Single Index Quantile Regression for Heteroscedastic Data

Eliana Christou
Department of Statistics, The Pennsylvania State University
(email: exc277@psu.edu)
and
Michael G. Akritas
Department of Statistics, The Pennsylvania State University
(email: mga@stat.psu.edu)

September 29, 2015
Abstract

Quantile regression (QR) is becoming increasingly popular due to its relevance in many scientific investigations. Linear and nonlinear QR models have been studied extensively, while recent research focuses on the single index quantile regression (SIQR) model. Compared to the single index mean regression problem, the fitting and the asymptotic theory of the SIQR model are more complicated due to the lack of closed form expressions for estimators of conditional quantiles. Consequently, the proposed methods are necessarily iterative. We propose a non-iterative estimation algorithm, and derive the asymptotic distribution of the proposed estimator under heteroscedasticity. For identifiability, we use a parametrization that sets the first coefficient to 1 instead of the typical condition which restricts the norm of the parametric component. This distinction is more than simply cosmetic as it affects, in a critical way, the correspondence between the estimator derived and the asymptotic theory.
1 INTRODUCTION

Ordinary least squares regression plays a prominent role in a wide variety of fields and is a very popular method for modeling the relationship between a \(d\)-dimensional vector of covariates \(X\) and the conditional mean of the response variable \(Y\) given \(X = x\). There are cases, however, where interest lies in certain conditional quantiles; see Koenker and Hallock (2001) for a practical implementation. When the error term is heteroscedastic, a direct approach for estimating conditional quantiles has a number of advantages. Koenker and Bassett (1978) proposed such a direct approach. Letting

\[Q_\tau(Y|x) \equiv Q_\tau(Y|X = x) = \inf\{y : P(Y \leq y|X = x) \geq \tau\}\]

denote the \(\tau\)-th conditional quantile of \(Y\) given \(X = x\), they considered the linear quantile regression model

\[Q_\tau(Y|x) = \beta'x\]  

(1.1)

and used the representation

\[Q_\tau(Y|x) = \arg\min_q \mathbb{E}(\rho_\tau(Y - q)|X = x),\]  

(1.2)

where, for \(0 < \tau < 1\), the function \(\rho_\tau(\cdot)\) is the loss function defined as

\[\rho_\tau(u) = (\tau - I(u < 0))u,\]

to define the estimator \(\hat{\beta}\) as

\[\hat{\beta} = \arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^n \rho_\tau(Y_i - b'X_i).\]

Thus, \(\hat{\beta}'x\) gives the estimator of the \(\tau\)-th conditional quantile under the linear quantile regression model. Observe that for \(\tau = 1/2\), the objective function is the \(L_1\) norm, that is

\[\arg\min_{b \in \mathbb{R}^d} \sum_{i=1}^n \frac{1}{2}|Y_i - b'X_i|,\]

which gives the estimated conditional median. It turns out that quantile regression inherits the well know robustness properties of the median regression; see Pollard (1991).

Koenker and Bassett (1978) studied the asymptotic statistical behavior of the estimated conditional regression quantiles, while Koenker (1994) studied confidence intervals for the regression quantiles, based on the asymptotic theory. He suggested three different construction methods for the confidence interval: the sparsity estimation which involves direct estimation of the sparsity
function, the inversion of rank tests which computes confidence intervals by inverting the rank score test, and the resampling method.

Because the linearity assumption of model (1.1) is quite strict, several authors considered the completely flexible nonparametric estimation of the conditional quantiles. Truong (1989) showed that, under conditions, local median estimators achieve the global optimal rates of Stone (1982) with respect to $L_m$ norms, $0 < m \leq \infty$. Chaudhuri (1991) constructed local polynomial estimators for conditional quantile functions and their derivatives, and also showed that they achieve the optimal nonparametric rates of convergence of Stone (1982) under mild conditions. A local Bahadur type representation was also established by Chaudhuri (1991) when the kernel function is uniform, and this result was later extended to general kernels by Hong (2003). Fan, Hu and Truong (1994) considered a general convex loss function, that includes the mean, median, quantiles, and other robust functionals, and constructed local linear estimators. See also Yu and Jones (1998) who proposed inverting a local linear conditional distribution estimator. Takeuchi, Le, Sears and Smola (2006) presented a nonparametric version of a quantile estimator, which can be obtained by solving a simple quadratic programming problem and provide uniform convergence results. Kong, Linton and Xia (2010) extended Chaudhuri’s (1991) and Hong’s (2003) pointwise Bahadur representation result by deriving a strong uniform (with respect to $x$) Bahadur representation also for dependent observations. Guerre and Sabbah (2012) investigated the bias and the weak Bahadur representation of a local polynomial estimator of the conditional quantile function and its derivatives uniformly with respect to the quantile level, the covariates and the smoothing parameter. Also, they showed that the local polynomial quantile estimator achieves the global optimal rates of Stone (1982) for the $L_m$ and uniform norms.

The rate of convergence of completely nonparametric estimators of conditional quantiles, however, decreases with increasing dimensionality of the covariate vector. This motivated the study of a number of semiparametric models, and of variable selection methods, for quantile regression. Koenker (2011) considered the additive model for quantile regression which includes both parametric and nonparametric components. Lin, Bondell, Zhang and Zou (2013) considered variable selection for nonparametric quantile regression via smoothing spline ANOVA (SS-ANOVA). See Su and Zhang (2012) for a literature review.

The single index quantile regression (SIQR) model has received particular attention. It
specifies that

\[ Q_\tau(Y|x) = Q_{\tau, \beta_1}(Y|\beta_1^t x), \]  

(1.3)

where for any \( d \)-dimensional vector \( b_1 \),

\[ Q_{\tau, b_1}(Y|b_1^t x) = \inf \{ y : P(Y \leq y|b_1^t X = b_1^t x) \geq \tau \}. \]  

(1.4)

For identifiability one imposes certain conditions on \( \beta_1 \), the most common of which is to assume that \( ||\beta_1|| = 1 \), with its first coordinate positive. In this paper, we propose the parametrization which assumes that \( \beta_1 = (1, \beta_1')' \), \( \beta \in \mathbb{R}^{d-1} \); this parametrization is also used in the \texttt{R} package \texttt{np} for the single index mean regression (SIMR) model. This distinction is more than simply cosmetic as it affects, in a critical way, the correspondence between the estimator derived and the asymptotic theory. The advantages of the proposed parametrization are demonstrated in the simulations.

Existing literature considers the SIQR model under homoscedasticity (Wu, Yu and Yu 2010), restricted heteroscedasticity (Chaudhuri, Doksum and Samarov 1997) and general heteroscedasticity (Kong and Xia 2012). Chaudhuri et al. (1997) considered the average derivative quantile regression estimator which, under a SIQR model where the variance function depends only on \( \beta_1^t x \), estimates the direction of \( \beta_1 \). Wu et al. (2010) and Kong and Xia (2012) estimate \( \beta_1 \) by minimizing an objective function. Compared with SIMR, (see Li and Racine 2007, Chapter 8), the main complication faced by this approach lies in the lack of a closed form expression for the estimator of conditional quantiles. Thus, the proposed methods are necessarily iterative. Wu et al. (2010) proposed an algorithm which, starting from an initial value \( b_0^1 \) for the parametric component, iteratively estimates the nonparametric component and its derivative using local linear quantile regression, and the parametric component using (essentially) linear quantile regression. Kong and Xia (2012) criticized the convergence properties of the algorithm in Wu et al. (2010) and proposed an improved iterative algorithm by introducing a penalty term that assures its almost sure convergence. (The outliers in the boxplot of the Wu et al. (2010) estimator shown in Section 4.2 are probably a consequence of the iteration issues of their algorithm.) In addition, Kong and Xia (2012) allowed general heteroscedasticity, but the covariance function of the limiting normal distribution they obtained depends on the true value of the parametric component in an explicit manner.

In this paper we propose a \textit{non-iterative} method, based on minimization of an objective function, for estimating the parametric component of the SIQR model. The proposed estimator is
shown to have an asymptotically normal distribution, with a simple expression for the covariance matrix, under general heteroscedasticity.

In Section 2 we present the proposed estimator, while in Section 3 we present the main results, that include the \( \sqrt{n} \)-consistency and the asymptotic normality of the estimated parametric component. In Section 4 we present results from several simulation examples and a real data application on the Boston housing data. A discussion including conclusions and future work is given in Section 5.

2 THE PROPOSED ESTIMATOR

Let \( \{Y_i, X_i\}_{i=1}^n \) be independent and identically distributed observations that satisfy

\[
Y_i = Q_{\tau, \beta_1}(Y|\beta'_1 X_i) + \epsilon_i,
\]

where \( Q_{\tau, \beta_1}(Y|\beta'_1 X_i) \) is defined in (1.4), and the error term \( \epsilon_i \) satisfies \( Q_\tau(\epsilon_i|X_i) = 0 \). The quantities \( \beta_1 \) and \( \epsilon_i \) are specific to the \( \tau \)-th quantile, but we omit the subscript \( \tau \) for notational convenience. Note that (2.1) is an equivalent way of specifying the SIQR model (1.3). Relation (1.2) implies that the true parametric vector \( \beta \) satisfies

\[
\beta = \arg\min_{b} E(\rho_{\tau}(Y - Q_{\tau, \beta_1}(Y|\beta'_1 X))) \quad \text{where} \quad \beta_1 = (1, \beta'_1)' \tag{2.2}
\]

The sample level version of (2.2) consists of minimizing

\[
\sum_{i=1}^n \rho_{\tau}(Y_i - Q_{\tau, \beta_1}(Y|\beta'_1 X_i)).
\]

As in the single index mean regression (SIMR) problem, the unknown \( Q_{\tau, \beta_1}(Y|\beta'_1 X_i) \) must be replaced with an estimator. Unlike the SIMR problem, however, there is no closed form expression for the estimator of \( Q_{\tau, \beta_1}(Y|\beta'_1 X_i) \), and this has led to iterative algorithms for estimating \( \beta \); see the literature review in Section 1.

To overcome this difficulty, we define, for any given \( b \in \mathbb{R}^{d-1} \), the function \( g(u|b) : \mathbb{R} \to \mathbb{R} \) as

\[
g(u|b) = E(Q_{\tau}(Y|X)|b'_1 X = u), \quad \text{where} \quad b_1 = (1, b')'.
\]

Noting that \( Q_{\tau}(Y|X) = Q_{\tau, \beta_1}(Y|\beta'_1 X) = g(\beta'_1 X|\beta) \), it follows that \( \beta \) also satisfies

\[
\beta = \arg\min_{b} E(\rho_{\tau}(Y - g(b'_1 X|b))). \tag{2.3}
\]

The sample level version of (2.3) consists of minimizing

\[
S_n(\tau, b) = \sum_{i=1}^n \rho_{\tau}(Y_i - g(b'_1 X_i|b)).
\]
Again, \( g(\cdot | \mathbf{b}) \) is unknown but it can be estimated, in a non-iterative fashion, by first obtaining estimators \( \hat{Q}_\tau(Y | \mathbf{X}_i) \), for \( i = 1, \ldots, n \), and forming the Nadaraya-Watson-type estimator

\[
\hat{g}_{NW}^Q(t | \mathbf{b}) = \sum_{i=1}^{n} \hat{Q}_\tau(Y | \mathbf{X}_i) K_2 \left( \frac{t - \mathbf{b}'_1 \mathbf{X}_i}{h_2} \right),
\]

where \( K_2(\cdot) \) is a univariate kernel function and \( h_2 \) is a bandwidth. The different methods for constructing nonparametric estimators \( \hat{Q}_\tau(Y | \mathbf{X}_i) \) are summarized in Racine and Li (2014), who also introduce a new direct method. In this paper we will use the local linear conditional quantile estimator, which was studied in Guerre and Sabbah (2012). Specifically, for a multivariate kernel function \( K_1(\mathbf{x}) = K_1(x_1, \ldots, x_d) \) and a univariate bandwidth \( h_1 \), let

\[
L_n((\alpha_0, \alpha_1); \tau, h_1, \mathbf{x}) = \frac{1}{nh_1^d} \sum_{i=1}^{n} \rho_\tau(Y_i - \alpha_0 - \alpha'_1(X_i - \mathbf{x})) K_1 \left( \frac{X_i - \mathbf{x}}{h_1} \right)
\]

and define \( \hat{Q}_\tau(Y | \mathbf{x}) \) as \( \hat{\alpha}_0(\tau; h_1, \mathbf{x}) \), where \( \hat{\alpha}_0(\tau; h_1, \mathbf{x}) \) is defined through

\[
(\hat{\alpha}_0(\tau; h_1, \mathbf{x}), \hat{\alpha}_1(\tau; h_1, \mathbf{x})) = \arg \min_{(\alpha_0, \alpha_1)} L_n((\alpha_0, \alpha_1); \tau, h_1, \mathbf{x}).
\]

**Remark 2.1.** For high dimensional data it is possible to improve the estimation of conditional quantiles by employing local variable selection methods; see Section 5 for a more detailed description. This methodology will be studied in a forthcoming paper dealing with variable selection in SIQR.

