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

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

January 16, 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 sequential 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, Michael Price, Liang Wang, and seminar participants at Northwestern University, University of Kentucky, University of Tennessee, 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. The R software for the method introduced in this paper is available upon request.

†University of Wisconsin-Milwaukee, Department of Economics, Bolton Hall 852, 3210 N. Maryland Ave., Milwaukee, WI 53201; and University of Iowa, W210 Pappajohn Business Building, 21 E. Market Street, Iowa City, IA 52242. E-mail: agalvaouwm.edu

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

§University of Tennessee, Department of Economics, 527A Stokely Management Center, Knoxville, Tennessee, 37996. 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 censored data, standard regression methods employed to estimate linear conditional mean models lead to inconsistent estimates of the parameters of interest. A classical model to deal with censored data is the Tobit model, introduced by Tobin (1958), as well as its variations such as the Tobit type II-IV models. Censored regression models are usually estimated using likelihood techniques and the general validity of this approach has been shown by Schnedler (2005) who also provides a method 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 and may be subject to shrinkage toward a common value as in the Gaussian random effects model. 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 (2011), 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 and group member specific errors in the group 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 in longitudinal studies. 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 the 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 by 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 for empirical research, two approaches are proposed. Both estimators are based on QR fixed effects models with a separation restriction on the censoring probability. The first method is a three-step estimator in which the first step selects observations by using a parametric regression to estimate the conditional censoring probability. The second method is a two-step estimator in which the first step selects a subset of observations by estimating nonparametrically the propensity score. The asymptotic properties of the estimators are derived by using sequential limits. This is an important innovation because it facilitates the derivation of the limiting results and can be utilized in future works with QR panel data models.

Monte Carlo simulations are conducted to evaluate the performance of the proposed methods in small samples. The simulations indicate that the estimators offer good finite-sample performance in terms of bias, mean squared error, and coverage probability of confidence interval. We also consider an empirical application to investigate whether the 1964 Civil Rights Act contributed to reduce the black-white wage gap. In particular, we divided the sample into young and mature black workers and estimated a quantile difference-in-difference model for each group. Our approach is easy to implement in practice and indicates that the 1964 Civil Rights Act had a small and insignificant effect on mature workers and low-income
young workers. However, the findings show that the policy dramatically reduced the earning gap among high-income young workers. 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 proposed estimators, and studies their large sample properties. Section 3 presents a series of simulations to investigate the finite sample properties of the estimators. An empirical illustration is considered in Section 4. Section 5 concludes.

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.\(^1\) Moreover, \( y^*_{it} \) is assumed to be conditionally independent of the censoring point \( C_{it} \), such that, conditional on covariates, \( x_{it} \), and an individual effect, \( \alpha_i \),

\[
P(y^*_{it} < y|x_{it}, \alpha_i, C_{it}) = P(y^*_{it} < y|x_{it}, \alpha_i).
\]

Thus, the quantile regression (QR) model is given by,

\[
Q_{y^*_{it}}(\tau|x_{it}, \alpha_{i0}) = \alpha_{i0}(\tau) + x_{it}^T \beta_0(\tau), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,
\]  

(2.1)

where \( x_{it} \) is a \( p \times 1 \) vector of regressors, \( \beta_0(\tau) \) is a \( p \times 1 \) vector of parameters, \( \alpha_{i0}(\tau) \) is a scalar individual effect for each \( i \), \( \tau \in (0,1) \) is a quantile of interest, and \( Q_{y^*_{it}}(\tau|x_{it}, \alpha_{i0}) = \inf\{y : Pr(y^*_{it} < y|x_{it}, \alpha_{i0}) \geq \tau\} \) is the conditional \( \tau \)-quantile of \( y^*_{it} \) given \( (x_{it}, \alpha_{i0}) \). In addition, \( \alpha_{i0}(\tau) \) is a quantile-specific individual effect, which is intended to capture individual specific sources of variability, or unobserved heterogeneity that was not adequately controlled by other covariates in the model. In general, each \( \alpha_{i0}(\tau) \) and \( \beta_0(\tau) \) can depend on \( \tau \), but we assume \( \tau \) to be fixed throughout the paper and suppress such a dependence for notational simplicity. The model is semiparametric in the sense that the functional form of the conditional distribution of \( y^*_{it} \) given \( (x_{it}, \alpha_{i0}) \) is left unspecified and no parametric assumption is made on the relation between \( x_{it} \) and \( \alpha_{i0} \).

---

\(^1\)The assumption that censoring points are known for all \( (i,t) \) is realistic in many (but clearly not all) applications. For instance, in the application offered in Section 4, the censoring point is known. In this paper, we only consider known censoring and leave unknown censoring for future research. Work on unobserved censoring for quantile regression cross-section models include, among others, Yang (1997), Honoré, Khan, and Powell (2002), Portnoy (2003), and Wang and Wang (2009).
Equivariance to monotone transformation is an important property of QR models. For a given monotone transformation \( \mathcal{I} \) of variable \( y^* \),
\[
Q_{\mathcal{I}c}(y^*_{it}) = \mathcal{I}(Q_{y^*_{it}}(\tau|x_{it}, \alpha_{i0})).
\]
The transformation equivariance of (2.1) naturally leads to a version of the Powell’s censored QR model,
\[
Q_{y^*_{it}}(\tau|x_{it}, \alpha_{i0}, C_{it}) = \max(C_{it}, \alpha_{i0} + x_{it}^\top \beta_0).
\]

We consider the fixed effects estimation of \( \beta_0 \), which is implemented by treating each individual effect as a parameter to be estimated. Throughout the paper, as in Hahn and Newey (2004) and Fernandez-Val (2005), we treat \( \alpha_i \) as fixed by conditioning on them. Thus, the structural parameter of interest, \( \beta_0 \), can be interpreted as representing the effect of \( x_{it} \) on the \( \tau \)-th conditional quantile function of the dependent variable, \( y_{it} \), while controlling for heterogeneity, here represented by \( \alpha_i \). 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 \( (\hat{\alpha}, \hat{\beta}) \) solving the following problem:
\[
\min_{\alpha, \beta \in A \times B} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(y_{it} - \max(C_{it}, \alpha_i + x_{it}^\top \beta)),
\]
where \( \alpha := (\alpha_1, ..., \alpha_N)^\top \) and \( \rho_\tau(u) = u(\tau - 1(u < 0)) \) denotes the loss function of Koenker and Bassett (1978).

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 non-convex process, and thus iterative linear programming methods are only guaranteed to converge to local minimum [see Fitzenberger (1997) and Khan and Powell (2001)]. Additional regressors, high censoring, and large sample 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 problem, 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
Denote \( \delta_{it} = 1(y_{it}^* > C_{it}) \) to indicate uncensored observations. Let \( u_{it} := y_{it}^* - \alpha_i - x_{it}^{\top} \beta_0 \), whose \( \tau \)-th conditional quantile given \((x_{it}, \alpha_i, C_{it})\) equals zero. Because \( \pi_0(\alpha_i, x_{it}, C_{it}) := P(\delta_{it} = 1|x_{it}, \alpha_i, C_{it}) = P(u_{it} > -\alpha_i - x_{it}^{\top} \beta_0 + C_{it}|x_{it}, \alpha_i, C_{it}) \) and \( P(u_{it} > 0|x_{it}, \alpha_i, C_{it}) = 1 - \tau \), 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

\[
\sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau}(y_{it} - \alpha_i - x_{it}^{\top} \beta_0)1(\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau) \quad (2.5)
\]

This development suggests that to obtain a censored QR estimator, one can simply apply fixed effects QR to the subset \( \{ (i, t) : \pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau \} \), including all the observations, even censored ones, for which the true \( \tau \)-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 \( \delta_{it} \) and regressors. From this step, the fitted function would be used to determine the observations to be included in a panel data QR.

The concept of multi-step estimator for censored QR in cross-section models dates back to Buchinsky and Hahn (1998) and Khan and Powell (2001). Buchinsky and Hahn (1998) proposed first estimating the propensity score \( \pi_0(x_i) = P(\delta_i = 1|x_i) \) by a nonparametric kernel regression, then selecting a subset of the whole sample, where \( \{ i : \pi_0(x_i) > 1 - \tau \} \), [i.e., those observations \( i \) where the conditional quantile line is above the censoring point 0], and then using a QR on the selected sample. Analogously, Khan and Powell (2001) proposed using any of the following three methods to perform the first-stage selection: maximum score estimators, nonparametric kernel propensity score, and nonparametric locally linear conditional quantile. The two-stage estimators are somewhat less efficient than the Powell estimator because of smoothing and trimming. However, more recently, Tang, Wang, He, and Zhu (2012) developed an estimation procedure that can achieve the same asymptotic efficiency as Powell’s estimator, as long as the conditional censoring probability can be estimated consistently at an appropriate nonparametric rate and the estimated function satisfies smoothness conditions.

The suggested first stages described above are attractive but practical only in low dimen-
sions, and have slow convergence rates. Local kernel smoothers apply to (sufficiently) continuous variables only, whereas many practical applications have many (sufficiently) discrete covariates. To overcome these shortcomings, Chernozhukov and Hong (2002) used 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. They use a fixed constant $c$ to avoid bias, and seek a further step to achieve efficiency.

In the next subsections we discuss procedures that yield a consistent estimator of $\beta_0$ in (2.1). We propose two methods for censored QR model with fixed effects that rely on applying standard QR to subsets of observations selected either parametrically or nonparametrically. The first method is a parametric 3-step estimator, and the second is a nonparametric 2-step estimator. These estimators extend, respectively, the work of Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012) from cross-section to panel data with fixed effects.

### 2.2 A parametric 3-step procedure

This section describes the steps required to compute the parametric 3-step estimator.

**Step 1**: Estimate a parametric classification fixed effects model,

$$P(\delta_{it} = 1|x_{it}, \alpha_i, C_{it}) = p(\tilde{X}_{it}^\top \gamma) = p(z_{it}^\top \alpha + \tilde{x}_{it}^\top \eta),$$

where $\delta_{it} = 1(y^*_{it} > C_{it})$, $p(\cdot)$ is the link function (for example logit), $z_{it}$ is an $N$-dimensional indicator variable for the individual effect $\alpha_i$, $\tilde{x}_{it}$ is a $\hat{p}$ dimensional desired transform of $x_{it}$, $\tilde{X}_{it} = (z_{it}^\top, \tilde{x}_{it}^\top)^\top$, and $\gamma = (\alpha^\top, \eta^\top)^\top$ is an $(N + \hat{p})$ dimensional vector with $\hat{p} > p$. Notice that $p(\tilde{X}_{it}^\top \hat{\gamma})$ gives an estimator of the true propensity score $\pi_0(\alpha_i, x_{it}, C_{it})$. The transforms of the covariates, $\tilde{x}_{it}$, could be equal to a vector that includes transformations of $x_{it}$, for instance, polynomials of $x_{it}$ as in Chernozhukov and Hong (2002). Since the true propensity score function $\pi_0(\alpha_i, x_{it}, C_{it})$ is unknown, step 1 gives an inconsistent estimator of $\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau$. In order to compensate for such an inconsistency, the classification rule is adjusted by a constant $c$. In other words, we select the sample $J_0 = \{ (i, t) : p(\tilde{X}_{it}^\top \hat{\gamma}) > 1 - \tau + c \}$, where $c$ is strictly between 0 and $\tau$ and not too small. Thus, we offset the potential inconsistency of the propensity score estimator with a more conservative classification rule by adding a trimming constant $c$, which is equivalent to say that $p(\tilde{X}_{it}^\top \hat{\gamma}_0) - c$ is a lower bound on
\[ \pi_0(\alpha_i, x_{it}, C_{it}), \] that is

