Estimation of Censored Quantile Regression for Panel Data with Fixed Effects*

Antonio F. Galvao† Carlos Lamarche‡ Luiz Renato Lima§

June 10, 2013

Abstract

This paper investigates estimation of censored quantile regression models with fixed effects. Standard available methods are not appropriate for estimation of a censored quantile regression model with a large number of parameters or with covariates correlated with unobserved individual heterogeneity. Motivated by these limitations, the paper proposes estimators that are obtained by applying fixed effects quantile regression to subsets of observations selected either parametrically or nonparametrically. We derive the limiting distribution of the new estimators under joint limits, and conduct Monte Carlo simulations to assess their small sample performance. An empirical application of the method to study the impact of the 1964 Civil Rights Act on the black-white earnings gap is considered.

Key Words: Quantile Regression; Panel Data; Censored; Civil Rights; Earnings Gap

*We are grateful to Ivan Canay, Joel Horowitz, Roger Koenker, Stephen Portnoy, Michael Price, Liang Wang, and seminar participants at the University of Illinois at Urbana-Champaign, Northwestern University, University of Kentucky, University of Tennessee, NY Camp Econometrics VIII, and the 21st meeting of the Midwest Econometric Group for useful comments and discussion regarding this paper. We also would like to thank the editor, the associate editor, and three anonymous referees for their careful reading and comments to improve the manuscript. We thank Marcelo Moreira for providing the data for the empirical section. Luiz Lima thanks financial support from CNPq-Brazil. The `R` software for the method introduced in this paper is available upon request.

†University of Iowa, Department of Economics, W210 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242. E-mail: antonio-galvao@uiowa.edu

‡University of Kentucky, Department of Economics, 335A Gatton College of Business and Economics, Lexington, KY 40506-0034. Email: clamarche@uky.edu

§University of Tennessee, Department of Economics, 527A Stokely Management Center, Knoxville, Tennessee, 37996; and Federal University of Paraiba-Brazil. Email: llima@utk.edu

1 Introduction

Censored observations are common in applied work. Standard examples are survey data on wealth and income. In order to obtain responses from wealthy individuals or households, some surveys only ask about the amount of wealth up to a given threshold, allowing wealthy individuals to simply indicate if their wealth is above a threshold. Due to the presence of censoring, standard regression methods employed to estimate linear conditional mean models lead to inconsistent estimates of the parameters of interest. Censored regression models are usually estimated using likelihood techniques. Schnedler (2005) shows the general validity of this approach and provides methods to find the likelihood for a broad class of applications.

When the interest lies on the effect of a given covariate on the location and scale parameters of the conditional distribution of a latent variable, likelihood methods can be replaced by quantile regression (QR) techniques. In this context, Powell (1984, 1986) proposed the celebrated Powell estimator based on the equivariance to monotone transformation property of quantiles. Despite its intuitive appeal, the slow convergence of the Powell estimator when the degree of censoring is high or when the number of estimated parameters is large limited the use of the method in empirical research. Motivated in part by such limitations, Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012) proposed simple, easily implementable, and well-behaved estimation procedures.

Recently, there has been a growing literature on estimation and testing of QR panel data models. Koenker (2004) introduced a general approach to estimation of QR models for longitudinal data. Individual specific (fixed) effects are treated as pure location shift parameters common to all conditional quantiles. Controlling for individual specific heterogeneity while exploring heterogeneous covariate effects within the QR framework, offers a more flexible approach to the analysis of panel data than the afforded by the classical Gaussian fixed and random effects estimators. In spite of the large literature on censored QR for cross-sectional models [see, e.g., Powell (1986), Fitzenberger (1997), Buchinsky and Hahn (1998), Bilias, Chen, and Ying (2000), Khan and Powell (2001), Honoré, Khan, and Powell (2002), Chernozhukov and Hong (2002), Portnoy (2003), Peng and Huang (2008), Lin, He, and Portnoy (2012), Tang, Wang, He, and Zhu (2012)], the literature on censored QR for panel data is still very limited. Honoré (1992) proposes estimators for trimmed least absolute deviation censored models with individual fixed effects, which do not parametrically specify the distribution of the error term. Chen and Khan (2008) consider an estimation procedure for median censored regression models that is robust to non-stationary errors in the longitudinal data context. Wang and Fygenson (2009) develop inference procedures for a QR panel data model, while accounting for issues associated with censoring and intra-subject correlation. More recently, Khan, Ponomareva, and Tamer (2011) analyze identification in a censored panel data model where the censoring can depend on both observable and unobservable variables in arbitrary ways under a median independence assumption.

This paper investigates estimation of a panel data censored QR model with fixed effects. In the analysis of panel data, it is natural to treat the individual effects as nuisance parameters in the model. Although one could estimate this model using the Powell estimator, it is well known that this estimation method suffers from computational instability [see, e.g., Buchinsky (1994) and Fitzenberger (1997)]. The Powell estimator simply does not perform well when the number of estimated parameters is large and the degree of censoring is high. Motivated by these limitations, the paper proposes estimators that are obtained by applying fixed effects QR to subsets of observations. We propose two-step estimators in which the first step selects a subset of observations by estimating a propensity score either parametrically or nonparametrically, and the second step applies fixed effects QR to the selected subset of observations. These estimators are simple to compute and easy to be implemented in panel data applications with a large number of subjects. We derive their asymptotic properties under joint limits, assuming that the conditional censoring probability satisfies smoothness conditions and can be estimated at an appropriate nonparametric rate. Finally, we suggest an alternative parametric estimator which can be employed in models with polychotomous independent variables, although it comes at the cost of employing additional steps.

Monte Carlo simulations are conducted to evaluate the finite sample performance of the proposed methods. The simulations indicate that the estimators offer good performance in terms of bias, mean squared error, and coverage probability of the confidence interval. We also consider an empirical application to investigate whether the 1964 Civil Rights Act contributed to reduce the black-white earning gap. Our approach shows that the policy had a small and insignificant effect among mature workers, while significantly reducing the earning gap among young workers at the upper quantiles. This finding is not uncovered by other competing methods that fail to simultaneously address censoring at the maximum taxable earnings and unobserved heterogeneity.

The paper is organized as follows. Section 2 presents the model, the estimators, and the large sample theory. Section 3 investigates the small sample performance of the methods. Section 4 extends the asymptotic results to allow for dependence across time. An empirical illustration is considered in Section 5. Section 6 concludes. The technical proofs are in the Appendix.

2 Censored quantile regression with fixed effects

2.1 The model

Let y_{it}^* denote the potentially left censored t-th response of the i-th individual and let $y_{it} = \max(C_{it}, y_{it}^*)$ be its corresponding observed value, where C_{it} is a known censoring point. Moreover, y_{it}^* is assumed to be conditionally independent of the censoring point C_{it}, such that, conditional on covariates, $\boldsymbol{x}_{it}$, and an individual effect, α_i, $\mathrm{P}(y_{it}^* < y|\boldsymbol{x}_{it}, \alpha_i, C_{it}) = \mathrm{P}(y_{it}^* < y|\boldsymbol{x}_{it}, \alpha_i)$. Given a quantile $\tau \in (0,1)$, we define the following quantile regression (QR) model,

$$y_{it}^* = \alpha_{i0}(\tau) + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0(\tau) + u_{it}, \quad i = 1, ...N, \; t = 1, ...T,$$

where $\boldsymbol{x}_{it}$ is a $p \times 1$ vector of regressors, $\boldsymbol{\beta}_0(\tau)$ is a $p \times 1$ vector of parameters, $\alpha_{i0}(\tau)$ is a scalar individual effect for each i, and u_{it} is the innovation term whose τ-th conditional quantile is zero. $Q_{y_{it}^*}(\tau|\boldsymbol{x}_{it}, \alpha_{i0}) = \inf\{y : \Pr(y_{it}^* < y|\boldsymbol{x}_{it}, \alpha_{i0}) \geq \tau\}$ is the conditional τ-quantile of y_{it}^* given $(\boldsymbol{x}_{it}, \alpha_{i0})$. The quantile-specific individual effect, $\alpha_{i0}(\tau)$, is intended to capture individual specific sources of variability, or unobserved heterogeneity that was not adequately controlled by other covariates. In general, each $\alpha_{i0}(\tau)$ and $\boldsymbol{\beta}_0(\tau)$ can depend on τ, but we assume τ to be fixed throughout the paper and suppress this dependence for notational simplicity. The model is semiparametric in the sense that the functional form of the conditional distribution of y_{it}^* given $(\boldsymbol{x}_{it}, \alpha_{i0})$ is left unspecified and no parametric assumption is made on the relation between $\boldsymbol{x}_{it}$ and α_{i0}. Thus, the QR model can be written as,

$$Q_{y_{it}^*}(\tau|\boldsymbol{x}_{it}, \alpha_{i0}) = \alpha_{i0} + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0. \tag{2.1}$$

Equivariance to monotone transformation is an important property of QR models. For a given monotone transformation $\Im_c(y)$ of variable y^*, $Q_{\Im_c(y_{it}^*)}(\tau|\boldsymbol{x}_{it}, \alpha_{i0}) = \Im_c(Q_{y_{it}^*}(\tau|\boldsymbol{x}_{it}, \alpha_{i0}))$. The transformation of (2.1) naturally leads to a version of the Powell's censored QR model,

$$Q_{y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_{i0}, C_{it}) = \max(C_{it}, \alpha_{i0} + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0). \tag{2.2}$$

We consider the fixed effects estimation of $\boldsymbol{\beta}_0$, which is implemented by treating each individual effect as a parameter to be estimated. Throughout the paper, as in Hahn and Newey (2004) and Fernández-Val (2005), we treat α_i as fixed by conditioning on them. Thus, the parameter of interest, $\boldsymbol{\beta}_0$, can be interpreted as representing the effect of $\boldsymbol{x}_{it}$ on the τ-th conditional quantile function of the dependent variable while controlling for heterogeneity, here represented by α_i. This model can be viewed as a conditional model. There are other conditional models available in the QR literature and we refer the reader to Kim and Yang (2011) for additional discussion on marginal and conditional quantile regression models. Following Powell (1986), if we control for fixed effects we could define the estimator $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ solving the following minimization problem:

$$Q_{1,N}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - \max(C_{it}, \alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta})), \tag{2.3}$$

where $\boldsymbol{\alpha} := (\alpha_1, ..., \alpha_N)^\top$ and $\rho_\tau(u) = u(\tau - 1(u < 0))$ denotes the loss function of Koenker and Bassett (1978). Throughout the paper, the number of individuals is denoted by N and the number of time periods is denoted by $T = T_N$ that depends on N. In what follows, we omit the subscript N of T_N. Hence, only N is explicitly shown in $Q_{1,N}(\boldsymbol{\alpha}, \boldsymbol{\beta})$.

In despite of its intuitive appeal, the Powell estimator has not become popular in empirical research due to its computational difficulty. The problem with estimating (2.3) is caused by its low frequency of convergence. The Powell estimator involves the minimization of a non-convex problem, and thus iterative linear programming methods are only guaranteed to converge to local minimum [see Fitzenberger (1997) and Khan and Powell (2001)]. Additional regressors, large proportions of censored observations, and large samples only worsen the problem. Furthermore, its finite sample performance has come into question, and has been addressed in simulation studies [see, e.g., Paarsch (1984)].

In order to overcome the above problems, we consider an alternative approach to estimate model (2.2). Following the arguments in Honoré, Khan, and Powell (2002) and Tang, Wang, He, and Zhu (2012), it can be shown that (2.3) is asymptotically equivalent to the minimizer of

$$Q_{2,N}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}^\top \boldsymbol{\beta}) 1(\alpha_{i0} + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0 > C_{it}). \tag{2.4}$$

Denote $\delta_{it} = 1(y_{it}^* > C_{it})$ to indicate uncensored observations. Let $u_{it} := y_{it}^* - \alpha_{i0} - \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0$, whose τ-th conditional quantile given $(\boldsymbol{x}_{it}, \alpha_i, C_{it})$ equals zero. Because $\pi_0(\alpha_i, \boldsymbol{x}_{it}, C_{it}) :=$

$\mathrm{P}(\delta_{it}=1|\boldsymbol{x}_{it},\alpha_i,C_{it})=\mathrm{P}(u_{it}>-\alpha_{i0}-\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_0+C_{it}|\boldsymbol{x}_{it},\alpha_i,C_{it})$ and $\mathrm{P}(u_{it}>0|\boldsymbol{x}_{it},\alpha_i,C_{it})=1-\tau$, and noticing that the restriction set selects those observations (i,t) where the conditional quantile line is above the censoring point C_{it}, the objective function (2.4) is equivalent to

$$Q_{3,N}(\boldsymbol{\alpha},\boldsymbol{\beta})=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(y_{it}-\alpha_i-\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta})1(\pi_0(\alpha_i,\boldsymbol{x}_{it},C_{it})>1-\tau). \tag{2.5}$$

This development suggests that to obtain a censored QR estimator, one can simply apply fixed effects QR to the subset $\{(i,t):\pi_0(\alpha_i,\boldsymbol{x}_{it},C_{it})>1-\tau\}$, including all the observations, even censored ones, for which the true τ-th conditional quantile is above the censoring point C_{it}. However, in applications the true propensity score function $\pi_0(\cdot)$ is unknown. Thus, a (feasible) estimator would first estimate $\pi_0(\cdot)$, only using the values of δ_{it} and regressors. From this step, the fitted function would be used to determine the observations to be included in a panel data QR.

2.2 The proposed methods

This section describes the proposed estimator for censored QR panel data. The estimator can be obtained in two steps. In what follows, we extend the results in Tang, Wang, He, and Zhu (2012) for the problem considered in this paper.

Step 1. Estimate $\pi_0(\alpha_i,\boldsymbol{x}_{it},C_{it})$ by using either a parametric or nonparametric regression method for binary data, and denote the estimated conditional probability as $\hat{\pi}(\alpha_i,\boldsymbol{x}_{it},C_{it})$. Determine the informative subset $J_T=\{(i,t):\hat{\pi}(\alpha_i,\boldsymbol{x}_{it},C_{it})>1-\tau+c_N\}$, where c_N is a pre-specified small positive value with $c_N\to 0$ as $N\to\infty$.

Step 2. Then $\boldsymbol{\theta}_0=(\boldsymbol{\alpha}_0^{\top},\boldsymbol{\beta}_0^{\top})^{\top}$ can be estimated by applying fixed effects QR to the subset J_T, i.e., $\hat{\boldsymbol{\theta}}=(\hat{\boldsymbol{\alpha}}^{\top},\hat{\boldsymbol{\beta}}^{\top})^{\top}$ is the minimizer of

$$Q_N(\boldsymbol{\alpha},\boldsymbol{\beta},\hat{\pi})=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(y_{it}-\boldsymbol{z}_{it}^{\top}\boldsymbol{\alpha}-\boldsymbol{x}_{it}^{\top}\boldsymbol{\beta})1\left(\hat{\pi}(\alpha_i,\boldsymbol{x}_{it},C_{it})>1-\tau+c_N\right), \tag{2.6}$$

where $\boldsymbol{z}_{it}$ is an N-dimensional indicator variable for the individual effect α_i.

