Econometric Theory, **38**, 2022, 497–535.
doi:10.1017/S0266466621000232

NONPARAMETRIC WEIGHTED AVERAGE QUANTILE DERIVATIVE

YING-YING LEE
University of California Irvine

The weighted average quantile derivative (AQD) is the expected value of the partial derivative of the conditional quantile function (CQF) weighted by a function of the covariates. We consider two weighting functions: a known function chosen by researchers and the density function of the covariates that is parallel to the average mean derivative in Powell, Stock, and Stoker (1989, *Econometrica* 57, 1403–1430). The AQD summarizes the marginal response of the covariates on the CQF and defines a nonparametric quantile regression coefficient. In semiparametric single-index and partially linear models, the AQD identifies the coefficients up to scale. In nonparametric nonseparable structural models, the AQD conveys an average structural effect under certain independence assumptions. Including a stochastic trimming function, the proposed two-step estimator is root-n-consistent for the AQD defined by the *entire* support of the covariates. To facilitate tractable asymptotic analysis, a key preliminary result is a new Bahadur-type linear representation of the generalized inverse kernel-based CQF estimator uniformly over the covariates in an *expanding* compact set and over the quantile levels. The weak convergence to Gaussian processes applies to the differentiable nonlinear functionals of the quantile processes.

1. INTRODUCTION

The *weighted average quantile derivative* (AQD) is the weighted expected value of the partial derivatives of the conditional quantile function (CQF), defined as

$$\beta_W(\tau) \equiv \mathbb{E}[\nabla Q(\tau|X)W(X)], \tag{1}$$

where $Q(\tau|X)$ is the τth CQF of the dependent variable Y given the continuous covariates X. The weighting function $W(X)$ defines the AQD that conveys information of the distributional impacts of the covariates on the response variable. Thus, the AQD defines a *nonparametric quantile regression (QR) parameter* that summarizes the marginal effect of X on the τth CQF. Policy makers may specify $W(X)$ for counterfactual analysis to evaluate distributional policy effects (e.g.,

This paper is based on the second chapter of my PhD dissertation. I am grateful to Bruce Hansen and Jack Porter for invaluable comments and guidance. I also thank Chris Tabor, Matias Cattaneo, David Jacho-Chávez, Alexandre Poirier, Emmanuel Guerre, Ingrid van Keilegom, and Efang Kong for helpful comments and discussion. Finally, I want to thank the Editor Peter C.B. Phillips, the Co-Editor Michael Jansson, and two anonymous referees whose comments have significantly improved this paper. Address correspondence to Ying-Ying Lee, Department of Economics, University of California Irvine, 3151 Social Science Plaza, Irvine, CA 92697, USA; e-mail: yingying.lee@uci.edu.

Chernozhukov, Fernández-Val, and Melly, 2013). For $W(X) = f(X)$ the density of the covariates, we estimate the data-driven density weight nonparametrically. We show that the proposed estimators of $\beta_W(\tau)$ are $\sqrt{n}$-consistent and weakly converges to a Gaussian process indexed by τ. It follows that we can conduct joint inference on the Hadamard-differentiable functionals of the conditional or average quantile process, which can be nonlinear functionals. In an intermediate step, we provide a new Bahadur-type linear representation for the generalized inverse kernel-based CQF estimator uniformly over the covariates in an *expanding* compact set and over the quantile levels, which can be of independent interest.

The AQD is a nonparametric alternative to the linear QR by Koenker and Bassett (1978), just as the average mean derivative (AMD), $\mathbb{E}[\nabla \mathbb{E}[Y|X]]$, is a nonparametric alternative to the ordinary least-squares (OLS) estimation. Powell, Stock, and Stoker (1989) introduce the density-weighted AMD $\mathbb{E}[\nabla \mathbb{E}[Y|X]f(X)]$, which has received a lot of attention: Härdle and Stoker (1989), Powell and Stoker (1996), Nishiyama and Robinson (2000), Schafgans and Zinde-Walsh (2010), Cattaneo, Crump, and Jansson (2010, 2013, 2014a, 2014b), and Cattaneo and Jansson (2018), to mention just a few. When the economic theory implies some semiparametric single-index and partially linear models, the AQD identifies the coefficient up to scale, e.g., Chaudhuri, Doksum, and Samarov (1997), Lee (2003), and Hoderlein and Mammen (2009). The AQD gives a simple picture of the impacts of the covariates on the outcome distribution and is more robust against possible extreme values than the mean estimators. In nonparametric nonseparable structural models, the derivative of the CQF has causal interpretation of continuous quantile treatment effects, under certain conditional independence assumptions, e.g., Chesher (2003), Chernozhukov and Hansen (2005), Hoderlein and Mammen (2007), Matzkin (2007), and Sasaki (2015), among others. The AQD is a simple summary statistic for the quantile treatment effects by averaging over the covariates.

The paper is concerned with the estimation and inference of the weighted AQD in (1):

$$\beta_W(\tau) = -\mathbb{E}\left[Q(\tau|X)\left(\nabla W(X) + W(X)\frac{\nabla f(X)}{f(X)}\right)\right], \tag{2}$$

where the equality follows by integration by parts and assuming the covariates have zero density on the boundary. Hence, the weighted AQD can be interpreted as a weighted average CQF. More generally, we focus on the *weighted average quantile response* (AQR)

$$\beta_\phi(\tau) \equiv \mathbb{E}\big[Q(\tau|X)\phi(X)\big]. \tag{3}$$

When the weight $\phi(X) = -\nabla W(X) - W(X)\nabla f(X)/f(X)$, $\beta_\phi(\tau) = \beta_W(\tau)$ is the weighted AQD in (2). We propose a two-step estimator for $\beta_W(\tau)$: The first step is leave-one-out nonparametric kernel-based estimation of the unknown functions $f(x)$, $\nabla f(x)$, and $Q(\tau|X)$. The CQF is estimated by a generalized inverse of the estimated conditional distribution function, $\hat{Q}(\tau|X) \equiv \inf\{y : \hat{F}_Y(y|X) \geq \tau\}$, where $\hat{F}_Y(y|X)$ is a smoothed local constant regression estimator. In the second

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

step, the expectation is replaced by its sample analog involving a *stochastic trimming function* to account for estimating the CQF near the boundary with small density. A stochastic trimming function selects a compact interior support of the covariates by removing the observation i if $\hat{f}_{XY}(X_i, \hat{Q}(\tau|X_i)) < \delta$, where the trimming parameter δ vanishes at an appropriate rate as the sample size grows. Hence, our estimator is consistent for $\beta_\phi(\tau)$ defined by the *entire* support of X.

We present three main results. First, to the best of our knowledge, this is the first paper to provide the limit theory for the weighted AQD defined by the whole support of X, which has zero density on the boundary and can be unbounded. Our limit theorems also cover the case when the estimand is defined by a compact interior support of X using a nonvanishing trimming parameter, commonly used in previous research discussed below. Our tractable approach may be applied to other multistep estimation problems based on a preliminary nonparametric kernel-based estimator, where the stochastic trimming is required but often avoided due to technical complication.

The second result is the *density-weighted* AQD by choosing $W(X) = f(X)$. The density weight inherits the spirit of the AMD in Powell et al. (1989). The density-weighted AQD has a simplified expression in β_ϕ with $\phi(X) = -2\nabla f(X)$ by eliminating the denominator in (2). Consequently, compared with the AQD with a *known* weight in (2), the density-weighted AQD estimator allows the trimming parameter to vanish at a faster rate, so the estimator trims away less observations in finite samples and also assumes weaker smoothness conditions on the distributions. We also provide an optimal bandwidth that minimizes the mean squared error (MSE), using the results of Powell and Stoker (1996).

The third result is a Bahadur-type linear representation of the CQF estimator that is uniform over values of the covariates in a sequence of *expanding* compact interior support and over quantile levels in a compact subset of $(0, 1)$. The new Bahadur-type representation allows us to use the stochastic trimming function to select the interior support of the covariates and the quantile. Hence, it is particularly useful when the CQF is involved in a multistep estimation procedure, for example, the first-price auction in Guerre, Perrigne, and Vuong (2000) and Marmer and Shneyerov (2012) and the quantile correlated random coefficients panel data model in Graham et al. (2018). We also derive the weak convergence of the conditional quantile process estimator.

Now, we discuss our contributions of the above three results to the related literature. The weighted AQD in (1) and (2) are first estimated by Chaudhuri et al. (1997) using local polynomial estimators. Lee (2003) estimates (1) in a partially linear model. Recently, Belloni et al. (2019) develop a nonparametric series framework and perform inference on linear functionals of the CQF, including the AQD with a known weight. We add to the literature by (i) an explicit expression of the first-order bias that could be useful for robust inference in finite samples. (ii) Our asymptotic analysis accounts for the estimation error of the density weight. (iii) Using a stochastic trimming function, we are able to conduct inference on the average quantile process defined by the whole support of X. In contrast,

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

the abovementioned papers all employ a *fixed trimming function* for technical simplification without losing their main focus. Consequently, their estimands are defined by the interior support of the regressors. For the AQD to serve as a nonparametric QR coefficient or in some economic applications, their estimands would be different objects from the AQD defined by the entire support. Our stochastic trimming approach has two advantages of consistency and efficiency over the fixed trimming approach. In a nonparametric model, our estimator reaches the semiparametric efficiency bound (Newey, 1990). In the semiparametric single-index and partially linear models, a trimming function and the weighting function do not affect the consistency of the estimators. But when the optimal weight in terms of efficiency is concerned, a fixed trimming parameter results in efficiency loss, as noted in Lee (2003).

We contribute a new tractable approach to handle the stochastic trimming in asymptotic analysis. Toward this end, we apply the result of the nonparametric kernel-based estimators in Hansen (2008) and Cattaneo et al. (2013). In particular, to account for the entire support of the covariates, we face two challenges: The first issue is the "denominator problem" for estimating the CQF when the density is small near the boundary. Second, we derive the limit theory by plugging a Bahadur-type representation for the CQF estimator into the two-step AQD estimator, which becomes a U-statistic. The linear representation of $\hat{Q}(\tau|X)$ is uniform over X in a compact inner subset of the support, rather than the entire support. The key to overcome these two problems is to incorporate a Bahadur-type representation of $\hat{Q}(\tau|X)$ on an expanding compact interior support with a stochastic trimming function. The smoothed estimator $\hat{F}_Y(y|x)$ utilizes the uniform convergence results on expanding interior support in Hansen (2008) and Cattaneo et al. (2013). The trimming function selects the expanding compact interior support by controlling the lower bound of the joint density $f_{XY}(X, Q(\tau|X))$ converging to zero. Then, we derive a Bahadur representation for the generalized inverse estimator of the CQF uniformly over X in this trimmed interior support at an appropriate rate as the sample grows, depending on a tradeoff between the tail behavior of the distribution and the estimation error from the CQF. Therefore, the trimmed compact interior support, where we have the uniform linear representation for $\hat{Q}(\tau|X)$, is expanding to the entire support as the sample grows.

Finally, our third result contributes a Bahadur representation of the generalized inverse smoothed estimator for the CQF. Bhattacharya and Gangopadhyay (1990) provide a Bahadur representation uniformly over the bandwidth, and Dabrowska (1992) derives the uniformity over the quantiles. For the local polynomial estimator of the CQF, Chaudhuri et al. (1997) and Kong, Linton, and Xia (2010) derive Bahadur representations for uniformity in the covariates X, Qu and Yoon's (2015) result is uniform over the quantile τ, and Guerre and Sabbah (2012) and Fan and Guerre (2016) provide the uniformity in X and τ. To extend their Bahadur representations to uniformity on expanding interior supports, the uniform convergence rate is penalized by the lower bound of the density at a slower rate, as noted in Hansen (2008). As a result, compared with our local constant

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

estimator, using a local linear quantile estimator trims more observations in finite samples.

The paper proceeds as follows. In Section 2, we discuss applications of the AQD. Section 3 introduces the estimators. In Section 4, we first show a uniform linear representation and weak convergence for the nonparametric kernel-based CQF estimation. The AQD estimators are $\sqrt{n}$-consistent and asymptotically normal. We suggest a consistent estimator for the asymptotic covariance matrix and an optimal bandwidth choice. In Section 5, by a simple simulation study, we compare the proposed AQD estimator with the AMD estimator in Powell et al. (1989), the linear QR estimator in Koenker and Bassett (1978), and the OLS for the semiparametric partially linear models. We also implement several bootstrap-based confidence intervals (CIs). All proofs are in the Appendix.

2. APPLICATIONS

We discuss applications of the AQD by starting with a general nonparametric nonseparable structural model. We demonstrate how the AQD captures informative causal features under certain conditional independence assumptions. By imposing further assumptions on the data generating processes (DGPs) and the structural equations, the AQD estimates the coefficients (up to scale) in semiparametric QR models. Another application relates to the counterfactual distribution or decomposition analysis literature. We discuss some earlier work that is most related to ours; more details are in their references therein.

Consider the general setting in Newey and Stoker (1993), $Y = \phi(X, e)$, where e captures the unobserved individual heterogeneity and could be multidimensional. Let $X = (X_1, X_2')'$ and X_1 be conditionally independent of e given X_2. Hoderlein and Mammen (2007) and Sasaki (2015) investigate causal interpretation of the derivative of the CQF, which identifies a weighted average of heterogeneous structural partial effects among the subpopulation of individuals at the conditional quantile of interest.

A common identification strategy assumes that the structural function ϕ is strictly increasing in the scalar unobservable e. Then, the CQF of Y given X identifies the structural function ϕ up to a normalization on $Q_e(\tau|X_2)$,

$$Q(\tau|X) = \phi(X, Q_e(\tau|X)) = \phi(X, Q_e(\tau|X_2)). \tag{4}$$

Therefore, the partial derivative of the CQF with respect to X_1, $\partial Q(\tau|X)/\partial X_1$, identifies the structural derivative, $\partial \phi(X, Q_e(\tau|X_2))/\partial X_1$, which is the causal effect of X_1 while leaving the value of the unobserved variable e unchanged at $Q_e(\tau|X_2)$.

Further assume quantile independence and normalization such that (4) yields $\phi(X, Q_e(\tau)) = \phi(X, \tau)$, which is the τth *quantile treatment response* defined in Chernozhukov and Hansen (2005). It follows that for a nonseparable single-index

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

model $Y = \phi(X'\beta_0, e)$, $\beta_W(\tau)$ identifies the index coefficient β_0 up to scale, e.g., Chaudhuri et al. (1997) and a rank estimator in Khan (2001). Then, the structural function ϕ can be further estimated by a nonparametric QR of Y on the estimated index $X'\hat{\beta}_W(\tau)$.[1]

The partially linear and single-index models relax restrictive parametric assumptions and ease the curse of dimension in nonparametric estimation. For a single-index QR model, $Y = \phi_\tau(X'\beta_\tau) + e_\tau$ and $Q_{e_\tau}(\tau|X) = 0$. The weighted AQD identifies the coefficient β_τ up to scale, $\beta_W(\tau) = \beta_\tau \mathbb{E}[\phi'_\tau(X'\beta_\tau)W(X)]$. For example, Wu, Yu, and Yu (2010) propose an iterative algorithm, and Kong and Xia (2012) propose an adaptive estimation procedure.

The weighted AQD relates to the counterfactual distribution literature by the choice of the weighting function $W(X)$; see Chernozhukov et al. (2013) and the references therein. For example, the policy maker may change the covariate distribution exogenously to some probability density function $f^*(X)$, as in Rothe (2010), and consider a simple counterfactual, "what would the AQD have been if individuals' attributes had been distributed by $f^*(X)$?" By choosing $W(X) = f^*(X)/f(X)$, the weighted AQD $\beta_W(\tau) = \int \nabla Q_Y(\tau|x) f^*(x) dx = \mathbb{E}[\nabla Q_Y(\tau|X) f^*(X)/f(X)] = -\mathbb{E}[Q_Y(\tau|X)\nabla f^*(X)/f(X)]$ is the *counterfactual AQD*. Our estimator for β_W is directly applicable when $W(X)$ is known. The limit theory for an estimated $W(X)$ can be modified using the general results in the Appendix.

3. ESTIMATION

The data consist of n observations $(X_i', Y_i)'$, $i = 1, \ldots, n$, which is an independently and identically distributed random sample from a distribution $F_{XY}(X, Y)$. The τth CQF of Y given X is $Q(\tau|X) \equiv \inf\{y : F_Y(y|X) \geq \tau\}$, where $F_Y(y|X)$ is the conditional cumulative distribution function (CDF) of Y given X.

We propose two-step estimators for three estimands: the *weighted AQR* $\beta_\phi(\tau)$ in (3), the *weighted AQD* $\beta_W(\tau)$ in (2), and the *density-weighted AQD*

$$\beta_f(\tau) = \mathbb{E}[\nabla Q(\tau|X) f(X)] = -2\mathbb{E}[Q(\tau|X)\nabla f(X)]. \tag{5}$$

The first step is leave-one-out nonparametric estimation of the unknown functions. The second step is the sample analog involving a stochastic trimming function $\mathbf{1}\{X_i \in \hat{\mathcal{S}}\}$:

$$\hat{\beta}_\phi(\tau) = \frac{1}{n}\sum_{i=1}^{n} \hat{Q}(\tau|X_i)\phi(X_i)\mathbf{1}\{X_i \in \hat{\mathcal{S}}\}, \tag{6}$$

[1] This specification includes many models as special cases, for an example of a selection model where X_1, X_2, and $Y = X_1'\beta_1 + e_1$ are observed only if the unobserved $Z_2^* = X_2'\beta_2 + e_2 > 0$. Assuming (e_1, e_2) is independent of (X_1, X_2), $Q_Y(\tau|X_1, X_2, Z_2^* > 0) = X_1'\beta_1 + Q_{e_1}(\tau|Z_2^* > 0)$. If X_2 has no variables in common with X_1, then the AQD identifies the structural parameter β_1 and the selection parameter β_2 up to scale. If $X_1'\beta_1$ and $X_2'\beta_2$ are the same, then it is the truncated Tobit model, as discussed in Stoker (1986).

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

$$\hat{\beta}_W(\tau) = -\frac{1}{n}\sum_{i=1}^{n} \hat{Q}(\tau|X_i)\left(\nabla W(X_i) + \frac{\nabla \hat{f}(X_i)}{\hat{f}(X_i)} W(X_i)\right)\mathbf{1}\{X_i \in \hat{\mathcal{S}}\}, \tag{7}$$

$$\hat{\beta}_f(\tau) = -\frac{2}{n}\sum_{i=1}^{n} \hat{Q}(\tau|X_i)\nabla \hat{f}(X_i)\mathbf{1}\{X_i \in \hat{\mathcal{S}}\}, \tag{8}$$

for a quantile level $\tau \in \mathcal{T} = [\varepsilon, 1-\varepsilon]$ for $\varepsilon \in (0, 1/2)$. Following Powell et al. (1989), a more interpretable rescaled density-weighted AQD is defined as $\beta_s \equiv \beta_f / \mathbb{E}[f(X)]$, with a normalized density weight $W(X) = f(X)/\mathbb{E}[f(X)]$. The scaled estimator is $\hat{\beta}_s(\tau) = \hat{\beta}_f(\tau)/\left(n^{-1}\sum_{i=1}^{n}\hat{f}(X_i)\right)$.

Next, we describe each component in the estimators. The leave-one-out kernel estimator for the density function of X at X_i is $\hat{f}(X_i) = (|H_1|(n-1))^{-1}\sum_{j\neq i} K(H_1^{-1}(X_j - X_i))$, where $K(u) = \Pi_{s=1}^{d} k(u_s)$ is a ν_1th-order multivariate product kernel and the bandwidth matrix H_1 is the $d \times d$ identity matrix multiplied by h_1, a positive sequence of n. The covariates X can be normalized by the standard deviations, so that the bandwidths are equal to the same h_1 for all components of X for simplicity.

The CQF is estimated by inverting the estimated conditional CDF, $\hat{Q}(\tau|X) \equiv \inf\{y : \hat{F}_Y(y|X) \geq \tau\}$, where

$$\hat{F}_Y(y|X_i) = \frac{1}{|H|(n-1)}\frac{1}{\hat{f}(X_i)}\sum_{j\neq i} G\left(\frac{y - Y_j}{h_0}\right) K(H^{-1}(X_j - X_i))$$

with a kernel of order ν and a bandwidth matrix H. The indicator function $\mathbf{1}\{Y_j \leq y\}$ for the dependent variable is smoothed by a cumulative kernel $G(z) = \int^z g(t)dt$ with a second-order kernel g and bandwidth h_0.