Thus, the proposed estimator is obtained by

\[
\hat{\beta} = \arg \min_{\mathbf{b} \in \Theta} \hat{S}_n(\tau, \mathbf{b}),
\]

where \( \Theta \subset \mathbb{R}^{d-1} \) is a compact set, \( \beta \) is in the interior of \( \Theta \), and

\[
\hat{S}_n(\tau, \mathbf{b}) = \sum_{i=1}^{n} \rho_\tau \left( Y_i - \hat{g}_{NW}^Q(\mathbf{b}'_1 \mathbf{X}_i | \mathbf{b}) \right).
\]

For technical reasons that have to do with the uniform convergence of the Nadaraya-Watson estimator, a trimming function is usually introduced in the objective function (2.7). To avoid complicating the notation, we will assume that the support \( X_0 \) of \( \mathbf{X} \) is compact and the density \( f_\mathbf{b} \) of \( \mathbf{b}'_1 \mathbf{X} \) stays bounded away from zero on \( \mathcal{T}_\mathbf{b} = \{ t : t = \mathbf{b}'_1 \mathbf{x}, \mathbf{x} \in X_0 \} \), uniformly in \( \mathbf{b} \in \Theta \).

### 3 MAIN RESULTS

The first two results, which have to do with the uniform, in both \( t \) and \( \mathbf{b} \), consistency of \( \hat{g}_{NW}^Q(t | \mathbf{b}) \), and the \( \sqrt{n} \)-consistency of \( \hat{\beta} \), are needed for the proof of Theorem 3.3.
PROPOSITION 3.1. Let \( \hat{g}_Q^{NW}(t|b) \) be defined in (2.4). Assume that for some \( k > 2 \), 
\[ E|Q_\tau(Y|X)|^k < \infty \] and \( \sup_{t \in \tau_b} E(|Q_\tau(Y|X)|^k|b_i'X = t)f_b(t) < \infty \) holds for all \( b \in \Theta \), where \( \tau_b = \{ t : t = b_i'x, x \in \mathcal{X}_0 \} \), and that \( Q_\tau(Y|X) \) is in \( H_s(\mathcal{X}_0) \) for some \( s \) with \( [s] = 1 \), where \( H_s(\mathcal{X}_0) \) is defined in Appendix A. Under Assumptions GS1-GS3 and Assumptions A1-A5 given in Appendix A, we have
\[
\sup_{b \in \Theta, t \in \tau_b} |\hat{g}_Q^{NW}(t|b) - g(t|b)| = O_p \left( a_n^* + a_n + h_2^2 \right),
\]
where \( a_n^* = (\log n/n)^{s/(2s+d)} \) and \( a_n = (\log n/(nh_2))^{1/2} \).

The proof of Proposition 3.1 is given in Section B.2.

PROPOSITION 3.2. Let \( \hat{\beta} \) be as defined in (2.6). Then, under the assumptions of Proposition 3.1 and Assumptions A6 and A7 given in Appendix A, \( \hat{\beta} \) is \( \sqrt{n} \)-consistent estimator of \( \beta \).

The proof of Proposition 3.2 is given in Section B.3.

THEOREM 3.3. Let \( \hat{\beta} \) be as defined in (2.6). Then, under the assumptions of Proposition 3.1 and Assumptions A6 and A7 given in Appendix A,
\[
\sqrt{n}(\hat{\beta} - \beta) = -V^{-1}W_n + o_p(1),
\]
where
\[
V = E\left( (g'(\beta'_iX|\beta))^2(X_{-1} - E(X_{-1}|\beta'_iX))(X_{-1} - E(X_{-1}|\beta'_iX))'f_i(X(0|X)) \right), \tag{3.1}
\]
for \( X_{-1} \) the \((d-1)\)-dimensional vector consisting of coordinates \( 2, ..., d \) of \( X \), and
\[
W_n = -n^{-1/2} \sum_{i=1}^n \rho_i(Y^*_i)g'(\beta'_iX|\beta)(X_{i,-1} - E(X_{-1}|\beta'_iX)), \tag{3.2}
\]
for \( Y^*_i = Y_i - \hat{g}_Q^{NW}(\beta'_iX_i|\beta) \). Furthermore,
\[
\sqrt{n}(\hat{\beta} - \beta) \overset{d}{\to} N\left(0, \tau(1-\tau)V^{-1}\Sigma V^{-1}\right),
\]
where
\[
\Sigma = E\left( (g'(\beta'_iX|\beta))^2(X_{-1} - E(X_{-1}|\beta'_iX))(X_{-1} - E(X_{-1}|\beta'_iX))' \right) \tag{3.3}
\]

The proof of Theorem 3.3 is given in Section B.4.

Next, let
\[
(\hat{a}, \hat{b}) = \arg\min_{a, b} \sum_{i=1}^n \rho_i(Y_i - a - b(1, \hat{\beta}')'(X_i - x))K_{\delta}\left(\frac{(1, \hat{\beta}')'(X_i - x)}{\delta}\right),
\]

where $K_3(\cdot)$ is a univariate kernel function and $h_3$ is a bandwidth, and define
\[
\hat{Q}^{LL}_\tau(Y|x) = \hat{Q}^{LL}_\tau(Y|1,\hat{\beta}^\prime)x = \hat{a},
\] (3.4)
as the quantile estimator based on the assumption of the SIQR model (1.3).

**COROLLARY 3.4.** Let $\hat{Q}^{LL}_\tau(Y|x)$ be as defined in (3.4), where $K_3(\cdot)$ is a symmetric, second order and density kernel with a compact support and a bounded first derivative that satisfies
\[
|\int t^j K_3^2(t)dt| < \infty \text{ for } j = 0, 1, 2.
\]
Then, under the assumptions of Proposition 3.1 and Assumptions A6-A8 given in Appendix A,
\[
\sqrt{n}h_3(\hat{Q}^{LL}_\tau(Y|x) - Q_\tau(Y|x) - h_3^2 I(\beta_1^\prime x)) \xrightarrow{d} N(0, \omega^2(\beta_1^\prime x)),
\]
where $I(\beta_1^\prime x) = (1/2)g''(\beta_1^\prime x|\beta)\int t^2 K_3(t)dt$ and
\[
\omega^2(\beta_1^\prime x) = \frac{\tau(1-\tau)\int K_3^2(t)dt}{f_\beta(\beta_1^\prime x)(f_\epsilon|\beta(0|\beta_1^\prime x))^2},
\]
for $h_3 \to 0$ and $nh_3 \to \infty$ as $n \to \infty$.

The corollary follows from the fact that the proposed estimator $\hat{\beta}$ is $\sqrt{n}$ consistent and the proof of Theorem 2 of Wu et al. (2010).

**Remark 3.5.** It can be shown that, when using the parametrization adopted here, the asymptotic normality result for the parametric vector of Wu et al. (2010) is a special case (pertaining to the homoscedastic case) of Theorem 3.3. Oberhofer and Haupt (2015), considered nonlinear quantile regression (thus known link function), in a fixed design with heteroscedastic errors which are allowed to be weakly dependent. Taking these differences into consideration, the form of the limiting covariance matrix they obtained is also related to that of Theorem 3.3. Finally, Kong and Xia (2012), who considered the heteroscedastic case and use a penalty term to ensure the convergence of their iterative algorithm, obtain a covariance matrix that depends on the true value of the parametric component in an explicit manner and whose form is not directly comparable to that of the previous literature or ours.

### 4 NUMERICAL STUDIES

This section contains simulation results and the analysis of a real data set contrasting the proposed estimator with that of Wu et al. (2010).
4.1 Computational Remarks

For the computation of the proposed estimator, the quantile estimators \( \hat{Q}_\tau(Y|x) \) are multivariate local linear conditional quantile estimators, computed using an extension of the code for the function \texttt{lprq} in the \texttt{R} package \texttt{quantreg} (which applies only to univariate covariates). The bandwidth \( h_1 \) is selected to be the rule-of-thumb bandwidth for the local linear estimator derived in Yu and Jones (1998). The bandwidth \( h_2 \) used in (2.4) is selected according to the Cross-Validation criterion. The function \texttt{n1rq} of the same \texttt{R} package was used for minimizing the objective function (2.7). The resulting code is available from the author. For the computation of the Wu et al. (2010) estimator we used the code provided by these authors.

Because the two estimators being compared are derived using different parametrizations for ensuring identifiability, e.g., (2.2) versus the constraint to have a norm of one, for the sake of comparison we found necessary to introduce two modifications of the Wu et al. (2010) estimator. For the first modification, we divide their estimator by its first component, and for the second modification we use our proposed parametrization in conjunction with the iterative algorithm of Wu et al. (2010) for estimating the remaining \( d-1 \) coefficients. In what follows, NWQR denotes the proposed estimator, and WYY, WYY-1 and WYY-2 denote, respectively, the estimator in Wu et al. (2010) and its first and second modification. All simulation results use a sample size of \( n = 400 \) and are based on \( N = 100 \) iterations.

4.2 Simulation Results

Example 1 (Asymmetric Homoscedastic Errors): Here the data are generated according to the model

\[
Y = 5 \cos(\beta_1'X) + \exp(-((\beta_1'X)^2) + \varepsilon,
\]

where \( X = (X_1, X_2)' \), \( X_i \sim U(0,1) \) are iid, \( \beta_1 = (1, 2)' \), the residual \( \varepsilon \) follows an exponential distribution with mean 2 and \( X_i \)'s and \( \varepsilon \) are mutually independent. In this example, we fit the single index median regression model using the proposed method and that of Wu et al. (2010).

The boxplot presented in Figure 1 shows 100 coefficient estimates using the proposed methodology and the two modifications of the Wu et al. (2010) estimator. Observe that the boxplot for NWQR is more closely concentrated around the true value of 2 than the other two boxplots, while the boxplot for WYY-1 displays the widest variability around the true value. The
observed outliers in the boxplot for the WYY-2 estimator (which uses the proposed parametrization) is probably a consequence of the iterative algorithm which stops after a maximum number of iterations.

Figure 1: Boxplot of estimated parametric coefficient for Model (4.1) for the three estimators; the true $\beta$ is 2.

