10-2020

Unconditional quantile regression with high-dimensional data

Yuya SASAKI

Takuya URA

Yichong ZHANG
Singapore Management University, yczhang@smu.edu.sg

Unconditional Quantile Regression with High-Dimensional Data*

Yuya Sasaki†
Department of Economics
Vanderbilt University

Takuya Ura‡
Department of Economics
University of California, Davis

Yichong Zhang§
School of Economics
Singapore Management University

Abstract

Credible counterfactual analysis often requires high-dimensional controls. This paper considers estimation and inference for heterogeneous counterfactual effects with high-dimensional data. We propose a novel doubly robust score for double/debiased estimation and inference for the unconditional quantile regression (Firpo, Fortin, and Lemieux, 2009) as a measure of heterogeneous counterfactual marginal effects. We propose a multiplier bootstrap inference and develop asymptotic theories to guarantee that the bootstrap works. Simulation studies support our theories. Applying the proposed method to Job Corps survey data, we find that i) the marginal effects of counterfactually extending the duration of the exposure to the Job Corps program are globally positive across quantiles robustly regardless of definitions of the treatment and outcome variables and that ii) these counterfactual effects are larger for higher potential earners than lower potential earners robustly regardless of whether we define the outcome as the level or its logarithm.

Keywords: counterfactual analysis, double/debiased machine learning, doubly robust score

*The first arXiv date: July 27, 2020. We thank Layal Lettry for useful advice on the Job Corps survey data. We would like to thank Colin Cameron for very helpful comments. Yichong Zhang acknowledges the financial support from Singapore Ministry of Education Tier 2 grant under grant MOE2018-T2-2-169 and the Lee Kong Chian fellowship. The usual disclaimer applies.

†Y. Sasaki: Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819. Email: yuya.sasaki@vanderbilt.edu

‡T. Ura: Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616. Email: takura@ucdavis.edu

§Y. Zhang: School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore. Email: yczhang@smu.edu.sg

1 Introduction

Analysis of an outcome response to a counterfactual shift in the covariate distribution is of interest in policy studies. Such a counterfactual analysis requires accounting for the Oaxaca-Blinder decomposition of heterogeneous outcome distributions into structural heterogeneity ($F_{Y|X}$) and distributional heterogeneity (F_X); see Fortin, Lemieux, and Firpo (2011) for a review. To mitigate confoundedness in causal effects for conducting a credible counterfactual analysis, it is crucial for researchers to control a structure ($F_{Y|X}$) with rich information about X while applying a counterfactual shift in the distribution of X. According to Athey, Imbens, and Wager (2016), "[t]he unconfoundedness assumption is often more plausible if a large number of pre-treatment variables are included in the analysis." In this light, a researcher ideally wants to use high-dimensional X in data.

Motivated by this feature of causal inference and the recently increasing availability of high-dimensional data, we develop a novel theory and method of estimation and inference for heterogeneous counterfactual effects with high-dimensional controls based on machine learning techniques. The existing literature features a number of alternative approaches and frameworks of counterfactual analysis. Among others, we focus on the unconditional quantile partial effect (UQPE; Firpo, Fortin, and Lemieux, 2009) in the unconditional quantile regression based on the re-centered influence function (RIF) of Firpo et al. (2009) for two reasons: (i) its advantage of providing "a simple way of performing detailed decompositions" (Fortin et al., 2011, p. 76) and (ii) its popularity.[1] This parameter measures the marginal effect of counterfactually shifting the distribution of a coordinate of X on population quantiles of an outcome.[2]

The UQPE is expressed in the potential outcome framework as follows. Let

$$Y = Y(X),$$

where Y is the observed outcome and $Y(x)$ is the potential outcome under $X = x$. The UQPE with respect to the first coordinate, X_1, of X is defined by

$$UQPE(\tau) = \frac{\partial}{\partial \varepsilon} Q_\tau(Y(X_1 + \varepsilon, X_{-1}))\bigg|_{\varepsilon=0}, \tag{1}$$

where $X = (X_1, X_{-1})$ and $Q_\tau(\cdot)$ is the τ-th quantile operator. (For notational simplicity, we will focus on the change from X to $(X_1 + \varepsilon, X_{-1})$ throughout this paper, while our analysis can be generalized to the change in any fixed direction.) The UQPE

[1] As of September 30, 2020, Firpo et al. (2009) have attracted 1767 Google Scholar citations.

[2] In addition to the unconditional quantile regression framework of Firpo et al. (2009), which we focus on in this paper, we remark that there is another important branch of the literature on counterfactual inference under fixed distributional changes. See, for example, Machado and Mata (2005); Melly (2005); Rothe (2010); Chernozhukov, Fernández-Val, and Melly (2013); and Hsu, Lai, and Lieli (2020). See also Frölich and Melly (2013) for unconditional quantile treatment effects under endogeneity.

measures the change in the outcome quantile when the distribution of X changes infinitesimally in the direction of the first coordinate.

While the RIF regression approach is indeed simpler to implement than alternative methods of counterfactual analysis (Fortin et al., 2011), an estimation of the UQPE still requires a three-step procedure. The first step is an estimation of unconditional quantiles. The second step implements the RIF regression. The third step integrates the RIF regression estimates to in turn estimate the UQPE. Firpo et al. (2009) provide an estimation procedure for the case of low-dimensional data. If we accommodate high-dimensional controls with the aforementioned motivation, then the second step will require some machine learning of the high-dimensional RIF regression, and hence the traditional techniques to incorporate estimation errors of the second step into the third step no longer apply. To overcome this challenge, we construct a novel doubly robust score for estimation of the UQPE. The key insight for the construction is the identification result in Firpo et al. (2009, p. 958) that the UQPE has the same structure as the average derivative estimator, whose influence function in the presence of nonparametric preliminary estimation has been well studied in the existing literature (e.g., Newey, 1994; see also Newey and Ruud, 2005). With this doubly robust score, we obtain a Z-estimation criterion with robustness against perturbations in functional nuisance parameters as in Belloni, Chernozhukov, and Kato (2014) and Belloni, Chernozhukov, Chetverikov, and Wei (2018a), and can thereby use the double/debiased machine learning approach (e.g., Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, an 2017; Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018a; Chernozhukov, Escanciano, Ichimura, Newey, and Robins, 2018b), which allows one to obtain the asymptotic distribution of a UQPE estimator, independently of the second-step estimation as far as it satisfies only mild convergence rate conditions, as is the case with major machine learners such as Lasso, ridge, elastic nets, and neural networks, among others.

To accommodate a wide range of machine learners, possibly including those that may be developed in the future, we first present a general method with the main theory based on high-level assumptions. In addition, to provide a readily applicable method for practitioners, we also present a specific method that is easy to implement and accompanying lower-level assumptions that are easy to interpret. Namely, focusing on Lasso estimators of the nuisance parameters, we provide a concrete estimation procedure, present primitive conditions to verify our high-level assumptions, prescribe tuning parameter choice rules, and derive required convergence rate properties following Belloni, Chernozhukov, Fernández-Val, and Hansen (2017). While the Lasso enjoys restricted entropy of function spaces via variable selection, other machine learners may have more complex or even unknown functional forms, implying that the entropies of the classes of functions they belong to can be larger than what is required by our high-level assumptions. To accommodate general machine learners, we thus propose a method of kernel convolution that, combined with the cross-fitting technique, weakens the required entropy condition.

This paper is related to a series of recent papers that propose the estimation of treatment effect parameters via the nonparametric regressions of continuous variables. Abrevaya, Hsu, and Lieli (2015), Lee, Okui, and Whang (2017), and Fan, Hsu, Lieli, and Zhang (2019) consider estimation and inference for average treatment effects of a binary treatment conditionally on possibly continuous covariates. The latter two references also develop doubly robust estimators, as in this paper. Furthermore, Kennedy, Ma, McHugh, and Smal (2017), Semenova and Chernozhukov (2017), Su, Ura, and Zhang (2019), Zimmert and Lechner (2019), and Colangelo and Lee (2020) consider an expectation or quantiles of $Y(x_1 + \varepsilon, X_{-1})$ given a value x_1 of X_1. None of these existing results directly apply to the UQPE, since the UQPE does *not* restrict the subpopulation of interest to those with fixed x_1; see (1). A naïve integration of these existing estimators with respect to the conditional distribution of X_1 given $Y = Q_\tau(Y)$ will not lead to a doubly robust estimation of the UQPE either. Therefore, it requires to develop a novel method, as we have done in this paper.

Prior to this work, the use of doubly robust or locally robust methods for causal inference has been considered by an extensive body of literature including, but not limited to, Imbens (1992), Robins, Mark, and Newey (1992), Robins and Rotnitzky (1995), Hahn (1998), Van der Laan and Robins (2003), Hirano, Imbens, and Ridder (2003), Van Der Laan and Rubin (2006), Firpo (2007), Tsiatis (2007), Wooldridge (2007), Chen, Hong, and Tarozzi (2008), Graham (2011), Van der Laan and Rose (2011), Graham, Pinto, and Egel (2012), Farrell (2015), Graham, Pinto, and Egel (2016), Belloni et al. (2017), Chernozhukov et al. (2017), Kennedy et al. (2017), Lee et al. (2017), Robins, Li, Mukherjee, Tc (2017), Semenova and Chernozhukov (2017), Słoczyński and Wooldridge (2018), Wager and Athey (2018), Sant'Anna and Zhao (2018), Fan et al. (2019), Rothe and Firpo (2019), Su et al. (2019), Zimmert and Lechner (2019), Colangelo and Lee (2020), and Sasaki and Ura (2020), among many others. More recent papers in this list are motivated similarly to this paper and use the double robustness or local robustness to accommodate machine learning of high-dimensional preliminary functions. This vast literature investigates various causal parameters in a variety of model frameworks, but to our best knowledge, none has investigated the UQPE.

This paper also contributes to the literature on high-dimensional econometrics and machine learning. Namely, we follow and extend the existing literature in a few significant directions. First, we follow Chernozhukov et al. (2018a,b) and apply a cross-fitting technique for estimation. We complement their results by considering the uniformly valid inference. Second, this paper is related to Belloni et al. (2014) and Belloni et al. (2018a), which consider uniformly valid inference in Z-estimation with high-dimensional data. We complement their results by providing high-level assumptions for the cross-fitting estimation.[3]

The paper is organized as follows. Section 2 proposes a doubly robust score for the UQPE and discusses a multiplier bootstrap method of inference. We present an

[3]Belloni et al. (2018a) mention using a cross-splitting technique to relax their assumptions but do not provide formal results.

asymptotic theory for the estimator and its multiplier bootstrap counterpart. Section 3 introduces preliminary first-stage estimators. Section 4 presents Monte Carlo simulation studies. Section 5 presents an application to the Job Corps data. Section 6 concludes. The appendix collects all the proofs and auxiliary results.

2 Doubly Robust Score and Bootstrap Inference for $UQPE(\tau)$

In this section, we develop a new score for a doubly robust estimation of the UQPE. We then present a uniform asymptotic linear representation for this doubly robust UQPE estimator and its multiplier bootstrap counterpart. While we focus on a general framework in this section, specific estimation procedures with lower-level primitive conditions will follow in Section 3.1. It is worthwhile to mention here that our analysis allows the dimensionality of X to depend on the sample size N and diverge as $N \to \infty$. This feature is important as the unconfoundedness, the key assumption for the identification with observational data, is more plausible with a larger number of conditioning variables, as emphasized at the beginning of this paper.

Following Firpo et al. (2009), we can rewrite our parameter of interest, defined in (1), as a function of identifiable objects. Namely, if $X \equiv (X_1, X_{-1})$ and $\{Y(x_1, x_{-1}) : (x_1, x_{-1})\}$ are independent, then

$$UQPE(\tau) = -\frac{\theta(\tau)}{f_Y(q_\tau)}, \tag{2}$$

where q_τ is the τ-th quantile of Y and

$$\theta(\tau) = \int \frac{\partial F_{Y|X=x}(q_\tau)}{\partial x_1} dF_X(x). \tag{3}$$

Although this equation is shown in Firpo et al. (2009, Corollary 1), we also present its proof in Appendix B for the sake of completeness and for the convenience of readers.

2.1 Doubly Robust Score

We could estimate $\theta(\tau)$ based on the equality (3) and some estimator for $F_{Y|X}(\cdot)$. When X is high-dimensional, this direct estimation of $\theta(\tau)$ can result in a large bias, a large variance, or both. Instead, we propose constructing an estimator for $\theta(\tau)$ based on another representation:

$$\begin{aligned}\theta(\tau) &= \int \frac{\partial F_{Y|X=x}(q_\tau)}{\partial x_1} dF_X(x) - \int \omega(x)(1\{y \le q_\tau\} - m_0(x, q_\tau)) dF_{Y,X}(y, x) \\ &= \int m_1(x, q_\tau) - \omega(x)(1\{y \le q_\tau\} - m_0(x, q_\tau)) dF_{Y,X}(y, x),\end{aligned} \tag{4}$$

where $\omega(x) = \frac{\partial}{\partial x_1} \log f_{X_1|X_{-1}=x_{-1}}(x_1)$, $m_0(x,q) = F_{Y|X=x}(q)$ and $m_1(x,q) = \partial m_0(x,q)/\partial x_1$. This representation comes from the influence adjustment term for the average derivative estimator (Newey, 1994, p.1369). Namely, in the moment (4), the term $\int \omega(x)(1\{y \leq q_\tau\} - m_0(x,q_\tau))dF_{Y,X}(y,x)$ adjusts the estimation error from nonparametric preliminary estimation.

The advantage of (4) over (3) is that the moment (4) is doubly robust in the sense that

$$\theta(\tau) = \int (\tilde{m}_1(x,q_\tau) - \omega(x)(1\{y \leq q_\tau\} - \tilde{m}_0(x,q_\tau)))\, dF_{Y,X}(y,x) \tag{5}$$

and

$$\theta(\tau) = \int (m_1(x,q_\tau) - \tilde{\omega}(x)(1\{y \leq q_\tau\} - m_0(x,q_\tau)))\, dF_{Y,X}(y,x) \tag{6}$$

hold for a set of values that the high-dimensional nuisance parameters ($\tilde{\omega}(x)$, $\tilde{m}_0(x,q)$, $\tilde{m}_1(x,q)$) take as far as some regularity conditions are satisfied. A precise statement and its proof are found in Appendix A. Note that $(\tilde{\omega}(x), \tilde{m}_0(x,q), \tilde{m}_1(x,q))$ in (5) and (6) can be different from the true value $(\omega(x), m_0(x,q), m_1(x,q))$. Thus, the estimation error for $(\omega(x), m_0(x,q), m_1(x,q))$ does not have a first-order asymptotic influence on the estimation error for θ.

Based on the moment condition (4), we propose to estimate $\theta(\tau)$ by a cross-fitting approach (Chernozhukov et al., 2018a, Definition 3.2). We split the sample of size N into a random partition $\{I_1, \cdots, I_L\}$ of approximately equal size. For simplicity, let $|I_l| = n$ for every l so that $N = nL$. In this section, we assume that, for every index $l \in \{1, ..., L\}$ of fold, we can construct an estimator $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$ by using all the observations except those in I_l. Section 3.1 provides a concrete example of $(\hat{\omega}(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$ based on the Lasso regularization. Letting $\hat{q}_\tau$ be the full sample τ-th empirical quantile of Y, we estimate $\theta(\tau)$ by

$$\hat{\theta}(\tau) = \frac{1}{L}\sum_{l=1}^{L} \frac{1}{n}\sum_{i \in I_l} \left[\hat{m}_{1,l}(X_i,\hat{q}_\tau) - \hat{\omega}_l(X_i)(1\{Y_i \leq \hat{q}_\tau\} - \hat{m}_{0,l}(X_i,\hat{q}_\tau))\right]. \tag{7}$$

With this estimator for $\theta(\tau)$, our proposed estimator for $UQPE(\tau)$ is

$$\widehat{UQPE}(\tau) = -\frac{\hat{\theta}(\tau)}{\hat{f}_Y(\hat{q}_\tau)},$$

where

$$\hat{f}_Y(y) = \sum_{i=1}^{N} \frac{1}{Nh_1} K_1\left(\frac{Y_i - y}{h_1}\right)$$

for some kernel function $K_1(\cdot)$ and a bandwidth parameter h_1.

2.2 Bootstrap Inference

For an inference about $UQPE(\tau)$, we propose the multiplier bootstrap without recalculating the preliminary estimators $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$ in each bootstrap iteration. Using independent standard normal random variables $\{\eta_i\}_{i=1}^N$ that are independent of the data, we compute the bootstrap estimator $\widehat{UQPE}^*(\tau)$ in the following steps. The bootstrap estimator for q_τ is $\hat{q}_\tau^*$ defined by the h_N^*-th order statistic of Y_i, where h_N^* is the integer part of $1 + \sum_{i=1}^N (\tau + \eta_i(\tau - \mathbf{1}\{Y_i \leq \hat{q}_\tau\}))$. The bootstrap estimator for f_Y is

$$\hat{f}_Y^*(y) = \sum_{i=1}^N \frac{(\eta_i + 1)}{\sum_{i=1}^N (\eta_i + 1)} \frac{1}{h_1} K_1 \left(\frac{Y_i - y}{h_1} \right),$$

and the bootstrap estimator for $\theta(\tau)$ is

$$\hat{\theta}^*(\tau) = \frac{1}{L} \sum_{l=1}^L \frac{1}{\sum_{i \in I_l} (\eta_i + 1)} \sum_{i \in I_l} (\eta_i + 1) \left[\hat{m}_{1,l}(X_i, \hat{q}_\tau^*) - \hat{\omega}_l(X_i)(1\{Y_i \leq \hat{q}_\tau^*\} - \hat{m}_{0,l}(X_i, \hat{q}_\tau^*)) \right].$$

With these components, the bootstrap estimator $\widehat{UQPE}^*(\tau)$ is given by

$$\widehat{UQPE}^*(\tau) = -\frac{\hat{\theta}^*(\tau)}{\hat{f}_Y^*(\hat{q}_\tau^*)}.$$

We can use the above multiplier bootstrap method to conduct various types of inference. For example, we can construct a pointwise confidence interval for $UQPE(\tau)$. Denote by $CI(\tau)$ the interval whose lower (resp. upper) bound is the $\alpha/2$ (resp. $(1-\alpha/2)$) quantile of $\widehat{UQPE}^*(\tau)$ conditional on the data.