The CQF estimator by the generalized inverse is monotone in τ by construction. However, the CDF estimator is not increasing in y when we use a bias-reducing or higher-order kernel K. Chernozhukov, Fernández-Val, and Galichon (2010) propose a generic rearrangement method to get a monotonized version of the estimate $\tilde{F}_Y(y|X_i)$, which preserves the same asymptotics as $\hat{F}_Y(y|X_i)$. Then, the CQF can be estimated by $\hat{Q}(\tau|X_i) \equiv \inf_y\{\tilde{F}_Y(y|X_i) \geq \tau\}$ in practice.[2]

The trimming function $\mathbf{1}\{X_i \in \hat{\mathcal{S}}\}$ is defined by a small enough positive trimming parameter δ, which can be a constant or a positive sequence converging to zero, and $\mathcal{S} \equiv \{x : f_{XY}(x, Q(\tau|x)) \geq \delta\}$. To estimate the AQR and the AQD at a particular quantile level τ, let $\hat{\mathcal{S}} \equiv \left\{x : \hat{f}_{XY}(x, \hat{Q}(\tau|x)) \geq \delta\right\}$. The trimming parameter δ can be a constant that defines an interior support of X or a sequence $\delta = \delta_n$ converging to zero that defines a sequence of expanding sets converging to the

[2]Note that the weighted AQD $\beta_W(\tau)$ is a nonparametric object of interest, which is a summary statistic of the marginal effect of X on the CQF. We do not assume that $x'\beta_W(\tau)$ is the CQF and is monotone in τ for all x. However, quantile crossings occur in the single-index and partially linear models in Section 2, which is a fundamental problem for (semi)parametric QR models. In particular, Phillips (2015) shows that quantile crossings are inevitably present with positive probability in quantile predictive regressions.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

entire support $\mathcal{X} \subseteq R^d$.[3] The conditions on the tuning parameters $h, h_0, h_1, \nu, \nu_1, \delta$ are specified in Assumptions 3 and 4 in Section 4.

The derivative can only be calculated for the continuous covariates. When the covariates contain discrete components, the same estimation works for each point in a finite set of the realized values of the discrete components.

4. ASYMPTOTIC PROPERTIES

We provide a Bahadur-type representation for the CQF estimator uniformly in the quantiles and the covariates in Section 4.1. We also show the weak convergence of the conditional quantile process indexed by the quantile, for a given value of the regressors. Section 4.2 presents the limit theories for our estimators of the AQR β_ϕ, the *density-weighted AQD* β_f, and the *weighted AQD* β_W. In Section 4.3, we provide a consistent estimator of the asymptotic covariance matrix. In Section 4.4, we provide an optimal bandwidth choice. Limits are taken as $n \to \infty$ unless otherwise noted. We start with regularity assumptions. The joint density of $(X', Y)'$, denoted by $f_{XY}(X, Y)$, is with respect to the Lebesgue measure on $\mathcal{X} \times \mathcal{Y} \subseteq \mathcal{R}^{d+1}$. Denote the boundary of $\mathcal{X}$ by $\partial\mathcal{X}$.

Assumption 1..

(i) $\mathcal{X} \times \mathcal{Y}$ is convex. $f(x)$ is uniformly bounded above. $\lim_{x\to\partial\mathcal{X}} f(x) = 0$. $\lim_{x\to\partial\mathcal{X}} Q(\tau|x)$ exists for all $\tau \in [\varepsilon, 1-\varepsilon]$, for some small $\varepsilon > 0$.

(ii) For $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $\tau \in [0, 1]$, the partial derivatives of $F_Y(y|x)$ and $Q(\tau|x)$ with respect to x of order p_x are uniformly continuous in x and bounded; $\partial^3 F_Y(y|x)/\partial y^3$ is uniformly continuous in y and bounded.

Assumption 2 (Kernel)**.** (K) The kernel function k is Lipschitz continuous, bounded, symmetric, with convex bounded support, and of order ν, i.e., $\kappa_j \equiv \int x^j k(x)dx = 0$, for $j < \nu$ and $\kappa_\nu \in (0, \infty)$. The first derivative $k'(x)$ is bounded and integrable.

(G) The second-order kernel g is bounded and symmetric. When g has an unbounded support and $\mathcal{Y} \subset \mathcal{R}$, there exist some positive constants $C, L < \infty$ and $m > 4$ such that $|g(u)| \le C|u|^{-m}$, for $|u| > L$.

Assumption 2(K) imposes standard kernel conditions, e.g., Powell et al. (1989) and Hansen (2008). Assumption 2(G) is used to characterize the first-order bias from smoothing the indicator function with $G\big((y - Y_j)/h_0\big)$. The commonly used Gaussian kernel satisfies Assumption 2(G) that restricts the tail behavior of the kernel function g with an unbounded support when Y does not have a full support on $\mathcal{R}$.

[3]For the denominator problem, Robinson (1988), Härdle and Stoker (1989), Lavergne and Vuong (1996), Ichimura and Todd (2007), and Escanciano, Jacho-Chávez, and Lewbel (2014), among others, use a similar stochastic trimming approach by bounding the density of X away from zero. Because we are dealing with an additional problem of estimating the CQF and its uniform linear representation, trimming on $f(X)$ is not sufficient. Similarly, for the series estimator in Belloni et al. (2019), their AQD is defined by a *fixed* interior support where $f_{XY}(x, Q(\tau|x))$ is bounded away from zero.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

4.1. Conditional Quantile Function

We give a condition on the trimming parameter δ in the trimming function $\mathbf{1}\{X_i \in \hat{\mathcal{S}}\}$ via an expanding set $\mathcal{C}_n \equiv \{x : \|x\| \leq c_n\}$, where a positive sequence $c_n \to \infty$ and $\|x\| = |x'x|^{1/2}$. Then, we build on and extend the uniform convergence results of kernel-based estimators on an expanding interior support in Hansen (2008) and Cattaneo et al. (2013) to specify the condition on c_n. Specifically, for any positive sequence $\delta = \delta_n \to 0$, define $\mathcal{S} = \mathcal{S}_n \equiv \{x : \inf_{\tau\in\mathcal{T}} f_{XY}(x, Q(\tau|x)) \geq \delta\}$ that approaches $\mathcal{X}$ as $n \to \infty$. There exists such an expanding set $\mathcal{C}_n$ equal to the convex hull of $\mathcal{S}$. We can show that $\delta = \inf_{\tau\in\mathcal{T}} f_{XY}(\bar{x}, Q(\tau|\bar{x}))$ with $\bar{x}$ on the boundary of $\mathcal{C}_n$, i.e., $\|\bar{x}\| = c_n$. When $\mathcal{X}$ is convex, we can write $\delta = \inf_{x\in\mathcal{C}_n, \tau\in\mathcal{T}} f_{XY}(x, Q(\tau|x))$.[4]

We remark that the convex support condition in Assumption 1(i) is used to conveniently specify the condition on the trimming parameter δ and is for notational simplicity. Nonetheless, the convex support assumption is not uncommon in the literature; for example, it can be implied by other smoothness conditions on $f(x)$, $f_Y(y|x)$, or $Q(\tau|x)$ as in Guerre and Sabbah (2012) and Qu and Yoon (2015), among others. We can allow a nonconvex support by letting Assumption 1 be local following Qu and Yoon (2015) in the sense that the restrictions are on neighborhoods surrounding $\mathcal{X} \times \mathcal{Y}$ rather than on the support of (X', Y) by a slight abuse of notation. Then, our results can be applied to the case when the support is nonconvex and is a union of convex sets.

THEOREM 1 (Bahadur representation). *Let Assumptions 1 and 2 hold with* $p_x \geq \nu$. *Let* $\delta^{-1}\left(\sqrt{\log n/(nh^d)} + h_0^2 + h^\nu\right) \to 0$. *Let* δ *be a constant or a sequence* $\delta = \delta_n = \inf_{\tau\in\mathcal{T}} f_{XY}(\bar{x}, Q(\tau|\bar{x})) \to 0$ *with* $\|\bar{x}\| = c_n \to \infty$ *and* $\limsup_{n\to\infty} \log(c_n)/\log n < \infty$. *Then,*

$$\sup_{\substack{x\in\mathcal{S}\\ \tau\in\mathcal{T}}} \left|\hat{Q}(\tau|x) - Q(\tau|x)\right| = O_p\left(\frac{1}{\delta}\left(\sqrt{\frac{\log n}{nh^d}} + h_0^2 + h^\nu\right)\right).$$

Furthermore, let $\delta^{-1}\sqrt{\log n/(nh^d h_0)} \to 0$. *Then, for any* $\tau \in \mathcal{T}$ *and* $x \in \mathcal{S}$,

$$\hat{Q}(\tau|x) - Q(\tau|x) = \frac{1}{n|H|} \frac{\sum_{i=1}^n K\left(H^{-1}(X_i - x)\right)\left(\tau - G\left(\frac{Q(\tau|x) - Y_i}{h_0}\right)\right)}{f_{XY}(x, Q(\tau|x))} + R_n(\tau, x), \tag{9}$$

[4] We can normalize X such that the support $\mathcal{X}$ is centered at zero without loss of generality. To see the relationship between δ and c_n, let $\mathcal{T} = \{\tau\}$ be a singleton for simplicity. Since $\mathcal{S}$ is compact, $\bar{x}$ on the boundary of the convex hull of $\mathcal{S}$ is also on the boundary of $\mathcal{S}$, and hence, $f_{XY}(\bar{x}, Q(\tau|\bar{x})) \geq \delta$. Suppose to the contrary that the inequality is strict, i.e., $f_{XY}(\bar{x}, Q(\tau|\bar{x})) = a\delta > \delta$ for a constant $a > 1$. By the assumption $\lim_{x\to\partial\mathcal{X}} f(x) = 0$, $\lim_{x\to\partial\mathcal{X}} f_{XY}(x, Q(\tau|x)) = 0$, and hence, $\mathcal{S}$ is a strict subset of $\mathcal{X}$, for any $\delta > 0$. So $\mathcal{X} \cap \mathcal{S}^c$ is not empty. By continuity, there exists $\eta > 0$ such that for $x \in \{x \in \mathcal{X} \cap \mathcal{S}^c : \|x - \bar{x}\| \leq \eta\}$, $|f_{XY}(x, Q(\tau|x)) - f_{XY}(\bar{x}, Q(\tau|\bar{x}))| \leq \delta$. So $f_{XY}(x, Q(\tau|x)) \geq f_{XY}(\bar{x}, Q(\tau|\bar{x})) - \delta = (a-1)\delta > \delta$ that contradicts $x \in \mathcal{X} \cap \mathcal{S}^c$. Therefore, $f_{XY}(\bar{x}, Q(\tau|\bar{x})) = \delta$. When $\mathcal{X}$ is convex, $\mathcal{S} = \mathcal{C}_n$ for n large enough, and hence, $\delta = \inf_{x\in\mathcal{C}_n} f_{XY}(x, Q(\tau|x))$ by construction.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

where the remainder term $R_n(\tau,x)$ satisfies

$$\sup_{\substack{x\in\mathcal{S}\\ \tau\in\mathcal{T}}}|R_n(\tau,x)| = O_p\left(\frac{\log n}{\delta^2 nh^d}\left(\frac{1}{\sqrt{h_0}}+\frac{1}{\delta}\right)\right) + O_p\left(\frac{h^\nu+h_0^2}{\delta^2}\left(\sqrt{\frac{\log n}{nh^d h_0}}+\frac{h^\nu+h_0^2}{\delta}\right)\right). \tag{10}$$

The linear representation is useful for analyzing large sample properties of the final plug-in estimators under different applications. The CQF estimator inherits the uniform convergence rate of the conditional CDF estimator $\hat{F}_Y(y|X)$. The Bahadur representation allows for the optimal rate for estimating the CQF. In the uniform convergence rate of the remainder term R_n, the second part in (10) is of smaller order if the estimator is undersmoothed by assuming $\sqrt{nh^d}(h_0^2+h^\nu)\to 0$.

Theorem 2 below shows the weak convergence of the empirical conditional quantile process $\{x\mapsto \hat{Q}(\tau|x):\tau\in\mathcal{T}\}$. When the interest is to conduct inference on the CQF, Theorem 2 enables the inference method in Fan and Liu (2016), who develop a new CI interval from any conditional quantile process estimator that weakly converges to a Gaussian process.

THEOREM 2 (Weak convergence). *Let the conditions in Theorem 1 hold. Furthermore, let $\sqrt{nh^d}(h_0^2+h^\nu)\to 0$. Then, for any $x\in\mathcal{S}$, $\sqrt{nh^d}\big(\hat{Q}(\cdot|x)-Q(\cdot|x)\big)\Longrightarrow \mathbb{G}(\cdot|x)$ that is a zero-mean Gaussian process $\mathbb{G}(\cdot|x)$ with covariance*

$$\mathrm{Cov}\left(\mathbb{G}(\tau_1|x),\mathbb{G}(\tau_2|x)\right)=\frac{\tau_1(1-\tau_2)}{f(x)\,f_Y\big(Q(\tau_1|x)\big|x\big)\,f_Y\big(Q(\tau_2|x)\big|x\big)}\left(\int k(v)^2dv\right)^d,$$

for any $\tau_1\le\tau_2\in\mathcal{T}$.

4.2. Weighted Average Quantile Response

We first establish asymptotic linearity of the estimator for the weighted AQR with a known weight $\phi(X)$, $\beta_\phi(\tau)=\mathbb{E}\big[Q(\tau|X)\phi(X)\big]$ in (6). Theorem 3 also provides the preliminary results to analyze $\hat{\beta}_f$ and $\hat{\beta}_W$, where the weight is estimated. Let $\partial_k^l g(x)$ denote the lth-order partial derivative of a generic function $g(x)$ with respect to the kth component of x and $\partial_k g(x)\equiv\partial_k^1 g(x)$.

THEOREM 3 (Weighted AQR). *Let the conditions in Theorem 1 hold with $p_x\ge\nu+1$. For a measurable function $\phi:\mathcal{X}\mapsto\mathcal{R}^q$, assume the p_xth-order derivative of $\phi(x)$ to be uniformly continuous and bounded. Let the following conditions hold.*

(i) $\mathbb{E}\big[\|\phi(X)\|^2\left(\inf_{\tau\in\mathcal{T}}f_{XY}(X,Q(\tau|X))\right)^{-2}\big]<\infty$, $nh^{2d}\to\infty$, *and* $\sqrt{n}(h_0^2+h^\nu)\to C\in[0,\infty)$.

(ii) $\sup_{x\in\mathcal{S},\tau\in\mathcal{T}}\|\phi(x)\||R_n(\tau,x)|=o_p\left(n^{-1/2}\right)$, *where the bound of R_n is given in Theorem 1.*

(iii) $\mathbb{E}[\|Q(\tau|X)\phi(X)\|\boldsymbol{1}\{X\notin\mathcal{S}\}]=o(n^{-1/2})$.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

Then, uniformly in $\tau \in \mathcal{T}$,

$$\sqrt{n}\left(\hat{\beta}_\phi(\tau) - \beta_\phi(\tau) - \text{Bias}_\phi(\tau; h, h_0)\right)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\phi(X_i)}{f_Y(Q(\tau|X_i)|X_i)} (\tau - \mathbf{1}\{Y_i \le Q(\tau|X_i)\}) + Q(\tau|X_i)\phi(X_i) - \beta_\phi(\tau) \right)$$
$$+ o_p(1), \textit{where}$$
$$\text{Bias}_\phi(\tau; h, h_0) \equiv \mathbb{E}\left[\frac{-\phi(X)}{f_Y(y|X)} \left(h_0^2 \frac{\kappa_{G2}}{2} \frac{\partial f_Y(y|X)}{\partial y} + \frac{h^\nu \kappa_\nu}{f(X)} \sum_{l=1}^{\nu} \frac{1}{l!\,(\nu - l)!} \sum_{k=1}^{d} \partial_k^l F_Y(y|X) \partial_k^{\nu-l} f(X) \right) \Bigg|_{y=Q(\tau|X)} \right] \tag{11}$$

and $\kappa_{G2} \equiv \int G'(z) z^2 dz$.

The condition (i) in Theorem 3 is due to the U-process theory in Sherman (1994). The bandwidths are large enough to achieve asymptotic linearity. The bandwidths are small enough with possibly higher-order kernels to control the bias to vanish at a rate no slower than $\sqrt{n}$. Consequently, the nonparametric estimations of the density and CQF are undersmoothed, which is conventional in semiparametric estimation.

Our asymptotic theorems apply to both cases using fixed trimming and vanishing trimming approaches. The asymptotic linear representation in Theorem 3 coincides with Chaudhuri et al. (1997), where the weight W serves as a trimming function to define a compact interior of the support of X. We further characterize the first-order bias from the CQF estimate, which vanishes at a faster rate under the condition $\sqrt{n}(h_0^2 + h^\nu) \to 0$.

The trimming parameter δ vanishes to zero at a rate specified by the conditions (ii) and (iii) in Theorem 3. The condition (ii) ensures δ to be large enough to control the sampling variation of estimating the CQF. The condition (iii) controls the trimming bias that depends on the tail behavior of the distribution f_{XY} and ensures that δ is small enough for the estimator to approach the entire support. The trimming parameter δ trades off the estimation variance and the trimming bias,[5] while the bandwidths h and h_0 trade off the variance and bias of the preliminary estimators.

We use a standard approach to derive the limit theory by plugging a Bahadur-type representation for the CQF estimator into the two-step estimator, which becomes a U-statistic. The linear representation of $\hat{Q}(\tau|X)$ given in Theorem 1 has the joint density $f_{XY}(x, Q(\tau|x))$ in the denominator. Heuristically, we control the remainder terms in the linear representation of the final estimator to be of smaller

[5] Similar assumptions have been used in Lavergne and Vuong (1996), Ichimura and Todd (2007), and Khan and Tamer (2010), for example.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

order based on $1/f_{XY}(x, Q(\tau|x)) < 1/\delta$. In a semiparametric problem that involves the CQF, the conditional density $f_Y(Q(\tau|x)|x)$ is commonly assumed to be bounded away from zero for identification or for technical simplification. Assumption 1 implies that there exist some constants δ and c such that $\{x : f_{XY}(x, Q(\tau|x)) > \delta\} = \{x : f_Y(Q(\tau|x)|x) > c\}$. Therefore, using the conditional density, $f_Y(Q(\tau|x)|x)$ is equivalent to using the joint density $f_{XY}(x, Q(\tau|x))$ in our asymptotic analysis. Exploiting the assumption $f_Y(Q(\tau|x)|x) > c$ would not improve the large sample properties by our approach.

Next, we present the asymptotic properties for the density-weighted AQD with an estimated weight $\nabla\hat{f}(X)$ in (8) $\hat{\beta}_f(\tau)$ and the scaled AQD $\hat{\beta}_s(\tau) = \hat{\beta}_f(\tau)/\left(n^{-1}\sum_{i=1}^n \hat{f}(X_i)\right)$. Assumption 3 gives specific conditions for the tuning parameters for $\hat{\beta}_f$ and are sufficient for the conditions in Theorem 3. Let $B_n \equiv \{X : f_{XY}(X, Q(\tau|X)) < \delta\}$. When uniformity over $\tau \in \mathcal{T}$ is considered, let $B_n \equiv \{X : \sup_{\tau\in\mathcal{T}} f_{XY}(X, Q(\tau|X)) < \delta\}$.

Assumption 3 (Bandwidth—$\hat{\beta}_f$)**.** The positive sequences h, h_1, h_0, δ satisfy $\sqrt{n}(h_0^2 + h^\nu + h_1^{\nu_1}) \to C \in [0, \infty)$, $\delta^4 nh^{2d}h_0 \to \infty$, $\delta^6 nh^{2d} \to \infty$, $nh_1^{2d+2} \to \infty$, and $\delta^2 nh^d h_1^{d+2} \to \infty$. The trimming parameter $\delta = \delta_n$ satisfies $\int_{B_n} \left\|Q(\tau|X)\nabla f(X)\right\| f(X)dX = o(n^{-1/2})$.

