsto \frac{1}{\theta}$

```
f4_cond(x1,x2,theta)=f4/fx1;f4_cond.factor()
```

$(x_1,x_2,\theta) \mapsto -\frac{x_2-1}{\theta}$

Finally, conditional exp. of x2, x1 given:

```
hold=false
Exp(x1,theta)=myint(f2_cond(x1,x2,theta)*x2,x2,-theta,1-
theta)+myint(f3_cond(x1,x2,theta)*x2,x2,1-
theta,0)+myint(f4_cond(x1,x2,theta)*x2,x2,0,1)
Exp(x1,theta).factor()
```

$-\frac{1}{2}\,\theta + \frac{1}{2}$

Test: overall integral of conditional density f-n should be "=1":

```
Exp1(x1,x2,theta)=myint(f2_cond(x1,x2,theta),x2,-theta,1-theta)
Exp2(x1,x2,theta)=myint(f3_cond(x1,x2,theta),x2,1-theta,0)
Exp3(x1,x2,theta)=myint(f4_cond(x1,x2,theta),x2,0,1)
(Exp1(x1,x2,theta)+Exp2(x1,x2,theta)+Exp3(x1,x2,theta)).factor()
```

1

OK

8. Sage computations 3

small_upper_triangular_part_out

$1 - 1/\theta <= x_1 <= 1 - 1/\theta$

Upper part - rectangular prism (smalltriangular part left out)

```
w1,w2,w0,x1,x2,theta=var('w1','w2','w0','x1','x2','theta')
```

Define the integration f-n:

```
def myint(a,b,c,d,hold=false):
    if hold==true:
        return sage.calculus.calculus.dummy_integrate(a,b,c,d)
    else:
        return(integral(a,b,c,d))
```

Define diff f-n:

```
def mydiff(a,b,hold):
    if hold==true:
        return(sage.calculus.calculus.dummy_diff(a,b))
    else:
        return(diff(a,b))
```

$F_2(x_1,x_2): \quad 0 \leq x_1 \leq 1 - 1/\theta, \quad -\theta \leq x_2 \leq -\theta x_1$

```
hold=true;print 'F2=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,(-x2)/theta,1,hold))
hold=false
F2(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,(-
x2)/theta,1,hold)
show(F2(x1,x2,theta))
```

F2=

$$\int_{-\frac{x_2}{\theta}}^{1} \int_{\frac{w_1 - x_1}{\theta}}^{1} \int_{0}^{\theta w_1 + x_2} 1\, dw_2\, dw_0\, dw_1$$

$$\frac{3(2\theta + 2x_1 - 1)x_2 + 3\theta^2 + 3\theta x_1 - 2\theta}{6\theta} + \frac{3\left(\theta^2 + \theta x_1\right)x_2^2 + x_2^3}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F2 /dx1 dx2';show(mydiff(mydiff(F2,x1,true),x2,true))
hold=false
print('f2=');f2=mydiff(mydiff(F2,x1,true),x2,true);show(f2(x1,x2,theta))
```

Diff: d^2F2 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto \frac{1}{\theta} + \frac{x_2}{\theta^2}$$

f2=

$$\frac{1}{\theta} + \frac{x_2}{\theta^2}$$

$F_3(x_1,x_2): \quad 0 \leq x_1 \leq 1 - 1/\theta, \quad -\theta x_1 \leq x_2 \leq 1 - \theta$

```
hold=true;print 'F3=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,(-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F3(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,(-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F3(x1,x2,theta))
```

F3=

$$\int_{x_1}^{1} \int_{\frac{w_1 - x_1}{\theta}}^{1} \int_{0}^{\theta w_1 + x_2} 1\, dw_2\, dw_0\, dw_1 + \int_{-\frac{x_2}{\theta}}^{x_1} \int_{0}^{1} \int_{0}^{\theta w_1 + x_2} 1\, dw_2\, dw_0\, dw_1$$

$$\frac{1}{2}\theta x_1^2 + x_1x_2 + \frac{x_2^2}{2\theta} - \frac{3\theta^2x_1^2 + \theta x_1^3 + 3\left(2\theta x_1 + x_1^2\right)x_2}{6\theta} + \frac{3(2\theta + 2x_1 - 1)x_2 + 3\theta^2 + 3\theta x_1 - 2\theta}{6\theta}$$