Another example of inference is a confidence band for $\{UQPE(\tau) : \tau \in \Upsilon\}$ for some closed interval $\Upsilon \subset (0,1)$. Let $c_\Upsilon(1-\alpha)$ denote the $(1-\alpha)$ quantile of $\sup_{\tau \in \Upsilon} \left| (\widehat{UQPE}^*(\tau) - \widehat{UQPE}(\tau)) / \hat{\sigma}(\tau) \right|$ conditional on the data, where $\hat{\sigma}(\tau)$ is some estimator of the standard error of $\widehat{UQPE}(\tau)$ for $\tau \in \Upsilon$. Let CB_Υ denote the band on Υ whose lower and upper bounds at $\tau \in \Upsilon$ are $\widehat{UQPE}(\tau) \pm \hat{\sigma}(\tau) c_\Upsilon(1-\alpha)$.

2.3 Asymptotic Theory

In this section, we investigate the asymptotic properties of the estimator $\widehat{UQPE}(\tau)$ and the bootstrap estimator $\widehat{UQPE}^*(\tau)$ introduced in the previous two subsections. As in Section 2.2, the uniformity over τ is relevant to applications (e.g., analysis of heterogeneous counterfactual effects across τ), and therefore, in this section, we aim to control the residuals for the linear expansion uniformly over $\tau \in \Upsilon$ for some closed interval $\Upsilon \subset (0,1)$. Let $\mathcal{Q} = \{q_\tau : \tau \in \Upsilon\}$, and let $\mathcal{Q}^\varepsilon$ denote the ε enlargement of $\mathcal{Q}$.

Assumption 1.

1. *For every* $\tau \in \Upsilon$*,* $F_{Y(X_1+\varepsilon, X_{-1})}(q)$ *is differentiable with respect to* ε *in a neighborhood of zero for every* q *in a neighborhood of* q_τ*, and* $Q_\tau(Y(X_1+\varepsilon, X_{-1}))$ *is well defined and is differentiable with respect to* ε *in a neighborhood of zero.*

2. X_1 *and* $\{Y(x_1, X_{-1}) : x_1\}$ *are conditionally independent given* X_{-1}*.*

3. $\int \left[\sup_{q\in\mathcal{Q}} |m_1(x, q_\tau)|\right]^{2+\delta} dF_X(x)$ *and* $\int |\omega(x)|^{2+\delta} dF_X(x)$ *are finite.*

4. *For every* x_{-1} *in the support of* X_{-1}*, the conditional distribution of* X_1 *given* $X_{-1} = x_{-1}$ *has a probability density function, denoted by* $f_{X_1|X_{-1}}$*, which is continuously differentiable everywhere and is zero on the boundary of the support of the conditional distribution of* X_1*.*

5. $m_1(x, q)$ *and* $m_0(x, q)$ *are differentiable with respect to* q *for* $q \in \mathcal{Q}$*, and the derivatives are bounded in absolute value uniformly over* $x \in SuppX$ *and* $q \in \mathcal{Q}$*.*

6. $f_Y(y)$ *is three times differentiable on* $\mathcal{Q}^\varepsilon$ *with all the derivatives uniformly bounded.* $f_Y(q_\tau) > 0$ *for every* $\tau \in \Upsilon$*.*

This assumption is on the model primitives of this paper. Assumptions 1.1 and 1.3–1.6 impose regularity in terms of the smoothness of various functions representing features of the data. Assumption 1.2 imposes the unconfoundedness, which is a key assumption in causal inference. While this assumption may be implausible with traditional low-dimensional models, it tends to be more plausible as X_{-1} contains a richer set of controls (i.e., as the dimension of X_{-1} increases). As emphasized in the introduction, this is the main motivation for our investigation of extended models allowing for high dimensions of X_{-1} in this paper.

Assumption 2. *For every index* $l \in \{1, ..., L\}$ *of folds, there exist sequences* ν_N*,* A_N*,* π_N *such that the following conditions hold with probability approaching one:*

$$\sup_Q N(\{\hat{m}_j(x,q) : q \in \mathcal{Q}\}, e_Q, \varepsilon ||G_l^{(j)}||_{Q,2}) \lesssim \left(\frac{A_N}{\varepsilon}\right)^{\nu_N} \quad \text{for every } \varepsilon \in (0,1], \tag{8}$$

$$\sup_{q\in\mathcal{Q}} \int |\hat{m}_{1,l}(x,q) - m_1(x,q)|^2 \, dF_X(x) = O_P(\pi_N^2), \tag{9}$$

$$\int |\hat{\omega}_l(x) - \omega(x)|^2 \, dF_X(x) = O_P(\pi_N^2), \tag{10}$$

$$\sup_{q\in\mathcal{Q}} \int ||\hat{\omega}_l(x)\hat{m}_{0,l}(x,q) - \omega(x)m_0(x,q)|^2 \, dF_X(x) = O_P(\pi_N^2), \tag{11}$$

$$\int \left[\sup_{q\in\mathcal{Q}} |\hat{m}_{1,l}(x,q)|\right]^{2+\delta} dF_X(x) = O_P(1), \tag{12}$$

$$\int \left[\sup_{q\in\mathcal{Q}} |\hat{\omega}_l(x)(1+|\hat{m}_{0,l}(x,q)|)| \right]^{2+\delta} dF_X(x) = O_P(1), \tag{13}$$

$$\sup_{q\in\mathcal{Q}} \left| \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q) \right) dF_X(x) \right| = o_P(N^{-1/2}), \tag{14}$$

$$\sup_{q\in\mathcal{Q}} \left| \int (\hat{\omega}_l(x)-\omega(x))(\hat{m}_{0,l}(x,q)-m_0(x,q)) dF_X(x) \right| = o_P(N^{-1/2}), \tag{15}$$

$$\pi_N^2 \nu_N \log(A_N/\pi_N) = o(1), \qquad \text{and} \qquad \nu_N N^{-\frac{\delta}{4+2\delta}} \log(A_N/\pi_N) = o(1), \tag{16}$$

where, in (8)*,* $N(\cdot)$ *is the covering number,* $G_l^{(j)}$ *is the envelope for* $\{\hat{m}_j(x,q) : q \in \mathcal{Q}\}$*,* $e_Q(f,g) = ||f-g||_{Q,2}$ *for the probability measure* Q*, and the supremum is taken over all finitely discrete probability measures.*

Several comments are in order. First, this assumption consists of a list of high-level conditions that should be satisfied by the preliminary estimator $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q))$. While we state these high-level conditions here for the sake of accommodating a general class of preliminary estimators, Section 3.1 demonstrates that these conditions are satisfied in particular for a concrete estimator that we propose. Second, (8) is the entropy condition for the classes of functions $\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}\}$. We require this condition because (1) we want to derive the linear expansion for $\hat{\theta}(\tau)$ that is uniform in τ and (2) $\hat{m}_{j,l}(x, \hat{q}_\tau)$ has the estimated $\hat{q}_\tau$ inside for $j = 0, 1$. We will directly verify (8) and (16) for the Lasso estimator in Section 3.1 and general machine learning estimators via a kernel convolution technique in Section 3.2. Third, it is worth mentioning that the $o_P(N^{-1/2})$-consistency conditions in (14) and (15) are feasible. The term (14) is zero if we construct $\hat{m}_1(x,q)$ by $\hat{m}_1(x,q) = \frac{\partial}{\partial x_1}\hat{m}_0(x,q)$. The term (15) is $o_P(N^{-1/2})$ and therefore negligible, as long as $\hat{\omega}_l(x)$ and $\hat{m}_{0,l}(x,q)$ are $o_P(N^{-1/4})$-consistent (in the L^2 norm with respect to x). These $o_P(N^{-1/4})$ conditions are achievable even if X is high-dimensional. In addition, the product structure in (15) allows for the trade-off between how fast $\hat{\omega}_l(x)$ converges and how fast $\hat{m}_{0,l}(x,q)$ converges.

Assumption 3. *1.* $K_1(\cdot)$ *is a second-order symmetric kernel function with a compact support. 2.* $h_1 = c_1 N^{-H}$ *for some positive constant* c_1 *and some* $1/2 > H \geq 1/5$*.*

Assumption 3.1 states requirements for the kernel function. Assumption 3.2 describes admissible rates at which the bandwidth parameter tends to zero. The next theorem presents asymptotic expansions for $\widehat{UQPE}(\tau)$ and $\widehat{UQPE}^*(\tau)$ under the above assumptions.

Theorem 1. *If Assumptions 1–3 hold, then*

$$\widehat{UQPE}(\tau) - UQPE(\tau) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IF}_i(\tau) + \frac{\theta(\tau) f_Y^{(2)}(q_\tau)(\int u^2 K_1(u)du)h_1^2}{2 f_Y^2(q_\tau)} + R(\tau) \tag{17}$$

$$\widehat{UQPE}^*(\tau) - \widehat{UQPE}(\tau) = \frac{1}{N}\sum_{i=1}^{N} \eta_i \cdot \mathrm{IF}_i(\tau) + R^*(\tau), \tag{18}$$

where the residuals are $o_P(N^{-1/2})$ uniformly in τ, i.e.,

$$\sup_{\tau\in\Upsilon} \max\{|R(\tau)|, |R^*(\tau)|\} = o_P(N^{-1/2}),$$

the influence function is

$$\mathrm{IF}_i(\tau) = \frac{\theta(\tau)}{f_Y^2(q_\tau)h_1} K_1\left(\frac{Y_i - q_\tau}{h_1}\right) + \frac{\theta(\tau) f_Y^{(1)}(q_\tau)}{f_Y^3(q_\tau)}(\tau - \mathbf{1}\{Y_i \le q_\tau\})$$
$$- \frac{m_1(X_i, q_\tau) - \theta(\tau) - \omega(X_i)(1\{Y_i \le q_\tau\} - m_0(X_i, q_\tau)) + \frac{\partial}{\partial q} m_1(X_i, q_\tau)\frac{\tau - 1\{Y_i \le q_\tau\}}{f_Y(q_\tau)}}{f_Y(q_\tau)},$$

and $f_Y^{(1)}(\cdot)$ and $f_Y^{(2)}(\cdot)$ are the first and second derivatives of $f_Y(\cdot)$, respectively.

This theorem first establishes the uniform influence function representation in (17) for the estimator $\widehat{UQPE}$ with the influence function $\mathrm{IF}_i(\tau)$. Second, it also establishes the multiplier counterpart in (18). These two results together imply that we can simulate the limit process of $r_N(\widehat{UQPE}(\cdot) - UQPE(\cdot))$ by the process of $r_N \cdot \frac{1}{N}\sum_{i=1}^{N} \eta_i \cdot \mathrm{IF}_i(\cdot)$ conditionally on the data. The leading term of the score function is $\frac{\theta(\tau)}{f_Y^2(q_\tau)h_1} K_1\left(\frac{Y_i - q_\tau}{h_1}\right)$, which has a slower convergence rate than the other terms. However, this does not imply that the doubly robust method is unnecessary. In fact, we only require the convergence rate of the nuisance estimators to be $o(N^{-1/4})$, which is still slower than the nonparametric convergence rate of $\widehat{UQPE}(\tau)$. Without the help of the doubly robust method, the estimation error of the nuisance parameters would dominate the leading term in the score function. In addition, Theorem 1 shows the bootstrap estimator can mimic not only the leading term but also the other faster convergent terms in the score function. This indicates that the bootstrap method can provide a higher-order approximation of the variance of the original estimator. That said, a comparison between (17) and (18) makes it clear that the bootstrap cannot approximate the bias term in the kernel estimation. In practice, we recommend an undersmoothing, that is, taking the bandwidth h smaller than the MSE-optimal rate as in Assumption 3 so that the bias is asymptotically negligible.

The following corollary summarizes the validity for the bootstrap inference.

Corollary 1. *Suppose Assumptions 1–3 hold and* $\sqrt{Nh_1} = o(h_1^{-2})$*, then*

$$\mathbb{P}(UQPE(\tau) \in CI(\tau)) \to 1 - \alpha.$$

If, in addition, $\sup_{\tau \in \Upsilon} \left| \sqrt{Nh_1}\hat{\sigma}(\tau) - \sqrt{h_1 Var(\mathrm{IF}_i(\tau))} \right| = o_P(\log^{-1/2}(N))$ *holds and* $h_1 Var(\mathrm{IF}_i(\tau))$ *is bounded away from zero, then*

$$\mathbb{P}(\{UQPE(\tau) : \tau \in \Upsilon\} \in CB_{\Upsilon}) \to 1 - \alpha.$$

Corollary 1 is a direct consequence of the linear expansions in Theorem 1 and the strong approximation theory developed by Chernozhukov, Chetverikov, and Kato (2014a,b). Its proof is omitted for brevity. To compute $\hat{\sigma}(\tau)$, we can use either the plug-in method or the bootstrap method. For these methods, the convergence rate of $\sqrt{Nh_1}\hat{\sigma}(\tau)$ is polynomial in N, which implies $o_P(\log^{-1/2}(N))$.

3 Preliminary First-Stage Estimators

The general theory presented in Section 2 presumes general machine learners for preliminary first-stage estimation and therefore uses high-level conditions stated in Assumption 2. In the current section, we present two additional results to supplement this general framework. First in Section 3.1, to provide empirical practitioners with a readily applicable method, we focus on the Lasso preliminary first-stage estimation, propose a concrete estimation and inference procedure that is easy to implement, and present lower-level sufficient conditions for Assumption 2 that are easy to interpret in the context of this specific estimation procedure. Because the Lasso regularization achieves variable selection, the entropy of the function class to which the estimators belong is small enough to directly satisfy our general assumptions. Second, in Section 3.2, to accommodate a general class of machine learners as well as the Lasso, we propose a method of kernel deconvolution so that the entropy of the function classes to which the convoluted machine learning estimates belong is reduced so that our general assumption can be satisfied for these general machine learners.

3.1 Lasso Preliminary Estimator

In this section, we use the Lasso regularization to construct $(\hat{\omega}_l(x), \hat{m}_{0,l}(x,q))$ for every index $l \in \{1, ..., L\}$ of fold and derive a low-level sufficient condition for Assumption 2 in Section 2.3. In this paper, we use $\hat{m}_{1,l}(x,q)$ defined by

$$\hat{m}_{1,l}(x,q) = \frac{\partial}{\partial x_1} \hat{m}_{0,l}(x,q),$$

which immediately implies (14) in Assumption 2.

To estimate $\omega(X)$, we consider the location scale model

$$X_1 = \mu(X_{-1}) + \sigma(X_{-1})U \quad \text{where } U \sim N(0,1). \tag{19}$$

By the shape of the normal distribution,

$$\omega(X) = \frac{-(X_1 - \mu(X_{-1}))}{\sigma^2(X_{-1})} \tag{20}$$

and

$$\mu(X_{-1}) = \frac{Q_{0.25}(X_1|X_{-1}) + Q_{0.75}(X_1|X_{-1})}{2}, \ \sigma(X_{-1}) = \frac{Q_{0.25}(X_1|X_{-1}) - Q_{0.75}(X_1|X_{-1})}{z_{0.25} - z_{0.75}},$$

where $Q_\tau(X_1|X_{-1})$ is the conditional τ-th quantile of X_1 given X_{-1} and z_τ is the τ-th standard normal critical value. It is possible to generalize the model (19) to

$$g(X_1) = \mu(X_{-1}) + \sigma(X_{-1})U,$$

where $g(\cdot)$ is some known transformation and U may follow a non-Gaussian (but known) distribution.

We now discuss the detailed estimation procedure for $\omega(x)$ based on (20).[4] Consider the approximately sparse linear model for $Q_\tau(X_1|X_{-1})$:

$$Q_\tau(X_1|X_{-1}) = h(X_{-1})^T\gamma_\tau + r_Q(X_{-1}, \tau),$$

where $h(X_{-1})$ is a p_h-dimensional vector and Assumption 4 (to be stated below) specifies the conditions for the sparsity and the approximation error $r_Q(x_{-1}, \tau)$. We estimate γ_τ by the Lasso penalized quantile regression

$$\hat{\gamma}_{\tau,l} = \arg\min_{\gamma} \frac{1}{n(L-1)} \sum_{i \in I_l^c} \rho_\tau(X_{1,i} - h(X_{-1,i})^T\gamma) + \frac{\lambda}{n(L-1)} ||\Xi_\tau \gamma||_1,$$

where $\rho_\tau(u) = u(\tau - 1\{u \leq \tau\})$ is the check function and Ξ_τ is the $p_h \times p_h$ diagonal matrix whose jth diagonal entry is $\sqrt{\frac{1}{n(L-1)} \sum_{i \in I_l^c} \tau(1-\tau)h_j^2(X_{-1,i})}$. The regularization parameter follows Belloni et al. (2017, p. 261):

$$\lambda = 1.1\Phi^{-1}(1 - (0.1/\log(N))/(p_h \vee N))(n(L-1))^{1/2}.$$

[4] It is possible to estimate the conditional CDF of X_1 given X_{-1} via a logistic Lasso regression proposed by Belloni et al. (2017) and then use a numerical derivative to estimate $\omega(x) = \frac{\partial}{\partial x_1} \log f_{X_1|X_{-1}=x_{-1}}(x_1)$. We refer interested readers to Belloni, Chernozhukov, and Kato (2018b) for more detail. Both methods involve parametric assumptions on the error term. Although the numerical derivative method does not assume the location-scale model, it requires more tuning parameters than our method and needs to estimate $\omega(x)$ for each x_1 separately. As introduced below, our method is easier to implement, however, because it can construct an estimator for the function $x \mapsto \omega(x)$ by estimating the $(2 \cdot p_h)$-dimensional parameters for $(\mu(x_{-1}), \sigma(x_{-1}))$.

