On the Multivariate Conditional Quantiles Estimation

Malak Lotfy Musa Abu Musa

Supervised by

Prof. Dr. Raid B. Salha

statistic

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Mathematics

November/2016
On the Multivariate Conditional Quantiles Estimation

حوال تقدير المئينات الشرطية المتعددة

أقر بأن ما اشتملت عليه هذه الرسالة إنما هو نتاج جهدي الخاص، باستثناء ما تمت الإشارة إليه حيئما ورد، وأن هذه الرسالة ككل أو أي جزء منها لم يقدم من قبل الآخرين لنيل درجة أو لقب علمي أو بحثي لدى أي مؤسسة تعليمية أو بحثية أخرى.

Declaration

I understand the nature of plagiarism, and I am aware of the University's policy on this.

The work provided in this thesis, unless otherwise referenced, is the researcher's own work, and has not been submitted by others elsewhere for any other degree or qualification.

Student's name: ملك لطفي موسى أبو موسى
Signature: ملك أبو موسى
Date: 2/11/2016

اسم الطالب: 
توقيع: 
التاريخ: 

ملك لطفي موسى أبو موسى
نتيجة الحكم على أطروحة ماجستير

بناءً على موافقة شئون البحث العلمي والدراسات العليا بالجامعة الإسلامية بغزة على تشكيل لجنة الحكم على أطروحة الباحثة/ ملكة نظمى موسى أبو موسى ليل درجة الماجستير في كلية العلوم قسم الرياضيات وموضوعها:

 حول تقدير المئينات الشرطية المتعددة

On the Multivariate Conditional Quantiles Estimation

وبعد المناقشة العلمية التي تمت اليوم الأربعاء 29 صفر 1438هـ الموافق 30/11/2016م الساعة الحادية عشر صباحاً بمبني اللحيدان، اجتمعت لجنة الحكم على الأطروحة والمكونة من:

أ.د. رائد بشير صالحة
د. عصام بشير مهدي
د. حازم اسماعيل الشيخ أحمد

وبعد المداولات أوصت اللجنة بمنح الباحثة درجة الماجستير في كلية العلوم قسم الرياضيات.

وإن تمنح هذه الدرجة فإنها توصي بها في خدمة بلدها ووطنيها.

وإن تمنح هذه الدرجة فإنها توصي بها بقدوة الله ولزوم طاعته وأن تصر على عملها في

الله وإلى الله التوفيق

نائب الرئيس لشؤون البحث العلمي والدراسات العليا

أ.د. عبد الروؤف علي المناعمة
Abstract

We are interested in the area of nonparametric prediction, therefore, we will study the relationship between a current observation and past observations, where the conditional density function and the conditional quantiles play an important role.

In this thesis, we study the kernel estimation of the conditional probability density function, the conditional cumulative density function, and the univariate conditional quantile, then we generalize it for the case of multivariate conditional quantiles by considering multivariate conditional quantiles based on norm minimization. For the multivariate conditional quantiles, we studied the conditions under which we have driven the asymptotic consistency of the kernel estimation of the conditional quantile.

Finally, we use the conditional quantile estimation in some applications, by considering the prediction intervals of a financial position with multiple assets.
الملخص

الدراسة في هذا البحث تهتم بالنتيئات اللامعممية، وبدالة الكثافة الشرطية، والمئينات الشرطية التي تلعب دوراً مهماً في هذا البحث.

الهدف الرئيسي للبحث هو دراسة تقدير النواة لدالة الكثافة الاحتمالية الشرطية، ودالة الكثافة التراكمية المشروطة، ومئينات الشرطية الوحيدة ثم تعميمها لحالة المئينات الشرطية المتعددة على أساس تقليل القاعدة.

في حالة المئينات الشرطية المتعددة درسنا الظروف التي أثبتت حالة الاختلاف التقاربي لتقدير نواة المئين الشرطي.

أخيراً نستخدم تقدير المئينات الشرطية في بعض التطبيقات باستخدام فترات التنبؤ للمركز المالي متعدد الأصول.
Dedication

To My Parents.

To My brother and sister.

To My Friends.

To all Knowledge Seekers.
Acknowledgment

First of all, I am awarding my great thanks for the Almighty Allah who all the time helps me and grants me the power and courage to finish this study and give me the success in my live.
I would like to express my grateful sincere to my family especially my Parents and my brother for give me confidence, support, and help me to reach this level of learning.
I would like to express my sincere to professor Dr. Raid Salha, my supervisors, for his grateful efforts, support, and continuous supervision through preparing this thesis.
I would also like to thank my thesis committee members, Dr.Hazem Al Sheikh Ahmed and Dr. Esam Mahdi, for their contributions to this work.
and I am also thankful to all my friends for their kind advises, and encouragement.
Finally. I pray to god to accept this work.
# Table of Contents

Declaration ........................................................................................................... i
Abstract in English ............................................................................................ ii
Abstract in Arabic .............................................................................................. iii
Dedication ............................................................................................................ iv
Acknowledgment ................................................................................................. v
Table of Contents ................................................................................................. vi
List of Tables ......................................................................................................... viii
List of Figures ....................................................................................................... ix
List of Abbreviations ........................................................................................... x
List of Symbols ..................................................................................................... xi

## Chapter 1 Preliminaries ...................................................................................... 4
  1.1 Basic Definitions and Notations ................................................................. 4
  1.2 Estimation .................................................................................................. 6
  1.3 Kernel Density Estimation of the Pdf ....................................................... 14
  1.4 Kernel Density Estimation of the Cdf ....................................................... 15
  1.5 Properties of the Kernel Estimator ......................................................... 16
  1.6 Optimal Bandwidth .................................................................................. 19

## Chapter 2 Univariate Conditional Quantile .................................................... 22
  2.1 Important of Quantiles ............................................................................ 22
  2.2 The Nadaraya-Watson Estimator .............................................................. 24
  2.3 Estimating the Univariate Conditional Quantile ...................................... 25
  2.4 Asymptotic Properties of the N-W Estimator of the Cdf ....................... 27

## Chapter 3 Multivariate Conditional Quantiles ............................................... 31
  3.1 Important of Multivariate Conditional Quantile ....................................... 31
  3.2 The Mean as A minimization Problem ..................................................... 32
  3.3 The Conditional Quantile as A minimization Problem ............................ 34
  3.4 Multivariate Quantile Based on A norm Minimization ............................. 36
  3.5 The Nadaraya-Watson Estimator of the Multivariate Conditional Quantile .... 38
  3.6 Consistency of the Multivariate Nadaraya-Watson Estimator ................. 41

## Chapter 4 Application ...................................................................................... 51
  4.1 Prediction intervals for A bivariate Time Series ....................................... 51
  4.2 Discussion and Conclusion ....................................................................... 59
The Reference List .................................................................................................................. 60
List of Tables

Table (4.1): 90% C.I. for the last 8 observation of the IBM ........................................ 55
Table (4.2): 90% C.I. for the last 8 observation of the SP500 ....................................... 56
Table (4.3): 95% C.I. for the last 8 observation of the IBM ......................................... 57
Table (4.4): 95% C.I. for the last 8 observation of the SP500 ....................................... 58
List of Figures

Figure (4.1): Time plot of the rescaled IBM stock. ............................................... 53
Figure (4.2): Time plot of the rescaled SP500 stock. ........................................... 53
Figure (4.3): Scatterplot of the rescaled IBM stock versus the rescaled SP500 stock. .................................................................................................................. 54
Figure (4.4): Scatterplot of the squares of the rescaled IBM stock versus the squares of the rescaled SP500 stock. ................................................................. 54
Figure (4.5): 90% C.I. for the last 8 observation of the IBM. .............................. 55
Figure (4.6): 90% C.I. for the last 8 observation of the SP500. ............................ 56
Figure (4.7): 95% C.I. for the last 8 observation of the IBM ............................... 57
Figure (4.8): 95% C.I. for the last 8 observation of the SP500 ............................ 58
**List of Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>arg</td>
<td>Argument</td>
</tr>
<tr>
<td>Cdf</td>
<td>Cumulative distribution function</td>
</tr>
<tr>
<td>Cov</td>
<td>Covariance</td>
</tr>
<tr>
<td>C.I.</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>i.i.d.</td>
<td>Independent and identically distributed</td>
</tr>
<tr>
<td>Inf</td>
<td>Infinum</td>
</tr>
<tr>
<td>Min</td>
<td>Minimum</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Error</td>
</tr>
<tr>
<td>o</td>
<td>Small oh</td>
</tr>
<tr>
<td>O</td>
<td>Big oh</td>
</tr>
<tr>
<td>Pdf</td>
<td>Probability density function</td>
</tr>
<tr>
<td>Var.</td>
<td>Variance</td>
</tr>
<tr>
<td>N-W</td>
<td>Nadaraya Watson</td>
</tr>
<tr>
<td>M-NW</td>
<td>Multivariate Nadaraya-Watson</td>
</tr>
<tr>
<td>DK</td>
<td>Double Kernel</td>
</tr>
<tr>
<td>w.p</td>
<td>With probability</td>
</tr>
</tbody>
</table>
On the Multivariate Conditional Quantiles Estimation

December 10, 2016
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_\alpha(x)$</td>
<td>the population conditional quantile.</td>
</tr>
<tr>
<td>$\rho_\alpha(u)$</td>
<td>check function.</td>
</tr>
<tr>
<td>$q_{\alpha,n}(x)$</td>
<td>the sample conditional quantile.</td>
</tr>
<tr>
<td>$|\cdot|$</td>
<td>the $l^p$–norm function.</td>
</tr>
<tr>
<td>$|\cdot|_{p,\alpha}$</td>
<td>the norm like function.</td>
</tr>
<tr>
<td>$\theta_\alpha(x)$</td>
<td>the population multivariate conditional –quantile.</td>
</tr>
<tr>
<td>$\theta_{\alpha,n}(x)$</td>
<td>the sample multivariate conditional -quantile.</td>
</tr>
</tbody>
</table>
Preface

The probability density function is a fundamental concept in statistics. Suppose we have a set of observed data points assumed to be a sample from an unknown probability density function $f$, the construction of an estimate of the density function from observed data is known as density estimation.

There are many methods for statistical estimation of the density function, these methods are divided into two kinds, parametric estimation and nonparametric estimation. The classical approach for estimating the density function is called parametric density estimation. In this approach, one assumes that the data are drawn from a known parametric distribution which depends only on finitely many parameters, and one uses the data to estimate the unknown values of these parameters.

For example, the normal distribution depends on two parameters, the mean $\mu$, and the variance $\sigma^2$. The density function $f$ could be estimated by finding estimates of $\mu$ and $\sigma^2$ from the data, and substituting these estimates into the formula for the normal density. Parametric estimates usually depend only on few parameters, therefore they are suitable even for small sample sizes $n$.

On the other hand, nonparametric estimation, in this case we do not assume a priori a known parametric distribution, but the data themselves are allowed to decide which function fits them best, without the restrictions imposed by the parametric estimation, see Tarter and Kronmal (1976)[29], Silverman (1986)[28], and Hardle, et al. (1997)[15]. There are several reasons for using the nonparametric smoothing approaches.

1. they can be employed as a convenient and succinct means of displaying the features of a data set and hence to aid practical parametric model building.
2. they can be used for diagnostic checking of an estimated parametric model.

3. one may want to conduct inference under only the minimal restrictions imposed in fully nonparametric structures.

For more details see Engle and McFadden(1994,Chapter 38)[11].

The main subject of this thesis, is the kernel estimation of the probability density function. In this thesis, we will study the conditional quantile. Then we will generalize it for the case of multivariate conditional quantile.

There are many approaches to extend the univariate quantiles to the multivariate case:[27]

1. Multivariate quantile functions based on depth functions.

2. Multivariate quantiles based on norm minimization.

3. Multivariate quantiles as inverse mappings.

4. Data-based multivariate quantiles based on gradients.

The estimator we introduce in Section 3.4 is based on a norm minimization approach. Also we will study some theoretical properties of the kernel estimator for the conditional density function, univariate conditional quantile, and the multivariate conditional quantiles.

This thesis will consist of the following chapters

**Chapter 1. Preliminaries**

This chapter contains notations, some basic definitions, and facts that we need in the remaining of this thesis. Also, it will contain an introduction to the kernel estimation.

**Chapter 2. Univariate Conditional Quantiles**

In this chapter we will study the univariate conditional quantile, for this purpose, we will study the Nadaraya Watson estimator of the conditional pdf, and conditional cdf. In this chapter, we study the asymptotic properties of the Nadaraya-Watson estimator of the conditional quantile estimation.
Chapter 3. Multivariate Conditional Quantiles
This chapter is the main chapter of the thesis. In this chapter, we will introduce the multivariate conditional quantiles, multivariate quantiles based on a norm minimization, and we will prove the consistency of the multivariate Nadaraya-Watson estimator.

Chapter 4 (Applications)
In this chapter, we will give some applications of the NW estimator of the multivariate conditional quantile to real data. We will use it to construct 90% and 95% prediction intervals for a bivariate time series consists of the IBM and SP500 time series.
Chapter 1
Preliminaries
Chapter 1

Preliminaries

This chapter contains notations, some basic definitions, and facts that we need in the remaining of this thesis.

In Section 1.1, we introduce some basic definitions and notations. In section 1.2, we present some well known estimators of the density function. Section 1.3, introduces the kernel density estimator of the probability density function (pdf). In Section 1.4, we introduce the kernel density estimator of the cumulative density function (cdf). In the next section, we summarize some properties of the kernels. Finally, in Section 1.6, we present the problems of the optimal bandwidth selection.

1.1 Basic Definitions and Notations

In this section we will introduce some basic definitions and theorems that will be helpful in the remaining of this thesis.

Definition 1.1.1. Compact space [10]

A space is compact if and only if each open cover has a finite subcover.
Every closed interval is a compact.