For further comparison, Table 1 reports the observed mean values and standard deviations for the three estimators, as well as the mean square error, $\hat{R}(\hat{\beta})$, and the mean check based
absolute residuals, $R_r(\hat{\beta}_1)$, which are defined as

$$R(\hat{\beta}) = \frac{1}{N} \sum_{i=1}^{N} (\hat{\beta}_i - 2)^2, \quad R_r(\hat{\beta}_1) = \frac{1}{n} \sum_{i=1}^{n} \rho_r(Y_i - \hat{Q}_{\tau,\hat{\beta}_1}(Y | \hat{\beta}_1' X_i)) \quad (4.2)$$

respectively, for each of the three estimators, where $N$ denotes the number of simulation iterations, and $\hat{Q}_{\tau,\hat{\beta}_1}(Y | \hat{\beta}_1' X_i)$ is defined in (3.4). The findings in Table 1 can be further confirm the conclusions drawn from the boxplot in Figure 1. We observe that NWQR gives the smallest bias and the smallest value of $R(\hat{\beta})$ and $R_r(\hat{\beta}_1)$, followed by WYY-2, which uses the proposed constraint, and then WYY-1.

Table 1: mean values and standard errors (in parenthesis), $R(\hat{\beta})$ and $R_r(\hat{\beta}_1)$ defined in (4.2), for Model (4.1).

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}$</th>
<th>$R(\hat{\beta})$</th>
<th>$R_r(\hat{\beta}_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWQR</td>
<td>2.0199 (0.1619)</td>
<td>0.0264</td>
<td>0.6855</td>
</tr>
<tr>
<td>WYY-1</td>
<td>2.1328 (0.7730)</td>
<td>0.6092</td>
<td>0.6941</td>
</tr>
<tr>
<td>WYY-2</td>
<td>1.9763 (0.2834)</td>
<td>0.0801</td>
<td>0.6908</td>
</tr>
</tbody>
</table>

Finally, we compare the coverage of 95% confidence intervals based on NWQR, WYY-2, and the second estimated coefficient of Wu et al. (2010) (which estimates $2/\sqrt{5}$). The WYY-1 estimator (i.e., $\hat{\beta}_1/\hat{\beta}_1$) was not considered due to the additional complication presented by the ratio. The resulting coverage probabilities are 0.96, 0.91 and 0.68 for NWQR, WYY-2 and the second estimated coefficient of Wu et al. (2010), respectively. The p-values corresponding to 0.91 and 0.68 are 0.07 and $3.02 \times 10^{-35}$, respectively. The conclusion is that the marginal standard error formula given in Wu et al. (2010) for the second component of $\beta_1$ is appropriate for the parametrization used in the present paper.

Example 2 (Symmetric Homoscedastic Errors): Here the data are generated according to the model

$$Y = \exp(\beta_1' X) + \epsilon, \quad (4.3)$$

where $X = (X_1, X_2)'$, $X_i \sim N(0, 1)$ are iid, $\beta_1' = (1, 2)'$ and the residual $\epsilon$ follows a standard normal distribution. In this example, we fit the single index quantile regression model for five different quantile levels, $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, using the proposed method and that of Wu et al. (2010).
The boxplots for Model (4.3) are not presented here since they reveal similar conclusions as in Example 1. Table 2 presents the observed mean values and standard deviations for the three estimators, as well as the mean square error, $R(\hat{\beta})$, and the mean check based absolute residuals, $R_\tau(\hat{\beta}_1)$. NWQR is more closely concentrated around 2 and gives the smallest $R(\hat{\beta})$ and $R_\tau(\hat{\beta}_1)$ values for all quantile levels. Note that for $\tau = 0.9$, the NWQR estimator has a mean value of 2.0768, while WYY-2 has a mean value of 2.0287. This is due to an outlier presented for NWQR, while it still gives the smallest bias. Also, observe the very large values of $\hat{\beta}$ and $R_\tau(\hat{\beta}_1)$ for $\tau = 0.5$ for WYY-1. This is due to a very extreme outlier (around 140) which can be observed from the boxplot and which is a consequence of the iterative algorithm. Table 2 also reports the coverage of 95% confidence intervals based on NWQR, WYY-2, and the second estimated coefficient of Wu et al. (2010) denoted by WYY in the table. Similar conclusions as in Example 1 can be drawn regarding the performance of these confidence intervals with the additional remark that the coverage probability of the WYY-2 intervals deteriorates for the 75th and 90th percentiles.

Example 3 (Heteroscedastic Errors): Here the data are generated according to the model

$$
Y = \sin(2\pi \beta_1'X) + \frac{(1 + (\beta_2'X)^2)}{4}\epsilon,
$$

(4.4)

where $X = (X_1, X_2)'$, $X_i \sim U(0, 1)$ are iid, $\beta_1 = (1, 2)'$, $\beta_2 = (1, 1)'$, and the residual $\epsilon$ follows an exponential distribution with mean 1. We fit the single index quantile regression model for five different quantile levels, $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, using the proposed method and that of Wu et al. (2010). Table 3 reports the coverage of 95% confidence intervals based on NWQR, WYY-2, and WYY. The mean values and standard deviations, as well as $R(\hat{\beta})$ and $R_\tau(\hat{\beta}_1)$ for the three estimators are not reported here since they reveal the same trends as in the previous two examples.

From Table 3 we observe that the coverage probabilities of the NWQR intervals are close to the true nominal value of 0.95, while those of the WYY-2 estimator tend to be smaller than the nominal value, probably a consequence of the heteroscedasticity.

4.3 Boston Housing Data

For this example we consider an application regarding Boston housing data. The data contains 506 observations on 14 variables, for which the dependent variable of interest is $\text{medv}$, the me-
Table 2: mean values and standard errors (in parenthesis), \( R(\hat{\beta}) \) and \( R_\tau(\hat{\beta}_1) \) defined in (4.2) for Model (4.3). Also, coverage probability for NWQR, WYY-2 and the second estimated coefficient of Wu et al. (2010), denoted by WYY.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>NWQR</th>
<th>WYY-1</th>
<th>WYY-2</th>
<th>NWQR</th>
<th>WYY-1</th>
<th>WYY-2</th>
<th>NWQR</th>
<th>WYY-1</th>
<th>WYY-2</th>
<th>NWQR</th>
<th>WYY-1</th>
<th>WYY-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2.0085</td>
<td>1.9521</td>
<td>1.9699</td>
<td>0.0037</td>
<td>0.0652</td>
<td>0.0260</td>
<td>0.1682</td>
<td>0.1783</td>
<td>0.1720</td>
<td>0.99</td>
<td>0.93</td>
<td>0.89</td>
</tr>
<tr>
<td>0.25</td>
<td>2.0132</td>
<td>1.9890</td>
<td>1.9899</td>
<td>0.0024</td>
<td>0.0503</td>
<td>0.0170</td>
<td>0.3096</td>
<td>0.3273</td>
<td>0.3160</td>
<td>0.97</td>
<td>0.90</td>
<td>0.84</td>
</tr>
<tr>
<td>0.5</td>
<td>2.0149</td>
<td>3.4450</td>
<td>2.0158</td>
<td>0.0025</td>
<td>192.6956</td>
<td>0.1537</td>
<td>0.3909</td>
<td>0.4154</td>
<td>0.4031</td>
<td>0.97</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>0.75</td>
<td>2.0290</td>
<td>1.9363</td>
<td>2.0525</td>
<td>0.0055</td>
<td>0.1084</td>
<td>0.4518</td>
<td>0.3113</td>
<td>0.3479</td>
<td>0.3310</td>
<td>0.97</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td>0.9</td>
<td>2.0768</td>
<td>1.9592</td>
<td>2.0287</td>
<td>0.1569</td>
<td>0.3469</td>
<td>0.5197</td>
<td>0.1761</td>
<td>0.2265</td>
<td>0.2050</td>
<td>0.95</td>
<td>0.70</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Table 3: Coverage probability for NWQR, WYY-2, and the second estimated coefficient of Wu et al. (2010), denoted by WYY, for Model (4.4).

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>NWQR</th>
<th>WYY-2</th>
<th>WYY</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.97</td>
<td>0.86</td>
<td>0.89</td>
</tr>
<tr>
<td>0.25</td>
<td>0.95</td>
<td>0.93</td>
<td>0.84</td>
</tr>
<tr>
<td>0.5</td>
<td>0.93</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>0.75</td>
<td>0.97</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>0.9</td>
<td>0.97</td>
<td>0.64</td>
<td>0.45</td>
</tr>
</tbody>
</table>

dian value of owner-occupied homes in $1000s, and the other thirteen variables are statistical measurements on the 506 census tracts in suburban Boston from the 1970 census. The data was
originally published by Harrison and Rubinfeld (1978). This data set can be found in the MASS library in R.

Quantile regression is appropriate for this data set because the response variable is the median price of homes and because $y$-values larger than or equal to $50,000 have been recorded as $50,000. As was noted in Chaudhuri et al. (1997, p. 724), “such a truncation in the upper tail of the response makes quantile regression, which is not influenced very much by extreme values of the response, a very appropriate methodology”.

Due to the collinearity in the data set, Breiman and Friedman (1985) applied their alternating conditional expectation (ACE) method for selecting the relevant variables and selected the four covariates RM, TAX, PTRATIO, and LSTAT, for which the description is given below. Many regression studies (Opsomer and Ruppert 1998; Yu and Lu 2004; Wu et al. 2010) have used this data set and, using a logarithmic transformation on the covariates TAX and LSTAT, found potential relationship between the response medv and these four covariates. Opsomer and Ruppert (1998) considered mean regression and fitted the additive model after removing the observations with outliers on the covariates TAX and LSTAT. Yu and Lu (2004) fitted an additive quantile regression model and Wu et al. (2010) considered the single index quantile regression model. In addition, many studies (Chaudhuri et al. 1997; Wu et al. 2010) considered the relationship between medv and the three covariates RM, LSTAT, DIS, for which the description is given below. Chaudhuri et al. (1997) considered the average derivative quantile regression, while Wu et al. (2010) considered the single index quantile regression.

We apply our methodology using the above two sets of predictors. First, consider the four covariates:

- **RM**: average number of rooms per house in the area
- **TAX**: full-value property tax (in dollar) per $10,000
- **PTRATIO**: pupil-teacher ratio by town
- **LSTAT**: percentage of the population having lower economic status in the area.

Following previous studies, we take logarithmic transformations on TAX and LSTAT, and center the dependent variable. Let $X_1$, $X_2$, $X_3$ and $X_4$ denote the standardized RM, log(TAX), PTRATIO and log(LSTAT), respectively, and set $X = (X_1, X_2, X_3, X_4)$. We consider the single index quantile
regression models:

\[ Q_\tau(\text{medv}|X) = g(X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 | \theta), \]  

for \( \tau = 0.1, 0.25, 0.5, 0.75, 0.9 \), which assume that RM is a significant predictor. An analysis assuming different significant predictor gave the same results. Note that the dependence of \( g \) and \( \beta \) on \( \tau \) is not made explicit.

Table 4 gives the estimators and their standard errors, resulting from the proposed methodology. The conclusions derived from this table are: a) PTRATIO seems to have a significant contribution only on the 90th quantile; b) none of TAX, PTRATIO, LSTAT appear to have a significant contribution on the 50th and 75th quantiles; c) TAX is more significant than LSTAT for the 10th and 25th quantiles; d) LSTAT is more significant than TAX for the 90th quantile; e) RM seems to have an opposite effect than that of the other predictors. Conclusions a)-d) differ from those in the aforementioned literature. The difference with the conclusions in Yu and Lu (2004) is largely because they considered only the absolute value of the coefficients instead of the coefficient’s \( t \)-values. Similarly, the conclusions of Wu et al. (2010) regarding the relative significance of the predictors are based on the absolute value of their coefficients, even though they did compute standard errors (based on bootstrap instead of their variance formulas). Finally, the results of Opsomer and Ruppert (1998) are not directly comparable to ours because they considered mean regression using the additive model.

