Smoothing Quantile Regressions

Marcelo Fernandes, Emmanuel Guerre, and Eduardo Horta

QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left are hyperlinked to the location of the query in your paper.

The title and author names are listed on this sheet as they will be published, both on your paper and on the Table of Contents. Please review and ensure the information is correct and advise us if any changes need to be made. In addition, please review your paper as a whole for typographical and essential corrections.

Your PDF proof has been enabled so that you can comment on the proof directly using Adobe Acrobat. For further information on marking corrections using Acrobat, please visit http://journalauthors.tandf.co.uk/production/acrobat.asp; https://authorservices.taylorandfrancis.com/how-to-correct-proofs-with-adobe/

The CrossRef database (www.crossref.org/) has been used to validate the references.

AUTHOR QUERIES

Q1 Please check whether the author names (first name followed by last name) and affiliations are correct as presented in the proofs.

Q2 Please provide the Department name and full postal address for all affiliations.

Q3 To maintain sequential order, figures and its citations are renumbered. Please check and provide revision if this is inaccurate.

Q4 Please provide a short description of the supplementary materials.

Q5 Please update the reference "de Castro et al. (2019)" with volume number and page range.

Smoothing Quantile Regressions

Q1
Marcelo FERNANDES
Sao Paulo School of Economics, FGV (*marcelo.fernandes@fgv.br*)

Emmanuel GUERRE
University of Kent (*e.o.guerre@kent.ac.uk*)

Q2 **Eduardo HORTA**
Universidade Federal do Rio Grande do Sul (*eduardo.horta@ufrgs.br*)

We propose to smooth the objective function, rather than only the indicator on the check function, in a linear quantile regression context. Not only does the resulting smoothed quantile regression estimator yield a lower mean squared error and a more accurate Bahadur–Kiefer representation than the standard estimator, but it is also asymptotically differentiable. We exploit the latter to propose a quantile density estimator that does not suffer from the curse of dimensionality. This means estimating the conditional density function without worrying about the dimension of the covariate vector. It also allows for two-stage efficient quantile regression estimation. Our asymptotic theory holds uniformly with respect to the bandwidth and quantile level. Finally, we propose a rule of thumb for choosing the smoothing bandwidth that should approximate well the optimal bandwidth. Simulations confirm that our smoothed quantile regression estimator indeed performs very well in finite samples. Supplementary materials for this article are available online.

KEY WORDS: Asymptotic expansion; Bahadur–Kiefer representation; Conditional quantile; Convolution-based smoothing; Data-driven bandwidth.

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.

There is a price to pay, though. The objective function that the standard QR estimator aims to minimize is not smooth. As established in Bassett and Koenker (1982), it follows that the paths of this estimator have jumps, even if the underlying quantile function is very regular. Because the objective function does not have second derivatives, statistical inference is not straightforward and involves ancillary estimation of nuisance parameters (namely, the asymptotic covariance matrix depends on the population conditional density evaluated at the true quantile). See the discussions in Koenker (1994), Buchinsky (1995), Koenker (2005), Goh and Knight (2009), and Fan and Liu (2016), among others. Unsurprisingly, the literature now 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 (Whang 2006; Otsu 2008), strong approximation methods (Portnoy 2012), nonstandard inference (Goh and Knight 2009), and other nonparametric approaches (Fan and Liu 2016; Mammen, Van Keilegom, and Yu 2017).

Further, the asymptotic normality of the standard QR estimator relies on Bahadur–Kiefer representations with poor rates

of convergence. The latter is at best of order $n^{-1/4}$ for iid (homoscedastic) errors (see Koenker and Portnoy 1987; Knight 2001; Jurečková, Sen, and Picek 2012). This means that the first-order *linear* Gaussian approximation for the distribution of the QR estimators could well fail in finite samples. Portnoy (2012), however, obtains a nearly $n^{-1/2}$ rate using a nonlinear approximation, explaining perhaps why QR inference is actually rather good in finite samples despite the poor rates of the Bahadur–Kiefer remainder.

In this article, we employ a convolution-type smoothing of the objective function that helps produce a continuous QR estimator, which is not only less irregular and variable than the standard estimator but also more linear, in the sense that the stochastic order of the Bahadur–Kiefer remainder term is much closer to $n^{-1/2}$ for a proper bandwidth choice. As the smoothing we propose ensures the twice differentiability of the objective function with respect to the parameter vector, we can readily estimate the asymptotic covariance matrix of our smoothed QR estimator using a standard sandwich formula. In fact, the second derivative of our kernel-smoothed objective function indeed coincides with the usual kernel-based covariance matrix estimators in the standard QR setup.

Quantile estimation aims to offer a global picture of the distribution. Bearing this in mind, we contemplate an asymptotic theory that holds uniformly with respect to quantile levels, or in a functional sense. This is in line with the functional

central limit theorem that Koenker and Xiao (2002) establish for specification-testing purposes. Apart from uniformity in quantile level, we also focus on bandwidth uniformity in consistency and asymptotic distribution results to provide some robustness against bandwidth snooping. Practical applications of smoothing procedures may indeed involve several bandwidth choices, exposing practitioners to the risk of biasing inference by selecting a bandwidth that agrees with prior beliefs. As a result, our asymptotic theory considers uniformity in both quantile level and bandwidth, allowing for data-driven choices for the latter. This is in sharp contrast with related works on smoothing, such as Horowitz (1998) and Kaplan and Sun (2017). In particular, they restrict attention to a given quantile level and to deterministic bandwidths when studying higher-order accuracy for confidence interval coverage and Type I test errors, respectively.

Our main contributions are as follows. First, smoothing the QR objective function induces bias in the QR estimation. We derive the order of such bias, showing that it is negligible with respect to $\sqrt{n}$ for the optimal bandwidth rate, as well as for the data-driven bandwidth we use in the Monte Carlo experiments. Second, we establish that the smoothed QR estimator has a smaller Bahadur–Kiefer linearization error than the standard QR estimator as long as the bandwidth remains large enough. Third, we find that, for a proper bandwidth choice, our smoothed QR estimator yields a lower asymptotic mean squared error (AMSE) than the standard estimator. Moreover, we derive the optimal bandwidth for our smoothed QR estimator. Fourth, we provide a covariance matrix estimator that captures the variance improvement term. Fifth, we provide a functional central limit theorem analogous to Koenker and Machado's (1999) that holds uniformly with respect to the bandwidth. Sixth, we show that our smoothed QR slope estimator is asymptotically differentiable. We exploit this feature to come up with conditional quantile density function (qdf) and probability density function (pdf) estimators that do not suffer from the curse of dimensionality under correct QR specification. In particular, this allows us to propose an efficient QR estimator using sample splitting weights based on the qdf estimator.

Our simulations examine confidence intervals using a random bandwidth. Although it is as usual hard to estimate it due to the derivative of the conditional pdf that appears in the bias term, we show how to circumvent that by resorting to Silverman's (1986) rule-of-thumb bandwidth for the standard QR residuals. The simulation experiment also illustrates the estimation of the qdf and pdf in presence of a reasonably large number of explanatory variables.

We are obviously not the first to sail in the smoothing 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. Xiang (1994) and Ralescu (1997) provide Bahadur–Kiefer representations for smoothed quantile estimators. Azzalini (1981) shows that the former smoothed estimator 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. For discussions on the relative deficiency of sample

quantiles with respect to kernel-based quantile estimators, see also Falk (1984) and Kozek (2005). Although the above quantile estimators are all continuous, only a few of them consider covariates in a quantile regression fashion. Volgushev, Chao, and Cheng (2019) allow for covariates, ensuring quantile-level differentiability using spline smoothing.

The asymptotic covariance matrix of quantile estimators involves an unknown pdf (Koenker 2005), making inference a bit harder.[1] Smoothing the objective function helps in this dimension for it allows estimating the asymptotic covariance matrix using a standard sandwich formula (Kaplan and Sun 2017). In turn, we focus more on the implications of smoothing the objective function, notably in terms of smoothness of the QR estimator, and on quantile level and bandwidth uniformity.

There is notwithstanding an intimate connection between the above papers on inference and ours. Although the smoothed objective function of Horowitz (1998) differs from ours, his theoretical arguments easily extends to our framework. In addition, the first-order conditions that Whang (2006) employs to derive his empirical likelihood QR estimator is identical to the ones our smoothed QR estimator solves. Interestingly, the same applies to Kaplan and Sun's (2017) smoothed estimating equations (SEE) if restricting attention to QR without instrumental variables. As us, they argue that their smoothing approach can improve on Horowitz's (1998) smoothed QR estimator. See also Kaido and Wuthrich (2018) for numerical algorithms that connect standard SEE and QR estimations.

Finally, our article also relates with the literature on qdf estimation. Parzen (1979) is the first to notice the important role that the quantile density plays. See Guerre and Sabbah (2012) for a recent application in the econometrics of auctions, as well as Roth and Wied (2018) and references therein for related estimation methods.[2] We propose a novel efficient QR estimator based on a first-step qdf estimation, whose second-order behavior avoids the curse of dimensionality, as opposed to the estimation procedures of Newey and Powell (1990), Zhao (2001), Otsu (2008), and Komunjer and Vuong (2010).

The remainder of this article proceeds as follows. Section 2 introduces our smoothed QR estimator. Section 3 describes the main assumptions, the estimation procedure, the asymptotic covariance matrix estimator, and main results. 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 the proofs of our main results, whereas we relegate the proofs of the intermediary results to an online appendix. The latter also includes an in-depth comparison with Horowitz's (1998) smoothing approach.

2. SMOOTHED QR ESTIMATION AND EXTENSIONS

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

---

[1] For instance, Hall and Sheather (1988) study higher-order accuracy of confidence intervals. See also Portnoy (2012) and references therein.

[2] In particular, Roth and Wied (2018) develop a qdf estimator based on the estimation of the second derivative of the population QR objective function.

Y given the covariate $X = x$ is such that

$$Q(\tau \mid x) = x'\beta(\tau), \quad \tau \in (0, 1), \qquad (1)$$

with $Q(\tau \mid x) := \inf\left\{q : F(q \mid x) \geq \tau\right\}$ and $F(\cdot \mid x)$ denoting the conditional cdf of Y given $X = x$, with density $f(\cdot \mid x)$. We henceforth assume the correct specification of the QR model in (1).

Koenker and Bassett (1978) define the (population) objective function of the quantile regression as

$$R(b; \tau) := \mathbb{E}\left[\rho_\tau\left(e(b)\right)\right] = \int \rho_\tau(t)\, dF(t; b), \qquad (2)$$

where $e(b) := Y - X'b$, $F(t; b) = \Pr\left[e(b) \leq t\right]$, and $\rho_\tau(u) := u\left[\tau - \mathbb{I}(u < 0)\right]$ is the usual check function with $\mathbb{I}(A)$ 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 $\widehat{\beta}(\tau)$ minimizes the sample analog based on the empirical distribution, namely,

$$\widehat{R}(b; \tau) := \frac{1}{n} \sum_{i=1}^{n} \rho_\tau\left(e_i(b)\right) = \int \rho_\tau(t)\, d\widehat{F}(t; b), \qquad (3)$$

where $e_i(b) := Y_i - X_i'b$, and $\widehat{F}(\cdot; b)$ 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 $\widehat{f}_h(v; b) := \frac{1}{n} \sum_{i=1}^{n} k_h\left(v - e_i(b)\right)$ and $\widehat{F}_h(t; b) := \int_{-\infty}^{t} \widehat{f}_h(v; b)\, dv$, respectively. We then apply kernel smoothing to the empirical objective function in (3), yielding

$$\widehat{R}_h(b; \tau) := \int \rho_\tau(t)\, d\widehat{F}_h(t; b) = \int \rho_\tau(t) \widehat{f}_h(t; b)\, dt. \qquad (4)$$

Accordingly, the resulting smoothed QR estimator is

$$\widehat{\beta}_h(\tau) := \arg\min_{b \in \mathbb{R}^d} \widehat{R}_h(b; \tau). \qquad (5)$$

By smoothing the integrating measure of the objective function, we ensure that the mapping $b \mapsto \widehat{R}_h(b; \tau)$ is twice continuously differentiable, in opposition to the nonsmoothness of the standard objective function $b \mapsto \widehat{R}(b; \tau)$. The differentiability of the objective function is very convenient for at least two reasons. First, the smoothness of the objective function entails the regularity of the resulting QR estimator. Second, differentiability allows us to estimate the asymptotic covariance matrix of our quantile slope coefficient estimates in a canonical manner (see, e.g., Newey and McFadden 1994).

Finally, observe that

$$\widehat{R}_h(b; \tau) = (1 - \tau) \int_{-\infty}^{0} \widehat{F}_h(v; b)\, dv + \tau \int_{0}^{\infty} \left(1 - \widehat{F}_h(v; b)\right) dv,$$

so that the first- and second-order derivatives of $\widehat{R}_h(b; \tau)$ with respect to b are, respectively,

$$\widehat{R}_h^{(1)}(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} X_i \left[K\left(-\frac{e_i(b)}{h} \right) - \tau \right] \text{ and}$$

$$\widehat{R}_h^{(2)}(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} X_i X_i' k_h\left(-e_i(b) \right),$$

where $K(t) := \int_{-\infty}^{t} k(v)\, dv$.

2.1. Asymptotic Inference

The availability of the second-order derivative $\widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau); \tau\right)$ allows us to propose asymptotic confidence intervals for $\beta(\tau)$ in a standard fashion. We show in Section 3.4 that $\sqrt{n}\left(\widehat{\beta}_h(\tau) - \beta(\tau)\right)$ weakly converges to a Gaussian distribution with mean zero and covariance matrix $\Sigma(\tau) := D^{-1}(\tau)V(\tau)D^{-1}(\tau)$, where $V(\tau) := \tau(1 - \tau)\mathbb{E}(XX')$ and

$$D(\tau) := R^{(2)}\left(\beta(\tau); \tau\right) = \mathbb{E}\left[XX'f\left(X'\beta(\tau) \mid X\right)\right]$$

is the Hessian of the objective function evaluated at the true parameter. In addition, we show that

$$\widehat{\Sigma}_h(\tau) := \widehat{D}_h^{-1}(\tau)\widehat{V}_h(\tau)\widehat{D}_h^{-1}(\tau),$$

with $\widehat{D}_h(\tau) := \widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau); \tau\right)$ and

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

is a consistent estimator of $\Sigma(\tau)$.[3] This means that we may compute a $(1 - \alpha)$−confidence interval for the kth QR coefficient $\beta_k(\tau)$ as

$$\text{CI}_{1-\alpha}\left(\beta_k(\tau)\right) := \widehat{\beta}_{k,h}(\tau) \pm \frac{z_{\alpha/2}\, \widehat{\sigma}_{k,h}(\tau)}{\sqrt{n}},$$

where $\widehat{\sigma}_{k,h}(\tau)$ is the square root of the kth diagonal entry of $\widehat{\Sigma}_h(\tau)$, $\widehat{\beta}_{k,h}(\tau)$ is the kth element of the smoothed QR estimator, and z_α is the α quantile of a standard Gaussian distribution.