Definition 1.1.2. Indicator function [26]

If $A$ is any set, we define the indicator function $I_A$ of the set $A$ to be the function given by
\[ I_A = \begin{cases} 
1 & \text{if } x \in A, \\
0 & \text{if } x \notin A. 
\end{cases} \]

**Definition 1.1.3. (Converge in Probability) [16].**

Let \( X_n \) be a sequence of random variables and let \( X \) be a random variable defined on a sample space. We say \( X_n \) converges in probability to \( X \) if for all \( \epsilon > 0 \), we have

\[
\lim_{n \to \infty} P[|X_n - X| \geq \epsilon] = 0, \tag{1.1.1}
\]

or equivalently,

\[
\lim_{n \to \infty} P[|X_n - X| < \epsilon] = 1. \tag{1.1.2}
\]

If so, we write \( X_n \xrightarrow{p} X \).

**Definition 1.1.4. Converge in Distribution [16].**

Let \( X_n \) be a sequence of random variables and let \( X \) be a random variable. Let \( F_{X_n} \) and \( F_X \) be the cdfs of \( X_n \) and \( X \) respectively. Let \( C(F_X) \) denote the set of all points where \( F_X \) is continuous. We say that \( X_n \) converge in distribution to \( X \) if

\[
\lim_{n \to \infty} F_{X_n}(x) = F_X(x), \quad \text{for all } x \in C(F_X). \tag{1.1.3}
\]

We denote this convergence by

\[ X_n \xrightarrow{d} X \]

**Definition 1.1.5. Converge with probability 1 [16]**

Let \( \{X_n\}_{n=1}^{\infty} \) be a sequence of random variables on \((\Omega, L, P)\). We say that \( X_n \) converge almost surly to a random variable \( X \) (\( X_n \xrightarrow{a.s.} X \)) or Converge with probability 1 to \( X \) or \( X_n \) converge strongly to \( X \) if and only if

\[
P(\{w : X_n(w) \to X(w), \text{ as } n \to \infty\}) = 1,
\]

or equivalent, for all \( \epsilon > 0 \), there exists \( N \in \mathbb{N} \)

\[
P(|X_n - X| < \epsilon, \ n \geq N) = 1.
\]

**Definition 1.1.6. Order Notation \( O \) and \( o \) [31].**

Let \( a_n \) and \( b_n \) each be sequences of real numbers. Then we say that \( a_n \) is of order...
$b_n$ or $a_n$ is big oh $b_n$) as $n \to \infty$ and write $a_n = O(b_n)$ as $n \to \infty$, if and only if

\[ \lim_{n \to \infty} \sup |\frac{a_n}{b_n}| < \infty. \]

In other words, $a_n = O(b_n)$ if $|\frac{a_n}{b_n}|$ remains bounded as $n \to \infty$.

We say that $a_n$ is of small order $b_n$ and write $a_n = o(b_n)$ as $n \to \infty$, if and only if

\[ \lim_{n \to \infty} \frac{|a_n|}{b_n} = 0. \]

Taylor expansion is important mathematical tool for obtaining asymptotic approximations in kernel smoothing and allows us to approximate function values close to a given point in term of higher-order derivatives at that point (provided the derivatives exists).

**Theorem 1.1.1. Taylor’s Theorem** [31]

Suppose that $f$ is real-valued function defined on $\mathbb{R}$ and let $x \in \mathbb{R}$. Assume that $f$ has $p$ continuous derivatives in an interval $(x - \delta, x + \delta)$ for some $\delta > 0$. Then for any sequence $\alpha_n$ converging to zero,

\[ f(x + \alpha_n) = \sum_{j=0}^{p} \frac{\alpha_n^j}{j!} f^{(j)}(x) + o(\alpha_n^p). \]

**Definition 1.1.7. Borel set**

A Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, and countable intersection.

## 1.2 Estimation

The purpose of inferential statistical conclusion of community properties sample drawn from it, when you use the sample data statistic to infer from the community because we don’t have all the facts, community urge for practical way we can trust the fact required within a given dependent on the nature of the desired community appreciation transactions parameter trying to access values numerical to community through sample data drawn from it at random. Statistical inference is divided into two sections:
• Statistical Estimation.

• Hypothesis testing.

The main purpose of this thesis is the first section, the statistical estimation. The pdf is a fundamental concept in statistics. Consider any random variable $X$ that has pdf $f$. Specifying the function $f$ gives a natural description of the distribution of $X$, and allows probabilities associated with $X$ to be found from the relation.

$$P(a < X < b) = \int_{a}^{b} f(x) \, dx,$$

for any real constants $a$ and $b$ with $a < b$.

**Definition 1.2.1. Estimator**[13] An estimator is a statistic, often expressed as a formula, that tells how to calculate the value of an estimate based on the measurements contained in sample.

**Definition 1.2.2.** [13]. Let $X$ be a random variable with pdf with parameter $\theta$. Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution of $X$ and let $\hat{\theta}$ denotes an estimator of $\theta$. We say $\hat{\theta}$ is an unbiased estimator of $\theta$ if

$$E(\hat{\theta}) = \theta.$$

If $\hat{\theta}$ is not unbiased, we say that $\hat{\theta}$ is a biased estimator of $\theta$.

**Example 1.2.1.** $S^2$ is unbiased estimator for $\sigma^2$. 
Proof.

\[
E(S^2) = E\left(\frac{1}{n-1} \sum (X_i - \bar{x})^2\right)
= \frac{1}{n-1} E(\sum (X_i - \bar{x})^2)
= \frac{1}{n-1} E(\sum X_i - \mu + \mu - \bar{x})^2
= \frac{1}{n-1} \left[ E(\sum (X_i - \mu)^2 - 2(\bar{x} - \mu) \sum (X_i - \mu) + \sum (\bar{x} - \mu)^2) \right]
= \frac{1}{n-1} \left[ nE(X_i - \mu)^2 - 2nE(\bar{x} - \mu)^2 + nE(\bar{x} - \mu)^2 \right]
= \frac{1}{n-1} \left[ n\sigma^2 - nE(\bar{x} - \mu)^2 \right]
= \frac{1}{n-1} \left[ n\sigma^2 - \frac{\sigma^2}{n} \right]
= \frac{\sigma^2}{n-1} \left[ n - 1 \right]
= \sigma^2
\]

\[\square\]

but \(S\) is biased estimator for \(\sigma\)

**Definition 1.2.3.** [13]. If \(\hat{\theta}\) is an unbiased estimator of \(\theta\) and

\[
\text{Var}(\hat{\theta}) = \frac{1}{nE \left[ \left( \frac{\partial \ln f(X_1)}{\partial \theta} \right)^2 \right]} \tag{1.2.1}
\]

then \(\hat{\theta}\) is called a minimum variance unbiased estimator (efficient) of \(\theta\).

**Example 1.2.2.** \(\bar{x}\) is a minimum variance unbiased estimator for \(\mu\) in normal population.

Proof.

\[
\text{Var}(\bar{x}) = \frac{\sigma^2}{n}
\]
\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)
\]

\[
\ln f(x) = -\ln \sigma - \frac{1}{2} \ln 2\pi - \frac{1}{2}(\frac{x-\mu}{\sigma})^2
\]

\[
\frac{\partial \ln f(x)}{\partial \mu} = -\frac{1}{2}(\frac{x-\mu}{\sigma^2})(\frac{-1}{\sigma}) = \frac{x-\mu}{\sigma^2}
\]

\[
E(\frac{x-\mu}{\sigma^2})^2 = \frac{1}{\sigma^4} E(x-\mu)^2 = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}
\]

\[
\frac{1}{nE(\frac{\partial}{\partial \mu} \ln f(x))^2} = \frac{1}{n(\frac{1}{\sigma^2})} = \frac{\sigma^2}{n}
\]

Definition 1.2.4. [13]. The statistic \( \hat{\theta}_n \) is a **consistent estimator** of the parameter \( \theta \) if and only if for each \( \varepsilon > 0 \)

\[
\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1.
\]

(1.2.2)

**Example 1.2.3.** \( \bar{X} \) is a consistent estimator for \( \mu \) in normal population.

**Proof.**

\[
\lim_{n \to \infty} P(|\bar{x} - \mu| < c) \geq \ln 1 - \frac{\sigma^2}{nc} = 1
\]

then,

\[
\lim_{n \to \infty} P(|\bar{x} - \mu| < c) = 1
\]

Hence, \( \bar{x} \) is consistent estimator for \( \mu \).

**Theorem 1.2.1.** [13]. If \( \hat{\theta}_n \) is an unbiased estimator of \( \theta \) and \( \text{Var}(\hat{\theta}) \to 0 \), as \( n \to \infty \),
then \( \hat{\theta}_n \) is a consistent estimator of \( \theta \).

**Example 1.2.4.** \( S^2 \) is consistent estimator for \( \sigma^2 \) in normal population.

Definition 1.2.5. [13]. The statistic \( \hat{\theta} \) is a **sufficient estimator** of the parameter \( \theta \) if, and only if for each value of \( \hat{\theta} \) the conditional probability distribution or density of the random sample \( X_1, X_2, \ldots, X_n \) given \( \hat{\theta} = \theta \) is independent of \( \theta \).
Example 1.2.5. $\bar{x}$ is sufficient estimator for $\mu$ in normal population.

There are two types of density estimation:

1. Parametric Estimation.


**Parametric Estimation**

The parametric approach for estimating $f(x)$ is to assume that $f(x)$ is a member of some parametric family of distributions, e.g. $N(\mu, \sigma^2)$, and then to estimate the parameters of the assumed distribution from the data. For example, fitting a normal distribution leads to the estimator

$$f_n(x) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(\frac{(x - \mu_n)^2}{2\sigma_n^2}\right), x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma_n > 0,$$

where,

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

and,

$$\sigma_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu_n)^2.$$

This approach has advantages as long as the distributional assumption is correct, or at least is not seriously wrong. It is easy to apply and it yields (relatively) stable estimates. The main disadvantage of the parametric approach is lack of exibility. Each parametric family of distributions imposes restrictions on the shapes that $f(x)$ can have. For example the density function of the normal distribution is symmetrical and bell-shaped, and therefore is unsuitable for representing skewed densities or bimodal densities.

**Methods of finding parametric Estimator**:

There are two main methods of parametric estimation, the method of moments and the method of maximum likelihood function.

**The Method of Moments**

In statistics, the method of moments is a method of estimating of population parameters such as mean, variance, median, etc. (which need not be moments ), by equating sample moments with unobservable population moments and then solving
those equations for the quantities to be estimated.

The Method of Maximum Likelihood

The maximum likelihood method which depends on finding the value of the unknown parameter $\theta$ that maximize the joint distribution $f(x_1, x_2, ..., x_n; \theta)$.

**Definition 1.2.6.** If $x_1, x_2, ..., x_n$ are the values of the random sample from a population with the parameter $\theta$, the likelihood function of the sample is given by

$$L(\theta) = f(x_1, x_2, ..., x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

as $x_1, x_2, ..., x_n$ are independent random the Maximum likelihood Method.

The Maximum likelihood Method for finding an estimator of $\theta$, consist of finding the estimator $\hat{\theta}$ which make the function $L(\hat{\theta})$ is maximum . That is to find $\theta$ by finding

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0.$$

**Example 1.2.6.** If $x_1, x_2, ..., x_n$ are the values of an independent random sample of size $n$, from the Bernoulli population.

$$f(x) = \theta^x(1 - \theta)^{1-x}, \; x = 0, 1, \; 0 < \theta < 1$$

$$L(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1 - \theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i}(1 - \theta)^{n - \sum_{i=1}^{n} x_i}$$

$$LnL(\theta) = \sum x_i \ln(\theta) + (n - \sum x_i) \ln(1 - \theta)$$

$$\frac{\partial LnL(\theta)}{\partial \theta} = \sum \frac{x_i}{\theta} - \frac{n - \sum x_i}{1 - \theta} = 0$$

$$\sum \frac{x_i}{\theta} = \frac{n - \sum x_i}{1 - \theta}$$

$$\sum x_i - \theta \sum x_i = n\theta - \theta \sum x_i$$

$$\sum x_i = n\theta$$

$$\hat{\theta} = \frac{\sum x_i}{n} = \bar{x}$$

**Non Parametric Estimation**

If the data that we study come from unknown distribution i.e. The density function $f(x)$ is unknown, then we must estimate the density function. This estimation
is called a nonparametric estimation, there are many nonparametric statistical objects of potential interest, including density functions (univariate and multivariate), density derivatives, conditional density functions, conditional distribution functions, regression functions, median functions, quantile functions, and variance functions. Many nonparametric problems are generalizations of univariate density estimation. For obtaining a nonparametric estimation of a pdf there many methods. Three common methods are:

- Histogram
- The Naive Estimator
- Kernel Density Estimation

The simplest form of non-parametric Density estimation is the histogram, where the histogram requires two parameters to be defined bin width and starting position of the first bin, divide the sample space into a number of bins and approximate the density at the center of each bin by the fraction of points in the training data that fall into the corresponding bin. Histogram has been widely used to estimate the density function. Given an origin $x_0$ and a bandwidth $h$, we define the bins of the histogram to be the intervals $[x_0 + mh, x_0 + (m + 1)h], m \in Z - \{0\}$. The intervals have been chosen closed on the left and open on the right for definiteness.

**Definition 1.2.7. Histogram Estimator**[31] Let $X_1, X_2, ..., X_n$ be a random sample from unknown pdf $f(x)$, the histogram estimator of the density function $f(x)$ is defined by:

$$f_n(x) = \frac{1}{nh} (\text{no. of } X_i \text{ in same bin as } x).$$

(1.2.3)

Histogram is a very simple form of density estimation, but has several weakness:

1. The density estimate depends on the starting position of the bins For multivariate data, the density estimate is also affected by the orientation of the bins.

2. The discontinuities of the estimate are not due to the underlying density, they are only an artifact of the chosen bin locations these discontinuities make it very difficult to grasp the structure of the data.
3. A much more serious problem is the curse of dimensionality, since the number of bins grows exponentially with the number of dimensions, in high dimensions we would require a very large number of examples or else most of the bins would be empty.

4. These issues make the histogram unsuitable for most practical applications except for quick visualizations in one or two dimensions.

To overcome the weakness of the histogram method there is another method which is naive estimator.