Based on $\hat{\gamma}_{\tau,l}$, we can estimate $Q_\tau(X_1|X_{-1})$ by

$$\hat{Q}_{\tau,l}(X_1|X_{-1}) = h(X_{-1})^T\hat{\gamma}_{\tau,l}.$$

Now we can estimate $\omega(X)$ by

$$\hat{\omega}_l(X) = \frac{-(X_1 - \hat{\mu}_l(X_{-1}))}{\hat{\sigma}_l^2(X_{-1})},$$

where

$$\hat{\mu}_l(X_{-1}) = \frac{\hat{Q}_{0.25,l}(X_1|X_{-1}) + \hat{Q}_{0.75,l}(X_1|X_{-1})}{2},\ \hat{\sigma}_l(X_{-1}) = \frac{\hat{Q}_{0.25,l}(X_1|X_{-1}) - \hat{Q}_{0.75,l}(X_1|X_{-1})}{z_{0.25} - z_{0.75}}.$$

We estimate $m_0(x,q)$ by the logistic Lasso regression using data in I_l^c. With the standard logistic CDF denoted by Λ, we consider the approximately sparse logistic regression model for $m_0(x,q)$:

$$m_0(X,q) = \Lambda(b(X)^T\beta_q) + r_m(X,q),$$

where $b(X)$ is a p_b-dimensional vector and Assumption 4 (to be stated below) specifies the conditions for the sparsity and the approximation error $r_m(x,q)$. We estimate β_q by the Lasso penalized logistic regression

$$\tilde{\beta}_{q,l} = \arg\min_\beta \frac{1}{n(L-1)}\sum_{i\in I_l^c} M(1\{Y_i \leq q\}, b(X_i);\beta) + \frac{\lambda}{n(L-1)}||\Psi_q\beta||_1, \qquad (21)$$

where $M(\cdot)$ is the logistic likelihood and Ψ_q is a diagonal matrix with penalty loadings defined in the next paragraph. The regularization parameter is

$$\lambda = 1.1\Phi^{-1}(1-(0.1/\log(N))/(p_b \vee N))(n(L-1))^{1/2}.$$

We recommend using the post-Lasso estimator for β_q defined by

$$\hat{\beta}_{q,l} = \arg\min_{\beta\in\mathbb{R}^p:\mathrm{Supp}(\beta)\subset\mathrm{Supp}(\tilde{\beta}_{q,l})\cup S_1} \frac{1}{n(L-1)}\sum_{i\in I_l^c} M(1\{Y_i\leq q\}, b(X_i);\beta),$$

where $S_1 \subset \{1,\ldots,p_b\}$ represents the set of covariates researchers want to include in the post-Lasso regression. In the context of the UQPE with respect to X_1, it is intuitive to include X_1 in the regression. The post-Lasso estimator can do so by setting $1 \in S_1$, whereas the Lasso estimator $\tilde{\beta}_{q,l}$ may exclude X_1 from the regression. With $\hat{\beta}_{q,l}$, we can estimate $\hat{m}_{0,l}(x,q)$ as

$$\hat{m}_{0,l}(x,q) = \Lambda(b(X)^T\hat{\beta}_{q,l}).$$

The penalty loading matrix $\Psi_q = \text{diag}(\psi_{q,1}, \cdots, \psi_{q,p_b})$ in (21) needs to be estimated. Ideally, we would like to use the infeasible penalty loading

$$\bar{\psi}_{q,j} = \sqrt{\frac{1}{n(L-1)} \sum_{i \in I_l^c} \left(1\{Y_i \leq q\} - m_0(X_i, q)\right)^2 b_j^2(X_i)}.$$

Since $m_0(X, q)$ is unknown, Belloni et al. (2017) propose the following iterative algorithm to obtain the feasible version of the loading matrix:

1. We start the algorithm with $\psi_{q,j}^0 = \sqrt{\frac{1}{n(L-1)} \sum_{i \in I_l^c} 1\{Y_i \leq q\} b_j^2(X_i)}$.

2. For $k = 0, \cdots, K-1$ for some fixed positive integer K, we can compute $\tilde{\beta}_q^k$ by (21) with $\tilde{\Psi}_q^k = \text{diag}(\psi_{q,1}^k, \cdots, \psi_{q,p_b}^k)$, and construct

$$\psi_{q,j}^{k+1} = \sqrt{\frac{1}{n(L-1)} \sum_{i \in I_l^c} \left(1\{Y_i \leq q\} - \Lambda(b(X_i)^T \tilde{\beta}_q^k)\right)^2 b_j^2(X_i)}.$$

3. The final penalty loading matrix $\Psi_q^K = \text{diag}(\psi_{q,1}^K, \cdots, \psi_{q,p_b}^K)$ will be used for Ψ_q in (21).

We provide a sufficient condition under which the estimator, $(\hat{\omega}_l(x), \hat{m}_{0,l}(x, q))$, defined above satisfies Assumption 2 in Section 2.3.

Assumption 4.

1. *(Conditional distribution) Suppose* $X_1 | X_{-1} \sim N(\mu(X_{-1}), \sigma^2(X_{-1}))$.

2. *(Boundedness) The following statements hold for positive constants* $\delta, \overline{c}, \underline{c}$*: (i)* $\underline{c} \leq \sigma(x_{-1}) \leq \overline{c}$ *for every* $x_{-1} \in Supp(X_{-1})$*. (ii)* $\underline{c} \leq \mathbb{E} b_j(X)^2 \leq \overline{c}$ *for every* $j = 1, \ldots, p$*. (iii)* $\sup_{x \in Supp(X), q \in \mathcal{Q}^\varepsilon} |m_1(x, q)| \leq \overline{c}$*. (iv)* $\sup_{q \in \mathcal{Q}} ||\frac{\partial}{\partial x_1} b(X)^T \beta_q||_{\mathbb{P}, \infty} \leq \overline{c}$*. (v)* $||\omega(X)||_{\mathbb{P}, 2+\delta} < \overline{c}$.

3. *(Restricted eigenvalue condition) There are positive constants* $\overline{c}, \underline{c}$ *and a sequence* $m_N \to \infty$ *such that, with probability approaching one,*

$$\underline{c} \leq \inf_{\beta \neq 0, ||\beta||_0 \leq m_N} \frac{||b(X)^T \beta||_{\mathbb{P}_n, 2}}{||\beta||_2} \leq \sup_{\beta \neq 0, ||\beta||_0 \leq m_N} \frac{||b(X)^T \beta||_{\mathbb{P}_n, 2}}{||\beta||_2} \leq \overline{c},$$

$$\underline{c} \leq \inf_{\beta \neq 0, ||\beta||_0 \leq m_N} \frac{||\frac{\partial}{\partial x_1} b(X)^T \beta||_{\mathbb{P}_n, 2}}{||\beta||_2} \leq \sup_{\beta \neq 0, ||\beta||_0 \leq m_N} \frac{||\frac{\partial}{\partial x_1} b(X)^T \beta||_{\mathbb{P}_n, 2}}{||\beta||_2} \leq \overline{c},$$

$$\underline{c} \leq \inf_{\gamma \neq 0, ||\gamma||_0 \leq m_N} \frac{||h(X_{-1})^T \gamma||_{\mathbb{P}_n, 2}}{||\gamma||_2} \leq \sup_{\gamma \neq 0, ||\gamma||_0 \leq m_N} \frac{||h(X_{-1})^T \gamma||_{\mathbb{P}_n, 2}}{||\gamma||_2} \leq \overline{c},$$

$$\sup_{\beta\neq 0,||\beta||_0\le m_N}\left|\frac{||\frac{\partial}{\partial x_1}b(X)^T\beta||_{\mathbb{P}_n,2}}{||\frac{\partial}{\partial x_1}b(X)^T\beta||_{\mathbb{P},2}}-1\right|+\sup_{\beta\neq 0,||\beta||_0\le m_N}\left|\frac{||\frac{\partial}{\partial x_1}b(X)^T\beta||_{\mathbb{P}_n,2}}{||\frac{\partial}{\partial x_1}b(X)^T\beta||_{\mathbb{P},2}}-1\right|$$
$$+\sup_{\gamma\neq 0,||\gamma||_0\le m_N}\left|\frac{||h(X_{-1})^T\gamma||_{\mathbb{P}_n,2}}{||h(X_{-1})^T\gamma||_{\mathbb{P},2}}-1\right|=o_P(1),$$

where $||v||_0$ *denotes the the number of nonzero coordinates of vector* v*.*

4. *(Sparsity)* $\max(||\gamma_{0.25}||_0, ||\gamma_{0.75}||_0, \sup_{q\in\mathcal{Q}^\varepsilon}||\beta_q||_0)\le s$ *for a sequence* $s=s_N$ *satisfying* $s=o(m_N)$*,* $\zeta_N^2 s^2\log(p)=o(N)$*,* $\zeta_N^{4/(2+\delta)}s^{(6+2\delta)/(2+\delta)}\log^2(p)=o(N)$*, and* $s\log(p)=o(N^{\delta/(4+\delta)})$*, where* $p=\max(p_h,p_b)$ *and*

$$\zeta_N=\max(||\max_{j=1,\dots,p_b}|b_j(X)|||_{\mathbb{P},\infty}, ||\max_{j=1,\dots,p_b}|\frac{\partial}{\partial x_1}b_j(X)|||_{\mathbb{P},\infty}, ||\max_{j=1,\dots,p_h}|h_j(X_{-1})|||_{\mathbb{P},\infty}).$$

5. *(Approximation Error)*

$$\sup_{q\in\mathcal{Q}^\varepsilon}\left\|\frac{\partial}{\partial x_1}r_m(X,q)\right\|_{\mathbb{P},2}+\|r_Q(X_{-1},0.25)\|_{\mathbb{P},2}+\|r_Q(X_{-1},0.75)\|_{\mathbb{P},2}=O((s\log(p)/N)^{1/2})$$

$$\sup_{q\in\mathcal{Q}^\varepsilon}||\frac{\partial}{\partial x_1}r_m(X,q)||_{\mathbb{P},\infty}+||r_Q(X_{-1},0.25)|||_{\mathbb{P},\infty}+||r_Q(X_{-1},0.75)||_{\mathbb{P},\infty}=O((\log(p)s^2\zeta_N^2/N)^{1/2}).$$

Several remarks are in order. First, Assumption 4.2 is the common regularity condition. Second, Assumptions 4.3 and 4.5 are common in the literature of logistic and quantile regressions with ω_1 penalty. See, for instance, Belloni and Chernozhukov (2011), Belloni et al. (2017), and Belloni et al. (2018b), among others. Third, Assumption 4.4 is due to the fact that $\mathbb{E}\omega^{2+\delta}(X)<\infty$. If all the moments of $\omega(x)$ are finite, then Assumption 4.4 reduces to $\zeta_N^2 s^2\log(p)=o(N)$ and $s^2\log^2(p)=o(N)$ up to some logarithmic factor. Fourth, the quantile regression requires that the conditional quantile is bounded and bounded away from zero. Such condition holds automatically in our setup. Let $\phi(\cdot)$ be the standard normal PDF. Assumptions 4.1 and 4.2 imply that

$$f_{X_1|X_{-1}=x_{-1}}(x_1)=\frac{1}{\sigma(x_{-1})}\phi\left(\frac{x_1-\mu(x_{-1})}{\sigma(x_{-1})}\right)$$

and that

$$\frac{\partial}{\partial x_1}f_{X_1|X_{-1}=x_{-1}}(x_1)=-\frac{x_1-\mu(x_{-1})}{\sigma^3(x_{-1})}\phi\left(\frac{x_1-\mu(x_{-1})}{\sigma(x_{-1})}\right)$$

which are uniformly bounded in absolute value over the support of X. In addition, we note that

$$f_{X_1|X_{-1}=x_{-1}}(Q_\tau(X_1|X_{-1}=x_{-1}))=\frac{1}{\sigma(x_{-1})}\phi(z_\tau).$$

As $\sigma(x_{-1})$ is uniformly bounded away from zero, the previous display implies that $f_{X_1|X_{-1}=x_{-1}}(Q_\tau(X_1|X_{-1}=x_{-1}))$ is bounded away from zero uniformly for all $x_{-1} \in \text{Supp}(X_{-1})$ for $\tau = 0.25$ and 0.75.

With Assumption 4, we can demonstrate that the high-level conditions in Assumption 2 hold for the Lasso regularized estimator proposed in this section. The formal statement is as follows.

Theorem 2. *If Assumptions 1 and 4 hold, then Assumption 2 holds with* $\pi_N = \left(\zeta_N^{4/(\delta+2)} s^{(4+\delta)/(2+\delta)} \log(p)/N\right)^{1/2}$, $\nu_N = s$, *and* $A_N = p$.

3.2 General Preliminary Machine Learning Estimators

In this section, we propose a kernel convolution method to smooth general machine learning estimators $\hat{m}_{j,l}(x,q)$ over q. This convolution benefits the theoretical arguments for the uniform consistency over q because the resulting convolution is Lipschitz continuous, as shown in the proof of Theorem 3. Chernozhukov, Fernández-Val, and Kowalski (2015) use a similar idea. For a generic machine learning estimator $\hat{m}_{0,l}(x,q)$, the entropy of the class of functions $\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}\}$ for $j = 0, 1$ and $l \in \{1, \ldots, L\}$ is usually unknown. This kernel smoothing method provides one way to introduce smoothness to $\hat{m}_{j,l}(x,q)$ over q and thus reduces the entropy of $\{\hat{m}_{j,l}(x,q) : q \in \mathcal{Q}\}$.

Assumption 5.

1. *Both* $m_0(x,q)$ *and* $m_1(x,q)$ *are* $2k$*-th order differentiable with respect to* q*, and all the derivatives are bounded uniformly over* x*.*

2. $K_2(\cdot)$ *is a symmetric function with bounded support,* $\int K_2(u)du = 1$, $\int u^j K_2(u)du = 0$ *for* $j = 1, \cdots, 2k-1$, $sup_u|K_2(u)| < \infty$ *and* $\int u^{2k}|K_2(u)|du < \infty$. $h_2 = c_2 N^{\frac{-1}{2(2k+1)}}$ *for some positive constant* c_2*.*

We use the higher-order kernel to fully exploit the smoothness of $m_0(x,q)$ and reduce the bias caused by the kernel convolution method. We further assume that the errors of the initial machine learning estimators $\{\breve{m}_{j,l}(x,q)\}_{j=0,1,l\in\{1,\ldots,L\}}$ and $\{\hat{\omega}_l(x)\}_{l\in\{1,\ldots,L\}}$ satisfy the following conditions.

Assumption 6. *For every subsample index* $l \in \{1, ..., L\}$*, there exists a vanishing sequence* ρ_N *such that*

$$\sup_{q\in\mathcal{Q}^\epsilon} \|\breve{m}_{j,l}(x,q) - m_j(x,q)\|_{\mathbb{P},\infty} = O_P(h_2\rho_N),\ j = 0, 1, \tag{22}$$

$$\sup_{q\in\mathcal{Q}^\epsilon} \int |\breve{m}_{1,l}(x,q) - m_1(x,q)|^2 dF_X(x) = O_P(h_2^2\rho_N^2), \tag{23}$$

$$\int |\hat{\omega}_l(x) - \omega(x)|^2 dF_X(x) = O_P(h_2^2\rho_N^2), \tag{24}$$

$$\int \left[\sup_{q\in\mathcal{Q}^\epsilon} |\breve{m}_{1,l}(x,q)|\right]^{2+\delta} dF_X(x) = O_P(1), \tag{25}$$

$$\int \left[\sup_{q\in\mathcal{Q}^\epsilon} |\hat{\omega}_l(x)(1+|\breve{m}_{0,l}(x,q)|)|\right]^{2+\delta} dF_X(x) = O_P(1), \tag{26}$$

$$\sup_{q\in\mathcal{Q}^\epsilon} \left|\int \left(\breve{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\breve{m}_{0,l}(x,q)\right) dF_X(x)\right| = o_P(N^{-1/2}), \tag{27}$$

$$\sup_{q\in\mathcal{Q}^\epsilon} \left|\int (\hat{\omega}_l(x)-\omega(x))(\breve{m}_{0,l}(x,q)-m_0(x,q))dF_X(x)\right| = o_P(N^{-1/2}). \tag{28}$$

Our final first-stage estimator of $(m_0(x,q), m_1(x,q), \omega(x))$ is $(\hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q), \hat{\omega}_l(x))$, where

$$\hat{m}_{0,l}(x,q) = \int \frac{\breve{m}_{0,l}(x,t)}{h_2} K_2\left(\frac{q-t}{h_2}\right) dt,\ j=0,1.$$

The next theorem shows that the high-level conditions in Assumption 2 hold for $(\hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q), \hat{\omega}_l(x))$.

Theorem 3. *Suppose Assumptions 1, 5 and 6 hold, then $(\hat{m}_{0,l}(x,q), \hat{m}_{1,l}(x,q), \hat{\omega}_l(x))$ satisfy Assumption 2 with $\nu_N = 1$, $A_N = 1$, and $\pi_N = h_2\rho_N + h_2^{2k}$.*

We can likewise apply the kernel convolution method to the Lasso preliminary estimators proposed in Section 3.1. Suppose ζ_N is of logarithmic rate and $\mathbb{E}\omega^4(X) < \infty$ so that $\delta = 2$, then Assumption 4.4 reduces to that $s\log(p) = o(N^{1/3})$ (up to some logarithmic rate). Assumption 6 requires $s^2\log(p) = o(N^{\frac{2k}{2k+1}})$ and $s^2\log^2(p) = o(N)$ (up to some logarithmic rate), which is weaker than Assumption 4.4 for all $k \geq 1$. It is also interesting to see that the rate restriction of the sparsity level s and dimensionality p can be relaxed by imposing extra smoothness. The cost for the weaker condition is that we have to introduce one tuning parameter. In our simulation studies and empirical illustration, we implement the Lasso preliminary estimation without the kernel convolution, and the simulation results are impeccable even without the kernel convolution.

4 Simulation Studies

In this section, we use Monte Carlo simulations to study the finite sample performance of the proposed method of estimation and inference for the UQPE.

Consider a set of alternative data-generating designs as follows. The outcome variable is generated according to the partial linear high-dimensional model

$$Y \mid X \sim N(g(X_1) + \sum_{j=2}^{p} \alpha_j X_j, 1),$$

where the function $g(\cdot)$ is defined in the following three ways:

$$\begin{aligned} &\text{DGP 1}: \quad && g(x) = 1.00 \cdot x \\ &\text{DGP 2}: \quad && g(x) = 1.00 \cdot x - 0.10 \cdot x^2 \\ &\text{DGP 3}: \quad && g(x) = 1.00 \cdot x - 0.10 \cdot x^2 + 0.01 \cdot x^3 \end{aligned}$$

The high-dimensional controls $(X_1, ..., X_p)^T$ are generated by

$$X_1 \mid (X_2, ..., X_p) \sim N\left(\sum_{j=2}^{p} \gamma_j X_j, 1\right) \quad \text{and} \quad (X_2, ..., X_p) \sim N(0, \Sigma_{p-1}),$$

where Σ_{p-1} is the $(p-1) \times (p-1)$ variance-covariance matrix whose (r, c)-element is $0.5^{2(|r-c|+1)}$. Note that this data-generating process (DGP) induces dependence of the control X_1 of main interest on the rest of the $p-1$ controls $(X_2, \ldots, X_p)^T$, as well as the dependence among the $p-1$ controls $(X_2, \ldots, X_p)^T$. For the high-dimensional parameter vectors in the above data-generating model, we set $(\alpha_2, \ldots, \alpha_p)^T = (\gamma_2, \ldots, \gamma_p)^T = (0.5^2, 0.5^3, ..., 0.5^p)^T$.