\[ a.s. \ p(\mathbf{X}_{it}^\top \gamma_0) - c < \pi_0(\alpha_i, x_{it}, C_{it}), \]

where \( \gamma_0 \equiv \text{plim} \ \hat{\gamma} \),

and the foregoing construction will follow Chernozhukov and Hong (2002) by assuming that the estimator \( \hat{\gamma} \) is reasonable and converges to a value \( \gamma_0 \) that minimizes a sensible distance between \( \pi_0(\alpha_i, x_{it}, C_{it}) \) and the model \( p(\mathbf{X}_{it}^\top \gamma) \). In the empirical section, we used both polynomial logistic and linear models as link functions. The goal of the first step is to select some, and not necessarily the largest, subset of observations where \( \pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau \), that is, where the quantile line \( \alpha_{i0} + x_{it}^\top \beta_0 \) is above the censoring point \( C_{it} \) so as to obtain a consistent but inefficient estimator \( \hat{\beta}_0 \). Chernozhukov and Hong (2002) suggest that a sensible rule for choosing \( c \) is to compare the size of the selected sample \( J(c) = \{(i, t) : p(\mathbf{X}_{it}^\top \hat{\gamma}) > 1 - \tau + c \} \) for \( c = 0 \) and other values. Choosing \( c = q \) th quantile of all \( p(\mathbf{X}_{it}^\top \hat{\gamma}) \) such that \( p(\mathbf{X}_{it}^\top \hat{\gamma}) > 1 - \tau \) appears to be sound, because it gives a control of percentage of observations from \( J(0) \) to be thrown out: \( \#J(c)/\#J(0) = (1 - q) \times 100\% \).

**Remark 1.** Our interest is to estimate equation (2.6) which depends on a set of individual specific effects \( (\alpha_i) \). Thus, there are two potential sources of inconsistency. The first one is related to the presence of fixed effects. Such an estimation may be affected by the incidental parameter problem [Neyman and Scott (1948) and Hsiao (2003)]. The second problem is related to the fact that the true propensity is unknown [Chernozhukov and Hong (2002), p.874]. In this paper, we derive the asymptotic properties of the estimator using large \( T \) asymptotics, which removes the first source of inconsistency.\(^2\) The second concern is addressed by choosing the trimming constant \( c \) as discussed above. Although our theoretical results are obtained under large \( T \) asymptotics, our simulations presented below indicate that the estimator performs well in cases with a small number of observations on each individual.

In general, the goal of step 1 is to select some, and not necessarily the largest subset of observations. For this task to be carried out, it suffices, but not necessary, that Assumption A3 (to be presented in the next section) holds. There are several models that could be estimated in step 1. The simplest model would be a linear probability model (LPM), with advantage that LPM does not suffer from incidental parameters problem. Hsiao (2003)

\[^2\]Hahn and Newey (2004) investigate the asymptotic properties of nonlinear models using MLE estimator for large panel data. They show that the incidental parameter problem disappears when \( T \) grows faster than \( N \).
describes alternative models as logit and maximum score estimators. In addition, as we described before, under large panel asymptotics, the MLE with smooth objective functions is available [Hahn and Newey (2004)]. The computation of the binary response model with fixed effects is feasible as demonstrated by Greene (2001). In a number of models of interest to practitioners, the estimation of the fixed effects model is quite feasible even in panels with large numbers of groups.\footnote{The important point is that the models are fully parametric, and all parameters of interest are estimable. The estimation is carried using Newton’s method of maximizing the log likelihood. In practice this is a very convenient manner to compute the desired probabilities.}

Step 2. Obtain the initial (but inefficient) estimator \( \hat{\theta}^0 = (\hat{\alpha}^{0\top}, \hat{\beta}^{0\top})^\top \) by solving a panel quantile regression problem defined as,

\[
\min_{\alpha, \beta} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau} \left( y_{it} - z_{it}^{\top} \alpha - x_{it}^{\top} \beta \right) 1(p(\hat{X}_{it}^\top \gamma) > 1 - \tau + c),
\]

(2.7)

where \( z_{it} \) denotes an indicator variable for the individual specific effect \( \alpha_i \).\footnote{Notice that \( z_{it} \) is an indicator variable for the individual effect \( \alpha_i \). Recall that equation (2.6) is estimated using the whole sample whereas equation (2.7) is estimated using observations from sample \( J_0 \) and \( J_1 \). Therefore, the dimension of the vector \( \alpha \) in steps 1 and 2 can be different.}

Solving the problem described in (2.7) is equivalent to solve the following optimization,

\[
\min_{\alpha, \beta} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \rho_{\tau} \left( y_{it} - z_{it}^{\top} \alpha - x_{it}^{\top} \beta \right) 1(p(\hat{X}_{it}^\top \gamma) > 1 - \tau + c).
\]

(2.8)

Next, we select \( J_1 = \left\{ (i, t) : \hat{\alpha}_i^0 + \hat{x}_{it}^\top \hat{\beta}^0 > \delta_{NT} + C_{it} \right\} \), where \( \delta_{NT} \) is a small positive number such that \( \delta_{NT} \downarrow 0 \) and \( \sqrt{NT} \times \delta_{NT} \) is bounded. This step asymptotically selects those observations that have a fixed effect \( \alpha_i \) and covariates values \( x_{it} \) such that \( \alpha_{i0} + x_{it}^\top \beta_0 > C_{it} \), building up the efficiency of the next step 3.

Step 3. Solve the quantile regression problem (2.7) with \( J_1 \) in place of \( J_0 \) if \( J_0 \subset J_1 \). Denote this estimator as \( \hat{\theta}^1 = (\hat{\alpha}^{1\top}, \hat{\beta}^{1\top})^\top \).

2.3 A nonparametric 2-step procedure

The 3-step parametric procedure described above assumes an envelope restriction on the censoring probability, which requires that the parametric probability model \( p(\hat{X}_{it}^\top \gamma) \) is not severely misspecified. However, in some applications, the misspecification of \( p(\cdot) \) could be severe invalidating the envelope condition. In this case, a recent paper by Tang, Wang,
He, and Zhu (2012) shows that this problem can be overcome by using a 2-step estimator in which the propensity score is nonparametrically estimated in the first step. If the parametric form of the true propensity score is unknown, then one can obtain a consistent estimator of \( \pi_0(\alpha_i, x_{it}, C_{it}) \) by applying nonparametric methods (e.g., generalized linear regression with spline approximation, generalized additive models, or maximum score with series function approximation).\(^5\) In what follows, we extend the results in Tang, Wang, He, and Zhu (2012) for the problem considered in this paper.

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

**Step 2.** Then \( \theta_0 = (\alpha_0^\top, \beta_0^\top)^\top \) can be estimated by applying fixed effects QR to the subset \( J_T \), i.e., \( \hat{\theta} \) is the minimizer of

\[
\min_{\alpha, \beta} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \rho_{\tau}(y_{it} - z_{it}^\top \alpha - x_{it}^\top \beta)1(\hat{\pi}(\alpha_i, x_{it}, C_{it}) > 1 - \tau + c_T).
\]

Therefore, both nonparametric 2-step and parametric 3-step estimators are based on informative subsets selected in their respective step 1, and they both depend on a trimming parameter. The first step in the parametric estimator represents a lower bound of the true propensity score, which implies that the QR estimator obtained in the second step will be consistent, but inefficient. For this reason, a third step is necessary to build efficiency. On the other hand, the first step in the nonparametric estimator consistently estimates the true propensity score, which implies that the QR estimator obtained in the second step is not only consistent but also efficient, yielding no need for a third step. This performance of the 2-step estimator is obtained at the expenses of additional assumptions that control smoothness and rate of convergence of \( \hat{\pi}(\alpha_i, x_{it}, C_{it}) \) employed in the first step.

**Remark 2.** A penalized quantile regression might be used to improve the performance of the parametric 3-step and nonparametric 2-step estimators in small samples. As demonstrated in Koenker (2004), a penalized estimator can improve the efficiency of the fixed effects quantile regression estimator. The potential gains depend on carefully choosing the tuning parameter \( \lambda \). Then, data-driven approaches to \( \lambda \) selection in censored models are of

\(^5\)See e.g. Li and Racine (2007) for binary dependent variable panel data models.
fundamental interest. In the Monte Carlo section, we describe the estimation procedure and show numerical evidence that suggests that the approach works in practice.

2.4 Large Sample Properties

This section investigates the large sample properties of the proposed parametric 3-step and nonparametric 2-step estimators. While these methods are similar to the ones proposed by Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012), which have been developed for cross-sectional models, the existence of the individual fixed effects parameter, $\alpha$, in equations (2.8) and (2.9), 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 when the time dimension is short 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, and to ensure the asymptotic normality, the requirements on the diverging rates of $N$ and $T$ are very stringent, say $N^2(\log N)^3/T \rightarrow 0$.

Motivated by the large time-series dimension requirement in the QR for panel data models, we propose to use an alternative asymptotics employing sequential limits to derive the asymptotic properties of the censored QR panel with fixed effects. The sequential asymptotics is defined as $T$ diverging to infinity first, and then $N$. We do not specify the exact relationship between $N$ and $T$, although we maintain that $T$ depends on $N$. For notational simplicity, we suppress this dependence. For a detailed discussion on sequential asymptotics for panel data, see Phillips and Moon (1999, 2000). In what follows, we adopt the following notation: $(T,N)_{seq} \rightarrow \infty$ means that first $T \rightarrow \infty$ and then $N \rightarrow \infty$.

We consider the following regularity conditions:

(A1) $\{(x_{it}, y_{it}', C_{it})\}$ are independent across subjects and independently and identically distributed (i.i.d.) for each $i$ and all $t \geq 1$. The support of the distribution of $(x_{it}, C_{it})$, $\mathcal{X}$, is compact.
(A2) Let \( u_{it} := y_{it}^* - \alpha_{i0} - x_{it}^\top \beta_0 \) and \( \alpha := (\alpha_1, \ldots, \alpha_N)^\top \). \( (\alpha, \beta) \in A^N \times B \), where \( A \) is a compact subset of \( \mathbb{R} \), \( A^N \) is the product of \( n \) copies of \( A \), and \( B \) is a compact subset of \( \mathbb{R}^p \). \( F_i(u|x) \) is defined as the conditional distribution function of \( u_{it} \) given \( x_{it} = x \). Assume that \( F_i(u|x) \) has conditional densities, \( 0 < f_i(u|x) < \infty \).

(A3) \( p(\cdot) \) is strictly increasing and second order differentiable. Let \( \gamma = (\alpha^T, \eta^T)^\top \) and \( \pi_0(\alpha_i, x_{it}, C_{it}) \) be the true propensity score function. There exist constants \( c > 0 \) and \( v > 0 \) such that \( p(\hat{X}_{it}^\top \gamma) > 1 - \tau + c \) implies \( \pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau + v \) a.s., for any \( \gamma \) in a compact neighborhood \( \Gamma \) of \( \gamma_0 \), which is the unique maximizer of

\[
\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[ \delta_{it} \log p(\hat{X}_{it}^\top \gamma) + (1 - \delta_{it}) \log(1 - p(\hat{X}_{it}^\top \gamma)) \right].
\]

Each element of the Hessian matrix has finite mean uniformly for \( \gamma \in \Gamma \), and \( \hat{X}_{it} \) is a known function of \( X_{it} \) and \( C_{it} \), with \( X_{it} = (z_{it}^\top, x_{it}^\top)^\top \).

