Modeling, estimation and simulation into two statistical models : quantile regression and blind deconvolution

Josephine Merhi Bleik

Par **Josephine MERHI BLEIK**

Modeling, estimation and simulation into two statistical models: quantile regression and blind deconvolution

Thèse présentée
pour l'obtention du grade
de Docteur de l'UTC

Soutenue le 6 septembre 2019
Spécialité : Mathématiques Appliquées : Laboratoire de Mathématiques Appliquées de Compiègne (Unité de recherche EA-2222)

D2506

Modeling, estimation and simulation into two statistical models: quantile regression and blind deconvolution

A thesis submitted in fulfillment of the requirements

for the degree of Doctor of Applied Mathematics

Spécialité : Mathématiques Appliquées

by Josephine MERHI BLEIK

at the

Université de technologie de Compiègne

Département de Génie Informatique

Laboratoire de Mathématiques Appliquées de Compiègne

(LMAC)

Soutenue le 6 septembre 2019

Supervisor: Pr. Ghislaine Gayraud, Université de technologie de Compiègne

Reporters: Pr. Alexandre BERRED, Université du Havre

Pr. Sophie DABO-NIANG, Université de Lille

Examiners: Dr. Vlad Stephan Barbu, Université de Rouen

Pr. Salim Bouzebda, Université de technologie de Compiègne

"All models are wrong, but . . . some of them are useful."

Georges Box

Abstract

This thesis is dedicated to the estimation of two statistical models: the simultaneous regression quantiles model and the blind deconvolution model. It therefore consists of two parts. In the first part, we are interested in estimating several quantiles simultaneously in a regression context via the Bayesian approach. Assuming that the error term has an asymmetric Laplace distribution and using the relation between two distinct quantiles of this distribution, we propose a simple fully Bayesian method that satisfies the noncrossing property of quantiles. For implementation, we use Metropolis-Hastings within Gibbs algorithm to sample unknown parameters from their full conditional distribution. The performance and the competitiveness of the underlying method with other alternatives are shown in simulated examples.

In the second part, we focus on recovering both the inverse filter and the noise level of a noisy blind deconvolution model in a parametric setting. After the characterization of both the true noise level and inverse filter, we provide a new estimation procedure that is simpler to implement compared with other existing methods. As well, we consider the estimation of the unknown discrete distribution of the input signal. We derive strong consistency and asymptotic normality for all our estimates. Including a comparison with another method, we perform a consistent simulation study that demonstrates empirically the computational performance of our estimation procedures.

Résumé

Cette thèse est consacrée à l'estimation de deux modèles statistiques : le modèle des quantiles de régression simultanés et le modèle de déconvolution aveugle. Elle se compose donc de deux parties. Dans la première partie, nous nous intéressons à l'estimation simultanée de plusieurs quantiles de régression par l'approche Bayésienne. En supposant que le terme d'erreur suit la distribution de Laplace asymétrique et en utilisant la relation entre deux quantiles distincts de cette distribution, nous proposons une méthode simple entièrement Bayésienne qui satisfait la propriété non croisée des quantiles. Pour la mise en œuvre, nous utilisons l'algorithme de Gibbs avec une étape de Metropolis-Hastings pour simuler les paramètres inconnus suivant leur distribution conditionnelle a posteriori. Nous montrons la performance et la compétitivité de la méthode sous-jacente par rapport à d'autres méthodes en fournissant des exemples de simulation.

Dans la deuxième partie, nous nous concentrons sur la restoration du filtre inverse et du niveau de bruit d'un modèle de déconvolution aveugle bruyant dans un environnement paramétrique. Après la caractérisation du niveau de bruit et du filtre inverse, nous proposons une nouvelle procédure

d'estimation plus simple à mettre en œuvre que les autres méthodes existantes. De plus, nous considérons l'estimation de la distribution discrète inconnue du signal d'entrée. Nous obtenons une forte cohérence et une normalité asymptotique pour toutes nos estimations. En incluant une comparaison avec une autre méthode, nous effectuons une étude de simulation cohérente qui démontre empiriquement la performance informatique de nos procédures d'estimation.

Acknowledgements

To my life-coach, my late brother Mohamad Ali Merhi Bleik: because I owe it all to you. Many Thanks!

A very special gratitude goes out to my supervisor Ghislaine Gayraud for its patience and support in overcoming numerous obstacles I have faced in my research and for all the fruitful discussions we had.
I am also grateful to my thesis reporters for reviewing my work. I would like to thank, in particular, Elisabeth Gautherat, with whom I'd had a joint work in the signal processing field that was new to me but very interesting. I would like to thank her for her kindness and her always smiling face and also for accepting being present the day of my defense as guest in the jury.
I would like to thank Salim Bouzebda, for his unfailing support and encouragement, and for the opportunity to participate in teaching his course. I would like to thank him for the nice recommendation he made of me and for participating in the examination of my research work and being member of the jury.
I would like to thank all the laboratory staff, professors and all the PhD students, especially Tarik Fahlaoui and Wang Gao with whom I spent the most of times in the lab, sometimes exchanging ideas and sometimes trying to give help when someone needs it. It was great sharing laboratory with all of you during last years.

I am grateful to my mother and my father, who have provided me through moral and emotional support in my life. I am also grateful to my brothers Mohamad Said and Abed Al Rahman, my sister Jamal, my childhood friend Léa Malas. I am also grateful to my friends in Compiegne Joanna Akrouche, Youssef Rahme and Sam Taoum who have supported me along the way and with whom I spent the happiest moments.

A very special mention I would like to make to my fellow man Hani Al Hajjar who was always there for me in happy and sad moments, under work pressure and work success, ... at any time. I was always wanting to let him know what I was doing in my research work and how I was doing it, even though he probably never understood what it was all about, that made me feel good anyway! When unmotivated, he was there to push me and straighten up my mood .... Briefly, it was fantastic to have you by my sides along the way.

And finally, I thank all those who attended and who travelled to attend my defense and also those who could not make it but supported me over the distance.

Thank you all for all your encouragement!

Contents

Abstract iii

Résumé v

Acknowledgements vii

General introduction 1

I Bayesian quantile regression 7

1 Literature overview on quantile regression 9

I Introduction 9

II Single quantile regression 10

II.1 Frequentist quantile regression 13

II.2 Bayesian quantile regression 17

II.3 Crossing quantile problem 21

III Simultaneous quantile regression 24

III.1 Frequentist simultaneous quantile regression 24

III.2 Bayesian simultaneous quantile regression 28

IV Our contributions 33

2 Bayesian estimation of joint quantile regression 35

I Introduction 35

II Simultaneous Bayesian estimation of quantiles 37
II.1 Model 37
II.2 Bayesian procedure 38
II.2.1 Likelihood 38
II.2.2 Priors 39
III Computations 42
III.1 Full conditional distributions 43
III.2 Algorithm 47
IV Empirical comparison in linear case 48
IV.1 Numerical example 49
IV.1.1 Choice of $\tilde{\Sigma}_0$ and $\check{\Sigma}_0$ 49
IV.1.2 Estimation results 50
V Simulation study 52
V.1 Design 1: Univariate linear quantile regression 54
V.2 Design 2: Multivariate Linear quantile regression . . . 58
V.3 Design 3: Non parametric quantile regression 60
VI Conclusion 63

II Blind deconvolution 65

3 Literature overview on blind deconvolution 67

I Introduction 67
II Convolution system 68
II.1 Signal types 69
II.2 Filter types 71
III The deconvolution 72
III.1 Inverse filter 74

III.2 Ill-posed inverse problem 75
III.3 Blind deconvolution 75
IV Blind deconvolution approaches 77
V Our contribution 80

4 A new estimation procedure in noisy blind deconvolution model 83
I Introduction 83
II Assumptions and characterization 84
II.1 Assumptions 84
II.2 Constrast function 86
III Estimation procedures 90
IV Asymptotic Results 93
V Simulation study 96
V.1 Parameter settings 97
V.2 Illustration of the empirical contrast function roots . . 97
V.3 Implementation 98
V.3.1 Starting points 99
V.4 The results 100
V.5 Computational comparison 107
VI Discussion and conclusion 109

General conclusion 111
VII Proofs of Chapter 4 115
VII.1 Proof of Proposition 1 115
VII.2 Proof of the asymptotic results 118

Appendix 114

List of Figures

1.1 Estimated quantiles at $\tau = 0.1$ and $\tau = 0.2$ along the period 1950-2018 fitted by single quantile regression for North Atlantic Tropical storm data. 22

1.2 B-spline estimates of quantiles regression at $\tau = 0.15, 0.25, 0.5, 0.6$ for motorcycle data. 23

2.1 Trace and density plot (top), autocorrelation plot (middle) and $\hat{R}$ evolution through iterations (bottom) of $\hat{\sigma}$ (left panel) and $\hat{p}$ (right panel). 55

2.2 Trace and density plot (top), autocorrelation plot (middle) and $\hat{R}$ evolution (bottom) of the intercept $\hat{\beta}_{0.2}^{(0)}$ (left panel) and the slope $\hat{\beta}_{0.3}^{(0)}$ of the $0.2-$th and $0.3-$th quantile respectively. . . . 56

2.3 Trace and density plot (top), autocorrelation plot (middle) and $\hat{R}$ evolution (bottom) of the intercept $\hat{\beta}_{0.4}^{(0)}$ (left panel) of the $0.4-$th quantile and the slope $\hat{\beta}_p$. 57

2.4 Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for SBQR method 62

2.5 Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for F&R method 62

2.6 Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for M&ST method 63

3.1 Convolution schema 69

3.2 The deconvolution schema 73

3.3 Blind deconvolution schema 76

4.1 Graph of G_n with $k(n) = 1$. Second order autoregressive, $\sigma_0 = 0.05$. 98

List of Tables

2.1 Estimation results and crossing evaluation of overparameterized and parsimonious models for $n = 500, 1000, 2000$ and for different prior covariance matrices (ξ =300, 100, 0.1) and $\Sigma_0 = Id$, where $\Sigma_0 \in \{\tilde{\Sigma}_0, \check{\Sigma}_0\}$). 51

2.2 Table of criteria: crossing loss and RMISE 59

2.3 Table of criteria: crossing loss and RMISE for nomal covariate case 60

2.4 Table of criteria: crossing loss and RMISE for uniform covariate case 60

2.5 RMISE at different quantile levels computed for M&ST, F&R and SBQR methods 61

4.1 Estimates of the mixture model for $\sigma_0 = 0.05$, $k(n) = 1$ and different n 101

4.2 Estimates of the mixture model for $\sigma_0 = 1$, $k(n) = 1$ and different n. 101

4.3 Estimates of the mixture model for $\sigma_0 = 0.05$, $k(n) = 2$ and different n. 102

4.4 Estimates of the mixture model for $\sigma_0 = 1$, $k(n) = 2$ and different n. 102

4.5 Estimates of autoregressive model for $\sigma_0 = 0.05$, $k(n) = 1$ and different n 103
4.6 Estimates of autoregressive model for $\sigma_0 = 1$, $k(n) = 1$ and different n. 104
4.7 Estimates of autoregressive model for $\sigma_0 = 0.05$, $k(n) = 2$ and different n. 104
4.8 Estimates of autoregressive model for $k(n) = 2$ for different n 105
4.9 Estimates of moving average model for $k(n) = 4$ and $\sigma_0 = 0.05$ 106
4.10 Estimates of moving average model for $k(n) = 4$ and $\sigma_0 = 1$. 107
4.11 Results of SNR and the parameter estimation for AR(2) with $\sigma_0 = 0.1, 1$ by both $\mathcal{M}_1$ and $\mathcal{M}_2$. 108

General introduction

By nature, human beings are used to building and operating mental models of the reality that surrounds them. Every time we walk or cross the street, predict that it will rain or make any decision, we use predictive models that we have built based on the availability of factors or our past experiences.

In science, it is particularly interesting to study physical phenomena whose characteristics or properties change over time, e.g., the movement of the planets, the growth of a planet, the growth of a child, the transmission of disease, seismic waves propagation, the customer flow in a bank, the evolution of a stock market share, etc. These phenomena are called dynamic systems. More formally, a dynamic system is a combination of physical elements that interact with each other according to certain principles or rules.
Now, there is a question that arises here that is "*why we study such a system*"? To answer this, let us take the example of growth curves used in health booklet. They show how the distribution of weight, or height, varies with age. More precisely, they represent certain percentiles (traditionally the 3rd, 25th, 75th and 97th) of these age-conditional distributions. They thus make it possible to verify that a child's growth is "normal" by placing him or her in the distribution corresponding to his or her age.
This little example shows how this study can make our lives easier by detecting problems that can happen to us and perhaps finding the solutions.

On the other hand, the dynamic systems can be also defined as an entity in which certain action or operation is applied on an input X and provides a certain output Y as a response. We gave this latter definition because almost all daily observed phenomena have important dynamic aspects that are described by a set of inputs and outputs, e.g.: the electrical engine where the input is the current and the output is the rotation at a certain speed, a computer where the input can be a typed text on the keyboard and the output is the text displayed on the screen, etc.
Oldly, the scientists studied and analyzed these dynamic systems in an intuitive way or by using a prototype. However, the relationship between inputs and outputs is often complicated or uncertain, e.g., controlling a tanker ship or a car on ice. In addition, the experimentation prototype has several disadvantages in the sense of cost, slowness, unfeasibility and risk.

The alternative to this prototype lies in the formal study of dynamic systems, which has been, by history, the basis for many advances in understanding reality and in the realization of technologies. Many of these formalisms have been proposed in the literature to represent the dynamic behavior of a problem; on the top of them are mathematical formalisms such as the simple equations and the differential equations. Indeed, when the underlying law of a dynamic system is defined by a mathematical formalism, we have a mathematical model. One should note that a model never represents reality in all its details but provides an approximation that is much better if we consider more precise laws and if we have good quality-of-data.

Establishing a model requires in advance to establish the objective of modeling, e.g., the representation of a biological phenomenon, the simulation of an electrical network or the prediction of the future behavior of a physical system, etc. Sometimes, it requires a deep knowledge of the domain, but other times it does not, e.g., when the model is established automatically by a machine or computer. Another requirement is that the model must not be over-complicated. This requirement is very important and represents a very delicate topic that is highly addressed by the modelers. It is justified by the principle of parsimony, also known as the Ockham razor, which is one of the fundamental principles of science. This principle means that models should not be made excessively complicated if this is not necessary. In other words, among two models with the same predictive power, the simplest one must be chosen.

What we just mentioned above does not represent the goal of the modeling. In fact, modeling aims first at understanding the reality so that the model can provide additional knowledge on the mechanisms of the studied phenomenon. It aims also at generating solutions that provide examples of the system behavior. It is then interesting to study and analyze the behavior of the solutions according to the parameters or the structure of the model (stability analysis, sensitivity and robustness). But, before providing solutions, the modeling designs rules or assumptions, also known as the control system, that are capable of bringing the system to the desired state or structure.

Once a model is developed and validated, it can be used to answer the following question:

what will happen if we apply this operation or that modification to the system?

However, for most real systems, the answer to the above question is very complex to be evaluated analytically. It is, therefore, necessary to proceed numerically via simulation to estimate the characteristics of the model. By simulation, we mean the imitation of the behavior of a real process or system over time. Through this, we can collect a whole series of data, as if we were using the real system. Synthetically, the generated data can be used to study the behavior and performance of the system, and the result of the simulation can be useful for several purposes, such as prediction, analysis and decision-making. It should be noted that in most cases, the modeler is not only interested in a solution but also in a family of solutions (e.g., according to a parameter or initial conditions). The reasons for the simulation success nowadays are the availability of increasing computational power, the availability of softwares and tools designed specifically for simulation and finally, the need to manipulate increasingly complex models.

In this thesis, we focus on the study of two mathematical models. The first is the quantile regression model, which is a typical model in popular use today that allows to study the behavior of the distribution of a response variable given explanatory ones; this model fits exactly the example of the growth curves in health booklet we gave previously. We will be specifically interested in the estimation of simultaneous regression quantiles in a Bayesian framework, the subject that is not very well-developed to date. Also, we will address the complexity ("the principle of parsimony") of this model.
The second model is a blind deconvolution model that describes a functioning of an input/output system. The study of this model aims at restoring the

system input from noisy observations (outputs).

In this way, this thesis will be made up of two parts. The first part will be devoted to the quantile regression model and will consist of two chapters. Chapter 1 presents the state of art, in the frequentist and Bayesian framework, of the regression of simple quantile as well as the of simultaneous quantiles, the subject that we address in Chapter 2 in a Bayesian framework. The second part deals with a blind deconvolution model in the frequentist setting and consists of Chapters 3 and 4. In Chapter 3 we will recall some important notions in signal processing that are used in the estimation methods of blind deconvolution model. We will also provide an overview on some methods of blind deconvolution. In Chapter 4 we propose a new estimation procedure[1] for noisy blind deconvolution model where we focus on characterizing the underlying system and using this characterizations as assumptions for the model.

Results of simulations and/or real data will be presented throughout this thesis. In particular, Chapter 2 in Part 1 will concern the empirical validation of our proposed simultaneous regression quantiles estimation against existing methods and the same is for Chapter 4 where the simulation results show the performance of our proposed method for noisy blind deconvolution.

[1] This procedure is proposed in joint work with Emmanuelle Gautherat.

Part I

Bayesian quantile regression

1 Literature overview on quantile regression

I Introduction

Since the seminal work of Koenker and Basset [33], quantile regression has received increasing attention from both theoretical and empirical point of view. As introduced in [33], it is a statistical procedure intended to estimate and conduct inference about conditional quantile of a response distribution given a set of covariates.

In regression, when we deal with highly skewed conditional distribution, traditional mean regression may not explore interesting aspects of the relationship between the response variable and the available covariates. Furthermore, in the presence of outliers, the evaluation of the response average becomes much more complicated. These situations are often encountered in economics, business, epidemiology and many other fields.

An alternative to the traditional mean regression is quantile regression which provides a more complete description of the conditional response distribution. Quantile regression was pioneered by Koenker and Basset [33] for single quantile fitting in the frequentist framework. Based on their theory, many

other approaches have emerged in both frequentist and Bayesian frameworks. In the first instance, most of the approaches were concerned with single quantile regression. And, when interested in multiple regression quantiles, these single quantile approaches are used repeatedly. However, this may lead to the embarrassing crossing quantiles phenomenon (see [11], [57]). Since then, multiple quantile regression approaches, in both frequentist and Bayesian context, have been developed taking into consideration the crossing problem primarily caused by the violation of the monotonicity property of quantiles.

This issue has also motivated us to address the crossing problem of quantiles by developing a new estimation method for simultaneous quantiles. This method will be presented in details in Chapter 2 of this manuscript. But, before that, it is worth to mention some existent approaches for both single and multiple quantile regression in both frequentist and Bayesian frameworks and this will be addressed in this chapter.

II Single quantile regression

As mentioned before, estimating regression quantiles has become more appealing for some data when the mean regression does not hold or the objective of interest is no longer the mean or, also, when the distribution of the error term is non-Gaussian (see [33]).

Now, we shall introduce quantile regression as extension of unconditional quantiles theory. First, let Y be a univariate random variable with unknown

distribution function F and $Y_1, \ldots, Y_n$ be n variables having the same distribution as Y. Then, the $\tau-$th quantile is defined as

$$q_\tau = \inf\{y;\, F(y) \geq \tau\} = F^-(\tau),$$

where F^- denotes the generalized inverse of F. If Y is a continuous random variable, F^- coincides with F^{-1}, the inverse function of F, and then

$$q_\tau = F^{-1}(\tau).$$

It is well-known that q_τ is the solution of the following minimization problem (see [33])

$$q_\tau = \underset{\mu \in \mathbb{R}}{\operatorname{argmin}}\, E[\rho_\tau(Y - \mu)],$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}_{u<0})$ is the loss function beeing differentiable for all $u \neq 0$.

Proof. To minimize the expected loss $E[\rho_\tau(Y - \mu)]$, we take the derivative with respect to μ and set it to 0:

$$\begin{aligned} (\tau - 1) \int_{-\infty}^{\mu} dF(y) + \tau \int_{\mu}^{+\infty} dF(y) &= 0 \\ \tau - F(\mu) &= 0. \end{aligned}$$

Then, the solution is the $\tau-$th quantile, $q_\tau = F^-(\tau)$.

□

Since F is unknown, we take the empirical distribution function defined as

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{Y_i \leq y}.$$

Then, the M-estimator $\hat{\mu}$, defined as

$$\hat{\mu} = \underset{\mu \in \mathbb{R}}{\operatorname{argmin}} \frac{1}{n} \sum_i \rho_\tau(Y_i - \mu), \tag{1.1}$$

is a natural estimator of μ and it corresponds to the $\tau-$th quantile given by:

$$\hat{q}_\tau = \inf\{y;\, F_n(y) \geq \tau\}.$$

If we take the ordered values $\{y_{(1)}, \ldots, y_{(n)}\}$, $\hat{q}_\tau$ coincides with the $\lfloor n\tau \rfloor-$th order statistic, i.e., $\hat{q}_\tau = y_{(\lfloor n\tau \rfloor)}$, where $\lfloor n\tau \rfloor$ denotes the greatest integer less than or equal to $n\tau$, which means that $\hat{q}_\tau$ satisfies the fundamental property of quantiles [12]. Indeed, if we denote by n_-, n^+ and n_0 respectively, the number of y_i that lie below, above and on $\hat{q}_\tau$. Then, $n_- \leq n\tau \leq n_- + n_0$ and $n^+ \leq n(1-\tau) \leq n^+ + n_0$. Moreover, if Y has continous distribution, then $\frac{n_-}{n} \underset{n\to\infty}{\to} \tau$ (see [64]).

Let us turn now to the quantile regression model, defined by:

$$Y = q_\tau(\mathbf{X}) + \epsilon, \tag{1.2}$$

where,

- Y is the response variable,
- $\mathbf{X}$ denotes an explanatory variable whose distribution is supported by $\chi^d \subset \mathbb{R}^d$, $d \geq 1$,
- $\epsilon \sim F$, where F is an unknonwn distribution such that $P(\epsilon < 0) = \tau$, which implies that that $q_\tau(\mathbf{X})$ corresponds to the conditional $\tau-$th quantile of $Y|\mathbf{X}$.

Suppose that $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ are n pairs of independent random variables following the same distribution as $(\mathbf{X}, Y)$. If we consider that q_τ belongs to some class of functions $\mathcal{Q}$, then mimicking (1.1), we estimate q_τ by solving the following minimization problem:

$$\underset{q_\tau \in \mathcal{Q}}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(\mathbf{X}_i)). \tag{1.3}$$

This context is the germ of quantile regression, as introduced from the frequentist point of view of Koenker and Bassett [33], and will be considered for the rest of this chapter.

Next, we review different approaches, in both frequentist and Bayesian settings in the literature, that estimate single regression quantiles.

II.1 Frequentist quantile regression

Indeed, Koenker and Bassett's theory [33] is fundamental for all further quantile regression approaches in both frequentist and Bayesian frameworks. However, different techniques have been developed in the literature to infer regression quantiles. In this section, we focus on the most popular estimation approaches of quantile regression from the frequentist point of view.

For linear quantile regression, the $\tau-$th conditional quantile is expressed as

$$q_\tau(\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}_\tau,$$

where $\mathbf{X}$ is a vector valued in χ^d and $\boldsymbol{\beta}_\tau \in \mathbb{R}^d$. The minimization problem (1.3) then becomes

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta}\in\mathbb{R}^d}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i^\top \boldsymbol{\beta}). \tag{1.4}$$

Koenker and Bassett proposed, in [33], a practical technique that enables to easily compute linear regression quantiles. This technique consists on solving Problem (1.4) using linear programming so that it is reformulated as

$$\min_{(\boldsymbol{\beta},\mathbf{u},\mathbf{v})\in\mathbb{R}^d\times\mathbb{R}_+^{2n}} \{\tau \mathbf{1}_n' \mathbf{u} + (1-\tau)\mathbf{v}\mathbf{1}_n' | \mathbf{X}_n\boldsymbol{\beta} + \mathbf{u} - \mathbf{v} = \mathbf{Y}_n\}, \tag{1.5}$$

where $\mathbf{X}_n$ is $(n \times d)$-matrix whose rows are $\mathbf{X}_1, \ldots, \mathbf{X}_n$, $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, $\mathbf{u}$ and $\mathbf{v}$ correspond to the positive and negative parts of the vector of residuals $\mathbf{Y}_n - \mathbf{X}_n\boldsymbol{\beta}$. The codes are publicly available in R (R Core Team, 2017) from the package *quantreg* (see [32]).

Alternatively, it is shown in Gutenbrunner and Jureckova [25], that linear quantiles can also be computed by solving the corresponding dual problem to (1.5)

$$\max_{\hat{\boldsymbol{a}}_\tau\in[0,1]^n} \mathbf{Y}_n^\top \hat{\boldsymbol{a}}_\tau | \mathbf{X}_n^\top \hat{\boldsymbol{a}}_\tau = (1-\tau)\mathbf{X}_n^\top \mathbf{1_n},$$

where $\hat{\boldsymbol{a}}_\tau$ is a vector of length n whose elements are defined as

$$\hat{a}_{i\tau} = \begin{cases} 1 & if \quad Y_i > \mathbf{X}_i^\top \boldsymbol{\beta}_\tau, \\ 0 & if \quad Y_i < \mathbf{X}_i^\top \boldsymbol{\beta}_\tau. \end{cases}$$

The solution $\hat{\boldsymbol{a}}_\tau$ of this dual problem is called regression rank-scores. Later on, the method in [78] considered a distribution-free approach to construct confidence bands of the conditional quantile of $Y|\mathbf{X}$, $q_\tau(\mathbf{X}) = \mathbf{X}^\top\beta_\tau = \tilde{\mathbf{X}}^\top\beta +$

$F^{-1}(\tau)$, $\tilde{\mathbf{X}}$ is the covariable vector without including the constant, by directly constructing confidence bands of $F^{-1}(\tau)$, where, here, F is the error distribution and is supposed to be continuous.

From theoretical point of view, the asymptotic properties for quantile coefficient estimators such as asymptotic normality and consistency are given in [25], [33], [35] and [78].

In nonparametric case, we recall that the minimization problem is the following

$$\min_{q_\tau \in \mathcal{Q}} \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(\mathbf{X}_i)); \tag{1.6}$$

where $\mathcal{Q}$ is a class of functions mapping from χ^d to $\mathbb{R}$. In this setting, we quote two approaches that are commonly used for quantile regression fitting: the splines method and the local polynomial method.

In spline methods, $\mathcal{Q}$ denotes the space of spline functions of order r with a sequence of K knots. Note that r and K are both related to the smoothness of the spline function. In general, cubic splines are used, i.e, $r = 3$ but it remains to choose k, which is critical and needs to be carefully addressed. Indeed, a large K reduces the empirical loss but leads also to overfitting, i.e., small bias and large variance. On the other hand, a small K leads to low variance but high bias. Thus, the optimal choice of K is the best trade-off between bias and variance. An alternative approach considered in [15] consists on adding a penalty term to the minimization problem (1.6), i.e.,

$$\min_{q_\tau \in \mathcal{Q}} \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(X_i)) + \lambda J(q_\tau) \tag{1.7}$$

where λ is an extra parameter that controls the smoothness of the resulting spline and $J(q_\tau)$ is the penalty term. Various penalty terms were considered in the literature: for univariate covariable, the method in [50] followed the traditional spline smoothing by using the quadratic penalty, $J(q_\tau) = \int |q''_\tau(x)|^2 dx$. Unfortunately, the use of quadratic penalty moves the problem away from linear to quadratic programming, adding further complexity to the optimisation process. Alternatively, another method in [34] used the penalty $J(q_\tau) = \int_0^1 |q''_\tau(x)| dx$ where $q_\tau \in \mathcal{Q}$ is a linear spline function, which enables to express (1.7) as a linear programming problem. An extension to multivariate covariable is found in [27] where the same penalty, $J(q_\tau) = \int |q''_\tau(\mathbf{x})| d\mathbf{x}$, as in [34], is used for constraint B-splines smoothing in linear and quadratic fits.

Like other nonparametric smoothing approaches, the smoothing parameter λ plays also an important role in determining the trade-off between data good fitting and the penalty. For λ too large, there will be much more penalty on the estimator and thus there is oversmoothness. On the contrary, for small λ, the penalty has little impact and therefore we risk undersmoothness. A common way to calibrate λ is by cross-validation. The method in [50] proposed robust cross-validation addressing the tuning of smoothness parameter of quantile splines. To reduce the computation cost, a generalization of the approximate cross-validation is proposed in [77] proposed.

When assuming that $q_\tau \in \mathcal{C}^1$, local polynomial methods are commonly used. Considering the univariate case, the idea is to locally approximate the quantile function, $q_\tau(X)$ given as

$$\begin{aligned} q_\tau(x) &\approx q_\tau(u) + (x-u)q'_\tau(u) \\ &\approx \beta_0 + \beta_1(x-u), \end{aligned}$$

for u is the neighborhood of x. Therefore, the minimization problem (1.6) becomes

$$\min_{\beta\in\mathbb{R}^2} \sum_i \omega_i(x)\rho_\tau(Y_i - \beta_0 - \beta_1(X_i - x)),$$

with $\omega_i(x) = K_h(\frac{x_i - x}{h})$ is a weight function, K_h is a kernel function and h is the bandwidth parameter that controls the smoothness of the estimated quantile function. As in spline methods, one of the main problems of this nonparametric technique is the choice of h. The method in [72] proposed an optimal choice of h as the minimizer of the empirical mean squared error.

II.2 Bayesian quantile regression

Bayesian methods have been successfully emerged to study quantile regression. This has been of interest since the posterior distributions correspond to some updated knowledge on the parameters and can be summarized through Bayesian estimators or provide credible intervals. However, the most challenging problem, in Bayesian quantile regression, is that the likelihood is not usually available unless assuming a given distribution for the error terms. But, whatever the error distribution is, it is still possible to consider misspecified model via a "likelihood tool" expressed in terms of the quantity of interest. This is exactly the case of the Asymmetric Laplace Distribution (ALD) whose probability density function is given by:

$$f(y;\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\frac{\rho_\tau(y-\mu)}{\sigma}\right\},$$

where μ is the location parameter that corresponds to the $\tau-$th quantile of the distribution, σ is the scale parameter and τ is the skewness parameter. We denote this Asymmetric Laplace Distribution by $\mathcal{ALD}(\mu, \sigma, \tau)$.

It is well-known that MCMC methods are commonly used to conduct Bayesian inference. Through these methods, samples from the posterior distribution are generated and then summarized through Bayesian estimators such as the posterior maximum (MAP) or the posterior mean. The most commonly used algorithms in MCMC are the Gibbs sampler based on conditional sampling and the Metropolis-Hastings algorithm based on acceptance-rejection procedure. In the following, we will quote methods that use these algorithms for Bayesian quantile regression.

In [74], Yu and Moyeed showed that, using the ALD as a "likelihood tool", the maximum likelihood estimator for μ (i.e., the $\tau-$th quantile), corresponds to the minimizer of (1.1). For linear quantile regression ($q_\tau(\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}_\tau$), they misspecified Model (2.1) under the ALD and showed that even for improper priors on the coefficient $\boldsymbol{\beta}_\tau$, the Bayesian approach yields a proper posterior distribution. At that instant, the authors were unable to sample directly from the posterior distributions of the quantile coefficients due to the complexity of the likelihood function. They used, therefore, Metropolis-Hastings algorithm for posterior samples. Even though the ALD is not the true underlying error distribution, the Bayesian approach has successfully worked for fitting quantile regression since it provides satisfactory results. For this reason, the ALD has emerged successfully as a natural way for Bayesian quantile regression.

Since the resulting posterior density of the quantile estimates is not analytically tractable under the ALD, Kozumi and Kobayashi [39] proposed to use, for linear quantile regression, the location-scale mixture representation for the ALD (see also [38]). They showed that this representation enables to compute tractable full conditional distributions of the unknown parameters, which allows to develop Gibbs sampling algorithm for posterior inference. Let us see how full conditional distributions are computed using the location-scale mixture representation of the ALD.