2.2. Smoothness of $\widehat{\beta}_h(\cdot)$ and Extensions

Given that $\widehat{\beta}_h(\tau)$ satisfies the first-order condition $\widehat{R}_h^{(1)}\left(\widehat{\beta}_h(\tau); \tau\right) = 0$, it follows from the implicit function theorem that $\widehat{\beta}_h(\tau)$ is continuously differentiable with respect to τ with

$$\frac{\partial \widehat{\beta}_h(\tau)}{\partial \tau} = -\left[\widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau); \tau\right)\right]^{-1} \frac{\partial \widehat{R}_h^{(1)}\left(\widehat{\beta}_h(\tau); \tau\right)}{\partial \tau}$$
$$= \left[\widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau); \tau\right)\right]^{-1} \bar{X} =: \widehat{\beta}_h^{(1)}(\tau) \qquad (6)$$

---

[3] Alternatively, one could simply employ the sample analog of $\tau(1-\tau)\mathbb{E}[XX']$ to estimate $V(\tau)$, which would lead to Powell's (1991) estimator of the asymptotic covariance of the standard QR estimator. See also Angrist, Chernozhukov, and Fernández-Val (2006) and Kato (2012). However, we expect that our variance estimator to entail better finite-sample properties and, accordingly, shorter confidence intervals.
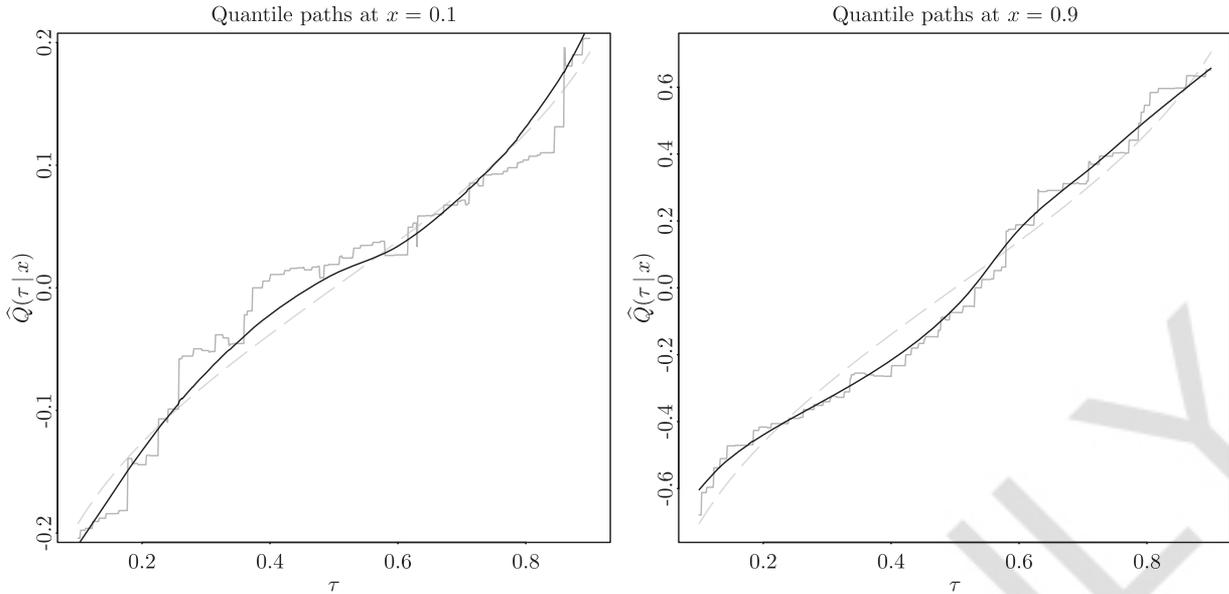
Figure 1. Examples of quantile paths implied by the standard and smoothed QR estimates. We consider a sample of 100 observations from $Y = (0.1+0.5X)\epsilon$, where ϵ is a standard Gaussian error independent of $X \sim U(0,1)$. The left and right panels compare the conditional quantiles of Y, respectively, given $X = 0.1$ and $X = 0.9$ (dashed line) with their corresponding standard (gray) and convolution-type kernel (black) QR estimates. We employ the rule-of-thumb bandwidth h_{ROT} for the smoothed QR estimator as in Section 4.

as long as the Hessian has an inverse, which holds with a probability tending to 1. Because the Hessian matrix is asymptotically positive, it also follows that the QR function evaluated at the average covariate, $\bar{X}'\widehat{\beta}_h(\tau)$, is strictly increasing in τ.

Figure 1 compares two different paths of our smoothed QR estimator with the ones of the standard QR estimator. The quantile paths implied by the convolution-type kernel QR estimates are not only much smoother but much less variable than the conditional quantiles based on the standard QR estimates. Theorem 2 indeed shows that our smoothed QR estimator is continuous with a probability tending to one, while the standard QR estimator is a step function. Theorem 3 also establishes that smoothing reduces the variance of the QR estimator. Even though Figure 1 evaluates the conditional quantiles at $X = 0.1$ and $X = 0.9$, rather than at $\bar{X}$, the quantile paths based on our smoothed QR estimator look increasing such that there is no need to apply Chernozhukov, Fernandez-Val, and Galichon's (2010) rearrangement procedure. This is most likely due to the smoothness of the convolution-type kernel estimator, inasmuch as the nonmonotonicity of the standard QR estimator is mostly due to small jumps.

We next consider two extensions based on the qdf given by

$$q(\tau|x) = \frac{\partial Q(\tau|x)}{\partial \tau} = \frac{1}{f(Q(\tau|x)|x)}.$$

As noted by Guerre and Sabbah (2017), the curve $\tau \mapsto (Q(\tau|x), 1/q(\tau|x)) =: \mathtt{f}(\tau|x)$ is the graph $y \mapsto (y, f(y|x))$ of the conditional pdf $f(\cdot|x)$. The qdf plays a major role in first-price auctions (Guerre and Sabbah 2012) and in semiparametric efficient QR estimation (see, among others, Newey and Powell 1990; Zhao 2001). In particular, the efficient QR slope estimator $\tilde{b}_q(\tau) = \arg\min_b \tilde{R}_q(b;\tau)$ is infeasible because

$$\tilde{R}_q(b;\tau) = \frac{1}{n}\sum_{i=1}^{n} \frac{\rho_\tau(Y_i - X_i'b)}{q(\tau|X_i)} \tag{7}$$

depends on the conditional qdf. Newey and Powell (1990) and Zhao (2001) propose to estimate the conditional qdf using kernel methods. However, their estimators are bound to perform poorly if the covariate dimension is large due to the curse of dimensionality. The same drawback also applies to the efficient estimators put forth by Otsu (2008) and Komunjer and Vuong (2010).

2.2.1. pdf Curve Estimation. The linear QR model dictates that $q(\tau|x) = x'\beta^{(1)}(\tau)$, which we may estimate using $\widehat{\beta}_h^{(1)}(\tau) = \left[\widehat{R}_h^{(2)}(\widehat{\beta}_h(\tau);\tau)\right]^{-1}\bar{X}$ in (6). The resulting conditional qdf estimator, $\widehat{q}_h(\tau|x) := x'\widehat{\beta}_h^{(1)}(\tau)$, does not suffer from the curse of dimensionality because it makes use of the linear QR structure. The pdf curve estimator

$$\widehat{\mathtt{f}}_h(\tau|x) := \left(x'\widehat{\beta}_h(\tau), \frac{1}{x'\widehat{\beta}_h^{(1)}(\tau)}\right) \tag{8}$$

converges to $\mathtt{f}(\tau|x)$ at the rate of univariate kernel density estimator. If $\widehat{q}_h(\cdot|x) > 0$, the pdf estimator is positive and integrates to one given that $\int_0^1 1/\widehat{q}_h(\tau|x)\, d\widehat{Q}_h(\tau|x) = \int_0^1 d\tau = 1$. This means that the pdf curve estimator is strict. Figure 2 depicts an example of pdf curve estimation with three covariates using a sample of 200 observations. It also exhibits the QR and qdf estimates we use for computing the pdf estimator. The pdf curve estimator seems to capture reasonably well the underlying density despite the high dimensionality of the estimation problem relative to the sample size.

2.2.2. Semiparametric Efficiency. As in Newey and Powell (1990) and Zhao (2001), we resort to sample splitting to obtain efficient QR slope estimators. To this end, we minimize a feasible counterpart of the objective function $\tilde{R}_q(b;\tau)$ in (7), namely,
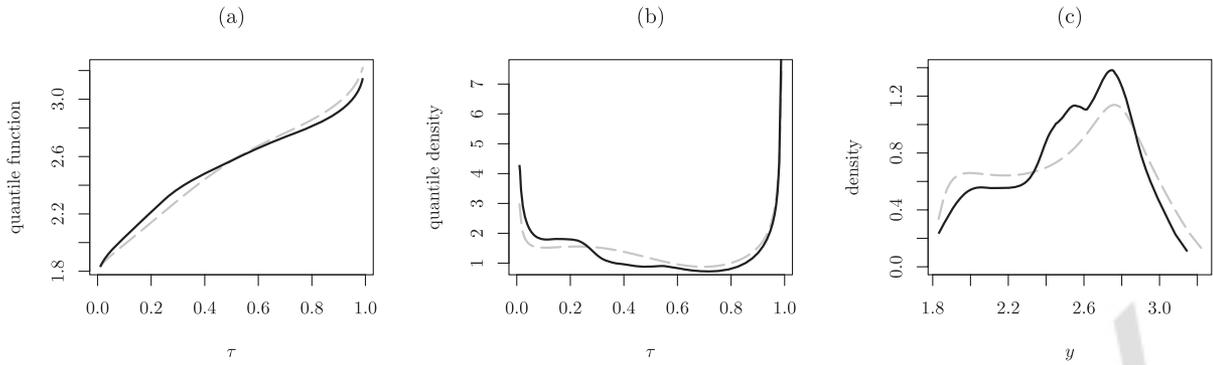
Figure 2. Conditional quantile function, qdf, and pdf estimation examples. We sample 200 observations from $Q(\tau|x) = \beta_0(U) + \beta_1(U)X_1 + \beta_2(U)X_2 + \beta_3(U)X_3$, where U, X_1, X_2, and X_3 are independent and uniformly distributed variables on the unit interval $[0, 1]$, $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are, respectively, given by the Beta$(1, 16)$ and Beta$(32, 32)$ quantile functions, $\beta_2(\tau) = 1$, and $\beta_3(\tau) = \frac{(2\pi+8)\tau-(\cos(2\pi\tau)-1)}{2\pi+8}$. Panels (a) to (c) are about the conditional quantile function, qdf, and pdf of Y given $x_1 = 0.5$, $x_2 = x_3 = 0.9$, respectively. True functions are in gray dashed lines, whereas black solid lines correspond to estimates using h_{ROT} as bandwidth.

$$\check{R}_q(b; \tau) = \frac{1}{n-m} \sum_{i=m+1}^{n} \frac{\rho_\tau\left(Y_i - X_i'b\right)}{\check{q}_h(\tau|X_i)}, \qquad (9)$$

where $\check{q}_h(\tau|x) = x'\check{\beta}_h^{(1)}(\tau)$ considers only the first $m < n$ observations. One could alternatively use either a smoothed version of the check function $\rho_\tau(\cdot)$ or a version of Newey and Powell's (1990) one-step estimator using the weights $\check{q}_h(\tau|X_i)$.

Theoretical results for pdf graph and for the efficient QR in (9) readily follow from the asymptotic theory we establish for the smoothed QR estimator. We state them in Section 3.5.

2.3. Alternative Smoothed Objective Function

In the absence of covariates, the first-order condition $\widehat{R}_h^{(1)}\big(\widehat{\beta}_h(\tau); \tau\big) = 0$ that results in our smoothed QR estimator coincides with the first-order condition in Nadaraya (1964), namely, $\widehat{F}_h\big(\widehat{\beta}_h(\tau)\big) = \tau$. This confirms that the smoothing we apply to the objective function boils down to estimating the conditional cumulative distribution function using a kernel approach. In turn, the smoothed objective function $\widehat{\mathfrak{R}}_h(b; \tau)$ put forth by Horowitz (1998) replaces the indicator in the check function by a kernel counterpart: $\widehat{\mathfrak{R}}_h(b; \tau) := \frac{1}{n} \sum_{i=1}^{n} e_i(b)\Big[\tau - K\big(-e_i(b)/h\big)\Big]$. As noted by Kaplan and Sun (2017), the first-order derivative of this objective function is

$$\widehat{\mathfrak{R}}_h^{(1)}(b; \tau) = \frac{1}{n} \sum_{i=1}^{n} X_i \left[K\left(-\frac{e_i(b)}{h}\right) - \tau - k\left(-\frac{e_i(b)}{h}\right) \frac{e_i(b)}{h} \right]$$

$$= \widehat{R}_h^{(1)}(b; \tau) - \frac{1}{n} \sum_{i=1}^{n} X_i k\left(-\frac{e_i(b)}{h}\right) \frac{e_i(b)}{h}. \qquad (10)$$

The resulting first-order condition differs from the one we obtain because of the additional term in (10). We show in the online appendix that, as a consequence, Horowitz's (1998) smoothed QR estimator exhibits larger bias and variance than ours.

3. ASYMPTOTIC THEORY

We start with some notation. Let $\|\cdot\|$ denote the Euclidean norm of a matrix, namely, $\|A\| = \sqrt{\operatorname{tr}(AA')}$. We denote by $f(\cdot \mid x)$ the conditional probability density function of Y given $X = x$, with jth partial derivative given by $f^{(j)}(y \mid x) := \frac{\partial^j}{\partial y^j} f(y \mid x)$. Similarly, let $q(\tau \mid x) := \frac{\partial}{\partial \tau} Q(\tau \mid x)$. 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 this session with some inference implications.

