Econometric Theory, **32**, 2016, 458–497.
doi:10.1017/S0266466614000887

SEMIPARAMETRIC ESTIMATION OF PARTIALLY LINEAR TRANSFORMATION MODELS UNDER CONDITIONAL QUANTILE RESTRICTION

ZHENGYU ZHANG
Shanghai University of Finance and Economics and Shanghai Academy of Social Sciences

This article is concerned with semiparametric estimation of a partially linear transformation model under conditional quantile restriction with no parametric restriction imposed either on the link functional form or on the error term distribution. We describe for the finite-dimensional parameter a $\sqrt{n}$-consistent estimator which combines the features of Chen (2010)'s maximum integrated score estimator as well as Lee (2003)'s average quantile regression. We show the remaining two infinite-dimensional unknown functions in the model can be separately identified and propose estimators for these functions based on the marginal integration method. Furthermore, a simple approach is proposed to estimate the average partial quantile effect. Two important extensions, i.e., random censoring as well as estimating a transformation model with an endogenous regressor are also considered.

1. INTRODUCTION

Since the seminal work by Box and Cox (1964), taking transformations of the data has been a useful approach to model estimation and inference, particularly in the analysis of survival data. In the literature of economics, survival analysis is also called duration analysis, which is relevant in a variety of applications. This is because many time-to-event variables are of interest to researchers conducting empirical studies in economics, finance, and management. These time-to-event variables can be the length of an unemployment spell or a strike in labor economics, the time between purchases of a durable good in marketing, the duration of business cycles or recessions in macroeconomics and the survival time in biostatistics. See Lancaster (1990) and Van den Berg (2001) for extensive surveys on these topics. A baseline transformation model can be specified as

$$H(Y^*) = X'\beta_0 - U, \tag{1.1}$$

I am grateful to the co-editor Oliver Linton and an anonymous referee for their constructive comments. Zhengyu Zhang is also affiliated with the Key Laboratory of Mathematical Economics (SUFE), Ministry of Education. The research is supported by the National Science Foundation (Grant No. 71271139). Address correspondence to Zhengyu Zhang, Shanghai University of Finance and Economics, 777, Guoding Road, 200433 Shanghai, China; e-mail: zy.zhang@mail.shufe.edu.cn

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

where Y^* is the latent scalar dependent variable, subject to censoring of certain type; $H(\cdot)$ is an unknown monotone increasing function; X is a p_X-dimensional vector of covariates; β_0 is a conformable vector of coefficients; and U is the error term.[1] This linear transformation model (1.1) forms a rich class and incorporates the proportional hazards model (when U follows the extreme value distribution) and the proportional odds model (when U follows the logistic distribution) as special cases. In addition, if U follows the standard normal distribution, it naturally generalizes the classical Box–Cox model. The statistics and econometrics literature studying variations of this transformation model is quite extensive and can be classified by the set of assumptions that a given paper imposes on the link functional form, the error term distribution, as well as the censoring mechanism. Recent development in econometrics shows increased interest in distribution-free estimation and inference of the model, which requires no parametric specification either for the error term distribution or the link functional form, and can accommodate general forms of censoring. These studies include Asparouhova, Golanski, Kasprzyk, Sherman, and Asparouhov (2002), Abrevaya (1999, 2003), Horowitz (1996), Cavanagh and Sherman (1998), Chen (2002, 2010), Chen, Jin, and Ying (2002) and Khan and Tamer (2007), and Shin (2010), to mention only a few.

Despite these accomplishments, linearity of covariate effect after transformation may be too strong. There are several occasions calling for relaxation of this strong assumption. First, in duration analysis, it is more often than not that at least one covariate effect might be nonlinear. For example, in the analysis of unemployment spell data from continuous wage and benefit history unemployment insurance administrative records (Meyer, 1990), the covariate age showed a strong nonlinear effect on the individual hazards.[2] In general, modeling time-to-event variables through a partially linear structure can be viewed as an effective scheme for lessening the possibility of delivering misleading statistical inference due to model misspecification. Second, for nonlinear econometric models such as (1.1), a nontrivial and important issue is the presence of endogenous regressors. However, the existing literature lacks formal studies concerning estimation of a transformation model with endogeneity. When (1.1) contains an endogenous regressor, the endogeneity can be adequately corrected via estimating a partially linear transformation model, by adopting a control function approach as in Newey, Powell, and Vella (1999), Imbens and Newey (2009), and Lee (2007).

The discussion above thus motivates us to consider estimation of a general class of partially linear transformation models, specified as

$$H(Y^*) = X'\beta_0 + m(Z) - U, \tag{1.2}$$

where $m(\cdot)$ is an unknown function of $Z \in \mathbb{R}^{p_Z}$. In comparison with (1.1), (1.2) involves two nonparametric components $H(\cdot)$ and $m(\cdot)$ that need to be estimated under different constraints: the monotonicity on the former and the smoothness on the latter. It is intriguing how to take both constraints into account seamlessly in the estimation. Although there is a rich literature, particularly in biostatistics,

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

on studying the proportional hazards model with nonlinear or partially linear covariate effects (Fan, Gijbels, and King, 1997; Huang, 1999; Huang, Kooperberg, Stone, and Truong, 2000; Cai, Fan, Jiang, and Zhou, 2007, 2008; among others) based on the partial likelihood function, relatively fewer methods have been developed for a general semiparametric partially linear transformation model where both the link function and the error term distribution are unspecified. Among them available, Ma and Kosorok (2005) studied the penalized log-likelihood estimation of a partially linear transformation model with current status data. Linton, Sperlich, and Van Keilegom (2008) studied the identification and estimation of a much more general class of semiparametric transformation models, which combine a parametric transformation with the flexibility of an additive nonparametric regression function. More recently, Lu and Zhang (2010) proposed a martingale representation based estimation procedure for (1.2). Abrevaya and Shin (2011) provided a rank based methodology for a generalized partially linear regression model including (1.2) as a special case. All these studies, however, require independence between the error term and the regressors.

In this article, we propose an alternative estimation procedure for (1.2). Instead of assuming U is independent of (X, Z), we impose a conditional quantile restriction, i.e., the τth quantile of U conditional on (X, Z) is zero for any given $0 < \tau < 1$. To the best of our knowledge, there has been no such study that considers quantile regression estimation of a partially linear transformation model. The advantage of quantile restriction relative to the prior independence restriction is well known: the former allows for an arbitrary form of dependence between regressors and error term as long as a conditional quantile independence assumption is satisfied. In addition, quantile regression enables researchers to learn more than a pure location shift impact of the regressors upon the transformed dependent variable, thus providing an effective way to capture unobservable heterogeneity, an issue of particular interest to many applied researchers.

Specifically, we consider (1.2) in which the dependent variable Y is observed according to a general two-sided fixed censoring mechanism, i.e.,

$$Y = \underline{c} I\{Y^* \le \underline{c}\} + Y^* I\{\underline{c} < Y^* < \overline{c}\} + \overline{c} I\{Y^* \ge \overline{c}\}, \tag{1.3}$$

with $-\infty < \underline{c} < \overline{c} < +\infty$.[3] The primary interest of this paper is to develop a comprehensive as well as computationally feasible procedure for estimating both finite and infinite dimensional parameters with desirable convergence rates. We propose for β_0 a two-stage procedure, which combines the features of the maximum integrated score estimator (MISE) proposed by Chen (2010) for a linear transformation model and the average quantile regression (AQR) developed by Lee (2003). While MISE ably exploits the monotonicity constraint imposed on the link function, AQR provides a general principle for achieving $\sqrt{n}$ convergence rate for β_0. Given the estimate of β_0, we consider identification and estimation of two infinite dimensional components in the model, i.e., $H(\cdot)$ and $m(\cdot)$. Applying the marginal integration method, we show the two nonparametric components can be

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

separately identified and the resulting estimators converge at a certain desirable nonparametric rate. Consistent estimation of these structural parameters means that we can only learn about the marginal effect of covariates on the conditional quantile function of the dependent variable after transformation. In practice, it is common to care about the average partial quantile effect, the marginal effect of X on the latent dependent variable Y^*, a quantity involving all unknown structural parameters in the model. We propose a novel and computationally simple approach to estimating this average quantile effect, which obviates the need for nonparametric estimation of $H'(\cdot)$ and $H^{-1}(\cdot)$. Moreover, we show that the proposed procedure can readily extend to random censoring, as well as deal with endogeneity.

The rest of the paper is structured as follows. The estimation procedure is motivated and described in Section 2. Section 3 introduces the regularity conditions and establishes large sample properties for the proposed estimators. Extension to randomly censored data is considered in Section 4. Estimation of a transformation model with an endogenous regressor is discussed in Section 5. Monte Carlo results concerning the finite sample performance of these estimators are reported in Section 6. Section 7 concludes the article and all the technical details are collected in the Appendix.

2. ESTIMATION PROCEDURE

This section studies estimation of (1.2) in which the dependent variable Y is observed according to (1.3). Section 2.1 focuses on estimating the finite dimensional parameter β_0. Estimation of the unknown functions $H(\cdot)$ and $m(\cdot)$ will be considered in Section 2.2. Based on these results, we come up with a computationally simple estimator for average partial quantile effect in Section 2.3.

2.1. Estimation of β_0

Our strategy for estimating β_0 builds on the maximum integrated score estimator (MISE) proposed by Chen (2010) for a linear transformation model. The MISE is motivated by the following observation: under monotonicity of $H(\cdot)$, (1.2) is equivalent to a sequence of binary choice models through a discretization technique. That is, given $c \in (\underline{c}, \overline{c})$ and defining $D(c) = I\{Y > c\}$, we have the following binary choice model for each $c \in (\underline{c}, \overline{c})$ due to the monotonicity of $H(\cdot)$, i.e.,

$$D(c) = I\{Y > c\} = I\{Y^* > c\} = I\{X'\beta_0 + m(Z) - H(c) > U\}. \tag{2.1.1}$$

It then follows from the τth zero quantile restriction on U conditional on (X, Z) that

$$\begin{aligned} E\big(D(c)\big|X, Z\big) - \tau = F_{U|XZ}\big(X'\beta_0 + m(Z) - H(c)\big|X, Z\big) - \tau > 0 \\ \Leftrightarrow X'\beta_0 + m(Z) - H(c) > 0, \end{aligned} \tag{2.1.2}$$

where $F_{U|XZ}(\cdot|\cdot, \cdot)$ is the distribution of U conditional on (X, Z).

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

To motivate our estimator, let us assume for the moment that the nonlinear part $m(\cdot)$ is known to the researcher. If so, Chen (2010) proposed estimating β_0 like this: first, based on the rank condition (2.1.2), Horowitz (1992)'s smoothed maximum score (SMS) method applies and for any given c, a consistent estimator for $(\beta_0, H(c))$ can be obtained by maximizing

$$\frac{1}{n}\sum_{i=1}^{n}\left(D_i(c)-\tau\right)K\left(\frac{X_i'\beta+m(Z_i)-a}{h}\right) \tag{2.1.3}$$

with respect to β and a, where $K(v)=\int_{-\infty}^{v}k(s)ds$ is an integrated kernel function and h is a bandwidth, decreasing to zero as $n\to\infty$. Note that in (2.1.3), $K(\cdot/h)$ is used to mimic the behavior of an indicator function as h diminishes to zero. By fixing c, Horowitz's SMS method gives an estimator for β_0 with its convergence rate much slower than the usual parametric $\sqrt{n}$ rate. To restore the desired $\sqrt{n}$ convergence, Chen (2010) proposed integrating (2.1.3) over the interval $(\underline{c},\overline{c})$ and estimating β_0 by maximizing the integrated criterion function

$$Q_n(\beta)=\int\frac{1}{n}\sum_{i=1}^{n}\left(D_i(c)-\tau\right)K\left(\frac{X_i'\beta+m(Z_i)-H_n(c,\beta)}{h}\right)\omega(c)dc, \tag{2.1.4}$$

where $H_n(c,\beta)$ is profiled out from maximizing (2.1.3) for any given β and c, i.e.,

$$H_n(c,\beta)=\arg\max_a \tfrac{1}{n}\textstyle\sum_{i=1}^{n}\left(D_i(c)-\tau\right)K\left(\frac{X_i'\beta+m(Z_i)-a}{h}\right),$$

and $\omega(\cdot)$ is a nonnegative weighting function. For example, letting

$$\omega(c)=I\left\{\underline{c}\le c\le\overline{c}\right\}, \tag{2.1.5}$$

amounts to imposing a uniform weighting scheme. Following Chen (2010), it is straightforward to show that the above (infeasible) MISE that maximizes (2.1.4) is $\sqrt{n}$-consistent.

The above procedure is not operational due to the lack of knowledge about $m(\cdot)$. To overcome this difficulty, let us go back to the rank condition (2.1.2). In the presence of a nonlinear component $m(\cdot)$, the principle of Horowitz (1992)'s SMS method still applies but a local version should be adopted. That is, for any $z\in\mathbb{R}^{p_Z}$ and $c\in(\underline{c},\overline{c})$, we could locally estimate $(\beta_0, m(z)-H(c))$ by $(\widehat{\beta}_n(c,z),\widehat{\mathcal{H}}_n(c,z))$ that maximizes the following local smoothed maximum score (LSMS) criterion function

$$q_n(c,z,\beta,a)=\frac{1}{nh_2^{p_Z}}\sum_{i=1}^{n}\left(D_i(c)-\tau\right)K_1\left(\frac{X_i'\beta+a}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right), \tag{2.1.6}$$

with respect to β and a, where $K_1(s)=\int_{-\infty}^{s}k_1(v)dv$ is an integrated kernel function,

$$\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)=\prod_{l=1}^{p_Z}k_2\left(\frac{Z_{li}-z_l}{h_2}\right),$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Z_{li} and z_l are the lth components of Z_i and z respectively, $k_2(\cdot)$ is another univariate density kernel function, and h_1 and h_2 are two bandwidths. The main rationale behind (2.1.6) is that with a carefully chosen kernel $k_2(\cdot)$ and a sufficiently small h_2, only the observations lying within a small neighborhood around z contribute to the objective function. As a result, the maximizer of (2.1.6) will be a consistent estimator for $(\beta_0, m(z) - H(c))$. However, due to localizing around (c, z), this LSMS estimator converges to β_0 at a rate slower than $\sqrt{n}$. Next we take two steps to improve its convergence rate. First we apply Chen (2010)'s integration method and obtain $\widehat{\beta}_n(z)$ by integrating out c, that maximizes

$$Q_n(z,\beta) = \int \frac{1}{nh_2^{pz}} \sum_{i=1}^{n} \big(D_i(c) - \tau\big) K_1\left(\frac{X_i'\beta + \mathcal{H}_n(c,z,\beta)}{h_1}\right) \times \mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right) \omega(c)dc \tag{2.1.7}$$

with respect to β for any given z, where $\mathcal{H}_n(c,z,\beta)$ maximizes (2.1.6) with respect to a for any given (c,z,β). In general, these $\widehat{\beta}_n(z)$'s, due to their localizing around z, are still expected to have a convergence rate slower than $\sqrt{n}$. To restore the $\sqrt{n}$ convergence rate, we follow the principle of average quantile regression (AQR) developed by Lee (2003) and average these $\widehat{\beta}_n(z)$'s over $\mathcal{Z}$, the support of Z, to obtain

$$\widehat{\beta}_n = \sum_{i=1}^{n} t(Z_i)\widehat{\beta}_n(Z_i) \Big/ \sum_{i=1}^{n} t(Z_i), \tag{2.1.8}$$

where $t(z)$ is a trimming function, aiming to delete those inaccurate local estimates from the averaging process. In Section 3, we show (2.1.8) is $\sqrt{n}$-consistent and derive its limiting distribution.