Let ω be an exponential latent variable with $1/\sigma$ as scale parameter, $\omega \sim \mathcal{E}(1/\sigma)$, and Z be a standard normal variable. The variables ω and Z are independent. Therefore, if $\epsilon \sim \mathcal{ALD}(0, \sigma, \tau)$, we can represent ϵ as a mixture of Gaussian variable:

$$\epsilon = \gamma_\tau \omega + \delta_\tau \sqrt{\sigma \omega} z,$$

where $\gamma_\tau = \frac{1-2\tau}{\tau(1-\tau)}$ and $\delta_\tau^2 = \frac{2}{\tau(1-\tau)}$. With this result, the distribution of the error, given ω, is Gaussian, i.e., $\epsilon|\omega \sim \mathcal{N}(\gamma_\tau \omega, \sigma \delta_\tau^2 \omega)$. Hence, in the linear case, we can rewrite Model (2.1) as

$$Y = \mathbf{X}^\top \boldsymbol{\beta}_\tau + \gamma_\tau \omega + \epsilon^*, \tag{1.8}$$

with $\epsilon^*|\omega \sim \mathcal{N}(0, \delta_\tau^2 \sigma \omega)$. With the classical conjugate Gaussian prior for $\boldsymbol{\beta}_\tau$, namely $\boldsymbol{\beta}_\tau \sim \mathcal{N}_{(d+1)}(\boldsymbol{\beta}_0, \Sigma_0)$, and the inverse gamma for σ, $\sigma \sim \mathcal{IG}(a, b), a > 0, b > 0$ where $\mathcal{IG}(a, b)$ and assuming that $\omega \sim \mathcal{E}(1/\sigma)$, the full conditional distributions of the parameters are then

$$\boldsymbol{\beta}_\tau | \mathbf{Y}_n, \mathbf{X}_n, \boldsymbol{\omega}_n, \sigma \sim \mathcal{N}_{(d+1)}(\hat{\boldsymbol{\beta}}_0, \hat{\Sigma}_0),$$

$$\omega^{-1}|\mathbf{Y}_n, \mathbf{X}_n, \boldsymbol{\beta}_\tau, \sigma \sim \mathcal{GIG}(\psi, \phi),$$

$$\sigma|\mathbf{Y}_n, \mathbf{X}_n, \boldsymbol{\omega}_n, \boldsymbol{\beta}_\tau, \sigma \sim \mathcal{IG}(\hat{a}, \hat{b}),$$

where $\boldsymbol{\omega}_n = (\omega_1 \ldots, \omega_n)$, $\mathbf{Y}_n = (Y_1, \ldots, Y_n)$, $\mathbf{X}_n$ is the $n \times (d+1)$ covariable matrix, $\hat{\Sigma}_0 = (\mathbf{X}_n' V \mathbf{X}_n + \Sigma_0^{-1})^{-1}$, $\hat{\boldsymbol{\beta}}_0 = \hat{\Sigma}_0(\mathbf{X}_n V(\mathbf{Y}_n - \gamma_\tau \boldsymbol{\omega}_n) + \Sigma_0^{-1}\boldsymbol{\beta}_0)$, $\mathcal{GIG}(\psi, \phi)$ denotes the Generalized Inverse Gaussian distribution with $\psi = \frac{\sqrt{\gamma_\tau^2 + 2\delta_\tau^2}}{\sum_{i=1}^n |y_i - \mathbf{x}_i \boldsymbol{\beta}_\tau|}$ and $\phi = \frac{\gamma_\tau^2 + 2\delta_\tau^2}{\delta_\tau^2 \sigma}$ as shape and scale parameters, $\hat{a} = a + \frac{n}{2}$ and $\hat{b} = b_0 + \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i \boldsymbol{\beta}_\tau - \gamma_\tau \omega_i)^2}{\delta_\tau^2 \omega_i}$.

Many further Bayesian quantile regression approaches used this representation for the sake of computing full conditional distributions like in [2], [4], [5], [7] and [75].

Indeed, the methods in [39] and [75] extended Yu and Moyeed's method [74] to Tobit model, when dealing with censored data; For continuous dependent variable case, we quote the methods in [2] that used adaptive lasso quantile regression for variable selection, in [4] that generalized quantile regression to deal with dichotomous response data and in [5] that generalized the adaptative lasso in [2] also for dichotomous data.

Note that we also have used the Gaussian representation of the ALD in our Bayesian inference (see Chapter 2).

In nonparametric Bayesian quantile regression, there are different approaches, some of which used the ALD tool and others estimated the likelihood. For example, Kottas and Gelfand [36] estimated the error distribution using Dirichlet Process Mixture. Specifically, they used a mixture of uniform distributions for the case of $\tau = 1/2$, whereas Kottas et al. [37] used Mixture of multivariate normal distributions for any $\tau \in [0, 1]$.

Other likelihood tools are studied in the literature such as the empirical likelihood in [40]. Yang and He [70] used this empirical likelihood for quantile regression and showed that the resultant posterior from any fixed prior is asymptotically normal.

We can also find in the literature Bayesian nonparametric models for the error distribution like the idea of polya tree used to develop quantile pyramid process [28].

II.3 Crossing quantile problem

All these aforementioned methods, from both frequentist and Bayesian point of views, are well adapted to infer a single quantile regression. However, there are many applications where tackling several quantiles at the same time is needed. Unfortunately, repeatedly applying approaches that deal with single quantile regression, in a separate fashion, may not be satisfactory. The problem here is that the separate estimation of a set of quantiles may break the monotonicity property of the conditional quantiles. In other terms, in such case, quantile curves may cross when, for a given covariable, the predicted value of the response variable at level τ_1 is smaller than its value at level τ_2 when τ_2 is lower than τ_1.

To illustrate the problem of crossing quantiles, let us consider quantile regression for a real application.

In North Atlantic basin, the analysis of the strongest tropical storms [1] includes fitting linear quantile regression models to the maximum wind speed (Wmax) against the year (Year) of storm occurrence over a range of τ values in $[0, 1]$. Figure 1.1 shows the plot of the North Atlantic storm data during the

[1] The data is available from http://weather.unisys.com/hurricanes

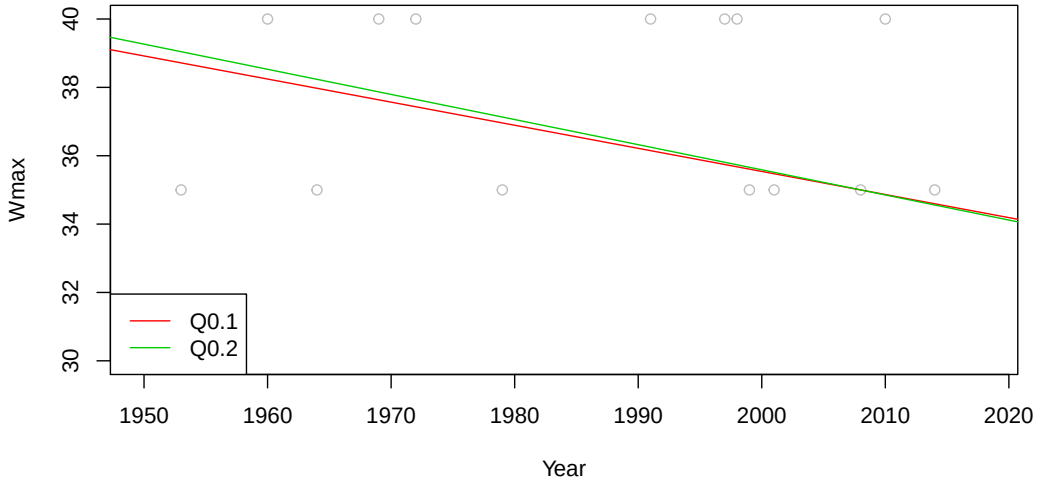

FIGURE 1.1: Estimated quantiles at $\tau = 0.1$ and $\tau = 0.2$ along the period 1950-2018 fitted by single quantile regression for North Atlantic Tropical storm data.

period 1950-2018. The estimated quantiles have been plotted at $\tau = 0.1$ and $\tau = 0.2$ using the single quantile technique of Koenker and Basset [33]. We use *quantreg* package in R software ([32]) for implementing the underlying method.

Figure 1.2 illustrates nonparametric estimates of conditional quantile functions at $\tau = 0.15, 0.25, 0.5, 0.6$ for motorcycle data[2] using spline method. The data consists of the measurements of the acceleration of a motorcycle rider's head as a function of time in the first moment after an impact. We fit cubic spline functions separately with 15 internal knots. We use *splines* and *quantreg* packages in R for implementation. The fitted conditional quantile functions that have close orders (0.15, 0.25) and (0.5, 0.6) cross in different regions: below 10 milliseconds, between 20 and 30 milliseconds and above 55 milliseconds.

[2]The motorcycle data is available in *MASS* package in R

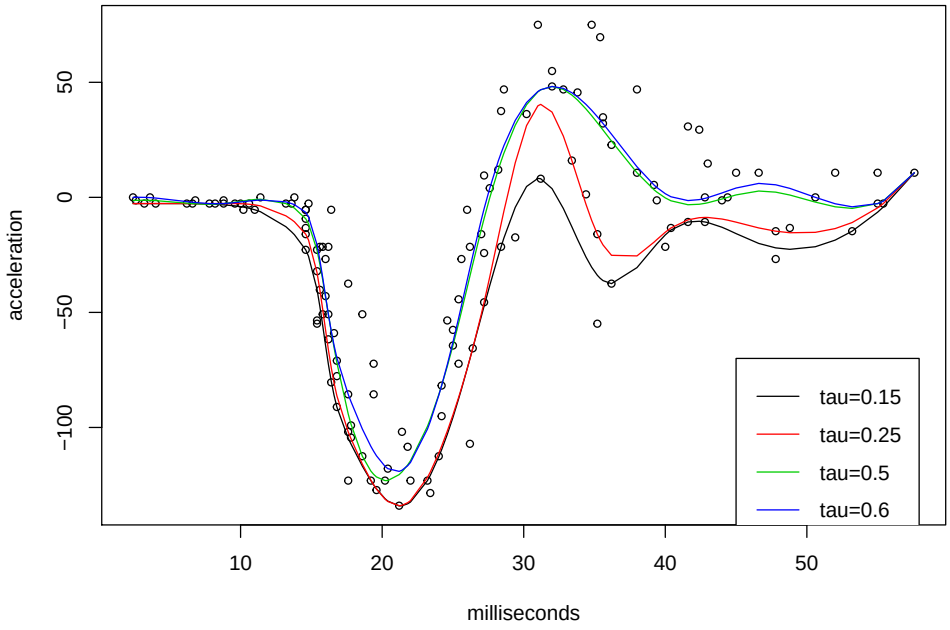

FIGURE 1.2: B-spline estimates of quantiles regression at $\tau = 0.15, 0.25, 0.5, 0.6$ for motorcycle data.

Therefore, it is to be noted that the estimated quantile curves cross each other, in both linear and nonlinear examples presented above, when quantile levels are very close and hence, the monotonicity property of quantiles regression is not fulfilled.

Neverthless, Yu and Jones [72], who proposed a local polynomial method to fit a single quantile regression, has also proposed a double kernel approach whose empirical results showed that, when repeatedly estimating several quantiles, the quantile estimators do not cross. However, this method is computationally expansive.

Other approaches, in both frequentist and Bayesian frameworks, have addressed the crossing problem seeking the simultaneous estimation of multiple quantiles. We will present some of these approaches is the next section.

III Simultaneous quantile regression

Since in real applications, it is often needed to estimate different quantiles of the response distribution simultaneously, the basic requirement for this concept is that quantile estimates respect the natural ordering of quantile levels. For this reason, noncrossing quantile regression approaches have received much attention from both frequentist and Bayesian point of views.

In this section, we present the most relevant frequentist and Bayesian approaches, in the literature, that preserve the noncrossing property of quantile estimates.

III.1 Frequentist simultaneous quantile regression

In the literature, different methods addressed the crossing issue from a frequentist point of view and most of them are derived from the classical optimization problem of quantile regression,

$$\min_{q_\tau \in \mathcal{Q}} \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(\mathbf{X}_i)), \tag{1.9}$$

under additional noncrossing constraints.

This traced back to He [26] who considered linear quantiles for a location-scale model

$$Y = \mathbf{X}^\top \boldsymbol{\beta} + (\mathbf{X}^\top \gamma)e,$$

where e is any error distribution whose median is zero and $\mathbf{X}$ is a multidimensional covariate. Based on relating all quantile functions to the conditional median, He computed the median of Y on $\mathbf{X}$ to obtain $\hat{\boldsymbol{\beta}}$ and thus the absolute values of the residuals $|r_i|$, with $r_i = y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}$, then he regressed

$|r_i|$ on $\mathbf{x}_i$ to obtain the median regression coefficient $\hat{\gamma}$ and the fitted values, $s_i = |\mathbf{x}^\mathsf{T}\hat{\gamma}e|$, of the residuals. The conditional quantile estimate is then $\hat{q}_\tau(\mathbf{x}) = \mathbf{x}^\mathsf{T}(\hat{\boldsymbol{\beta}} + c_\tau\hat{\gamma})$ where $c_\tau = \underset{c}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau(r_i - cs_i)$. Under the assumption that $s_i > 0$, He derived basic properties of the estimates, $\hat{\boldsymbol{\beta}}$ and $\hat{\gamma}$, based on the work of Koenker and Bassett [33] such as the consistency and the monotonicity property of conditional quantiles. However, even for linear regression quantiles, the corresponding model of He can be much more general.

Thus, a more general development of noncrossing linear regression quantiles is addressed further by Liu et al. [46] who proposed a stepwise method based on additional constraints in the estimation procedure. Indeed, with a current quantile regression function at a given quantile level, they added constraints of the form:

$$\hat{q}_{\tau_k}(x) + \delta_0 \leq q_{\tau_{k+1}}(x), k = 1, \ldots, s, \tag{1.10}$$

where δ_0 is some given small positive number, so that the next quantile regression function does not cross with the current one. They also extended this to nonparametric quantiles considering $q_\tau(x)$ as a kernel function. An embarrassing drawback of this method is that, as noted by Bondell et al. [8], the estimation of quantile curves depends on the grid of quantiles.

Fortunatly, Bondell et al. showed in [8] that, for linear quantile functions and for $\mathbf{x} \in [0,1]^d$, the number of constraints can be reduced when employing a new reparametrization in their model while ensuring the noncrossing. This reparametrization is based on transforming quantile coefficients, $\boldsymbol{\beta}_{\tau_1}, \ldots, \boldsymbol{\beta}_{\tau_s}$, into $\boldsymbol{\gamma}_{\tau_1}, \ldots, \boldsymbol{\gamma}_{\tau_s}$, where $\boldsymbol{\gamma}_{\tau_1} = (\gamma_{0,\tau_1}, \ldots, \gamma_{d,\tau_1}) = \beta_{\tau_1}$ and $\boldsymbol{\gamma}_{\tau_k} = (\gamma_{0,\tau_k}, \ldots, \gamma_{d,\tau_k}) = \beta_{\tau_k} - \beta_{\tau_{k-1}}$, for $k = 2, \ldots, s$. Thus, setting γ^-_{j,τ_k} and γ^+_{j,τ_k} to the negative and the positive part of γ_{j,τ_k}, $j = 1, \ldots, d$, the constraints

in (1.10) are reduced to $\gamma_{0,\tau_k} - \sum_{j=1}^{d} \gamma_{j,\tau_k}^{-} \geq 0$, to ensure noncrossing. An extension to nonparametric quantiles is also studied in [8] using smoothing splines.

As commonly known, spline methods are the most popular, in the literature, for nonparametric quantile regression. We quote Muggeo et al.'s approach [48] who proposed a stepwise estimation procedure using cubic spline to ensure both monotonicity and noncrossing of the estimated quantile curves. In their work, the authors considered the regression quantile $q_{\tau_k}(\mathbf{x}, z) = \mathbf{x}^{\top}\boldsymbol{\beta}_{\tau_k} + \sum_{j=1}^{J} b_{j,k}B_j(z)$, $k = 1, \ldots, s$, where $b_{j,k}$ and B_j are respectively the spline coefficients and basis functions, J is number of basis functions and z is a covariate of interest. Then, they imposed strong inequality constraints on quantile coefficients, $\theta_{\tau_k} < \theta_{\tau_{k+1}}$, where $\theta_{\tau_k} = (\boldsymbol{\beta}_{\tau_k}, b_{1,k}, \ldots, b_{J,k})$. To guarantee the smoothness of the estimated quantile curves, they used low rank B-spline basis with L_2 penalty on their coefficients since, as discussed in Bondell et al. [8], the quantile smoothing parameters increase the amount of computation without improving the estimation performance.
In this thesis, our proposed simultaneous quantile regression, will be compared with Muggeo et al.'s method for nonparametric case. We have used available code in *quantregGrowth* package in R software (see [47] for more details). We will detail this in chapter 2.

There are also methods based on considering interesting functional spaces for regression quantiles. Among others, Sangnier et al. [57] proposed a simultaneous estimation of multiple quantiles based on vector-valued reproducing kernel Hilbert space (RKHS). They supposed that $q_{\tau}(\mathbf{X}) \in H = \{f + b, f \in$

$\mathcal{K}_K, \|f\|_{\mathcal{K}_K} \leq c, b \in \mathbb{R}^s\}$ where, b is the intercept, $\mathcal{K}_K$ is an RKHS associated to the kernel valued matrix K, $||.||_{\mathcal{K}_K}$ is the norm on $\mathcal{K}_K$ and c is a positive real number. The choice of the kernel matrix K is very important since it controls the regularity of the estimation of f by $\|f\|_{\mathcal{K}_K} \leq c$. In order to minimize the burden of crossing, they considered a decomposable kernel matrix, $K : (\mathbf{x}, \mathbf{x}') \to k(\mathbf{x}, \mathbf{x}')B$ where $k : \chi \times \chi \to \mathbb{R}$ is a scalar kernel and $B = \left(\exp\{-\gamma(\tau_j - \tau_i)^2\}\right)_{1 \leq i,j \leq s}$ is a symmetric and semi-definite positive (s.d.p.) $(p \times p)$-matrix. Here, the matrix B encodes the relationship between the different conditional quantile functions and thus, controls the noncrossing of quantile curves. In this thesis, our proposed method for nonparametric Bayesian quantile regression, presented in Chapter 2, uses as prior the kernel valued matrix used by Sangnier et al..

Besides splines and RKHS methods, one can find in the literature many other approaches that are not based on solving the commonly known optimization problem for quantile regression. They are, instead, based on the estimation of the conditional distribution function F in a first step then invert it to obtain noncrossing quantile estimates.

Dette and Volgushev [13] proposed to invert the estimate of the conditional distribution function, say $\hat{F}$, and monotonizing it with respect to τ. This monotonization is based on a non-decreasing rearrangement when introducing an increasing distribution function G so that the estimate of the conditional quantile is $\hat{F}^{-1}(\tau) = G^{-1} \circ \hat{H}_x^{-1}(\tau)$ where

$$\hat{H}_x^{-1}(\tau) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{\tau} k_d \left(\frac{\hat{F}(G^{-1}(\frac{i}{N})) - u}{h_d} \right) du$$

is a smoothed non-decreasing function, k_d is a symmetric density and h_d denotes a bandwidth converging to 0.

Later on, Chernozhukov et al. improved this, in [11], by only rearranging non-monotonic quantile functions $\hat{q}^*_{\tau_k}(x)$, for $k \in \{1, \ldots, s\}$, and this is done by inverting the estimate of the conditional distribution given by $\hat{F}(y|x) = \int_0^1 (1_{\{\hat{q}^*_{\tau_k}(x) \leq y\}}) du$. In this work, it is proved that the rearranged estimates have, theoretically, smaller error than the initial ones.

Also in the context of inverting the conditional distribution function, we cite among others, Yu and Jones [72] who proposed a kernel-weighted local linear CDF and Yu [73] who extended this to nonlinear additive regression using a rule of thumb to impose constraints for monotonicity.

Next, we shall present, from the literature, an overview on Bayesian methods that estimate regression quantiles simultaneously.

III.2 Bayesian simultaneous quantile regression

In the Bayesian framework, noncrossing quantile estimates are mostly provided through simultaneous quantile inference. In the literature, some of the approaches consist on approximating the conditional distribution while others assume the ALD as a likelihood tool for the Bayesian procedure.

In what follows, we present some of these Bayesian approaches. We start with those dealing with non-ALD likelihood and then, we go on those tacking the ALD as a likelihood tool.

Some of the non-ALD methods are based on modeling the noncrossing quantile function and then deriving approximately the likelihood of the model. For linear regression quantiles, where $q_\tau(\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}(\tau)$, Reich et al. [55] proposed a two-stage method, where at the first stage, they used Koenker and Basset's method [33] to estimate quantile coefficients. Then, they analyzed the obtained estimates, $\hat{\boldsymbol{\beta}}(\tau)$, in a second stage, using Bernstein polynomial basis function, $\hat{\boldsymbol{\beta}}(\tau) = \sum_{k=1}^{M} B_k(\tau)\alpha_{j,k}$, $j = 1, \ldots, d$, where B_k are the basis functions, α_{jk} denote their coefficients and M is their number. As in [8], the reparametrized $\alpha_{j,k}$ to $\delta_{j,k}$ then modeled the $\delta_{j,k}$ with respect to a latent variable $\delta^*_{j,k}$ with Gaussian prior so that $\sum_{j=1}^{d} X_{ij}\delta_{j,k} \geq 0$ and thus, $\hat{q}_\tau(\mathbf{X}) = \mathbf{X}^\top \hat{\boldsymbol{\beta}}(\tau)$ is nondecreasing in τ. This second stage is built once assuming that the first stage quantile coefficient estimates follow a Normal distribution with covariance obtained from asymptotic results, and mean modeled through the Bernstein polynomial basis. In this method, the likelihood is approximated by computing conditional quantiles, $q_\tau(\mathbf{X})$, on a grid of quantile levels, τ_k, $k = 1, \ldots, 100$, equally spaced in $[0, 1]$, and tacking $p(y_i/\mathbf{x}_i) \approx \frac{1}{q_{\tau_{k+1}}(\mathbf{x}_i) - q_{\tau_k}(\mathbf{x}_i)}$. However, this is not a closed form for the likelihood which leads to computational difficulty. Later, Reich et al., in [53], considered Gaussian model and used a piecewise Gaussian basis function to model $q_\tau(\mathbf{X})$ and by doing so, in one stage, they used a Gaussian prior for the quantile function over all quantile levels. Fortunately, this choice of basis function enables to have a closed-form of the likelihood and also to obtain smooth quantile functions. A little later, Reich and Smith considered, in [54], a location-scale model for quantile regression given as $Y_i = \mathbf{X}_i^\top \boldsymbol{\alpha}_0 + (\mathbf{X}_i^\top \boldsymbol{\alpha}_1)\epsilon_i$, so that the regression quantile is $q_\tau(\mathbf{X}) = \mathbf{X}_i^\top \boldsymbol{\alpha}_0 + (\mathbf{X}_i^\top \boldsymbol{\alpha}_1) q_0(\tau)$, where $q_0(\tau)$ denotes the $\tau-$th quantile of ϵ. To ensure quantile monotonicity, they used

basis functions to model $q_0(\tau)$ on a grid of quantile levels and choose the centered Gaussian prior on the quantile coefficients. This enabled them to express the likelihood through a closed-form and thus to proceed with MCMC sampling.

Also, for a univariate explanatory variable with a bounded and convex domain, Tokdar showed in [65] that a linear quantile, $q_\tau(X) = \beta_0(\tau) + X\beta(\tau)$, is monotonically increasing if and only if $\beta_0(\tau)$ and $\beta(\tau)$ are linear combination of two monotonic increasing functions in $\tau \in [0,1]$, $\eta_1(\tau)$ and $\eta_2(\tau)$. Using logistic transformation, they reparametrized $\eta_1(\tau)$ and $\eta_2(\tau)$ to a smooth Gaussian process to ensure the quantile monotonicity and then specify classical priors on the parameters to continue the Bayesian formulation. However, the likelihood cannot be derived by inverting the quantile function but it is approximated on a grid of quantile levels. And recently, Yang and Tokdar [69] extended this method to the multidimensional case with arbitrary bounded convex explanatory space.

Unlike these aforementioned methods, there are also other approaches in the literature that use pseudo-likelihoods for the Bayesian procedure. Jeffreys [30] was the first who proposed the so-called substitution likelihood to deal with Bayesian median regression. This was extended further by Dunson and Taylor [14] for quantile regression where the substitution likelihood, characterized by $\theta = (\theta_1, \ldots, \theta_s)$, the vector of s quantiles, is given by:

$$L(\theta) = \begin{pmatrix} n \\ u_1(\theta), \ldots, u_{s+1}(\theta) \end{pmatrix} \prod_{l=1}^{s+1} \Delta\tau_l^{u_l(\theta)}$$

where $\begin{pmatrix} n \\ u_1(\theta), \ldots, u_{s+1}(\theta) \end{pmatrix} = \frac{n!}{u_1(\theta)!\ldots u_{s+1}(\theta)!}$, $\theta_{s+1} = \infty$, $\Delta\tau_1 = \tau_1$, $\Delta\tau_l = \tau_l - \tau_{l-1}$, $l = 2, \ldots, s$ and $u_l(\theta) = \sum_{i=1}^{n} 1_{]\theta_{l-1}, \theta_l]}(y_i)$. Considering linear regression quantiles, $\theta_i = \mathbf{X}^\top \boldsymbol{\beta}_i$, $i = 1, \ldots, s$, they specified a truncated Gaussian prior for the $\boldsymbol{\beta}_i$ such that θ_i belong to an order-restriction space, where the restriction implies $\mathbf{X}^\top \boldsymbol{\beta}_1 < \ldots < \mathbf{X}^\top \boldsymbol{\beta}_s$ and then they used Metropolis-Hastings algorithm for posterior computation.

Now it is the time to mention the broadly emerged Bayesian approaches based on the ALD as a likelihood tool, especially, when Yu and Moyeed [74] argued that it yields to satisfactory results even if the ALD is a misspecification. Among others, Sriram et al. [62] gave a more formal justification addressing the ALD issue, in Bayesian quantile regression, showing posterior consistency under some general conditions. However, this result does not mean that the estimates credible intervals constructed from the posterior are valid. This issue was handled later by Sriram et al. [63] where they showed that, asymptotically, it is suitable that 95% of the credible interval contains the true parameters 95% of the time. It is also shown that if the sample size increases, the length of the credible interval decreases. Shortly after, Yang et al. [71] and Sriram [61] proposed a correction to MCMC iterations in order to asymptotically build valid intervals. We note that the inference in [63] is based on pseudo-ALD given by

$$L(\mathbf{q}, \boldsymbol{\sigma}) = \prod_{j=1}^{s} \frac{1}{\sigma_j^n} \exp\left\{ - \sum_{i=1}^{n} \sum_{j=1}^{s} \frac{\rho_{\tau_j}(Y_i - q_{\tau_j}(\mathbf{X}_i))}{\sigma_j} \right\};$$

the embarrassing issue in this method is that this pseudo-ALD does not coincide with any probability density function and this is, actually, in line with the optimization problem addressed in the frequentist context.

Later, Rodrigues and Fan [56] proposed a two-stage method for both linear and nonparametric quantile regression. In the first stage, the authors used the standard Bayesian quantile regression [74] fitted separately at different quantile levels. The second stage consists of using Gaussian process prior for quantiles in order to control noncrossing constraints. Here they used the induced quantile notion, i.e., any quantile $q_\tau(\mathbf{X})$ can be derived from the location parameter, $q_p(\mathbf{X})$, of the ALD ($\mathcal{ALD}(q_p(\mathbf{X}), \sigma, p)$, $p = p_1, \ldots, p_s$, $p \neq \tau$);

$$q_\tau(\mathbf{X}|p) = \begin{cases} q_p(\mathbf{X}) + \frac{\sigma}{1-p}\log(\frac{\tau}{p}) & if \quad 0 \leq \tau \leq p \\ q_p(\mathbf{X}) - \frac{\sigma}{p}\log(\frac{1-\tau}{1-p}) & if \quad p \leq \tau \leq 1. \end{cases}$$

An obvious property of the estimators is that all of them are monotonically ordered. The authors provided also asymptotic property that these estimators are consistent. However, this method is not fully Bayesian and still affected by the first outputs of the first stage since different likelihoods are used for initial estimates. In Chapter 2, we will compare our proposed approach, also based on the induced quantile idea, with that of Rodrigues and Fan [56].

These cited approaches do not really infer regression quantiles simultaneously. Indeed, the simultaneous estimation of regression quantiles takes place when the likelihood is expressed through all the quantiles of interest so that it includes sufficient information for a further simultaneous update. Moreover, the approaches consisting of two steps are not truly fully Bayesian.

IV Our contributions

In this part of this thesis, we introduce a fully Bayesian method that estimates different quantiles simultaneously. Unlike, the other Bayesian methods, this proposed method is not based on misspecification, so that the ALD is the true underlying response distribution. We used the relation between quantiles of the ALD distribution and Metropolis-Hastings within Gibbs algorithm for implementation. The proposed method provides parallel quantile functions and thus the noncrossing is a natural result. It remains to be verified that the constants of the regression quantiles respect the order of quantile levels. We do not provide theoretical justification for this issue, but our numerical results show, using specific empirical criteria, that the estimated regression quantiles respect quantile level ordering, hence the good performance of the method.

2 Bayesian estimation of joint quantile regression

I Introduction

One should note that the advantages of simultaneous quantile fitting go much beyond avoiding the crossing issue. Although monotone estimates is a minimum requirement for drawing a plausible inference and thus, simply maintaining noncrossing surely ensures quantile monotonicity. Quantile fitted curves will overall be separated from one another, especially for quantiles at the tails. As data is scarce in these regions, and information from other parts of the distribution is not shared, extreme quantile estimates may present higher errors. Therefore, borrowing information across quantile levels occurs more efficiently when a simultaneous fitting is addressed.

As mentioned in Chapter 1, many Bayesian approaches that estimate multiple quantiles using the ALD distribution consist on misspecification, i.e., the ALD is not the true underlying distribution of $Y|X$, and besides, they are not fully Bayesian approaches, such as all two-stage approaches ([53], [55], [56]). Moreover, it is evident that, when the ALD is the true distribution, all methods should provide parallel estimated curves due to the relation between the

quantiles of the ALD [76]. We, in this chapter, propose a fully Bayesian approach for the simultaneous estimation of multiple quantile curves and, we restrict the case to assume that the errors follow an ALD distribution. We will implement the method and study its performance empirically using MCMC methods, in particular, Metropolis-Hastings within Gibbs sampler. We show that using the link between two quantiles of the ALD distribution, this approach practically provides parallel and thus noncrossing estimated quantile curves for both linear and nonparametric cases. Even though noncrossing is guaranteed, it is essential to make sure that the quantile constants respect the quantile level ordering. This will be addressed empirically through numerical study.

This Chapter is organized as follows. Section II describes the whole method presenting the general model and establishing the relationship between the quantiles of the ALD distribution that is used in the estimation procedure. Also, prior distributions are specified for the linear and nonparametric quantile regression model. For the linear case, we present two different versions of the model, one with overparameterization and the other with much more parsimony. The full conditional distributions are tackled in Section III for the different cases and the different model versions. We also give the algorithm associated with this method and we present, in particular, the location-scale mixture of the ALD distribution. Section IV is dedicated to the comparison between the overparameterized and the parsimonious models in the linear case. To illustrate the performance and the flexibility of our method, simulated examples are studied in Section V. Concluding remarks are stated in Section VI.

II Simultaneous Bayesian estimation of quantiles

II.1 Model

We recall that we consider the quantile regression model given by

$$Y_i = q_p(\mathbf{X}_i) + \epsilon_i, \quad i = 1, \dots, n, \tag{2.1}$$

with $(\epsilon_i)_i \overset{i.i.d.}{\sim} \mathcal{ALD}(\mu = 0, \sigma, p)$ and where μ, σ and p respectively are the location, the scale and the asymmetry parameters. This leads to assume that the response variable Y_i, given the covariable $\mathbf{X}_i$, follows an ALD distribution;

$$Y_i | (q_p(\mathbf{X}_i), \sigma, p) \sim \mathcal{ALD}(q_p(\mathbf{X}_i), \sigma, p) \quad \forall i = 1, \dots, n,$$

whose probability density function (p.d.f.) is given by

$$f(y_i; q_p(\mathbf{x}_i), \sigma, p) = \frac{p(1-p)}{\sigma} \exp\left\{ -\frac{\rho_p(y_i - q_p(\mathbf{x}_i))}{\sigma} \right\}, \tag{2.2}$$

with $\rho_p(u) = u(p - \mathbb{1}_{u<0})$ is the loss function. The explanatory variables $\mathbf{X}_i$'s are supposed to be i.i.d. distributed according to an arbitrary continuous distribution, $\mathbb{P}_{\mathbf{X}}$, whose support is $\mathcal{X}^d$, $\mathcal{X} \subset \mathbb{R}$, $d \geq 1$. In what follows, the function $q_p(\mathbf{X})$, the paramters σ, p and $\mathbb{P}_{\mathbf{X}}$ are supposed to be unknown.

In all subsequent sections, any quantity in bold represents a vector, capital letters denote random variables or vectors whereas lowercase letters denote observed values of the corresponding random vectors.

II.2 Bayesian procedure

II.2.1 Likelihood

To infer simultaneously s distinct quantiles with a fully Bayesian approach, say $q_{\tau_1}, \ldots, q_{\tau_s}$ with distinct orders $\tau_1, \ldots, \tau_s$, of the conditional distribution of $Y|\mathbf{X}$, one needs to characterize the likelihood through all these s unknown quantiles.

As explained below, this is done first by partitioning the whole sample into s sub-samples, which requires a sufficient number of observations, second by using the relation between any two quantiles of the ALD distribution, and third by rewriting the whole likelihood into s terms where the j-th term only depends on q_{τ_j}.

1. Consider the following partition of the whole sample into s sub-samples:

$$(\mathbf{X}_{I_j}, \mathbf{Y}_{I_j}) = \left\{(\mathbf{X}_i, Y_i)_{i \in I_j}\right\}, \forall j \in \{1, \ldots, s\},$$

   where $I_j = \{(j-1)r + 1, \ldots, jr\}$ with $I_j \bigcap I_k = \emptyset$, $\forall j \neq k$ and $j, k \in \{1, \ldots, s\}$; we suppose, without loss of generality, that $r = \dfrac{n}{s}$ is integer.