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, that is, the support of X is a compact subset of $\bar{\mathbb{R}}_{+*}^d$. The matrix $\mathbb{E}[XX']$ is full rank.

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

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

Q2 There exists an integer $s \geq 1$ such that the derivative $f^{(s)}(\cdot \mid \cdot)$ is uniformly continuous in the sense that $\lim_{\epsilon \to 0} \sup_{(x,y) \in \mathbb{R}^{d+1}} \sup_{t: |t| \leq \epsilon} \left| f^{(s)}(y + t \mid x) - f^{(s)}(y \mid x) \right| = 0$, as well as such that $\sup_{(x,y) \in \mathbb{R}^{d+1}} \left| f^{(j)}(y \mid x) \right| < \infty$ and $\lim_{y \to \pm\infty} f^{(j)}(y \mid x) = 0$ for all $j = 0, \ldots, s$.

Assumption K. The kernel function k and bandwidth h satisfy

K1 The kernel $k : \mathbb{R} \to \mathbb{R}$ is even, integrable, twice differentiable with bounded first and second derivatives, and such that $\int k(z) \, dz = 1$ and $0 < \int_0^\infty K(z) \left[1 - K(z)\right] dz < \infty$. In addition, for s as in Assumption Q2, $\int \left| z^{s+1} k(z) \right| dz < \infty$, and k is orthogonal to all nonconstant monomials of degree up to s, that is, $\int z^j k(z) \, dz = 0$ for $j = 1, \dots, s$, and $\int z^{s+1} k(z) \, dz \neq 0$.

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

Some remarks are in order. First, observe that $R^{(2)}(b; \tau) = \mathbb{E}\big[XX'f(X'b \,|\, X)\big]$ is positive definite for all b and any τ under Assumptions Q1 and X. This means that $D^{-1}(\tau)$ exists for every τ. 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 τ given that $q(\tau \,|\, x) = 1/f\big(Q(\tau \,|\, x) \,|\, x\big)$. 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 $\int_0^\infty K(z) \left[1 - K(z)\right] dz$ is positive will become clear in Theorem 3. It makes sense because it guarantees that our smoothed QR estimator dominates the standard QR estimator in the AMSE sense. Fifth, in view that Assumption K does not preclude high-order kernels, $\widehat{f}_h(\cdot; b)$ is not necessarily a density, even if it is a consistent estimator of the density of the error term $e(b)$. It is nonetheless possible to show $\widehat{R}_h(b; \tau) = \frac{1}{n} \sum_{i=1}^n \rho_\tau * k_h\big(e_i(b)\big)$, 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 that the bandwidth h implicitly depends on the sample size n through its lower and upper limits in Assumption K2. This is paramount because we wish to entertain data-driven bandwidths.

3.2. Bias and Bahadur–Kiefer Representation

In this section, we study the order of the asymptotic bias and obtain a Bahadur–Kiefer representation for the stochastic error of $\widehat{\beta}_h(\tau)$. As for the former, it is indeed expected from popular wisdom that smoothing the QR objective function should induce bias in finite samples. More formally, $\widehat{\beta}_h(\tau)$ actually estimates $\beta_h(\tau) := \arg\min_{b \in \mathbb{R}^d} R_h(b; \tau)$, with $R_h(b; \tau) := \mathbb{E}[\widehat{R}_h(b; \tau)]$. This yields $\beta_h(\tau) - \beta(\tau)$ as the bias term and $\widehat{\beta}_h(\tau) - \beta_h(\tau)$ as the stochastic error of our smoothed QR estimator. Our first result shows that the smoothing bias shrinks to zero as the sample size grows.

Theorem 1. For $\bar{h}_n$ small enough, $\beta_h(\tau)$ is unique under Assumptions X, Q, and K for every $\tau \in [\underline{\tau}, \bar{\tau}]$ and such that $\beta_h(\tau) = \beta(\tau) - h^{s+1} B(\tau) + o(h^{s+1})$ uniformly over $(\tau, h) \in [\underline{\tau}, \bar{\tau}] \times [\underline{h}_n, \bar{h}_n]$, with $B(\tau) = \frac{\int z^{s+1} k(z) \, dz}{(s+1)!} D^{-1}(\tau) \, \mathbb{E}\big[Xf^{(s)}\big(X'\beta(\tau) \,|\, X\big)\big]$.

It is interesting to discuss the implications of Theorem 1 to the particular case of a standard linear regression model $Y_i = X_i'\beta + \varepsilon_i$ with iid errors independent of the covariates. Let $X_i = (1, \tilde{X}_i)'$, with $\tilde{X}_i \in \mathbb{R}$, so that the conditional pdf of Y_i given $X_i = x$ is $f(y|x) = f_\varepsilon\big(y - x'\beta(\tau)\big)$. It then follows that $D(\tau) =$ $f_\varepsilon(0)\mathbb{E}[XX']$ for all τ and $B(\tau) \propto \mathbb{E}^{-1}[XX']\mathbb{E}[X] = (1, 0)'$, so that the first term in the bias appears only for the intercept. More generally, Theorem 1 settles the issue of possible side effects of smoothing in that $\beta_h(\tau)$ eventually becomes uniformly close to the true parameter $\beta(\tau)$.

The next result derives some convenient expansions for the stochastic error $\widehat{\beta}_h(\tau) - \beta_h(\tau)$. For this purpose, let $\widehat{S}_h(\tau) := \widehat{R}_h^{(1)}\big(\beta_h(\tau); \tau\big)$ and $D_h(\tau) := R_h^{(2)}\big(\beta_h(\tau); \tau\big)$. Note that the first-order condition $R_h^{(1)}\big(\beta_h(\tau); \tau\big) = 0$ implies that the score term $\widehat{S}_h(\tau)$ has zero mean, and hence, the stochastic error in the Bahadur–Kiefer representation (11) is asymptotically centered.

Theorem 2. Under Assumptions X, Q, and K, $\widehat{\beta}_h(\cdot)$ is unique and continuous over $(\tau, h) \in [\underline{\tau}, \bar{\tau}] \times [\underline{h}_n, \bar{h}_n]$ with probability tending to one, satisfying the following two representations

$$\sqrt{n}\left(\widehat{\beta}_h(\tau) - \beta_h(\tau)\right) = -\sqrt{n} D_h^{-1}(\tau)\, \widehat{S}_h(\tau) + O_p(\varrho_n(h)) \tag{11}$$

$$= -\sqrt{n} \widehat{D}_h^{-1}(\tau)\widehat{S}_h(\tau) + O_p(1/\sqrt{n}) \tag{12}$$

with $\varrho_n(h) = \sqrt{\ln n/(nh)}$ and both remainder terms uniform with respect to $(\tau, h) \in [\underline{\tau}, \bar{\tau}] \times [\underline{h}_n, \bar{h}_n]$.

Unlike the standard QR objective function, $\widehat{R}_h(\cdot; \tau)$ is not necessarily convex because higher-order kernels may take negative values. We address this issue by first showing that $\widehat{\beta}_h(\tau)$ is close to $\beta_h(\tau)$, uniformly in τ and h, and then proving that the smoothed objective function $\widehat{R}_h(\cdot; \tau)$ is asymptotically strictly convex in the vicinity of $\beta_h(\tau)$. This follows from the convergence of $\widehat{R}_h^{(2)}(b; \tau)$ to $R^{(2)}(b; \tau)$, uniformly for b in any compact set and also in τ and h, as established using a powerful concentration inequality from Massart (2007). The first-order condition $\widehat{R}_h\big(\widehat{\beta}_h(\tau); \tau\big) = 0$ implies the following integral representation

$$\widehat{\beta}_h(\tau) - \beta_h(\tau)$$
$$= -\left[\int_0^1 \widehat{R}_h^{(2)}\big(\beta_h(\tau) + u[\widehat{\beta}_h(\tau) - \beta_h(\tau)]; \tau\big) \, du\right]^{-1} \widehat{S}_h^{(1)}(\tau)$$

for the stochastic error. Deriving the order of the score uniformly in τ and h via a concentration inequality entails the representations (11) and (12). One could also obtain higher-order expansions for the stochastic error along the same lines by establishing uniform convergence of higher-order derivatives of the smooth objective function through concentration inequalities.

The Bahadur–Kiefer representation in (11) shows that $\widehat{\beta}_h(\tau)$ is, in a sense, more linear than the standard QR estimator $\widehat{\beta}(\tau)$. Knight (2001) and Jurečková, Sen, and Picek (2012) show that, in many cases of interest, the Bahadur–Kiefer representation for the standard QR estimator is given by $\sqrt{n}\big(\widehat{\beta}(\tau) - \beta(\tau)\big) = -\sqrt{n}\, D^{-1}(\tau)\widehat{S}(\tau) + O_p(n^{-1/4})$, with $\widehat{S}(\tau) = \widehat{R}^{(1)}\big(\beta(\tau); \tau\big)$. In contrast, the remainder term in (11) is of order nearly $O_p(n^{-1/2})$ for proper bandwidth choices if centering around $\beta_h(\tau)$. The asymptotically negligible bias given by the difference between $\beta_h(\tau)$ and $\beta(\tau)$ is the price we pay for improving the rate of the Bahadur–Kiefer representation.

The nonlinear approximation in (12) obtains a smaller remainder term of order $n^{-1/2}$ by replacing the deterministic

standardization $D_h^{-1}(\tau)$ in (11) with $\widehat{D}_h^{-1}(\tau)$. This slightly improves the rate of the remainder term in Portnoy's (2012) approximation. One could also obtain higher-order approximations involving quadratic terms under stronger bandwidth rate conditions. As in Horowitz (1998), this would however exclude bandwidths of the optimal order $n^{-1/5}$ for a kernel of order $s + 1 = 2$, and hence, we do not follow such alternative. Lemma 4 in the Appendix shows that (12) also holds for $\widehat{R}^{(2)}(\beta_h(\tau); \tau)$ in lieu of $\widehat{D}_h(\tau)$. Using the standardization $\widehat{D}_h(\tau)$ instead of the infeasible $\widehat{R}^{(2)}(\beta_h(\tau); \tau)$ is important for practical purposes, such as computing Wald statistics.

It also follows from Theorems 1 and 2 that $\widehat{\beta}_h(\cdot)$ offers a fair global picture of $\beta(\cdot)$ in that

$$\left\| \widehat{\beta}_h(\tau) - \beta(\tau) \right\| = O_p\left(\frac{1}{\sqrt{n}} + h^{s+1} \right), \qquad (13)$$

uniformly for $\tau \in [\underline{\tau}, \bar{\tau}]$ and $h \in [\underline{h}_n, \bar{h}_n]$. Accordingly, the remainder in (13) is of order $O_p(n^{-1/2})$ for any $h \leq O\left(n^{-1/(2(s+1))}\right)$. The latter restriction is quite light, holding not only for the infeasible AMSE-optimal bandwidth in Theorem 4 but also for the corresponding rule-of-thumb bandwidth that we suggest in Section 3.3.

Lastly, it is important to stress the major role that uniformity plays here. It ensures that, if a random (possibly data-driven) bandwidth process $\{\widehat{h}(\tau); \tau \in [\underline{\tau}, \bar{\tau}]\}$ has sample paths in $[\underline{h}_n, \bar{h}_n]$ with a sufficiently high probability, then (11) remains valid even if we replace h with a data-driven bandwidth $\widehat{h}(\tau)$. The next result states this property in a rigorous manner.

Corollary 1. If $\{\widehat{h}(\tau); \tau \in [\underline{\tau}, \bar{\tau}]\}$ satisfies $\Pr(\widehat{h}(\tau) \in [\underline{h}_n, \bar{h}_n] \text{ for all } \tau) \to 1$, then both (11) and (12) hold with $\widehat{h}(\tau)$ in place of h, uniformly in τ.

The asymptotic theory so far posits that our smoothed QR estimator entails a better Bahadur–Kiefer representation than 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 characterizing the AMSE of our convolution-type kernel QR estimator as well as the bandwidth choice that minimizes it.

3.3. Asymptotic Mean Squared Error

The asymptotic covariance matrix of $\widehat{\beta}_h(\tau)$ comes from the leading term of its Bahadur–Kiefer linear representation in (11). The next result not only characterizes this asymptotic covariance matrix but also shows that it is smaller than the asymptotic covariance matrix of $\widehat{\beta}(\tau)$. In what follows, let $\Sigma_h(\tau) := \mathbb{V}\left(\sqrt{n} D_h^{-1}(\tau) \widehat{S}_h(\tau) \right)$.

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

$$\Sigma_h(\tau) = \Sigma(\tau) - c_k h D^{-1}(\tau) + o(h), \qquad (14)$$

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

Theorem 3 shows that the asymptotic covariance matrix of the smoothed QR estimator is equal to the asymptotic

Table 1. Examples of Gaussian-type kernels

Gaussian-type kernel	Kernel constant	Order $(s + 1)$
$k(x) = \phi(x)$	$c_k = \frac{1}{\sqrt{\pi}}$	2
$k(x) = \frac{3}{2}\left(1 - \frac{x^2}{3}\right)\phi(x)$	$c_k = \frac{7}{16\sqrt{\pi}}$	4
$k(x) = \frac{15}{8}\left(1 - \frac{2x^2}{3} + \frac{x^4}{15}\right)\phi(x)$	$c_k = \frac{321}{1024\sqrt{\pi}}$	6
$k(x) = \frac{35}{16}\left(1 - x^2 + \frac{x^4}{5} - \frac{x^6}{105}\right)\phi(x)$	$c_k = \frac{4175}{16384\sqrt{\pi}}$	8

NOTE: We provide examples of kernel functions that satisfy Assumption K1 and their corresponding kernel constants c_k. We denote by ϕ the standard Gaussian density.

covariance matrix of the standard QR estimator minus a term $c_k h D^{-1}(\tau)$ induced by smoothing. Table 1 reports the values of c_k for Gaussian-type kernels that satisfy Assumption K1. The order $\ln n/(nh)$ of the squared Bahadur–Kiefer remainder term is negligible with respect of the order h of the smoothing-related term given that $1/h = o(\sqrt{n/\ln n})$ by Assumption K2.