Here c_N is added to exclude boundary cases from the subset used in (2.6). The rate of c_N required for establishing asymptotic properties is given and discussed in assumption B3 below. The practical choice of c_N is discussed in the simulation section. If the parametric form of $\pi_0(\alpha_i,\boldsymbol{x}_{it},C_{it})$ is known, a consistent estimator of $\hat{\pi}(\alpha_i,\boldsymbol{x}_{it},C_{it})$ can be obtained by applying parametric methods to data. For instance, standard probability methods can be applied under the assumption that the censoring probability follows a parametric classification fixed effects model, $\mathrm{P}(\delta_{it} = 1 | \boldsymbol{X}_{it}, C_{it}) = p(\boldsymbol{X}_{it}^{\top}\boldsymbol{\gamma})$, where $\delta_{it} = 1(y_{it}^{*} > C_{it})$, $p(\cdot)$ is a known link function, $\boldsymbol{X}_{it} = (\boldsymbol{z}_{it}^{\top}, \boldsymbol{x}_{it}^{\top})^{\top}$, and $\boldsymbol{\gamma} = (\boldsymbol{\alpha}^{\top}, \boldsymbol{\beta}^{\top})^{\top}$ is a $(N+p)$-dimensional vector. When the parametric form of the true propensity score is unknown, then one can obtain a consistent estimator of $\pi_0(\cdot)$ by applying nonparametric or semiparametric methods (e.g., generalized linear regression with spline approximation, generalized additive models, or maximum score with series function approximation). We refer the reader to Li and Racine (2007) for estimation of binary dependent variable panel models. In the next section, the asymptotic properties of the 2-step estimator are derived assuming that the estimated censoring probability, $\hat{\pi}(\cdot)$, satisfies some smoothness conditions and converges to $\pi_0(\cdot)$ at the uniform rate of $T^{-1/4}$. Thus, we allow both parametric and nonparametric estimation of $\pi_0(\cdot)$.

Although the suggested nonparametric methods for the first stage are attractive, they are practical only in low dimensions, have slow convergence rates, might not allow for categorical data, and rely on additive separability on the fixed effects or nonlinear difference techniques (see e.g. Chernozhukov, Fernández-Val, Hahn, and Newey (2013) for nonseparable panel models). Local kernel smoothers apply to (sufficiently) continuous variables only, whereas many practical applications have many (sufficiently) discrete covariates. To overcome these shortcomings, in a cross-sectional context, Chernozhukov and Hong (2002) use parametric regression to estimate the conditional censoring probability, which may give inconsistent estimation of $\pi_0(\cdot)$. They assume an envelope restriction on the censoring probability, requiring that the misspecification of the parametric model is not severe, use a fixed constant d to avoid bias, and seek a further step to achieve efficiency. Thus, for the mentioned reasons, we also consider a 3-step estimator for censored QR model.

The 3-step estimator has the advantage that it allows for some misspecification in the propensity score, can be used in models that are nonseparable in α_i and $\boldsymbol{x}_{it}$, and is simple to be implemented in practice, demanding shorter T relative to the 2-step estimator. The estimator is computed using the following steps. The first step selects the sample $J_0 = \big\{(i,t) : p(\boldsymbol{X}_{it}^{\top}\hat{\boldsymbol{\gamma}}) > 1 - \tau + d\big\}$, where d is strictly between 0 and τ and $p(\cdot)$ is a parametric link function, for instance a logit function. The goal of the first step is to select some, and not necessarily the largest, subset of observations where $\pi_0(\alpha_i, \boldsymbol{x}_{it}, C_{it}) > 1 - \tau$, that is, where the quantile line $\alpha_{i0} + \boldsymbol{x}_{it}^{\top}\boldsymbol{\beta}_0$ is above the censoring point C_{it}. The second step applies fixed effects QR to the subset J_0, selecting the subset $J_1 = \left\{(i,t) : \hat{\alpha}_i^0 + \boldsymbol{x}_{it}^\top \hat{\boldsymbol{\beta}}^0 > \delta_{NT} + C_{it}\right\}$, where δ_{NT} is a small positive number such that $\delta_{NT} \downarrow 0$ and $\sqrt{NT} \times \delta_{NT}$ is bounded, and $\hat{\boldsymbol{\theta}}^0 = (\hat{\boldsymbol{\alpha}}^{0\top}, \hat{\boldsymbol{\beta}}^{0\top})^\top$ is the second stage estimator of $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\beta}_0^\top)^\top$. Lastly, we seek a third step by solving a fixed effects QR problem on the subset J_1 if $J_0 \subset J_1$. We denote this estimator as the 3-step estimator, $\hat{\boldsymbol{\theta}}^1 = (\hat{\boldsymbol{\alpha}}^{1\top}, \hat{\boldsymbol{\beta}}^{1\top})^\top$.

Naturally, the 2-step and 3-step procedures have advantages and disadvantages when applied to panel data. On the one hand, in practice, the parametric form of the propensity score is unknown and estimation might be subject to misspecification. On the other hand, the nonparametric 2-step estimator requires relatively larger T and additional assumptions that control the degree of smoothness, and the 3-step estimator requires additional assumptions and parametric estimation in the first step. We investigate the finite sample properties of the estimators in the simulation section.

2.3 Large sample properties

This section investigates the asymptotic properties of the proposed 2-step estimator. The asymptotic results together with the required assumptions for the 3-step estimator are provided in the Supplemental Appendix. While the 2-step framework is similar to the one proposed by Tang, Wang, He, and Zhu (2012), which has been developed for cross-sectional models, the existence of the individual fixed effects parameter, $\boldsymbol{\alpha}$, in equation (2.6), whose dimension N tends to infinity, raises some new issues for the asymptotic analysis of the quantile regression (QR) estimators. As first noted by Neyman and Scott (1948), leaving the individual heterogeneity unrestricted in a nonlinear or dynamic panel model generally results in inconsistent estimators of the common parameters due to the incidental parameters problem; that is, noise in the estimation of the individual specific effects leads to inconsistent estimates of the common parameters due to the nonlinearity of the problem. In this respect, QR panel data suffers from this problem. To overcome this drawback, it has become standard in the panel QR literature, to employ a large N and T asymptotics, as for example in Koenker (2004) and Kato, Galvao, and Montes-Rojas (2012). The latter work derives the asymptotic properties of the panel QR estimator under joint limits. We employ the same joint asymptotics in this paper. Following Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012), we set in this section $C_{it} = 0$ since any model with known censoring C_{it} can be reduced to a model with a fixed censoring at 0.

Let $\|\pi - \pi_0\|_\infty = \sup_w |\pi(\boldsymbol{w}) - \pi_0(\boldsymbol{w})|$ for a given $\pi(\cdot)$ and a generic vector $\boldsymbol{w}$. We consider the following regularity conditions for consistency of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$.

A1: $\{(\boldsymbol{x}_{it}, y_{it}^*)\}$ are independent across subjects and independently and identically distributed (i.i.d.) for each i and all $t \geq 1$.

A2: $\sup_{i \geq 1} \mathrm{E}[||\boldsymbol{x}_{i1}||^{2s}] < \infty$ and some real $s \geq 1$.

Let $u_{it} := y_{it}^* - \alpha_{i0} - \boldsymbol{x}_{it}^\top \boldsymbol{\beta}_0$, and $\pi_{i0}(\boldsymbol{x}_{it}) := \pi_0(\alpha_i, \boldsymbol{x}_{it})$. $F_i(u|\boldsymbol{x})$ is defined as the conditional distribution function of u_{it} given $\boldsymbol{x}_{it} := \boldsymbol{x}$. Assume that $F_i(u|\boldsymbol{x})$ has density $f_i(u|\boldsymbol{x})$. Let $f_i(u)$ denote the marginal density of u_{it}.

A3: For each $\delta > 0$,

$$\epsilon_\delta := \inf_{i \geq 1} \inf_{|\alpha| + ||\boldsymbol{\beta}||_1 = \delta} \mathrm{E}\left[\int_0^{\alpha + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}} \{F_i(s|\boldsymbol{x}_{i1}) - \tau\}\, ds 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau\}\right] > 0.$$

A4: For any $\epsilon_N \to 0$, $\sup_{||\pi - \pi_0||_\infty \leq \epsilon_N} \frac{1}{N} \sum_{i=1}^N \mathrm{E}[1\{|\pi_i(\boldsymbol{x}_{it}) - (1 - \tau + c_N)| < \epsilon_N\}] = O(\epsilon_N)$.

A5: $\lim_{N \to \infty} c_N = 0$.

Conditions A1 and A2 are standard in the QR panel literature, and are the same as the ones used in Kato, Galvao, and Montes-Rojas (2012). In condition A1, we exclude the temporal dependence to focus on the simplest case first and to highlight the difficulties arising from panel data models with fixed effects and censored observations. The temporal independence is also assumed in Hahn and Newey (2004), Fernández-Val (2005), and Canay (2011). Nevertheless, the results are extended in Section 4 to the dependent case under suitable mixing conditions as in Hahn and Kuersteiner (2011). Condition A3 represents an identification condition, and corresponds to condition A3 of Kato, Galvao, and Montes-Rojas (2012). Condition A4 is the same as assumption A5.3 in Tang, Wang, He, and Zhu (2012) and requires that $\pi_0(\cdot)$ is nonflat around $1 - \tau$. This is standard in the literature with two step estimators. Finally, A5 is required for establishing consistency and is a restriction on c_N which serves to avoid boundary situations. This condition is largely employed in the literature on censored QR (e.g. Buchinsky and Hahn (1998), Khan and Powell (2001), Chernozhukov and Hong (2002), Tang, Wang, He, and Zhu (2012)). Now we state the result for consistency.

Theorem 1. *Assume* $\sup_i ||\hat{\pi}_i - \pi_{i0}||_\infty = o_p(1)$. *Under Assumptions A1–A5, as* $N/T^s \to 0$, $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *is consistent.*

The result in Theorem 1 shows that the 2-step estimator is consistent. The condition on T in Theorem 1 is the same as that in Theorems 1-2 of Fernández-Val (2005) and Theorem 3.1 in Kato, Galvao, and Montes-Rojas (2012).

For the estimator $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ to converge weakly in distribution, we make the following assumptions.

B1: The covariates $\boldsymbol{x}_{it}$ has a bounded, convex support $\mathbb{R}_x$ and a density function f_{x_i}, which is bounded away from zero and infinity uniformly over $\boldsymbol{x}$ and i. In addition, $\inf_i \lambda_{\min}(\mathrm{E}[\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^\top]) > 0$, where $\lambda_{\min}$ is the smallest eigenvalue.

B2: For any nonnegative sequence $\epsilon_N \to 0$ and N large enough, $\lambda_{\min,\epsilon_N}$, the smallest eigenvalue of the matrix $\inf_i \lambda_{\min}(\mathrm{E}[\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^\top] f_i(0|\boldsymbol{x}_{i1}) 1\{\alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0 > \epsilon_N\}) > \lambda_0 > 0$. There exists a constant $\zeta > 0$ such that for any $\epsilon_N \to 0$, $\sup_{|\alpha_i - \alpha_{i0}| + ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq \zeta} \mathrm{E}[1\{|\alpha_i + \boldsymbol{x}^\top \boldsymbol{\beta}| < \epsilon_N\}] = O(\epsilon_N)$.

B3: $c_N \to 0$ and $T^{1/4} c_N$ is greater than some positive constant c^*.

B4: For any positive sequence $\epsilon_N \to 0$ with $\epsilon_N / c_N \to 1$ and any $\boldsymbol{x}_{i1}$, $\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + \epsilon_N$ implies $\alpha_{i0} + \boldsymbol{x}_{i1}^\top \beta_0 > \epsilon_N^*$ for some ϵ_N^* such that $\epsilon_N = O(\epsilon_N^*)$.

B5: $P(\pi_0(\boldsymbol{x}), \hat{\pi}(\boldsymbol{x}) \in C_c^{p+\alpha}(\mathcal{R}_{\boldsymbol{x}})) \to 1$ for some positive $\alpha \in (0, 1]$ and finite c, where $C_c^{p+\alpha}(\mathcal{R}_{\boldsymbol{x}})$ is the set of all continuous functions $h : \mathcal{R}_{\boldsymbol{x}} \to \mathbb{R}$ with $\|h\|_{\infty, p+\alpha} \leq c$.

B6: For any positive $\epsilon_N \to 0$ with $\max\{\max_{1 \leq i \leq N} |\alpha_i - \alpha_{i0}|, ||\boldsymbol{\beta} - \boldsymbol{\beta}_0||\} \leq \epsilon_N$,

$$
\begin{aligned}
&\mathrm{E}[1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}] = -\boldsymbol{D}_{N1}^*(\alpha_i - \alpha_{i0}) - \boldsymbol{D}_{N2}^*(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\
&\mathrm{E}[\boldsymbol{x}_{i1} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}] = -\boldsymbol{D}_{N3}^*(\alpha_i - \alpha_{i0}) - \boldsymbol{D}_{N4}^*(\boldsymbol{\beta} - \boldsymbol{\beta}_0)
\end{aligned}
$$

where $\boldsymbol{D}_{N1}^*$, $\boldsymbol{D}_{N2}^*$, $\boldsymbol{D}_{N3}^*$, and $\boldsymbol{D}_{N4}^*$ are positive semi-definite matrices satisfying $0 \leq \min\{\lambda_{\min}(\boldsymbol{D}_{Nj}^*)\}$ $\leq \max_j\{\lambda_{\max_j}(\boldsymbol{D}_{Nj}^*)\} < \infty$. The limiting forms of the following matrices are positive definite:

$$
\begin{aligned}
\boldsymbol{V} &= \tau(1-\tau) \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \mathrm{E}\{(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})^\top 1(\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau)\}, \\
\boldsymbol{\Lambda} &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N [\boldsymbol{B}_i - \boldsymbol{A}_i a_i^{-1} \boldsymbol{A}_i^\top] 1(\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau),
\end{aligned}
$$

where, $a_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{it}) 1(\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau)]$, $\boldsymbol{A}_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{it}) \boldsymbol{x}_{it} 1(\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau)]$, $\boldsymbol{B}_i := \mathrm{E}[f_i(0|\boldsymbol{x}_{it}) \boldsymbol{x}_{it} \boldsymbol{x}_{it}^\top 1(\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau)]$.

These conditions are similar to those in Tang, Wang, He, and Zhu (2012). B1 assumes a bounded support for convenience and is similar to A1 in Tang, Wang, He, and Zhu (2012). It can be relaxed under additional conditions on the smoothness of the propensity score function. B2 is parallel to A3 and assumption R.2 of Powell (1986) and is a standard condition in censored QR. B3 is similar to assumption A4 in Tang, Wang, He, and Zhu (2012) and is used to avoid boundary conditions. Assumption B4 is similar to A5.1 in Tang, Wang, He, and Zhu (2012) which basically requires that the derivative of $\pi_i(\boldsymbol{x}_{i1})$ is bounded and the true quantile line is not flat. B5 is analogous to A5.2 in Tang, Wang, He, and Zhu (2012). Finally, B6 is similar to A6 in Tang, Wang, He, and Zhu (2012). The following result states convergence in distribution.