THEOREM 4 (Density-weighted AQD). *Let the conditions in Theorem 1 hold with $p_x \geq \max\{\nu+1, \nu_1+2\}$ and $\phi(X) = -2\nabla f(X)$. Let Assumption 3 hold. Define the influence function*

$$r_f(Z_i;\tau) \equiv \frac{2\nabla f(X_i)}{f_Y(Q(\tau|X_i)|X_i)}\Big(\mathbf{1}\{Y_i \leq Q(\tau|X_i)\} - \tau\Big) + 2f(X_i)\nabla Q(\tau|X_i) - 2\beta_f(\tau).$$

1. *Then,* $\sqrt{n}\left(\hat{\beta}_f(\tau) - \beta_f(\tau) - \mathrm{Bias}_f(\tau; h, h_0, h_1)\right) = n^{-1/2}\sum_{i=1}^n r_f(Z_i;\tau) + o_p(1)$, *uniformly in* $\tau \in \mathcal{T}$, *where* $\mathrm{Bias}_f(\tau; h, h_0, h_1) \equiv -2h_1^{\nu_1}\kappa_{\nu_1}(\nu_1!)^{-1}\sum_{k=1}^d \mathbb{E}\left[Q(\tau|X)\partial_k^{\nu_1}\nabla f(X)\right] + \mathrm{Bias}_\phi(\tau; h, h_0)$ *and* $\mathrm{Bias}_\phi(\tau; h, h_0)$ *is defined in (11) with* $\phi(X) = -2\nabla f(X)$.
2. *Let* $\sqrt{n}(h_0^2 + h^\nu + h_1^{\nu_1}) \to 0$ *such that* $\mathrm{Bias}_f(\tau; h, h_0, h_1) = o(n^{-1/2})$. *Then, the empirical process indexed by* $\tau \in \mathcal{T}$ *converges weakly to a zero-mean Gaussian process* $\sqrt{n}\left(\hat{\beta}_f(\cdot) - \beta_f(\cdot)\right) \Longrightarrow \mathbb{G}_f(\cdot)$ *with covariance*

$$\begin{aligned}&\mathrm{Cov}\left(\mathbb{G}_f(\tau_1), \mathbb{G}_f(\tau_2)\right)\\ &\quad= 4\tau_1(1-\tau_2)\mathbb{E}\left[\frac{\nabla f(X)\nabla f(X)'}{f_Y(Q(\tau_1|X)|X)f_Y(Q(\tau_2|X)|X)}\right]\\ &\quad+ 4\mathbb{E}\left[\left(f(X)\nabla Q(\tau_1|X) - \beta_f(\tau_1)\right)\left(f(X)\nabla Q(\tau_2|X) - \beta_f(\tau_2)\right)\right],\end{aligned}$$

for any $\tau_1 \leq \tau_2 \in \mathcal{T}$. *For the scaled density-weighted AQD,* $\sqrt{n}\left(\hat{\beta}_s(\cdot) - \beta_s(\cdot)\right) = n^{-1/2}\sum_{i=1}^n \left(r_f(Z_i;\cdot)/\mathbb{E}[f(X)] - 2\left(W(X_i) - 1\right)\beta^*(\cdot)\right) + o_p(1) \Longrightarrow \mathbb{G}^*(\cdot)$ *that is a*

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

zero-mean Gaussian process with covariance

$$\begin{aligned}\operatorname{Cov}\left(\mathbb{G}^*(\tau_1),\mathbb{G}^*(\tau_2)\right) &= 4\tau_1(1-\tau_2)\mathbb{E}\left[\frac{\nabla W(X)\nabla W(X)'}{f_Y(Q(\tau_1|X)|X)f_Y(Q(\tau_2|X)|X)}\right]\\ &\quad+4\mathbb{E}\Big[\left(W(X)\nabla Q(\tau_1|X)-\beta^*(\tau_1)+W(X)-1\right)\\ &\quad\times\left(W(X)\nabla Q(\tau_2|X)-\beta^*(\tau_2)+W(X)-1\right)\Big],\end{aligned}$$

for any $\tau_1\le\tau_2\in\mathcal{T}$.

Now, we consider the general weighted AQD $\hat{\beta}_W(\tau)$ in (7). Assumption 4 below gives specific conditions for the tuning parameters for $\hat{\beta}_W$.

Assumption 4 (Bandwidth—$\hat{\beta}_W$)**.** The positive sequences h,h_1,h_0, δ satisfy $\sqrt{n}(h_0^2+h^\nu+h_1^{\nu_1})\to C\in[0,\infty)$, $\delta^6nh^{2d}h_0\to\infty$, $\delta^8nh^{2d}\to\infty$, $nh_1^{2d+2}\to\infty$, $\delta^4nh^dh_1^{d+2}\to\infty$, and $\delta^6nh^dh_1^d\to\infty$. The trimming parameter $\delta=\delta_n$ satisfies $\int_{B_n}\left\|Q(\tau|X)\nabla\left(W(X)f(X)\right)\right\|dX=o(n^{-1/2})$.

THEOREM 5 (Weighted AQD). *Let the conditions in Theorem 1 hold with* $p_x\ge\max\{\nu+1,\nu_1+2\}$ *and* $\phi(X)=-\nabla\left(W(X)f(X)\right)/f(X)$. *Let Assumption 4 hold. Assume the* (ν_1+1)*th-order derivative of* $W(X)$ *to be uniformly continuous and bounded. Assume* $\mathbb{E}\left[W(X)^2/f(X)^2\right]<\infty$ *and* $\mathbb{E}\left[W(X)^2(\nabla f(X))^2/f(X)^4\right]<\infty$.

1. *Then,* $\sqrt{n}\left(\hat{\beta}_W(\tau)-\beta_W(\tau)-\operatorname{Bias}_w(\tau;h,h_0,h_1)\right)=n^{-1/2}\sum_{i=1}^n r_w(Z_i)+o_p(1)$, *uniformly in* $\tau\in\mathcal{T}$, *where the influence function is*

$$\begin{aligned}r_w(Z_i)\equiv&\frac{\nabla\left(W(X_i)f(X_i)\right)}{f_Y(Q(\tau|X_i)|X_i)f(X_i)}\left(\mathbf{1}\{Y_i\le Q(\tau|X_i)\}-\tau\right)\\&+W(X_i)\nabla Q(\tau|X_i)-\beta_W(\tau)\end{aligned}$$

   and $\operatorname{Bias}_w(\tau;h,h_0,h_1)\equiv\operatorname{Bias}_\phi(\tau;h,h_0)+h_1^{\nu_1}\kappa_{\nu_1}(\nu_1!)^{-1}\sum_{k=1}^d\mathbb{E}\big[Q(\tau|X)W(X)/f(X)^2\times\left(\nabla f(X)\partial_k^{\nu_1}f(X)-f(X)\partial_k^{\nu_1}\nabla f(X)\right)\big]$, *where* $\operatorname{Bias}_\phi(\tau;h,h_0)$ *is defined in (11) with* $\phi(X)=-\nabla\left(W(X)f(X)\right)/f(X)$.
2. *Let* $\sqrt{n}(h_0^2+h^\nu+h_1^{\nu_1})\to0$ *such that* $\operatorname{Bias}_w(\tau;h,h_0,h_1)=o(n^{-1/2})$. *Then, the empirical process indexed by* $\tau\in\mathcal{T}$ *converges weakly to a zero-mean Gaussian process* $\sqrt{n}\left(\hat{\beta}_W(\cdot)-\beta_W(\cdot)\right)\Longrightarrow\mathbb{G}_w(\cdot)$ *with covariance*

$$\begin{aligned}&\operatorname{Cov}\left(\mathbb{G}_w(\tau_1),\mathbb{G}_w(\tau_2)\right)\\&\quad=\tau_1(1-\tau_2)\mathbb{E}\left[\frac{\nabla\left(W(X)f(X)\right)\nabla\left(W(X)f(X)\right)'}{f_{XY}(X,Q(\tau_1|X))f_{XY}(X,Q(\tau_2|X))}\right]\\&\qquad+\mathbb{E}\left[\left(W(X)\nabla Q(\tau_1|X)-\beta_W(\tau_1)\right)\left(W(X)\nabla Q(\tau_2|X)-\beta_W(\tau_2)\right)\right],\end{aligned}$$

   for any $\tau_1\le\tau_2\in\mathcal{T}$.

Our theorems show that the proposed estimators are asymptotically linear and weakly converge to Gaussian processes. It is worth noting the difference in

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

estimating the density-weighted AQDβ_f in Theorem 4 and the general weighted AQD β_W in Theorem 5. First, estimating the density weight contributes an additional term $f(X)\nabla Q(\tau|X) - \beta_f(\tau)$. Thus, the estimation error of the density function is not ignorable to perform correct inference. Second, because the density appears in the denominator in β_W, the trimming bound δ vanishes at a slower rate. Thus, $\hat{\beta}_W$ trims more observations in finite samples than $\hat{\beta}_f$ does. Third, the smoothness conditions are weaker for $\hat{\beta}_f$ than for $\hat{\beta}_W$.

Remark 1 (Efficiency)**.** Since our estimands are explicit functions of the distribution, Newey (1990) implies that these nonparametric estimators reach the semiparametric efficiency bounds when the distribution is unrestricted. That is, the influence function of any asymptotically linear and regular estimator for our estimand is unique and hence efficient. It follows that our estimators reach the efficiency bounds of the weighted AQR and the (density-)weighted AQD, respectively. Other nonparametric estimations for the first-step unknown functions, such as series or local polynomial, will give the same asymptotic distribution.[6]

Remark 2 (Choice of tuning parameters)**.** Although the tail distribution conditions are not testable, we provide an example that satisfies the condition $\limsup_{n\to\infty} \log(c_n)/\log n < \infty$ in Theorem 1. When the tail of the joint distribution of (X, Y) decays at an exponential rate, we can choose $c_n \propto (\log n)^q$, for some $q > 0$, and $\delta = \delta_n \propto n^{-b}$.[7]

An alternative set of sufficient conditions for the nonparametric tuning parameters in Theorems 4 and 5 is to let the positive sequences vanish at a polynomial rate, $h \propto n^{-a}$, $h_1 \propto n^{-c}$, $h_0 \propto n^{-\eta}$, and $\delta \propto n^{-b}$, for some positive constants, a, b, c, η:

For $\hat{\beta}_f$, choose $\nu > \frac{4d}{3}$, $a \in [\frac{1}{2\nu}, \frac{3}{8d})$, $\nu_1 > \max\left\{d+1, \frac{d+2}{2-2ad}\right\}$, $c \in \left[\frac{1}{2\nu_1}, \min\left\{\frac{1}{2d+2}, \frac{1-ad}{d+2}\right\}\right)$, $\eta \in \left[\frac{1}{4}, 1-2ad\right)$, and $b < \min\left\{\frac{1}{4}(1-2ad-\eta), \frac{1}{2}(1-ad-c(d+2)), \frac{1}{6}(1-2ad)\right\}$.

For $\hat{\beta}_W$, choose $\nu > \frac{4d}{3}$, $a \in [\frac{1}{2\nu}, \frac{3}{8d})$, $\nu_1 > \max\left\{d+1, \frac{d+2}{2-2ad}\right\}$, $c \in \left[\frac{1}{2\nu_1}, \min\left\{\frac{1}{2d+2}, \frac{1-ad}{d+2}\right\}\right)$, $\eta \in [\frac{1}{4}, 1-2ad)$, and $b < \min\left\{\frac{1}{6}(1-2ad-\eta), \frac{1}{4}(1-ad-c(d+2)), \frac{1}{6}(1-ad-cd), \frac{1}{8}(1-2ad)\right\}$.

These sufficient conditions suggest an upper bound of the convergence rate of the tuning parameter $\delta \propto n^{-b}$, i.e., δ cannot be too small, so that we can control the first-step estimation error. On the other hand, the condition

[6] Specifically, Newey and Stoker (1993) calculate the efficiency bounds for the weighted average derivative for general loss functions, including conditional mean and quantiles, where the weighting function is a known function. By proceeding as in the proof of Theorem 3.1 in Newey and Stoker (1993), we can calculate the efficiency bounds for the density-weighted average quantile/mean derivatives where the density weight is estimated. We can verify that the estimators proposed in this paper and in Powell et al. (1989) are semiparametrically efficient, as implied by the result in Newey (1990). Since the proof closely follows Theorem 3.1 in Newey and Stoker (1993), we do not repeat the details to save space.

[7] Suppose the joint distribution of $Z = (X', Y)'$ to be proportional to $e^{-\|z\|^p}$, for some $p > 0$. The bandwidth assumption requires the trimming parameter δ to be bounded above by $\inf_{\|x\| \le c_n} f_{XY}(x, Q(\tau|x)) \propto e^{-(\log n)^{qp}}$ that is larger than n^{-b} by letting $qp < 1$. A smaller p results in a larger q, meaning that we could use more observations when f_{XY} has a fatter tail.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

$\int_{B_n} \|Q(\tau|X)\nabla f(X)\| f(X)dX = o(n^{-1/2})$ in Assumptions 3 and 4 suggests that δ cannot be too big, so that the trimming bias is of smaller order. To see these conditions are feasible in practice, we give a set of tuning parameters for $\hat{\beta}_f$ in our Monte Carlo simulations.

Remark 3 (Functionals of the weighted average quantile)**.** Once the weak convergence of the quantile process is established in the above theorems, we can extend the results to the Hadamard-differentiable functionals of the quantile process $\Gamma(\beta)$, which can be nonlinear functionals of the distributions. The functional delta method, e.g., Theorem 20.8 in van der Vaart (2000), implies the limit distribution and uniform inference on functionals of the quantile process. For example, the interquantile change $\Gamma(\beta) = \beta(.75) - \beta(.25)$ or the Average Response $\Gamma(\beta) = \int_{\tau_1}^{\tau_2} \beta(\tau)d\tau$, for $0 < \tau_1 < \tau_2 < 1$.

We may estimate the weighted Aggregate Response $\int_0^1 \beta_\phi(\tau)d\tau$ over the entire quantile range $[0, 1]$ by trimming on the quantile levels such that $[\varepsilon, 1-\varepsilon]$ expands to $[0, 1]$. Specifically, it is sufficient to let $\varepsilon = \varepsilon_n = o(n^{-1/2})$.[8] This result may be applicable to the first-price auctions in Marmer and Shneyerov (2012).

Remark 4 (Alternative small bandwidth asymptotics and bootstrap)**.** There are recently developed resampling methods for two-step semiparametric estimators. Cattaneo and Jansson (2018) develop an alternative asymptotic theory to the conventional empirical process theory that relies on the usual stochastic equicontinuity condition and is used in this paper. They allow for low precision of the first-step kernel-based estimators due to a small bandwidth and account for the resulting undersmoothing bias. They show that some nonparametric bootstrap methods automatically correct for such bias. In our simulation study, we examine the robustness of the nonparametric bootstrap as well as our normal distributional approximation, with respect to bandwidth choice.

4.3. Asymptotic Covariance Matrix

An asymptotically pivotal test statistic, a CI, or the corresponding hypothesis test can be constructed by a studentized version of the estimator using Slusky's theorem with a consistent covariance matrix estimator. The covariance matrix could be consistently estimated as the sample variance of uniformly consistent estimators of the influence function. We provide a covariance matrix estimator that is composed of preliminary estimators already used in the primary AQD estimator. So we do not need additional estimation for the asymptotic covariance, such as estimating the derivative of the CQF $\nabla Q(\tau|X_i)$.

[8] Note that the minimum and maximum of the quantile levels ε and $1-\varepsilon$ do not enter the uniform convergence rate of the remainder term in the Bahadur representation in (9). The condition for the convergence rate of ε only depends on the tail distributions by $\int_{[0,\varepsilon)\cup[1-\varepsilon,1]} \mathbb{E}[Q(\tau|X)\phi(X)]d\tau = o_p(n^{-1/2})$. Thus, assuming uniform bounded $Q(\tau|X)\phi(X)$, it suffices to let $\varepsilon = \varepsilon_n = o(n^{-1/2})$.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

We utilize the projection structure of the U-statistic, following Härdle and Stoker (1989) for the AMD. Define

$$\hat{r}_{Ii}(\tau) = \frac{\hat{\phi}(X_i)}{\hat{f}_Y(\hat{Q}(\tau|X_i)|X_i)}\big(\mathbf{1}\{Y_i \le \hat{Q}(\tau|X_i)\} - \tau\big)\mathbf{1}\Big\{X_i \in \hat{\mathcal{S}}\Big\},$$

$$\hat{r}_{\mathit{II}i}(\tau) = \frac{-1}{n-1}\sum_{j\neq i}\frac{1}{h_1^{d+1}}\nabla K\big(H^{-1}(X_i - X_j)\big) \times \Big(\hat{\psi}(X_i;\tau) - \hat{\psi}(X_j;\tau)\mathbf{1}\Big\{X_j \in \hat{\mathcal{S}}\Big\}\Big)\mathbf{1}\Big\{X_i \in \hat{\mathcal{S}}\Big\},$$

$$\hat{r}_{\mathit{III}i}(\tau) = \frac{-1}{n-1}\sum_{j\neq i}\frac{1}{h_1^{d}}K\big(H^{-1}(X_i - X_j)\big) \times \Big(\hat{\gamma}(X_i;\tau) + \hat{\gamma}(X_j;\tau)\mathbf{1}\Big\{X_j \in \hat{\mathcal{S}}\Big\}\Big)\mathbf{1}\Big\{X_i \in \hat{\mathcal{S}}\Big\},$$

where $\hat{\phi}(X_i)$, $\hat{\psi}(X_i;\tau)$, and $\hat{\gamma}(X_i;\tau)$ are defined specifically for $\hat{\beta}_f$ and $\hat{\beta}_W$ as follows.

We estimate the asymptotic covariance of $\hat{\beta}_f$ by

$$\widehat{\mathrm{Cov}}\big(\mathbb{G}_f(\tau_1), \mathbb{G}_f(\tau_2)\big) = \frac{1}{n}\sum_{i=1}^{n}\hat{r}_f(Z_i;\tau_1)\hat{r}_f(Z_i;\tau_2)' - \bar{r}(\tau_1)\bar{r}(\tau_2)',$$

for $\tau_1 \le \tau_2 \in \mathcal{T}$, where $\hat{r}_f(Z_i;\tau) = \hat{r}_{Ii}(\tau) + \hat{r}_{\mathit{II}i}(\tau) - \hat{\beta}_f(\tau)$, $\bar{r}(\tau) = n^{-1}\sum_{i=1}^{n}\hat{r}_f(Z_i;\tau)$, and letting $\hat{\phi}(X) = 2\nabla\hat{f}(X)$ in $\hat{r}_{Ii}(\tau)$ and $\hat{\psi}(X;\tau) = 2\hat{Q}(\tau|X)$ in $\hat{r}_{\mathit{II}i}(\tau)$.

We estimate the asymptotic covariance of $\hat{\beta}_w$ by

$$\widehat{\mathrm{Cov}}\left(\mathbb{G}_w(\tau_1), \mathbb{G}_w(\tau_2)\right) = \frac{1}{n}\sum_{i=1}^{n}\hat{r}_w(Z_i;\tau_1)\hat{r}_w(Z_i;\tau_2)' - \bar{r}_w(\tau_1)\bar{r}_w'(\tau_2),$$

where $\hat{r}_w(Z_i;\tau) = \hat{r}_{Ii}(\tau) + \hat{r}_{\mathit{II}i}(\tau) + \hat{r}_{\mathit{III}i}(\tau) - \hat{\beta}_w(\tau)$, $\bar{r}_w(\tau) = n^{-1}\sum_{i=1}^{n}\hat{r}_w(Z_i;\tau)$, and letting $\hat{\phi}(X) = \nabla(W(X)\hat{f}(X))/\hat{f}(X)$ in $\hat{r}_{Ii}(\tau)$, $\hat{\psi}(X;\tau) = \hat{Q}(\tau|X)W(X)/\hat{f}(X)$ in $\hat{r}_{\mathit{II}i}(\tau)$, and $\hat{\gamma}(X;\tau) = \hat{Q}(\tau|X)W(X)\nabla\hat{f}(X)/\hat{f}(X)^2$ in $\hat{r}_{\mathit{III}i}(\tau)$. These preliminary estimators of $f(X_i), \nabla f(X_i)$, $Q(\tau|X_i)$, and $f_Y(Q(\tau|X_i)|X_i)$ are already used in the primary estimators, so we do not need additional estimation.

THEOREM 6. *Let Assumptions 1–4 hold. Then, for any* $\tau_1 \le \tau_2 \in \mathcal{T}$, $\widehat{\mathrm{Cov}}\big(\mathbb{G}_f(\tau_1), \mathbb{G}_f(\tau_2)\big)$ *and* $\widehat{\mathrm{Cov}}\left(\mathbb{G}_w(\tau_1), \mathbb{G}_w(\tau_2)\right)$ *are consistent for* $\mathrm{Cov}(\mathbb{G}_f(\tau_1), \mathbb{G}_f(\tau_2))$ *and* $\mathrm{Cov}(\mathbb{G}_w(\tau_1), \mathbb{G}_w(\tau_2))$, *respectively.*

The influence function of the scaled AQD estimator $\hat{\beta}_s \equiv \hat{\beta}_f/\hat{\alpha}$ with $\hat{\alpha} = n^{-1}\sum_{i=1}^{n}\hat{f}(X_i)$ can be estimated by $\hat{r}_{si}(\tau) \equiv \big(\hat{r}_f(Z_i;\tau) - 2\big(\hat{f}(X_i) - \hat{\alpha}\big)\hat{\beta}_s(\tau)\big)\big/\hat{\alpha}$. Then, the asymptotic covariance matrix of the scaled AQD estimator can be estimated by $n^{-1}\sum_{i=1}^{n}\hat{r}_{si}(\tau_1)\hat{r}_{si}'(\tau_2) - \bar{r}_s(\tau_1)\bar{r}_s'(\tau_2)$, where $\bar{r}_s(\tau) = n^{-1}\sum_{i=1}^{n}\hat{r}_{si}(\tau)$, for any $\tau_1 \le \tau_2 \in \mathcal{T}$.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

4.4. Optimal Bandwidth Choice

We consider a linear combination of the density-weighted AQD $a'\hat{\beta}_f$, where $a \in \mathcal{R}^d$ and $a'a = 1$. For example, when $a = (1, 0, \ldots, 0)'$, $a'\hat{\beta}_f$ estimates the density-weighted AQD with respect to the first component of X_i. We modify the estimator proposed by Powell and Stoker (1996) for the optimal bandwidth h in K that minimizes the leading terms of the asymptotic MSE of $a'\hat{\beta}_f$. We can apply this approach for $\hat{\beta}_\phi$ and $\hat{\beta}_W$. We may also extend this approach to choose the trimming parameter δ and h_0.