Remark 2.1. Our method resembles Lee's AQR in that both intend to locally estimate β_0 at a convergence rate slower than $\sqrt{n}$ at the first stage and then restore the $\sqrt{n}$-consistency through averaging in the second stage. To obtain the local estimates, Lee (2003) employed the local polynomial quantile regression (LPQR), in contrast to our LSMS estimator used here. However, one major drawback associated with LPQR is that it does not seem flexible enough to accommodate many limited dependent variable models, e.g., models with censored dependent variables. Instead, we adopt the LSMS method as our first stage estimator. It turns out that this different selection of local estimates in the first stage makes our method more flexible and thus has potentially wider applications than Lee's AQR.[4]

2.2. Estimation of $H(\cdot)$ and $m(\cdot)$

Based on the $\sqrt{n}$-consistent estimate of β_0, we then turn to estimating $H(\cdot)$ and $m(\cdot)$ respectively. To begin with, we impose two normalization restrictions as follows:

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

(i) $\int \omega(c)dc = 1$ and (ii) $\int H(c)\omega(c)dc = 0$.

The first restriction turns the weighing function (2.1.5) into

$$\omega(c) = \frac{1}{\overline{c} - \underline{c}} I\left\{\underline{c} \le c \le \overline{c}\right\}.$$

Given (i), (ii) is simply a location normalization on $H(\cdot)$, thus is innocuous since (1.2) is observationally equivalent to

$$H\left(Y^*\right) - a = X'\beta_0 + (m(Z) - a) - U, \tag{2.2.1}$$

for any $a \in \mathbb{R}$. Applying the LSMS method, it is straightforward to understand that for any given $c \in (\underline{c}, \overline{c})$ and $z \in \mathcal{Z}$, $m(z) - H(c)$ can be consistently estimated by $\widehat{\mathcal{H}}_n(c, z)$ that maximizes

$$q_n(c, z, \widehat{\beta}_n, a) = \frac{1}{nh_2^{p_Z}} \sum_{i=1}^{n} \left(D_i(c) - \tau\right) K_1\left(\frac{X_i'\widehat{\beta}_n + a}{h_1}\right) \mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right), \tag{2.2.2}$$

with respect to a, where $\widehat{\beta}_n$ is any $\sqrt{n}$-consistent estimate of β_0. We propose estimating $m(z)$ for any given $z \in \mathcal{Z}$ by marginal integration, i.e.,

$$\widehat{m}_n(z) = \int \widehat{\mathcal{H}}_n(c, z)\omega(c)dc, \tag{2.2.3}$$

that is rationalized by

$$\int \left(m(z) - H(c)\right)\omega(c)dc = m(z)\int \omega(c)dc - \int H(c)\omega(c)dc = m(z). \tag{2.2.4}$$

Finally, it follows from the observation

$$E_Z\left(m(Z) - H(c)\right) = E_Z\left(m(Z)\right) - H(c) \tag{2.2.5}$$

that we can estimate $H(c)$ for any given $c \in (\underline{c}, \overline{c})$ by

$$\widehat{H}_n(c) = \frac{1}{n}\sum_{i=1}^{n} \widehat{m}_n(Z_i) - \frac{1}{n}\sum_{i=1}^{n} \widehat{\mathcal{H}}_n(c, Z_i). \tag{2.2.6}$$

Remark 2.2. If $\int \omega(c)dc = 1$ but $\int H(c)\omega(c)dc \neq 0$, then it follows from (2.2.4) that $m(z)$ can be identified only up to a location shift. Even in this case, it follows from (2.2.5) that $H(c)$ can still be identified exactly.

Remark 2.3. We shall show in Section 3 that $\widehat{m}_n(z)$ converges to $m(z)$ pointwise at the rate $\sqrt{nh_2^{p_Z}}$ and $\widehat{H}_n(c)$ to $H(c)$ at the rate $\sqrt{nh_1}$. Obviously, these have achieved the fastest rates of nonparametric estimates for p_Z-variate and univariate functions, respectively.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

2.3. Average Partial Quantile Effect

Consistent estimation of β_0 means that we can learn about the marginal effect of X on the quantile function of the dependent variable after transformation conditional on the regressors. In practice, one may also be concerned with the average partial quantile effect of X on the latent dependent variable Y^*. Without loss of generality, we consider the average partial median effect, i.e., $\tau = 0.5$, which is defined as

$$\vartheta_0 = E\left\{\frac{\partial \text{Med}\left(Y^* \middle| X, Z\right)}{\partial X}\right\}.$$

Since the median of $H(Y^*)$ conditional on (X, Z) is $X'\beta_0 + m(Z)$, the conditional median of Y^* given (X, Z) is $\text{Med}\left(Y^* | X, Z\right) = H^{-1}\left(X'\beta_0 + m(Z)\right)$. Thus

$$\begin{aligned}\vartheta_0 &= E\left\{\frac{\partial H^{-1}\left(X'\beta_0 + m(Z)\right)}{\partial X}\right\} \\ &= E\left\{\frac{1}{H'\left(H^{-1}\left(X'\beta_0 + m(Z)\right)\right)}\right\}\beta_0, \end{aligned} \tag{2.3.1}$$

which implies that we can estimate ϑ by

$$\widehat{\vartheta}_n = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\widehat{H}_n'\left(\widehat{H}_n^{-1}\left(X_i'\widehat{\beta}_n + \widehat{m}_n(Z_i)\right)\right)}\widehat{\beta}_n,$$

where $\widehat{H}_n^{-1}(\cdot)$ and $\widehat{H}_n'(\cdot)$ are nonparametric estimates of $H^{-1}(\cdot)$ and $H'(\cdot)$ respectively. Theoretically, both $\widehat{H}_n^{-1}(\cdot)$ and $\widehat{H}_n'(\cdot)$ can be recovered from the nonparametric estimate of $H(\cdot)$ given by (2.2.6). However, this direct plug-in procedure inevitably causes substantial computational burden and makes the subsequent asymptotic analysis a formidable task. Below we propose a computationally simple estimator for ϑ_0 that obviates the need for nonparametric estimation of $H^{-1}(\cdot)$ and $H'(\cdot)$. The key observation behind our estimator is

$$\begin{aligned}&\int \frac{1}{h_1}k_1\left(\frac{X'\beta_0 + m(Z) - H(c)}{h_1}\right)dc \\ &\quad = {}^{\sharp_1}\int \frac{1}{h_1}k_1(s)\,dH^{-1}\left(X'\beta_0 + m(Z) - h_1 s\right) \\ &\quad = {}^{\sharp_2}\frac{1}{H'\left(H^{-1}\left(X'\beta_0 + m(Z)\right)\right)} + O(h_1^{\kappa}),\end{aligned}$$

where $\sharp_1$ follows from the change of variables $X'\beta_0 + m(Z) - H(c) = h_1 s$, $\sharp_2$ follows if we assume $H(\cdot)$ is κ times continuously differentiable, $k_1(\cdot)$ is a κth

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

order bias reduction kernel, i.e., $\int k_1(v)v^r dv = 0$ for $0 < r < \kappa$. Thus a simpler estimator for ϑ_0 is

$$\widehat{\vartheta}_n = \frac{1}{n}\sum_{i=1}^{n}\int \frac{1}{h_1}k_1\left(\frac{X_i'\widehat{\beta}_n + \widehat{m}_n(Z_i) - \widehat{H}_n(c)}{h_1}\right)dc \cdot \widehat{\beta}_n. \tag{2.3.2}$$

Later we show $\widehat{\vartheta}_n$ is $\sqrt{n}$-consistent and derive its limiting distribution.

3. LARGE SAMPLE PROPERTIES

This section studies the large sample properties of the estimators introduced in the preceding section. To proceed, we introduce the following notations. It is well known (e.g., from Horowitz, 1996) that for models subject to unknown transformation, β_0 can be identified up to scale and no intercept is identified. Thus the regressor vector X should not contain a constant term. In addition, identification of β_0 requires that X should contain at least one component whose probability distribution conditional on the remaining components is absolutely continuous with respect to the Lebesgue measure. For these reasons, we rearrange the components of $X = (X_1, X_2')'$ so that its first component X_1 satisfies this condition. For scale normalization, let $\beta_0 = (\beta_{10}, \beta_{20}')'$, with $|\beta_{10}| = 1$ and β_{20} being a $(p_X - 1)$-dimensional coefficient vector for the remaining components. Let $\mathcal{X} \subset \mathbb{R}^{p_X}$ and $\mathcal{Z} \subset \mathbb{R}^{p_Z}$ signify the support of X and Z respectively.

Assumption 3.1.

(i) The distribution of U conditional on (X, Z) is absolutely continuous with respect to the Lebesgue measure and the corresponding conditional density is continuous and positive almost everywhere.

(ii) The τth quantile of U conditional on (X, Z) is zero.

Assumption 3.2.

(i) The support of X is not contained in any proper linear subspace of $\mathbb{R}^{p_X}$ for almost every Z.

(ii) The distribution of X_1 conditional on X_2 and Z has a positive density almost everywhere with respect to the Lebesgue measure.

(iii) The distribution of Z conditional on X is absolutely continuous with respect to the Lebesgue measure almost everywhere in X and the density of Z conditional on X, i.e., $f_{Z|X}(\cdot|\cdot)$ is everywhere positive.

(iv) $|\beta_{10}| = 1$.

Assumption 3.3. $\{Y_i, X_i, Z_i\}$, $i = 1, \ldots, n$ is a random sample of (Y, X, Z) generated by (1.2) and (1.3). $H(\cdot)$ is a strictly increasing function.

Assumption 3.4. $\mathcal{B} = \{-1, 1\} \times \mathcal{B}_2$ is a compact subset of $\mathbb{R}^{p_X}$ that contains β_0 as the interior; $\mathcal{M}_H$ is a compact subset on the real line that contains $m(z) - H(c)$ as the interior for each $z \in \mathcal{Z}$ and c with $\omega(c) > 0$.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Assumption 3.5.

(i) The density of U conditional on (X, Z), i.e., $f_{U|XZ}(u|x,z)$ is bounded away from zero in a neighborhood of $u=0$ for almost all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$; $f_{U|XZ}(u|x_1,x_2,z)$ is κ times continuously differentiable with respect to u, x_1, and z with all the partial derivatives uniformly bounded.

(ii) $f_{X_1Z|X_2}(x_1,z|x_2)$ is κ times continuously differentiable with respect to x_1 and z with all the partial derivatives uniformly bounded.

(iii) $H(c)$ and $m(z)$ are κ times continuously differentiable in c or z with all the partial derivatives uniformly bounded.

Assumption 3.6.

(i) The weighting function $\omega(\cdot)$ is κ times continuously differentiable with uniformly bounded derivatives, and nonnegative over a bounded support.

(ii) The trimming function $t(z) = I\{z \in \mathcal{Z}_1\}$, where $\mathcal{Z}_1$ is a compact subset of $\mathcal{Z}$.

Assumption 3.7.

(i) $K_1(s) = \int_{-\infty}^{s} k_1(v)dv$. Both k_1 and k_2 are continuously differentiable, κth order bias reduction kernel, i.e.,

$$\int k_l(v)dv = 1, \qquad \int k_l(v)v^r dv = 0, \quad \text{for } 1 \le r < \kappa,\ l = 1,2.$$

(ii) $nh_1h_2^{2p_Z} \to \infty$, $nh_2^{2\kappa} \to 0$, $nh_1^{2\kappa}/h_2^{2p_Z} \to 0$, $\varrho_{1n}\varrho_{2n} = o(n^{-1/2})$, where

$$\varrho_{1n} = h_1^{\kappa} + h_2^{\kappa} + \left(\ln n/nh_1h_2^{p_Z}\right)^{1/2}, \quad \varrho_{2n} = h_1^{\kappa} + h_2^{\kappa} + \left(\ln n/nh_1^3h_2^{p_Z}\right)^{1/2}.$$

Define

$$\mathcal{H}(c,z,\beta) = \arg\max_{a\in\mathbb{R}} q(c,z,\beta,a),$$

and

$$q(c,z,\beta,a) = E\left\{\left(D(c)-\tau\right)I\left\{X'\beta + a > 0\right\}\middle| Z = z\right\}.$$

Assumption 3.8. For all $z \in \mathcal{Z}$, the matrix

$$\sigma(z) = \int \left\{\alpha_{XX}(c,z) - \alpha_{OO}(c,z)\frac{\partial\mathcal{H}(c,z,\beta_0)}{\partial\beta_2}\frac{\partial\mathcal{H}(c,z,\beta_0)}{\partial\beta_2'}\right\}\omega(c)dc$$

is nonsingular, in which

$$\alpha_{XX}(c,z) = E\left\{f_{UX_1Z|X_2}\left(0, -X_2'\beta_{20} - m(z) + H(c), z\middle| X_2\right)X_2X_2'\right\},$$

and

$$\alpha_{OO}(c,z) = E\left\{f_{UX_1Z|X_2}\left(0, -X_2'\beta_{20} - m(z) + H(c), z\middle| X_2\right)\right\}.$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Assumption 3.1 is standard for quantile regression. Assumption 3.2(i) and (ii) is necessary for attaining identification of β_0 by the local maximum score method. Assumption 3.2(iii) requires that all Z's components are continuously distributed, aiming to ease the expositional burden in the subsequent analysis. This condition can be relaxed to permit discrete components in Z, requiring that at least one component of Z is continuously distributed. Assumption 3.2(iv) places a scale normalization restriction on β_0. Assumption 3.3 describes the data generating process. Assumption 3.4 is standard for extremum estimate. Assumption 3.5 imposes smoothness conditions on $f_{U|XZ}$, $f_{X_1Z|X_2}$, $m(\cdot)$, and $H(\cdot)$. Assumption 3.6 places restrictions on the weighting function and the trimming function. For the trimming function $t(z)$, a $\mathcal{Z}_1$ that is too small can induce the loss of efficiency, whereas a $\mathcal{Z}_1$ that is too large allows the estimator to be unduly influenced by the tail behavior of the distribution of Z. Assumption 3.7 is concerned with selection of the kernel functions and the bandwidths. If one assumes $h_1 = A_1 \times n^{-\eta_1}$ and $h_2 = A_2 \times n^{-\eta_2}$ with η_1 and η_2 some positive numbers, then Assumption 3.7(ii) implies that $p_Z < \kappa(\kappa-2)/(\kappa+2)$. Thus, in general, kernel functions of higher order should be employed if the number of variables in Z goes large. Assumption 3.8 ensures that the variance of the asymptotic distribution of the resulting estimator is well defined.

PROPOSITION 3.1. *Under Assumptions 3.1–3.8, $\widehat{\beta}_n$ as defined by (2.1.8) is consistent and has the limiting distribution as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\beta}_{2n} - \beta_{20}\right) \to^d \mathcal{N}\left(0, V_\beta\right),$$

with $V_\beta = E\left(\psi_{\beta i}\psi'_{\beta i}\right)$,

$$\psi_{\beta i} = t^*(Z_i)\left(I\left\{U_i < 0\right\} - \tau\right) f_Z(Z_i)\sigma^{-1}(Z_i) \times \left(X_{2i} + \frac{\alpha_{XO}\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right), Z_i\right)}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right), Z_i\right)}\right) \times \frac{\omega\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)},$$

$$t^*(z) = I\left\{z \in \mathcal{Z}_1\right\}/Pr(Z \in \mathcal{Z}_1),$$

$$\alpha_{XO}(c,z) = E\left\{f_{UX_1Z|X_2}\left(0, -X_2'\beta_{20} - m(z) + H(c), z \middle| X_2\right)X_2\right\}.$$

For the purpose of carrying out large sample inference, consistent estimate of V_β needs to be constructed. It follows from the proof of Proposition 3.1 that $\psi_{\beta i}$ can be estimated by

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

$$\widehat{\psi}_{\beta i} = \frac{I\{Z_i \in \mathcal{Z}_1\}}{\sum_{j\neq i} I\{Z_j \in \mathcal{Z}_1\}} \sum_{j\neq i} I\left\{Z_j \in \mathcal{Z}_1\right\} \frac{1}{h_2^{pz}} \mathbf{k}_2\left(\frac{Z_i - Z_j}{h_2}\right) \widehat{\sigma}_n^{-1}(Z_j)$$
$$\times \int \left(D_i(c) - \tau\right) \frac{1}{h_1} k_1\left(\frac{X_i'\widehat{\beta}_n + \mathcal{H}_n(c, Z_j, \widehat{\beta}_n)}{h_1}\right)$$
$$\times \left(X_{2i} + \frac{\widehat{\alpha}_{XO}(c, Z_j)}{\widehat{\alpha}_{OO}(c, Z_j)}\right) \omega(c) dc,$$

where

$$\widehat{\alpha}_{OO}(c,z) = \int \frac{1}{nh_1^2 h_2^{pz}} \sum_{i=1}^n \left(D_i(c) - \tau\right) k_1'\left(\frac{X_i'\widehat{\beta}_n + \mathcal{H}_n(c,z,\widehat{\beta}_n)}{h_1}\right) \mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right),$$
$$\widehat{\alpha}_{XO}(c,z) = \int \frac{1}{nh_1^2 h_2^{pz}} \sum_{i=1}^n \left(D_i(c) - \tau\right) k_1'\left(\frac{X_i'\widehat{\beta}_n + \mathcal{H}_n(c,z,\widehat{\beta}_n)}{h_1}\right) \mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right) X_{2i},$$

and

$$\widehat{\sigma}_n(z) = \int \frac{1}{nh_1^2 h_2^{pz}} \sum_{i=1}^n \left(D_i(c) - \tau\right) k_1'\left(\frac{X_i'\widehat{\beta}_n + \mathcal{H}_n(c,z,\widehat{\beta}_n)}{h_1}\right) \mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right)$$
$$\times \left(X_{2i}X_{2i}' - \frac{\partial \mathcal{H}_n(c,z,\widehat{\beta}_n)}{\partial \beta_2} \frac{\partial \mathcal{H}_n(c,z,\widehat{\beta}_n)}{\partial \beta_2'}\right) \omega(c) dc.$$

To establish the asymptotic properties for $\widehat{m}_n(z)$ and $\widehat{H}_n(c)$, we maintain the following assumption:

Assumption 3.9.