Theorem 2. *Assume* $\sup_i ||\hat{\pi}_i - \pi_{i0}||_\infty = o_p(T^{-1/4})$. *Under conditions of Theorem 1 and B1–B6, as* $N^2(\log N)^3/T \to 0$,

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Lambda}^{-1}\boldsymbol{V}\boldsymbol{\Lambda}^{-1}).$$

The restriction that T grows at most polynomially in N is the same as in Kato, Galvao, and Montes-Rojas (2012). This condition is used only to "kill" the remainder term in the derivation of the asymptotic results. It serves as a warning device to practitioners on the type of situations where the asymptotics are likely to provide a good approximation in practice. Nevertheless, the large T requirement is unusual in several panel data sets in economics and finance. In this respect, the Monte Carlo simulations presented below assess the finite sample performance of the estimators and provide evidence of good small-sample performance. The simulation results confirm the asymptotic theory prediction that the bias decreases as T increases. In addition, even if the asymptotic theory requires relatively large T, the simulations show evidence that the bias is small for moderate T.

The components of the asymptotic covariance matrices in Theorem 2 that need to be estimated include a_i, $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$. Following Powell (1986), the matrices can be estimated by their sample counterpart. For instance, a_i can be estimated as

$$\hat{a}_i = \frac{1}{2Tg_N}\sum_{t=1}^{T} 1(|\hat{u}(\tau)| \leq g_N)1(\hat{\pi}(\alpha_i, \boldsymbol{x}_{it}) > 1 - \tau + c_N), \tag{2.7}$$

where $\hat{u}(\tau)$ has the τ-th conditional quantile at zero, the constant $c_N \to 0$, and g_N is an appropriately chosen bandwidth, with $g_N \to 0$ and $NTg_N^2 \to \infty$. Note also that $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ can be estimated similarly. The consistency of these asymptotic covariance matrix estimators is standard and will not be discussed further in this paper.

Remark 1. In the theorems above, we provide results based on joint asymptotics for the nonparametric 2-step estimator, and derive the requirements on the sample growth for the asymptotic properties. The large sample results for the 3-step estimator are similar and are given in the Supplemental Appendix. We show that the 2-step and 3-step estimators are asymptotically equivalent. In addition, all the asymptotic results hold for fixed N and $T \to \infty$.

3 Monte Carlo

In this section, we use Monte Carlo simulations to assess the finite sample performance of the estimators. We report results for empirical bias, root mean squared error (RMSE), and coverage probability for confidence interval with nominal level 0.95. We define the latent variable as, $y_{it}^* = \alpha_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \left[1 + \left(x_{1,it} + x_{2,it} + x_{1,it}^2 + x_{2,it}^2\right) \cdot \zeta\right] \cdot u_{it}$, where $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (10, -2)^\top$ is the parameter of interest, ζ modulates the amount of heteroscedasticity and $u_{it} \sim iid\mathcal{N}(0,1)$. We performed simulations with $\zeta \in \{0, 0.5\}$ and $u_{it} \sim t_3$, but we only report the case of Normal heteroscedastic errors to save space. In this case, we consider a parameter of interest $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta} + \zeta F_u^{-1}(\tau)$. We draw $\boldsymbol{x}_{it} \in \mathcal{X} \subset \mathbb{R}^2$ from independent standard normal distributions, truncated as $\{\boldsymbol{x}_{it} : \|x_{it}\|_\infty < 2\}$. The fixed effect, α_i, is generated as $\alpha_i = v_i + \varphi \sum_t (x_{1,it} + x_{2,it})$, with $v_i \sim \mathcal{N}(0,1)$. The censored variable is defined as $y_{it} = \max(y_{it}^*, C_{it})$, with C_{it} taking the value -0.95 or -1.45. These choices yield roughly 50% and 45% of censoring, respectively. Since we are considering left-censored observations, we estimate the model for $\tau \in \{0.25, 0.5\}$. Finally, we consider different sample sizes, setting the number of replications to 1000.

In the experiments, we consider six estimators. The first one is the Omniscient estimator which assumes knowledge of y_{it}^*. The second one is the parametric 3-step estimator, labeled 3-step, in which $(x_{1,it}, x_{2,it})$ and $(x_{1,it}^2, x_{2,it}^2)$ are used in the parametric (logit) estimation of the propensity score in the first step. Following Chernozhukov and Hong (2002), the cutoff value d is equal to the 0.1-th quantile of all $p(\boldsymbol{X}_{it}^\top \hat{\boldsymbol{\gamma}})$'s such that $p(\boldsymbol{X}_{it}^\top \hat{\boldsymbol{\gamma}}) > 1 - \tau$. In the second step, the parameter δ_{NT} is selected as the $1/3(NT)^{-1/3}$-th quantile of the estimated quantile function $\hat{\alpha}_i^0(\tau) + \boldsymbol{x}_{it}' \hat{\boldsymbol{\beta}}^0(\tau)$. We consider two versions of the 2-step estimator. The parametric 2-step (p2-step) estimates the propensity score in the first step using parametric logit regression. The nonparametric 2-step estimator (labeled n2-step) uses generalized additive methods for a logistic regression in the first step with $c = (NT)^{-1/5}\tau$, as

Sample Size		Censor	Quantile	Estimators					
N	T	point		Omniscient	3-step	p2-step	n2-step	Powell	Naive
100	15	–0.95	0.5	0.000	0.034	–0.059	–0.054	0.529	–1.650
				(0.100)	(0.198)	(0.302)	(0.297)	(1.006)	(4.730)
				[0.950]	[0.938]	[0.804]	[0.819]	[0.612]	[0.000]
100	50	–0.95	0.5	–0.001	0.009	–0.031	–0.023	0.496	–1.733
				(0.053)	(0.099)	(0.138)	(0.126)	(0.891)	(4.810)
				[0.953]	[0.945]	[0.844]	[0.889]	[0.517]	[0.000]
100	15	–1.45	0.5	0.000	0.030	–0.057	–0.050	0.529	–1.561
				(0.100)	(0.191)	(0.289)	(0.281)	(0.965)	(4.492)
				[0.950]	[0.942]	[0.876]	[0.820]	[0.604]	[0.000]
100	50	–1.45	0.5	–0.001	0.007	–0.029	–0.022	0.486	–1.653
				(0.053)	(0.096)	(0.133)	(0.121)	(0.851)	(4.582)
				[0.953]	[0.942]	[0.854]	[0.883]	[0.513]	[0.000]
100	15	–0.95	0.25	0.021	0.096	–0.113	–0.102	0.764	–1.641
				(0.115)	(0.407)	(0.577)	(0.568)	(1.478)	(5.035)
				[0.948]	[0.921]	[0.790]	[0.806]	[0.000]	[0.000]
100	50	–0.95	0.25	0.003	0.022	–0.026	–0.020	0.818	–1.673
				(0.059)	(0.148)	(0.210)	(0.198)	(1.566)	(5.091)
				[0.953]	[0.937]	[0.824]	[0.861]	[0.000]	[0.000]
100	15	–1.45	0.25	0.021	0.094	–0.102	–0.090	0.768	–1.540
				(0.115)	(0.382)	(0.532)	(0.518)	(1.414)	(4.780)
				[0.948]	[0.919]	[0.796]	[0.809]	[0.000]	[0.000]
100	50	–1.45	0.25	0.003	0.021	–0.024	–0.017	0.808	–1.568
				(0.059)	(0.139)	(0.198)	(0.186)	(1.479)	(4.797)
				[0.953]	[0.938]	[0.824]	[0.861]	[0.000]	[0.000]

Table 3.1: *Monte Carlo simulation results for* $\tau = \{0.5, 0.25\}$ *quantile and* $\varphi = 0.5$ *in the case of Normal heteroscedastic errors. The table shows the bias, RMSE (in parentheses), and coverage [in brackets].*

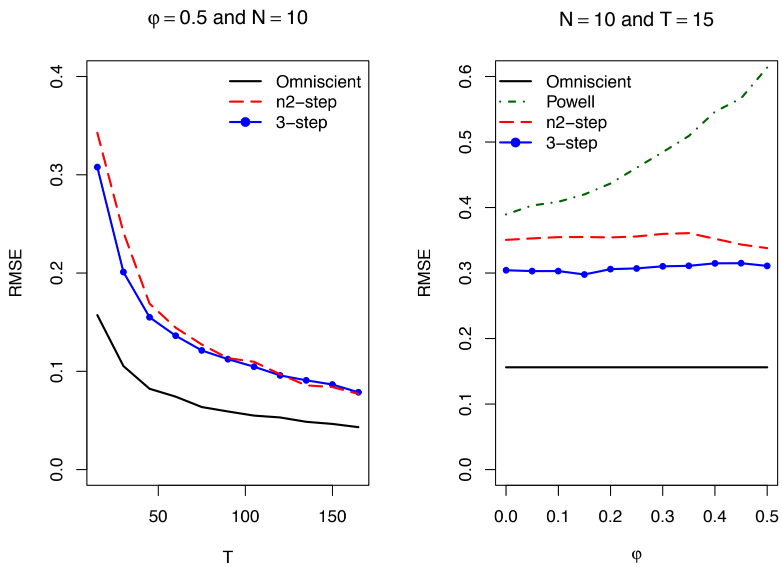

Figure 3.1: *Small sample performance of the estimators when T or φ increases. The case of $\varphi = 0$ represents the random effects case.*

in Tang, Wang, He, and Zhu (2012). The fifth estimator is a version of the Powell estimator without fixed effects, and lastly, a "naive" estimator that assumes that the observations are uncensored were also considered.

Table 3.1 shows, as expected, that the Omniscient estimator performs better than any other estimator. At the 0.5 quantile, the 3-step and p2-step estimators are slightly biased for small T, but their biases decrease substantially when T increases. The results for n2-step also show small biases, which tend to disappear as T increases. In terms of empirical coverage, the 3-step estimator performs well and produces empirical coverage close to the nominal 0.95. The bottom block of Table 3.1 presents results for the model estimated at $\tau = 0.25$. The results are somewhat analogous to the ones presented at the upper part of Table 3.1. When compared with the case for $\tau = 0.50$, we find that, in general, the bias of the estimators are slightly larger, but as in the previous case, the biases decrease substantially when T increases.

In order to shed light on the performance of the n2-step *vis-à-vis* the 3-step, Figure 3.1 offers the RMSE of the estimators from short-N simulations when varying the time series T (left panel) and φ (right panel). In these simulations, we only considered $C = -0.95$.

N	T	C	Non-linear Link Function 3-step Bias	RMSE	p2-step Bias	RMSE	Linear Probability Model 3-step Bias	RMSE	p2-step Bias	RMSE
1000	15	–0.95	0.023	0.068	–0.057	0.243	0.012	0.061	–0.105	0.374
1000	50	–0.95	0.007	0.032	–0.028	0.104	0.002	0.032	–0.056	0.213

Table 3.2: *Monte Carlo results for $\tau = \varphi = 0.5$ in the case of Normal heteroscedastic errors.*

The left panel shows that although the 3-step outperforms the n2-step for small T, the equivalence between them is achieved as T increases. The right panel shows the RMSE of the proposed estimators when the correlation between α_i and $\boldsymbol{x}_{it}$ is changed. When $\varphi = 0$, the Powell estimator performs relatively well, similarly to the estimators proposed in the paper. The converse is not true since the performance of the Powell estimator deteriorates quickly as φ differs from zero, which corresponds to the case where it is important to account for individual heterogeneity. In contrast, the performance of the proposed estimators remains unaffected for different values of φ.

To investigate how sensitive are the parametric estimators to the choice of a logit model in the first stage, we conduct simulations where we use a Linear Probability Model (LPM) to estimate propensity scores. Table 3.2 presents results for $N = 1000$ which is similar to the number of subjects considered in the empirical section. The results suggest that the LPM performs well in panel data models with large N and moderate T.

4 Extension: dependence case

We extend the results in Theorems 1 and 2 to the case where we allow for dependence across time while maintaining independence across individuals. The following assumptions are needed for this case.

E1: $\{(\boldsymbol{x}_{it}, y_{it}^*), t \geq 1\}$ is stationary and β-mixing for each fixed i, and independent across i. Let $\beta_i(j)$ denote the β -mixing coefficients of $\{(\boldsymbol{x}_{it}, y_{it}^*), t \geq 1\}$. Then, there exists constants $a \in (0, 1)$ and $B > 0$ such that $\sup_{i\geq 1} \beta_i(j) \leq Ba^j$ for all $j \geq 1$.

E2: Let $f_{i,j}(u_1, u_{1+j}|\boldsymbol{x}_1, \boldsymbol{x}_{1+j})$ denote the conditional density of (u_1, u_{1+j}) given $(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i,1+j}) = (\boldsymbol{x}_1, \boldsymbol{x}_{1+j})$. There exists a constant $C_f' > 0$ such that $f_{i,j}(u_1, u_{1+j}|\boldsymbol{x}_1, \boldsymbol{x}_{1+j}) \leq C_f'$ uniformly over $(u_1, u_{1+j}, \boldsymbol{x}_1, \boldsymbol{x}_{1+j})$ for all $i \geq 1$ and $j \geq 1$.

E3: Let $\widetilde{\boldsymbol{V}}_{Ni}$ denote the covariance matrix of the term $T^{-1/2}\sum_{t=1}^{T}\{\tau - I(u_{it} \leq 0)\}(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})1(\pi_{i0}(\boldsymbol{x}_{it}) > 1-\tau+c_N)$, then the limit $\widetilde{\boldsymbol{V}} := \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\widetilde{\boldsymbol{V}}_{Ni}$ exists and is nonsingular.

Condition E1 is similar to condition 1 of Hahn and Kuersteiner (2011) and Kato, Galvao, and Montes-Rojas (2012). Condition E2 imposes a new restriction on the conditional densities, but it is also standard as in Kato, Galvao, and Montes-Rojas (2012). Finally, E3 defines the long-run variance-covariance matrix.

Theorem 3. *Under Conditions E1-E3, A3 and B1-B6,* $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ *is consistent provided that* $(\log N)^2/T \to 0$ *and* $\sup_i ||\hat{\pi}_i - \pi_{i0}||_\infty = o_p(1)$*. Moreover, if* $N^2(\log N)^3/T \to 0$ *and* $\sup_i ||\hat{\pi}_i - \pi_{i0}||_\infty = o_p(T^{-1/4})$*, then we have that* $\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{V}}\boldsymbol{\Lambda}^{-1})$.

The proof for this result is given in the Supplemental Appendix.

5 An empirical application

Using data from Chay and Powell (2001), this section investigates relative earnings of black workers in the southern states of the United States in the period between 1957 and 1971. The black-white earnings differentials is an important research area in economics with a very large literature on the subject [see e.g. Brown (1984), Altonji and Blank (1999), Heckman, Lyons, and Todd (2000), Lang (2007)]. We apply our quantile method to a difference-in-difference model of earnings in which the parameter of interest measures the black-white earning gap after the introduction of the Title VII of the Civil Rights Act of 1964. This policy prohibited discrimination by employers on the basis of race and gender.

This paper employs data from the Current Population Survey. In a joint project of the Census Bureau and the Social Security Administration (SSA), respondents to the 1973 and 1978 March Current Population Surveys were matched by their Social Security numbers to their Social Security earnings histories. The data contains information on earnings of 1314 workers over 15 years, of which over 50% are censored at the maximum taxable earnings level for Social Security. Following Levine and Mitchell (1988), we consider two labor groups: young workers (ages 22-30 in 1957) and mature workers (ages 31-43 in 1957). This allows us to investigate whether the policy has a differential effect on the age structure of the workers.

