Nonparametric Prediction: Some Selected Topics
Zerom Godefay, D.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 4

Additive Conditional Quantiles

4.1 Introduction

Let \( \{(X_i, Y_i; i \geq 1)\} \) be a strictly stationary sequence of random vectors taking values in \( \mathbb{R}^d \times \mathbb{R} \) where \( d \geq 2 \). Naturally, this set-up includes the case where the pairs \((X_i, Y_i)\) are independently and identically distributed (i.i.d.). Suppose \( Y \) denotes the response variable which depends on the vector of stochastic covariates \( X = (X_1, \ldots, X_d)^T \) where \( T \) denotes the transpose of a matrix or vector. In the time series context \( X \) typically denotes a vector of lagged values of \( Y \). For a fixed \( \alpha \in (0, 1) \), the \( \alpha \)-conditional quantile of \( Y \) given \( x = (x_1, \ldots, x_d)^T \) is defined as the value \( \theta_{\alpha}(x) \) such that \( F(\theta_{\alpha}(x) | x) = \alpha \), where \( F(\cdot | x) \) denotes the conditional distribution of \( Y \) given \( X = x \). Equivalently, \( \theta_{\alpha}(x) \) can also be viewed as any solution to the problem \( \theta_{\alpha}(x) = \arg \min_{a \in \mathbb{R}} E\{\rho_{\alpha}(Y - a) | X = x\} \) where \( \rho_{\alpha}(z) = |z| + (2\alpha - 1)z \) is the so-called check function.

When the relationship between \( X \) and \( Y \) evolves across the distribution of \( Y \), conditional quantiles constitute parsimonious ways of describing the whole distribution. If one rather considers only the conditional mean, much information contained in the data would be lost because of the implicit assumption that possible differences in terms of the impact of \( X \) along the conditional distribution of \( Y \) are unimportant. For example, in Section 4.5, we consider the relationship between the value of owner-occupied homes (\( Y \)) in the Boston area and four potentially relevant variables. The empirical analysis suggests that the nature of the effect of the covariates does vary along the quantiles of \( Y \). Surely, such information cannot be offered by statistical models which deal only with the conditional mean. Thus, conditional quantiles can be useful tools in offering unique insights as to the nature of the relationship between a response and potential covariates, possibly with interesting implications to the issue at hand.

Suppose \( n \) observations of \((X, Y)\) denoted by \( \{(X_1, Y_1), \ldots, (X_n, Y_n)\} \) where \( X, = \)
Chapter 4: Additive Conditional Quantiles

$(X_{i1}, \ldots, X_{id})^T$ $(i = 1, \ldots, n)$ are available. In practice, the problem is how to use the information contained in the sample in order to estimate $\theta_\alpha(x)$. When dealing with this estimation problem, by far the most common approach is to model the relationship between $\theta_\alpha(x)$ and $x$ by the linear model $\theta_\alpha(x) = \beta_\alpha^T x$ where $\beta_\alpha = (\beta^{(1)}_\alpha, \ldots, \beta^{(d)}_\alpha)^T$. In this way, the issue of estimating the conditional quantile becomes a problem of estimating the finite-dimensional Euclidean parameter. This linear quantile approach was first introduced by Koenker and Bassett (1978) and has since then become increasingly popular in econometric applications. In many practical situations, however, the linear model for the conditional quantile may not be “rich” enough to capture the underlying relationship between the quantile of response variable $Y$ and its associated covariates $X$. Indeed some or all components can be highly nonlinear.

The other alternative to estimating the conditional quantile is the nonparametric approach, see, e.g., Chauduri (1991), Fan, Hu and Troung (1994), and De Gooyier, Gannoun and Zeron (2001a,b). Here the only assumption required is $\theta_\alpha(\cdot)$ to be suitably smooth function instead of assuming a finite dimensional linear parametric model for $\theta_\alpha(\cdot)$. In theory nonparametric conditional quantile estimates are still consistent for large $d$. But in practice, because of the “curse of dimensionality”, obtaining reliable estimates of conditional quantiles nonparametrically is difficult. Further, in high-dimensions, because the conditional quantile is a high-dimensional surface, its visual display is difficult making it less useful directly for exploratory purposes unlike the one-dimensional case.

Motivated by the above considerations, we suggest a conditional quantile set up which logically extends the linear quantile approach by allowing arbitrary smooth (rather than just linear) functions of the covariates, i.e.

$$
\theta_\alpha(x) = \delta + \sum_{u=1}^{d} \theta_u(x_u)
$$

where $\delta$ is a constant and $\theta_u(x_u)$ $(u = 1, \ldots, d)$ are $\alpha$th quantile functions of $Y$ related to each of the covariates. From data analytic point of view, the advantage of the additive set up is that since each covariate is represented separately, it retains an important interpretation feature of the linear model, i.e. the nature of the effect of a covariate on the conditional quantile response surface does not depend on the values of the other covariates. In practice this means that once the additive model is fitted to the data, we can plot the $d$ coordinate functions separately to examine the roles of the covariates in predicting the $\alpha$th conditional quantile; see, e.g., Section 4.5. When the purpose at hand is time series prediction, the credibility of quantile predictions heavily depend on accuracy of the estimate of $\theta_\alpha(\cdot)$. Under the additive set-up, the optimal rate for estimating $\theta_\alpha(\cdot)$ is $n^{-2/5}$.
as opposed to the usual optimal rate $n^{-2/(4+d)}$. Thus additivity can serve as a dimension reduction tool. By circumventing the curse of dimensionality problem, the accuracy of the estimates of $\theta_a(\cdot)$ is increased.

In this chapter we address the problem of estimating the additive conditional quantile components $\theta_u(x_u) \ (u = 1, \ldots , d)$ in (4.1). In particular, we introduce a nonparametric method that can help compute all the individual components with a one-dimensional nonparametric rate. The method is based on an extension of the works of Fan, Härdle and Mammen (1998) and Cai and Fan (2000) to the context of additive conditional quantiles. In practice, the backfitting algorithm (an iterative method) of Hastie and Tibshirani (1990) has been widely used to estimate additive models. In the context of the conditional mean, the backfitting method has been evaluated on numerous datasets and been refined quite considerably since its introduction. In contrast to this method, the estimator proposed in this chapter can be computed in a single-step avoiding the need for iterations. In this sense, it facilitates fast computer implementation making it convenient for "routine" data analysis.

The plan of this chapter is as follows. In Section 4.2, the proposed method is discussed. Here a simulation study is also conducted in order to assess the performance of the method and to help elaborate some estimation-related issues. In Section 4.3 the asymptotic behaviour of the estimator is formulated. In addition various important issues related to the theory are discussed. A special application of the methodology in Section 4.2 to partially linear models is discussed in Section 4.4. In Section 4.5, we give an empirical example to illustrate the methodology. Section 4.6 provides a sketch proof of the main theoretical result of the chapter. We give some concluding remarks in Section 4.7.

