Smoothing quantile regressions

Marcelo Fernandes
Emmanuel Guerre
Eduardo Horta
Os artigos dos Textos para Discussão da Escola de Economia de São Paulo da Fundação Getulio Vargas são de inteira responsabilidade dos autores e não refletem necessariamente a opinião da FGV-EESP. É permitida a reprodução total ou parcial dos artigos, desde que creditada a fonte.

Escola de Economia de São Paulo da Fundação Getulio Vargas FGV-EESP
www.eesp.fgv.br
Smoothing quantile regressions

Marcelo Fernandes
Sao Paulo School of Economics, FGV
and Queen Mary University of London
marcelo.fernandes@fgv.br

Emmanuel Guerre
School of Economics and Finance
Queen Mary University of London
e.guerre@qmul.ac.uk

Eduardo Horta
Department of Statistics
Universidade Federal do Rio Grande do Sul
eduardo.horta@ufrgs.br

This version: June 1, 2017

Abstract: We propose to smooth the entire objective function rather than only the check function in a linear quantile regression context. We derive a uniform Bahadur-Kiefer representation for the resulting convolution-type kernel estimator that demonstrates it improves on the extant quantile regression estimators in the literature. In addition, we also show that it is straightforward to compute asymptotic standard errors for the quantile regression coefficient estimates as well as to implement Wald-type tests. Simulations confirm that our smoothed quantile regression estimator performs very well in finite samples.

JEL classification numbers: G11, G12, C12, C14

Keywords: asymptotic expansion, Bahadur-Kiefer representation, conditional quantile, convolution-based smoothing, data-driven bandwidth.

Acknowledgments: We are grateful to Aureo de Paulo for valuable comments. We thanks the financial support from FAPESP (2013/22930-0) and CNPq (302272/2014-3).
1 Introduction

Quantile regression (QR) enjoys some very appealing features. Apart from enabling some very flexible patterns of partial effects, quantile regressions are also interesting because they satisfy some equivariance and robustness principles. See Koenker and Bassett (1978) and Koenker (2005) for theoretical aspects; and Koenker (2000), Buchinsky (1998), Koenker and Hallock (2001), Koenker (2005) and references therein for applications. However, there is a price to pay. The objective function that the standard quantile regression estimator aims to minimize is not smooth and, as a result, statistical inference is not straightforward for two reasons. First, it is hard to compute asymptotic confidence intervals because it involves ancillary estimation of nuisance parameters (namely, the asymptotic covariance matrix depends on the population conditional density evaluated at the true quantile). Second, the asymptotic normality of the standard QR estimator relies on Bahadur-Kiefer representations with very poor rates of convergence. See Koenker and Portnoy (1987), Chaudhuri (1991), He and Shao (1996), Knight (2001), Guerre and Sabbah (2012), Portnoy (2012), Kong, Linton and Xia (2013), and Mammen et al. (2013). As a matter of fact, it is very hard to establish the precise order of magnitude of the remainder term in such expansions. For instance, it is possible to show that it is at best of order \( n^{-1/4} \) for iid errors (see Koenker and Portnoy, 1987; Knight, 2001; Jurečková, Sen and Picek, 2012). This means that the first-order Gaussian approximation for the distribution of the QR estimators is likely to fail in finite samples.

In this paper, we tackle these two shortcomings of the standard QR estimator by means of a convolution-type kernel smoothing of the QR objection function. We are obviously not the first to sail in this direction. In particular, Nadaraya (1964) and Parzen (1979) estimate unconditional quantiles respectively by inverting a smoothed estimator of the cumulative distribution function (cdf) and by smoothing the sample quantile function. Azzalini (1981) shows that the former dominates sample quantiles at the second order (see also Cheung and Lee, 2010), with Sheather and Marron (1990) establishing a similar result for the latter. Falk (1984) discusses the relative

---

1 See, among others, the discussion in Koenker (1994), Buchinsky (1995), Koenker (2005), Goh and Knight (2009), and Fan and Liu (2016). Note that the literature boasts a wide array of techniques to tackle this issue, including bootstrapping (Horowitz, 1998; Machado and Parente, 2005), MCMC methods (Chernozhukov and Hong, 2003), empirical likelihood (Otsu, 2008; Whang, 2006), strong approximation methods (Portnoy, 2012), nonstandard inference (Goh and Knight, 2009), and other nonparametric approaches (Mammen, Van Keilegom and Yu, 2013; Fan and Liu, 2016).
deficiency of sample quantiles with respect to kernel-based quantile estimators, whereas Xiang (1994), Mack (1987) and Ralescu (1997) provide Bahadur-Kiefer representations for smoothed quantile estimators. Bearing this strand of the literature in mind, our contribution is essentially to extend Nadaraya’s (1964) approach to handle conditional quantiles in a multivariate setting (rather than unconditional quantiles). We show that, by kernel smoothing the conditional cdf, we end up with a QR estimator that not only improves on the Bahadur-Kiefer approximation of the standard QR estimators, but also allows for straightforward computation of asymptotic confidence intervals.

Kernel smoothing is naturally very popular in the literature dealing with nonparametric estimation of conditional quantile functions. See, for instance, Stute (1986), Samanta (1989), Bhattacharya and Gangopadhyay (1990), Mehra, Rao and Upadrasta (1991), Yu (1999), and Mammen et al. (2013). Despite several papers that propose smoothing methods for (semi-)parametric quantile regression models, there are only a few that attempt to generalize the smoothed quantile estimators of Nadaraya (1964) and Parzen (1979), though. For instance, Horowitz (1998) smooth the check function of the quantile regression to obtain a smoothed QR estimator. It turns out, however, that this leads neither to Nadaraya’s (1964) nor to Parzen’s (1979) smoothed estimators in the unconditional case. Accordingly, it is not necessarily the case that the second-order improvements obtained by Azzalini (1981) hold. Indeed, it is possible to show that Horowitz’s (1998) smoothed QR estimator does not improve on the standard estimator of the QR coefficients. More recently, Kaplan and Sun (2017) independently develop an instrumental variables quantile regression testing framework based on smoothed estimating equations (SEE). The latter are analogous to the first-order conditions that our smoothed QR estimator solves. As us, they argue that this approach can improve on Horowitz (1998) smoothed QR estimator. Using SEE for estimation purpose would bring about some technical difficulties, though: e.g., a compact parametric space and uniqueness of the parameter vector satisfying the population SEE. Unlike us, Kaplan and Sun (2017) do not establish bandwidth uniform results, and hence they are not able to handle data-driven bandwidths in their asymptotic theory.

Our main contributions are as follows. First, we show that smoothing the entire objective function in a smoothed QR estimator that is more linear than the standard QR estimator, in
the sense that the stochastic order of the Bahadur-Kiefer remainder term is much closer to \( n^{-1/2} \).

Second, we establish that the smoothing bias is asymptotically negligible, providing a precise rate of convergence. This means that our smoothed estimator is \( \sqrt{n} \)-consistent and asymptotically normal. Third, we find that the asymptotic mean squared error (AMSE) of our convolution-type kernel QR estimator is lower than the AMSE of the standard QR estimator for a proper bandwidth choice. This essentially generalizes Azzalini’s (1981) result for unconditional quantiles to conditional quantiles. Fourth, we provide a consistent estimator of the asymptotic covariance matrix that does not depend on ancillary estimation of nuisance parameters, allowing us to compute asymptotic confidence intervals and Wald-type tests in a straightforward manner. Finally, our asymptotic theory is uniform both in the quantile level and in the bandwidth parameter. The former is important because it allows one to recover the conditional cdf of the response (except perhaps for the tails) at any given level of the covariate. In turn, the latter is crucial to justify data-driven bandwidth choices that may depend even on quantile level and covariates.

The remainder of this paper is as follows. Section 2 introduces our smoothed QR estimator. Section 3 describes the main assumptions and results, including the Bahadur-Kiefer representation and AMSE of our estimator. It also discusses how to estimate the asymptotic covariance matrix in a consistent fashion in order to implement Wald-type inference. Section 4 assesses by means of a simulation study the performance of our kernel-based QR estimator relative to Koenker and Bassett’s (1978) and Horowitz’s (1998) estimators. Section 5 offers some concluding remarks. Appendix A collects all technical proofs.

## 2 The smoothed quantile regression estimator

Let \((Y_i, X_i)\), with \(i = 1, \ldots, n\), denote an iid sample from \((Y, X) \in \mathbb{R} \times \mathbb{R}^d\), where the conditional quantile of the response \(Y\) given the covariate \(X = x\) is such that

\[
Q(\tau \mid x) = x'\beta(\tau), \quad \tau \in (0, 1),
\]

with \(Q(\tau \mid x) := \inf \{ q : F(q \mid x) \geq \tau \} \) and \(F(\cdot \mid x)\) denoting the conditional cdf of \(Y\) given \(X = x\). Koenker and Bassett (1978) define the objective function of the quantile regression as

\[
R(b; \tau) := \mathbb{E} \left[ \rho_\tau (e(b)) \right] = \int \rho_\tau (t) dF(t; b),
\]
where \( e(b) := Y - X'b \), \( F(t; b) = \Pr[e(b) \leq t] \), and \( \rho_r(u) := u[\tau - I(u < 0)] \) is the usual check function with \( I(A) \) denoting the indicator function that takes value 1 if \( A \) is true, zero otherwise. As the true parameter \( \beta(\tau) \) minimizes (2), Koenker and Bassett’s (1978) standard QR estimator \( \hat{\beta}(\tau) \) minimizes the sample analog based on the empirical distribution, namely,

\[
\hat{R}(b; \tau) := \frac{1}{n} \sum_{i=1}^{n} \rho_r(e_i(b)) = \int \rho_r(t) \, d\hat{F}(t; b),
\]

(3)

where \( e_i(b) := Y_i - X_i'b \), and \( \hat{F}(\cdot; b) \) denotes the empirical distribution function of \( e_i(b) \).

The right-hand side of equation (2) suggests that one may obtain a different QR estimator by varying the integrating measure of the check function integral, that is to say, by changing the estimator of the cdf. Instead of employing the empirical distribution as in the standard QR estimator, we shall consider a kernel-type cdf estimator as the integrating measure in a similar fashion to what Nadaraya (1964) proposes for smoothing the unconditional quantile estimator.

Consider a bandwidth \( h > 0 \) that shrinks to zero as the sample size grows and a smooth kernel function \( k \) such that \( \int k(v) \, dv = 1 \). Letting \( k_h(v) = \frac{1}{h} k(v/h) \), the kernel density and distribution estimators are given by \( \hat{f}_h(v; b) := \frac{1}{n} \sum_{i=1}^{n} k_h(v-e_i(b)) \) and \( \hat{F}_h(t; b) := \int_{-\infty}^{t} \hat{f}_h(v; b) \, dv \), respectively.

We then apply kernel smoothing to the empirical objective function in (3), yielding

\[
\hat{R}_h(b; \tau) := \int \rho_r(t) \, d\hat{F}_h(t; b) = \int \rho_r(t) \, \hat{f}_h(t; b) \, dt.
\]

(4)

Accordingly, the resulting smoothed QR estimator is

\[
\hat{\beta}_h(\tau) := \arg \min_{b \in \mathbb{R}^d} \hat{R}_h(b; \tau).
\]

(5)

By smoothing the integrating measure of the objective function, we ensure that the map \( b \mapsto \hat{R}_h(b; \tau) \) is twice continuously differentiable. This is in stark contrast with the nonsmoothness of the standard objective function \( b \mapsto \hat{R}(b; \tau) \). The differentiability of the objective function is very convenient. First, it enables us to compute \( \hat{\beta}_h(\tau) \) using standard Newton-Raphson algorithms, therefore avoiding the usual computational concerns that arise in the context of standard quantile regression. Second, differentiability circumvents the usual inferential issues that plague the standard QR estimator because it allows us to estimate the covariance matrix of our quantile slope coefficient estimates in a canonical manner (see, for instance, Newey and McFadden, 1994). In what follows, we discuss inference a bit further.
We first observe that
\[ \hat{R}_h(b; \tau) = (1 - \tau) \int_{-\infty}^{0} \hat{F}_h(v; b) \, dv + \tau \int_{0}^{\infty} (1 - \hat{F}_h(v; b)) \, dv, \]
and hence the first-order derivative of \( \hat{R}_h(b; \tau) \) with respect to \( b \) is
\[ \hat{R}_h^{(1)}(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} X_i \left[ K \left( -\frac{e_i(b)}{h} \right) - \tau \right], \]
where \( K(t) := \int_{-\infty}^{t} k(v) \, dv \). In the same fashion, the second-order derivative of \( \hat{R}_h(b; \tau) \) with respect to \( b \) is
\[ \hat{R}_h^{(2)}(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} X_i X_i' k_h(-e_i(b)), \]
ensuring enough smoothness for inference. In particular, smoothness allows us to construct asymptotic confidence intervals for \( \hat{\beta}_h(\tau) \) in a standard fashion.