We follow the general estimation and inference approach outlined in Section 2 together with the Lasso preliminary estimator introduced in Section 3.1. We set $h(x_{-1}) = (x_{-1}^T, (x_{-1}^2)^T, (x_{-1}^3)^T)^T$ for estimation of ω_l, and set $b(x) = (x^T, (x^2)^T, (x^3)^T)^T$ for estimation of m_0 and m_1. For the choice of h_1, we undersmooth the rule-of-thumb optimal choice as $h_1 = 1.06\sigma(Y)N^{-1/5-0.01}$. For each design, we use 500 iterations of Monte Carlo simulations to compute the mean, bias, and root mean square error (RMSE) of the estimate, as well as the 95% uniform coverage over the set $[0.20, 0.80]$ of quantiles. To evaluate the bias, RMSE, and the 95% uniform coverage, we first numerically approximate the true UQPE by large-sample Monte Carlo simulations. Across sets of Monte Carlo simulations, we vary the DGP $\in$ {DGP 1, DGP 2, DGP 3}, the sample size $N \in \{250, 500\}$, and the dimension $p \in \{50, 100\}$.

Table 1 summarizes the simulation results. We can make the following three observations in these results. First, the bias of our UQPE estimator is small, especially relative to the RMSE. This feature of the results supports the fact that our estimator mitigates the bias via the use of the doubly robust score and sample splitting. Second, the RMSE decreases as the sample size increases. Third, the 95% uniform coverage frequencies are close to the nominal probability, namely, 0.95. This feature of the results supports our theory on the asymptotic validity of the bootstrap inference. From these simulation results, we confirm the main theoretical properties of the proposed method of estimation and inference for the UQPE across alternative data-generating processes. In addition to the simulation designs introduced above, we also experimented with other designs, and the simulation results are very similar and support the main theoretical properties of our proposed method as well.

DGP	N	p	L	τ	True UQPE	Estimates Mean	Bias	RMSE	95% Cover Point	Unif.
1	500	100	5	0.20	1.00	1.04	0.04	0.30	0.92	0.95
				0.40	1.00	1.04	0.04	0.37	0.91	
				0.60	1.00	1.05	0.05	0.42	0.92	
				0.80	1.00	1.01	0.01	0.50	0.93	
	1000	200	5	0.20	1.00	1.02	0.02	0.10	0.92	0.96
				0.40	1.00	1.03	0.03	0.09	0.93	
				0.60	1.00	1.03	0.03	0.09	0.93	
				0.80	1.00	1.02	0.02	0.11	0.92	
2	500	100	5	0.20	1.12	1.15	0.03	0.31	0.93	0.94
				0.40	1.03	1.07	0.04	0.37	0.90	
				0.60	0.96	1.01	0.05	0.44	0.91	
				0.80	0.88	0.89	0.01	0.47	0.93	
	1000	200	5	0.20	1.12	1.14	0.02	0.11	0.93	0.96
				0.40	1.03	1.05	0.02	0.09	0.94	
				0.60	0.95	0.98	0.03	0.09	0.93	
				0.80	0.87	0.90	0.03	0.10	0.92	
3	500	100	5	0.20	1.14	1.17	0.03	0.32	0.92	0.94
				0.40	1.04	1.09	0.04	0.37	0.91	
				0.60	0.97	1.03	0.05	0.44	0.92	
				0.80	0.90	0.92	0.01	0.48	0.92	
	1000	200	5	0.20	1.14	1.17	0.02	0.11	0.94	0.96
				0.40	1.04	1.07	0.02	0.09	0.94	
				0.60	0.97	1.00	0.03	0.09	0.93	
				0.80	0.90	0.93	0.03	0.10	0.91	

Table 1: Monte Carlo simulation results under approximate sparsity. The true UQPE is numerically computed with a large-sample Monte Carlo. The 95% coverage is uniform over the set $[0.20, 0.80]$.

5 Heterogeneous Counterfactual Marginal Effects of Job Corps Training

Applying our proposed method, we analyze heterogeneous counterfactual marginal effects of Job Corps training on labor outcomes in this section. Job Corps is the largest training program for disadvantaged youth in the United States. A number of economists have analyzed the causal effects of this job training program on labor, health, and behavioral outcomes. Schochet, Burghardt, and McConnell (2008) are the first to provide an intensive study of the survey data associated with Job Corps and find average effects of the program on a variety of labor and behavioral outcomes. Flores and Flores-Lagunes (2009) study the causal effects by accounting for the endogeneity of work experiences based on unconfoundedness given a set of observed controls. Flores, Flores-Lagunes, Gonzalez, and Neumann (2012) consider the labor effects of the duration of exposure to the program as a continuous treatment. Huber (2014) accounts for endogenous selection in employment as the mediator on health outcomes based on selection on observables given a set of observed controls. Frölich and Huber (2017) use the instrumental variables approach to disentangle the indirect effects through work hours and the direct effects of the program. Hsu, Huber, Lee, and Pipoz (2018) consider the duration of exposure to the program as a continuous treatment and study its effects on behavioral outcomes with employment status as a mediator using a large set of observed controls.

While the rich set of interesting empirical findings have been reported about the treatment effects of Job Corps, an analysis of heterogeneous counterfactual effects is missing in the literature to the best of our knowledge, despite its potential relevance to designing effective program policies and schemes. For instance, natural questions may arise about whether higher (respectively, lower) potential earners would benefit more (respectively, less) from counterfactually extending the duration of the training program. Since the entrance interview in Job Corps provides some information regarding the human capital of prospective trainees, answers to these empirical questions may possibly help the program designers to devise more efficient policies and schemes for the training programs. As such, we are interested in heterogeneous counterfactual marginal effects of the duration of the exposure to the program, as a continuous treatment variable, on labor outcomes measured by hourly wages. As in some of the preceding papers in this literature discussed above, we identify and estimate the causal effects based on unconfoundedness given a large set of observed controls by taking advantage of our machine-learning-based method. For the outcome variable, we consider hourly wages. For the continuous treatment variables, we consider two seemingly similar but different measures: the duration in days of participation in Job Corps and the duration in days of actually taking classes in Job Corps. As will be shown shortly, these two definitions entail qualitatively different empirical findings. We use 42 observed controls (and their powers) on which the unconfoundedness is assumed. Table 2 shows the summary statistics of our data. Different sets of observations are missing across different variables, and hence we use the intersection of observations that are non-missing across all the variables in use for our analysis. After dropping the missing observations, we are left with $n = 481$ when we define the duration of participation in Job Corps as the treatment, while we are left with $n = 368$ when we define the duration of actually taking classes in Job Corps. Note that the dimension of covariates is relatively large given these effective sample sizes, and hence high-dimensional econometric methods are indispensable.

Using the same computer program as the one used for simulation studies presented in Section 4, we obtain estimates, pointwise 95% confidence intervals, and uniform 95% confidence bands for $UQPE(\tau)$ for $\tau \in [0.20, 0.80]$. Table 3 summarizes the results. The row groups (I) and (II) report results for days in Job Corps as the treatment variable, while the row groups (III) and (IV) report results for days of

		25th Percentile	Median	Mean	75th Percentile	Non-Missing
Outcome Y	Hourly wage	4.750	5.340	5.892	6.500	7606
Treatment X_1	Days in Job Corps	54.0	129.0	153.4	237.0	4748
	Days taking classes	41.0	91.0	120.2	179.0	4207
Controls X_{-1}	Age	17.00	18.00	18.43	20.00	14653
	Female	0.000	0.000	0.396	1.000	14653
	White	0.000	0.000	0.303	1.000	14327
	Black	0.000	1.000	0.504	1.000	14327
	Hispanic origin	0.000	0.000	0.184	0.000	14288
	Native language is English	1.000	1.000	0.855	1.000	14327
	Years of education	9.00	10.00	10.24	11.00	14327
	Other job trainings	0.000	0.000	0.339	1.000	13500
	Mother's education	11.00	12.00	11.53	11.53	11599
	Mother worked	1.000	1.000	0.752	1.000	14223
	Father's education	11.00	12.00	11.50	12.00	8774
	Father worked	0.000	1.000	0.665	1.000	12906
	Received welfare	0.000	1.000	0.563	1.000	14327
	Head of household	0.000	0.000	0.123	0.000	14327
	Number of people in household	2.000	3.000	3.890	5.000	14327
	Married	0.000	0.000	0.021	0.000	14327
	Separated	0.000	0.000	0.017	0.000	14327
	Divorced	0.000	0.000	0.007	0.000	14327
	Living with spouse	0.000	0.000	0.014	0.000	14235
	Child	0.000	0.000	0.266	1.000	13500
	Number of children	0.000	0.000	0.347	0.000	13500
	Past work experience	0.000	1.000	0.648	1.000	14327
	Past hours of work per week	0.000	24.00	25.15	40.00	14299
	Past hourly wage	4.250	5.000	5.142	5.500	7884
	Expected wage after training	7.000	9.000	9.910	11.000	6561
	Public housing or subsidy	0.000	0.000	0.200	0.000	14327
	Own house	0.000	0.000	0.411	1.000	11457
	Have contributed to mortgage	0.000	0.000	0.255	1.000	13951
	Past AFDC	0.000	0.000	0.301	1.000	14327
	Past SSI or SSA	0.000	0.000	0.251	1.000	14327
	Past food stamps	0.000	0.000	0.438	1.000	14327
	Past family income $\geq$ \$12K	0.000	1.000	0.576	1.000	14327
	In good health	1.000	1.000	0.871	1.000	14327
	Physical or emotional problem	0.000	0.000	0.049	0.000	14327
	Smoke	0.000	1.000	0.537	1.000	14327
	Alcohol	0.000	1.000	0.584	1.000	14327
	Marijuana or hashish	0.000	0.000	0.369	1.000	14327
	Cocaine	0.000	0.000	0.033	0.000	14327
	Heroin/opium/methadone	0.000	0.000	0.012	0.000	14327
	LSD/peyote/psilocybin	0.000	0.000	0.055	0.000	14327
	Arrested	0.000	0.000	0.266	1.000	14327
	Number of times arrested	0.000	0.000	0.537	1.000	14218

Table 2: Summary statistics of data.

	Outcome	Treatment	τ	$\widehat{UQPE}(\tau)$	Pointwise 95% CI	Uniform 95% CB
(I)	Hourly	Days in	0.2	0.00118	[0.00097 0.00138]	[0.00087 0.00148]
	wage	Job Corps	0.4	0.00074	[0.00061 0.00087]	[0.00054 0.00094]
			0.6	0.00229	[0.00192 0.00266]	[0.00174 0.00284]
			0.8	0.00446	[0.00331 0.00561]	[0.00274 0.00618]
(II)	Log	Days in	0.2	0.00020	[0.00016 0.00024]	[0.00013 0.00027]
	hourly	Job Corps	0.4	0.00017	[0.00013 0.00021]	[0.00010 0.00024]
	wage		0.6	0.00037	[0.00030 0.00044]	[0.00025 0.00049]
			0.8	0.00062	[0.00046 0.00078]	[0.00035 0.00089]
(III)	Hourly	Days in	0.2	0.00327	[0.00278 0.00377]	[0.00253 0.00402]
	wage	Job Corps	0.4	0.00203	[0.00169 0.00239]	[0.00151 0.00257]
		classes	0.6	0.00110	[0.00082 0.00137]	[0.00068 0.00151]
			0.8	0.00267	[0.00133 0.00401]	[0.00064 0.00470]
(IV)	Log	Days in	0.2	0.00056	[0.00045 0.00067]	[0.00038 0.00074]
	hourly	Job Corps	0.4	0.00053	[0.00038 0.00068]	[0.00028 0.00077]
	wage	classes	0.6	0.00017	[0.00012 0.00022]	[0.00010 0.00026]
			0.8	0.00037	[0.00017 0.00057]	[0.00003 0.00070]

Table 3: Heterogeneous counterfactual marginal effects of days in Job Corps using $p = 42$ controls. The row groups (I) and (II) report results for days in Job Corps as the treatment variable, while the row groups (III) and (IV) report results for days of taking classes in Job Corps as the treatment variable. The row groups (I) and (III) report results for the hourly wage as the outcome variable, while the row groups (II) and (IV) report results for the logarithm of the hourly wage as the outcome variable. Row groups (I) and (II) are based on $n = 481$, while row groups (III) and (IV) are based on $n = 368$. The results are based on $L = 5$ folds of sample splitting.

taking classes in Job Corps as the treatment variable. The row groups (I) and (III) report results for the hourly wage as the outcome variable, while the row groups (II) and (IV) report results for the logarithm of the hourly wage as the outcome variable.

Overall, the magnitudes of the estimates are consistent with those from prior studies,[5] and we also obtain the following new findings. First, observe that the UQPE is globally positive in the table, and the pointwise 95% confidence intervals and uniform 95% confidence bands are all contained in the set of positive reals. These results indicate that the counterfactual marginal effects of our interest are globally positive across the heterogeneous subpopulations. We next look into the heterogeneity of these effects. Observe in (I) that the UQPE of days in Job Corps on hourly wages is smaller for $\tau = 0.2$ than that for $\tau = 0.8$ and that the uniform 95% confidence band does not intersect across these two quantile points. These results indicate that,

[5] In the row group (I) in Table 3 for instance, the *daily* marginal effects range from 0.0007 to 0.0044 dollars. This is consistent with the 0.22 difference in average hourly wages between the treatment and control groups (Schochet, Burghardt, and McConnell, 2008, Table 3), where the average number of days in Job Corps for the treated group is 153.4 (Table 2).

although the effects are globally positive, the levels of the effects are heterogeneous. However, we should be careful in analyzing this point because the larger effects for the subpopulation of higher potential earners (i.e., higher quantiles) could simply result from the scale effect. Heterogeneity in causal effects across different quantiles often vanishes once we take the logarithm of the outcome variable. Therefore, we next consider the row group (II), where the outcome variable is defined as the logarithm of the hourly wage. Notice that, even in this row group, we continue to observe the same qualitative pattern as that in the row group (I). Namely, the UQPE is smaller for $\tau = 0.2$ than that for $\tau = 0.8$, and the uniform 95% confidence band does not intersect across these two quantile points. These results indicate that the subpopulation of higher potential earners or those with more innate human capital would benefit more from marginally extending the duration of the training program than the subpopulation of lower potential earners or those with less innate human capital. On the other hand, if we turn to row groups (III) and (IV), where the treatment variable is now defined as days of taking classes in Job Corps, then we no longer observe the aforementioned pattern of heterogeneous counterfactual marginal effects, and both the pointwise 95% confidence intervals and the uniform 95% confidence bands largely overlap across different quantiles.

In summary, we obtain the following three new findings about counterfactual marginal effects of the duration of exposure to Job Corps training on the hourly wage. First, the effects are globally positive for all the subpopulations under consideration, regardless of the definition of the treatment variable and the definition of the outcome variable. Second, the counterfactual marginal effects of days in Job Corps are heterogeneous with larger effects for the subpopulation of higher potential earners. This result holds robustly regardless of whether we define the outcome variable as the hourly wage or the logarithm of it. Third, we fail to detect the aforementioned pattern of heterogeneous counterfactual marginal effects once we define the treatment variable as days of taking classes in Job Corps. The last two points imply that some features of the Job Corps program other than merely taking classes may well be the source of heterogeneous benefits to the different trainees.

6 Conclusion

Credible counterfactual analysis requires the unconfoundedness condition to be plausibly satisfied, and high-dimensional controls are preferred to this end. On the other hand, existing methods of estimation and inference for heterogeneous counterfactual effects are not compatible with high-dimensional settings. In this paper, we therefore propose a novel doubly robust score for double/debiased estimation and inference for the UQPE as a measure of heterogeneous counterfactual marginal effects. For a concrete implementation procedure, we propose a multiplier bootstrap inference method for the double/debiased estimator. Asymptotic theories are presented to guarantee that the bootstrap method works. Lower-level sufficient conditions for our assumptions tailored to the Lasso double/debiased estimator are also provided so that a user can check the plausibility of the required assumptions in terms of primitive conditions involving only non-stochastic population objects. In addition, we also propose a method of kernel convolution to accommodate general machine learners.

Applying the proposed method of estimation and inference to survey data of Job Corps, the largest training program for disadvantaged youth in the United States, we obtain the following two empirical findings. First, the marginal effects of counterfactually extending the duration of the exposure to the Job Corps program are globally positive across quantiles robustly regardless of the definitions of the treatment variable (days in Job Corps and days taking classes) and regardless of whether we define the outcome as the level or its logarithm. Second, these counterfactual effects are larger for higher potential earners than lower potential earners robustly regardless of whether we define the outcome as the level or its logarithm.

References

Abrevaya, J., Y.-C. Hsu, and R. P. Lieli (2015): "Estimating conditional average treatment effects," *Journal of Business & Economic Statistics*, 33, 485–505.

Athey, S., G. W. Imbens, and S. Wager (2016): "Approximate residual balancing: De-biased inference of average treatment effects in high dimensions," *arXiv preprint arXiv:1604.07125*.

Belloni, A. and V. Chernozhukov (2011): "ℓ_1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39, 82–130.

Belloni, A., V. Chernozhukov, D. Chetverikov, and Y. Wei (2018a): "Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework," *The Annals of Statistics*, 46, 3643.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen (2017): "Program Evaluation with High-dimensional Data," *Econometrica*, 85, 233–298.

Belloni, A., V. Chernozhukov, and K. Kato (2014): "Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems," *Biometrika*, 102, 77–94.

——— (2018b): "Valid Post-Selection Inference in High-Dimensional Approximately Sparse Quantile Regression Models," *Journal of the American Statistical Association*, 114, 749–758.

Chen, X., H. Hong, and A. Tarozzi (2008): "Semiparametric efficiency in GMM models with auxiliary data," *The Annals of Statistics*, 36, 808–843.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, and W. Newey (2017): "Double/debiased/neyman machine learning of treatment effects," *American Economic Review*, 107, 261–265.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018a): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21, C1–C68.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014a): "Anti-concentration and Honest, Adaptive Confidence Bands," *The Annals of Statistics*, 42, 1787–1818.

——— (2014b): "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics*, 42, 1564–1597.

Chernozhukov, V., J. C. Escanciano, H. Ichimura, W. K. Newey, and J. M. Robins (2018b): "Locally robust semiparametric estimation," *arXiv preprint arXiv:1608.00033*.

Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (2015): "Quantile regression with censoring and endogeneity," *Journal of Econometrics*, 186, 201–221.

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

Colangelo, K. and Y.-Y. Lee (2020): "Double debiased machine learning nonparametric inference with continuous treatments," *arXiv preprint arXiv:2004.03036*.

Fan, Q., Y.-C. Hsu, R. P. Lieli, and Y. Zhang (2019): "Estimation of conditional average treatment effects with high-dimensional data," *arXiv preprint arXiv:1908.02399*.

Farrell, M. H. (2015): "Robust inference on average treatment effects with possibly more covariates than observations," *Journal of Econometrics*, 189, 1–23.

Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75, 259–276.

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional quantile regressions," *Econometrica*, 77, 953–973.

Flores, C. A. and A. Flores-Lagunes (2009): "Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness," Working paper.

Flores, C. A., A. Flores-Lagunes, A. Gonzalez, and T. C. Neumann (2012): "Estimating the effects of length of exposure to instruction in a training program: the case of job corps," *Review of Economics and Statistics*, 94, 153–171.

Fortin, N., T. Lemieux, and S. Firpo (2011): "Decomposition methods in economics," in *Handbook of labor economics*, Elsevier, vol. 4, 1–102.

Frölich, M. and M. Huber (2017): "Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 1645–1666.

Frölich, M. and B. Melly (2013): "Unconditional quantile treatment effects under endogeneity," *Journal of Business & Economic Statistics*, 31, 346–357.

Graham, B. S. (2011): "Efficiency bounds for missing data models with semiparametric restrictions," *Econometrica*, 79, 437–452.

Graham, B. S., C. C. d. X. Pinto, and D. Egel (2012): "Inverse probability tilting for moment condition models with missing data," *The Review of Economic Studies*, 79, 1053–1079.

——— (2016): "Efficient estimation of data combination models by the method of auxiliary-to-study tilting (AST)," *Journal of Business & Economic Statistics*, 34, 288–301.

Hahn, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 66, 315–331.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161–1189.

Hsu, Y.-C., M. Huber, Y.-Y. Lee, and L. Pipoz (2018): "Direct and indirect effects of continuous treatments based on generalized propensity score weighting," Tech. rep., Université de Fribourg.

Hsu, Y.-C., T.-C. Lai, and R. P. Lieli (2020): "Counterfactual Treatment Effects: Estimation and Inference," *Journal of Business & Economic Statistics*, forthcoming.

Huber, M. (2014): "Identifying causal mechanisms (primarily) based on inverse probability weighting," *Journal of Applied Econometrics*, 29, 920–943.

Imbens, G. W. (1992): "An efficient method of moments estimator for discrete choice models with choice-based sampling," *Econometrica*, 60, 1187–1214.

Kennedy, E. H., Z. Ma, M. D. McHugh, and D. S. Small (2017): "Nonparametric methods for doubly robust estimation of continuous treatment effects," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 1229–1245.

Lee, S., R. Okui, and Y.-J. Whang (2017): "Doubly robust uniform confidence band for the conditional average treatment effect function," *Journal of Applied Econometrics*, 32, 1207–1225.

Machado, J. A. and J. Mata (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20, 445–465.

Melly, B. (2005): "Decomposition of differences in distribution using quantile regression," *Labour economics*, 12, 577–590.

Newey, W. and P. Ruud (2005): "Density weighted linear least squares," in *Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg*, ed. by D. Andrews and J. Stock, Cambridge University Press, 554–573.

Newey, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, 62, 1349–1382.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199.

Robins, J. M., L. Li, R. Mukherjee, E. T. Tchetgen, and A. van der Vaart (2017): "Minimax estimation of a functional on a structured high-dimensional model," *The Annals of Statistics*, 45, 1951–1987.

Robins, J. M., S. D. Mark, and W. K. Newey (1992): "Estimating exposure effects by modelling the expectation of exposure conditional on confounders," *Biometrics*, 48, 479–495.

Robins, J. M. and A. Rotnitzky (1995): "Semiparametric efficiency in multivariate regression models with missing data," *Journal of the American Statistical Association*, 90, 122–129.

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

Rothe, C. and S. Firpo (2019): "Properties of doubly robust estimators when nuisance functions are estimated nonparametrically," *Econometric Theory*, 35, 1048–1087.

Sant'Anna, P. H. and J. B. Zhao (2018): "Doubly robust difference-in-differences estimators," *arXiv:1812.01723*.

SASAKI, Y. AND T. URA (2020): "Estimation and Inference for Policy Relevant Treatment Effects," *arXiv preprint arXiv:1805.11503*.

SCHOCHET, P. Z., J. BURGHARDT, AND S. MCCONNELL (2008): "Does job corps work? Impact findings from the national job corps study," *American Economic Review*, 98, 1864–86.

SEMENOVA, V. AND V. CHERNOZHUKOV (2017): "Estimation and Inference about Conditional Average Treatment Effect and Other Structural Functions," *arXiv:1702.06240*.

SŁOCZYŃSKI, T. AND J. M. WOOLDRIDGE (2018): "A general double robustness result for estimating average treatment effects," *Econometric Theory*, 34, 112–133.

SU, L., T. URA, AND Y. ZHANG (2019): "Non-separable models with high-dimensional data," *Journal of Econometrics*, 212, 646–677.

TSIATIS, A. (2007): *Semiparametric theory and missing data*, Springer Science & Business Media.

VAN DER LAAN, M. J. AND J. M. ROBINS (2003): *Unified methods for censored longitudinal data and causality*, Springer Science & Business Media.

VAN DER LAAN, M. J. AND S. ROSE (2011): *Targeted learning: causal inference for observational and experimental data*, Springer Science & Business Media.

VAN DER LAAN, M. J. AND D. RUBIN (2006): "Targeted maximum likelihood learning," *The International Journal of Biostatistics*, 2.

VAN DER VAART, A. W. AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.

WAGER, S. AND S. ATHEY (2018): "Estimation and inference of heterogeneous treatment effects using random forests," *Journal of the American Statistical Association*, 113, 1228–1242.

WOOLDRIDGE, J. M. (2007): "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 141, 1281–1301.

ZIMMERT, M. AND M. LECHNER (2019): "Nonparametric estimation of causal heterogeneity under high-dimensional confounding," *arXiv preprint arXiv:1908.08779*.

A Double Robustness

The double robustness of (4) follows from Chernozhukov et al. (2018b, Theorem 3). In this section, for the sake of completeness, we demonstrate that (4) is doubly robust.

Lemma 1 (Double Robustness)**.** *Suppose Assumption 1 holds. If*

(i) $\int |\tilde{m}_1(x,q_\tau)|dF_X(x)$, $\int |\tilde{\omega}(x)1\{y \leq q_\tau\}|dF_{Y,X}(y,x)$, $\int |\tilde{\omega}(x)m_0(x,q_\tau)|dF_X(x)$, *and* $\int |\omega(x)\tilde{m}_0(x,q_\tau)|dF_X(x)$ *are finite;*

(ii) for every x_{-1} *in the support of* X_{-1}*, the mappings* $x_1 \mapsto (m_0(x,q_\tau) - \tilde{m}_0(x,q_\tau))$ *and* $x_1 \mapsto f_{X_1|X_{-1}=x_{-1}}(x_1)$ *are continuously differentiable with* $(m_0(x,q_\tau) - \tilde{m}_0(x,q_\tau))f_{X_1|X_{-1}=x_{-1}}(x_1) \to 0$ *as* $x_1 \to \pm\infty$*; and*

(iii) $\int \tilde{m}_1(x,q_\tau)dF_X(x) = \int \frac{\partial}{\partial x_1}\tilde{m}_0(x,q_\tau)dF_X(x)$*;*

then (5) *and* (6) *hold.*

In the above lemma, conditions (i) and (ii) are regularity conditions for the nuisance parameter values. Condition (iii) is satisfied if $\tilde{m}_1(x,q_\tau) = \frac{\partial}{\partial x_1}\tilde{m}_0(x,q_\tau)$. It is reasonable since $\tilde{m}_0(x,q_\tau)$ is a value for $m_0(x,q_\tau)$ and $\tilde{m}_1(x,q_\tau)$ is a value for $m_1(x,q_\tau) = \frac{\partial}{\partial x_1}m_0(x,q_\tau)$.

Proof. Note that (5) follows from

$$
\begin{aligned}
&\int \left(\tilde{m}_1(x,q_\tau) - \omega(x)(1\{y \leq q_\tau\} - \tilde{m}_0(x,q_\tau))\right) dF_{Y,X}(y,x) \\
&= \int \tilde{m}_1(x,q_\tau)dF_X(x) - \iint (m_0(x,q_\tau) - \tilde{m}_0(x,q_\tau)) \left(\frac{\partial}{\partial x_1} f_{X_1|X_{-1}=x_{-1}}(x_1)\right) dx_1 dF_{X_{-1}}(x_{-1}) \\
&= \int \tilde{m}_1(x,q_\tau)dF_X(x) + \iint \left(m_1(x,q_\tau) - \left(\frac{\partial}{\partial x_1}\tilde{m}_0(x,q_\tau)\right)\right) \left(f_{X_1|X_{-1}=x_{-1}}(x_1)\right) dx_1 dF_{X_{-1}}(x_{-1}) \\
&= \int m_1(x,q_\tau)dF_X(x) \\
&= \theta(\tau),
\end{aligned}
$$

where the first equality follows from Fubini's theorem, and the second equality follows from integration by parts. Next, (6) follows from

$$
\begin{aligned}
&\int \left(m_1(x,q_\tau) - \tilde{\omega}(x)(1\{y \leq q_\tau\} - m_0(x,q_\tau))\right) dF_{Y,X}(y,x) \\
&= \int m_1(x,q_\tau)dF_X(x) - \iint \tilde{\omega}(x)(m_0(x,q_\tau) - m_0(x,q_\tau)) f_{X_1|X_{-1}=x_{-1}} dx_1 dF_{X_{-1}}(x_{-1}) \\
&= \int m_1(x,q_\tau)dF_X(x) \\
&= \theta(\tau).
\end{aligned}
$$

This completes a proof of the lemma. □

B Proof of the Results in the Main Text

B.1 Proof of (2)

The proof of (2) is available in Firpo et al. (2009, Corollary 1), but we include it for the sake of completeness and for convenience of readers.

Lemma 2. *If Assumption 1 is satisfied, then* (2) *holds.*

Proof. Since $F_{Y(X_1+\varepsilon,X_{-1})}(Q_\tau(Y(X_1+\varepsilon,X_{-1}))) = \tau$, we have

$$\begin{aligned}
0 &= \frac{\partial}{\partial \varepsilon}(F_{Y(X_1+\varepsilon,X_{-1})}(Q_\tau(Y(X_1+\varepsilon,X_{-1})))\Big|_{\varepsilon=0} \\
&= \left(\frac{\partial}{\partial \varepsilon}F_{Y(X_1+\varepsilon,X_{-1})}\Big|_{\varepsilon=0}\right)(Q_\tau(Y(X))) + f_{Y(X)}(Q_\tau(Y(X)))\left(\frac{\partial}{\partial \varepsilon}Q_\tau(Y(X_1+\varepsilon,X_{-1}))\Big|_{\varepsilon=0}\right) \\
&= \left(\frac{\partial}{\partial \varepsilon}F_{Y(X_1+\varepsilon,X_{-1})}\Big|_{\varepsilon=0}\right)(Q_\tau(Y(X))) + f_{Y(X)}(Q_\tau(Y(X)))UQPE(\tau)
\end{aligned}$$

Therefore, $UQPE(\tau) = -\frac{\left(\frac{\partial}{\partial \varepsilon}F_{Y(X_1+\varepsilon,X_{-1})}\big|_{\varepsilon=0}\right)(Q_\tau(Y(X)))}{f_{Y(X)}(Q_\tau(Y(X)))}$ follows. The conditional independence between $Y(x_1+\varepsilon,x_{-1})$ and X_1 given X_{-1} implies $F_{Y(X_1+\varepsilon,X_{-1})}(y) = \int F_{Y(x_1+\varepsilon,x_{-1})|X=x}(y)dF_X(x) = \int F_{Y(x_1+\varepsilon,x_{-1})|X=(x_1+\varepsilon,x_{-1})}(y)dF_X(x) = \int F_{Y|X=(x_1+\varepsilon,x_{-1})}(y)dF_X(x)$. Thus, $\frac{\partial}{\partial \varepsilon}F_{Y(X_1+\varepsilon,X_{-1})}(y)\big|_{\varepsilon=0} = \int m_1(x,y)dF_X(x)$, which yields the statement of this lemma. □

B.2 Proof of (4)

Lemma 3. *Equation* (4) *holds under Assumption 1.*

Proof. This statement follows from

$$\begin{aligned}
&\mathbb{E}[m_1(X,q_\tau) - \theta - \omega(X)(1\{Y \le q_\tau\} - m_0(X,q_\tau))] \\
&= -\int \omega(x)(1\{y \le q_\tau\} - m_0(x,q_\tau))dF_{Y,X}(y,x) \\
&= -\int \omega(x)(\int 1\{y \le q_\tau\}dF_{Y|X=x}(y) - m_0(x,q_\tau))dF_X(x) \\
&= 0,
\end{aligned}$$

where the first equality follows from the definition of θ, the second equality comes from the law of iterated expectations, and the last equality follows from the definition of $m_0(x,q)$. □

B.3 Proof of Theorem 1

For a proof of this theorem, we let $\mathbb{P}_N f$, $\mathbb{P}_{n,l} f$, $\mathbb{P}_l f$, and $\mathbb{P} f$ denote $\frac{1}{N}\sum_{i=1}^N f(Z_i)$, $\frac{1}{n}\sum_{i\in I_l} f(Z_i)$, $\mathbb{E}(f(Z_i)|\{Z_j\}_{j\in I_l^c})$, and $\mathbb{E}f$, respectively. For a vector $v = (v_1, \cdots, v_k)$, $\text{diag}(v)$ denotes the a diagonal matrix with the diagonal being v. We write $a_N \lesssim b_N$ for two positive sequences a_N and b_N if there exists a constant independent of n such that $a_N \leq cb_N$. The constant c may vary in different contexts. For any estimator $\hat{\theta}$, we follow the empirical processes literature and denote $\mathbb{E}f(X, \hat{\theta})$ as $\mathbb{E}f(X, \theta)$ evaluated at $\theta = \hat{\theta}$.

The proof of Theorem 1 is divided into three sections. In Section B.3.1, we prove several technical lemmas that will be used later. In Section B.3.2, we derive the linear expansion of $\widehat{UQPE}(\tau)$. In Section B.3.3, we derive the linear expansion of $\widehat{UQPE}^*(\tau)$.

B.3.1 Useful Lemmas

Define $\phi_i(q) = m_1(X_i, q) - \omega(X_i)(1\{Y_i \leq q\} - m_0(X_i, q)) - \theta(\tau)$ and $\hat{\phi}_{i,l}(q) = \hat{m}_{1,l}(X_i, q) - \hat{\omega}_l(X_i)(1\{Y_i \leq q\} - \hat{m}_{0,l}(X_i, q)) - \theta(\tau)$.

Lemma 4. *Under the assumptions in Theorem 1, $\frac{1}{L}\sum_{l=1}^L \mathbb{P}_l(\hat{\phi}_{i,l}(\hat{q}_\tau) - \phi_i(\hat{q}_\tau)) = o_P(N^{-1/2})$ for any estimator $(\hat{\omega}_l(x), \hat{m}_{0,l}(x, q), \hat{m}_{1,l}(x, q))$ of $(\omega(x), m_0(x, q), m_1(x, q))$ and any quantile index $\tau \in \Upsilon$.*

Proof. Using the law of iterated expectations and $m_0(x, q) = \int 1\{y \leq q\} dF_{Y|X=x}(y)$, we have

$$\begin{aligned}
&\int (m_1(x, q) - \omega(x)(1\{y \leq q\} - m_0(x, q)))\, dF_{Y,X}(y, x) \\
&- \int (\hat{m}_{1,l}(x, q) - \hat{\omega}_l(x)(1\{y \leq q\} - \hat{m}_{0,l}(x, q)))\, dF_{Y,X}(y, x) \\
&= \int (\hat{m}_{1,l}(x, q) - m_1(x, q))\, dF_X(x) + \int \omega(x)(\hat{m}_{0,l}(x, q) - m_0(x, q)) dF_X(x) \\
&\quad + \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x, q) - m_0(x, q)) dF_X(x).
\end{aligned}$$

The integration by parts implies

$$\begin{aligned}
&\int \omega(x)(\hat{m}_{0,l}(x, q) - m_0(x, q)) f_{X_1|X_{-1}=x_{-1}}(x_1) dx_1 \\
&= -\int \left(\frac{\partial}{\partial x_1}\hat{m}_{0,l}(x, q) - \frac{\partial}{\partial x_1} m_0(x, q)\right) f_{X_1|X_{-1}=x_{-1}}(x_1) dx_1,
\end{aligned}$$

where $(\hat{m}_{0,l}(x, q) - m_0(x, q)) f_{X_1|X_{-1}=x_{-1}}(x_1)$ disappears on the boundary of x_1. Then

$$\int (m_1(x, q) - \omega(x)(1\{y \leq q\} - m_0(x, q)))\, dF_{Y,X}(y, x)$$

$$- \int \left(\hat{m}_{1,l}(x,q) - \hat{\omega}_l(x)(1\{y \le q\} - \hat{m}_{0,l}(x,q))\right) dF_{Y,X}(y,x)$$

$$= \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q)\right) dF_X(x) + \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x).$$

Because $\sup_{\tau\in\Upsilon} |\hat{q}_\tau - q_\tau| = o_P(N^{-1/2})$, we have, with probability approaching one (w.p.a.1),

$$\begin{aligned}
|\mathbb{P}_l(\hat{\phi}_{i,l}(\hat{q}_\tau) - \phi_i(\hat{q}_\tau))| \le & \sup_{q\in\mathcal{Q}} \left| \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q)\right) dF_X(x) \right| \\
& + \sup_{q\in\mathcal{Q}} \left| \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \right| \\
= & o_P(N^{-1/2}),
\end{aligned}$$

where the last equality holds due to (14) and (15). □

Lemma 5. *Let $\tilde{\eta}_i = 1$ for every $i = 1,\ldots,N$ or if $\tilde{\eta}_i = 1+\eta_i$ for every $i = 1,\ldots,N$. If the assumptions in Theorem 1 hold, then* $\sup_{l\in\{1,\ldots,L\},q\in\mathcal{Q}} |(\mathbb{P}_{n,l} - \mathbb{P}_l)\tilde{\eta}_i(\hat{\phi}_{i,l}(q) - \phi_i(q))| = o_P(N^{-1/2})$.