```
hold=true
print 'Diff: d^2F3 /dx1 dx2';show(mydiff(mydiff(F3,x1,true),x2,true))
hold=false
print('f3=')
f3=mydiff(mydiff(F3,x1,true),x2,true);
show(f3(x1,x2,theta))
```

Diff: d^2F3 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto -\frac{\theta + x_1}{\theta} + \frac{1}{\theta} + 1$$

f3=

$$-\frac{\theta + x_1}{\theta} + \frac{1}{\theta} + 1$$

$F_4(x_1, x_2):\quad 0 \le x_1 \le 1 - 1/\theta, \quad 1 - \theta \le x_2 \le 1 - \theta x_1$

```
hold=true;print 'F4=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-
x2/theta,x1,hold)+myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,x1,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,(1-x2)/theta,1,hold))
hold=false
F4(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-
x2/theta,x1,hold)+myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,x1,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,(1-x2)/theta,1,hold)
show(F4(x1,x2,theta))
```

F4=

$$\int_{x_1}^{-\frac{x_2-1}{\theta}} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^{\theta w_1 + x_2} 1\,dw_2\,dw_0\,dw_1 + \int_{-\frac{x_2}{\theta}}^{x_1} \int_0^1 \int_0^{\theta w_1 + x_2} 1\,dw_2\,dw_0\,dw_1 + \int_{-\frac{x_2-1}{\theta}}^{1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^1 1\,dw_2\,dw_0\,dw_1$$

$$\frac{1}{2}\theta x_1^2 + x_1x_2 + \frac{x_2^2}{2\theta} + \frac{2\theta + 2x_1 - 1}{2\theta} - \frac{3\theta^2x_1^2 + \theta x_1^3 + 3\left(2\theta x_1 + x_1^2\right)x_2}{6\theta} + \frac{2\left(\theta^2 + \theta x_1 - 1\right)x_2 - 2\theta^2 - 2\theta x_1 + x_2^2 + 1}{2\theta^3} - \frac{3\left(\theta^2 + \theta x_1\right)x_2^2 - 3\theta^2 + x_2^3}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F4 /dx1 dx2';show(mydiff(mydiff(F4,x1,true),x2,true))
hold=false
print('f4=')
f4=mydiff(mydiff(F4,x1,true),x2,true);
show(f4(x1,x2,theta))
```

Diff: d^2F4 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto -\frac{\theta + x_1}{\theta} - \frac{x_2}{\theta^2} + \frac{1}{\theta^2} + 1$$

f4=

$$-\frac{\theta + x_1}{\theta} - \frac{x_2}{\theta^2} + \frac{1}{\theta^2} + 1$$

$F_5(x_1, x_2):\quad 0 \le x_1 \le 1 - 1/\theta, \quad 1 - \theta x_1 \le x_2 \le 0$

```
hold=true;print 'F5=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-x2/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F5(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-x2/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F5(x1,x2,theta))
```

F5=

$$\int_{-\frac{x_2}{\theta}}^{-\frac{x_2-1}{\theta}} \int_0^1 \int_0^{\theta w_1 + x_2} 1\,dw_2\,dw_0\,dw_1 + \int_{x_1}^{1} \int_{\frac{w_1-x_1}{\theta}}^{1} \int_0^1 1\,dw_2\,dw_0\,dw_1 + \int_{-\frac{x_2-1}{\theta}}^{x_1} \int_0^1 \int_0^1 1\,dw_2\,dw_0\,dw_1$$

$$\frac{x_2^2}{2\theta} + x_1 + \frac{x_2 - 1}{\theta} - \frac{x_2^2 - 1}{2\theta} + \frac{2\theta + 2x_1 - 1}{2\theta} - \frac{2\theta x_1 + x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F5 /dx1 dx2';show(mydiff(mydiff(F5,x1,true),x2,true))
hold=false
print('f5=')
f5=mydiff(mydiff(F5,x1,true),x2,true);
show(f5(x1,x2,theta))
```

Diff: d^2F5 /dx1 dx2

$$(x_1,x_2,\theta) \mapsto 0$$

f5=

$$0$$

$F_6(x_1,x_2): \quad 0 \le x_1 \le 1-1/\theta, \quad 0 \le x_2 \le 1$