We show in Section 3.4 that \( \sqrt{n}(\hat{\beta}_h(\tau) - \beta(\tau)) \) weakly converges to a Gaussian distribution with mean zero and covariance matrix \( \Sigma(\tau) := D^{-1}(\tau)V(\tau)D^{-1}(\tau) \), where \( V(\tau) := \tau(1 - \tau)E(X'X) \) and \( D(\tau) := R^{(2)}(\beta(\tau); \tau) = E[X'f(X'\beta(\tau)|X)] \) is the Hessian of the objective function evaluated at the true parameter. In addition, we show that
\[ \hat{\Sigma}_h(\tau) := \hat{D}_h^{-1}(\tau)\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau) \]
with \( \hat{D}_h(\tau) := \hat{R}_h^{(2)}(\hat{\beta}_h(\tau); \tau) \) and
\[ \hat{V}_h(\tau) := \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \left[ K \left( -\frac{e_i(\hat{\beta}_h(\tau))}{h} \right) - \tau \right]^2, \]
is a consistent estimator of \( \Sigma(\tau) \). This means that we may compute a \((1 - \alpha)-\)confidence interval for the \( k \)th QR coefficient \( \beta_k(\tau) \) as
\[ CI_{1-\alpha}(\beta_k(\tau)) := \hat{\beta}_{k,h}(\tau) \pm \frac{z_{\alpha/2}}{\sqrt{n}} \hat{\sigma}_{k,h}(\tau), \]
where \( \hat{\sigma}_{k,h}(\tau) \) is the square root of the \( k \)th diagonal entry of \( \hat{\Sigma}_h(\tau) \), \( \hat{\beta}_{k,h}(\tau) \) is the \( k \)th element of the smoothed QR estimator, and \( z_{\alpha} \) is the \( \alpha \) quantile of a standard Gaussian distribution.

Before moving to the asymptotic theory, it is useful to consider for a moment the one-sample scenario with \( d = 1 \) and \( X = 1 \). The first-order condition \( \hat{R}_h^{(1)}(\hat{\beta}_h(\tau); \tau) = 0 \) solved by our smoothed estimator reduces to the first-order condition in Nadaraya (1964), namely, \( \hat{F}_h(\hat{\beta}_h(\tau)) = \tau. \)
This confirms that the smoothing we apply to the objective function essentially boils down to
estimating the conditional cumulative distribution function using a kernel. This is in contrast with
the smoothing put forth by Horowitz (1998), who replaces the indicator function in \( \hat{R}(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} e_i(b) (\tau - I[e_i(b) < 0]) \) by a kernel counterpart.

In particular, Horowitz’s smoothed objective function reads
\[
\tilde{R}_h(b; \tau) := \frac{1}{n} \sum_{i=1}^{n} e_i(b) \left[ \tau - K \left( \frac{-e_i(b)}{h} \right) \right],
\]
whose first-order derivative with respect to \( b \) is equal to \( \hat{R}_h^{(1)}(b; \tau) \) plus an additional term that
depends explicitly on the kernel \( k \). This means that his first-order condition involves a quantity
analogous to a kernel-based estimator of the probability density function. As a result, while the
second-order derivative of our convolution-type kernel QR estimator is similar to a kernel density
estimator, Horowitz’s involves terms that relate to a kernel estimator of the derivative of a proba-
bility density function. As kernel density estimators converge at a faster rate than kernel derivative
estimators, we expect our estimator to have better higher-order properties than Horowitz’s. In ad-
dition, we will push this argument a bit further to show that, in linear asymptotic approximations,
the variance of Horowitz’s smoothed QR estimator is greater than the variance of our smoothed
estimator.

3 Asymptotic theory

We start with some notation. Let \( \| \cdot \| \) denote the Euclidean norm of a vector or a matrix, namely,
\[ \| A \| = \sqrt{\text{tr}(AA^T)} \]. We denote by \( f(\cdot | x) \) the conditional probability density function of \( Y \) given
\( X = x \), with \( j \)th partial derivative given by \( f^{(j)}(y | x) := \frac{\partial^j}{\partial y^j} f(y | x) \). Similarly, let \( Q^{(1)}(\tau | x) := \frac{\partial}{\partial \tau} Q(\tau | x) \). Also, define \( \lfloor s \rfloor \) as the lower integer part of any positive real number \( s \), that is to say,
the unique integer number satisfying \( \lfloor s \rfloor < s \leq \lfloor s \rfloor + 1 \). Note that the latter implies that \( \lfloor s \rfloor = s - 1 \)
if \( s \) is integer. It is perhaps easier to interpret \( \hat{\beta}_h(\tau) \) as an estimator not of \( \beta(\tau) \), but rather of
\( \beta(\tau) \) := \( \arg \min_{b \in \mathbb{R}^d} R_h(b; \tau) \), with \( R_h(b; \tau) := \mathbb{E}[\hat{R}_h(b; \tau)] \). We shall refer to \( \beta_h(\tau) \) as the smoothed
parameter. Finally, it is also convenient to introduce the quantity \( \hat{S}_h(\tau) := \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \) and to
set \( D_h(\tau) := R_h^{(2)}(\hat{\beta}_h(\tau); \tau) \).

In what follows, we first discuss the assumptions we require to work out the asymptotic theory
and then derive the asymptotic mean squared error and Bahadur-Kiefer representation for our
smoothed QR estimator. We wrap up the session with some inference implications. In particular,
we show not only how to construct asymptotic valid confidence intervals, but also how to implement Wald-type tests.

### 3.1 Assumptions

In this section, we discuss the conditions under which we derive the asymptotic theory. Apart from standard technical conditions on the covariates and kernel function, we essentially require the conditional quantile and density functions to be smooth enough. In particular, we assume the following conditions.

**Assumption X** The components of $X$ are positive, bounded random variables, i.e. the support of $X$ is a bounded subset of $\mathbb{R}^d_+$. The matrix $E(XX')$ is full rank.

**Assumption Q** The conditional quantile and density functions $Q(\tau | x)$ and $f(y | x)$ satisfy

Q1 The map $\tau \mapsto \beta(\tau)$ is continuously differentiable over $(0, 1)$. The conditional density $f(y | x)$ is continuous and strictly positive over $\mathbb{R} \times \text{supp}(X)$.

Q2 There are some $s \geq 1$ and $L > 0$ such that $f^{(s)}(\cdot | x)$ exists and $\sup_{x, y} |f^{(j)}(y | x)| \leq L$, with $\lim_{y \to \pm \infty} f^{(j)}(y | x) = 0$, for all $j = 0, \ldots, \lfloor s \rfloor$. Moreover, it holds for all $x \in \text{supp}(X)$ and all $y, w \in \mathbb{R}$ that $|f^{(\lfloor s \rfloor)}(y | x) - f^{(\lfloor s \rfloor)}(y + w | x)| \leq L |w|^{s-\lfloor s \rfloor}$.

**Assumption K** The kernel function $k$ and bandwidth $h$ satisfy

K1 The kernel $k : \mathbb{R} \to \mathbb{R}$ is even, integrable, piecewise differentiable with a bounded derivative, and such that $\int k(z) \, dz = 1$ and $0 < \int_0^\infty K(z) \, dz < \infty$. In addition, for $s$ as in Assumption Q2, $\int |z^{s+1}k(z)| \, dz < \infty$, and $k$ is orthogonal to all nonconstant monomials of degree up to $\lfloor s \rfloor$, i.e., $\int z^j k(z) \, dz = 0$ for $j = 1, \ldots, \lfloor s \rfloor + 1$.

K2 $h \in [h_n, \tilde{h}_n]$ with $1/h_n = O(n/\ln^3 n)$ and $\tilde{h}_n = o(1)$.

Some remarks are in order. First, observe that $R^{(2)}(b; \tau) = \mathbb{E}[XX'f(X'b | X)]$ is positive definite for all $b$ and any $\tau$ under Assumptions Q1 and X. This means that $D^{-1}(\tau)$ exists for every $\tau$. Second, Assumption Q1 also ensures that $\tau \mapsto Q(\tau | x)$ is strictly increasing over $(0, 1)$, with a strictly positive derivative with respect to $\tau$ given that $Q^{(1)}(\tau | x) = 1/f(Q(\tau | x) | x)$. Third, we
assume that \(Y\) has support on the real line merely for notational simplicity. It is straightforward to relax it with some minor adaptations. Fourth, the reason why we choose a kernel such that 
\[
\int_0^\infty K(z) [1 - K(z)] \, dz \quad \text{is positive will become clear in Theorem 3. It makes sense because it guarantees that our convolution-type kernel QR estimator dominates the standard QR estimator in the AMSE sense.}
\]
Fifth, in view that Assumption K does not preclude high-order kernels, \(\hat{f}_h (\cdot; b)\) is not necessarily a density, even if it is a consistent estimator of the the density of the error term \(e(b)\). It is nonetheless possible to show 
\[
\hat{R}_h (b; \tau) = \frac{1}{n} \sum_{i=1}^n \rho \ast k_h (e_i (b)),
\]
where * is the convolution operation. This means that, technically speaking, it is probably more rigorous to interpret our approach as a convolution-type smoothing.

Finally, it is also important to clarify two points. The first is that, for any integer \(s\) in Assumption Q2, the order of differentiability of \(f(\cdot | x)\) is \([s] = s - 1\) and not \(s\) as one may think at first. The second is that the bandwidth \(h\) implicitly depends on the sample size \(n\) through its lower and upper limits in Assumption K2. This is obviously paramount because we wish to entertain data-driven bandwidths.

### 3.2 Bahadur-Kiefer representation

In this section, we study the order of the Bahadur-Kiefer representation for \(\sqrt{n} \left( \hat{\beta}_h (\tau) - \beta (\tau) \right)\). We proceed in two steps. We first establish that the difference between \(\beta_h (\tau)\) and \(\beta (\tau)\) is asymptotically negligible and then derive a Bahadur-Kiefer representation for \(\sqrt{n} \left( \hat{\beta}_h (\tau) - \beta_h (\tau) \right)\). The first step is important because it is well known that smoothing entails bias in finite samples. In fact, one could well argue that the smoothed QR estimator \(\hat{\beta}_h (\tau)\) is actually an (unbiased) estimator of \(\beta_h (\tau) := \arg \min_{b \in \mathbb{R}^d} R_h (b; \tau)\), with \(R_h (b; \tau) := E \left[ \hat{R}_h (b; \tau) \right]\). Out first result shows that the smoothing bias indeed shrinks to zero as the sample size grows.

**Theorem 1** For \(\bar{h}_n\) small enough, \(\beta_h (\tau)\) is under Assumptions X, Q and K unique for every \(\tau \in [\bar{\tau}, \bar{\tau}]\) and such that \(\beta_h (\tau) = \beta (\tau) + O (h^{s+1})\) uniformly for \((\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [\bar{h}_n, \bar{h}_n]\). Additionally, if \(s\) is an integer number and the conditional density \(f(\cdot | x)\) is \(s\)-times continuously differentiable for all \(x\), then \(\beta_h (\tau) = \beta (\tau) - h^{s+1} B (\tau) + o (h^{s+1})\), where 
\[
B (\tau) = \frac{\int_{\mathbb{R}^d} z^{s+1} k(z) \, dz}{(s+1)!} D^{-1} (\tau) E \left[ X f^{(s)} (X' \beta (\tau) | X) \right].
\]

Theorem 1 settles the issue of possible side effects of smoothing in that \(\beta_h (\tau)\) is eventually uniformly close to the true parameter \(\beta (\tau)\). This also serves as a further justification for the
standardization in the Bahadur-Kiefer representation we document next.

**Theorem 2** Under Assumptions $X$, $Q$ and $K$, $\hat{\beta}_h(\tau)$ not only is unique for $(\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [h_n, \bar{h}_n]$ with probability tending to 1, but also satisfies the following representation uniformly with respect to $(\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [\bar{h}_n, \bar{h}_n]$:

$$\sqrt{n} \left( \hat{\beta}_h(\tau) - \beta(\tau) \right) = -\sqrt{n} D^{-1}(\tau) \hat{S}_h(\tau) + O_p(g_n(h)), \quad (6)$$

where $g_n(h) = \sqrt{\ln n/(nh)}$.

Theorems 1 and 2 are relevant for mainly two reasons. First, they show that $\hat{\beta}_h(\tau)$ is, in a sense, more linear than the standard QR estimator $\hat{\beta}(\tau)$. Knight (2001), Jurečková et al. (2012) and Portnoy (2012) show that, in many cases of interest, the Bahadur-Kiefer representation for the standard QR estimator is given by

$$\sqrt{n} \left( \hat{\beta}(\tau) - \beta(\tau) \right) = -\sqrt{n} D^{-1}(\tau) \hat{S}(\tau) + O_p(n^{-1/4}),$$

with $\hat{S}(\tau) = \hat{R}^{(1)}(\beta(\tau); \tau)$. In contrast, the remainder term in (6) is of order nearly $O_p(n^{-1/2})$ for proper bandwidth choices. Second, they imply that $\hat{\beta}_h(\cdot)$ offers a fair global picture of $\beta(\cdot)$ in that

$$\|\hat{\beta}_h(\tau) - \beta(\tau)\| = O_p \left( \frac{1}{\sqrt{n}} + h^{s+1} \right), \quad (7)$$

uniformly for $\tau \in [\bar{\tau}, \bar{\tau}]$ and $h \in [h_n, \bar{h}_n]$. Accordingly, the remainder in (7) is of order $O_p(n^{-1/2})$ for any $h \leq O \left( n^{-1/(2(s+1))} \right)$.