The optimal bandwidth of $a'\hat{\beta}_f$ is estimated by $\hat{h}_{opt} = \left(d\hat{\mathsf{V}}/\left(\nu\hat{\mathsf{B}}^2\right)\right)^{1/(2\nu+d)} n^{-2/(2\nu+d)}$, where $\hat{\mathsf{V}}$ estimates the leading variance associated with the bandwidth h and $\hat{\mathsf{B}}$ estimates the leading bias. As the first-order variance of $\hat{\beta}_f$ is $O(n^{-1/2})$, by choosing this optimal bandwidth h_{opt}, the bias is first-order asymptotically negligible. In addition, the first-step estimators are undersmoothed.

Define $\hat{\mathsf{V}} = h_v^d (n(n-1))^{-1} \sum_{i=1}^n \sum_{j \neq i} \hat{p}(Z_i, Z_j; \lambda)^2$, where

$$\begin{aligned}\hat{p}(Z_i, Z_j; \lambda) \equiv & \left(\hat{Q}(\tau|X_j) - \hat{Q}(\tau|X_i)\right) a' \nabla K\left(\frac{X_i - X_j}{h_v}\right) h_v^{-(d+1)} \mathbf{1}_i \\ & - K\left(H_v^{-1}(X_i - X_j)\right) h_v^{-d} \mathbf{1}_i \\ & \times \left(\frac{a'\nabla \hat{f}(X_i)}{\hat{f}_{XY}(X_i, \hat{Q}(\tau|X_i))} \left(\tau - G\left(\frac{\hat{Q}(\tau|X_i) - Y_j}{h_0}\right)\right) \right. \\ & \left. + \frac{a'\nabla \hat{f}(X_j)}{\hat{f}_{XY}(X_j, \hat{Q}(\tau|X_j))} \left(\tau - G\left(\frac{\hat{Q}(\tau|X_j) - Y_i}{h_0}\right)\right) \right)\end{aligned}$$

and the preliminary estimators $\hat{Q}(\tau|X)$, $\nabla\hat{f}(X)$, and $\hat{f}_{XY}$ use a νth-order kernel with bandwidth $h = h_1 = h_v$. We consider a fixed trimming function $\mathbf{1}_i$ with a constant δ, to simplify the application of the results in Powell and Stoker (1996).

Consider estimating the leading bias of $\hat{\beta}_f$. For a positive constant $u \neq 1$ and a preliminary bandwidth $h = h_1 = h_b$, let $\hat{\beta}_{f,h_b}$ and $\hat{\beta}_{f,uh_b}$ be the estimators using the bandwidths h_b and uh_b, respectively. Let $\nu = \nu_1$ for simplicity. Define $\hat{\mathsf{B}} = a'\left(\hat{\beta}_{f,uh_b} - \hat{\beta}_{f,h_b}\right)/\left((uh_b)^\nu - h_b^\nu\right)$.

COROLLARY 1. *Let the conditions in Theorem 4 hold.*

(i) *Then, the bandwidth that minimizes the leading terms associated with h in the asymptotic MSE of* $\hat{\beta}_f$ *is* $h_{opt} = \left(d\mathsf{V}/\left(\nu\mathsf{B}^2\right)\right)^{1/(2\nu+d)} n^{-2/(2\nu+d)}$, *where* $\mathsf{V} = \int \mathbb{E}\left[(\nabla Q(\tau|X)u)^2 f(X)\right](a'\nabla K(u))^2 du + 2\tau(1-\tau)\mathbb{E}\left[(a'\nabla f(X))^2 f(X) f_{XY}(X, Q(\tau|X))^{-2}\right] \times \int K(u)^2 du$ *with* $u \in \mathcal{R}^d$, *and* $\mathsf{B} \equiv 2\kappa_\nu \sum_{k=1}^d a_k \mathbb{E}\left[\partial_k f(X) f_{XY}(X, Q(\tau|X))^{-1} \sum_{l=1}^\nu (l!(\nu-l)!)^{-1} \partial_k^l F_Y(Q(\tau|X)|X) \partial_k^{\nu-l} f(X) - (\nu!)^{-1} Q(\tau|X) \partial_k^\nu \nabla f(X)\right]$.

(ii) *Furthermore, let* $h_v, h_b \to 0$, $nh_v^{3d} \to \infty$, *and* $nh_b^{2\nu+d} \to \infty$. *Then,* $\hat{\mathsf{V}}$ *and* $\hat{\mathsf{B}}$ *are consistent estimators of* V *and* B, *respectively, and* $\hat{h}_{opt} - h_{opt} = o_p(n^{-2/(2\nu+d)})$.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

We may consider a further bias correction using the above simple bias estimator, i.e., $\hat{\beta}_f - h^v \hat{\mathsf{B}}$. Such robust bias-corrected inference may allow for a wider range of bandwidth choices h in practice, as discussed in Calonico, Cattaneo, and Farrell (2018), and is left for future research.

5. MONTE CARLO SIMULATIONS

We compare the finite-sample performance of our scaled density-weighted AQD estimator $\hat{\beta}_s$ with the AMD in Powell et al. (1989), the conventional Koenker and Bassett (1978) linear QR (labeled by KB), and the OLS. We consider partially linear models with homogenous and heterogenous errors for the DGPs. Both AMD and AQD identify the coefficient of interest in the partially linear model. The linear OLS and KB estimators suffer from misspecification. We find that the scaled

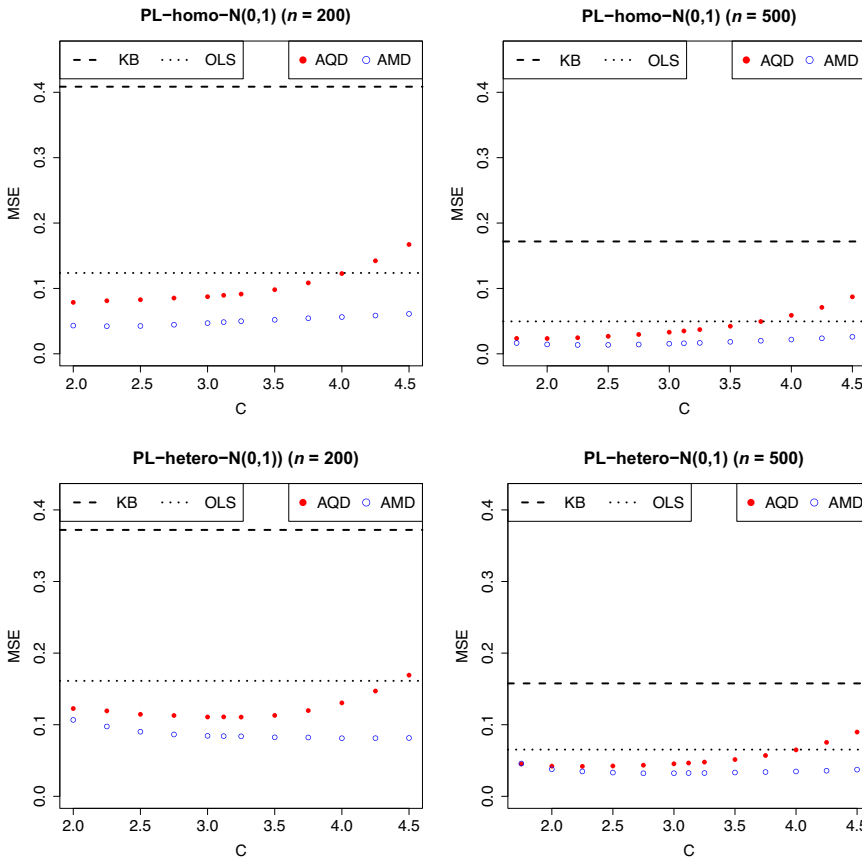

FIGURE 1. Partially linear model with $\mathcal{N}(0,1)$ error.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

density-weighted AQD estimator $\hat{\beta}_s$ outperforms the AMD when the outcome distribution has fat tails. The results are rather robust with a range of bandwidths.

We consider four DGPs that are modified from the experiments in Lee (2003).

1. Partially linear model with homoscedastic error (PL-homo):
   $Y = X_1 + X_2 + 30\exp(-X_1^2)/\sqrt{2\pi} + e$.
2. Partially linear model with heteroskedastic error (PL-hetero):
   $Y = X_1 + X_2 + 30\exp(-X_1^2)/\sqrt{2\pi} + 2\exp((X_1 + X_2)/4)e$.

We consider two error distributions: $e \sim \mathcal{N}(0,1)$ and $e \sim t(2)$ for a fat-tailed distribution. The regressors X_1 and X_2 are jointly normal with mean zero, variance one, and covariance 0.5. Thus, the regressors have unbounded support. The parameter of interest is the coefficient of X_2, i.e., the true parameter is 1.

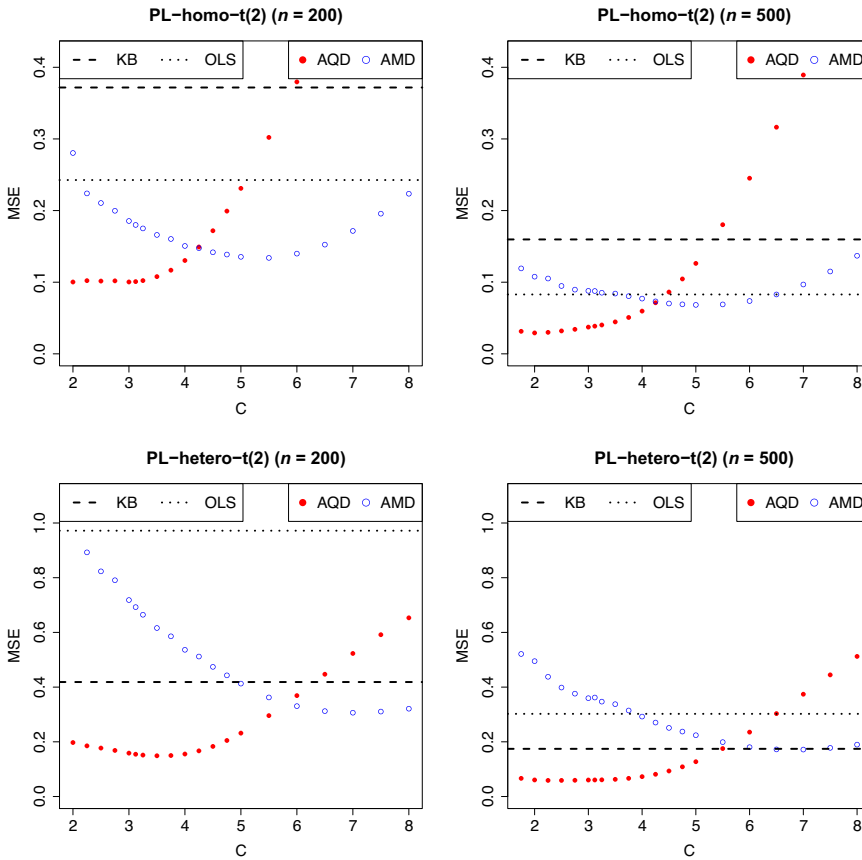

FIGURE 2. Partially linear model with $t(2)$ error.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

TABLE 1. (PL-hetero-N(0,1): $n = 500$, 500 bootstrap samples, 500 replications for each experiment) The bandwidth $h = C\sigma_x 500^{-0.15}$, where σ_x is the interquantile range of X_2. The first row reports the simulation bias relative to simulation standard error. The second row reports the simulation MSE. The next four rows report the coverage rates of the Normal, Symmetric, Percentile, and Efron 95% CIs, respectively. The top panel is for the results of AQD, and the bottom panel is for AMD

Bandwidth constant	C	1.5	2	2.5	3	3.5	4
AQD	BIAS/SE	−0.220	−0.171	−0.180	−0.323	−0.606	−1.043
	MSE	0.051	0.040	0.039	0.040	0.045	0.058
	Normal	0.990	0.972	0.962	0.952	0.936	0.854
	Symmetric	0.990	0.978	0.966	0.956	0.936	0.852
	Percentile	0.986	0.974	0.960	0.950	0.920	0.842
	Efron	0.980	0.966	0.952	0.950	0.938	0.852
AMD	BIAS/SE	0.075	0.042	−0.020	−0.064	−0.173	−0.315
	MSE	0.052	0.036	0.031	0.029	0.029	0.030
	Normal	0.996	0.976	0.974	0.966	0.964	0.952
	Symmetric	0.996	0.978	0.974	0.972	0.962	0.956
	Percentile	0.994	0.978	0.970	0.968	0.956	0.956
	Efron	0.984	0.968	0.964	0.962	0.960	0.948

We use the fourth-order Epanechnikov kernel. Under Assumption 3, we choose the trimming bound $\delta \propto n^{-0.02}$ and trim 5% of the sample at the tails. The bandwidths are $h_1 = h = C\sigma_x n^{-0.15}$ and $h_0 = C\sigma_y n^{-0.3}$, where the powers satisfy Assumption 3, and σ_x and σ_y are the interquantile range of X and Y, respectively (Silverman, 1986). Figures 1 and 2 report the MSEs against a range of C. There are 1,000 replications in each experiment. We compute the optimal bandwidth h_{opt} for $(0,1)\hat{\beta}_f$ proposed in Section 4.4 as a reference bandwidth. For the DGP PL-hetero-N(0,1), the corresponding optimal constant $C = 2.39$ for $n = 200$ and $C = 1.76$ for $n = 500$.[9] The theoretical optimal bandwidth that minimizes the MSE appears to agree with the simulation results in the lower panel of Figure 1.

For the normal error in Figure 1, AMD and AQD outperform the linear estimators, OLS and KB. For the fat-tailed error in Figure 2, the QRs (AQD and KB) outperform the mean regressions (AMD and OLS). The optimal bandwidth that minimizes the MSE for the AQD is smaller than that of the AMD. This is because AQD involves additional nonparametric estimation of the CQF, and the nonparametric estimator is more undersmoothed. When the bandwidth is around

[9] Specifically, we numerically compute $\hat{\mathrm{V}}$ described in Section 4.4 with a preliminary bandwidth $h_v = 3.12\sigma_x n^{-0.15}$, where 3.12 is from the Silverman rule-of-thumb bandwidth and $n = 5,000$. For $\hat{\mathrm{B}}$, we choose $u = 0.5$ and $h_b = 3.12\sigma_x n^{-0.07}$. The conditions in Corollary 1 hold.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

the MSE optimal bandwidth, the nonparametric estimators perform well in finite samples. Overall, AQD outperforms the linear KB.

Table 1 reports the coverage rates of three bootstrap-based CIs, following the definitions in Cattaneo and Jansson (2018). In standard notation, the superscript $*$ denotes the bootstrap analog computed under the bootstrap distribution conditional on the data. Let $q^*_{n,\alpha} = \inf\{q \in \mathcal{R} : P^*[\hat{\beta}^*_s - \hat{\beta}_s \leq q] \leq \alpha\}$ and $Q^*_{n,\alpha} = \inf\{Q \in \mathcal{R} : P^*[|\hat{\beta}^*_s - \hat{\beta}_s| \leq Q] \leq \alpha\}$. Then, Efron CI $= \left[\hat{\beta}_s + q^*_{n,\alpha/2}, \hat{\beta}_s + q^*_{n,1-\alpha/2}\right]$, Percentile CI $= \left[\hat{\beta}_s - q^*_{n,1-\alpha/2}, \hat{\beta}_s + q^*_{n,\alpha/2}\right]$, and Symmetric CI $= \left[\hat{\beta}_s - Q^*_{n,1-\alpha}, \hat{\beta}_s + Q^*_{n,1-\alpha}\right]$. We also consider Normal CI $= \left[\hat{\beta}_s - \hat{q}_{n,1-\alpha/2}, \hat{\beta}_s + \hat{q}_{n,\alpha/2}\right]$, where $\hat{q}_{n,\alpha} = \Phi^{-1}(\alpha)se$, Φ is the standard normal CDF, and we use the bootstrap standard error of $\hat{\beta}_s$ for se.

The simulation results are mostly in line with the theoretical findings in Cattaneo and Jansson (2018) and this paper. For most cases, Symmetric CI has the largest coverage rates, whereas the Efron CI has the smallest coverage rates, as predicted by Cattaneo and Jansson (2018). Normal CI is comparable with the bootstrap-based CIs, and all inference procedures perform reasonably. One possible explanation of this result is that the small-bandwidth bias studied in Cattaneo and Jansson (2018) is relatively small in this DGP.

6. CONCLUSION AND OUTLOOK

We estimate weighted AQDs via a weighted average CQF. We show that our estimators are asymptotic linear uniformly over the quantile index and converge weakly to Gaussian processes. We also characterize the leading bias. More generally, this paper is concerned with one of the semiparametric estimation problems based on a preliminary nonparametric estimator and involving a stochastic trimming function. We demonstrate a novel application of the uniform convergence results of nonparametric kernel-based estimators on expanding interior supports in Hansen (2008) and Cattaneo et al. (2013), so that our asymptotic analysis is tractable to account for the stochastic trimming problem.

There are several important directions for future research. The criteria of choosing the bandwidths and trimming parameter for finite samples are to be investigated. For the AMD, Cattaneo et al. (2010, 2013, 2014a, 2014b) and Cattaneo and Jansson (2018) develop several methods for robust inference in terms of the bandwidth choice.[10] There is recent development on the trimming parameter for the inverse probability weighting estimator of treatment effects, e.g., Ma and Wang (2019) and Sasaki and Ura (2021) propose inference methods that account for the trimming bias. Since our estimands are more complex, estimating the

[10]For example, Cattaneo et al. (2013) propose a generalized Jackknife estimator for the unweighted AMD, where the first-step estimator enters the m-estimator nonlinearly. They correct for the nonlinear bias and characterize a quadratic expansion. Consequently, they assume weaker-than-usual conditions on moments, bandwidths, and kernel order. Our first-step estimators $f(X)$ and $F_Y(y|X)$ enter the weighted AQD nonlinearly, and we characterize their biases. However, the quadratic expansion is more complicated in our problem due to the CQF estimation.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

trimming bias and a data-driven trimming parameter are out of the scope of this paper and left for future research. As with the rich literature of the AMD, more theoretical and empirical research on the AQD could be expected.

APPENDIX

The Appendix is organized as follows. We first state the notations and assumptions. Then, we present preliminary lemmas, whose proofs are in Appendix C. The asymptotic theorems for the weighted AQD estimator $\hat{\beta}$ are first derived for the infeasible estimator $\tilde{\beta} \equiv -n^{-1}\sum_{i=1}^{n}\hat{Q}(\tau|X_i)\hat{\phi}(X_i)\mathbf{1}\{X_i \in \mathcal{S}\}$, trimmed based on the true density. Then, Lemma 4 shows $\sqrt{n}(\tilde{\beta}-\hat{\beta}) = o_p(1)$ uniformly over τ. Appendix A.1 presents the proofs of Theorems 1 and 2. Appendix B presents the proofs of Theorems 3–6.

Notations. Let $Z = (X', Y)'$. $f'_Y(y|X)$ denotes the derivative with respect to y that should not be confusing with the transpose of a matrix A'. C denotes a generic constant. For an $m \times n$ matrix A, we use Frobenius norm: $\|A\| = trace(A'A)^{1/2}$. Let $\|\cdot\|_\infty$ be the sup-norm for a function, i.e., $\|f\|_\infty = \sup_{x\in\mathcal{X}}|f(x)|$. *s.o.* denotes smaller-order terms. *w.p.a.*1 means with probability approaching one. Denote the product of two functions $A(X)$ and $B(X)$ by $A(\cdot)B(\cdot)(X) \equiv (AB)(X) \equiv A(X)B(X)$. For simplicity, $Q_i \equiv Q(\tau|X_i)$, $G_{ij} \equiv G\big(\frac{Q(\tau|X_i)-Y_j}{h_0}\big)$, $K_{ij} \equiv K\big(H^{-1}(X_i - X_j)\big)$, $f_i \equiv f(X_i)$, and $\mathbf{1}_i \equiv \mathbf{1}\{X_i \in \mathcal{S}\}$. For some notations in the following proof, we omit τ and X for brevity without loss of clarity; for example, $Q \equiv Q(\tau|x)$ and $\bar{Q} \equiv \bar{Q}(\tau|x)$.

LEMMA 1 (Uniform convergence rate). *Let Assumptions 1 and 2 hold with* $p_X \geq \nu$. *Define* $\mathcal{C}_n \equiv \{x : \|x\| \leq c_n\}$, *where* c_n *satisfies* $\limsup_{n\to\infty}\log(c_n)/\log n < \infty$.