### 4.2 Methodology

#### 4.2.1 Definition

Let $X^u = (X_1, \ldots , X_{u-1}, x_u, X_{u+1}, \ldots , X_d)^T$ and $X^{-u} = (X_1, \ldots , X_{u-1}, X_{u+1}, \ldots , X_d)^T$ $(u = 1, \ldots , d)$. For identification, we assume without loss of generality that $E\{\theta_u^a(X_u)\} = 0$. Suppose $W_u(\cdot)$: $\mathbb{R}^{(d-1)} \to \mathbb{R}$ be a known weight function such that $E\{W_u(X^{-u})\} = 1$. From (4.1), we can see that

\[ E\{\theta_a(X^u)W_u(X^{-u})\} = \delta_u + \theta_u(x_u) = \theta_u^*(x_u) \tag{4.2} \]

where

\[ \delta_u = \delta + \sum_{j \neq u} E\{\theta_j(x_j)W_u(X^{-u})\}. \]
Thus, up to a constant, $\theta_u(x_u)$ can be constructed through averaging the conditional quantile surface with respect to the variable $X^{-u}$. This in turn suggests a direct estimation procedure. The constant term is not related to the final estimator, since $\theta_u(x_u)$, in practice, is centered to have mean zero for identifiability purpose. The weight function $W_u(\cdot)$ is introduced in order to improve efficiency of the estimates of $\theta_u(x_u)$ in an asymptotic sense (see the remarks in Section 4.3). In practice, the weight function can also be defined in such a way that it screens out extreme observations.

From given observations $\{(X_i, Y_i) : i = 1, \ldots, n\}$, the estimator of $\theta_u(x_u)$ is defined as

$$
\hat{\theta}_u^*(x_u) = n^{-1} \sum_{i=1}^n \hat{\theta}_u(X_i^u) W_u(X_i^{-u})
$$

(4.3)

$$
\hat{\theta}_u(x_u) = \hat{\theta}_u^*(x_u) - n^{-1} \sum_{i=1}^n \hat{\theta}_u^*(X_{iu})
$$

(4.4)

where $X_i^u = (X_{1,i}, \ldots, X_{u-1,i}, x_u, X_{u+1,i}, \ldots, X_{d,i})^T$, $X_i^{-u} = (X_{1,i}, \ldots, X_{u-1,i}, X_{u+1,i}, \ldots, X_{d,i})^T$, and the high-dimensional pilot estimator, $\hat{\theta}_u(X_i^u)$ solves

$$
\hat{F}(\hat{\theta}_u(X_i^u)|X_i^u) = \alpha
$$

where $\hat{F}(\cdot)$ is an estimator of $F(\cdot|X_i^u)$. In principle, $\hat{\theta}_u(X_i^u)$ can also be any other consistent estimator of $\theta_u(X_i^u)$.

From the above definition, the basic idea for estimating $\theta_u(x_u)$ is to first compute the high-dimensional conditional quantile surface $\hat{\theta}_u(X_i^u)$ and then take the weighted average of the estimated surface over variables $X^{-u}$ to stabilize the variance. Thus, unlike the backfitting approach, our method of estimating additive conditional quantile components is a “direct” one requiring no iterations.

### 4.2.2 Estimation of $F(\cdot|x)$

Here we outline an estimation procedure for $F(\cdot|x)$ which takes into account the underlying additive nature of the conditional quantile. To fix ideas, we consider the case of $d=2$. Then the additive set-up (4.1) reduces to

$$
\theta_\alpha(x) = \delta + \theta_1(x_1) + \theta_2(x_2).
$$

(4.5)

In particular, we focus on $u = 1$. One can follow similar steps for $u = 2$. We consider the approximation of $F(t|x)$ by a linear term near a fixed point $x_1$ and a constant term near a fixed point $x_2$ as follows,

$$
F(t|x) \approx a(x) + b(x)(X_1 - x_1).
$$
Note that unlike the usual local linear approach, it is only for the chosen direction of interest, i.e. $u = 1$, that we use the local linear approximation. In the “nuisance” direction, i.e. $u = 2$, the local constant approximation is used. Doing so can help facilitate the easy implementation of the conditional distribution function estimator that we are going to introduce shortly. With the help of a Monte Carlo study, we shall see later that doing a local linear approximation can improve the accuracy of the additive conditional quantile component of interest.

Given $n$-sample observations, the local linear estimator of $F(t|x)$ is defined as $\hat{a}(x) = \hat{a}$, where $(\hat{a}, \hat{b})$ minimize

$$
\sum_{i=1}^{n} \left( I_{Y_i \leq t} - a - b(X_{1,i} - x_1) \right)^2 K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})
$$

where $K_{h_{1,n}}(\cdot) = K(\cdot/h_{1,n})/h_{1,n}$ and $L_{h_{2,n}}(\cdot) = L(\cdot/h_{2,n})/h_{2,n}$ with $K(\cdot)$ and $L(\cdot)$ being kernel functions in $\mathbb{R}^1$ and $h_{1,n}$ and $h_{2,n}$ are respective bandwidths. The notation $1_A$ denotes the indicator function for set $A$.

Despite the attractive properties it possesses, the local linear method is not always guaranteed to give distribution function estimates that are monotone increasing or constrained between 0 and 1. Because our approach of estimating the conditional quantiles requires inverting the conditional distribution, the properties of positivity and monotonicity are particularly necessary. Here we re-introduce the re-weighted Nadaraya-Watson smoother (RNW) which was discussed in Chapter 3 but adapted to the conditional distribution. The main advantage of the RNW smoother is that while preserving the superior properties of the local linear approach, it gives distribution function estimates that are between 0 and 1 and also monotonic.

We now proceed to a rather brief construction of the smoother. All the necessary details were already being discussed in Section 3.2. Let $\tau_i(x)$ denote probability like factors with the properties that $\tau_i(x) \geq 0$, $\sum_{i=1}^{n} \tau_i(x) = 1$, and

$$
\sum_{i=1}^{n} \tau_i(x)(X_{1,i} - x_1) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i}) = 0. \quad (4.6)
$$

In practice, we need to choose the factors $\tau_i(x)$. First, introduce the empirical log-likelihood function as follows

$$
L = \sum_{i=1}^{n} \log(\tau_i(x)).
$$

By maximizing $L$ subject to the constraints (the maximum can be found by Lagrange multipliers), one easily obtains

$$
\tau_i(x) = n^{-1} \{1 + \lambda(X_{1,i} - x_1) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})\}^{-1} \quad (4.7)
$$
where $\lambda$ is a function of the data and of $x$. Using these weights, the RNW estimator of the conditional distribution function is defined as

$$\hat{F}(t|x) = \frac{\sum_{i=1}^{n} 1(Y_i \leq t) \tau_i(x) K_{h_{1,n}}(x_i - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})}{\sum_{i=1}^{n} \tau_i(x) K_{h_{1,n}}(x_i - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})}.$$  (4.8)

It is also easy to see that, $\lambda$ can be computed as a unique minimizer of

$$-\sum_{i=1}^{n} \log\{1 + \lambda(x_{1,i} - x_1) K_{h_{1,n}}(x_i - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})\}.$$