```
hold=true;print 'F6=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F6(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F6(x1,x2,theta))
```

F6=

$$\int_0^{-\frac{x_2-1}{\theta}} \int_0^1 \int_0^{\theta w_1+x_2} 1\,dw_2\,dw_0\,dw_1 + \int_{x_1}^1 \int_{\frac{w_1-x_1}{\theta}}^1 \int_0^1 1\,dw_2\,dw_0\,dw_1 + \int_{-\frac{x_2-1}{\theta}}^{x_1} \int_0^1 \int_0^1 1\,dw_2\,dw_0\,dw_1$$

$$x_1 + \frac{x_2-1}{\theta} - \frac{x_2^2-1}{2\theta} + \frac{2\theta+2x_1-1}{2\theta} - \frac{2\theta x_1+x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F6 /dx1 dx2';show(mydiff(mydiff(F6,x1,true),x2,true))
hold=false
print('f6=')
f6=mydiff(mydiff(F6,x1,true),x2,true);
show(f6(x1,x2,theta))
```

Diff: d^2F6 /dx1 dx2

$$(x_1,x_2,\theta) \mapsto 0$$

f6=

$$0$$

$F_7(x_1,x_2): \quad 0 \le x_1 \le 1-1/\theta, \quad 1 \le x_2$

```
hold=true;print 'F7=';
show(myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,0,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,
(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F7(x1,x2,theta)=myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,0,x1,hold)+myint(myint(myint(1,w2,0,1,ho
(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F7(x1,x2,theta))
```

F7=

$$\int_{x_1}^1 \int_{\frac{w_1-x_1}{\theta}}^1 \int_0^1 1\,dw_2\,dw_0\,dw_1 + \int_0^{x_1} \int_0^1 \int_0^1 1\,dw_2\,dw_0\,dw_1$$

$$x_1 + \frac{2\theta+2x_1-1}{2\theta} - \frac{2\theta x_1+x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F7 /dx1 dx2';show(mydiff(mydiff(F7,x1,true),x2,true))
hold=false
print('f7=')
f7=mydiff(mydiff(F7,x1,true),x2,true);
show(f7(x1,x2,theta))
```

Diff: d^2F7 /dx1 dx2

$$(x_1,x_2,\theta) \mapsto 0$$

f7=

0

For cond. density we need f_{x_1} (in region $0 \leq x_1 < 1$)

```
print 'fx1:';fx1(x1,x2,theta)=(1-x1)/theta;show(fx1)
```

fx1:

$$(x_1, x_2, \theta) \mapsto -\frac{x_1 - 1}{\theta}$$

Conditional density function (piecewise)

```
f2_cond(x1,x2,theta)=f2/fx1;f2_cond.factor()
```

$$(x_1, x_2, \theta) \mapsto -\frac{\theta + x_2}{(x_1 - 1)\theta}$$

```
f3_cond(x1,x2,theta)=f3/fx1;f3_cond.factor()
```

1

```
f4_cond(x1,x2,theta)=f4/fx1;f4_cond.factor()
```

$$(x_1, x_2, \theta) \mapsto \frac{\theta x_1 + x_2 - 1}{(x_1 - 1)\theta}$$

Finally, $E(x_2|x_1)$:

```
hold=false
Exp(x1,theta)=myint(f2_cond(x1,x2,theta)*x2,x2,-theta,-theta*x1)+myint(f3_cond(x1,x2,theta)*x2,x2,-
theta*x1,1-theta)+myint(f4_cond(x1,x2,theta)*x2,x2,1-theta,1-theta*x1)
Exp(x1,theta).factor()
```

$$-\frac{1}{2}\,\theta x_1 - \frac{1}{2}\,\theta + \frac{1}{2}$$

Test: overall integral of conditional density f-n should be "=1":

```
Exp1(x1,x2,theta)=myint(f2_cond(x1,x2,theta),x2,-theta,-theta*x1)
Exp2(x1,x2,theta)=myint(f3_cond(x1,x2,theta),x2,-theta*x1,1-theta)
Exp3(x1,x2,theta)=myint(f4_cond(x1,x2,theta),x2,1-theta,1-theta*x1)
(Exp1(x1,x2,theta)+Exp2(x1,x2,theta)+Exp3(x1,x2,theta)).factor()
```

1

OK

9. Sage computations 4

large_upper_triangular_part_out

$0 <= x_1 <= 1 - 1/\theta$

Upper part - rectangular prism (large triangular part left out)