It is sometimes possible to refine the asymptotic covariance matrix expansion in Theorem 3 using strong approximation tools. For instance, assume there is no covariate and that Y_i is uniform over $[0, 1]$. In this case, there is no bias in that $\beta_h(\tau) = \tau$ and $D_h(\tau) = D(\tau) = 1$ for h small enough. The score function then reads

$$\widehat{S}_h(\tau) = \int \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\mathbb{I}(Y_i \leq \tau + ht) - (\tau + ht) \right] k(t) \, dt,$$

using $\int k(t) \, dt = 1$ and $\int t k(t) \, dt = 0$. The Komlós–Major–Tusnády strong approximation result shows that one may reconstruct the sample jointly with a sequence of Brownian bridges $B_n(\cdot)$ such that $\sup_{y \in [0,1]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\mathbb{I}(Y_i \leq \tau + ht) - (\tau + ht) \right] - B_n(t) \right| = O_p\left(\frac{\ln n}{\sqrt{n}} \right)$; see Pollard (2002). It follows that, if k has a compact support,

$$\widehat{S}_h(\tau) = \int B_n(\tau + ht) k(t) \, dt + O_p\left(\frac{\ln n}{\sqrt{n}} \right) \qquad (15)$$

$$= B_n(\tau) + \int \left[B_n(\tau + ht) - B_n(\tau) \right] k(t) \, dt + O_p\left(\frac{\ln n}{\sqrt{n}} \right).$$

As the integral term that comes from smoothing vanishes for $h = 0$, $B_n(\tau)$ gives the limit distribution of the standard QR estimator.

The covariance between $B_n(\tau)$ and $B_n(\tau + ht) - B_n(\tau)$ is negative of order h. This implies a negative covariance between $B_n(\tau)$ and the integral term in (15). As the variance of the latter is of smaller order than this negative covariance, the asymptotic variance of $\widehat{S}_h(\tau)$ is smaller than $\mathbb{V}(B_n(\tau)) = \tau(1 - \tau)$ by a term of order h. Theorem 3 also implies that the distribution of $D_h(\tau)^{-1}\widehat{S}_h(\tau)$ is a centered normal with variance $\tau(1 - \tau) - c_k h$ up to the order $\ln n/\sqrt{n}$. Because the second term in (14) is negative, the asymptotic variance of $\widehat{\beta}_h(\tau)$ is indeed smaller than the asymptotic variance of $\widehat{\beta}(\tau)$.

We next focus on obtaining the bandwidth h_λ^* that minimizes the asymptotic mean squared error of $\lambda' \widehat{\beta}_h(\tau)$ for a given $\lambda \in \mathbb{R}^d$: $\text{AMSE}\left(\lambda' \widehat{\beta}_h(\tau) \right) =$

$\mathbb{E}\left[\lambda'\left(\beta_h(\tau) - D_h^{-1}(\tau)\widehat{S}_h(\tau) - \beta(\tau)\right)\right]^2$. This approximates the mean squared error $\mathrm{MSE}\left(\lambda'\widehat{\beta}_h(\tau)\right) = \mathbb{E}\left[\lambda'\left(\widehat{\beta}_h(\tau) - \beta(\tau)\right)\right]^2$ by essentially ignoring the asymptotically negligible remainder term of the Bahadur–Kiefer representation. To do so, we require that the bandwidth is such that $1/h = o\left(\sqrt{n/\ln n}\right)$; see Theorems 2 and 3.

Theorem 4. Let Assumptions X, Q, and K hold. If $\lambda'B(\tau) \neq 0$, and the conditional density $f(\cdot\,|\,x)$ is s-times continuously differentiable for all x, then $\mathrm{AMSE}\left(\lambda'\widehat{\beta}_h(\tau)\right)$ is minimal for

$$h_\lambda^* = \left(\frac{c_k\,\lambda'D^{-1}(\tau)\lambda}{2n(s+1)\left(\lambda'B(\tau)\right)^2}\right)^{\frac{1}{2s+1}}, \tag{16}$$

and equal to $\mathrm{AMSE}\left(\lambda'\widehat{\beta}_{h_\lambda^*}(\tau)\right) = \frac{1}{n}\lambda'\left[\Sigma(\tau) - c_k\,h_\lambda^*\,\frac{2s+1}{2s+2}\,D^{-1}(\tau)\right]\lambda + o(h_\lambda^*/n)$.

The optimal bandwidth we derive in Theorem 4 obviously depends on λ. This is not the case of Kaplan and Sun's (2017) bandwidth choice, which minimizes $\mathbb{E}\left[\left\|\widehat{\beta}(\tau) - \beta(\tau)\right\|^2\right]$. Although their optimal bandwidth has the same order of h_λ^* in (16), it obviously does not depend on λ given that they focus on the global estimation of $\beta(\tau)$. As such, it is suboptimal for a linear combination of the elements of $\beta(\tau)$, as we consider here. This matters for instance in the particular case of a standard linear regression in view that the optimal bandwidth is well defined only for the intercept as the leading term of the bias vanishes for the slope coefficients.

It is easy to appreciate that Theorem 4 remains valid for any bandwidth $h_\lambda = [1 + o(1)]h_\lambda^*$. Although entertaining a plug-in bandwidth based on the nonparametric estimation of $D(\tau)$ and $B(\tau)$ is certainly feasible, it is perhaps not very advisable due to the presence of the pdf derivative in the bias term. We circumvent this issue by means of a rule-of-thumb approach. In particular, it follows from the objective function in (4) that it is as if our convolution-kernel QR estimator were using a kernel-based estimator of the cdf of the standard QR residuals. This naturally leads to using Silverman's (1986) rule-of-thumb bandwidth based on residual dispersion measures (e.g., sample standard deviation or interquantile range).

3.4. Inference

In this section, we derive a functional limit distribution of the smoothed QR estimator that holds for data-driven bandwidths. We also discuss how to consistently estimate the asymptotic covariance matrix of $\widehat{\beta}_h(\tau)$ to make inference. In particular, we establish not only that $\widehat{\Sigma}_h(\tau) := \widehat{D}_h^{-1}(\tau)\widehat{V}_h(\tau)\widehat{D}_h^{-1}(\tau)$, with $\widehat{D}_h(\tau) := \widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau);\tau\right)$ and

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

is a consistent estimator of $\Sigma(\tau)$, but also the asymptotic normality of our convolution-type kernel QR estimator.

Theorem 5. Let Assumptions X, Q, and K hold.

(G) $\left\{\sqrt{n}\left(\widehat{\beta}_h(\tau) - \beta_h(\tau)\right) : \tau \in [\underline{\tau}, \bar{\tau}]\right\}$ converges in distribution to a centered Gaussian process $\left\{W(\tau) : \tau \in [\underline{\tau}, \bar{\tau}]\right\}$ with covariance structure

$$\mathbb{V}\left(W(\tau), W(\varsigma)\right) = (\tau \wedge \varsigma - \tau\varsigma)D(\tau)^{-1}\mathbb{E}(XX')D(\varsigma)^{-1},$$
$$\tau, \varsigma \in [\underline{\tau}, \bar{\tau}]. \tag{17}$$

(S) Consider a data-driven bandwidth $\widehat{h}_n$ which satisfies $\widehat{h}_n = h_0\left\{1 + o_p(1/\sqrt{h_n\ln n})\right\}$ for some deterministic h_n in $[\underline{h}_n, \bar{h}_n]$. $\left\{\sqrt{n}\left(\widehat{\beta}_{\widehat{h}_n}(\tau) - \beta_{\widehat{h}_n}(\tau)\right) : \tau \in [\underline{\tau}, \bar{\tau}]\right\}$ converges in distribution to $\left\{W(\tau) : \tau \in [\underline{\tau}, \bar{\tau}]\right\}$.

(U) If $\underline{h}_n = o(1/\ln n)$, the convergence result in **(G)** is uniform in that

$$\sup_{h \in [\underline{h}_n, \bar{h}_n]}\left\|\sqrt{n}\left(\widehat{\beta}_h(\tau) - \beta_h(\tau) - \left[\widehat{\beta}_{h_n}(\tau) - \beta_{h_n}(\tau)\right]\right)\right\| = o_p(1)$$

for any arbitrary deterministic sequence h_n in $[\underline{h}_n, \bar{h}_n]$.

(V) $\widehat{\Sigma}_h(\tau) = \Sigma_h(\tau) + O_p\left(\sqrt{\ln n/(nh)}\right) + o(h^s)$ *uniformly for* $(\tau, h) \in [\underline{\tau}, \bar{\tau}] \times [\underline{h}_n, \bar{h}_n]$.

Theorem 5(G) shows that the functional limit distribution of the smoothed and standard QR estimators are identical. Theorem 5(S) posits that such a functional central limit theorem (FCLT) holds for data-driven bandwidths that converge to a deterministic counterpart. Theorem 5(U) states that the FCLT holds uniformly with respect to the bandwidth. Theorem 5(V) implies not only that the confidence intervals of Section 2.1 asymptotically have the desired level, but also that $\widehat{\Sigma}_h(\tau)$ is a consistent estimator of the asymptotic covariance matrix of both standard and smoothed QR estimators. Under Assumption K2, as $s \geq 1$,

$$\widehat{\Sigma}_h(\tau) = \Sigma_h(\tau) + o_p(h) = \Sigma(\tau) - c_k\,h\,D^{-1}(\tau) + o_p(h)$$

by Theorem 3. This means that $\widehat{\Sigma}_h(\tau)$ captures the improvement in precision due to smoothing and, accordingly, the length of a confidence interval based on $\widehat{\Sigma}_h(\tau)$ for any element of $\beta(\tau)$ is asymptotically smaller than the length of a confidence interval based on the naive variance estimator $\tau(1 - \tau)\widehat{D}_h^{-1}(\tau)\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\widehat{D}_h^{-1}(\tau)$.

Note that both (S) and (U) make use of the implicit function theorem applied to the first-order condition $\widehat{R}_h^{(1)}\left(\widehat{\beta}_h(\tau);\tau\right) = 0$ to obtain

$$\frac{\partial\widehat{\beta}_h(\tau)}{\partial h} = -\left[\widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau);\tau\right)\right]^{-1}\frac{\partial\widehat{R}_h^{(1)}\left(\widehat{\beta}_h(\tau);\tau\right)}{\partial h}.$$

We then establish a convergence rate that holds uniformly in h for $\widehat{R}_h^{(2)}\left(\widehat{\beta}_h(\tau);\tau\right)$ and $\frac{\partial}{\partial h}\widehat{R}_h^{(1)}\left(\widehat{\beta}_h(\tau);\tau\right)$, leading to

$$\sqrt{n}\left(\widehat{\beta}_{h_1}(\tau) - \beta_{h_1}(\tau) - \left[\widehat{\beta}_{h_0}(\tau) - \beta_{h_0}(\tau)\right]\right)$$
$$= \sqrt{n}\int_{h_0}^{h_1}\left[\frac{\partial\widehat{\beta}_h(\tau)}{\partial h} - \frac{\partial\beta_h(\tau)}{\partial h}\right]dh$$
$$= \sqrt{n}\,O_p\left(\int_{h_0}^{h_1}\sqrt{\frac{\ln n}{nh}}\,dh\right) = O_p(\sqrt{\ln n})\left|\sqrt{h_1} - \sqrt{h_0}\right|,$$

uniformly for $(h_0, h_1) \in [\underline{h}_n, \bar{h}_n]^2$.

3.5. Extensions

3.5.1. pdf Curve Estimation. The next result states the uniform consistency of the pdf curve estimator. A distinctive feature of the estimation procedure we propose is that consistency holds even when the covariate value x is not in the support of X. The key for this ability to extrapolate lies on the underlying quantile regression specification, which we assume correct.

Proposition 1. Under Assumptions X, Q, and K, $\|\widehat{\mathtt{f}}_h(\tau|x) - \mathtt{f}(\tau|x)\| = o(h^s) + O_p\big(\sqrt{\ln n/(nh)}\big)$ uniformly for x in any compact set, τ in $[\underline{\tau}, \overline{\tau}]$, and h in $[\underline{h}_n, \overline{h}_n]$.

Proposition 1 shows that the dimension of the covariate vector X does not affect the order $\sqrt{\ln n/(nh)}$ of the stochastic estimation error of $\widehat{\mathtt{f}}_h(\tau|x)$, so that the curse of dimensionality does not apply. This result obviously depends on the correct specification of the quantile regression model. The bias in Proposition 1 is of order $o(h^s)$ due to the Hessian bias in Lemma 1. The latter arises instead of the usual $O(h^s)$ bias order because we calibrate the kernel order for the QR estimation, which involves a function with $s + 1$ derivatives. In contrast, it is the estimation of the pdf, which has only s derivatives, that drives the bias in the Hessian term.

As it turns out, the quantile density estimator $\widehat{q}_h(\tau|x)$ drives the nonparametric consistency rate. Alternatively, we could estimate the pdf by $\widehat{f}_h(y|x) = 1/\widehat{q}_h\big(\widehat{F}_h(y|x)|x\big)$, where $\widehat{F}_h(y|x)$ is a conditional cdf estimator. The consistency rate of such $\widehat{f}_h(y|x)$ is the same as in Proposition 1 if $\widehat{F}_h(y|x)$ ensues from inverting $x'\widehat{\beta}_h(\tau)$ for y in its range.

3.5.1.1. Semiparametric Efficiency. We next establish the asymptotic equivalence of the feasible QR estimator $\check{b}_q(\tau)$ in (9) to its unfeasible counterpart $\tilde{b}_q(\tau)$ in (7). In addition, we also document the asymptotic normality and efficiency of both estimators. Before stating the result, recall that Newey and Powell (1990) show that $\Sigma_q(\tau) := \tau(1-\tau)D_q^{-1}(\tau)$, with

$$D_q(\tau) := \mathbb{E}\left[\frac{XX'}{q^2(\tau|X)}\right] = \mathbb{E}\left\{XX'f^2[X'\beta(\tau)|X]\right\},$$

is the asymptotic covariance matrix of the asymptotically efficient QR estimator.

Proposition 2. Let $m = o(n^{1/2})$ and $h = o(1)$, with $1/h = o\big((m/\ln m)^{1/3}\big)$. Assumptions X, Q, and K1 ensure that $\sqrt{n}\big(\check{b}_q(\tau) - \tilde{b}_q(\tau)\big) = O_p(\varrho_q + \varrho_s + \varrho_{bk})$, where $\varrho_q = h^s + \sqrt{\ln m/(mh)}$, $\varrho_s = \sqrt{m/n}$, and $\varrho_{bk} = (\ln n)^{3/4}\, n^{-1/4}$. In addition, both $\sqrt{n}\big(\check{b}_q(\tau) - \beta(\tau)\big)$ and $\sqrt{n}\big(\tilde{b}_q(\tau) - \beta(\tau)\big)$ converge in distribution to $\mathcal{N}\big(0, \Sigma_q(\tau)\big)$.