**The Naive Estimator.**

Refinement of the histogram method, is the naive estimator, it is equivalent to a histogram where the estimation point \( x \) is used as the center of the bin, and the bin width is \( 2h \). If the random variable \( X \) has density function \( f \), then

\[
f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h).
\]

For any given \( h \), we can estimate \( P(x - h < X < x + h) \) by the proportion of the sample falling in the interval \((x - h, x + h)\). Thus a natural estimator \( f_n \) of the density function is given by choosing a small number \( h \) and setting

\[
f_n(x) = \frac{1}{2nh} \text{(no. of } X_i \text{ falling in } (x - h, x + h)).
\]

(1.2.4)

This estimator is called the Naive estimator.

We can rewrite Equation (1.2.4) by the weighting function \( w \):

\[
w(x) = \begin{cases} 
\frac{1}{2}, & |x| < 1; \\
0, & \text{otherwise};
\end{cases}
\]

(1.2.5)

Using this notation, we can express the naive estimator as

\[
f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} w(\frac{x - X_i}{h}).
\]

where \( X_i \) are the data samples.

The Naive have some disadvantage:

- It is continues but has jumps at the points of \((x_i \pm h)\).
• It has zero derivative everywhere.

The kernel gives different weights a wording to the proximity at the data around the point \( x \). Refinement of naive estimator by replacing the weight function \( W \) by kernel function \( K \).

**The Kernel Estimator.**

In statistics, kernel Density Estimation (KDE) is a non-parametric way to estimate the (pdf) of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made based on a finite data sample.

### 1.3 Kernel Density Estimation of the Pdf

We present the kernel density estimation of the pdf and review some important definitions and aspects in this area. In statistics, kernel Density Estimation (KDE) is a non-parametric way to estimate the (pdf) of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made based on a finite data sample.

**Definition 1.3.1. Kernel Estimator\[^{[31]}\]** Let \( X_1; X_2; \ldots; X_n \) be a random sample from unknown pdf \( f(x) \), the kernel estimator of the density function \( f(x) \) is defined by Rosenblatt(1956)\[^{[25]}\] generalize Naive estimator to the Kernel form

\[
f_{nh}(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.3.1}
\]

where \( K(.) \) is called a kernel function.

**Definition 1.3.2. (Kernel Estimator of a Probability Density Function)\[^{[28]}\]**

Suppose that \( X_1, \ldots, X_n \) is a random sample of data from an unknown continuous distribution with pdf \( f(x) \) and cumulative distribution function (cdf) \( F(x) \), the kernel estimator of a probability density function is defined as

\[
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.3.2}
\]
where the bandwidth $h = h_n$ is a sequence of positive numbers that converging to zero and $K(\cdot)$ is the kernel function considers to be both symmetric and satisfies

$$\int_{-\infty}^{\infty} K(x)dx = 1$$

The density estimates derived using such kernels can fail to be probability densities, because they can be negative for some values of $x$. Typically, $K$ is chosen to be a symmetric pdf. There is a large body of literature on choosing $K$ and $h$ well, where ”well” means that the estimate converges asymptotically as rapidly as possible in some suitable norm on pdf.

A slightly more compact formula for the kernel estimator can be obtained by introducing the recalling notation $K_h(u) = h^{-1}K(u/h)$. This allows us to write

$$\hat{f}(x; h) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i).$$

### 1.4 Kernel Density Estimation of the Cdf

In this section, we present the kernel estimator for the cdf $\hat{F}(x)$.

**Definition 1.4.1.** The kernel estimator of the cdf is defined by :

$$\hat{F}(x) = \int_{-\infty}^{x} \hat{f}(u)du = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{x} K(\frac{u - X_i}{h})du.$$

**Lemma 1.4.1.** The Mean-Squared Error of $\hat{f}(x)$ is given by :

$$MSE(\hat{f}(x)) = \frac{\mu_r^2(K)}{(r!)^2} f(x)^2 h^{2r} + \frac{f(x)R(K)}{nh} \tag{1.4.1}$$

where,

$$\mu_j(K) = \int_{-\infty}^{\infty} u^j K(u)du,$$

$r$ is the order of the kernel and

$$R(K) = \int_{-\infty}^{\infty} K^2(u)du.$$

**Proof.** See[23] \qed
Remark 1.4.1. In Equation (1.4.1), the first term (squared bias) is increasing in $h$ and the second term (the variance) is decreasing in $nh$ and hence to make the $MSE(\hat{f}(x))$ to decline as $n \to \infty$ we have to make both of these terms small, which meaning that as $n \to \infty$ we must have $h \to 0$ and $nh \to \infty$. That is, the bandwidth $h$ must decrease, but not at a rate faster than sample size $n$.

Remark 1.4.2. Consider that the kernels are of the second order $r = 2$ and the assumptions(C) that we will need are summarized below :

(C1) The unknown density function $f(x)$ has continuous second derivative $f^{(2)}(x)$.
(C2) The bandwidth $h = h_n$ is a sequence of positive numbers and satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$ (see Remark 1.4.1).
(C3) The kernel $K$ is a bounded probability density function of order 2 and symmetric about the zero.

By lemma 1.4.1, Definition 1.4.1 and under the assumptions $C_1$, $C_2$ and $C_3$ we have $\hat{F}(x) \overset{P}{\to} F(x)$.

1.5 Properties of the Kernel Estimator

In this section, we will introduce some important properties of the kernel. A kernel is a piecewise continuous function, symmetric around zero, even function and integrating to one, i.e.

$$K(x) = K(-x), \int_{-\infty}^{\infty} K(x)dx = 1.$$ 

The kernel function need not have bounded support and in most applications $K$ is a positive pdf.

Definition 1.5.1. [5] A kernel function $K$ is said to be of order $p$, if its first nonzero moment is $\mu_p \neq 0$, i.e. if

$$\mu_j(K) = 0, j = 1, 2, ..., p - 1; \mu_p(K) \neq 0;$$

where

$$\mu_j(K) = \int_{-\infty}^{\infty} y^j K(y)dy.$$ (1.5.1)
we consider the following conditions:

- The unknown density function $f(x)$ has continuous second derivative $f''(x)$
- The bandwidth $h = h_n$ satisfies $\lim_{n \to \infty} h = 0$, and $\lim_{n \to \infty} nh = \infty$
- The kernel $K$ is a bounded pdf of order 2 and symmetric about the origin. $\int_{-\infty}^{\infty} zK(z)dz = 0$, and $\int_{-\infty}^{\infty} z^2K(z)dz \neq 0 < \infty$.

**Definition 1.5.2.** [5] The Bias of an estimator $f_n(x)$ of a density $f(x)$ is the difference between the expected value of $f_n(x)$ and $f(x)$.

**Theorem 1.5.1.** Let $X$ be a random variable having a density $f$; then the bias of $\hat{f}(x)$ can be expressed as

$$E\hat{f}(x; h) - f(x) = \frac{1}{2}h^2\mu_2(K)f''(x) + o(h^2).$$  \hspace{1cm} (1.5.2)

where $\int K(z)dz = 1$, $\int zK(z)dz = 0$, $\int z^2K(z)dz < \infty$, and $\mu_2(K) = \int z^2K(z)dz$.

**Proof.**

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

then,

$$E\hat{f}(x, h) = E\left(\frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right)$$

$$= \frac{1}{nh} \sum_{i=1}^{n} E K\left(\frac{x - X_i}{h}\right)$$

$$= \frac{1}{nh} nE K\left(\frac{x - X}{h}\right)$$

$$= \frac{1}{h} \int K\left(\frac{x - y}{h}\right)f(y)dy.$$  

(when we take $\frac{x - y}{h} = z$ then we have $x - y = hz$ then $y = x - hz$ and $dy = -hdz$)

hence,

$$E\hat{f}(x, h) = \frac{1}{h} \int K(z)f(x - hz)(-h)dz$$

$$= \int K(z)f(x - hz)dz.$$
(Now expanding \(f(x - hz)\) in a Taylor series about \(x\) to obtain

\[
f(x - hz) = f(x) - hz f'(x) + \frac{1}{2} h^2 z^2 f''(x) + o(h^2)
\]

\[
E(\hat{f}(x, h)) = \int K(z)[f(x) - hz f'(x) + \frac{1}{2} h^2 z^2 f''(x)]dz + o(h^2)
\]

\[
= f(x) \int K(z)dz - hf'(x) \int zK(z)dz + \frac{h^2}{2} f''(x) \int z^2dz + o(h^2)
\]

This leads to:

\[
E\hat{f}(x; h) = f(x) + \frac{1}{2} h^2 f''(x) \int z^2 K(z)dz + o(h^2)
\]

then we have,

\[
E\hat{f}(x; h) - f(x) = \frac{1}{2} h^2 \mu_2(K) f''(x) + o(h^2).
\]

Notice that the bias is of order \(h^2\) which implies that \(\hat{f}(x; h)\) is asymptotically unbiased.

Now, we will find \(\text{Var}(\hat{f}(x; h))\)

**Theorem 1.5.2.** Let \(X\) be a random variable having a density \(f\); then

\[
\text{Var}(\hat{f}(x)) = ((nh)^{-1})R(K)f(x) + o((nh)^{-1})
\]

where, \(R(K) = \int_{-\infty}^{\infty} K^2(x)dx\).

**Proof.**

\[
\text{Var}(\hat{f}(x; h)) = \text{Var}((nh)^{-1} \sum K(\frac{x - X_i}{h}))
\]

\[
= (n^2h^2)^{-1} \sum \text{Var}(K(\frac{x - X_i}{h}))
\]

\[
= \frac{1}{nh^2} \text{Var}(K(\frac{x - X_i}{h}))
\]

\[
= \frac{1}{nh^2} \left(E(K^2(\frac{x - X_i}{h})) - (EK(\frac{x - X_i}{h}))^2\right)
\]

\[
= \frac{1}{nh^2} E(K^2(\frac{x - X}{h})) - \frac{1}{n} \hat{f}(x; h)^2
\]

\[
= \frac{1}{nh^2} \int K^2(\frac{x - u}{h})f(u)du - \frac{1}{n} \hat{f}(x; h)^2
\]
(now, let $\frac{x-u}{h} = z$ then $hz = x - u$ then $u = x - hz$, $du = -hdz$)

Then we have

$$Var \hat{f}(x; h) = (nh)^{-1} \int K(z)^2 f(x - hz) dz - n^{-1} \{E \hat{f}(x; h)\}^2$$

$$= (nh)^{-1} \int K(z)^2 \{f(x) + o(1)\} dz - n^{-1} \{f(x) + o(1)\}^2$$

$$= (nh)^{-1} \int K(z)^2 dz f(x) + o\{(nh)^{-1}\}.$$

$$\approx (nh)^{-1} R(K) f(x) + o((nh)^{-1})$$

\[\square\]

### 1.6 Optimal Bandwidth

The problem of selection the bandwidth is very important in kernel density estimation. Choice of appropriate bandwidth is critical to the performance of most nonparametric density estimators. When the bandwidth is very small, the estimate will be very close to the original data. The estimate will be almost unbiased, but it will have large variation under repeated sampling. If the bandwidth is very large, the estimate will be very smooth, lying close to the mean of all the data. Such an estimate will have small variance, but it will be highly biased. There are many rules for bandwidth selection, for example Normal Scale Rules, Over-smoothed bandwidth selection rules, Least Squares Cross-Validation, Biased Cross-Validation, Estimation of density function- als and Plug-In Bandwidth Selection. For more details see Leove, Silverman and Wand.

we shall use two types of errors criteria. The mean square error (MSE) is used to measure the error when estimating the density function at a single point. It is defined by

$$MSE\{f_n(x)\} = E\{f_n(x) - f(x)\}^2. \quad (1.6.1)$$
We can write the MSE as a sum of the squared bias and the variance at $x$,

$$MSE(f_n(x)) = \{Ef_n(x) - f(x)\}^2 + Var(f_n(x)). \quad (1.6.2)$$

A second type of criteria measures the error when estimating the density over the whole real line. The most well known of this type is the mean integral square error (MISE) introduced by Rosenblatt (1956)[25]. The MISE is defined as

$$MISE(f_n) = E\int_{-\infty}^{\infty} \{f_n(x) - f(x)\}^2 dx. \quad (1.6.3)$$

By changing the order of integration we have,

$$MISE(f_n) = \int_{-\infty}^{\infty} MSE\{f_n(x)\} dx = \int_{-\infty}^{\infty} \{Ef_n(x) - f(x)\}^2 dx + \int_{-\infty}^{\infty} Var(f_n(x)) dx. \quad (1.6.4)$$

Equation (1.6.4) gives the MISE as a sum of the integral squared bias and the integral variance. Substituting (1.5.2) and (1.5.4) we conclude that

$$MISE(f_n) = AMISE(f_n) + o\{h^4 + (nh)^{-1}\}, \quad (1.6.5)$$

where AMISE is the asymptotic mean integral squared error given by

$$AMISE(f_n) = \frac{1}{4} h^4 \mu_2(K)^2 R(f^{(2)}) + (nh)^{-1} R(K), \quad (1.6.6)$$

see Wand and Jones (1995)[31].

The natural way for choosing $h$ is to plot out several curves and choose the estimate that best matches one prior (subjective) ideas. However, this method is not practical in pattern recognition since we typically have high-dimensional data. Assume a standard density function and find the value of the bandwidth that minimizes the integral of the square error (MISE)

$$h_{MISE} = \arg\min E[\int (f_n(x) - f(x))^2 dx]. \quad (1.6.7)$$

If we assume that the true distribution is Gaussian and we use a Gaussian kernel, the bandwidth $h$ is computed using the following equation from Silverman.

$$h^* = 1.06SN^{-\frac{1}{2}} \quad (1.6.8)$$
where $S$ is the sample standard deviation and $N$ is the number of training examples.

By differentiating (1.6.6) with respect to $h$ we can find the optimal bandwidth with respect to AMISE criterion. This yields, the optimal bandwidth is given by

$$h_{op} = \left[ \frac{R(K)}{\mu_2(K)^2 R(f^n)n} \right].$$

(1.6.9)

Therefore if we substitute (1.6.9) into (1.6.6), we obtain the smallest value of AMISE for estimating $f$ using the kernel $K$.

$$\inf_{h>0} AMISE\{f_n\} = \frac{5}{4} \left\{ \mu_2 K^2 R(K)^4 R(f^n) \right\}^{\frac{1}{2}} n^{-\frac{4}{5}}.$$

Notice that in 1.6.9 the optimal bandwidth depends on the unknown density being estimated, so we can not use (1.6.9) directly to find the optimal bandwidth $h_{opt}$.