Variable	OLS1	OLS2	QR1	QR2	PH	POR	CLAD
				Young Workers			
Black-white gap	–0.303	–0.155	–0.436	–0.202	–0.476	–0.471	–0.450
	(0.025)	(0.036)	(0.013)	(0.024)	(0.025)	(0.029)	(0.025)
Black-white gap after	0.056	0.027	0.126	0.046	0.125	0.120	0.128
1964 Civil Rights Act	(0.039)	(0.058)	(0.021)	(0.038)	(0.036)	(0.045)	(0.035)
Number of observations	10170	4886	10170	4886	10170	10170	10170
				Mature Workers			
Black-white gap	–0.203	–0.160	–0.209	–0.121	–0.252	–0.251	–0.185
	(0.020)	(0.033)	(0.011)	(0.020)	(0.027)	(0.022)	(0.020)
Black-white gap after	0.093	0.108	0.023	0.019	0.082	0.084	0.058
1964 Civil Rights Act	(0.031)	(0.052)	(0.018)	(0.030)	(0.032)	(0.025)	(0.030)
Number of observations	14490	6153	14490	6153	14490	14490	14490

Table 5.1: *Black-white earnings differentials. All quantile models are estimated at the median. Standard errors are presented in parentheses.*

We estimate the following censored quantile regression model with fixed effects,

$$Q_{y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i, C_{it}) = \min(C_{it}, \alpha_i(\tau) + \boldsymbol{x}_{it}^\top \boldsymbol{\beta}(\tau)), \tag{5.1}$$

where Q_y is the conditional quantile of the natural logarithm of earnings and $\boldsymbol{x}$ includes race (1 = black, 0 = white), an indicator for the period after the policy (1 = after 1964, 0 = before 1964), and an interaction term for race in the period after the policy (1 = black after 1964, 0 otherwise). The model includes other control variables as described in the Supplemental Appendix. The effect of interest is the black-white gap after the 1964 Civil Rights Act. For comparison, we estimate models without individual specific intercepts, $Q_{y_{it}}(\tau|\boldsymbol{x}_{it}, C_{it}) = \min(C_{it}, \boldsymbol{x}_{it}^\top \boldsymbol{\beta}(\tau))$, and without modeling the censored data in a quantile regression model for earnings, $Q_{y_{it}}(\tau|\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}^\top \boldsymbol{\beta}(\tau)$.

Table 5.1 presents results for the parameter of interest. The first column (labeled OLS1) shows standard least squares estimates, while the second column (OLS2) present least squares estimates obtained from a sample that drops the top-censored observations. The third and fourth columns present quantile regression estimates of the parameter of interest at the conditional median. QR1 is the standard quantile regression estimator and QR2 is the quantile regression estimator used on a sample that does not include top-censored observations. The next two columns (labeled PH and POR) present results obtained from Peng and Huang's (2008) method and Portnoy's (2003) censored quantile regression estimator. The last column, labeled CLAD, presents results from a version of Powell's semi-parametric estimator.

The OLS1 and QR1 estimators in Table 5.1 are suspected to deliver biased results due to the presence of censoring and unobserved individual heterogeneity. The estimates obtained by using OLS2 and QR2 are also biased because they only consider uncensored observations. PH, POR and CLAD address censoring but ignore individual heterogeneity possibly correlated with the covariates. The analysis of the results reported in Table 5.1 indicates that the median effects of the 1964 Civil Rights Act are small among mature workers but significant for young workers.

However, a complete analysis can only be obtained if we investigate the effect of the 1964 Civil Act on other quantiles of the conditional earnings distribution. This is presented in Figure 5.1, which shows three types of estimators: the first one does not account either for the presence of censoring or for fixed effects (QR1, QR2); the second type accounts for censoring but not for fixed effects (PH, PO, CLAD). In particular, recall that CLAD is exactly the estimator proposed by Wang and Fygenson (2009) which has the advantage of allowing identification of time-invariant effects. The last estimator, the 3-step, is the only one that accounts for both censoring and fixed effects and can be employed to estimate model (5.1) with dichotomous independent variables.

Due to the top coded observations, the estimates of upper quantiles obtained from QR1 would be biased towards zero. This is exactly what we see in Figure 5.1 where the graph showing the coefficient estimates obtained from QR1 are approaching zero as we go across quantiles. It is interesting to see that censoring does not seem to be the only issue at the upper quantiles, because the curves associated with PH, POR and CLAD tend to be concave. In order to avoid the potential bias caused by endogenous individual effects and censoring, we employ the 3-step estimator. Unlike the conclusion obtained using QR1 or CLAD for instance, we notice that the effect of the 1964 Civil Rights Act is increasing and significant at the upper quantiles of the conditional earnings distribution of young workers. Indeed, our simulations showed that under random effects, the n2-step estimator, the 3-step estimator, and the Powell estimator have similar performance because they take care of the censoring, but under fixed effects, only the methods proposed in this paper have satisfactory performance. Therefore, any empirical difference between the 3-step and Powell estimators may reflect the presence of endogeneity. Competing quantile regression methods fail to uncover the large effect in the upper tail of the conditional earnings distribution.

Chay and Powell (2001) use a semiparametric censored regression model to investigate

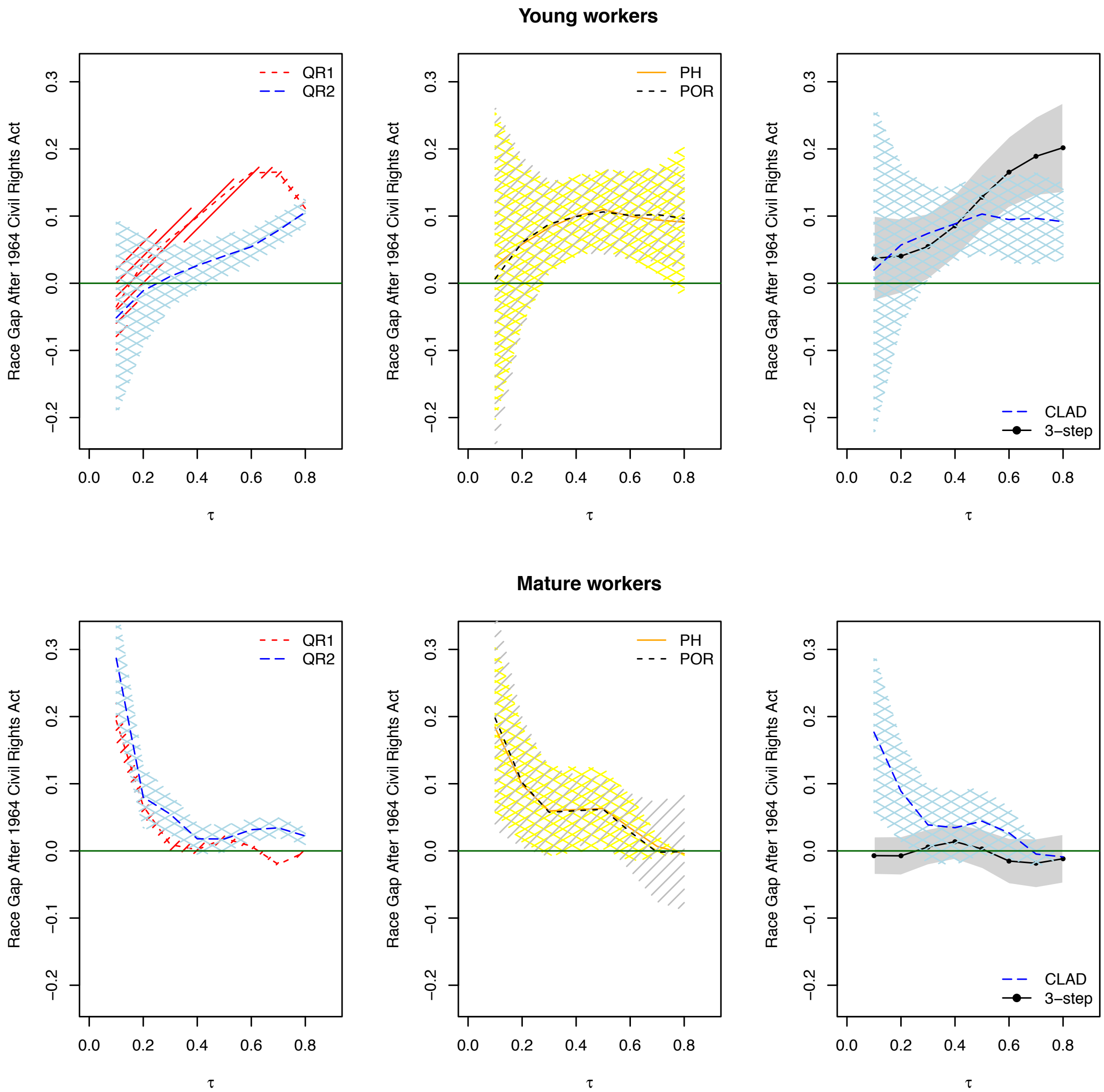

Figure 5.1: *Quantile effects of the black-white gap after the Civil Rights Act of 1964. The areas represent 90 percent pointwise confidence intervals.*

the black-white wage gap, and find significant earnings convergence among black and white man after the passage of the 1964 Civil Rights Act. We shed more light on the debate by revisiting this question and applying the quantile regression estimator. The proposed method offers a flexible approach to the analysis of censored panel data since one is able to control for individual specific intercepts while exploring heterogeneous covariate effects on the response variable. Our analysis contributes to the black-white earnings gap debate with two new conclusions: (i) the 1964 Civil Rights Act had no effect on the earnings distribution of mature workers, only affecting young workers; (ii) among the young workers to whom the policy had a significant effect, the ones at the upper quantiles of the distribution were more benefited. Thus, as a policy to reduce income inequality, we interpret this evidence as suggesting that the 1964 Civil Rights Act was beneficial to the group of black workers who need it less.

6 Conclusions

In this paper, we have introduced quantile regression methods to estimate censored panel models with individual specific fixed effects. We proposed methods that are obtained by applying fixed effects quantile regression to subsets of observations selected either parametrically or nonparametrically. We used the new estimator to reassess the effect of the 1964 Civil Rights Act on the black-white earnings gap. This policy prohibited discrimination against black and female workers and aimed to reduce the race income gap in the United States. Possible topics for future research include the case where C_{it} is a latent variable potentially dependent on covariates, inference in the presence of dependence and bootstrap methods for the proposed estimators.

References

Altonji, J., and R. Blank (1999): "Race and Gender in the Labor Market," in *Handbook of Labor Economics Vol 3*, ed. by O. Ashenfelter, and D. Card, pp. 3143–3259. North-Holland, Amsterdam.

Bilias, Y., S. Chen, and Z. Ying (2000): "Simple Resampling Methods for Censored Regression Quantiles," *Journal of Econometrics*, 99, 373–386.

Bousquet, O. (2002): "A Bennett Concentration Inequality and Its Application to Suprema of Empirical Processes," *Comptes Rendus Mathematique*, 334, 495–500.

Brown, C. (1984): "Black-White Earnings Ratios since the Civil Rights Acts of 1964: The Importance of Labor Market Dropouts," *Quarterly Journal of Economics*, 91, 31–44.

Buchinsky, M. (1994): "Changes in U.S. Wage Structure 1963-87: An Application of Quantile Regression," *Econometrica*, 62, 405–458.

Buchinsky, M., and J. Hahn (1998): "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653–671.

Canay, I. A. (2011): "A Simple Approach to Quantile Regression for Panel Data," *The Econometrics Journal*, 14, 368–386.

Chay, K. Y., and J. L. Powell (2001): "Semiparametric Censored Regression Models," *Journal of Economic Perspectives*, 15, 29–42.

Chen, S., and S. Khan (2008): "Semiparametric Estimation of Nonstationary Censored Panel Data Models with Time Varying Factor Loads," *Econometric Theory*, 24, 1149–1173.

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

Chernozhukov, V., and H. Hong (2002): "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of the American Statistical Association*, 97, 872–882.

Fernández-Val, I. (2005): "Bias Correction in Panel Data Models with Individual Specific Parameters," Mimeo.

Fitzenberger, B. (1997): "A Guide to Censored Quantile Regressions," in *Handbook of Statistics, Robust Inference*, ed. by G. Maddala, and C. Rao, pp. 405–437. North-Holland, Amsterdam.

Hahn, J., and G. Kuersteiner (2011): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," *Econometric Theory*, 27, 1152–1191.

Hahn, J., and W. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319.

Heckman, J. J., T. M. Lyons, and P. E. Todd (2000): "Understanding Black-White Wage Differentials, 1960-1990," *American Economic Review, Papers and Proceedings*, 90, 344–349.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533–565.

Honoré, B. E., S. Khan, and J. L. Powell (2002): "Quantile Regression Under Random Censoring,," *Journal of Econometrics*, 109, 67–105.

Kato, K., A. F. Galvao, and G. Montes-Rojas (2012): "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics*, 170, 76–91.

Khan, S., M. Ponomareva, and E. Tamer (2011): "Identification of Panel Data Models with Endogenous Censoring," Mimeo.

Khan, S., and J. L. Powell (2001): "Two-Step Estimation of Semiparametric Censored Regression Models," *Journal of Econometrics*, 100, 319–355.

Kim, M.-O., and Y. Yang (2011): "Semiparametric Approach to a Random Effects Quantile Regression Model," *Journal of the American Statistical Association*, 106, 1405–1417.

Knight, K. (1989): "Limit Theory for Autoregressive Parameter Estimates in an Infinite Variance Random Walk," *The Canadian Journal of Statistics*, 17, 261–278.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

Koenker, R., and G. W. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–49.

Lang, K. (2007): *Poverty and Discrimination*. Princeton University Press, New Jersey.

Levine, P. B., and O. S. Mitchell (1988): "The Baby Boom's Legacy: Relative Wages in the Twenty-First Century," *American Economic Review*, 72, 66–69.

Li, Q., and J. S. Racine (2007): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press, New Jersey.

LIN, G., X. HE, AND S. PORTNOY (2012): "Quantile Regression with Doubly Censored Data," *Computational Statistics & Data Analysis*, 56, 797–812.

NEYMAN, J., AND E. L. SCOTT (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16, 1–32.

PAARSCH, H. J. (1984): "A Monte Carlo Comparison of Estimators for Censored Regression Models," *Journal of Econometrics*, 24, 197–213.

PENG, L., AND Y. HUANG (2008): "Survival Analysis With Quantile Regression Models," *Journal of the American Statistical Association*, 103, 637–649.

PORTNOY, S. L. (2003): "Censored Regression Quantiles," *Journal of the American Statistical Association*, 98, 1001–1012.

POWELL, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

——— (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

SCHNEDLER, W. (2005): "Likelihood Estimation for Censored Random Vectors," *Econometric Reviews*, 24, 195–217.

TANG, Y., H. J. WANG, X. HE, AND Z. ZHU (2012): "An Informative Subset-Based Estimator for Censored Quantile Regression," *Test*, 21, 635–655.

VAN DER VAART, A., AND J. A. WELLNER (1996): *Weak convergence and empirical processes*. Springer-Verlag Press, New York, New York.

WANG, H. J., AND M. FYGENSON (2009): "Inference for Censored Quantile Regression in Longitudinal Studies," *The Annals of Statistics*, 37, 756–781.