### 4.2.3 A Monte Carlo experiment

In this small-scale simulation experiment, we investigate if the proposed method (4.4) is capable of recovering conditional quantile components, $\theta_{\alpha}(x_n)$, when the full-dimensional conditional quantile, $\theta_{\alpha}(\cdot)$ is purely additive over the covariates. As an illustration, we consider a univariate time series process $\{Z_t\}$ which is generated by the following nonlinear autoregressive (AR) model

$$Z_t = \theta_1(Z_{t-1}) + \theta_2(Z_{t-2}) + \theta_3(Z_{t-3}) + 0.4\varepsilon_t,$$  (4.9)

where

$$\begin{align*}
\theta_1(x_1) &= x_1 \exp(-0.9x_1^2), \\
\theta_2(x_2) &= 0.6x_2, \\
\theta_3(x_3) &= \sin(0.5\pi x_3)
\end{align*}$$

and $\{\varepsilon_t\}$ are i.i.d. $N(0,1)$. To put (4.9) in a regression set-up, we define $Y_t = Z_t$ and $X_t = (X_{1,t}, X_{2,t}, X_{3,t})^T$ where $X_{k,t} = Z_{t-k}$ ($k = 1, 2, 3; t > k$). Clearly, the $\alpha$ quantile of $Y_t$ conditional on $X_t = x$ where $x = (x_1, x_2, x_3)^T$ is simply

$$\begin{align*}
\theta_{\alpha}(x) = \delta + \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3)
\end{align*}$$

where the constant $\delta = 0.4\Phi^{-1}(\alpha)$. We consider the case of $\alpha = 0.5$ (or the median). In this particular example, it does not matter what $\alpha$-level we take as the conditional quantile components for all $\alpha$'s are the same up to a constant.

From (4.9), 25 samples of size $n = 200$ are generated. Thus the experiment is carried out 25 times. In the computations a Gaussian kernel is used. In this example, we do not employ any automatic bandwidth selection to choose the three bandwidths involved. After
Figure 4.1: Estimated conditional additive quantiles $\hat{\theta}_u(x_u)$ (vertical axis) versus $x_u$ (horizontal axis) ($u = 1, 2, 3$). The first and second columns correspond, respectively, to RNW based estimates and Nadaraya-Watson based estimates. The first, second, and third rows correspond, respectively, to $x_1$, $x_2$, and $x_3$. Solid lines are true components and dotted lines show estimates.
experimenting on various combinations, the bandwidth choice $h_{1,n} = 0.27$, $h_{2,n} = 0.61$ and $h_{3,n} = 0.27$ gives a good fit to the true conditional quantile components.

Figure 4.1 shows the RNW based estimates $\hat{\theta}_u(x_u) \ (u = 1, 2, 3)$ for 25 replications along with the true quantile components; see, Figures 4.1(a), (c), and (e). To help appreciate the role of the local linear approximation in the direction of interest, we also displayed estimates of the components; see, Figures 4.1(b), (d), and (f), where the pilot $\hat{\theta}(x)$ is computed via the local constant smoother of the conditional distribution. Overall, regardless of how the pilot is estimated, the proposed method performs fairly well in identifying the conditional quantile components when in fact the underlying conditional quantile is additive. But in terms of accuracy of the component estimates, one can well see that the local linear approximation as reflected in the RNW smoother pays off. Note how the local constant based estimates deviate consistently. While on the contrary, the RNW estimates are centered around the true components and hence indicating their favourable bias characteristics; see also Remark 2 in Section 4.3.

4.3 Asymptotic behaviour of the estimator

The asymptotic behaviour of the additive conditional quantile estimator is derived under $\alpha$-mixing condition; see Section 3.3. To simplify discussion, we shall consider the case where $X_i$ is bivariate ($d = 2$) random vector. Thus, the additive set-up is as in (4.5). Here we derive the asymptotic theory for the first additive component $\theta_1(x_1)$. One can follow similar steps to work on the asymptotic theory for the second component. Let $x' = (x_1, X_2, x_{12})^T$. From (4.3), the estimator of interest is given by

$$\hat{\theta}_1(x_1) = n^{-1} \sum_{i=1}^n \hat{\theta}_1(x_i) W(X_{2i}).$$

where $\hat{F}(\theta_1(x_1)|x') = \alpha$. For notational convenience the subscript 1 has been dropped from $W_1(X_{2i})$.

We suppose that the conditional distribution function $F(\cdot|x)$ admits a unique conditional quantile at a point $\theta_\alpha(x)$. Define $\sigma^2_\alpha(x) = \text{Var}(Z|x)$ where $Z = 1_{\{Y \leq t\}}$ for some $t \in \mathbb{R}$. Let $p(x)$ be the joint density of $X$, $p_1(x_1)$ and $p_2(x_2)$ be the marginal densities of $X_1$ and $X_2$, respectively, and $f(t|x)$ be the conditional density of $Y$ given $X = x$.

Under Conditions A.1-A.9 given in Section 4.6 and letting $n \to \infty$ the following theorem extends the additive conditional mean result of Fan et al. (1998) and Cai and Fan (2000) to the additive conditional quantile context.
Theorem: Assume that the Conditions A.1-A.9 are satisfied. Define the constants \(k_1 = \int u^2 K(u) du\) and \(k_2 = \int K^2(u) du\). If the bandwidths are chosen such that \(h_{1,n} \rightarrow 0\), \(h_{2,n} \rightarrow 0\) in such a way that \(n h_{1,n}^5 = O(1)\), \(h_{2,n} = o(h_{1,n})\), and \(n h_{1,n} h_{2,n} \rightarrow \infty\), then

\[(n h_{1,n})^{1/2} \left\{ \hat{\theta}_1^*(x_1) - \theta_1^*(x_1) - \text{Bias}(\hat{\theta}_1(x_1)) \right\} \xrightarrow{D} N(0, v(x_1)) \quad (4.11)\]

where the bias and the variance are given respectively by

\[\text{Bias}(\hat{\theta}_1^*(x_1)) = \frac{1}{2} h_{1,n}^2 k_1 \theta_1^{(2)}(x_1) \quad (4.12)\]

and

\[v(x_1) = k_2 p_1(x_1) \alpha(1 - \alpha) E \left( \Gamma^2(X) | X_1 = x_1 \right) \quad (4.13)\]

where

\[\Gamma(x) = \frac{p_2(x_2) W(x_2)}{p(x) f(\theta_\alpha(x)|x)}. \quad (4.14)\]

Remark 1: The above Theorem says that the function estimate \(\hat{\theta}_1^*(x_1)\) is asymptotically normal at rate \(n^{2/5}\). Thus the estimator works in producing an \(n^{2/5}\) rate of convergence out of an \(n^{2/6}\)-consistent pilot, i.e. \(\hat{\theta}_\alpha(x)\).

Remark 2: Note that the bias of \(\hat{\theta}_1^*(x_1)\) in (4.12) is exactly the asymptotic bias of an additive estimator if the local linear method was used to estimate the pilot \(\theta_\alpha(x)\). In this sense the RNW smoother is able to reproduce the superior bias property of the local linear method. For performance in small samples, see the simulation example in Section 4.2.