1. *Define* $\delta = \delta_n \equiv \inf_{x\in\mathcal{C}_n} f(x)$. *Let* $\delta^{-1}\left(\sqrt{\log n/(nh^d)} + h_0^2 + h^\nu\right) \to 0$. *Then,*

$$\sup_{x\in\mathcal{C}_n, y\in\mathcal{Y}}\left|\hat{F}_Y(y|x) - F_Y(y|x)\right| = O_p\left(\frac{1}{\delta}\left(\sqrt{\frac{\log n}{nh^d}} + h_0^2 + h^\nu\right)\right).$$

2. *Define* $\mathcal{Y}_n \equiv \{\|y\| \leq c_n\}$ *and* $\delta = \delta_n \equiv \inf_{x\in\mathcal{C}_n, y\in\mathcal{Y}_n} f_{XY}(x,y)$. *Let* $\delta^{-1}\left(\sqrt{\log n/(nh^d h_0)} + h_0^2 + h^\nu\right) \to 0$. *Then,*

$$\sup_{x\in\mathcal{C}_n, y\in\mathcal{Y}_n}\left|\frac{\partial}{\partial y}\hat{F}_Y(y|x) - f_Y(y|x)\right| = O_p\left(\frac{1}{\delta}\left(\sqrt{\frac{\log n}{nh^d h_0}} + h_0^2 + h^\nu\right)\right).$$

LEMMA 2. *Let Assumption 1 and 2 hold with* $p_X \geq \nu_1 + 2$. *For a measurable function* $\psi : \mathcal{X}\times\mathcal{T} \mapsto \mathcal{R}$, *assume the* (ν_1+1)*th-order derivative of* $\psi(x;\tau)$ *with respect to* x *to be uniformly continuous and bounded, for any* $x \in \mathcal{S}$ *and* $\tau \in \mathcal{T}$. *Assume* $\{X \mapsto \psi(X;\tau) : \tau \in \mathcal{T}\}$ *to be euclidean and* $\mathbb{E}\left[\left(\sup_{\tau\in\mathcal{T}}\psi(X;\tau)\right)^2\right] < \infty$. *Let* $nh_1^{2d+2} \to \infty$ *and* $\sqrt{n}h_1^{\nu_1} \to c \in [0,\infty)$. *Then, uniformly in* $\tau \in \mathcal{T}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\psi(X_i;\tau)\nabla\hat{f}(X_i)\mathbf{1}\{X_i\in\mathcal{S}\} - \mathbb{E}\big[\psi(X;\tau)\nabla f(X)\big] - \mathrm{Bias}_\psi(X_i;\tau)\Big)$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(-f(X_i) \nabla \psi(X_i;\tau) + \mathbb{E}\big[f(X) \nabla \psi(X;\tau)\big] \Big) + o_p(1), \text{where}$$

$$\text{Bias}_{\psi}(X;\tau) \equiv \frac{h_1^{\nu_1} \kappa_{\nu_1}}{\nu_1!} \sum_{k=1}^{d} \mathbb{E}\big[\psi(X;\tau) \partial_k^{\nu_1} \nabla f(X)\big].$$

LEMMA 3. *Let Assumptions 1 and 2 hold with* $p_x \geq \nu_1 + 1$. *For a measurable function* $\gamma : \mathcal{X} \times \mathcal{T} \mapsto \mathcal{R}$, *assume the* (ν_1+1)*th-order derivative of* $\gamma(x;\tau)$ *with respect to* x *to be uniformly continuous and bounded, for any* $x \in \mathcal{S}$ *and* $\tau \in \mathcal{T}$. *Assume* $\{X \mapsto \gamma(X;\tau) : \tau \in \mathcal{T}\}$ *to be euclidean and* $\mathbb{E}\Big[\big(\sup_{\tau \in \mathcal{T}} \gamma(X;\tau)\big)^2\Big] < \infty$. *Let* $nh_1^{2d} \to \infty$ *and* $\sqrt{n} h_1^{\nu_1} \to c \in [0,\infty)$. *Then, uniformly in* $\tau \in \mathcal{T}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\gamma(X_i;\tau)(\hat{f}(X_i) - f(X_i)) \mathbf{1}\{X_i \in \mathcal{S}\} - \text{Bias}_{\gamma}(X;\tau) \Big)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\gamma(X_i;\tau) f(X_i) - \mathbb{E}[\gamma(X;\tau) f(X)] \Big) + o_p(1), \text{where}$$

$$\text{Bias}_{\gamma}(X;\tau) \equiv \frac{h_1^{\nu_1} \kappa_{\nu_1}}{\nu_1!} \sum_{k=1}^{d} \mathbb{E}\Big[\gamma(X;\tau) \partial_k^{\nu_1} f(X)\Big].$$

LEMMA 4 (Trimming). *Let all assumptions in Theorem 3 hold. Let* $\sup_{x \in \mathcal{S}, \tau \in \mathcal{T}} |\hat{\phi}(x;\tau) - \phi(x;\tau)| = o_p(1)$. *Then,* $n^{-1/2} \sum_{i=1}^{n} \hat{Q}(\tau|X_i) \hat{\phi}(X_i;\tau) \Big(\mathbf{1}\big\{X_i \in \hat{\mathcal{S}}\big\} - \boldsymbol{1}\big\{X_i \in \mathcal{S}\big\} \Big) = o_p(1)$ *uniformly in* $\tau \in \mathcal{T}$.

LEMMA 5. *Let the conditions in Theorem 3 hold. Denote* $p(Z_i, Z_j; \lambda) \equiv \frac{\phi_i \boldsymbol{I}_i}{f_i f_Y(Q_i|X_i)} \frac{1}{|H|} K_{ij} (\tau - G_{ij})$. *Then,*

$$\mathbb{E}\Big[p(Z_i, Z_j; \lambda) \Big| Z_i\Big] = \frac{-\phi_i \mathbf{1}_i}{f_i f_Y(Q_i|X_i)} \Big\{ h^{\nu} \kappa_{\nu} \sum_{l=1}^{\nu} \frac{1}{l!\,(\nu - l)!} \sum_{k=1}^{d} \partial_k^l F_Y(Q_i|X_i) \partial_k^{\nu-l} f(X_i)$$
$$+ \frac{h_0^2}{2} \kappa_{G2} f_Y'(Q_i|X_i) f_i - h^{\nu} h_0^2 R_I(X_i) + O\big(h^{\nu+1} + h_0^3\big) \Big\},$$

where $R_I(X_i) \equiv \frac{\kappa_{G2}}{2} \kappa_{\nu} \sum_{l=0}^{\nu} \frac{1}{l!(\nu-l)!} \sum_{k=1}^{d} \partial_k^l f_Y'(Q_i|X_i) \partial_k^{\nu-l} f(X_i)$. *In addition,*

$$\mathbb{E}\left[\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} p(Z_i, Z_j; \lambda) \right]$$

$$= -\mathbb{E}\Big[\frac{\phi_i}{f_i f_Y(Q_i|X_i)} \Big(h^{\nu} \kappa_{\nu} \sum_{l=1}^{\nu} \frac{1}{l!\,(\nu - l)!} \sum_{k=1}^{d} \partial_k^l F_Y(Q_i|X_i) \partial_k^{\nu-l} f(X_i) + \frac{h_0^2}{2} \kappa_{G2} f_Y'(Q_i|X_i) f_i \Big)$$
$$+ h^{\nu} h_0^2 R_I(X_i) \Big] + o\big(h^{\nu} + h_0^2\big).$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

A. Proofs of Theorems 1 and 2 in Section 4.1

Proof of Theorem 1. The proofs use the following preliminary results (i) and (ii). Suppose that, for any $x \in \mathcal{S}$, there exists a compact convex set $\mathcal{Y}_x \equiv [Q(\varepsilon|x), Q(1-\varepsilon|x)] \subset \mathcal{Y}$ such that $\inf_{x\in\mathcal{S}} \inf_{y\in\mathcal{Y}_x} f_{XY}(x,y) \geq \delta$.

(i) $\hat{Q}(\tau|x) \in [Q(\varepsilon|x), Q(1-\varepsilon|x)] = \mathcal{Y}_x$ w.p.a.1, for any $x \in \mathcal{S}$ and $\tau \in \mathcal{T}$.
(ii) $\sup_{x\in\mathcal{S},\tau\in\mathcal{T}} \left(f_{XY}(x, \hat{Q}(\tau|x))\right)^{-1} = O_p(\delta^{-1})$.

To briefly discuss these results, Lemma 1 implies $\eta \in (0, \min\{\tau-\varepsilon, 1-\varepsilon-\tau\})$ such that $\sup_{y\in\mathcal{Y},x\in\mathcal{C}_n} |\hat{F}_Y(y|x) - F_Y(y|x)| < \eta$, *w.p.a.*1. So $\sup_{x\in\mathcal{S},\tau\in\mathcal{T}} |\hat{F}_Y(\hat{Q}(\tau|x)|x) - F_Y(\hat{Q}(\tau|x)|x)| < \eta w.p.a.1$. Since $\hat{F}_Y(\hat{Q}(\tau|x)|x) = \tau$, we observe $F_Y(Q(\varepsilon|x)|x) = \varepsilon < \tau - \eta < F_Y(\hat{Q}(\tau|x)|x) < \tau + \eta < 1 - \varepsilon = F_Y(Q(1-\varepsilon|x)|x)$. Monotonicity of $F_Y(y|x)$ in y implies (i). Then, $\inf_{x\in\mathcal{S},\tau\in\mathcal{T}} f_{XY}(x, \hat{Q}(\tau|x)) \geq \delta$, *w.p.a.*1 that implies (ii). When $\mathcal{Y}_x$ degenerates to a point, i.e., $Q(\tau|x) = Q(\varepsilon|x) = Q(1-\varepsilon|x)$, for $\tau \in \mathcal{T}$, we modify (i) $\hat{Q}(\tau|x) \xrightarrow{p} Q(\varepsilon|x)$. Then (ii) follows.

Uniform convergence rate. For any $x \in \mathcal{S}$ and $\tau \in \mathcal{T}$, a Taylor series expansion yields $F_Y\big(\hat{Q}(\tau|x)\big|x\big) = F_Y\big(Q(\tau|x)\big|x\big) + f_Y\big(\bar{Q}(\tau|x)\big|x\big)\big(\hat{Q}(\tau|x) - Q(\tau|x)\big)$, where $\bar{Q}(\tau|x)$ is on the line segment between $Q(\tau|x)$ and $\hat{Q}(\tau|x)$. We claim

$$\sup_{\substack{\tau\in\mathcal{T}\\ x\in\mathcal{S}}} \left|\hat{Q}(\tau|x) - Q(\tau|x)\right| = \sup_{\substack{\tau\in\mathcal{T}\\ x\in\mathcal{S}}} \left|\frac{\big(F_Y\big(\hat{Q}(\tau|x)\big|x\big) - \tau\big) f(x)}{f_{XY}\big(x, \bar{Q}(\tau|x)\big)}\right|$$

$$= O_p\left(\frac{1}{\delta} \sup_{\substack{y\in\mathcal{R}\\ x\in\mathcal{S}}} \left|\big(F_Y\big(y|x\big) - \hat{F}_Y\big(y|x\big)\big) f(x)\right|\right) = O_p\left(\frac{1}{\delta}\left(\sqrt{\frac{\log n}{nh^d}} + h_0^2 + h^\nu\right)\right)$$

by the following reasons: For the numerator, $\hat{F}_Y(\hat{Q}(\tau|x)|x) = \tau = F_Y(Q(\tau|x)|x)$ by construction. Then, use the result for $\Psi = F_Y(y|x)f(x)$ in Proof of Lemma 1(1). For the denominator, the above results (i) and (ii) imply $\sup_{x\in\mathcal{S},\tau\in\mathcal{T}} \left(f_{XY}(x, \bar{Q}(\tau|x))\right)^{-1} = O_p(\delta^{-1})$.

Bahadur representation. Since $\hat{F}_Y(y|x)$ is smooth in y, we can expand $\hat{F}_Y(\hat{Q}(\tau|x)|x)$ around $Q(\tau|x)$ by a Taylor series expansion, for any $x \in \mathcal{S}$ and $\tau \in \mathcal{T}$:

$$\tau = \hat{F}_Y\big(\hat{Q}(\tau|x)\big|x\big) = \hat{F}_Y\big(Q(\tau|x)\big|x\big) + \hat{f}_Y\big(Q(\tau|x)\big|x\big)\big(\hat{Q}(\tau|x) - Q(\tau|x)\big) + \frac{1}{2}\hat{f}'_Y\big(\bar{Q}(\tau|x)\big|x\big)\big(\hat{Q}(\tau|x) - Q(\tau|x)\big)^2, \tag{12}$$

where $\bar{Q}(\tau|x)$ is on the line segment between $Q(\tau|x)$ and $\hat{Q}(\tau|x)$. To simplify notations without loss of clarity, we sometimes omit τ and x, e.g., $Q \equiv Q(\tau|x)$. From (12),

$$\begin{aligned}
&\hat{Q}(\tau|x) - Q(\tau|x)\\
&= \frac{\tau - \hat{F}_Y(Q(\tau|x)|x)}{f_Y(Q(\tau|x)|x)} + \big(\tau - \hat{F}_Y(Q(\tau|x)|x)\big)\left(\frac{1}{\hat{f}_Y(Q(\tau|x)|x)} - \frac{1}{f_Y(Q(\tau|x)|x)}\right)\\
&\quad - \frac{1}{2}\frac{\hat{f}'_Y(\bar{Q}(\tau|x)|x)}{\hat{f}_Y(Q(\tau|x)|x)}(\hat{Q}(\tau|x) - Q(\tau|x))^2\\
&= \frac{A_\tau(x)}{f_{XY}(x, Q)}
\end{aligned}$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

$$+\frac{A_\tau(x)}{f_{XY}(x,Q)}\underbrace{\Big(\frac{f(x)-\hat f(x)}{\hat f(x)}\Big)}_{\equiv B(x)}+\frac{A_\tau(x)}{f_{XY}(x,Q)}\underbrace{\Big(\frac{f_Y(Q(\tau|x)|x)-\hat f_Y(Q(\tau|x)|x)}{\hat f_Y(Q(\tau|x)|x)}\Big)}_{\equiv C_\tau(x)} \tag{13}$$

$$+\frac{A_\tau(x)}{f_{XY}(x,Q)}B(x)C_\tau(x)-\frac{1}{2}\frac{\hat f_Y'(\bar Q(\tau|x)|x)}{\hat f_Y(Q(\tau|x)|x)}(\hat Q(\tau|x)-Q(\tau|x))^2, \tag{14}$$

where $A_\tau(x)\equiv n^{-1}\sum_{j=1}^n K_h(x-X_j)\Big(\tau-G\big(\frac{Q(\tau|x)-Y_j}{h_0}\big)\Big)$. By Proof of Lemma 1(1), $A_\tau(x)=\tau\Big(\hat f(x)-f(x)\Big)-\Big(\hat\Psi(Q(\tau|x),x)-\tau f(x)\Big)$. So $\sup_{x\in\mathcal{S},\,\tau\in\mathcal{T}}|A_\tau(x)|=O_p\Big(\sqrt{\frac{\log n}{nh^d}}+h_0^2+h^\nu\Big)$.

$$C_\tau(x)=\frac{f_Y(Q|x)-\hat f_Y(Q|x)}{f_Y(Q|x)}+\Big(f_Y(Q|x)-\hat f_Y(Q|x)\Big)\Big(\frac{1}{\hat f_Y(Q|x)}-\frac{1}{f_Y(Q|x)}\Big),$$

where the first leading term is

$$\begin{aligned}&\frac{f_{XY}(x,Q)-\hat f_{XY}(x,Q)}{f_{XY}(x,Q)}+\frac{\hat f_{XY}(x,Q)}{f_{XY}(x,Q)}\Big(1-\frac{f(x)}{\hat f(x)}\Big)\\&=\frac{1}{f_{XY}(x,Q)}\Big(f_{XY}(x,Q)-\hat f_{XY}(x,Q)\Big)-B(x)+\Big(\frac{\hat f_{XY}(x,Q)}{f_{XY}(x,Q)}-1\Big)\Big(1-\frac{f(x)}{\hat f(x)}\Big)\\&=O_p\left(\frac{1}{\delta}\left(\sqrt{\frac{\log n}{nh^dh_0}}+h_0^2+h^\nu\right)\right)-B(x)\\&\quad+O_p\left(\frac{1}{\delta^2}\left(\sqrt{\frac{\log n}{nh^dh_0}}+h_0^2+h^\nu\right)\left(\sqrt{\frac{\log n}{nh^d}}+h^\nu\right)\right)\end{aligned}$$

by Lemma 1(2). So the leading term of $A_\tau(x)\big(B(x)+C_\tau(x)\big)/f_{XY}(x,Q)$ in (13) is

$$O_p\left(\frac{1}{\delta^2}\left(\sqrt{\frac{\log n}{nh^d}}+h^\nu+h_0^2\right)\left(\sqrt{\frac{\log n}{nh^dh_0}}+h^\nu+h_0^2\right)\right). \tag{15}$$

The last term in (14)

$$\sup_{x\in\mathcal{S},\,\tau\in\mathcal{T}}\left|-\frac{1}{2}\frac{\hat f_Y'(\bar Q(\tau|x)|x)}{\hat f_Y(Q(\tau|x)|x)}(\hat Q(\tau|x)-Q(\tau|x))^2\right|\le O_p\left(\frac{1}{\delta^3}\left(\sqrt{\frac{\log n}{nh^d}}+h_0^2+h^\nu\right)^2\right) \tag{16}$$

by the following reasons: First, the result (ii) and Assumption 1 imply $f_Y'(\bar Q(\tau|x)|x)$ is uniformly bounded w.p.a.1. Second, for any $\varepsilon>0$, there is a constant c_f such that $|\hat f_{XY}(x,Q(\tau|x))-f_{XY}(x,Q(\tau|x))|\mathbf{1}\{x\in\mathcal{S},\tau\in\mathcal{T}\}\le c_f\big((n^{1-\varepsilon}h_0h^d)^{-1/2}+h_0^2+h^\nu\big)\equiv c_{2n}$, *w.p.a.*1. So $\inf_{x\in\mathcal{S},\,\tau\in\mathcal{T}}\hat f_{XY}(x,Q(\tau|x))\ge\delta-c_{2n}$, *w.p.a.*1. We obtain the bound of R_n in (10) by collecting the remainder terms (15) and (16). ■

Proof of Theorem 2. For all $\omega\in\Omega$ and $x\in\mathcal{S}$, the triangular array $f_{ni}(\omega,\tau)\equiv(n|H|)^{-1/2}K\big(H^{-1}(X_i(\omega)-x)\big)\big(\tau-G\big(\frac{Q(\tau|x)-Y_i(\omega)}{h_0}\big)\big)\big/f_{XY}(x,Q(\tau|x))$ are independent

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

within rows. Define the $n \times 1$ vector $f_n(\omega,\tau) \equiv \left(f_{n1}(\omega,\tau), ldots, f_{nn}(\omega,\tau)\right)'$ and the random set $\mathcal{F}_{n\omega} \equiv \left\{f_n(\omega,\tau) : \tau \in \mathcal{T}\right\}$. In the following, we check the conditions (i)–(v) for the functional CLT, Theorem 10.6 in Pollard (1990).

(i) The triangular array processes $\{f_{ni}(\omega,\tau)\}$ are manageable with respect to the envelopes $F_{ni}(\omega) \equiv \delta^{-1}(nh^d)^{-1/2}K\left(H^{-1}(X_i(\omega)-x)\right)$. First, $\{X \mapsto f_Y(Q(\tau|X)|X) : \tau \in \mathcal{T}\}$ and $\{(X,Y) \mapsto G((Q(\tau|X)-Y)/h_0) : \tau \in \mathcal{T}\}$ are euclidean and manageable by Lemma 2.13 and Example 2.10 in Pakes and Pollard (1989) and p. 221 in Kosorok (2008). In addition, $F_n(\omega) \equiv (F_{n1},\ldots,F_{nn})'$ is an $\mathcal{R}^n$-valued function on the underlying probability space. Then (i) is proved by applying Lemma E1 in Andrews and Shi (2013).