2. To characterize the likelihood through all quantiles of interest, we use the relation that links any $\tau-$th quantile to the $p-$th quantile of the $\mathcal{ALD}(q_p(\mathbf{X}), \sigma, p)$ (see for e.g. [76] and [1]):

$$q_\tau(\mathbf{X}) = q_p(\mathbf{X}) + \sigma g(\tau, p), \tag{2.3}$$

   where $g(\tau, p) = \frac{1}{1-p}\log(\frac{\tau}{p})\mathbb{1}_{0<\tau\leq p} - \frac{1}{p}\log(\frac{1-\tau}{1-p})\mathbb{1}_{p<\tau<1}$.

3. From Equation (2.3), we rewrite the model given by (2.1) as follows:

$$\mathbf{Y}_{I_j} = \mathbf{q}_{\tau_j,I_j} - \sigma \mathbf{g}_{\tau_j,p} + \boldsymbol{\epsilon}_{I_j},\ \forall j = 1,\dots,s, \tag{2.4}$$

where

$$\begin{aligned}
\mathbf{Y}_{I_j} &= (Y_{i_j})_{i_j \in I_j},\\
\mathbf{q}_{\tau_j,I_j} &= (q_{\tau_j}(\mathbf{X}_{i_j}))_{i_j \in I_j},\\
\mathbf{g}_{\tau_j,p} &= (g(\tau_j,p)\mathbb{1}_{i_j \in I_j})_{i_j \in I_j},\\
\boldsymbol{\epsilon}_{I_j} &= (\epsilon_{i_j})_{i_j \in I_j}.
\end{aligned}$$

Then, the likelihood associated to the model given by Equation (2.1) is rewritten as the product of s likelihoods, each one only depends on one sub-sample:

$$\begin{aligned}
&L(\mathbf{q}_{\tau_{1,I_1}},\dots,\mathbf{q}_{\tau_{s,I_s}},\sigma,p;\mathbf{x}_{I_1},\dots,\mathbf{x}_{I_s},\mathbf{y}_{I_1},\dots,\mathbf{y}_{I_s}) =\\
&\prod_{j=1}^{s}\prod_{i_j \in I_j} \frac{p(1-p)}{\sigma} \exp\left\{-\frac{\rho_p\left(y_{i_j} - (q_{\tau_j}(\mathbf{x}_{i_j}) - \sigma g(\tau_j,p))\right)}{\sigma}\right\}.
\end{aligned}$$

II.2.2 Priors

In this section, we specify the prior distributions for all unknown quantities. It worth mentioning that even $\mathbb{P}_{\mathbf{X}}$ is not our primary interest, it is unknown, and therefore it should require a prior; but since any prior on $\mathbb{P}_{\mathbf{X}}$ that is independent of the prior on $((\mathbf{q}_{\tau_j,I_j})_{j=1,\dots,s},\sigma,p)$, would disappear upon marginalization of the posterior of $(\mathbb{P}_{\mathbf{X}},(\mathbf{q}_{\tau_j,I_j})_{j=1,\dots,s},\sigma,p)$ relatively to $\mathbb{P}_{\mathbf{X}}$, we drop it in the sequel. Thus, it suffices to choose a prior distribution for σ, p and $\mathbf{q}_{\tau_j,I_j}$ for $j = 1,\dots,s$; we note that our choice of priors is quite classical.

1. For the quantiles, we distinguish between the linear and nonparametric settings.

   (a) **Linear case** i.e. $q_{\tau_j}(\mathbf{X}) = \mathbf{X}^\top \boldsymbol{\beta}_{\tau_j} = \beta^{(0)}_{\tau_j} + \mathbf{X}^\top \boldsymbol{\beta}^{(x)}_{\tau_j}$, $j \in \{1, \ldots, s\}$, where $\mathbf{X}^\top$ denotes the transpose of $\mathbf{X}$ and $\boldsymbol{\beta}^{(x)}_{\tau_j} \in \mathbb{R}^d$.

   We propose to study two model versions: one ignores that the model is under ALD assumption so that the slopes, $\boldsymbol{\beta}^{(x)}_{\tau_1}, \ldots, \boldsymbol{\beta}^{(x)}_{\tau_s}$, of the different quantiles, are distinct and the model is overparameterized, and the second version take into account the ALD assumption which means that, according to Equation (2.3), all quantiles have the same slope that is equal to the slope of the $p-$th quantile so that the quantiles curves are parallel and the model is much more parsimonious. Therefore, Model (2.1) will be rewritten, into these two versions, with more details in Section III where the full conditional distributions are computed for each. Before reaching that, we shall specify prior distributions on parameters for both the overparameterized and the parsimonious models.

   i. *Overparameterization*

      Let us consider the unknown parameter vector $\tilde{\boldsymbol{\beta}} = (\beta^{(0)}_{\tau_1}, \boldsymbol{\beta}^{(x)}_{\tau_1}, \ldots, \beta^{(0)}_{\tau_s}, \boldsymbol{\beta}^{(x)}_{\tau_s})$ of dimension $s(d+1)$. For this parameter vector, we choose a zero-mean Gaussian distribution as prior:

      $$\tilde{\boldsymbol{\beta}} \sim \mathcal{N}_{s(d+1)}(0, \tilde{\Sigma}_0),$$

      with $\tilde{\Sigma}_0$ is an $s(d+1) \times s(d+1)$ invertible and positive definite square matrix.

   ii. *Parsimony*

Due to Equation (2.3), one should note that distinct quantiles differ only by the intercept $\beta_{\tau_j}^{(0)}$, so that all quantiles of interest have the same slope ($\boldsymbol{\beta}_{\tau_1}^{(x)} = \ldots = \boldsymbol{\beta}_{\tau_s}^{(x)} = \boldsymbol{\beta}_p^{(x)}$); then we consider the unknown vector parameter $\check{\boldsymbol{\beta}} = (\beta_{\tau_1}^{(0)}, \ldots, \beta_{\tau_s}^{(0)}, \boldsymbol{\beta_p})$ of dimension $s + d$ with Gaussian prior distribution,

$$\check{\boldsymbol{\beta}} \sim \mathcal{N}_{s+d}(0, \check{\Sigma}_0),$$

where $\check{\Sigma}_0$ is a positive definite square matrix of dimension $s + d$.

(b) **Nonparametric case**

We set h to be the function defined on $[0,1] \times \mathcal{X}^d$, by $h(\tau, \mathbf{x}) = q_\tau(\mathbf{x})$. We put a Gaussian process prior on h, i.e., $h \sim \mathcal{GP}(0, k)$ with a zero-mean function and a covariance function $k : [0,1]^2 \times \mathcal{X}^{2d} \to \mathbb{R}$; following [57], we choose k to be decomposable taking the following form:

$$k((\tau_i, \mathbf{x}), (\tau_j, \mathbf{x}')) = k_x(\mathbf{x}, \mathbf{x}') \exp(-c(\tau_j - \tau_k)^2), \tag{2.5}$$

where $k_x(\mathbf{x}, \mathbf{x}') = \exp\left(-b\|\mathbf{x} - \mathbf{x}'\|^2\right)$, b and c are positive hyperparameters and $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$. Here, $k((\tau_i, \mathbf{x}), (\tau_j, \mathbf{x}'))$, $i \neq j$, encodes the relation between the conditional quantiles $q_{\tau_i}(\mathbf{x})$ and $q_{\tau_j}(\mathbf{x}')$. As explained by [57], if $c \to 0$, the quantile curves are parallel so they do not cross. If $c \to +\infty$, the quantiles are learned independently and then they may cross. Therefore, the choice of c is important to control the occurrence of crossing.

2. $\sigma \sim \mathcal{IG}(a_0, b_0), \quad a_0 > 0,\ b_0 > 0$, where $\mathcal{IG}$ denotes the Inverse Gamma distribution with positive hyperparameters a_0 and b_0.

3. $p \sim Beta(\alpha_0, \lambda_0), \quad \alpha_0, \lambda_0 > 0$.

III Computations

Now, we are interested in generating posterior samples of the conditional quantiles. Since this is not possible directly from the posterior distribution, we are going to use MCMC methods. In particular, as shown by [39], to explicitly compute the full conditional distributions required for Gibbs sampler, we shall make use of the location-scale mixture of the ALD distribution [38].

Let ω be an exponential latent variable with parameter $1/\sigma$, denoted by $\mathcal{E}(1/\sigma)$, and Z be a standard normal variable such that ω and Z are independent. If ϵ has an ALD distribution, $\epsilon \sim \mathcal{ALD}(0, \sigma, p)$, then it can be represented as a mixture of normal variable,

$$\epsilon = \gamma_p \omega + \delta_p \sqrt{\sigma \omega} z, \tag{2.6}$$

where $\gamma_p = \dfrac{1-2p}{p(1-p)}$ and $\delta_p^2 = \dfrac{2}{p(1-p)}$. Exploiting this augmented data structure, the model defined by the system of equations given by (2.4) admits, conditionally on $\boldsymbol{\omega}_{I_j} = (\omega_i)_{i \in I_j} \overset{i.i.d.}{\sim} \mathcal{E}(1/\sigma)$, the following Gaussian representation:

$$\mathbf{Y}_{I_j} = \mathbf{q}_{\tau_j, I_j} - \sigma \mathbf{g}_{\tau_j, p} + \gamma_p \boldsymbol{\omega}_{I_j} + \delta_p \sqrt{\sigma \boldsymbol{\omega}_{I_j}} \boldsymbol{z}_{I_j}, \quad j = 1, \ldots, s, \tag{2.7}$$

where $\boldsymbol{z}_{I_j} = (z_i)_{i \in I_j} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.

This location-scale mixture representation of the ALD allows to easily sample regression coefficients through Gibbs sampler ([39], [7]).

III.1 Full conditional distributions

From the equations given in (2.7) and the priors defined in Section II.2.2, we follow [7] to derive the full conditional distributions of $\boldsymbol{\omega}_n = (\omega_1, \ldots, \omega_n)$ and $(\mathbf{q}_{\tau_{j,I_j}})_{j=1,\ldots,s}$. Yet, the full conditionals of σ and p are not tractable. To handle this, we include a Metropolis-Hasting step to the Gibbs sampler.

For further details, we set $\mathbf{Y}_n = (Y_i)_{i=1,\ldots,n}$ to be the vector of observations and $X_n = (\mathbf{X}_i)_{i=1,\ldots,n}$ to be the covariable matrix.

1. For all $j = 1, \ldots, s$, and $i \in I_j$, set $v_i = \omega_i^{-1}$ and considering the distribution $\omega_i \sim \mathcal{E}(1/\sigma)$ as a prior on ω_i, one gets

$$v_i | (y_i, q_{\tau_j}(\mathbf{x}_i), \sigma, p) \sim \mathcal{IG}auss(\Psi_i, \phi_p)$$

with

$$\Psi_i = \sqrt{\frac{\gamma_p^2 + 2\delta_p^2}{(y_i - q_{\tau_j}(\mathbf{x}_i) + \sigma g(\tau_j, p))^2}}, \; \phi_p = \frac{\gamma_p^2 + 2\delta_p^2}{\delta_p^2 \sigma},$$

where $\mathcal{IG}auss$ stands for the Inverse Gaussian distribution with $\Psi_i > 0$ and $\phi_p > 0$ as location and shape parameters.

2. The conjugate Gaussian prior on quantiles provides Gaussian full conditional distributions in both parametric and nonparametric cases.

   (a) **Linear case.** In order to derive the full conditional distributions for overparameterized and parsimonious models, let us first introduce the notation $\boldsymbol{g}_n = (g(\tau_j, p)\mathbb{1}_{i \in I_j})_{1 \leq j \leq s}$ used in both models.

i. *Overparameterization*

Setting $\tilde{\mathbf{X}} = (\tilde{X}_{i,k})_{1\leq i\leq n, 1\leq k\leq s(d+1)}$ to be the $n \times s(d+1)$ design matrix defined by

$$\tilde{X}_{i,k} = \mathbb{1}_{i\in I_j}\left(\mathbb{1}_{k=(d+1)(j-1)+1} + X_{i,k}\mathbb{1}_{k\in\{(d+1)(j-1)+2,\ldots,(d+1)j\}}\right),$$

allows to rewrite the systems of equations (2.7) into the following format:

$$\mathbf{Y}_n = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} - \sigma \boldsymbol{g}_n + \gamma_p \boldsymbol{\omega}_n + \boldsymbol{\epsilon}_n^*,$$

with $\boldsymbol{\epsilon}_n^* \sim \mathcal{N}_n(0, \Sigma^*)$ and where $\Sigma^* = diag(\delta_p^2 \sigma \boldsymbol{\omega}_n)$. We note that the matrix $\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}$ is not invertible.

With this zero-mean Gaussian prior distribution, the full conditional distribution of $\tilde{\boldsymbol{\beta}}$ is given by:

$$\tilde{\boldsymbol{\beta}}|(\tilde{\boldsymbol{x}}, \boldsymbol{y}_n, \boldsymbol{\omega}_n, \sigma, p) \sim \mathcal{N}_{s(d+1)}(\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{\beta}}}, \hat{\Sigma}_{\tilde{\boldsymbol{\beta}}})$$

with

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{\beta}}} &= \left(\tilde{\boldsymbol{x}}^\top\Sigma^*\tilde{\boldsymbol{x}} + \tilde{\Sigma}_0^{-1}\right)^{-1}\tilde{\boldsymbol{x}}^\top\Sigma^*(\boldsymbol{y}_n + \sigma\boldsymbol{g}_n - \gamma_p\boldsymbol{\omega}_n), \\
\hat{\Sigma}_{\tilde{\boldsymbol{\beta}}} &= \left(\tilde{\boldsymbol{x}}^\top\Sigma^*\tilde{\boldsymbol{x}} + \tilde{\Sigma}_0^{-1}\right)^{-1}.
\end{aligned}$$

ii. *Parsimony*

Here, a different design matrix with different dimension is considered. Set $\check{\mathbf{X}} = (\check{X}_{i,l})_{1\leq i\leq n, 1\leq l\leq s+d}$, the underlying $n \times$

$(s+d)$ design matrix defined by

$$\check{X}_{i,l} = \begin{cases} \mathbb{1}_{i \in I_l} & if \qquad l \in 1, \dots s, \\ X_{i,l-s} & if \quad l \in s+1, \dots, s+d \end{cases}$$

Then, the system of equations (2.7) can be rewritten as:

$$\mathbf{Y}_n = \check{\mathbf{X}}\check{\boldsymbol{\beta}} - \sigma \boldsymbol{g}_n + \gamma_p \boldsymbol{\omega}_n + \boldsymbol{\epsilon}_n^*,$$

with $\boldsymbol{\epsilon}_n^* \sim \mathcal{N}_n(0, \Sigma^*)$ and where $\Sigma^* = diag(\delta_p^2 \sigma \boldsymbol{\omega}_n)$. Implemented similarly to that in the overparameterization case, the full conditional distribution on $\check{\boldsymbol{\beta}}$ is Gaussian:

$$\check{\boldsymbol{\beta}} | (\check{\boldsymbol{x}}, \boldsymbol{y}_n, \boldsymbol{\omega}_n, \sigma, p) \sim \mathcal{N}_{s+d}(\hat{\boldsymbol{\mu}}_{\check{\boldsymbol{\beta}}}, \hat{\Sigma}_{\check{\boldsymbol{\beta}}}),$$

with

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\check{\boldsymbol{\beta}}} &= \left(\check{\boldsymbol{x}}^\mathsf{T} \Sigma^* \check{\boldsymbol{x}} + \check{\Sigma}_0^{-1} \right)^{-1} \check{\boldsymbol{x}}^\mathsf{T} \Sigma^* (\boldsymbol{y}_n + \sigma \boldsymbol{g}_n - \gamma_p \boldsymbol{\omega}_n), \\ \hat{\Sigma}_{\check{\boldsymbol{\beta}}} &= \left(\check{\boldsymbol{x}}^\mathsf{T} \Sigma^* \check{\boldsymbol{x}} + \check{\Sigma}_0^{-1} \right)^{-1}. \end{aligned}$$

(b) **Nonparametric case.** We shall use another extra notations:

$$\begin{aligned} \boldsymbol{h}_n &= \left(\boldsymbol{h}_{\tau_j, I_j} \right)_{j=1,\dots,s} \\ &= \left(\left(h(\tau_j, \mathbf{x}_i) \right)_{i \in I_j} \right)_{j=1,\dots,s}, \\ \boldsymbol{k_n(\tau, x)} &= \left((k(\tau_j, \boldsymbol{x}_{i_j})_{i_j \in I_j})_{j=1,\dots,s} \right). \end{aligned}$$

This allows to rewrite the system of equations (2.7) into the following vector format:

$$\mathbf{Y}_n = \boldsymbol{h}_n - \sigma \boldsymbol{g}_n + \gamma_p \boldsymbol{\omega}_n + \boldsymbol{\epsilon}_n^*. \tag{2.8}$$

Combining (2.8) with the Gaussian process prior on h leads to the following joint distribution of $(\mathbf{Y}_n, \boldsymbol{h}_n)$ conditional on $(x_n, \boldsymbol{\omega}_n, \sigma, p)$:

$$\begin{pmatrix} \mathbf{Y}_n \\ \boldsymbol{h}_n \end{pmatrix} \sim \mathcal{N}_{2n}\left(\begin{pmatrix} -\sigma \boldsymbol{g}_n + \gamma_p \boldsymbol{\omega}_n \\ \mathbf{0}_n \end{pmatrix}, \begin{pmatrix} \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) + \Sigma^* & \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) \\ \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) & \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) \end{pmatrix} \right).$$

Finally, classical calculations lead to the desired full conditional distribution:

$$\boldsymbol{h}_n | (\mathbf{x}_n, \boldsymbol{y}_n, \boldsymbol{\omega}_n, \sigma, p) \sim \mathcal{N}(\hat{\boldsymbol{\mu}}_{\boldsymbol{h}}, \hat{\Sigma}_h),$$

where

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\boldsymbol{h}} &= \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x})(\boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) + \Sigma^*)^{-1}(\boldsymbol{y}_n - \gamma_p \boldsymbol{\omega}_n + \sigma \boldsymbol{g}_n), \\ \hat{\Sigma}_h &= \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) - \boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x})(\boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}) + \Sigma^*)^{-1}\boldsymbol{k_n}(\boldsymbol{\tau}, \boldsymbol{x}). \end{aligned}$$

3. The full conditional distribution, π_σ of $\sigma | \mathbf{y}_n, \mathbf{x}_n, \boldsymbol{\omega}_n, p$ is proportional to

$$\frac{1}{(\sqrt{\sigma})^{n+a_0+1}} \exp\left\{ -\frac{1}{2}(\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \boldsymbol{g}_n - \gamma_p \boldsymbol{\omega}_n)^\mathsf{T} (\Sigma^*)^{-1} (\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \boldsymbol{g}_n - \gamma_p \boldsymbol{\omega}_n) \right\}, \tag{2.9}$$

where $\mathbf{q}_{\boldsymbol{\tau},n} = \left(\mathbf{q}_{\tau_j, I_j} \right)_{j=1,\ldots,s}$.

4. The full conditional, π_p of $p|\mathbf{y}_n, \mathbf{x}_n, \boldsymbol{\omega}_n, \sigma$ is proportional to

$$p^{\alpha_0}(1-p)^{\lambda 0} \exp\left\{-\frac{1}{2}(\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \boldsymbol{g}_n - \gamma_p \boldsymbol{\omega}_n)^\top (\Sigma^*)^{-1}(\mathbf{Y}_n - \mathbf{q}_{\boldsymbol{\tau},n} + \sigma \boldsymbol{g}_n - \gamma_p \boldsymbol{\omega}_n)\right\}. \quad (2.10)$$

III.2 Algorithm

Since it is not possible to generate σ and p directly from their full conditional distribution by equations (2.9) and (2.10), they are simulated by incorporating a random walk Metropolis-Hastings step within the Gibbs sampler, as described below.

For a subset $E \subset \mathbb{R}$, denote by $\mathcal{N}_E(\cdot,\cdot)$ the truncated version on E of the corresponding Gaussian distribution. Note that the choice of the truncated Gaussian distribution is classical and provides high Metropolis-Hastings ratio.

Algorithm 1 Metropolis-Hastings step

1: **Initialization** $t = 0$: $\sigma^{(0)} = \sigma_0$, $p^{(0)} = p_0$.
2: **Loop** $t = t + 1$:

1.
   - $\sigma_{prop}|\sigma^{(t)} \sim \mathcal{N}_{]0,\infty]}(\sigma^{(t)}, sd_\sigma)$,
   - $\alpha_\sigma = \min\{1, P_\sigma Q_\sigma\}$
   - $u_1 \sim \mathcal{U}_{[0,1]} \Rightarrow \sigma^{(t+1)} = \begin{cases} \sigma_{prop} & if \quad \alpha > u_1 \\ \sigma^{(t)} & otherwise \end{cases}$

2.
   - $p_{prop}|p^{(t)} \sim \mathcal{N}_{]0,1[}(p^{(t)}, sd_p)$,
   - $\alpha_p = \min\{1, P_p Q_p\}$
   - $u_2 \sim \mathcal{U}_{[0,1]} \Rightarrow p^{(t+1)} = \begin{cases} p_{prop} & if \quad \alpha > u_2 \\ p^{(t)} & otherwise \end{cases}$

It worth mentioning that the values of the scale parameters sd_σ and sd_p are calibrated to quickly achieve the equilibrium of the random walk Metropolis-Hastings step; in fact, they are chosen neither too small nor too large so that the acceptance rate becomes practically stable.

The conditional posterior ratios of σ and p, are given by

$$\begin{aligned} P_\sigma &= \frac{\pi_\sigma\left(\sigma_{prop}|\mathbf{y}_n,\mathbf{x}_n,((\mathbf{q}_{\tau_{j,I_j}})_{j=1,\dots,s})^{(t+1)},\boldsymbol{\omega}_n^{(t+1)},p^{(t)}\right)}{\pi_\sigma\left(\sigma^{(t)}|\mathbf{y}_n,\mathbf{x}_n,((\mathbf{q}_{\tau_{j,I_j}})_{j=1,\dots,s})^{(t+1)},\boldsymbol{\omega}_n^{(t+1)},p^{(t)}\right)} \\ P_p &= \frac{\pi_p\left(p_{prop}|\mathbf{y}_n,\mathbf{x}_n,((\mathbf{q}_{\tau_{j,I_j}})_{j=1,\dots,s})^{(t+1)},\boldsymbol{\omega}_n^{(t+1)},p^{(t)}\right)}{\pi_p\left(p^{(t)}|\mathbf{y}_n,\mathbf{x}_n,((\mathbf{q}_{\tau_{j,I_j}})_{j=1,\dots,s})^{(t+1)},\boldsymbol{\omega}_n^{(t+1)},p^{(t)}\right)}, \end{aligned}$$

where $\pi_\sigma(\cdot|\cdot)$ and $\pi_p(\cdot|\cdot)$ are respectively given up to a constant in equations (2.9) and (2.10), and the transition probabilities are given by

$$Q_\sigma = \frac{f_{\mathcal{N}_{[0,\infty[}(\sigma^{(t)},sd_\sigma)}}{f_{\mathcal{N}_{[0,\infty[}(\sigma_{prop},sd_\sigma)}} = \frac{1-\phi(\frac{-\sigma_{prop}}{sd_\sigma})}{1-\phi(\frac{-\sigma^{(t)}}{sd_\sigma})},$$

$$Q_p = \frac{f_{\mathcal{N}_{[0,1[}(p^{(t)},sd_p)}}{f_{\mathcal{N}_{[0,1[}(p_{prop},sd_p)}} = \frac{\phi(\frac{1-p_{prop}}{sd_p})-\phi(\frac{-p_{prop}}{sd_p})}{(\phi(\frac{1-p^{(t)}}{sd_p})-\phi(\frac{-p^{(t)}}{sd_p}))},$$

where $f_{\mathcal{N}_E(\cdot,\cdot)}$ denotes the pdf of $\mathcal{N}_E(\cdot,\cdot)$ and $\phi(.)$ denotes the cumulative distribution function of the standard normal distribution.

IV Empirical comparison in linear case

As mentioned in the previous sections, we propose to study the linear quantile regression by two different ways, overparameterization and parsimony approaches. In this section, we shall show, through a numerical example, that the parsimony approach provides better results than that with overparameterization.

IV.1 Numerical example

We generate n observations issued from the model $Y = 1 + 2X + \epsilon$ with $\epsilon \sim \mathcal{ALD}(0, \sigma = 0.5, p = 0.25)$ and $X \sim \mathcal{U}_{[-1,1]}$. We consider three quantiles of interest of orders τ_1 =0.1, τ_2 =0.15 and τ_3 =0.25. We denote by $\boldsymbol{\beta_\tau} = (\beta_\tau^{(0)}, \beta_\tau^{(1)})$, $\tau \in \{\tau_1, \tau_2, \tau_3\}$, the quantiles coefficients where $\beta_\tau^{(0)}$ denotes the intercept while $\beta_\tau^{(1)}$ is the slope. The true quantile coefficients are $\boldsymbol{\beta_{0.1}}$ =(0.389, 2), $\boldsymbol{\beta_{0.15}}$ =(0.659, 2) and $\boldsymbol{\beta_{0.25}}$ =(1,2), respectively.

For both approaches, we evaluate the estimation performance for different values of the number of observations, n = 500, 1000, 2000. Runs are made with 10000 iterations, one third of which is burn-in. For the hyperparameters, we set: $a_0 = 1$ and $b_0 = 0.01$ for the Gamma prior distribution of σ, $\alpha_0 = 2$ and $\lambda_0 = 2$ for the Beta prior of p. We will discuss in next subsection the form of the covariances matrices, $\tilde{\Sigma}_0$ and $\check{\Sigma}_0$ of the Gaussian prior distribution of $\tilde{\boldsymbol{\beta}}$ and $\check{\boldsymbol{\beta}}$, respectively. The proposals parameters of the Metropolis-Hasting step are chosen to be $sd_\sigma = 0.03$ and $sd_p = 0.01$. For the output of the algorithm, we denote by $\hat{\theta}$, the posterior mean of the parameter θ, $\hat{\theta} \in \{\hat{\beta}_\tau^{(0)}, \hat{\beta}_\tau^{(1)}, \hat{\sigma}, \hat{p}\}$, $\tau = \tau_1, \tau_2$, and $\hat{q}(X)_\tau = \hat{\beta}_\tau^{(0)} + X\hat{\beta}_\tau^{(1)}$ the resulting posterior mean, estimate of the $\tau-$th quantile function.

IV.1.1 Choice of $\tilde{\Sigma}_0$ and $\check{\Sigma}_0$

Essentially, the prior covariance matrices $\tilde{\Sigma}_0$ and $\check{\Sigma}_0$ encode the linear relation between the components of $\tilde{\boldsymbol{\beta}}$ and $\check{\boldsymbol{\beta}}$ respectively, and thus, the linear relation between the conditional quantiles. Due to equation (2.3), for p fixed, the absolute value of the distance between two intercepts, $\beta_{\tau_k}^{(0)}$ and $\beta_{\tau_l}^{(0)}$, $k, l \in 1, 2, 3$, is equal to $\sigma|g(\tau_k, p) - g(\tau_l, p)|$. Therefore, due to the expression of g, a possible choice for the covariance matrices, $\tilde{\Sigma}_0$ and $\check{\Sigma}_0$, is

such that, $cov(\beta_{\tau_k}^{(0)}, \beta_{\tau_l}^{(0)}) = \xi |\log\left(\frac{\tau_k}{\tau_j}\right)|$ where ξ is a positive real number, $Var(\beta_{\tau_k}^{(0)}) = Var(\beta_{\tau_k}^{(1)}) = 1$ and, all slopes are independent from each other and from the intercepts.

Another choice for $\tilde{\Sigma}_0$ and $\check{\Sigma}_0$ is the classical identity matrix, ($\tilde{\Sigma}_0 = Id_6$ and $\check{\Sigma}_0 = Id_4$). These choices will be compared to each other through the estimation results given in the next subsection.

IV.1.2 Estimation results

For estimated quantile curves, the crossing is evaluated through the *crossing loss* criterion (see [57]) that measures how big is the difference between $\hat{q}_{\tau_l}(X)$ and $\hat{q}_{\tau_k}(X)$, when $\hat{q}_{\tau_l}(X) < \hat{q}_{\tau_k}(X)$ $\tau_k < \tau_l, k, l \in \{1, 2, 3\}$, that is,

$$crossing\ loss = \frac{1}{n}\sum_{i=1}^{n} \max\left(0, \hat{q}_{\tau_k}(X_i) - \hat{q}_{\tau_l}(X_i)\right), \tau_k < \tau_l. \tag{2.11}$$

The numerical results are presented in Table 2.1 below. Speaking of the choice of the covariance matrix, according to Table 2.1, it is clear that, for both overparameterization and parsimony, there are no significant changes in the estimation of the different parameters whatever is the value of ξ. The same occurs when $\tilde{\Sigma}_0 = Id_6$ and $\check{\Sigma}_0 = Id_4$. However, in the case of overparameterization, the choice of ξ has an impact on crossing occurrence when n is small. As shown in Table 2.1, for n =500, the *crossing loss* is equal to 0.0005 for ξ =300 and it decreases to be negligible (3.6×10^{-8}) for ξ =100; for this latter choice, the total number of crossing (the number of data point below $\hat{q}_{\tau_k}(X)$ that are above $\hat{q}_{\tau_l}(X)$ when $\tau_k < \tau_l$) is small relative to n. Yet, for a very small value of ξ (0.1), the *crossing loss* is null, which means that no crossing occurs. The same occurs for $\tilde{\Sigma}_0 = Id_6$. In the case of parsimony, the

noncrossing holds whatever n is small or large and whatever the form of $\check{\Sigma}_0$ is.

		n	$\hat{p}$	$\hat{\sigma}$	$\hat{\beta}_{0.1}$	$\hat{\beta}_{0.15}$	$\hat{\beta}_{0.25}$	*crossing loss*
Overparametrization		500	0.2472	0.5223	(0.1993, 1.7777)	(0.8208 ,2.0886)	(0.9896, 1.9013)	5×10^{-4}
	$\xi = 300$	1000	0.256	0.5228	(0.2762,1.7956)	(0.5834,1.8537)	(0.921,2.0546)	0
		2000	0.2469	0.4951	(0.4189, 1.954)	(0.6997, 1.9334)	(0.9677, 1.9243)	0
		500	0.2463	0.522	(0.1992, 1.7773)	(0.8157, 2.0789)	(0.989, 1.9054)	3.6×10^{-8}
	$\xi = 100$	1000	0.2557	0.5238	(0.2734,1.7851)	(0.5875,1.8552)	(0.9222,2.0621)	0
		2000	0.249	0.4986	(0.412, 1.9427)	(0.6982, 1.9353)	(0.9684, 1.9183)	0
		500	0.245	0.5199	(0.2077, 1.7051)	(0.8015, 2.0116)	(0.976, 1.855)	0
	$\xi = 0.1$	1000	0.2563	0.522	(0.2728,1.7608)	(0.5856,1.8326)	(0.9187,2.0384)	0
		2000	0.2466	0.495	(0.4193, 1.9312)	(0.699, 1.9253)	(0.9668, 1.908)	0
		500	0.2471	0.526	(0.1944, 1.688)	(0.7972, 2.0066)	(0.9788, 1.8609)	0
	Id_6	1000	0.2559	0.5231	(0.2764,1.75)	(0.5829,1.8328)	(0.9173,2.0398)	0
		2000	0.2482	0.4961	(0.4157, 1.9266)	(0.6961, 1.9235)	(0.9683, 1.906)	0
Parsimony		500	0.2377	0.5091	(0.4146, 1.8584)	(0.7078, 1.8584)	(0.7985, 1.8584)	0
	$\xi = 300$	1000	0.2552	0.525	(0.3242,1.9438)	(0.6546,1.9438)	(1.1321,1.9438)	0
		2000	0.2475	0.496	(0.4168, 1.9311)	(0.6182,1.9311)	(1.0471,1.9311)	0
		500	0.2361	0.5083	(0.4187, 1.8564)	(0.7098, 1.8564)	(0.7944, 1.8564)	0
	$\xi = 100$	1000	0.254	0.5219	(0.3299,1.9432)	(0.6568,1.9432)	(1.13141.9432)	0
		2000	0.2466	0.4943	(0.4196, 1.9301)	(0.6182,1.9301)	(1.0468,1.9301)	0
		500	0.2366	0.5086	(0.4146, 1.8569)	(0.7012, 1.8569)	(0.7858, 1.8569)	0
	$\xi = 0.1$	1000	0.2549	0.5244	(0.3254,1.9423)	(0.6514,1.9423)	(1.12265,1.9423)	0
		2000	0.2473	0.4954	(0.417, 1.9304)	(0.6161, 1.9304)	(1.0453, 1.9304)	0
		500	0.238	0.5109	(0.4108 ,1.8559)	(0.6991,1.8559)	(0.7859,1.8559)	0
	Id_4	1000	0.2543	0.5225	(0.3259,1.9421)	(0.6534,1.9421)	(1.1258,1.9421)	0
		2000	0.251	0.4955	(0.4165, 1.9287)	(0.6145, 1.9287)	(1.0488, 1.9287)	0

TABLE 2.1: Estimation results and crossing evaluation of overparameterized and parsimonious models for $n = 500, 1000, 2000$ and for different prior covariance matrices (ξ =300, 100, 0.1) and $\Sigma_0 = Id$, where $\Sigma_0 \in \{\tilde{\Sigma}_0, \check{\Sigma}_0\}$).