(i) $\int \omega(c)dc = 1, \qquad \int H(c)\omega(c)dc = 0.$
(ii) $h_1/h_2^{pz} \to 0.$

PROPOSITION 3.2. *Under Assumptions 3.1–3.9, $\widehat{m}_n(z)$ as defined by (2.2.3) is consistent and has the limiting distribution as $n \to \infty$,*

$$\sqrt{nh_2^{pz}}\left(\widehat{m}_n(z) - m(z)\right) \to^d \mathcal{N}(0, V_m),$$

where

$$V_m = \tau(1-\tau)\varsigma_2^{pz} E_X\left\{\left[\frac{\omega\left(H^{-1}\left(X'\beta_0 + m(z)\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X'\beta_0 + m(z)\right), z\right) H'\left(H^{-1}\left(X'\beta_0 + m(z)\right)\right)}\right]^2 f_{Z|X}(z|X)\right\}$$

with $\varsigma_l = \int k_l^2(v)dv$, for $l = 1, 2$. Moreover, $\widehat{H}_n(c)$ as defined by (2.2.6) is consistent and has the limiting distribution as $n \to \infty$,

$$\sqrt{nh_1}\left(\widehat{H}_n(c) - H(c)\right) \to^d \mathcal{N}(0, V_H),$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

where

$$V_H = \tau(1-\tau)\varsigma_1 E_{X_2 Z}\left\{\left[\frac{f_Z(Z)}{\alpha_{OO}(c,Z)}\right]^2 f_{X_1|X_2Z}\left(H(c)-m(Z)-X_2'\beta_{20}\middle|X_2,Z\right)\right\}.$$

The next proposition summarizes the limiting distribution of the average partial quantile effect estimator (2.3.2).

PROPOSITION 3.3. *Under Assumptions 3.1–3.9, $\widehat{\vartheta}_n$ as defined by (2.3.2) is a consistent estimator for ϑ_0 and has the limiting distribution as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta_0\right) \to^d \mathcal{N}(0, V_\vartheta),$$

where $V_\vartheta = E\left(\psi_{\vartheta i}\psi_{\vartheta i}'\right)$,

$$\psi_{\vartheta i} = \left(\left\{\Pi\left(X_i'\beta_0 + m(Z_i)\right) - \theta_0\right\} + E\left\{\Pi'\left(X'\beta_0 + m(Z)\right)X_2'\right\}\psi_{\beta i} + \psi_{\theta i}^1 + \psi_{\theta i}^2\right)\beta_0 + \theta_0\psi_{\beta i},$$

$$\Pi(y) = 1\Big/H'\left(H^{-1}(y)\right), \qquad \theta_0 = E\left\{\frac{1}{H'\left(H^{-1}\left(X'\beta_0 + m(Z)\right)\right)}\right\},$$

$$\psi_{\theta i}^1 = E_X\left\{\Pi'\left(X'\beta_0 + m(Z)\right)\right\} f_Z(Z_i)\frac{I\left\{U_i < 0\right\} - \tau}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right), Z_i\right)} \times \frac{\omega\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)},$$

$$\psi_{\theta i}^2 = \frac{\Gamma\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right), Z_i\right)}\left(I\left\{U_i < 0\right\} - \tau\right) f_Z(Z_i),$$

and

$$\Gamma(c) = E_{X_2 Z}\left\{f_{X_1|X_2Z}'\left(H(c) - X_2'\beta_{20} - m(Z)\middle|X_2, Z\right)\right\}.$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

4. RANDOM CENSORING

The estimation strategy proposed above can readily extend to randomly censored data, which is common in duration and survival analysis. In this case, the dependent variable typically represents right-censored survival time and the realization of the censored variable is observed only when the observation is censored. Without loss of generality, let us assume that the observations consist of a random sample of (Y, δ, X, Z) with

$$Y = \min\{Y^*, C\}, \qquad \delta = I\{Y \le C\}, \tag{4.1}$$

the latent dependent variable Y^* is generated by (1.2) and the random variable C is assumed to be independent of (U, X, Z). Then similar reasoning to (2.1.1) and (2.1.2) gives that for each c,

$$\begin{aligned} E\left(D(c)\middle|X,Z\right) &= E\left(I\left\{Y > c\right\}\middle|X,Z\right) \\ &= E\left(I\left\{\min\left\{Y^*, C\right\} > c\right\}\middle|X,Z\right) \\ &= E\left(I\left\{C > c\right\}\right)E\left(I\left\{Y^* > c\right\}\middle|X,Z\right) \\ &= G(c)E\left(I\left\{Y^* > c\right\}\middle|X,Z\right), \end{aligned}$$

where $G(c) = 1 - F_C(c)$, $F_C(\cdot)$ is the marginal cumulative distribution function of C and thus

$$E\left(\frac{D(c)}{G(c)}\middle|X,Z\right) > \tau \Leftrightarrow X'\beta_0 + m(Z) - H(c) > 0. \tag{4.2}$$

The rank condition (4.2), together with the preceding procedure defined for the fixed censoring case, suggests that the AQR estimator for β_0 is defined by

$$\widehat{\beta}_n^r = \sum_{i=1}^n t(Z_i)\widehat{\beta}_n^r(Z_i) \Big/ \sum_{i=1}^n t(Z_i), \tag{4.3}$$

where $\widehat{\beta}_n^r(z)$ maximizes

$$\begin{aligned} Q_n^r(z,\beta) = \int \frac{1}{nh_2^{pz}} \sum_{i=1}^n \left(\frac{D_i(c)}{\widehat{G}_n(c)} - \tau\right) \\ \times K_1\left(\frac{X_i'\beta + \mathcal{H}_n^r(c,z,\beta)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right)\omega(c)dc \end{aligned} \tag{4.4}$$

with respect to β, $\widehat{G}_n(c)$ is the Kaplan–Meier estimator of $G(c)$, and $\mathcal{H}_n^r(c,z,\beta)$ maximizes

$$q_n^r(c,z,\beta,a) = \frac{1}{nh_2^{pz}} \sum_{i=1}^n \left(\frac{D_i(c)}{\widehat{G}_n(c)} - \tau\right) K_1\left(\frac{X_i'\beta + a}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right), \tag{4.5}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

with respect to a for any given c, z, and β. Let $\widehat{\mathcal{H}}_n^r(c,z)$ denote the maximizer of

$$q_n^r\left(c,z,\widehat{\beta}_n^r,a\right)=\frac{1}{nh_2^{pz}}\sum_{i=1}^{n}\left(\frac{D_i(c)}{\widehat{G}_n(c)}-\tau\right)K_1\left(\frac{X_i'\widehat{\beta}_n^r+a}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right) \tag{4.6}$$

with respect to a for any given (c,z). Moreover, we can estimate $m(z)$ for any given $z\in\mathcal{Z}$ by

$$\widehat{m}_n^r(z)=\int\widehat{\mathcal{H}}_n^r(c,z)\omega(c)dc \tag{4.7}$$

and estimate $H(c)$ for any given $c\in(\underline{c},\overline{c})$ by

$$\widehat{H}_n^r(c)=\frac{1}{n}\sum_{i=1}^{n}\widehat{m}_n^r(Z_i)-\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathcal{H}}_n^r(c,Z_i). \tag{4.8}$$

Without loss of generality, we present below only the limiting distribution of $\widehat{\beta}_n^r$ while asymptotic properties of other components can be analyzed similarly.

Assumption 3.3*.

(i) $\{Y_i,X_i,Z_i,\delta_i\}$, $i=1,\ldots,n$ is a random sample of (Y,X,Z,δ) generated by (1.2) and (4.1). $H(\cdot)$ is a strictly increasing function.

(ii) The censoring variable C is independent of (X,Z,U) and continuously distributed with a positive density on an interval contained in the support of $\omega(\cdot)$.

PROPOSITION 4.1. *Under Assumptions 3.1, 3.2, 3.3*, and 3.4–3.8, $\widehat{\beta}_n^r$ as defined by (4.3) is consistent and has the limiting distribution as $n\to\infty$,*

$$\sqrt{n}\left(\widehat{\beta}_{2n}^r-\beta_{20}\right)\to^d\mathcal{N}\left(0,V_\beta^r\right),$$

where $V_\beta^r=E\left(\psi_{\beta i}^r\psi_{\beta i}^{r\prime}\right)$, $\psi_{\beta i}^r=\psi_{\beta,1i}^r+\psi_{\beta,2i}^r$,

$$\psi_{\beta,1i}^r=t^*(Z_i)\left(\frac{I\left\{H(Y_i)>X_i'\beta_0+m(Z_i)\right\}}{G\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)}-\tau\right)\sigma^{-1}(Z_i)f_Z(Z_i)$$
$$\times\left(X_{2i}+\frac{\alpha_{XO}\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right),Z_i\right)}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right),Z_i\right)}\right)$$
$$\times\frac{\omega\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)},$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

$$\psi^r_{\beta,2i} = \frac{1}{Pr(Z \in \mathcal{Z})} \int \frac{E_Z\{t(Z)\sigma^{-1}(Z)\alpha^*_{RX}(s,Z)\}}{\pi(s)} dM_i(s),$$

$$\pi(s) = Pr(Y \geq s),$$

$$M_i(s) = I\left\{Y_i \leq s, \delta_i = 0\right\} - \int_0^s I\left\{Y_i \geq v\right\} d\Lambda(v),$$

$\Lambda(\cdot)$ *is the cumulative hazard function of* C,

$$\alpha^*_{RX}(s,z) = \int_s^{+\infty} \alpha_{RX}(c,z)\omega(c)dc$$

and

$$\alpha_{RX}(c,z) = \tau E\left\{f_{X_1 Z|X_2}\left(H(c) - X_2'\beta_{20} - m(z), z \middle| X_2\right) X_2\right\}.$$

In comparison with the fixed censoring case, there is an extra term $\psi^r_{\beta,2i}$ in the asymptotic linear representation due to the presence of a Kaplan–Meier estimator $\widehat{G}_n(\cdot)$ in (4.4). In addition, consistent estimate of V^r_β can be constructed in a way similar to V_β in Section 3.

5. ENDOGENEITY

The estimation strategy proposed in Section 2 can be used to deal with the problem of endogeneity. To see this, consider a linear transformation model in triangular simultaneous system, i.e.,

$$H(Y^*) = X'\beta_0 - U, \tag{5.1}$$

and

$$X_E = W'\gamma_0 + Z, \tag{5.2}$$

where X_E is a scalar endogenous variable in $X = (X_I', X_E)'$, $W = (W_1', X_I')'$ is a vector of observed instruments, and Z is the unobserved residual. For identification it is assumed that there is at least one component of W_1 that is not included in X_I. Without loss of generality, let us focus on median regression only. When the zero median restriction, i.e., $\text{Med}(U|X) = 0$ is violated, the existing estimation procedure, e.g., Chen (2010) for is not applicable any more. Our basic idea to address this problem is to adopt a control function approach, resulting in a partially linear structure studied by the preceding sections. To operationalize this idea, we make a distributional exclusion assumption as in Newey et al. (1999), Imbens and Newey (2009), and Lee (2007), i.e.,

$$\text{Med}\left(U \middle| Z\right) = \text{Med}\left(U \middle| W, Z\right). \tag{5.3}$$

Essentially, (5.3) imposes a conditional exogeneity assumption on the instruments W, i.e., W is median independent of U conditional on Z, which is weaker than

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

requiring that W be unconditionally independent of U. Combining (5.2) and (5.3),

$$\text{Med}\left(U\middle|W, X_E, Z\right) =^{\sharp} \text{Med}\left(U\middle|W, Z\right) = \text{Med}\left(U\middle|Z\right) = -m(Z), \tag{5.4}$$

where $\sharp$ holds since there is a one-to-one correspondence between Z and X_E given W. Here $m(Z)$ is regarded as an unknown nuisance (control) function. (5.4) together with (5.1) suggests that we could estimate

$$H\left(Y^*\right) = X'\beta_0 + m(Z) - \left(U + m(Z)\right) \tag{5.5}$$

to infer about β_0 as $\text{Med}\left(U + m(Z)\middle|X, Z, W_1\right) = \text{Med}\left(U + m(Z)\middle|W, X_E, Z\right) = 0$.

To fix the idea, let us consider the fixed censoring case, i.e., Y is observed according to (1.3). The random censoring case can be analyzed similarly. Based on the estimation strategy proposed in Section 2, we can estimate β_0 by

$$\widehat{\beta}_n^e = \sum_{i=1}^n t\left(\widehat{Z}_i\right)\widehat{\beta}_n^e\left(\widehat{Z}_i\right) \Big/ \sum_{i=1}^n t\left(\widehat{Z}_i\right), \tag{5.6}$$

where $\widehat{\beta}_n^e(z)$ maximizes

$$Q_n^e(z,\beta) = \int \frac{1}{nh_2}\sum_{i=1}^n \left(D_i(c) - \tau\right) K_1\left(\frac{X_i'\beta + \mathcal{H}_n^e(c,z,\beta)}{h_1}\right) \times k_2\left(\frac{\widehat{Z}_i - z}{h_2}\right)\omega(c)dc \tag{5.7}$$

with respect to β, $\widehat{Z}_i = X_{Ei} - W_i'\widehat{\gamma}_n$ with $\widehat{\gamma}_n$ any $\sqrt{n}$-consistent estimator for γ_0, and $\mathcal{H}_n^e(c,z,\beta)$ maximizes

$$q_n^e(c,z,\beta,a) = \frac{1}{nh_2}\sum_{i=1}^n \left(D_i(c) - \tau\right) K_1\left(\frac{X_i'\beta + a}{h_1}\right) k_2\left(\frac{\widehat{Z}_i - z}{h_2}\right), \tag{5.8}$$

with respect to a for any given c, z, and β. To formally derive the limiting distribution of $\widehat{\beta}_n^e$, let us introduce the following assumptions.

Assumption 3.3.**

(i) $\left\{Y_i, X_i, W_i\right\}$, $i = 1, \ldots, n$ is a random sample of (Y, X, W) generated by (5.1), (5.2), and (1.3). $H(\cdot)$ is a strictly increasing function.