Lastly, it is important to stress the major role that uniformity plays here. It ensures that, if a stochastic process $\{\hat{h}(\tau); \tau \in [\bar{\tau}, \bar{\tau}]\}$ has sample paths in $[h_n, \bar{h}_n]$ with a sufficiently high probability, then (6) remains valid even if we replace $h$ with a data-driven bandwidth $\hat{h}(\tau)$. The next result states this property in a rigorous manner.

**Corollary 1** If $\{\hat{h}(\tau); \tau \in [\bar{\tau}, \bar{\tau}]\}$ satisfy $\Pr(\hat{h}(\tau) \in [h_n, \bar{h}_n] \text{ for all } \tau) \to 1$, then (6) holds with $\hat{h}(\tau)$ in place of $h$, uniformly in $\tau$.

The asymptotic theory so far posits that our smoothed QR estimator entails a better Bahadur-Kiefer representation that the standard QR estimator and that we may employ a data-driven bandwidth that depends on the quantile level and covariates. In the next section, we complement the asymptotic theory by computing the AMSE of our convolution-type kernel QR estimator.
3.3 Asymptotic mean squared error

The previous section shows that our smoothed QR estimator has no asymptotic bias. This means it suffices to compute the asymptotic variance in order to obtain the AMSE. The next result not only documents the asymptotic variance of \( \hat{\beta}_h(\tau) \), but also shows that it is smaller than the asymptotic variance of \( \hat{\beta}(\tau) \).

**Theorem 3**  **Assumptions X, Q and K ensure that**

\[
\forall \left( \sqrt{n} D_h^{-1}(\tau) \hat{S}_h(\tau) \right) = \Sigma(\tau) - c_k h D^{-1}(\tau) + O(h^{2\wedge s}),
\]

with \( c_k = 2 \int_0^{\infty} K(y)[1 - K(y)] dy > 0 \), uniformly with respect to \((\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [\bar{h}_n, \bar{h}_n]\).

Table 1 reports the values of \( c_k \) for Gaussian-type kernels that satisfy Assumption K1. Because the second term in (8) is negative, the asymptotic variance of \( \hat{\beta}_h(\tau) \) is indeed smaller than the asymptotic variance of \( \hat{\beta}(\tau) \).

**Table 1: Examples of Gaussian-type kernels**

We provide examples of kernel functions that satisfy Assumption K1 and their corresponding kernel constants \( c_k \). We denote by \( \phi \) the standard Gaussian density.

<table>
<thead>
<tr>
<th>Gaussian-type kernel</th>
<th>kernel constant</th>
<th>order</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k(x) = \phi(x) )</td>
<td>( c_k = \frac{1}{\sqrt{\pi}} )</td>
<td>2</td>
</tr>
<tr>
<td>( k(x) = \frac{3}{2} \left( 1 - \frac{x^2}{3} \right) \phi(x) )</td>
<td>( c_k = \frac{7}{16\sqrt{\pi}} )</td>
<td>4</td>
</tr>
<tr>
<td>( k(x) = \frac{15}{8} \left( 1 - \frac{2x^2}{3} + \frac{x^4}{15} \right) \phi(x) )</td>
<td>( c_k = \frac{321}{1024\sqrt{\pi}} )</td>
<td>6</td>
</tr>
<tr>
<td>( k(x) = \frac{35}{16} \left( 1 - \frac{x^2}{2} + \frac{x^4}{15} - \frac{x^6}{105} \right) \phi(x) )</td>
<td>( c_k = \frac{4175}{16384\sqrt{\pi}} )</td>
<td>8</td>
</tr>
</tbody>
</table>

We next focus on obtaining the bandwidth \( h^* \) that minimizes the asymptotic mean squared error of \( \lambda' \hat{\beta}_h(\tau) \):

\[
\text{AMSE}(\lambda' \hat{\beta}_h(\tau)) = \mathbb{E} \left[ \lambda' (\hat{\beta}_h(\tau) - D_h^{-1}(\tau) \hat{S}_h(\tau) - \beta(\tau)) \right]^2.
\]

This quantity approximates the mean squared error \( \text{MSE}(\lambda' \hat{\beta}_h(\tau)) = \mathbb{E} \left[ \lambda' (\hat{\beta}_h(\tau) - \beta(\tau)) \right]^2 \) by essentially ignoring the asymptotically negligible remainder term of the Bahadur-Kiefer representation.
Theorem 4  Let Assumptions X, Q and K hold. If $s$ is integer and the conditional density $f(\cdot | x)$ is $s$-times continuously differentiable for all $x$, then $\text{AMSE}(\lambda \hat{\beta}_h(\tau))$ is minimal for

$$h^* = \left( \frac{c_k \lambda' D^{-1}(\tau) \lambda}{2n(s+1)(\lambda B(\tau))^2} \right)^{\frac{1}{s+1}},$$

and equal to $\text{AMSE}(\lambda' \hat{\beta}_{h^*}(\tau)) = \frac{1}{n} \lambda' (\Sigma(\tau) - c_k h^* D^{-1}(\tau)) \lambda + o(n^{-1}).$

Before turning our attention to statistical inference, it is worth emphasizing that Theorem 4 remains valid for any bandwidth of the form $h = [1 + o(1)] h^*$.

### 3.4 Inference

The previous section provides the expression for the asymptotic covariance matrix of our kernel-based QR estimator. This section discusses how to consistently estimate the asymptotic covariance matrix of $\hat{\beta}_h(\tau)$. Recall that $\hat{\Sigma}_h(\tau) := \hat{D}_h^{-1}(\tau) \hat{V}_h(\tau) \hat{D}_h^{-1}(\tau)$, with $\hat{D}_h(\tau) := \hat{R}_h^{(2)}(\hat{\beta}_h(\tau); \tau)$ and

$$\hat{V}_h(\tau) := \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \left[ K \left( - \frac{e_i(\hat{\beta}_h(\tau))}{h} \right) - \tau \right]^2.$$

The next result establishes not only that $\hat{\Sigma}_h(\tau)$ is a consistent estimator of $\Sigma(\tau)$, but also the asymptotic normality of our smoothed QR estimator.

**Theorem 5**  It follows from Assumptions X, Q and K that

$$\hat{\Sigma}_h(\tau) = \Sigma(\tau) + O_p \left( \frac{1}{h \sqrt{n}} + h \right)$$

uniformly in $(\tau, h) \in [\tau, \bar{\tau}] \times [h_n, \bar{h}_n]$. Now, if $\sqrt{n} h_n^{s+1} \to 0$, then $\sqrt{n} \hat{\Sigma}_h^{-1/2}(\tau) (\hat{\beta}_h(\tau) - \beta(\tau))$ weakly converges to a standard Gaussian distribution as $n \to \infty$.

The first result in Theorem 5 ensures that one may construct asymptotic confidence intervals and implement Wald-type tests in a canonical fashion. The second result in Theorem 5 is analogous to the asymptotic standard normality of $\sqrt{n} \Sigma(\tau)^{-1/2}(\hat{\beta}(\tau) - \beta(\tau))$. As in Azzalini (1981), the difference is at the second order. Our asymptotic derivations indeed show that our convolution-type kernel QR estimator entails a second-order refinement to the standard QR estimator in view that it obtains a lower AMSE for a a properly chosen bandwidth.
4 Monte Carlo study

To assess how well the asymptotic theory reflects the performance of our estimator in finite samples, we run some simulations for a simple median linear regression: \( Y = X'\beta + \epsilon \), where \( X = (1, \tilde{X}) \), with \( \tilde{X} \sim U[1,5] \), and \( \beta \equiv \beta(1/2) = (1,1) \). We entertain three different specifications for the error distribution. The first is asymmetric, positing that \( \epsilon = Z - \ln 2/\sqrt{2} \), where \( Z \) comes from an exponential distribution with parameter \( 1/\sqrt{2} \). The second specification displays heavy tails, with \( \epsilon = \sqrt{2/3} Z \), where \( Z \) is t-student with 3 degrees of freedom. The third distribution exhibits conditional heteroskedasticity, with \( \epsilon = \frac{1}{4}(1+\tilde{X}) Z \), where \( Z \sim N(0,1) \). The last two specifications are as in Horowitz (1998), Whang (2006) and Kaplan and Sun (2017). The parameters are such that the conditional quantile of \( \epsilon \) given \( X \) is zero in every specification.

In each of the 100,000 replications, we sample \( n \in \{100,1000\} \) observations from the above models and then compute Koenker and Bassett’s (1978) standard median regression estimator (MR), Horowitz’s (1998) smoothed median regression estimator (SMR), and our convolution-type kernel estimator (CKMR). We also compute the empirical coverage of their asymptotic confidence intervals at the 90%, 95% and 99% levels. We compute the smoothed estimators using a standard Gaussian kernel and a bandwidth grid with values ranging from 0.08 to 0.80, with increments of 0.02. In addition, we also evaluate the smoothed estimators at Silverman’s (1986) rule-of-thumb bandwidth \( h_{ROT} = 1.06 \hat{s} n^{-1/5} \), where \( \hat{s} \) is the sample standard deviation of the residuals from the standard quantile regression fit to the data. We compute the standard errors for the standard QR estimator as in Koenker (2005, Sections 3.4.2 and 4.10.1), whereas the standard errors for the smoothed median regression estimator are as in Horowitz (1998, Section 2). Finally, we estimate the standard errors for the convolution-type kernel median regression estimator using the square root of the diagonal entries of \( \hat{\Sigma}_h(1/2) \).

Figures 1 to 3 show how each median regression estimator performs across the different distributions and sample sizes. For simplicity, they focus exclusively on the slope coefficient. For every figure, the upper panel considers a sample size of \( n = 100 \) observations, whereas the lower

---

3 The parameters are also such that the unconditional variance of \( \epsilon \) is equal to two, i.e., \( V(\epsilon) = 2 \), in both heavy-tailed and heteroskedastic models.

4 A code in \texttt{R} based on the \texttt{optim} optimization function is readily available from the authors upon request.

5 The results for the median regression intercepts are very similar as one would expect, and hence we omit them to conserve on space. They are obviously available from the authors upon request.
panel contemplates the larger sample sizes of $n = 1,000$ observations. In each panel, plot (a) displays the mean squared error of the smoothed estimators relative to the standard median regression estimator. Plots (b) to (d) then depict the empirical coverage of the asymptotic confidence intervals for the MR, SMR and CKMR estimators at the 90%, 95% and 99% levels. We gauge the latter as the proportion of replications in which the absolute value of the t-statistic is below the corresponding percentile in the standard normal distribution (namely, 1.64, 1.96 and 2.58 at the 90%, 95% and 99% confidence level, respectively). The smoothed estimators obviously depend on the bandwidth choice, and hence we plot their results as a function of the bandwidth $h$ as well as single out the corresponding outcomes for the average rule-of-thumb bandwidth $\bar{h}_{ROT}$ across the 100,000 replications. The inside tick marks in the horizontal axis of each plot depict the deciles of the distribution of the rule-of-thumb bandwidth across replications. The latter distribution seems symmetric in every instance, with dispersion reducing drastically once we move from a sample size of 100 to 1,000 observations.

Figure 1 reports the estimation results for the asymmetric specification based on an exponential distribution. Smoothing seems to pay off as both SMR and CKMR typically entail lower MSE than the standard MR estimator. Interestingly, the relative mean squared errors of the smoothed estimators do not seem to change much as a function of the bandwidth as the sample size increases. However, the same does not apply if we focus exclusively on the relative MSE evaluated at the rule-of-thumb bandwidth. MSE improvement slightly decreases for the convolution-type kernel estimator, whereas it essentially shrinks to zero for Horowitz’s (1998) smoothed median regression estimator. In turn, the confidence intervals match very well their nominal values for every estimator, though Horowitz’s smoothed estimator seems to perform a bit worse, especially for the larger sample size. All in all, our convolution-type kernel estimator performs better than the latter for any reasonable bandwidth value (say, for any $h$ not much higher than the maximum value of the rule-of-thumb bandwidth across the 100,000 replications) in that MSE is not only lower, but also more robust to variations in the bandwidth value.

Figures 2 and 3 display similar relative MSE patterns respectively for the models with heavy tails and conditional heteroskedasticity. The empirical coverage of the asymptotic confidence intervals still match fairly well their nominal values. The confidence intervals of the standard QR estimator
exhibit good coverage in every instance, whereas the CKMR confidence intervals are a bit too tight for the smaller sample size. As before, the SMR coverage still deteriorates substantially as the sample size increases from 100 to 1,000 observations. It is interesting to note the that rule-of-thumb bandwidth works much better than it does for the first specification with asymmetric errors. This is not very surprising because the rule of thumb assumes a Gaussian distribution as reference. The t-student distribution is symmetric and bell-shaped, just as the mixture of normals implied by the conditional heteroskedastic model. As they are much closer to Gaussianity than the exponential distribution, the rule of thumb should indeed yield a better approximation of the optimal bandwidth than it does in the asymmetric case.