In order to compare the performance of the proposed estimator with that of Wu et al. (2010), we consider the lack-of-fit statistic \( R_\tau(\hat{\beta}_1) \) as defined in (4.2) in connection with (3.4). From the left panel of Table 6, we note that both methods perform similarly, with the proposed estimator achieving somewhat smaller values for all quantiles.

Figure 2 presents the estimated conditional quantiles, along with scatter plots of \( y \) and the estimated indices. Note that there is an indication of heteroscedasticity, an assumption that is not made in the work of Wu et al. (2010).

Next, we consider the three covariates RM, LSTAT and DIS where:

- **RM**: average number of rooms per house in the area
- **LSTAT**: percentage of the population having lower economic status in the area
- **DIS**: weighted distance to five Boston employment centers from houses of the area.
Table 4: Parametric vector estimates and standard errors (in parenthesis) for Boston housing data for Models (4.5) with five different quantile levels.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>RM</th>
<th>log(TAX)</th>
<th>PTRATIO</th>
<th>log(LSTAT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
<td>-1.1274</td>
<td>-0.6098</td>
<td>-1.1366</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2431)</td>
<td>(0.3521)</td>
<td>(0.4038)</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>-0.7052</td>
<td>-0.5828</td>
<td>-1.2833</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.2325)</td>
<td>(0.4518)</td>
<td>(0.4861)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>-0.4984</td>
<td>-0.4497</td>
<td>-1.0197</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3286)</td>
<td>(0.6176)</td>
<td>(0.6765)</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>-0.3323</td>
<td>-0.3961</td>
<td>-0.8692</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3585)</td>
<td>(0.6414)</td>
<td>(0.7184)</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>-0.6809</td>
<td>-1.6961</td>
<td>-5.2774</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0452)</td>
<td>(0.0995)</td>
<td>(0.0936)</td>
</tr>
</tbody>
</table>

Let $X_1$, $X_2$, and $X_3$ denote the standardized $\text{RM}$, $\text{LSTAT}$ and $\text{DIS}$, respectively, and consider the single index quantile regression models:

$$Q_\tau(\text{medv}|X) = g(X_1 + \beta_2 X_2 + \beta_3 X_3 | \beta)$$

for $\tau = 0.1, 0.25, 0.5, 0.75, 0.9$, which assume that $\text{RM}$ is a significant predictor, an assumption which is confirmed by the previous studies. An analysis assuming that $\text{LSTAT}$ is the significant predictor gave the same results.

Table 5 gives the estimators and their standard errors, resulting from the proposed methodology. The conclusions derived from this table are: a) $\text{LSTAT}$ seems to be the most important covariate for all quantile levels by comparing the coefficient’s $t$-values; b) the effect of $\text{LSTAT}$ is essentially stable across different quantiles; c) $\text{DIS}$ seems to have a significant contribution only on the 90th quantile. Conclusions a) and b) are in conformity with the ones of Chaudhuri et al. (1997) and Wu et al. (2010), with the difference that the aforementioned authors draw their conclusions by comparing only the absolute values of the normalized coefficients without calculating the coefficient’s $t$-values. Our conclusion c), however, is a more definitive statement regarding the relevance of $\text{DIS}$, since none of the above investigators mentioned anything about the significance of $\text{DIS}$.
Figure 2: Estimated single index quantile regression for Boston housing data for Models (4.5). The dots are the observations and the curve is the estimated link function.

The right panel of Table 6 gives $R_\tau(\hat{\beta}_1)$ for the five different quantile levels and the two different methods. We remark that Wu et al. (2010) display the $R_\tau(\hat{\beta}_1)$ values resulting from the quantile average derivative estimate (qADE) of Chaudhuri et al. (1997). Because these values are considerably higher, they are not displayed in Table 6 for either model. For $\tau = 0.1, 0.25, 0.5, 0.75$ the two methods give similar $R_\tau(\hat{\beta}_1)$ values with NWQR resulting in somewhat larger values for $\tau = 0.1, 0.25$ and somewhat smaller for $\tau = 0.5, 0.75$; for $\tau = 0.9$, however, NWQR results in considerably lower $R_\tau(\hat{\beta}_1)$ value. In terms of comparing the two models, Model (4.5) seems
Table 5: Parametric vector estimates and standard errors (in parenthesis) for Boston housing data for Models (4.6) with five different quantile levels.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>RM</th>
<th>LSTAT</th>
<th>DIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
<td>-2.6814</td>
<td>-0.0622</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3261)</td>
<td>(0.2562)</td>
</tr>
<tr>
<td>0.25</td>
<td>1</td>
<td>-2.8518</td>
<td>-0.2340</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3944)</td>
<td>(0.3745)</td>
</tr>
<tr>
<td>0.5</td>
<td>1</td>
<td>-2.2955</td>
<td>-0.2867</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.6222)</td>
<td>(0.4697)</td>
</tr>
<tr>
<td>0.75</td>
<td>1</td>
<td>-2.3256</td>
<td>-0.4659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.4384)</td>
<td>(0.3328)</td>
</tr>
<tr>
<td>0.9</td>
<td>1</td>
<td>-2.3858</td>
<td>-0.5138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.3776)</td>
<td>(0.2442)</td>
</tr>
</tbody>
</table>

better for $\tau = 0.1, 0.25, 0.5$ according to both WYY and NWQR, while Model (4.6) seems better for $\tau = 0.75$ according to WYY and for $\tau = 0.75, 0.9$ for NWQR. Lastly, our method suggests that the best fit corresponds to the lower quantile $\tau = 0.1$ for both models, a conclusion which is in conformity with the aforementioned literature.

Table 6: Comparison of average check absolute residuals $R_\tau(\hat{\beta}_1)$ defined in (4.2) for Models (4.5) and (4.6); WYY denotes the method proposed by Wu et al. (2010) and NWQR denotes the proposed estimator.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>WYY</th>
<th>NWQR</th>
<th>WYY</th>
<th>NWQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.102</td>
<td>1.093</td>
<td>1.228</td>
<td>1.245</td>
</tr>
<tr>
<td>0.25</td>
<td>2.105</td>
<td>2.085</td>
<td>2.229</td>
<td>2.243</td>
</tr>
<tr>
<td>0.5</td>
<td>2.845</td>
<td>2.786</td>
<td>2.874</td>
<td>2.853</td>
</tr>
<tr>
<td>0.75</td>
<td>2.577</td>
<td>2.482</td>
<td>2.490</td>
<td>2.421</td>
</tr>
<tr>
<td>0.9</td>
<td>1.749</td>
<td>1.709</td>
<td>3.320</td>
<td>1.509</td>
</tr>
</tbody>
</table>
Finally, Figure 3 presents the estimated conditional quantiles, along with the scatter plots of \( y \) and the estimated indices for the different quantiles. Again, there is an evidence of heteroscedasticity.

![Figure 3: Estimated single index quantile regression for Boston housing data for Models (4.6). The dots are the observations and the curve is the estimated link function.](image)
5 DISCUSSION

In this work, we have considered modeling conditional quantiles because they are of primary interest in many scientific investigations. Moreover, they provide a more complete picture of the conditional distribution and share nice properties such as robustness. The existing literature on SIQR relies on iterative algorithms for estimating the parametric component of the model. Such algorithms either have convergence issues or produce an estimator whose limiting covariance matrix has an undesirable form. We have proposed a method that directly estimates the parametric component non-iteratively and have derived its asymptotic distribution under heteroscedastic errors. Moreover, we use a parametrization that corresponds to the asymptotic theory in the sense that, as demonstrated by simulations, correct coverage of the confidence intervals is achieved only for the proposed parametrization. In fact, to the best of our knowledge the simulations reported in Section 4.2 are the first showing coverage probabilities of confidence intervals based on the asymptotic distribution of the estimators. An illustration using the Boston housing data has been presented. In future work we will consider the use of penalty terms both for local variable selection to estimate $Q_{\tau}(Y|x)$ and for producing the estimator of the parametric component of the SIQR model. Specifically, for the local variable selection procedure, we will consider minimizing, with respect to $(\alpha_0, \alpha_1)$, the objective function

$$L_n((\alpha_0, \alpha_1); \tau, h_1, x) = \sum_{i=1}^n \rho_\tau(Y_i - \alpha_0 - (X_i - x)'\alpha_1) \prod_{j=1}^d K_1 \left( \frac{X_{ij} - x_j}{h_1} \right) + nh_1^d \lambda_1 \sum_{j=1}^d \hat{w}_j(x)|\alpha_{1j}|,$$

where $K_1(t)$ is a univariate kernel function, $h_1$ is a bandwidth, $\lambda_1$ is a tuning parameter and $\hat{w}_j(x) = |\partial \hat{Q}_\tau(Y|x)/\partial x_j|^{-1}$ for $j = 1, \ldots, d$, where $\hat{Q}_\tau(Y|x)$ is the local linear conditional quantile estimator defined in connection with (2.5). The estimated vector $\hat{\alpha}_1$ can be used for local variable selection in the sense that its nonzero elements correspond to the locally relevant variables. Let $X^*$ denote the vector of relevant variables, i.e., the vector with components $X_j$, with $j \in \hat{A}_n(x) = \{j \in (1, \ldots, d) : \hat{\alpha}_{1j} \neq 0\}$, and let $\hat{d}^*$ be the dimension of $X^*$. Then, the the local linear conditional quantile estimator $\hat{Q}^{VS}_\tau(Y|x)$ is obtained as $\hat{Q}^{VS}_\tau(Y|x) = c_0$, where

$$(c_0, c_1) = \arg \min_{c_0, c_1} \sum_{i=1}^n \rho_\tau(Y_i - c_0 - c_1'(X_i^* - x^*))K^* \left( \frac{X_i^* - x^*}{h^*} \right)$$

and $K^*(\cdot)$ is a $\hat{d}^*$-dimensional kernel function and the bandwidth $h^*$ is, for simplicity, univariate. Thus, $\hat{Q}^{VS}_\tau(Y|x)$ is a penalized version of the classical local linear estimator, using the relevant
variables. Finally, estimation of $\beta$ will also be done using a penalty function, and the asymptotic properties of the resulting estimator will be studied.

A APPENDIX: ASSUMPTIONS

Notation: We say that a function $m(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ has the order of smoothness $s$ on $\mathcal{X}_0$, denoted by $m(\cdot) \in H_s(\mathcal{X}_0)$, if it is differentiable up to order $[s]$, where $[s]$ denotes the lowest integer part of $s$, and there exists a constant $L > 0$, such that for all $v = (v_1, ..., v_d)$ with $|v| = v_1 + ... + v_d = [s]$, all $\tau$ in $[\underline{\tau}, \overline{\tau}]$, and all $x, x'$ in $\mathcal{X}_0$,

$$|D^v m(x) - D^v m(x')| \leq L \|x - x'\|^{s-[s]},$$

where $D^v m(x)$ denotes the partial derivative $\partial^{(v)} m(x)/\partial x_1^{v_1}...x_d^{v_d}$.

Assumptions GS1-GS3 come from the work of Guerre and Sabbah (2012) and are necessary for the uniform consistency of $\hat{Q}_\tau(Y|x)$ defined in connection with (2.5).

Assumption GS1: The distribution of $X$ has a probability density function $f_X(\cdot)$ with respect to the Lebesgue measure, which is strictly positive and continuously differentiable over the compact support $\mathcal{X}_0$ of $X$.