We first calculate the following results. As $h_0 \to 0$,

$$\begin{aligned}
&\mathbb{E}\left[G\Big(\frac{Q(\tau|X_i)-Y_j}{h_0}\Big)\Big|X_i,X_j\right] \\
&= G\Big(\frac{Q_i-\bar{y}}{h_0}\Big) + \int_{\mathcal{V}} g(v)F_Y(Q_i-h_0v|X_j)dv \\
&= G\Big(\frac{Q_i-\bar{y}}{h_0}\Big) \\
&\quad + \int_{-\infty}^{\infty} g(v)\Big(F_Y(Q_i|V_j) - h_0vf_Y(Q_i|X_j) + \frac{h_0^2}{2}v^2f'_Y(Q_i|X_j) + \frac{h_0^3}{3!}v^3f''_Y(\bar{Q}_i|X_j)\Big)dv \\
&\quad - \int_{v\notin\mathcal{V}} g(v)\Big(F_Y(Q_i|X_j) - h_0vf_Y(Q_i|X_j) + \frac{h_0^2}{2}v^2f'_Y(Q_i|X_j) + \frac{h_0^3}{3!}v^3f''_Y(\bar{Q}_i|X_j)\Big)dv \qquad \textbf{(17)}
\end{aligned}$$

$$= F_Y(Q_i|X_j) + \frac{h_0^2}{2}f'_Y(Q_i|X_j)\kappa_{G2} + O(h_0^3), \qquad \textbf{(18)}$$

where $\mathcal{V} \equiv \left[\frac{Q_i-\bar{y}}{h_0}, \frac{Q_i-\underline{y}}{h_0}\right]$. The second equality is a Taylor series expansion around Q_i, and $\bar{Q}_i$ is on the line segment between Q_i and $Q_i - h_0v$. When g has a bounded support, $G(z/h_0)$ is zero for a small enough h_0 and for any negative z. Thus, when g has a bounded support or when the support of Y is $\mathcal{R}$, i.e., $\mathcal{V} = \mathcal{R}$, the term (17) is zero for a small enough h_0. When $\bar{y}$ or $\underline{y}$ is bounded and g has an unbounded support, Assumption 2(G) implies that $G(z/h_0) = o(h_0^3)$, for any $z < 0$. Thus, the first term of (17) is $-F_Y(Q_i|X_j)\Big(1 - G\Big(\frac{Q_i-\underline{y}}{h_0}\Big) + G\Big(\frac{Q_i-\bar{y}}{h_0}\Big)\Big) = o(h_0^3)$. Similarly, the second and third terms are $o(h_0^3)$ by integration by parts. The last term of (17) is $o(h_0^3)$ by the uniform continuity of $f'_Y(y|X)$ in y and the dominated convergence theorem.

By a similar argument, for $q_1 \le q_2$,

$$\begin{aligned}
&\mathbb{E}\left[G\left(\frac{q_1-Y}{h_0}\right)G\left(\frac{q_2-Y}{h_0}\right)\Big|X\right] \\
&= \int_{\underline{y}}^{\bar{y}}\left(\frac{1}{h_0}g\left(\frac{q_1-y}{h_0}\right)G\left(\frac{q_2-y}{h_0}\right) + \frac{1}{h_0}g\left(\frac{q_2-y}{h_0}\right)G\left(\frac{q_1-y}{h_0}\right)\right)F_Y(y|X)dy \\
&\quad + G\left(\frac{q_1-y}{h_0}\right)G\left(\frac{q_2-y}{h_0}\right)F_Y(y|X)\Big|_{\underline{y}}^{\bar{y}} = F_Y(q_1|X) + O(h_0^2). \qquad \textbf{(19)}
\end{aligned}$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

(ii) Define $\mathcal{Z}_n(\tau) = \sum_{i=1}^n \big(f_{ni}(\tau) - \mathbb{E}f_{ni}(\tau)\big)$. The covariance kernel of the limiting Gaussian process is $\lim_{n\to\infty} \mathbb{E}\big[\mathcal{Z}_n(\tau_1)\mathcal{Z}_n(\tau_2)\big] = \lim_{n\to\infty} n\mathbb{E}[f_n(\omega,\tau_1)f_n(\omega,\tau_2)] - n\mathbb{E}[f_n(\omega,\tau_1] \times \mathbb{E}[f_n(\omega,\tau_2)] = f(x)\tau_1(1-\tau_2)\int K^2(v)dv$.

(iii) $\sum_{i=1}^n \mathbb{E}\big[F_{ni}^2\big] = f(x)\int K(v)^2 dv$.

(iv) For any $\varepsilon > 0$, $\sum_{i=1}^n \mathbb{E}\big[F_{ni}^2\,\mathbf{1}\{F_{ni} > \varepsilon\}\big] \to 0$. This is because $\mathbf{1}\{F_{ni} > \varepsilon\} = 0$ for n large enough, by assuming K is bounded and $\sqrt{nh^d} \to \infty$.

(v) $n\mathbb{E}\big[\big|f_{ni}(\tau_1) - f_{ni}(\tau_2)\big|^2\big] \to \rho(\tau_1,\tau_2)^2 \equiv f(x)(\tau_2-\tau_1)(1+\tau_1-\tau_2)\int K(v)^2dv$, uniformly in τ_1,τ_2. Therefore, uniformly in τ_1,τ_2, $\rho_n(\tau_1,\tau_2) \equiv \big(\sum_{i=1}^n \mathbb{E}\big[\big|f_{ni}(\tau_1) - f_{ni}(\tau_2)\big|^2\big]\big)^{1/2} \to \rho(\tau_1,\tau_2)$.

■

B. Proofs of Theorems 3–6 in Section 4.2

We use the U-process theorems in Sherman (1994) to prove Lemma 2, Lemma 3, and Theorem 3. Then, the proofs of Theorems 4 and 5 build on these results. We start with an overview of the proof of the U-process theorems.

Denote $\lambda \equiv (\tau, h, h_0, \delta) \in \Lambda \equiv \mathcal{T}\times\mathcal{R}_+\times\mathcal{R}_+\times\mathcal{R}_+$. Let $(Z_i,Z_j) \in \mathcal{Z}^2 \equiv \mathcal{Z}\otimes\mathcal{Z}$ from the product measure $\mathbb{P}^2 \equiv \mathbb{P}\otimes\mathbb{P}$. Let $\mathcal{F} \equiv \{(Z_i,Z_j) \mapsto p(Z_i,Z_j;\lambda) : \lambda \in \Lambda\}$ be a class of measurable functions on $\mathcal{Z}^2$. The collection $\{U_n p : p \in \mathcal{F}\}$ is a U-process of order 2 indexed by $\mathcal{F}$ where, for each $p \in \mathcal{F}$,

$$
\begin{aligned}
U_n p &\equiv \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i} p(Z_i,Z_j;\lambda) \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[p(Z_i,Z;\lambda)|Z_i] + \frac{1}{n}\sum_{j=1}^n \mathbb{E}[p(Z,Z_j;\lambda)|Z_j] - \mathbb{P}^2[p(Z_i,Z_j;\lambda)] + U_n^2, \\
U_n^2 &\equiv \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i} r(Z_i,Z_j;\lambda),
\end{aligned}
$$

and $r(Z_i,Z_j;\lambda) \equiv p(Z_i,Z_j;\lambda) - \mathbb{E}[p(Z_i,Z;\lambda)|Z_i] - \mathbb{E}[p(Z,Z_j;\lambda)|Z_j] + \mathbb{P}^2[p(Z_i,Z_j;\lambda)]$. The proof involves the following steps.

Step 1. We show $\sup_{\mathcal{F}}|U_n^2| = o_p(n^{-1/2})$ by Corollary 4 in Sherman (1994).

Step 2 [Projection]. Calculate the projection $r_{ni}(\tau) \equiv \mathbb{E}[p(Z_i,Z;\lambda)|Z_i] + \mathbb{E}[p(Z,Z_i;\lambda)|Z_i]$. Find the influence function $r_i(\tau)$ such that $n^{-1/2}\sum_{i=1}^n r_{ni}(\tau) = n^{-1/2}\sum_{i=1}^n r_i(\tau) + o_p(1)$, uniformly in $\tau \in \mathcal{T}$.

Step 3 [Bias]. Calculate $\mathbb{E}[U_n p] = \mathbb{P}^2 p$, uniformly in $\tau \in \mathcal{T}$. For the weak convergence result, let the asymptotic bias converge to zero at a rate faster than root-n.

Step 4 [Weak convergence]. By van der Vaart (2000), (i) monotonic and smooth function classes are Donsker and (ii) the Cartesian product of two Donsker classes of functions is also a Donsker class. By Donsker's theorem, we complete the proof.

Proof of Theorem 3. By Theorem 1,

$$
\frac{1}{n}\sum_{i=1}^n (\hat{Q}_i - Q_i)\phi_i \mathbf{1}_i = U_n p + \frac{1}{n}\sum_{i=1}^n \phi_i \mathbf{1}_i R_n(\tau, X_i), \tag{20}
$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

where the second-order U-statistic

$$U_n p \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \underbrace{\frac{\phi(X_i) K\big(H^{-1}(X_i - X_j)\big)}{f(X_i) f_Y(Q(\tau|X_i)|X_i)|H|} \Big(\tau - G\Big(\frac{Q(\tau|X_i) - Y_j}{h_0}\Big)\Big) \mathbf{1}_i}_{\equiv p(Z_i, Z_j; \tau, h, h_0, \delta)}$$

and the second term in (20) is $O_p\big(n^{-1/2} \sum_{i=1}^{n} \|\phi_i\| \mathbf{1}_i |R_n(\tau, X_i)|\big) = O_p\big(\sqrt{n} \sup_{X \in \mathcal{S}, \tau \in \mathcal{T}} \|\phi(X)\| \, |R_n(X)|\big) = o_p(1)$ by the condition (ii).

Step 1. We claim $\mathcal{F}_1 \equiv \{(Z_i, Z_j) \mapsto h^d p(Z_i, Z_j; \lambda) : \lambda \in \Lambda\}$ is euclidean for the envelope $F_1(Z_i, Z_j) = C\phi(X_i)/(\inf_{\tau \in \mathcal{T}} f_{XY}(X_i, Q(\tau|X_i)))$ that satisfies $\mathbb{P}^2 F^2 = C\mathbb{E}[\|\phi(X_i)\|^2/(\inf_{\tau \in \mathcal{T}} f_{XY}(X_i, Q(\tau|X_i)))^2] < \infty$. The classes $\{X \mapsto Q(\tau|X) : \tau \in \mathcal{T}\}$, $\{X \mapsto K((X-x)/h) : h > 0\}$, $\{(X, Y) \mapsto G((Q(\tau|X) - Y)/h_0) : \tau \in \mathcal{T}, h_0 > 0\}$ and $\{\mathbf{1}\{\inf_{\tau \in \mathcal{T}} f_{XY}(X, Q(\tau|X)) \geq \delta\} : \delta > 0\}$ are euclidean by Lemma 2.13 and Example 2.10 in Pakes and Pollard (1989) and p. 221 in Kosorok (2008). Then, Lemma 2.14 in Pakes and Pollard (1989) implies $\mathcal{F}_1$ is euclidean.

Thus, we can apply Corollary 4 in Sherman (1994). Lemma 6 in Sherman (1994) implies the class of $\mathbb{P}$-degenerate functions of order 2 $\{(Z_i, Z_j) \mapsto h^d r(Z_i, Z_j; \lambda) : \lambda \in \Lambda\}$ is euclidean.

Step 2 [Projection]. We show that uniformly in $\tau \in \mathcal{T}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_{ni} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (r_{1ni} + r_{2ni}) \overset{\text{(i)}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_{2ni} + o_p(1)$$
$$\overset{\text{(ii)}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_{3ni} + o_p(1) \overset{\text{(iii)}}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_{3i} + o_p(1),$$

where $r_{1ni} \equiv \mathbb{E}[p(Z_i, Z; \lambda)|Z_i]$, $r_{2ni} \equiv \mathbb{E}[p(Z, Z_i; \tau)|Z_i]$,

$$r_{3ni} \equiv \frac{\phi(X_i)}{f_Y\big(Q(\tau|X_i)|X_i\big)} \Big(\tau - G\Big(\frac{Q(\tau|X_i) - Y_i}{h_0}\Big)\Big),$$

$r_{3i} \equiv A(X_i)\big(\tau - \mathbf{1}\{Y_i \leq Q(\tau|X_i)\}\big)$, and $A(X_i) \equiv \phi(X_i)/f_Y(Q(\tau|X_i)|X_i)$. We prove equalities (i), (ii), and (iii) in the above equation in the following.

(i) We claim $\sup_{\tau \in \mathcal{T}} n^{-1/2} \sum_{i=1}^{n} r_{1ni} = o_p(1)$. We calculate $r_{1ni} \equiv \mathbb{E}[p(Z_i, Z; \lambda)|Z_i]$ in Lemma 5. Furthermore, by the condition $\mathbb{E}[\|\phi(X_i)\|^2/(\inf_{\tau \in \mathcal{T}} f_{XY}(X_i, Q(\tau|X_i)))^2] < \infty$, one can show $\mathbb{E}[\|r_{1ni}\|^2] = O\big((h_0^2 + h^\nu)^2\big) = o(1)$. Chebyshev's inequality implies the claim.

(ii) We claim $\sup_{\tau \in \mathcal{T}} n^{-1/2} \sum_{i=1}^{n} (r_{2ni} - r_{3ni}) = o_p(1)$.

$$\begin{aligned}
r_{2ni} &\equiv \mathbb{E}[p(Z, Z_i; \tau)|Z_i] \\
&= \mathbb{E}\Big[\frac{\phi(X_j)\mathbf{1}X_j}{f(X_j) f_Y(Q_j|X_j)} \frac{1}{|H|} K_{ji}(\tau - G_{ji}) \Big| Z_i\Big] \\
&= \int_{\mathcal{S}} \frac{\phi(X_j)}{f_Y(Q_j|X_j)} \frac{1}{|H|} K_{ji}\Big(\tau - G\Big(\frac{Q_j - Y_i}{h_0}\Big)\Big) dX_j \\
&= \int \frac{\phi(X_i + uh) K(u)}{f_Y\big(Q(\tau|X_i + uh)|X_i + uh\big)} \Big(\tau - G\Big(\frac{Q(\tau|X_i + uh) - Y_i}{h_0}\Big)\Big) \mathbf{1}\{X_i + uh \in \mathcal{S}\} du \\
&= r_{3ni} + O(h^\nu).
\end{aligned}$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

By the argument for equality (i), it suffices that $\mathbb{E}[\|r_{2ni} - r_{3ni}\|^2] = O(h^{2\nu}) = o(1)$.

(iii) We claim $T_n \equiv n^{-1/2}\sum_{i=1}^n \big(r_{3ni} - r_{3i} - \mathbb{E}[r_{3ni} - r_{3i}]\big) = o_p(1)$, uniformly in $\tau \in \mathcal{T}$.

$$\mathbb{E}\Big[\|r_{3ni} - r_{3i}\|^2\Big] = \mathbb{E}\Big[\|A(X_i)\|^2\Big(G\Big(\frac{Q(\tau|X_i) - Y_i}{h_0}\Big) - \mathbf{1}\{Y_i \le Q(\tau|X_i)\}\Big)^2\Big] = O(h_0)$$

by (19) and

$$\begin{aligned}&\int_{\underline{y}}^{Q_i} G\Big(\frac{Q_i - Y_i}{h_0}\Big) f_{Y|X}(y|X_i)dy \\ &= G\Big(\frac{Q_i - Y_i}{h_0}\Big)F_{Y|X}(y|X_i)\Big|_{\underline{y}}^{Q_i} + \int_{\underline{y}}^{Q_i} \frac{1}{h_0} g\Big(\frac{Q_i - Y_i}{h_0}\Big)F_{Y|X}(y|X_i)dy \\ &= G(0)\tau + \tau/2 + f_Y(Q_i|X_i)O(h_0).\end{aligned}$$

Furthermore, using (18), $\mathbb{E}[r_{3ni} - r_{3i}] = O(h_0^2)$. Thus, $\mathbb{E}\big[\|T_n\|^2\big] = o(1)$. The claim follows the argument for equality (i).

For the stochastic trimming function, by Proof of Lemma 4, $n^{-1}\sum_{i=1}^n \hat{Q}_i\phi_i \mathbf{1}\{X_i \in \mathcal{S}\} - n^{-1}\sum_{i=1}^n \hat{Q}_i\phi_i\mathbf{1}\{X_i \in \hat{\mathcal{S}}\} = o_p(n^{-1/2})$.

Step 3 [Bias]. The bias $\mathbb{E}[U_n p] = O\big(h^\nu + h_0^2\big)$ by Lemma 5. ■

Proof of Theorem 4. *Density-weighted AQD.*

$$\begin{aligned}\tilde{\beta}_f - \beta_f &= -\frac{2}{n}\sum_{i=1}^n \hat{Q}(\tau|X_i)\nabla\hat{f}_i\mathbf{1}_i - \beta_f \\ &= \underbrace{-\frac{2}{n}\sum_{i=1}^n (\hat{Q}_i - Q_i)\nabla f_i\mathbf{1}_i}_{\equiv(I)} \underbrace{-\frac{2}{n}\sum_{i=1}^n \big(Q_i\nabla\hat{f}_i\mathbf{1}_i - \mathbb{E}[Q_i\nabla f_i]\big)}_{\equiv(II)} \underbrace{-\frac{2}{n}\sum_{i=1}^n \big(\hat{Q}_i - Q_i\big)\big(\nabla\hat{f}_i - \nabla f_i\big)\mathbf{1}_i}_{\equiv s.o.}.\end{aligned}$$

Decompose $(I) = n^{-1}\sum_{i=1}^n \big(-2\hat{Q}_i\nabla f_i\mathbf{1}_i - \beta_f\big) + \big(2Q_i\nabla f_i\mathbf{1}_i + \beta_f\big)$, where the influence function for the first part is given in Theorem 3 with $\phi(X) = -2\nabla f(X)$. The influence function for (II) is given in Lemma 2 with $\psi(X;\tau) = -2Q(\tau|X)$. For the third term *s.o.*,

$$\sup_{X_i \in S}\Big\|-2\big(\hat{Q}_i - Q_i\big)\big(\nabla\hat{f}_i - \nabla f_i\big)\Big\| = O_p\Big(\frac{1}{\delta}\Big(\frac{\log n}{nh^d}\Big)^{1/2}\Big(\frac{\log n}{nh_1^{d+2}}\Big)^{1/2}\Big) = o_p(n^{-1/2})$$

by Lemma 1, Theorem 6 in Hansen (2008), and Assumption 3.

Combining the results and $\sqrt{n}(\tilde{\beta}_f - \hat{\beta}_f) = o_p(1)$ in Lemma 4, we obtain $r_f(Z_i;\tau)$. Combining the bias terms in Theorem 3 and Lemma 2, we obtain $\text{Bias}_f(\tau;h,h_0,h_1)$.

As argued in Step 4, we obtain the weak convergence, with the covariance $\text{Cov}\,(\mathbb{G}(\tau_1), \mathbb{G}(\tau_2)) = \mathbb{E}\big[r_f(Z;\tau_1)r_f(Z;\tau_2)'\big]$.

Scaled AQD. By Theorem 4 with $\gamma(X_i;\tau) = 1$, $\hat{\alpha} - \alpha \equiv n^{-1}\sum_{i=1}^n r_\alpha(X_i) + o_p(n^{-1/2})$, where the influence function $r_\alpha(X_i) = 2\big(f(X_i) - \mathbb{E}f(X)\big)$, as shown in Powell and Stoker

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

(1996). Then, $\sqrt{n}\left(\hat{\beta}_s - \beta_s\right)$

$$= \frac{\sqrt{n}}{\hat{\alpha}\alpha}\left(\hat{\beta}_f\alpha - \beta_f\hat{\alpha}\right) = \frac{\sqrt{n}}{\hat{\alpha}\alpha}\left((\hat{\beta}_f - \beta_f)\alpha - \beta_f(\hat{\alpha} - \alpha)\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(r_f(Z_i) - r_\alpha(X_i)\beta_s\right)/\hat{\alpha} + o_p(1).$$

As argued in Step 4, we obtain the second part of Theorem 4. ■

Proof of Theorem 5.

$$\hat{\beta}_W - \beta_W = \underbrace{-\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Q}_i - Q_i\right)\frac{\nabla(W_i f_i)}{f_i}\mathbf{1}_i}_{(I)} - \frac{1}{n}\sum_{i=1}^{n} Q_i\frac{\nabla(W_i\hat{f}_i)}{\hat{f}_i}\mathbf{1}_i - \beta_W + s.o.1, \tag{21}$$

where $s.o.1 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Q}_i - Q_i\right)\left(\frac{\nabla(W_i\hat{f}_i)}{\hat{f}_i} - \frac{\nabla(W_i f_i)}{f_i}\right)\mathbf{1}_i = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Q}_i - Q_i\right)W_i\left(\frac{\nabla\hat{f}_i - \nabla f_i}{\hat{f}_i} + \nabla f_i\left(\frac{1}{\hat{f}_i} - \frac{1}{f_i}\right)\right)\mathbf{1}_i$. So by Assumption 4,

$$\sqrt{n}\|s.o.1\| = O_p\left(\sqrt{n}\|\hat{Q} - \hat{Q}\|_\infty\delta^{-1}\left(\|\nabla\hat{f} - \nabla f\|_\infty + \delta^{-1}\|\hat{f} - f\|_\infty\right)\right)$$
$$= O_p\left(\frac{\log n}{\delta^2\sqrt{nh_1^d h^d}}\left(h_1^{-1} + \delta^{-1}\right)\right) = o_p(1).$$

The influence function of (I) is implied by Theorem 3 with $\phi(X) = -\nabla(W(X)f(X))/f(X)$ by assuming $\delta^6 nh^{2d}h_0 \to \infty$, $\delta^8 nh^{2d} \to \infty$, and uniformly bounded functions.