(A4) There exists a \( \hat{\gamma} = (\hat{\alpha}^\top, \hat{\eta}^\top)^\top \) in an open neighborhood of \( \gamma_0 \) for which \( p(\hat{X}_{it}^\top \hat{\gamma} > v) \) is Lipschitz in \( \hat{\gamma} \) uniformly in \( v \), with \( \hat{X}_{it} \) including \( \hat{x}_{it}, z_{it} \) and \( x_{it} \).

(A5) The limiting forms of the following matrices are positive definite:

\[
V = \tau(1 - \tau) \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \mathbb{E} \{ (x_{it} - A_i a_i^{-1})(x_{it} - A_i a_i^{-1})^\top (\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau) \},
\]

\[
\Lambda = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N [a_i^{-1} A_i A_i^\top - B_i] (\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau),
\]

where, \( a_i := \mathbb{E}[f_i(0|X_{it})1(\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau)], A_i := \mathbb{E}[f_i(0|X_{it})x_{it}1(\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau)], B_i := \mathbb{E}[f_i(0|X_{it})x_{it}x_{it}^\top 1(\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau)]. \]

The conditions imposed above are usual in the QR literature, in particular Chernozhukov and Hong (2002) assume similar conditions. Assumption A1 is an extended version of condition a.1 in Chernozhukov and Hong (2002) and condition A1 in Koenker (2004). Assumption A2 imposes restrictions on the conditional distributions and densities and is familiar in

\[\text{We assume that the censoring points are observable. In a random sample with fixed (or observable random) censoring, observations on the triple } (y_{it}, x_{it}, C_{it}) \text{ are assumed to be available. Nevertheless, with unobserved random censoring, the observations are of the form } (y_{it}, x_{it}, d_{it}), \text{ where } d_{it} \text{ is a binary variable indicating whether the dependent variable is uncensored, } d_{it} = 1[y_{it} < C_{it}], \text{ and the censoring variables } C_{it} \text{ are not known for all } (i, t). \]
the QR literature. It has been considered in Chernozhukov and Hong (2002) and Koenker (2004), among others. Assumption A3 is an identification condition. The first part of A3 is similar to assumption (c) in Chernozhukov and Hong (2002). Under model 2.1 and this condition, as shown in the appendix, the model is identified. In addition, condition A3 is needed to show that the first step of the 3-step estimator does not affect its asymptotic distribution. Assumption A4 indicates that the distribution of \( \alpha_i + \mathbf{x}_i^T \eta \) changes smoothly with respect to changes in \((\alpha_i^T, \eta^T)\). Finally, Assumption A5 is standard in the QR literature [see, e.g., Koenker (2005)] and is concerned with the asymptotic covariance matrix of the 3-step estimator.

The following result derives the asymptotic properties of the parametric 3-step estimator.

**Theorem 1.** Under regularity conditions A1–A4, provided that the sequence parameter \( \delta_{NT} \downarrow 0 \) and \( \delta_{NT} \times \sqrt{NT} = O(1) \), as \((T, N) \) seq \( \rightarrow \infty \), the 3-step quantile regression slope estimator for model (2.1), \( \hat{\beta}^1 \), is consistent. Moreover, under the assumptions A1–A5, we have

\[
\sqrt{NT}(\hat{\beta}^1 - \beta_0) \overset{d}{\rightarrow} N(0, \Lambda^{-1}V\Lambda^{-1}).
\]

**Proof.** See Appendix A.1. \(\square\)

Theorem 1 states the asymptotic properties for the estimator described in step 3 of Section 2.2 above, \( \hat{\beta}^1 \). However, as a byproduct of the theorem we derive the asymptotic properties of \( \hat{\beta}^0 \) as well. The details and the proof are provided in Appendix A. Furthermore, using Theorem 1, the joint distribution of several censored QR slope estimators for a model with fixed effects can be obtained.

**Corollary 1.** Under the regularity conditions of Theorem 1, \( \hat{\beta}^1(\tau_j) \) for \( j \leq J \) is asymptotically normal with covariance matrix \((\tau_l \wedge \tau_j - \tau_l \tau_j) \Lambda(\tau_l)^{-1} V(\tau_l, \tau_j) \Lambda(\tau_j)^{-1} \), where,

\[
V(\tau_l, \tau_j) = E\{(x_{il} - A_i(\tau_l) a_i(\tau_l)^{-1})(x_{il} - A_i(\tau_j) a_i(\tau_j)^{-1})^T 1(\tau(\alpha_i, x_{il}, C_{il}) > (\tau_l \wedge \tau_j - \tau_l \tau_j))\},
\]

\[
\Lambda(\tau_l) = E\{[a_i(\tau_l)^{-1} A_i(\tau_l)] A_i(\tau_l)^T - B_i(\tau_l)] 1(\tau(\alpha_i, x_{il}, C_{il}) > 1 - \tau_l)\},
\]

and 
\[ a_i(\tau_k) := E[f_i(0|x_{il}) 1(\pi(\alpha_i, x_{il}, C_{il}) > 1 - \tau_k)] , \quad A_i(\tau_k) := E[f_i(0|x_{il}) x_{il} 1(\pi(\alpha_i, x_{il}, C_{il}) > 1 - \tau_k)] , \quad B_i(\tau_k) := E[f_i(0|x_{il}) x_{il} x_{il}^T 1(\pi(\alpha_i, x_{il}, C_{il}) > 1 - \tau_k) , \] where \( k \in (l, j) \).

**Remark 3.** The components of the asymptotic covariance matrices in Theorem 1 and Corollary 1 that need to be estimated include \( a_i, A_i \) and \( B_i \). Following Powell (1986), the matrices
can be estimated by their sample counterpart. For instance, in the case of \( \hat{\beta}^1 \), \( a_i \) can be estimated as
\[
\hat{a}_i = \frac{1}{2Tg_N} \sum_{t=1}^{T} 1(|\hat{u}(\tau)| \leq g_N)1(X_{it}^\top \hat{\theta}^1(\tau) > C_{it} + \delta_{NT}),
\]
where \( \hat{u}(\tau) \) has the \( \tau \)-th conditional quantile at zero, \( \delta_{NT} \) is a known constant, and \( g_N \) is an appropriately chosen bandwidth, with \( g_N \to 0 \) and \( NTg_N^2 \to \infty \). Note also that \( A_i \) and \( B_i \) can be estimated similarly. The consistency of these asymptotic covariance matrix estimators is standard and will not be discussed further in this paper.

**Remark 4.** Under the regularity conditions of Theorem 1, provided that \( C_{it} \to -\infty \), the estimator \( \hat{\beta}^1 \) is asymptotically normal with covariance matrix \( \tau(1 - \tau)\Lambda^{-1}V\Lambda^{-1} \), where \( V := E\{(x_{it} - A_i a_i^{-1})(x_{it} - A_i a_i^{-1})^\top\} \) and \( \Lambda := E\{a_i^{-1} A_i A_i^\top - B_i\} \), with \( a_i := E[f_i(0|X_{it})] \), \( A_i := E[f_i(0|X_{it}) x_{it}] \), and \( B_i := E[f_i(0|X_{it}) x_{it} x_{it}^\top] \). This result is similar to the one derived in Kato, Galvao, and Montes-Rojas (2012), which is obtained by imposing conditions on \( N \) and \( T \) jointly going to infinity. In other words, the results obtained by Kato, Galvao, and Montes-Rojas (2012) can also be obtained via sequential asymptotics.

Now we turn to the regularity conditions of the nonparametric 2-step estimator. Let \( \|\pi - \pi_0\|_\infty = \sup_u |\pi(u) - \pi_0(u)| \) for a given \( \pi(\cdot) \) and a generic vector \( u \).

**B1** \( \{(x_{it}, y_{it}^*, C_{it})\} \) are independent across subjects and independently and identically distributed (i.i.d.) for each \( i \) and all \( t \geq 1 \). There exist constants \( M_1 \) and \( M_2 \) such that \( \max \|x_{it}\| < M_1 \) and \( \max |C_{it}| < M_2 \). The density function \( f_x(\cdot) \) is bounded away from zero and infinity uniformly over the support. In addition, \( E(x_{it} x_{it}^\top) \) is positive definite for each \( i \).

**B2** The conditional distribution of \( u_{it} := y_{it} - \alpha_{it} - x_{it}^\top \beta_0 \) given \( x_{it} = x \), \( F_i(u|x) \), has the continuous derivative with respect to \( u \) such that \( 0 < f_i(u|x) \leq \bar{f}_0 < \infty \).

**B3** For any positive integer \( i \) and any positive \( \epsilon_T \to 0 \) with \( T \) large enough, \( \lambda_{\min,\epsilon_T,i} \), the smallest eigenvalue of the matrix
\[
E X_{it} X_{it}^\top f_i(X_{it}^\top \theta_0 | X_{it}) 1(X_{it}^\top \theta_0 > C_{it} + \epsilon_T),
\]
with \( \theta_0 = (\alpha_0^\top, \beta_0^\top)^\top \) and \( X_{it} = (z_{it}^\top, x_{it}^\top)^\top \), satisfies \( \lambda_{\min,\epsilon_T,i} > \lambda > 0 \). There is a constant \( \zeta > 0 \) such that for any \( i \) and positive \( \epsilon_T > 0 \), \( \sup_{\|\theta - \theta_0\|_2 \leq \zeta} E \{1(|X_{it}^\top \theta| < C_{it} + \epsilon_T)\} = O(\epsilon_T) \).

13
(B4) \( c_T \to 0 \) and \( c_T T^{1/4} > c^* \) for some positive constant.

(B5.1) For any \( \epsilon_T \to 0 \) with \( \epsilon_T/\epsilon_T \to 1 \), \( \pi_0(\alpha, x_{it}, C_{it}) > 1 - \tau + \epsilon_T \) implies \( X_{it}^\top \theta_0 > C_{it} + \epsilon_T \) for some \( \epsilon_T^* \) with \( \epsilon_T = O(\epsilon_T^*) \).

(B5.2) Let \( \alpha_0 = (\alpha_1, \ldots, \alpha_N)\top, \theta_0 = (\alpha_0, \beta_0) \in \mathcal{A}^N \times \mathcal{B} \), where \( \mathcal{A} \) is a compact subset of \( \mathbb{R} \), \( \mathcal{A}^N \) is the product of \( n \) copies of \( \mathcal{A} \), and \( \mathcal{B} \) is a compact subset of \( \mathbb{R}^p \). Let \( \int_0^\infty \sqrt{\log N(\epsilon^2, \mathcal{H}, || \cdot ||_{\mathcal{H}})} d\epsilon < \infty \), where \( \mathcal{H} \) is the parameter space to which \( \pi_0 \) and \( \hat{\pi} \) belongs.