Both $\check{b}_q(\tau)$ and $\tilde{b}_q(\tau)$ are asymptotically efficient, with asymptotic covariance matrix $\Sigma_q(\tau)$. To estimate the latter, one may employ a weighted version of the Hessian $\widehat{R}_h^{(2)}(b; \tau)$. The rate at which the difference $\sqrt{n}\big(\check{b}_q(\tau) - \tilde{b}_q(\tau)\big)$ shrinks to zero involves three components. The first is the consistency rate ϱ_q of the qdf estimator $\check{q}_h(\tau|x)$ in Proposition 1. Using standard kernel estimators as in Otsu (2008) and Komunjer and Vuong (2010) or k-NN smoothing as in Newey and Powell (1990)

and Zhao (2001) would also involve a variance term of order $1/(mh^d)$ due to the curse of dimensionality. The rate ϱ_s comes from the sample splitting scheme, whereas ϱ_{bk} arises from the Bahadur–Kiefer linear approximation.

A natural extension of Proposition 2 is to consider smoothed versions of the objective functions corresponding to the efficient feasible and infeasible estimators, to obtain better AMSE performances.

4. MONTE CARLO STUDY

To assess how well the asymptotic theory reflects the performance of our estimator in finite samples, we run simulations for a 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 five different specifications for the error distribution. The first three are asymmetric, with errors coming from exponential, Gumbel (Type I extreme value), and χ_3^2 distributions. The fourth specification displays heavy tails, with errors following a t-student with 3 degrees of freedom. The fifth distribution exhibits conditional heteroscedasticity: $\epsilon = \frac{1}{4}(1 + \tilde{X})Z$, with $Z \sim \mathcal{N}(0, 1)$. We recenter, if necessary, the error distributions to ensure median zero and, except to the χ_3^2 distribution, also rescale them to obtain unconditional variance equal to two. Apart from the exponential distribution, the other specifications are as in Horowitz (1998), Whang (2006), and Kaplan and Sun (2017).

In each of the 100,000 replications, we sample $n \in \{100, 250, 500, 1000\}$ observations 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 asymptotically valid confidence intervals at the 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, 1.96 and 2.58 at the 95% and 99% confidence levels, respectively).

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 them at Silverman's (1986) rule-of-thumb bandwidth $h_{\text{ROT}} = 1.06\widehat{s}\,n^{-1/5}$, where $\widehat{s}$ is the minimum between the sample standard deviation and the interquartile range (divided by 1.38898) of the standard MR residuals. We compute the MR standard errors by a pairs bootstrap procedure as usual in the literature,[4] whereas the SMR standard errors are as in Horowitz (1998, sec. 2). Finally, we estimate the standard errors for the CKMR estimator using the square root of the diagonal entries of $\widehat{\Sigma}_h(1/2)$.

Figures 3–7 show how each estimator performs across distributions and sample sizes. For simplicity, we focus exclusively on the slope coefficient given that we find little difference between the intercept estimates. For every figure, the first row displays the relative mean squared error (RMSE) of the

---

[4]We omit results using standard errors as in Koenker (2005, secs. 3.4.2 and 4.10.1) because bootstrapping always entails better empirical coverage. They are available from the authors upon request.
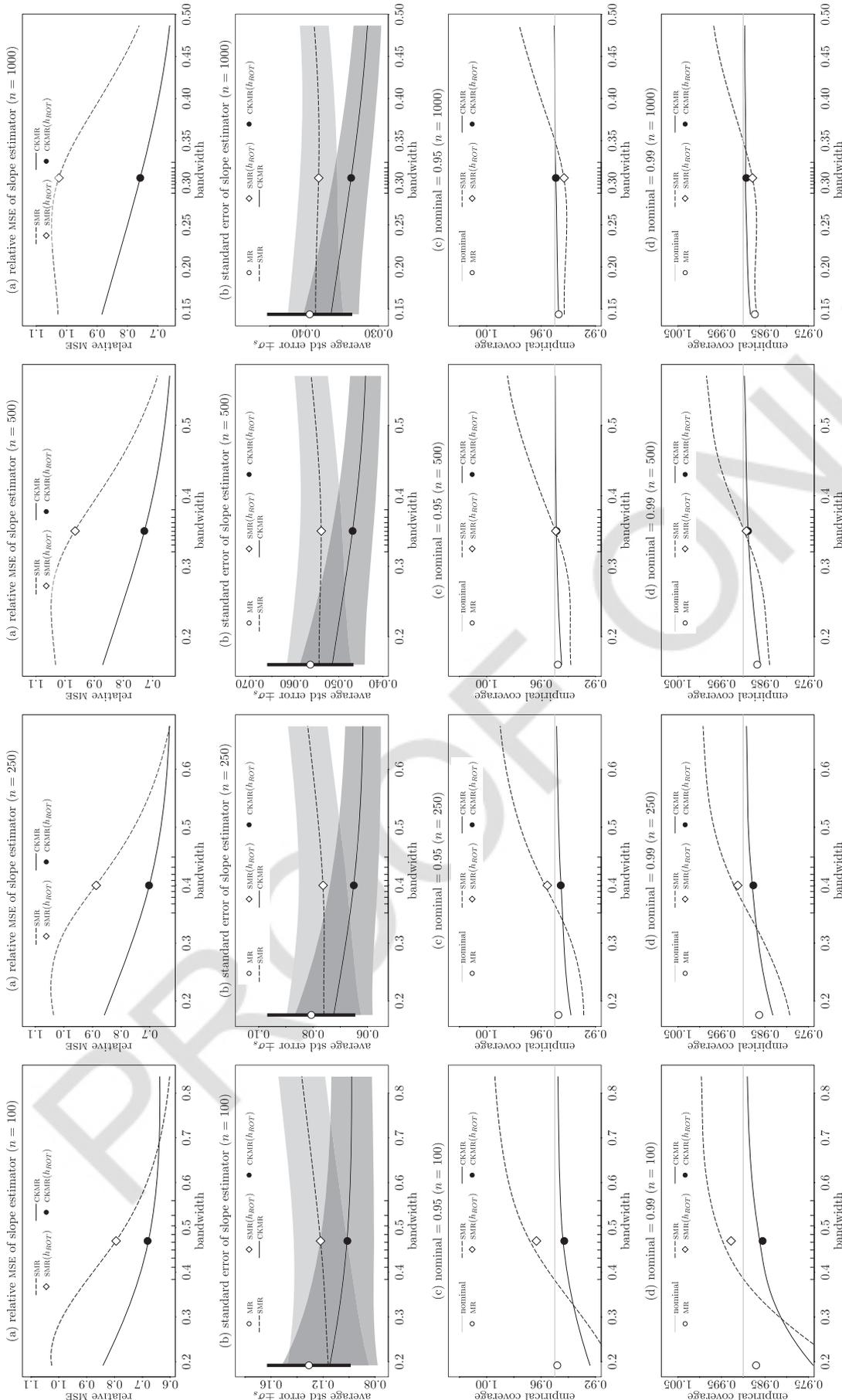
Figure 3. Performance of the MR estimators with exponential errors. The first row of plots displays the mean squared error relative to the standard MR estimator, whereas the second row depicts the standard error of the slope estimator, within a one-standard-deviation band. The third and fourth rows focus, respectively, on the empirical coverage of the asymptotic confidence intervals at the 95% and 99% levels, as a function of the bandwidth h. We consider sample sizes from 100 to 1000 observations. MR corresponds to Koenker and Bassett's (1978) standard estimator, SMR to Horowitz's (1998) smoothed estimator, and CKMR to our convolution-type kernel estimator. Finally, h_{ROT} denotes the rule-of-thumb bandwidth, whose results we place arbitrarily at its average value across the 100,000 replications.

Figure 4. Performance of the MR estimators with Gumbel errors. The first row of plots displays the mean squared error relative to the standard MR estimator, whereas the second row depicts the standard error of the slope estimator, within a one-standard-deviation band. The third and fourth rows focus, respectively, on the empirical coverage of the asymptotic confidence intervals at the 95% and 99% levels, as a function of the bandwidth h. We consider sample sizes from 100 to 1000 observations. MR corresponds to Koenker and Bassett's (1978) standard estimator, SMR to Horowitz's (1998) smoothed estimator, and CKMR to our convolution-type kernel estimator. Finally, h_{ROT} denotes the rule-of-thumb bandwidth, whose results we place arbitrarily at its average value across the 100,000 replications.

Figure 5. Performance of the MR estimators with chi-squared errors. The first row of plots displays the mean squared error relative to the standard MR estimator, whereas the second row depicts the standard error of the slope estimator, within a one-standard-deviation band. The third and fourth rows focus, respectively, on the empirical coverage of the asymptotic confidence intervals at the 95% and 99% levels, as a function of the bandwidth h. We consider sample sizes from 100 to 1000 observations. MR corresponds to Koenker and Bassett's (1978) standard estimator, SMR to Horowitz's (1998) smoothed estimator, and CKMR to our convolution-type kernel estimator. Finally, h_{ROT} denotes the rule-of-thumb bandwidth, whose results we place arbitrarily at its average value across the 100,000 replications.

Figure 6. Performance of the MR estimators with *t*-student errors. The first row of plots displays the mean squared error relative to the standard MR estimator, whereas the second row depicts the standard error of the slope estimator, within a one-standard-deviation band. The third and fourth rows focus, respectively, on the empirical coverage of the asymptotic confidence intervals at the 95% and 99% levels, as a function of the bandwidth *h*. We consider sample sizes from 100 to 1000 observations. MR corresponds to Koenker and Bassett's (1978) standard estimator, SMR to Horowitz's (1998) smoothed estimator, and CKMR to our convolution-type kernel estimator. Finally, h_{ROT} denotes the rule-of-thumb bandwidth, whose results we place arbitrarily at its average value across the 100,000 replications.

Figure 7. Performance of the MR estimators with heteroscedastic errors. The first row of plots displays the mean squared error relative to the standard MR estimator, whereas the second row depicts the standard error of the slope estimator, within a one-standard-deviation band. The third and fourth rows focus, respectively, on the empirical coverage of the asymptotic confidence intervals at the 95% and 99% levels, as a function of the bandwidth h. We consider sample sizes from 100 to 1000 observations. MR corresponds to Koenker and Bassett's (1978) standard estimator, SMR to Horowitz's (1998) smoothed estimator, and CKMR to our convolution-type kernel estimator. Finally, h_{ROT} denotes the rule-of-thumb bandwidth, whose results we place arbitrarily at its average value across the 100,000 replications.

smoothed estimators with respect to the standard MR estimator across different sample sizes, whereas the second row depicts the standard error of the slope estimators, within their one-standard-deviation band. The third and fourth rows portray, respectively, the empirical coverage of the asymptotic confidence intervals at the 95% and 99% levels.

The smoothed estimators obviously depend on the bandwidth choice, and hence, we plot their results as a function of the bandwidth h, singling out the corresponding outcomes for the rule-of-thumb bandwidth h_{ROT}. As the value of the latter changes across replications, we arbitrarily place the results at the average value of h_{ROT} across the 100,000 replications. The inside tick marks in the horizontal axis depict the deciles of the distribution of the rule-of-thumb bandwidth across replications. The distribution seems symmetric in every instance, with dispersion reducing drastically once we move from a sample size of 100 to 1000 observations.

Smoothing seems to pay off as both SMR and CKMR entail lower MSE than the standard MR estimator for every distribution. Interestingly, the RMSE of the CKMR estimator is less sensitive to the sample size and bandwidth value, especially around the rule-of-thumb choice, than the RMSE of the SMR estimator. For instance, if errors are exponential as in Figure 3, the CKMR estimator using the rule-of-thumb bandwidth yields a RMSE of about 70% for the smaller sample sizes, and between 75% and 80% for the sample sizes of 500 and 1000 observations. Conversely, the RMSE of the SMR estimator changes from 80% to 90% as we increase the sample size from 100 to 250 observations and then cease to improve the standard MR estimator in terms of MSE for the larger sample sizes.

Regardless of the specification we consider, the key to the CKMR success lies on the much lower standard errors. It is apparent from Figures 3 to 7 that the standard errors of the SMR estimator that uses the rule-of-thumb bandwidth are very close to the bootstrap-based standard errors of the MR estimator. They are slightly lower for the exponential and chi-squared errors, but slightly larger for the Gumbel, t-student, and heteroscedastic specifications.

The confidence intervals match very well their nominal values for every estimator at the 95% and 99% levels. This is particularly reassuring for the CKMR estimator, whose confidence intervals are significantly narrower due to the smaller standard errors. In addition, the empirical coverage of the CKMR estimator is very stable across bandwidth values, especially for samples with 250 or more observations. In contrast, the empirical coverage of the SMR estimator deteriorates for larger bandwidth values regardless of the error distribution and sample size. All in all, the CKMR estimator performs very well relative to the MR and SMR estimators, with lower MSE and tighter (asymptotic) confidence intervals that are nonetheless valid even in small samples.

4.1.   Conditional qdf and pdf Estimation

Next, we complement the discussion in Section 2.2 by illustrating the qdf and pdf estimation of the QR model $Y = \beta_0(U) + \beta_1(U)X_1 + \beta_2(U)X_2 + \beta_3(U)X_3$, where U, X_1, X_2, and X_3 are independent and uniformly distributed variables on the unit interval $[0, 1]$, $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are, respectively, given by the $\mathsf{Beta}(1, 16)$ and $\mathsf{Beta}(32, 32)$ quantile functions, $\beta_2(\tau) = 1$, and $\beta_3(\tau) = \frac{(2\pi+8)\tau - (\cos(2\pi\tau)-1)}{2\pi+8}$. The large number of regressors relative to the sample size of 1000 observations does not bode well for standard nonparametric estimation due to the curse of dimensionality.