Assumption GS2: The cumulative distribution function $F_{Y|X}(\cdot|\cdot)$ of $Y$ given $X$ has a continuous probability density function $f_{Y|X}(y|x)$ with respect to the Lebesgue measure, which is strictly positive for $y$ in $\mathbb{R}$ and $x$ in $\mathcal{X}_0$. The partial derivative $\partial F_{Y|X}(y|x)/\partial x$ is continuous over $\mathbb{R} \times \mathcal{X}_0$. There is a $L_0 > 0$, such that

$$|f_{Y|X}(y|x) - f_{Y|X}(y'|x')| \leq L_0 \| (x, y) - (x', y') \|$$

for all $(x, y), (x', y')$ of $\mathcal{X}_0 \times \mathcal{X}_0$.

Assumption GS3: The nonnegative kernel function $K_1(\cdot)$ is Lipschitz over $\mathbb{R}^d$, has a compact support and satisfies $\int K_1(z)dz = 1$. For some $K > 0$, $K_1(z) \geq KI(z \in B(0,1))$ where $B(0,1)$ is the closed unit ball. The bandwidth is in $[\underline{h_1}, \overline{h_1}]$ with $0 < \underline{h_1} \leq \overline{h_1} < \infty$, $\lim_{n \rightarrow \infty} \overline{h_1} = 0$ and $\lim_{n \rightarrow \infty}(\log n)/(nh_1^d) = 0$.

Assumption A1: Let $F_{\epsilon|X}(\cdot|x)$ denotes the distribution function of $\epsilon$ given $X = x$, which satisfies $F_{\epsilon|X}(0|x) = P(\epsilon \leq 0|X = x) = \tau$ for $0 < \tau < 1$, and has a density $f_{\epsilon|X}(t|x)$ which is continuous in $x$ for each $t$.

Assumption A2: The density $f_b(t)$ of $b_1'X$ (usually called the “marginal density of $X$ in direction $b_1$”) is uniformly bounded for $t \in \mathcal{T}_b = \{ t : t = b_1'x, x \in \mathcal{X}_0 \}$ and all $b \in \Theta$.
that is \( \sup_{b \in \Theta, t \in \mathbb{T}_b} f_b(t) < \infty \), and bounded away from 0, uniformly in \( \mathbb{T}_b \) and \( \Theta \), that is \( \inf_{b \in \Theta, t \in \mathbb{T}_b} f_b(t) > 0 \). Moreover, the density function of \( b'X \) is uniformly continuous for \( b \) in a neighborhood of \( \beta \).

**Assumption A3:** Let \( K_2(t) : \mathbb{R} \to \mathbb{R} \) be a univariate, symmetric, second order kernel function, for which \( K_2(t) \) satisfies:

(a) \( |K_2(t)| \leq K < \infty \), \( \int_{\mathbb{R}} |K_2(t)| dt \leq \mu < \infty \), \( \int_{\mathbb{R}} t^2|K_2(t)| dt < \infty \), and

(b) for some \( \Lambda_1 < \infty \) and \( \Lambda_2 < \infty \), either \( K_2(t) = 0 \) for \( |t| > \Lambda_2 \) and for all \( t, t' \in \mathbb{R} \) \( |K_2(t) - K_2(t')| \leq \Lambda_1 |t - t'| \), or \( K_2(t) \) is differentiable, \( |(\partial/\partial t)K_2(t)| \leq \Lambda_1 \), and for some \( \nu > 1 \), \( |(\partial/\partial t)K_2(t)| \leq \Lambda_1 |t|^{-\nu} \) for \( |t| > \Lambda_2 \).

**Assumption A4:** The first two derivatives of \( f_b(t) \) and \( \Psi(t|b) = f_b(t)g(t|b) \) are uniformly continuous and are bounded uniformly in \( b \).

**Assumption A5:** The bandwidth \( h_2 = o(1) \) satisfies \( \log n/(nh_2) = o(1) \).

**Assumption A6:** Define \( \varphi(t|x) = \mathbb{E}(\rho_r(\epsilon + t)|X = x) \), for which the expectation and differentation can be interchanged, and let the first, second, and third derivatives of \( \varphi(t|x) \) with respect to \( t \) exist, such that \( \varphi(0|x), \varphi'(0|x), \varphi''(0|x) \), as functions of \( x \), are bounded and continuous in a neighborhood of \( x \) for all small \( t \), and \( \varphi'''(0|x) \) is bounded as a function of \( t \). Moreover, \( \varphi'''(t|x) \) is continuous as a function of \( t \) in a neighborhood of \( 0 \) uniformly in \( x \).

**Assumption A7:** The function \( Q_{\tau, \beta_1}(Y|t) \), defined in (1.3), is smooth in a neighborhood of \( \beta \), for \( t \in \mathbb{T}_{\beta} \), such that the first and second derivatives with respect to \( t \) exist and are bounded.

**Assumption A8:** The conditional density function of \( \epsilon \) given \( b'X \), \( f_{\epsilon|b}(\epsilon|t) \) is continuous in \( t \) for each \( \epsilon \). Moreover, there exist positive constants \( \varepsilon \) and \( \delta \) and a positive function \( G(\epsilon|t) \) such that

\[
\sup_{|t_n - t| \leq \varepsilon} f_{\epsilon|b}(\epsilon|t) \leq G(\epsilon|t), \quad \int |\rho_r'(\epsilon + Q_{\tau, \beta_1}(Y|\beta_1'x) - Q_{\tau, \beta_1}(Y|t)|^{2+\delta}G(\epsilon|t)\,d\epsilon < \infty, \quad \text{and}
\]

\[
\int (\rho_r(\epsilon + Q_{\tau}(Y|x) - u) - \rho_r(\epsilon + Q_{\tau}(Y|x)) - \rho_r'(\epsilon + Q_{\tau}(Y|x))u)^2 G(\epsilon|t)\,d\epsilon = o(u^2)
\]

as \( u \to 0 \).

**Comments on Assumption A:** Assumption A1 allows to work under possible dependence between the covariate \( X \) and the error term \( \epsilon \). Assumptions A2-A5 come from the work of Hansen (2008), in order to ensure the uniform convergence of the density estimator \( \hat{f}_{b_1}(t) \) and of \( \hat{\Psi}(t|b) \); see proof of Proposition 3.1. Assumption A6 imposes smoothness conditions on \( \varphi(\cdot|x) \), since \( \rho_r(\cdot) \) is actually not differentiable at 0. Finally, A7 is a common assumption for the link
function and Assumption A8 is required by the dominated convergence theorem for the proof of Corollary 3.4; see Fan et al. (1994) and Wu et al. (2010).

**B APPENDIX: PROOF OF MAIN RESULTS**

For the study of the asymptotic properties for the parametric vector, we consider an equivalent objective function. Observe that by adding and subtracting the quantity \( \hat{g}^{NW}_Q \beta ' X_i | \beta \) in the objective function \( \hat{S}_n(\tau, b) \) in (2.7), we get

\[
\hat{S}_n(\tau, b) = \sum_{i=1}^{n} \rho(\tau Y_i^* - \tilde{g}(X_i | b, \beta)) = \sum_{i=1}^{n} \rho(\tau Y_i^* - \tilde{g}(X_i | b, \beta)),
\]

where \( Y_i^* = Y_i - \hat{g}^{NW}_Q \beta ' X_i | \beta \) and, for any \( \gamma \in \mathbb{R}^{d-1} \) such that \( \gamma + \beta \in \Theta \), we define

\[
\tilde{g}(X_i | \gamma + \beta, \beta) = \hat{g}^{NW}_Q ((\gamma + \beta)' X_i | \gamma + \beta) - \hat{g}^{NW}_Q (\beta ' X_i | \beta),
\]

where, according to the convention used, \( (\gamma + \beta)' = (1, (\gamma + \beta))' \). For the sake of convenience in the derivation of the asymptotic results we replace the objective function (B.1) with

\[
A_n(\tau, \sqrt{n}(b - \beta)) = \sum_{i=1}^{n} (\rho(\tau Y_i^* - \tilde{g}(X_i | b, \beta)) - \rho(\tau Y_i^*)).
\]

For what follows, \( P(\cdot | X) \) and \( E(\cdot | X) \) will denote the conditional probability and conditional expectation, respectively, on the design matrix \( X \).

**B.1 Some Lemmas**

**LEMMA B.1. (Corollary 1 (ii)-Guerre and Sabbah (2012)):** Assume that \( Q_\tau(Y | x) \) is in \( H_s(\mathcal{X}_0) \) for some \( s \) with \( \lfloor s \rfloor = 1 \), where \( H_s(\mathcal{X}_0) \) is defined in Appendix A. Suppose that Assumptions GS1-GS3, given in the Appendix A, hold and let \( \hat{Q}_\tau(Y | x) \) defined in connection with (2.5). Then for all \( \tau \) in an interval \( [\underline{\tau}, \overline{\tau}] \),

\[
\sup_{x \in \mathcal{X}_0} | \hat{Q}_\tau(Y | x) - Q_\tau(Y | x) | = O_p(\frac{\log n}{n})^s/(2s+d) = O_p(a^*_n)
\]

if \( h_1 \) is asymptotically proportional to \( (\log n/n)^{1/(2s+d)} \).

**Proof:** See Guerre and Sabbah (2012). ■

**LEMMA B.2. (Convexity Lemma - Pollard (1991)):** Let \( \{A_n(u) : u \in U\} \) be a sequence of real valued random convex functions defined on a convex, open subset \( U \subseteq \mathbb{R}^d \). Suppose \( A(u) \)
is a real-valued function on $U$ for which $A_n(u) \to A(u)$ in probability, for each $u \in U$. Then, for each compact subset $K$ of $U$,
\[
\sup_{u \in K} |A_n(u) - A(u)| \xrightarrow{p} 0.
\]
The function $A(\cdot)$ is necessarily convex on $U$.


**LEMMA B.3.** (Quadratic Approximation Lemma - Hjort and Pollard (1993)): Suppose $A_n(u)$ is convex and can be represented as $(1/2)u'Vu + U_n'u + C_n + r_n(u)$, where $V$ is symmetric and positive definite, $U_n$ is stochastically bounded, $C_n$ is arbitrary, and $r_n(u)$ goes to zero in probability for each $u$. Then, $\alpha_n$, the argmin of $A_n$ is only $o_p(1)$ away from $\beta_n = -V^{-1}U_n$, the argmin of $(1/2)u'Vu + U_n'u + C_n$. If also $U_n \xrightarrow{d} U$, then $\alpha_n \xrightarrow{d} -V^{-1}U$.


**LEMMA B.4.** (Uniform Law of Large Numbers for Triangular Arrays): Suppose (a) $\Theta$ is compact, (b) $g(X_{ni}, \theta)$ is continuous at each $\theta \in \Theta$ with probability one, (c) $g(X_{ni}, \theta)$ is dominated by a function $G(X_{ni})$, i.e. $|g(X_{ni}, \theta)| \leq G(X_{ni})$, and (d) $\sup_n E G(X_{ni}) < \infty$. Then,
\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} (g(X_{ni}, \theta) - E g(X_{ni}, \theta)) \right| \xrightarrow{p} 0.
\]


**LEMMA B.5.** Let $\hat{g}_Q^{NW}(t|b)$ be as defined in (2.4). Then, under the assumptions of Proposition 3.1,
\[
\nabla_b \hat{g}_Q^{NW}(b_f|b) = g'(\beta'_f X|\beta)(X_{-1} - E(X_{-1} X'_f)) + O_p(n^{-1})
\]
for $b$ in a $\sqrt{n}$-neighborhood of $\beta$.