A Appendix: Proofs

For notational purposes, let $\pi_{i0}(\boldsymbol{x}_{it}) \equiv \pi_0(\alpha_i, \boldsymbol{x}_{it})$. We usually supress arguments of the functions $\psi(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) = \tau - 1(y_{it} < \alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta})$ and $\rho_\tau(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) = \rho_\tau(y_{it} - \alpha_i - \boldsymbol{x}_{it}^\top \boldsymbol{\beta})$ for notational simplicity. Therefore, $\psi(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) \equiv \psi(\cdot; \alpha_i, \boldsymbol{\beta})$ and $\rho_\tau(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) \equiv \rho_\tau(\cdot; \alpha_i, \boldsymbol{\beta})$. Following Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012), we assume $C_{it} = C = 0$ throughout the proofs.

Proof of Theorem 1. The proof of consistency is an application of three auxiliary lemmas and is divided in two steps. First, we show that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - (\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) \xrightarrow{p} 0$, where $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \operatorname{argmin} Q_N(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\pi})$ and $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) = \operatorname{argmin} Q_{3,N}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Second, we demonstrate that $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) \xrightarrow{p} (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$. Therefore, we conclude that $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) \xrightarrow{p} (\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$.

Step 1: The asymptotic equivalence of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ and $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$. This step is an application of two lemmas. To this end, in Lemma 1, we first show that the objection functions $Q_N(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\pi})$ and $Q_{3,N}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ are asymptotically equivalent uniformly in $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Then, we apply Lemma 2, which shows that the uniform asymptotic equivalence of the objective functions implies the asymptotic equivalence of the minimizers of the objective functions. Therefore, $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) - (\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}}) \xrightarrow{p} 0$.

Step 2: The consistency of $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$ is shown in Lemma 3. $\square$

Lemma 1. *Under the assumptions of Theorem 1,* $\sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} |Q_N(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\pi}) - Q_{3,N}(\boldsymbol{\alpha}, \boldsymbol{\beta})| = o_p(1)$.

Proof. To see this notice that,

$$
\begin{aligned}
&(Q_N(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{\pi}) - Q_{3,N}(\boldsymbol{\alpha}, \boldsymbol{\beta}))^2 \\
=&\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\rho_\tau(\cdot;\alpha_i,\boldsymbol{\beta}) - \rho_\tau(\cdot;\alpha_{i0},\boldsymbol{\beta}_0))(1\{\hat{\pi}_i(\boldsymbol{x}_{it}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1-\tau\})\Big)^2 \\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\rho_\tau(\cdot;\alpha_i,\boldsymbol{\beta}) - \rho_\tau(\cdot;\alpha_{i0},\boldsymbol{\beta}_0)\right)^2 \times \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(1\{\hat{\pi}_i(\boldsymbol{x}_{it}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1-\tau\})^2 \\
\leq&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} 9\left((\alpha_i - \alpha_{i0}) + \boldsymbol{x}_{it}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)^2 \times \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|1\{\hat{\pi}_i(\boldsymbol{x}_{it}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1-\tau\}| \\
=&\left(9\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\left((\alpha_i - \alpha_{i0}) + \boldsymbol{x}_{it}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)^2 + o_p(1)\right) \times o_p(1) = 0.
\end{aligned}
$$

The first inequality uses Cauchy-Schwarz inequality and the second inequality uses the identity of Knight (1989). The first term of the last line uses the Weak Law of Large Numbers for independent data and condition A2. To see that the second term is $o_p(1)$, we first do the following calculation:

$$
\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}|1\{\hat{\pi}_i(\boldsymbol{x}_{it})>1-\tau+c_N\}-1\{\pi_{i0}(\boldsymbol{x}_{it})>1-\tau\}|\\
=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}1\{\hat{\pi}_i(\boldsymbol{x}_{it})-c_N\le 1-\tau<\pi_{i0}(\boldsymbol{x}_{it}) \text{ or } \pi_{i0}(\boldsymbol{x}_{it})\le 1-\tau<\hat{\pi}_i(\boldsymbol{x}_{it})-c_N\}\\
=&\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[1\{\pi_i(\boldsymbol{x}_{it})-c_N\le 1-\tau<\pi_{i0}(\boldsymbol{x}_{it}) \text{ or } \pi_{i0}(\boldsymbol{x}_{it})\le 1-\tau<\pi_i(\boldsymbol{x}_{it})-c_N\}]_{\pi_i=\hat{\pi}_i}+o_p(1)\\
\le&\lim_{N\to\infty}\sup_{||\pi-\pi_0||<\epsilon_N}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[1\{\pi_i(\boldsymbol{x}_{it})-c_N\le 1-\tau<\pi_{i0}(\boldsymbol{x}_{it}) \text{ or } \pi_{i0}(\boldsymbol{x}_{it})\le 1-\tau<\pi_i(\boldsymbol{x}_{it})-c_N\}]+o_p(1)\\
\le&\lim_{N\to\infty}\sup_{||\pi-\pi_0||<\epsilon_N}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[1\{\pi_i(\boldsymbol{x}_{it})-(1-\tau+c_N)\le\epsilon_N\}]+o_p(1)=o_p(1).
\end{aligned}
$$

Condition A4 says that if π is very close to π_0, then it is very unlikely that π is close to $1-\tau+c_N$. However, from the previous expression, the only way for $\hat{\pi}$ to be close to π_0 is to get close to $1-\tau+c_N$. $\square$

Lemma 2. *Let $S_N(\boldsymbol{\theta},\pi)$ be a convex function in $\boldsymbol{\theta}$. Suppose $\sup_{\boldsymbol{\theta}}|S_N(\boldsymbol{\theta},\hat{\pi})-S_N(\boldsymbol{\theta},\pi_0)|=o_p(1)$. For any $\delta>0$, $S_N(\hat{\boldsymbol{\theta}}_1,\hat{\pi})<\inf_{||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_1||>\delta}S_N(\boldsymbol{\theta},\hat{\pi})$, and $S_N(\hat{\boldsymbol{\theta}}_2,\pi_0)<\inf_{||\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_2||>\delta}S_N(\boldsymbol{\theta},\pi_0)$, then $||\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\theta}}_2||=o_p(1)$.*

Proof. For a fixed δ, we note that if $||\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\theta}}_2||>\delta$, then $S_N(\hat{\boldsymbol{\theta}}_1,\hat{\pi})<S_N(\hat{\boldsymbol{\theta}}_2,\hat{\pi})=S_N(\hat{\boldsymbol{\theta}}_2,\pi_0)+o_p(1)$. The inequality is due to the definition of $\hat{\boldsymbol{\theta}}_1$, while the equality is due to the uniform asymptotic equivalence condition $\sup_{\boldsymbol{\theta}}|S_N(\boldsymbol{\theta},\hat{\pi})-S_N(\boldsymbol{\theta},\pi_0)|=o_p(1)$.

Note that the event relation $\{||\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\theta}}_2||>\delta\}\subset\{S_N(\hat{\boldsymbol{\theta}}_1,\pi_0)>S_N(\hat{\boldsymbol{\theta}}_2,\pi_0)+\epsilon(\delta)\}$, with $\epsilon(\delta)>0$. Therefore, $\mathrm{P}\{||\hat{\boldsymbol{\theta}}_1-\hat{\boldsymbol{\theta}}_2||>\delta\}\le\mathrm{P}\{S_N(\hat{\boldsymbol{\theta}}_1,\pi_0)>S_N(\hat{\boldsymbol{\theta}}_2,\pi_0)+\epsilon(\delta)\}$. But $S_N(\hat{\boldsymbol{\theta}}_1,\pi_0)-S_N(\hat{\boldsymbol{\theta}}_2,\pi_0)<S_N(\hat{\boldsymbol{\theta}}_1,\pi_0)-S_N(\hat{\boldsymbol{\theta}}_1,\hat{\pi})+o_p(1)\le\sup_{\boldsymbol{\theta}}|S_N(\boldsymbol{\theta},\hat{\pi})-S_N(\boldsymbol{\theta},\pi_0)|+o_p(1)=o_p(1)$. $\square$

Lemma 3. *Under the assumptions of Theorem 1, as $N/T^s\to 0$ for some real $s\ge 1$, the minimizer of $Q_{3,N}(\boldsymbol{\alpha},\boldsymbol{\beta})$, $(\tilde{\boldsymbol{\alpha}},\tilde{\boldsymbol{\beta}})$, is a consistent estimator of $(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0)$.*

Proof. Denote $\mathbb{M}_{Ni}(\alpha_i,\boldsymbol{\beta})=\frac{1}{T}\sum_{t=1}^{T}\rho_\tau(y_{it}-\alpha_i-\boldsymbol{x}_{it}^\top\boldsymbol{\beta})1\{\pi_{i0}(\boldsymbol{x}_{it})>1-\tau\}$ and $\Delta_{Ni}(\alpha_i,\boldsymbol{\beta})=\mathbb{M}_{Ni}(\alpha_i,\boldsymbol{\beta})-\mathbb{M}_{Ni}(\alpha_{i0},\boldsymbol{\beta}_0)$. For each $\delta>0$, define $B_i(\delta):=\{(\boldsymbol{\alpha},\boldsymbol{\beta}):|\alpha_i-\alpha_{i0}|+||\boldsymbol{\beta}-\boldsymbol{\beta}_0||_1\le\delta\}$ and $\partial B_i(\delta):=\{(\boldsymbol{\alpha},\boldsymbol{\beta}):|\alpha_i-\alpha_{i0}|+||\boldsymbol{\beta}-\boldsymbol{\beta}_0||_1=\delta\}$.

Step 1. The consistency of $\tilde{\boldsymbol{\beta}}$ Fix any $\delta>0$. For each $(\alpha_i,\boldsymbol{\beta})\notin B_i(\delta)$, define $\bar{\alpha}_i=r_i\alpha_i+(1-r_i)\alpha_{i0}$, $\bar{\boldsymbol{\beta}}_i=r_i\boldsymbol{\beta}+(1-r_i)\boldsymbol{\beta}_0$, where $r_i=\frac{\delta}{|\alpha_i-\alpha_{i0}|+||\boldsymbol{\beta}-\boldsymbol{\beta}_0||_1}$. Note that the objective function is convex. Therefore, we follow the steps in Kato, Galvao, and Montes-Rojas (2012).

$$
\begin{aligned}
r_i[\mathbb{M}_{Ni}(\alpha_i, \boldsymbol{\beta}) - \mathbb{M}_{Ni}(\alpha_{i0}, \boldsymbol{\beta}_0)] \geq & \mathbb{M}_{Ni}(\bar{\alpha}_i, \bar{\boldsymbol{\beta}}_i) - \mathbb{M}_{Ni}(\alpha_{i0}, \boldsymbol{\beta}_0) \\
= & \mathrm{E}[\Delta_{Ni}(\bar{\alpha}_i, \bar{\boldsymbol{\beta}}_i)] + (\Delta_{Ni}(\bar{\alpha}_i, \bar{\boldsymbol{\beta}}_i) - \mathrm{E}[\Delta_{Ni}(\bar{\alpha}_i, \bar{\boldsymbol{\beta}}_i)]). \qquad \text{(A.1)}
\end{aligned}
$$

Using Knight's identity,

$$
\mathrm{E}[\Delta_{Ni}(\alpha_i, \boldsymbol{\beta}_i)] = \mathrm{E}\left[\int_0^{(\alpha_i - \alpha_{i0}) + \boldsymbol{x}_{i1}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)} [F_i(s|\boldsymbol{x}_{i1}) - \tau]\, ds 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau\}\right].
$$

By condition A3, the first term of (A.1) is greater or equal to ϵ_δ for all $1 \leq i \leq N$. Therefore, we have

$$
\begin{aligned}
& \{||\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||_1 > \delta\} \subset \{\mathbb{M}_{Ni}(\alpha_i, \boldsymbol{\beta}) \leq \mathbb{M}_{Ni}(\alpha_{i0}, \boldsymbol{\beta}_0), \text{for some } i \text{ and } (\alpha_i, \boldsymbol{\beta})\} \\
\subset & \left\{ \max_{1 \leq i \leq N} \sup_{\alpha_i, \boldsymbol{\beta} \in B_i(\delta)} |\Delta_{Ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{Ni}(\alpha_i, \boldsymbol{\beta})]| \geq \epsilon_\delta \right\}.
\end{aligned}
$$

Therefore, it suffices to show that for every $\epsilon > 0$,

$$
\lim_{N \to \infty} \mathrm{P}\left\{ \max_{1 \leq i \leq N} \sup_{\alpha_i, \boldsymbol{\beta} \in B_i(\delta)} |\Delta_{Ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{Ni}(\alpha_i, \boldsymbol{\beta})]| \geq \epsilon \right\} = 0,
$$

whose sufficient condition is

$$
\max_{1 \leq i \leq N} \mathrm{P}\left\{ \sup_{\alpha_i, \boldsymbol{\beta} \in B_i(\delta)} |\Delta_{Ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{Ni}(\alpha_i, \boldsymbol{\beta})]| \geq \epsilon \right\} = o(N^{-1}).
$$

Without loss of generality, assume $\alpha_{i0} = 0$ and $\boldsymbol{\beta} = 0$. Then all the balls $B_i(\delta)$ are the same and therefore are denoted by $B(\delta)$. Let $g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u, \boldsymbol{x})$ denote $(\rho_\tau(u - \boldsymbol{\alpha} - \boldsymbol{x}^\top\boldsymbol{\beta}) - \rho_\tau(u))1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau\}$. We have that $|g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u, \boldsymbol{x}) - g_{\check{\alpha},\check{\boldsymbol{\beta}}}(u, \boldsymbol{x})| \leq C(1 + ||\boldsymbol{x}||_1)(|\boldsymbol{\alpha} - \check{\boldsymbol{\alpha}}| + ||\boldsymbol{\beta} - \check{\boldsymbol{\beta}}||_1)$ for some constant $C > 0$. Let $L(\boldsymbol{x}) := C(1 + ||\boldsymbol{x}||_1)$ and $\kappa := \sup_{i \geq 1} \mathrm{E}[L(\boldsymbol{x})]$. Since $B(\delta)$ is a compact subset, there exist K ℓ_1-balls with centers $(\alpha^{(j)}, \boldsymbol{\beta}^{(j)})$, $j = 1, ..., K$ and radius $\frac{\epsilon}{7\kappa}$ such that the collection of these balls covers $B(\delta)$. Note that K is independent of i and can be chosen such that $K = K(\epsilon) = O(\epsilon^{-p-1})$ as $\epsilon \to 0$. For each $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \in B(\delta)$, there is $j \in \{1, ..., K\}$ such that $|g_{\boldsymbol{\alpha},\boldsymbol{\beta}}(u, \boldsymbol{x}) - g_{\alpha^{(j)},\boldsymbol{\beta}^{(j)}}(u, \boldsymbol{x})| \leq L(\boldsymbol{x})\epsilon/(7\kappa)$, which leads to

$$
|\Delta_{Ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{Ni}(\alpha_i, \boldsymbol{\beta})]| \leq |\Delta_{Ni}(\alpha_i^{(j)}, \boldsymbol{\beta}^{(j)}) - \mathrm{E}[\Delta_{Ni}(\alpha_i^{(j)}, \boldsymbol{\beta}^{(j)})]| + \frac{\epsilon}{7\kappa}\left|\frac{1}{T}\sum_{t=1}^T \{L(\boldsymbol{x}_{it}) - \mathrm{E}[L(\boldsymbol{x}_{it})]\}\right| + \frac{2\epsilon}{7}
$$

and therefore

$$
\begin{aligned}
& \mathrm{P}\left\{ \sup_{(\boldsymbol{\alpha},\boldsymbol{\beta}) \in B(\delta)} |\Delta_{Ni}(\alpha_i, \boldsymbol{\beta}) - \mathrm{E}[\Delta_{Ni}(\alpha_i, \boldsymbol{\beta})]| > \epsilon \right\} \\
\leq & \sum_{j=1}^K \mathrm{P}\left\{ |\Delta_{Ni}(\alpha_i^{(j)}, \boldsymbol{\beta}^{(j)}) - \mathrm{E}[\Delta_{Ni}(\alpha_i^{(j)}, \boldsymbol{\beta}^{(j)})]| > \frac{\epsilon}{3} \right\} + \mathrm{P}\left\{ \frac{1}{T}\left|\sum_{t=1}^T \{L(\boldsymbol{x}_{it}) - \mathrm{E}[L(\boldsymbol{x}_{i1})]\}\right| > \frac{7\kappa}{3} \right\}.
\end{aligned}
$$

Since $\sup_{i\geq 1} \mathrm{E}[L^{2s}(\boldsymbol{x}_{i1})] < \infty$ by condition A2, application of the Marcinkiewicz-Zygmund inequality implies that both terms on the right side of the previous inequality are $O(T^s)$ uniformly over $1 \leq i \leq N$, and therefore $o(N^{-1})$.