Also from 1.6.9 we can conclude the following useful conclusions.

- The optimal bandwidth will converge to zero as the sample size increases, but at very slow rate.

- The optimal bandwidth is inversely proportional to $R(f^n)^{\frac{1}{2}}$. Since $R(f^n)$ measures the curvature of $f$, this means that for a density function with little curvature, the optimal bandwidth will be large. Conversely, if the density function has a large curvature, the optimal bandwidth will be small.

In this chapter, we introduced some basic definitions and theorems that will need it in this thesis, and we studied definition of estimation, it’s types, and it’s common methods, then we presented kernel density estimation of the pdf, kernel density estimation of the cdf, properties of the kernel estimator, and also we studied the optimal bandwidth. In the next chapter, we will study the univariate conditional quantiles estimation.
Chapter 2
Univariate Conditional Quantile
Chapter 2

Univariate Conditional Quantiles

This chapter consists of four sections. In Section 2.1, we introduce the important of the quantile, and give historical notes. In the next section, we present the Nadaraya-Watson estimator. Section 2.3 talking about the estimation of the univariate conditional quantile. In Section 2.4, we present the asymptotic properties of the Nadaraya-Watson estimator of the cdf.

2.1 Importance of Quantiles

In statistics and the theory of probability, quantiles are outpoints dividing the range of a probability distribution into continuous intervals with equal probabilities, or dividing the observations in a sample in the same way. There is one less quantile than the number of groups created. Thus quartiles are the three cut points that will divide a data set into four equal-size groups. Common quantiles have special names: for instance quartile, decile (creating 10 groups). The groups created are termed halves, thirds, quarters, etc., though sometimes the terms for the quantile are used for the groups created, rather than for the cut points. $q$—Quantiles are values that partition a finite set of values into $q$ subsets of (nearly) equal sizes. There are $q-1$ of the $q$—quantiles, one for each integer $k$ satisfying $0 < k < q$. In some cases the value of a quantile may not be uniquely determined, as can be the case for the median (2-quantile) of a uniform probability distribution on a set of even size.

Quantile estimation is of interest in many application settings For example, in
computing tables of critical values associated with complicated hypothesis tests in which $F$ cannot be computed analytically in closed form, it may be necessary to resort to Monte Carlo methodology as a means of calculating such critical values, these values are defined as quantiles of an appropriately defined test statistic $x$. A second setting in which quantile estimation arises naturally is in the manufacturing context, in which a supplies may wish to compute a "promise data" by which the company can guarantee with high probability, delivery of the required product to its customers. Such a computation involves calculating an appropriately defined quantile associated with the (random) time required to process an order, from the instant of order placement to order fulfillment [14].

Quantiles are very important statistics information used to describe the distribution of datasets. Given the quantiles of a dataset, we can easily know the distribution of the dataset, which is a fundamental problem in data analysis. However, quite often, computing quantiles directly is inappropriate due to the memory limitations. Further, in many settings such as data streaming and sensor network model, even the data size is unpredictable. Although the quantiles computation has been widely studied, it was mostly in the sequential setting [33].

The median is the best example of a quantile, the sample median can be defined as the middle value, i.e the sample median splits the data into two parts with an equal number of data points in each. Let $Y$ be a random variable with cdf $F_Y(y)$, usually, the sample median is taken as an estimator of the population median $m$; a quantity which splits the distribution into two halves in the sense that, $P(Y \leq m) = P(Y \geq m) = \frac{1}{2}$. In particular, for a continuous random variable, $m$ is a solution to the equation $F_Y(y) = \frac{1}{2}$.

More generally, the 25% and 75% sample quartiles can be defined as values that split the data in proportion of one and three-quarters, and vice versa. Similarly in the continuous case, the population lower quartile and upper quartile are the solutions to the equations $F(y) = \frac{1}{4}$ and $F(y) = \frac{3}{4}$ respectively. Generally, for a proportion $\alpha$, $(0 < \alpha < 1)$, in the continuous case, the 100$\alpha$% quantile of $F$ is the value $y$ which solves $F(y) = \alpha$. Note that we assume this value is unique [27].
2.2 The Nadaraya-Watson Estimator

In this section, we will study some basic information about Nadaraya-Watson (N-W) estimator to use this later. It is a popular nonparametric method for estimating the conditional density function \( f(y|x) \). We will consider the kernel estimation of the conditional cumulative distribution function (cdf).

let \((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\) be a random sample from a distribution with a conditional probability density function (pdf) \( f(y|x) \) then the cdf \( F(y|x) \) is given by:

\[
F(y|x) = \int_{-\infty}^{y} f(u|x) \, du,
\]

where

\[
f(y|x) = \frac{f(x,y)}{f_X(x)}.
\]

To estimate \( f(y|x) \), the numerator and denominator separability must be estimated by using kernel estimators.

\[
\hat{f}(y|x) = \frac{\hat{f}(x,y)}{\hat{f}_X(x)}
\]

\[
= \frac{\frac{1}{n} \sum_{i=1}^{n} K_h(x-X_i)K_h(y-Y_i)}{\frac{1}{n} \sum_{i=1}^{n} K_h(x-X_i)}
\]

\[
= \frac{\sum_{i=1}^{n} K_h(x-X_i)K_h(y-Y_i)}{\sum_{i=1}^{n} K_h(x-X_i)},
\]

where \( K \) is a kernel function and \( h \) is sequence of positive number converging to zero and it is called bandwidth, and \( K_h(x) = K(x/h)/h \). Now the estimator of the conditional cdf is given by:

\[
\hat{F}(y|x) = \int_{-\infty}^{y} \hat{f}(u|x) \, du
\]

\[
= \frac{\sum_{i=1}^{n} K_h(x-X_i) \int_{-\infty}^{y} K_h(u-Y_i) \, du}{\sum_{i=1}^{n} K_h(x-X_i)}
\]

Now, there are two ways to estimate the conditional cdf \( F(y|x) \). Firstly, by using the indicator function \( I(Y_i \leq y) \), which is called the Nadaraya-Watson estimator,

\[
\hat{F}_{NW}(y|x) = \sum_{i=1}^{n} w_i I(Y_i \leq y),
\] (2.2.1)
where $w_i$ is the nonzero weight function and given by,

$$w_i(x) = \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)},$$

secondly, we can estimate $F(y|x)$ by using equation (2.2.1) by replacing the indicator function by the continuous distribution function $\Omega(\frac{Y_i - y}{h_2})$, which is called the double kernel estimator

$$\hat{F}_{DK}(y|x) = \sum_{i=1}^{n} w_i(x)\Omega(\frac{Y_i - y}{h_2}),$$

where

$$\Omega(y) = \int_{-\infty}^{y} W(u)du$$

is a distribution function with associated density function $W(u)$.

**Remark 2.2.1.**

$$0 \leq \hat{F}_{NW}(y|x) \leq 1,$$

because, if $y \leq Y_i$ for all $i = 1, 2, ..., n$ then $I(Y_i \leq y) = 0$ for all $i$, then,

$$\hat{F}_{NW}(y|x) = \sum_{i=1}^{n} w_i(x)I(Y_i \leq y) = 0.$$

If $y$ lies between the $Y_i$'s, i.e. there are some $Y_i$ less than or equal $y$ but not at all, then $I(Y_i \leq y) = 0$ for some $i$, and $I(Y_i \leq y) = 1$ for the other, then

$$\hat{F}_{NW}(y|x) = \sum_{i=1}^{n} w_i(x)I(Y_i \leq y) < 1$$

If $y \geq Y_i$ for all $i$ then $I(Y_i \leq y) = 1$ for all $i$ then

$$\hat{F}_{NW}(y|x) = 1$$

### 2.3 Estimating the Univariate Conditional Quantile

In this section the univariate conditional quantile and its kernel estimation by using the NW estimator, will be studied.
Definition 2.3.1. (The Quantile)  
Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a distribution with cdf $F_Y(y)$ then the $\alpha$-th quantile is denoted by $q_\alpha$ where,

$$q_\alpha = \inf\{y : F_Y(y) \geq \alpha\} = F_Y^{-1}(\alpha)$$

Definition 2.3.2. (The Conditional Quantile)  
Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample from a distribution with a conditional cdf $F(y|x)$, then the $\alpha$-th conditional quantile $q_\alpha(x)$, is defined by

$$q_\alpha(x) = \inf\{y : F(y|x) \geq \alpha\},$$

where, $0 < \alpha < 1$.

Note that when $\alpha = 0.5$, we have the definition of the median.

Now, to estimate $q_\alpha(x)$, we need to estimate $F(y|x)$ by using the NW estimator of the cdf.

Definition 2.3.3. The Nadaraya-Watson Estimator of the cdf[27]  
The distribution function $F(y|x)$ may be estimated nonparametrically by,

$$\hat{F}_{NW}(y|x) = \sum_{i=1}^{n} w_i(x) I(Y_i \leq y),$$

where $I$ is the indicator function and the weight function given by,

$$w_i(x) = \frac{K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)},$$

where, $K$ is a kernel function and $h$ is a sequence of positive constants called the bandwidth.

Now, the estimator for the conditional quantile is given by,

Definition 2.3.4. [27]  
The estimator $q_{\alpha,n}(x)$ for the conditional quantile $q_\alpha(x)$ is defined by,

$$q_{\alpha,n}(x) = \arg\min_{\alpha} \sum_{i=1}^{n} \rho_\alpha(Y_i - a) K\left(\frac{x - X_i}{h}\right),$$

where $\rho_\alpha$ is the check function, and its given by,

$$\rho_\alpha(u) = \alpha u I_{[0,\infty)}(u) - (1 - \alpha) u I_{(-\infty,0)}(u) = \frac{|u| + (2\alpha - 1)u}{2}.$$
When we taking $\alpha = \frac{1}{2}$ gives $\rho_{0.5}(u) = \frac{|u|}{2}$.

**Lemma 2.3.1.** For some $\alpha$, $0 < \alpha < 1$, we have

$$\rho_{\alpha}(u) = \alpha u I_{[0, \infty)}(u) - (1 - \alpha)u I_{(-\infty, 0)}(u) = \frac{|u| + (2\alpha - 1)u}{2}$$

**Proof.** The left side:

$$\alpha u I_{[0, \infty)}(u) - (1 - \alpha)u I_{(-\infty, 0)}(u) = \begin{cases} 
\alpha u, & u \geq 0; \\
(\alpha - 1)u, & u < 0.
\end{cases}$$

The right side:

$$\frac{|u| + (2\alpha - 1)u}{2} = \begin{cases} 
\frac{u + (2\alpha - 1)u}{2} = \frac{2\alpha u}{2} = \alpha u, & u \geq 0; \\
\frac{-u + (2\alpha - 1)u}{2} = \frac{2(\alpha - 1)u}{2} = (\alpha - 1)u, & u < 0.
\end{cases}$$

Then the left side equal the right side. \(\square\)

### 2.4 Asymptotic Properties of the N-W Estimator of the Cdf

In this section the asymptotic properties of the NW estimator of the cdf are discussed. We will prove The bias of the NW estimator of the cdf,

$$\hat{F}(y|x) = \int_{-\infty}^{y} \hat{f}(y|x)dy$$

where,

$$\hat{f}(y|x) = \frac{1}{nh^2} \sum_{i=1}^{n} K(\frac{x - X_i}{h}) K(\frac{y - Y_i}{h})$$

$$= \frac{1}{h} \sum_{i=1}^{n} w_i(x) K(\frac{y - Y_i}{h})$$

then,
\[ E\hat{f}(y|x) = \frac{1}{h} \sum_{i=1}^{n} w_i(x) EK\left(\frac{y-Y_i}{h}\right) \]
\[ = \frac{1}{h} \sum_{i=1}^{n} w_i(x) \int K\left(\frac{y-u}{h}\right) f(u)du \]
\[ = \frac{1}{h} \sum_{i=1}^{n} w_i(x) \int K(z) f(y-hz)(-h)dz \]
\[ = \sum_{i=1}^{n} w_i(x) \int K(z) f(y-hz)dz \]

Expanding \( f(y-hz) \) by using Taylor series about \( y \) we obtain,
\[ f(y-hz) = f(y) - hzf'(y) + \frac{1}{2}h^2z^2f''(y) + o(h^2) \]
then we have,
\[ E\hat{f}(y|x) = \sum_{i=1}^{n} w_i(x) \int K(z)[f(y) - hzf'(y|x) + \frac{1}{2}h^2z^2f''(y|x) + o(h^2)] \]
\[ = \sum_{i=1}^{n} w_i(x)[f(y) \int K(z)dz - hf'(y|x) \int zK(z)dz + \frac{1}{2}h^2f''(y|x) \int z^2K(z)dz + o(h^2)] \]
this leads to;
\[ E\hat{f}(y|x) = \sum_{i=1}^{n} w_i(x)[f(y) + \frac{1}{2}h^2f''(y|x) \int z^2K(z) + o(h^2)], \]
where we have used, \( \int K(z)dz = 1 \), \( \int zK(z)dz = 0 \), and \( \int z^2K(z)dz < \infty \), let \( \mu_2(K) = \int z^2K(z)dz \), then we have,
\[ E\hat{f}(y|x) = \sum_{i=1}^{n} w_i(x)[f(y|x) + \frac{1}{2}h^2f''(y|x)\mu_2 + o(h^2)], \]
this implies that,
\[ E\hat{f}(y|x) - f(y|x) = \frac{1}{2}h^2f''(y|x)\mu_2 + o(h^2). \]
Now,
\[ E\hat{F}(y|x) = \int_{-\infty}^{\infty} [f(y|x) + \frac{1}{2}h^2f''(y|x)\mu_2 + o(h^2)]dy \]
\[ = F(y|x) + \frac{1}{2}h^2\mu_2f'(y|x) + o(h^2). \]
this implies that,

\[
E \hat{F}(y|x) - F(y|x) = \frac{1}{2} h^2 \mu_2 f'(y|x) + o(h^2).
\]