Here, expressing the covariance matrix of the prior Gaussian distribution with respect to the quantile orders does not have any effect on the quality of the estimation; indeed, the information that it keeps about the quantile

coefficients is already involved in the model due to the relation, given by equation (2.3), between the quantiles of the ALD distribution. This allows using the identity matrix as a simple and classic choice for the covariance matrix of the prior Gaussian distribution. Besides that, for small n, results of the estimation based on overparameterization are not good compared to that of the estimation based on parsimony. Table 2.1 shows that the estimated values of the quantile coefficients are not close to the true values when n is small (n =500, 1000) while with parsimony, the estimated quantile coefficients are close enough from the true ones even for small n.

Based on all these results, we can say that the estimation based on the overparameterization guarantees neither good estimation, for small n, nor parallel estimated quantile curves unlike the estimation with fewer parameters. Since our main objective, in this chapter, is to ensure noncrossing in simultaneous quantile regression, the estimation based on considering the same slope for all quantiles, we adopted the estimation with parsimony that provides good parallel estimators of quantile curves since it is suitable to guarantee the noncrossing.

V Simulation study

In this section, we study the performance of our method in both linear and nonparametric quantile regression cases. For the model given by (2.1), we shall consider three different designs for the $p-$th quantile:

1. Univariate linear quantile: $q_p(X) = 1 + 2X$, with $X \sim \mathcal{U}_{[-1,1]}$;

2. Multivariate linear quantile: $q_p(\mathbf{X}) = 1 + \mathbf{X}^\top \boldsymbol{\beta}_p$, with $\boldsymbol{\beta}_p \in \mathbb{R}^d$, $d = 10$ and either $X_l \sim \mathcal{U}_{[-1,1]}$ or $X_l \sim \mathcal{N}(0,1)$, $\forall l \in \{1,\dots,d\}$;

3. Non linear quantile: for $X \sim \mathcal{U}_{[0,1]}$,

$$\begin{aligned} q_p(X) \;=\; & \cos(\tfrac{5}{2}\pi X \exp\{-\tfrac{3}{2}X\}) + \left[\tfrac{1}{4}\exp\{2(X-0.5)\} - \exp\{-1\}(\tfrac{1}{4}+\tfrac{X}{2})\right] \mathbb{1}_{(X<0.5)} \\ & + \left[\tfrac{1}{4}\exp\{-2(X-0.5)\} - (\tfrac{1}{4}+\tfrac{X}{2})\exp\{-1\} - (\tfrac{1}{2}-X)\right] \mathbb{1}_{(X\geq 0.5)} \end{aligned} ;$$

For all designs, we generate independently 300 observations issued from the model defined in Equation (2.1). All runs of Metropolis-Hastings within Gibbs algorithm consist in 20000 iterations, one third of which is burn-in. The prior hyperparameters are chosen as follows: for the inverse Gamma on σ, $a_0 = 1$ and $b_0 = 0.01$, for the Beta prior on p, $\alpha_0 = 2$ and $\lambda_0 = 2$ and $\check{\Sigma}_0$ is set to be the identity matrix for linear quantile and $c = 0.1$, $b = 5$ for the nonparametric case. The choice of $c = 0.1$ and $b = 5$ is a typical one; indeed, whatever the value of τ is, $c = 0.1$ minimizes the empirical root mean integrated square error, RMISE, i.e., $c = argmin \sqrt{\frac{1}{n}\sum_{i=1}^{n}(q_\tau(\mathbf{X}_i) - \hat{q}_\tau(\mathbf{X}_i))^2}$, where $\hat{q}_\tau$ stands for the posterior mean quantile regression. In order to test the robustness of our procedure with respect to the model parameters, different values of σ and p are considered for the three designs.

The first design is a very simple example and is carried out just to check the convergence of the algorithm from different tools: the $\hat{R}$ of Gelman and Rubin diagnostic, the autocorrelation analysis and the posterior plots of the different parameters.

Through the second design, we use the crossing loss criterion (see [57]) to illustrate the noncrossing performance of different approaches: our method,

denoted by "SBQR", the frequentist single quantile method of [33], denoted by "K&B", the Bayesian single quantile regression method of [74], denoted by "Y&M" and the simultaneous Bayesian method of [54], denoted by "R&S"; in addition we use the RMISE to evaluate their estimation performance.

These other methods are performed using available codes in R Core Team (2017): *rq* function available in *quantreg* package [32] for "K&B", *bayesQR* function in bayesQR package [6] for "Y&M" and *qreg* function in *BSquare* package [60] for "R&S".

For design 3, we compare our "SBQR" method with both the nonparametric quantile regression method of [48], denoted by "M&ST", and the simultaneous noncrossing method of [1], denoted by "F&R". We have implemented "M&ST" with *quantregGrowth* R package (see [47]) and "F&R" with an own made code in R. We use features of the RMISE criterion to show how well perform "SBQR" among the other considered methods.

Note that $n = 300$, the number of observations, is chosen carefully in order to illustrate the intended objectives in all designs.

V.1 Design 1: Univariate linear quantile regression

The i.i.d. sample $\epsilon_1, \ldots, \epsilon_n$ is generated according to $\mathcal{ALD}(0, \sigma = 0.1, p = 0.25)$. We propose to infer three quantiles that are close, namely quantiles of order $\tau = 0.2, 0.3$ and 0.4 respectively.

We fix $sd_\sigma = sd_p = 0.1$. To evaluate the convergence of our algorithm, we use three different seeds and starting parameters values to run three different chains and calculate the Gelman's $\hat{R}$ convergence diagnostic. Besides, we use

other convergence diagnostics such as the autocorrelation analysis and the posterior plots.

As shown in the top panel of figures 2.1, 2.2 and 2.3, all posterior distributions shrink at the true parameters value. Furthermore, in the middle panel of figures 2.1, 2.2 and 2.3, the decrease of the empirical autocorrelation of posterior samples proves that the underlying chains are stationary. The bottom panels of Fig 2.1, 2.2 and 2.3 show that the $\hat{R}$ goes to 1 through the iterations, which confirms the convergence of the algorithm.

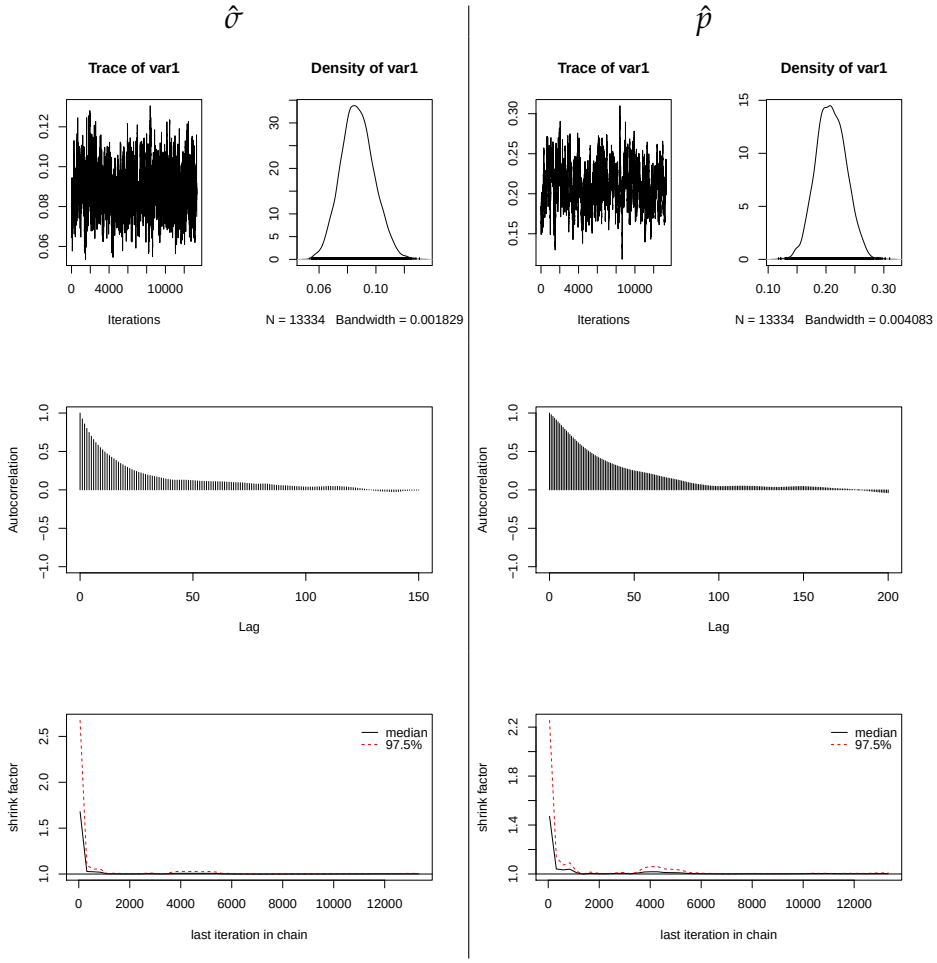

FIGURE 2.1: Trace and density plot (top), autocorrelation plot (middle) and $\hat{R}$ evolution through iterations (bottom) of $\hat{\sigma}$ (left panel) and $\hat{p}$ (right panel).

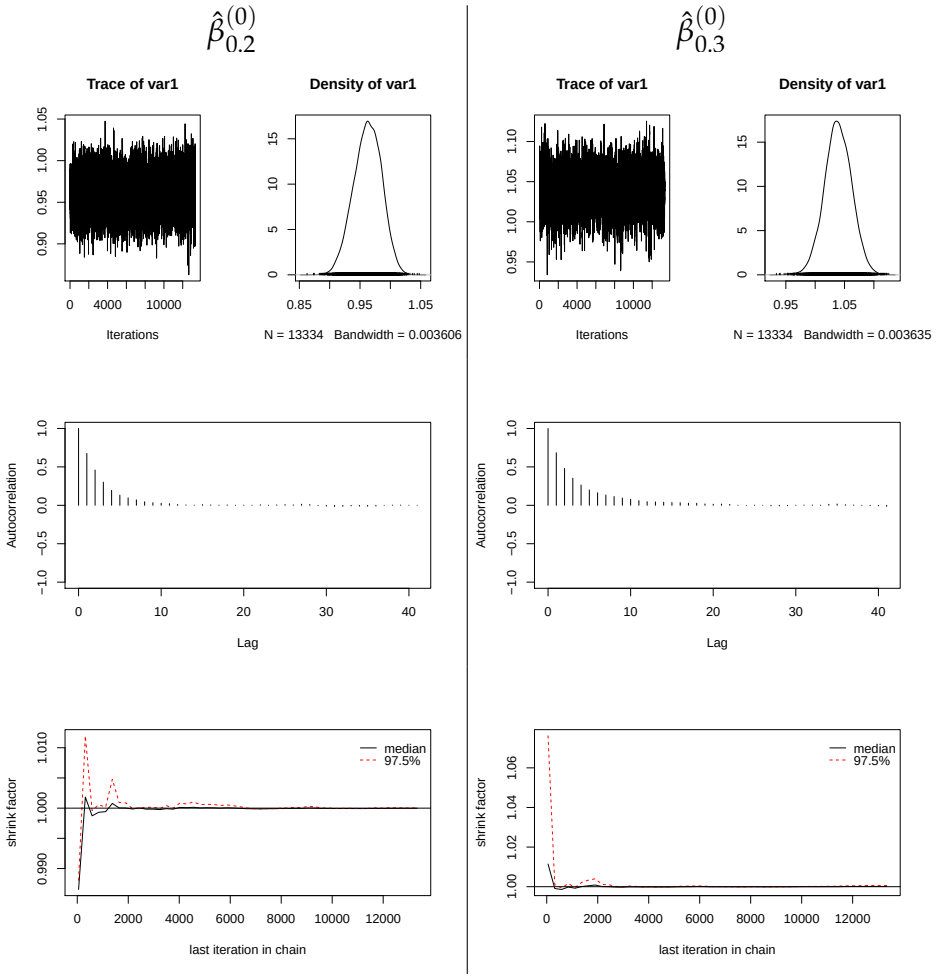

FIGURE 2.2: Trace and density plot (top), autocorrelation plot (middle) and $\hat{R}$ evolution (bottom) of the intercept $\hat{\beta}^{(0)}_{0.2}$ (left panel) and the slope $\hat{\beta}^{(0)}_{0.3}$ of the 0.2−th and 0.3−th quantile respectively.

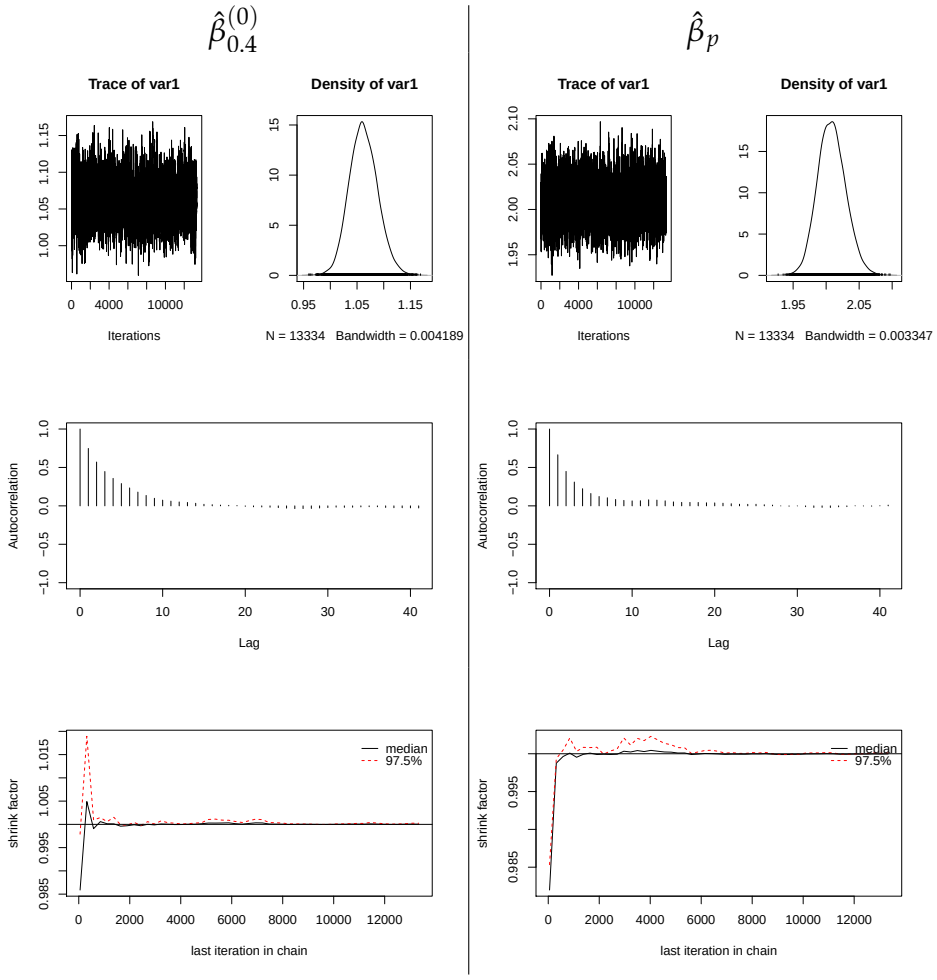

FIGURE 2.3: Trace and density plot (top), autocorrelation plot (middle) and $\hat{R}$ evolution (bottom) of the intercept $\hat{\beta}_{0.4}^{(0)}$ (left panel) of the $0.4-$th quantile and the slope $\hat{\beta}_p$.

V.2 Design 2: Multivariate Linear quantile regression

The second design is dedicated to the multivariate linear case; hence we consider the model given by (2.1) with $\epsilon_1, \ldots, \epsilon_n \overset{i.i.d.}{\sim} \mathcal{ALD}(\mu = 0, \sigma = 0.5, p = 0.25)$, $\mathbf{X} \in \mathbb{R}^d$, $d = 10$, $X_k \sim \mathcal{U}_{[-1,1]}$, $k = 1, \ldots, d$ and $\boldsymbol{\beta_p}$ =(1.6, 2.2, 2.8, 3.4, 4, 4.5, 5.1, 5.7, 6.3, 6.9).

As commonly known, crossing quantiles is a practical problem that often occurs when there is a large number of covariates. We propose to infer the quantiles of order $\tau_1 = 0.1$ and $\tau_2 = 0.12$, since they are close and to study the crossing phenomenon throughout the following four methods: "K&B", "Y&M", "R&S" and "SBQR".

To achieve the desired posterior distribution through MCMC methods, we perform Y&M and R&S with a different number o iteration: 1000 for Y&M and 10000 for R&S. For "R&S", we use the logistic base distribution with 4 basis functions. For "SBQR", we fix $sd_\sigma = sd_p = 0.01$ and $\check{\Sigma}_0$ to be the identity matrix.

To compare the methods, we make use of the crossing loss criterion (see [57]) that measures how far $\hat{q}_{0.12}(\boldsymbol{X})$ goes below $\hat{q}_{0.10}(\boldsymbol{X})$;

$$crossing\ loss = \frac{1}{n}\sum_{i=1}^{n} \max\left(0, \hat{q}_{0.10}(\boldsymbol{X}_i) - \hat{q}_{0.12}(\boldsymbol{X}_i)\right). \tag{2.12}$$

For a given approach, the less the crossing loss is, the better is the method. As shown in Table 2.2, the crossing loss, by "K&B", is significantly of 0.43%, which corresponds to 34 data points of $\hat{q}_{0.1}(\boldsymbol{X})$ that are above $\hat{q}_{0.12}(\boldsymbol{X})$. This percentage is considerably weakened when applying the separate Bayesian

method "Y&M" (0.38%), but still have crossing quantiles (26 data points of $\hat{q}_{0.1}(\boldsymbol{X})$ are above $\hat{q}_{0.12}(\boldsymbol{X})$). However, for simultaneous estimation methods, as our proposed "SBQR" or "R&S" methods, the crossing loss becomes zero; this means that the simultaneous quantile estimation has the potential to make quantile crossing vanish. Moreover, the estimated quantiles are in the right order. While simultaneous approaches control the monotonicity property of quantiles, in a certain sense, separate approaches do not, and they provide more than 8% of violation according to results previously discussed (11.33% of violation by "K&B" and 8.66% of violation by "Y&M").

Method	nb.crossing	crossing loss	RMISE(0.10)	RMISE(0.12)
"K&B"	34	0.0043	0.3565	0.3218
"Y&M"	26	0.0038	0.3344	0.3362
"R&S"	0	0	0.2449	0.2849
"SBQR"	0	0	0.1989	0.2383

TABLE 2.2: Table of criteria: crossing loss and RMISE

To go further, we compute the RMISE for each quantile order and for all methods. Among the latter, the "SBQR" has roughly the smallest RMISE in both quantile levels $\tau = 0.1$ and $\tau = 0.12$. The fact is that "R&S" may over smooth when estimating simultaneously the quantiles and here, for this linear case, the over smoothness is reflected on the flexibility of the method (high RMISE).

To see if the covariable support has an impact on results, we consider another simulation set in which $\mathbf{X} \sim \mathcal{N}_d(\mathbf{0}_d, Id_d)$, i.e., support($\mathbf{X}$)$= \mathbb{R}^d$. Table 2.3 shows that "SBQR" behaves like in the previous example: a zero crossing loss and the smallest RMISE among all the methods.

Method	nb.crossing	crossing loss	RMISE(0.10)	RMISE(0.12)
"K&B"	38	0.0037	0.2356	0.2151
"Y&M"	5	0.0005	0.4569	0.4391
"R&S"	0	0	0.2865	0.3189
"SBQR"	0	0	0.2218	0.2102

TABLE 2.3: Table of criteria: crossing loss and RMISE for nomal covariate case

We also consider another case with a different pair of quantile orders: $\tau_1 = 0.7$ and $\tau_2 = 0.8$. We turn back to the support $[-1, 1]^d$ for $\mathbf{X}$. Table 2.4 shows similar results as the ones obtained for $\tau_1 = 0.10$ and $\tau_2 = 0.12$. Thus, "SBQR" still have the best behaviour among the other methods in term of crossing loss and RMISE.

Method	nb.crossing	crossing loss	RMISE(0.7)	RMISE(0.8)
"K&B"	6	0.0041	0.4255	0.7177
"Y&M"	1	0.0009	0.403	0.7013
"R&S"	0	0	0.224	0.3791
"SBQR"	0	0	0.2065	0.2708

TABLE 2.4: Table of criteria: crossing loss and RMISE for uniform covariate case

It should be noted that the estimation of σ and p by "SBQR" is quite good in terms of the closeness of their estimated values to the true ones in the different treated cases.

V.3 Design 3: Non parametric quantile regression

Considering the third design with $X \sim \mathcal{U}_{[0,1]}$ and $\epsilon \sim \mathcal{ALD}(\mu = 0, \sigma = 0.05, p = 0.75)$, we are interested in estimating quantile functions for orders τ =0.10, 0.12, 0.15 and 0.20. We fix $sd_p = 0.05$, $sd_\sigma = 0.005$ and we compare our "SBQR" method with two others: the "M&ST" method with a three order

cubic B-splines and the "F&R" approach with smoothness parameter value equal to 0.1.

For each value of τ, we evaluate the performance of these methods through the RMISE criterion. Table 2.5 shows that the RMISE values are significantly smaller for "SBQR" than the ones for "M&ST" and "F&R". It worth noting then, that our method is significantly better than the two others in quantiles estimation; in addition, it provides a good estimation of σ ($\hat{\sigma} = 0.0535$) and p ($\hat{p} = 0.7744$).

Method	RMISE(0.10)	RMISE(0.12)	RMISE(0.15)	RMISE(0.20)
"M&ST"	0.1416	0.1397	0.1072	0.0774
"F&R"	0.1520	0.1033	0.0821	0.0664
"SBQR"	0.0692	0.0768	0.0777	0.0660

TABLE 2.5: RMISE at different quantile levels computed for M&ST, F&R and SBQR methods

Figure 2.4 gives the quantile curves estimators. While our "SBQR" method (left panel) provides quantile curves estimates that are close to the true ones (red dashed lines), the $0.1-$th quantile curve estimate given by "F&R" method (middle panel) is fairly distant from the true curve especially when $x \in [0.2, 0.6]$. The same happens for "M&ST" method when $x \in [0.7, 1]$. However, there is no scarred crossing by any of these three methods since they are tackling simultaneous quantile estimation techniques.

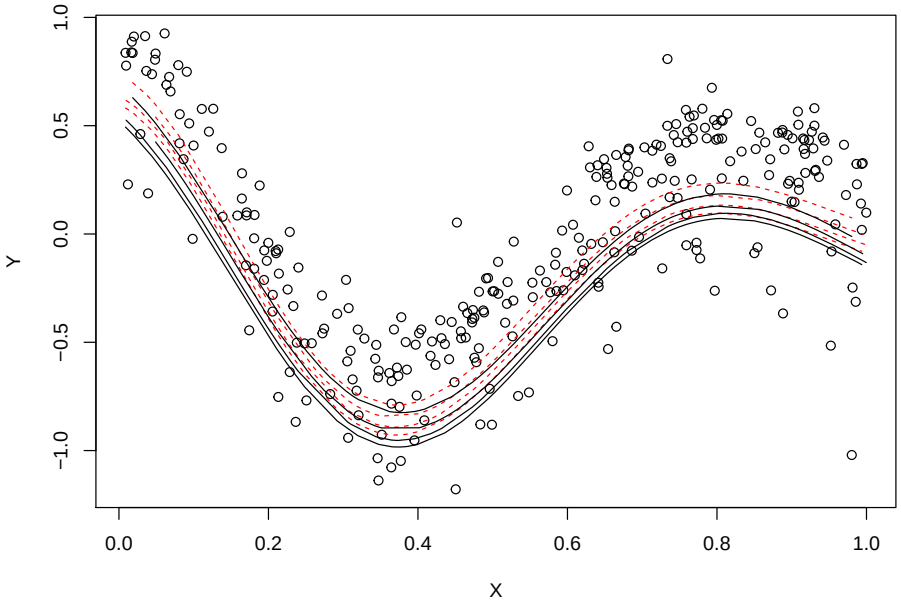

FIGURE 2.4: Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for SBQR method

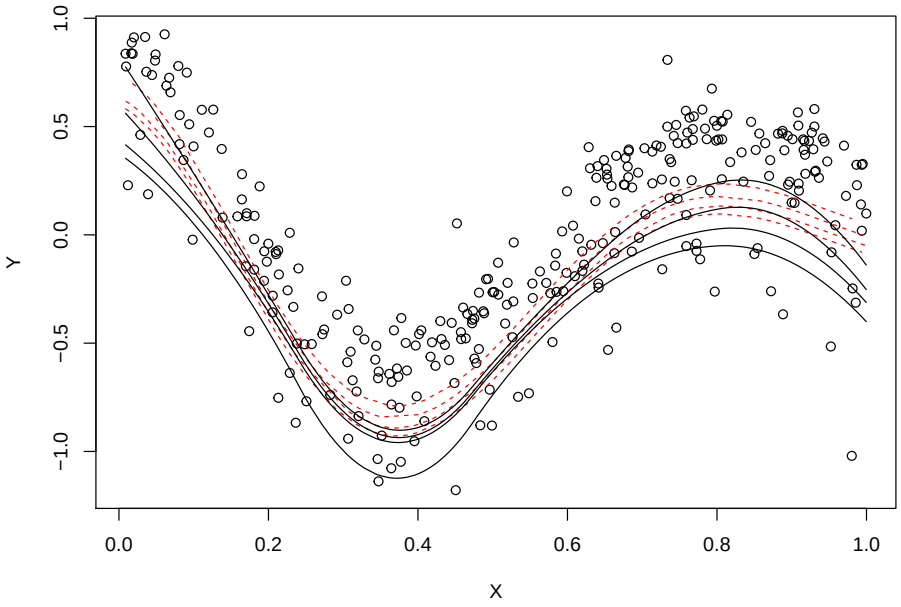

FIGURE 2.5: Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for F&R method

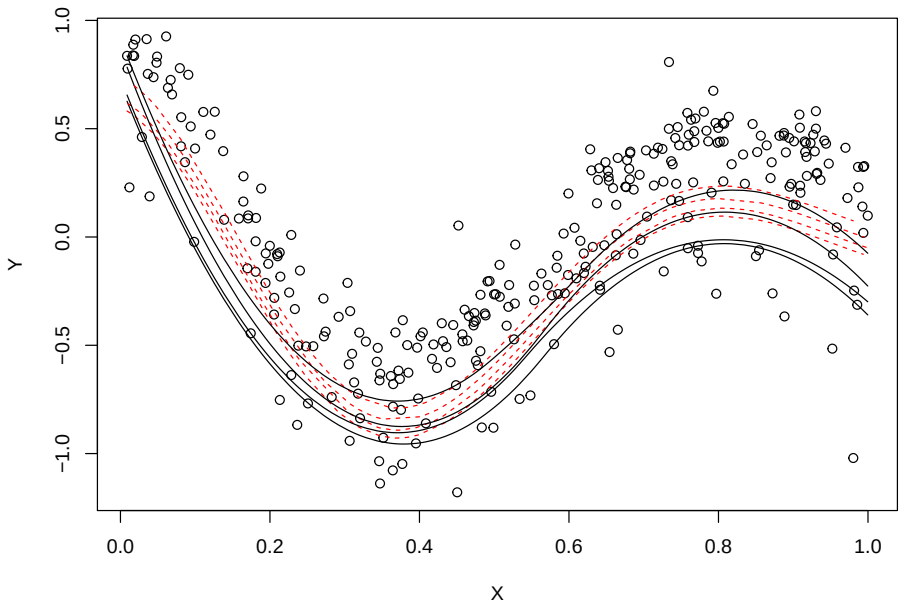

FIGURE 2.6: Estimated quantile curves (black solid lines) against the true ones (red dashed lines) for M&ST method

The issue that makes "F&R"'s method less flexible in simultaneous quantiles fitting is that the second stage final estimators are quite affected by the first stage output that may be badly estimated. For "M&ST", the estimators are constructed iteratively, when solving the minimization problem, by adding constraints so that each subsequent quantile function do not cross with the previous one; this may cause overestimation of quantile curves and thus, it can explain the flexibility of the underlying method.

VI Conclusion

This proposed estimation procedure "SBQR" for simultaneous Bayesian quantile regression guarantees the fundamental property of noncrossing. Asuming that the ALD is the underlying data distribution, this method enables to

characterize the likelihood function by all quantiles of interest using the relation that links two distinct quantiles. A Metropolis-Hastings within Gibbs algorithm is used for the implementation after using the mixture normal representation of the ALD distribution. By studying the linear case for this simultaneous quantile regression, we have faced two versions of modeling: one based on overparameterization, when assuming that the quantiles have distinct slopes, and the other based on more fewer parameters (the parsimony) when we assume that all quantiles have the same slope. We have shown that with the parsimonious model the proposed method provides better results than with the overparameterized model. Our simulation studies show good results that reflect the good performance of the method and guarantee the convergence of the algorithm. Against the crossing problem of estimated quantiles, our method has good performance compared with single quantile estimation methods like [33] and [74] methods. From the RMISE point of view, our method is very competitive in both parametric and nonparametric cases. Moreover, it is the first step for a the only fully Bayesian estimation of multiple quantiles.

Part II

Blind deconvolution

3 Literature overview on blind deconvolution

I Introduction

Signal processing has become an essential science nowadays. All measurements and information processing applications use signal processing techniques to extract the desired information. Thus, signal processing, initially designed to extract the signal during measurements, is widely applied in various and varied applications. A popular one is the restoration of a signal distorted in transmission through a communication channel, called also deconvolution. We can quote, for example, the identification of a telephone channel in digital communication, the removal of the effects of microphones and channels in speech recognition, the deblurring of distorted image, the dereverberation of acoustic recordings, seismic exploration of the ground, etc.... In this field, a received signal Y_t can be considered as the convolution of the transmitted signal X_t with the transmission channel u_t, also called filter. In signal processing, the process of recovery of an input signal convolved with a filter is known as deconvolution. The deconvolution provides then a prototypical example of an inverse problem. When only the output signal Y_t is known, the process is called blind deconvolution and the objective is to identify the transmission channel (the filter) and the input signal from the

output signal recorded at the receiver.

The objective of this chapter, including the blind deconvolution, is to review the main theoretical notions of signal processing that have been useful throughout this work. We expect that this chapter will provide those unfamiliar with these concepts with all the information necessary for a good understanding of this thesis work. We will then present the principles of blind deconvolution as well as some methods from different points of view that are used for this.

II Convolution system

We will first focus on the direct problem, i.e., the convolution.

Most of the signals we perceive are modified not only by the environments they pass through but also by our means of perception. For example, two close but distinct stars, observed through a telescope, can only be identified as two distinct entities if they are distant enough from each other, i.e., if the distance between the two stars is higher than the resolution of the device (telescope). Similarly, in seismology, the response of a seismometer, with a small shock, will never give a very small electrical signal at the output, but an oscillating signal, certainly transient and of considerable duration compared with the input signal.
In these two examples, an original signal (star, hammer stroke) has been modified and transformed by a physical system (telescope, seismometer) which finally gives us an output signal different from the original signal. Figure 3.1 below, describes this phenomenon.

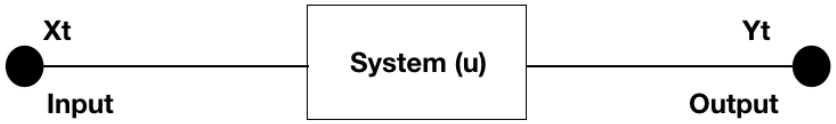

FIGURE 3.1: Convolution schema

In general, a system can be considered as an operator applied to one or more input signals X_t, involving some physical or mathematical transformations, to provide one or more output signals Y_t. In signal processing, this operation is called filtering. The corresponding convolution model to the system described above is given by

$$Y_t = (u * X)_t, \tag{3.1}$$

where X_t is the input signal, u denotes the filter, $*$ denotes the mathematical convolution operator and Y_t is the observed output signal. A system is generally characterized by the number and the type of the input signal X_t, either continuous ($t \in \mathbb{R}$) or discrete ($t \in \mathbb{Z}$), as well as by the type of the filter u (see Section II.2).

II.1 Signal types

Most of the physical signals that surround us are continuous and called analog signals or continuous-time signal. An analog signal is represented as a smooth function x of t where t is defined on $\mathbb{R}^+$, the time space.

However, with the progress of informatics, digital computers and the powerful computational tools provided today, the traditional analytical analysis

of analog signals has given way to the digital analysis of discrete signals. Indeed, as shown by Shanon [59], an analog signal with maximum frequencies $Fmax$ can be represented, without loss of information, by a discrete sequence of analog signal values taken at regular times nT_e, with $n \in \mathbb{N}$, if and only if the sampling frequency $F_e = 1/T_e$ of the digital signal is at least greater than $2Fmax$:

$$X_t \approx X_n \Leftrightarrow F_e \geq 2Fmax.$$

T_e and F_e are respectively the sampling step and frequency. Even though modelling an analog signal by a discrete sequence is an approximation, we can still build the analog signal using its discretized version. We note that this process is guided with a certain error that usually we try to make as small as possible ([66]).

According to Fourier (1807), a signal can be seen as a linear combination of elementary sine waves (signals). A sine signal is entirely defined by three characteristics which are:

- its amplitude A,
- its frequency F or its period $T = 1/F$,
- its phase ϕ,

and it is represented then as

$$X_t = A \sin(\frac{2\pi}{T}t + \phi).$$

The underlying signal can be associated with its following complex representation,

$$Ae^{j(\frac{2\pi}{T}t+\phi)},$$

where $e^{j\theta} = cos(\theta) + jsin(\theta)$ (De Moivre's formula). The Complex exponential notation is the most used as it facilitates mathematical operations on signals.