Step 2. The consistency of $\tilde{\boldsymbol{\alpha}}$. For each i, $\tilde{\alpha}_i = \operatorname{argmin}_{\boldsymbol{\alpha}} \mathbb{M}_{Ni}(\boldsymbol{\alpha}, \tilde{\boldsymbol{\beta}})$. Fix $\delta > 0$. For each α_i with $|\alpha_i - \alpha_{i0}| > \delta$, define $\bar{\alpha}_i = r_i\alpha_i + (1-r_i)\alpha_{i0}$, where $r_i = \frac{\delta}{|\alpha_i - \alpha_{i0}|}$. Due to the convexity of the objective function, we have

$$
\begin{aligned}
&r_i(\mathbb{M}_{Ni}(\alpha_i, \tilde{\boldsymbol{\beta}}) - \mathbb{M}_{Ni}(\alpha_{i0}, \tilde{\boldsymbol{\beta}})) \geq \mathbb{M}_{Ni}(\bar{\alpha}_i, \tilde{\boldsymbol{\beta}}) - \mathbb{M}_{Ni}(\alpha_{i0}, \tilde{\boldsymbol{\beta}}) \\
=&\mathbb{M}_{Ni}(\bar{\alpha}_i, \tilde{\boldsymbol{\beta}}) - \mathbb{M}_{Ni}(\alpha_{i0}, \boldsymbol{\beta}_0) - [\mathbb{M}_{Ni}(\alpha_{i0}, \tilde{\boldsymbol{\beta}}) - \mathbb{M}_{Ni}(\alpha_{i0}, \boldsymbol{\beta}_0)] \\
=&\{\Delta_{Ni}(\bar{\alpha}_i, \tilde{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{Ni}(\bar{\alpha}_i, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}\} - \{\Delta_{Ni}(\alpha_{i0}, \tilde{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{Ni}(\alpha_{i0}, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}\} \\
&+ \mathrm{E}[\Delta_{Ni}(\bar{\alpha}_i, \boldsymbol{\beta}_0)] + \mathrm{E}[\Delta_{Ni}(\bar{\alpha}_i, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} - \mathrm{E}[\Delta_{Ni}(\bar{\alpha}_i, \boldsymbol{\beta}_0)] + \mathrm{E}[\Delta_{Ni}(\alpha_{i0}, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}.
\end{aligned}
$$

From condition A3 the third term on the right side is greater than ϵ_δ. Thus, we obtain the inclusion relation

$$
\begin{aligned}
&\{|\tilde{\alpha}_i - \alpha_{i0}| > \delta \text{ for some } i\} \subset \{\mathbb{M}_{Ni}(\alpha_i, \tilde{\boldsymbol{\beta}}) \leq \mathbb{M}_{Ni}(\alpha_{i0}, \tilde{\boldsymbol{\beta}}) \text{ for some } i \text{ and } \alpha_i \text{ such that } |\alpha_i - \alpha_{i0}| > \delta|\} \\
\subseteq&\{\max_{1\leq i\leq N} \sup_{|\alpha - \alpha_{i0}|\leq\delta} |\Delta_{Ni}(\alpha, \tilde{\boldsymbol{\beta}}) - \mathrm{E}[\Delta_{Ni}(\alpha, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}| \geq \frac{\epsilon_\delta}{4}\} \\
&\cup \{\max_{1\leq i\leq N} \sup_{|\alpha - \alpha_{i0}|\leq\delta} |\mathrm{E}[\Delta_{Ni}(\alpha, \boldsymbol{\beta})]|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} - \mathrm{E}[\Delta_{Ni}(\alpha, \boldsymbol{\beta}_0)]| \geq \frac{\epsilon_\delta}{4}\} := A_{1N} \cup A_{2N}.
\end{aligned}
$$

Because $\tilde{\boldsymbol{\beta}}$ is consistent and especially $\tilde{\boldsymbol{\beta}} = O_p(1)$, then $\mathrm{P}(A_{1N}) \to 0$. Also, since

$$
|\mathrm{E}[\Delta_{Ni}(\alpha, \boldsymbol{\beta})] - \mathrm{E}[\Delta_{Ni}(\alpha, \boldsymbol{\beta}_0)]| \leq 2\mathrm{E}[||\boldsymbol{x}_{i1}||]||\boldsymbol{\beta} - \boldsymbol{\beta}_0||,
$$

and $\sup_{i\geq 1} \mathrm{E}[||\boldsymbol{x}_{i1}||] \leq 1 + \sup_{i\geq 1} \mathrm{E}[||\boldsymbol{x}_{i1}||^{2s}] < \infty$ by condition A2, consistency of $\tilde{\boldsymbol{\beta}}$ implies that $\mathrm{P}(A_{2N}) \to 0$. $\square$

Proof of Theorem 2. Recall that $\psi(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) = \tau - 1(y_{it} < \alpha_i + \boldsymbol{x}_{it}^\top \boldsymbol{\beta})$. Define,

$$
\begin{aligned}
\mathbb{H}_{Ni}^{(1)}(\alpha_i, \boldsymbol{\beta}, \pi_i) :=&\frac{1}{T}\sum_{t=1}^{T} \psi(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta}) 1\{\pi_i(\boldsymbol{x}_{it}) > 1 - \tau + c_N\} \\
\mathbb{H}_{N}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) :=&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \psi(y_{it}, \boldsymbol{x}_{it}; \alpha_i, \boldsymbol{\beta})\boldsymbol{x}_{it} 1\{\pi_i(\boldsymbol{x}_{it}) > 1 - \tau + c_N\} \\
H_{Ni}^{(1)}(\alpha_i, \boldsymbol{\beta}, \pi_i) :=&\mathrm{E}[\mathbb{H}_{Ni}^{(1)}(\alpha_i, \boldsymbol{\beta}, \pi_i)], \quad \text{and} \quad H_{N}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) := \mathrm{E}[\mathbb{H}_{N}^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi})].
\end{aligned}
$$

We divide the proof into several steps.

Step 1: Zeros of the estimating equations. By the computational property of the QR estimator, we have $\max_{1\le i\le N}|\mathbb{H}^{(1)}_{Ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}},\hat{\pi}_i)| = O_p(T^{-1})$. To see this, note that

$$\begin{aligned}|\mathbb{H}^{(1)}_{Ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}},\hat{\pi}_i)| &= \left|\frac{1}{T}\sum_{t=1}^{T}\psi(y_{it},\boldsymbol{x}_{it};\hat{\alpha}_i,\hat{\boldsymbol{\beta}})1\{\hat{\pi}_i(\boldsymbol{x}_{it})>1-\tau+c_N\}\right| \\ &\le \left|\sum_{t=1}^{T}1\{y_{it}=\hat{\alpha}_i+\boldsymbol{x}_{it}^\top\hat{\boldsymbol{\beta}}\}\right|\max_{i,t}\frac{1\{\hat{\pi}_i(\boldsymbol{x}_{it})>1-\tau+c_N\}}{T} = O_p(T^{-1}).\end{aligned}$$

$$\begin{aligned}|\mathbb{H}^{(2)}_{N}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat{\pi})| &= \left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi(y_{it},\boldsymbol{x}_{it};\hat{\alpha}_i,\hat{\boldsymbol{\beta}})\boldsymbol{x}_{it}1\{\hat{\pi}_i(\boldsymbol{x}_{it})>1-\tau+c_N\}\right| \\ &\le \left|\sum_{i=1}^{N}\sum_{t=1}^{T}1\{y_{it}=\hat{\alpha}_i+\boldsymbol{x}_{it}^\top\hat{\boldsymbol{\beta}}\}\right|\max_{i,t}\frac{||\boldsymbol{x}_{it}||1\{\hat{\pi}_i(\boldsymbol{x}_{it})>1-\tau+c_N\}}{NT} = O_p(T^{-1}).\end{aligned}$$

Step 2: Asymptotic equicontinuity. Take $\delta_N\to 0$ such that $\max_{1\le i\le N}|\hat{\alpha}_i-\alpha_{i0}|\vee ||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|| = O_p(\delta_N)$. We shall show that

$$\begin{aligned}&\left|\left|\frac{1}{N}\sum_{i=1}^{N}m_i\{\mathbb{H}^{(1)}_{Ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}},\hat{\pi}_i)-H^{(1)}_{Ni}(\hat{\alpha}_i,\hat{\boldsymbol{\beta}},\hat{\pi}_i)-\mathbb{H}^{(1)}_{Ni}(\alpha_{i0},\boldsymbol{\beta}_0,\pi_{i0})\}\right|\right| = O_p(d_N)\\ &\left|\left|\mathbb{H}^{(2)}_{N}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\pi}})-H^{(2)}_{N}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\pi}})-\mathbb{H}^{(2)}_{N}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\boldsymbol{\pi}_0)\right|\right| = O_p(d_N)\end{aligned}$$

where m_i is any sequence bounded over i, and $d_N = T^{-1}|\log(\delta_N\vee T^{-1/4})|\vee T^{-1/2}(\delta_N^{1/2}\vee T^{-1/8})|\log(\delta_N\vee T^{-1/4})|$.

We only prove the first equation since that of the second is analogous. Without loss of generality, we assume $\alpha_{i0}=\alpha_0$, $\boldsymbol{\beta}=\boldsymbol{\beta}_0$, and $\pi_{i0}=\pi_0$. Put $g_{\alpha,\boldsymbol{\beta},\pi}=1\{y\le\alpha+\boldsymbol{x}^\top\boldsymbol{\beta}\}1\{\pi(\boldsymbol{x})>1-\tau+c_N\}-1\{y\le\alpha_0+\boldsymbol{x}^\top\boldsymbol{\beta}_0\}1\{\pi_0(\boldsymbol{x})>1-\tau+c_N\}$, $\mathcal{G}_\delta=\{g_{\alpha,\boldsymbol{\beta},\pi}:|\alpha-\alpha_0|\le\delta,||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\le\delta,||\pi-\pi_0||_\infty\le\delta\}$, and $\xi_{it}=(u_{it},\boldsymbol{x}_{it})$. It suffices to show that

$$\max_{1\le i\le N}\mathrm{E}\left[\left|\left|\sum_{t=1}^{T}\{g(\xi_{it})-\mathrm{E}[g(\xi_{i1})]\}\right|\right|_{\mathcal{G}_{\delta_N}}\right] = O(d_N T).$$

To this end, we apply Proposition B.1 of Kato, Galvao, Montes-Rojas (2012) to the class of functions $\tilde{\mathcal{G}}_{i,\delta_N}:=\{g-\mathrm{E}[g(\xi_{i1})]:g\in\mathcal{G}_{\delta_N}\}$. Note that $\tilde{\mathcal{G}}_{i,\delta_N}$ is pointwise measurable and each of the element is bounded by 2. Because of Lemmas 2.6.15, 2.6.18, 2.6.7, and 2.7.1 of van der Vaart and Wellner (1996) condition B5, an estimate of an upper bound $N(\tilde{\mathcal{G}}_\infty,L_2(Q),2\epsilon)$ of the class $\tilde{\mathcal{G}}_\infty:=\{g_{\alpha,\boldsymbol{\beta},\pi}:\alpha\in\mathbb{R},\boldsymbol{\beta}\in\mathbb{R}^p,\pi\in\Pi\}$ is $(A/\epsilon)^v$ for some constant $A>3e^{1/2}$ and $v>1$, every $0<\epsilon<1$ and probability measure Q. Therefore, $N(\tilde{\mathcal{G}}_{i,\delta_N},L_2(Q),2\epsilon)\le(A/\epsilon)^v$

independent of i and N. Combining the fact that

$$\begin{aligned}
\mathrm{E}[g_{\alpha,\boldsymbol{\beta},\pi}(\xi_{i1})^2] =&\mathrm{E}[|1\{y \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}\}1\{\pi(\boldsymbol{x}) > 1-\tau+c_N\} - 1\{y \le \alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|] \\
=&\mathrm{E}[|\mathrm{P}\{y^* \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}|\boldsymbol{x}\}1\{\alpha + \boldsymbol{x}^\top\boldsymbol{\beta} > 0\}1\{\pi(\boldsymbol{x}) > 1-\tau+c_N\} \\
&- \mathrm{P}\{y^* \le \alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0|\boldsymbol{x}\}1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|] \\
\le&\mathrm{E}[|\mathrm{P}\{y^* \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}|\boldsymbol{x}\}1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\} \\
&- \mathrm{P}\{y^* \le \alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0|\boldsymbol{x}\}1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|] \\
&+ \mathrm{E}[|\mathrm{P}\{y^* \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}|\boldsymbol{x}\}1\{\alpha + \boldsymbol{x}^\top\boldsymbol{\beta} > 0\}1\{\pi(\boldsymbol{x}) > 1-\tau+c_N\} \\
&- \mathrm{P}\{y^* \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}|\boldsymbol{x}\}1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|].
\end{aligned}$$

The first term equals $\mathrm{E}[f(0|\boldsymbol{x})(\alpha-\alpha_0) + f(0|\boldsymbol{x})\boldsymbol{x}^\top(\boldsymbol{\beta}-\boldsymbol{\beta}_0)] + o(\delta_N)$. The second term

$$\begin{aligned}
&\mathrm{E}[|\mathrm{P}\{y^* \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}|\boldsymbol{x}\}1\{\alpha + \boldsymbol{x}^\top\boldsymbol{\beta} > 0\}1\{\pi(\boldsymbol{x}) > 1-\tau+c_N\} \\
&- \mathrm{P}\{y^* \le \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}|\boldsymbol{x}\}1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|] \\
\le&\mathrm{E}[|1\{\alpha + \boldsymbol{x}^\top\boldsymbol{\beta} > 0\}1\{\pi(\boldsymbol{x}) > 1-\tau+c_N\} - 1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|] \\
\le&\mathrm{E}[|1\{\pi(\boldsymbol{x}) > 1-\tau+c_N\} - 1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N\}|] + \mathrm{E}[|1\{\alpha + \boldsymbol{x}^\top\boldsymbol{\beta} > 0\} - 1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0\}|] \\
=&\mathrm{E}[1\{\pi(\boldsymbol{x}) > 1-\tau+c_N \ge \pi_0(\boldsymbol{x})\} + 1\{\pi_0(\boldsymbol{x}) > 1-\tau+c_N \ge \pi(\boldsymbol{x})\}] \\
&+ \mathrm{E}[1\{\alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0 > 0 \ge \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}\} + 1\{\alpha + \boldsymbol{x}^\top\boldsymbol{\beta} > 0 \ge \alpha_0 + \boldsymbol{x}^\top\boldsymbol{\beta}_0\}] = o(T^{-1/4} \vee \delta_N).
\end{aligned}$$

Now all the conditions of Proposition B.1 are satisfied and we obtain the conclusion.