**Theorem 2.4.1.** [9] Let \( Y_i \) be an independent random variables. The variance of the estimator \( \hat{F}_n(y|x) \) is given by

\[
\text{Var}[\hat{F}_n(y|x)] = \sum_{i=1}^{n} \frac{K^2(\frac{x-X_i}{h_n})}{[\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})]^2} |F(Y_i) - F^2(Y_i)|.
\]

**Theorem 2.4.2.** [9] Let \((X_1, Y_1), ..., (X_n, Y_n)\) be a random sample from a distribution with pdf \( f(y|x) \) and cdf \( F(y|x) \) and if the following conditions are satisfied:

1. \( h_n \) is sequence of positive number satisfies the following:
   - \( h_n \to 0; \) for \( n \to \infty \);
   - \( nh_n \to \infty; \) for \( n \to \infty \)

2. The kernel \( K \) is a Borel function and satisfies the following:
   - \( K \) has a compact support;
   - \( K \) is symmetric;
   - \( K \) is Lipschitz-continuous;
   - \( \int K(u)du = 1; \)
   - \( K \) is bounded;

3. For a fixed \( y \in \mathbb{R} \) there exists \( F''(y|x) = \frac{\partial^2 F(y|x)}{\partial x^2} \) in a neighborhood of \( x \):

   then it holds for \( n \to \infty \) and \( x \in (h_n, 1 - h_n) \):

\[
\text{MSE}(\hat{F}_n(y|x)) \approx \left[ \frac{h^2}{2} F''(y|x) \int u^2 K(u)du \right]^2 + \frac{1}{nh_n} (F(y|x) - F''(y|x)) \int K^2(u)du
\]

**Proof.** see [9]  

\[29\]
Theorem 2.4.3. [9] Let the conditions of the last theorem be satisfied and let
\( F_{n,x}(q_{a,n}(x)) = F_x(q_a(x)) = \alpha \) be unique. Then it hold;

\[
MSE[q_{a,n}(x)] = \left[ \frac{1}{2} h_n^2 \left( \frac{F_{n,x}^{2,0}(q_{a,n}(x)|x)}{f(q_a(x)|x)} \right) \int u^2 K(u)du \right]^2 
+ \frac{1}{n h_n f^2(q_a(x))} \int K^2(u)du.
\]

Proof. see [9] \qed

Theorem 2.4.4. [9] Let the condition of the last theorem be satisfied and let \( nh_n^5 \to 0 \), for \( n \to \infty \). Then it holds

\[
(nh_n)^{\frac{1}{2}} (q_{a,n}(x) - q_a(x)) \overset{d}{\to} N \left( \frac{1}{2} (nh_n)^{\frac{1}{2}} \frac{F_{n,x}^{2,0}(q_{a,n}(x)|x)}{f(q_a(x)|x)} \int u^2 K(u)du, \frac{\alpha(1 - \alpha)}{f^2(q_a(x)|x) \int K^2(u)du} \right);
\]

\[
\overset{d}{\to} N(0, \frac{\alpha(1 - \alpha)}{f^2(q_a(x)|x) \int K^2(u)du})
\]

Proof. see [9] \qed

In this chapter, we presented precis of important of the quantile, and we studied the Nadaraya-Watson estimator, then we presented estimating the univariate conditional quantile. Also we studied the asymptotic properties of the N-W estimator of the cdf. In the next chapter we will generalized the results of this chapter for the case of multivariate conditional quantile.
Chapter 3
Multivariate Conditional Quantiles
Chapter 3

Multivariate Conditional Quantiles

This chapter is the main chapter of the thesis. In this chapter, we will introduce the multivariate conditional quantiles, multivariate quantiles based on a norm minimization, and we will prove the consistency of the multivariate Nadaraya-Watson estimator. This chapter consists of six sections. In Section 3.1, we introduce the multivariate conditional quantile. In the next two section, we present the mean as a minimization problem, and the quantile as a minimization problem. Section 3.4 talk about multivariate quantile based on a norm minimization. In section 3.5, we will study the Nadaraya-Watson estimator of the multivariate conditional quantile. Section 3.6 comprises consistency of the multivariate Nadaraya-Watson estimator.

3.1 Important of Multivariate Conditional Quantile

Multivariate quantiles have been defined by a number of researchers and can be estimated by different methods. Quantiles play an important role in statistical analysis of many areas such as economics, finance, and coastal engineering. The problem is often to estimate the quantiles of a variable conditional on the values of other variables. There have been several approaches to quantile functions for multivariate distributions. Multivariate nonparametric density estimation is an often used pilot tool for examining the structure of data multivariate time series arise when several
time series are observed simultaneously over time. A multivariate time series consists of multiple single series referred to as components, see Tsay (2002) [30] and De Gooijer, et al. (2004)[8]. When the individual series are related to each other, there is a need for jointly analyzing the series rather than treating each one separately. By so doing, one hopes to improve the accuracy of the predictions by utilizing the additional information available from the related series in the predictions of each other. Our approach of solving prediction problems is via generalizing the well-known univariate conditional quantile definition into a multivariate setting. we introduce a multivariate conditional quantile notion which extends the definition of quantiles by Abdous and Theodorescu (1992)[1]. We also propose a nonparametric estimator for our quantile definition which is shown to be consistent for \( \alpha \)-mixing processes. There are many approaches to extend the univariate quantile to the multivariate case:

1. Multivariate quantile functions based on depth functions.
2. Multivariate quantiles based on norm minimization.
3. Multivariate quantiles as inverse mappings.
4. Data-based multivariate quantiles based on gradients.

In this chapter, we will study only the second point in Section 3.4

### 3.2 The Mean as A minimization Problem

In this section, we will show that the sample mean can be defined as a problem of minimization a sum of squared residuals.

**Definition 3.2.1.** [10] Let \( X \) and \( Y \) be a continuous random variables with joint pdf \( f(x,y) \). Then the regression mean function or (conditional mean) of \( Y \) given \( X = x, E(Y|X = x) \), is defined as follows

\[
m(x) = E(Y|X = x) = \int_{-\infty}^{\infty} yf(y|x)dy = \frac{\int_{-\infty}^{\infty} yf(x,y)dy}{\int_{-\infty}^{\infty} f(x,y)dy}
\]
The regression mean function estimator of $Y$ given $X = x, \hat{m}(x)$, is defined as follows

$$\hat{m}(x) = \int_{-\infty}^{\infty} y \hat{f}(y|x)dy = \frac{\int_{-\infty}^{\infty} y \hat{f}(x,y)dy}{\int_{-\infty}^{\infty} \hat{f}(x,y)dy}.$$ 

**Theorem 3.2.1.** Suppose we have a random sample $Y_1, Y_2, ..., Y_n$, the unconditional mean $E(Y) = \mu$, can be obtained as the minimization of the following

$$\arg \min_{b \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - b)^2$$

when $b = \mu$

**Proof.** [6] Suppose we measure the distance between a random sample $Y_i$ and a constant $b$ by $(Y_i - b)^2$. The closer $b$ to $Y_i$, the smaller this quantity is. We can now determine the value of $b$ that minimizes $E(Y_i - b)^2$ and, hence, will provide us with a good predictor of $Y_i$.

$$E(Y_i - b)^2 = E(Y_i - EY_i + EY_i - b)^2$$

$$= E((Y_i - EY_i) + (EY_i - b))^2$$

$$= E((Y_i - EY_i)^2 + (EY_i - b)^2 + 2E((Y_i - EY_i)(EY_i - b)),$$

where we have expanded the square.

Now, note that

$$E((Y_i - EY_i)(EY_i - b)) = (EY_i - b)E(Y_i - EY_i) = 0,$$

since $(EY_i - b)$ is constant and comes out of the expectation, and $E(Y_i - EY_i) = EY_i - EY_i = 0$. This means that

$$E(Y_i - b)^2 = E((Y_i - EY_i)^2 + (EY_i - b)^2,$$

we have no control over the first term on the right-hand side, and the second term, which is always greater than or equal to 0, can be made equal to 0 by choosing $b = \mu = EY_i$. Hence,

$$\min_{b} E(Y_i - b)^2 = E(Y_i - EY_i)^2,$$

then

$$\arg \min_{b \in \mathbb{R}} \sum_{i=1}^{n} (Y_i - b)^2 = \sum_{i=1}^{n} (Y_i - EY_i)^2$$

33
then,

\[ b = EY_i = \mu. \]

In the next section we will find solution of the theorem conditional quantile as a minimization problem.

### 3.3 The Conditional Quantile as A minimization Problem

Conditional quantiles are used e.g. in the calculation of Value-at-Risk in the presence of conditional information and in conditional quantile regression. By characterizing the conditional quantile as a solution to a minimization problem it is possible to use numerical minimization methods to numerically find the conditional quantile. In the previous section we studied the mean as minimization, similarly, we can define the median as the solution to the problem of minimizing a sum of absolute residuals.

**Example 3.3.1.** To define the median as minimization problem we will solve:

\[
\arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^{n} |Y_i - \theta|.
\]

*Proof.* [6]

\[
E|Y_i - \theta| = \int_{-\infty}^{\infty} |Y_i - \theta| f_Y(y) dy
\]

\[
= \int_{-\infty}^{\theta} (Y_i - \theta) f(y) dy + \int_{\theta}^{\infty} (Y_i - \theta) f(y) dy
\]

\[
= \int_{-\infty}^{\theta} f(y) dy - \int_{\theta}^{\infty} f(y) dy
\]

then

\[
\int_{-\infty}^{\theta} f(y) dy = \int_{\theta}^{\infty} f(y) dy
\]
The definition of the median is \( p(Y \leq \theta) = p(Y \geq \theta) = 0.5 \).

The symmetry of the piecewise linear absolute value function implies that the minimization of the sum of absolute residuals equates the number of the positive and negative residuals. The median regression estimates the conditional median of \( Y \) given \( X = x \) and corresponds to the minimization of \( E(|Y - \theta| |X = x) \) over \( \theta \).

The associated loss function is \( r(u) = |u| \). We can take the loss function to be \( \rho_{0.5}(u) = 0.5|u| \), because the half positive equal half negative. We may write \( \rho_{0.5}(u) \) in the form:

\[
\rho_{0.5}(u) = 0.5uI_{[0,\infty)}(u) - (1 - 0.5)uI_{(-\infty,0)}(u),
\]

where \( I_A(u) \) is the indicator function of the set \( A \).

**Remark 3.3.1.** When we replace 0.5 by some \( \alpha \), \( 0 < \alpha < 1 \), we have

\[
\rho_{\alpha}(u) = \alpha uI_{[0,\infty)}(u) - (1 - \alpha)uI_{(-\infty,0)}(u) = \frac{|u|+(2\alpha-1)u}{2}
\]

**Proof.** The left side:

\[
\alpha uI_{[0,\infty)}(u) - (1 - \alpha)uI_{(-\infty,0)}(u) = \begin{cases} 
\alpha u, & u \geq 0; \\
(\alpha - 1)u, & u < 0.
\end{cases}
\]

The right side:

\[
\frac{|u|+(2\alpha-1)u}{2} = \begin{cases} 
\frac{u+(2\alpha-1)u}{2} = \frac{2\alpha u}{2} = \alpha u, & u \geq 0; \\
\frac{-u+(2\alpha-1)u}{2} = \frac{2\alpha u - 2u}{2} = (\alpha - 1)u, & u < 0.
\end{cases}
\]

Then the left side equal the right side. \( \square \)

We call \( \rho_{\alpha}(u) \) the check function. Now, when we define the 100\( \alpha \)% conditional quantile \( q_{\alpha}(x) \) at \( x \), as the value of \( \theta \) that minimizes \( E[\rho_{\alpha}(Y - \theta) | X = x] \), where \( \rho_{\alpha}(u) \) defines as the above lemma. Then the definition of the \( \alpha \)th conditional quantile \( q_{\alpha}(x) \) is as:

\[
q_{\alpha}(x) = \arg\min E[\rho_{\alpha}(Y - a) | X = x].
\]
3.4 Multivariate Quantile Based on A norm Minimization

In this section we will study the multivariate quantile based on a norm minimization, but in the beginning, we will introduce some basic definitions.

**Definition 3.4.1.** Suppose we have a complex vector space $V$. A **norm** is a function $f : V \rightarrow \mathbb{R}$ which satisfies:

1. $f(x) \geq 0$ for all $x \in V$
2. $f(x + y) \leq f(x) + f(y)$ for all $x, y \in V$
3. $f(\lambda x) = |\lambda| f(x)$ for all $\lambda \in \mathbb{C}$ and $x \in V$
4. $f(x) = 0$ if and only if $x = 0$

We usually write a norm by $||x||$.

Examples of the most important norms are as follows:

- The **2-norm** or Euclidean norm:
  
  $$ ||x||_2 = \left( \sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}} $$

- The **1-norm**:
  
  $$ ||x||_1 = \left( \sum_{i=1}^{n} |x_i| \right) $$

- For any integer $p \geq 1$ we have the **$p$-norm**:
  
  $$ ||x||_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} $$

- The **$\infty$-norm**, also called the sup-norm:
  
  $$ ||x||_{\infty} = \max_i |x_i| $$

  This notation is used because $||x||_{\infty} = \lim_{p \to \infty} ||x||_p$
Definition 3.4.2. A seminorm on a vector space $X$ over $K$ is a function $\|\| : X \rightarrow R$ such that the following properties hold:

1. $\|x\| \geq 0$ for every $x \in X$;
2. $\|\lambda x\| = |\lambda|\|x\|$ for every $x \in X$ and every $\lambda \in K$;
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$.

Remark 3.4.1. The difference between a seminorm and a norm is that a seminorm does not have to be definite; it is possible that $\|x\| = 0$ even if $x$ is not 0.

Example 3.4.1. The function $\|x\| = |x_1|$, $x = (x_1, x_2) \in R^2$, is a seminorm on $R^2$, but not a norm since $\|x\| = 0$ only implies that $x_1 = 0$.