Proof: Using the interchange of the limit and differentiation (Rudin 1964, Theorem 7.17) and the uniform consistency of $\hat{g}_Q^{NW}$ derived in Proposition 3.1, enough to show that
\[
\nabla_b g(b_f|b) = g'(\beta'_f X|\beta)(X_{-1} - E(X_{-1} X'_f)) + O_p(n^{-1})
\]
for $b$ in a $\sqrt{n}$-neighborhood of $\beta$. Using Taylor expansion, we have that
\[
g(b_f|\beta) = g(\beta'_f X|\beta) + (b_1 - \beta_1)'x g'(\beta'_f X|\beta) + O_p(n^{-1}).
\]
Using the definition of $g$ and relation (B.4), we get that

$$g(b'_1|x|b) = E(Q_\gamma(Y|X)|b'_1X) = E(g(\beta'_1X|\beta)|b'_1X)$$

$$= E \left( g(b'_1X|\beta) - (b_1 - \beta_1)'Xg'(\beta'_1X|\beta) + O_p(n^{-1})|b'_1X) \right)$$

$$= g(b'_1X|\beta) - (b_1 - \beta_1)'E(Xg'(\beta'_1X|\beta)|b'_1X) + O_p(n^{-1})$$

$$= g(\beta'_1X|\beta) + (b_1 - \beta_1)'xg'(\beta'_1X|\beta)$$

$$- (b_1 - \beta_1)'g'(\beta'_1X|\beta)E(X|b'_1X) + O_p(n^{-1}),$$

where the last equality follows from substituting the first term with relation (B.4). Finally, using the fact that $E(X|b'_1X) = E(X|\beta'_1X) + O_p(n^{-1/2})$ for $b$ in a $\sqrt{n}$-neighborhood of $\beta$, we get

$$g(b'_1X|b) - g(\beta'_1X|\beta) = (b_1 - \beta_1)'g'(\beta'_1X|\beta)(x - E(X|\beta'_1X)) + O_p(n^{-1})$$

$$= (b - \beta)'g'(\beta'_1X|\beta)(x - E(X|\beta'_1X)) + O_p(n^{-1})$$

and (B.3) follows.

\[\square\]

**LEMMA B.6.** Let $A_n(\tau, \gamma)$ defined in (B.2) for $\gamma \in \mathbb{R}^{d-1}$ such that $\gamma + \beta \in \Theta$. Then, under the assumptions of Proposition 3.1 and Assumptions A6 and A7 given in Appendix A, we have the following quadratic approximation, uniformly in $\gamma$ in a compact set,

$$A_n(\tau, \gamma) = \frac{1}{2} \gamma' V \gamma + W_n' \gamma + o_p(1), \quad (B.5)$$

where $V$ and $W_n$ are defined in (3.1) and (3.2) respectively.

**Proof:** For the proof define $H$ to be a class of bounded functions $\eta : \mathbb{R}^d \to \mathbb{R}$, whose value at $(t, \beta)' \in \mathbb{R}^d$ can be written as $\eta(t|\beta)$, in the non-separable space $L^\infty(t, \beta) = \{(t, \beta)' : \mathbb{R}^d \to \mathbb{R} : \|\eta\|_{(t, \beta)} := \sup_{(t, \beta)' \in \mathbb{R}^d} |\eta(t|\beta)| < \infty\}$, and having bounded and continuous partial derivatives, where the first and second derivatives with respect to $t$ exist and are bounded. Thus, $H$ includes $g(t|\beta)$, as well as $\hat{g}_Q^{NW}(t|\beta)$ for $n$ large enough. Define,

$$A_n(\eta, \tau, \sqrt{n}(b - \beta)) = \sum_{i=1}^n \left( \rho_+ (e_i(\beta, \eta) - \tilde{\eta}(X_i|b, \beta)) - \rho_+(e_i(\beta, \eta)) \right), \quad (B.6)$$

where $e_i(\beta, \eta) = Y_i - \eta(\beta'_1X_i|\beta)$ and $\tilde{\eta}(X_i|b, \beta) = \eta(b'_1X_i|b) - \eta(\beta'_1X_i|\beta)$. Showing that the quadratic approximation holds uniform in $\eta \in H$, it also holds for replacing $\eta$ with $\hat{g}_Q^{NW}$, where (B.6) reduces to the objective function defined in (B.2).
Write $A_n(\eta, \tau, \gamma)$ as
\[
\mathbb{E} \left( A_n(\eta, \tau, \gamma) \mid X \right) - \sum_{i=1}^{n} \left( \rho'_\tau(e_i(\beta, \eta)) - \mathbb{E}(\rho'_\tau(e_i(\beta, \eta)) \mid X) \right) \left( X_i | \gamma/\sqrt{n} + \beta \right) \\
+ R_n(\eta, \tau, \gamma),
\]
where $X$ denotes the design matrix, and $R_n(\eta, \tau, \gamma)$ is the remainder term defined by (B.7). Let $\varphi(t|X = x) = \mathbb{E}(\rho_\tau(\epsilon + t) \mid X = x)$ be as defined in Assumption A6, and observe that
\[
\mathbb{E} \left( A_n(\eta, \tau, \gamma) \mid X \right) = \sum_{i=1}^{n} \mathbb{E} \left( \rho_\tau \left( e_i(\beta, \eta) - \tilde{\eta} \left( X_i | \gamma/\sqrt{n} + \beta \right) \right) - \rho_\tau(e_i(\beta, \eta)) \right) \mid X
\]
\[
= \sum_{i=1}^{n} \mathbb{E} \left( \rho_\tau \left( e_i + g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) - \tilde{\eta} \left( X_i | \gamma/\sqrt{n} + \beta \right) \right) \right) \mid X
\]
\[
= \sum_{i=1}^{n} \left( \varphi(g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) - \tilde{\eta} \left( X_i | \gamma/\sqrt{n} + \beta \right) \mid X) \right)
\]
\[
+ \frac{1}{2} \sum_{i=1}^{n} (\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta))^2 \varphi'' \left( g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) \right) \mid X
\]
\[
- \frac{1}{6} \sum_{i=1}^{n} (\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta))^3 \varphi''' \left( g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) + u_i | X \right)
\]
\[
= - \sum_{i=1}^{n} \mathbb{E}(\rho'_\tau(e_i(\beta, \eta)) \mid X) \tilde{\eta} \left( X_i | \gamma/\sqrt{n} + \beta, \beta \right)
\]
\[
+ \frac{1}{2} \sum_{i=1}^{n} (\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta))^2 \varphi'' \left( g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) \right) + o_p(1),
\]
(B.8)

for $u_i$ in a neighborhood of $-|\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta)|$ and $|\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta)|$, where the last equality uses $\varphi'(g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) \mid X) = \mathbb{E}(\rho'_\tau(e_i(\beta, \eta)) \mid X)$. The last term of the last equality is $o_p(1)$ uniformly in $\eta \in H$, which follows by noting that
\[
\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta) = \eta((\gamma/\sqrt{n} + \beta)|X_i | \gamma/\sqrt{n} + \beta) - \eta(\beta'_i X_i | \beta)
\]
\[
= \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \bigg|_{b = \beta + t_n},
\]
(B.9)

for $\|t_n\|$ been between $-\|\gamma\| / \sqrt{n}$ and $\|\gamma\| / \sqrt{n}$, and writing
\[
\sup_{\eta \in H} \left| \sum_{i=1}^{n} (\tilde{\eta}(X_i | \gamma/\sqrt{n} + \beta, \beta))^3 \varphi''' \left( g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) + u_i | X \right) \right|
\]
\[
\leq \sum_{i=1}^{n} \sup_{\eta \in H} \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \bigg|_{b = \beta + t_n} \right|^3 \varphi''' \left( g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) + u_i | X \right) \right|
\]
\[
= o_p(1),
\]

27
where the last equality follows by the bounded partial derivatives of \( \eta \) and the boundedness of \( \varphi'''(\cdot | X = x) \).

Following, we will show that \( \sup_{\eta \in \mathcal{H}} | R_n(\eta, \tau, \gamma) | = o_p(1) \), where \( R_n(\eta, \tau, \gamma) \) is defined in (B.7). Obviously \( R_n(\eta, \tau, \gamma) \) is centered and can be written as

\[
R_n(\eta, \tau, \gamma) = \sum_{i=1}^{n} (R_{ni}(\eta, \tau, \gamma) - \mathbb{E}(R_{ni}(\eta, \tau, \gamma) | X_i)) ,
\]

where

\[
R_{ni}(\eta, \tau, \gamma) = \rho_{\tau}(e_i(\beta, \eta) - \tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)) - \rho_{\tau}(e_i(\beta, \eta)) + \rho'_{\tau}(e_i(\beta, \eta)) \tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta).
\]

The uniformity will follow from the Uniform Law of Large Numbers for Triangular Arrays, restated in Lemma B.4, by considering \( nR_{ni}(\eta, \tau, \gamma) \) which is continuous in \( \eta \) with probability one. Note that, by using the definition of \( \rho_{\tau}(\cdot) \) and \( \rho'_{\tau}(\cdot) \), \( R_{ni}(\eta, \tau, \gamma) \) can be equivalently written as

\[
R_{ni}(\eta, \tau, \gamma) = (e_i(\beta, \eta) - \tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)) \times (I(e_i(\beta, \eta) < 0) - I(e_i(\beta, \eta) < \tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)))
\]

which is bounded by

\[
|R_{ni}(\eta, \tau, \gamma)| \leq |e_i(\beta, \eta) - \tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)| I(|e_i(\beta, \eta)| \leq |\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)|)
\]

\[
\leq 2|\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)| I(|e_i(\beta, \eta)| \leq |\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)|),
\]

where the first inequality follows from the fact that

\[
|I(e_i(\beta, \eta) < 0) - I(e_i(\beta, \eta) < \tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta))| \\
\leq I(-|\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)| \leq e_i(\beta, \eta) < |\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)|).
\]

Following, note that \( nR_{ni}(\eta, \tau, \gamma) \) is dominated by

\[
|nR_{ni}(\eta, \tau, \gamma)| \leq 2n |\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)| I(|e_i(\beta, \eta)| \leq |\tilde{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta)|)
\]

\[
(B.9) = 2n \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b_i'X_i | b) \right|_{b=\beta+t_n} \\
\times I \left( |e_i(\beta, \eta)| \leq \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b_i'X_i | b) \right|_{b=\beta+t_n} \right)
\]

28
\[ \leq 2\sqrt{n} \sup_{\eta \in H} \left| \gamma' \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \]
\[ \times I \left( |\epsilon_i(\beta, \eta)| \leq \sup_{\eta \in H} \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \right) \]
\[ \leq 2\sqrt{n} \sup_{\eta \in H} \left| \gamma' \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \]
\[ \times I \left( |\epsilon_i - \eta^*_i(\beta) + g(\beta'_i X_i | \beta)| \leq \sup_{\eta \in H} \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \right) \]
\[ = \tilde{R}_{ni}(\tau, \gamma), \]

where \( \eta^*_i(\beta) = \arg\inf_{\eta \in H} |\epsilon_i - \eta(\beta'_i X_i | \beta) + g(\beta'_i X_i | \beta)| \). Then, using the fact that

\[ E \left( I \left( |\epsilon_i - \eta^*_i(\beta) + g(\beta'_i X_i | \beta)| \leq \sup_{\eta \in H} \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \right) \right) \]
\[ = F_{\epsilon_i, X_i} \left( \eta^*_i(\beta) - g(\beta'_i X_i | \beta) \right) + \sup_{\eta \in H} \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \]
\[ - F_{\epsilon_i, X_i} \left( \eta^*_i(\beta) - g(\beta'_i X_i | \beta) \right) - \sup_{\eta \in H} \left| \frac{\gamma'}{\sqrt{n}} \nabla_b \eta(b'_i X_i | b) \right|_{b=\beta+t_n} \]
\[ = O(n^{-1/2}), \]

almost surely, it is easy to see that \( \sup_n E(\tilde{R}_{ni}(\tau, \gamma) | X) = O(1) \), almost surely, and therefore,

\[ \sup_{\eta \in H} \left| \frac{1}{n} \sum_{i=1}^{n} (nR_{ni}(\eta, \tau, \gamma) - E(nR_{ni}(\eta, \tau, \gamma) | X)) \right| \]
\[ = \sup_{\eta \in H} \left| \sum_{i=1}^{n} (R_{ni}(\eta, \tau, \gamma) - E(R_{ni}(\eta, \tau, \gamma) | X)) \right| = o_p(1). \]