Step 3: Expansion of $H_{Ni}^{(1)}(\alpha_i, \boldsymbol{\beta}, \pi_i)$ and $H_N^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi})$. Rewrite $H_{Ni}^{(1)}(\alpha_i, \boldsymbol{\beta}, \pi_i)$ as

$$H_{Ni}^{(1)}(\alpha_i, \boldsymbol{\beta}, \pi_i) = b_1(\alpha_i, \boldsymbol{\beta}) + b_2(\pi_i) + b_3(\alpha_i, \boldsymbol{\beta}, \pi_i),$$

where

$$\begin{aligned}
b_1(\alpha_i, \boldsymbol{\beta}) &= \mathrm{E}[\psi(\cdot; \alpha_i, \boldsymbol{\beta})1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}], \\
b_2(\pi_i) &= \mathrm{E}[\psi(\cdot; \alpha_{i0}, \boldsymbol{\beta}_0)(1\{\pi_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})], \\
b_3(\alpha_i, \boldsymbol{\beta}, \pi_i) &= \mathrm{E}[(\psi(\cdot; \alpha_i, \boldsymbol{\beta}) - \psi(\cdot; \alpha_{i0}, \boldsymbol{\beta}_0))1\{\pi_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})],
\end{aligned}$$

and $H_N^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi})$ as

$$H_N^{(2)}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) = d_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) + d_2(\pi_i) + d_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \pi),$$

where

$$\begin{aligned}
d_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{1}{N}\sum_{i=1}^N \mathrm{E}[\psi(\cdot; \alpha_i, \boldsymbol{\beta})1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}\boldsymbol{x}_{i1}] \\
d_2(\boldsymbol{\pi}) &= \frac{1}{N}\sum_{i=1}^N \mathrm{E}[\psi(\cdot; \alpha_{i0}, \boldsymbol{\beta}_0)(1\{\pi_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})\boldsymbol{x}_{i1}] \\
d_3(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi}) &= \frac{1}{N}\sum_{i=1}^N \mathrm{E}[(\psi(\cdot; \alpha_i, \boldsymbol{\beta}) - \psi(\cdot; \alpha_{i0}, \boldsymbol{\beta}_0))(1\{\pi_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})\boldsymbol{x}_{i1}].
\end{aligned}$$

We simplify each of the terms. For $b_1(\alpha_i, \boldsymbol{\beta})$,

$$\begin{aligned}
b_1(\alpha_i, \boldsymbol{\beta}) =& \mathrm{E}[\psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_i, \boldsymbol{\beta}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}] \\
=& \mathrm{E}[(\tau - \mathrm{P}\{y_{i1} < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} | \boldsymbol{x}_{i1}\}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}] \\
=& \mathrm{E}[(\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} | \boldsymbol{x}_{i1}\} - \mathrm{P}\{y_{i1} < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} | \boldsymbol{x}_{i1}\}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}] \\
=& - \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}(\alpha_i - \alpha_{i0})] \\
& - \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) \boldsymbol{x}_{i1}^\top 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)] + o((\alpha_i - \alpha_{i0}) + ||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) \\
& + \mathrm{E}[\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}\} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}].
\end{aligned}$$

Let $a_i = \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) 1\{\alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0 > 0\}]$. Note that by condition B1

$$\begin{aligned}
& (a_i - \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})(\alpha_i - \alpha_{i0}) \\
=& \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})(\alpha_i - \alpha_{i0})] \\
=& \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) 1\{1-\tau < \pi_{i0}(\boldsymbol{x}_{i1}) \le 1-\tau+c_N\}(\alpha_i - \alpha_{i0})] = O(c_N|\alpha_i - \alpha_{i0}|) = o(|\alpha_i - \alpha_{i0}|).
\end{aligned}$$

Let $\boldsymbol{A}_i = \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) \boldsymbol{x}_{i1} 1\{\alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0 > 0\}]$. Note that, using B1

$$\begin{aligned}
& (\boldsymbol{A}_i - \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) \boldsymbol{x}_{i1} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\
=& \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) \boldsymbol{x}_{i1}(1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})(\boldsymbol{\beta} - \boldsymbol{\beta}_0)] \\
=& \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) \boldsymbol{x}_{i1} 1\{1-\tau < \pi_{i0}(\boldsymbol{x}_{i1}) \le 1-\tau+c_N\}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)] = O(c_N||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) = o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||).
\end{aligned}$$

For the third term in $b_1(\cdot)$ and B6,

$$\begin{aligned}
& \mathrm{E}[\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}\} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}] \\
=& \mathrm{E}[\mathrm{P}\{y_{i1}^* < \alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0\} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}] \\
& + \mathrm{E}[(\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}\} - \mathrm{P}\{y_{i1}^* < \alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0\}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}] \\
=& \tau \mathrm{E}[1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}] \\
& + \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(\alpha_i - \alpha_{i0}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}] + o(\alpha_i - \alpha_{i0}) \\
& + \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) \boldsymbol{x}_{i1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \le 0\}] + o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) \\
=& - \tau\left(\boldsymbol{D}_{N1}^*(\alpha_i - \alpha_{i0}) + \boldsymbol{D}_{N2}^*(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right) + o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) + o(\alpha_i - \alpha_{i0}).
\end{aligned}$$

Thus, $b_1(\alpha_i, \boldsymbol{\beta}) = -(a_i + \tau \boldsymbol{D}_{N1}^*)(\alpha_i - \alpha_{i0}) - (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N2}^*)(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) + o(\alpha_i - \alpha_{i0})$.

For $b_2(\pi_i)$, because $\sup_i ||\pi_i - \pi_{i0}||_\infty = o_p(T^{-1/4})$, and $T^{1/4} c_N > c^* > 0$, then $\pi_i(\boldsymbol{x}) > 1 - \tau + c_N$ implies $\pi_{i0}(\boldsymbol{x}) > 1 - \tau$, and therefore $\alpha_{i0} + \boldsymbol{x}^\top \boldsymbol{\beta}_0 > 0$. It follows that

$$\begin{aligned}
b_2(\pi_i) =& \mathrm{E}[\psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_{i0}, \boldsymbol{\beta}_0)(1\{\pi_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})] \\
=& \mathrm{E}[\psi(y_{i1}^*, \boldsymbol{x}_{i1}; \alpha_{i0}, \boldsymbol{\beta}_0)(1\{\pi_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})] = 0.
\end{aligned}$$

Regarding $b_3(\alpha_i, \boldsymbol{\beta}, \pi_i)$,

$$\begin{aligned}||b_3(\alpha_i, \boldsymbol{\beta}, \pi_i)|| =&\mathrm{E}[(\psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_i, \boldsymbol{\beta}) - \psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_{i0}, \boldsymbol{\beta}_0))\\ &\times (1\{\pi_i(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\})]\\ \leq&\mathrm{E}[|\mathrm{P}\{y_{i1} < \alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0 | \boldsymbol{x}_{i1}\} - \mathrm{P}\{y_{i1} < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} | \boldsymbol{x}_{i1}\}|]\\ &\times (1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N \geq \pi(\boldsymbol{x}_{i1})\} + 1\{\pi_i(\boldsymbol{x}_{i1}) > 1 - \tau + c_N \geq \pi_0(\boldsymbol{x}_{i1})\})]\\ \leq&\mathrm{E}[(1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N \geq \pi(\boldsymbol{x}_{i1})\} + 1\{\pi_i(\boldsymbol{x}_{i1}) > 1 - \tau + c_N \geq \pi_0(\boldsymbol{x}_{i1})\})].\end{aligned}$$

It follows that $\sup_{|\alpha - \alpha_{i0}| + ||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \leq \epsilon_N, ||\pi - \pi_0||_\infty = o_p(T^{-1/4})} ||b_3(\alpha_i, \boldsymbol{\beta}, \pi_i)|| = o(T^{-1/4})$.

Similarly to $b_1(\alpha_i, \boldsymbol{\beta})$, for $d_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$,

$$\begin{aligned}&d_1(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[\boldsymbol{x}_{i1}\psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_i, \boldsymbol{\beta}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\}]\\ =&\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[(\tau - \mathrm{P}\{y_{i1} < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} | \boldsymbol{x}_{i1}\}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} \boldsymbol{x}_{i1}]\\ =&-\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[f_i(0|\boldsymbol{x}_{i1}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\}(\alpha_i - \alpha_{i0})\boldsymbol{x}_{i1}]\\ &-\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^\top 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)] + o(\max_{1\leq i\leq N}\{\alpha_i - \alpha_{i0}\}) + ||\boldsymbol{\beta} - \boldsymbol{\beta}_0||)\\ &+\frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}\} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}\boldsymbol{x}_{i1}].\end{aligned}$$

Using the same argument as before, note that

$$(\boldsymbol{A}_i - \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\})(\alpha_i - \alpha_{i0}) = O(c_N||\alpha_i - \alpha_{i0}||) = o(||\alpha_i - \alpha_{i0}||).$$

Let $\boldsymbol{B}_i = \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^\top 1\{\alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0 > 0\}]$. By the same derivation, we have that by B1

$$(\boldsymbol{B}_i - \mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^\top 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\}](\boldsymbol{\beta} - \boldsymbol{\beta}_0) = O(c_N||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) = o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||).$$

For the third term in $d_1(\cdot)$,

$$\begin{aligned}&\mathrm{E}[\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}\} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}\boldsymbol{x}_{i1}]\\ =&\mathrm{E}[\mathrm{P}\{y_{i1}^* < \alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0\} 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}\boldsymbol{x}_{i1}]\\ &+\mathrm{E}[(\mathrm{P}\{y_{i1}^* < \alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}\} - \mathrm{P}\{y_{i1}^* < \alpha_{i0} + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta}_0\}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}\boldsymbol{x}_{i1}]\\ =&\tau\mathrm{E}[1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}\boldsymbol{x}_{i1}]\\ &+\mathrm{E}[f_i(0|\boldsymbol{x}_{i1})(\alpha_i - \alpha_{i0}) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}\boldsymbol{x}_{i1}] + o(\alpha_i - \alpha_{i0})\\ &+\mathrm{E}[f_i(0|\boldsymbol{x}_{i1})\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau + c_N\} 1\{\alpha_i + \boldsymbol{x}_{i1}^\top \boldsymbol{\beta} \leq 0\}] + o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||)\\ =&-\tau\left(\boldsymbol{D}_{N3}^*(\alpha_i - \alpha_{i0}) + \boldsymbol{D}_{N4}^*(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right) + o(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) + o(\alpha_i - \alpha_{i0}).\end{aligned}$$

Thus, $d_1(\boldsymbol{\alpha},\boldsymbol{\beta}) = -\frac{1}{N}\sum_{i=1}^N(\boldsymbol{A}_i+\tau\boldsymbol{D}_{N3}^*)(\alpha_i-\alpha_{i0}) - \frac{1}{N}\sum_{i=1}^N(\boldsymbol{B}_i+\tau\boldsymbol{D}_{N4}^*)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+o(||\boldsymbol{\beta}-\boldsymbol{\beta}_0||)+o(\max\{\alpha_i-\alpha_{i0}\})$. Using the same arguments than for $b_2(\pi_i)=0$, we have that $d_2(\boldsymbol{\pi})=0$. For $d_3(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi})$,

$$
\begin{aligned}
||d_3(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi})|| =&\frac{1}{N}\sum_{i=1}^N \mathrm{E}[(\psi(y_{i1},\boldsymbol{x}_{i1};\alpha_i,\boldsymbol{\beta})-\psi(y_{i1},\boldsymbol{x}_{i1};\alpha_{i0},\boldsymbol{\beta}_0))\\
&\times(1\{\pi_i(\boldsymbol{x}_{i1})>1-\tau+c_N\}-1\{\pi_{i0}(\boldsymbol{x}_{i1})>1-\tau+c_N\})]\\
\le&\frac{1}{N}\sum_{i=1}^N \mathrm{E}[|\mathrm{P}\{y_{i1}<\alpha_{i0}+\boldsymbol{x}_{i1}^\top\boldsymbol{\beta}_0|\boldsymbol{x}_{i1}\}-\mathrm{P}\{y_{i1}<\alpha_i+\boldsymbol{x}_{i1}^\top\boldsymbol{\beta}|\boldsymbol{x}_{i1}\}|]\\
&\times(1\{\pi_{i0}(\boldsymbol{x}_{i1})>1-\tau+c_N\ge\pi(\boldsymbol{x}_{i1})\}+1\{\pi_i(\boldsymbol{x}_{i1})>1-\tau+c_N\ge\pi_0(\boldsymbol{x}_{i1})\})]\\
\le&\frac{1}{N}\sum_{i=1}^N \mathrm{E}[(1\{\pi_{i0}(\boldsymbol{x}_{i1})>1-\tau+c_N\ge\pi(\boldsymbol{x}_{i1})\}+1\{\pi_i(\boldsymbol{x}_{i1})>1-\tau+c_N\ge\pi_0(\boldsymbol{x}_{i1})\})].
\end{aligned}
$$

It follows that $\sup_{|\alpha-\alpha_{i0}|+||\boldsymbol{\beta}-\boldsymbol{\beta}_0||\le\epsilon_N,||\pi-\pi_0||_\infty=o_p(T^{-1/4})}||d_3(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\pi})||=o(T^{-1/4})$.