The second term in (21) is further decomposed

$$-\frac{1}{n}\sum_{i=1}^{n} Q_i\frac{\nabla(W_i\hat{f}_i)}{\hat{f}_i}\mathbf{1}_i = -\frac{1}{n}\sum_{i=1}^{n} Q_i\nabla W_i\mathbf{1}_i$$
$$\underbrace{-\frac{1}{n}\sum_{i=1}^{n} Q_i\frac{W_i}{f_i}\nabla\hat{f}_i\mathbf{1}_i}_{(II)} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} Q_i\frac{W_i\nabla f_i}{f_i^2}\left(\hat{f}_i - f_i\right)\mathbf{1}_i}_{(III)} + s.o.2,$$

where $s.o.2 = \frac{1}{n}\sum_{i=1}^{n}\frac{Q_i}{f_i}\left(\hat{f}_i - f_i\right)\left(\frac{\nabla(W_i\hat{f}_i)}{\hat{f}_i} - \frac{\nabla(W_i f_i)}{f_i}\right)\mathbf{1}_i$. So $\|s.o.2\| = o_p(\|s.o.1\|)$.

The influence function of (II) is given by Lemma 2 with $\psi(X;\tau) = -Q(\tau|X)W(X)/f(X)$,

$$\sqrt{n}\left((II) + \mathbb{E}\left[\nabla f\frac{QW}{f}\right]\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(f_i\nabla\left(\frac{Q_i W_i}{f_i}\right) - \mathbb{E}\left[f_i\nabla\left(\frac{Q_i W_i}{f_i}\right)\right]\right) + o_p(1).$$

The influence function for (III) is given by Lemma 3 with $\gamma(X;\tau) = Q(\tau|X)W(X)\frac{\nabla f(X)}{f(X)^2}$,

$$\sqrt{n}(III) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{Q_i W_i}{f_i}\nabla f_i - \mathbb{E}\left[\frac{Q_i W_i}{f_i}\nabla f_i\right]\right) + o_p(1).$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

Therefore,

$$-\frac{1}{n}\sum_{i=1}^{n} Q_i \mathbf{1}_i \frac{\nabla(W_i \hat{f}_i)}{\hat{f}_i} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nabla Q_i W_i + o_p(1).$$

As argued in Step 4, Donsker's theorem implies the weak convergence. Combining the bias terms in Theorem 3 and Lemmas 2 and 3, we obtain $\text{Bias}_w(\tau; h, h_0, h_1)$. ■

Proof of Theorem 6. We follow Härdle and Stoker (1989) using the projection structure of the U-statistic. For $r_{\text{II}i}$ in Appendix C, let $\xi_{nij} = p(X_i, X_j; \lambda) + p(X_j, X_i; \lambda)$. The limit of $r_n(X_i; \lambda) = \mathbb{E}[\xi_{nij}|Z_i]$, denoted by $r_i \equiv r(X_i; \tau)$, is estimated by the sample analog, $\hat{r}_i \equiv (n-1)^{-1}\sum_{j\neq i}\hat{\xi}_{nij}$, where $\hat{\xi}_{nij}$ is obtained by a first-step nonparametric estimation of the unknown functions.

By a similar argument for trimming in Appendix C, it suffices to show consistency of $n^{-1}\sum \hat{r}_i\hat{r}_i'\mathbf{1}_i$ for $\mathbb{E}[rr']$ and $n^{-1}\sum \hat{r}_i\mathbf{1}_i$ for $\mathbb{E}[r]$. First, we show $\sup_i |\hat{r}_i - r_i|\mathbf{1}_i = o_p(1)$.

$$\sup_i |\hat{r}_i - r_i|\mathbf{1}_i \leq \sup_i \left|\frac{1}{n-1}\sum_{j\neq i}(\hat{\xi}_{nij} - \xi_{nij})\right|\mathbf{1}_i + \sup_i \left|\frac{1}{n-1}\sum_{j\neq i}\xi_{nij} - \mathbb{E}[\xi_{nij}|Z_i]\right|\mathbf{1}_i$$
$$+ \sup_i \left|r_n(X_i;\lambda) - r(X_i;\tau)\right|\mathbf{1}_i = o_p(1)$$

by the uniform convergence of the nonparametric estimation, the law of large numbers, and the proof of Lemma 2.

By similar arguments, we obtain $\sup_i |\hat{r}_{\beta i} - r_{\beta i}|\mathbf{1}_i = o_p(1)$. Let $r_{fi} \equiv r_f(Z_i; \tau)$. Since the variance of r_{fi} exists and $Pr(f(X) \leq \delta) = o(1)$,

$$n^{-1}\sum_{i=1}^{n} \hat{r}_{fi}\hat{r}_{fi}'\mathbf{1}_i - \mathbb{E}[r_{fi}r_{fi}']$$
$$= n^{-1}\sum_{i=1}^{n}(\hat{r}_{fi} - r_{fi})(\hat{r}_{fi} - r_{fi})'\mathbf{1}_i + n^{-1}\sum_{i=1}^{n} r_{fi}(\hat{r}_{fi} - r_{fi})'\mathbf{1}_i + n^{-1}\sum_{i=1}^{n}(\hat{r}_{fi} - r_{fi})\hat{r}_{fi}'\mathbf{1}_i$$
$$- n^{-1}\sum_{i=1}^{n}\hat{r}_{fi}\hat{r}_{fi}'(1 - \mathbf{1}_i) + n^{-1}\sum_{i=1}^{n} r_{fi}r_{fi}' - \mathbb{E}[r_{fi}r_{fi}'] = o_p(1).$$

The same arguments prove $\widehat{\text{Cov}}(\mathbb{G}_w(\tau_1), \mathbb{G}_w(\tau_2))$ is consistent. ■

Proof of Corollary 1. The results are implied by Propositions 4.1 and 4.2 in Powell and Stoker (1996), so we verify their conditions, note the additional complication due to estimating the CQF and density derivative, and do not repeat the proofs.

(i) $\mathsf{B}h^{\nu}$ includes the leading terms associated with h in Bias_f in Theorem 4. The derivation of Bias_f implies Assumption 1 in Powell and Stoker (1996).

Consider the variance. Following the proof of Theorem 4, (20) in Theorem 3 and (24) in Lemma 2 imply

$$p(Z_i, Z_j; \lambda) \equiv \frac{-2a'\nabla f(X_i)K\big(H^{-1}(X_i - X_j)\big)}{f(X_i)f_Y(Q(\tau|X_i)|X_i)|H|}\Big(\tau - G\Big(\frac{Q(\tau|X_i) - Y_j}{h_0}\Big)\Big)\mathbf{1}_i$$
$$- 2Q(\tau|X_i)\mathbf{1}_i a'\nabla K\Big(\frac{X_i - X_j}{h}\Big)\frac{1}{h^{d+1}}. \tag{22}$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

So $a'\hat{\beta}_{f,h} = (n(n-1))^{-1}\sum_{i=1}^{n}\sum_{j\neq i}p(Z_i,Z_j;\lambda) + s.o.$ for some smaller-order term $s.o = o_p(n^{-1/2})$ due to $\hat{Q}(\tau|X_i)$ and $\nabla\hat{f}(X_i)$. The standard formulation of U-statistic implies that the terms associated with h in the variance of $\hat{\beta}_{f,h}$ are dominated by $2n^{-2}\mathbb{E}\left[\tilde{p}(Z_i,Z_j;\lambda)^2\right]$ for a symmetric $\tilde{p}(Z_i,Z_j;\lambda) = (p(Z_i,Z_j;\lambda) + p(Z_j,Z_i;\lambda))/2.\hat{p}(Z_i,Z_j;\lambda)$ is a plug-in estimator of $\tilde{p}(Z_i,Z_j;\lambda)$. By (18) and (19), some algebra yields $\mathbb{E}\left[\tilde{p}(Z_i,Z_j;\lambda)^2\right] = \mathsf{V}h^{-d} + o(h^{-d})$. Thus, Assumption 2 in Powell and Stoker (1996) holds. Their Proposition 4.1 implies the result.

(ii) Consider the consistency of $\hat{\mathsf{V}}$. The proof of Theorem 4, Theorem 1, and $nh_v^{3d} \to \infty$implies $s.o. = o_p(n^{-1/2})$. It is straightforward to show that $\mathbb{E}[p(Z_i,Z_j;\lambda)^4] = O_p(h_v^{-3d})$ (i.e., $\eta = d$ and $\gamma = d$ in the notations and equation (4.39) of Powell and Stoker (1996)). Then, Proposition 4.2 of Powell and Stoker (1996) implies the result.

■

C. Proof of Lemmas

Proof of Lemma 1. This proof modifies the proofs of Theorems 2 and 6 in Hansen (2008) and Lemma B(1) in Cattaneo et al. (2013) that amends a truncation argument in the proof of Theorem 2 in Hansen (2008).

1. $\hat{F}_Y(y|x) \equiv \hat{\Psi}(y,x)/\hat{f}(x) = \frac{\hat{\Psi}(y,x)/f(x)}{\hat{f}(x)/f(x)}$, where $\hat{\Psi}(y,x) \equiv (nh^d)^{-1}\sum_{i=1}^{n}K(H^{-1}(X_i - x))G\left(\frac{y-Y_i}{h_0}\right)$, for $x \in \mathcal{C}_n$ and $y \in \mathcal{Y}$.Consider the uniformity over $\mathcal{Y}\times\mathcal{C}_n$. Construct a grid using regions of the form $B_j = \{y : |y - y_j| \leq a_n h\} \times \{x : |x - x_j| \leq a_n h\}$. Since we have a bounded dependent variable $G\left(\frac{y-Y_i}{h_0}\right) \in (0,1)$, the argument for uniform bound is the same, for example, in (A.8) of Hansen (2008) $\left|K(x_2)G\left(\frac{y_2-Y}{h_0}\right) - K(x_1)G\left(\frac{y_1-Y}{h_0}\right)\right| \leq \zeta K^*(x_1)$, for all $\|(y_1,x_1) - (y_2,x_2)\| \leq \zeta \leq L$. Therefore, Hansen's proof of Theorem 2 gives $\sup_{x\in\mathcal{C}_n, y\in\mathcal{Y}_n}|\hat{\Psi}(y,x) - \mathbb{E}[\hat{\Psi}(y,x)]| = O_p\left(\left(\log n/\left(nh^d\right)\right)^{1/2}\right)$. By change of variables, the smoothness assumptions, and the dominated convergence theorem, for any $y \in \mathcal{Y}_n$ and $x \in \mathcal{C}_n$,

$$\begin{aligned}\mathbb{E}\left[\hat{\Psi}(y,x)\right] &= \frac{1}{h^d}\mathbb{E}\left[K(H^{-1}(X-x))\mathbb{E}\left[G\left(\frac{y-Y}{h_0}\right)\middle|x\right]\right]\\ &= \frac{1}{h^d}\mathbb{E}\left[K(H^{-1}(X-x))\left(F_Y(y|X) + \frac{h_0^2}{2}\kappa_{G2}f_Y'(y|X) + o(h_0^2)\right)\right]\\ &= F_Y(y|x)f(x) + O(h^\nu + h_0^2).\end{aligned}$$

Thus, $\sup_{x\in\mathcal{C}_n, y\in\mathcal{Y}}|\hat{\Psi}(y,x) - \Psi(y,x)| = O_p(a^\dagger)$, where $a^\dagger \equiv \left(\frac{\log n}{nh^d}\right)^{1/2} + h_0^2 + h^\nu$ and $\Psi(y,x) \equiv F_Y(y|x)f(x)$. Theorem 6 in Hansen (2008) gives

$$\sup_{x\in\mathcal{C}_n}|\hat{f}(x) - f(x)| = O_p(a^*), \text{ where } a^* \equiv \left(\frac{\log n}{nh^d}\right)^{1/2} + h^\nu. \tag{23}$$

Therefore, uniformly in $y \in \mathcal{Y}$ and $x \in \mathcal{C}_n$, $\hat{F}_Y(y|x) = \frac{\hat{\Psi}(y,x)/f(x)}{\hat{f}(x)/f(x)} = \frac{F_Y(y|x) + O_p(a^\dagger\delta^{-1})}{1 + O_p(a^*\delta^{-1})} = F_Y(y|x) + O_p(a^\dagger\delta^{-1})$.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

2. Similarly, $\frac{\partial}{\partial y}\hat{F}_Y(y|x) \equiv \hat{f}_Y(y|x) \equiv \hat{\Psi}(y,x)/\hat{f}(x)$, where $\hat{\Psi}(y,x) \equiv (nh^d h_0)^{-1}\sum_{i=1}^n g(\frac{y-Y_i}{h_0})\,K(H^{-1}(X_i-x))$. Theorem 6 in Hansen (2008) implies $\sup_{x\in\mathcal{C}_n, y\in\mathcal{Y}}\left|\hat{\Psi}(y,x) - \mathbb{E}\left[\hat{\Psi}(y,x)\right]\right| = O_p\left(\left(\frac{\log n}{nh^d h_0}\right)^{1/2}\right)$. A similar calculation yields $\text{Bias}_\phi(\hat{\Psi}(y,x)) = O(h^\nu + h_0^2)$.

■

Proof of Lemma 2. We suppress the subscript in h_1 and ν_1 for notational simplicity. Denote $\lambda \equiv (\tau, h) \in \mathcal{T}\times\mathcal{R}_+$.

$$U_n \equiv \frac{1}{n}\sum_{i=1}^n \psi(X_i;\tau)\nabla\hat{f}_i\mathbf{1}_i = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n p(X_i,X_j;\lambda), \text{ where}$$

$$p(X_i,X_j;\lambda) \equiv \psi(X_i;\tau)\mathbf{1}_i\nabla K\left(\frac{X_i-X_j}{h}\right)\frac{1}{h^{d+1}}. \tag{24}$$

The proof follows the procedure outlined at the beginning of Appendix B.

Step 1. Using the same arguments in Proof of Theorem 3, we show $\mathcal{F}_1 \equiv \{(Z_i,Z_j)\mapsto h^{d+1}p(Z_i,Z_j;\lambda) : \lambda\in\Lambda\}$ is euclidean for the envelope $F_1(Z_i,Z_j) = \sup_{\tau\in\mathcal{T}}\psi(X_i;\tau)$ satisfying $\mathbb{P}^2F_1^2 < \infty$. Assume $nh^{2d+2}\to\infty$ and apply Corollary 4 in Sherman (1994). Then, Lemma 6 in Sherman (1994) implies the class of $\mathbb{P}$-degenerate functions of order 2 $\{(Z_i,Z_j)\mapsto h^{d+1}r(Z_i,Z_j;\lambda) : \lambda\in\Lambda\}$ is euclidean.

Step 2.

$$\begin{aligned}
r_n(X_i;\lambda) &\equiv \mathbb{E}[p(X_i,X_j;\lambda)|X_i] + \mathbb{E}[p(X_j,X_i;\lambda)|X_i]\\
&= \int_{\mathcal{X}}\frac{1}{h^{d+1}}\nabla K_{ij}f(X_j)dX_j\psi_i\mathbf{1}_i + \int_{\mathcal{X}}\frac{1}{h^{d+1}}\nabla K_{ji}\psi_j f(X_j)\mathbf{1}_j dX_j\\
&= \psi_i\mathbf{1}_i\left(\frac{-1}{h^d}K_{ij}f(X_j)\Big|_{\mathcal{X}} + \frac{1}{h^d}\int_{\mathcal{X}}K_{ij}\nabla f(X_j)dX_j\right)\\
&\quad + \frac{1}{h^d}\left(K_{ij}f(X_j)\psi_j\Big|_{\mathcal{X}} - \int_{\mathcal{X}}K_{ji}\nabla(f_j\psi_j)dX_j\right)\\
&= \psi_i\mathbf{1}_i\int K(V)\nabla f(X_i+HV)dV - \int K(V)\nabla(f\psi)(X_i+HV)\mathbf{1}\{X_i+HV\in\mathcal{S}\}dV\\
&= r(X_i;\tau) + O_p(h^\nu).
\end{aligned}$$

Let $r(X_i;\tau) \equiv \psi(X_i;\tau)\nabla f_i - \nabla(\psi(X_i;\tau)f_i) = -f_i\nabla\psi(X_i;\tau)$. The conditions imply $\mathbb{E}\|r_n(X;\lambda) - r(X_i;\tau)\|^2 = \mathbb{E}[\psi(X;\tau)^2]O(h_1^{2\nu_1}) = o(1)$, so $n^{-1/2}\sum_{i=1}^n(r_n(X_i;\lambda) - r(X_i;\tau)) = o_p(1)$.

Let $r(X_i;\tau) = 2\gamma(X_i;\tau)f(X_i)$. By similar arguments in Proof of Theorem 3, $\mathbb{E}\|r_n(X;\lambda) - r(X_i;\tau)\|^2 = O\left(h_1^{2\nu_1}\right) = o(1)$, so $n^{-1/2}\sum_{i=1}^n(r_n(X_i;\lambda) - r(X_i;\tau)) = o_p(1)$.

Step 3 [Bias]. Assuming $p_x > \nu + 2$,

$$\begin{aligned}
\mathbb{P}^2p(X_i,X_j;\lambda) &= \mathbb{E}\left[\psi(X;\tau)\mathbf{1}\{X\in\mathcal{S}\}\int K(u)\nabla f(X+uh)du\right]\\
&= \mathbb{E}\left[\psi(X;\tau)\nabla f(X)\right] + \frac{h^\nu\kappa_\nu}{\nu!}\sum_{k=1}^d\mathbb{E}\left[\psi(X;\tau)\partial_k^\nu\nabla f(X)\right] + O_p(h^{\nu+1}).
\end{aligned}$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

Thus, $n^{-1}\sum_{i=1}^n \psi(X_i;\tau)\nabla\hat{f}_i\mathbf{1}_i = n^{-1}\sum_{i=1}^n -f_i\nabla\psi_i - \mathbb{E}[\psi\nabla f]$. By integration by parts, $-\mathbb{E}[\psi\nabla f] = \mathbb{E}[\nabla(\psi f)] = \mathbb{E}[\nabla\psi f] + \mathbb{E}[\psi\nabla f)]$. Thus, $-2\mathbb{E}[\psi\nabla f] = \mathbb{E}[\nabla\psi f]$. ■

Proof of Lemma 3. By the same steps as in the proof of Lemma 2, define $p(X_i,X_j;\lambda) \equiv K_h(X_i - X_j)\gamma(X_i;\tau)\mathbf{1}_i$. We only note the difference without repeating each step.

Step 2.

$$\begin{aligned} r_n(X_i;\lambda) &\equiv \mathbb{E}[p(X_i,X;\lambda)|X_i] + \mathbb{E}[p(X,X_i;\lambda)|X_i] \\ &= \int_{\mathcal{X}} \frac{1}{h^d} K_{ij} f(X_j) dX_j \gamma_i \mathbf{1}_i + \int_{\mathcal{X}} \frac{1}{h^d} K_{ij}\gamma_j f_j \mathbf{1}_j dX_j \\ &= \gamma_i \mathbf{1}_i \int K(V) f(X_i + HV) dV + \int K(V)\big(\gamma f\big)(X_i + HV)\mathbf{1}\{X_i + HV \in \mathcal{S}\} dV. \end{aligned}$$

Let $r(X_i;\tau) = 2\gamma(X_i;\tau)f(X_i)$. The conditions imply $\mathbb{E}\|r_n(X;\lambda) - r(X_i;\tau)\|^2 = O\left(h_1^{2\nu_1}\right) = o(1)$, so $n^{-1/2}\sum_{i=1}^n (r_n(X_i;\lambda) - r(X_i;\tau)) = o_p(1)$. Then, $n^{-1}\sum_{i=1}^n \gamma(X_i;\tau)\hat{f}(X_i)\mathbf{1}_i - \mathbb{E}[\gamma(X;\tau)\, f(X)] = n^{-1}\sum_{i=1}^n 2\gamma(X_i;\tau)f(X_i) - 2\mathbb{E}[\gamma(X;\tau)f(X)] + o_p(n^{-1/2})$.