Proof. Define $\mathcal{M}_l(M)$ the set of $(\tilde{m}_1(x,q), \tilde{m}_0(x,q), \tilde{\omega}(x))$ which satisfies

$$\begin{aligned}
& \{\tilde{m}_j(x,q) : q \in \mathcal{Q}\} \subset \{\hat{m}_j(x,q) : q \in \mathcal{Q}\},\ j = 0,1 \\
& \sup_{q\in\mathcal{Q}} \int |\tilde{m}_1(x,q) - m_1(x,q)|^2 dF_X(x) \le M\pi_N, \\
& \int |\tilde{\omega}(x) - \omega(x)|^2 dF_X(x) \le M\pi_N, \\
& \sup_{q\in\mathcal{Q}} \int |\tilde{\omega}(x)\tilde{m}_0(x,q) - \omega(x)m_0(x,q)|^2 dF_X(x) \le M\pi_N, \\
& \int \left[\sup_{q\in\mathcal{Q}} |\tilde{m}_1(x,q)| + \sup_{q\in\mathcal{Q}} |m_1(x,q)|\right]^{2+\delta} dF_X(x) \le M, \quad \text{and} \\
& \int \left[\sup_{q\in\mathcal{Q}} |\tilde{\omega}_l(x)(1 + |\tilde{m}_0(x,q)|)| + \sup_{q\in\mathcal{Q}} |\omega(x)(1 + m_{0,l}(x,q))|\right]^{2+\delta} dF_X(x) \le M.
\end{aligned}$$

Define $F_l(X_i) = |\tilde{\eta}_i| \sup_{q\in\mathcal{Q}} |\hat{\omega}_l(x)(1+|\hat{m}_{0,l}(x,q)|)| + |\tilde{\eta}_i| \sup_{q\in\mathcal{Q}} |\omega(x)(1+m_{0,l}(x,q))| + |\tilde{\eta}_i| \sup_{q\in\mathcal{Q}} |\hat{m}_{1,l}(x,q)| + |\tilde{\eta}_i| \sup_{q\in\mathcal{Q}} |m_1(x,q)|$, and

$$\mathcal{F}_l = \left\{ \begin{array}{l} \tilde{\eta}_i(\hat{m}_{1,l}(X_i,q) - \hat{\omega}_l(X_i)(1\{Y_i \le q\} - \hat{m}_{0,l}(X_i,q))) \\ -\tilde{\eta}_i(m_1(X_i,q) - \omega(X_i)(1\{Y_i \le q\} - m_0(X_i,q))) \end{array} : q \in \mathcal{Q} \right\}.$$

By Assumption 2, for any $\delta > 0$, we can find a sufficiently large constant $M > 0$ such that $(\hat{m}_{1,l}, \hat{m}_{0,l}, \hat{\omega}_l) \in \mathcal{M}_l(M)$ occurs with probability greater than $1-\delta$. Conditional on $\{(\hat{m}_{1,l}, \hat{m}_{0,l}, \hat{\omega}_l) \in \mathcal{M}_l(M)\}$ and $\{X_i, Y_i\}_{i \in I_l^c}$, we can treat $\hat{m}_{1,l}, \hat{m}_{0,l}, \hat{\omega}_l$ as fixed, and $\mathbb{P}_l F_l^{2+\delta} < \infty$. In addition, by Van der Vaart and Wellner (1996, Theorem 2.7.11) and the fact that $\sup_Q N(\{\hat{m}_j(x,q) : q \in \mathcal{Q}\}, e_Q, \varepsilon ||G_l^{(j)}||_{Q,2}) \lesssim \left(\frac{A_N}{\varepsilon}\right)^{\nu_N}$, we have $\sup_Q N(\mathcal{F}_l, e_Q, \varepsilon ||F_l||_{Q,2}) \lesssim \left(\frac{A_N}{\varepsilon}\right)^{\nu_N}$. Furthermore, note that

$$\begin{aligned}\sup_{f \in \mathcal{F}_l} \mathbb{P}_l f^2 \leq & \sup_{q \in \mathcal{Q}} \mathbb{P}_l \left[|\hat{m}_{1,l}(X,q) - m_1(X,q)| + |\hat{\omega}_l(X) - \omega(X)| + |\omega(X) m_0(X,q) - \hat{\omega}_l(X)\hat{m}_{0,l}(X,q)|\right]^2 \\ \lesssim & \pi_N^2.\end{aligned}$$

Then, by Chernozhukov et al. (2014b, Corollary 5.1), we have

$$\begin{aligned}\mathbb{P}_l \sup_{q \in \mathcal{Q}} |(\mathbb{P}_{n,l} - \mathbb{P}_l)(\hat{\phi}_{i,l}(q) - \phi_i(q))| &= \mathbb{P}_l ||(\mathbb{P}_{n,l} - \mathbb{P}_l)||_{\mathcal{F}_l} \\ &\lesssim \sqrt{\frac{\pi_N^2 \nu_N}{N} \log\left(\frac{A_N ||F||_{\mathbb{P}_l,2}}{\pi_N}\right)} + \frac{\nu_N || \max_i F(X_i)||_{\mathbb{P}_l,2}}{N} \log\left(\frac{A_N ||F||_{\mathbb{P}_l,2}}{\pi_N}\right).\end{aligned}$$

Because $\mathbb{E} F_l^{2+\delta} < \infty$, we have $|| \max_i F_l||_{\mathbb{P}_l,2} = O(N^{1/(2+\delta)})$ on $\{(\hat{m}_{1,l}, \hat{m}_{0,l}, \hat{\omega}_l) \in \mathcal{M}(\varepsilon, M)\}$.[6] Then, by letting n be sufficiently large, we have

$$\begin{aligned}\mathbb{P}_l \sup_{q \in \mathcal{Q}} |(\mathbb{P}_{n,l} - \mathbb{P}_l)(\hat{\phi}_{i,l}(q) - \phi_i(q))| &\lesssim \pi_N \nu_N^{1/2} N^{-1/2} \log^{1/2}(A_N/\pi_N) + \nu_N N^{-(1+\delta)/(2+\delta)} \log(A_N/\pi_N) \\ &= o(N^{-1/2}).\end{aligned}$$

This leads to the desired result. ☐

Lemma 6. *Under the assumptions in Theorem 1,* $\sup_{l \in \{1,\ldots,L\}, \tau \in \Upsilon} |(\mathbb{P}_{n,l} - \mathbb{P}_l)(\phi_i(\hat{q}_\tau) - \phi_i(q_\tau))| = o_P(N^{-1/2})$ *and* $\sup_{l \in \{1,\ldots,L\}, \tau \in \Upsilon} |(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i + 1)(\phi_i(\hat{q}^*_\tau) - \phi_i(q_\tau))| = o_P(N^{-1/2})$.

Proof. We know that $\sup_{\tau \in \Upsilon} |\hat{q}_\tau - q_\tau| = O_P(N^{-1/2})$ and $\sup_{\tau \in \Upsilon} |\hat{q}^*_\tau - q_\tau| = O_P(N^{-1/2})$.[7] These conditions imply that, for any $\varepsilon > 0$, there exists a constant $M > 0$ such that

$$\mathbb{P}\left(\sup_{\tau \in \Upsilon} |\hat{q}^*(\tau) - q_\tau| \leq M N^{-1/2}, \sup_{\tau \in \Upsilon} |\hat{q}_\tau - q_\tau| \leq M N^{-1/2}\right) \geq 1 - \varepsilon.$$

Next, we show

$$\sup_{|v| \leq M, \tau \in \Upsilon} |(\mathbb{P}_{n,l} - \mathbb{P}_l)\tilde{\eta}_i(\phi_i(q_\tau + v N^{-1/2}) - \phi_i(q_\tau))| = o_P(N^{-1/2}).$$

---

[6] If $\{X_i\}$ is sequence of i.i.d. nonnegative random variables with $\mathbb{E} X_i^{2+\delta} \leq M$, then $[\mathbb{E}(\max_{i=1,\ldots,N} X_i)^2]^{1/2} \lesssim N^{\frac{1}{2+\delta}}$. It is shown as follows. Note that $\mathbb{E}(\max_{i=1,\ldots,N} X_i)^2 = 2\int_0^\infty x \mathbb{P}(\max_{i=1,\ldots,N} X_i) > x) dx = 2\int_0^{\alpha_N} x \mathbb{P}(\max_{i=1,\ldots,N} X_i) > x) dx + 2\int_{\alpha_N}^\infty x \mathbb{P}(\max_{i=1,\ldots,N} X_i) > x) dx \leq \alpha_N^2 + 2N \int_{\alpha_N}^\infty \frac{\mathbb{E} X^{2+\delta}}{X^{1+\delta}} dx \leq \alpha_N^2 + \frac{2MN}{\delta \alpha_N^\delta}$. We can obtain the desired result by taking $\alpha_N = N^{\frac{1}{2+\delta}}$.

[7] See Section B.3.3 for more detail.

Let $\mathcal{F} = \{\tilde{\eta}_i \left(\phi_i(q_\tau + vN^{-1/2}) - \phi_i(q_\tau)\right) : |v| \leq M, \tau \in \Upsilon\}$ with envelope $F(X_i) = |\tilde{\eta}_i| \left[\sup_{q\in\mathcal{Q}} |m_1(x,q)| + \sup_{q\in\mathcal{Q}} |\omega(x)(1 + m_{0,l}(x,q))|\right]$. Note that $\mathcal{F}$ is nested in $\{\tilde{\eta}_i \left(\phi_i(q_1) - \phi_i(q_2)\right) : q_1, q_2 \in \mathbb{R}\}$. Because $m_j(x,q)$ is Lipschitz continuous in q and $\{1\{Y \leq q\} : q \in \mathbb{R}\}$ is a VC class with VC index 2, we have $J(\theta) = \int_0^\theta \sqrt{1 + \log(\sup_Q N(\mathcal{F}, e_Q, \varepsilon ||F||_{Q,2})} d\varepsilon \lesssim \theta\sqrt{\log(a/\theta)}$ for some constant $a > 0$.

Last,

$$\sup_{f\in\mathcal{F}} \mathbb{P}_l f^2 \leq \mathbb{P}_l \sup_{\tau\in\Upsilon, |v|\leq M} \Big\{ \left|m_1(X, q_\tau + vN^{-1/2}) - m_1(X, q_\tau)\right| + |\omega(X)| \left[\left|m_0(X, q_\tau + vN^{-1/2}) - m_0(X, q_\tau)\right| + \left|1\{Y \leq q_\tau\} - 1\{Y \leq q_\tau + vN^{-1/2}\}\right|\right] \Big\}^2 \lesssim N^{-1/2}.$$

Then, by Chernozhukov et al. (2014b, Corollary 5.1), we have

$$\begin{aligned}
&\mathbb{P}_l \sup_{|v|\leq M, \tau\in\Upsilon} |(\mathbb{P}_{n,l} - \mathbb{P}_l)(\phi_i(q_\tau + vN^{-1/2}) - \phi_i(q_\tau))| \\
=& \mathbb{P}_l ||(\mathbb{P}_{n,l} - \mathbb{P}_l)||_{\mathcal{F}_l} \\
\lesssim& \sqrt{\frac{1}{N^{3/2}} \log\left(a||F||_{\mathbb{P}_l,2} N\right)} + \frac{|| \max_i F(X_i)||_{\mathbb{P}_l,2}}{N} \log\left(a||F||_{\mathbb{P}_l,2} N\right) = o(N^{-1/2}).
\end{aligned}$$

Therefore the statement of this lemma holds. □

Lemma 7. *Under the assumptions in Theorem 1,*

$$\sup_{\tau\in\Upsilon} \left| \hat{\theta}(\tau) - \theta(\tau) - \frac{1}{N} \sum_{i=1}^N \left[\phi_i(q_\tau) + \frac{\frac{\partial}{\partial q} \mathbb{E} m_1(X, q_\tau)}{f_Y(q_\tau)} (\tau - 1\{Y_i \leq q_\tau\}) \right] \right| = o_P(N^{-1/2}).$$

Proof. Taking $\tilde{\eta}_i = 1$ and by (7) and Lemmas 4, 5, and 6, we have

$$\begin{aligned}
\hat{\theta}(\tau) - \theta(\tau) &= \frac{1}{L} \sum_{l=1}^L \mathbb{P}_{n,l} \hat{\phi}_{i,l}(\hat{q}_\tau) \\
&= \frac{1}{L} \sum_{l=1}^L (\mathbb{P}_{n,l} - \mathbb{P}_l) \phi_i(q_\tau) + \frac{1}{L} \sum_{l=1}^L \mathbb{P}_l \phi_i(\hat{q}_\tau) + \frac{1}{L} \sum_{l=1}^L (\mathbb{P}_{n,l} - \mathbb{P}_l)(\hat{\phi}_{i,l}(\hat{q}_\tau) - \phi_i(\hat{q}_\tau)) \\
&+ \frac{1}{L} \sum_{l=1}^L \mathbb{P}_l (\hat{\phi}_{i,l}(\hat{q}_\tau) - \phi_i(\hat{q}_\tau)) + \frac{1}{L} \sum_{l=1}^L (\mathbb{P}_{n,l} - \mathbb{P}_l)(\phi_i(\hat{q}_\tau) - \phi_i(q_\tau)).
\end{aligned}$$

Rearranging the above equation, we have $\hat{\theta}(\tau) - \theta(\tau) = \frac{1}{L} \sum_{l=1}^L (\mathbb{P}_{n,l} - \mathbb{P}_l)(\phi_i(q_\tau)) + \frac{1}{L} \sum_{l=1}^L \mathbb{P}_l \phi_{i,l}(\hat{q}_\tau) + o_P(N^{-1/2})$ where the $o_P(N^{-1/2})$ term holds uniformly over $\tau \in \Upsilon$.

By $\mathbb{E}m_1(X, q_\tau) = 0$ and the usual delta method,

$$\mathbb{P}_l\phi_i(\hat{q}_\tau) = (\mathbb{E}m_1(X,\hat{q}_\tau) - \mathbb{E}m_1(X,q_\tau)) = \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X,q_\tau)}{f_Y(q_\tau)}\frac{1}{N}\sum_{i=1}^{N}(\tau - 1\{Y_i \leq q_\tau\}) + o_P(N^{-1/2})$$

where the $o_P(N^{-1/2})$ term holds uniformly over $l = 1, \ldots, L$ and $\tau \in \Upsilon$. Therefore the statement of this lemma holds. □

B.3.2 Linear Expansion of $\widehat{UQPE}(\tau)$

Note that $\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau) = A_1(\tau) + A_2(\tau) + A_3(\tau)$, where $A_1(\tau) \equiv (\mathbb{P}_N - \mathbb{P})\frac{1}{h_1}K_1\left(\frac{Y_i - \hat{q}_\tau}{h_1}\right)$, $A_2(\tau) \equiv \mathbb{P}\frac{1}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau}{h_1}\right) - f_Y(\hat{q}_\tau)$ and $A_3(\tau) \equiv f_Y(\hat{q}_\tau) - f_Y(q_\tau)$. Below we will analyze $A_1(\tau)$, $A_2(\tau)$, and $A_3(\tau)$, and then derive the linear expansion of $\widehat{UQPE}(\tau)$.

First, we will analyze $A_1(\tau)$. Let $R_1(\tau) = A_1(\tau) - (\mathbb{P}_N - \mathbb{P})\frac{1}{h_1}K_1\left(\frac{Y_i - q_\tau}{h_1}\right)$. Because $\sup_{\tau\in\Upsilon}|\hat{q}_\tau - q_\tau| = O_P(N^{-1/2})$. For any $\varepsilon > 0$, there exists a constant $M > 0$ such that, with probability greater than $1 - \varepsilon$,

$$\sup_{\tau\in\Upsilon}|R_1(\tau)| \leq \sup_{q\in\mathcal{Q},|v|\leq M}\left|(\mathbb{P}_N - \mathbb{P})\left(\frac{1}{h_1}K_1\left(\frac{Y_i - q - v/\sqrt{N}}{h_1}\right) - \frac{1}{h_1}K_1\left(\frac{Y_i - q}{h_1}\right)\right)\right|.$$

In the following, we aim to bound $\sup_{q\in\mathcal{Q},|v|\leq M}\left|(\mathbb{P}_N - \mathbb{P})\frac{\tilde{\eta}_i}{h_1}\left[K_1\left(\frac{Y_i-q-v/\sqrt{N}}{h_1}\right) - K_1\left(\frac{Y_i-q}{h_1}\right)\right]\right|$. Consider the class of functions $\mathcal{F} = \left\{\frac{\tilde{\eta}_i}{h_1}\left[K_1\left(\frac{y-q-v/\sqrt{N}}{h_1}\right) - K_1\left(\frac{y-q}{h_1}\right)\right] : q \in \mathcal{Q}, |v| \leq M\right\}$ with an envelope function $F_i = C|\tilde{\eta}_i|/h$ for some constant $C > 0$ such that $||\max_{i=1,\ldots,N} F_i||_{\mathbb{P},2} \lesssim \sqrt{\log(N)}$. We note that $\mathcal{F}$ is a VC-class with a fixed VC index and

$$\sup_{f\in\mathcal{F}}\mathbb{P}f^2 = \sup_{q\in\mathcal{Q},|v|\leq M}\int\left(K_1(u - \frac{v}{\sqrt{N}h_1}) - K_1(u)\right)^2 f_Y(q + h_1u)du \lesssim 1/(Nh_1^2).$$

Therefore, Chernozhukov et al. (2014b, Corollary 5.1) implies

$$\mathbb{E}\sup_{q\in\mathcal{Q},|v|\leq M}\left|(\mathbb{P}_N - \mathbb{P})\left(\frac{\tilde{\eta}_i}{h_1}\left[K_1\left(\frac{Y_i - q - v/\sqrt{N}}{h_1}\right) - K_1\left(\frac{Y_i - q}{h_1}\right)\right]\right)\right| \lesssim \frac{\sqrt{\log(N)}}{Nh_1} + \frac{\log(N)^{3/2}}{Nh_1},$$

and thus, $\sup_{\tau\in\Upsilon}|R_1(\tau)| = o_P(N^{-1/2})$.

Second, we will analyze $A_2(\tau)$. Let $R_2(\tau) = A_2(\tau) - \frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2$. By the Taylor expansion, we have

$$\begin{aligned}\sup_{\tau\in\Upsilon}|R_2(\tau)| &\leq \sup_{\tau\in\Upsilon}\left|\int(f_Y(\hat{q}_\tau + uh_1) - f_Y(\hat{q}_\tau))K_1(u)du - \frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2\right| \\ &\lesssim \sup_{\tau\in\Upsilon}\frac{|f_Y^{(2)}(q_\tau) - f_Y^{(2)}(\tilde{q}_\tau)|(\int u^2K_1(u)du)}{2}h_1^2,\end{aligned}$$

where $\tilde{q}_\tau$ is between $\hat{q}_\tau$ and $\hat{q}_\tau + h_1$ such that $\sup_{\tau\in\Upsilon}|\tilde{q}_\tau - q_\tau| \le \sup_{\tau\in\Upsilon}|\tilde{q}_\tau - \hat{q}_\tau| + \sup_{\tau\in\Upsilon}|\hat{q}_\tau - q_\tau| = O_P(h_1 + N^{-1/2})$. Therefore, $\sup_{\tau\in\Upsilon}|R_2(\tau)| = O_P(h_1^3 + h_1 N^{-1/2}) = o_P(N^{-1/2})$.