(B5.3) For any \( \epsilon_T \to 0 \), \( \sup_{||\pi-\pi_0||\leq \epsilon_T} E[1\{|\pi(\alpha, x_{it}, C_{it}) - (1 - \tau + c_T)| < \epsilon_T\}] = O(\epsilon_T). \)

(B6) Let \( \theta_i = (\alpha_i, \beta_i)\top \). For any positive \( \epsilon_T \to 0 \) with \( ||\theta_i - \theta_{i0}|| \leq \epsilon_T \) and all \( i \),
\[
\left( \begin{array}{cccc}
D_{11}, 0, \cdots, 0, D_{12} \\
\vdots \\
0, \cdots, 0, D_{N1}, D_{N2} \\
\end{array} \right) = K_2 \left( \begin{array}{c}
\alpha_1 - \alpha_{i0} \\
\cdots \\
\alpha_N - \alpha_{N0}
\end{array} \right)
\]
with \( D_{i1} = E[1\{|\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau + c_T\}] \), and \( D_{i2} = E[1\{|\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau + c_T\}] \), and \( D = E[1\{|\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau + c_T\}] \), where \( K_2 \) is semi-positive definite matrix satisfying \( 0 \leq \lambda_{min}(K_2) \leq \lambda_{max}(K_2) < \infty \) for large \( T \).

(B7) The limiting forms of the following matrices are positive definite:
\[
V = \tau(1 - \tau) \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N E\left\{ (x_{it} - A_ia_i^{-1}) (x_{it} - A_ia_i^{-1})\top \right\} 1\{\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau\},
\]
\[
\Lambda = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \left[ a_i^{-1} A_i A_i\top - B_i \right] 1\{\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau\},
\]
where \( a_i := E[f_i(0)|X_{it}]1\{\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau\} \), \( A_i := E[f_i(0)|X_{it}]x_{it}1\{\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau\} \), and \( B_i := E[f_i(0)|X_{it}]x_{it}x_{it}\top 1\{\pi_0(\alpha_i, x_{it}, C_{it}) > 1 - \tau\} \).

The conditions imposed above are standard in the QR literature, in particular they are similar to those in Tang, Wang, He, and Zhu (2012). Assumption B1 is an extended version of condition A1 in Tang, Wang, He, and Zhu (2012), a.1 in Chernozhukov and Hong (2002), and condition A1 in Koenker (2004). Assumption B2 is common in the QR literature and similar to the previous A2. This condition has been considered in Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012), among others. Assumption B3 is similar to A3 in Tang, Wang, He, and Zhu (2012). Assumptions B4 and B5.1-B5.3 are parallel to A4
to A5.1-A5.3 in Tang, Wang, He, and Zhu (2012). They impose restrictions on the functions \( \pi_0 \) and \( \hat{\pi} \) which are used in the nonparametric statistics literature. This condition is similar to (3.3) in Chen, Linton, and van Keilegom (2003) and allows for many nonparametric estimators of \( \pi_0 \), although it restricts smoothness of the functions. Finally, B6 and B7 regard the asymptotic variance-covariance matrix of the estimator.

The result for the nonparametric 2-step estimator can be stated as follows:

**Theorem 2.** Under Assumptions B1–B5 and \( \| \hat{\pi} - \pi_0 \|_\infty = o_p(1) \), as \((N,T)_{\text{seq}} \to \infty \), the 2-step estimator for model (2.1), \( \hat{\beta} \), is consistent. In addition, under Assumptions B1–B7, and \( \| \hat{\pi} - \pi_0 \|_\infty = o_p(T^{-1/4}) \), we have,

\[
\sqrt{NT}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Lambda^{-1}V\Lambda^{-1}).
\]

**Proof.** See Appendix A.2

The result in Theorem 2 shows that both parametric 3-step and nonparametric 2-step estimators are asymptotically equivalent, with same asymptotic variance-covariance matrix. This result is in line with those in Tang, Wang, He, and Zhu (2012) and Ackerberg, Chen, and Hahn (2012).

In the theorems above, we provide results based on sequential asymptotics. As discussed in Phillips and Moon (1999, 2000), this view of double indexes asymptotics simplifies the proofs substantially, and provides valuable insights of the results, although sometimes sequential limits can give misleading asymptotic results. In contrast, the view of simultaneous asymptotics is more general, but it is substantially more difficult to derive and apply under stronger conditions. In our case, the cost is related to the requirements on the diverging rates of \( T \) and \( N \). In the results listed in Kato, Galvao, and Montes-Rojas (2012), one can notice that for the standard QR with fixed effects and simultaneous asymptotics, roughly speaking, the best can be done is to require \( T \) tending to infinity faster than \( N^2(\log N)^3 \). Thus, given the stringent requirement on the growth rate of \( T \) under the joint limits, we believe the use of sequential asymptotics is an important tool and provides useful approximations for censored QR panel fixed effects analysis, and dramatically decreases the complexity of the proofs.

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 $T$.

3 Monte Carlo

In this section, we use Monte Carlo simulations to assess the finite sample performance of the two proposed estimators. We report results for empirical bias, root mean squared error (RMSE), and coverage probability for confidence interval with nominal level 0.95. A general version of the model is considered in the simulation experiments, where we define the latent variable as,

$$y_{it}^* = \alpha_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \left[1 + (x_{1,it} + x_{2,it} + x_{1,it}^2 + x_{2,it}^2) \cdot \zeta \right] \cdot u_{it},$$

where $\alpha_i$ captures the individual-specific intercepts, $\beta = (\beta_1, \beta_2)^\top$ is the parameter vector associated with the variable $x_{it} = (x_{1,it}, x_{2,it})^\top$, $u_{it} \sim iid$ is the innovation term, and $\zeta$ modulates the amount of heteroscedasticity. We performed simulations with $\zeta \in \{0, 0.5\}$, but we only report the case with heteroscedasticity to save space. Notice that a fixed effect model with homoscedasticity is obtained when $\zeta = 0$. We draw $x_{it} \in \mathcal{X} \subset \mathbb{R}^2$ from independent standard normal distributions, truncated as $\{x_{it} : \|x_{it}\|_\infty < 2\}$. In addition, we generate the innovation, $u$, from two different distributions: a standard normal distribution ($\mathcal{N}$) and a $t$-student distribution with 3 degrees of freedom ($t_3$). The fixed effect, $\alpha_i$, is correlated with the independent variables and generated considering the following equation:

$$\alpha_i = v_i + \varphi \sum_{t=1}^T (x_{1,it} + x_{2,it}),$$

with the distribution of the variable $v_i$ assumed to be standard Normal. In what follows, we set $\varphi \in \{0, 1/2\}$. When $\varphi \neq 0$, this method of generating $\alpha_i$ ensures that the classical random effects estimators are biased because the individual effects and the explanatory variables are correlated.

The censored variable is defined as $y_{it} = \max(y_{it}^*, C_{it})$. For simplicity, we use fixed censoring in the simulations. We explore the effects of different proportions of censoring, where

---

7 The results for the case $\zeta = 0$ are available upon request.

8 Robustness checks show that the choice of the Gaussian random variable does not influence our results.
the known censoring point, \(C_{it}\), is either –0.95 or –1.45 in the variants of the simulations. These choices yield roughly 50% and 45% of censoring, respectively. Since we are considering left-censored observations, we estimated the model for \(\tau \in \{0.25, 0.5\}\), because the median and left-tail quartiles are potentially more affected by the presence of left-censored observations. Finally, we analyze several different sample size configurations for \(T = \{15, 50\}\) and \(N = \{50, 100\}\), setting the number of replications to 1000 and \((\beta_1, \beta_2) = (10, -2)\).\(^9\)

In the experiments, we consider five estimators. The first one is the Omniscient estimator, which is obtained by fitting the quantile regression model with the latent response variable and thus serves as a golden standard. The second one is the parametric 3-step estimator, labeled 3-step, in which \((x_{1,it}, x_{2,it})\) and \((x_{1,2,it}, x_{2,2,it})\) are used in the parametric (logit) estimation of the propensity score in the first step. Following Chernozhukov and Hong (2002), the cutoff value \(c\) is equal to the 0.1-th quantile of all \(p(\hat{\mathbf{X}}_{it}^\top \hat{\gamma})\)'s such that \(p(\mathbf{X}_{it}^\top \hat{\gamma}) > 1 - \tau\). In the second step, the parameter \(\delta_{NT}\) was selected as the \(1/3(NT)^{-1/3}\)-th quantile of the estimated quantile function \(\hat{\alpha}_i^0(\tau) + \mathbf{x}^\top_{it} \hat{\beta}^0(\tau)\). The third estimator is the proposed nonparametric 2-step estimator (labeled 2-step) which, as discussed previously, is asymptotically equivalent to the 3-step. For this estimator, the first step is carried out by using generalized additive methods for a logistic regression using \(c = (NT)^{-1/5}\), as in Tang, Wang, He, and Zhu (2012). A version of the Powell estimator and a “naive” estimator were considered. The Powell estimator does not control for fixed effects, and the naive estimator assumes that the observations are uncensored.

Table 3.1 displays the mean bias, RMSE and empirical coverage of the confidence interval of the estimators employed to estimate the location-scale model \((\zeta = 1/2)\) at the median quantile, \(\tau = 0.5\). The confidence intervals were constructed by applying standard quantile regression routines obtained in the R package \texttt{quantreg} [Koenker (2012)]. In both cases of the parametric 3-step and the nonparametric 2-step, the kernel density estimation approach proposed by Powell (1991) is applied to a subset of observations. The upper block of the table collects results for Normal innovations and the lower block for \(t_3\).

We can see from the upper part of Table 3.1 that the Omniscient estimator performs better than any other estimator. This is expected because it uses full information from the latent variable \(y_{it}^*\). In practice, however, the researcher only observes \(y_{it} = \max(C_{it}, y_{it}^*)\). The 3-step estimator is slightly biased for small \(T\), but the bias decreases substantially when