(ii) $\widehat{\gamma}_n$ is a $\sqrt{n}$-consistent estimator for γ_0 and permits the following asymptotic representation

$$\widehat{\gamma}_n - \gamma_0 = \frac{1}{n}\sum_{i=1}^n \psi_{\gamma i} + o_p\left(n^{-1/2}\right),$$

where $\psi_{\gamma i} = \psi_\gamma\left(X_{Ei}, W_i\right)$, $E\psi_{\gamma i} = 0$, and $E\left\|\psi_{\gamma i}\right\|^2 < \infty$.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

(iii) $\widehat{Z}_i = X_{Ei} - W_i'\widehat{\gamma}_n$.

(iv) The τth quantile of U conditional on Z is independent of W.

Assumption 3.7.** $nh_1^4h_2^4 \to \infty$, $nh_2^{2\kappa} \to 0$, $nh_1^{2\kappa}/h_2^2 \to 0$, $\overline{\varrho}_{1n}\overline{\varrho}_{2n} = o\left(n^{-1/2}\right)$, where

$$\overline{\varrho}_{1n} = h_1^\kappa + h_2^\kappa + \left(\ln n/nh_1h_2\right)^{1/2} + n^{-1/2}h_2^{-2}h_1^{-1},$$

$$\overline{\varrho}_{2n} = h_1^\kappa + h_2^\kappa + \left(\ln n/nh_1^3h_2\right)^{1/2} + n^{-1/2}h_2^{-2}h_1^{-2}.$$

PROPOSITION 5.1. *Under Assumptions 3.1, 3.2, 3.3**, 3.4–3.6, 3.7(i), 3.7**, and 3.8, $\widehat{\beta}_n^e$ as defined by (5.6) is consistent and has the limiting distribution as $n \to \infty$,*

$$\sqrt{n}\left(\widehat{\beta}_{2n}^e - \beta_{20}\right) \to^d \mathcal{N}\left(0, V_\beta^e\right),$$

where $V_\beta^e = E\left(\psi_{\beta i}^e \psi_{\beta i}^{e\prime}\right)$,

$$\begin{aligned}\psi_{\beta i}^e = {} & t^*(Z_i)\left(I\left\{U_i < 0\right\} - \tau\right)\sigma^{-1}(Z_i) f_Z(Z_i) \\ & \times \left(X_{2i} + \frac{\alpha_{XO}\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right), Z_i\right)}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right), Z_i\right)}\right) \\ & \times \frac{\omega\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0 + m(Z_i)\right)\right)} \\ & + \frac{1}{Pr(Z \in \mathcal{Z}_1)} E_Z\left\{t(Z)\sigma^{-1}(Z)\alpha_{XW}^*(Z)\right\}\psi_{\gamma i}\end{aligned}$$

with

$$\alpha_{XW}^*(z) = \int \alpha_{XW}(c, z)\omega(c)dc,$$

$$\alpha_{XW}(c, z) = E\left\{f_{UX_1Z|X_2}\left(0, H(c) - X_2'\beta_{20} - m(z), z \middle| X_2\right) X_2 W'\right\}.$$

In comparison with the limiting distribution of $\widehat{\beta}_{2n}$ derived in Proposition 3.1, there is an extra term containing $\psi_{\gamma i}$ in the asymptotic linear representation for $\widehat{\beta}_{2n}^e$ due to the fact that the regressor Z_i is replaced by its empirical residual $\widehat{Z}_i$.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

6. MONTE CARLO SIMULATION

In this section, we conduct a small scale Monte Carlo experiment to evaluate the finite sample performance of some estimators proposed above. The following six data generation processes are considered in the simulation.

DGP 1: $Y^* = X_1 + X_2 + m_1(Z) - U$, with $m_1(z) = \exp(z)/(1+\exp(z))$;
DGP 2: $Y^* = X_1 + X_2 + m_2(Z) - U$, with $m_2(z) = 0.5 + 0.25\sin(0.5\pi z)$;
DGP 3: $H(Y^*, \lambda) = X_1 + X_2 + m_1(Z) - U$, with $\lambda = 0.5$ and $m_1(z) = \exp(z)/(1+\exp(z))$;
DGP 4: $H(Y^*, \lambda) = X_1 + X_2 + m_2(Z) - U$, with $\lambda = 0.5$ and $m_2(z) = 0.5 + 0.25\sin(0.5\pi z)$;
DGP 5: $H(Y^*, \lambda) = X_1 + X_2 + m_1(Z) - U$, with $\lambda = 2$ and $m_1(z) = \exp(z)/(1+\exp(z))$;
DGP 6: $H(Y^*, \lambda) = X_1 + X_2 + m_2(Z) - U$, with $\lambda = 2$ and $m_2(z) = 0.5 + 0.25\sin(0.5\pi z)$,

where $H(y, \lambda) = \frac{|y|^\lambda \operatorname{sgn}(y) - 1}{\lambda}$ is a modified Box–Cox transformation, $U = \frac{|X_1| + Z}{2} \times \epsilon$, X_1, X_2, and ϵ are i.i.d. drawn from the standard normal distribution $\mathcal{N}(0,1)$, Z is i.i.d. drawn from the uniform distribution on $[0,4]$. All these six models contain heteroscedastic error terms. Being regarded as transformation models (i.e., both the functional form of H and the true value of λ are unknown to the researcher), their coefficients can be identified only up to scale. Thus we only consider the estimation of $\beta_{20} = 1$ by placing a normalization restriction on the coefficient on X_1, i.e., $\beta_{10} = 1$. Without loss of generality, only the median regression is considered.

Let Y be generated by (1.3) with the left cutoff point $\underline{c}$ being the 20% empirical quantile of Y^* and the right cutoff point $\overline{c}$ the 80% empirical quantile of Y^*, thus resulting in about 40% degree of censoring. In order to implement our AQR estimator, we have to select the kernel functions $K_1(\cdot)$ and $k_2(\cdot)$, the bandwidth parameters h_1 and h_2, as well as a weighting function $\omega(\cdot)$. In particular, with $p_Z = 1$, Assumption 3.7 requires that $\kappa \geq 4$. For this reason, we choose $K_1(\cdot)$ to be the following fourth order integrated kernel function

$$K_1(s) = \left[\frac{1}{2} + \frac{105}{64}\left(s - \frac{5}{3}s^3 + \frac{7}{5}s^5 - \frac{3}{7}s^7\right)\right] I\{|s| \leq 1\} + I\{|s| > 1\}$$

and choose $k_2(\cdot)$ to be

$$k_2(s) = \left(\frac{15}{8} - \frac{5}{4}s^2 + \frac{1}{8}s^4\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{s^2}{2}\right)$$

a Gaussian kernel of order six. We refer to the estimators using the higher order kernels as AQR-1 below. As a second possibility, we simply compute the estimators using the Gaussian kernel (i.e., using the standard normal c.d.f. as $K_1(\cdot)$ and the standard normal p.d.f. as $k_2(\cdot)$), instead of using the higher-order kernel functions described above. This is motivated by the frequently made observation

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

that while a higher-order kernel might be required from a theoretical point of view in many semiparametric applications, the resulting estimators often tend to have inferior finite sample properties compared to those based on standard kernels (Marron, 1994; Jones and Signorini, 1997). Thus, although strictly speaking not compatible with our asymptotic analysis, we also consider such lower order kernel based estimators in our simulations, which are referred as AQR-2 below.

The bandwidth conditions required by our assumptions are only qualitative. Since optimal selection of the bandwidths is difficult to derive, a simple Silverman (1986)'s rule of thumb is adopted. We select $h_1 = \widehat{\sigma}_S n^{-9.5/56}$ and $h_2 = \widehat{\sigma}_Z n^{-1/7}$, where $\widehat{\sigma}_S$ and $\widehat{\sigma}_Z$ are the sample standard deviations of $X_i'\breve{\beta}_n + \breve{m}_n(Z_i)$'s and Z_i's respectively, with $\breve{\beta}_n$ and $\breve{m}_n(z)$ retrieved in the preliminary estimation using $(h_1, h_2) = (n^{-9.5/56}, n^{-1/7})$. Although it is straightforward to check that these order conditions for bandwidths satisfy Assumption 3.7, their selection of course is arbitrary in that they depend on the bandwidths selected originally. Even so, it has the benefit of being easy to implement while performing reasonably well compared to other choices used in preliminary experiments.

We choose the uniform weighting scheme $\omega(c) = I\{\underline{c} < c < \overline{c}\}/(\overline{c} - \underline{c})$. The trimming function used in the experiments is $t(z) = I\{\underline{z} \le z \le \overline{z}\}$, where $\underline{z}$ and $\overline{z}$ are 10% and 90% empirical quantiles of Z_i's respectively. The sample size considered in the simulation consists of $n \in \{400, 800\}$. In addition to estimating β_{20}, we also consider the estimation of the average partial median effect (APE) of X_2 on Y^* as defined by (2.3.1), a quantity that relies on $H(\cdot)$ as well as $m(\cdot)$. To this end, we estimate $m(z)$ evaluated at $z = Z_i$'s for $\underline{z} < Z_i < \overline{z}$ by (2.2.3) and then estimate $H(c)$ over an equidistantly spaced grid of $(\underline{c}, \overline{c})$ by (2.2.6). It is worth noting that according to Remark 2.2, while (2.2.6) provides a consistent estimate of $H(c)$, (2.2.3) is only a consistent estimator for

$$m(z) - \int H(c)\omega(c)dc,$$

but not $m(z)$ unless $\int H(c)\omega(c)dc = 0$. Even this, we can always recover the exact estimate of $m(z)$ with a location shift correction term

$$\int \widehat{H}_n(c)\omega(c)dc.$$

We report empirical bias and empirical standard deviation for each estimator based on 1,000 repetitions. The simulation results are summarized in Table 1. The results indicate that our estimator performs well in finite samples for each DGP considered. The estimators for β_{20} are slightly biased while the APE estimators have moderately upward bias. The magnitude of these biases decreases as the sample goes large. Their empirical standard deviations decline with the sample size and the magnitude of such decline is consistent with $\sqrt{n}$-asymptotics, as expected by the theory. In addition, compared with AQR-1 group, the estimators in AQR-2 group using Gaussian kernel usually have moderately larger bias but substantially lower variability, suggesting certain bias-variance tradeoff between using high order and Gaussian kernels.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

TABLE 1. Simulation results

			DGP1	DGP2	DGP3	DGP4	DGP5	DGP6
$n = 400$								
AQR-1	β_2	Bias	−.0530	−.0636	.0525	−.0362	.0438	.0459
		SD	.326	.284	.263	.337	.293	.274
	APE	Bias	.0504	.0723	.0675	.0731	.0614	.0784
		SD	.424	.370	.326	.414	.372	.342
AQR-2	β_2	Bias	−.0564	−.0658	.0546	−.0437	.0451	.0483
		SD	.289	.267	.242	.316	.272	.251
	APE	Bias	.0552	.0748	.0711	.0755	.0621	.0764
		SD	.346	.336	.287	.385	.347	.313
$n = 800$								
AQR-1	β_2	Bias	−.0346	−.0473	.0335	−.0227	.0362	.0316
		SD	.258	.228	.203	.253	.227	.214
	APE	Bias	.0363	.0588	.0624	.0626	.0541	.0704
		SD	.315	.277	.254	.313	.278	.257
AQR-2	β_2	Bias	−.0371	−.0504	.0569	−.0238	.0370	.0339
		SD	.229	.205	.181	.226	.207	.193
	APE	Bias	.0373	.0605	.0642	.0637	.0549	.0702
		SD	.298	.257	.242	.298	.269	.232

7. CONCLUSION

In this article, we consider semiparametric estimation of a general class of partially linear transformation models under conditional quantile restriction. The model is said to be general since no parametric restriction is imposed either on the link functional form or on the error term distribution. In addition, the dependent variable is allowed subject to either fixed or random censoring. In extending Chen (2010)'s maximum integrated score estimator for linear transformation model to the current setup, this article contributes to the literature in several meaningful ways. First, it provides a comprehensive multistage procedure for estimating various structural parameters in a transformation model with nonlinear covariate effects. In addition, we propose a novel and computationally simple approach to estimating the average quantile effect, as a summary statistic measuring unconditional marginal impact of a covariate on the latent dependent variable at a given quantile index. Second, we show the proposed method can be used to deal with endogeneity, a nontrivial and important issue for many nonlinear econometric models, by adopting the control function approach as in Lee (2007).

NOTES

1. Throughout the paper, any parameter with subscript zero represents the true one that generates the data.

2. The empirical results in Meyer (1990) showed that the hazard falls almost monotonically with age. Those of age 17–24 have the highest hazard while those aged 55 and over have the lowest.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

3. This general design incorporates the case of no censoring as a special one by setting $\overline{c}$ sufficiently large and $\underline{c}$ sufficiently small.

4. It is straightforward to see that our LSMS-based AQR described above is applicable to partially linear censored quantile regression model ($H(y) = y$) as a special case of (1.2). However, since the LPQR does not seem robust against fixed censoring, it is difficult to extend Lee's AQR to the occasion with the dependent variable generated by (1.3).

REFERENCES

Abrevaya, J. (1999) Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics* 93, 203–228.

Abrevaya, J. (2003) Pairwise difference rank estimation of the transformation model. *Journal of Business and Economic Statistics* 21, 437–447.

Abrevaya, J. & Y. Shin (2011) Rank estimation of partially linear index models. *Econometrics Journal* 14, 409–437.

Asparouhova, E., R. Golanski, K. Kasprzyk, R.P. Sherman, & T. Asparouhov (2002) Rank estimators for a transformation model. *Econometric Theory* 18, 1099–1120.

Box, G.E.P. & D.R. Cox (1964) An analysis of transformations. *Journal of Royal Statistical Society, Series B* 26, 211–252.

Cai, J., J. Fan, J. Jiang, & H. Zhou (2007) Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association* 102, 538–551.

Cai, J., J. Fan, J. Jiang, & H. Zhou (2008) Partially linear hazard regression with varying coefficients for multivariate survival data. *Journal of the Royal Statistical Society, Series B* 70, 141–158.

Carroll, R.J., J. Fan, I. Gijbels, & M.P. Wand (1997) Generalized partially linear single index models. *Journal of the American Statistical Association* 92, 477–489.

Cavanagh, C. & R.P. Sherman (1998) Rank estimators for monotonic index models. *Journal of Econometrics* 84, 351–381.

Chen, S. (2002) Rank estimation of transformation models. *Econometrica* 70, 1683–1696.

Chen, S. (2010) An integrated maximum score estimator for a generalized censored quantile regression model. *Journal of Econometrics* 155, 90–98.

Chen, K., Z. Jin, & Z. Ying (2002) Semiparametric analysis of transformation models with censored data. *Biometrika* 89, 659–668.

Fan, J., I. Gijbels, & M. King (1997) Local likelihood and local partial likelihood in hazard regression. *Annals of Statistics* 25, 1661–1690.

Fleming, T.R. & D.P. Harrington (1991) *Counting Processes and Survival Analysis*. Wiley.

Horowitz, J.L. (1992) A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.

Horowitz, J.L. (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* 64, 103–137.

Huang, J. (1999) Efficient estimation of the partially linear additive Cox model. *Annals of Statistics* 27, 1536–1563.

Huang, J., C. Kooperberg, C. Stone, & Y. Truong (2000) Functional ANOVA modeling for proportional hazards regression. *Annals of Statistics* 28, 961–999.

Imbens, G.W. & W.K. Newey (2009) Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77, 1481–1512.

Jones, M. & D. Signorini (1997) A comparison of higher-order bias kernel density estimators. *Journal of the American Statistical Association* 92, 1063–1073.

Khan, S. & E. Tamer (2007) Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics* 136, 251–280.