For a univariate random variable $X$ with $E|X| < \infty$, the $\alpha$th quantile, ($0 < \alpha < 1$) may be characterized as any value $\theta$ satisfying

$$\theta = \arg \min_{\theta} E\{|X - \theta| + (2\alpha - 1)(X - \theta)\}.$$ 

Abdous and Theodorescu (1992) [1], for $1 \leq p \leq \infty$ and $0 < \alpha < 1$, define the norm like functions

$$\|x\|_{p,\alpha} = \|(x_1, \ldots, x_d)\|_{p,\alpha} = ||x_1| + (2\alpha - 1)x_1| \ldots, \frac{|x_d| + (2\alpha - 1)x_d|}{2},$$ 

where $\|\|$ denotes the usual $l^p$-norm on $R^d$, and define the $\alpha$th quantile of $X \in R^d$ as the value $\theta_{p,\alpha}$ which satisfies the following condition

$$\theta_{p,\alpha} = \arg \min_{\theta} E\{\|X - \theta\|_{p,\alpha} - \|X\|_{p,\alpha}\}.$$ 

Although $\|\|_{p,\alpha}$ is not a norm on $R^d$, it has properties similar to those of a norm:

1. for all $x \in R^d$, $\|x\|_{p,\alpha} \geq 0$;
2. for all $x \in R^d$ and $a \in R$,

$$\|ax\|_{p,\alpha} = \begin{cases} |a|\|x\|_{p,1-\alpha} & a \leq 0; \\ a\|x\|_{p,\alpha} & a > 0. \end{cases}$$
3. for all \( x, y \in \mathbb{R}^d \), \( ||x + y||_{p,\alpha} \leq ||x||_{p,\alpha} + ||y||_{p,\alpha} \);

4. for all \( x \in \mathbb{R}^d \), \( ||x||_{p,\alpha} \leq \max\{\alpha, 1 - \alpha\}||x||_p < ||x|| \).

After we studied in this section the multivariate quantile based on a norm minimization, we will study the Nadaraya-Watson estimator of the multivariate conditional quantile in the next section.

### 3.5 The Nadaraya-Watson Estimator of the Multivariate Conditional Quantile

In this section, our aim is to introduce the Nadaraya-Watson estimator of the multivariate conditional quantile. De Gooijer, et al. (2004) [8] have used the Nadaraya-Watson (NW) estimator of the conditional distribution function to find an estimator for the multivariate conditional quantile. In the beginning we will write some definition to use it in this section and we will study just the second point. Let \( f \) be a function from \( \mathbb{R}^n \) to \( \mathbb{R} \), \( f : \mathbb{R}^n \to \mathbb{R} \). The domain of \( f \) is a set in \( \mathbb{R}^n \) defined by

\[
\text{dom}(f) = \{x \in \mathbb{R}^n | f(x) \text{ is well defined (finite)}\}
\]

**Definition 3.5.1. Convex function** A function \( f : M \to \mathbb{R} \) defined on a nonempty subset \( M \) of \( \mathbb{R}^n \) and taking real values is called convex, if:

1. the domain \( M \) of the function is convex;
2. for any \( x, y \in M \) and every \( \lambda \in [0,1] \) one has

\[
f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{3.5.1}
\]

If the above inequality is strict whenever \( x \neq y \) and \( 0 < \lambda < 1 \), \( f \) is called strictly convex.

**Examples of convex functions:**

**norms**, recall that a real-valued function \( ||x|| \) on \( \mathbb{R}^n \) is called a norm, if it is:

- nonnegative everywhere, for all \( x \in \mathbb{R}^n, ||x|| \geq 0 \);
- homogeneous, for all \( x \in \mathbb{R}^n \) and \( a \in \mathbb{R} \), \( ||ax|| = |a||x|| \);
- satisfies the triangle inequality, \( ||x + y|| \leq ||x|| + ||y|| \);
- \( ||x|| = 0 \) if and only if \( x = 0 \)

**Lemma 3.5.1.** Let \( ||x|| \) be a real-valued function on \( \mathbb{R}^n \) which is positively homogeneous of degree 1:

\[
||tx|| = t||x||, \forall x \in \mathbb{R}^n, t \geq 0.
\]

\( ||.|| \) is convex if and only if it is sub-additive:

\[
||x + y|| \leq ||x|| + ||y||, \forall x, y \in \mathbb{R}^n.
\]

In particular, a norm (which by definition is positively homogeneous of degree 1 and is sub-additive) is convex.

Our aim is to introduce a conditional quantile concept that generalizes the well-known univariate conditional quantile definition (see [20]) into a multivariate setting. Now let \((X, Y)\) be a random variable on \( \mathbb{R}^s \times \mathbb{R}^d \), \((s \geq 1, d \geq 2)\), let \( ||.|| \) denote any strictly convex norm \((||\alpha + \beta|| < ||\alpha|| + ||\beta|| \) whenever \( \alpha \) and \( \beta \) are not proportional\) on \( \mathbb{R}^d \). For instance, it could be the Euclidean norm, or any \( p \)-norm \((0 < p < \infty)\), defined by

\[
||(z_1, ..., z_d)||_p = \left( \sum_{i=1}^{d} |z_i|^p \right)^{\frac{1}{p}}.
\]

Further, let \( ||.||_{p,\alpha} : \mathbb{R}^d \rightarrow \mathbb{R} \), be the application defined by

\[
||x||_{p,\alpha} = ||(x_1, ..., x_d)||_{p,\alpha} = \left| \frac{x_1 + (2\alpha - 1)x_1}{2}, ..., \frac{x_d + (2\alpha - 1)x_d}{2} \right|_p,
\]

Although \( ||.||_{p,\alpha} \) is not a norm on \( \mathbb{R}^d \), it has properties similar to those of norm; see [1]. Furthermore, for notational simplicity, we write \( ||.||_\alpha \) for \( ||.||_{2,\alpha} \), and \( ||.|| \) for \( ||.||_2 \). For a fixed \( x \in \mathbb{R}^s \), define a vector function of \( \theta \), \( (\theta \in \mathbb{R}^d) \), by

\[
\varphi(\theta, x) = E(||Y - \theta||_\alpha - ||Y||_\alpha|X = x) = \int_{\mathbb{R}^d} (||y - \theta||_\alpha - ||y||_\alpha)Q(dy|x), \quad (3.5.2)
\]

where \( Q(.|x) \) is the conditional probability measure of \( Y \) given \( X = x \). Because \( ||\theta||_\alpha < ||\theta|| \), we have \( |\varphi(\theta, x)| < ||\theta|| \), for all \( \theta \in \mathbb{R}^d \). Thus, \( \varphi(\theta, x) \) is well defined.
Definition 3.5.2. An $\alpha$-multivariate conditional quantile, is the point $\theta_\alpha(x)$ which assumes the infimum

$$\varphi(\theta_\alpha(x), x) = \inf_{\theta \in \mathbb{R}^d} \varphi(\theta, x),$$

i.e.

$$\theta_\alpha(x) = \arg \min_{\theta \in \mathbb{R}^d} \varphi(\theta, x).$$  \hfill (3.5.3)

In fact, for $\alpha = 0.5$, $\theta_\alpha(x)$ reduces to the well-known multivariate conditional median. It follows from Theorem 2.17 of [19] that, unless the support of $Q(V|x)$ is included onto a straight line in $\mathbb{R}^d$, $\varphi(\theta, x)$ must be a continuous and strictly convex function of $\theta$, see [8]. This guarantees the existence and uniqueness of $\theta_\alpha(x)$. If the norm is not strictly convex, uniqueness of $\varphi(\theta, x)$ is not guaranteed, see [24]. Also, when $\varphi(\theta, x)$ is defined on an infinite dimensional space, it may not have a minimum, see [22]. So far we have given a definition for the conditional multivariate quantile. Now, we introduce a nonparametric estimator for $\theta_\alpha(x)$. The estimator of the conditional multivariate is denoted by $F_n(y|x)$ where,

$$F_n(y|x) = \frac{\sum_{i=1}^n K_h(x - x_i) I_{(-\infty,y]}(Y_i)}{\sum_{i=1}^n K_h(x - x_i)}, y \in \mathbb{R}^d$$

where, $K_h(.) = \frac{1}{h^d} K(./h)$, $K(.)$ is a multivariate kernel function and $h$ is a bandwidth.

For any Borel set $V \in \mathbb{R}^d$, let $Q_n(V|x) = \int_V F_n(dy|x)$ be the estimate of $Q(V|x)$. Then, for $\theta \in \mathbb{R}^d$, the natural estimate of $\varphi(\theta, x)$ denoted by $\varphi_n(\theta, x)$, can defined as

$$\varphi_n(\theta, x) = \int_{\mathbb{R}^d} (||y - \theta||_\alpha - ||y||_\alpha) F_n(dy|x)$$

$$= \sum_{j=1}^n (||Y_j - \theta||_\alpha - ||Y_j||_\alpha) \frac{K_{hn}(x - X_j)}{\sum_{j=1}^n K_{hn}(x - X_j)}. \hfill (3.5.4)$$

Finally, if we minimize $\varphi_n(\theta, x)$ instead of $\varphi(\theta, x)$, the minimizer is an estimator of $\theta_\alpha(x)$ Denoted by $\theta_{\alpha,n}(x)$ such an estimator is given by

$$\theta_{\alpha,n}(x) = \arg \min_{\theta \in \mathbb{R}^d} \sum_{j=1}^n (||Y_j - \theta||_\alpha - ||Y_j||_\alpha) K_{hn}(x - X_j) \hfill (3.5.5)$$

From an implementation point of view, there is a difficulty in calculating $\theta_{\alpha,n}(x)$ because it does not have an explicit representation.
3.6 Consistency of the Multivariate Nadaraya-Watson Estimator

In this section, we will prove the consistency of the multivariate Nadaraya-Watson estimator, under some usual assumption. Let $C$ denote a fixed compact subset of $\mathbb{R}^s$ on which the marginal density of $X$, denoted by $g$, is lower bounded by some positive constant. Below we impose some regularity conditions which are required to prove the theoretical results of this section.

(A1) (i) For fixed $x$, $g(x) > 0$ and $g(.)$ is continuous at $x$.
(ii) For fixed $y, x$, $0 < F(y|x) < 1$ and has continuous second-order derivative with respect to $x$.
(iii) let $g_{1,i}(.,.)$ be the joint density of $X_1, X_i$, for $i \geq 2$. Assume that for all $u, v$

$$|g_{1,i}(u, v) - g(u)g(v)| \leq M < \infty.$$  

(A2) The kernel $K : \mathbb{R}^s \rightarrow \mathbb{R}$ is continuous, bounded, non negative, and satisfying:

(i) $\sup_{u \in \mathbb{R}^s} |K(u)| < \infty$.
(ii) $\int_{\mathbb{R}^s} K(u)du = 1$.
(iii) $\int_{\mathbb{R}^s} u_i K(u)du = 0, \forall i \in \{1, ..., s\}$.
(iv) $||u||^s K(u) \rightarrow 0$ as $||u|| \rightarrow \infty$.
(v) $\int_{\mathbb{R}^s} u_i u_j K(u)du = 0, \forall i \in \{1, 2, ..., s\}$.
(vi) $\int_{\mathbb{R}^s} u_i^2 K(u)du < \infty, \forall i \in \{1, 2, ..., s\}$

(A3) The process $\{(X_i, Y_i)\}$ is $\alpha$-mixing with mixing coefficient satisfying that $\alpha(i) = O(i^{-(2+\delta)})$ for some $\delta > 0$.

(A4) (i) As $n \rightarrow \infty, h \rightarrow 0$ and $nh^s \rightarrow \infty$,
(ii) $nh^{s(1+2/\delta)} \rightarrow \infty$.

(A5) For any Borel $V \subset \mathbb{R}^d$ and for any $\theta \in \mathbb{R}^d$, the functions $Q(V|.)$ and $\varphi(\theta,.)$ are continuous on $C$. 

41
(A6) The function \( \theta_\alpha(.) \) satisfies an uniform uniqueness property over \( C \):

\[
\forall \epsilon > 0, \exists \delta > 0, \forall t : C \to \mathbb{R}^d :
\]

\[
\sup_{x \in C} ||\theta_\alpha(x) - t(x)|| \geq \epsilon \Rightarrow \sup_{x \in C} |\varphi(\theta_\alpha(x), x) - \varphi(t(x), x)| \geq \delta.
\]

**Remark 3.6.1.** Assumptions (A1) and (A2) are classical in nonparametric estimation. Assumption (A3), (A4) is needed to show the almost sure convergence of \( \sup_{y \in \mathbb{R}^d} |F_n(y|x) - F(y|x)| \) to 0.

Assumption (A5) simply implies uniform convergence of \( \varphi_n(\theta, .) \) to \( \varphi(\theta, .) \). The uniform uniqueness property (A6) was introduced by Collomb, et al. (1987)[7] in order to prove consistency of an estimate of the conditional mode.

**Theorem 3.6.1.** Assume that Assumptions (A.1)-(A.4) are satisfied, Then:

- with probability 1 (w.p. 1), we can find an integer \( N \geq 1 \), such that if \( n \geq N \) and \( x \in C, \theta_{\alpha,n}(x) \) exists and is unique;
- \( \theta_{\alpha,n}(x) \to \theta_{\alpha}(x) \) with probability, if \( n \to \infty \).

**Theorem 3.6.2.** Assume that Assumptions (A.1)-(A.6) are satisfied. Then w.p. 1, we have

\[
\sup_{x \in C} ||\theta_{\alpha,n}(x) - \theta_{\alpha}(x)|| \to 0, \text{ if } n \to \infty.
\]

Now, we will prove this theorem but, Before prove the theories we will write some lemmas and definitions that benefit us to prove.

**Lemma 3.6.1.** **Bochner Lemma** suppose that \( K \) is a Borel function satisfying the conditions:

1. \( \sup_{x \in \mathbb{R}^s} |K(x)| < \infty \).
2. \( \int_{\mathbb{R}^s} K(x)d\mathbf{x} < \infty \).
3. \( \lim_{||x|| \to \infty} ||x||^4 K(x) = 0 \).
Let $g(y)$ satisfy $\int_{\mathbb{R}} g(y) dy < \infty$. Let $\{h\}$ be a sequence of positive constants such that $\lim_{n \to \infty} h = 0$. Define $g_n(x) = \frac{1}{h} \int_{\mathbb{R}} K(\frac{x-y}{h}) g(x-y) dy$, then at every point $x$ of continuity of $g(\cdot)$,

$$\lim_{n \to \infty} g_n(x) = g(x) \int_{\mathbb{R}} K(y) dy.$$ 

**Lemma 3.6.2. (Borel-Cantelli Lemma).** Let $\{A_n\}$ be a sequence of events, and denote by $P(A_n)$ the probability that $A_n$ occurs, $n \geq 1$. Also, let $A$ denote the event that $A_n$ occur infinitely often (i.o). Then

$$\sum_{i=1}^{n} P(A_n) < \infty \Rightarrow P(A) = 0,$$

no matter whether the $A_n$ are independent or not.