In each of the 1000 replications, we estimate the pdf by means of (8) using the rule-of-thumb bandwidth h_{ROT}.[5] Figure 8 displays the true quantile function in Panel (a), the true qdf in Panel (b), and the true conditional density of Y given $(x_1, x_2, x_3) = (0.9, 0.5, 0.9)$ in Panel (c). The gray-shaded regions in Panels (a) and (b) correspond to the interquantile band from 0.05 to 0.95, that is, the region for which at most 5% of the replications lay below its lower bound and at most 5% above its upper bound, at each fixed τ.

Figure 8(a) suggests that the variability and bias of the QR estimator are not only quite small but also of a similar magnitude. Panel (b) reveals that the qdf estimates approximate well the poles in the tails, up to some boundary bias at $\tau = 0$, with a very narrow band regardless of the quantile level τ. Panel (c) displays the *paths* $\tau \mapsto \widehat{\mathsf{f}}_h(\tau|x)$ for $\tau \in \{0.01, 0.02, \ldots, 0.99\}$,

---

[5]It is perhaps worth mentioning that the optimal bandwidth for the QR estimation is not necessarily optimal for the qdf estimator.

| (a) | (b) | (c) |
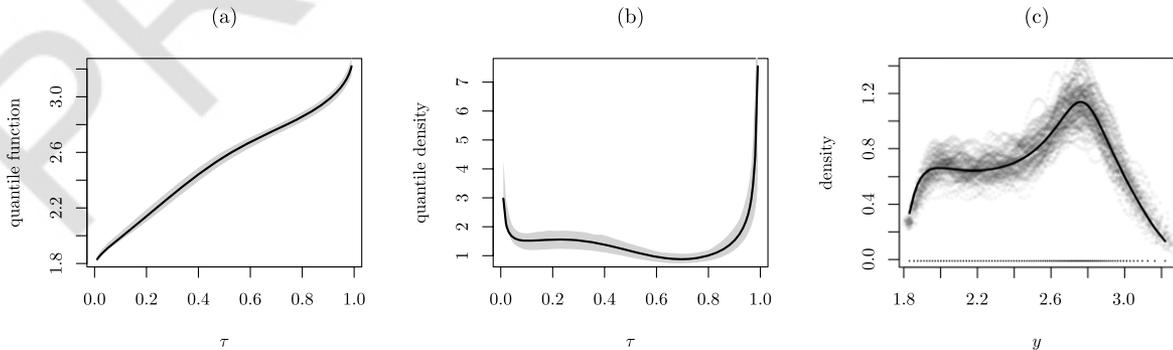
Figure 8. Conditional quantile function, qdf, and pdf estimates. We consider 1000 samples of 1000 observations from $Y = \beta_0(U)+\beta_1(U)X_1+\beta_2(U)X_2 + \beta_3(U)X_3$, where U, X_1, X_2, and X_3 are independent and uniformly distributed variables on the unit interval $[0, 1]$, $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are, respectively, given by the $\mathsf{Beta}(1, 16)$ and $\mathsf{Beta}(32, 32)$ quantile functions, $\beta_2(\tau) = 1$, and $\beta_3(\tau) = \frac{(2\pi+8)\tau-(\cos(2\pi\tau)-1)}{2\pi+8}$. Panels (a) to (c) are about the conditional quantile function, qdf, and pdf of Y given $(x_1, x_2, x_3) = (0.9, 0.5, 0.9)$, respectively. Gray shades refer to the [0.05,0.95]-interquantile bands of the corresponding estimators using h_{ROT} as bandwidth, whereas solid black lines depict the true functions.

corresponding to the 100 first realizations of $\widehat{f}_h$. We find a much higher variability for the pdf curve estimates, as typical in nonparametric estimation. As before, we evince some boundary bias at $\tau = 0$ and tighter interquantile bands in the tails. Altogether, the pdf estimator seems to perform reasonably well even for three conditioning variables.

5.   CONCLUDING REMARKS

This article proposes a convolution-type kernel QR estimator based upon smoothing the objective function. The resulting estimator improves on standard QR estimation both in terms of asymptotic mean square error (MSE) and Bahadur–Kiefer representation. Applying the implicit function theorem to the corresponding first-order condition gives way to a conditional pdf estimator that avoids the curse of dimensionality by exploiting the QR specification. We also show how to use this pdf estimator to come up with a two-step efficient QR estimator. Many of our asymptotic results are uniform with respect to the quantile level and bandwidth. The latter is convenient because it makes room for data-driven bandwidth choices, such as the rule-of-thumb bandwidth we propose. In particular, we establish a functional central limit theorem for our smoothed QR estimator that accommodates data-driven bandwidths.

Simulations show that the rule-of-thumb bandwidth works well in practice, yielding very good results in terms of confidence interval coverage. They also reveal that our smoothed QR estimator outperforms not only the standard QR estimator, but also Horowitz's (1998) alternative smoothed estimator. Finally, our Monte Carlo study also confirms that our QR-based pdf estimator does not suffer from the curse of dimensionality.

There are many extensions to consider in future research. First, uniformity in bandwidth is useful not only for adaptive bandwidth choices as in Lepski, Mammen, and Spokoiny (1997), but also for bandwidth-snooping-robust inference as in Armstrong and Kolesár (2018). Second, one could combine our QR-based pdf estimator in a data-rich environment with Belloni and Chernozhukov's (2011) lasso-type methods to select the relevant covariates and with Volgushev, Chao, and Cheng's (2019) to cope with large sample sizes. Third, one could exploit the connection between QR and quantile instrumental variable models (Kaido and Wuthrich 2018) to obtain smoother estimators and better inferential procedures (Chernozhukov and Hansen 2005; Jun 2008; Andrews and Mikusheva 2016; de Castro et al. 2019). Finally, one could also extend our smoothing approach to handle panel QR as in Galvão and Kato (2016).

A.   PROOFS OF MAIN RESULTS

Notice that $\widehat{R}(b; \tau)$ is integrable if and only if Y and X are integrable. This matters for the definition of $\beta(\tau)$ as the minimizer of $R(b; \tau) := \mathbb{E}[\widehat{R}(b; \tau)]$. It is convenient to assume that both $\widehat{R}(b; \tau)$ and $\widehat{R}_h(b; \tau)$ are integrable such that $R(b; \tau)$ and $R_h(b; \tau)$ are well defined. Otherwise, one should define $R_h(b; \tau)$ as $\mathbb{E}[\widehat{R}_h(b; \tau) - \widehat{R}_h(0; \tau)]$, and similarly for $R(b; \tau)$. These quantities are finite under Assumption X. Let $\mathcal{S}$ denote the set $\mathbb{R}^d \times [\underline{\tau}, \overline{\tau}] \times [\underline{h}_n, \overline{h}_n]$ to which (b, τ, h) belongs. Note that $\mathcal{S}$ depends on n through $[\underline{h}_n, \overline{h}_n]$. In what follows, C denotes

a generic constant that may vary from line to line. Lastly, the online appendix collects the proofs of the intermediary results.

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

(i) $\displaystyle \sup_{(b,\tau,h)\in\mathcal{S}} \left| \frac{R_h(b; \tau) - R(b; \tau)}{h^{s+1}} \right| = O(1);$

(ii) $\displaystyle \sup_{(b,\tau,h)\in\mathcal{S}} \left\| \frac{R_h^{(1)}(b; \tau) - R^{(1)}(b; \tau)}{h^{s+1}} \right\| = O(1);$

(iii) $\displaystyle \sup_{(b,\tau,h)\in\mathcal{S}} \left\| \frac{R_h^{(2)}(b; \tau) - R^{(2)}(b; \tau)}{h^s} \right\| = o(1);$

(iv) $\displaystyle \sup_{(\delta,b,\tau,h)\in\mathbb{R}^d\times\mathcal{S}} \left\| \frac{R_h^{(2)}(b + \delta; \tau) - R_h^{(2)}(b; \tau)}{\|\delta\|} \right\| = O(1).$

Proof of Lemma 1. See online appendix.   □

Proof of Theorem 1. Observe first that $R(\cdot; \tau)$ is strictly convex. To see this, we show that, 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 $v \in \mathbb{R}^d$,

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

by Assumptions X and Q. It follows from Lemma 1(i) and the continuity of $b \mapsto R^{(2)}(b; \tau) = R^{(2)}(b)$ that any potential minimizer $\beta_h(\tau)$ of $R_h(\cdot; \tau)$ must lie in a compact $\mathcal{B}$ for every admissible (τ, h). As Lemma 1(i) ensures that $R_h(\cdot; \tau)$ is convex over $\mathcal{B}$ for all admissible τ and h small enough, it follows that $\beta_h(\tau)$ is unique for $(h, \tau) \in [\underline{h}_n, \overline{h}_n] \times [\underline{\tau}, \overline{\tau}]$ and n large enough.

By the implicit function theorem, $\beta_h(\tau)$ is continuous in (τ, h) given that $R_h^{(1)}(b; \tau)$ and $R_h^{(2)}(b; \tau)$ are continuous in (b, τ, h) and that $R_h^{(1)}(\beta_h(\tau); \tau) = 0$. Let

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

and let $c := \inf f\Big(x'\big[\beta(\tau) + w\big(\beta_h(\tau) - \beta(\tau)\big)\big] \mid x\Big)$ for $(\tau, h, w, x) \in [\underline{\tau}, \overline{\tau}] \times [\underline{h}_n, \overline{h}_n] \times [0, 1] \times \mathrm{supp}(X)$. Assumptions X and Q ensure that $c > 0$, and so

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

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

$$R^{(1)}\big(\beta_h(\tau); \tau\big) - R_h^{(1)}\big(\beta_h(\tau); \tau\big) = R^{(1)}\big(\beta_h(\tau); \tau\big) - R^{(1)}\big(\beta(\tau); \tau\big)$$
$$= A(\tau, h)\big[\beta_h(\tau) - \beta(\tau)\big],$$

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

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

$$\left. \left\| \frac{R^{(1)}(\beta_h(\tau);\tau) - R_h^{(1)}(\beta_h(\tau);\tau)}{h^{s+1}} \right\| \right\} = O(1).$$

Now, a Taylor expansion obtains

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

by Lemma 1. It follows from $\int_0^1 w(1-w)^{s-1}\,dw = \frac{1}{(s+1)s}$ and (25) that

$$T_2 := \int_{-\infty}^{x'\beta(\tau)} \left\{ \mathbb{E}[k_h(v-Y)\,|\,x] - f(v\,|\,x) \right\}\,dv$$
$$= h^{s+1} \frac{\int z^{s+1} k(z)\,dz}{(s+1)!} f^{(s)}(x'\beta(\tau)\,|\,x) + o(h^{s+1}).$$

Using (24) then yields

$$R_h^{(1)}(\beta(\tau);\tau) = h^{s+1} \frac{\int z^{s+1} k(z)\,dz}{(s+1)!}$$
$$\times \int x f^{(s)}(x'\beta(\tau)\,|\,x)\,dF_X(x) + o(h^{s+1}),$$

completing the proof. □

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 $\underline{f}$ and $\bar{f}$ with $\underline{f} \leq \bar{f}$, let $[\underline{f},\bar{f}]$ denote the set of all functions g such that $\underline{f} \leq g \leq \bar{f}$. For a set $\mathcal{F}$ and a family $\{F_i\}$ of subsets of $\mathcal{F}$, we say that $\{F_i\}$ *covers* $\mathcal{F}$ if $\mathcal{F} \subset \bigcup_i F_i$.

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

(i) there are some positive constants σ and M such that $\mathbb{E}\left[|f(Z_i)|^2\right] \leq \sigma^2$ and $\sup_{z\in\mathcal{Z}} |f(z)| \leq M$ for all $f \in \mathcal{F}$;

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

It then follows that, for any $r \geq 0$,

$$\Pr\left(\sup_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(Z_i) - \mathbb{E}[f(Z_i)]) \geq \mathcal{H}_n + 7\sigma\sqrt{2r} + \frac{2Mr}{\sqrt{n}} \right)$$
$$\leq \exp(-r), \qquad (A.1)$$

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 (2007). □

Before proceeding to the next result, let us introduce some additional notation. In what follows, let $\sup_{(\tau,h)}$ denote the supremum over $(\tau,h) \in [\underline{\tau},\bar{\tau}] \times [\underline{h}_n,\bar{h}_n]$, which depends on n via $[\underline{h}_n,\bar{h}_n]$. We also use the same implicit notation for any other similar operator (e.g., infimum or union). Recalling that $\varrho_n^{-1}(h) := \sqrt{nh/(\ln n)}$, let

$$\mathcal{E}_n(r) := \left\{ \sqrt{n} \sup_{(\tau,h)} \varrho_n^{-1}(h) \left\| \widehat{\beta}_h(\tau) - \beta_h(\tau) + D_h^{-1}(\tau)\widehat{S}_h(\tau) \right\| \geq r^2 \right\}.$$

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

$$\sqrt{n}(\widehat{\beta}_h(\tau) - \beta_h(\tau)) = -\sqrt{n}D_h^{-1}(\tau)\widehat{S}_h(\tau) + \widehat{E}_h(\tau),$$

where the approximation error is such that $\|\widehat{E}_h(\tau)\| \leq \varrho_n(h)\,r^2$ uniformly for $(\tau,h) \in [\underline{\tau},\bar{\tau}] \times [\underline{h}_n,\bar{h}_n]$. In particular, if $\Pr(\mathcal{E}_n(r))$ is small for large r, then the representation (11) from Theorem 2 holds uniformly in τ and h.

Finally, let $\breve{A}$ denote the complementary set of A and

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

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

where the norms in $\mathcal{E}_n^1(r)$ and $\mathcal{E}_n^2(r)$ are the Euclidean and operator norms, respectively. We are now ready to state the functional exponential inequality we will apply in the remaining technical proofs.

Proposition 3. Given Assumptions X, Q, and K, $\widehat{\beta}_h(\cdot)$ is unique and continuous over $[\underline{\tau},\bar{\tau}] \times [\underline{h}_n,\bar{h}_n]$ with probability growing to one. There also exist positive constants C_0, C_1, and C_2 such that, for small enough ϵ, $1/r$ and $1/n$,

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

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

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

Proof of Proposition 3. See online appendix. □

Theorem 2 is an immediate corollary to Proposition 3, whose proof builds on Lemmas 2–4. It is sometimes convenient to consider the auxiliary objective functions $\widehat{\mathcal{R}}_h(b;\tau) := \widehat{R}_h(b;\tau) - \widehat{R}_h(\beta_h(\tau);\tau)$ and $\mathcal{R}_h(b;\tau) := \mathbb{E}[\widehat{\mathcal{R}}_h(b;\tau)]$. They are such that $\widehat{\beta}_h(\tau) = \arg\min_b \widehat{\mathcal{R}}_h(b;\tau)$ and $\beta_h(\tau) = \arg\min_b \mathcal{R}_h(b;\tau)$. Similarly, set $\breve{\mathcal{R}}(b;\tau) = \widehat{R}(b;\tau) - \widehat{R}(\beta(\tau);\tau)$ and $\mathcal{R}(b;\tau) := \mathbb{E}[\breve{\mathcal{R}}(b;\tau)]$. The next result shows that $\widehat{\beta}_h(\tau)$ is close to $\beta_h(\tau)$ uniformly for $(\tau,h) \in [\underline{\tau},\bar{\tau}] \times [\underline{h}_n,\bar{h}_n]$.