Lancaster, T. (1990) *The Econometric Analysis of Transition Data*. Cambridge University Press.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Lee, S. (2003) Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory* 19, 1–31.
Lee, S. (2007) Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics* 141, 1131–1158.
Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Linton, O., S. Sperlich, & I. Van Keilegom (2008) Estimation of a semiparametric transformation model. *Annals of Statistics* 36, 686–718.
Lu, W. & H. Zhang (2010) On estimation of partially linear transformation models. *Journal of the American Statistical Association* 105, 683–691.
Ma, S. & M.R. Kosorok (2005) Penalized log likelihood estimation for partly linear transformation models with current status data. *Annals of Statistics* 33, 2256–2290.
Manski, C.F. (1985) Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics* 27, 313–333.
Marron, J. (1994) Visual understanding of higher-order kernels. *Journal of Computational and Graphical Statistics* 3, 447–458.
Meyer, B. (1990) Unemployment insurance and unemployment spells. *Econometrica* 58, 757–782.
Newey, W. & D. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2111–2245. Elsevier.
Newey, W., J.L. Powell, & F. Vella (1999) Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67, 565–603.
Pollard, D. (1995) Uniform ratio limit theorems for empirical processes. *Scandinavian Journal of Statistics* 22, 271–278.
Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of weighted average derivatives. *Econometrica* 57, 1403–1430.
Shin, Y. (2010) Local rank estimation of transformation models with functional coefficients. *Econometric Theory* 26, 1807–1819.
Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.
Van den Berg, G.J. (2001) Duration models: Specification, identification and multiple durations. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5. North-Holland.

APPENDIX: Proofs

LEMMA A.1. *Let* $\mathcal{H}_n(c,z,\beta)$ *maximize* $q_n(c,z,\beta,a)$ *given by (2.1.6) and* $\mathcal{H}(c,z,\beta) = \arg\max_a q(c,z,\beta,a)$, *with*

$$q(c,z,\beta,a) = E\left\{\left(D(c)-\tau\right) I\left\{X'\beta + a > 0\right\} \middle| Z = z\right\}.$$

Then under Assumptions 3.1–3.5 and 3.7,

$$\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta) = O_p(\varrho_{1n}), \qquad \frac{\partial \mathcal{H}_n(c,z,\beta)}{\partial \beta_2} - \frac{\partial \mathcal{H}(c,z,\beta)}{\partial \beta_2} = O_p(\varrho_{2n})$$

and

$$\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta) = \frac{\alpha_{OO}^{-1}(c,z,\beta)}{nh_1h_2^{p_Z}} \sum_{i=1}^{n} \left(D_i(c)-\tau\right) k_1\left(\frac{X_i'\beta + \mathcal{H}(c,z,\beta)}{h_1}\right) \times \mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right) + o_p\left(n^{-1/2}\right),$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, *and* $\beta \in \mathcal{B}$, *with*

$$\alpha_{OO}(c,z,\beta) = \int \varphi\Big(X_2'(\beta_{20}-\beta_2)+\mathcal{H}(c,z)-\mathcal{H}(c,z,\beta), -\mathcal{H}(c,z,\beta)-X_2'\beta_2, X_2, z\Big) dF_{X_2},$$

$\mathcal{H}(c,z) = m(z) - H(c)$ *and*

$$\varphi(s_1,s_2,x_2,z) = -\frac{\partial\left\{\left(F_{U|XZ}\left(v+s_1\middle|v+s_2,x_2,z\right)-\tau\right) f_{X_1Z|X_2}\left(v+s_2,z\middle|x_2\right)\right\}}{\partial v}\bigg|_{v=0}.$$

Proof. The proof is similar to Lemma A.1 in Chen (2010). Based on the uniform law of large numbers in Pollard (1995), $q_n(c,z,\beta,a) - q(c,z,\beta,a) = o_p(1)$ uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, and $(\beta, a) \in \mathcal{B} \times \mathcal{M}_H$. Then it follows from Carroll, Fan, Gijbels, and Wand (1997)'s Lemma A.1 that

$$\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta) = o_p(1) \tag{A.1}$$

uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, and $\beta \in \mathcal{B}$. To sharpen the order of $\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta)$, Taylor expansion of the first order condition that $\mathcal{H}_n(c,z,\beta)$ should satisfy, i.e.,

$$\frac{1}{nh_1h_2^{p_Z}}\sum_{i=1}^n \Big(D_i(c)-\tau\Big) k_1\left(\frac{X_i'\beta+\mathcal{H}_n(c,z,\beta)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right) = 0$$

gives that

$$\begin{aligned}\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta) &= \underbrace{\left[-\frac{1}{nh_1^2h_2^{p_Z}}\sum_{i=1}^n \Big(D_i(c)-\tau\Big) k_1'\left(\frac{X_i'\beta+\overline{\mathcal{H}}_n(c,z,\beta)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\right]^{-1}}_{\dagger_1} \\ &\quad \times \underbrace{\frac{1}{nh_1h_2^{p_Z}}\sum_{i=1}^n \Big(D_i(c)-\tau\Big) k_1\left(\frac{X_i'\beta+\mathcal{H}(c,z,\beta)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)}_{\dagger_2},\end{aligned}$$

where $\overline{\mathcal{H}}_n(c,z,\beta)$ lies between $\mathcal{H}_n(c,z,\beta)$ and $\mathcal{H}(c,z,\beta)$. By the change of variables and (A.1), it is straightforward to show that $\dagger_1 = \alpha_{OO}(c,z,\beta) + o_p(1)$, uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, and $\beta \in \mathcal{B}$. Moreover, using the similar argument to Theorem 2.6 in Li and Racine (2007, Chap. 2) and the fact that both k_1 and k_2 are κth order bias reduction kernel, we can show $\dagger_2 = O_p\left(h_1^\kappa + h_2^\kappa + \left(\ln n/nh_1h_2^{p_Z}\right)^{1/2}\right)$, uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, and $\beta \in \mathcal{B}$. Consequently,

$$\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta) = O_p(\varrho_{1n}) \quad \text{and} \quad \frac{\partial \mathcal{H}_n(c,z,\beta)}{\partial \beta_2} - \frac{\partial \mathcal{H}(c,z,\beta)}{\partial \beta_2} = O_p(\varrho_{2n})$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, and $\beta \in \mathcal{B}$. Based on these preliminary results,

$$\mathcal{H}_n(c,z,\beta) - \mathcal{H}(c,z,\beta) - \alpha_{OO}^{-1}(c,z,\beta)\frac{1}{nh_1h_2^{p_Z}}\sum_{i=1}^{n}\left(D_i(c)-\tau\right)$$
$$\times k_1\left(\frac{X_i'\beta + \mathcal{H}(c,z,\beta)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i - z}{h_2}\right) = O_p(\varrho_{1n}\varrho_{2n}) = o_p\left(n^{-1/2}\right),$$

uniformly in $c \in (\underline{c}, \overline{c})$, $z \in \mathcal{Z}$, and $\beta \in \mathcal{B}$. ■

Proof of Proposition 3.1. We prove the consistency by verifying the following conditions of Newey and McFadden (1994, Thm. 2.1) successively: (a) $Q(z,\beta)$ is uniquely maximized at β_0 for any $z \in \mathcal{Z}$, where

$$Q(z,\beta) = \int q(c,z,\beta)\omega(c)dc,$$

with

$$q(c,z,\beta) = E\left\{\left(D(c)-\tau\right)I\left\{X'\beta + \mathcal{H}(c,z,\beta) > 0\right\}\middle| Z = z\right\},$$
$$\mathcal{H}(c,z,\beta) = \arg\max_a q(c,z,\beta,a),$$

and

$$q(c,z,\beta,a) = E\left\{\left(D(c)-\tau\right)I\left\{X'\beta + a > 0\right\}\middle| Z = z\right\}.$$

(b) $Q_n(z,\beta)$ is maximized over a compact set $\mathcal{B} \subset \mathbb{R}^{p_X}$ that contains the true parameters as the interior. (c) $Q(z,\beta)$ is continuous in β. (d) $Q_n(z,\beta)$ converges to $Q(z,\beta)$ in probability uniformly in $\beta \in \mathcal{B}$ and $z \in \mathcal{Z}$.

Among them, (b) and (c) are trivial by Assumptions 3.4 and 3.2 respectively. For (a), notice that by definition, $\mathcal{H}(c,z,\beta_0) = \mathcal{H}(c,z) = m(z) - H(c)$. It follows from the rank condition (2.1.2) that $q(c,z,\beta)$ is maximized at $\beta = \beta_0$. Moreover, since $\mathcal{H}(c,z,\beta)$ is continuous in β, one can follow Manski (1985)'s identification argument to show that

$$q(c,z,\beta) < q(c,z,\beta_0) = q(c,z,\beta_0, m(z) - H(c))$$

for any $\beta \neq \beta_0$. Then

$$Q(z,\beta_0) - Q(z,\beta) = \int\left(q(c,z,\beta_0) - q(c,z,\beta)\right)\omega(c)dc > 0$$

for any $\beta \neq \beta_0$. For (d), it follows from Horowitz (1992)'s Lemma 4 and the uniform law of large numbers (Pollard, 1995) that

$$q_n(c,z,\beta,a) - q(c,z,\beta,a) = o_p(1)$$

uniformly in $c \in (\underline{c}, \overline{c})$, $(\beta, a) \in \mathcal{B} \times \mathcal{M}_H$, and $z \in \mathcal{Z}$. Then

$$Q_n(z,\beta) - Q(z,\beta) = \int\left(q_n(c,z,\beta) - q(c,z,\beta)\right)\omega(c)dc$$
$$= \int\left(\sup_{a\in\mathcal{M}_H} q_n(c,z,\beta,a) - \sup_{a\in\mathcal{M}_H} q(c,z,\beta,a)\right)\omega(c)dc = o_p(1)$$

uniformly in $\beta \in \mathcal{B}$ and $z \in \mathcal{Z}$ due to the nonnegativity and boundedness of $\omega(\cdot)$.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

To derive the limiting distribution for $\widehat{\beta}_{2n}$, we first establish the asymptotic representation for $\widehat{\beta}_{2n}(z) - \beta_{20}$. Taylor expansion of the first order condition that $\widehat{\beta}_{2n}(z)$ should satisfy, i.e., $\partial Q_n(z, \widehat{\beta}_n(z))/\partial\beta_2 = 0$ gives

$$\widehat{\beta}_{2n}(z) - \beta_{20} = \left[-\frac{\partial^2 Q_n(z, \overline{\beta}_n(z))}{\partial\beta_2 \partial\beta_2'} \right]^{-1} \times \frac{\partial Q_n(z, \beta_0)}{\partial\beta_2},$$

where $\overline{\beta}_n(z)$ lies between $\widehat{\beta}_n(z)$ and β_0. First consider $\frac{\partial^2 Q_n(z, \overline{\beta}_n(z))}{\partial\beta_2 \partial\beta_2'}$. Notice

$$\begin{aligned}
\frac{\partial Q_n(z,\beta)}{\partial\beta_2} &= \int \frac{1}{nh_1 h_2^{p_Z}} \sum_{i=1}^{n} (D_i(c) - \tau) k_1 \left(\frac{X_i'\beta + \mathcal{H}_n(c,z,\beta)}{h_1} \right) \\
&\quad \times \mathbf{k}_2 \left(\frac{Z_i - z}{h_2} \right) \left(X_{2i} + \frac{\partial \mathcal{H}_n(c,z,\beta)}{\partial\beta_2} \right) \omega(c) dc \\
&= \int \frac{1}{nh_1 h_2^{p_Z}} \sum_{i=1}^{n} (D_i(c) - \tau) k_1 \left(\frac{X_i'\beta + \mathcal{H}_n(c,z,\beta)}{h_1} \right) \\
&\quad \times \mathbf{k}_2 \left(\frac{Z_i - z}{h_2} \right) X_{2i} \omega(c) dc
\end{aligned}$$

since

$$\frac{1}{nh_1 h_2^{p_Z}} \sum_{i=1}^{n} (D_i(c) - \tau) k_1 \left(\frac{X_i'\beta + \mathcal{H}_n(c,z,\beta)}{h_1} \right) \mathbf{k}_2 \left(\frac{Z_i - z}{h_2} \right) = 0 \tag{A.2}$$

by the definition of $\mathcal{H}_n(c,z,\beta)$. Thus

$$\begin{aligned}
\frac{\partial^2 Q_n(z, \overline{\beta}_n(z))}{\partial\beta_2 \partial\beta_2'} &= \int \frac{1}{nh_1^2 h_2^{p_Z}} \sum_{i=1}^{n} (D_i(c) - \tau) k_1' \left(\frac{X_i'\overline{\beta}_n(z) + \mathcal{H}_n(c,z,\overline{\beta}_n(z))}{h_1} \right) \\
&\quad \times \mathbf{k}_2 \left(\frac{Z_i - z}{h_2} \right) X_{2i} \left(X_{2i} + \frac{\partial \mathcal{H}_n(c,z,\overline{\beta}_n(z))}{\partial\beta_2} \right)' \omega(c) dc \\
&=^{\dagger_3} \int \frac{1}{nh_1^2 h_2^{p_Z}} \sum_{i=1}^{n} (D_i(c) - \tau) k_1' \left(\frac{X_i'\overline{\beta}_n(z) + \mathcal{H}_n(c,z,\overline{\beta}_n(z))}{h_1} \right) \\
&\quad \times \mathbf{k}_2 \left(\frac{Z_i - z}{h_2} \right) \left(X_{2i} X_{2i}' - \frac{\partial \mathcal{H}_n(c,z,\overline{\beta}_n(z))}{\partial\beta_2} \frac{\partial \mathcal{H}_n(c,z,\overline{\beta}_n(z))}{\partial\beta_2'} \right) \\
&\quad \times \omega(c) dc \\
&=^{\dagger_4} \int \frac{1}{nh_1^2 h_2^{p_Z}} \sum_{i=1}^{n} (D_i(c) - \tau) k_1' \left(\frac{X_i'\overline{\beta}_n(z) + \mathcal{H}_n(c,z,\overline{\beta}_n(z))}{h_1} \right) \\
&\quad \times \mathbf{k}_2 \left(\frac{Z_i - z}{h_2} \right) \left(X_{2i} X_{2i}' - \frac{\partial \mathcal{H}(c,z,\overline{\beta}_n(z))}{\partial\beta_2} \frac{\partial \mathcal{H}(c,z,\overline{\beta}_n(z))}{\partial\beta_2'} \right) \\
&\quad \times \omega(c) dc =^{\dagger_5} \sigma(z) + o_p(1),
\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

where $\dagger_3$ follows from

$$\frac{1}{nh_1^2h_2^{pz}}\sum_{i=1}^{n}(D_i(c)-\tau)k_1'\left(\frac{X_i'\beta+\mathcal{H}_n(c,z,\beta)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\left(X_{2i}+\frac{\partial\mathcal{H}_n(c,z,\beta)}{\partial\beta_2}\right)=0$$

by differentiating (A.2) with respect to β_2; $\dagger_4$ follows from the consistency of $\frac{\partial\mathcal{H}_n(c,z,\beta)}{\partial\beta_2}$ uniformly in $\beta\in\mathcal{B}$ by Lemma A.1 and $\dagger_5$ follows from $\overline{\beta}_n(z)-\beta_0=o_p(1)$ and similar reasoning to the proof of Lemma A.1. Next,

$$\begin{aligned}
\frac{\partial Q_n(z,\beta_0)}{\partial\beta_2} &= \int\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}(D_i(c)-\tau)k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n(c,z,\beta_0)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\omega(c)dc\\
&=^{\dagger_6}\int\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}(D_i(c)-\tau)k_1\left(\frac{X_i'\beta_0+\mathcal{H}(c,z)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\\
&\quad X_{2i}\omega(c)dc+\int\frac{1}{nh_1^2h_2^{pz}}\sum_{i=1}^{n}(D_i(c)-\tau)k_1'\left(\frac{X_i'\beta_0+\overline{\mathcal{H}}_n(c,z,\beta_0)}{h_1}\right)\\
&\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\Big(\mathcal{H}_n(c,z,\beta_0)-\mathcal{H}(c,z)\Big)\omega(c)dc\\
&=^{\dagger_7}\int\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}(D_i(c)-\tau)k_1\left(\frac{X_i'\beta_0+\mathcal{H}(c,z)}{h_1}\right)\\
&\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\left(X_{2i}+\frac{\alpha_{XO}(c,z)}{\alpha_{OO}(c,z)}\right)\omega(c)dc+o_p(n^{-1/2})\\
&=^{\dagger_8}\frac{1}{nh_2^{pz}}\sum_{i=1}^{n}\Big(I\{U_i<0\}-\tau\Big)\\
&\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\left(X_{2i}+\frac{\alpha_{XO}\Big(H^{-1}\big(X_i'\beta_0+m(z),z\big)\Big)}{\alpha_{OO}\Big(H^{-1}\big(X_i'\beta_0+m(z),z\big)\Big)}\right)\\
&\quad\times\frac{\omega\Big(H^{-1}\big(X_i'\beta_0+m(z)\big)\Big)}{H'\Big(H^{-1}\big(X_i'\beta_0+m(z)\big)\Big)}+O_p\Big(h_1^{\kappa}/h_2^{pz}+h_2^{\kappa}\Big)+o_p(n^{-1/2})\\
&=\frac{1}{nh_2^{pz}}\sum_{i=1}^{n}\Big(I\{U_i<0\}-\tau\Big)\\
&\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\left(X_{2i}+\frac{\alpha_{XO}\Big(H^{-1}\big(X_i'\beta_0+m(z),z\big)\Big)}{\alpha_{OO}\Big(H^{-1}\big(X_i'\beta_0+m(z),z\big)\Big)}\right)\\
&\quad\times\frac{\omega\Big(H^{-1}\big(X_i'\beta_0+m(z)\big)\Big)}{H'\Big(H^{-1}\big(X_i'\beta_0+m(z)\big)\Big)}+o_p(n^{-1/2}),
\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