5 Concluding Remarks

This paper proposes smoothing the entire objective function of a linear quantile regression. The resulting convolution-type kernel estimator has a uniform Bahadur-Kiefer representation that improves on the extant QR estimators. In addition, we also show that it is straightforward to compute asymptotic standard errors for the QR coefficient estimates as well as to conduct Wald-type inference. We run a simple Monte Carlo study to assess the relative performance of our convolution-type kernel QR estimator in finite samples. We find that it outperforms not only the standard QR estimator, but also the smoothed QR estimator of Horowitz (1998) for reasonable bandwidth choices.

It would be interesting to further examine data-driven bandwidth implementation as in, for instance, Lepski, Mammen and Spokoiny (1997). There are also many possible extensions: e.g., adapting our convolution-type kernel smoothing to panel quantile regressions (Galvão and Kato, 2016) and extending the framework to deal with endogenous quantile regressions (Chernozhukov and Hansen, 2005, 2006). For instance, Chernozhukov and Hansen’s (2008) robust inferential procedure for the linear quantile model with endogeneity relies on a first-step instrumental variables quantile regression. We conjecture that employing our convolution-type kernel smoothing in the first-step QR would improve the accuracy of the asymptotic confidence intervals for the second-step structural quantile regression slope coefficients.

---

6 In this vein, see also de Castro, Galvão and Kaplan (2017), who extend Kaplan and Sun’s (2017) smoothing ideas to the nonlinear quantile model with endogeneity and weakly dependent data.
A Technical Appendix: Proofs

Notice that \( \hat{R}(b; \tau) \) is integrable if and only if \( Y \) and \( X \) are integrable. This matters when defining \( \beta(\tau) \) as the minimizer of \( R(b; \tau) := \mathbb{E} \left[ \hat{R}(b; \tau) \right] \). It is convenient to assume that both \( \hat{R}(b; \tau) \) and \( \hat{R}_h(b; \tau) \) are integrable such that \( R(b; \tau) \) and \( R_h(b; \tau) \) are well defined. Otherwise, one should define \( R_h(b; \tau) = \mathbb{E} \left[ \hat{R}_h(b; \tau) - \hat{R}_h(0; \tau) \right] \), and similarly for \( R(b; \tau) \). These quantities are finite under Assumption X. Let \( S \) denote the set \( \mathbb{R}^d \times [\tau, \tau] \times [b_n, \bar{b}_n] \) to which \( (b, \tau, h) \) belongs. Note that \( S \) depends on \( n \) through \( [b_n, \bar{b}_n] \). In what follows, \( C \) denotes a generic constant that may vary from line to line.

A.1 Smoothing bias

To study the bias of our smoothed QR estimator, we make use of the following result.

**Lemma 1** Assumptions X, Q2 and K1 ensure that

\[
\begin{align*}
(i) \quad & \sup_{(b, \tau, h) \in S} \left| \frac{R_h(b; \tau) - R(b; \tau)}{h^{s+1}} \right| = O(1); \\
(ii) \quad & \sup_{(b, \tau, h) \in S} \left\| \frac{R_h^{(1)}(b; \tau) - R(b; \tau)}{h^{s+1}} \right\| = O(1); \\
(iii) \quad & \sup_{(b, \tau, h) \in S} \left\| \frac{R_h^{(2)}(b; \tau) - R^{(2)}(b; \tau)}{h^s} \right\| = O(1); \\
(iv) \quad & \sup_{(\delta, b, \tau, h) \in \mathbb{R}^d \times S} \left\| \frac{R_h^{(2)}(b + \delta; \tau) - R_h^{(2)}(b; \tau)}{\|\delta\|} \right\| = O(1).
\end{align*}
\]

**Proof of Lemma 1** Assume that \(|s| \geq 1\). This is without loss of generality given that it is straightforward to deal with \(|s| = 0\) in a similar fashion by imposing the Hölder condition on \( f(\cdot \mid x) \) instead of restricting \( f^{(\lfloor s \rfloor)}(\cdot \mid x) \) as below.

(i) Under Assumption Q2, a Taylor expansion with integral remainder gives

\[
f(v + hz \mid x) = \sum_{\ell=0}^{\lfloor s \rfloor - 1} \frac{f^{(\ell)}(v \mid x)}{\ell!} (hz)^\ell + \frac{(hz)^{\lfloor s \rfloor}}{([s] - 1)!} \int_0^1 f^{(\lfloor s \rfloor)}(v + whz \mid x)(1 - w)^{[s]-1} dw.
\]

In turn, Assumption K1 ensures that

\[
\mathbb{E}[k_h(v - Y) \mid x] - f(v \mid x) = \int k_h(v - y) f(y \mid x) dy - f(v \mid x)
\]

\[
= \int k(z) (f(v + hz \mid x) - f(v \mid x)) dz
\]

\[
= \int_0^1 (1 - w)^{\lfloor s \rfloor - 1} \int \frac{(hz)^{\lfloor s \rfloor}}{([s] - 1)!} k(z) f^{(\lfloor s \rfloor)}(v + whz \mid x) dz dw
\]

\[
= \int_0^1 (1 - w)^{\lfloor s \rfloor - 1} \int \frac{(hz)^{\lfloor s \rfloor}}{([s] - 1)!} k(z) \left[ f^{(\lfloor s \rfloor)}(v + whz \mid x) - f^{(\lfloor s \rfloor)}(v \mid x) \right] dz \mathbb{E}(\text{d}(h))
\]

16
through a change of variables $y = v + hz$. Now, the check function is such that
\[
\int \rho_r(v) \, dG(v) = (1 - \tau) \int_{-\infty}^{0} G(v) \, dv + \tau \int_{0}^{\infty} [1 - G(v)] \, dv
\]
for any arbitrary cdf $G$, and hence
\[
R(b; \tau) = \int \left\{ (1 - \tau) \int_{-\infty}^{0} f(v \mid x) \, dv \, dt + \tau \int_{0}^{\infty} \int_{t+\tau}^{\infty} f(v \mid x) \, dv \, dt \right\} \, dF_X(x),
\]
where $F_X(x)$ is the cdf of $X$. Similarly,
\[
R_h(b; \tau) = \int \left\{ (1 - \tau) \int_{-\infty}^{0} \mathbb{E}[k_h(v - Y) \mid x] \, dv \, dt + \tau \int_{0}^{\infty} \int_{t+\tau}^{\infty} \mathbb{E}[k_h(v - Y) \mid x] \, dv \, dt \right\} \, dF_X(x).
\]
It follows from (9) that
\[
L_1 := \left| \int_{-\infty}^{0} \int_{-\infty}^{t+\tau} \left\{ \mathbb{E}[k_h(v - Y) \mid x] - f(v \mid x) \right\} \, dv \, dt \right|
\]
\[
= \left| \int_{0}^{1} (1 - w)|s|^{-1} \int \frac{(hz)^{|s|}}{(|s| - 1)!} k(z) \int_{-\infty}^{0} \int_{-\infty}^{t+\tau} \left[ f(|s|)(v + whz \mid x) - f(|s|)(v \mid x) \right] \, dv \, dt \, mbox{box} \, dz \, dw \right|
\]
\[
= \left| \int_{0}^{1} (1 - w)|s|^{-1} \int \frac{(hz)^{|s|}}{(|s| - 1)!} k(z) \left[ f(|s|)(x'b + whz \mid x) - f(|s|)(x'b \mid x) \right] \, dz \, dw \right| \leq C h^{|s|+1},
\]
given that $\int |z|^{|s|+1}k(z) \, dz < \infty$ by Assumption K1 and that $f(|s|)(\cdot \mid \cdot)$ is Lipschitz. Analogously, $\left| \int_{0}^{1} \int_{t+\tau}^{\infty} \mathbb{E}[k_h(v - Y) \mid x] - f(v \mid x) \, dv \, dt \right| \leq C h^{|s|+1}$, establishing the result.

(ii) By the definitions of $R(b; \tau)$ and $R_h(b; \tau)$, it follows from the Lebesgue Dominated Convergence Theorem that
\[
R^{(1)}(b; \tau) = \mathbb{E}\{X[F(X'b \mid X) - \tau]\} = \int x \left( \int_{-\infty}^{x'b} f(y \mid x) \, dy - \tau \right) \, dF_X(x),
\]
and that
\[
R_h^{(1)}(b; \tau) = \mathbb{E}\left\{ X \left[ K \left( \frac{X'b - Y}{h} \right) - \tau \right] \right\} = \int x \left( \int_{-\infty}^{x'b} \mathbb{E}[k_h(v - Y) \mid x] \, dv - \tau \right) \, dF_X(x). \tag{10}
\]
In view that $\int z^{|s|}k(z) \, dz = \int z^{|s|+1}k(z) \, dz = 0$ and $\int |z|^{|s|+1}k(z) \, dz < \infty$, integrating (9) yields
\[
L_2 := \left| \int_{-\infty}^{x'b} \mathbb{E}[k_h(v - Y) \mid x] - f(v \mid x) \, dv \right|
\]
\[
= \left| \int_{0}^{1} (1 - w)|s|^{-1} \int \frac{(hz)^{|s|}}{(|s| - 1)!} k(z) \int_{-\infty}^{x'b} \left[ f(|s|)(v + whz \mid x) - f(|s|)(v \mid x) \right] \, dv \, dz \, dw \right|
\]
\[
= \left| \int_{0}^{1} (1 - w)|s|^{-1} \int \frac{(hz)^{|s|}}{(|s| - 1)!} k(z) \left[ f(|s|)(x'b + whz \mid x) - f(|s|)(x'b \mid x) \right] \, dz \, dw \right|
\]
\[
= \left| \int_{0}^{1} w(1 - w)|s|^{-1} \int \frac{(hz)^{|s|+1}}{(|s| - 1)!} k(z) \int_{-\infty}^{0} f(|s|)(x'b + twhz \mid x) - f(|s|)(x'b \mid x) \, dt \, dz \, dw \right| \leq C h^{|s|+4}.
\]
by the Hölder condition on $f^{(1)}$. It then suffices to impose Assumption X to establish the result.

(iii) Differentiating $R^{(1)}(b; \tau)$ with respect to $b$ results in

$$R^{(2)}(b; \tau) = \mathbb{E}[XX'f(X'b | X)] = \int xx'f(x'b | x) \, dF_X(x)$$

and, likewise,

$$R^{(2)}_h(b; \tau) = \mathbb{E}[XX'k_h(X'b - Y)] = \int xx'\mathbb{E}[k_h(x'b - Y) | x] \, dF_X(x).$$

Setting $v = x'b$ in (9) then yields

$$\left\| R^{(2)}_h(b; \tau) - R^{(2)}(b; \tau) \right\| \leq C |\mathbb{E}[k_h(v - Y) | x] - f(v | x)| \leq C h^s,$$

under Assumptions X and Q2, as stated.

(iv) Recall that

$$R^{(2)}_h(b; \tau) = \mathbb{E}[XX'k_h(X'b - Y)] = \int k(z) \int xx'f(x'b + hz | x) \, dF_X(x) \, dz.$$

Under Assumption Q2, it ensues from $f(\cdot | \cdot)$ being Lipschitz for $|s| \geq 1$ that

$$\left\| R^{(2)}_h(b + \delta; \tau) - R^{(2)}_h(b; \tau) \right\| \leq C \int |k(z)| \int \|xx'\| |x'\delta| \, dF_X(x) \, dz \leq C \|\delta\|,$$

uniformly in $b, \delta, \tau$, completing the proof.

**Proof of Theorem 1**

It follows from the convexity of $b \mapsto R_h(b; \tau)$ that $\beta_h(\tau)$ is well defined as long as $\tilde{h}_n$ is small enough. Let

$$A(\tau, h) := \int_0^1 R^{(2)}(\beta(\tau) + w(\beta_h(\tau) - \beta(\tau)); \tau) \, dw$$

and let $c := \inf f \left( x' \left[ \beta(\tau) + w(\beta_h(\tau) - \beta(\tau)) \right] | x \right)$ for $(\tau, h, w, x) \in [\bar{\tau}, \bar{\tau}] \times [b_n, \bar{h}_n] \times [0, 1] \times \text{supp}(X)$.