Lemma 3. Suppose that Assumptions X, Q, and K hold. If n is large enough, there is a positive constant C such that, for any $\eta \in [1/\ln n, 1]$,

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

Proof of Lemma 3. See online appendix. □

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,

$$\Pr\left(\sup_{(\tau, h)} \left\| \sqrt{n} \widehat{S}_h(\tau) \right\| \geq C_1(1 + r) \right) \leq C_0 \exp\left(-r^2\right),$$

$$\Pr\left(\sup_{(\tau, h)} \sup_{\{b: \|b - \beta_h(\tau)\| \leq \eta\}} \left\| \sqrt{\frac{nh}{\ln n}} \left(\widehat{R}_h^{(2)}(b, \tau) - R_h^{(2)}(b, \tau) \right) \right\| \right.$$

$$\left. \geq C_1(1 + r) \right) \leq C_0 \exp\left(-r \ln n\right).$$

Moreover, for any compact subset $\mathcal{B}$ of $\mathbb{R}^{d+1}$,

$$\sup_{(b_0, b_1) \in \mathcal{B}^2} \sup_{(\tau, h)} \frac{\left\| \widehat{R}_h^{(2)}(b_1, \tau) - \widehat{R}_h^{(2)}(b_0, \tau) \right\|}{\|b_1 - b_0\|} = O_p(1),$$

$$\sup_{(b_0, b_1) \in \mathcal{B}^2} \sup_{(\tau, h)} \frac{\left\| \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \left\{ \left[K(-e_i(b_1)/h) - \tau \right]^2 \right. \right.}{\left. \left. - \left[K(-e_i(b_0)/h) - \tau \right]^2 \right\} \right\|}{\|b_1 - b_0\|} = O_p(1),$$

$$\sup_{b \in \mathcal{B}} \sup_{(\tau, h)} \left\| \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \left[K(-e_i(b)/h) - \tau \right]^2 \right. \\ \left. - \mathbb{E}\left[X_i X_i' \left[K(-e_i(b)/h) - \tau \right]^2 \right] \right\| = O_p(1/\sqrt{n}).$$

Proof of Lemma 4. See online appendix. $\qquad\qquad\square$

Proof of Theorem 2. As uniqueness holds, to establish (11), it suffices to fix ϵ and increase r to make $\Pr(\mathcal{E}_n(r)) \leq \Pr\left(\mathcal{E}_n(r) \cap \left[\check{\mathcal{E}}_n^1(r) \cap \check{\mathcal{E}}_n^2(r) \right] \right) + \Pr\left(\mathcal{E}_n^1(r) \right) + \Pr\left(\mathcal{E}_n^2(r) \right)$ arbitrarily small for n large enough. As for (12), it readily follows from (28) in the online appendix and Lemma 4 in view that the latter and (11) imply that $\sup_{\tau, h} \left\| \widehat{\beta}_h(\tau) - \beta_h(\tau) \right\| = O_p(1/\sqrt{n})$. $\qquad\square$

A.3. Asymptotic Variance and Mean Squared Error

Proof of Theorem 3. We first show that the expansion

$$\mathbb{V}\left(\sqrt{n} D^{-1}(\tau) \widehat{S}_h(\tau) \right) = \Sigma(\tau) - c_k h D^{-1}(\tau) + O(h^2) \qquad (A.2)$$

holds uniformly with respect to $(\tau, h) \in [\underline{\tau}, \bar{\tau}] \times [\underline{h}_n, \bar{h}_n]$. Given that $\mathbb{E}\left[\widehat{R}_h^{(1)}(\beta_h(\tau); \tau) \right] = 0$,

$$\mathbb{V}\left(\sqrt{n} \widehat{S}_h(\tau) \right) = \mathbb{V}\left(\sqrt{n} \widehat{R}_h^{(1)}(\beta_h(\tau); \tau) \right)$$

$$= \mathbb{V}\left(X \left[K\left(\frac{X' \beta_h(\tau) - Y}{h} \right) - \tau \right] \right)$$

$$= \mathbb{E}\left(XX' \left[K\left(\frac{X' \beta_h(\tau) - Y}{h} \right) - \tau \right]^2 \right)$$

$$= \mathbb{E}\left[XX' K^2 \left(\frac{X' \beta_h(\tau) - Y}{h} \right) \right]$$

$$- 2\tau \mathbb{E}\left[XX' K\left(\frac{X' \beta_h(\tau) - Y}{h} \right) \right] + \tau^2 \mathbb{E}(XX').$$

Along similar lines to the proof of Lemma 1, it follows from Assumptions Q2 and K that

$$\mathbb{E}\left[K\left(\frac{X' \beta_h(\tau) - Y}{h} \right) \middle| X = x \right]$$

$$= \int K\left(\frac{x' \beta_h(\tau) - y}{h} \right) f(y \mid x) \, \mathrm{d}y$$

$$= \frac{1}{h} \int k\left(\frac{x' \beta_h(\tau) - y}{h} \right) F(y \mid x) \, \mathrm{d}y$$

$$= F(x' \beta_h(\tau) \mid x) + \int \left[F(x' \beta_h(\tau) - hz \mid x) - F(x' \beta_h(\tau) \mid x) \right] k(z) \, \mathrm{d}z$$

$$= \tau + O(h^{s+1}), \qquad (A.3)$$

using integration by parts and Theorem 1, in view that $x' \beta(\tau) = F^{-1}(\tau \mid x)$ by definition. Let now $\mathbf{K}(z) = 2 k(z) K(z) = \frac{\mathrm{d}}{\mathrm{d}z} K^2(z)$, so that $\int \mathbf{K}(z) \, \mathrm{d}z = \lim_{z \to \infty} K^2(z) = 1$. As before, this leads to

$$\mathbb{E}\left[K^2 \left(\frac{X' \beta_h(\tau) - Y}{h} \right) \middle| X = x \right]$$

$$= \frac{1}{h} \int \mathbf{K}\left(\frac{x' \beta_h(\tau) - y}{h} \right) F(y \mid x) \, \mathrm{d}y$$

$$= \tau + O(h^{s+1}) + \int \left[F\left(x' \beta_h(\tau) - hz \mid x \right) - F\left(x' \beta_h(\tau) \mid x \right) \right] \mathbf{K}(z) \, \mathrm{d}z$$

$$= \tau + O(h^{s+1}) - h \left[f(x' \beta_h(\tau) \mid x) + O(h) \right] \int z \, \mathbf{K}(z) \, \mathrm{d}z$$

$$= \tau + O(h^{s+1}) - h \left[f(x' \beta(\tau) \mid x) + O(h^{(s+1)}) + O(h) \right] \int z \, \mathbf{K}(z) \, \mathrm{d}z$$

$$= \tau - h f(x' \beta(\tau) \mid x) \int z \, \mathbf{K}(z) \, \mathrm{d}z + O(h^2). \qquad (A.4)$$

The variance expansion (A.2) then follows by noticing that, as $K(-z) = 1 - K(z)$,

$$\int z \, \mathbf{K}(z) \, \mathrm{d}z = 2 \int z \, k(z) \, K(z) \, \mathrm{d}z$$

$$= \int_{-\infty}^{0} z \, \mathrm{d}K^2(z) + \int_{0}^{\infty} z \, \mathrm{d}\left[K^2(z) - 1 \right]$$

$$= -\int_{-\infty}^{0} K^2(z) \, \mathrm{d}z + \int_{0}^{\infty} \left[1 - K^2(z) \right] \mathrm{d}z$$

$$= \int_{0}^{\infty} \left\{ -\left[1 - K(z) \right]^2 + 1 - K^2(z) \right\} \mathrm{d}z$$

$$= 2 \int_{0}^{\infty} K(z)[1 - K(z)] \, \mathrm{d}z.$$

Finally, the local Lipschitz property of matrix inversion ensures that

$$\mathbb{V}\left(\sqrt{n} D_h^{-1}(\tau) \widehat{S}_h(\tau) \right) = \Sigma(\tau) - c_k h D^{-1}(\tau) + O(h^2)$$

$$+ \left[D_h^{-1}(\tau) \mathbb{V}(\sqrt{n} \widehat{S}_h(\tau)) D_h^{-1}(\tau) - D^{-1}(\tau) \mathbb{V}(\sqrt{n} \widehat{S}_h(\tau)) D^{-1}(\tau) \right].$$

given (A.2), Lemma 1, and Theorem 1. To establish (14), it suffices to observe that, by Lemma 1, the norm of the last term within brackets is at most equal to

$$\left\| D_h^{-1}(\tau) - D^{-1}(\tau) \right\| \left\| \mathbb{V}(\sqrt{n} \widehat{S}_h(\tau)) \right\| \left(\left\| D_h^{-1}(\tau) \right\| + \left\| D^{-1}(\tau) \right\| \right)$$

$$\leq C \left\| D_h^{-1}(\tau) - D^{-1}(\tau) \right\| = o(h).$$

$$\square$$

Proof of Theorem 4. It follows from Theorem 1, (14) and $\mathbb{E}\big[\widehat{S}_h(\tau)\big] = 0$ that

$$\text{AMSE}\big(\lambda'\widehat{\beta}_h(\tau)\big) = h^{2s+2}\big[\lambda'B(\tau)\big]^2$$
$$+ \frac{1}{n}\,\lambda'\Big(\Sigma(\tau) - c_k\,h\,D^{-1}(\tau)\Big)\lambda + \underbrace{O(h^2) + o(h^s)}_{=o(h)}.$$

Letting $g(h) = h^{2s+2}\big[\lambda'B(\tau)\big]^2 - n^{-1}c_k\,h\,\lambda'D^{-1}(\tau)\lambda$ and then differentiating with respect to h yields

$$g'(h) = (2s+2)\,h^{2s+1}\big[\lambda'B(\tau)\big]^2 - \frac{1}{n}\,c_k\,\lambda'D^{-1}(\tau)\lambda.$$

Solving for h_λ^* such that $g'(h_\lambda^*) = 0$ yields the desired result and AMSE expansion. $\qquad\square$

A.4. Weak Convergence and Asymptotic Covariance Estimator

In this section, we first establish the weak convergence of $\big\{\sqrt{n}\widehat{S}_h : \tau \in [\underline{\tau}, \overline{\tau}]\big\}$ to a Gaussian process in Lemma 5 and then derive the results (G), (V), and (U) in Theorem 5.

Lemma 5. Under Assumptions X, Q and K, $\big\{\sqrt{n}\widehat{S}_h : \tau \in [\underline{\tau}, \overline{\tau}]\big\}$ converges in distribution to a tight, centered Gaussian process $\big\{W_*(\tau) : \tau \in [\underline{\tau}, \overline{\tau}]\big\}$ with covariance structure

$$\mathbb{V}\big(W_*(\tau), W_*(\varsigma)\big) = (\tau \wedge \varsigma - \tau\varsigma)\,\mathbb{E}(XX'), \qquad \tau, \varsigma \in [\underline{\tau}, \overline{\tau}], \text{ (A.5)}$$

Proof of Lemma 5. See online appendix. $\qquad\square$

Proof of Theorem 5. (G) is a corollary to Theorem 2 and Lemma 5, whereas (V) directly follows from the fact that $\widehat{\beta}_h(\tau) = \beta_h(\tau) + O_p(n^{-1/2})$ uniformly and from Lemma 4. To establish (U), recall that $\widehat{\mathcal{R}}_h^{(1)}\big(\widehat{\beta}_h(\tau); \tau\big) = \mathcal{R}_h^{(1)}\big(\beta_h(\tau); \tau\big) = 0$ and $\widehat{\beta}_h(\tau) = \beta_h(\tau) + O_p(n^{-1/2})$ uniformly. By the implicit function theorem, Lemmas 1 and 4 yield that, with probability tending to one,

$$\frac{\partial\widehat{\beta}_h}{\partial h} = -\Big[\widehat{\mathcal{R}}_h^{(2)}\big(\widehat{\beta}_h(\tau); \tau\big)\Big]^{-1}\frac{\partial}{\partial h}\,\widehat{\mathcal{R}}_h^{(1)}\big(\widehat{\beta}_h(\tau); \tau\big)$$
$$= -\Big[\widehat{\mathcal{R}}_h^{(2)}\big(\widehat{\beta}_h(\tau); \tau\big)\Big]^{-1}\frac{1}{nh}\sum_{i=1}^{n}X_i\,\tilde{k}\bigg(-\frac{e_i\big(\widehat{\beta}_h(\tau)\big)}{h}\bigg),$$
$$\frac{\partial\beta_h}{\partial h} = -\Big[\mathcal{R}_h^{(2)}\big(\beta_h(\tau); \tau\big)\Big]^{-1}\frac{\partial}{\partial h}\,\mathcal{R}_h^{(1)}\big(\beta_h(\tau); \tau\big)$$
$$= -\Big[\mathcal{R}_h^{(2)}\big(\beta_h(\tau); \tau\big)\Big]^{-1}\frac{1}{h}\,\mathbb{E}\bigg[X_i\,\tilde{k}\bigg(-\frac{e_i\big(\beta_h(\tau)\big)}{h}\bigg)\bigg],$$