where $\dagger_6$ follows from the Taylor expansion with respect to $\mathcal{H}_n(c,z,\beta_0)$ around $\mathcal{H}(c,z,\beta_0)=\mathcal{H}(c,z)$, $\dagger_7$ follows from Lemma A.1, and $\dagger_8$ follows from the change of variables $c=H^{-1}(X_i'\beta_0+m(z)-h_1v)$. In summary, we have established

$$\begin{aligned}\widehat{\beta}_{2n}(z)-\beta_{20}=\frac{1}{nh_2^{p_Z}}\sum_{i=1}^{n}\big(I\{U_i<0\}-\tau\big)&\\ \times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\sigma^{-1}(z)&\left(X_{2i}+\frac{\alpha_{XO}\big(H^{-1}(X_i'\beta_0+m(z),z)\big)}{\alpha_{OO}\big(H^{-1}(X_i'\beta_0+m(z),z)\big)}\right)\\ \times\frac{\omega\big(H^{-1}(X_i'\beta_0+m(z))\big)}{H'\big(H^{-1}(X_i'\beta_0+m(z))\big)}&+o_p(n^{-1/2}).\end{aligned}$$

Plugging the above linear representation for $\widehat{\beta}_{2n}(z)-\beta_{20}$ into

$$\begin{aligned}\widehat{\beta}_{2n}-\beta_{20}&=\frac{\frac{1}{n}\sum_{i=1}^{n}t(Z_i)\widehat{\beta}_{2n}(Z_i)}{\frac{1}{n}\sum_{i=1}^{n}t(Z_i)}-\beta_{20}\\ &=\frac{1}{n\Pr(Z\in\mathcal{Z}_1)}\sum_{i=1}^{n}t(Z_i)\big(\widehat{\beta}_{2n}(Z_i)-\beta_{20}\big)+o_p(n^{-1/2})\end{aligned}$$

gives

$$\begin{aligned}\widehat{\beta}_{2n}-\beta_{20}=\frac{1}{n(n-1)h_2^{p_Z}\Pr(Z\in\mathcal{Z}_1)}&\sum_{i=1}^{n}\sum_{j\neq i}t(Z_i)\left(I\{U_j<0\}-\tau\right)\\ \times\mathbf{k}_2\left(\frac{Z_i-Z_j}{h_2}\right)\sigma^{-1}(Z_i)&\left(X_{2j}+\frac{\alpha_{XO}\left(H^{-1}(X_j'\beta_0+m(Z_i),Z_i)\right)}{\alpha_{OO}\left(H^{-1}(X_j'\beta_0+m(Z_i),Z_i)\right)}\right)\\ \times\frac{\omega\left(H^{-1}(X_j'\beta_0+m(Z_i))\right)}{H'\left(H^{-1}(X_j'\beta_0+m(Z_i))\right)}&+o_p(n^{-1/2})\\ =\frac{1}{n(n-1)\Pr(Z\in\mathcal{Z}_1)}&\sum_{i=1}^{n}\sum_{j\neq i}\mathcal{F}_n(\xi_i,\xi_j)+o_p(n^{-1/2}),\end{aligned}$$

where $\xi_i=(X_i,Z_i,U_i)$ and

$$\begin{aligned}\mathcal{F}_n(\xi_i,\xi_j)=t(Z_i)\left(I\{U_j<0\}-\tau\right)&h_2^{-p_Z}\mathbf{k}_2\left(\frac{Z_i-Z_j}{h_2}\right)\sigma^{-1}(Z_i)\\ \times\left(X_{2j}+\frac{\alpha_{XO}\left(H^{-1}(X_j'\beta_0+m(Z_i),Z_i)\right)}{\alpha_{OO}\left(H^{-1}(X_j'\beta_0+m(Z_i),Z_i)\right)}\right)&\frac{\omega\left(H^{-1}(X_j'\beta_0+m(Z_i))\right)}{H'\left(H^{-1}(X_j'\beta_0+m(Z_i))\right)}.\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

According to Powell, Stock, and Stoker (1989), the condition sufficient for U-statistic projection theorem is

$$E\left\|\mathcal{F}_n(\xi_i,\xi_j)\right\|^2 = o(n). \tag{A.3}$$

Notice that the only term in $\mathcal{F}_n$ that becomes unbounded as $n\to\infty$ is $h_2^{-p_Z}\mathbf{k}_2\left(\frac{Z_i-Z_j}{h_2}\right)$, which has the order $O\left(h_2^{-p_Z}\right)$. Thus (A.3) holds by Assumption 3.7(ii). The desired result follows from the U-statistic projection

$$\begin{aligned}\widehat{\beta}_{2n}-\beta_{20} = \frac{1}{n}\sum_{i=1}^{n}\frac{t(Z_i)}{\Pr(Z\in\mathcal{Z}_1)}\Big(I\{U_i<0\}-\tau\Big)\sigma^{-1}(Z_i)f_Z(Z_i)\\ \times\left(X_{2i}+\frac{\alpha_{XO}\Big(H^{-1}\big(X_i'\beta_0+m(Z_i),Z_i\big)\Big)}{\alpha_{OO}\Big(H^{-1}\big(X_i'\beta_0+m(Z_i),Z_i\big)\Big)}\right)\\ \times\frac{\omega\Big(H^{-1}\big(X_i'\beta_0+m(Z_i)\big)\Big)}{H'\Big(H^{-1}\big(X_i'\beta_0+m(Z_i)\big)\Big)}+o_p(n^{-1/2})\\ = \frac{1}{n}\sum_{i=1}^{n}\psi_{\beta i}+o_p(n^{-1/2}).\end{aligned}$$

■

Proof of Proposition 3.2. First, by rank condition (2.1.2), Assumptions 3.1–3.5, and applying similar argument as Horowitz (1992)'s Theorem 1, we can show $\widehat{\mathcal{H}}_n(c,z)$ that maximizes the smoothed maximum score criterion function (2.2.2) is a consistent estimator for $\mathcal{H}(c,z)=m(z)-H(c)$ for any given $z\in\mathcal{Z}$ and $c\in(\underline{c},\overline{c})$. Thus we first establish the asymptotic representation for $\widehat{\mathcal{H}}_n(c,z)-\mathcal{H}(c,z)$. Taylor expansion of the first order condition that $\widehat{\mathcal{H}}_n(c,z)$ should satisfy, i.e.,

$$\begin{aligned}\left.\frac{\partial q_n(c,z,\widehat{\beta}_n,a)}{\partial a}\right|_{a=\widehat{\mathcal{H}}_n(c,z)} = \frac{1}{nh_1h_2^{p_Z}}\sum_{i=1}^{n}\big(D_i(c)-\tau\big)k_1\left(\frac{X_i'\widehat{\beta}_n+\widehat{\mathcal{H}}_n(c,z)}{h_1}\right)\\ \times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)=0\end{aligned}$$

gives

$$\begin{aligned}\widehat{\mathcal{H}}_n(c,z)-\mathcal{H}(c,z) = \left[\frac{1}{nh_1^2h_2^{p_Z}}\sum_{i=1}^{n}\big(D_i(c)-\tau\big)k_1'\left(\frac{X_i'\widehat{\beta}_n+\overline{\mathcal{H}}_n(c,z)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\right]^{-1}\\ \times\frac{1}{nh_1h_2^{p_Z}}\sum_{i=1}^{n}\big(D_i(c)-\tau\big)k_1\left(\frac{X_i'\widehat{\beta}_n+\mathcal{H}(c,z)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right),\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

where $\overline{\mathcal{H}}_n(c,z)$ lies between $\mathcal{H}(c,z)$ and $\widehat{\mathcal{H}}_n(c,z)$. Applying Lemma A.1, we get

$$\widehat{\mathcal{H}}_n(c,z)-\mathcal{H}(c,z)=\frac{1}{nh_1h_2^{p_Z}\alpha_{OO}(c,z)}\sum_{i=1}^{n}\left(D_i(c)-\tau\right)k_1\left(\frac{X_i'\beta_0+\mathcal{H}(c,z)}{h_1}\right)\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)+o_p\left(n^{-1/2}\right).$$

To obtain the limiting distribution for $\widehat{m}_n(z)-m(z)$, we plugging the above representation for $\widehat{\mathcal{H}}_n(c,z)-\mathcal{H}(c,z)$ into (2.2.3) gives

$$\begin{aligned}\widehat{m}_n(z)-m(z) &= \int\left(\widehat{\mathcal{H}}_n(c,z)-\mathcal{H}(c,z)\right)\omega(c)dc\\ &= \frac{1}{nh_1h_2^{p_Z}}\sum_{i=1}^{n}\int\frac{D_i(c)-\tau}{\alpha_{OO}(c,z)}k_1\left(\frac{X_i'\beta_0+\mathcal{H}(c,z)}{h_1}\right)\\ &\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\omega(c)dc+o_p\left(n^{-1/2}\right)\\ &=^{\ddagger_1}\frac{1}{nh_2^{p_Z}}\sum_{i=1}^{n}\frac{I\{U_i\le 0\}-\tau}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0+m(z),z\right)\right)}\\ &\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\frac{\omega\left(H^{-1}\left(X_i'\beta_0+m(z)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0+m(z)\right)\right)}+O_p\left(h_1^{\kappa}/h_2^{p_Z}+h_2^{\kappa}\right)\\ &\quad+o_p\left(n^{-1/2}\right),\end{aligned}$$

where $\ddagger_1$ follows from the change of variables $c=H^{-1}(X_i'\beta_0+m(z)-h_1v)$ and noting that both k_1 and k_2 are κth order bias reduction kernels. Thus

$$\begin{aligned}\sqrt{nh_2^{p_Z}}\left(\widehat{m}_n(z)-m(z)\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{I\{U_i\le 0\}-\tau}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0+m(z),z\right)\right)}\frac{1}{\sqrt{h_2^{p_Z}}}\\ &\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)\frac{\omega\left(H^{-1}\left(X_i'\beta_0+m(z)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0+m(z)\right)\right)}\\ &\quad+\left(O_p\left(\sqrt{n}h_1^{\kappa}/h_2^{p_Z}+\sqrt{n}h_2^{\kappa}\right)+o_p(1)\right)\cdot O_p\left(h_2^{p_Z/2}\right)\\ &=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{mi}+o_p(1)\end{aligned}$$

with $E\psi_{mi}=0$ and

$$\begin{aligned}\mathrm{Var}(\psi_{mi}) &= \tau(1-\tau)\left(\int k_2^2(v)dv\right)^{p_Z}\\ &\quad\times E_X\left\{\left[\frac{\omega\left(H^{-1}\left(X'\beta_0+m(z)\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X'\beta_0+m(z),z\right)\right)H'\left(H^{-1}\left(X'\beta_0+m(z)\right)\right)}\right]^2 f_{Z|X}(z|X)\right\}.\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

To establish the limiting distribution for $\widehat{H}_n(c) - H(c)$, we note by construction,

$$\sqrt{nh_1}\left(\widehat{H}_n(c) - H(c)\right) = \sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{m}_n(Z_i) - E(m(Z))\right] - \sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{\mathcal{H}}_n(c, Z_i) - E(\mathcal{H}(c, Z))\right],$$

where $E\mathcal{H}(c, Z) = Em(Z) - H(c)$. It follows from $\sqrt{nh_2^{p_Z}}\left(\widehat{m}_n(z) - m(z)\right) = O_p(1)$ for each $z \in \mathcal{Z}$ that

$$\sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{m}_n(Z_i) - E(m(Z))\right] = \sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{m}_n(Z_i) - m(Z_i))\right] + \sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[m(Z_i) - E(m(Z))\right] = o_p(1)$$

as long as $h_1/h_2^{p_Z} \to 0$. Thus applying the U-statistic projection theorem, we have

$$\begin{aligned}
&\sqrt{nh_1}\left(\widehat{H}_n(c) - H(c)\right)\\
&\quad = -\sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{\mathcal{H}}_n(c, Z_i) - E(\mathcal{H}(c, Z))\right] + o_p(1)\\
&\quad = -\sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{\mathcal{H}}_n(c, Z_i) - \mathcal{H}(c, Z_i)\right]\\
&\qquad -\sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\mathcal{H}(c, Z_i) - E\mathcal{H}(c, Z)\right] + o_p(1)\\
&\quad = -\sqrt{\frac{h_1}{n}}\sum_{i=1}^{n}\left[\widehat{\mathcal{H}}_n(c, Z_i) - \mathcal{H}(c, Z_i)\right] + o_p(1)\\
&\quad = -\frac{1}{\sqrt{nh_1}(n-1)h_2^{p_Z}}\sum_{j\neq i}\frac{D_j(c) - \tau}{\alpha_{OO}(c, Z_i)}k_1\left(\frac{X_j'\beta_0 + \mathcal{H}(c, Z_i)}{h_1}\right)\\
&\qquad \times \mathbf{k}_2\left(\frac{Z_i - Z_j}{h_2}\right) + o_p(1)\\
&\quad = \overset{\ddagger_2}{} -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{D_j(c) - \tau}{\alpha_{OO}\left(c, Z_j\right)}f_Z\left(Z_j\right)\frac{1}{\sqrt{h_1}}k_1\left(\frac{X_j'\beta_0 + m\left(Z_j\right) - H(c)}{h_1}\right)\\
&\qquad + o_p(1) = -\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{\psi}_{Hj} + o_p(1).
\end{aligned}$$

Now define

$$\psi_{Hj} = \frac{I\{U_j < 0\} - \tau}{\alpha_{OO}\left(c, Z_j\right)}f_Z\left(Z_j\right)\frac{1}{\sqrt{h_1}}k_1\left(\frac{X_j'\beta_0 + m\left(Z_j\right) - H(c)}{h_1}\right).$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