\(^9\)We performed simulations with \(N = \{500, 1314\}\). The results were similar to those with \(N = 100\). We omit these results to save space.
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Censor</th>
<th>Error Term</th>
<th>Omniscient</th>
<th>3-step</th>
<th>2-Step</th>
<th>Powell</th>
<th>Naive</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>T</td>
<td>point</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>-0.95</td>
<td>N</td>
<td>0.002</td>
<td>0.045</td>
<td>-0.046</td>
<td>0.543</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.143)</td>
<td>(0.297)</td>
<td>(0.355)</td>
<td>(1.230)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.952]</td>
<td>[0.944]</td>
<td>[0.883]</td>
<td>[0.771]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.95</td>
<td>N</td>
<td>-0.001</td>
<td>0.010</td>
<td>-0.020</td>
<td>0.533</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.076)</td>
<td>(0.136)</td>
<td>(0.155)</td>
<td>(1.096)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.951]</td>
<td>[0.942]</td>
<td>[0.918]</td>
<td>[0.686]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>15</td>
<td>-0.95</td>
<td>N</td>
<td>0.000</td>
<td>0.034</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.100)</td>
<td>(0.198)</td>
<td>(0.297)</td>
<td>(1.006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
<td>[0.938]</td>
<td>[0.819]</td>
<td>[0.612]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>-0.95</td>
<td>N</td>
<td>-0.001</td>
<td>0.009</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.053)</td>
<td>(0.099)</td>
<td>(0.126)</td>
<td>(0.891)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.953]</td>
<td>[0.945]</td>
<td>[0.889]</td>
<td>[0.517]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>15</td>
<td>-1.45</td>
<td>N</td>
<td>0.002</td>
<td>0.038</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.143)</td>
<td>(0.282)</td>
<td>(0.338)</td>
<td>(1.154)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.952]</td>
<td>[0.950]</td>
<td>[0.887]</td>
<td>[0.686]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>-1.45</td>
<td>N</td>
<td>-0.001</td>
<td>0.010</td>
<td>-0.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.076)</td>
<td>(0.133)</td>
<td>(0.149)</td>
<td>(1.041)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.951]</td>
<td>[0.944]</td>
<td>[0.915]</td>
<td>[0.686]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>15</td>
<td>-1.45</td>
<td>N</td>
<td>0.000</td>
<td>0.030</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.100)</td>
<td>(0.191)</td>
<td>(0.281)</td>
<td>(0.965)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
<td>[0.942]</td>
<td>[0.820]</td>
<td>[0.604]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>-1.45</td>
<td>N</td>
<td>-0.001</td>
<td>0.007</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.053)</td>
<td>(0.096)</td>
<td>(0.121)</td>
<td>(0.851)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.953]</td>
<td>[0.942]</td>
<td>[0.883]</td>
<td>[0.513]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>15</td>
<td>-0.95</td>
<td>t_3</td>
<td>0.000</td>
<td>0.048</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.155)</td>
<td>(0.325)</td>
<td>(0.396)</td>
<td>(1.280)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.959]</td>
<td>[0.940]</td>
<td>[0.903]</td>
<td>[0.794]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>-0.95</td>
<td>t_3</td>
<td>-0.001</td>
<td>0.009</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.083)</td>
<td>(0.151)</td>
<td>(0.174)</td>
<td>(1.119)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
<td>[0.948]</td>
<td>[0.918]</td>
<td>[0.690]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>15</td>
<td>-0.95</td>
<td>t_3</td>
<td>0.002</td>
<td>0.043</td>
<td>-0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.108)</td>
<td>(0.225)</td>
<td>(0.323)</td>
<td>(0.998)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.946]</td>
<td>[0.940]</td>
<td>[0.834]</td>
<td>[0.656]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>-0.95</td>
<td>t_3</td>
<td>0.000</td>
<td>0.008</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.060)</td>
<td>(0.107)</td>
<td>(0.132)</td>
<td>(0.949)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.952]</td>
<td>[0.950]</td>
<td>[0.898]</td>
<td>[0.551]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>15</td>
<td>-1.45</td>
<td>t_3</td>
<td>0.000</td>
<td>0.043</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.155)</td>
<td>(0.307)</td>
<td>(0.374)</td>
<td>(1.224)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.959]</td>
<td>[0.944]</td>
<td>[0.904]</td>
<td>[0.783]</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>50</td>
<td>-1.45</td>
<td>t_3</td>
<td>-0.001</td>
<td>0.007</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.083)</td>
<td>(0.144)</td>
<td>(0.167)</td>
<td>(1.071)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
<td>[0.944]</td>
<td>[0.917]</td>
<td>[0.676]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>15</td>
<td>-1.45</td>
<td>t_3</td>
<td>0.002</td>
<td>0.042</td>
<td>-0.052</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.108)</td>
<td>(0.214)</td>
<td>(0.309)</td>
<td>(0.950)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.946]</td>
<td>[0.937]</td>
<td>[0.837]</td>
<td>[0.658]</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>50</td>
<td>-1.45</td>
<td>t_3</td>
<td>0.000</td>
<td>0.008</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.060)</td>
<td>(0.103)</td>
<td>(0.127)</td>
<td>(0.904)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.952]</td>
<td>[0.950]</td>
<td>[0.891]</td>
<td>[0.534]</td>
</tr>
</tbody>
</table>

Table 3.1: Monte Carlo results for the location-scale model at $\tau = 0.5$, and $\varphi = 1/2$. The table shows the bias, RMSE (in parentheses), and coverage (in brackets).
As $T$ increases. In addition, one can see that, for a fixed sample, the bias of the estimator decreases as the censoring point $C$ goes from $-0.95$ to $-1.45$. In terms of empirical coverage, the 3-step estimator performs very well and produces empirical coverage close to the nominal 0.95.

The results for the nonparametric 2-step also show small biases, which tend to disappear as $T$ increases. The nonparametric 2-step seems to require more time-series observations to start performing as well as the parametric 3-step. However, as the number of time series observations increases, the 2-step starts performing close to the parametric 3-step, confirming the theoretical results obtained in the last section. For example, for the Normal distribution case with $N = T = 50$ and $C = -1.45$, both 3-step and 2-step perform similarly. Finally, the empirical confidence interval coverage for the 2-step is very good and becomes closer to the nominal 95% as sample size increases, with the 2-step estimator requiring again a large $T$ relative to the 3-step.

The bottom block on Table 3.1 describes the results for the location-scale case ($\zeta = 1/2$) and $t_3$ distribution. The results are similar to those in the upper block of Table 3.1, and show that among all the estimators that are obtained by using the censored observations $y_{it}$’s, the one with the worst performance is the Naive. It is severely biased with large RMSE and empirical coverage of zero. Remember that this estimator naively assumes that $y_{it}^* = y_{it}$, that is, it is obtained by using all observations as if none were censored. The second worst estimator is the modified Powell estimator. In the model proposed by Wang and Fygenson (2009), $\alpha_i$ is an i.i.d. random effect, thus the Powell estimator performed adequately. However, when $\alpha_i$ is allowed to be correlated with $x_{it}$ and one does not account for individual effects in the estimation procedure, the Powell estimator breaks down, presenting the second largest bias and RMSE. Moreover, one could notice that the bias does not decrease when the time dimension increases because of the omission of the fixed effects.

In order to shed more light on the performance of the nonparametric 2-step vis-à-vis the parametric 3-step, Figure 3.1 offers the RMSE of the estimators from short-$N$ simulations when varying the time series $T$ (left panel) and $\varphi$ (right panel). In these simulations, we only considered Gaussian innovations with $C = -0.95$. The left panel shows that although the 3-step outperforms the 2-step for small $T$, the equivalence between them is achieved as $T$ increases, and they both approach the Omniscient estimator for large $T$. The right panel shows the RMSE of the proposed estimators when the correlation between $\alpha_i$ and
$x_{it}$ is changed. When $\varphi = 0$, the RMSE of the new estimators are close to Powell’s, which suggest that the new methods can be used at almost no cost in terms of higher RMSE. This is expected because the version of Powell estimator implemented in Wang and Fygenson (2009) seeks to estimate a marginal model and not a quantile model which conditions on individual effects. The results show that under no model misspecification, 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 $\varphi$ 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 $\varphi$.

Table 3.2 presents results for the location-scale model ($\zeta = 1/2$) estimated at $\tau = 0.25$. While the upper block of the table shows results when $u$ is distributed as Gaussian, the lower block presents results when $u$ is distributed as $t_3$. The results are somewhat analogous to the ones presented in Table 3.1. When compared with the case for $\tau = 0.50$, we find

Figure 3.1: Small sample performance of the 3-step and 2-step estimators when the number of time series observations increases. The figure also shows their performance under different degrees of correlation between $\alpha$ and $x$. The case of $\varphi = 0$ represents the random effects case.
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Censor</th>
<th>Error</th>
<th>Estimators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Term</td>
<td>Omniscient</td>
</tr>
<tr>
<td>N</td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>-0.95</td>
<td>[0.019]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.156]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.954]</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>-0.95</td>
<td>[0.005]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.085]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>-0.95</td>
<td>[0.021]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.115]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.948]</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>-0.95</td>
<td>[0.003]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.059]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.953]</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>-1.45</td>
<td>[0.019]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.156]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.954]</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>-1.45</td>
<td>[0.005]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.085]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>-1.45</td>
<td>[0.021]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.115]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.948]</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>-1.45</td>
<td>[0.003]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.059]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.953]</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>-0.95</td>
<td>[0.035]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.197]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>-0.95</td>
<td>[0.009]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.105]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.944]</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>-0.95</td>
<td>[0.035]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.141]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.953]</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>-0.95</td>
<td>[0.013]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.074]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.942]</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
<td>-1.45</td>
<td>[0.035]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.197]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.950]</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>-1.45</td>
<td>[0.009]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.105]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.944]</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>-1.45</td>
<td>[0.035]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.141]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.953]</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>-1.45</td>
<td>[0.013]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.074]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[0.942]</td>
</tr>
</tbody>
</table>

Table 3.2: Monte Carlo results for the location-scale model at $\tau = 0.25$, and $\varphi = 1/2$. The table shows the bias, RMSE (in parentheses), and coverage (in brackets).
Table 3.3: Monte Carlo results for the parametric 3-step estimator with different parametric models estimated in a first stage. Results are reported at $\tau = 0.50$ and $\varphi = 1/2$. The table shows the bias and RMSE (in parentheses).
that, in general, the bias of the estimators are slightly larger, but as in the previous case, the biases decrease substantially when $T$ increases. The results also show excellent coverage probability for the 3-step. Finally, as in the previous cases, Naive and Powell are severely biased, and their performance do not improve as the sample size increases. The lower block of Table 3.2 shows the performance of the estimators in the case of $u$ distributed as $t_3$. The results are qualitatively similar to those in Table 3.1. When compared with the results for the Normal case in Table 3.2, one can see that the bias and RMSE for the 3-step and 2-step estimators are larger for the heavy tail case. But once again, in general, they decrease with the number of time series observations.
To investigate how sensitive is the 3-step estimator 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. Results for mean bias and RMSE are displayed in Table 3.3. These simulations use Normal innovations and $\tau = 0.5$. We extend the designs by increasing the number of cross-sectional units. The table presents results for $N = \{500, 1000\}$ which are similar to the number of subjects considered in the empirical section. In addition, we report results for the location shift model ($\zeta = 0$) in the upper part of the table and location-scale shift model ($\zeta = 1/2$) in the lower part of the table. For convenience, we report the results for the second and third steps of the 3-step estimator as well as results using the logit link function. One can see that there are small biases for $T = 15$, which quickly decrease as $T$ increases to 50. The results of the table suggest that the LPM performs well in large panel data.

High dimensional models similar in spirit to the ones considered in Table 3.3 invite an opportunity for employing the penalized version of estimator as remarked in Section 2. The parametric 3-step and the nonparametric 2-step estimators can be easily accommodated to the penalized estimation of models with a large $N$ number of individual specific location shifts. Koenker (2004) suggests to jointly estimate slope parameters and individual location shifts when the number of observations on each subject is small. This alternative approach can be easily implemented by solving a penalized quantile regression problem, which can be seen as an alternative last step of the parametric 3-step or nonparametric 2-step estimators. For instance, one might replace (2.9) by:

$$\min_{\alpha, \beta} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{g=1}^{G} \left( \omega_g \rho_{\tau_g}(y_{it} - z_{it}^\top \alpha - x_{it}^\top \beta(\tau_g)) + \lambda \|\alpha\|_1 \right) 1 \left( \hat{\tau}(\alpha_i, x_{it}, C_{it}) > 1 - \tau_g + c_T \right),$$

where $\omega_g$ is a relative weight given to the $g$-th quantile, $G$ is the number of quantiles $\{\tau_1, \tau_2, ..., \tau_G\}$ to be estimated, $\lambda \|\alpha\|_1$ is a penalty term and $\lambda$ is a tuning parameter. Koenker (2004) points out that the choice of the weights, $\omega_g$, and the associated quantiles $\tau_g$, is somewhat analogous to the choice of discretely weighted $L$-statistics, as for example in Mosteller (1946). An alternative less efficient, yet practical choice, is to ignore the potential gains and estimate models with equal weights defined as $\omega_g = G^{-1}$ for all $g$. One can similarly define

\footnote{In our model, the individual effects include the intercept term and it depends on the quantile. Thus, the individual effects depend on the quantile too. Koenker (2004) used a different approach, where the individual specific intercepts are restricted to be the same across the quantiles. This procedure can be implemented using weighted quantile regression, as proposed initially by Koenker (1984). It is important to note that both models are identical for our purposes of estimating a single fixed quantile.}
an (optional) penalized third step in Section 2.3.