Finally, the method we have developed (see Chapter 4) deals with complex-valued digital signals so from now on, we focus on digital signal processing.

II.2 Filter types

In a convolutional system, a filter controls the weighting of past (causal part u_t, $t > 0$) and future (anti-causal part u_t, $t < 0$) inputs. In digital signal processing, there are two types of filters depending on the support of the impulse response[1]:

- **Finite impulse response filter (FIR):** it corresponds to an the impulse response that includes $L + 1$ components. In this case the convolution model is given as

$$Y_t = \sum_{k=k_0}^{k_0+L} u_k X_{t-k},\ k_0 \in \mathbb{Z}, L > 0. \tag{3.2}$$

  For a FIR u with $k \in \mathbb{N}$, the model is called "L order moving average" denoted by MA(L). Indeed, for such a system, the output signal is given as the weighted average of the current input signal and the L passed input signals. Thus, this system is causal and it is well adapted when

[1] The impulse response Y_t defines the response of a linear time-invariant system when the input signal is an impulse, i.e., $X_t = \delta(t)$.

modeling signals with limited (finite) duration, for e.g., in seismology, the source function (signal) of seismic motion.

- **Infinite impulse response filter (IIR):** it extends the FIR to an infinite number of components. In this case the convolution Model (3.2) becomes

$$Y_t = \sum_{k \in \mathbb{Z}} u_k X_{t-k}. \tag{3.3}$$

Since an IIR filter has an infinite number of components, it can only be implemented by an approximation that uses finite number of components.

A particular example of IIR filter is the M order autoregressive model, AR(M), defined as

$$Y_t = \sum_{k=1}^{M} \theta_k Y_{t-k} + X_t. \tag{3.4}$$

For this model, the output signal is calculated from the current input signal X_t and M previous output signals weighted by the sequence of the coefficients θ_k. It has also a causal representation as an MA(∞):

$$Y_t = \sum_{k=0}^{+\infty} u_k X_{t-k}.$$

The advantage of such systems is that they are able to model signals with infinite duration from only a finite number of parameters θ_k.

III The deconvolution

Let us move on the deconvolution problem. As we mentioned previously, the deconvolution is the inverse problem associated to the convolution. It

consists on recovering the input signal X_t from the output signal and the filter u (see Figure 3.2 below).

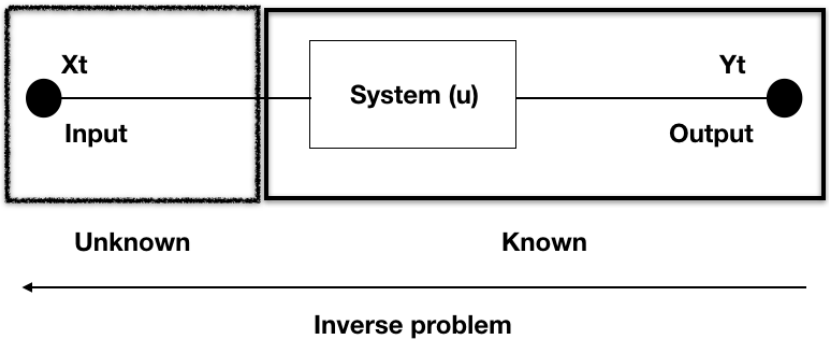

FIGURE 3.2: The deconvolution schema

Hence, inverting the convolution model in (3.1) leads to the deconvolution model, defined by

$$X_t = (Y * \theta)_t,$$

where $\theta = (\theta_t)_{t\in\mathbb{Z}}$ denotes the inverse of the filter $u = (u_t)_{t\in\mathbb{Z}}$, i.e., $(\theta_t)_{t\in\mathbb{Z}}$ satisfies

$$\sum_{t\in\mathbb{Z}} u_t\theta_{k-t} = \delta_k,$$

where δ_k is the Kroneker symbol.

Addressing the deconvolution problem with known inverse filter is called equalization; it is used mainly for the reception systems in telecommunications to correct the transmission channel, by directly identifying the inverse filter without identifying the filter itself. We shall present, in the next, some properties of the inverse filter.

III.1 Inverse filter

To guarantee the invertibility of u, it suffices that $x \mapsto \sum_t u_t e^{itx}$ is continuous in $\mathbb{C}$ and not vanishing on $[0, 2\pi]$ (see [17]). In addition, the latter condition implies that both u and θ are in $l_1(\mathbb{Z})$.

It is proved also in [42], for a moving average model, that any non-zero filter that has no characteristic roots on the unit circle, admits an inverse.

Indeed, taking B to be the backward shift operator, i.e., $B^k X = X_{t-k}$ and denoting by B^- to be the forward shift , i.e., $B^{-k} X_t = X_{t+k}$, we can then rewrite Model (3.2) as

$$Y_t = \phi(B)X,$$

where $\phi(B) = 1 + u_1 B + u_2 B^2 + \ldots + u_L B^L$. We denote by $\beta_i \in \mathbb{C}, i \in 1, \ldots L$, the roots of $\phi(B)$, so that

$$\phi(B) = \prod_{i=1}^{L}(B - \beta_i) = -\prod_{i=1}^{L}(\beta_i - B) = \prod_{i=1}^{L}(-\beta_i)\prod_{i=1}^{L}(1 - \frac{B}{\beta_i}).$$

Inverting $\phi(B)$ requires finding the inverse of $\left(1 - \frac{B}{\beta_i}\right)$. Hence, using Taylor series, one can gets

1. If $|\beta_i| = 1 \rightarrow \frac{1}{|\beta_i|} = 1$ the inverse of $\left(1 - \frac{B}{\beta_i}\right)^{-1}$ do not exist,
2. If $|\beta_i| > 1 \rightarrow \frac{1}{|\beta_i|} < 1 \rightarrow \left(1 - \frac{B}{\beta_i}\right)^{-1} = \sum_{k=0}^{\infty}\left(\frac{1}{\beta_i}\right)^k B^k$; this Taylor serie converges and thus, $\left(1 - \frac{B}{\beta_i}\right)^{-1}$ exists,
3. If $|\beta_i| < 1 \rightarrow \frac{1}{|\beta_i|} > 1 \rightarrow \left(1 - \frac{B}{\beta_i}\right) = \frac{1}{\beta_i}(\beta_i - B) = -\frac{B}{\beta_i}\left(1 - \beta_i B^{-1}\right)$ $\left(1 - \frac{B}{\beta_i}\right)^{-1} = -\frac{\beta_i}{B}\left(1 - \beta_i B^{-1}\right)^{-1} = -\frac{\beta_i}{B}\sum_{k=0}^{\infty}\beta_i^k B^{-k} = -\sum_{k=1}^{\infty}\beta_i^k B^{-k} =$ $-\sum_{k=-\infty}^{-1}\left(\frac{1}{\beta_i}\right)^k B^k$; it is convergent and thus, $\left(1 - \frac{B}{\beta_i}\right)^{-1}$ exists.

However, the convolution of $(Y_t)_{t\in\mathbb{Z}}$ with θ provides an infinite order autoregressive process, which is non-causal [42] only if all the roots of its characteristic polynomial $\phi(B)$ are outside the unit circle. We note also that the autoregressive models are always invertible.

III.2 Ill-posed inverse problem

It is well known that an inverse problem is well-posed if its solution satisfies the following three conditions: existence, singularity and stability. However, this is not the case for the deconvolution since it is non-identifiable and thus it does not have a single solution. This is due to the fact that, through the filtering operation, we lose information on the input signal which makes its exact restoration impossible.

On the other hand, any little perturbation (the presence of noise) on the output signal leads to different deconvolution signals (input signals) with poor quality.

Consequently, the filtering operation is often accompanied by a certain loss of information on the system's input signal. Thus, regularizing the deconvolution is important to extract as much information as possible.

III.3 Blind deconvolution

Suggested by its name, blind deconvolution is a deconvolution problem. It is called blind when neither the filter nor the input signal is known (see Figure

3.3). Thus, blind deconvolution is an ill-posed inverse problem (see [3] and [68]).

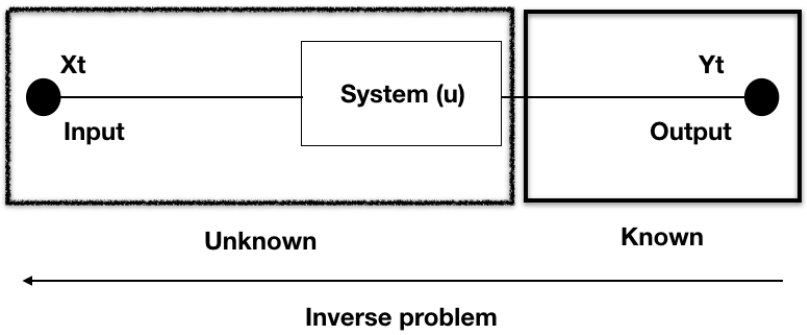

FIGURE 3.3: Blind deconvolution schema

Technically, blind deconvolution consists in estimating two quantities (the input signals and the filter) from a single signal. In this case, it is known that this inversion admits an infinity of solutions, unless we introduce additional information (assumptions).

Thus, depending on the assumptions made about the system or the input signal, we distinguish two different methods of blind deconvolution that are commonly used :

- the spectral deconvolution that uses empirical spectral functions (see [41], [44]),
- the high order moment methods that use the statistical properties of the random input signal (see [31]),

These techniques require that the source signal $(X_t)_{t\in\mathbb{Z}}$ is either a sequence of i.i.d. random variables or a stationary process with finite moments. However, when X is discrete we can not apply these methods to restore it.

Fortunately, there are other methods in the literature that use extra assumptions when dealing with discrete signals. We present some of them in the next section.

IV Blind deconvolution approaches

According to the literature, blind deconvolution is studied either by physics or by probabilistic perspectives. In physics, the reconstruction of the input signal is carried out using statistical tools without estimating the filter. Indeed, these techniques are based on physical and observed measurements, for e.g., the autocorrelation and cross-correlation measurements in phase retrieval[2] (see [67], [29]), in order to recover the underlying signal. In the presence of noise, denoising tools are also used before the signal deblurring restoration step (see [9]).

On the other hand, the probabilistic assessments focus on the estimation of the filter or its inverse in order to recover the distribution of the input signal. In this setting, as we have mentioned in Section III.2, the blind deconvolution model is ill-posed and does not yield a unique solution and thus, there is a need to consider extra assumptions on the model, which is also the case when the model involve additional noise.

The noisy blind deconvolution is defined as follows,

$$Y_t = (u \star X)_t + \sigma_0 W_t = \sum_{k\in\mathbb{Z}} u_k X_{t-k} + \sigma_0 W_t, \ \forall t \in \mathbb{Z} \tag{3.5}$$

[2]The phase retrieval is the problem of recovering one or multiple input signals from the autocorrelation measurements.

where $(X_t)_{t\in\mathbb{Z}}$ is an unobservable input signal, $u = (u_t)_{t\in\mathbb{Z}} \in l_1(\mathbb{Z})$ is an unknown filter and $(\sigma_0\ W_t)_{t\in\mathbb{Z}}$ is a noisy process with $\sigma_0 > 0$ is the noise level.

Because of the illness of Model (3.5), particular attention has restricted the study to the discrete signals with finite support, called alphabet in signal theory (see [49], [3], [10], [17], [18], [16], [24]).

In this setting, the probabilistic research is split into two groups. The first group considers (3.5) as a Bayesian model and focuses on developing Monte Carlo algorithms to calculate the posterior distribution of the input signal. We can refer to [10], [23], [45], [42] and [68] for this setting.

In [10], Chen & Liu addressed the problem of estimation of discrete-valued signals that are blurred by an unknown FIR filter and contaminated by an additive Gaussian white noise with unknown variance. They assumed that the input signals are stationary Markov chains with known state space (alphabet) but unknown initial and transition probabilities. They used Gibbs sampler to calculate the estimates of the unknown quantities.

This has been generalized in [45] by the introduction of a new sequential importing sampling to implement the Bayesian approach. In these two works, the authors advocate fully Bayesian approaches that treat the filter as an additional random quantity that is treated through its posterior distribution.

In contrast, Li and Shedden [42] used the order statistics method to estimate the filter in a separate step before the signal restoration. In [68], Wipf and Zhang used variational Bayesian strategy to find the maximum a posterior (MAP) of the distribution of $X, u|Y$ in the case of Gaussian noise. They used a particular joint prior prior on the input signal X, the filter u and the noise

level σ_0.

The second group uses frequentist approaches based on convolving the inverse filter with the observed signal Y_t in order to restore X_t.They aim to jointly estimate the inverse filter and the input signal; this is usually done by minimizing a specific cost function.

In the non-noisy case, Li [43] proposed to minimize in $s \in l_1(\mathbb{Z})$ the following cost function

$$g(s) = E\left(\prod_{i=1}^{p}(Z(s)_t - a_i)^2\right), \tag{3.6}$$

in s that denotes a possible inverse filter, where $Z(s)_t = (Y * s)_t$ and the a_i denote the p points of the support of X_t. Li dealt with real or complex valued signal, independent but not necessarily identically distributed and assumed that the signal support is finite, known and with known p. Under these assumptions, he proved the consistency of the inverse filter estimate.

This work has been extended further by Gamboa and Gassiat [17] for discrete real signal not necessarily independent and identically distributed. When the alphabet is unknown, they used the discreteness of the signal as prior information to prove the consistency of the inverse filter and the cardinal of the alphabet estimates.

Further, Gassiat and Gautherat [19] generalized Gamboa and Gassiat's method [17] for noisy model, i.e., when the output signal is distorted by an additive noisy $\sigma_0 W_t$ with unknown level σ_0. Therefore, they argued that the estimate of the inverse filter and the noise level $(\hat{\theta}, \hat{\sigma})$ is minimiser of an empirical criterion build using the pseudo moments of $Z(s)$, i.e., $E(\Phi(Z(s)))$ where

$\Phi = (1, \Phi_1, \ldots, \Phi_{2p})$ is a set of real or complex functions;

$$(\hat{\theta}, \hat{\sigma}) = \underset{(\theta,\sigma)\in(\Theta\times\mathbb{R}^+)}{\operatorname{argmin}} J_n(s,\sigma), \quad with \quad J_n(s,\sigma) = H_n(s,\sigma) + \delta(n)^2\sigma, \quad (3.7)$$

where $H_n(s,\sigma)$ is the estimate of the determinant of a Hankel or Toeplitz matrix build using the pseudo-moments of $Z(s)$ and $\delta(n)$ is sequence of positive real numbers such that $\delta(n) \underset{n\to+\infty}{\to} 0$. By taking advantage of the discreteness of X_t and the presence of additive noise, the authors proved that the estimates of u and σ_0 behave asymptotically better than in Li [43]. In [20] the same authors studied the convergence rate of both the inverse filter and the signal estimates in a parametric framework for noisy and non-noisy casess. Later, Gautherat [22] estimated the support points and their corresponding probability, and established their asymptotic distribution. The problem in this method is computational since the algorithm they used requires a starting point that is near to the true value to avoid a local minimizer.

V Our contribution

This part of this thesis is devoted to present the advantage of providing an estimation procedure for the inverse filter, the noise level, the signal support points and their probabilities under some assumptions. In Chapter 4, we propose a joint work with E. Gautherat on a new estimation procedure based on her work in [22] that addresses this problem. In our work, we consider the noisy blind deconvolution model assuming that the distribution of the noise is Gaussian. Unlike the methods mentioned above, our procedure is more exact since it consists on finding a critical zero of an empirical function using

Hankel matrix built from the pseudo-moments as in [19] and [20]. Theoretically, we have proven the consistency and the asymptotic distribution of all the estimates. Moreover, from a practical point of view, we propose a good algorithm that finds appropriate starting points which enables to provides better estimators of the unknown quantities.

4 A new estimation procedure in noisy blind deconvolution model

I Introduction

In this chapter, we provide a new estimation procedure for the inverse filter and the noise level of a noisy blind deconvolution model. This procedure basically enables to restore the unknown discrete distribution of the input signal by finding the roots of an empirical function of the inverse filter and the noise level.

Let us consider the noisy blind deconvolution model given by:

$$Y_t \;=\; (u \star X)_t + \sigma_0 W_t = \sum_{k\in\mathbb{Z}} u_k X_{t-k} + \sigma_0 W_t,\ \forall t \in \mathbb{Z} \tag{4.1}$$

where $(X_t)_{t\in\mathbb{Z}}$ is an unobservable finite discrete complex-valued process, $u = (u_t)_{t\in\mathbb{Z}} \in l_1(\mathbb{Z})$ is an unknown invertible filter and $(\sigma_0\ W_t)_{t\in\mathbb{Z}}$ is a noisy process with unknown level $\sigma_0 > 0$. We denote by $\theta = (\theta_t)_{t\in\mathbb{Z}}$ the inverse of the filter u. Indeed, we suppose that $x \mapsto \sum_t u_t e^{itx}$ is continuous and does not vanish on $[0, 2\pi]$ so that u is invertible as well as both u and θ are in $l_1(\mathbb{Z})$. Based on n observations $(Y_1, \ldots, Y_n)$ issued from Model (4.1), the objective is to recover the distribution of the input process $(X_t)_{t\in\mathbb{Z}}$ that requires the

estimation of both the noise level σ_0 and the filter u.

The Chapter is organized as follows. In Section II, we present the assumptions on the model defined by (4.1) and we give some characterizations on the noise level, the inverse filter and the distribution of the input signal. These characterizations are used in Section III to present our estimation procedure. In Section IV, we present our theoretical results about the strong consistency and the asymptotic distribution of all the estimates. The proofs are given later in Section VII. In Section V we present our simulation study where we empirically demonstrate the performance of our estimation procedure including a comparison of our numerical results with those given in [19]. Finally, we make discussion and give some concluding remarks, in Section VI.

II Assumptions and characterization

All along this chapter, we are going to use the following notations. Random variables without index, like X, W, $\tilde{W}$ ..., are used to refer to any X_t, W_t, $\tilde{W}_t$, ..., with $t \in \mathbb{Z}$. The notation $\|\cdot\|$ denotes either the l_2-norm or the Euclidean norm depending on the natural space to which the element " $\cdot$ " belongs.

II.1 Assumptions

We consider three types of assumptions on the output signal, the noise and the inverse filter.

(H1) X is a discrete complex-valued random variable with finite support $\mathcal{A}$ and distributed according to Π.

b. The cardinality of A, denoted $card(\mathcal{A}) = p \geq 2$, is known.

c. $\mathcal{A} = \{a_1, \ldots, a_p\} \in \mathbb{C}^p$ is unknown. The a_j's of $\mathcal{A}$ are ordered lexicographically, i.e., with notations $\mathcal{R}e$ and $\mathcal{I}m$ for real and imaginary parts of a complex number for any j, k in $\{1, \ldots, p\}$,

$$a_j < a_k \iff \begin{cases} \mathcal{R}e(a_j) < \mathcal{R}e(a_k) \\ \text{or} \\ \mathcal{R}e(a_j) = \mathcal{R}e(a_k), \quad \mathcal{I}m(a_j) < \mathcal{I}m(a_k) \end{cases}.$$

d. For all positive n, $\forall j_1, \ldots, j_n \in \{1, \ldots, p\}$, $\mathbb{P}(X_1 = a_{j_1}, \ldots, X_n = a_{j_n}) > 0$.

e. $\Pi = (\pi_1, \ldots, \pi_p)$ unknown such that $\pi_j = \mathbb{P}(X_0 = a_j)$, $\forall j \in \{1, \ldots, p\}$.

(H2) W is a Gaussian complex-valued random variable.

a. $W = W^R + iW^I$, where W^R and W^I are real-valued random variables with 0-mean and $1/2$-variance.

b. $(W_t^R)_{t\in\mathbb{Z}}$ and $(W_t^I)_{t\in\mathbb{Z}}$ are independent and they are sequences of independent random variables.

c. W^R and W^I are both independent of X.

(H3) θ belongs to a parametrized set Θ.

a. $\theta \in \Theta = \{s(\xi) \in l_1(\mathbb{Z}), \xi \in \mathcal{K}\}$, where $\mathcal{K} \subset \mathbb{R}^d$ is compact and $s : \mathcal{K} \mapsto l_1(\mathbb{Z})$.

b. $s \in \mathcal{C}^1(\mathcal{K})$ is known while ξ is unknown.

c. $s : \mathcal{K} \mapsto l_1(\mathbb{Z})$ is injective and satisfies

$$\forall \xi, \tilde{\xi} \in \mathcal{K} \text{ such that } s_k(\xi) = r s_{k-l}(\tilde{\xi}),\ \forall k \in \mathbb{Z} \implies \begin{cases} r = 1, \\ l = 0 \\ \xi = \tilde{\xi} \end{cases}.$$

Remark 1. In Assumption **(H3)**, the set of inverse filters Θ is expressed through an unknown vector ξ. Since the true inverse filter is assumed to belong to Θ, it means that it exists $\xi_0 \in \mathcal{K}$ such that $\theta = s(\xi_0)$. Hence, estimating the inverse filter θ reduces to estimate the vector-valued parameter ξ_0.

Remark 2. Moreover since s is supposed to be injective from $\mathcal{K}$ in $l_1(\mathbb{Z})$ and $\theta \in \Theta$, it entails that ξ_0 satisfying $s(\xi_0) = \theta$ is in the interior of $\mathcal{K}$. In addition, Assumption **(H3) c.** guarantees the identifiability of the model; in particular, it avoids problems of scale ($r = 1$) and delay ($l = 0$).

II.2 Constrast function

For any s satisfying **(H3)**, let us define $D(s)$ the $(p+1) \times (p+1)$ Hankel matrix whose entries are the conjugate moments of the filtered observations $(s \star Y)_t$.

First, to fix the idea, we consider the non noisy case, i.e., $\sigma_0 = 0$. On the one hand under the process $(s \star Y)_t = (s \star u \star X)_t$ takes at most p support points if and only if $s = \theta$ up to scale and delay. On the other hand when $s = \theta$, the determinant of $D(s)$ is null if and only if X has at most p support points, i.e.,

$$\begin{aligned} \det(D(\theta)) &= \det\left((\mathbb{E}(X_t^k \overline{X_t}^j)_{j,k=0,\dots,p}\right) = 0 \\ &\Updownarrow \\ X_t &\quad \text{has at most } p \text{ support points} \end{aligned} \tag{4.2}$$

where $\overline{X_t}$ denotes the conjugate of X_t. Hence, it is natural to define the estimate of θ as the argument that makes null the determinant of the empirical counterpart of $D(s)$ (see Gamboa and Gassiat [17]).

Back to the noisy case, i.e., $\sigma_0 \neq 0$ is unknown, we aim to consider a criteria that is similar to (4.2) and able to characterize both the noise level σ_0 and the inverse filter θ.

Due to the infinite divisibility of the Gaussian distribution, we note that for any $\sigma > 0$ and any $\tilde{W}$ and $\tilde{\tilde{W}}$ as independent copies of W, we have

$$(s \star Y)_t \overset{distrib}{=} (s \star u \star X)_t + i^{\mathbb{1}_{\sigma>\sigma_0}} |\sigma_0^2 - \sigma^2|^{1/2} (s \star \tilde{W})_t + \sigma (s \star \tilde{\tilde{W}})_t. \tag{4.3}$$

Then, we can define $\tilde{D}(\sigma, s)$ the $(p+1) \times (p+1)$-matrix whose element at the $j-$th row and $k-$th column is set to be

$$\begin{aligned} &\mathbb{E}[((s \star u \star X)_t + i^{\mathbb{1}_{\sigma>\sigma_0}} |\sigma_0^2 - \sigma^2|^{1/2} (s \star \tilde{W})_t)^k \\ &(\overline{(s \star u \star X)_t} + i^{\mathbb{1}_{\sigma>\sigma_0}} |\sigma_0^2 - \sigma^2|^{1/2} \overline{(s \star \tilde{\tilde{W}})_t})^j \ ] \quad . \end{aligned} \tag{4.4}$$

Again due to (4.2), it is obvious to note that

$$\begin{array}{ccc} \det(\tilde{D}(\sigma_0, \theta)) & = & 0 \\ & \Updownarrow & \\ X_t \quad \text{has at most } p \text{ support points} & & \end{array}$$

Hence, the natural criteria to characterize both σ_0 and θ, is the real-valued contrast function J defined by:

$$\forall \sigma \geq 0, \quad \forall s \in \Theta, \quad J(\sigma, s) \quad = \quad \det(\tilde{D}(\sigma, s)). \tag{4.5}$$

Note that due to the definition of $\tilde{D}(\sigma, s)$, the contrast function J is based on the filtered non-observable process X_t that cannot be directly usable in the estimation problem. Note also that (4.5) guarantees that the pair (σ_0, θ) is a zero of J but possibly not unique.

Hence, the next step is to express $\tilde{D}(\sigma, s)$ as a transformation of $D(s)$ (since $D(s)$ depends only on the available process $(s \star Y)_t$) in a way that both the true noise level and the inverse filter are precisely characterised.

To do this, we define $\tilde{d}(\sigma, s)$ the $(p+1)^2$-vector whose $(j(p+1)+(k+1))$-th component is the entry at row j and column k of the matrix $\tilde{D}(s, \sigma)$; in the same way we define the $(p+1)^2$-vector $d(s)$ whose components are obtained from the entries of $D(s)$. Then, from (4.3) and (4.5), simple calculations leads to

$$\tilde{d}(\sigma, s) = A(\sigma \|s\|) d(s), \tag{4.6}$$

where $A(\sigma\|s\|)$ is a $(p+1)^2 \times (p+1)^2$-matrix whose properties are listed in Lemma 1 below. It is worth to mention that the derivation of the closed form of A extends the work of Gassiat and Gautherat [19]. In particular, it leads to get an explicit expression of the contrast function for any positive σ.

Lemma 1. Assuming **(H2)**,

1. $\forall \sigma > 0, \forall s \in \Theta$, the $(p+1)^2 \times (p+1)^2$-matrix $A(\sigma\|s\|)$ is a lower triangular block matrix with $(p+1) \times (p+1)$ sub-matrices, whose element

at row k and column l in the (j,m)-th block is defined by

$$\begin{cases} (-1)^{j-m}\binom{k}{l}\binom{j}{m}\gamma_{j-m,l-k}(\sigma\|s\|)^{k-l+j-m}, & \text{for } j \geq m \text{ and } k \geq l \\ 0 & \text{otherwise,} \end{cases}$$

with $\gamma_{m,l} = \mathbb{E}((W_1)^l(\overline{W_1})^m) = \mathbb{1}_{m=l}m!$ and $m!$ is factorial m.

In particular, $A(\sigma\|s\|)$ has one on its diagonal;

2. $\forall s \in \Theta$, $\sigma \mapsto A(\sigma\|s\|)$ is $\mathcal{C}^\infty$ on $\mathbb{R}_+^*$;

3. $\forall \sigma \in \mathbb{R}_+^*$, $\|s\| \mapsto A(\sigma\|s\|)$ is $\mathcal{C}^\infty$ on $\mathbb{R}_+^*$.

The fact that A has an explicit form for any positive σ, is the key point of the new procedure we propose to estimate the level of noise and the inverse filter. Hence, it is worthwhile to mention that for any $\sigma > 0$, any $s \in \Theta$, $A(\sigma\|s\|)$ is a Hankel matrix whose entries are conjugate moments of the filtered Gaussian $\sigma(s \star \tilde{\tilde{W}})_t$ (see (4.4)). Actually, in previous papers (see for e.g., [20] and [21, 22]), since A was not explicitly given, then no explicit expression for $J(s,\sigma)$ was available; this has led to *argmin*-type estimation procedure based on a penalized contrast function.

Actually, because in previous works there is no knowledge about A for all $\sigma > \sigma_0$, authors had to deal with t Moreover, like this contrast function is very sharp, the minimum of square of its is very difficult to found. With this new characterization, an other way is now possible.

Finally recall the characterization of all true quantities θ, σ_0, a and Π through the function J. They were established in Gassiat and Gautherat [20] for θ and σ_0 and in Gautherat [21, 22] for a and Π in a more general setting.

Properties 1.

1. Under assumptions **(H1) c−e** and **(H2) a, c**, the true level of noise σ_0 and the true inverse filter $\theta = s(\xi_0)$ satisfy

$$\forall \sigma < \sigma_0, \qquad J(\sigma, s(\xi)) > 0, \forall \xi \in \mathcal{K} \tag{4.7}$$

$$J(\sigma_0, s(\xi)) = 0 \quad \text{iff} \quad s(\xi) = \theta \text{ up to scale an delay.} \tag{4.8}$$

2. Under **(H1) c, d**, the a_i's are the complex roots of the equation: $\sum_{j=0}^{p} v_j^\star a^j = 0$, where $v^\star$ denotes the eigenvector associated with the smallest eigenvalue of $\tilde{D}(\sigma_0, \theta)$.

3. Under **(H1) a, c, d**, the distribution $\Pi = (\pi_1, \ldots, \pi_p)$ is uniquely determined as the solution of the following linear system: $\mathbb{E}(X_0^k) = \sum_{j=1}^{p} \pi_j a_j^k, \ \ \forall k = 0, \ldots, p-1.$

Note that assertion 1 in Properties 1 implies that

$$\sigma_0 = \min\{\sigma > 0; \exists s \in l_1(\mathbb{Z}) : \ J(\sigma, s) = 0\}.$$

III Estimation procedures

Recall that we have at hand $n \in \mathbb{N}^*$ observations $Y_1, \ldots, Y_n$ issued from Model (4.1). For all $\xi \in \mathcal{K}$, let us consider $\check{s}(\xi)$ the truncated sequence of $s(\xi)$ defined as $\check{s}_k(\xi) = s_k(\xi)\mathbb{1}_{|k| \le k(n)}, \ \forall k \in \mathbb{Z}$, where $k(n)$ is a sequence of nonnegative integers increasing with n.

For any ξ, denote also by $d_n(\check{s}(\xi))$ the empirical conjugate moment vector of dimension $(p+1)^2$, whose general term is defined as the empirical version

of the $(p+1)^2$-vector $d(\bar{s}(\xi))$, i.e. for $j,k=0,\dots,p$,

$$d_{j(p+1)+k+1,n}(\check{s}(\xi)) = \frac{1}{n-2k(n)} \sum_{t=1+k(n)}^{n-k(n)} ((\check{s}(\xi) \star Y)_t)^k (\overline{(\check{s}(\xi) \star Y)_t})^j.$$

Then, similarly to (4.6), for all $\xi \in \mathcal{K}$ and $\sigma \geq 0$, define the $(p+1)^2$-vector $\tilde{d}_n(\sigma, \check{s}(\xi))$:

$$\tilde{d}_n(\sigma, \check{s}(\xi)) \quad = \quad A(\sigma \|\check{s}(\xi)\|) d_n(\check{s}(\xi)), \tag{4.9}$$

where $A(\cdot)$ is the matrix defined in Lemma 1. Next, define J_n as the empirical version of J (see (4.5)):

$$\forall \xi \in \mathcal{K}, \forall \sigma \geq 0, \quad J_n(\sigma, \check{s}(\xi)) = \det(\tilde{D}_n(\sigma, \check{s}(\xi))),$$

where $\tilde{D}_n(\sigma, \check{s}(\xi))$ is the $((p+1) \times (p+1))$-matrix with element $\tilde{d}_{j(p+1)+k+1,n}(\sigma, \check{s}(\xi))$ at row j and column k.

Mimicking relation (4.8) in Properties 1, the estimators of σ_0 and ξ_0 are defined as a zero of the function J_n whereas those of the a_j's and π_j's are plug-in estimates using assertions 2 and 3 in Properties 1.

In the definition below, we precisely define all of our estimates.

Definition 1.

1. $(\widehat{\sigma}_0, \widehat{\xi}_0)$ is the solution of $\begin{cases} J_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) = 0, \\ \widehat{\sigma}_0 = \min \{\sigma \in \mathbb{R}+; \exists \xi \in \mathcal{K} : \ J_n(\sigma, \check{s}(\xi)) = 0\}. \end{cases}$
2. $\widehat{\theta} = \check{s}(\widehat{\xi}_0)$.
3. The components of $\widehat{a} = (\widehat{a}_1, \dots, \widehat{a}_p) \in \mathbb{C}^p$ are the roots of the equation $\sum_{j=0}^{p} \widehat{v}_j^\star \widehat{a}^j$ in $\mathbb{C}[X]$, where $\widehat{v}^\star = (\widehat{v}_0^\star, \dots, \widehat{v}_p^\star)$ is the eigenvector associated

with the smallest eigenvalue of the matrix $\tilde{D}_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0))$. $(\widehat{a}_1, \ldots, \widehat{a}_p)$ are given in lexicographic order.

4. $\widehat{\Pi} = (\widehat{\pi}_1, \ldots, \widehat{\pi}_p)$ is uniquely determined as the solution in $[0,1]^p$ of $\tilde{d}_{k+1,n}(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) = \sum_{j=1}^{p} \widehat{\pi}_j \widehat{a}_j^k,\ \forall k = 0, \ldots, p-1$.

Roughly speaking, the idea is to look for the parameters that make the input predictors most concentrated.