By straightforward calculation,

$$
\begin{aligned}
&E\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\widetilde{\psi}_{Hj}-\psi_{Hj}\right)\\
&\quad=E\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{\psi}_{Hj}=\sqrt{nh_1}\int\frac{D(c)-\tau}{\alpha_{OO}(c,Z)}f_Z(Z)\frac{1}{h_1}k_1\left(\frac{X'\beta_0+m(Z)-H(c)}{h_1}\right)dF_{UXZ}\\
&\quad=\sqrt{nh_1}\int\frac{I\left\{X'\beta_0+m(Z)-H(c)>U\right\}-\tau}{\alpha_{OO}(c,Z)}f_Z(Z)\frac{1}{h_1}k_1\\
&\qquad\times\left(\frac{X'\beta_0+m(Z)-H(c)}{h_1}\right)dF_{UXZ}\\
&\quad=^{\ddagger_3}\sqrt{nh_1}\int\frac{I\left\{h_1s>U\right\}-\tau}{\alpha_{OO}(c,Z)}f_Z(Z)k_1(s)\,f_{UX_1X_2Z}\\
&\qquad\times\left(U,h_1s-X_2'\beta_{20}-m(Z)+H(c),X_2,Z\right)dUdsdX_2dZ\\
&\quad=^{\ddagger_4}\sqrt{nh_1}\left\{\int\frac{I\left\{0>U\right\}-\tau}{\alpha_{OO}(c,Z)}f_Z(Z)f_{UX_1X_2Z}\right.\\
&\qquad\left.\times\left(U,H(c)-X_2'\beta_{20}-m(Z),X_2,Z\right)dUdsdX_2dZ+o\left(h_1^{\kappa}\right)\right\}\\
&\quad=^{\ddagger_5}o\left(\sqrt{n}h_1^{\kappa+1/2}\right)=o(1),
\end{aligned}
$$

where $\ddagger_3$ follows from the change of variables $X_1+X_2'\beta_{20}+m(Z)-H(c)=h_1s$, $\ddagger_4$ follows from the fact that k_1 is a κth order bias reduction kernel, and $\ddagger_5$ follows because of the zero τth quantile restriction on U conditional on (X,Z). Moreover, by similar but more tedious computation, we can show $\mathrm{Var}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left(\widetilde{\psi}_{Hj}-\psi_{Hj}\right)=o(1)$. Thus we conclude

$$
\sqrt{nh_1}\left(\widehat{H}_n(c)-H(c)\right)=-\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\psi_{Hj}+o_p(1)
$$

with $E\left(\psi_{Hj}\right)=0$ and

$$
\begin{aligned}
\mathrm{Var}\left(\psi_{Hj}\right)=\tau(1-\tau)&\left(\int k_1^2(v)dv\right)E_{X_2Z}\\
&\times\left\{\left[\frac{f_Z(Z)}{\alpha_{OO}(c,Z)}\right]^2f_{X_1|X_2Z}\left(H(c)-m(Z)-X_2'\beta_{20}\middle|X_2,Z\right)\right\}.
\end{aligned}
$$

■

Proof of Proposition 3.3. Let θ_0 and $\widehat{\theta}_n$ signify $E\left[\frac{1}{H'\left(H^{-1}(X'\beta_0+m(Z))\right)}\right]$ and its estimator

$$
\frac{1}{n}\sum_{i=1}^{n}\int\frac{1}{h_1}k_1\left(\frac{X_i'\widehat{\beta}_n+\widehat{m}_n(Z_i)-\widehat{H}_n(c)}{h_1}\right)dc
$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

respectively. Since

$$\sqrt{n}\left(\widehat{\vartheta}_n-\vartheta_0\right)=\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)\widehat{\beta}_n+\theta_0\sqrt{n}\left(\widehat{\beta}_n-\beta_0\right), \tag{A.4}$$

then it suffices to consider the limiting distribution for $\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)$. Taylor expansion of $\widehat{\theta}_n$ with respect to $\left(\widehat{\beta}_n,\widehat{m}_n(z),\widehat{H}_n(c)\right)$ around their probability limit $\left(\beta_0,m(z),H(c)\right)$ gives

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)=&\underbrace{\sqrt{n}\left(\frac{1}{nh_1}\sum_{i=1}^{n}\int k_1\left(\frac{X_i'\beta_0+m(Z_i)-H(c)}{h_1}\right)dc-\theta_0\right)}_{\ddagger_6}\\&+\underbrace{\frac{1}{nh_1^2}\sum_{i=1}^{n}\int k_1'\left(\frac{X_i'\overline{\beta}_n+\overline{m}_n(Z_i)-\overline{H}_n(c)}{h_1}\right)X_{2i}'dc\times\sqrt{n}\left(\widehat{\beta}_{2n}-\beta_{20}\right)}_{\ddagger_7}\\&+\underbrace{\frac{1}{\sqrt{n}h_1^2}\sum_{i=1}^{n}\int k_1'\left(\frac{X_i'\overline{\beta}_n+\overline{m}_n(Z_i)-\overline{H}_n(c)}{h_1}\right)dc\left(\widehat{m}_n(Z_i)-m(Z_i)\right)}_{\ddagger_8}\\&-\underbrace{\frac{1}{\sqrt{n}h_1^2}\sum_{i=1}^{n}\int k_1'\left(\frac{X_i'\overline{\beta}_n+\overline{m}_n(Z_i)-\overline{H}_n(c)}{h_1}\right)\left(\widehat{H}_n(c)-H(c)\right)dc}_{\ddagger_9}.\end{aligned}$$

For $\ddagger_6$, it follows from the change of variables $X_i'\beta_0+m(Z_i)-H(c)=h_1s$ and the κth order bias reduction kernel k_1 that

$$\begin{aligned}\ddagger_6&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{1}{H'\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)}-\theta_0\right]+O_p\left(\sqrt{n}h_1^{\kappa}\right)\\&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\Pi\left(X_i'\beta_0+m(Z_i)\right)-\theta_0\right]+o_p(1),\end{aligned}$$

where $\Pi(y)=1/H'\left(H^{-1}(y)\right)$. For $\ddagger_7$, it again follows from similar change of variables, $\int k_1'(s)ds=0$, $\int k_1'(s)sds=-1$ and the κth order bias reduction kernel k_1 that

$$\begin{aligned}\int\frac{1}{h_1^2}k_1'\left(\frac{X'\beta_0+m(Z)-H(c)}{h_1}\right)dc&=\int\frac{1}{h_1}k_1'(s)\,\Pi\left(X'\beta_0+m(Z)-h_1s\right)ds\\&=\Pi'\left(X'\beta_0+m(Z)\right)+O\left(h_1^{\kappa}\right).\end{aligned}$$

Thus we can show

$$\ddagger_7=E\left\{\Pi'\left(X'\beta_0+m(Z)\right)X_2'\right\}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\beta i}+o_p(1).$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

For $\ddagger_8$, using the same argument for $\ddagger_7$ and plugging in the representation for $\widehat{m}_n(z)-m(z)$ established in Proposition 3.2, we arrive at

$$\begin{aligned}\ddagger_8 &= \frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\Pi'\left(X_i'\beta_0+m(Z_i)\right)\frac{I\{U_j\le 0\}-\tau}{\alpha_{OO}\left(H^{-1}\left(X_j'\beta_0+m(Z_i),Z_i\right)\right)}\\ &\quad\times\frac{1}{h_2^{p_Z}}\mathbf{k}_2\left(\frac{Z_j-Z_i}{h_2}\right)\frac{\omega\left(H^{-1}\left(X_j'\beta_0+m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_j'\beta_0+m(Z_i)\right)\right)}+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}E_X\left[\Pi'(X'\beta_0+m(Z_j))\right]f_Z(Z_j)\frac{I\{U_j\le 0\}-\tau}{\alpha_{OO}\left(H^{-1}\left(X_j'\beta_0+m(Z_j),Z_j\right)\right)}\\ &\quad\times\frac{\omega\left(H^{-1}\left(X_j'\beta_0+m(Z_j)\right)\right)}{H'\left(H^{-1}\left(X_j'\beta_0+m(Z_j)\right)\right)}+o_p(1)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta i}^{1}+o_p(1).\end{aligned}$$

For $\ddagger_9$, plugging in the representation for $\widehat{H}_n(c)-H(c)$ yields

$$\begin{aligned}\ddagger_9 &= \frac{1}{\sqrt{n}h_1}\int\frac{1}{nh_1^2}\sum_{i=1}^{n}k_1'\left(\frac{X_i'\overline{\beta}_n+\overline{m}_n(Z_i)-\overline{H}_n(c)}{h_1}\right)\sum_{j=1}^{n}\frac{I\{U_j<0\}-\tau}{\alpha_{OO}(c,Z_j)}k_1\\ &\quad\times\left(\frac{X_j'\beta_0+m(Z_j)-H(c)}{h_1}\right)dc\times f_Z(Z_j)+o_p(1)\\ &= \frac{1}{\sqrt{n}h_1}\sum_{j=1}^{n}\int\frac{\Gamma(c)}{\alpha_{OO}(c,Z_j)}k_1\left(\frac{X_j'\beta_0+m(Z_j)-H(c)}{h_1}\right)dc\\ &\quad\times\left(I\{U_j<0\}-\tau\right)f_Z(Z_j)+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\Gamma\left(H^{-1}\left(X_j'\beta_0+m(Z_j)\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_j'\beta_0+m(Z_j)\right),Z_j\right)}\left(I\{U_j<0\}-\tau\right)f_Z(Z_j)+o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta i}^{2}+o_p(1),\end{aligned}$$

where $\Gamma(c)=E_{X_2Z}\left[f'_{X_1|X_2Z}\left(H(c)-X_2'\beta_{20}-m(Z)|X_2,Z\right)\right]$ with $f'_{X_1|X_2Z}(\cdot|\cdot,\cdot)$ is the derivative of $f_{X_1|X_2Z}$ with respect to its first argument. The desired limiting distribution of $\sqrt{n}\left(\widehat{\vartheta}_n-\vartheta_0\right)$ directly follows from (A.4), all the asymptotic representations established above, and the limiting distribution of $\sqrt{n}\left(\widehat{\beta}_n-\beta_0\right)$ in Proposition 3.1. ■

Proof of Proposition 4.1. In parallel with the proof of Proposition 3.1, the consistency can be established by checking a set of conditions in Newey and McFadden (1994). To derive the limiting distribution of $\widehat{\beta}_{2n}^{r}$, we first establish the representation for $\widehat{\beta}_{2n}^{r}(z)-\beta_{20}$.

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Taylor expansion of the first order condition that $\widehat{\beta}_{2n}^{r}(z)$ should satisfy, i.e., $\partial Q_n^r\left(z,\widehat{\beta}_n^r(z)\right)/\partial\beta_2 = 0$ gives

$$\widehat{\beta}_{2n}^{r}(z)-\beta_{20} = \left[-\frac{\partial^2 Q_n^r\left(z,\overline{\beta}_n(z)\right)}{\partial\beta_2\partial\beta_2'}\right]^{-1}\times\frac{\partial Q_n^r(z,\beta_0)}{\partial\beta_2},$$

where $\overline{\beta}_n(z)$ lies between $\widehat{\beta}_n^r(z)$ and β_0. Similar to the fixed censoring case, we can show

$$\frac{\partial^2 Q_n^r\left(z,\overline{\beta}_n(z)\right)}{\partial\beta_2\partial\beta_2'} = \sigma(z)+o_p(1). \tag{A.5}$$

Notice

$$\begin{aligned}
&\frac{\partial Q_n^r(z,\beta_0)}{\partial\beta_2}\\
&= \int\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}\left(\frac{D_i(c)}{\widehat{G}_n(c)}-\tau\right)k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^r(c,z,\beta_0)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\omega(c)dc\\
&= \underbrace{\int\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}\left(\frac{D_i(c)}{G(c)}-\tau\right)k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^r(c,z,\beta_0)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\omega(c)dc}_{T_1}\\
&\quad+\underbrace{\int\frac{G(c)-\widehat{G}_n(c)}{G(c)}\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}\frac{D_i(c)}{\widehat{G}_n(c)}k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^r(c,z,\beta_0)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\omega(c)dc}_{T_2}.
\end{aligned}$$

By similar arguments as the proof of Proposition 3.1, we can show that

$$\begin{aligned}
T_1 &= \frac{1}{nh_2^{pz}}\sum_{i=1}^{n}\left(\frac{I\left\{H(Y_i)>X_i'\beta_0+m(Z_i)\right\}}{G\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)}-\tau\right)\\
&\quad\times\left(X_{2i}+\frac{\alpha_{XO}\left(H^{-1}\left(X_i'\beta_0+m(z),z\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0+m(z),z\right)\right)}\right)\times\frac{\omega\left(H^{-1}\left(X_i'\beta_0+m(z)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0+m(z)\right)\right)}\\
&\quad\times\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)+o_p\left(n^{-1/2}\right).
\end{aligned}$$

For T_2, it follows from the $\sqrt{n}$-consistency of $G_n(c)$ and similar argument to (A.5) that

$$\begin{aligned}
&\frac{1}{nh_1h_2^{pz}}\sum_{i=1}^{n}\frac{D_i(c)}{\widehat{G}_n(c)}k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^r(c,z,\beta_0)}{h_1}\right)\mathbf{k}_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\\
&\quad=\tau E\left\{f_{X_1Z|X_2}\left(-X_2'\beta_{20}-m(z)+H(c),z\middle|X_2\right)X_2\right\}+o_p(1)=\alpha_{RX}(c,z)+o_p(1).
\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Moreover, according to Fleming and Harrington (1991), Kaplan–Meier estimator permits the following stochastic integral representation:

$$\frac{G(c)-G_n(c)}{G(c)} = \frac{1}{n}\sum_{i=1}^{n}\int_0^c \frac{1}{\pi(s)} dM_i(s) + o_p\left(n^{-1/2}\right)$$

with $\pi(s) = \Pr(Y \geq s)$, $M_i(s) = I\{Y_i \leq s, \delta_i = 0\} - \int_0^s I\{Y_i \geq v\} d\Lambda(v)$, Λ is the cumulative hazard function of C. Combining these results together, we get

$$\begin{aligned} T_2 &= \frac{1}{n}\sum_{i=1}^{n}\int \alpha_{RX}(c,z)\int_0^c \frac{1}{\pi(s)} dM_i(s)\omega(c)dc + o_p\left(n^{-1/2}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\int \frac{\alpha^*_{RX}(s,z)}{\pi(s)} dM_i(s) + o_p\left(n^{-1/2}\right), \end{aligned}$$

with

$$\alpha^*_{RX}(s,z) = \int_s^{+\infty} \alpha_{RX}(c,z)\omega(c)dc.$$

Finally, applying the U-statistic projection theorem, we have

$$\begin{aligned} &\widehat{\beta}^r_{2n} - \beta_{20} \\ &= \frac{1}{n(n-1)h_2^{p_Z}\Pr(Z\in\mathcal{Z}_1)}\sum_{i=1}^{n}\sum_{j\neq i} t(Z_i)\left(\frac{I\left\{H(Y_j) > X_j'\beta_0 + m(Z_j)\right\}}{G\left(H^{-1}\left(X_j'\beta_0 + m(Z_j)\right)\right)} - \tau\right)\mathbf{k}_2\left(\frac{Z_i - Z_j}{h_2}\right) \\ &\quad \times\sigma^{-1}(Z_i)\left(X_{2j} + \frac{\alpha_{XO}\left(H^{-1}\left(X_j'\beta_0 + m(Z_i), Z_i\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_j'\beta_0 + m(Z_i), Z_i\right)\right)}\right)\frac{\omega\left(H^{-1}\left(X_j'\beta_0 + m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_j'\beta_0 + m(Z_i)\right)\right)} \\ &\quad + \frac{1}{n\Pr(Z\in\mathcal{Z}_1)}\sum_{i=1}^{n} t(Z_i)\sigma^{-1}(Z_i)\frac{1}{n}\sum_{j=1}^{n}\int \frac{\alpha^*_{RX}(s,Z_i)}{\pi(s)} dM_j(s) + o_p\left(n^{-1/2}\right) \\ &= \frac{1}{n\Pr(Z\in\mathcal{Z}_1)}\sum_{j=1}^{n} t(Z_j)\left(\frac{I\left\{H(Y_j) > X_j'\beta_0 + m(Z_j)\right\}}{G\left(H^{-1}\left(X_j'\beta_0 + m(Z_j)\right)\right)} - \tau\right)\sigma^{-1}(Z_j) f_Z(Z_j) \\ &\quad \times\left(X_{2j} + \frac{\alpha_{XO}\left(H^{-1}\left(X_j'\beta_0 + m(Z_j), Z_j\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_j'\beta_0 + m(Z_j), Z_j\right)\right)}\right)\frac{\omega\left(H^{-1}\left(X_j'\beta_0 + m(Z_j)\right)\right)}{H'\left(H^{-1}\left(X_j'\beta_0 + m(Z_j)\right)\right)} \\ &\quad + \frac{1}{n\Pr(Z\in\mathcal{Z}_1)}\sum_{j=1}^{n}\int \frac{E_Z\left[t(Z)\sigma^{-1}(Z)\alpha^*_{RX}(s,Z)\right]}{\pi(s)} dM_j(s) + o_p\left(n^{-1/2}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\psi^r_{\beta,1i} + \psi^r_{\beta,2i}\right) + o_p\left(n^{-1/2}\right). \end{aligned}$$