with $\tilde{k}(t) = tk(t)$. As in the proof of Lemma 4, $\sup_{(\tau,h)\in[\underline{\tau},\overline{\tau}]\times[\underline{h}_n,\overline{h}_n]}\big\|\widehat{\beta}_h(\tau) - \beta_h(\tau)\big\| = O_p(n^{-1/2})$ implies not only that $\sup_{(\tau,h)\in[\underline{\tau},\overline{\tau}]\times[\underline{h}_n,\overline{h}_n]}\sqrt{\frac{nh}{\ln n}}\,\big\|\frac{\partial}{\partial h}\widehat{\mathcal{R}}_h^{(1)}\big(\widehat{\beta}_h(\tau); \tau\big) - \frac{\partial}{\partial h}\mathcal{R}_h^{(1)}\big(\widehat{\beta}_h(\tau); \tau\big)\big\| = O_p(1)$ but also that $\sup_{(\tau,h)\in[\underline{\tau},\overline{\tau}]\times[\underline{h}_n,\overline{h}_n]}\sqrt{\frac{nh}{\ln n}}\,\big\|\frac{\partial\widehat{\beta}_h(\tau)}{\partial h} - \frac{\partial\beta_h(\tau)}{\partial h}\big\| = O_p(1)$. This means that, uniformly in (h_0, h_1, τ),

$$\widehat{\beta}_{h_1}(\tau) - \beta_{h_1}(\tau) - \big[\widehat{\beta}_{h_0}(\tau) - \beta_{h_0}(\tau)\big] = \int_{h_0}^{h_1}\bigg[\frac{\partial\widehat{\beta}_h(\tau)}{\partial h} - \frac{\partial\beta_h(\tau)}{\partial h}\bigg]\mathrm{d}h$$
$$= O_p\bigg(\sqrt{\frac{\ln n}{n}}\bigg)\int_{h_0}^{h_1}h^{-1/2}\,\mathrm{d}h$$
$$= O_p\bigg(\sqrt{\frac{\ln n}{n}}\bigg) \times \big(\sqrt{h_1} - \sqrt{h_0}\big),$$

completing the proof. $\qquad\square$

A.5. The pdf Curve and Efficient QR Estimators

Proof of Proposition 1. Applying the implicit function theorem to the first-order condition $R^{(1)}(\beta(\tau); \tau)$ yields $\beta^{(1)}(\tau) = D^{-1}(\tau)\mathbb{E}[X]$. It then follows from Lemmas 1 and 4, Assumption Q2 and (13) that

$$\sup_{\tau,h}\Big\|\widehat{R}_h^{(2)}\big(\widehat{\beta}_h(\tau); \tau\big) - R^{(2)}\big(\beta(\tau); \tau\big)\Big\|$$
$$\leq o(h^s) + O_p\bigg(\sqrt{\frac{\ln n}{nh}}\bigg) + \sup_{\tau,h}\Big\|R^{(2)}\big(\widehat{\beta}_h(\tau); \tau\big) - R^{(2)}\big(\beta(\tau); \tau\big)\Big\|$$
$$= o(h^s) + O_p\big(\sqrt{\ln n/(nh)}\big).$$

The result readily follows. $\qquad\square$

Proof of Proposition 2. Let $\mathcal{F}_m$ denote the σ-field generated by the first m observations and

$$\widehat{S}(\tau; n, m, q) := \frac{1}{\sqrt{n-m}}\sum_{i=m+1}^{n}X_i\,\frac{\mathbb{I}\big[Y_i \leq Q(\tau|X_i)\big] - \tau}{q(\tau|X_i)}.$$

By following the same argument as in Guerre and Sabbah (2012), we find that, conditional on $\mathcal{F}_m$, the Bahadur–Kiefer representations of $\tilde{b}(\tau)$ and $\check{b}(\tau)$ are

$$\sqrt{n}\Big(\tilde{b}(\tau) - \beta(\tau)\Big) = -D_q^{-1}(\tau)\,\widehat{S}(\tau; n, 0, q) + O_p\Big((\ln n)^{3/4}\,n^{-1/4}\Big),$$
$$\sqrt{n}\Big(\check{b}(\tau) - \beta(\tau)\Big) = -D_{\check{q}_h}^{-1}(\tau)\,\widehat{S}(\tau; n, m, \check{q}_h) + O_p\Big((\ln n)^{3/4}\,n^{-1/4}\Big),$$

using Proposition 3 for $\check{b}(\tau)$ to ensure that the weights $\check{q}_h(\tau|X_i)$ stay, uniformly in i, asymptotically bounded away from 0 and infinity. Note that Proposition 3 implies that $\big\|D_{\check{q}_h}^{-1}(\tau) - D_q^{-1}(\tau)\big\| = O_p(\varrho_q)$. As the covariates are bounded by assumption, it follows from Proposition 3 that

$$\mathbb{E}\Big[\big(\widehat{S}(\tau; n, m, \check{q}_h) - \widehat{S}(\tau; n, 0, q)\big)^2\,\big|\,\mathcal{F}_m\Big]$$
$$\leq \frac{m}{n} + \mathbb{E}\Bigg[\bigg(\widehat{S}(\tau; n, m, \check{q}_h) - \sqrt{1 - \frac{m}{n}}\,\widehat{S}(\tau; n, m, q)\bigg)^2\,\bigg|\,\mathcal{F}_m\Bigg]$$
$$\leq C\,\frac{m}{n} + 2\,\mathbb{E}\Big[\big(\widehat{S}(\tau; n, m, \check{q}_h) - \widehat{S}(\tau; n, m, q)\big)^2\,\big|\,\mathcal{F}_m\Big]$$
$$\leq C\Bigg[\frac{m}{n} + \sum_{i=m+1}^{n}\bigg(\frac{\check{q}_h(\tau|X_i) - q(\tau|X_i)}{\check{q}_h(\tau|X_i)q(\tau|X_i)}\bigg)^2\Bigg]$$
$$= O(\varrho_s^2) + O_p(\varrho_q^2).$$

This implies the first part of Proposition 2. The CLT easily follows. $\qquad\square$

REFERENCES

Andrews, I., and Mikusheva, A. (2016), "Conditional Inference With a Functional Nuisance Parameter," *Econometrica*, 84, 1571–1612. [16]

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006), "Quantile Regression Under Misspecification, With an Application to the U.S. Wage Structure," *Econometrica*, 74, 539–563. [3]

Armstrong, T. B., and Kolesár, M. (2018), "A Simple Adjustment for Bandwidth Snooping," *Review of Economic Studies*, 85, 732–765. [16]

Azzalini, A. (1981), "A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method," *Biometrika*, 68, 326–328. [2]

Bassett, G., and Koenker, R. (1982), "An Empirical Quantile Function for Linear Models With iid Errors," *Journal of the American Statistical Association*, 77, 407–415. [1]

Belloni, A., and Chernozhukov, V. (2011), "ℓ_1-Penalized Quantile Regression in High-Dimensional Sparse Models," *The Annals of Statistics*, 39, 82–130. [16]

Buchinsky, M. (1995), "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models a Monte Carlo Study," *Journal of Econometrics*, 68, 303–338. [1]

——— (1998), "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research," *Journal of Human Resources*, 33, 88–126. [1]

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2010), "Quantile and Probability Curves Without Crossing," *Econometrica*, 78, 1093–1125. [4]

Chernozhukov, V., and Hansen, C. (2005), "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–262. [16]

Chernozhukov, V., and Hong, H. (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346. [1]

Cheung, K., and Lee, S. M. (2010), "Bootstrap Variance Estimation for Nadaraya Quantile Estimator," *Test*, 19, 131–145. [2]

de Castro, L. I., Galvão, A. F., Kaplan, D. M., and Liu, X. (2019), "Smoothed GMM for Quantiles Models," *Journal of Econometrics*. [16]

Falk, M. (1984), "Relative Deficiency of Kernel Type Estimators of Quantiles," *The Annals of Statistics*, 12, 261–268. [2]

Fan, Y., and Liu, R. (2016), "A Direct Approach to Inference in Nonparametric and Semiparametric Quantile Models," *Journal of Econometrics*, 191, 196–216. [1]

Galvão, A. F., and Kato, K. (2016), "Smoothed Quantile Regression for Panel Data," *Journal of Econometrics*, 193, 92–112. [16]

Goh, S. C., and Knight, K. (2009), "Nonstandard Quantile-Regression Inference," *Econometric Theory*, 25, 1415–1432. [1]

Guerre, E., and Sabbah, C. (2012), "Uniform Bias Study and Bahadur Representation for Local Polynomial Estimators of the Conditional Quantile Function," *Econometric Theory*, 28, 87–129. [2,4,19]

Guerre, E., and Sabbah, C. (2017), "An Augmented Local Polynomial Quantile Framework for Dependence Analysis," QMUL and Université de Lille 2. [4]

Hall, P., and Sheather, S. J. (1988), "On the Distribution of a Studentized Quantile," *Journal of the Royal Statistical Society*, Series B, 50, 381–391. [2]

Horowitz, J. L. (1998), "Bootstrap Methods for Median Regression Models," *Econometrica*, 66, 1327–1351. [1,2,5,7,9,10,11,12,13,14,16]

Jun, S. J. (2008), "Weak Identification Robust Tests in an Instrumental Quantile Model," *Journal of Econometrics*, 144, 118–138. [16]

Jurečková, J., Sen, P. K., and Picek, J. (2012), *Methodology in Robust and Nonparametric Statistics*, Boca Raton, FL: CRC/Chapman-Hall. [1,6]

Kaido, H., and Wuthrich, K. (2018), "Decentralization Estimators for Instrumental Variable Quantile Regression Models," arXiv no. 1812.10925. [2,16]

Kaplan, D. M., and Sun, Y. (2017), "Smoothed Estimating Equations for Instrumental Variables Quantile Regression," *Econometric Theory*, 33, 105–157. [2,5,8,9]

Kato, K. (2012), "Asymptotic Normality of Powell's Kernel Estimator," *Annals of the Institute of Statistical Mathematics*, 64, 255–273. [3]

Knight, K. (2001), "Comparing Conditional Quantile Estimators: First and Second Order Considerations," Technical Report, University of Toronto. [1,6]

Koenker, R. (1994), "Confidence Intervals for Regression Quantiles," in *Proceedings of the 5th Prague Symposium on Asymptotic Statistics*, eds. P. Mandl and M. Huskova, Springer, pp. 349–359. [1]

——— (2000), "Galton, Edgeworth, Frisch, and Prospects for Quantile Regression in Econometrics," *Journal of Econometrics*, 95, 347–374. [1]

——— (2005), *Quantile Regression*, Cambridge: Cambridge University Press. [1,2,9]

Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50. [1,2,3,9,10,11,12,13,14]

Koenker, R., and Hallock, K. (2001), "Quantile Regression: An Introduction," *Journal of Economic Perspectives*, 15, 43–56. [1]

Koenker, R., and Machado, J. A. F. (1999), "Goodness of Fit and Related Inference Processes for Quantile Regression," *Journal of the American Statistical Association*, 94, 1296–1310. [2]

Koenker, R., and Portnoy, S. (1987), "L-Estimation for Linear Models," *Journal of the American Statistical Association*, 82, 851–857. [1]

Koenker, R., and Xiao, Z. (2002), "Inference on the Quantile Regression Process," *Econometrica*, 70, 1583–1612. [2]

Komunjer, I., and Vuong, Q. (2010), "Efficient Estimation in Dynamic Conditional Quantile Models," *Journal of Econometrics*, 157, 272–285. [2,4,9]

Kozek, A. S. (2005), "How to Combine M-Estimators to Estimate Quantiles and a Score Function," *Sankhyā*, 67, 277–294. [2]

Lepski, O. V., Mammen, E., and Spokoiny, V. G. (1997), "Optimal Spatial Adaptation to Inhomogeneous Smoothness: An Approach Based on Kernel Estimates With Variable Bandwidth Selectors," *The Annals of Statistics*, 25, 929–947. [16]

Machado, J. A. F., and Parente, P. (2005), "Bootstrap Estimation of Covariance Matrices via the Percentile Method," *The Econometrics Journal*, 8, 70–78. [1]

Mammen, E., Van Keilegom, I., and Yu, K. (2017), "Expansions for Moments of Regression Quantiles With Application to Nonparametric Testing," arXiv no. 1306.6179. [1]

Massart, P. (2007), *Concentration Inequalities and Model Selection* (Vol. 1896), Berlin, Heidelberg: Springer. [6,17]

Nadaraya, E. A. (1964), "Some New Estimates for Distribution Functions," *Theory of Probability and Its Applications*, 9, 497–500. [2,3,5]

Newey, W. K., and McFadden, D. L. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics* (Vol. 4), eds. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier, Chapter 36, pp. 2111–2245. [3]

Newey, W. K., and Powell, J. L. (1990), "Efficient Estimation of Linear and Type I Censored Regression Models Under Conditional Quantile Restrictions," *Econometric Theory*, 6, 295–317. [2,4,5,9]

Otsu, T. (2008), "Conditional Empirical Likelihood Estimation and Inference for Quantile Regression Models," *Journal of Econometrics*, 142, 508–538. [1,2,4,9]

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105–121. [2]

Pollard, D. (2002), *A User's Guide to Measure Theoretic Probability* (Vol. 8), Cambridge: Cambridge University Press. [7]

Portnoy, S. (2012), "Nearly Root-n Approximation for Regression Quantile Processes," *The Annals of Statistics*, 40, 1714–1736. [1,2,7]

Powell, J. L. (1991), *Estimation of Monotonic Regression Models Under Quantile Restrictions*, Cambridge: Cambridge University Press, pp. 357–384. [3]

Ralescu, S. S. (1997), "A Bahadur–Kiefer Law for the Nadaraya Empiric-Quantile Processes," *Theory of Probability and Its Applications*, 41, 296–306. [2]

Roth, C., and Wied, D. (2018), "Estimating Derivatives of Function-Valued Parameters in a Class of Moment Condition Models," University of Mannheim and University of Cologne. [2]

Sheather, S. J., and Marron, J. S. (1990), "Kernel Quantile Estimators," *Journal of the American Statistical Association*, 85, 410–416. [2]

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Boca Raton, FL: CRC/Chapman and Hall. [2,8,9]

Volgushev, S., Chao, S.-K., and Cheng, G. (2019), "Distributed Inference for Quantile Regression Processes," *The Annals of Statistics*, 47, 1634–1662. [2,16]

Whang, Y.-J. (2006), "Smoothed Empirical Likelihood Methods for Quantile Regression Models," *Econometric Theory*, 22, 173–205. [1,2,9]

Xiang, X. (1994), "Bahadur Representation of Kernel Quantile Estimators," *Scandinavian Journal of Statistics*, 21, 169–178. [2]

Zhao, Q. (2001), "Asymptotically Efficient Median Regression in the Presence of Heteroskedasticity of Unknown Form," *Econometric Theory*, 17, 765–784. [2,4,9]