To summarize,

$$
\begin{aligned}
H_{Ni}^{(1)}(\alpha_i,\boldsymbol{\beta},\pi_i)=&-(a_i+\tau\boldsymbol{D}_{N1}^*)(\alpha_i-\alpha_{i0})-(\boldsymbol{A}_i+\tau\boldsymbol{D}_{N2}^*)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
&+o(||\boldsymbol{\beta}-\boldsymbol{\beta}_0||)+o(\alpha_i-\alpha_{i0})+o(T^{-1/4})\\
H_N^{(2)}(\boldsymbol{\alpha},\boldsymbol{\beta},\pi)=&-\frac{1}{N}\sum_{i=1}^N(\boldsymbol{A}_i+\tau\boldsymbol{D}_{N3}^*)(\alpha_i-\alpha_{i0})-\frac{1}{N}\sum_{i=1}^N(\boldsymbol{B}_i+\tau\boldsymbol{D}_{N4}^*)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\
&+o(||\boldsymbol{\beta}-\boldsymbol{\beta}_0||)+o(\max_{1\le i\le N}\{\alpha_i-\alpha_{i0}\})+o(T^{-1/4}).
\end{aligned}
$$

Step 4: Representation of $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0$. From Steps 1 and 2, we have that

$$
\frac{1}{N}\sum_{i=1}^N m_i\{\mathbb{H}_{Ni}^{(1)}(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i)-H_{Ni}^{(1)}(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i)-\mathbb{H}_{Ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0,\pi_{i0})\}=O_p(d_N)
$$

$$
H_N^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat\pi)=\mathbb{H}_N^{(2)}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}},\hat\pi)-\mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\pi_0)+O_p(d_N)=O_p(T^{-1})+\mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\pi_0)+O_p(d_N).
$$

Hence,

$$
\begin{aligned}
&\mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0,\boldsymbol{\beta}_0,\pi_0)+O_p(d_N)=-\frac{1}{N}\sum_{i=1}^N(\boldsymbol{A}_i+\tau\boldsymbol{D}_{N3}^*)(\hat\alpha_i-\alpha_{i0})-\frac{1}{N}\sum_{i=1}^N(\boldsymbol{B}_i+\tau\boldsymbol{D}_{N4}^*)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\\
&+o_p(||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||)+o_p(\max_{1\le i\le N}\{\hat\alpha_i-\alpha_{i0}\})+o(T^{-1/4}).
\end{aligned}
$$

Solving for $\hat\alpha_i-\alpha_{i0}$ and $\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0$, we obtain

$$
\begin{aligned}
\hat\alpha_i-\alpha_{i0}=&-(a_i+\tau\boldsymbol{D}_{N1}^*)^{-1}\left[H_{Ni}^{(1)}(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i)-(\boldsymbol{A}_i+\tau\boldsymbol{D}_{N2}^*)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right]\\
&+o_p(||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||)+o_p(\hat\alpha_i-\alpha_{i0})+o_p(T^{-1/4})
\end{aligned}
$$

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = & - \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} \left[H_N^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\pi}) - \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N3}^*)(\hat{\alpha}_i - \alpha_{i0}) \right] \\
& + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + o_p(\max_{1 \le i \le N} \{\hat{\alpha}_i - \alpha_{i0}\}) + o_p(T^{-1/4}),
\end{aligned}
$$

and plugging $\hat{\alpha}_i - \alpha_{i0}$ into the second equation, we have

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = & - \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} H_N^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\pi}) \\
& + \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N3}^*)(a_i + \tau \boldsymbol{D}_{N1}^*)^{-1} H_{Ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}, \hat{\pi}_i) \\
& + \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N3}^*)(a_i + \tau \boldsymbol{D}_{N1}^*)^{-1} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N2}^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
& + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + o_p(\max\{\hat{\alpha}_i - \alpha_{i0}\}) + o_p(T^{-1/4}).
\end{aligned}
$$

Solve for $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$, and using the results in Steps 1 and 2 we have

$$
\begin{aligned}
& \left(I - \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N3}^*)(a_i + \tau \boldsymbol{D}_{N1}^*)^{-1} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N2}^*) \right) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
= & \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} \mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \pi_0) + O_p(T^{-1}) + O_p(d_N) \\
& + \left(\frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*) \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N3}^*)(a_i + \tau \boldsymbol{D}_{N1}^*)^{-1} \mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0}) \\
& + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + O_p(\max\{(\hat{\alpha}_i - \alpha_{i0})^2\}) + o_p(T^{-1/4}).
\end{aligned}
$$

Step 5: Rate of $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$. From above, we have

$$
||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| = O_p(\max\{(\alpha_i - \alpha_{i0})^2\}) + o_p(T^{-1/4}).
$$

Therefore, $\max\{|\hat{\alpha}_i - \alpha_{i0}|\}$ is bounded with probability approaching one by

$$
k\{\max_{1 \le i \le N} |\mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0})| + \max_{1 \le i \le N} |\mathbb{H}_{Ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}, \hat{\pi}_i) - H_{Ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}, \hat{\pi}_i) - \mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0})|\} + o_p(T^{-1/4})
$$

where k is a constant. First, observe that for any $K > 0$,

$$
\mathrm{P}\{\max_{1 \le i \le N} |\mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0})| > (T/\log N)^{-1/2} K\} \le \sum_{i=1}^{N} \mathrm{P}\{|\mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0})| > (T/\log N)^{-1/2} K\}
$$

and the right side is bounded by $2N^{1-K^2/2}$ by Hoeffding's inequality. This implies that $\max_{1\le i\le N}|\mathbb{H}_{Ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0,\pi_{i0})| = O_p((T/\log N)^{-1/2})$. We next show that

$$\max_{1\le i\le N}|\mathbb{H}_{Ni}^{(1)}(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i) - H_{Ni}^{(1)}(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i) - \mathbb{H}_{Ni}^{(1)}(\alpha_{i0},\boldsymbol{\beta}_0,\pi_{i0})| = o_p((T/\log N)^{-1/2}).$$

We assume $\alpha_{i0}=\alpha_0$, $\boldsymbol{\beta}=\boldsymbol{\beta}_0$, and $\pi_{i0}=\pi_0$. Let $\mathcal{G}_\delta$ and ξ_{it} be the same as before. Because of the consistency of $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat\pi$, it suffices to show that for every $\epsilon>0$, there is a sufficiently small $\delta>0$ such that

$$\max_{1\le i\le N}\mathrm{P}\left\{\left|\left|\sum_{t=1}^{T}\{g(\xi_{it})-\mathrm{E}[g(\xi_{it})]\}\right|\right|_{\mathcal{G}_\delta} > (T\log N)^{1/2}\epsilon\right\} = o(N^{-1}).$$

To this end, we make use of Bousquet's version of Talagrand's inequality (see Bousquet (2002)). Fix $\epsilon>0$. Put $Z_i := ||\sum_{t=1}^{T}\{g(\xi_{it})-\mathrm{E}[g(\xi_{it})]\}||_{\mathcal{G}_\delta}$. By Proposition B.2 of Kato, Galvao, and Montes-Rojas (2012), for all $s>0$, with probability at least $1-e^{-s^2}$, we have

$$Z_i \le \mathrm{E}[Z_i] + s\sqrt{2\{T(\delta+(NT)^{-1/4})+4\mathrm{E}[Z_i]\}} + \frac{2s^2}{3}.$$

Take $s=\sqrt{2\log N}$. Then, there exist a constant δ and N_0 independent of i and N such that the right side of the previous inequality is smaller than $(T\log N)^{1/2}$ for all $N>N_0$. This implies that $\max_{1\le i\le N}\mathrm{P}\{Z_i>(T\log N)^{1/2}\epsilon\}\le N^{-2}$. Therefore, we have $\max_{1\le i\le N}|\hat\alpha_i-\alpha_{i0}| = O_p((T/\log N)^{-1/2})+o_p(T^{-1/4}) = o_p(T^{-1/4})$. For the second result, we have $||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|| = o_p((T/\log N)^{-1/2}\vee T^{-1/4}) = o_p(T^{-1/4})$.

Step 6: Rates improvement. By Assumptions B3 and B4, $\pi_{i0}(\boldsymbol{x})>1-\tau+c_N$ implies $\alpha_{i0}+\boldsymbol{x}^\top\boldsymbol{\beta}_0>d_N$, where $d_NT^{1/4}$ is greater than some constant. Now $\max\{\max_{1\le i\le N}|\hat\alpha_i-\alpha_{i0}|, ||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||\} = o_p(T^{-1/4})$. Therefore,

$$\begin{aligned}
&\boldsymbol{D}_{N1}^*(\hat\alpha_i-\alpha_{i0}) = o_p(\hat\alpha_i-\alpha_{i0}), \quad \boldsymbol{D}_{N2}^*(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = o_p(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0),\\
&\boldsymbol{D}_{N3}^*(\hat\alpha_i-\alpha_{i0}) = o_p(\hat\alpha_i-\alpha_{i0}), \quad \boldsymbol{D}_{N4}^*(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = o_p(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0).
\end{aligned}$$

Since $T^{1/4}c_N>c^*>0$ and $||\hat\pi-\pi_0|| = o_p(T^{-1/4})$, we have both $\pi_{i0}(\boldsymbol{x})>1-\tau+c_N$ and $\hat\pi_i(\boldsymbol{x})>1-\tau+c_N$ imply $\alpha_{i0}+\boldsymbol{x}^\top\boldsymbol{\beta}_0>0$ and $\hat\alpha_i+\boldsymbol{x}^\top\hat{\boldsymbol{\beta}}>0$. Therefore, evaluating $b_3(\cdot)$ at $(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i)$ we obtain

$$\begin{aligned}
b_3(\hat\alpha_i,\hat{\boldsymbol{\beta}},\hat\pi_i) =& \mathrm{E}[(\psi(y_{i1},\boldsymbol{x}_{i1};\alpha_i,\boldsymbol{\beta})-\psi(y_{i1},\boldsymbol{x}_{i1};\alpha_{i0},\boldsymbol{\beta}_0))\\
&\times(1\{\hat\pi_i(\boldsymbol{x}_{i1})>1-\tau+c_N\}-1\{\pi_{i0}(\boldsymbol{x}_{i1})>1-\tau+c_N\})]\\
=&\mathrm{E}[\boldsymbol{x}_{i1}^\top f_i(0|\boldsymbol{x}_{i1})\{\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0+o_p(||\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0||)\} + f_i(0|\boldsymbol{x}_{i1})\{\hat\alpha_i-\alpha_{i0}+o_p(\hat\alpha_i-\alpha_{i0})\}\\
&\times(1\{\hat\pi_i(\boldsymbol{x}_{i1})>1-\tau+c_N\}-1\{\pi_{i0}(\boldsymbol{x}_{i1})>1-\tau+c_N\})]\\
=&o_p(|\hat\alpha_i-\alpha_{i0}|\vee|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|).
\end{aligned}$$

Similarly, evaluating $d_3(\cdot)$ at $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\pi})$

$$\begin{aligned} d_3(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\pi}) =& \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}[(\psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_i, \boldsymbol{\beta}) - \psi(y_{i1}, \boldsymbol{x}_{i1}; \alpha_{i0}, \boldsymbol{\beta}_0)) \\ & \times (1\{\hat{\pi}_i(\boldsymbol{x}_{i1}) > 1-\tau+c_N\} - 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1-\tau+c_N\})] \\ =& o_p(\max|\hat{\alpha}_i - \alpha_{i0}| \vee |\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|). \end{aligned}$$

To summarize,

$$\begin{aligned} H_{Ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}, \hat{\pi}_i) =& -(a_i + \tau \boldsymbol{D}_{N1}^*)(\hat{\alpha}_i - \alpha_{i0}) - (\boldsymbol{A}_i + \tau \boldsymbol{D}_{N2}^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + o_p(\hat{\alpha}_i - \alpha_{i0}) \\ H_N^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\pi}) =& -\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{A}_i + \tau \boldsymbol{D}_{N3}^*)(\hat{\alpha}_i - \alpha_{i0}) - \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{B}_i + \tau \boldsymbol{D}_{N4}^*)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ & + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + o_p(\max_{1 \le i \le N}\{\hat{\alpha}_i - \alpha_{i0}\}). \end{aligned}$$

Therefore,

$$\begin{aligned} & \left(I - \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{B}_i\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i a_i^{-1}\boldsymbol{A}_i\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ =& -\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{B}_i\right)^{-1}\left[H_N^{(2)}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{\pi}) + \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i a_i^{-1} H_{Ni}^{(1)}(\hat{\alpha}_i, \hat{\boldsymbol{\beta}}, \hat{\pi}_i)\right] + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + O_p(\max\{(\hat{\alpha}_i - \alpha_{i0})^2\}) \\ =& \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{B}_i\right)^{-1}\left[\mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \pi_0) + O_p(T^{-1}) + O_p(d_N) - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i a_i^{-1}\mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0})\right] \\ & + o_p(||\boldsymbol{\beta} - \boldsymbol{\beta}_0||) + O_p(\max\{(\alpha_i - \alpha_{i0})^2\}). \end{aligned}$$

After going over Steps 2–5 again without the term $o_p(T^{-1/4})$, we obtain $\max|\hat{\alpha}_i - \alpha_{i0}| \vee ||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0|| = O_p((T/\log N)^{-1/2})$. Therefore, we can set $d_N = \frac{1}{T}|\log \delta_N| \vee \frac{\delta_N^{1/2}}{T^{1/2}}|\log \delta_N|^{1/2}$.

Finally,

$$\begin{aligned} & \left(I - \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{B}_i\right)^{-1} \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i a_i^{-1}\boldsymbol{A}_i\right)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{B}_i\right)^{-1}\mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \pi_0) + O_p(d_N) \\ & - \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{B}_i\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i a_i^{-1}\mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0}) + o_p(||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0||) + O_p(\max\{(\hat{\alpha}_i - \alpha_{i0})^2\}). \end{aligned}$$

Step 7: Weak convergence As $N^2(\log N)^3/T \to 0$,

$$\left(\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{B}_i - \boldsymbol{A}_i a_i^{-1}\boldsymbol{A}_i)\right)\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{NT}(\mathbb{H}_N^{(2)}(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \pi_0) - \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{A}_i a_i^{-1}\mathbb{H}_{Ni}^{(1)}(\alpha_{i0}, \boldsymbol{\beta}_0, \pi_{i0})) + o_p(1)$$

$$=\sqrt{NT}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi(y_{it}, \boldsymbol{x}_{it}; \alpha_{i0}, \boldsymbol{\beta}_0)\boldsymbol{x}_{it}1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau + c_N\}\right.$$

$$\left.-\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{A}_i a_i^{-1}\psi(y_{it}, \boldsymbol{x}_{it}; \alpha_{i0}, \boldsymbol{\beta}_0)1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau + c_N\}\right) + o_p(1)$$

$$=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tau - 1\{y_{it} \le \alpha_{i0} + \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0\})(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau + c_N\} + o_p(1).$$

Let $\boldsymbol{V} = \tau(1-\tau)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[(\boldsymbol{x}_{i1} - \boldsymbol{A}_i a_i^{-1})(\boldsymbol{x}_{i1} - \boldsymbol{A}_i a_i^{-1})^\top 1\{\pi_{i0}(\boldsymbol{x}_{i1}) > 1 - \tau\}] = \tau(1-\tau)\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}[(\boldsymbol{x}_{i1} - \boldsymbol{A}_i a_i^{-1})(\boldsymbol{x}_{i1} - \boldsymbol{A}_i a_i^{-1})^\top 1\{\alpha_{i0} + \boldsymbol{x}_{i1}^\top\boldsymbol{\beta}_0 > 0\}]$. Note that

$$Cov\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tau - 1\{y_{it} \le \alpha_{i0} + \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0\})(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau + c_N\}\right\}$$

$$=\tau(1-\tau)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})^\top 1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau + c_N\}.$$

By A4, $\frac{1}{T}\sum_{t=1}^{T}1\{1 - \tau < \pi_{i0}(\boldsymbol{x}_{it}) \le 1 - \tau + c_N\} = O_p(c_N) = o_p(1)$, which implies

$$Cov\left\{\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tau - 1\{y_{it} \le \alpha_{i0} + \boldsymbol{x}_{it}^\top\boldsymbol{\beta}_0\})(\boldsymbol{x}_{it} - \boldsymbol{A}_i a_i^{-1})1\{\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau + c_N\}\right\} \to \boldsymbol{V}.$$

By B6, $\boldsymbol{\Lambda} = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}[\boldsymbol{B}_i - \boldsymbol{A}_i a_i^{-1}\boldsymbol{A}_i^\top]1(\pi_{i0}(\boldsymbol{x}_{it}) > 1 - \tau)$, and CLT it follows that

$$\sqrt{NT}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \boldsymbol{\Lambda}^{-1}\boldsymbol{V}\boldsymbol{\Lambda}^{-1}).$$

□