If the map $(\tau, h) \mapsto \beta_h(\tau)$ is continuous, then Assumption Q1 ensures that $c > 0$, and so

$$v' A(\tau, h) v = \int_0^1 \left( v' x \right)^2 f \left( x' \left[ \beta(\tau) + w(\beta_h(\tau) - \beta(\tau)) \right] | x \right) \, dF_X(x) \, dw \geq c v \mathbb{E}(XX') v > 0$$

by Assumption X. This means that the eigenvalues of $A(\tau, h)$ are bounded away from zero, uniformly in $(\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [b_n, \bar{h}_n]$. However, since $R^{(1)}_{h}(\beta_h(\tau); \tau) = R^{(1)}(\beta(\tau); \tau) = 0$, a Taylor expansion with integral remainder leads to

$$R^{(1)}(\beta_h(\tau); \tau) - R^{(1)}_h(\beta_h(\tau); \tau) = R^{(1)}(\beta_h(\tau); \tau) - R^{(1)}(\beta(\tau); \tau)$$

$$= A(\tau, h) [\beta_h(\tau) - \beta(\tau)],$$

and hence, by (ii) in Lemma 1, $\beta_h(\tau) - \beta(\tau) = O(h^{s+1})$ given that

$$\left\| \frac{\beta_h(\tau) - \beta(\tau)}{h^{s+1}} \right\| \leq \sup_{(\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [b_n, \bar{h}_n]} \left\{ \left\| A(\tau, h)^{-1} \right\| \left\| \frac{R^{(1)}(\beta_h(\tau); \tau) - R^{(1)}_h(\beta_h(\tau); \tau)}{h^{s+1}} \right\| \right\} = O(1).$$

18
It now remains to establish the continuity of \((\tau, h) \mapsto \beta_h(\tau)\). First, observe that \(b \mapsto R^{(2)}(b; \tau) := \mathbb{E}[XX'f(X'b) | X]\) is continuous and does not depend on \(\tau\). Moreover, for any \(b \in \mathbb{R}^d\), \(R^{(2)}(b; \tau)\) is an element of the open set \(M^+_d\) of positive-definite \(d \times d\) matrices. Indeed, for any \(v \in \mathbb{R}^d\),

\[
v' R^{(2)}(b; \tau)v = \int (v'x)^2 f(x'b | x) \, dF_X(x) \geq v' \mathbb{E}(XX')v \inf_{x \in \text{supp}(X)} f(x'b | x) > 0
\]

by Assumption Q1. For some \(C > 0\), \((iii)\) in Lemma 1 ensures that

\[
\left\| R^{(2)}_h(\beta_h(\tau); \tau) - R^{(2)}_h(\beta_h(\tau); \tau) \right\| \leq C \bar{h}_n^s
\]

uniformly in \((\tau, h) \in [\tau, \bar{\tau}] \times [\bar{h}_n, \bar{h}_n] \). This means that, for a proper choice of \(\bar{h}_n\), \(R^{(2)}_h(\beta_h(\tau); \tau) \in M^+_d\) uniformly in \((\tau, h)\). Accordingly, \((\tau, h) \mapsto \beta_h(\tau)\) is unique and continuous by a straightforward application of the Implicit Function Theorem to the first-order condition \(R^{(1)}_h(\beta_h(\tau); \tau) = 0\). In addition, Taylor expanding the latter results in

\[
-R^{(1)}_h(\beta(\tau); \tau) = R^{(1)}_h(\beta_h(\tau); \tau) - R^{(1)}_h(\beta_h(\tau); \tau) = \left[ R^{(2)}(\beta(\tau); \tau) + O(h^s) \right] [\beta_h(\tau) - \beta(\tau)] + o(h^{s+1}).
\]

by \((iii)\) in Lemma 1. Bearing in mind that \(s = [s] + 1\), \(\int w^2(1-w)^{[s]-1} \, dw = \frac{2}{[s]([s]+1)([s]+2)}\), and

\[
f^{([s])}(x'\beta(\tau) + twhz | x) - f^{([s])}(x'\beta(\tau) | x) = f^{(s)}(x'\beta(\tau) | x) twhz + o(h),
\]

it follows from \((11)\) that

\[
T_2 := \int \left\{ \mathbb{E}[k_h(v - Y) | x] - f(v | x) \right\} dv = h^{s+1} f^{(s)}(x'\beta(\tau) | x) \int_0^1 w^2(1-w)^{[s]-1} \, dw \int_0^{\frac{z^{s+1}}{([s]-1)!}} k(z) \, dz \int_0^1 t \, dt + o(h^{s+1})
\]

\[
= h^{s+1} \int \frac{z^{s+1} k(z) \, dz}{(s+1)!} f^{(s)}(x'\beta(\tau) | x) + o(h^{s+1}).
\]

Using \((10)\) then yields

\[
R^{(1)}_h(\beta(\tau); \tau) = h^{s+1} \int \frac{z^{s+1} k(z) \, dz}{(s+1)!} \int x f^{(s)}(x'\beta(\tau) | x) \, dF_X(x) + o(h^{s+1}),
\]

completing the proof.

\[\textbf{A.2 Bahadur-Kiefer representation}\]

This section makes use of a powerful functional exponential inequality by Massart (2007). For the sake of completeness, we state a version of Massart’s result as a Lemma. For real-valued functions \(f\) and \(\bar{f}\) with \(f \leq \bar{f}\), let \([f, \bar{f}]\) denote the set of all functions \(g\) such that \(f \leq g \leq \bar{f}\). For a set \(\mathcal{F}\) and a family \(\{F_i\}\) of subsets of \(\mathcal{F}\), we say that \(\{F_i\}\) covers \(\mathcal{F}\) if \(\mathcal{F} \subseteq \bigcup_i F_i\).

\[\textbf{Lemma 2} \quad \text{Let} \ Z_i \ \text{be an iid sequence of random variables taking values in the measurable space} \ Z, \ \text{and let} \ \mathcal{F} \ \text{be a class of real valued, measurable functions on} \ Z. \ \text{Assume that}\]

\[\text{(L2a) there are some positive constants} \ \sigma \ \text{and} \ M \ \text{such that} \ \mathbb{E}\left[|f(Z_i)|^2\right] \leq \sigma^2 \ \text{and} \ \sup_{z \in Z} |f(z)| \leq M \ \text{for all} \ f \in \mathcal{F};\]
(L2b) for each \( \delta > 0 \), there exists a set of brackets \( \{ [f_j, \bar{f}_j]; j = 1, \ldots, J(\delta) \} \), for some integer \( J(\delta) > 1 \), that covers \( F \) such that \( \mathbb{E} \left[ |\bar{f}_j(z) - f_j(z)|^2 \right] \leq \delta^2 \) and \( \sup_{z \in Z} |\bar{f}_j(z) - f_j(z)| \leq M \) for every \( j = 1, \ldots, J(\delta) \).

Under (L2a) and (L2b),

\[
\Pr \left( \sup_{f \in F} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(Z_i) - \mathbb{E}[f(Z_i)]) \geq \mathcal{H}_n + 7\sigma \sqrt{2r} + \frac{2Mr}{\sqrt{n}} \right) \leq \exp(-r), \quad \text{for any } r \geq 0, \tag{12}
\]

where

\[
\mathcal{H}_n := 27 \left( \int_0^\sigma H^{1/2}(u) \, du + \frac{2(\sigma + M)H(\sigma)}{\sqrt{n}} \right)
\]

and \( H \) is any nonnegative measurable function of \( \delta > 0 \) satisfying \( H(\delta) \geq \ln J(\delta) \).

**Proof of Lemma 2**  See Corollary 6.9 in Massart’s (2007).

Before proceeding to the next result, let us introduce some additional notation. In what follows, let \( \sup_{(\tau,h)} \) denote the supremum over \((\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [h_n, \bar{h}_n] \), which depends on \( n \) via \([h_n, \bar{h}_n] \).

We also use the same implicit notation for any other similar operator (e.g., infimum or union). Recalling that \( g^{-1}_n(h) := \sqrt{nh}/(\ln n) \), let

\[
\mathcal{E}_n(r) := \left\{ \sqrt{n} \sup_{(\tau,h)} g^{-1}_n(h) \left\| \hat{\beta}_h(\tau) - \beta_h(\tau) + D^{-1}_h(\tau)\hat{S}_h(\tau) \right\| \geq r^2 \right\}.
\]

Notice that the event \( \mathcal{E}_n(r) \) depends on the sample size \( n \) and on the tail parameter \( r \), but neither on \( \tau \) nor on \( h \). On the complementary set of \( \mathcal{E}_n(r) \), it holds that

\[
\sqrt{n}(\hat{\beta}_h(\tau) - \beta_h(\tau)) = -\sqrt{n} D^{-1}_h(\tau)\hat{S}_h(\tau) + \hat{E}_h(\tau),
\]

where the approximation error is such that \( \|\hat{E}_h(\tau)\| \leq g_n(h) r^2 \) uniformly in \( \tau \in [\bar{\tau}, \bar{\tau}] \) and \( h \in [h_n, \bar{h}_n] \). In particular, if \( \Pr(\mathcal{E}_n(r)) \) is small for large \( r \), then the representation \( \hat{\beta}_h(\tau) \) from Theorem 2 holds uniformly in \( \tau \) and \( h \).

Finally, let \( \tilde{A} \) denote the complementary set of \( A \) and

\[
\mathcal{E}^1_n(r) := \left\{ \sup_{(\tau,h)} \left\| \sqrt{n} \hat{S}_h(\tau) \right\| \geq r \right\},
\]

\[
\mathcal{E}^2_n(r) := \left\{ \sup_{(\tau,h)} \sup_{\{b; \|b - \beta_h(\tau)\| \leq 1\}} \left\| \sqrt{\frac{nh}{\ln n}} \left( \hat{R}_h(2)(b; \tau) - R_h^2(b; \tau) \right) \right\| \geq r \right\},
\]

where the norm in \( \mathcal{E}^1_n(r) \) and \( \mathcal{E}^2_n(r) \) are the Euclidean and trace norms, respectively.

We are now ready to state the functional exponential inequality we will apply in the remaining technical proofs.

**Proposition 1**  Given Assumptions X, Q and K, \( \hat{\beta}_h(\tau) \) is unique for all \((\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [h_n, \bar{h}_n] \) with probability growing to 1. There also exist positive constants \( C_0, C_1 \) and \( C_2 \) such that, for \( \epsilon, 1/r \) and \( 1/n \) small enough,

---

\[ ^5 \] The trace norm is such that, for any matrix \( A \) and conformable vector \( v \), \( \|Av\| \leq \|A\|\|v\| \).

---

20
(i) $Pr\left(\mathcal{E}_n(r) \cap \mathcal{E}_n^1(r) \cap \mathcal{E}_n^2(r)\right) \leq C_0 \exp (-nc/C_0);

(ii) $Pr\left(\mathcal{E}_n^1(r)\right) \leq C_1 \exp (-r^2/C_1);

(iii) $Pr\left(\mathcal{E}_n^2(r)\right) \leq C_2 \exp (-r \ln n/C_2)$.

Theorem 2 is an immediate corollary from Proposition 1, whose proof relies on Lemmata 2 to 4. For some proofs, it is convenient to consider the auxiliary objective functions $\hat{R}_h(b; \tau) := R_h(b; \tau) - \hat{R}_h(\beta_h(\tau); \tau)$ and $\hat{R}_h(b; \tau) := E[\hat{R}_h(b; \tau)]$, which are such that $\hat{\beta}_h(\tau) = \arg \min_h \hat{R}_h(b; \tau)$ and $\beta_h(\tau) = \arg \min_b R_h(b; \tau)$. Similarly, set $\hat{R}(b; \tau) = \hat{R}(b; \tau) - \hat{R}(\beta(\tau); \tau)$ and $\hat{R}(b; \tau) := E[\hat{R}(b; \tau)]$.

The next result shows that $\hat{\beta}_h(\tau)$ is close to $\beta_h(\tau)$ uniformly for $(\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [h_n, \bar{h}_n]$.

**Lemma 3** Suppose that Assumptions X, Q and K hold. If $n$ is large enough, there are positive constants $C_0$ and $C_1$ such that

$$Pr\left(\sup_{(\tau, h)} \|\hat{\beta}_h(\tau) - \beta_h(\tau)\| \geq \eta \right) \leq C_0 \exp (-n\eta^4/C_1).$$

for any $\eta \in [1/\ln n, 1]$.

**Proof of Lemma 3** For $\eta > 0$,

$$\left\{ \sup_{(\tau, h)} \|\hat{\beta}_h(\tau) - \beta_h(\tau)\| \geq 2\eta \right\} = \bigcup_{(\tau, h)} \left\{ \|\hat{\beta}_h(\tau) - \beta_h(\tau)\| \geq 2\eta \right\}$$

$$\subset \bigcup_{(\tau, h)} \left\{ \inf_{\{b: \|b - \beta_h(\tau)\| \geq 2\eta\}} \hat{R}_h(b; \tau) \leq \inf_{\{b: \|b - \beta_h(\tau)\| \leq 2\eta\}} \hat{R}_h(b; \tau) \right\}$$

$$\subset \bigcup_{(\tau, h)} \left\{ \inf_{\{b: \|b - \beta_h(\tau)\| \geq 2\eta\}} \hat{R}_h(b; \tau) \leq \hat{R}_h(\beta_h(\tau); \tau) \right\}$$

$$= \bigcup_{(\tau, h)} \left\{ \inf_{\{b: \|b - \beta_h(\tau)\| \geq 2\eta\}} \hat{R}_h(b; \tau) \leq 0 \right\},$$

given that $\hat{R}_h(\beta_h(\tau); \tau) = 0$. Theorem 1 ensures that

$$\{ b : \|b - \beta_h(\tau)\| \geq 2\eta \} \subset \{ b : \|b - \beta(\tau)\| + \sup_{(\tau, h)} \|\beta_h(\tau) - \beta(\tau)\| \geq 2\eta \}$$

$$\subset \{ b : \|b - \beta(\tau)\| + O(\hat{k}_h^{r+1}) \geq 2\eta \}$$

$$\subset \{ b : \|b - \beta(\tau)\| \geq \eta \}$$

for all $(\tau, h)$ provided that $n$ is large enough. This means that

$$\left\{ \sup_{(\tau, h)} \|\hat{\beta}_h(\tau) - \beta_h(\tau)\| \geq 2\eta \right\} \subset \bigcup_{(\tau, h)} \left\{ \inf_{\{b: \|b - \beta(\tau)\| \geq \eta\}} \hat{R}_h(b; \tau) \leq 0 \right\}.$$