Third, we will analyze $A_3(\tau)$. By the delta method, we have

$$A_3(\tau) = f_Y^{(1)}(q_\tau)(\hat{q}_\tau - q_\tau) + R_3'(\tau) = \frac{f_Y^{(1)}(q_\tau)}{f_Y(q_\tau)}\left[\frac{1}{N}\sum_{i=1}^{N}(\tau - \mathbf{1}\{Y_i \le q_\tau\})\right] + R_3(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_3'(\tau)| + \sup_{\tau\in\Upsilon}|R_3(\tau)| = o_P(N^{-1/2})$.

Last, we will derive the linear expansion of $\widehat{UQPE}(\tau)$. Combining the analyses of $A_1(\tau)$, $A_2(\tau)$, and $A_3(\tau)$, we have

$$\begin{aligned}\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau) = (\mathbb{P}_N - \mathbb{P})&\left[\frac{1}{h_1}K_1\left(\frac{Y_i - q_\tau}{h_1}\right) + \frac{f_Y^{(1)}(q_\tau)}{f_Y(q_\tau)}(\tau - \mathbf{1}\{Y_i \le q_\tau\})\right]\\ &+ \frac{f_Y^{(2)}(q_\tau)(\int u^2 K_1(u)du)}{2}h_1^2 + R_4(\tau),\end{aligned} \tag{29}$$

where $\sup_{\tau\in\Upsilon}|R_4(\tau)| = o_P(N^{-1/2})$. Based on (29), we have

$$\sup_{\tau\in\Upsilon}|\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau)| = O_P(\log^{1/2}(N)(Nh_1)^{-1/2} + h_1^2),$$

$$\sup_{\tau\in\Upsilon}\left|\frac{(\hat{\theta}(\tau) - \theta(\tau))(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))}{\hat{f}_Y(\hat{q}_\tau)f_Y(q_\tau)}\right| = o_P(N^{-1/2}),$$

and

$$\sup_{\tau\in\Upsilon}\left|\frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y(\hat{q}_\tau)}\right| = O_P(\log(N)(Nh_1)^{-1} + h_1^4) = o_P(N^{-1/2}).$$

Therefore

$$\begin{aligned}\widehat{UQPE}(\tau) - UQPE(\tau) =& -\frac{\hat{\theta}(\tau) - \theta(\tau)}{f_Y(q_\tau)} + \frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))}{f_Y^2(q_\tau)}\\ &+ \frac{(\hat{\theta}(\tau) - \theta(\tau))(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))}{\hat{f}_Y(\hat{q}_\tau)f_Y(q_\tau)} - \frac{\theta(\tau)(\hat{f}_Y(\hat{q}_\tau) - f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y(\hat{q}_\tau)}\\ =&\frac{1}{N}\sum_{i=1}^{N}\mathrm{IF}_i(\tau) + \frac{\theta(\tau)f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2f_Y^2(q_\tau)} + R(\tau),\end{aligned} \tag{30}$$

where $\sup_{\tau\in\Upsilon}|R(\tau)| = o_P(N^{-1/2})$.

B.3.3 Linear Expansion of $\widehat{UQPE}^*(\tau)$

First, we will derive the linear expansion of $\hat{q}^*_\tau$. Note that $\hat{q}^*_\tau$ is the optimizer of the objective function $\sum_{i=1}^N \rho_\tau(Y_i - q) - q\sum_{i=1}^N \eta_i(\tau - \mathbf{1}\{Y_i \leq \hat{q}_\tau\})$. Define the local parameter as $\hat{u} = \sqrt{N}(\hat{q}^*_\tau - q_\tau)$. Then

$$\hat{u} = \arg\min_u \sum_{i=1}^N \rho_\tau(Y_i - q_\tau - uN^{-1/2}) - uN^{-1/2}\sum_{i=1}^N \eta_i(\tau - \mathbf{1}\{Y_i \leq \hat{q}_\tau\})$$

Note that $u \mapsto \sum_{i=1}^N \rho_\tau(Y_i - q_\tau - uN^{-1/2}) - uN^{-1/2}\sum_{i=1}^N \eta_i(\tau - \mathbf{1}\{Y_i \leq \hat{q}_\tau\})$ is convex in u for any $\tau \in \Upsilon$. By the Knight's identity, we can show that

$$\left[\sum_{i=1}^N \rho_\tau(Y_i - q_\tau - uN^{-1/2}) - uN^{-1/2}\sum_{i=1}^N \eta_i(\tau - \mathbf{1}\{Y_i \leq \hat{q}_\tau\})\right] - \left[-\frac{u}{\sqrt{N}}\sum_{i=1}^N(\eta_i+1)(\tau - \mathbf{1}\{Y_i \leq q_\tau\}) + \frac{f_Y(q_\tau)u^2}{2}\right]$$

is $o_P(1)$ pointwise in u. Then, by the convexity lemma (Pollard, 1991), we have

$$\hat{q}^*_\tau - q_\tau = \frac{1}{Nf_Y(q_\tau)}\sum_{i=1}^N(\eta_i+1)(\tau - \mathbf{1}\{Y_i \leq q_\tau\}) + R^*_1(\tau), \tag{31}$$

where $\sup_{\tau\in\Upsilon}|R^*_1(\tau)| = o_P(N^{-1/2})$.

Second, we will derive the linear expansion of $\hat{\theta}^*(\tau)$. Let $\hat{n}_l = \sum_{i\in I_l}(\eta_i + 1)$. Then,

$$\begin{aligned}
\hat{\theta}^*(\tau) - \theta(\tau) =& \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}\mathbb{P}_{n,l}(\eta_i+1)\hat{\phi}_{i,l}(\hat{q}^*_\tau) \\
=& \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i+1)\phi_i(\hat{q}^*_\tau) + \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}\mathbb{P}_l\hat{\phi}_{i,l}(\hat{q}^*_\tau) + R^*_1(\tau) \\
=& \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i+1)\phi_i(\hat{q}^*_\tau) + \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) + R^*_2(\tau) \\
=& \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}(\mathbb{P}_{n,l} - \mathbb{P}_l)(\eta_i+1)\phi_i(q_\tau) + \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) + R^*_3(\tau) \\
=& (\mathbb{P}_N - \mathbb{P})(\eta_i+1)\phi_i(q_\tau) + \frac{1}{L}\sum_{l=1}^L \frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}^*_\tau) + R^*_4(\tau),
\end{aligned} \tag{32}$$

where $\sup_{\tau\in\Upsilon}|R_j(\tau)| = o_P(N^{-1/2})$ for $j = 1, \cdots, 4$, the second equality is due to Lemma 5 and $\mathbb{P}_l\eta_i\hat{\phi}_{i,l}(\hat{q}^*_\tau) = (\mathbb{P}_l\eta_i)(\mathbb{P}_l\hat{\phi}_{i,l}(\hat{q}^*_\tau)) = 0$, the third equality is due to

Lemma 4, the fourth equality is due to Lemma 6 and the fact that $\sup_{\tau\in\Upsilon}|\hat{q}_\tau^* - q_\tau| = O_P(N^{-1/2})$, and the fifth equality holds because $\sup_{\tau\in\Upsilon}|(\mathbb{P}_{n,l}-\mathbb{P}_l)(\eta_i+1)\phi_i(q_\tau)| = o_P(N^{-1/2})$ and $\hat{n}_l/n = 1+o_P(1)$. For the second term on the RHS of (32), we have

$$\frac{1}{L}\sum_{l=1}^{L}\frac{n}{\hat{n}_l}\mathbb{P}_l\phi_i(\hat{q}_\tau^*) = \left(\frac{1}{L}\sum_{l=1}^{L}\frac{n}{\hat{n}_l}\right)(\mathbb{E}m_1(X,\hat{q}_\tau^*) - \mathbb{E}m_1(X,q_\tau))$$
$$= \frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X,q_\tau)}{f_Y(q_\tau)}\left[\sum_{i=1}^{N}\frac{(\eta_i+1)}{N}(\tau - 1\{Y_i\le q_\tau\})\right] + o_P(N^{-1/2}), \quad (33)$$

where the last equality is due to the delta method and (31). Combining (32) and (33), we have

$$\hat{\theta}^*(\tau)-\theta(\tau) = \frac{1}{N}\sum_{i=1}^{N}(\eta_i+1)\Bigg[m_1(X_i,q_\tau)-\theta(\tau)-\omega(X_i)(1\{Y_i\le q_\tau\}-m_0(X_i,q_\tau))$$
$$+\frac{\frac{\partial}{\partial q}\mathbb{E}m_1(X,q_\tau)}{f_Y(q_\tau)}(\tau-1\{Y_i\le q_\tau\})\Bigg] + R_N^*(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_N^*(\tau)| = o_P(N^{-1/2})$.

Third, we will derive the linear expansion of $\hat{f}_Y^*(\hat{q}_\tau^*)$. Let $\hat{N} = \sum_{i=1}^N(\eta_i+1)$. Note that $\hat{f}_Y^*(\hat{q}_\tau^*) - f_Y(q_\tau) = \frac{N}{\hat{N}}(\mathbb{P}_N-\mathbb{P})\frac{(1+\eta_i)}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau^*}{h_1}\right) + \frac{N}{\hat{N}}\left(\mathbb{P}\frac{1}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau^*}{h_1}\right) - f_Y(\hat{q}_\tau^*)\right) + \frac{N}{\hat{N}}\left(f_Y(\hat{q}_\tau^*)-f_Y(q_\tau)\right)$. Following the same argument in the proof in Section B.3.2 and the fact that $\left|\frac{N}{\hat{N}}-1\right| = O_P(N^{-1/2})$, we have

$$\frac{N}{\hat{N}}(\mathbb{P}_N-\mathbb{P})\frac{(1+\eta_i)}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau^*}{h_1}\right) = (\mathbb{P}_N-\mathbb{P})\frac{(1+\eta_i)}{h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right) + R_1^*(\tau),$$
$$\frac{N}{\hat{N}}\left(\mathbb{P}\frac{1}{h_1}K_1\left(\frac{Y_i-\hat{q}_\tau^*}{h_1}\right) - f_Y(\hat{q}_\tau^*)\right) = \frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2} + R_2^*(\tau),$$

and

$$\frac{N}{\hat{N}}\left(f_Y(\hat{q}_\tau^*)-f_Y(q_\tau)\right) = f_Y^{(1)}(q_\tau)(\hat{q}_\tau^*-q_\tau) + R_3^*(\tau)$$
$$= \frac{f_Y^{(1)}(q_\tau)}{f_Y(q_\tau)}\left[\frac{1}{N}\sum_{i=1}^{N}(\eta_i+1)(\tau-\mathbf{1}\{Y_i\le q_\tau\})\right] + R_4^*(\tau),$$

where $\sup_{\tau\in\Upsilon,j=1,\cdots,4}|R_j^*(\tau)| = o_P(N^{-1/2})$. This implies

$$\hat{f}_Y^*(\hat{q}_\tau^*) - f_Y(q_\tau) = (\mathbb{P}_N-\mathbb{P})(\eta_i+1)\left[\frac{1}{h_1}K_1\left(\frac{Y_i-q_\tau}{h_1}\right) + \frac{f_Y^{(1)}(q_\tau)}{f_Y(q_\tau)}(\tau-\mathbf{1}\{Y_i\le q_\tau\})\right]$$
$$+\frac{f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)}{2}h_1^2 + R_5^*(\tau),$$

where $\sup_{\tau\in\Upsilon}|R_5^*(\tau)| = o_P(N^{-1/2})$.

Last, we will derive the linear expansion of $\widehat{UQPE}^*(\tau)$. Based on the above arguments, we have

$$\begin{aligned}\widehat{UQPE}^*(\tau) - UQPE(\tau) =& -\frac{\hat{\theta}^*(\tau)-\theta(\tau)}{f_Y(q_\tau)} + \frac{\theta(\tau)(\hat{f}_Y^*(\hat{q}_\tau^*)-f_Y(q_\tau))}{f_Y^2(q_\tau)} \\ &+ \frac{(\hat{\theta}^*(\tau)-\theta(\tau))(\hat{f}_Y^*(\hat{q}_\tau^*)-f_Y(q_\tau))}{\hat{f}_Y^*(\hat{q}_\tau^*)f_Y(q_\tau)} - \frac{\theta(\tau)(\hat{f}_Y^*(\hat{q}_\tau^*)-f_Y(q_\tau))^2}{f_Y^2(q_\tau)\hat{f}_Y^*(\hat{q}_\tau^*)} \\ =& \frac{1}{N}\sum_{i=1}^N (1+\eta_i)\mathrm{IF}_i(\tau) + \frac{\theta(\tau)f_Y^{(2)}(q_\tau)(\int u^2K_1(u)du)h_1^2}{2f_Y^2(q_\tau)} + R_6^*(\tau)\end{aligned} \tag{34}$$

where $\sup_{\tau\in\Upsilon}|R_6^*(\tau)| = o_P(N^{-1/2})$. Taking difference between (30) and (34), we have $\widehat{UQPE}^*(\tau) - \widehat{UQPE}(\tau) = \frac{1}{N}\sum_{i=1}^N \eta_i\mathrm{IF}_i(\tau) + R^*(\tau)$, where $\sup_{\tau\in\Upsilon}|R^*(\tau)| = o_P(N^{-1/2})$.

B.4 Proof of Theorem 2

In the proof of Theorem 2, we use the notations

$$\mathcal{G}_l^{(0)} = \{\Lambda(b(X)^T\beta) : \beta\in\mathbb{R}^p, ||\beta||_0 \le Ms\},$$

and

$$\mathcal{G}_l^{(1)} = \left\{\Lambda(b(X)^T\beta)(1-\Lambda(b(X)^T\beta))\left(\frac{\partial}{\partial x_1}b(X)^T\beta\right) : \beta\in\mathbb{R}^p, ||\beta||_0\le Ms, \left\|\left(\frac{\partial}{\partial x_1}b(X)^T\beta\right)\right\|_{\mathbb{P},\infty} \le M,\right\},$$

where M is a sufficiently large constant.

Lemma 8. *Under the assumptions in Theorem 2, (i)* $\sup_{l\in\{1,\dots,L\},q\in\mathcal{Q}}||\hat{\beta}_{q,l}-\beta_q||_1 = O_P\left(\sqrt{\frac{s^2\log(p)}{N}}\right)$, *(ii)* $\sup_{l\in\{1,\dots,L\},q\in\mathcal{Q}}||\hat{m}_{0,l}(x,q)-m_0(x,q)||_{\mathbb{P},\infty} = O_P\left(\sqrt{\frac{\zeta_N^2 s^2\log(p)}{N}}\right)$, *(iii)* $\sup_{l\in\{1,\dots,L\},q\in\mathcal{Q}}||\hat{m}_{0,l}(x,q)-m_0(x,q)||_{\mathbb{P},2} = O_P\left(\sqrt{\frac{s\log(p)}{N}}\right)$, *(vi)* $\sup_{l\in\{1,\dots,L\},q\in\mathcal{Q}}||\hat{m}_{1,l}(x,q)-m_1(x,q)||_{\mathbb{P},\infty} = O_P\left(\sqrt{\frac{\zeta_N^2 s^2\log(p)}{N}}\right)$, *(v)* $\sup_{l\in\{1,\dots,L\},q\in\mathcal{Q}^\varepsilon}||\hat{m}_{1,l}(x,q)-m_1(x,q)||_{\mathbb{P},2} = O_P\left(\sqrt{\frac{s\log(p)}{N}}\right)$, *(vi)* $\sup_{l\in\{1,\dots,L\}}||\hat{\omega}_l(x)-\omega_l(x)||_{\mathbb{P},2} = O_P\left(\sqrt{\frac{s\log(p)}{N}}\right)$, *where, in all the statements, the norm in the left hand side is with respect to x and the stochastic convergence O_P in the right hand side is with respect to the randomness of the estimators.*

Proof. The first three results have been established by Belloni et al. (2017). For the fourth result, we have

$$
\begin{aligned}
&|\hat{m}_{1,l}(x,q) - m_1(x,q)| \\
\leq &\left|\Lambda(b(X)^T\hat{\beta}_{q,l})(1-\Lambda(b(X)^T\hat{\beta}_{q,l}))\frac{\partial}{\partial x_1}b(X)^T\hat{\beta}_{q,l} - \Lambda(b(X)^T\beta_q)(1-\Lambda(b(X)^T\beta_q))\frac{\partial}{\partial x_1}b(X)^T\beta_q\right| \\
&+\left|\frac{\partial}{\partial x_1}r_m(x,q)\right| \\
\leq &\left|\frac{\partial}{\partial x_1}b(X)^T(\hat{\beta}_{q,l}-\beta_q)\right| + \left|\Lambda(b(X)^T\hat{\beta}_{q,l}) - \Lambda(b(X)^T\beta_q)\right| + \left|\frac{\partial}{\partial x_1}r_m(x,q)\right| \\
\leq &\left\|\frac{\partial}{\partial x_1}b(X)\right\|_{L_\infty}||\hat{\beta}_{q,l}-\beta_q||_1 + |\hat{m}_{0,l}(x,q) - m_0(x,q)| + \left|\frac{\partial}{\partial x_1}r_m(x,q)\right|,
\end{aligned}
$$

where the first inequality is due to the triangle inequality and Assumption 4, and the second inequality is due to the facts that $\Lambda(\cdot)(1-\Lambda(\cdot))$ is bounded, $f(\lambda) = \lambda(1-\lambda)$ is Lipschitz-1 continuous in λ, and $\sup_{q\in\mathcal{Q}}||\frac{\partial}{\partial x_1}b(X)^T\beta_q||_{\mathbb{P},\infty} < \overline{c}$. Taking $\sup_{l\in\{1,\ldots,L\},q\in\mathcal{Q},x\in\mathrm{Supp}(X)}$ on both sides, we have $\sup_{l\in\{1,\ldots,L\},q\in\mathcal{Q}}||\hat{m}_{1,l}(x,q)-m_1(x,q)||_{L_\infty(\mathbb{P})} = O_P\left(\sqrt{\frac{\zeta_N^2 s^2\log(p)}{N}}\right)$.