Small sample evidence indicates that there are gains of shrinking the fixed effects in models where \( N \) ranges from 50 to 2500 and \( T = 5 \). Figure 3.2 briefly illustrates the advantage of the approach when \( N = 50 \), considering the quantiles and associated weights and intraclass correlation similar to those in Koenker (2004). By carefully choosing the tuning parameter and the weights in each quantile, it is possible to reduce the RMSE of the parametric 3-step and nonparametric 2-step estimators. For instance, as illustrated in Figure 3.2, the performance of the fixed effects estimator improves by selecting \( \lambda \) to minimize RMSE [Lamarche (2010)].

4 An Empirical Application

This section illustrates the use of the quantile regression approach that simultaneously handles censored data and unobserved heterogeneity. Using data from Chay and Powell (2001), we investigate 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. The findings suggest that the black-white gap was reduced after the policy was introduced, having a significant effect among young workers but a negligible effect among mature workers. We also find that while the 1964 Civil Rights Act has a moderate effect in the lower tail, it has a large effect at the upper tail of the conditional distribution of earnings among young workers. This might be interpreted as suggesting that the Civil Rights Act benefited high-income young black workers more than other black workers.

\[^{11}\]Generalized Cross Validation (GCV) is not well suited for quantile regression because it is based on residuals [He, Ng, and Portnoy (1998)].
<table>
<thead>
<tr>
<th>Variable</th>
<th>All Workers</th>
<th>Young Workers</th>
<th>Mature Workers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  Std Dev</td>
<td>Mean  Std Dev</td>
<td>Mean  Std Dev</td>
</tr>
<tr>
<td>Censored observations</td>
<td>0.552 0.497</td>
<td>0.519 0.497</td>
<td>0.575 0.494</td>
</tr>
<tr>
<td>Black</td>
<td>0.087 0.282</td>
<td>0.087 0.282</td>
<td>0.087 0.281</td>
</tr>
<tr>
<td>Age</td>
<td>38.546 7.468</td>
<td>32.431 4.878</td>
<td>42.838 5.793</td>
</tr>
<tr>
<td>Number of years</td>
<td>15 -</td>
<td>15 -</td>
<td>15 -</td>
</tr>
<tr>
<td>Number of workers</td>
<td>1314 -</td>
<td>542 -</td>
<td>772 -</td>
</tr>
<tr>
<td>Number of observations</td>
<td>24660 -</td>
<td>10170 -</td>
<td>14490 -</td>
</tr>
</tbody>
</table>

Table 4.1: Descriptive statistics for the earnings data. The groups analyzed are young workers (ages 22-30 in 1957) and mature workers (31-43 in 1957).

### 4.1 Data description

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. As a result, we could collect data on race, gender, education, age and earnings of 1314 workers over 15 years (Table 4.1). 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.

Table 4.1 reports the summary statistics for the variables. It shows that the earnings data is censored. A significant number of observations are top-censored at the maximum taxable earnings level for Social Security. Moreover, the tax ceiling changed over time, increasing from $15,000 in the period before the introduction of the Civil Rights Act to a little below $20,000 in the period after. The table shows that there is more than 50% censoring, and therefore, the upper quantiles of earnings are constant at the censoring point. This suggests that the estimation of the upper conditional quantiles by the standard quantile regression method is expected to be biased towards zero.

### 4.2 Model

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

\[
Q_{yi}(\tau|x_{it}, \alpha_i, C_{it}) = \min(C_{it}, \alpha_i(\tau) + x_{it}^\top \beta(\tau)),
\]

(4.1)
where $Q_y$ is the conditional quantile of the natural logarithm of earnings and $\mathbf{x}$ is a vector of effects of interest that includes an intercept, age, a quadratic term on age, schooling, and 14 indicator variables to control for time effects. In addition to these covariates, we include the variables of interest: race, an indicator for the period after the Civil Rights Act, and an interaction term for race in the period after the Civil Rights Act. The term $\alpha$ is used to control for individual unobserved heterogeneity. Notice that this empirical exercise is conducted with right-censored observations whereas the theory and Monte Carlo simulations were based on left-censored data. This implies that the samples $J_0$ and $J_1$ are built using a reversed inequality sign. We also adjust the computation of the asymptotic distribution of 3-step estimator to deal with right-censored observations. For comparison, we estimate models without individual specific intercepts,

$$Q_{y\mid \mathbf{x}_{it}, C_{it}}(\tau) = \min(C_{it}, \mathbf{x}_{it}^\top \beta(\tau)), \quad (4.2)$$

and without modeling the censored data in a quantile regression model for earnings,

$$Q_{y\mid \mathbf{x}_{it}}(\tau) = \mathbf{x}_{it}^\top \beta(\tau). \quad (4.3)$$

### 4.3 Empirical Results

We first apply several standard approaches to estimate models 4.2 and 4.3. Table 4.2 presents results for the parameter of interest. The table presents estimates from ordinary least squares (OLS), quantile regression (QR), and censored least absolute deviations (CLAD). While the first column (labeled OLS1) includes standard least squares estimates, 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. These approaches were employed by Chay and Powell (2001).

It is interesting to compare the results at the conditional mean (first two columns in Table 4.2) and the conditional median (last five columns). All estimators listed in Table
Table 4.2: Black-white earnings differentials. All quantile models are estimated at the median. Standard errors are presented in parentheses.

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS1</th>
<th>OLS2</th>
<th>QR1</th>
<th>QR2</th>
<th>PH</th>
<th>POR</th>
<th>CLAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black-white gap</td>
<td>-0.304</td>
<td>0.453</td>
<td>-0.213</td>
<td>-0.476</td>
<td>-0.471</td>
<td>-0.436</td>
<td></td>
</tr>
<tr>
<td>(0.025)</td>
<td>(0.036)</td>
<td>(0.003)</td>
<td>(0.020)</td>
<td>(0.030)</td>
<td>(0.026)</td>
<td>(0.060)</td>
<td></td>
</tr>
<tr>
<td>After 1964 Civil Rights Act</td>
<td>0.208</td>
<td>0.317</td>
<td>0.326</td>
<td>0.453</td>
<td>0.115</td>
<td>0.115</td>
<td>0.238</td>
</tr>
<tr>
<td>(0.036)</td>
<td>(0.078)</td>
<td>(0.015)</td>
<td>(0.049)</td>
<td>(0.036)</td>
<td>(0.028)</td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>Black-white gap after</td>
<td>0.057</td>
<td>0.030</td>
<td>0.129</td>
<td>0.055</td>
<td>0.125</td>
<td>0.120</td>
<td>0.126</td>
</tr>
<tr>
<td>(0.036)</td>
<td>(0.058)</td>
<td>(0.015)</td>
<td>(0.033)</td>
<td>(0.035)</td>
<td>(0.042)</td>
<td>(0.044)</td>
<td></td>
</tr>
<tr>
<td>1964 Civil Rights Act</td>
<td>(0.036)</td>
<td>(0.058)</td>
<td>(0.015)</td>
<td>(0.033)</td>
<td>(0.035)</td>
<td>(0.042)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>10170</td>
<td>4886</td>
<td>10170</td>
<td>4886</td>
<td>10170</td>
<td>10170</td>
<td>10170</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mature Workers</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Black-white gap</td>
<td>-0.202</td>
<td>-0.160</td>
<td>-0.212</td>
<td>-0.143</td>
<td>-0.252</td>
<td>-0.251</td>
<td>-0.206</td>
</tr>
<tr>
<td>(0.020)</td>
<td>(0.033)</td>
<td>(0.027)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.021)</td>
<td>(0.068)</td>
<td></td>
</tr>
<tr>
<td>After 1964 Civil Rights Act</td>
<td>0.195</td>
<td>0.384</td>
<td>0.269</td>
<td>0.350</td>
<td>0.166</td>
<td>0.164</td>
<td>0.250</td>
</tr>
<tr>
<td>(0.023)</td>
<td>(0.061)</td>
<td>(0.006)</td>
<td>(0.036)</td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>Black-white gap after</td>
<td>0.092</td>
<td>0.107</td>
<td>0.020</td>
<td>0.035</td>
<td>0.083</td>
<td>0.084</td>
<td>0.020</td>
</tr>
<tr>
<td>(0.031)</td>
<td>(0.052)</td>
<td>(0.041)</td>
<td>(0.036)</td>
<td>(0.034)</td>
<td>(0.024)</td>
<td>(0.044)</td>
<td></td>
</tr>
<tr>
<td>1964 Civil Rights Act</td>
<td>(0.031)</td>
<td>(0.052)</td>
<td>(0.041)</td>
<td>(0.036)</td>
<td>(0.034)</td>
<td>(0.024)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Number of observations</td>
<td>14490</td>
<td>6153</td>
<td>14490</td>
<td>6153</td>
<td>14490</td>
<td>14490</td>
<td>14490</td>
</tr>
</tbody>
</table>

4.2 are suspected to deliver biased results due to the presence of unobserved individual heterogeneity. In the case of young workers, OLS1 estimates at the conditional mean are significantly different than the ones obtained at the conditional median using QR1 and CLAD. The estimates obtained by using OLS2 and QR2 are also biased because they only consider uncensored observations, which yields a distribution that is shifted to the left. PH, POR and CLAD address censoring but ignore individual heterogeneity possibly correlated with the covariates. The analysis of the results reported in Table 4.2 indicates that the median effects of the 1964 Civil Rights Act are small among mature workers but (weakly) 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 4.1 where the QR1, QR2, PH, PO, CLAD and 3-step estimators are used to estimate the effect of the policy on various quantiles. Note that we have here 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.

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 4.1 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 are strictly concave. In order to avoid the potential bias caused by endogenous individual effects and censoring, we employ the 3-step estimator. The 2-step estimator offers similar conclusions to the 3-step estimator, and therefore, nonparametric 2-step results are not reported to save space. Unlike the conclusion obtained using QR2 or CLAD for instance, we notice that the effect of the 1964 Civil Rights Act is monotonically increasing and significant at the quantiles of the conditional earnings distribution of young workers. Indeed, our simulations showed that under random effects, both 3-step and Powell estimators have a similar performance because they take care of the censoring, but under fixed effects, only the 3-step has a good performance. Therefore, any empirical difference between the two estimators may reflect the presence of endogeneity.

Chay and Powell (2001) use a semiparametric censored regression model to investigate 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 3-step 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 the distribution of 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.