■

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Proof of Proposition 5.1. Notice that $Q_n^e(z,\beta)$ differs from $Q_n(z,\beta)$ in that the (control) variable Z_i's are not directly observed but replaced by their empirical counterpart $\widehat{Z}_i$'s. We begin the proof with calculating the asymptotic stochastic order of $\mathcal{H}_n^e(c,z,\beta) - \mathcal{H}(c,z,\beta)$ and $\partial\mathcal{H}_n^e(c,z,\beta)/\partial\beta_2 - \partial\mathcal{H}(c,z,\beta)/\partial\beta_2$ respectively, where $\mathcal{H}_n^e(c,z,\beta)$ is defined by (5.8) and $\mathcal{H}(c,z,\beta)$ is defined in Lemma A.1. Using the $\sqrt{n}$-consistency of $\widehat{\gamma}_n$,

$$
\begin{aligned}
&q_n^e(c,z,\beta,a) - q_n(c,z,\beta,a)\\
&\quad= \frac{1}{nh_2}\sum_{i=1}^n (D_i(c)-\tau)\,K_1\left(\frac{X_i'\beta+a}{h_1}\right)\left(k_2\left(\frac{\widehat{Z}_i - z}{h_2}\right) - k_2\left(\frac{Z_i - z}{h_2}\right)\right)\\
&\quad= \frac{1}{nh_2^2}\sum_{i=1}^n (D_i(c)-\tau)\,K_1\left(\frac{X_i'\beta+a}{h_1}\right)k_2'\left(\frac{\overline{Z}_i - z}{h_2}\right)(\widehat{Z}_i - Z_i)\\
&\quad= \frac{1}{n\sqrt{n}h_2^2}\sum_{i=1}^n (D_i(c)-\tau)\,K_1\left(\frac{X_i'\beta+a}{h_1}\right)k_2'\left(\frac{\overline{Z}_i - z}{h_2}\right)W_i \times \sqrt{n}(\widehat{\gamma}_n - \gamma_0)\\
&\quad= O_p\left(n^{-1/2}h_2^{-2}\right) = o_p(1),
\end{aligned}
$$

with $\overline{Z}_i$ lying between $\widehat{Z}_i$ and z. Moreover, $q_n(c,z,\beta,a) - q(c,z,\beta,a) = o_p(1)$ uniformly in $c \in (\underline{c},\overline{c})$, $z \in \mathcal{Z}$, and $(\beta,a) \in \mathcal{B}\times\mathcal{M}_H$, then we can conclude by Carroll et al. (1997)'s Lemma A.1 that

$$
\mathcal{H}_n^e(c,z,\beta) - \mathcal{H}(c,z,\beta) = o_p(1)
$$

uniformly in $c \in (\underline{c},\overline{c})$, $z \in \mathcal{Z}$, and $\beta \in \mathcal{B}$. To sharpen the stochastic order of $\mathcal{H}_n^e(c,z,\beta) - \mathcal{H}(c,z,\beta)$, Taylor expansion of the first order condition that $\mathcal{H}_n^e(c,z,\beta)$ should satisfy, i.e.,

$$
\frac{1}{nh_1h_2}\sum_{i=1}^n \big(D_i(c)-\tau\big)k_1\left(\frac{X_i'\beta + \mathcal{H}_n^e\big(c,z,\beta\big)}{h_1}\right)k_2\left(\frac{\widehat{Z}_i - z}{h_2}\right) = 0
$$

gives that

$$
\begin{aligned}
&\mathcal{H}_n^e(c,z,\beta) - \mathcal{H}(c,z,\beta)\\
&= \underbrace{\left[-\frac{1}{nh_1^2h_2}\sum_{i=1}^n \big(D_i(c)-\tau\big)k_1'\left(\frac{X_i'\beta + \overline{\mathcal{H}}_n^e(c,z,\beta)}{h_1}\right)k_2\left(\frac{\widehat{Z}_i - z}{h_2}\right)\right]^{-1}}_{\flat_1}\\
&\quad\times \underbrace{\frac{1}{nh_1h_2}\sum_{i=1}^n \big(D_i(c)-\tau\big)k_1\left(\frac{X_i'\beta + \mathcal{H}(c,z,\beta)}{h_1}\right)k_2\left(\frac{\widehat{Z}_i - z}{h_2}\right)}_{\flat_2},
\end{aligned}
$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

where $\overline{\mathcal{H}}_n^e(c,z,\beta)$ lies between $\mathcal{H}_n^e(c,z,\beta)$ and $\mathcal{H}(c,z,\beta)$. Since

$$\begin{aligned}
&\frac{1}{nh_1^2h_2}\sum_{i=1}^n\left(D_i(c)-\tau\right)k_1'\left(\frac{X_i'\beta+\overline{\mathcal{H}}_n^e(c,z,\beta)}{h_1}\right)\left(k_2\left(\frac{\widehat{Z}_i-z}{h_2}\right)-k_2\left(\frac{Z_i-z}{h_2}\right)\right)\\
&\quad=\frac{1}{nh_1^2h_2^2}\sum_{i=1}^n\left(D_i(c)-\tau\right)k_1'\left(\frac{X_i'\beta+\overline{\mathcal{H}}_n^e(c,z,\beta)}{h_1}\right)k_2'\left(\frac{\overline{Z}_i-z}{h_2}\right)\left(\widehat{Z}_i-Z_i\right)\\
&\quad=O_p\left(n^{-1/2}h_2^{-2}h_1^{-2}\right)=o_p(1),
\end{aligned}$$

we can demonstrate by similar reasoning to Lemma A.1 that $\flat_1=\alpha_{OO}(c,z,\beta)+o_p(1)$, uniformly in $c\in(\underline{c},\overline{c})$, $z\in\mathcal{Z}$, and $\beta\in\mathcal{B}$. Moreover, it follows from the fact that both k_1 and k_2 are κth order bias reduction kernel, that $\flat_2=O_p\left(h_1^\kappa+h_2^\kappa+\left(\ln n/nh_1h_2\right)^{1/2}+n^{-1/2}h_2^{-2}h_1^{-1}\right)$, uniformly in $c\in(\underline{c},\overline{c})$, $z\in\mathcal{Z}$, and $\beta\in\mathcal{B}$. Consequently,

$$\begin{aligned}
\mathcal{H}_n^e(c,z,\beta)-\mathcal{H}(c,z,\beta)&=O_p\left(h_1^\kappa+h_2^\kappa+\left(\ln n/nh_1h_2\right)^{1/2}+n^{-1/2}h_2^{-2}h_1^{-1}\right)\\
&=O_p\left(\overline{\varrho}_{1n}\right).
\end{aligned}\tag{A.6}$$

Similarly, we can show

$$\begin{aligned}
\frac{\partial\mathcal{H}_n^e(c,z,\beta)}{\partial\beta_2}-\frac{\partial\mathcal{H}(c,z,\beta)}{\partial\beta_2}&=O_p\left(h_1^\kappa+h_2^\kappa+\left(\ln n/nh_1^3h_2\right)^{1/2}+n^{-1/2}h_2^{-2}h_1^{-2}\right)\\
&=O_p\left(\overline{\varrho}_{2n}\right)
\end{aligned}\tag{A.7}$$

uniformly in $c\in(\underline{c},\overline{c})$, $z\in\mathcal{Z}$, and $\beta\in\mathcal{B}$. Moreover,

$$\begin{aligned}
&\mathcal{H}_n^e(c,z,\beta)-\mathcal{H}(c,z,\beta)-\alpha_{OO}^{-1}(c,z,\beta)\frac{1}{nh_1h_2}\sum_{i=1}^n\left(D_i(c)-\tau\right)\\
&\quad\times k_1\left(\frac{X_i'\beta+\mathcal{H}(c,z,\beta)}{h_1}\right)k_2\left(\frac{Z_i-z}{h_2}\right)=o_p\left(n^{-1/2}\right),
\end{aligned}\tag{A.8}$$

uniformly in $c\in(\underline{c},\overline{c})$, $z\in\mathcal{Z}$, and $\beta\in\mathcal{B}$.

Taylor expansion of the first order condition that $\widehat{\beta}_{2n}^e(z)$ should satisfy, i.e., $\partial Q_n^e\left(z,\widehat{\beta}_n^e(z)\right)/\partial\beta_2=0$ gives

$$\widehat{\beta}_{2n}^e(z)-\beta_{20}=\left[-\frac{\partial^2Q_n^e\left(z,\overline{\beta}_n^e(z)\right)}{\partial\beta_2\partial\beta_2'}\right]^{-1}\times\frac{\partial Q_n^e(z,\beta_0)}{\partial\beta_2},$$

where $\overline{\beta}_n^e(z)$ lies between $\widehat{\beta}_n^e(z)$ and β_0. Based on (A.6)–(A.8) and following the same manner as we establish the asymptotic properties of $\widehat{\beta}_n(z)$ in Proposition 3.1, we can show that

$$\frac{\partial^2Q_n^e\left(z,\overline{\beta}_n^e(z)\right)}{\partial\beta_2\partial\beta_2'}=\sigma(z)+o_p(1).$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

To analyze $\frac{\partial Q_n^e(z,\beta_0)}{\partial \beta_2}$, we decompose it into

$$\begin{aligned}
&\frac{\partial Q_n^e(z,\beta_0)}{\partial \beta_2} \\
&= \underbrace{\int \frac{1}{nh_1h_2}\sum_{i=1}^n (D_i(c)-\tau)k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^e(c,z,\beta_0)}{h_1}\right)k_2\left(\frac{Z_i-z}{h_2}\right)X_{2i}\omega(c)dc}_{\flat_3} \\
&= \underbrace{\int \frac{1}{nh_1h_2}\sum_{i=1}^n (D_i(c)-\tau)k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^e(c,z,\beta_0)}{h_1}\right)\left(k_2\left(\frac{\widehat{Z}_i-z}{h_2}\right)-k_2\left(\frac{Z_i-z}{h_2}\right)\right)X_{2i}\omega(c)dc}_{\flat_4}.
\end{aligned}$$

For $\flat_3$, using (A.8) and applying the same argument as in Proposition 3.1 give

$$\begin{aligned}
\flat_3 &= \frac{1}{nh_2}\sum_{i=1}^n\left(I\{U_i<0\}-\tau\right)k_2\left(\frac{Z_i-z}{h_2}\right)\left(X_{2i}+\frac{\alpha_{XO}\left(H^{-1}(X_i'\beta_0+m(z),z)\right)}{\alpha_{OO}\left(H^{-1}(X_i'\beta_0+m(z),z)\right)}\right) \\
&\quad\times\frac{\omega\left(H^{-1}(X_i'\beta_0+m(z))\right)}{H'\left(H^{-1}(X_i'\beta_0+m(z))\right)}+o_p\left(n^{-1/2}\right).
\end{aligned}$$

For $\flat_4$, applying the mean value theorem and the argument similar to Lemma A.1 gives

$$\begin{aligned}
\flat_4 &= \int \frac{1}{nh_1h_2^2}\sum_{i=1}^n (D_i(c)-\tau)k_1\left(\frac{X_i'\beta_0+\mathcal{H}_n^e(c,z,\beta_0)}{h_1}\right)k_2' \\
&\quad\times\left(\frac{\overline{Z}_i-z}{h_2}\right)X_{2i}W_i'\omega(c)dc\times(\widehat{\gamma}_n-\gamma_0) \\
&= \int\alpha_{XW}(c,z)\omega(c)dc\times\frac{1}{n}\sum_{i=1}^n\psi_{\gamma i}+o_p\left(n^{-1/2}\right)=\alpha_{XW}^*(z)\times\frac{1}{n}\sum_{i=1}^n\psi_{\gamma i}+o_p\left(n^{-1/2}\right).
\end{aligned}$$

In summary, we have established

$$\begin{aligned}
&\widehat{\beta}_{2n}^e(z)-\beta_{20} \\
&\quad=\frac{1}{nh_2}\sum_{i=1}^n\left(I\{U_i<0\}-\tau\right)k_2\left(\frac{Z_i-z}{h_2}\right)\sigma^{-1}(z) \\
&\qquad\times\left(X_{2i}+\frac{\alpha_{XO}\left(H^{-1}(X_i'\beta_0+m(z),z)\right)}{\alpha_{OO}\left(H^{-1}(X_i'\beta_0+m(z),z)\right)}\right)\times\frac{\omega\left(H^{-1}(X_i'\beta_0+m(z))\right)}{H'\left(H^{-1}(X_i'\beta_0+m(z))\right)} \\
&\qquad+\sigma^{-1}(z)\alpha_{XW}^*(z)\cdot\frac{1}{n}\sum_{i=1}^n\psi_{\gamma i}+o_p\left(n^{-1/2}\right).
\end{aligned}$$

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press

Plugging the above linear representation for $\widehat{\beta}^{e}_{2n}(z)-\beta_{20}$ into

$$\widehat{\beta}^{e}_{2n}-\beta_{20}=\frac{\frac{1}{n}\sum_{i=1}^{n}t(\widehat{Z}_i)\widehat{\beta}^{e}_{2n}(\widehat{Z}_i)}{\frac{1}{n}\sum_{i=1}^{n}t(\widehat{Z}_i)}$$

$$-\beta_{20}=\frac{1}{n\Pr(Z\in\mathcal{Z}_1)}\sum_{i=1}^{n}t(Z_i)\left(\widehat{\beta}^{e}_{2n}(Z_i)-\beta_{20}\right)+o_p\left(n^{-1/2}\right)$$

gives

$$\begin{aligned}
&\widehat{\beta}^{e}_{2n}-\beta_{20}\\
&=\frac{1}{n(n-1)h_2\Pr(Z\in\mathcal{Z}_1)}\sum_{i=1}^{n}\sum_{j\neq i}t(Z_i)\left(I\left\{U_j<0\right\}-\tau\right)k_2\left(\frac{Z_i-Z_j}{h_2}\right)\\
&\quad\times\sigma^{-1}(Z_i)\left(X_{2j}+\frac{\alpha_{XO}\left(H^{-1}\left(X_j'\beta_0+m(Z_i),Z_i\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_j'\beta_0+m(Z_i),Z_i\right)\right)}\right)\frac{\omega\left(H^{-1}\left(X_j'\beta_0+m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_j'\beta_0+m(Z_i)\right)\right)}\\
&\quad+\frac{1}{n\Pr(Z\in\mathcal{Z}_1)}\sum_{i=1}^{n}t(Z_i)\sigma^{-1}(Z_i)\alpha^{*}_{XW}(c,Z_i)\cdot\frac{1}{n}\sum_{j=1}^{n}\psi_{\gamma j}+o_p\left(n^{-1/2}\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\frac{t(Z_i)}{\Pr(Z\in\mathcal{Z}_1)}\left(I\left\{U_i<0\right\}-\tau\right)\sigma^{-1}(Z_i)f_Z(Z_i)\\
&\quad\times\left(X_{2i}+\frac{\alpha_{XO}\left(H^{-1}\left(X_i'\beta_0+m(Z_i),Z_i\right)\right)}{\alpha_{OO}\left(H^{-1}\left(X_i'\beta_0+m(Z_i),Z_i\right)\right)}\right)\times\frac{\omega\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)}{H'\left(H^{-1}\left(X_i'\beta_0+m(Z_i)\right)\right)}\\
&\quad+\frac{E_Z\left[t(Z)\sigma^{-1}(Z)\alpha^{*}_{XW}(Z)\right]}{\Pr(Z\in\mathcal{Z}_1)}\times\frac{1}{n}\sum_{i=1}^{n}\psi_{\gamma i}+o_p\left(n^{-1/2}\right)\\
&=\frac{1}{n}\sum_{i=1}^{n}\psi^{e}_{\beta i}+o_p\left(n^{-1/2}\right).
\end{aligned}$$

■

https://doi.org/10.1017/S0266466614000887 Published online by Cambridge University Press