It should be noted that our estimation procedure for σ and ξ differs from the well-known Z-estimation; indeed, even if $d_n(\check{s}(\xi))$ is the empirical version of the conjugate moment vector $d(s(\xi))$ and $A(\sigma\|s(\xi)\|)$ is an Hankel matrix with elements defined as expectation, the vector $\tilde{d}_n(\sigma, \check{s}(\xi))$ does not correspond to the empirical counterpart of $\tilde{d}(\sigma, s(\xi))$ since it involves $A(\sigma\|\check{s}(\xi)\|)$ which is not the empirical counterpart of $A(\sigma\|s(\xi)\|)$.

We end this part by proving the existence of $(\widehat{\sigma}_0, \widehat{\xi}_0)$ that satisfies the first item in Definition 1 for any positive n. Actually we are able to prove this existence only for some values of p; hence for some p we will give the proof of this existence whereas for other p values, we conjecture this existence. Nevertheless, it should be noted that for a value of p covered by the conjecture, we illustrate with simulations that the function $\sigma \mapsto J_n(\sigma, \check{s}(\xi))$ admits a positive zero even for ξ different from the true one.

Proposition 1. Assume **(H2)**. For any $n \in \mathbb{N}^*$, for any integer $p \geq 2$ such that $\frac{p(p+1)}{2}$ is an odd number, it exists $(\widehat{\sigma}_0, \widehat{\xi}_0)$ solution of the system in Assertion 1 of Definition 1. The proof of this is given is Appendix.

Conjecture 1. For any $n \in \mathbb{N}^*$, for any integer $p \geq 2$ such that $\frac{p(p+1)}{2}$ is an even number, it exists $(\widehat{\sigma}_0, \widehat{\xi}_0\ )$ solution of the system in Assertion 1 of Definition 1.

IV Asymptotic Results

Up to now, for any function F with one or more arguments, set $\partial^r F(y)$ be the value at y of the r-th differential of F and set $\partial^{r_{i_1},\ldots,r_{i_l}}_{i_1,\ldots,i_l} F(y)$ be the value at y of the $(\sum_{k=1}^{l} r_{i_k})$-th partial derivative of F, where r_{i_k} is the order of the derivative with respect to its i_k-th coordinate. For v a column vector, denote by v^T its transpose.

Some other extra assumptions are needed to establish the consistency and the asymptotic distribution of all our estimates.

(H3) d. $k(n) = o(\sqrt{n})$, $\sum_{|k|>k(n)} |s_k(\xi)| = o(\frac{1}{\sqrt{n}}) \forall \xi \in \mathcal{K}$ as $n \to +\infty$

(H3) e. The application $\xi \in \mathcal{K} \mapsto s(\xi)$ is twice continuously differentiable. For any $i = 1, \ldots, d$, $(\partial^1_i s_k(\xi_0))_{k\in\mathbb{Z}}$ and $(\partial^2_i s_k(\xi_0))_{k\in\mathbb{Z}}$ are in $l_1(\mathbb{Z})$. Moreover, $\left((\partial^1_1 s_k(\xi_0))_{k\in\mathbb{Z}}, \ldots, (\partial^1_d s_k(\xi_0))_{k\in\mathbb{Z}}\right)$ and $(s_k(\xi_0))_{k\in\mathbb{Z}}$ are linearly independent.

(H4) $\sqrt{n}\left(d_n(\check{s}(\xi_0)) - d(s(\xi_0)), \partial^1(d_n \circ \check{s})(\xi_0) - \partial^1(d \circ s)(\xi_0)\right) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma)$. Denote by Γ_1 the asymptotic covariance of $\sqrt{n}\left(d_n(\check{s}(\xi_0)) - d(s(\xi_0))\right)$.

The proof of all results are postponed in Appendix.

Theorem 1. Suppose that assumptions **(H1)-(H3) a-d** hold, then as n goes to infinity, $\widehat{\sigma}_0$ converges almost surely (a.s.) to σ_0 and $\left\|\widehat{\xi}_0 - \xi_0\right\|$ converges a.s. to 0.

Corollary 1. Suppose that assumptions **(H1)-(H3) a-d** hold, as n goes to infinity, then both $\|\widehat{a} - a\|$ and $\left\|\widehat{\Pi} - \Pi\right\|$ converge a.s. to 0.

Before giving the asymptotic distribution of our estimates, let us introduce an extra notation. Denote by $v(b)$ the vector in $\mathbb{C}^{p+1} \cap \{\|\cdot\| = 1\}$ associated with $b \in \mathbb{C}^p$ such that $\sum_{j=0}^{p} v_j(b) b^j = 0$; in particular, note that $v(a) = v^*$.

Theorem 2. Under assumptions **(H1)-(H4)**, as n tends to infinity, then $\sqrt{n}\left(\widehat{\xi}_0 - \xi \; ; \; \widehat{\sigma}_0 - \sigma_0\right)^{\mathsf{T}}$ converges in distribution to the $(d+1)$-centered Gaussian distribution with variance-covariance matrix Γ_{σ_0,ξ_0} given by

$$N\partial^1 h(\tilde{d}(\sigma_0,\theta)) A(\sigma_0\|\theta\|)\ \Gamma_1\ \left(N\partial^1 h(\tilde{d}(\sigma_0,\theta)) A(\sigma_0\|\theta\|)\right)^{\mathsf{T}},$$

where

$$\begin{aligned} \alpha &= -\partial_1^1 J(\sigma_0,\theta) \\ N &= \frac{1}{\alpha}\begin{pmatrix} \partial_2^2 J(\sigma_0, s(\xi_0))^{-1}\ \partial_{1,2}^{1,1} J(\sigma_0, s(\xi_0)) \\ 1 \end{pmatrix}, \\ h(d) &= det\left(\left(d_{j(p+1)+i+1}\right)_{i,j=0,\ldots,p}\right). \end{aligned}$$

Corollary 2. Suppose that **(H1)-(H4)** hold, as n tends to infinity, one obtains

- $\sqrt{n}(\hat{a} - a)$ converges in distribution to the centered Gaussian distribution with variance-covariance matrix Γ_a,
- $\sqrt{n}(\hat{\Pi} - \Pi)$ converges in distribution to the centered Gaussian distribution with variance-covariance matrix Γ_Π,

where

$$\begin{aligned}\Gamma_a &= \frac{1}{4|v_p^\star|^4} C^{-1}\, B\, R\, A(\sigma_0\|\theta\|)\, \Gamma_1 A(\sigma_0\|\theta\|)^\mathsf{T} \bar{B}^\mathsf{T}\, \bar{R}^\mathsf{T} C^{-1},\\ \Gamma_\Pi &= G\, R\, A(\sigma_0\|\theta\|)\, \Gamma_1 A(\sigma_0\|\theta\|)^\mathsf{T}\, \bar{R}^\mathsf{T}\, \overline{G}^\mathsf{T},\end{aligned}$$

with

- $C = \mathrm{diag}(K_1, \ldots, K_p)$, with $K_j = \mathbb{E}\,(\prod_{i=1, i\neq j}^{p} |X_0 - a_i|^2)$,
- B is a $p \times (p+1)^2$-matrix whose element at row l and column $i(p+1)+j+1$ is $B_{l,i(p+1)+j+1} = \left(\overline{\partial_l^1 v_i(a)}\, v_j^* + \partial_l^1 v_j(a) \overline{v}_i^*\right), \forall l \in \{1, \ldots, p\}$ and $(i,j) \in \{0, \ldots, p\}^2$,
- $R = \left(\mathrm{Id}_{(p+1)^2} + \partial_{1,2}^{1,1} \tilde{d}(\sigma_0, s(\xi_0)) \partial^1 h(\tilde{d}(\sigma_0, s(\xi_0)))N\right)$ where Id_d is the identity matrix of size d,
- $G = L^{-1}(Proj + F\frac{C^{-1}}{2|v_p^*|^2} B)$,
- $L = \left(a_j^i\right)_{i=0,\ldots,p-1; j=1,\ldots,p}$, where i denotes the rows and j the columns,
- $F = \left(0_p^\mathsf{T} \left(\pi_j\, i\, a_j^{i-1}\right)_{i=2,\ldots,p,\ j=1,\ldots,p} \right)'$, where i denotes the rows, j denotes the columns,
- $Proj$ is the projection of $\mathbb{C}^{(p+1)^2}$ on $\mathbb{C}^p$, i.e., $\forall w \in \mathbb{C}^{(p+1)^2}$, with $Proj(w) = v,\ v = (w_1, \ldots, w_p)'$.

Remark 3. It is worth noting that our estimation procedures allow to derive the asymptotic distribution of the vector $(\widehat{\sigma}_0 - \sigma_0, \widehat{\xi}_0 - \xi_0)$ whereas the penalized estimation methods do not (see [20] and relation (16) in [22] for the asymptotic distribution of $(\widehat{\xi}_0 - \xi_0)$ and $(\widehat{\sigma}_0 - \sigma_0)$, respectively).

V Simulation study

In this section, we illustrate the performance of our estimation procedure over simulated data with $p = 3$. We note that in this case $\frac{p(p+1)}{2} = 6$ is even (see conjecture 1). So, we consider the three classical models: the mixture model, the second order autoregressive model and the second order moving average model.

To avoid scale and delay problems on $\widehat{\theta} = (\widehat{\theta}_k)_{|k|\leq k(n)}$, we fix them by setting $\|\widehat{\theta}\| = 1$ and $\widehat{\theta}_{-k(n)} \geq \widehat{\theta}_t, \forall |t| \leq k(n)$.

- **M1 : Mixture model**

$$Y_t = X_t + \sigma_0 W_t,$$

where the true inverse filter θ has only one component; $\theta_t = s_t(\xi_0) = \mathbb{1}_0(t)$, $\forall t \in \mathbb{Z}$.

- **M2 : Second order Autoregressive model AR(2)**

$$Y_t = \sum_{k=-\infty}^{0} u_k X_{t-k} + \sigma_0 W_t,$$

which leads to the finite true inverse filter such that

$$X_t = \sum_{l\in\mathbb{Z}} \theta_{t-l} \sum_{k=-\infty}^{0} u_k X_{t-k} \Leftrightarrow X_t = \sum_{k=0}^{2} \theta_k \sum_{k=-\infty}^{0} u_k X_{t-k},$$

with $(\theta_0, \theta_1, \theta_2) = (s_1(\xi_0), s_2(\xi_0), s_3(\xi_0)) = (0.8571, -0.2857, 0.4286)$.

- **M3 : Second order Moving Average model MA(2)**

$$Y_t = \sum_{k=0}^{2} u_k X_{t-k} + \sigma_0 W_t,$$

which leads to the following infinite inverse filter $(\theta_0, \theta_1, \theta_2, \ldots)$ =(0.829, 0.497, 0.232, 0.099, 0.041, 0.017, 0.007, 0.003, 0.001, 0,...).

V.1 Parameter settings

We generate data from models **M1, M2** and **M3** with two values of σ_0, 0.05 and 1.To illustrate the asymptotic performance of our method, we deal with different sample size ($n = 50; 100; 500; 1000; 2000$). We choose the truncation parameter, $k(n), = 1, 2$ for **M1** and **M2**, and $k(n) = 4$ for **M3**. This choice of truncation is based on the fact that the mixture model **M1** is over-estimated when $k(n) \geq 1$, since it involves only one filter component. The same occurs with Model **M2** when $k(n) \geq 2$. For the moving average model **M3**, there is always under-estimation since the inverse filter is infinite; therefore, we only present results for $k(n) = 4$, which means that we only consider the first nine components of the inverse filter, i.e., $\theta = (\theta_i)_{i=1,\ldots,9}$, since for $i > 9$ the components are negligible.

Before discussing the obtained simulation results, we first provide a graphical illustration of the roots of the function J_n.

V.2 Illustration of the empirical contrast function roots

Taking the second order autoregressive model **M2** with $\sigma_0 = 0.05$, $k(n) = 1$ and fixed ξ, we show graphically that there exists a root for the function $J_n(\sigma, \check{s}(\xi))$ as conjectured for $p = 3$, since $\frac{p(p+1)}{2}$ is even.

Figure 4.1 presents the plot, for different n, of the function $\sigma \to G_n(\sigma) = \text{sign}(J_n(\sigma, \check{s}(\xi))) \times \log(|J_n(\sigma, \check{s}(\xi))| + 1)$, which has the same roots as the function $\sigma \to J_n(\sigma, \check{s}(\xi))$. For the left panel, ξ corresponds to the true ξ_0,

i.e., $\check{s}(\xi_0) = (0.8571, -0.2857, 0.4286)$ while for the right panel, ξ differs from ξ_0 and is such that $\check{s}(\xi) = (1, 0, 0)$.

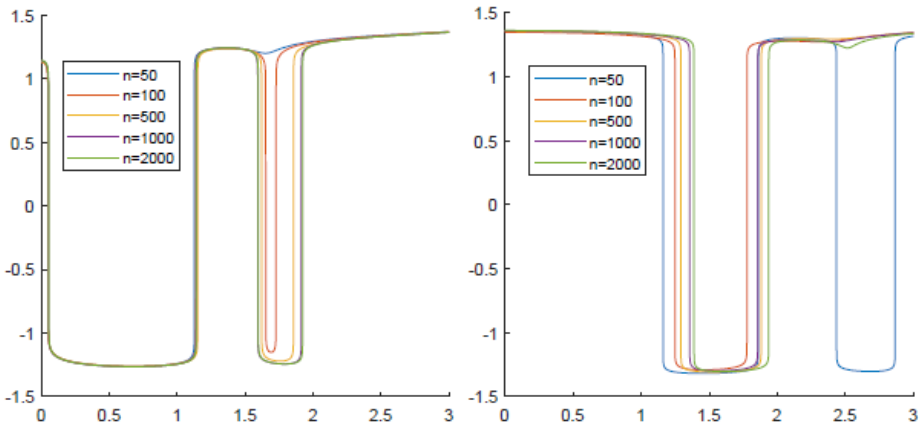

FIGURE 4.1: Graph of G_n with $k(n) = 1$. Second order autoregressive, $\sigma_0 = 0.05$.

The two graphics show the existence of zeros of the function G_n; for the left panel, it is worthwhile to note the convergence is achieved very quickly and accurately which is not the case for the right one since that the chosen ξ is far from the true value.

V.3 Implementation

For each model, we use the Mont Carlo technique to estimate the inverse filer θ, the noise level σ_0, the support points a_i, $i = 1, 2, 3$, and their probabilities π_i. We run $N = 100$ independent simulations of n observations; each run provides an estimate for all parameters. For negative estimate of π_i, $\forall i \in \{1, 2, 3\}$, we discard the simulation. We denote by $Nelim$ the total number of discarded simulations. The final estimate of each parameter is then the average $\widehat{E}$ of the estimated values issued from the $N - Nelim$ runs. We also compute the variability of each estimator using the empirical standard deviation std.

V.3.1 Starting points

Through the algorithm described above, we use the function "fsolve" in MATLAB version 2017b to implement our estimation procedure. This function solves non nonlinear systems given a starting point as input argument and searches for a zero near to the chosen starting point. This choice is then crucial since it has impact on the estimation results. For this reason, we have developed an algorithm that searches a good starting point by only using the data.

Algorithm 2 Initial value Algorithm

1: **Initialization**

- $\sigma_{step} \simeq 10^{-3}$
- $s = 0, \sigma_s = \sigma_{init} = 0$
- $\sigma_{max} = 1000\,\sigma_{step}$
- $\Theta = \{\theta \in \mathbb{R}^{2k(n)+1}, \|\theta\| = 1, \theta(1) > |\theta(i)|, \forall i = 2, \ldots, 2k(n)+1\}$
- $\theta_{init} = \theta_c = \theta_0 \in \Theta$

2: **Loop 1** While $\sigma_s < \sigma_{max}$ and $\theta_{init} = \theta_0$

1. $\sigma_{s+1} = \sigma_s + \sigma_{step}$
2. $t = 0$
3. **Loop 2** While $t < T$ and $\theta_c = \theta_0$
   (a) $t = t + 1$
   (b) $\theta_t \sim U_{[-1/2,1/2]^{2k(n)+1}}$ such that $\theta_t \in \Theta$
   (c) $(\sigma^*, \theta^*, J_n^*) = fsolve(J_n, (\sigma_{s+1}, \theta_t))$
   (d) if $|J_n^*| \leq 10^{-4}$ and $\theta^* \in \Theta$ then $\theta_c = \theta_t$
4. $\theta_{init} = \theta_c$, s=s+1

3: if $\sigma_s = \sigma_{max}$ then $(\sigma_{init}, \theta_{init})$ is a "non convenient initial value"
else $\sigma_{init} = \sigma_s$ and $(\sigma_{init}, \theta_{init})$ is the "chosen initial value"

In this algorithm, starting with a very small fixed σ, we randomly simulate θ on the unit sphere until we get $J_n(\sigma, \theta)$ "close" to zero. Otherwise, we slightly

increase σ and repeat the process; this algorithm is given below and the obtained numerical results are stated in Tables 4.1-4.10.

V.4 The results

- **Mixture model (M1)**

  As shown in Table 4.1, for $\sigma_0 = 0.05$ and $k(n) = 1$, the estimation of (σ, θ) becomes better as long as n increases whereas the support points and their corresponding probabilities have good estimates spacialy for small n. This result is valid also for $k(n) = 2$ as it is shown in Table 4.3. In addition, for both cases, $k(n) = 1$ and $k(n) = 2$, and when n is large, there is no discarded simulations ($Nelim = 0$).

  However, Table 4.2 shows that, when $\sigma = 1$ and $k(n) = 1$, the number of discarded simulations is significant even for large n ($Nelim = 26$). Table 4.4 shows the same result for $k(n) = 2$ since $Nelim = 23$ for $n = 2000$. In addition to that, the estimation of σ, θ, a_i and π_i, $i = 1, \ldots, 3$, is less good than that with small σ, and this is probably due to the significant variability in the estimation (see *std* column in Tables 4.2 and 4.4).

  Subsequently, for any positive $k(n)$, this mixture model is always overestimated. Therefore, it requires a large number of observations to achieve the convergence in the estimation. Therefore, the more truncated, the more we have to increase n in order to provide good estimate.

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.001	0.001	0.001	0.001	0.001
$\theta_{0,init}$	-	0.716	0.788	0.765	0.847	0.967
$\theta_{1,init}$	-	-0.436	-0.435	0.436	-0.444	0.213
$\theta_{2,init}$	-	-0.436	-0.435	0.436	-0.293	-0.139
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	0.05	0.248 ± 0.275	0.136 ± 0.194	0.053 ± 0.002	0.0528 ± 0.003	0.059 ± 0.009
$\hat{\theta}_0$	1	0.924 ± 0.119	0.969 ± 0.084	0.900 ± 0.00003	1 ± 0.00005	1 ± 0.0001
$\hat{\theta}_1$	0	0.113 ± 0.244	0.050 ± 0.170	0.0008 ± 0.005	-0.000006 ± 0.007	-0.002 ± 0.016
$\hat{\theta}_2$	0	0.102 ± 0.224	0.025 ± 0.151	-0.0001 ± 0.007	-0.0008 ± 0.006	0.001 ± 0.007
$\hat{a}_1$	-2 -i	-2.200 -0.997 i ± 0.379	-2.083 -1.012 i ± 0.254	-2 -1 i ± 0.009	-2 -1 i ± 0.010	-2 -1 i ± 0.015
$\hat{a}_2$	-1 +3 i	-1.036 + 2.948 i ± 0.447	-1.020 +2.993 i ± 0.251	-1 + 3.001 i ± 0.010	-0.999 + 2.999 i ± 0.011	-0.999 + 2.999 i ± 0.012
$\hat{a}_3$	4+i	3.431 + 1.130 i ± 1.317	3.821 +1.046 i ± 0.740	4 + 1 i ± 0.012	4 + 1 i ± 0.010	4 + 1 i ± 0.012
$\hat{\pi}_1$	0.6	0.550 ± 0.088	0.584 ± 0.060	0.600 ± 0.021	0.600 ± 0.014	0.603 ± 0.010
$\hat{\pi}_2$	0.25	0.274 ± 0.080	0.258 ± 0.060	0.250 ± 0.017	0.251 ± 0.014	0.247 ± 0.009
$\hat{\pi}_3$	0.15	0.176 ± 0.071	0.158 ± 0.041	0.151 ± 0.015	0.149 ± 0.010	0.149 ± 0.007
Nelim	-	4	1	0	0	0

TABLE 4.1: Estimates of the mixture model for $\sigma_0 = 0.05$, $k(n) = 1$ and different n

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.25	0.25	0.25	0.25	0.25
$\theta_{0,init}$	-	0.830	0.852	0.957	0.948	0.867
$\theta_{1,init}$	-	-0.225	-0.312	-0.172	-0.037	-0.091
$\theta_{2,init}$	-	0.511	0.420	-0.233	0.316	0.490
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	1	0.950 ± 0.115	0.996 ± 0.268	1.008 ± 0.068	1.044 ± 0.169	1.167 ± 0.281
$\hat{\theta}_0$	1	0.848 ± 0.125	0.902 ± 0.112	0.991 ± 0.032	0.988 ± 0.041	0.962 ± 0.062
$\hat{\theta}_1$	0	0.181 ± 0.323	0.058 ± 0.277	-0.009 ± 0.094	-0.009 ± 0.063	-0.046 ± 0.217
$\hat{\theta}_2$	0	0.169 ± 0.320	0.122 ± 0.283	-0.007 ± 0.093	0.0006 ± 0.139	0.015 ± 0.146
$\hat{a}_1$	-2 - i	-2.566 -0.968 i ± 1.251	-2.309 -1.033 i ± 0.595	-2.005 -0.994 i ± 0.172	-1.998 -1 i ± 0.153	-2.043 -1.086 i ± 0.536
$\hat{a}_2$	-1+3 i	-1.237 + 2.416 i ± 1.505	-1.128 + 2.742 i ± 1.036	-1 + 2.948 i ± 0.394	-0.942 + 2.855 i ± 0.554	-0.860 + 2.535 i ± 0.8882
$\hat{a}_3$	4 + i	3.118 + 1.303 i ± 1.660	3.472 + 1.183 i ± 1.233	3.943 + 1.040 i ± 0.262	3.883 + 0.972 i ± 0.671	3.574 + 0.933 i ± 1.076
$\hat{\pi}_1$	0.6	0.481 ± 0.125	0.527 ± 0.105	0.593 ± 0.051	0.592 ± 0.037	0.553 ± 0.118
$\hat{\pi}_2$	0.25	0.331 ± 0.120	0.304 ± 0.097	0.257 ± 0.050	0.252 ± 0.025	0.296 ± 0.120
$\hat{\pi}_3$	0.15	0.188 ± 0.089	0.169 ± 0.053	0.150 ± 0.017	0.156 ± 0.039	0.152 ± 0.033
Nelim	-	18	15	8	19	26

TABLE 4.2: Estimates of the mixture model for $\sigma_0 = 1$, $k(n) = 1$ and different n.

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.001	0.001	0.001	0.001	0.001
$\theta_{0,init}$	-	0.981	0.868	0.872	0.865	0.863
$\theta_{1,init}$	-	-0.112	0.127	0.402	0.304	0.079
$\theta_{2,init}$	-	-0.148	0.078	-0.238	0.132	0.177
$\theta_{3,init}$	-	0.057	-0.123	-0.143	-0.220	-0.320
$\theta_{4,init}$	-	0.022	-0.457	0.040	-0.308	0.338
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	0.05	0.063 ± 0.067	0.061 ± 0.294	0.062 ± 0.094	0.052 ± 0.002	0.052 ± 0.002
$\hat{\theta}_0$	1	0.995 ± 0.044	0.909 ± 0.141	0.995 ± 0.052	1 ± 0.00002	1 ± 0.00002
$\hat{\theta}_1$	0	0.003 ± 0.042	0.070 ± 0.179	0.004 ± 0.045	-0.001 ± 0.005	-0.0004 ± 0.004
$\hat{\theta}_2$	0	0.004 ± 0.045	0.0548 ± 0.161	0.004 ± 0.0476	0.0002 ± 0.004	-0.00007 ± 0.004
$\hat{\theta}_3$	0	0.005 ± 0.030	0.055 ± 0.154	0.006 ± 0.043	0.0001 ± 0.004	0.0001 ± 0.004
$\hat{\theta}_4$	0	0.004 ± 0.052	0.114 ± 0.225	0.004 ± 0.041	-0.0003 ± 0.004	0.0004 ± 0.005
$\hat{a}_1$	-2 - i	-2.016 -1.0012 i $\pm$ 0.178	-2.262 -0.952 i $\pm$ 0.500	-2.014 -0.996 i $\pm$ 0.133	-1.999 -1 i $\pm$ 0.009	-2 -1 i $\pm$ 0.009
$\hat{a}_2$	-1 + 3 i	-1.016 + 2.992 i $\pm$ 0.184	-1.047 + 2.947 i $\pm$ 0.462	-1.007 + 2.989 i $\pm$ 0.128	-0.999 + 3 i $\pm$ 0.009	-1 + 3 i $\pm$ 0.009
$\hat{a}_3$	4 + i	3.940 + 0.996 i $\pm$ 0.604	3.401 + 1.146 i $\pm$ 1.271	3.970 + 1.013 i $\pm$ 0.317	4 + 1 i $\pm$ 0.011	4 + 1 i $\pm$ 0.010
$\hat{\pi}_1$	0.6	0.589 ± 0.072	0.568 ± 0.0699	0.599 ± 0.027	0.601 ± 0.017	0.600 ± 0.012
$\hat{\pi}_2$	0.25	0.251 ± 0.061	0.260 ± 0.071	0.252 ± 0.028	0.250 ± 0.015	0.251 ± 0.009
$\hat{\pi}_3$	0.15	0.160 ± 0.051	0.172 ± 0.074	0.149 ± 0.016	0.149 ± 0.0125	0.150 ± 0.008
Nelim	-	1	2	0	0	0

TABLE 4.3: Estimates of the mixture model for $\sigma_0 = 0.05$, $k(n) = 2$ and different n.

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.25	0.25	0.25	0.25	0.25
$\theta_{0,init}$	-	0.986	0.753	0.898	0.918	0.843
$\theta_{1,init}$	-	-0.045	0.252	0.171	-0.278	-0.021
$\theta_{2,init}$	-	-0.031	0.234	-0.227	0.231	-0.071
$\theta_{3,init}$	-	0.156	0.403	0.333	-0.117	-0.454
$\theta_{4,init}$	-	0.005	-0.389	0.035	-0.119	0.277
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	1	0.919 ± 0.104	1.013 ± 0.096	1.021 ± 0.095	1.054 ± 0.167	1.045 ± 0.131
$\hat{\theta}_0$	1	0.686 ± 0.122	0.765 ± 0.155	0.973 ± 0.070	0.972 ± 0.076	0.983 ± 0.047
$\hat{\theta}_1$	0	0.125 ± 0.346	0.162 ± 0.268	0.022 ± 0.109	-0.006 ± 0.125	0.008 ± 0.074
$\hat{\theta}_2$	0	0.159 ± 0.279	0.174 ± 0.255	0.017 ± 0.118	-0.003 ± 0.102	0.011 ± 0.079
$\hat{\theta}_3$	0	0.137 ± 0.309	0.127 ± 0.291	0.027 ± 0.119	-0.009 ± 0.110	0.012 ± 0.123
$\hat{\theta}_4$	0	0.196 ± 0.357	0.148 ± 0.279	-0.004 ± 0.083	0.0198 ± 0.112	0.006 ± 0.069
$\hat{a}_1$	- 2 - i	-3.270 -0.703 i $\pm$ 1.874	-2.877 -0.726 i $\pm$ 1.822	-2.071 -1.016 i $\pm$ 0.329	-2.042 -1.018 i $\pm$ 0.302	-2.034 -1.008 i $\pm$ 0.242
$\hat{a}_2$	-1 + 3 i	-1.614 + 2.226 i $\pm$ 2.426	-1.609 + 2.279 i $\pm$ 1.649	-1.063 + 2.908 i $\pm$ 0.390	-0.945 + 2.827 i $\pm$ 0.639	-0.960 + 2.877 $\pm$ 0.496
$\hat{a}_3$	4 + i	2.742 + 1.1446 i $\pm$ 2.689	2.919 + 1.574 i $\pm$ 1.747	3.847 + 1.045 i $\pm$ 0.558	3.888 + 1.010 i $\pm$ 0.517	3.882 + 1.012 i $\pm$ 0.490
$\hat{\pi}_1$	0.6	0.383 ± 0.142	0.452 ± 0.124	0.582 ± 0.053	0.574 ± 0.082	0.585 ± 0.075
$\hat{\pi}_2$	0.25	0.408 ± 0.145	0.364 ± 0.124	0.267 ± 0.055	0.272 ± 0.072	0.263 ± 0.060
$\hat{\pi}_3$	0.15	0.210 ± 0.102	0.184 ± 0.072	0.152 ± 0.019	0.154 ± 0.032	0.152 ± 0.028
Nelim	-	5	10	1	16	23

TABLE 4.4: Estimates of the mixture model for $\sigma_0 = 1$, $k(n) = 2$ and different n.

- **Second order Autoregressive model (M2)**

  The estimation of the different parameters of the autoregressive model is quite good when $k(n) = 1$ and for both $\sigma = 0.05$ (see Table 4.5) and $\sigma = 1$ (see Table 4.6). The problem here is when we over-estimate the parameters i.e. when $k(n) = 2$, and when n is small, the estimation is damaged. However, it is improved when n increases. As in the mixture model, discarded simulations are noticed when σ_0 is large ($\sigma_0 = 1$) for both $k(n) = 1$ and $k(n) = 2$.

  We can conclude that our estimation procedure fits well the autoregressive model since convergence is attained easily when $k(n) = 1$ and when $k(n) = 2$ with large n.

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.0010	0.0010	0.0010	0.0010	0.0010
$\theta_{0,init}$	-	0.6914	0.9184	0.7423	0.7558	0.9268
$\theta_{1,init}$	-	-0.2476	0.0526	-0.4589	0.6548	-0.0818
$\theta_{2,init}$	-	-0.6787	-0.3922	-0.4882	-0.0111	0.3665
		$\hat{E} \pm$ std				
$\hat{\sigma}$	0.05	0.0489 ± 0.0048	0.0499 ± 0.0037	0.0511 ± 0.0024	0.0510 ± 0.0013	0.05225 ± 0.0038
$\hat{\theta}_0$	0.8571	0.8570 ± 0.0021	0.8574 ± 0.0017	0.8572 ± 0.0018	0.8572 ± 0.0015	0.8572 ± 0.0015
$\hat{\theta}_1$	-0.2857	-0.2855 ± 0.0029	-0.2852 ± 0.0032	-0.2854 ± 0.0029	-0.2854 ± 0.0024	-0.2865 ± 0.0034
$\hat{\theta}_2$	0.4286	0.4289 ± 0.0034	0.4283 ± 0.0026	0.4286 ± 0.0028	0.4285 ± 0.0023	0.4279 ± 0.0042
$\hat{a}_3$	-2 - i	-2.001 -0.998 i ± 0.0144	-2.001 - i ± 0.011	-2.0004 -0.9997 i ± 0.0067	-2.0003 - i ± 0.0052	-1.999 - i ± 0.0062
$\hat{a}_2$	-1 +3 i	-0.999 + 2.999 i ± 0.0251	-1.002 + 3.001 i ± 0.016	-1.001 + 3 i ± 0.0097	-1 + 3 i ± 0.0077	-0.999 + 2.999 i ± 0.0075
$\hat{a}_1$	4 + i	4.001 + 0.999 i ± 0.035	4 + 1.001 i ± 0.023	4 + 1.001 i ± 0.013	4 + 1 i ± 0.01	4.002 + 1 i ± 0.012
$\hat{\pi}_1$	0.6	0.5981 ± 0.064	0.6060 ± 0.0518	0.6008 ± 0.0232	0.5993 ± 0.0139	0.6016 ± 0.0117
$\hat{\pi}_2$	0.25	0.2502 ± 0.0637	0.2537 ± 0.0441	0.2496 ± 0.0204	0.2493 ± 0.0145	0.2489 ± 0.01
$\hat{\pi}_3$	0.15	0.1517 ± 0.0502	0.1403 ± 0.0361	0.1496 ± 0.0171	0.1514 ± 0.0113	0.1495 ± 0.0082
Nelim	-	0	0	0	0	0

TABLE 4.5: Estimates of autoregressive model for $\sigma_0 = 0.05$, $k(n) = 1$ and different n

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.25	0.25	0.5	0.25	0.25
$\theta_{0,init}$	-	0.889	0.968	0.967	0.772	0.909
$\theta_{1,init}$	-	-0.126	-0.246	-0.230	-0.014	-0.216
$\theta_{2,init}$	-	-0.440	-0.056	-0.111	-0.635	-0.357
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	1	0.923 ± 0.125	0.967 ± 0.077	1.003 ± 0.089	1.024 ± 0.076	1.013 ± 0.0363
$\hat{\theta}_0$	0.8571	0.794 ± 0.069	0.818 ± 0.058	0.838 ± 0.028	0.836 ± 0.043	0.838 ± 0.034
$\hat{\theta}_1$	-0.2857	-0.210 ± 0.295	-0.288 ± 0.145	-0.294 ± 0.089	-0.291 ± 0.114	-0.304 ± 0.0350
$\hat{\theta}_2$	0.4286	0.379 ± 0.302	0.448 ± 0.149	0.439 ± 0.101	0.430 ± 0.129	0.450 ± 0.040
$\hat{a}_3$	-2 - i	-2.475 -0.892 i $\pm$ 2.213	-2.118 -0.824 i $\pm$ 1.254	-2.065 -0.930 i $\pm$ 1.019	-2.083 -1.015 i $\pm$ 0.716	-1.981 -0.982 i $\pm$ 0.098
$\hat{a}_2$	-1 + 3 i	-1.247 + 2.583 i $\pm$ 2.472	-1.067 + 2.590 i $\pm$ 1.246	-1.039 + 2.824 i $\pm$ 0.907	-0.977 + 2.858 i $\pm$ 1.349	-0.967 + 2.923 i $\pm$ 0.198
$\hat{a}_1$	4 + i	4.044 + 1.020 i $\pm$ 1.808	3.721 + 1.029 i $\pm$ 0.887	3.928 + 0.981 i $\pm$ 0.452	3.961 + 1.039 i $\pm$ 0.532	3.904 + 0.1 i $\pm$ 0.197
$\hat{\pi}_1$	0.6	0.535 ± 0.111	0.561 ± 0.080	0.592 ± 0.050	0.594 ± 0.041	0.595 ± 0.015
$\hat{\pi}_2$	0.25	0.296 ± 0.097	0.281 ± 0.067	0.256 ± 0.035	0.254 ± 0.022	0.254 ± 0.014
$\hat{\pi}_3$	0.15	0.170 ± 0.065	0.157 ± 0.038	0.152 ± 0.025	0.152 ± 0.033	0.151 ± 0.008
Nelim	-	2	0	3	0	1

TABLE 4.6: Estimates of autoregressive model for $\sigma_0 = 1$, $k(n) = 1$ and different n.