Lastly, using Figure 4.2, we present additional empirical evidence obtained from a short-$T$ panel data. The previous results were based on a panel with $T = 15$, which is convenient because (i) the evidence presented in Section 3 indicates that 3-step performs well in mod-
Figure 4.1: Quantile effects of the black-white gap after the Civil Rights Act of 1964. 3-step is the censored quantile regression with fixed effects estimator. The continuous dotted line shows 3-step estimates and grey area a 95 percent pointwise confidence interval. QR1 denotes the classical quantile regression estimator and QR2 is the quantile regression estimator used on a sample that does not include top-censored observations. CLAD is the version of Powell’s estimator introduced in Wang and Fygenson (2009), POR denotes the censored quantile regression estimator proposed by Portnoy (2003) and PH denotes Peng and Huang (2008) estimator.
Figure 4.2: Quantile effects of the black-white gap after the Civil Rights Act of 1964 using Chay and Powell’s (2001) sample. 3-step is the censored quantile regression with fixed effects estimator. The continuous dotted line shows 3-step estimates and grey area a 95 percent pointwise confidence interval. QR1 denotes the classical quantile regression estimator and QR2 is the quantile regression estimator used on a sample that does not include top-censored observations. CLAD is the version of Powell’s estimator introduced in Wang and Fygenson (2009), POR denotes the censored quantile regression estimator proposed by Portnoy (2003) and PH denotes Peng and Huang (2008) estimator.
erate $T$-large $N$ panels, and (ii) the estimation of $N$ nuisance parameter may not lead to significant biases in the estimates of the effect of the policy. It is important, however, to evaluate how sensible are our conclusions to employing Chay and Powell (2001)'s sample of workers, which was constructed considering only four years of data. We then focus on the years 1963 and 1964 (the period before 1965, which is the year when the Title VII of the Civil Rights Act went into effect) and 1970 and 1971. The two labor groups are similarly defined. Our results shown in Figure 4.2 continue to indicate that (i) the black-white gap was reduced after the policy was introduced, having a significant effect among young workers but a negligible effect among mature workers; (ii) while the policy has a moderate effect in the lower tail of the conditional earnings distribution of young workers, it has a large effect at the upper tail; (iii) competing quantile regression methods fail to uncover the large effect in the upper tail.

5 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. Two estimators are investigated. The first is a parametric 3-step estimator, whereas the second one is a nonparametric 2-step estimator. The proposed estimators extend, respectively, the works of Chernozhukov and Hong (2002) and Tang, Wang, He, and Zhu (2012) to the panel data with fixed effects. We have studied the asymptotic properties of the estimators, and derived their limiting properties using sequential limits. We use Monte Carlo simulations to assess the finite sample performance of the estimators. The results show evidence that when the data are censored and individual fixed effects are correlated with covariates, both the 3-step and 2-step estimators perform well, presenting a small bias, low RMSE and a good coverage rate.

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. Our results indicate that the policy had a beneficial effect among young workers, while the mature workers were not affected. Among young workers, the 1964 Civil Rights Act has a moderate effect in the lower tail, but a more pronounced effect at the upper tail of the earnings distribution,
suggesting that the policy benefited more the young black worker with high income rather than any other group of the black working population.

Although the model developed here is quite general, we have not considered the case where $C_u$ is a latent variable potentially dependent on covariates. We assume that the censoring points are observable. Making the censoring dependent on covariates creates an endogeneity problem that has been recently discussed by Khan and Tamer (2009). Introducing covariate dependent censoring in our paper would force us to consider a quantile minimum distance estimator and a different asymptotic theory, which would result in a different manuscript for future research. It should be also noted that we do not offer a specific recommendation for tuning parameter selection on the penalized approach, although we provide evidence that suggests that shrinkage of the fixed effects may be associated with important gains in the case of censored data. Finally, the literature on quantile regression panel data models would also benefit from future studies about whether bootstrap is valid for the proposed method.

References


### A Appendix: Proofs

#### A.1 Parametric 3-step

##### A.1.1 Consistency

We start by showing consistency of $\hat{\beta}^0$ and $\hat{\beta}^1$ as $(T, N)_{\text{seq}} \to \infty$. In what follows we follow Chernozhukov and Hong (2002, p.880) and consider $\delta_{NT}$ as a parameter sequence. Recall that $\theta = (\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$ and let $\Theta = \mathcal{A} \times \mathcal{B}$.

Under conditions A1–A4, the consistency of $\hat{\beta}^0$ and $\hat{\beta}^1$ follows directly from the fact that $(\hat{\gamma}^T, \hat{\theta}_0^0, \hat{\theta}_0^1, \delta_{NT})^T \overset{p}{\to} (\gamma_0^T, \theta_0^0, \theta_0^1, 0)^T \in \Gamma \times \Theta \times \Theta \times \Delta$, a compact set, as $(N, T)_{\text{seq}} \to \infty$.

In fact, $(\hat{\gamma}^T, \hat{\theta}_0^0, \hat{\theta}_0^1, \delta_{NT})$ can be thought of as a $Z$-estimator with the following estimating
equations
\[
\Psi_{NT}(\gamma, \theta^0, \theta^1, \delta) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\delta_{it}}{p(\hat{X}_{it}^\top \gamma)} - \frac{1-\delta_{it}}{1-p(\hat{X}_{it}^\top \gamma)} \right) p' \left( \hat{X}_{it}^\top \gamma \right) \hat{X}_{it} \\
\psi_T(y_{it} - \hat{X}_{it}^\top \theta^0) \hat{X}_{it} 1(p(\hat{X}_{it}^\top \gamma) > 1 - \tau + c) = o_p(1),
\]
where \( \hat{X}_{it} = (z_{it}^\top, x_{it}^\top)^\top \), \( z_{it} \) denotes an indicator variable for the individual effect \( \alpha_i \), and \( \dot{X}_{it} = (z_{it}^\top, \dot{x}_{it}^\top)^\top \).

To see this result, note that \( \gamma \) is the only unknown in the first set of equations, \( \theta^0 \) and \( \gamma \) are the only unknowns from the second set of equations, and \( \theta^1 \) and \( \theta^0 \) are the only unknowns from the third set of equations. Therefore, we could solve the system of equations by solving \( \gamma \) from the first set of equations, then plug the solution of \( \gamma \) into the second set of equations and solve it for \( \theta^0 \). The third set of equation indeed consists of a single trivial equation. All it says is that the solution \( \delta_{NT} \rightarrow 0 \) as \( T \rightarrow \infty \). And finally we plug the solutions of \( \theta^0 \) and \( \delta \) into the fourth set of equations, and solve for \( \theta^1 \). Therefore, we can derive the consistency of \( (\hat{\gamma}^\top, \hat{\theta}^0, \hat{\theta}^1, \delta_{NT})^\top \) simultaneously.

For fixed \( N \), we send \( T \) to infinity, and verify the conditions of Theorem 5.9 of van der Vaart (1998). We need to show that the model is identified, and the law of large numbers holds uniformly in a compact parameter set.

The uniform convergence of the estimating equations can be verified elementwise using Corollary 3.1 of Newey (1991). For the \( j \)th estimating equation, since the summands in the braces of the following display are i.i.d.,
\[
\frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{\delta_{it}}{p(\hat{X}_{it}^\top \gamma)} - \frac{1-\delta_{it}}{1-p(\hat{X}_{it}^\top \gamma)} \right] p' \left( \hat{X}_{it}^\top \gamma \right) \hat{X}_{itj} \right\}
\]
and the law of large numbers holds for each \( \gamma \). Differentiating the \( j \)th element of the score with respect to \( \gamma_k \), the \( k \)th element of \( \gamma \), we obtain the \( (j, k) \)th element of the Hessian matrix of the parametric classification (probability) model, which by Condition A3 is assumed to be finite for \( \gamma \in \Gamma \). Therefore, all the conditions of Corollary 3.1 of Newey (1991) are verified, and we obtain
\[
\sup_{\gamma \in \Gamma} \left| \frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \left( \frac{\delta_{it} p' \left( \hat{X}_{it}^\top \gamma \right)}{p(\hat{X}_{it}^\top \gamma)} \hat{X}_{itj} - \frac{(1-\delta_{it}) p' \left( \hat{X}_{it}^\top \gamma \right)}{1-p(\hat{X}_{it}^\top \gamma)} \hat{X}_{itj} \right) \right| \rightarrow 0.
\]
Now we analyze the second set of the estimating equations,
\[
\frac{1}{T} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \psi_{\tau}(y_{it} - X_{it}^{\top} \theta^0) X_{it} 1(p(X_{it}^{\top} \gamma) > 1 - \tau + c) = o_p(1).
\]
Note that \( \{1(y_{it} - X_{it}^{\top} \theta^0) > 0|\theta^0 \in \Theta\} \) and \( \{1(p(X_{it}^{\top} \gamma) > 1 - \tau + c|\gamma \in \Gamma\} \) are type I classes of Andrews (1994), or VC classes, therefore, they are Donsker. By Example 19.20 of van der Vaart (1998),
\[
\left\{ \frac{1}{N} \sum_{i=1}^{N} \psi_{\tau}(y_{it} - X_{it}^{\top} \theta^0) X_{it} 1(p(X_{it}^{\top} \gamma) > 1 - \tau + c) | (\gamma, \theta^0) \in \Gamma \times \Theta \right\}
\]
is a Donsker class, and hence a Glivenko-Cantelli class.

Similarly, we could verify that
\[
\left\{ \frac{1}{N} \sum_{i=1}^{N} \psi_{\tau}(y_{it} - X_{it}^{\top} \theta^1) X_{it} 1(X_{it}^{\top} \theta^0 > C_{it} + \delta) | (\theta^0, \theta^1, \delta \in \Theta \times \Theta \times \Delta) \right\}
\]
is also a Donsker and Glivenko-Cantelli class.

Now we established the uniform consistency of the estimating equations. Notice that under model (2.1), \( P(y_{it}^* < X_{it}^{\top} \theta_0) = \tau \) where \( X_{it} = (z_{it}, x_{it}^{\top})^{\top} \), and \( \theta_0 = (\alpha_0^{\top}, \beta_0^{\top})^{\top} \). Thus, under the QR censored model and conditions A1–A4, as in Chernozhukov and Hong (2002), we have that
\[
E[\psi_{\tau}(y_{it} - X_{it}^{\top} \theta_0) X_{it} 1(p(X_{it}^{\top} \gamma_0) > 1 - \tau + c)] = E[\psi_{\tau}(y_{it} - X_{it}^{\top} \theta_0) X_{it} 1(X_{it}^{\top} \theta_0 > C_{it} + v)]
\]
\[
= E[\{X_{it} 1(X_{it}^{\top} \theta_0 > C_{it} + v) \} \{\tau - 1(y_{it} < X_{it}^{\top} \theta_0)\}]
\]
\[
= 0,
\]
because \( P(y_{it} < X_{it}^{\top} \theta_0 | X_{it}, C_{it}, X_{it}^{\top} \theta_0 > C_{it} + v) = \tau \) \( \{X_{it}^{\top} \theta_0 > C_{it}, y_{it} - X_{it}^{\top} \theta_0\} \) has \( \tau \)th conditional quantile at 0. Therefore, under model 2.1 and conditions A1–A4, \( \gamma_0 \) and \( \theta_0 \) are identified, and all the conditions of Theorem 5.9 of van der Vaart (1998) are verified.