**Definition 3.6.1.** A collection of functions $\mathcal{S}$ mapping $C$ into $\mathbb{R}$ is uniformly equicontinuous if for each $\epsilon > 0$, there exists $\delta > 0$ such that for all $f \in \mathcal{S}$ and for $\theta$ and $\beta$ satisfying $||\theta - \beta|| \leq \delta$ we have:

$$||f(\theta) - f(\beta)|| \leq \epsilon.$$ 

**Theorem 3.6.3. (Ascoli's theorem)** Let $\{H_n\}$ be a sequence of deterministic functions from $C$ to $\mathbb{R}$, where $C$ is a compact subset of a Euclidean space. Then $\{H_n\}$ converges uniformly to a function $H : C \to \mathbb{R}$ if and only if

1. $\{H_n\}$ converges point wise to $H$, and

2. $\{H_n\}$ is uniformly equicontinuous and $H$ is a continuous function.

**Lemma 3.6.3.** Under the assumptions (A1)-(A5), we have

1. $\lim_{||\theta|| \to \infty} \sup_{x \in C} |\tilde{\varphi}(\theta, x)| - 1| = 0.$

2. $\lim_{||\theta|| \to \infty} \sup_{n \geq 1} \sup_{x \in C} |\tilde{\varphi}^n(\theta, x)| - 1| = 0.$

**Proof.** An adaption of the proof of Lemma 1 in Berlinet, et al. (2001a) [3] gives the proof of the lemma, see Berlinet, et al. (2001b)[2].

Before we prove the two mains theorems, we prove the following theorem.
Theorem 3.6.4. Under the assumptions (A1)-(A4), we have that

\[ F_n(y|x) - F(y|x) = B(y|x) + O_p(h^2) + o_p(h^2) + O_p\left((nh^s)^{-\frac{1}{2}}\right), \]

where

\[ B(y|x) = \frac{h^2}{2} \frac{\partial^2 F(y|x)}{\partial x^2_1} \int_R u_1 K_1(u_1) du_1. \]

First, we introduce some notation.

For \( x = (x_1, ..., x_s)^T \), let

\[ K(x) = \prod_{i=1}^s K_i(x_i), \]

where \( K(.) \) is the multivariate kernel and \( K_i(.) \) is the univariate kernel in the \( x_i \) direction. Also, define \( k_1, k_2 \) and \( k_3 \) as follows

\[ K_1 = \int_R u_1 K_1(u_1) du_1, \]
\[ K_2 = \int_R u_1^2 K_1^2(u_1) du_1, \]

and

\[ K_3 = \int_{R^{s-1}} K_2^s(u_2) ... K_2^2(u_s) du_2 ... du_s. \]

Let \( C_1 \) be a generic constant that might take different values at different places.

Let

\[ J_1 = \sqrt{h^s/n} \sum_{i=1}^n \varepsilon_i K_h(x - X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i, \]
\[ \xi_i = \sqrt{h^s} \varepsilon_i K_h(x - X_i), \]
\[ \varepsilon_i = I_{(-\infty,y]}(Y_i) - F(y|X_i). \]

Lemma 3.6.4. Under the assumptions of the theorem,

\[ \text{Var}(J_1) \to v_0 F(y|x)[1 - F(y|x)] g(x) = \sigma^2(y|x) g^2(x) = \theta^2(y|x), \]

where

\[ v_0 = \int_{R^s} K^2(u) du. \]

Proof.

\[ E(\varepsilon_i|X_i) = \int_{R^d} (I_{(-\infty,y]}(v) - F(y|X_i)) f(v|X_i) dv \]
\[ = \int_{-\infty}^y f(v|X_i) dv - F(y|X_i) = F(y|X_i) - F(y|X_i) = 0. \]
Therefore,
\[ E\xi_i = 0. \]
\[ \text{Var}(J_i) = E(\xi_i^2) + \sum_{i=2}^{n} (1 - \frac{i-1}{n})\text{Cov}(\xi_1, \xi_i). \] (3.6.1)
then,
\[ E(\xi_i^2) = \int_{-\infty}^{\infty} f(v|X_i)dv - 2F(y|X_i) \int_{-\infty}^{\infty} f(v|X_i)dv + F^2(y|X_i) \]
\[ = F(y|X_i) - 2F(y|X_i) + F^2(y|X_i). \]

This implies that
\[ E(\xi_i^2) = F(y|X_i)[1 - F(y|X_i)]. \] (3.6.2)
\[ h^sEK_h^2(x - X_i) = h^s \int_{R^d} K_h^2(u)g(u)du \]
\[ = \frac{1}{h^s} \int_{R^d} K^2(\frac{u}{h})g(x - u)du, \]
by Bochner lemma we obtain that
\[ h^sEK_h^2(x - X_i) = g(x)v_0 + o_p(1). \] (3.6.3)

Therefore from 3.6.2, 3.6.3, we have that
\[ E(\xi_i^2) = \theta^2(y|x) + o_p(1). \] (3.6.4)

Choose \( d_n = O(h^s/(1+2)) \) and decompose the second term on the right-hand side of 3.6.1 into two terms as follows
\[ \sum_{i=2}^{n} = \sum_{i=2}^{d_n} + \sum_{i=d_n+1}^{n} = J_{11} + J_{12} \]
for some constant \( C_3 \), and by condition A1(iii), we get
\[ |\text{Cov}(\xi_1, \xi_i)| = |\int_{R^{2d}} \xi_1 \xi_i g_{1,i}(u, v)du dv - \int_{R^d} \xi_1 g(u)du \int_{R^d} \xi_i g(v)dv| \]
\[ \leq C_3h^s \int_{R^{2d}} K(u)K(v)g_{1,i}(x - hu, x - hv)du dv \]
\[ - \int_{R^d} K(u)g(x - hu)du \int_{R^d} g(x - hv)dv | \]
\[ \leq C_3h^s \int_{R^{2d}} |K(u)K(v)||g_{1,i}(x - hu, x - hv) \]
\[ - g(x - hu)g(x - hv)|du dv \leq C_1h^s. \]
Therefore

\[ J_{11} = O_p(d_nh^s) = o_p(1). \quad (3.6.5) \]

\[ |\xi_i| = h^s |\varepsilon_i K_h(x - X_i)| \]

\[ = h^{-s} |\varepsilon_i K(x - X_i)| \leq C_1 h^{-\frac{s}{2}}. \]

Then it follows from Theorem 17.2.1 of Ibragimov and Linnik (1971)[17]

\[ |Cov(\xi_1, \xi_i)| \leq C_1 h^{-s} \alpha(i - 1). \]

Therefore

\[
J_{12} \leq C_1 h^{-s} \sum_{i=d_n+1}^{n} \alpha(i-1) \leq C_1 h^{-s} \sum_{i=d_n}^{n} \alpha(i) \\
\leq C_1 h^{-s} \sum_{i=d_n}^{n} i^{-(2+\delta)} \leq C_1 h^{-s} \sum_{i=d_n}^{n} d_n^{-(2+\delta)} \\
\leq C_1 h^{-s} d_n^{-(2+\delta)} \leq C_1 h^{-s} d_n^{-(1+\delta)} \\
\leq C_1 h^{-s} (h^{-\frac{s}{2}})^{-(1+\delta)} \\
\leq C_1 h^{-s} (\frac{1}{1+\delta})^{-(1+\delta)} = o_p(1), \quad (3.6.6)
\]

since \( \frac{1+\delta}{1+\delta/2} > 1 \). From 3.6.4, 3.6.5 and 3.6.6, the proof of the lemma is completed.

\[ \square \]

Now from The estimator of the conditional multivariate in the previous section, we have that

\[
F_n(y|x) - F(y|x) = \frac{\sum_{i=1}^{n}[I_{(\infty, y]}(Y_i) - F(y|x)]K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \\
= \{ (nh^s)^{-\frac{s}{2}} J_1 + J_2 \} J_3^{-1} \{ 1 + o_p(1) \}, \quad (3.6.7)
\]

where

\[ J_2 = \frac{1}{n} \sum_{i=1}^{n} [F(y|X_i) - F(y|x)] K_h(x - X_i), \]

\[ J_3 = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i). \]

**Lemma 3.6.5.** Under the assumptions of the theorem, we have

1. \( J_2 = g(x) B(y|x) + O_p(h^2) + o_p(h^2), \)
2. \( J_3 = g(x) + o_p(1) \).

Proof. see [27] \qed

Now from (3.6.7) and Lemma 3.6.5, we have that

\[
(nh^s)^2[F_n(y|x) - F(y|x) - B(y|x) + O_p(h^2) + o_p(h^2)] = g^{-1}(x)J_1 + o_p(1).
\]  (3.6.8)

An application of Lemma 3.6.4 to (3.6.8) completes the proof of Theorem 3.6.4 Now, we will prove our main results. Our proof will follow the main steps of the proof of De Gooijer, et al. (2004)[8].

Proof of theorem 3.6.1

(i) Existence and uniqueness:

\[
|\varphi_n(\theta, x)| = |\int_{\mathbb{R}^d} (||y - \theta||_\alpha - ||y||_\alpha)F_n(dy|x)|
\]

\[
\leq \int_{\mathbb{R}^d} ||y - \theta||_\alpha - ||y||_\alpha F_n(dy|x)
\]

\[
= \int_{\mathbb{R}^d} ||\theta||_\alpha F_n(dy|x)
\]

\[
< \int_{\mathbb{R}^d} ||\theta|| F_n(dy|x)
\]

\[
= ||\theta||.
\]

This implies that \( |\varphi_n(\theta, x)| < ||\theta|| \). Thus, \( \varphi_n(\theta, x) \) is well defined.

\[
|\varphi_n(\theta, x) - \varphi_n(\beta, x)| = |\int_{\mathbb{R}^d} (||y - \theta||_\alpha - ||y||_\alpha)F_n(dy|x) - \int_{\mathbb{R}^d} (||y - \beta||_\alpha - ||y||_\alpha)F_n(dy|x)|
\]

\[
= |\int_{\mathbb{R}^d} (||y - \theta||_\alpha - ||y - \beta||_\alpha)F_n(dy|x)|
\]

\[
= |\int_{\mathbb{R}^d} (||z + \beta - \theta||_\alpha - ||z||_\alpha)F_n(dy|x)|
\]

\[
= |\varphi_n(\theta - \beta, x)|
\]

\[
< ||\theta - \beta||_\alpha < ||\theta - \beta||.
\]

Therefore, \( \varphi_n(\cdot, x) \) is continuous because it is Lipschitzian. It is also convex because it is the integral of a convex function. Now, because \( \mathbb{R}^d \) is finite-dimensional, the set of quantiles is not empty. In fact, it is a closed convex set. Further, from Berlinet, et al. (2001)[4] Lemma 2, \( Q_n(\cdot|x) \) is not carried by a straight line in \( \mathbb{R}^d \). Consequently,
according to Theorem 2.17 of Kemperman (1987)[19], \( \varphi_n(.,x) \) possesses a unique \( \alpha \) conditional quantile.

(ii) Consistency:
First, we have to prove that for a fixed \( x \in \mathbb{R}^s, \theta_{\alpha,n}(x) \) is weakly convergent to \( \theta(x) \).
Using Stute (1986), and Kemperman (1987)[19], it suffices to prove that

\[
\sup_{y \in \mathbb{R}^d} |F_n(y|x) - F(y|x)| \to 0 \text{ a.s.} \quad (3.6.9)
\]

From Theorem 3.6.4, we have for all \( x \) and \( y \), \( F_n(y|x) \to F(y|x) \) in probability. Since \( F(y|x) \) is a distribution function, it follows from Borel-Cantelli lemma, that \( 3.6.9 \) is true. Therefore using the convexity of \( \varphi(.,x) \) and \( \varphi_n(.,x) \) (see Kemperman (1987)[19] and De Gooijer, et al. (2004))\[8\] the minimizer \( \theta_{\alpha,n}(x) \) of \( \varphi_n(.,x) \) converges in probability to the minimizer \( \theta_{\alpha}(x) \) of \( \varphi(.,x) \).

**Proof. Proof of theorem 3.6.2** We have w.p.1, and for all \( \theta \in \mathbb{R}^d, i \geq 1 \)

\[
|||Y_i - \theta||_\alpha - ||Y_i||_\alpha| \leq ||\theta||. \quad (3.6.10)
\]

Using 3.6.10, (A2) and (A5), we have w.p.1 (Berlinet, et al. (2001))\[4\]

\[
\sup_{x \in C} |\varphi(\theta, x) - \varphi_n(\theta, x)| \to 0, i.f, n \to \infty. \quad (3.6.11)
\]

But w.p.1, if \( n \geq 1, x \in C \) and \( \theta, \beta \in \mathbb{R}^d \), we have that

\[
|\varphi_n(\theta, x) - \varphi_n(\beta, x)| \leq ||\theta - \beta||, |\varphi(\theta, x) - \varphi(\beta, x)| \leq ||\theta - \beta||. \quad (3.6.12)
\]

From 3.6.12 and w.p.1, the sequence of the functions \( (\varphi_n(.,x), n \geq 1) \) is equicontinuous, and this property is independent of \( x \in C \). Therefore using (3.6.11) and scoli's Theorem, we get that, w.p.1, if \( A > 0 \):

\[
\sup_{||\theta||<A} \sup_{x \in C} |\varphi_n(\theta, x) - \varphi(\theta, x)| \to 0, i.f n \to \infty. \quad (3.6.13)
\]

Now, we want to prove that, w.p.1, one can find \( r > 0 \), and \( N \geq 1 \), such that

\[
\sup_{n \geq N} \sup_{x \in C} ||\theta_{\alpha,n}(x)|| \leq r, \sup_{x \in C} ||\theta_{\alpha}(x)|| \leq r \quad (3.6.14)
\]
From Lemma 3.6.3(2), one can find, w.p.1, $r_1 > 0$ such that if $||\theta|| > r_1, \forall n \geq 1,$ and $\forall x \in C$:

$$\varphi_n(\theta, x) \geq \frac{1}{2}||\theta||.$$ (3.6.15)

Assume now that there exists $n \geq N$ and $x \in C$ such that $||\theta_{\alpha,n}(x)|| > r_1$. Then, according to 3.6.15

$$\varphi_n(\theta_{\alpha,n}(x), x) \geq \frac{1}{2}||\theta_{\alpha,n}(x)||.$$ (3.6.16)

But by the definition of $\theta_{\alpha,n}(x)$:

$$\varphi_n(\theta_{\alpha,n}(x), x) = \inf_{\theta \in \mathbb{R}^d} \varphi_n(\theta, x) \leq \varphi_n(0, x) = 0.$$

This is impossible. Hence w.p.1, and for all $n \geq N$ sup$x \geq N$ sup$||\theta_{\alpha,n}(x)|| \leq r_1$.