As $t \mapsto \rho_r(t)$ is 1-Lipschitz, it follows from

$$\hat{R}_h(b; \tau) = \frac{1}{nh} \sum_{i=1}^n \int \rho_r(t) k \left( \frac{t - (Y_i - X'_i(b))}{h} \right) dt = \frac{1}{n} \sum_{i=1}^n \int \rho_r(Y_i - X'_i(b + hz)) k(z) dz$$

21
that
\[
\left| \hat{R}_h(b; \tau) - \tilde{R}(b; \tau) \right| = \left| \frac{1}{n} \sum_{i=1}^{n} \int_{\tau} \left[ \rho_r(Y_i - X_i' + hz) - \rho_r(Y_i - X_i'b) \right] k(z) \, dz \right| \leq h \int |z k(z)| \, dz < \infty,
\]
for all \( b, \tau \) and \( h \) by Assumption K1. Theorem 1 and the Lipschitz property of \( b \mapsto \hat{R}(b; \tau) \) then ensures that \( \hat{R}_h(b; \tau) \geq \tilde{R}(b; \tau) - Ch \) uniformly in \( b \) and \( \tau \), so that
\[
\left\{ \sup_{(\tau, h)} \left\| \hat{\beta}_h(\tau) - \beta_h(\tau) \right\| \geq 2 \eta \right\} \subseteq \bigcup_{(\tau, h)} \left\{ \inf_{\{b: \|b - \beta(\tau)\| \geq \eta\}} \hat{R}(b; \tau) \leq C h \right\}.
\]
The next step is a convexity argument. We first perform the change of variables \( b = \beta(\tau) + \rho u \) with \( \|u\| = 1 \) and \( \rho \geq \eta \). In view that \( b \mapsto \hat{R}(b; \tau) \) is convex with \( \hat{R}(\beta(\tau); \tau) = 0 \),
\[
\frac{\eta}{\rho} \hat{R}(\beta(\tau) + \rho u; \tau) = \frac{\eta}{\rho} \hat{R}(\beta(\tau) + \rho u; \tau) + \left( 1 - \frac{\eta}{\rho} \right) \hat{R}(\beta(\tau); \tau) \geq \hat{R}(\beta(\tau) + \eta u; \tau).
\]
It follows from the above inequality that
\[
\left\{ \inf_{\{b: \|b - \beta(\tau)\| \geq \eta\}} \hat{R}(b; \tau) \leq C h \right\} \subseteq \left\{ \inf_{\{b: \|b - \beta(\tau)\| = \eta\}} \hat{R}(b; \tau) \leq C h \right\},
\]
and hence
\[
\bigcup_{(\tau, h)} \left\{ \left\| \hat{\beta}_h(\tau) - \beta_h(\tau) \right\| \geq 2 \eta \right\} \subseteq \bigcup_{\tau} \left\{ \inf_{\{b: \|b - \beta(\tau)\| = \eta\}} \hat{R}(b; \tau) \leq C \hat{h}_n \right\}
\]
\[
\subseteq \left\{ \inf_{\tau} \left\{ \inf_{\{b: \|b - \beta(\tau)\| = \eta\}} \left[ \hat{R}(b; \tau) - R(b; \tau) \right] \leq C \hat{h}_n - \inf_{\tau} \left\{ \inf_{\{b: \|b - \beta(\tau)\| = \eta\}} R(b; \tau) \right\} \right\} \right\}. \quad (2)
\]
We next provide an upper bound for \( C \hat{h}_n - \inf_{\tau \in [\bar{\tau}, \tau]} \inf_{\{b: \|b - \beta(\tau)\| = \eta\}} R(b; \tau) \). Because the eigenvalues of \( R^{(2)}(b; \tau) \) are bounded away from 0 uniformly in \( b \), with \( \|b - \beta(\tau)\| \leq 1 \) and \( \tau \in [\bar{\tau}, \tau] \), a second-order Taylor expansion of \( R(b; \tau) = R(b; \tau) - R(\beta(\tau); \tau) \) gives way to
\[
R(b; \tau) = 0 + R^{(1)}(\beta(\tau), \tau)'(b - \beta(\tau)) + (b - \beta(\tau))' \left[ \int_{0}^{1} (1 - t)R^{(2)}(\beta(\tau) + t[b - \beta(\tau)]; \tau) \, dt \right] (b - \beta(\tau)) \geq C \eta^2
\]
for all \( b \) with \( \|b - \beta(\tau)\| = \eta \). It follows that, for any \( \eta_2 = \eta - \epsilon_2 < \eta \) with conformable \( \epsilon_2 \) and \( \hat{h}_n \) small enough,
\[
\bigcup_{(\tau, h)} \left\{ \left\| \hat{\beta}_h(\tau) - \beta_h(\tau) \right\| \geq 2 \eta \right\} \subseteq \left\{ \sup_{\tau \in [\bar{\tau}, \tau]} \sup_{\{b: \|b - \beta(\tau)\| = \eta\}} \text{abs} \hat{R}(b; \tau) - R(b; \tau) \geq C \eta_2^2 \right\}.
\]
Now, let \( Z_i = (Y_i, X_i')' \) and \( \theta = (\tau, b)' \) and \( f(Z_i, \theta) = \rho_r(Y_i - X_i'b) - \rho_r(Y_i - X_i'\beta(\tau)) \), so that
\[
\hat{R}(b; \tau) - R(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} \left\{ f(Z_i, \theta) - \mathbb{E}[f(Z_i, \theta)] \right\}.
\]
Under Assumption X, it follows from \( \eta \leq 1 \) that, for all \( b \) with \( \|b - \beta(\tau)\| = \eta \) and \( \tau \in [\bar{\tau}, \tau] \),
\[
|f(Z_i, \theta)| \leq \|X_i\| \|b - \beta(\tau)\| \leq C,
\]
22
which implies that $\mathbb{V}(f(Z_i, \theta)) \leq \sigma^2 \leq C$. Observe also that pairing Assumption X with the Lipschitz conditions on $\tau \mapsto \beta(\tau)$ in Assumption Q1 and on $\tau \mapsto \rho_r(u)$ entails, for all admissible $z$,\[ |f(z, \theta_1) - f(z, \theta_2)| \leq C \|\theta_1 - \theta_2\|, \]where $\|\theta\|^2 = \|b\|^2 + |\tau|^2$. Next, for $\delta > 0$, let $\theta_j$, with $j = 1, \ldots, J(\delta) \leq C \delta^{-(d+1)}$, be such that\[ \Theta = \{\theta = (b, \tau) : \tau \in [\tau, \tau], \|b - \beta(\tau)\| = \eta_1\} \subset \bigcup_{j=1}^{J(\delta)} B(\theta_j, \delta), \]where $B(\theta_j, \delta)$ is the $\|\cdot\|$-ball with center $\theta_j$ and radius $\delta$. Define $f_j(\cdot)$ and $\tilde{f}_j(\cdot)$ respectively as $f_j(z) := \inf_{\theta \in B(\theta_j, \delta)} f(z, \theta)$ and $\tilde{f}_j(z) = \sup_{\theta \in B(\theta_j, \delta)} f(z, \theta)$, so that $\{f(\cdot, \theta) : \theta \in B(\theta_j, \delta)\} \subset [f_j, \tilde{f}_j]$.

Let $\mathcal{F}_\Theta := \{f(\cdot, \theta) : \theta \in \Theta\} \subset \bigcup_{j=1}^{J(\delta)} [f_j, \tilde{f}_j]$. Observe also that (13) implies $|\tilde{f}_j(z) - f_j(z)| \leq C\delta \leq C$ and $\mathbb{E}\left[|f_j(Z_i) - f_j(Z_i)|^2\right] \leq C\delta^2$. By (L2a) and (L2b), it follows from (12) that setting $H(\delta) = -(d+1) \ln \delta + C$ leads to\[ \Pr\left(\sup_{\theta \in \Theta} \left|\hat{\mathcal{R}}(b; \tau) - \mathcal{R}(b; \tau)\right| \geq C \frac{1 + \sqrt{r + \sqrt{n}}}{\sqrt{n}}\right) \leq \exp(-r). \]

This means that, for $n$ large enough with respect to $\eta_2^2$,
\[ \Pr\left(\sup_{\tau} \sup_{\{b : \|b - \beta(\tau)\| = \eta_1\}} \left|\hat{\mathcal{R}}(b; \tau) - \mathcal{R}(b; \tau)\right| \geq C\eta_2^2\right) \leq C \exp\left(-n C\eta_2^4\right), \]
and hence
\[ \Pr\left(\sup_{(\tau, b)} \left|\hat{\beta}_h(\tau) - \beta_h(\tau)\right| \geq 2\eta\right) \leq C \exp\left(-n C\eta_2^4\right), \]
completing the proof. \hfill \blackslug

**Lemma 4** Suppose that Assumptions X, Q, and K hold and consider $r > 0$ and $\eta \in (0, 1]$. As long as $n$ is large enough,
\[ \Pr\left(\sup_{(\tau, b)} \left\|\sqrt{n}\hat{S}_h(\tau)\right\| \geq C_1(1 + r)\right) \leq C_0 \exp\left(-r^2\right), \]
\[ \Pr\left(\sup_{(\tau, b)} \sup_{\{b : \|b - \beta_h(\tau)\| \leq \eta\}} \left\|\sqrt{\frac{nh}{\ln n}} \left(\hat{R}_h(2)(b, \tau) - R_h(2)(b, \tau)\right)\right\| \geq C_1(1 + r)\right) \leq C_0 \exp\left(-r \ln n\right). \]

**Proof of Lemma 4** We start with the first deviation probability. As $\hat{R}_h^{(1)}[\beta_h(\tau), \tau] = 0$,
\[ \sup_{(\tau, b)} \left\|\sqrt{n}\hat{R}_h^{(1)}[\beta_h(\tau), \tau]\right\| \leq \sup_{(\tau, b)} \sup_{\{b : \|b - \beta_h(\tau)\| \leq \eta\}} \left\|\sqrt{n} \left(\hat{R}_h^{(1)}(b, \tau) - R_h^{(1)}(b, \tau)\right)\right\|. \]
However,
\[ \hat{R}_h^{(1)}(b, \tau) = \frac{\partial}{\partial b} \left[ \frac{1}{n} \sum_{i=1}^{n} \int \rho_r(Y_i - X'_i b + h z) k(z) \, dz \right] = \frac{1}{n} \sum_{i=1}^{n} X_i \left[ \int 1(Y_i - X'_i b + h z < 0) k(z) \, dz - \tau \right]. \]
implying that \( \hat{R}_h^{(1)}(b, \tau) = \sum_{i=1}^n f(Z_i, \theta)/n \), with
\[
f(Z_i, \theta) = X_i \left[ \mathbb{I}(Y_i - X'_i b + h z < 0) k(z) \, dz - \tau \right],
\]
for \( Z_i = (Y_i, X'_i) \) and \( \theta \in \Theta := \{(b', h, \tau) : (r, h) \in [\tau, \bar{\tau}] \times [b_n, \bar{b}_n], \|b - \beta_h(\tau)\| \leq \eta\} \). We bound each of the entries of \( \hat{R}_h^{(1)}(b, \tau) \), so that there is no loss of generality in assuming that \( X_i \) is univariate. Note that \( |f(Z_i, \theta)| \leq C \) and \( \mathbb{V}(f(Z_i, \theta)) \leq \sigma^2 \leq C \) as well as that, for all \( \theta_1 \) and \( \theta_2 \),
\[
|f(Z_i, \theta_2) - f(Z_i, \theta_1)| \leq C.
\]
Let \( \|\theta\|^2 = |b|^2 + |h|^2 + |\tau|^2 \) and let \( \mathcal{B}(\theta, \delta^2) \) denote the \( \|\cdot\|\)-ball with center \( \theta \) and radius \( \delta^2 \). Assumption X ensures that, for any \( \theta_1 \) and \( \theta_2 \) in \( \mathcal{B}(\theta, \delta^2) \),
\[
|f(Z_i, \theta_2) - f(Z_i, \theta_1)| \leq C \left[ \int \mathbb{I}(Y_i - X'_i b + h z \in [-C \delta^2, C \delta^2]) \, |k(z)| \, dz + \delta^2 \right]. \tag{14}
\]
Consider a covering of \( \Theta \) with \( J(\delta) \leq C \delta^{-2(d+1)} \) balls \( \mathcal{B}(\theta_j, \delta^2) \), and let \( f_j(z) := \inf_{\theta \in \mathcal{B}(\theta_j, \delta)} f(z, \theta) \) and \( \bar{f}_j(z) = \sup_{\theta \in \mathcal{B}(\theta_j, \delta)} f(z, \theta) \), so that \( \{f(\cdot, \theta) : \theta \in \mathcal{B}(\theta_j, \delta)\} \subset [f_j, \bar{f}_j] \) and \( \mathcal{F}_\theta := \{f(\cdot, \theta) : \theta \in \Theta\} \subset \bigcup_{j=1}^{J(\delta)} [f_j, \bar{f}_j] \). Equation \((14)\) gives that, uniformly in \( j \) and \( \delta^2 \leq \sigma^2 \),
\[
\mathbb{E}\left[|\bar{f}_j(Z_i) - f_j(Z_i)|^2\right] \leq C \delta^4 + C \mathbb{E}\left[\int \mathbb{I}(Y_i - X'_i b + h z \in [-C \delta^2, C \delta^2]) \, |k(z)| \, dz\right]^2.
\]
Applying the Cauchy-Schwarz inequality under Assumptions K and Q2 then gives way to
\[
L_4 := \mathbb{E}\left[\int \mathbb{I}(Y_i - X'_i b + h z \in [-C \delta^2, C \delta^2]) \, |k(z)| \, dz\right]^2
\leq \mathbb{E}\left[\int \mathbb{I}(Y_i - X'_i b + h z \in [-C \delta^2, C \delta^2]) \, |k(z)| \, dz\right] \times \int |k(z)| \, dz
\leq \int \Pr(Y_i \in X_i b + h z \in [-C \delta^2, C \delta^2] | X_i = x) \, |k(z)| \, dz \times \int |k(z)| \, dz
\leq C \delta^2,
\]
implying that
\[
\mathbb{E}\left[|\bar{f}_j(Z_i) - f_j(Z_i)|^2\right] \leq C (\delta^4 + \delta^2) \leq C \delta^2
\]
uniformly in \( j \) and \( \delta^2 \leq \sigma^2 \). As a result, both (L2a) and (L2b) hold with \( \ln H(\delta) = -2(d+1) \ln \delta + C \), so that \((12)\) gives
\[
\Pr\left(\sup_{\theta \in \Theta} \|\sqrt{n} \left( \hat{R}_h^{(1)}(b, \tau) - R_h^{(1)}(b, \tau) \right) \| \geq C \left( \sqrt{T} + 1 + r/\sqrt{n} \right) \right) \leq 2 \exp(-r).
\]
Accordingly, the first bound holds for \( n \) large enough. As for the second bound, there is no loss of generality to assume that \( X_i \) is unidimensional. Note that \( \sqrt{nh/\ln n} \hat{R}_h^{(2)}(b, \tau) = \sum_{i=1}^n g(Z_i, \theta)/\sqrt{n} \), with
\[
g(Z_i, \theta) := \sqrt{\frac{1}{h \ln n}} X_i^2 k \left( \frac{X'_i b - Y_i}{h} \right).
\]
Assumptions K and X ensures that, uniformly in \( \theta \in \Theta \),
\[
|g(Z_i, \theta)| \leq C \sqrt{\frac{1}{h \ln n}} \leq C O(\sqrt{n}) \ln^2 n = M_n/2.
\]
It also follows from Assumption Q2 that, uniformly in \( \theta \in \Theta \),