Remark 3: The optimal bandwidth that minimizes the asymptotic MSE is given by,

\[h_1^* = \left[ \frac{v(x_1)}{\mu_2(K) \theta_1^*(x_1)} \right]^{(1/5)} n^{-1/5}. \quad (4.15)\]

Corollary: If the weight function \(W(\cdot)\) is chosen such that it minimizes the variance \(v(x_1)\), then under all the conditions of the Theorem,

\[(n h_{1,n})^{1/2} \left\{ \hat{\theta}_1^*(x_1) - \theta_1^*(x_1) - \text{Bias}(\hat{\theta}_1^*(x_1)) \right\} \xrightarrow{D} N(0, v^*(x_1)) \quad (4.16)\]

where

\[v^*(x_1) = \frac{k_2 \alpha(1 - \alpha)}{p_1(x_1) E(f^2(\theta_\alpha(X)|X) | X_1 = x_1)}.\]
Remark 4: By using Lagrange multiplier method and noting the constraint, \( \int_{\mathbb{R}} W(x_2) p_2(x_2)dx_2 = 1 \), we can see that the optimal weight is

\[
W(X_2) = \frac{\int f^2(\theta_\alpha(x_1, X_2)|x_1, X_2) p(x_1, X_2)}{E(\int f^2(\theta_\alpha(X)|X)|X_1 = x_1) p_1(x_1)p_2(X_2)}.
\]

(4.17)

Substituting this optimal weight into \( v(x_1) \), the Corollary follows. From the Corollary, with the variance term, being free of the weight function \( W(\cdot) \), looks pretty much similar to the asymptotic variance term of a univariate conditional quantile kernel smoother; see, e.g., Lemma 1 in Chapter 5.

Remark 5: The direct estimator presented in this chapter involves a weight \( W(\cdot) \). The alternative is to just take the unweighted average. Within the latter category, the works by Linton and Nielsen (1995), Masry and Tjøstheim (1997), and Cai and Masry (2000) can be mentioned; all in the conditional mean context. The problem of the unweighted estimator is that it may not be necessarily efficient; see, e.g., Linton (1997). One possible solution is to introduce a weight function (as done here) in order to create room for efficiency improvement in the sense of reducing the variance of the additive component estimator. In fact, as can be noted from the Corollary, when the weight is chosen as in (4.17), the additive estimator achieves the most efficiency possible. Linton (1997) and Cai (2001) also proposed a two-stage estimator as an alternative approach.

Remark 6: With a detailed simulation study, Sperlich, Linton and Härdle (1997) concluded that the “direct” method (but with no weight) as applied to the conditional mean under-performs vis-à-vis the backfitting method when there is high correlation among covariates. But we can note from (4.17) that

\[
W(X_2) \propto \frac{p(x_1, X_2)}{p_1(x_1)p_2(X_2)}.
\]

In this sense the weight has the role of taking care of the dependence between the covariates and hence may help improve accuracy. Of course, a practically working way of approximating the weight function from data needs further research.

Remark 7: It is possible to generalize the Theorem and the Corollary for dimension \( d > 2 \). To that end, we need conditions on the bandwidths as follows. Assume a common bandwidth, \( h_{2,n} \), for the rest of the covariate vector, i.e. \( (X_2, X_3, \ldots, X_d)^T \). Then, the required conditions on the bandwidths are: (i) \( nh_{1,n}^5 = O(1) \), (ii) \( h_{2,n} = o(h_{1,n}) \), and (iii) \( nh_{1,n}h_{2,n}^{-1} \to \infty \). Here, one should, however, note that when the univariate optimal order, i.e. \( nh_{1,n}^5 = O(1) \) or \( h_{1,n} = O(1)n^{-1/3} \), is imposed, the two other conditions ((ii)
Additive Conditional Quantiles

and (iii)) are satisfied only if \( d < 5 \). The implication of this observation is that the additive conditional quantile method introduced in this chapter still suffers the "curse of dimensionality" problem when \( d \geq 5 \). In applications, this dimension restriction can be relaxed by considering an \( r \)th \((r > 2)\) order-kernel with a bandwidth \( h_{1,n} = O(1)n^{-1/2r+1} \).

### 4.4 Application

Let \((X, Z, Y)\) be a \( \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R} \)-valued observable random variables. In terms of previous notations, \( d = p + q \). Suppose we have the following specification of the conditional quantile,

\[
\theta_\alpha(X, Z) = \beta^T X + \sum_{j=1}^{q} \theta_j(Z_j)
\]

where \( \beta \) is a \( \mathbb{R}^p \)-valued unknown parameter, and \( \theta_j(\cdot) \)'s are unknown real functions. In the context of the conditional mean, the above set-up, usually called a partially linear model, has become increasingly popular in econometric applications. Often, the interest lies in the precise estimation of \( \beta \) because the model is built in such a way that the components of \( \beta \) have interesting economic significance, and some hypotheses of interest may be expressible purely in terms of \( \beta \).

In the context of conditional quantiles, we hardly see such a partially linear specification. In fact, in most applied work concerning conditional quantiles, all candidate explanatory variables are assumed to affect the response in a linear fashion. In terms of (4.18), this implies all the \( \theta_j(\cdot) \)'s are also considered linear. Due to this model misspecification, the parameters of the true linear components \( \beta \) can be highly inconsistent. Here we show how \( \beta \) can be estimated by allowing \( \theta_j(\cdot) \)'s to take any unknown form, not necessarily linear. To that end, the nonparametric additive quantile estimation procedure suggested in Section 4.2 can be incorporated into a two-stage approach as follows,

1. Calculate all the \((p + q)\) components using the additive quantile estimation in (4.3), i.e., obtain \( \hat{\theta}_u^*(x_u) \) \((u = 1, \ldots, p)\) and \( \hat{\theta}_u^*(z_v) \) \((v = 1, \ldots, q)\).

2. Once these nonparametric components are calculated, estimate \( \beta \) by solving the following linear quantile regression problem

\[
\min_{(\beta_0, \beta)} \sum_{i=1}^{n} \rho_\alpha \left( Y_i - \sum_{v=1}^{q} \hat{\theta}_v^*(Z_{v,i}) - \beta_0 - \beta^T X \right).
\]

\( \rho_\alpha \) is a convex and bounded loss function, which is often the Huber loss or the \( \alpha \)-trimmed loss.
We may recall from (4.2) that \( \{\theta_1^*(z_1), \theta_2^*(z_2), \ldots, \theta_q^*(z_q)\} \) over-estimate the corresponding \( \{\theta_1(z_1), \theta_2(z_2), \ldots, \theta_q(z_q)\} \) by an amount \( \delta_1 + \delta_2 + \ldots + \delta_q \). Note that since (4.19) involves an intercept term \( \beta_0 \), the slope \( \beta \) will not be affected. To implement (4.19), one can use commercially available software packages.