```
w1,w2,w0,x1,x2,theta=var('w1','w2','w0','x1','x2','theta')
```

Define the integration f-n:

```
def myint(a,b,c,d,hold=false):
    if hold==true:
        return sage.calculus.calculus.dummy_integrate(a,b,c,d)
    else:
        return(integral(a,b,c,d))
```

Define diff f-n:

```
def mydiff(a,b,hold):
    if hold==true:
        return(sage.calculus.calculus.dummy_diff(a,b))
    else:
        return(diff(a,b))
```

$F_2(x_1, x_2): \quad 0 \le x_1 \le 1 - 1/\theta, \quad -\theta \le x_2 \le 1 - \theta$

```
hold=true;print 'F2=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,(-x2)/theta,1,hold))
hold=false
F2(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,(-
x2)/theta,1,hold)
show(F2(x1,x2,theta))
```

F2=

$$\int_{-\frac{x_2}{\theta}}^{1} \int_{\frac{w_1 - x_1}{\theta}}^{1} \int_{0}^{\theta w_1 + x_2} 1 \, dw_2 \, dw_0 \, dw_1$$

$$\frac{3(2\theta + 2x_1 - 1)x_2 + 3\theta^2 + 3\theta x_1 - 2\theta}{6\theta} + \frac{3\left(\theta^2 + \theta x_1\right)x_2^2 + x_2^3}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F2 /dx1 dx2';show(mydiff(mydiff(F2,x1,true),x2,true))
hold=false
print('f2=');f2=mydiff(mydiff(F2,x1,true),x2,true);show(f2(x1,x2,theta))
```

Diff: d^2F2 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto \frac{1}{\theta} + \frac{x_2}{\theta^2}$$

f2=

$$\frac{1}{\theta} + \frac{x_2}{\theta^2}$$

$F_3(x_1, x_2): \quad 0 \le x_1 \le 1 - 1/\theta, \quad 1 - \theta \le x_2 \le -\theta x_1$

```
hold=true;print 'F3=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,(-x2)/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,(1-x2)/theta,1,hold))
hold=false
F3(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,(-x2)/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,(1-x2)/theta,1,hold)
show(F3(x1,x2,theta))
```

F3=

$$\int_{-\frac{x_2}{\theta}}^{-\frac{x_2-1}{\theta}} \int_{\frac{w_1 - x_1}{\theta}}^{1} \int_{0}^{\theta w_1 + x_2} 1 \, dw_2 \, dw_0 \, dw_1 + \int_{-\frac{x_2-1}{\theta}}^{1} \int_{\frac{w_1 - x_1}{\theta}}^{1} \int_{0}^{1} 1 \, dw_2 \, dw_0 \, dw_1$$

$$\frac{2\theta+2x_1-1}{2\theta}+\frac{3\left(\theta^2+\theta x_1\right)x_2^2+x_2^3}{6\theta^3}+\frac{2\left(\theta^2+\theta x_1-1\right)x_2-2\theta^2-2\theta x_1+x_2^2+1}{2\theta^3}-\frac{3\left(\theta^2+\theta x_1\right)x_2^2-3\theta^2+x_2^3-3\theta x_1-3x_2+2}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F3 /dx1 dx2';show(mydiff(mydiff(F3,x1,true),x2,true))
hold=false
print('f3=')
f3=mydiff(mydiff(F3,x1,true),x2,true);
show(f3(x1,x2,theta))
```

Diff: d^2F3 /dx1 dx2

$$(x_1,x_2,\theta)\mapsto\frac{1}{\theta^2}$$

f3=

$$\frac{1}{\theta^2}$$

$F_4(x_1,x_2):\quad 0\le x_1\le 1-1/\theta,\quad -\theta x_1\le x_2\le 1-\theta x_1$