\[
\forall (g(Z_i, \theta)) \leq \frac{C}{h \ln n} \int \int k \left( \frac{x' - y}{h} \right) f(y \mid x) \, dy \, dF_X(x)
\]

\[
= \frac{C}{ln n} \times \int \int k(v) f(x' + hv \mid x) \, dv \, dF_X(x) \leq \frac{C}{ln n} = \sigma_n^2.
\]

Assumption K posits that \( g(Z_i, \theta) \) is Lipschitz over \( \Theta \) with a polynomial in \( n \) Lipschitz coefficient, that is, for any \( \theta_1 \) and \( \theta_2 \) in \( \Theta \), \( |g(Z_i, \theta_2) - g(Z_i, \theta_1)| \leq C n^C \| \theta_2 - \theta_1 \| \). Consider a covering of \( \Theta \) with \( J(\delta/n^C) \leq C (\delta/n^C)^{- (d+1)} \) balls \( \mathcal{B}(\theta_1, \delta/n^C) \) and let \( g_j(z) := \inf_{\theta \in \mathcal{B}(\theta_j, \delta)} g(z, \theta) \) and \( \bar{g}_j(\cdot) := \sup_{\theta \in \mathcal{B}(\theta_j, \delta)} g(z, \theta) \). It then turns out that \( \{ g(z, \theta) \in \mathcal{B}(\theta_j, \delta) \} \subset [\bar{g}_j, \bar{g}_j] \) and then \( G_3 = \{ g(\cdot, \theta) : \theta \in \Theta \} \subset \bigcup_j J(\delta/n^C) \big[ \bar{g}_j, \bar{g}_j \big] \), with \( E \left[ \bar{g}(Z_i) - g(Z_i) \right] \leq C \delta^2 \). It then follows that (L2a) and (L2b) hold with

\[
\ln H(\delta) = -2(d + 1) (\ln \delta - C \ln n) + C,
\]

so that (12) results for any \( u > 0 \) in

\[
\Pr \left( \sup_{\theta \in \Theta} \left\| \frac{1}{\ln n} \left( \hat{R}_h^{(2)}(b, \tau) - R_h^{(2)}(b, \tau) \right) \right\| \geq C \left( 1 + \frac{\sqrt{u}}{\sqrt{n \ln n}} + \frac{u}{\ln n} \right) \right) \leq 2 \exp(-u).
\]

Setting \( u = t \ln n \) then yields the result.

**Proof of Proposition 1**

Let

\[
\mathcal{E}_n^3(\epsilon) := \left\{ \sup_{(\tau, \hat{\beta}_h)} \left\| \hat{\beta}_h(\tau) - \beta_h(\tau) \right\| \geq \epsilon^{1/4} \right\},
\]

which is such that \( \Pr (\mathcal{E}_n^3(\epsilon)) \leq C \exp(-C n \epsilon) \) by Lemma 3. The bounds for \( \Pr (\mathcal{E}_n^3(r)) \) and \( \Pr (\mathcal{E}_n^2(\epsilon)) \) follow from Lemma 4. In particular, \( \lim_{n \to \infty} \Pr (\mathcal{E}_n^2(\epsilon)) = 0 \), whereas Lemma 1 ensures under Assumption X that \( b \to \hat{R}_h(b; \tau) \) is strictly convex for \( b \) in a vicinity of \( \beta_h(\tau) \), for all \( \tau \in [\tau, \tilde{\tau}] \) with at least \( 1 - \Pr (\mathcal{E}_n^1(\epsilon)) \). But, by Lemma 3 and Theorem 1, all minimizers of \( \hat{R}_h(b; \tau) \) lie in such a vicinity with a probability tending to 1. This means that we can make \( 1 - \Pr (\mathcal{E}_n^1(\epsilon)) \) arbitrarily close to 1 by increasing \( r \), and hence \( \hat{\beta}_h(\tau) \) is unique with a probability going to 1 as \( n \) increases. It also follows that, in case \( \mathcal{E}_n^1(\epsilon), \mathcal{E}_n^2(\epsilon) \) and \( \mathcal{E}_n^3(\epsilon) \) are all true and \( n \) is large enough, \( \hat{\beta}_h(\tau) \) satisfies the first-order condition \( \hat{R}_h^{(1)}(\hat{\beta}_h(\tau); \tau) = 0 \) and, accordingly,

\[
-\hat{R}_h^{(1)}(\beta_h(\tau); \tau) = \hat{R}_h^{(1)}(\hat{\beta}_h(\tau); \tau) - \hat{R}_h^{(1)}(\beta_h(\tau); \tau) = \int_0^1 \hat{R}_h^{(2)}(\beta_h(\tau) + t[\hat{\beta}_h(\tau) - \beta_h(\tau)]; \tau) \, dt.
\]

Now, if \( \epsilon \in \mathcal{E}_n^3(\epsilon) \) is small enough, the eigenvalues of the above matrix are in \([1/C, C]\) for a large \( C \) provided that \( n \) is large enough, uniformly in \( \tau \) and \( h \). This means that

\[
\hat{\beta}_h(\tau) - \beta_h(\tau) = \left( \int_0^1 \hat{R}_h^{(2)}(\beta_h(\tau) + u[\hat{\beta}_h(\tau) - \beta_h(\tau)]; \tau) \, du \right)^{-1} \hat{R}_h^{(1)}(\beta_h(\tau); \tau).
\]
Lemma 1 then implies that
\[ P_2 := \left\| \sqrt{n} \left( \hat{\beta}_h(\tau) - \beta_h(\tau) \right) + \left[ R_h^{(2)}(\beta_h(\tau); \tau) \right]^{-1} \sqrt{n} \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right\| \]
\[ \leq C \left\| \int_0^1 \left[ \hat{R}_h^{(2)}(\beta_h(\tau) + u[\hat{\beta}_h(\tau) - \beta_h(\tau)]; \tau) \right] - R_h^{(2)}(\beta_h(\tau) + u[\beta_h(\tau) - \beta_h(\tau)]; \tau) \right\| \, du \left\| \sqrt{n} \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right\| \]
\[ + C \left\| \int_0^1 \left[ R_h^{(2)}(\beta_h(\tau) + u[\hat{\beta}_h(\tau) - \beta_h(\tau)]; \tau) \right] - R_h^{(2)}(\beta_h(\tau); \tau) \right\| \, du \left\| \sqrt{n} \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right\| \]
\[ \leq C \left\{ C^2 \sqrt{\frac{\ln n}{nh}} \tau^2 + C \left\| \hat{\beta}_h(\tau) - \beta_h(\tau) \right\| \left\| \sqrt{n} \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right\| \right\} \]
\[ \leq C \left\{ C^2 \sqrt{\frac{\ln n}{nh}} \tau^2 + C^2 n^{-1/2} \left\| \sqrt{n} \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right\|^2 \right\} \]
\[ \leq C \left( \sqrt{\frac{\ln n}{nh}} + \frac{1}{\sqrt{n}} \right) \tau^2 \]
on \hat{\epsilon}_n^1(r) and \hat{\epsilon}_n^2(r), implying that \hat{\epsilon}_n(r) holds as long as \( C \) is large enough. ■

**Proof of Theorem 2** As uniqueness holds, it suffices now to observe that one can make
\( Pr(\mathcal{E}_n(r)) \leq Pr(\mathcal{E}_n^1(r) \cap \mathcal{E}_n^2(r)) + Pr(\mathcal{E}_n^1(r)) + Pr(\mathcal{E}_n^2(r)) \) arbitrarily small for large \( n \) by fixing \( \epsilon \) and increasing \( \cdot \) ■

**A.3 Asymptotic variance and mean squared error**

**Proof of Theorem 3** We first show that the expansion
\[ \sqrt{n} \hat{D}^{-1}(\tau) \hat{S}_h(\tau) = \Sigma(\tau) - c_k h D^{-1}(\tau) + O(h^2) \] (15)
holds uniformly with respect to \( (\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [\bar{h}, \bar{h}] \). Given
\[ \mathbb{E} \left[ \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right] = 0, \]
\[ \mathbb{V} \left( \sqrt{n} \hat{S}_h(\tau) \right) = \mathbb{V} \left( \sqrt{n} \hat{R}_h^{(1)}(\beta_h(\tau); \tau) \right) \]
\[ = \mathbb{V} \left( X \left[ K \left( \frac{X' \beta_h(\tau) - Y}{h} \right) - \tau \right] \right) = \mathbb{E} \left( XX' \left[ K \left( \frac{X' \beta_h(\tau) - Y}{h} \right) - \tau \right] \right) \]
\[ = \mathbb{E} \left[ XX' K^2 \left( \frac{X' \beta_h(\tau) - Y}{h} \right) \right] - 2 \tau \mathbb{E} \left[ XX' K \left( \frac{X' \beta_h(\tau) - Y}{h} \right) \right] + \tau^2 \mathbb{E}(XX'). \]
Along similar lines to the proof of Lemma 1, it follows from Assumptions Q2 and K that
\[ \mathbb{E} \left[ K \left( \frac{X' \beta_h(\tau) - Y}{h} \right) \mid X = x \right] = \int K \left( \frac{x' \beta_h(\tau) - y}{h} \right) f(y \mid x) \, dy = \frac{1}{h} \int k \left( \frac{x' \beta_h(\tau) - y}{h} \right) F(y \mid x) \, dy \]
\[ = F(x' \beta_h(\tau) \mid x) + \int [F(x' \beta_h(\tau) - h z \mid x) - F(x' \beta_h(\tau) \mid x)] k(z) \, dz \]
\[ = \tau + O(h^{s+1}), \] (16)
using integration by parts and Theorem 1, in view that \( x' \beta(\tau) = F^{-1}(\tau \mid x) \) by definition. Let now
\[
K(z) = 2 k(z) K(z) = \frac{dK^2(z)}{dz},
\]
so that \( \int K(z) \, dz = \lim_{z \to \infty} K^2(z) = 1 \). As before, this leads to
\[
\begin{align*}
\mathbb{E}
\left[K \left( \frac{x' \beta_h(\tau) - Y}{h} \right)^2 \middle| X = x \right] &= \frac{1}{h} \int K \left( \frac{x' \beta_h(\tau) - y}{h} \right) F(y \mid x) \, dy \\
&= \tau + O(h^{k+1}) + \int \left[ F \left( x' \beta_h(\tau) - h z \mid x \right) - F \left( x' \beta_h(\tau) \mid x \right) \right] K(z) \, dz \\
&= \tau + O(h^{k+1}) - h \left[ f(x' \beta_h(\tau) \mid x) + O(h) \right] \int z K(z) \, dz \\
&= \tau + O(h^{k+1}) - h \left[ f(x' \beta(\tau) \mid x) + O(h^{k+1}) + O(h) \right] \int z K(z) \, dz \\
&= \tau - h f(x' \beta(\tau) \mid x) \int z K(z) \, dz + O(h^2).
\end{align*}
\]
The variance expansion \((15)\) then follows by noticing that, as \( K(-z) = 1 - K(z) \),
\[
\int z K(z) \, dz = 2 \int z k(z) K(z) \, dz = \int_{-\infty}^{0} z \, dK^2(z) + \int_{0}^{\infty} z \, d \left[ K^2(z) - 1 \right] \\
= - \int_{-\infty}^{0} K^2(z) \, dz + \int_{0}^{\infty} (1 - K^2(z)) \, dz = \int_{0}^{\infty} \left( 1 - K(z) \right)^2 + 1 - K^2(z) \right) \, dz \\
= 2 \int_{0}^{\infty} K(z)[1 - K(z)] \, dz.
\]
Finally, given \((15)\), Lemma 1 and Theorem 1, the local Lipschitz property of matrix inversion ensures that
\[
\forall \left( \sqrt{n} D^{-1}_h(\tau) \tilde{S}_h(\tau) D^{-1}_h(\tau) \right) = \Sigma(\tau) - c_k h D^{-1}(\tau) + O(h^2) \\
+ \left[ D^{-1}_h(\tau) \mathbb{V} \left( \sqrt{n} \tilde{S}_h(\tau) \right) D^{-1}_h(\tau) - D^{-1}(\tau) \mathbb{V} \left( \sqrt{n} \tilde{S}_h(\tau) \right) D^{-1}(\tau) \right] .
\]
To establish \((8)\), it suffices to observe that the norm of the last term within brackets is at most equal to
\[
\| D^{-1}_h(\tau) - D^{-1}(\tau) \| \left\| \mathbb{V} \left( \sqrt{n} \tilde{S}_h(\tau) \right) \right\| \left( \| D^{-1}_h(\tau) \| + \| D^{-1}(\tau) \| \right) \leq C \| D^{-1}_h(\tau) - D^{-1}(\tau) \| = O(h^s). \]