To appreciate the usefulness of a partially linear quantile set-up as opposed to a wrongly specified pure linear model in the estimation of the linear parameters, we present a small-scale simulation. Suppose we have the following model of \( Y \)

\[
Y_i = 0.6X_i + \theta_1(Z_{1,i}) + \theta_2(Z_{2,i}) + 0.4\varepsilon_i, \quad (i = 1, \ldots, n)
\]

(4.20)

where

\[
\theta_1(z_1) = z_1 \exp(-0.9z_1^2),
\]

\[
\theta_2(z_2) = \sin(0.5\pi z_2).
\]

The covariates \((X, Z_1, Z_2)\) each are uniformly distributed on \([0,1]\) and are independent from each other. The errors \(\{\varepsilon_i\}\) are \(i.i.d. \mathcal{N}(0,1)\) and are independent from the covariates. As in the example of Section 4.2, the conditional quantile is given by

\[
\theta_\alpha(X, Z) = 0.6X + \theta_1(Z_1) + \theta_2(Z_2).
\]

(4.21)

We shall consider \(\alpha = 0.5\) without loss of generality.

The objective is the estimation of the linear parameter, \(\beta = 0.6\). To this purpose we use two approaches. In the first case, we estimate a pure linear model in which all the components are assumed linear. Here there clearly is a misspecification and we want to see the impact. In the second approach, we use the right partially linear specification and we follow the two-stage approach to estimate the linear coefficient.

From (4.20), we generated 100 samples of size \(n=100\). To evaluate the accuracy of the two approaches, we compute at each replication, \(r\), the absolute deviation error for each method, \(|\hat{\beta}(r) - 0.6|\) and \(|\tilde{\beta}(r) - 0.6|\) where \(\hat{\beta}\) is an estimate of \(\beta\) of the partially linear model and where \(\tilde{\beta}\) is an estimate from the purely linear set-up. In Figure 4.2, the absolute errors from the 100 replications are displayed in the form of box plots.

From Figure 4.2, one can clearly notice that the estimate from the partially linear model (the first column) gives a more accurate approximation to the true \(\beta\). Interestingly, the improvement in accuracy is reflected both in terms of bias and variance. This suggests that ignoring specification errors can have a serious impact on the accuracy of linear parameter estimates. This may in turn result in poor inferences. At this stage we leave the investigation of the asymptotic behaviour of the linear parameter estimates from the two-stage procedure for future research. But under some regularity conditions, we conjunctur...
Figure 4.2: Box plots of 100 absolute error values. The first column is for the partially linear model (PL): $|\hat{\beta} - 0.6|$, and the second column is for the purely linear model (L): $|\hat{\beta} - 0.6|$.

Figure 4.3: A histogram of the 100 $\hat{\beta}$ values. The solid line represents a superimposed fitted normal density.

that $\hat{\beta}$ is $n^{1/2}$-consistent for $\beta$ and asymptotically normal. For example, Figure 4.3 shows the frequency distribution in the form of a histogram of the 100 $\beta$ estimates ($\hat{\beta}$) from the two-stage approach. Even with such a small sample size as $n = 100$, the distribution looks rather close to the normal approximation. Of course, we see a deviation from the true $\beta = 0.6$. In the first-stage of estimation of the additive quantile components we have used bandwidths which are not necessarily optimal. We think that with a better choice of the
Chapter 4: Additive Conditional Quantiles

bandwidths at the first stage, the bias of \( \hat{\beta} \) at the second stage can be improved. Further, it is interesting to see the performance of the two-stage approach for various values of \( n \).

4.5 Example

In this section we illustrate by way of an example how one may compute and interpret the additive conditional quantile modeling in the analysis of real data. The data we use is the Boston housing data-set.\(^1\) The data-set contains 14 variables on various criteria such as distance to urban amenities, pollution and crime which may affect house prices. Among others, this data-set has been analyzed by Breiman and Friedman (1985), Doksum and Samarov (1995), and Chauduri, Doksum and Samarov (1997). Sperlich et al. (1999) also used it to illustrate their additive conditional mean procedure. In our analysis, we consider four of the covariates, i.e. per capita crime rate by town \((X_1)\), average number of rooms per dwelling \((X_2)\), weighted distances of five Boston employment centers \((X_3)\), and \% lower status of the population \((X_4)\). The response variable \((Y)\) is the median value of owner-occupied homes in $1000's. There are a total of \( n = 490 \) observations on each variable. Before we do the computations, the covariates are pre-scaled in order to avoid extreme differences of spread in various coordinate directions. Specifically, the covariates are linearly transformed to have a unit variance-covariance matrix. The response \( Y \) is also standardized to have mean zero and variance 1.

Chaudhuri et al. (1997) used the same response \( Y \) and three of the covariates, i.e. \( X_2, X_3, \) and \( X_4 \) to study the effect of each of the covariates on the various quantiles across the distribution of \( Y \) to illustrate their average derivative quantile regression method. With the same objective as in Chauduri et al. (1997) but using the methodology introduced in Section 4.2, we explore the nature of the relationship between each of the 4 covariates and the quantiles of \( Y \) for \( \alpha = 0.1, 0.5, 0.9 \).

In this example, we take \( h_{1,n} = h_{2,n} = h_{3,n} = h_{4,n} \equiv h_n \) for all \( \alpha \)'s. Varying choices of the bandwidth \( h_n \) were tried in order to get a feeling for the effect of bandwidth selection on the resulting estimates. We observed fairly similar estimates when \( h_n \) varies between 0.75 and 1.5. We report the final results for \( h_n = 1 \). As in the simulation illustration, a Gaussian kernel is used.

In Figure 4.4, we display \( \hat{\theta}_u(x_u) \) \((u = 1, 2, 3, 4)\) at the three \( \alpha \)-levels. To avoid spurious features that may arise due to possible boundary problem, the shown plots correspond only to those \( x_u \) between 5\% and 95\% percentiles of \( X_u \). One can notice from the plots that the \( \hat{\theta}_u(x_u) \)'s are not monotonic in \( \alpha \) and hence may cross each other if we were to

\(^1\)The data were obtained from the website http://lib.stat.cmu.edu/datasets/boston
draw them together. This is not surprising as these plots only measure partial conditional quantile effects at a given $\alpha$. Actually one can make a parallel to linear quantile regression where coefficient estimates should not be necessarily monotonic in $\alpha$.

In the figures, we also report 90% confidence bands around the component estimates. Because the expression for the variance (4.13) is rather complicated, instead of using the asymptotic normality result, we suggest a simple bootstrap (without bias correction) to do the confidence bands. Below is the procedure for $u = 1$.

1. For a particular $\alpha$, compute the global error, $\hat{e}_{\alpha,i} = Y_i - \sum_{u=1}^{4} \hat{\theta}_u(X_{u,i})$ and center them to have mean zero.