Next, substituting the expression of \( E(A_n(\eta, \tau, \gamma) | X) \) derived in (B.8), to relation (B.7) and using the fact that \( \sup_{\eta \in H} |R_n(\eta, \tau, \gamma)| = o_p(1) \), we get, uniformly in \( \eta \in H \),

\[ A_n(\eta, \tau, \gamma) = \frac{1}{2} \sum_{i=1}^{n} (\overline{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta))^2 \varphi''(g(\beta'_i X_i | \beta) - \eta(\beta'_i X_i | \beta) | X) \]
\[ - \sum_{i=1}^{n} \rho'_i(\epsilon_i(\beta, \eta)) \overline{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta) + o_p(1). \quad (B.10) \]

Since expression (B.10) holds uniformly in \( \eta \in H \), where the class \( H \) contains \( \hat{\eta}_{Q}^{NW} \), we substitute \( \eta \) with \( \hat{\eta}_{Q}^{NW} \). Using the fact that \( A_n(\hat{\eta}_{Q}^{NW}, \tau, \gamma) \) reduces to \( A_n(\tau, \gamma) \) defined in (B.2), and relation

\[ \overline{\eta}(X_i | \gamma / \sqrt{n} + \beta, \beta) = (0, \gamma' / \sqrt{n}) \nabla \hat{\eta}_{Q}^{NW}(b'_i X_i | b) \bigg|_{b=\beta} + O_p(n^{-1}) \]
\[ = \frac{\gamma'}{\sqrt{n}} g'(\beta'_i X_i | \beta)(X_{i-1} - E(X_{i-1} | \beta'_i X_i)) + O_p(n^{-1}), \]

which follows form Lemma B.5, we get

\[ A_n(\tau, \gamma) = \frac{1}{2} \gamma' V_n \gamma + W_n' \gamma + r_n(\tau, \gamma), \quad (B.11) \]
where \( r_n(\tau, \gamma) = o_p(1) \),

\[
\mathbb{V}_n = \frac{1}{n} \sum_{i=1}^{n} (g'(\beta'_i X_i | \beta))^2 (X_{i,-1} - \mathbb{E} (X_{-1} | \beta'_i X)) (X_{i,-1} - \mathbb{E} (X_{-1} | \beta'_i X))' \times \varphi'' \left( g(\beta'_i X_i | \beta) - \hat{g}_Q^{NW}(\beta'_i X_i | \beta) \right),
\]

and \( W_n \) is defined in (3.2). Next, showing that \( (1/2)\mathbb{V}^\prime \mathbb{V} \gamma = (1/2)\mathbb{V}^\prime \mathbb{V} \gamma + o_p(1) \), where \( \mathbb{V} \) is defined in (3.1), relation (B.11) can be written as,

\[
A_n(\tau, \gamma) = \frac{1}{2} \mathbb{V}^\prime \mathbb{V} \gamma + W'_n \gamma + o_p(1).
\]

This is easy to prove by noting that \( \varphi'(0 | X = x) = \tau - F_{\epsilon | X}(0 | x) = 0 \), \( \varphi''(0 | X = x) = f_{\epsilon | X}(0 | x) \) (see the note below for a proof), and

\[
\varphi''(g(\beta'_i X | \beta) - \hat{g}_Q^{NW}(\beta'_i X | \beta) | X = x) = \varphi''(0 | X = x) + o_p(a_n^* + a_n + h_n^2),
\]

where the last equality follows from Proposition 3.1. Therefore,

\[
\mathbb{V}_n = \frac{1}{n} \sum_{i=1}^{n} (g'(\beta'_i X_i | \beta))^2 (X_{i,-1} - \mathbb{E} (X_{-1} | \beta'_i X)) (X_{i,-1} - \mathbb{E} (X_{-1} | \beta'_i X))' \varphi''(0 | X) + o_p(1)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} (g'(\beta'_i X_i | \beta))^2 (X_{i,-1} - \mathbb{E} (X_{-1} | \beta'_i X)) (X_{i,-1} - \mathbb{E} (X_{-1} | \beta'_i X))' f_{\epsilon | X}(0 | X) + o_p(1)
\]

\[
= \mathbb{V} + o_p(1).
\]

Finally, noting that \( W_n \) has bounded second moment (see Lemma B.7), and hence is stochastically bounded, the convex function \( A_n(\tau, \gamma) - W'_n \gamma \) converges in probability to the convex function \( (1/2)\mathbb{V}^\prime \mathbb{V} \gamma \). Therefore, it follows from the convexity lemma (Pollard 1991), restated in Lemma B.2, that for any compact set \( K \), \( \sup_{\gamma \in K} | r_n(\tau, \gamma) | = o_p(1) \). Thus, the quadratic approximation to the convex function \( A_n(\tau, \gamma) \) holds uniformly for \( \gamma \) in a compact set.

**Note:** The fact that \( \varphi'(0 | X = x) = \tau - F_{\epsilon | X}(0 | x) = 0 \) and \( \varphi''(0 | X = x) = f_{\epsilon | X}(0 | x) \) can be easily proved using simple calculations and Assumption A6,

\[
\varphi'(0 | X = x) = \left. \frac{\partial}{\partial s} \left( \int \rho_{\tau}(z + s) f_{\epsilon | X}(z | x) dz \right) \right|_{s=0} = \left. \frac{\partial}{\partial s} \rho_{\tau}(z + s) \right|_{s=0} f_{\epsilon | X}(z | x) dz
\]

\[
= \int \rho_{\tau}(z) f_{\epsilon | X}(z | x) dz = \int (\tau - I(z < 0)) f_{\epsilon | X}(z | x) dz = \tau - F_{\epsilon | X}(0 | x)
\]

and

\[
\varphi''(0 | X = x) = \left. \frac{\partial}{\partial s} \varphi'(s | x) \right|_{s=0} = \left. \frac{\partial}{\partial s} (\tau - F_{\epsilon | X}(-s | x)) \right|_{s=0} = f_{\epsilon | X}(0 | x).
\]
LEMMA B.7. Let $W_n = n^{-1} \sum_{i=1}^n \rho'_t(Y_i)g'(\beta'_i X_i|\beta)(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i))$. Then, under the assumptions of Proposition 3.1 and Assumptions A6 and A7 given in Appendix A,

$$P \left( \sqrt{n}((\tau(1-\tau))^{-1/2}W_n^*) \leq t|X \right) = \Phi(t) + o_p(1),$$

where $\Phi(t)$ denotes the standard normal cumulative distribution function.

Proof: For the proof, we will use the same technique as in the proof of Lemma B.6, that is, define $Z_i(\eta) = \rho'_t(c_i(\beta, \eta))g'(\beta'_i X_i|\beta)(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i))$, where recall that $c_i(\beta, \eta) = Y_i - \eta(\beta'_i X_i|\beta)$ for $\eta \in H$, and let $T_i(\eta) = Z_i(\eta) - E(Z_i(\eta)|X)$. Using the Berry-Esseen theorem (Berry 1941, and Esseen 1942), we will show that $n^{-1/2} \sum_{i=1}^n T_i(\eta)$ converges to a multivariate normal distribution, uniformly in $\eta \in H$. To do this, recall the Cramer-Wald theorem and, for any $t \in \mathbb{R}^{d-1}$, consider $t' T_i(\eta) = t'(Z_i(\eta) - E(Z_i(\eta)|X))$, where

$$E(Z_i(\eta)|X) = E(\rho'_t(c_i(\beta, \eta))|X) g'(\beta'_i X_i|\beta)(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i))$$

$$= (\tau - F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i)) g'(\beta'_i X_i|\beta)(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i)).$$

Following, conditionally on the design matrix $X$, $t' T_i(\eta), ..., t' T_n(\eta)$ are independent random variables, with $E(t' T_i(\eta)|X) = 0$ by definition, and

$$Var(t' T_i(\eta)|X) = t'(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i)) Var(\rho'_t(c_i(\beta, \eta))|X)(g'(\beta'_i X_i|\beta))^2$$

$$= t'(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i))' t$$

$$= t'(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i)) F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i) g'(\beta'_i X_i|\beta)^2$$

$$\times (1 - F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i)) (X_{i,-1} - E(X_{i,-1}|\beta'_i X_i))^t t$$

$$= \sigma^2_t(\eta).$$

Moreover,

$$E(\lvert t' T_i(\eta) \rvert^3 |X) = E \left( \lvert t'(Z_i(\eta) - E(Z_i(\eta)|X)) \rvert^3 |X \right)$$

$$= E \left( \lvert (\rho'_t(c_i(\beta, \eta)) - \tau + F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i)) g'(\beta'_i X_i|\beta) \right.$$\n
$$\times t'(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i)) \rvert^3 |X \right)$$

$$= E \left( \lvert F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i) - I(c_i(\beta, \eta) < 0) \rvert^3 |X \right)$$

$$\times \lvert g'(\beta'_i X_i|\beta) t'(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i)) \rvert^3$$

$$= \left( F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i) + F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i) \right)$$

$$(1 - F_{c_i|X}(\eta(\beta'_i X_i|\beta) - g(\beta'_i X_i|\beta)X_i)) \lvert g'(\beta'_i X_i|\beta) t'(X_{i,-1} - E(X_{i,-1}|\beta'_i X_i)) \rvert^3$$

$$= \rho_t(\eta) < \infty,$$
where the equality before the last follows from the fact that

\[
\mathbb{E} \left( \left( F_i | x_i \right) (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i) - I(e_i(\beta, \eta) < 0) \right)^3 | X_i \right)
\]

\[
= \int_{\eta(\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta)}^{\infty} F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i) f_{e_i | x_i} (e | X_i) \, de
\]

\[
= \int_{-\infty}^{\eta(\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta)} (F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i) - 1)^3 f_{e_i | x_i} (e | X_i) \, de
\]

\[
= F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i) (1 - F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i))
\]

\[
- (F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i) - 1)^3 F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta) | X_i).\]

Then, by Esseen (1942), conditionally on the design matrix \( X \),

\[
\left| \frac{1}{\sqrt{\sum_{i=1}^{n} \sigma_i^2(\eta)}} \sum_{i=1}^{n} t_i' T_i(\eta) \right| \leq t \mathbb{P}(t) - \Phi(t) \leq C_0 \left( \sum_{i=1}^{n} \sigma_i^2(\eta) \right)^{-3/2} \sum_{i=1}^{n} \rho_i(\eta),
\]

where it is easy to see that

\[
\sup_{\eta \in H} \left| \frac{1}{n} \sum_{i=1}^{n} \rho_i(\eta) \right| \leq \frac{1}{n} \sum_{i=1}^{n} \sup_{\eta \in H} |\rho_i(\eta)| = o(1), \tag{B.12}
\]
a.s. and

\[
\sup_{\eta \in H} \left| \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2(\eta) - \bar{\sigma}^2(\eta) \right| = o(1) \tag{B.13}
\]
a.s., where

\[
\bar{\sigma}^2(\eta) = t' \mathbb{E} \left( F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta)) (1 - F_{e_i | x_i} (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta)) \right)
\]

\[
= (\eta (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta))^2 (X_{-1} - \mathbb{E}(X_{-1} | \beta'_i X)) (X_{-1} - \mathbb{E}(X_{-1} | \beta'_i X))' t.
\]