Step 3 [Bias]. Assuming $p_x \geq \nu + 1$,

$$\begin{aligned} \mathbb{E}[r_n(X;\lambda)] &= \mathbb{E}\left[\gamma(X;\tau)\int K(u)f(X+uh)du\right] \\ &= \mathbb{E}\left[\gamma(X;\tau)f(X)\right] + \frac{h^\nu \kappa_\nu}{\nu!}\mathbb{E}\left[\gamma(X;\tau)\sum_{k=1}^d \partial_k^\nu f(X)\right] + O_p(h^{\nu+1}). \end{aligned}$$

■

Proof of Lemma 4. Following Lavergne and Vuong (1996), choose ϵ_n such that $\epsilon_n^{-1}\sup_{X\in\mathcal{S},\tau\in\mathcal{T}}|\hat{f}_{XY}(X,\hat{Q}(\tau|X)) - f_{XY}(X,Q(\tau|X))| = o_p(1)$ and $\epsilon_n/\delta = o(1)$, which exists because

$$\begin{aligned} &\sup_{X\in\mathcal{S},\tau\in\mathcal{T}} |\hat{f}_{XY}(X,\hat{Q}(\tau|X)) - f_{XY}(X,Q(\tau|X))| \\ &\leq \sup_{X\in\mathcal{S},\tau\in\mathcal{T}} |\hat{f}_{XY}(X,\hat{Q}(\tau|X)) - f_{XY}(X,\hat{Q}(\tau|X))| \\ &\quad + \sup_{X\in\mathcal{S},\tau\in\mathcal{T}} |f_{XY}(X,\hat{Q}(\tau|X)) - f_{XY}(X,Q(\tau|X))| \\ &= O_p((\log n/(nh^d))^{1/2}(h_0^{-1/2} + \delta^{-1})). \end{aligned}$$

Since $\epsilon_n/\delta = o(1)$, we can work with the bound $\delta + \epsilon_n$ instead of δ. Define $\mathcal{S}_c \equiv \{x : \inf_{\tau\in\mathcal{T}} f_{XY}(x,Q(\tau|x)) \geq \delta + \epsilon_n\}$ and $\mathcal{S}_{c-} \equiv \{x : \inf_{\tau\in\mathcal{T}} f_{XY}(x,Q(\tau|x)) \geq \delta - \epsilon_n\}$.

$$\begin{aligned} \sqrt{n}(\tilde{\beta} - \hat{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Q}_i\hat{\phi}_i\Big(\mathbf{1}\{X_i \in \mathcal{S}_c\} - \mathbf{1}\{X_i \in \hat{\mathcal{S}}\}\Big) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{Q}_i\hat{\phi}_i\Big(\mathbf{1}\{X_i \in \mathcal{S}_c, X_i \notin \hat{\mathcal{S}}\} - \mathbf{1}\{X_i \notin \mathcal{S}_c, X_i \in \hat{\mathcal{S}}\}\Big). \end{aligned}$$

For any $x \in \mathcal{X}$and $\tau \in \mathcal{T}$, the event $\left\{x \in \mathcal{S}_c, x \notin \hat{\mathcal{S}}\right\} \subseteq \left\{|\hat{f}_{XY}(x,\hat{Q}) - f_{XY}(x,Q)| > \epsilon_n, x \in \mathcal{S}_c\right\} \subseteq \left\{\sup_{x\in\mathcal{X},\tau\in\mathcal{T}}|\hat{f}_{XY}(x,\hat{Q}) - f_{XY}(x,Q)|\mathbf{1}\{x \in \mathcal{S}\} > \epsilon_n\right\}$has asymptotic probability zero.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

Hence, $\mathbf{1}\{X_i \in \mathcal{S}_c, X_i \notin \hat{\mathcal{S}}\} = 0 w.p.a.1$, for all i. So we need to consider the second term only. Define $I_{ni} \equiv \mathbf{1}\{X_i \notin \mathcal{S}_c, X_i \in \hat{\mathcal{S}}\}$.

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{Q}_i \hat{\phi}_i I_{ni} \right\| \\
&\leq \sup_{\tau\in\mathcal{T}} \left\| n^{-1/2} \sum_{i=1}^{n} (\hat{Q}_i - Q_i)\phi_i I_{ni} \right\| + \sup_{\tau\in\mathcal{T}} \left\| n^{-1/2} \sum_{i=1}^{n} Q_i (\hat{\phi}_i - \phi_i) I_{ni} \right\| \\
&\quad + \sup_{\tau\in\mathcal{T}} \left\| n^{-1/2} \sum_{i=1}^{n} (\hat{Q}_i - Q_i)(\hat{\phi}_i - \phi_i) I_{ni} \right\| + \sup_{\tau\in\mathcal{T}} \left\| n^{-1/2} \sum_{i=1}^{n} Q_i \phi_i I_{ni} \right\|. \qquad (25)
\end{aligned}$$

For the last term in (25),

$$\begin{aligned}
&\mathbb{E}\left[\left\| n^{-1/2} \sum_{i=1}^{n} Q_i \phi_i I_{ni} \right\|^2 \right] \\
&\leq n^{-1} \mathbb{E}\left[\left(\sum_{i=1}^{n} \left\| Q_i \phi_i \right\| \mathbf{1}\{X_i \notin \mathcal{S}_c\} \right)^2 \right] \\
&= \mathbb{E}\left[\left\| Q_i \phi_i \right\|^2 \mathbf{1}\{X_i \notin \mathcal{S}_c\} \right] + (n-1) \left(\mathbb{E}\left[\left\| Q_i \phi_i \right\| \mathbf{1}\{X_i \notin \mathcal{S}_c\} \right] \right)^2 = o(1),
\end{aligned}$$

where the first term is $o(1)$ by Lebesgue dominated convergence theorem with $\mathbb{E}\|Q_i\phi_i\|^2 \leq \infty$ and $\delta + \epsilon_n \to 0$. For the second term, $\int_{B_{\epsilon_n}} \|Q_i\phi_i\| f_i dX_i = o(n^{-1/2})$ in Assumption 3 or 4, where $B_{\epsilon_n} \equiv \{X : \sup_{\tau\in\mathcal{T}} f_{XY}(X, Q(\tau|X)) < \delta + \epsilon_n\}$.

Now, consider the first term of (25). The event $\left\{ x \notin \mathcal{S}_c, x \in \hat{\mathcal{S}}, |\hat{f}_{XY}(x, \hat{Q}) - f_{XY}(x, Q)| > \epsilon_n \right\} \subseteq \left\{ |\hat{f}_{XY}(x, \hat{Q}) - f_{XY}(x, Q)| > \epsilon_n \right\}$has asymptotic probability zero. Observe that $\{x \notin \mathcal{S}_c, x \in \hat{\mathcal{S}}, |\hat{f}_{XY}(x, \hat{Q}) - f_{XY}(x, Q)| \leq \epsilon_n\} \subseteq \{x \in \mathcal{S}_{c-}\}$.

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}} \left\| n^{-1/2} \sum_{i=1}^{n} \phi_i (\hat{Q}_i - Q_i) I_{ni} \right\| \\
&\leq \sup_{\tau\in\mathcal{T}} n^{-1/2} \sum_{i=1}^{n} \left\| \phi_i (\hat{Q}_i - Q_i) \right\| I_{ni} \\
&\leq \sup_{\tau\in\mathcal{T}, x\in\mathcal{S}} \left\{ \left| \hat{Q}(\tau|x) - Q(\tau|x) \right| \mathbf{1}\{x \in \mathcal{S}_{c-}\} \right\} n^{-1/2} \sum_{i=1}^{n} \|\phi_i\| I_{ni} \\
&= O_p(\delta^{-1}(nh^d)^{-1/2}) O_p(1) = o_p(1),
\end{aligned}$$

where $n^{-1/2} \sum_{i=1}^{n} \|\phi_i\| I_{ni} = O_p(1)$ by the central limiting theorem by $\mathbb{E}\|\phi_i\|^2 < \infty$, and the uniform convergence of $\hat{Q}_i$ is implied by Theorem 1 and Assumption 3 or 4. By the similar arguments, the remaining terms of (25) vanish in probability. ■

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

Proof of Lemma 5. Define the operator Γ_ν on a function $f:\mathcal{V}\to\mathcal{R}$, where $\mathcal{V}$ is an open and convex subset of $\mathcal{R}^d$ by

$$\Gamma_\nu f(X+HV)\equiv\Gamma_\nu f(\bar{X})=h\sum_{k=1}^{d}\partial_k f(X)V_k+\frac{h^2}{2}\sum_{k_1=1}^{d}\sum_{k_2=1}^{d}\Big[\partial_{k_1}\partial_{k_2}f(X)\Big]V_{k_1}V_{k_2}+\dots$$
$$+\frac{h^{\nu-1}}{(\nu-1)!}\sum_{k_1=1}^{d}\cdots\sum_{k_{\nu-1}=1}^{d}\Big[\partial_{k_1}\cdots\partial_{k_{\nu-1}}f(X)\Big]V_{k_1}\cdots V_{k_{\nu-1}}$$
$$+\frac{h^\nu}{\nu!}\sum_{k_1=1}^{d}\cdots\sum_{k_\nu=1}^{d}\Big[\partial_{k_1}\cdots\partial_{k_\nu}f(\bar{X})\Big]V_{k_1}\cdots V_{k_\nu},$$

where V_k is the kth component of the vector V and $\bar{X}$ is on the line segment of X and $X+HV$. Hence, Taylor's theorem expands $f(X+HV)=f(X)+\Gamma_\nu f(X+HV)$ for small H. By (18), for all X_i,

$$\mathbb{E}\left[\frac{1}{|H|}K_{ij}(G_{ij}-\tau)\Big|x_i\right]$$
$$=\int_{\mathcal{X}}\frac{1}{|H|}K_{ij}\left(-\tau+F_Y(Q_i|X_j)+\frac{h_0^2}{2}f_Y'(Q_i|X_j)\kappa_{G2}+O(h_0^3)\right)f(X_j)dX_j$$
$$=\int_{\mathcal{V}}K(V)\left(-\tau+F_Y(Q_i|X_i+HV)+\frac{h_0^2}{2}f_Y'(Q_i|X_i+HV)\kappa_{G2}+O(h_0^3)\right)f(X_i+HV)dV$$
$$=\int_{\mathcal{V}}K(V)\Big(-\tau+F_Y(Q_i|X_i)+\Gamma_\nu F_Y(Q_i|\bar{X}_i)$$
$$+\frac{h_0^2}{2}\kappa_{G2}\Big(f_Y'(Q_i|X_i)+\Gamma_\nu f_Y'(Q_i|\bar{X}_i)\Big)+O(h_0^3)\Big)$$
$$\Big(f(X_i)+\Gamma_\nu f(\bar{X}_i)\Big)dV$$
$$=\int K(V)\Big(\Gamma_\nu F_Y(Q_i|\bar{X}_i)f_i+\Gamma_\nu F_Y(Q_i|\bar{X}_i)\Gamma_\nu f(\bar{X}_i)+\frac{h_0^2}{2}\kappa_{G2}f_Y'(Q_i|X_i)\Gamma_\nu f(\bar{X}_i)$$
$$+\frac{h_0^2}{2}\kappa_{G2}\Gamma_\nu f_Y'(Q_i|\bar{X}_i)f_i+\frac{h_0^2}{2}\kappa_{G2}\Gamma_\nu f_Y'(Q_i|\bar{X}_i)\Gamma_\nu f(\bar{X}_i)\Big)dV+\frac{h_0^2}{2}\kappa_{G2}f_Y'(Q_i|X_i)f_i+O(h_0^3)$$
$$=\frac{h_0^2}{2}\kappa_{G2}f_Y'(Q_i|X_i)f_i+\frac{h^\nu}{\nu!}\kappa_\nu\sum_{k=1}^{d}\partial_k^\nu F_Y(Q_i|X_i)f_i+\int K(V)\Gamma_\nu F_Y(Q_i|\bar{X}_i)\Gamma_\nu f(\bar{X}_i)dV$$
$$+\frac{h_0^2}{2}\kappa_{G2}f_Y'(Q_i|X_i)\frac{h^\nu}{\nu!}\kappa_\nu\sum_{k=1}^{d}\partial_k^\nu f_i+\frac{h_0^2}{2}\kappa_{G2}f_i\frac{h^\nu}{\nu!}\kappa_\nu\sum_{k=1}^{d}\partial_k^\nu f_Y'(Q_i|X_i)$$
$$+\frac{h_0^2}{2}\kappa_{G2}\int K(V)\Gamma_\nu f_Y'(Q_i|\bar{X}_i)\Gamma_\nu f(\bar{X}_i)dV+O(h^{\nu+1}+h_0^3)$$
$$=\frac{h_0^2\kappa_{G2}}{2}f_Y'(Q_i|X_i)f_i+\frac{h^\nu\kappa_\nu}{\nu!}\sum_{k=1}^{d}\partial_k^\nu F_Y(Q_i|X_i)f_i$$
$$+\sum_{l=1}^{\nu-1}\frac{h^\nu\kappa_\nu}{l!\,(\nu-l)!}\sum_{k=1}^{d}\partial_k^l F_Y(Q_i|X_i)\partial_k^{\nu-l}f(X_i)+h^\nu h_0^2R_I(X_i)+O(h_0^3+h^{\nu+1}),$$

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

where $\bar{X}_i$ is on the line segment between X_i and $X_i + HV$, by the dominated convergence theorem and the uniform continuity of $\partial_{k_1} \cdots \partial_{k_\nu} f(X)$ and $\partial_{k_1} \cdots \partial_{k_\nu} f'_Y(y|X)$ for $k_1, \ldots, k_\nu \in \{1, \ldots, d\}$ in X.

Since $\mathbf{1}\{X_i \notin \mathcal{S}\} = o(1)$ by $\delta \to 0$ and the moments exist, by the dominated convergence theorem,

$$\begin{aligned}\mathbb{E}[U_n p] = -\mathbb{E}\Bigg[\frac{\phi_i}{f_i f_Y(Q_i|X_i)}\Bigg\{\frac{h_0^2}{2}\kappa_{G2} f'_Y(Q_i|X_i) f_i \\ + h^\nu \kappa_\nu \sum_{l=1}^{\nu} \frac{1}{l!\,(\nu - l)!} \sum_{k=1}^{d} \partial_k^l F_Y(Q_i|X_i) \partial_k^{\nu - l} f(X_i) \\ + h^\nu h_0^2 R_I(X_i)\Bigg\}\Bigg] + o(h^\nu + h_0^2) = O(h^\nu + h_0^2).\end{aligned}$$

■

REFERENCES

Andrews, D. W. K. & X. Shi (2013) Inference based on conditional moment inequalities. *Econometrica* 81(2), 609–666.

Belloni, A., V. Chernozhukov, D. Chetverikov, & I. Fernández-Val (2019) Conditional quantile processes based on series or many regressors. *Journal of Econometrics* 213(1), 4–29. Annals: In Honor of Roger Koenker.

Bhattacharya, P. K. & A. K. Gangopadhyay (1990) Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics* 18(3), 1400–1415.

Calonico, S., M. D. Cattaneo & M. H. Farrell (2018) On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113(522), 767–779.

Cattaneo, M. D., R. K. Crump & M. Jansson (2010) Robust data-driven inference for density-weighted average derivatives. *Journal of the American Statistical Association* 105(491), 1070–1083.

Cattaneo, M. D., R. K. Crump & M. Jansson (2013) Generalized Jackknife estimators of weighted average derivatives. *Journal of the American Statistical Association* 108(504), 1243–1256.

Cattaneo, M. D., R. K. Crump & M. Jansson (2014a) Bootstrapping density-weighted average derivatives. *Econometric Theory* 30(6), 1135–1164.

Cattaneo, M. D., R. K. Crump & M. Jansson (2014b) Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory* 30(1), 176–200.

Cattaneo, M. D. & M. Jansson (2018) Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica* 86(3), 955–995.

Chaudhuri, P., K. Doksum & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25(2), 715–744.

Chernozhukov, V., I. Fernández-Val & A. Galichon (2010) Quantile and probability curves without crossing. *Econometrica* 78(3), 1093–1125.

Chernozhukov, V., I. Fernández-Val & B. Melly (2013) Inference on counterfactual distributions. *Econometrica* 81(6), 2205–2268.

Chernozhukov, V. & C. Hansen (2005) An IV model of quantile treatment effects. *Econometrica* 73(1), 245–261.

Chesher, A. (2003) Identification in nonseparable models. *Econometrica* 71(5), 1405–1441.

Dabrowska, D. M. (1992) Nonparametric quantile regression with censored data. *Sankhy: The Indian Journal of Statistics, Series A (1961–2002)* 54(2), 252–259.

https://doi.org/10.1017/S0266466621000232 Published online by Cambridge University Press

Escanciano, J. C., D. T. Jacho-Chávez & A. Lewbel (2014) Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics* 178(3), 426–443.

Fan, Y. & E. Guerre (2016) Multivariate local polynomial estimators: Uniform boundary properties and asymptotic linear representation. *Essays in Honor of Aman Ullah, Advances in Econometrics* 14, 489–537.

Fan, Y. & R. Liu (2016) A direct approach to inference in nonparametric and semiparametric quantile models. *Journal of Econometrics* 191(1), 196–216.

Graham, B., J. Hahn, A. Poirier & J. Powell (2018) A quantile correlated random coefficients panel data model. *Journal of Econometrics* 206(2), 303–335.

Guerre, E., I. Perrigne & Q. Vuong (2000) Optimal nonparametric estimation of first-price auctions. *Econometrica* 68(3), 525–574.

Guerre, E. & C. Sabbah (2012) Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* 28, 87–129.

Hansen, B. E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24(3), 726–748.

Härdle, W. & T. M. Stoker (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* 84(408), 986–995.

Hoderlein, S. & E. Mammen (2007) Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75(5), 1513–1518.

Hoderlein, S. & E. Mammen (2009) Identification and estimation of local average derivatives in nonseparable models without monotonicity. *The Econometrics Journal* 12(1), 1–25.

Ichimura, H. & P.E. Todd (2007) Chapter 74: Implementing nonparametric and semiparametric estimators. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics.* Handbook of Econometrics, vol. 6. Elsevier.

Khan, S. (2001) Two-stage rank estimation of quantile index models. *Journal of Econometrics* 100(2), 319–355.

Khan, S. & E. Tamer (2010) Irregular identification, support conditions, and inverse weight estimation. *Econometrica* 78(6), 2021–2042.

Koenker, R. & G. Bassett (1978) Regression quantile. *Econometrica* 46, 33–50.

Kong, E., O. Linton & Y. Xia (2010) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* 26, 1529–1564.

Kong, E. & Y. Xia (2012) A single-index quantile regression model and its estimation. *Econometric Theory* 28, 730–768.

Kosorok, M. R. (2008) *Introduction to Empirical Processes and Semiparametric Inference* (Springer).

Lavergne, P. & Q. H. Vuong (1996) Nonparametric selection of regressors: The nonnested case. *Econometrica* 64(1), 207–219.

Lee, S. (2003) Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory* 19(1), 1–31.

Ma, X. & J. Wang (2019) Robust inference using inverse probability weighting. *Journal of the American Statistical Association* 115(532), 1851–1860.

Marmer, V. & A. Shneyerov (2012) Quantile-based nonparametric inference for first-price auctions. *Journal of Econometrics* 167(2), 345–357.

Matzkin, R. L. (2007) Nonparametric identification. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*. Elsevier, 5307–5368.

Newey, W. & T. M. Stoker (1993) Efficiency of weighted average derivative estimators and index models. *Econometrica* 61(5), 1199–1223.

Newey, W. K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5(2), 99–135.

Nishiyama, Y. & P. M. Robinson (2000) Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* 68(4), 931–980.

Pakes, A. & D. Pollard (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57(5), 1027–1057.

Phillips, P. C. (2015) Halbert White Jr. memorial JFEC lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics* 13(3), 521–555.

Pollard, D. (1990) Empirical Processes: Theory and Applications. NSF - CBMS Regional Conference Series in Probability and Statistics, Volume 2, IMS, Hayward, American Statistical Association, Alexandria.

Powell, J. L., J. H. Stock & T. M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrica* 57(6), 1403–1430.

Powell, J. L. & T. M. Stoker (1996) Optimal bandwidth choice for density-weighted averages. *Journal of Econometrics* 75(2), 291–316.

Qu, Z. & J. Yoon (2015) Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics* 185(1), 1–19.

Robinson, P. M. (1988) Root-n-consistent semiparametric regression. *Econometrica* 56(4), 931–954.

Rothe, C. (2010) Nonparametric estimation of distributional policy effects. *Journal of Econometrics* 155(1), 56–70.

Sasaki, Y. (2015) What do quantile regressions identify for general structural functions? *Econometric Theory* 31(5), 1102–1116.

Sasaki, Y. & T. Ura (2021) Estimation and inference for moments of ratios with robustness against large trimming bias. *Econometric Theory*, first published online 23 February 2021; https://doi.org/10.1017/S0266466621000025.

Schafgans, M. & V. Zinde-Walsh (2010) Smoothness adaptive average derivative estimation. *Econometrics Journal* 13(1), 40–62.

Sherman, R. (1994) Maximal inequalities for degenerate *U*-processes with applications to optimization estimators. *Annals of Statistics* 22(1), 439–459.

Silverman, B. (1986) *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall.

Stoker, T. M. (1986) Consistent estimation of scaled coefficients. *Econometrica* 54(6), 1461–1481.

van der Vaart, A. (2000) *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wu, T. Z., K. Yu & Y. Yu (2010) Single-index quantile regression. *Journal of Multivariate Analysis* 101(7), 1607–1621.