2. Resample from $\{\epsilon_{\alpha,i}\}$ to form $\{\epsilon_{\alpha,i}^*\}$.

3. Create

$$Y_{1,i}^* = \tilde{\theta}_1(x_i) + \epsilon_{\alpha,i}^*.$$  

4. For the bootstrap sample $(X_{1,i}, Y_{1,i}^*)$, use the following one-dimensional kernel smoothing$^{2}$

$$\tilde{\theta}_1(x_1) = \arg \min_{a \in \mathbb{R}} \sum_{i=1}^{n} \rho_a(Y_{i}^* - a)K_{h_{1,n}}(x_1 - X_{1,i}).$$

The minimization can be done via iteratively re-weighted least squares. We repeat the last 3 steps 100 times. Then we take pointwise the 5% empirical percentile of the 100 bootstrap estimates $\tilde{\theta}_1(x_1)$ as the lower confidence band. The upper confidence band is defined analogously.

**Empirical results**

Looking at Figure 4.4, there seems to be evidence of nonlinearity in the component estimates. We also see that the covariates affect $Y$ very differently, both in size and to some extent in sign, across the quantiles of $Y$. For example, $X_2$, i.e. the number of rooms per dwelling, has hardly an effect on the 0.1 quantile of $Y$ while at the median and the upper tails it seems to have significant contribution. We also see the impact of $X_2$ changing sign as we move from low to upper quantiles. But this may not be the real situation as the reported negative effect is very negligible. We see the effects of $X_1$ and $X_3$ decreasing

$^{2}$This is a local constant (Nadaraya-Watson) smoother for the check function characterization of the conditional quantile; see, e.g., Yu and Jones (1998).
Figure 4.4: Estimated conditional additive quantiles $\hat{\theta}_u(x_u)$ (vertical axis) versus $x_u$ (horizontal axis) ($u = 1, 2$). The dotted lines denote the upper confidence bands and the lower confidence bands.
Figure 4.4: (Continued) Estimated conditional additive quantiles $\hat{\theta}_u(x_u)$ (vertical axis) versus $x_u$ (horizontal axis) ($u = 3, 4$). The dotted lines denote the upper confidence bands and the lower confidence bands.
Chapter 4: Additive Conditional Quantiles

dramatically for increasing \( \alpha \) while the opposite is true for \( X_2 \) and \( X_4 \). In terms of relative contribution, one can look at the range of \( \hat{\theta}_u(x_u) \). Clearly, at all \( \alpha \) levels, \( X_4 \) has the largest effect on the quantile of \( Y \). Only at \( \alpha = 0.9 \), we see \( X_2 \) coming closer to \( X_4 \). Interestingly, one can also look at the widths of the pointwise confidence bands to determine the relative contribution of the covariates. For example, at \( \alpha = 0.9 \), all along, \( X_3 \) has the largest confidence band width. So it is the least important covariate at this \( \alpha \). Similarly, at \( \alpha = 0.5 \), \( X_1 \) has the largest width all along. So it is the least important covariate at this \( \alpha \). Thus, we can interpret the width of confidence bands of the component function estimates similar to that of linear quantile regression coefficient standard errors, i.e. the more relevant a covariate, the smaller the standard error and hence the narrower becomes its symmetric confidence interval.

4.6 Sketch of the proof

The proof is based on modifications and extensions of similar proofs in Fan et al. (1998) and Cai and Fan (2000). We will provide the main idea and skip some technical details, which are fairly routine in view of the proofs already documented in the mentioned references.

The theoretical results will be derived under the following set of conditions.

A.1 The functions \( W(\cdot) \) and \( \theta_2(\cdot) \) are bounded on the support \( D \) of \( W(\cdot) \). The weight function \( W(\cdot) \) is uniformly continuous with respect to \( x_2 \).

A.2 The kernel functions \( K(\cdot) \) and \( L(\cdot) \) are symmetric probability densities and have bounded supports.

A.3 For \( u_1 \) in a neighborhood of \( x_1 \) and \( u_2 \in D \), the joint density \( p(u_1, u_2) \) is bounded away from zero by a constant. Further, it has bounded partial derivatives up to order 2 with respect to \( u_1 \) and \( u_2 \).

A.4 The functions \( \sigma^2(u) \) and \( c(u) = E(|Z - F(t|u)|^{2+\delta}|X = u) \) are continuous at the point \( u_1 = x_1 \), and \( E(c(X)|\Gamma(X)|^{2+\delta}|X_1 = u_1) \) is bounded for \( u_1 \) in a neighborhood of \( x_1 \).

A.5 The joint conditional density \( f_{Y_i,Y_i}(x_i,x_1) \) of \( (Y_i,Y_i) \) given \( (X_1,X_i) \) satisfies for all \( i > 1 \) all values of arguments involved, \( f_{Y_i,Y_i}(x_i,x_1,y_1,y_2)(u,v)) \leq C < \infty \) for some positive constant \( C \). Note that, \( X_1 = (X_{1,i},X_{2,i})^T \).

A.6 The conditional distribution \( F(y|u) \) of \( Y \) given \( X = u \) is twice continuously differentiable at the point \( u_1 = x_1 \).
A.7 \( F(y|\mathbf{u}) \) has a conditional density \( f(y|\mathbf{u}) \) and \( f(y|\mathbf{u}) \) is continuous at \( \mathbf{u} \). Further, \( f(\theta_\alpha(\mathbf{u}))|\mathbf{u}) > 0 \).

A.8 The process \( \{(X_i,Y_i)\}_{i=-\infty}^\infty \) is strongly mixing with \( \sum_{\ell=1}^\infty e^{a[\alpha(\ell)]^{\delta/(2+\delta)}} < \infty \) for some \( a > \delta/(2+\delta) \), where \( \delta \) is given in Condition (A.4).

A.9 There is a sequence of positive integers satisfying satisfying \( v_n \to \infty \) and \( v_n = O(\sqrt{n\lambda_1}) \) such that \( (n/h_1,n)^{1/2}\alpha(n) \to 0 \).

**Proof:** As in Section 4.3, let \( \mathbf{x}' = (x_1, X_{2:})^T \). We also introduce the short-hand notation

\[
p^{(i,j)}(\mathbf{x}) = \frac{\partial^{i+j}p(\mathbf{x})}{\partial x_1^i \partial x_2^j}.
\]

From (4.2) we observe that

\[
E\{\theta(\alpha)(x_1, X_2)W(X_2)\} = \delta_1 + \theta(1)(x_1) = \theta_1^*(x_1).
\]

Now applying the central limit theorem for stationary \( \alpha \)-mixing sequences; see, e.g., McCabe and Tremayne (1993, p. 128), it is easy to see that

\[
\frac{1}{n} \sum_{i=1}^n \theta(\alpha)(x_1)W(X_{2,i}) = \theta_1^*(x_1) + O_p(n^{-1/2}).
\]

(4.23)

For notational simplicity, in the remaining part of the proof, we drop the subscript \( \alpha \) and the superscript \( * \) from the notations \( \hat{\theta}(\cdot), \theta(\cdot), \hat{\theta}_1(x_1), \) and \( \theta_1(x_1) \). These terms will be re-denoted as \( \hat{\theta}(\cdot), \theta(\cdot), \hat{\theta}_1(x_1), \) and \( \theta_1(x_1) \), respectively.