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.0010	0.0010	0.0010	0.0010	0.0010
$\theta_{0,init}$	-	0.8787	0.8148	0.8100	0.8438	0.9228
$\theta_{1,init}$	-	-0.2112	-0.4699	-0.456	-0.0356	-0.0625
$\theta_{2,init}$	-	-0.1224	0.1721	0.1684	0.4083	-0.2029
$\theta_{3,init}$	-	0.1569	0.2553	0.3273	0.0474	-0.0302
$\theta_{4,init}$	-	-0.379	-0.1434	-0.011	-0.3431	-0.3201
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	0.05	0.1993 ± 0.2339	0.0976 ± 0.1896	0.0811 ± 0.15671	0.07 ± 0.1088	0.0519 ± 0.0031
$\hat{\theta}_0$	0.8571	0.8123 ± 0.071	0.8470 ± 0.0419	0.8528 ± 0.0239	0.8533 ± 0.0258	0.8568 ± 0.0017
$\hat{\theta}_1$	-0.2857	-0.2873 ± 0.0609	-0.2876 ± 0.0197	-0.2897 ± 0.0227	-0.2883 ± 0.0159	-0.286 ± 0.0021
$\hat{\theta}_2$	0.4286	0.4034 ± 0.0866	0.416 ± 0.0774	0.4235 ± 0.0267	0.4256 ± 0.0219	0.429 ± 0.0026
$\hat{\theta}_3$	0	0.1118 ± 0.2001	0.0274 ± 0.1043	0.0132 ± 0.0702	0.007 ± 0.0426	-0.000004 ± 0.0026
$\hat{\theta}_4$	0	-0.0779 ± 0.1441	-0.0209 ± 0.0828	-0.0089 ± 0.0493	-0.0102 ± 0.0659	-0.000005 ± 0.0020
$\hat{a}_3$	-1 - i	-2.355 -1.187 i $\pm$ 0.722	-2.0686 -1.041 i $\pm$ 0.430	-2.019 -1.012 i $\pm$ 0.121	-2.023 -1.020 i $\pm$ 0.23	-2 -1 i $\pm$ 0.006
$\hat{a}_2$	-1 + 3 i	-1.077 + 3.328 i $\pm$ 1.254	-0.960 + 3 i $\pm$ 0.653	-1.002 + 3.012 i $\pm$ 0.095	-1.004 + 3.054 i $\pm$ 0.512	-1 + 2.999 i $\pm$ 0.008
$\hat{a}_3$	4 + i	4.155 + 1.215 i $\pm$ 0.861	4.231 + 0.808 i $\pm$ 3.12	4.012 + 1.02 i $\pm$ 0.145	4.023 + 1.012 i $\pm$ 0.177	3.998 + 1 i $\pm$ 0.012
$\hat{\pi}_1$	0.6	0.552 ± 0.092	0.591 ± 0.054	0.595 ± 0.027	0.6 ± 0.023	0.602 ± 0.012
$\hat{\pi}_2$	0.25	0.284 ± 0.077	0.259 ± 0.054	0.254 ± 0.025	0.249 ± 0.016	0.249 ± 0.009
$\hat{\pi}_3$	0.15	0.164 ± 0.060	0.150 ± 0.038	0.151 ± 0.015	0.151 ± 0.023	0.150 ± 0.007
Nelim	-	1	0	0	0	0

TABLE 4.7: Estimates of autoregressive model for $\sigma_0 = 0.05$, $k(n) = 2$ and different n.

	true values	$n = 50$	$n = 100$	$n = 500$ $n = 1000$	$n = 2000$	
σ_{init}	-	0.25	0.25	0.5	0.5	0.5
$\theta_{0,init}$	-	0.843	0.664	0.950	0.918	0.943
$\theta_{1,init}$	-	-0.195	0.400	0.202	0.124	0.109
$\theta_{2,init}$	-	0.374	-0.488	0.205	0.346	0.092
$\theta_{3,init}$	-	0.103	-0.266	-0.110	0.024	0.253
$\theta_{4,init}$	-	-0.318	0.299	-0.060	-0.150	0.162
		$\hat{E} \pm$ std				
$\hat{\sigma}$	1	0.905 ± 0.113	1.030 ± 0.111	1.196 ± 0.103	1.258 ± 0.127	1.258 ± 0.102
$\hat{\theta}_0$	0.8571	0.682 ± 0.080	0.697 ± 0.079	0.708 ± 0.060	0.686 ± 0.062	0.696 ± 0.055
$\hat{\theta}_1$	-0.2857	-0.317 ± 0.175	-0.349 ± 0.152	-0.368 ± 0.075	-0.414 ± 0.113	-0.366 ± 0.069
$\hat{\theta}_2$	0.4286	0.378 ± 0.109	0.350 ± 0.095	0.342 ± 0.110	0.343 ± 0.136	0.353 ± 0.121
$\hat{\theta}_3$	0	0.304 ± 0.188	0.285 ± 0.187	0.267 ± 0.180	0.185 ± 0.208	0.235 ± 0.236
$\hat{\theta}_4$	0	-0.272 ± 0.206	-0.303 ± 0.155	-0.301 ± 0.175	-0.328 ± 0.142	-0.278 ± 0.216
$\hat{a}_3$	-2 - i	-3.699 -0.811 i $\pm$ 3	-3.210 -1.273 i $\pm$ 2.055	-2.810 -1.160 i $\pm$ 2.928	-2.718 -1.517 i $\pm$ 1.087	-2.713 -1.385 i $\pm$ 0.980
$\hat{a}_2$	-1 + 3 i	-1.697 + 2.766 i $\pm$ 4.127	-1.464 + 3.659 i $\pm$ 2.926	-1.151 + 3.998 i $\pm$ 2.462	-1.060 + 3.778 i $\pm$ 1.850	-1.143 + 3.855 i $\pm$ 1.865
$\hat{a}_3$	4 + i	4.404 + 1.532 i $\pm$ 2.567	4.998 + 1.496 i $\pm$ 2.452	4.717 + 1.455 i $\pm$ 0.943	4.585 + 1.381 i $\pm$ 0.812	4.568 + 1.441 i $\pm$ 0.916
$\hat{\pi}_1$	0.6	0.409 ± 0.114	0.444 ± 0.093	0.488 ± 0.087	0.467 ± 0.060	0.499 ± 0.069
$\hat{\pi}_2$	0.25	0.361 ± 0.099	0.349 ± 0.087	0.318 ± 0.073	0.336 ± 0.063	0.316 ± 0.053
$\hat{\pi}_3$	0.15	0.229 ± 0.082	0.207 ± 0.075	0.194 ± 0.068	0.198 ± 0.055	0.186 ± 0.050
Nelim	-	7	2	3	3	5

TABLE 4.8: Estimates of autoregressive model for $k(n) = 2$ for different n

- **Moving average (M3)**

  Tables 4.9 and 4.10 below present the results of the estimation of the moving average model **M3** where the inverse filter is namely infinite, while only the first nine components are estimated ($k(n) = 4$) since, as of the tenth one, the components are significantly meaningless. It is very clear that for both small and large positive σ_0, the estimators of θ and σ_0 are very close to their true values. However, the estimates of the support points are far from the true ones but, their associated probabilities are close to their true values. In addition to that, the number of eliminated simulations, in both $\sigma_0 = 0.05$ and $\sigma_0 = 1$ cases, is not significant.

  This result means that the restoration of the input signal was not successful even though the inverse filter and noise level were well estimated. This can be explained by the truncation effect of an infinite

filter. Such a model was handled by Li and Shedden [42] that used the Bayesian approach where the posterior distribution of the signal is restored as well as estimating the unknown filter coefficient.

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	0.0001	0.0001	0.0001	0.0001	0.0001
$\theta_{0,init}$	-	0.6	0.6	0.6	0.6	0.6
$\theta_{1,init}$	-	0.4	0.4	0.4	0.4	0.4
$\theta_{2,init}$	-	0.2	0.2	0.2	0.2	0.2
$\theta_{3,init}$	-	0.1	0.1	0.1	0.1	0.1
$\theta_{4,init}$	-	0.05	0.05	0.05	0.05	0.05
$\theta_{5,init}$	-	0	0	0	0	0
$\theta_{6,init}$	-	0	0	0	0	0
$\theta_{7,init}$	-	0	0	0	0	0
$\theta_{8,init}$	-	0	0	0	0	0
		$\widehat{E} \pm$ std				
$\hat{\sigma}$	0.05	0.0452 ± 0.0069	0.0490 ± 0.0064	0.0509 ± 0.0028	0.0528 ± 0.0057	0.0525 ± 0.0040
$\hat{\theta}_0$	0.8288	0.8284 ± 0.0019	0.8285 ± 0.0013	0.8286 ± 0.0009	0.8287 ± 0.0014	0.8286 ± 0.0010
$\hat{\theta}_1$	0.4973	0.4973 ± 0.0019	0.4975 ± 0.0016	0.4974 ± 0.0013	0.4972 ± 0.0018	0.4975 ± 0.0015
$\hat{\theta}_2$	0.2321	0.2328 ± 0.0032	0.2325 ± 0.0018	0.2323 ± 0.0014	0.2324 ± 0.0022	0.2320 ± 0.0021
$\hat{\theta}_3$	0.0995	0.1002 ± 0.0036	0.0998 ± 0.0030	0.0996 ± 0.0013	0.0995 ± 0.0018	0.0996 ± 0.0015
$\hat{\theta}_4$	0.0411	0.0417 ± 0.0040	0.0411 ± 0.0026	0.0411 ± 0.0017	0.0412 ± 0.0023	0.0412 ± 0.0018
$\hat{\theta}_5$	0.0167	0.0169 ± 0.0042	0.0168 ± 0.0028	0.0167 ± 0.0015	0.0166 ± 0.0025	0.0172 ± 0.0023
$\hat{\theta}_6$	0.0067	0.0073 ± 0.0035	0.0069 ± 0.0025	0.0068 ± 0.0014	0.0069 ± 0.0022	0.0070 ± 0.0015
$\hat{\theta}_7$	0.0027	0.0035 ± 0.0030	0.0027 ± 0.0026	0.0026 ± 0.0014	0.0025 ± 0.0021	0.0024 ± 0.0017
$\hat{\theta}_8$	0.0011	0.0015 ± 0.0030	0.0008 ± 0.0019	0.0008 ± 0.0013	0.0010 ± 0.0029	0.0006 ± 0.0023
$\hat{a}_3$	-1 + 3 i	-3.3189 + 1.6523i ± 0.0316	-3.3151 + 1.6535i ± 0.0208	-3.3136 + 1.6576i ± 0.0095	-3.3141 + 1.6579i ± 0.0080	-3.3146 + 1.6573i± 0.0056
$\hat{a}_2$	-2 - i	-0.8320 - 2.4841i ± 0.0237	-0.8300 - 2.4862i ± 0.0179	-0.8286 - 2.4862i ± 0.0064	-0.8280 - 2.4861i ± 0.0056	-0.8283 - 2.4858i ± 0.0042
$\hat{a}_1$	4 + i	2.4817 + 4.1386i ± 0.0403	2.4818 + 4.1441i ± 0.0261	2.4868 + 4.1436i ± 0.0125	2.4874 + 4.1436i ± 0.0139	2.4861 + 4.1429i ± 0.0094
$\hat{\pi}_1$	0.25	0.2605 ± 0.0746	0.2497 ± 0.0410	0.2482 ± 0.0184	0.2513 ± 0.0142	0.2500 ± 0.0096
$\hat{\pi}_2$	0.60	0.5950 ± 0.0840	0.6001 ± 0.0467	0.5990 ± 0.0193	0.6010 ± 0.0166	0.6002 ± 0.0109
$\hat{\pi}_3$	0.15	0.1445 ± 0.0631	0.1502 ± 0.0384	0.1528 ± 0.0176	0.1478 ± 0.0112	0.1498 ± 0.0071
Nelim	-	0	0	0	0	0

TABLE 4.9: Estimates of moving average model for $k(n) = 4$ and $\sigma_0 = 0.05$

	true values	$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
σ_{init}	-	1	1	1	1	1
$\theta_{0,init}$	-	0.6	0.6	0.6	0.6	0.6
$\theta_{1,init}$	-	0.4	0.4	0.4	0.4	0.4
$\theta_{2,init}$	-	0.2	0.2	0.2	0.2	0.2
$\theta_{3,init}$	-	0.1	0.1	0.1	0.1	0.1
$\theta_{4,init}$	-	0.1	0.1	0.1	0.1	0.1
$\theta_{5,init}$	-	0.01	0.01	0.01	0.01	0.01
$\theta_{6,init}$	-	0.001	0.001	0.001	0.001	0.001
$\theta_{7,init}$	-	0.0001	0.0001	0.0001	0.0001	0.0001
$\theta_{8,init}$	-	0	0	0	0	0
		$\hat{E} \pm$ std				
$\hat{\sigma}$	1	1.0229 ± 0.0788	1.0388 ± 0.0954	1.0404 ± 0.0770	1.0362 ± 0.0410	1.0343 ± 0.0304
$\hat{\theta}_0$	0.8288	0.6988 ± 0.0939	0.7275 ± 0.0851	0.7772 ± 0.0393	0.7839 ± 0.0165	0.7850 ± 0.0102
$\hat{\theta}_1$	0.4973	0.4858 ± 0.0920	0.5048 ± 0.0624	0.5228 ± 0.0182	0.5231 ± 0.0091	0.5232 ± 0.0128
$\hat{\theta}_2$	0.2321	0.2739 ± 0.0780	0.2737 ± 0.0687	0.2688 ± 0.0243	0.2727 ± 0.0167	0.2725 ± 0.0163
$\hat{\theta}_3$	0.0995	0.1512 ± 0.0884	0.1426 ± 0.0564	0.1428 ± 0.0313	0.1529 ± 0.0175	0.1553 ± 0.0156
$\hat{\theta}_4$	0.0411	0.1063 ± 0.0803	0.1105 ± 0.0753	0.0959 ± 0.0566	0.0850 ± 0.0349	0.0802 ± 0.0370
$\hat{\theta}_5$	0.0167	0.0212 ± 0.0760	0.0179 ± 0.0690	0.0332 ± 0.0434	0.0432 ± 0.0232	0.0470 ± 0.0242
$\hat{\theta}_6$	0.0067	0.0208 ± 0.0788	0.0029 ± 0.0436	0.0020 ± 0.0279	-0.0003 ± 0.0163	-0.0007 ± 0.0129
$\hat{\theta}_7$	0.0027	0.0460 ± 0.1301	0.0302 ± 0.1082	0.0025 ± 0.0267	-0.0003 ± 0.0097	-0.0008 ± 0.0067
$\hat{\theta}_8$	0.0011	0.2021 ± 0.2395	0.1355 ± 0.2159	0.0202 ± 0.0971	0.0042 ± 0.0443	0.0006 ± 0.0063
$\hat{a}_3$	-1 + 3 i	-3.7502 + 0.9237i ± 2.0951	-3.6762 + 1.3120i ± 1.3953	-3.3293 + 1.5460i ± 0.4833	-3.2560 + 1.5074i ± 0.1609	-3.2505 + 1.5055i ± 0.0902
$\hat{a}_2$	-2 - i	-1.2785 - 2.8357i ± 1.9693	-1.1818 - 2.9028i ± 1.1911	-0.9631 - 2.5007i ± 0.3983	-0.9225 - 2.4467i ± 0.1465	-0.9269 - 2.4343i ± 0.0700
$\hat{a}_1$	4 + i	1.8358 + 3.8790i ± 2.4040	2.0122 + 3.9898i ± 1.5473	2.2033 + 3.9701i ± 0.4911	2.2277 + 3.8766i ± 0.1788	2.2076 + 3.8796i ± 0.1235
$\hat{\pi}_1$	0.25	0.3172 ± 0.1134	0.2997 ± 0.0842	0.2566 ± 0.0258	0.2540 ± 0.0149	0.2527 ± 0.0085
$\hat{\pi}_2$	0.60	0.5093 ± 0.1306	0.5352 ± 0.1086	0.5928 ± 0.0282	0.5952 ± 0.0163	0.5975 ± 0.0100
$\hat{\pi}_3$	0.15	0.1734 ± 0.0758	0.1650 ± 0.0512	0.1505 ± 0.0127	0.1508 ± 0.0084	0.1498 ± 0.0059
Nelim	-	0	1	0	0	0

TABLE 4.10: Estimates of moving average model for $k(n) = 4$ and $\sigma_0 = 1$

V.5 Computational comparison

In this section, we compare our estimation procedure, here named $\mathcal{M}_2$, with the estimation method proposed by Gassiat and Gautherat [19], named $\mathcal{M}_1$. To do so, we consider a second order autoregressive model with a real-valued signal, $p = 2$ and true inverse filter $\theta = (0.8571, -0.2857, 0.4286)$; so we adapt our method to the real case. We use Algorithm 2 to provide suitable starting points for implementing our method.

To evaluate the performance of both methods in signal restoration, we shall use the signal-to-noise ratio SNR[1], whose higher values generally mean that

[1]The signal-to-noise ratio is an indicator of the quality of the information transmission. It is the power ratio between signal and noise and it is generally expressed in decibels dB.

there is more useful information (the signal) than unwanted data (the noise).

Table 4.11 below presents the results of SNR and parameter estimation by both methods $\mathcal{M}_1$ and $\mathcal{M}_2$ for $\sigma_0 = 0.1$ and $\sigma_0 = 1$. According to this, when the noise level is small ($\sigma_0 = 0.1$) the SNR is high and is about 46. On the opposite, for a large noise level ($\sigma_0 = 0.1$) the SNR is small (SNR= 0.46), which means that there is more noise than signal and thus detecting the source signal becomes much more difficult. This is reflected in the unsatisfactory results by both $\mathcal{M}_1$ and $\mathcal{M}_2$. However, regarding the parameter estimation, our method $\mathcal{M}_2$ works better than method $\mathcal{M}_1$ when n increases in the case of $\sigma_0 = 1$. And, for $\sigma_0 = 0.1$, $\mathcal{M}_2$ is always better than the method $\mathcal{M}_1$ and performs very well in terms of estimated values and standard deviation, even for a small n.

σ_0	SNR	n	Method	$\widehat{E}(\widehat{\theta}_0) \pm std(\widehat{\theta}_0)$	$\widehat{E}(\widehat{\theta}_1) \pm std(\widehat{\theta}_1)$	$\widehat{E}(\widehat{\sigma}_0) \pm std(\widehat{\sigma}_0)$
0.1	46	100	$\mathcal{M}_1$	0.7500 ± 0.0647	-0.1767 ± 0.0668	0.0000 ± 0.0000
0.1	46	100	$\mathcal{M}_2$	0.8553 ± 0.0095	-0.2879 ± 0.0205	0.1004 ± 0.0100
0.1	46	500	$\mathcal{M}_1$	0.7591 ± 0.0140	-0.1860 ± 0.0154	0.0000 ± 0.0000
0.1	46	500	$\mathcal{M}_2$	0.8580 ± 0.0049	-0.2941 ± 0.0187	0.1035 ± 0.0031
0.1	46	1000	$\mathcal{M}_1$	0.7639 ± 0.0128	-0.1923 ± 0.0142	0.0000 ± 0.0000
0.1	46	1000	$\mathcal{M}_2$	0.8573 ± 0.0080	-0.2837 ± 0.0151	0.1009 ± 0.0139
1	0.46	500	$\mathcal{M}_1$	0.5913 ± 0.2537	-0.0916 ± 0.2652	0.9407 ± 0.1015
1	0.46	500	$\mathcal{M}_2$	0.1423 ± 0.6017	-0.4147 ± 0.2475	0.8607 ± 0.5653
1	0.46	5000	$\mathcal{M}_1$	0.7529 ± 0.1211	-0.2450 ± 0.1353	1.0022 ± 0.0612
1	0.46	5000	$\mathcal{M}_2$	0.5491 ± 0.5060	-0.3259 ± 0.2020	0.9646 ± 0.3915
1	0.46	15000	$\mathcal{M}_1$	0.7771 ± 0.1294	-0.2576 ± 0.1206	1.0302 ± 0.0651
1	0.46	15000	$\mathcal{M}_2$	0.7159 ± 0.3757	-0.2928 ± 0.1476	0.9844 ± 0.0387

TABLE 4.11: Results of SNR and the parameter estimation for AR(2) with $\sigma_0 = 0.1, 1$ by both $\mathcal{M}_1$ and $\mathcal{M}_2$.

We should note that the estimation method in Gassiat and Gautherat [19] needs to calibrate some parameters since it is based on the minimization of a penalized contrast function while our method does not. Moreover, $\mathcal{M}_1$ suffers from the problem of chosen the starting point near the true value and thus could provide a local minimum, which is not the case in our method, since Algorithm 2 allows to randomly and adaptively select a "good" starting point.

On the other hand, a noticeable drawback of our method is that it can provide a significant variability, i.e., large standard deviation.

VI Discussion and conclusion

- **Choice of the Hankel matrix**

  In our method, it is possible to deal with the Toeplitz matrix instead of the Hankel matrix of the $(s \star Y)_t$ since the characterizations of θ and σ_0 given by relations (4.8) and (4.7) also hold for the Toeplitz matrix. More generally for the same reason, it would be possible to consider any $(p+1)^2$-vector built on the moment of type $(\mathbb{E}((\phi_1((s(\xi) \star Y)_0))^k \overline{\phi_1((s(\xi) \star Y)_0)^j}])_{k,j=1,\dots,p}$ where ϕ_1 is any complex injective function defined on $\mathbb{C}$. The main difference would lie in the non-trivial determination of A defined in (4.6). It also could be extend to some entropy distance which allows to distinguish variables which have less than p point of support than the others (see [19]).

- **Gaussian noise**

  The Gaussian assumption on the noise could be relaxed; actually only an infinitely divisible distribution is needed to establish a relation like

(4.3); nevertheless, in any cases one should get an explicit expression of A which could be non trivial since in particular it requires the calculations of the $\gamma_{j,k}$'s.

Thus, we have proposed a method that focuses on the recovery of both the inverse filter and the noise level of a noisy blind deconvolution model in a parametric setting. With a deeper investigation into the characterization of this model, we have provided a new estimation procedure that is simpler to implement compared with other existing methods. Through this procedure, we have estimated discrete distribution of the input signal and also derived the strong consistency and asymptotic normality for all our estimates. We perform a consistent simulation study that demonstrates empirically the computational performance of our estimation procedures including a comparison with another method.

General conclusion

Ensuring the quantile monotonicity is a challenging problem in quantile regression, since it requires a general and unique model that simultaneously estimates multiple quantiles with different levels. In the first part of this thesis, we present different methods in both the frequentist and the Bayesian frameworks that addressed the monotonicity issue. In Chapter 2, we addressed the crossing problem of quantile under a Bayesian paradigm and we proposed new modelling strategy to solve this problem. We assumed that the error of the quantile regression model follows the asymmetric Laplace distribution (ALD) and we used the relation between any two quantiles of the ALD, which allows to characterise the likelihood of the model through all the quantiles of interest. By our estimation procedure and using Metropolis-Hastings within Gibbs algorithm, we fitted parametric as well as non parametric quantile functions in the linear and the non linear case respectively. Through simulation study, we showed that our method performs better than other existent ones in terms of convergence (see our obtained diagnostics results) and flexibility (based on our RMISE interpretation). So, this work is the first step for a first fully Bayesian estimation of multiple quantiles.

As future perspectives, we envisage extending this work to deal with any conditional distribution of $Y \mid X$, not necessarily ALD. Indeed, when we deal with other data distribution, the relation between any two quantiles is no longer the same as that of the ALD. Therefore, we believe that attention

should be considered in establishing a general relation that links any two quantiles of any distribution. Furthermore, It is also of great interest to theoretically study the asymptotic behaviour (posterior convergence) of our estimators resulting from this procedure, first under ALD model likelihood and then generalized this to any model distribution.

A little further from quantile regression but certainly in the modelling paradigm, we proposed, in the second part of this thesis, a new estimation procedure for noisy blind deconvolution model where the objective is to recover the input signal once the inverse filter and the noise level are estimated. The method we proposed consists first on characterizing the distribution of the input signal, the noise and the inverse filter. Second, it consists on providing a contrast function defined through the noise level and the inverse filter and finally find the roots of its empirical counterpart that correspond to the estimates of the inverse filter and the noise level. We further proved the existence of the roots of the underlying empirical function when the cardinal p of the input signal support is such that $\frac{p(p+1)}{2}$ is even. We also provide asymptotic properties of the obtained estimators.

For this part, our perspectives for the futur are the following: first, prove the existence of roots of the empirical function for any $p \geq 2$; this means prove the conjecture 1 to any value of $p \geq 2$ and second, from computational point of view, adjust our procedure so that we can compute an infinite inverse filters. Indeed, as shown by our simulation results, our proposed procedure was unable to reconstruct the input signal when we dealt with the second order Moving average model whose inverse filter is infinite.

Finally, we would like to end this manuscript by claiming that applying techniques, such as those we proposed in this thesis to either the simultaneous quantile regression or the blind deconvolution models, is very appealing, and we believe this is a promising research area in these fields. Additionally, we believe attention should be considered in developing interpretable and manageable models, as they are vital for expanding the knowledge to a greater audience so that it becomes an easily accessible tool.

Appendix

VII Proofs of Chapter 4

VII.1 Proof of Proposition 1

For the sake of clarity, let us introduce some notations. Let $\mathbb{E}_{mp_t}$ be the empirical mean with respect to t going from $1+k(n)$ to $n-k(n)$, i.e., $\mathbb{E}_{mp_t} = \frac{1}{n-2k(n)} \sum_{t=1+k(n)}^{n-k(n)}$ and set $\otimes_{l=0}^{p} \mathbb{E}_{mp_{t_l}} = \mathbb{E}_{mp_{t_0}} \times \ldots \times \mathbb{E}_{mp_{t_p}}$. The random variables $\tilde{W}^{(k)}$ for $k = 0, \ldots, p$ are independent versions of W and $\mathbb{E}^{\tilde{W}^{(k)}}$ denotes the expectation with respect to the distribution of $\tilde{W}^{(k)}$. Since we consider any $\xi \in \mathcal{K}$, we will sometimes omit the dependence on ξ and ofter we will use $\check{s}$ instead of $\check{s}(\xi)$. Set $(\check{s} \star Y)_{t_{k,j}} = (\check{s} \star Y)_{t_k} - (\check{s} \star Y)_{t_j}$ and $\tilde{W}^{(k,j)} = \tilde{W}^{(k)} - \tilde{W}^{(j)}$. For convenience, set $\tilde{d}_{j(p+1)+k+1,n} = \tilde{d}_{j,k,n}$ and recall that

$$J_n(\sigma, \check{s}(\xi)) = \sum_{\pi \in S_{p+1}} \epsilon(\pi) \prod_{l=0}^{p} \tilde{d}_{n,\pi(l),l}(\sigma, \check{s}(\xi)),$$

where the sum is taken over S_{p+1} the set of permutations of $\{0, \ldots, p\}$ and $\epsilon(\pi)$ is the sign of the permutation π.

The next step is to rewrite J_n in a convenient way; roughly speaking, we aim to exchange the determinant and the empirical mean, mimicking what happens when dealing with J. This rewriting is stated as the following lemma whose proof is postponed after ending the proof Proposition 1.

Lemma 2. Under **(H2)**, one obtains for any $s \in \Theta$, for any $\sigma > 0$ and for any $n \in \mathbb{N}^*$,

$$\begin{aligned} J_n(\sigma, \check{s}) &= \frac{1}{(p+1)!} \\ &\otimes_{l=0}^{p} \mathbb{E}_{mp_{t_l}} \mathbb{E}^{\tilde{W}^{(l)}} \prod_{0 \le j < k \le p} \left((i(\sigma - \sigma_0)\|\check{s}\| + i\sigma_0 \|\check{s}\|)\tilde{W}^{(k,j)} + (\check{s} \star Y)_{t_{k,j}} \right) \\ &\left((i(\sigma - \sigma_0)\|\check{s}\| + i\sigma_0 \|\check{s}\|)\overline{\tilde{W}^{(k,j)}} + \overline{(\check{s} \star Y)_{t_{k,j}}} \right). \end{aligned}$$

Since $\tilde{D}_n(\sigma, \check{s})$ is a Hermitian matrix, its determinant is real and due to Lemma 2, $\sigma \in \mathbb{R}_+ \mapsto J_n(\sigma, \check{s})$ is a polynomial in $\mathbb{R}[X]$ with degree at most $p(p+1)$. Due to Eq (4.9), we first derive c_0 the constant term of $\sigma \in \mathbb{R}_+ \mapsto J_n(\sigma, \check{s})$ as follows

$$\begin{aligned} c_0 &= J_n(0, \check{s}(\xi)) = det((d_{j(p+1)+k+1,n}(\check{s})_{j,k}) \\ &= \frac{1}{(p+1)!} \otimes_{l=0}^{p} \mathbb{E}_{mp_{t_l}} \prod_{0 \le j < k \le p} |(\check{s} \star Y)_{t_{k,j}}|^2 > 0 \text{ a.s.}, \end{aligned}$$

since $(\check{s} \star Y)_{t_{k,j}}$ has more than p support points.

Therefore, to prove Proposition 1, it suffices to show that there exists σ such that $J_n(\sigma, \check{s}) < 0$.

Next, let us study $c_{p(p+1)}$ the coefficient of the highest degree term of $\sigma \in \mathbb{R}_+ \mapsto J_n(\sigma, \check{s})$.

$$\begin{aligned} c_{p(p+1)} &= \frac{1}{(p+1)!} \otimes_{l=0}^{p} \mathbb{E}_{mp_{t_l}} \mathbb{E}^{\tilde{W}^{(l)}} \prod_{0 \le j < k \le p} i^2 \|\check{s}\|^2 |\tilde{W}^{(k,j)}|^2 \\ &= (-1)^{p(p+1)/2} (\|\check{s}\|^2)^{p(p+1)} det\left[(\mathbb{E}^{\tilde{W}} (\tilde{W}^{(j)} \overline{\tilde{W}^{(k)}}))_{j,k=0,\dots,p} \right] \\ &= (-1)^{p(p+1)/2} (\|\check{s}\|^2)^{p(p+1)} det\left[(\gamma_{j,k})_{j,k=0,\dots,p} \right] \\ &= (-1)^{p(p+1)/2} \|\check{s}\|^{2p(p+1)} \prod_{k=0}^{p} k!, \end{aligned}$$

since $det\left((\gamma_{j,k})_{j,k=0,\dots,p}\right) = \prod_{k=0}^{p} k! > 0$. It means that for p such that $p(p+1)/2$ odd, $c_{p(p+1)} < 0$ and hence $J_n(\sigma, \check{s}(\xi)) \overset{\sigma\to+\infty}{\longrightarrow} -\infty$ that achieves the proof of Proposition 1.