```
hold=true;print 'F4=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-
x2/theta,x1,hold)+myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,x1,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,(1-x2)/theta,1,hold))
hold=false
F4(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-
x2/theta,x1,hold)+myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,(w1-x1)/theta,1,hold),w1,x1,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,(1-x2)/theta,1,hold)
show(F4(x1,x2,theta))
```

F4=

$$\int_{x_1}^{-\frac{x_2-1}{\theta}}\int_{\frac{w_1-x_1}{\theta}}^{1}\int_0^{\theta w_1+x_2}1\,dw_2\,dw_0\,dw_1+\int_{-\frac{x_2}{\theta}}^{x_1}\int_0^1\int_0^{\theta w_1+x_2}1\,dw_2\,dw_0\,dw_1+\int_{-\frac{x_2-1}{\theta}}^{1}\int_{\frac{w_1-x_1}{\theta}}^{1}\int_0^1 1\,dw_2\,dw_0\,dw_1$$

$$\frac{1}{2}\theta x_1^2+x_1x_2+\frac{x_2^2}{2\theta}+\frac{2\theta+2x_1-1}{2\theta}-\frac{3\theta^2x_1^2+\theta x_1^3+3\left(2\theta x_1+x_1^2\right)x_2}{6\theta}+\frac{2\left(\theta^2+\theta x_1-1\right)x_2-2\theta^2-2\theta x_1+x_2^2+1}{2\theta^3}-\frac{3\left(\theta^2+\theta x_1\right)x_2^2-3\theta^2+x_2^3}{6\theta^3}$$

```
hold=true
print 'Diff: d^2F4 /dx1 dx2';show(mydiff(mydiff(F4,x1,true),x2,true))
hold=false
print('f4=')
f4=mydiff(mydiff(F4,x1,true),x2,true);
show(f4(x1,x2,theta))
```

Diff: d^2F4 /dx1 dx2

$$(x_1,x_2,\theta)\mapsto-\frac{\theta+x_1}{\theta}-\frac{x_2}{\theta^2}+\frac{1}{\theta^2}+1$$

f4=

$$-\frac{\theta+x_1}{\theta}-\frac{x_2}{\theta^2}+\frac{1}{\theta^2}+1$$

$F_5(x_1,x_2):\quad 0\le x_1\le 1-1/\theta,\quad 1-\theta x_1\le x_2\le 0$

```
hold=true;print 'F5=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-x2/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F5(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,-x2/theta,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F5(x1,x2,theta))
```

F5=

$$\int_{-\frac{x_2}{\theta}}^{-\frac{x_2-1}{\theta}}\int_0^1\int_0^{\theta w_1+x_2}1\,dw_2\,dw_0\,dw_1+\int_{x_1}^{1}\int_{\frac{w_1-x_1}{\theta}}^{1}\int_0^1 1\,dw_2\,dw_0\,dw_1+\int_{-\frac{x_2-1}{\theta}}^{x_1}\int_0^1\int_0^1 1\,dw_2\,dw_0\,dw_1$$

$$\frac{x_2^2}{2\theta}+x_1+\frac{x_2-1}{\theta}-\frac{x_2^2-1}{2\theta}+\frac{2\theta+2x_1-1}{2\theta}-\frac{2\theta x_1+x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F5 /dx1 dx2';show(mydiff(mydiff(F5,x1,true),x2,true))
hold=false
print('f5=')
f5=mydiff(mydiff(F5,x1,true),x2,true);
show(f5(x1,x2,theta))
```

Diff: d^2F5 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto 0$$

f5=

$$0$$

$F_6(x_1, x_2)$: $\quad 0 \le x_1 \le 1 - 1/\theta, \quad 0 \le x_2 \le 1$

```
hold=true;print 'F6=';
show(myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F6(x1,x2,theta)=myint(myint(myint(1,w2,0,theta*w1+x2,hold),w0,0,1,hold),w1,0,(1-
x2)/theta,hold)+myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,(1-
x2)/theta,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F6(x1,x2,theta))
```

F6=

$$\int_0^{-\frac{x_2-1}{\theta}} \int_0^1 \int_0^{\theta w_1 + x_2} 1\, dw_2\, dw_0\, dw_1 + \int_{x_1}^1 \int_{\frac{w_1 - x_1}{\theta}}^1 \int_0^1 1\, dw_2\, dw_0\, dw_1 + \int_{-\frac{x_2-1}{\theta}}^{x_1} \int_0^1 \int_0^1 1\, dw_2\, dw_0\, dw_1$$

$$x_1 + \frac{x_2 - 1}{\theta} - \frac{x_2^2 - 1}{2\theta} + \frac{2\theta + 2x_1 - 1}{2\theta} - \frac{2\theta x_1 + x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F6 /dx1 dx2';show(mydiff(mydiff(F6,x1,true),x2,true))
hold=false
print('f6=')
f6=mydiff(mydiff(F6,x1,true),x2,true);
show(f6(x1,x2,theta))
```

Diff: d^2F6 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto 0$$

f6=

$$0$$

$F_7(x_1, x_2)$: $\quad 0 \le x_1 \le 1 - 1/\theta, \quad 1 \le x_2$