**Proof of Theorem 4**

It follows from Theorem 1, \((8)\) and \( \mathbb{E} \left[ \tilde{S}_h(\tau) \right] = 0 \) that
\[
\text{AMSE} \left( \chi' \tilde{\beta}_h(\tau) \right) = h^{2s+2} \left[ \chi' \mathbb{B}(\tau) \right]^2 + \frac{1}{n} \chi' \left( \Sigma(\tau) - c_k h D^{-1}(\tau) \right) \lambda + O \left( \frac{h^{s+2}}{n} \right) + o \left( h^{2s+2} \right) .
\]
Letting \( g(h) = h^{2s+2} \left[ \chi' \mathbb{B}(\tau) \right]^2 - n^{-1} c_k h \chi' D^{-1}(\tau) \lambda \) and then differentiating with respect to \( h \) yields
\[
\begin{align*}
g'(h) &= (2s + 2) h^{2s+1} \left[ \chi' \mathbb{B}(\tau) \right]^2 - \frac{1}{n} c_k \chi' D^{-1}(\tau) \lambda .
\end{align*}
\]
Solving for \( h^* \) such that \( g'(h^*) = 0 \) yields the desired result and AMSE expansion.
A.4 Asymptotic covariance estimator

**Lemma 5** Under Assumptions $X$, $Q$ and $K$, it turns that (i) $\|\hat{D}_h(\tau) - D(\tau)\| = O_p\left(\sqrt{\frac{\ln n}{nh}} + \frac{1}{\sqrt{n}} + h^s\right)$, (ii) $\|\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)\| = O_p\left(\frac{\sqrt{\ln n}}{n} + \frac{1}{\sqrt{n}} + h^s\right)$, and (iii) $\|\hat{V}_h(\tau) - V(\tau)\| = O_p\left(\sqrt{\frac{\ln n}{n}} + \frac{1}{\sqrt{n}} + h\right)$ uniformly with respect to $(\tau, h) \in [\tau, \bar{\tau}] \times [h_n, \bar{h}_n]$.

**Proof of Lemma 5** As for item (i), note that

$$
\|\hat{D}_h(\tau) - D(\tau)\| \leq \|\hat{R}_h^{(2)}(\beta_h(\tau); \tau) - R_h^{(2)}(\beta(\tau); \tau)\| + \|R_h^{(2)}(\beta(\tau); \tau) - R_h^{(2)}(\beta(\tau); \tau)\|.
$$

Lemma 4 ensures that the first term on the right-hand side is $O_p\left(\sqrt{\ln n/(nh)}\right)$ uniformly for $(\tau, h) \in [\tau, \bar{\tau}] \times [h_n, \bar{h}_n]$, whereas

$$
\|R_h^{(2)}(\beta_h(\tau); \tau) - R_h^{(2)}(\beta(\tau); \tau)\| \leq \|R_h^{(2)}(\beta_h(\tau); \tau) - R_h^{(2)}(\beta(\tau); \tau)\| + \|R_h^{(2)}(\beta(\tau); \tau) - R_h^{(2)}(\beta(\tau); \tau)\| = O(h^s),
$$

by Lemma 1. In addition, Lemma 1 and Theorems 1 and 2 ensure that

$$
\|R_h^{(2)}(\beta_h(\tau); \tau) - R_h^{(2)}(\beta(\tau); \tau)\| \leq C \|\hat{\beta}_h(\tau) - \beta(\tau)\|
$$

$$
\leq C \left(\|\hat{\beta}_h(\tau) - \beta_h(\tau)\| + \|\beta_h(\tau) - \beta(\tau)\|\right) = O_p\left(n^{-1/2} + h^{s+1}\right),
$$

establishing (i). Item (ii) follows readily from (i) and the locally Lipschitz property of matrix inversion. As for item (iii), let $W(b; \tau) := \mathbb{E}\left(XX'[\|Y - X'b \leq 0\|^2] - \tau\right)$ and $\hat{W}_h(b; \tau) := \mathbb{E}\left[\hat{W}_h(b; \tau)\right]$, with

$$
\hat{W}_h(b; \tau) := \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \left[K \left(\frac{X_i'b - Y_i}{h}\right) - \tau\right]^2,
$$

so that $W(\beta(\tau); \tau) = V(\tau)$ and $\hat{W}_h(\beta_h(\tau); \tau) = \hat{V}_h(\tau)$. Now,

$$
\|\hat{V}_h(\tau) - V(\tau)\| = \|\hat{W}_h(\beta_h(\tau); \tau) - W(\beta(\tau); \tau)\|
$$

$$
\leq \|\hat{W}_h(\beta_h(\tau); \tau) - W(\beta_h(\tau); \tau)\| + \|W(\beta_h(\tau); \tau) - W(\beta(\tau); \tau)\|
$$

$$
\leq \|\hat{W}_h(\beta_h(\tau); \tau) - W(\beta_h(\tau); \tau)\| + \|W(\beta_h(\tau); \tau) - W(\beta(\tau); \tau)\|
$$

$$
+ C \mathbb{E}\left[\left(K \left(\frac{X'b - Y}{h}\right) - \tau\right)^2\right] - \tau(1 - \tau)
$$

$$
= \|\hat{W}_h(\beta_h(\tau); \tau) - W(\beta_h(\tau); \tau)\| + \|W(\beta_h(\tau); \tau) - W(\beta(\tau); \tau)\| + O(h)
$$

by Assumption X and equations (16) and (17). However, given that both $K$ and $K^2$ are Lipschitz, $\hat{W}_h(b; \tau)$ and $W_h(b; \tau)$ are also Lipschitz, with a Lipschitz constant of order $O\left(h^{-1}\right)$. It then follows from Assumption X and Theorem 3x that

$$
\|W_h(\beta_h(\tau); \tau) - W_h(\beta(\tau); \tau)\| \leq \frac{C}{h} \|\hat{\beta}_h(\tau) - \beta_h(\tau)\| = O_p\left(n^{-1/2}h^{-1}\right),
$$

28
and hence
\[
\|\hat{V}_h(\tau) - V(\tau)\| \leq \|\hat{W}_h(\hat{\beta}_h(\tau); \tau) - W_h(\hat{\beta}_h(\tau); \tau)\| + O_p\left(n^{-1/2}h^{-1}\right) + O(h).
\]
To complete the proof, it suffices to apply a similar argument to the proof of Lemma 4.

**Proof of Theorem 5** We have
\[
\|\hat{\Sigma}_h(\tau) - \Sigma(\tau)\| \leq \|\hat{D}_h^{-1}(\tau)\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau)\| \\
+ \|D^{-1}(\tau)\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)V(\tau)D^{-1}(\tau)\| \\
\leq \|\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)\| \|\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau)\| + \|D^{-1}(\tau)\| \|\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau) - V(\tau)D^{-1}(\tau)\| \\
\leq \|\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)\| \|\hat{V}_h(\tau)\hat{D}_h^{-1}(\tau)\| \\
+ \|D^{-1}(\tau)\| \left( \|\hat{D}_h^{-1}(\tau)\| \|\hat{V}_h(\tau) - V(\tau)\| + \|V(\tau)\| \|\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)\| \right).
\]
In view that \(\|D^{-1}(\tau)\|\) and \(\|V(\tau)\|\) are \(O(1)\) uniformly for \(\tau \in [\bar{\tau}, \bar{\tau}]\) and that \(\|\hat{V}_h(\tau)\|\) and \(\|\hat{D}_h^{-1}(\tau)\|\) are \(O_p(1)\) uniformly for \((\tau, h) \in [\bar{\tau}, \bar{\tau}] \times [\bar{h}_n, \bar{h}_n]\), it turns out that
\[
\|\hat{\Sigma}_h(\tau) - \Sigma(\tau)\| = O_p\left(\|\hat{D}_h^{-1}(\tau) - D^{-1}(\tau)\| + \|\hat{V}_h(\tau) - V(\tau)\|\right).
\]
The stated results then immediately follow from Lemma 5 and from the asymptotic normality of the leading term in the Bahadur-Kiefer representation, respectively.
References


Knight, K., 2001, Comparing conditional quantile estimators: First and second order considerations, Technical report, University of Toronto.


Koenker, R., 2005, Quantile Regression, Cambridge University Press.


Figure 1: Performance of the QR estimators with exponential errors
We display the mean squared error relative to the standard QR estimator, as well as the empirical coverage of their asymptotic confidence intervals at the 90%, 95% and 99% levels, as a function of the bandwidth $h$. We consider sample sizes of $n = 100$ and $n = 1,000$ observations. QR corresponds to Koenker and Bassett’s (1978) standard QR estimator, SQR to Horowitz’s (1998) smoothed QR estimator, and CKQR to our convolution-type kernel QR estimator. Finally, $h_{ROT}$ denotes the average rule-of-thumb bandwidth across the 100,000 replications.

$a = 100$

(a) Relative MSE of slope estimator

(b) Nominal = 0.90

(c) Nominal = 0.95

(d) Nominal = 0.99

$n = 1,000$

(a) Relative MSE of slope estimator

(b) Nominal = 0.90

(c) Nominal = 0.95

(d) Nominal = 0.99
Figure 2: Performance of the smoothed QR estimators with t-student errors
We display the mean squared error relative to the standard QR estimator, as well as the empirical coverage of their asymptotic confidence intervals at the 90%, 95% and 99% levels, as a function of the bandwidth \( h \). We consider sample sizes of \( n = 100 \) and \( n = 1,000 \) observations. QR corresponds to Koenker and Bassett’s (1978) standard QR estimator, SQR to Horowitz’s (1998) smoothed QR estimator, and CKQR to our convolution-type kernel QR estimator. Finally, \( h_{ROT} \) denotes the average rule-of-thumb bandwidth across the 100,000 replications.

\[ n = 100 \]

(a) Relative MSE of slope estimator

(b) Nominal = 0.90

(c) Nominal = 0.95

(d) Nominal = 0.99

\[ n = 1,000 \]

(a) Relative MSE of slope estimator

(b) Nominal = 0.90

(c) Nominal = 0.95

(d) Nominal = 0.99
Figure 3: Performance of the smoothed QR estimators with heteroskedastic errors
We display the mean squared error relative to the standard QR estimator, as well as the empirical coverage of their asymptotic confidence intervals at the 90%, 95% and 99% levels, as a function of the bandwidth $h$. We consider sample sizes of $n = 100$ and $n = 1,000$ observations. QR corresponds to Koenker and Bassett’s (1978) standard QR estimator, SQR to Horowitz’s (1998) smoothed QR estimator, and CKQR to our convolution-type kernel QR estimator. Finally, $h_{ROT}$ denotes the average rule-of-thumb bandwidth across the 100,000 replications.

$$n = 100$$

(a) Relative MSE of slope estimator

(b) Nominal = 0.90

(c) Nominal = 0.95

(d) Nominal = 0.99

$$n = 1,000$$

(a) Relative MSE of slope estimator

(b) Nominal = 0.90

(c) Nominal = 0.95

(d) Nominal = 0.99