The uniformity of (B.13) follows from the Uniform Strong Law of Large Numbers (Jennrich 1969) since \( \sigma_i^2(\eta) \) is dominated by \( t'(\eta (\beta'_i X_i | \beta))^2 (X_{-1} - \mathbb{E}(X_{-1} | \beta'_i X)) (X_{-1} - \mathbb{E}(X_{-1} | \beta'_i X))' t \), which has a bounded expectation given by \( t' \Sigma t \), where \( \Sigma \) is defined in (3.3). Therefore, conditionally on \( X \),

\[
\left| \frac{1}{\sqrt{\sum_{i=1}^{n} \sigma_i^2(\eta)}} \sum_{i=1}^{n} t_i' T_i(\eta) \right| \leq t \mathbb{P}(t) - \Phi(t) = o_p(1) \tag{B.14}
\]

uniformly in \( \eta \in H \). Since (B.14) holds uniformly in \( \eta \in H \), it also holds for \( \eta = \hat{g}_q^{NW} \), where

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E} \left( t_i' \hat{Z}_i (\hat{g}_q^{NW}) | X \right) = t' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \tau - F_{e_i | x_i} (\hat{g}_q^{NW} (\beta'_i X_i | \beta) - g(\beta'_i X_i | \beta)) \right)
\]

\[
\times g' (\beta'_i X_i | \beta) (X_{i,-1} - \mathbb{E}(X_{i,-1} | \beta'_i X)) = o_p(1), \tag{B.15}
\]
and the last equality follows from the fact that
\[ F_{c|X_i}(\hat{g}_{Q}^{NW}(\beta'_i|X_i|\beta) - g(\beta'_i|X_i|\beta)|X_i) = F_{c|X_i}(0|X_i) + o_p(a_n + a_n + h_2^2) = \tau + o_p(a_n + a_n + h_2^2) \] which holds uniformly (see Proposition 3.1) and that
\[ n^{-1} \sum_{i=1}^{n} g'((\beta'_i|X_i|\beta)(X_i-1 - E(X_{i-1}|\beta'_i|X_i))) \overset{P}{\to} E(g'((\beta'_i|X_i|\beta)(X_{i-1} - E(X_{i-1}|\beta'_i|X_i)))) = 0. \]

Furthermore,
\[
\frac{1}{n} \sum_{i=1}^{n} \sigma_{i}^{2}(\hat{g}_{Q}^{NW}) = t \frac{1}{n} \sum_{i=1}^{n} (X_{i-1} - E(X_{i-1}|\beta'_i|X_i))F_{c|X_i}(\hat{g}_{Q}^{NW}(\beta'_i|X_i|\beta) - g(\beta'_i|X_i|\beta)|X_i)
\]
\[
= t \sum_{i=1}^{n} (X_{i-1} - E(X_{i-1}|\beta'_i|X_i))(X_{i-1} - E(X_{i-1}|\beta'_i|X_i))'t
\]
\[
= t\tau(1 - \tau)\Sigma t + o_p(1). \quad (B.16)
\]

Therefore, using (B.14), (B.15), (B.16) and Slutsky’s theorem, we get that, conditionally on \(X_i\), \(\sqrt{n}W_n \overset{d}{\to} N_{d-1}(0, \tau(1 - \tau)\Sigma)\), where the unconditional case follows from the Dominated Convergence theorem and the almost sure convergence of (B.12) and (B.13).

\[ ■ \]

B.2 Proof of Proposition 3.1

We need to prove that \(\hat{g}_{Q}^{NW}(t|\beta)\) is uniformly consistent estimator of \(g(t|\beta)\), uniformly in \(\beta \in \Theta\) and in \(t \in \mathcal{T}_\beta\). Let \(K_{2,h_2}(\cdot) = K_2(\cdot/h_2)\), and write \(\hat{g}_{Q}^{NW}(t|\beta) = \hat{\Psi}(t|\beta)\hat{f}_b(t)\), where \(\hat{\Psi}(t|\beta) = (nh_2)^{-1} \sum_{i=1}^{n} \hat{Q}_\tau(Y|X_i)K_{2,h_2}(t - b'_iX_i)\) and \(\hat{f}_b(t) = (nh_2)^{-1} \sum_{i=1}^{n} K_{2,h_2}(t - b'_iX_i)\). To prove the uniform consistency, we consider the numerator and denominator separately.

For the denominator, we use Theorem 6 of Hansen (2008) [take his \(\beta = \infty\) and the mixing coefficients as \(\alpha_m = 0\)] to obtain
\[
\sup_{\beta \in \Theta, t \in \mathcal{T}_\beta} \left| \hat{f}_b(t) - f_b(t) \right| = O_p \left( \frac{\log n}{nh_2} \right)^{1/2} + h_2^2 = O_p(a_n + h_2^2). \quad (B.17)
\]

Next, we will show that \(\hat{\Psi}(t|\beta)\) is consistent estimator of \(\Psi(t|\beta) = g(t|\beta)f_b(t)\), uniformly in \(\beta \in \Theta\) and \(t \in \mathcal{T}_\beta\), and determine its rate of convergence. Let
\[
\Psi^*(t|\beta) = \frac{1}{nh_2} \sum_{i=1}^{n} Q_\tau(Y|X_i)K_2 \left( \frac{t - b'_iX_i}{h_2} \right)
\]
and note that
\[
\left| \hat{\Psi}(t|\beta) - \Psi^*(t|\beta) \right| = \left| \frac{1}{nh_2} \sum_{i=1}^{n} \left( \hat{Q}_\tau(Y|X_i) - Q_\tau(Y|X_i) \right)K_2 \left( \frac{t - b'_iX_i}{h_2} \right) \right|
\]
\[\leq \sup_{1 \leq i \leq n} \left| \hat{Q}_\tau(Y|X_i) - Q_\tau(Y|X_i) \right| \frac{1}{nh_2} \sum_{i=1}^{n} K_2 \left( \frac{t - b'_iX_i}{h_2} \right)
\]
\[= O_p \left( \left( \frac{\log n}{n} \right)^{s/(2s+d)} \right) \hat{f}_b(t),
\]

33
where the last equality follows from Lemma B.1. Therefore, by (B.17) and Assumption A2,

\[
\sup_{b \in \Theta, t \in \mathcal{T}_b} \left| \hat{\Psi}(t|b) - \Psi^*(t|b) \right| \leq O_p(a_n^*) \sup_{b \in \Theta, t \in \mathcal{T}_b} \hat{f}_b(t) \\
= O_p(a_n^*) \left( \sup_{b \in \Theta, t \in \mathcal{T}_b} f_b(t) + O_p(a_n + h_2^2) \right) \\
= O_p(a_n^*). \tag{B.18}
\]

Next, Theorem 2 of Hansen (2008) yields \( \sup_{b \in \Theta, t \in \mathcal{T}_b} |\Psi^*(t|b) - \mathbb{E}(\Psi^*(t|b))| = O_p(a_n) \), where, recalling the notation \( \hat{\Psi}(t|b) = g(t|b)f_b(t) \), and using Assumption A4,

\[
\mathbb{E}(\Psi^*(t|b)) = \frac{1}{h_2} \mathbb{E} \left( \mathbb{E}(Q(X|b')X) K_2 \left( \frac{t - b'X}{h_2} \right) \right) \\
= \frac{1}{h_2} \int g(u|b)K_2 \left( \frac{t - u}{h_2} \right) f_b(u) du \\
= \int \Psi(t - vh_2|b)K_2(v) dv \\
= \Psi(t|b) + O(h_2^2).
\]

Thus, \( \sup_{b \in \Theta, t \in \mathcal{T}_b} |\Psi^*(t|b) - \hat{\Psi}(t|b)| = O_p(a_n + h_2^2) \) which, together with (B.18) yields

\[
\sup_{b \in \Theta, t \in \mathcal{T}_b} \left| \hat{\Psi}(t|b) - \Psi(t|b) \right| = O_p(a_n^* + a_n + h_2^2). \tag{B.19}
\]

Therefore, using (B.17), (B.19) and Assumption A2, we get

\[
\left| \hat{\Psi}(t|b) - g(t|b) \right| = \frac{\hat{\Psi}(t|b) - \hat{\Psi}(t|b)}{f_b(t)} \left| \frac{\hat{f}_b(t) - f_b(t)}{\hat{f}_b(t)} \right| + \left| \frac{\hat{\Psi}(t|b) - \Psi(t|b)}{f_b(t)} \right| \\
= O_p(a_n + h_2^2) + O_p(a_n^* + a_n + h_2^2) \\
= O_p(a_n^* + a_n + h_2^2)
\]

uniformly in \( b \in \Theta \) and \( t \in \mathcal{T}_b \).

\[ \blacksquare \]

### B.3 Proof of Proposition 3.2

To prove the \( \sqrt{n} \)-consistency of \( \hat{\beta} \), enough to show that for any given \( \delta > 0 \), there exists a large constant \( C \) such that

\[
P \left( \inf_{nA_n(\tau, \phi) = 0} \frac{1}{n} A_n(\tau, 0) \right) \geq 1 - \delta. \tag{B.20}
\]
where $A_n(\tau, \gamma)$ defined in (B.2), and implies that with probability at least $1 - \delta$ there exists a local minimum in the ball $\{\beta + \gamma/\sqrt{n} : \|\gamma\| \leq C\}$. This in turn implies that there exists a local minimizer such that $\|\hat{\beta} - \beta\| = O_p(n^{-1/2})$.

The quadratic approximation derived in Lemma B.6, yields that

$$A_n(\tau, \gamma) - A_n(\tau, 0) = \frac{1}{2} \gamma' V \gamma + W_n' \gamma + o_p(1),$$

(B.21)

where $V$ and $W_n$ are defined in (3.1) and (3.2) respectively, for any $\gamma$ in a compact subset of $\mathbb{R}^{d-1}$. Therefore, the difference (B.21) is dominated by the quadratic term $(1/2)\gamma' V \gamma$ for $\gamma$ equal to sufficiently large $C$. Hence, (B.20) follows.

B.4 Proof of Theorem 3.3

The proof follows form the Quadratic Approximation Lemma (Hjort and Pollard 1993), restated in Lemma B.3.

Consider the scaled difference $\gamma = \sqrt{n}(b - \beta)$ and $\hat{\gamma}$ the solution of the objective function $A_n(\tau, \gamma)$ defined in (B.2). From the $\sqrt{n}$ consistency of $\hat{\beta}$, the quadratic approximation (B.5), derived in Lemma B.6, holds uniformly in $\gamma$ in a compact set. Using the convexity assumption, the minimizer $\hat{\gamma}$ of $A_n(\tau, \gamma)$ converges in probability to the minimizer $\hat{\gamma}^* = -V^{-1}W_n$. Thus, $\hat{\gamma} - \hat{\gamma}^* = o_p(1)$ and therefore,

$$\sqrt{n}(\hat{\beta} - \beta - S_n) = o_p(1),$$

(B.22)

where

$$S_n = -\frac{1}{\sqrt{n}} V^{-1} W_n = V^{-1} W_n^*.$$

The asymptotic normality of $W_n^*$ was derived in Lemma B.7 and it follows that

$$\sqrt{n}S_n = \sqrt{n}V^{-1}W_n^* \xrightarrow{d} N(0, \tau(1 - \tau)V^{-1}\Sigma V^{-1}),$$

(B.23)

where $\Sigma$ is defined in (3.3). Therefore, using (B.22) and (B.23) we get,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \tau(1 - \tau)V^{-1}\Sigma V^{-1}).$$

\[\blacksquare\]
References