Similarly,

$$
\begin{aligned}
&||\hat{m}_{1,l}(x,q) - m_1(x,q)||_{\mathbb{P},2} \\
\lesssim &||\frac{\partial}{\partial x_1}b(X)^T(\hat{\beta}_{q,l}-\beta_q)||_{\mathbb{P},2} + ||\Lambda(b(X)^T\hat{\beta}_{q,l}) - \Lambda(b(X)^T\beta_q)||_{\mathbb{P},2} + \left\|\frac{\partial}{\partial x_1}r_m(x,q)\right\|_{\mathbb{P},2} \\
\lesssim &(1+o_P(1))\left\|\frac{\partial}{\partial x_1}b(X)^T(\hat{\beta}_{q,l}-\beta_q)\right\|_{\mathbb{P}_n,2} + O_P\left(\sqrt{\frac{s\log(p)}{N}}\right) \\
\lesssim &(1+o_P(1))\overline{c}||\hat{\beta}_{q,l}-\beta_q||_2 + O_P\left(\sqrt{\frac{s\log(p)}{N}}\right) \\
= &O_P\left(\sqrt{\frac{s\log(p)}{N}}\right),
\end{aligned}
$$

where the second and third inequalities are due to Assumption 4.3.

For the last result, Belloni and Chernozhukov (2011) have established that

$$\sup_{l\in\{1,\ldots,L\}}||\hat{\sigma}_l(x_{-1}) - \sigma(x_{-1})||_{\mathbb{P},\infty} = O_P\left(\sqrt{\frac{\zeta_N^2 s^2\log(p)}{N}}\right), \tag{35}$$

$$\sup_{l\in\{1,\ldots,L\}}||\hat{\sigma}_l(x_{-1}) - \sigma(x_{-1})||_{\mathbb{P},2} = O_P\left(\sqrt{\frac{s\log(p)}{N}}\right), \tag{36}$$

$$\sup_{l\in\{1,\dots,L\}} ||\hat{\mu}_l(x_{-1})-\mu(x_{-1})||_{\mathbb{P},\infty} = O_P\left(\sqrt{\frac{\zeta_N^2 s^2\log(p)}{N}}\right), \tag{37}$$

and

$$\sup_{l\in\{1,\dots,L\}} ||\hat{\mu}_l(x_{-1})-\mu(x_{-1})||_{\mathbb{P},2} = O_P\left(\sqrt{\frac{s\log(p)}{N}}\right). \tag{38}$$

(Although Belloni and Chernozhukov (2011) focus on the high-dimensional quantile regression without approximation error, it is straightforward to extend their results to the case with approximation error as long as it is sufficiently small, as imposed in Assumption 4. Then, the general theory and proof techniques in Belloni et al. (2017) lead to (35)–(38).) Further note that $|\hat{\omega}_l(x)-\omega(x)| \le \left|\frac{\hat{\mu}_l(x_{-1})-\mu(x_{-1})}{\hat{\sigma}_l(x_{-1})}\right| + \left|\frac{(x_1-\mu(x_{-1}))(\hat{\sigma}_l(x_{-1})-\sigma(x_{-1}))}{\hat{\sigma}_l(x_{-1})\sigma(x_{-1})}\right|$. Since $\sigma(x_{-1}) \ge \underline{c} > 0$ for every $x_{-1} \in \text{Supp}(X_{-1})$, the equalities (35)–(38) imply

$$\begin{aligned}
||\hat{\omega}_l(x)-\omega(x)||_{\mathbb{P},2} \lesssim& ||\hat{\mu}_l(x_{-1})-\mu(x_{-1})||_{\mathbb{P},2} + \left[\mathbb{P}_l(X_1-\mu(X_{-1}))^2(\hat{\sigma}_l(X_{-1})-\sigma(X_{-1}))^2\right]^{1/2}\\
\lesssim& ||\hat{\mu}_l(x_{-1})-\mu(x_{-1})||_{\mathbb{P},2} + ||\sigma(X_{-1})(\hat{\sigma}_l(X_{-1})-\sigma(X_{-1}))||_{\mathbb{P},2}\\
=& O_P\left(\sqrt{\frac{s\log(p)}{N}}\right),
\end{aligned}$$

where the second inequality is due to the fact that $\mathbb{E}((X_1-\mu(X_{-1}))^2|X_{-1}) = \sigma^2(X_{-1})$. This complete a proof of the lemma. □

Now we will show (8)–(15) in Assumption 2. First, we will show (8). We note that Belloni et al. (2017) have shown $\sup_{q\in\mathcal{Q}} ||\hat{\beta}_{q,l}||_0 = O_P(s)$. In addition, by Lemma 8, $\sup_{q\in\mathcal{Q}} \left\|\frac{\partial}{\partial x_1} b(X)^T(\hat{\beta}_{q,l}-\beta_q)\right\|_{\mathbb{P},\infty} \le \zeta_N \sup_{q\in\mathcal{Q}} ||\hat{\beta}_{q,l}-\beta_q||_1 = o_P(1)$. This implies, w.p.a.1., $\sup_{q\in\mathcal{Q}} \left\|\frac{\partial}{\partial x_1} b(X)^T\hat{\beta}_{q,l}\right\|_{\mathbb{P},\infty} = O(1)$. Then, (8) directly follows the argument in the proof of Belloni et al. (2017, Theorem 5.1) with $\nu_N = s$ and $A_N = p$.

Second, Lemma 8 verifies (9) and (10) with the rate of convergence $\sqrt{\frac{s\log(p)}{N}}$.

Third, we will show (11). Note that $|\hat{\omega}_l(x)\hat{m}_{0,l}(x,q)-\omega(x)m_0(x,q)| \le |(\hat{\omega}_l(x)-\omega(x))||\hat{m}_{0,l}(x,q)| + |\omega(x)(\hat{m}_{0,l}(x,q)-m_0(x,q))|$. Then, we have

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}} \int |\hat{\omega}_l(x)\hat{m}_{0,l}(x,q)-\omega(x)m_0(x,q)|^2 dF_X(x)\\
\lesssim& \sup_{q\in\mathcal{Q}} \int (\hat{\omega}_l(x)-\omega(x))^2\hat{m}_{0,l}^2(x,q)dF_X(x) + \int \omega^2(x)(\hat{m}_{0,l}(x,q)-m_0(x,q))^2 dF_X(x)\\
\lesssim& \int (\hat{\omega}_l(x)-\omega(x))^2 dF_X(x)\\
&+ \left[\int \omega^{2+\delta}(x)dF_X(x)\right]^{2/(2+\delta)} \sup_{q\in\mathcal{Q}} \left[\int (\hat{m}_{0,l}(x,q)-m_0(x,q))^{2(2+\delta)/\delta} dF_X(x)\right]^{\delta/(2+\delta)}
\end{aligned}$$

$$=O_P\left(\zeta_N^{4/(2+\delta)}s^{(4+\delta)/(2+\delta)}\log(p)N^{-1}\right),$$

where the last inequality is due to the Hölder's inequality and the last equality holds due to the fact that $\int |U(x)|^{2(2+\delta)/\delta}dF_X(x) \leq \int |U(x)|^2 dF_X(x)||U(x)||_{\mathbb{P},\infty}^{4/\delta}$ and Lemma 8. Based on the above two steps, we can set $\pi_N = \sqrt{\zeta_N^{4/(2+\delta)}s^{(4+\delta)/(2+\delta)}\log(p)N^{-1}}$.

Fourth, (12) follows from the fact that $\sup_{q\in\mathcal{Q}}||\hat{m}_{1,l}(x,q) - m_1(x,q)||_{\mathbb{P},\infty} = o_P(1)$.

Fifth, (13) follows from Lemma 8, $\zeta_N^2 s^2\log(p) = o(N)$, $\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}|m_0(x,q)| \leq 1$, $\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}|\breve{m}_{0,l}(x,q) - \hat{m}_{0,l}(x,q)| = o_P(1)$, and $\mathbb{E}|\omega(X)|^{2+\delta} < \infty$.

Sixth, (14) follows from $\hat{m}_{1,l}(x,q) = \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q)$.

Seventh, we will show (15). By Lemma 8, we have

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}}\left|\int(\hat{\omega}_l(x)-\omega(x))(\hat{m}_{0,l}(x,q)-m_0(x,q))dF_X(x)\right| \\
\leq&\left(\int(\hat{\omega}_l(x)-\omega(x))^2dF_X(x)\right)^{1/2}\left(\sup_{q\in\mathcal{Q}}\int(\hat{m}_{0,l}(x,q)-m_0(x,q))^2dF_X(x)\right)^{1/2} \\
&= O_P\left(\frac{s\log(p)}{N}\right) = o_P(N^{-1/2}).
\end{aligned}$$

Lastly, (16) follows from $\zeta_N^{4/(2+\delta)}s^{1+(4+\delta)/(2+\delta)}\log^2(p) = o(N)$ and $s\log(p) = o(N^{\delta/(4+\delta)})$.

B.5 Proof of Theorem 3

We will show (8)–(15) in Assumption 2. First, we will show (8). To verify the first condition in Assumption 2, we note that

$$\begin{aligned}
&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\left|\frac{\partial}{\partial q}\hat{m}_{j,l}(x,q)\right| \\
=&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\left|\int\frac{\breve{m}_{j,l}(x,t)}{h_2^2}K_2^{(1)}\left(\frac{t-q}{h_2}\right)dt\right| \\
\leq&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\left|\int\frac{m_j(x,t)}{h_2}dK_2\left(\frac{t-q}{h_2}\right)\right| + \sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}\frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}\int d|K_2(u)| \\
\leq&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\left|\int\frac{\frac{\partial}{\partial q}m_j(x,t)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right| + \sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}\frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}\int d|K_2(u)| \\
\leq&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}|\frac{\partial}{\partial q}m_j(x,q)|\int|K_2(u)|du + \sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}\frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}\int d|K_2(u)| \\
<&\infty,
\end{aligned}$$

where the first inequality is due to the triangle inequality, the second equality is due to the integration by parts and the fact that the kernel function $K_2(\cdot)$ vanishes at the boundary, and the last inequality is due to the facts that $\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}|\frac{\partial}{\partial q}m_j(x,q)|$ is bounded and that $\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}\frac{|\breve{m}_{j,l}(x,q)-m_j(x,q)|}{h_2}=O_P(\rho_N)=o_P(1)$. Given the derivative $\frac{\partial}{\partial q}\hat{m}_{j,l}(x,q)$ is uniformly bounded w.p.a.1, there exists a constant M such that $|\hat{m}_{j,l}(x,q_1)-\hat{m}_{j,l}(x,q_2)|\leq M|q_1-q_2|$. The class of Lipschitz continuous functions is a VC-class with a fixed VC-index. This implies $\mu_N=A_N=1$.

Second, (9) follows from

$$\begin{aligned}
&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}|\breve{m}_{j,l}(x,q)-\hat{m}_{j,l}(x,q)|\\
=&\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\left|\int\frac{\breve{m}_{j,l}(x,t)-\breve{m}_{j,l}(x,q)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right|\\
\leq&\,2\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\int\frac{\sup_{t\in\mathcal{Q}^\epsilon}|\breve{m}_{j,l}(x,t)-m_j(x,t)|}{h_2}\left|K_2\left(\frac{t-q}{h_2}\right)\right|dt\\
&+\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}}\left|\int\frac{m_j(x,t)-m_j(x,q)}{h_2}K_2\left(\frac{t-q}{h_2}\right)dt\right|\\
\leq&\,2\sup_{x\in\text{Supp}(X),t\in\mathcal{Q}^\varepsilon}|\breve{m}_{j,l}(x,t)-m_j(x,t)|\int|K_2(u)|du\\
&+\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}|\frac{\partial^{2k}}{\partial q^{2k}}m_j(x,q)|h_2^{2k}\int u^{2k}|K_2(u)|du\\
=&\,O_P(h_2\rho_N+h_2^{2k}),
\end{aligned}$$

where the last inequality holds because of (22) and the fact that $\sup_{x\in\text{Supp}(X),q\in\mathcal{Q}^\varepsilon}|\frac{\partial^{2k}}{\partial q^{2k}}m_j(x,q)|<\infty$.

Therefore,

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}}\int|\hat{m}_{1,l}(x,q)-m_1(x,q)|^2dF_X(x)\\
\lesssim&\sup_{q\in\mathcal{Q}}\int|\hat{m}_{1,l}(x,q)-\breve{m}_1(x,q)|^2dF_X(x)+\sup_{q\in\mathcal{Q}}\int|\breve{m}_{1,l}(x,q)-m_1(x,q)|^2dF_X(x)\\
=&O_P(\rho_N^2h_2^2+h_2^{4k}).
\end{aligned}$$

Third, (10) is the same as (24).

Fourth, we will show (11). Note that

$$|\hat{\omega}_l(x)\hat{m}_{0,l}(x,q)-\omega(x)m_0(x,q)|\leq|(\hat{\omega}_l(x)-\omega(x))||\hat{m}_{0,l}(x,q)|+|\omega(x)(\hat{m}_{0,l}(x,q)-m_0(x,q))|.$$

Then, we have

$$\begin{aligned}
&\sup_{q\in\mathcal{Q}}\int|\hat{\omega}_l(x)\hat{m}_{0,l}(x,q)-\omega(x)m_0(x,q)|^2dF_X(x)\\
\lesssim&\sup_{q\in\mathcal{Q}}\int(\hat{\omega}_l(x)-\omega(x))^2\hat{m}_{0,l}^2(x,q)dF_X(x)+\int\omega^2(x)(\hat{m}_{0,l}(x,q)-m_0(x,q))^2dF_X(x)
\end{aligned}$$

$$\lesssim \int (\hat{\omega}_l(x) - \omega(x))^2 dF_X(x) + \int \omega^2(x) dF_X(x) \sup_{q\in\mathcal{Q}} ||\hat{m}_{0,l}(x,q) - m_0(x,q)||^2_{\mathbb{P},\infty}$$
$$= O_P(\rho_N^2 h_2^2 + h_2^{4k}),$$

where the last equality holds due to the fact that

$$\sup_{q\in\mathcal{Q}} ||\hat{m}_{0,l}(x,q) - m_0(x,q)||_{\mathbb{P},\infty} \le \sup_{q\in\mathcal{Q}} ||\hat{m}_{0,l}(x,q) - \breve{m}_{0,l}(x,q)||_{\mathbb{P},\infty} + \sup_{q\in\mathcal{Q}} ||\breve{m}_{0,l}(x,q) - m_0(x,q)||_{\mathbb{P},\infty}$$
$$= O_P(h_2\rho_N + h_2^{2k}).$$

Fifth, (12) holds because $\sup_{x\in\mathrm{Supp}(X),q\in\mathcal{Q}} |m_j(x,q)|$ is bounded for $j = 0,1$ and $\sup_{x\in\mathrm{Supp}(X),q\in\mathcal{Q}} |\breve{m}_{1,l}(x,q) - \hat{m}_{1,l}(x,q)| = o_P(1)$.

Sixth, (13) follows from $\sup_{x\in\mathrm{Supp}(X),q\in\mathcal{Q}} |m_0(x,q)| \le 1$, $\sup_{x\in\mathrm{Supp}(X),q\in\mathcal{Q}} |\breve{m}_{0,l}(x,q) - \hat{m}_{0,l}(x,q)| = o_P(1)$, and $\mathbb{E}|\omega(X)|^{2+\delta} < \infty$.

Seventh, (14) holds because

$$\sup_{q\in\mathcal{Q}} \left| \int \left(\hat{m}_{1,l}(x,q) - \frac{\partial}{\partial x_1}\hat{m}_{0,l}(x,q) \right) dF_X(x) \right|$$
$$\le \sup_{q\in\mathcal{Q}} \int \frac{1}{h_2} \sup_{t\in\mathcal{Q}} \left| \int \left(\breve{m}_{1,l}(x,t) - \frac{\partial}{\partial x_1}\breve{m}_{0,l}(x,t) \right) dF_X(x) \right| \left| K_2\left(\frac{q-t}{h_2}\right) \right| dt$$
$$= o_P(N^{-1/2}).$$

Eighth, we will show (15). Note that

$$\left| \int (\hat{\omega}_l(x) - \omega(x))(\hat{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \right|$$
$$\le \left| \int (\hat{\omega}_l(x) - \omega(x))(\breve{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \right|$$
$$+ \int \frac{\left| \int (\hat{\omega}_l(x) - \omega(x))(\breve{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \right|}{h_2} \left| K_2\left(\frac{t-q}{h_2}\right) \right| dt$$
$$+ \int \frac{\left| \int (\hat{\omega}_l(x) - \omega(x))(\breve{m}_{0,l}(x,t) - m_0(x,t)) dF_X(x) \right|}{h_2} \left| K_2\left(\frac{t-q}{h_2}\right) \right| dt$$
$$+ \int |\hat{\omega}_l(x) - \omega(x)| \left| \int \frac{m_0(x,t) - m_0(x,q)}{h_2} K_2\left(\frac{t-q}{h_2}\right) dt \right| dF_X(x). \tag{39}$$

By Assumption 6, we have

$$\sup_{q\in\mathcal{Q}} \left| \int (\hat{\omega}_l(x) - \omega(x))(\breve{m}_{0,l}(x,q) - m_0(x,q)) dF_X(x) \right| = o_P(N^{-1/2}).$$

For the second term on the RHS of (39), we have, by (28),

$$\sup_{q\in\mathcal{Q}} \int \frac{\left| \int (\hat{\omega}_l(x) - \omega(x))(\breve{m}_{0,l}(x,t) - m_0(x,t)) dF_X(x) \right|}{h_2} K_2\left(\frac{t-q}{h_2}\right) dt = o_P(N^{-1/2}).$$

Similarly, we can show the third term is $o_P(N^{-1/2})$ uniformly over $q \in \mathcal{Q}$ as well. For the fourth term on the RHS of (39), we have

$$\int |\hat{\omega}_l(x) - \omega(x)| \left| \int \frac{m_0(x,t) - m_0(x,q)}{h_2} K_2\left(\frac{t-q}{h_2}\right) dt \right| dF_X(x)$$
$$\leq \int |\hat{\omega}_l(x) - \omega(x)| \left| \frac{\partial^{2k}}{\partial q^{2k}} m_0(x,\tilde{q}) \right| h^{2k} dF_X(x) \int u^{2k} |K_2(u)| du = o_P(N^{-1/2}),$$

where we use the fact that $\sup_{x \in \mathrm{Supp}(X), q \in \mathcal{Q}^{\varepsilon}} |\frac{\partial^{2k}}{\partial q^{2k}} m_0(x,q)| < \infty$, $h_2^{2k} = O(N^{\frac{-k}{2k+1}})$, and

$$\int |\hat{\omega}_l(x) - \omega(x)| \, dF_X(x) \leq ||\hat{\omega}_l(x) - \omega(x)||_{\mathbb{P},2} = O_P(h_2 \rho_N) = o_P(N^{\frac{-1}{2(2k+1)}}).$$

Lastly, (16) follows from the choice of (ν_N, A_N, π_N).