Similarly there is a real number $r_1 > 0$ such that sup$||\theta_{\alpha}(x)|| \leq r_2$. We obtain (3.6.14) by putting $r = \max(r_1, r_2)$. Therefore,

$$\varphi_\alpha(x, x) = \inf_{\theta \in \mathbb{R}^d} \varphi_\alpha(\theta, x) = \inf_{||\theta|| \leq r} \varphi_\alpha(\theta, x),$$

$$\varphi_n(\theta_{\alpha,n}(x), x) = \inf_{\theta \in \mathbb{R}^d} \varphi_n(\theta, x) = \inf_{||\theta|| \leq r} \varphi_n(\theta, x)$$

Thus, w.p.1, if $n \geq N$

$$\sup_{x \in C} |\varphi(\theta_{\alpha}(x), x) - \varphi(\theta_{\alpha,n}(x), x)| \leq \sup_{x \in C} |\varphi(\theta_{\alpha}(x), x) - \varphi_n(\theta_{\alpha,n}(x), x)|$$

$$+ \sup_{x \in C} |\varphi_n(\theta_{\alpha,n}(x), x) - \varphi(\theta_{\alpha,n}(x), x)|$$

$$\leq \sup_{x \in C} \inf_{||\theta|| \leq r} |\varphi(\theta, x) - \inf_{||\theta|| \leq r} \varphi_n(\theta, x)|$$

$$+ \sup_{||\theta|| \leq r} \sup_{x \in C} |\varphi_n(\theta, x) - \varphi(\theta, x)|.$$ Using the assumptions (A1), (A2), (A4), (A5) and by (3.6.13) we have

$$\sup_{x \in C} |\varphi(\theta_{\alpha}(x), x) - \varphi(\theta_{\alpha,n}(x), x)| \to 0, if n \to \infty.$$ Then, by using assumption (A6) we get that w.p.1,

$$\sup_{x \in C} ||\theta_{\alpha}(x) - \theta_{\alpha,n}(x)|| \to 0, if n \to \infty.$$
conditional quantile based on a norm minimization. Then we studied the Nadaraya-Watson estimator of the multivariate conditional quantile. Also we studied the asymptotic consistency of the multivariate conditional quantile. We considered finished theoretical phase whose need it in practical phase.
Chapter 4
Application
Chapter 4

Application

This chapter consists of two sections. In Section 4.1, we use the M-NW to construct prediction intervals for a bivariate time series. Section 4.2 contains a discussion of the results of the thesis and some important conclusions.

4.1 Prediction Intervals for A Bivariate Time Series

The importance of effective risk management has never been greater. Recent financial disasters, like the stock market crash on the Wall Street in October 1987, have emphasized the need for accurate risk measures for financial institutions. The use of quantitative risk measures has become an essential management tool to be placed in parallel with models for returns. These measures are used for investment decisions, supervisory decisions and external regulation. In the fast paced financial world, effective risk measures must be as responsive to new as are other forecasts and must be easy to grasp even in complex situations. As a result, Value at Risk (VaR) has become the widely used measure of market risk in risk management employed by financial institutions and their regulators. See [12], [30] and [21].

In this chapter, we will give some applications of the NW estimator of the multivariate conditional quantile, we use our conditional M-NW estimator of Chapter 3 to estimate the prediction intervals for a bivariate time series.
Prediction intervals for the IBM and SP500 series

We illustrate the application of our conditional M-NW estimator by considering the prediction intervals of a financial position with multiple assets.

Consider the bivariate time series of the monthly log returns of the IBM stock and the SP500 index, from January 1926 to December 1999 consisting of 888 observations. The source of this data set is from [30]. Our goal here is to use the first 880 observations to estimate 90% and 95% prediction intervals for the last 8 observations nonparametrically, by using the M-NW estimator.

We rescaled the data such that they range from zero to one. Now, let $x_{1,t} = \{IBM\}_t$ and $x_{2,t} = \{SP500\}_t$. Thus $x_t = \{(x_{1,t}, x_{2,t})\}$ is a bivariate time series.

The two time series $x_{1,t}$ and $x_{2,t}$ are correlated. Figure 4.1 and Figure 4.2 show the time plot of the two series, while Figure 4.3 and Figure 4.4 show the scatterplots of the two series and their squares respectively.

Details of Calculation

We computed $\theta_{\alpha,n}(x)$ by finding the minimum at a finite sample of points $(\theta_1, \theta_2) \in [0,1] \times [0,1]$ of the function $S(\theta)$, where

$$S(\theta) = \sum_{t=1}^{(n-k)}(||Y_t - \theta||_\alpha - ||Y_t||_\alpha) K_h(x - X_t),$$

where $X_t$ and $Y_t$ are given by

$$X_t = x_t, t = 1, \ldots, (n-k), Y_t = x_t, t = (K + 1), \ldots, n, n = 888, K = 1, 2.$$ 

We used the bivariate Gaussian kernel function, and selected the bandwidths using the rule of thumb of Yu and Jones (1998)[32]. The primary bandwidth $h_{mean}$ is given by

$$h_{mean} = 4h_{op},$$

where $h_{op}$ is the optimal bandwidth for each time series, $h_{op} = 1.06sn^{-0.2}$ where $s$ is the sample standard division and $n$ is the sample size.

To see the performance of the M-NW, we calculate for each quantile the average number of observations which are less than or equal to it. The estimator performance is good as this average is close to the true $\alpha$. We report in Table 4.1, 4.2, 4.3 and 4.4 the true values of the last 8 observations and their 90% and 95% confidence intervals respectively.
Figure 4.1: Time plot of the rescaled IBM stock

Figure 4.2: Time plot of the rescaled SP500 stock
Figure 4.3: Scatterplot of the rescaled IBM stock versus the rescaled SP500 stock

Figure 4.4: Scatterplot of the squares of the rescaled IBM stock versus the squares of the rescaled SP500 stock
Figure 4.5: 90% C.I. for the last 8 observation of the IBM

Table 4.1: 90% C.I. for the last 8 observation of the IBM.

| 90% C.I. for IBM |  
|-----------------|---|---|---|
| 881  | 0.6251316 | 0.6751316 | 0.7151316 |
| 882  | 0.5751316 | 0.6811093 | 0.7551316 |
| 883  | 0.5251316 | 0.4560167 | 0.7751316 |
| 884  | 0.4751316 | 0.4889528 | 0.7851316 |
| 885  | 0.4251316 | 0.4542471 | 0.8051316 |
| 886  | 0.3751316 | 0.1577721 | 0.8151316 |
| 887  | 0.3551316 | 0.5832439 | 0.8351316 |
| 888  | 0.3551316 | 0.5777071 | 0.8651316 |

Note that, the confidence intervals contain the corresponding true values, except for the C.I for 883, 886 are very small compared the other observation.
Figure 4.6: 90% C.I. for the last 8 observation of the SP500

Table 4.2: 90% C.I. for the last 8 observation of the SP500.

<table>
<thead>
<tr>
<th>90% C.I. for SP500)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>881</td>
<td>0.4168489</td>
</tr>
<tr>
<td>882</td>
<td>0.4068489</td>
</tr>
<tr>
<td>883</td>
<td>0.4168489</td>
</tr>
<tr>
<td>884</td>
<td>0.4068489</td>
</tr>
<tr>
<td>885</td>
<td>0.3968489</td>
</tr>
<tr>
<td>886</td>
<td>0.3868489</td>
</tr>
<tr>
<td>887</td>
<td>0.3768489</td>
</tr>
<tr>
<td>888</td>
<td>0.3768489</td>
</tr>
</tbody>
</table>

Note that, the confidence intervals contain the corresponding true values, except for the C.I for 882 are very small compared the other observation.

The mean of the 90% C.I of the last 8 observation of the IBM is 0.33 and $h = 0.03025832$.

The mean of the 90% C.I of the last 8 observation of the SP500 is 0.24875 and $h = 0.02182918$. 

56
Figure 4.7: 95% C.I. for the last 8 observation of the IBM

Table 4.3: 95% C.I. for the last 8 observation of the IBM.

<table>
<thead>
<tr>
<th>95% C.I. for IBM</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>881</td>
<td>0.6251316</td>
<td>0.6751316</td>
<td>0.7151316</td>
</tr>
<tr>
<td>882</td>
<td>0.5751316</td>
<td>0.6811093</td>
<td>0.7551316</td>
</tr>
<tr>
<td>883</td>
<td>0.5251316</td>
<td>0.4560167</td>
<td>0.7951316</td>
</tr>
<tr>
<td>884</td>
<td>0.4751316</td>
<td>0.488528</td>
<td>0.8351316</td>
</tr>
<tr>
<td>885</td>
<td>0.4251316</td>
<td>0.4542471</td>
<td>0.8751316</td>
</tr>
<tr>
<td>886</td>
<td>0.3751316</td>
<td>0.1577721</td>
<td>0.9051316</td>
</tr>
<tr>
<td>887</td>
<td>0.3451316</td>
<td>0.5832439</td>
<td>0.9251316</td>
</tr>
<tr>
<td>888</td>
<td>0.3451316</td>
<td>0.5777071</td>
<td>0.9451316</td>
</tr>
</tbody>
</table>

Note that, the confidence intervals contain the corresponding true values, except for the C.I for 883, 886 are very small compared the other observation.
Figure 4.8: 95% C.I. for the last 8 observation of the SP500

Table 4.4: 95% C.I. for the last 8 observation of the SP500.

<table>
<thead>
<tr>
<th></th>
<th>95% C.I. for SP500</th>
</tr>
</thead>
<tbody>
<tr>
<td>881</td>
<td>0.4168489</td>
</tr>
<tr>
<td>881</td>
<td>0.4668489</td>
</tr>
<tr>
<td>881</td>
<td>0.5068489</td>
</tr>
<tr>
<td>882</td>
<td>0.3768489</td>
</tr>
<tr>
<td>882</td>
<td>0.5774273</td>
</tr>
<tr>
<td>882</td>
<td>0.5468489</td>
</tr>
<tr>
<td>883</td>
<td>0.3768489</td>
</tr>
<tr>
<td>883</td>
<td>0.4565564</td>
</tr>
<tr>
<td>883</td>
<td>0.5868489</td>
</tr>
<tr>
<td>884</td>
<td>0.3768489</td>
</tr>
<tr>
<td>884</td>
<td>0.4937072</td>
</tr>
<tr>
<td>884</td>
<td>0.6268489</td>
</tr>
<tr>
<td>885</td>
<td>0.3668489</td>
</tr>
<tr>
<td>885</td>
<td>0.4616538</td>
</tr>
<tr>
<td>885</td>
<td>0.6668489</td>
</tr>
<tr>
<td>886</td>
<td>0.3668489</td>
</tr>
<tr>
<td>886</td>
<td>0.5882350</td>
</tr>
<tr>
<td>886</td>
<td>0.7068489</td>
</tr>
<tr>
<td>887</td>
<td>0.3568489</td>
</tr>
<tr>
<td>887</td>
<td>0.5292267</td>
</tr>
<tr>
<td>887</td>
<td>0.7468489</td>
</tr>
<tr>
<td>888</td>
<td>0.3568489</td>
</tr>
<tr>
<td>888</td>
<td>0.5819740</td>
</tr>
<tr>
<td>888</td>
<td>0.7868489</td>
</tr>
</tbody>
</table>

Note that, the confidence intervals contain the corresponding true values, except for the C.I for 882 are very small compared the other observation.

The mean of the 95% C.I of the last 8 observation of the IBM is 0.3825 and $h = 0.03025832$.

The mean of the 95% C.I of the last 8 observation of the SP500 is 0.2725 and $h = 0.02182918$.

The result of the application indicate that the M-NW estimator perform good in constructing prediction intervals.
In this section we calculated $\theta_{\alpha,n}(.)$ by computing arg min in (3.5.5) over a finite grid. and we used our conditional M-NW estimator of Chapter 3 to prediction intervals for a bivariate time series.

4.2 Discussion and Conclusion

In this thesis, we study the N-W estimator and the conditional quantile plays an important role. Estimation of the conditional quantiles has gained particular attention during the recent three decades because of their useful application in various fields such as econometrics, finance, environmental sciences and medicine. For more details see [18]. We are proposed the asymptotic normality of the conditional quantiles. De Gooijer, et al. (2004)[8] introduced a multivariate conditional quantile notion which extends the definition of Abdous and Theodorescu (1992)[1]. They also proposed a nonparametric estimator for the multivariate conditional quantile, depending on the NW estimator of the conditional distribution function $F(.|x)$. Using the N-W estimator of $F(.|x)$, we introduced our M-NW estimator $\theta_{\alpha,n}(.)$ of the multivariate conditional quantile $\theta_{\alpha}(.)$. The consistency of the M-NW estimator has been shown. We study the application of our conditional M-NW estimator by considering the prediction intervals of a financial position with multiple assets. The result of the application indicate that the M-NW estimator perform good in constructing prediction intervals. We suggest that a new estimator of $\theta_{\alpha}(.)$ could rely upon the double kernel estimator of Yu and Jones (1998)[32]. We suggest that reweighted Nadaraya-Watson estimator used instead the NW estimator, and we can suggest that use the variable bandwidth instead that constant bandwidth.
Bibliography