Using the new notations and combining (4.10) and (4.23), we obtain

\[
\hat{\theta}_1(x_1) - \theta_1(x_1) = \frac{1}{n} \sum_{i=1}^n \left( \hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}) \right) W(X_{2,i}) + O_p(n^{-1/2}).
\]

(4.24)

Let \( \delta_n = \hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}) \). It is possible to show that the full-dimensional estimator, \( \hat{\theta}(\mathbf{x}) \), is consistent, i.e., \( \hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}) \to 0 \) in probability; see, e.g., Berhinet, Gannoum and Matzner-Løber (2001) and Chapter 5 of this thesis for a similar proof. Noting that \( F(\theta(\mathbf{x})|\mathbf{x}) = \hat{F}(\theta(\mathbf{x})|\mathbf{x}) = \alpha \), a direct use of Lemma 4 in Cai (2002) gives

\[
\hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}) = - \frac{\hat{F}(\theta(\mathbf{x})|\mathbf{x}) - F(\theta(\mathbf{x})|\mathbf{x})}{f(\theta(\mathbf{x})|\mathbf{x})} + o_p(1) + O_p(n\lambda_1^{-1}h_{1,n}^{-1}h_{2,n}^{-1}).
\]

(4.25)

Using the definition of \( \hat{F}(\cdot|\cdot) \) in (4.8),

\[
\hat{F}(\theta(\mathbf{x})|\mathbf{x}) = \frac{\sum_{i=1}^n Z_i \tau_1(\mathbf{x})K_{h_1,n}(x_1 - X_{1,i})L_{h_2,n}(x_2 - X_{2,i})}{\sum_{i=1}^n \tau_1(\mathbf{x})K_{h_1,n}(x_1 - X_{1,i})L_{h_2,n}(x_2 - X_{2,i})}
\]

(4.26)

where \( Z_i = 1_{\{Y_i < \theta(\mathbf{x})\}} \). Following similar arguments as in Section 3.4, one can show that

\[
\lambda = \frac{h_{1,n}h_{2,n}k_1p(1,0)(\mathbf{x})}{k_2k_3p(\mathbf{x})} \{1 + o_p(h_{1,n})\}.
\]
Substituting the above approximation of $\lambda$ into the definition of $\tau_i(x)$ (see (4.7)), one can see that

$$\tau_i(x) = n^{-1} b_i(x) \{1 + o_p(h_{1,n})\}$$

(4.27)

where

$$b_i(x) = \left[ 1 + \frac{h_{1,n} h_{2,n} k_1 p^{(1,0)}(x)}{k_2 k_3 p(x)} (X_{i,i} - x_1) K_{h_{1,n}}(x_1 - X_{i,i}) L_{h_{2,n}}(x_2 - X_{2,i}) \right]^{-1}.$$  

Define $\varepsilon_i = Z_i - F(\theta(x)|X_i)$. Now, using (4.26) and (4.27),

$$\hat{F}(\theta(x)|x) - F(\theta(x)|x) = ((nh_{1,n})^{-1/2} J_1 + J_2) J_3^{-1} \{1 + o_p(h_{1,n})\}$$

(4.28)

where

$$J_1(x) = n^{-1/2} h_{1,n}^{1/2} \sum_{i=1}^{n} b_i(x) \varepsilon_i K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i}),$$

$$J_2(x) = n^{-1} \sum_{i=1}^{n} \left[ F(\theta(x)|X_i) - F(\theta(x)|x) \right] b_i(x) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})$$

and

$$J_3(x) = n^{-1} \sum_{i=1}^{n} b_i(x) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i}).$$

First we deal with $J_2(.)$. Expanding $F(\theta(x)|X_i)$ in a Taylor series around $|\theta(x) - \theta(X_i)| \leq h_{2,n}$,

$$F(\theta(x)|X_i) = F(\theta(X_i)|X_i) + f(\theta(X_i)|X_i)[\theta(x) - \theta(X_i)] + O_p(h_{2,n}^2).$$

Further, note that $F(\theta(X_i)|X_i) = F(\theta(x)|x)$. Thus, $J_2(.)$ becomes

$$J_2(x) = f(\theta(x)|x)n^{-1} \sum_{i=1}^{n} (\theta(x) - \theta(X_i)) b_i(x) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i}) + O_p(h_{2,n}^2).$$

Now we exploit the additivity structure, i.e. we replace $\theta(x)$ by $\theta_1(x_1) + \theta_2(x_2)$. Similarly, $\theta(X_i)$ by $\theta_1(X_{1,i}) + \theta_2(X_{2,i})$. Then expanding $\theta_1(X_{1,i})$ to the second order about $x_1$ and also using condition (4.6), we obtain

$$J_2(x) = f(\theta(x)|x) \{J_{21} + J_{22}\} + O_p(h_{2,n}^2)$$

where

$$J_{21} = -\frac{1}{2} \theta_1''(x_1)n^{-1} \sum_{i=1}^{n} (X_{1,i} - x_1)^2 b_i(x) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i})$$

$$J_{22} = \frac{1}{2} \sum_{i=1}^{n} \left[ \sum_{j=1}^{n} (X_{1,i} - x_1)^2 b_i(x) K_{h_{1,n}}(x_1 - X_{1,i}) L_{h_{2,n}}(x_2 - X_{2,i}) \right].$$
and
\[ J_{22} = n^{-1} \sum_{i=1}^{n} [\theta_2(x_i) - \theta_2(X_{2,i})]b_i(x)K_{h_{1,n}}(x_1 - X_{1,i})L_{h_{2,n}}(x_2 - X_{2,i}). \]

One can use simple arguments to see that
\[ J_{21} = -\frac{1}{2} h_{1,n}^2 k_1 \theta_1^{(2)}(x_1)p(x) + o_p(h_{1,n}^2). \]
Similarly, \( J_{22} = O_p(h_{2,n}^2) \). Therefore,
\[ J_2(x) = -f(\theta(x)|x) \left\{ \frac{1}{2} h_{1,n}^2 k_1 \theta_1^{(2)}(x_1)p(x) + o_p(h_{1,n}^2) + O_p(h_{2,n}^2) \right\}. \] (4.29)

One can also show that
\[ J_3(x) = p(x) + o_p(1). \] (4.30)

Using (4.29) and (4.30), (4.28) reduces to
\[ \hat{F}(\theta(x)|x) - F(\theta(x)|x) = -f(\theta(x)|x) \left\{ \frac{1}{2} h_{1,n}^2 k_1 \theta_1^{(2)}(x_1) + o_p(h_{1,n}^2) + O_p(h_{2,n}^2) \right\} - \frac{1}{(nh_{1,n})^{1/2}} p^{-1}(x)J_1(x). \] (4.31)

Substituting (4.31) into (4.25), we see that
\[ \hat{\theta}(x) - \theta(x) = \frac{1}{2} h_{1,n}^2 k_1 \theta_1^{(2)}(x_1) + o_p(h_{1,n}^2) + O_p(h_{2,n}^2) + O_p(\delta_n(nh_{1,n}h_{2,n})^{-1}) + \frac{1}{(nh_{1,n})^{1/2}} f^{-1}(\theta(x)|x)p^{-1}(x)J_1(x). \] (4.32)