Proof of Lemma 2: Set

$$\begin{cases} \varphi(x,y) &= (i(\sigma-\sigma_0)\|\check{s}\| + i\sigma_0\|\check{s}\|)y + x \\ \phi(x,y) &= (i(\sigma-\sigma_0)\|\check{s}\| + i\sigma_0\|\check{s}\|\overline{y} + \overline{x} \\ M_t &= (\check{s}\star Y)_t \\ R &= \tilde{W}, \end{cases}$$

and note that $d_{j,k,n}(\sigma,\check{s}) = \mathbb{E}_{mp_t}\mathbb{E}^R\varphi^j(M_t,R)\phi^k(M_t,R)$. In addition, $R^{(k)}$ for $k = 0,\dots,p$ are independent versions of R and $\mathbb{E}^R$ is the expectation with respect to R. Denote by τ some permutation in S_{p+1}. Then,

$$\begin{aligned}
&J_n(\sigma,\check{s}) \\
&= det\left(\left(\mathbb{E}_{mp_t}\mathbb{E}^R\varphi^j(M_t,R)\phi^k(M_t,R)\right)_{j,k=0,\dots,p}\right) \\
&= det\left(\left(\mathbb{E}_{mp_{t_k}}\mathbb{E}^{R^{(k)}}\varphi^j(M_{t_k},R^{(k)})\phi^k(M_{t_k},R^{(k)})\right)_{j,k=0,\dots,p}\right) \\
&= det\left(\left(\mathbb{E}_{mp_{t_{\tau(k)}}}\mathbb{E}^{R^{(\tau(k))}}\varphi^j(M_{t_{\tau(k)}},R^{(\tau(k))})\phi^k(M_{t_{\tau(k)}},R^{(\tau(k))})\right)_{j,k=0,\dots,p}\right) \\
&= \sum_{\pi\in S_{p+1}}\epsilon(\pi)\prod_{k=0}^{p}\mathbb{E}_{mp_{t_{\tau(k)}}}\mathbb{E}^{R^{(\tau(k))}}\left[\varphi^{\pi(k)}(M_{t_{\tau(k)}},R^{(\tau(k))})\phi^k(M_{t_{\tau(k)}},R^{(\tau(k))})\right] \\
&= \sum_{\pi\in S_{p+1}}\epsilon(\pi)\otimes_{l=0}^{p}\left(\mathbb{E}_{mp_{t_{\tau(l)}}}\mathbb{E}^{R^{\tau(l)}}\right)\prod_{k=0}^{p}\left(\varphi^{\pi(k)}(M_{t_{\tau(k)}},R^{(\tau(k))})\right) \\
&\qquad\qquad \prod_{k'=0}^{p}\left(\phi^{k'}(M_{t_{\tau(k')}},R^{(\tau(k'))})\right) \\
&= \otimes_{l=0}^{p}\mathbb{E}_{mp_{t_{\tau(l)}}}\mathbb{E}^{R^{\tau(l)}}\left(\sum_{\pi\in S_{p+1}}\epsilon(\pi)\prod_{k=0}^{p}\varphi^{\pi(k)}(M_{t_{\tau(k)}},R^{(\tau(k))})\right) \\
&\qquad\qquad \prod_{k=0}^{p}\left(\phi^k(M_{t_{\tau(k)}},R^{(\tau(k))})\right) \\
&= \otimes_{l=0}^{p}\mathbb{E}_{mp_{t_{\tau(l)}}}\mathbb{E}^{R^{\tau(l)}}det\left((\varphi^j(M_{t_{\tau(k)}},R^{(\tau(k))}))_{j,k=0,\dots,p}\right)\prod_{k=0}^{p}\left(\phi^k(M_{t_{\tau(k)}},R^{(\tau(k))})\right) \\
&= \otimes_{l=0}^{p}\mathbb{E}_{mp_{t_l}}\mathbb{E}^{R^{(l)}}\epsilon(\tau)det\left((\varphi^j(M_{t_k},R^{(k)}))_{j,k=0,\dots,p}\right)\prod_{k=0}^{p}\left(\phi^k(M_{t_{\tau(k)}},R^{(\tau(k))})\right).
\end{aligned}$$

Taking the sum over all permutations τ in the right-hand side of the last equation leads to

$$\begin{aligned}
& \otimes_{l=0}^{p}\mathbb{E}_{mp_{t_l}}\mathbb{E}^{R^{(l)}} det\left((\varphi^j(M_{t_k},R^{(k)}))_{j,k=0,\dots,p}\right)\sum_{\tau\in S_{p+1}}\epsilon(\tau)\prod_{k=0}^{p}\phi^k(M_{t_{\tau(k)}},R^{(\tau(k))}) \\
= \; & \otimes_{l=0}^{p}\mathbb{E}_{mp_{t_l}}\mathbb{E}^{R^{(l)}} det\left((\varphi^j(M_{t_k},R^{(k)}))_{j,k=0,\dots,p}\right) det\left((\phi^k(N_{t_j},R^{(j)})_{j,k=0,\dots,p}\right) \\
= \; & \otimes_{l=0}^{p}\mathbb{E}_{mp_{t_l}}\mathbb{E}^{R^{(l)}}\prod_{0\le j<k\le p}\left(\varphi(M_{t_k},R^{(k)})-\varphi(M_{t_j},R^{(j)})\right)\left(\phi(M_{t_j},R^{(j)})-\phi(M_{t_k},R^{(k)})\right).
\end{aligned}$$

That achieves the proof of Lemma 2 since the cardinality of S_{p+1} is equal to $(p+1)!$

VII.2 Proof of the asymptotic results

We first give a very useful tool which is a combination of existing results obtained by Gautherat [21] (see Lemma 5.3.1 p.130) and Gassiat and Gautherat [20] (see Lemma 4.1. p. 1695):

Lemma 3. Under assumptions **(H1)-(H4)** , one gets

i) $\forall\sigma\in\mathbb{R}_+^*,\forall\xi\in\mathcal{K},\quad J_n(\sigma,\check{s}(\xi))\xrightarrow[n\to\infty]{a.s.}J(\sigma,s(\xi))$.

ii) $\forall n\in\mathbb{N},\forall\xi\in\mathcal{K},\sigma\in\mathbb{R}_+\mapsto J_n(\sigma,\check{s}(\xi))$ is differentiable and
$\partial_1^1 J_n(\sigma_0,\check{s}(\xi_0))\xrightarrow[n\to\infty]{a.s.}\partial_1^1 J(\sigma_0,\theta)=-\alpha<0$.

iii) The function $s(\xi)\in\Theta\mapsto\partial_1 J_n(\sigma_0,\check{s}(\xi))$ is continuous on Θ.

iv) $\forall n\in\mathbb{N}$, $(\sigma,\xi)\mapsto J_n(\sigma,\check{s}(\xi))$ is twice differentiable in (σ_0,ξ_0). The first and second derivatives of $(\sigma,\xi)\mapsto J_n(\sigma,\check{s}(\xi))$ calculated in (σ_0,ξ_0) converge almost surely to the first and second derivative of $(\sigma,\xi)\mapsto J(\sigma,s(\xi))$ calculated in (σ_0,ξ_0).

v) The asymptotic distribution of $\sqrt{n}\left(\partial_2^1 J_n(\sigma_0, \check{s}(\xi_0)), J_n(\sigma_0, \check{s}(\xi_0))\right)$ is a centered Gaussian vector whose covariance matrix is

$$\partial^1 h(d(\sigma_0,\theta))A(\sigma_0\|\theta\|)\ \Gamma_1\ A(\sigma_0\|\theta\|)^{\mathsf{T}}\partial^1 h(d(\sigma_0,\theta))^{\mathsf{T}}.$$

Proof of Lemma 3 The proof of **iv)** of Lemma 3 follows directly the one of Theorem 4.1 in Gassiat and Gautherat [20] whereas points **i)**, **ii)** and **v)** come directly from Lemma 4.1 in Gassiat and Gautherat [20] (the only changes are the complex-valued signal and noise). Point **iii)** is derived using the truncated version of s combined with the polynomial structure of J_n.

Proof of Theorem 1.

• **Consistency of $\widehat{\sigma}_0$.** Let $\mathcal{V}_{\sigma_0}$ be some neighborhood of σ_0 and consider any σ_1 in $\mathcal{V}_{\sigma_0}$ such that $\sigma_1 < \sigma_0$. From (4.7), σ_1 satisfies $J(\sigma_1, s(\xi)) > 0\ \forall \xi$. On the other hand, due to assertion **ii)** of Lemma 3, it exists σ_2 in $\mathcal{V}_{\sigma_0}$, such that both $\sigma_2 > \sigma_0$ and $J(\sigma_2,\theta) < 0$ are valid; hence for any $\sigma \in]\sigma_0, \sigma_2[$, $J(\sigma,\theta) < 0$.

Assertion **i)** in Lemma 3 implies that

$$\begin{aligned} J_n(\sigma_1, \check{s}(\xi_0)) &\xrightarrow[n\to\infty]{a.s.} J(\sigma_1,\theta), \\ J_n(\sigma_2, \check{s}(\xi_0)) &\xrightarrow[n\to\infty]{a.s.} J(\sigma_2,\theta). \end{aligned}$$

Let $0 < \epsilon < \inf\{J(\sigma_1,\theta), |J(\sigma_2,\theta)|\}$. Then, it exists a positive integer N_0 such that for all $n > N_0$, $J_n(\sigma_1, \check{s}(\xi_0)) > 0$ and $J_n(\sigma_2, \check{s}(\xi_0)) < 0$. Thus, from **ii)** in Lemma 3, $\forall n > N_0$, it exists $\tilde{\sigma}_n \in]\sigma_1, \sigma_2[$ satisfying $J_n(\tilde{\sigma}_n, \check{s}(\xi_0)) = 0$; we set $\tilde{\sigma}_n$ defined by $\tilde{\sigma}_n = \inf\{\sigma_n \in]\sigma_1, \sigma_2[:\ J_n(\sigma_n, \check{s}(\xi_0)) = 0\}$. From Assertions **i)** and **ii)** in Lemma 3 combined with a Taylor expansion of $\sigma \mapsto J_n(\sigma, \check{s}(\xi_0))$

evaluated at σ_0

$$J_n(\tilde{\sigma}_n, \check{s}(\xi_0)) = J_n(\sigma_0, \check{s}(\xi_0)) + (\tilde{\sigma}_n - \sigma_0)\, \partial_1^1 J_n(\sigma_0, \check{s}(\xi_0))(1 + o(1)),$$

leads to $\tilde{\sigma}_n \xrightarrow[n\to\infty]{a.s.} \sigma_0$. Since for all $n > N_0$ and all $\sigma < \sigma_1$, $J_n(\sigma, \check{s}(\xi_0)) > 0$, it implies that for n large enough ($n > N_0$) $\widehat{\sigma}_0$ is such that $\widehat{\sigma}_0 > \sigma_1$. In the sequel consider only large n, i.e., $n > N_0$. By definition of $\widehat{\sigma}_0$, one has $\widehat{\sigma}_0 < \tilde{\sigma}_n$. Since we consider only σ lying in the compact set $[\sigma_1, \sigma_2]$, any subsequence $\tilde{\sigma}_{0,n}$ of $\widehat{\sigma}_0$ converges and denote by $\tilde{\sigma}_0$ its limit that satisfies $\tilde{\sigma}_0 \leq \sigma_0$.

Suppose that $\tilde{\sigma}_0 < \sigma_0$. Since $J_n(\tilde{\sigma}_{0,n}, \check{s}(\xi_0)) = 0$ and $J_n(\tilde{\sigma}_{0,n}, \check{s}(\xi_0)) \xrightarrow[n\to\infty]{a.s.} J(\tilde{\sigma}_0, \theta)$, it follows that $J(\tilde{\sigma}_0, \theta) = 0$, which contradicts the definition of σ_0 (see relation (4.7)); hence $\tilde{\sigma}_0 = \sigma_0$. The consistency of $\widehat{\sigma}_0$ is then achieved since our argument remains valid whatever the subsequence $\tilde{\sigma}_{0,n}$ of $\widehat{\sigma}_0$ is.

• **Consistency of $\widehat{\tilde{\xi}}_0$.** Consider only σ's in $[\sigma_1, \sigma_2]$. Since $\mathcal{K}$ is a compact set, any subsequence $(\tilde{\tilde{\xi}}_n)_n$ of $\widehat{\tilde{\xi}}_0$ converges and denote by $\tilde{\tilde{\xi}}_0$ its limit in $\mathcal{K}$. Assertion **i)** in Lemma 3, the a.s.-convergence of $\widehat{\sigma}_0$ and the continuity of J_n and s lead to $J_n(\widehat{\sigma}_0, \check{s}(\tilde{\tilde{\xi}}_n)) \xrightarrow[n\to\infty]{a.s.} J(\sigma_0, s(\tilde{\tilde{\xi}}_0))$. This implies that $\tilde{\tilde{\xi}}_0$ is equal to ξ_0 since $J(\sigma_0, s(\tilde{\tilde{\xi}}_0)) = 0 \iff s(\tilde{\tilde{\xi}}_0) = \theta$. Using the same arguement as for $\widehat{\sigma}_0$, the consistency of $\widehat{\tilde{\xi}}_0$ in $l_2(\mathbb{Z})$ is achieved.

Suppose now there exists $\overline{\tilde{\xi}}_0$ an accumulation point which is different from $\tilde{\tilde{\xi}}_0$. Then it exists another subsequence $\overline{\tilde{\xi}}_n$ of $\widehat{\tilde{\xi}}_0$ which converges to $\overline{\tilde{\xi}}_0$. Using the same tricks as previously, one gets $s(\overline{\tilde{\xi}}_0) = \theta$ which proves the uniqueness of $\tilde{\tilde{\xi}}_0$.

This achieves the proof.

Proof of Corollary 1. This proof comes directly from the proof of Theorem 3.2 in the unpublished work Gautherat [22]. It requires only the consistency of $\widehat{\sigma}_0$ and $\widehat{\tilde{\xi}}_0$.

Proof of Theorem 2. In the sequel, all partial derivatives of functions $J_n(\sigma, \check{s}(\xi))$ and $J(\sigma, s(\xi))$ are considered with respect to either σ or the vector ξ. The definition of $\widehat{\sigma}_0$ leads to $J_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) = 0$ as well as $J_n^2(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) = 0$.
It entails that both $\partial_1^1 J_n^2(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) = 0$ and $\partial_2^1 J_n^2(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) = 0_d$. For simplicity's sake, set $J_n = J_n(\sigma_0, \check{s}(\xi_0))$, $\partial_i^{r_i}\partial_i^{r_i} J_n = J_n(\sigma_0, \check{s}(\xi_0))$, $i = 1, 2$ and $\partial_{i,j}^{r_i, r_j} J_n = \partial_{i,j}^{r_i, r_j} J_n(\sigma_0, \check{s}(\xi_0))$, $i, j = 1, 2$. We then apply the Delta method, introduced in [52] with J_n^2 evaluated in (σ_0, ξ_0), since

$$\begin{pmatrix} \partial_2^1 J_n^2(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) \\ \partial_1^1 J_n^2(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) \end{pmatrix} = \begin{pmatrix} 2J_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0))\partial_2^1 J_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) \\ 2J_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0))\partial_1^1 J_n(\widehat{\sigma}_0, \check{s}(\widehat{\xi}_0)) \end{pmatrix} = \begin{pmatrix} 0_d \\ 0 \end{pmatrix}.$$

Setting $A_n = \partial_2^1 J_n (\partial_2^1 J_n)^\mathsf{T} + J_n \partial_2^2 J_n$, $B_n = \partial_2^1 J_n \partial_1^1 J_n + J_n \partial_{1,2}^{1,1} J_n$ and $d_n = (\partial_1^1 J_n)^2 + J_n \partial_1^2 J_n$, we consider the following Taylor expansion at the first order of J_n^2 at (σ_0, ξ_0):

$$\begin{pmatrix} 0_d \\ 0 \end{pmatrix} = \begin{pmatrix} 2J_n\partial_2^1 J_n \\ 2J_n\partial_1^1 J_n \end{pmatrix} + \begin{pmatrix} A_n & B_n \\ B_n^\mathsf{T} & d_n \end{pmatrix} \begin{pmatrix} \widehat{\xi}_0 - \xi_0 \\ \widehat{\sigma}_0 - \sigma_0 \end{pmatrix} (1 + o(1))$$

Then, from the Schur complement (see [51], page 11, formula 0.7.2. or Searle [58]), we obtain

$$\begin{pmatrix} \widehat{\xi}_0 - \xi_0 \\ \widehat{\sigma}_0 - \sigma_0 \end{pmatrix} (1 + o(1))$$

$$= \begin{pmatrix} J_n A_n^{-1}\partial_2^1 J_n + \frac{J_n}{d_n - B_n^\mathsf{T} A_n^{-1} B_n} \left(A_n^{-1} B_n B_n^\mathsf{T} A_n^{-1} \partial_2^1 J_n - A_n^{-1} B_n \partial_1^1 J_n \right) \\ \frac{J_n}{d_n - B_n^\mathsf{T} A_n^{-1} B_n} \left(-B_n^\mathsf{T} A_n^{-1} \partial_2^1 J_n + \partial_1^1 J_n \right) \end{pmatrix}.$$

Rewrite the top-term in the previous right-hand equation as follows,

$$J_n A_n^{-1}\partial_2^1 J_n + \frac{J_n}{d_n - B_n^\mathsf{T} A_n^{-1} B_n}\left(A_n^{-1}B_n B_n^\mathsf{T} A_n^{-1}\partial_2^1 J_n - A_n^{-1}B_n\partial_1^1 J_n\right) = J_n(T_1+T_2+T_3),$$

where

$$\begin{aligned} T_1 &= A_n^{-1}(\mathrm{Id}_d - \frac{(\partial_1^1 J_n)^2}{d_n - B_n^\mathsf{T} A_n^{-1} B_n})\partial_2^1 J_n, \\ T_2 &= A_n^{-1}\frac{B_n B_n^\mathsf{T} A_n^{-1}\partial_2^1 J_n}{d_n - B_n^\mathsf{T} A_n^{-1} B_n}, \\ T_3 &= -A_n^{-1}\frac{\partial_1^1 J_n}{d_n - B_n^\mathsf{T} A_n^{-1} B_n}\partial_{1,2}^{1,1} J_n. \end{aligned}$$

Rewrite the approximation of the vector $(\widehat{\xi}_0 - \xi_0, \widehat{\sigma}_0 - \sigma_0)^\mathsf{T}$ as follows

$$J_n \begin{pmatrix} T_1 + T_2 + T_3 \\ \frac{1}{d_n - B_n^\mathsf{T} A_n^{-1} B_n}\left(-B_n^\mathsf{T} A_n^{-1}\partial_2^1 J_n + \partial_1^1 J_n\right) \end{pmatrix}.$$

Due to Lemma 3 and the continuity of J_n in ξ, one has $A_n \xrightarrow[n\to\infty]{\mathbb{P}-a.s.} 0$, $B_n \xrightarrow[n\to\infty]{\mathbb{P}-a.s.} 0$, $d_n \xrightarrow[n\to\infty]{\mathbb{P}-a.s.} (\partial_1^1 J(\sigma_0,\theta))^2 = \alpha^2 > 0$, $\frac{A_n}{J_n} \xrightarrow[n\to\infty]{\mathbb{P}} \partial_2^2 J(\sigma_0, s(\xi_0))$, $\frac{B_n}{J_n} \xrightarrow[n\to\infty]{distrib} W$, where W is a d-dimensional non degenerate random vector and $\frac{d_n}{J_n} \xrightarrow[n\to\infty]{\mathbb{P}-a.s.} \pm\infty$. As n large enough, it entails that

$$\sqrt{n}\begin{pmatrix} \widehat{\xi}_0 - \xi_0 \\ \widehat{\sigma}_0 - \sigma_0 \end{pmatrix} \quad \begin{matrix} distrib \\ = \end{matrix} \quad \sqrt{n}J_n \begin{pmatrix} \frac{(\partial_2^2 J(\sigma_0, s(\xi_0)))^{-1}\partial_{1,2}^{1,1} J(\sigma_0, s(\xi_0))}{\alpha} \\ \\ \frac{1}{\alpha} \end{pmatrix}. \quad (4.10)$$

Note that $h(\tilde{d}_n(\sigma_0, \check{s}(\xi_0))) = J_n(\sigma_0, \check{s}(\xi_0))$, where h is the determinant function. Then, due to Assumption **(H3) d**, and due to the Taylor expansion of h

at $\tilde{d}(\sigma_0, \theta)$, one obtains

$$J_n \;=\; \partial^1 h(\tilde{d}(\sigma_0,\theta)) A(\sigma_0\|\theta\|)(d_n(\check{s}(\xi_0)) - d(s(\xi_0)) + o(\frac{1}{\sqrt{n}})). \quad (4.11)$$

Set $M = A(\sigma_0\|\theta\|)\Gamma_1 A(\sigma_0\|\theta\|)^{\mathsf{T}}$ and

$N = \begin{pmatrix} \frac{1}{\alpha}\,(\partial_2^2 J(\sigma_0, s(\xi_0)))^{-1}\,\partial_{1,2}^{1,1} J(\sigma_0, s(\xi_0)) & \\ & \\ & \frac{1}{\alpha} \end{pmatrix}$, then due to (4.10), (4.11) and **v)** of Lemma 3 one gets,

$$\sqrt{n}\begin{pmatrix} \widehat{\xi}_0 - \xi_0 \\ \widehat{\sigma}_0 - \sigma_0 \end{pmatrix} \xrightarrow[n\to\infty]{distrib} \mathcal{N}\left(0_{d+1}, N\partial^1 h(\tilde{d}(\sigma_0,\theta))\; M\; \partial^1 h(\tilde{d}(\sigma_0,\theta))^{\mathsf{T}} N^{\mathsf{T}}\right).$$

Proof of Corollary 2. Following the proof of Theorem 3.3 in Gautherat [22], it remains to obtain an equivalent for $\sqrt{n}(\tilde{d}_n(\hat{\sigma}_0, \check{s}(\widehat{\xi}_0)) - \tilde{d}(\sigma_0, s(\xi_0)))$. As n large enough, this term is equivalent in distribution to

$$\left(\mathrm{Id}_{(p+1)^2} + \partial^1\tilde{d}(\sigma_0,\theta)\partial^1 h(\tilde{d}(\sigma_0,\theta))N\right) A(\sigma_0, \|\theta\|)\sqrt{n}(d_n(\check{s}(\xi_0)) - d(s(\xi_0))).$$

On the other hand, one has

$$\sqrt{n}(\widehat{a} - a) = \left(\frac{C^{-1}}{2|v_p^*|^2}B\right)\sqrt{n}(\tilde{d}_n(\hat{\sigma}_0, \check{s}(\widehat{\xi}_0)) - \tilde{d}(\sigma_0, s(\xi_0))), \quad (4.12)$$

$$\sqrt{n}(\widehat{\Pi} - \Pi) = L^{-1}\left(Proj + F\frac{C^{-1}}{2|v_p^*|^2}B\right)\sqrt{n}(\tilde{d}_n(\hat{\sigma}_0, \check{s}(\widehat{\xi}_0)) - \tilde{d}(\sigma_0, s(\xi_0))). \quad (4.13)$$

Relations (4.12) and (4.13) entail the results. All matrices used here are defined in the statement of both Theorem 2 or Corollary 2.

Bibliography

[1]

[2] R. Alhamzawi, K. Yu, and D. F. Benoit. Bayesian adaptive lasso quantile regression. *Statistical Modelling*, 12(3):279–297, 2012.

[3] M. Behr and A. Munk. Identifiability for blind source separation of multiple finite alphabet linear mixtures. *IEEE Transactions on Information Theory*, 63(9):5506–5517, 2017.

[4] D. F. Benoit and D. Van den Poel. Binary quantile regression: a bayesian approach based on the asymmetric laplace distribution. *Journal of Applied Econometrics*, 27(7):1174–1188, 2012.

[5] D. F. Benoit, R. Alhamzawi, and K. Yu. Bayesian lasso binary quantile regression. *Computational Statistics*, 28(6):2861–2873, 2013.

[6] D. F. Benoit, D. Van den Poel, et al. bayesqr: A bayesian approach to quantile regression. *Journal of Statistical Software*, 76(7):1–32, 2017.

[7] M. Bernardi, G. Gayraud, L. Petrella, et al. Bayesian tail risk interdependence using quantile regression. *Bayesian Analysis*, 10(3):553–603, 2015.

[8] H. D. Bondell, B. J. Reich, and H. Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838, 2010.

[9] A.-O. Boudraa, J.-C. Cexus, et al. Denoising via empirical mode decomposition.

[10] R. Chen and T.-H. Li. Blind restoration of linearly degraded discrete signals by gibbs sampling. *IEEE Transactions on Signal Processing*, 43(10): 2410–2413, 1995.

[11] V. Chernozhukov, I. Fernandez-Val, and A. Galichon. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2009.

[12] G. De Rossi and A. Harvey. Quantiles, expectiles and splines. *Journal of Econometrics*, 152(2):179–185, 2009.

[13] H. Dette and S. Volgushev. Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):609–627, 2008.

[14] D. B. Dunson and J. A. Taylor. Approximate bayesian inference for quantiles. *Nonparametric Statistics*, 17(3):385–400, 2005.

[15] P. H. Eilers and B. D. Marx. Flexible smoothing with b-splines and penalties. *Statistical science*, pages 89–102, 1996.

[16] F. Gamboa and E. Gassiat. Source separation when the input sources are discrete or have constant modulus. *IEEE transactions on signal processing*, 45(12):3062–3072, 1997.

[17] F. Gamboa, E. Gassiat, et al. Blind deconvolution of discrete linear systems. *The Annals of Statistics*, 24(5):1964–1981, 1996.

[18] F. Gamboa, E. Gassiat, et al. Bayesian methods and maximum entropy for ill-posed inverse problems. *The Annals of Statistics*, 25(1):328–350, 1997.

[19] E. Gassiat and E. Gautherat. Identification of noisy linear systems with discrete random input. *IEEE Transactions on Information Theory*, 44(5): 1941–1952, 1998.

[20] E. Gassiat and E. Gautherat. Speed of convergence for the blind deconvolution of a linear system with discrete random input. *Annals of statistics*, pages 1684–1705, 1999.

[21] E. Gautherat. *Déconvolution aveugle des systemes linéaires aléatoires discrets bruités ou non*. PhD thesis, 1997.

[22] E. Gautherat et al. Déconvolution aveugle bruitée: estimation de la distribution du processus source. Technical report, 2002.

[23] R. Giri and T. Zhang. Bayesian blind deconvolution with application to acoustic feedback path modeling. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 601–605. IEEE, 2017.

[24] J. H. Gunther and A. L. Swindlehurst. Recursive blind symbol estimation of convolutionally coded cochannel signals. *IEEE transactions on signal processing*, 48(4):956–965, 2000.

[25] C. Gutenbrunner and J. Jurecková. Regression rank scores and regression quantiles. *The Annals of Statistics*, pages 305–330, 1992.

[26] X. He. Quantile curves without crossing. *The American Statistician*, 51 (2):186–192, 1997.

[27] X. He and P. Ng. Quantile splines with several covariates. *Journal of Statistical Planning and Inference*, 75(2):343–352, 1999.

[28] N. L. Hjort, S. G. Walker, et al. Quantile pyramids for bayesian nonparametrics. *The Annals of Statistics*, 37(1):105–131, 2009.

[29] K. Jaganathan and B. Hassibi. Reconstruction of signals from their autocorrelation and cross-correlation vectors, with applications to phase retrieval and blind channel estimation. *arXiv preprint arXiv:1610.02620*, 2016.

[30] H. Jeffreys. Theory of probability, 1961.

[31] A. Khan and H. Yin. Parametric blind image deblurring with gradient based spectral kurtosis maximization. *Image Analysis & Stereology*, 37(3): 213–223, 2018.

[32] R. Koenker. Quantile regression in r: a vignette (2015). *URL: https://cran. r-project. org/web/packages/quantreg/vignettes/rq. pdf (accessed Feb 28, 2018)*, 2015.

[33] R. Koenker and J. G. Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[34] R. Koenker and S. Portnoy. Nonparametric estimation of conditional quantile functions. 1992.

[35] R. Koenker and Z. Xiao. Inference on the quantile regression process. *Econometrica*, 70(4):1583–1612, 2002.

[36] A. Kottas and A. E. Gelfand. Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association*, 96(456): 1458–1468, 2001.

[37] A. Kottas and M. KRNJAJIĆ. Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36(2):297–319, 2009.

[38] S. Kotz, T. Kozubowski, and K. Podgorski. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media, 2012.

[39] H. Kozumi and G. Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81 (11):1565–1578, 2011.

[40] N. A. Lazar. Bayesian empirical likelihood. *Biometrika*, 90(2):319–326, 2003.

[41] K. Lee, F. Krahmer, and J. Romberg. Spectral methods for passive imaging: Nonasymptotic performance and robustness. *SIAM Journal on Imaging Sciences*, 11(3):2110–2164, 2018.

[42] K.-C. Li and K. SHEDDEN. Monte carlo deconvolution of digital signals guided by the inverse filter. *Journal of the American Statistical Association*, 96(455):1014–1014, 2001.

[43] T.-H. Li. Blind deconvolution of linear systems with multilevel nonstationary inputs. *The Annals of Statistics*, pages 690–704, 1995.

[44] H. Liu, L. Yan, T. Huang, S. Liu, and Z. Zhang. Blind spectral signal deconvolution with sparsity regularization: an iteratively reweighted least-squares solution. *Circuits, Systems, and Signal Processing*, 36(1):435–446, 2017.

[45] J. S. Liu and R. Chen. Blind deconvolution via sequential imputations. *Journal of the american statistical association*, 90(430):567–576, 1995.

[46] Y. Liu and Y. Wu. Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and its Interface*, 2(3):299–310, 2009.

[47] V. M. Muggeo. quantreggrowth-package: Growth charts via regression quantiles. 2018.

[48] V. M. Muggeo, M. Sciandra, A. Tomasello, and S. Calvo. Estimating growth charts via nonparametric quantile regression: a practical framework with application in ecology. *Environmental and ecological statistics*, 20(4):519–531, 2013.

[49] M. Nagahara. Discrete signal reconstruction by sum of absolute values. *IEEE Signal Processing Letters*, 22(10):1575–1579, 2015.

[50] D. Nychka, G. Gray, P. Haaland, D. Martin, and M. O'connell. A nonparametric regression approach to syringe grading for quality improvement. *Journal of the American Statistical Association*, 90(432):1171–1178, 1995.

[51] S. Puntanen and G. P. Styan. Historical introduction: Issai schur and the early development of the schur complement. In *The Schur complement and its applications*, pages 1–16. Springer, 2005.

[52] C. R. Rao, C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao. *Linear statistical inference and its applications*, volume 2. Wiley New York, 1973.

[53] B. J. Reich. Spatiotemporal quantile regression for detecting distributional changes in environmental processes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(4):535–553, 2012.

[54] B. J. Reich and L. B. Smith. Bayesian quantile regression for censored data. *Biometrics*, 69(3):651–660, 2013.

[55] B. J. Reich, M. Fuentes, and D. B. Dunson. Bayesian spatial quantile regression. *Journal of the American Statistical Association*, 106(493):6–20, 2011.

[56] T. Rodrigues and Y. Fan. Regression adjustment for noncrossing bayesian quantile regression. *Journal of Computational and Graphical Statistics*, 26(2):275–284, 2017.

[57] M. Sangnier, O. Fercoq, and F. d'Alché Buc. Joint quantile regression in vector-valued rkhss. In *Advances in Neural Information Processing Systems*, pages 3693–3701, 2016.

[58] S. R. Searle and A. I. Khuri. *Matrix algebra useful for statistics*. John Wiley & Sons, 2017.

[59] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IEEE*, 86(2):447–457, 1998.

[60] L. Smith, B. Reich, and M. L. Smith. Package ,Äòbsquare,Äô. 2013.

[61] K. Sriram. A sandwich likelihood correction for bayesian quantile regression based on the misspecified asymmetric laplace density. *Statistics & Probability Letters*, 107:18–26, 2015.

[62] K. Sriram, R. Ramamoorthi, P. Ghosh, et al. Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis*, 8(2):479–504, 2013.

[63] K. Sriram, R. Ramamoorthi, and P. Ghosh. On bayesian quantile regression using a pseudo-joint asymmetric laplace likelihood. *Sankhya A*, 78 (1):87–104, 2016.

[64] I. Takeuchi, Q. V. Le, T. D. Sears, and A. J. Smola. Nonparametric quantile estimation. *Journal of Machine Learning Research*, 7(Jul):1231–1264, 2006.

[65] S. T. Tokdar, J. B. Kadane, et al. Simultaneous linear quantile regression: a semiparametric bayesian approach. *Bayesian Analysis*, 7(1):51–72, 2012.

[66] M. Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88 (4):569–587, 2000.

[67] P. Walk, P. Jung, G. E. Pfander, and B. Hassibi. Blind deconvolution with additional autocorrelations via convex programs. *arXiv preprint arXiv:1701.04890*, 2017.

[68] D. Wipf and H. Zhang. Revisiting bayesian blind deconvolution. *The Journal of Machine Learning Research*, 15(1):3595–3634, 2014.

[69] Y. Yang and S. T. Tokdar. Joint estimation of quantile planes over arbitrary predictor spaces. *Journal of the American Statistical Association*, 112 (519):1107–1120, 2017.

[70] Y. Yang, X. He, et al. Bayesian empirical likelihood for quantile regression. *The Annals of Statistics*, 40(2):1102–1131, 2012.

[71] Y. Yang, H. J. Wang, and X. He. Posterior inference in bayesian quantile regression with asymmetric laplace likelihood. *International Statistical Review*, 84(3):327–344, 2016.

[72] K. Yu and M. Jones. Local linear quantile regression. *Journal of the American statistical Association*, 93(441):228–237, 1998.

[73] K. Yu and Z. Lu. Local linear additive quantile regression. *Scandinavian Journal of Statistics*, 31(3):333–346, 2004.

[74] K. Yu and R. A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4):437–447, 2001.

[75] K. Yu and J. Stander. Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics*, 137(1):260–276, 2007.

[76] K. Yu and J. Zhang. A three-parameter asymmetric laplace distribution and its extension. *Communications in Statistics,ÄîTheory and Methods*, 34 (9-10):1867–1879, 2005.

[77] M. Yuan. Gacv for quantile smoothing splines. *Computational statistics & data analysis*, 50(3):813–829, 2006.

[78] K. Q. Zhou, S. L. Portnoy, et al. Direct use of regression quantiles to construct confidence sets in linear models. *The Annals of Statistics*, 24(1): 287–306, 1996.