```
hold=true;print 'F7=';
show(myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,0,x1,hold)+myint(myint(myint(1,w2,0,1,hold),w0,
(w1-x1)/theta,1,hold),w1,x1,1,hold))
hold=false
F7(x1,x2,theta)=myint(myint(myint(1,w2,0,1,hold),w0,0,1,hold),w1,0,x1,hold)+myint(myint(myint(1,w2,0,1,ho
(w1-x1)/theta,1,hold),w1,x1,1,hold)
show(F7(x1,x2,theta))
```

F7=

$$\int_{x_1}^1 \int_{\frac{w_1 - x_1}{\theta}}^1 \int_0^1 1\, dw_2\, dw_0\, dw_1 + \int_0^{x_1} \int_0^1 \int_0^1 1\, dw_2\, dw_0\, dw_1$$

$$x_1 + \frac{2\theta + 2x_1 - 1}{2\theta} - \frac{2\theta x_1 + x_1^2}{2\theta}$$

```
hold=true
print 'Diff: d^2F7 /dx1 dx2';show(mydiff(mydiff(F7,x1,true),x2,true))
hold=false
print('f7=')
f7=mydiff(mydiff(F7,x1,true),x2,true);
show(f7(x1,x2,theta))
```

Diff: d^2F7 /dx1 dx2

$$(x_1, x_2, \theta) \mapsto 0$$

```
f7=
```

0

For cond. density we need f_{x_1} (in region $0 \le x_1 < 1$)

```
print 'fx1:';fx1(x1,x2,theta)=(1-x1)/theta;show(fx1)
```

```
fx1:
```

$$(x_1,x_2,\theta) \mapsto -\frac{x_1-1}{\theta}$$

Conditional density function (piecewise)

```
f2_cond(x1,x2,theta)=f2/fx1;f2_cond.factor()
```

$(x_1,x_2,\theta) \mapsto -\frac{\theta+x_2}{(x_1-1)\theta}$

```
f3_cond(x1,x2,theta)=f3/fx1;f3_cond.factor()
```

$(x_1,x_2,\theta) \mapsto -\frac{1}{(x_1-1)\theta}$

```
f4_cond(x1,x2,theta)=f4/fx1;f4_cond.factor()
```

$(x_1,x_2,\theta) \mapsto \frac{\theta x_1+x_2-1}{(x_1-1)\theta}$

Finally, $E(x_2|x_1)$:

```
hold=false
Exp(x1,theta)=myint(f2_cond(x1,x2,theta)*x2,x2,-theta,1-theta)+myint(f3_cond(x1,x2,theta)*x2,x2,1-
theta,-theta*x1)+myint(f4_cond(x1,x2,theta)*x2,x2,-theta*x1,1-theta*x1)
Exp(x1,theta).factor()
```

$-\frac{1}{2}\,\theta x_1 - \frac{1}{2}\,\theta + \frac{1}{2}$

Test: overall integral of conditional density f-n should be "=1":

```
Exp1(x1,x2,theta)=myint(f2_cond(x1,x2,theta),x2,-theta,1-theta)
Exp2(x1,x2,theta)=myint(f3_cond(x1,x2,theta),x2,1-theta,-theta*x1)
Exp3(x1,x2,theta)=myint(f4_cond(x1,x2,theta),x2,-theta*x1,1-theta*x1)
(Exp1(x1,x2,theta)+Exp2(x1,x2,theta)+Exp3(x1,x2,theta)).factor()
```

1

OK

Non-exclusive licence to reproduce thesis and make thesis public

I, Agris Vaselāns (date of birth: 05.08.1988),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

Predictions by Non-Invertible ARMA Models,

supervised by Raul Kangro

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **20.05.2013**