Using the conditions on the bandwidths (see Section 4.3), we can further simplify (4.32) to
\[ \hat{\theta}(x) - \theta(x) = \frac{1}{2} h_{1,n}^2 k_1 \theta_1^{(2)}(x_1) + o_p(h_{1,n}^2) + (nh_{1,n})^{-1/2} f^{-1}(\theta(x)|x)p^{-1}(x)J_1(x). \] (4.33)

Now we substitute (4.33) into (4.24) to obtain
\[ \hat{\theta}_1(x_1) - \theta_1(x_1) - \text{Bias}(\hat{\theta}_1(x_1)) = (nh_{1,n})^{-1/2} T(x_1) + o_p(h_{1,n}^2 + (nh_{1,n})^{-1/2}) \] (4.34)
where \( \text{Bias}(\hat{\theta}_1(x_1)) \) is as defined in (4.12) and
\[ T(x_1) = n^{-1} \sum_{i=1}^{n} J_1(x_i)A(x_i) \]
with \( A(x) = W(x_2)/p(x)f(\theta(x)|x) \).
Suppose $\Gamma(.)$ is as defined in (4.14) and define
\[
\Gamma_n(x) = n^{-1} \sum_{i=1}^{n} L_{h_{2,n}}(x_2 - X_{2,i})A(x_i).
\]
By some algebra, we can see that
\[
T(x_1) = G(x_1) + G^*(x_1)
\]
where
\[
G(x_1) = n^{-1/2} \sum_{i=1}^{n} h^{1/2}_1 b_i(x^i)K_{h_{1,n}}(x_i - X_{1,i}) \Gamma(x^i) \varepsilon_i
\]
and
\[
G^*(x_1) = n^{-1/2} \sum_{i=1}^{n} h^{1/2}_1 b_i(x^i)K_{h_{1,n}}(x_i - X_{1,i}) \left\{ \Gamma_n(x^i) - \Gamma(x^i) \right\} \varepsilon_i.
\]

Now we deal with $G(.)$ and $G^*(.)$. By Lemma 2 of Cai and Fan (2000), it turns out that $G^*(x_1) = o_p((nh_{1,n})^{-1/2})$. Thus, (4.34) reduces to
\[
\hat{\theta}_1(x_1) - \theta_1(x_1) - \text{Bias}(\hat{\theta}_1(x_1)) = (nh_{1,n})^{-1/2}G(x_1) + o_p(h_{1,n}^2 + (nh_{1,n})^{-1/2}). \quad (4.35)
\]
Define $\Delta_i = h^{1/2}_1 b_i(x^i)\Gamma(x^i)\varepsilon_iK_{h_{1,n}}(x_i - X_{1,i})$. Then $G(x_1) = n^{-1/2} \sum_{i=1}^{n} \Delta_i$. Now note that showing the desired asymptotic normality result is the same as showing the asymptotic normality of $G(x_1)$. To achieve this, we can routinely apply the familiar Doob’s small block and large block procedure to the sum $n^{-1/2} \sum_{i=1}^{n} \Delta_i$; see Section 3.4 for more details. Finally, to complete the proof of the Theorem we derive $\text{Var}(G_1(x_1))$. Note by stationarity that
\[
\text{Var}(G(x_1)) = \text{Var}(\Delta_1) + 2 \sum_{i=2}^{n} \left( 1 - \frac{i}{n} \right) \text{Cov}(\Delta_1 \Delta_i)
\]
\[
= G_{11} + G_{12}.
\]
Using Condition A.7 and following Lemma 1 of Cai and Fan (2000), one can see that $G_{12} \to 0$. The line of the proof is the same as Masry’s technique of splitting outlined in Section 3.4. It remains to show $\text{Var}(\Delta_1)$. First, note that $E(\Delta_1) = 0$. By conditioning on $(X_{1,1}, X_{2,1}) = (u_1, u_2)$ and taking $t = \theta(x_1, u_2)$, one can see that
\[
E(\Delta_1^2) = h_{1,n} E \left\{ b^2(x_1, u_2)K^2_{h_{1,n}}(x_1 - u_1) \Gamma^2(x_1, u_2) \sigma^2_1(u_1, u_2)p(u_1, u_2)du_1 du_2 \right\}.
\]
By using change of variables, $u_1 = x_1 + h_{1,n}v_1$,

$$E(\Delta_i^2) = \int_{\mathbb{R}} \int_{\mathbb{R}} K^2(v_1)\Gamma^2(x_1, u_2)\sigma^2(x_1 + h_{1,n}v_1, u_2)p(x_1 + h_{1,n}v_1, u_2)dv_1du_2$$

$$\rightarrow k_2 \int_{\mathbb{R}} K^2(v_1)\Gamma^2(x_1, u_2)\sigma_i^2(x_1, u_2)p(x_1, u_2)du_2$$

$$= k_2 p_1(x_1)\alpha(1-\alpha)E\{\Gamma^2(X_1, X_2)|X_1 = x_1\}. $$

The last step follows from the fact that $\sigma_i^2(x_1, u_2) = \alpha(1-\alpha)$. Combining this with (4.35), the proof of the Theorem is complete.

### 4.7 Concluding remarks

In this chapter we investigated the estimation of the conditional quantile of a response when many covariates are involved. In particular, we modeled the conditional quantile of a response as a nonlinear additive function of relevant covariates. Within this set-up, we proposed a nonparametric smoother that can directly compute the unknown functions. In addition to the efficient identification of nonlinear conditional quantile components, the methodology of this chapter is applied to estimate linear parameters of a partially linear model. Throughout this chapter we only discussed how the conditional components can be estimated or fitted. Here we remark how we can adapt the additive set-up and its estimation to generate multi-step quantile predictions from a time series. Let $\{W_t; t \geq 1\}$ be a strictly stationary time series. Suppose, without any loss of generality, that $\{W_t\}$ is Markovian, say of order $m = 2$. Of interest is to predict future values $W_{N+H}$ from observed values $W_N, W_{N-1}, \ldots, W_1$ where $H$ is the prediction horizon. Suppose $H = 2$. Let $X_t = (W_t, W_{t+1})$ and $Y_t = W_{t+3}$ for $t = 1, \ldots, n$ with $n = N - 3$. Then, to obtain $\alpha$-quantile prediction of $W_{N+2}$, we have to estimate $\theta_\alpha(X_{N-1})$. The way to estimate the needed prediction is by assuming $\theta_\alpha(\cdot)$ to be additive in the designs. Thus,

$$\theta_\alpha(X_{N-1}) = \delta + \theta_1(W_{N-1}) + \theta_2(W_N).$$

Now using the data $(X_t, Y_t)$, we just have to use the approach in Section 4.2 to compute the individual components and then the full-dimensional $\theta_\alpha(\cdot)$ by adding the individual estimates.