Hence, we obtain that, as \((N, T)_{seq} \rightarrow \infty\),
\[
\hat{\beta}^0 \overset{p}{\rightarrow} \beta_0 \text{ and } \hat{\beta}^1 \overset{p}{\rightarrow} \beta_0.
\]

### A.1.2 Asymptotic normality

To show asymptotic normality of the estimator, as in Chernozhukov and Hong (2002), we start by deriving the asymptotic distribution of \( \hat{\beta}^0 \), then we proceed to the normality of \( \hat{\beta}^1 \).
For the proof of asymptotic normality of $\hat{\beta}^0$, we first produce the Bahadur representation of the QR estimator. Define

$$H_{IT}^{(1)}(\alpha, \beta; \gamma) = \frac{1}{T} \sum_{t=1}^{T} \{ \tau - 1(y_{it} \leq \alpha \beta + x_{it}^\top \beta) \} \{ p(X_{it}^\top \gamma) > 1 - \tau + c \}$$

$$H_{IT}^{(2)}(\alpha, \beta; \gamma) = E[H_{IT}^{(1)}(\alpha, \beta; \gamma)]$$

$$H_{NT}^{(2)}(\alpha, \beta; \gamma) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ \tau - 1(y_{it} \leq \alpha \beta + x_{it}^\top \beta) \} \{ p(X_{it}^\top \gamma) > 1 - \tau + c \}$$

where $\alpha = (\alpha_1, ..., \alpha_N)^\top$.

Under conditions A1–A5, expanding $H_{IT}^{(1)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma)$ and $H_{NT}^{(2)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma)$ around $(\alpha_0, \beta_0)$, and $(\alpha_0, \beta_0)$, respectively, we have

$$H_{IT}^{(1)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma) = -a_i^H(\gamma)(\hat{\alpha}_i - \alpha_0) - A_i^H(\gamma)(\hat{\beta}_i - \beta_0) + O_p((\hat{\alpha}_i - \alpha_0)^2 \vee ||\hat{\beta}_i - \beta_0||^2) \quad (A.1)$$

$$H_{NT}^{(2)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma) = -\frac{1}{N} \sum_{i=1}^{N} A_i^H(\gamma)(\hat{\alpha}_i - \alpha_0) - \frac{1}{N} \sum_{i=1}^{N} B_i^H(\gamma)(\hat{\beta}_i - \beta_0) + O_p((\hat{\alpha}_i - \alpha_0)^2 \vee ||\hat{\beta}_i - \beta_0||^2) \quad (A.2)$$

where

$$a_i^H(\gamma) := E[f_i(0|X_{it})1(p(X_{it}^\top \gamma) > 1 - \tau + c)]$$

$$A_i^H(\gamma) := E[f_i(0|X_{it})x_{it}1(p(X_{it}^\top \gamma) > 1 - \tau + c)]$$

$$B_i^H(\gamma) := E[f_i(0|X_{it})x_{it}x_{it}^\top 1(p(X_{it}^\top \gamma) > 1 - \tau + c)]$$

The remainder terms in equations (A.1) and (A.2) do not depend on $\gamma$; this is because $H_{IT}^{(1)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma)$ and $H_{NT}^{(2)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma)$ depend on $\gamma$ only through the indicator function which is uniformly bounded.

Solving for $(\hat{\alpha}_i - \alpha_0)$ from equation (A.1) we obtain

$$\hat{\alpha}_i - \alpha_0 = -a_i^H(\gamma)^{-1}H_{IT}^{(1)}(\hat{\alpha}_i, \hat{\beta}_i; \gamma) - a_i^H(\gamma)^{-1}A_i^H(\gamma)^\top(\hat{\beta}_i - \beta_0) + O_p((\hat{\alpha}_i - \alpha_0)^2 \vee ||\hat{\beta}_i - \beta_0||^2)$$
and plugging it in equation (A.2), we have

\[
H_{NT}^{(2)}(\hat{\alpha}^0, \hat{\beta}^0; \gamma) = \frac{1}{N} \sum_{i=1}^{N} A_i^H(\gamma) a_i^H(\gamma)^{-1} H_{iT}^{(1)}(\hat{\alpha}_i^0, \hat{\beta}_i^0; \gamma) + \frac{1}{N} \sum_{i=1}^{n} a_i^H(\gamma)^{-1} A_i^H(\gamma) A_i^H(\gamma)^\top (\hat{\beta}_i^0 - \beta_0)
- \frac{1}{N} \sum_{i=1}^{N} B_i^H(\gamma)(\hat{\beta}_i^0 - \beta_0) + O_p(\max_{1 \leq i \leq N}(\hat{\alpha}_i^0 - \alpha_0)^2 \vee ||\hat{\beta}_i^0 - \beta_0||^2).
\]

Solving for \((\hat{\beta}^0 - \beta_0)\), we obtain

\[
\hat{\beta}^0 - \beta_0 + o_p(||\hat{\beta}^0 - \beta_0||) = \left( \frac{1}{N} \sum_{i=1}^{N} (a_i^H(\gamma)^{-1} A_i^H(\gamma) A_i^H(\gamma)^\top - B_i^H(\gamma)) \right)^{-1} \left[ H_{NT}^{(2)}(\hat{\alpha}^0, \hat{\beta}^0; \gamma) - \frac{1}{N} \sum_{i=1}^{N} A_i^H(\gamma) a_i^H(\gamma)^{-1} H_{iT}^{(1)}(\hat{\alpha}_i^0, \hat{\beta}_i^0; \gamma) \right] + O_p(\max_{1 \leq i \leq N}(\hat{\alpha}_i^0 - \alpha_0)^2).
\]

Since \(\gamma \xrightarrow{p} \gamma_0\), by the continuous mapping theorem,

\[
\frac{1}{N} \sum_{i=1}^{N} [a_i^H(\gamma)^{-1} A_i^H(\gamma) A_i^H(\gamma)^\top - B_i^H(\gamma)] \xrightarrow{p} \frac{1}{N} \sum_{i=1}^{N} [a_i^H(\gamma_0)^{-1} A_i^H(\gamma_0) A_i^H(\gamma_0)^\top - B_i^H(\gamma_0)] := \Lambda_N,
\]
as \(T \to \infty\).

It is important to note that we can use the continuous mapping theorem stated before because \(a_i^H(\gamma)\) is continuous. We only verify that \(a_i^H(\gamma)\) is indeed continuous at \(\gamma\), since the verifications of the continuity property of \(A_i^H(\gamma)\) and \(B_i^H(\gamma)\) are similar. Recall the definition

\[
a_i^H(\gamma) = \mathbb{E}[f_i(0|X_{it})1(p(X_{it}^\top \gamma) > 1 - \tau + c)] = \int_{X_{it}^\top \gamma > 1 - \tau + c} f_i(0|X_{it}) G_X dX,
\]
the result follows directly from the continuity of Condition A2.

Now, note that

\[
H_{NT}^{(2)}(\hat{\alpha}^0, \hat{\beta}^0; \gamma) = \mathbb{H}_N^{(2)}(\alpha_0, \beta_0; \gamma_0) + \mathbb{H}_{NT}^{(2)}(\alpha_0^0, \beta_0^0; \gamma) - H_{NT}^{(2)}(\alpha_0, \beta_0; \gamma_0) + \mathbb{H}_{NT}^{(2)}(\alpha_0, \beta_0^0; \gamma) - \mathbb{H}_{NT}^{(2)}(\alpha_0, \beta_0; \gamma_0) + O_p(1/T).
\]

The equality follows from the computational property of the QR estimator \(||\mathbb{H}_{NT}^{(2)}(\hat{\alpha}^0, \hat{\beta}^0)|| = O_p(T^{-1})\).

Now we verify that \(\mathbb{H}_{NT}^{(2)}(\alpha, \beta; \gamma) - H_{NT}^{(2)}(\alpha, \beta; \gamma)\) is stochastically equicontinuous. To this end, we verify that \(\{[\tau - 1(y - X^\top \theta)]x1(p(X^\top \gamma) > 1 - \tau + c)|(\theta, \gamma) \in \Theta \times \Gamma\}\) is a
Donsker class. Note that this step was shown previously in the proof of the consistency of \( \hat{\beta}_0 \), by verifying that this function is a type I class as in Andrews (1994), or a VC class. Therefore,

\[
H_{NT}^{(2)}(\hat{\alpha}^0, \hat{\beta}^0; \hat{\gamma}) = \mathbb{H}_{NT}^{(2)}(\alpha_0, \beta_0; \gamma_0) + o_p(1/\sqrt{T}).
\]

Similarly, from the definitions and exactly same argument as in the proof for consistency, we have that

\[
\frac{1}{N} \sum_{i=1}^N A_i^H(\hat{\gamma}) a_i^H(\hat{\gamma})^{-1} H_{IT}^{(1)}(\hat{\alpha}^0, \hat{\beta}^0; \hat{\gamma}) = \frac{1}{N} \sum_{i=1}^N A_i^H(\gamma_0) a_i^H(\gamma_0)^{-1} \mathbb{H}^{(1)}_{IT}(\alpha_0, \beta_0; \gamma_0) + o_p(1/\sqrt{T}).
\]

Therefore,

\[
\hat{\beta}^0 - \beta_0 = \hat{\Lambda}^{-1}_N \left[ \mathbb{H}_{NT}^{(2)}(\alpha_0, \beta_0; \gamma_0) - \frac{1}{N} \sum_{i=1}^N A_i^H(\gamma_0)a_i^H(\gamma_0)^{-1} \mathbb{H}^{(1)}_{IT}(\alpha_0, \beta_0; \gamma_0) \right] + o_p(1/\sqrt{T})
\]

\[
= \hat{\Lambda}^{-1}_N \left[ \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^{T} \{ \tau - 1(y_{it} \leq \alpha_{i0} + x_{it}\beta_0) \} x_{it} 1(p(\hat{X}_{it}^T \gamma) > 1 - \tau + c) \right] + o_p(1/\sqrt{T})
\]

\[
\quad - \frac{1}{N} \sum_{i=1}^N A_i^H(\gamma_0)a_i^H(\gamma_0)^{-1} \frac{1}{T} \sum_{t=1}^{T} \{ \tau - 1(y_{it} \leq \alpha_{i0} + x_{it}\beta_0) \} 1(p(\hat{X}_{it}^T \gamma) > 1 - \tau + c) (x_{it} - A_i^H(\gamma_0)a_i^H(\gamma_0)^{-1})
\]

\[
+ o_p(1/\sqrt{T}).
\]

Thus, as \( T \to \infty \),

\[
\sqrt{T}(\hat{\beta}_0 - \beta) \overset{d}{\to} N(0, \hat{\Lambda}^{-1}_N \hat{V}_N \hat{\Lambda}^{-1}_N / N)
\]

where \( \hat{V}_N = \frac{1}{N} \sum_{i=1}^N \tau(1-\tau)E\{(x_{it} - A_i^H(\gamma_0)a_i^H(\gamma_0)^{-1})(x_{it} - A_i^H(\gamma_0)a_i^H(\gamma_0)^{-1})^\top 1(p(\hat{X}_{it}^T \gamma) > 1 - \tau + c)\} \), and \( \hat{\Lambda}_N := \frac{1}{N} \sum_{i=1}^N [a_i^H(\gamma_0)^{-1} A_i^H(\gamma_0) A_i^H(\gamma_0)^\top - B_i^H(\gamma_0)] \).

Now, letting \( N \to \infty \), we have

\[
\sqrt{NT}(\hat{\beta}^0 - \beta_0) \overset{d}{\to} N(0, \hat{\Lambda}^{-1} \hat{V} \hat{\Lambda}^{-1}).
\]

where \( \hat{V}_N \to \hat{V} \) and \( \hat{\Lambda}_N \to \hat{\Lambda} \) as \( N \to \infty \), with \( \hat{V} = \lim_{N \to \infty} \hat{V}_N \) and \( \hat{\Lambda} = \lim_{N \to \infty} \hat{\Lambda}_N \) being positive definite matrices.

To complete the proof of asymptotic normality, we need to derive the limiting distribution of \( \hat{\beta}^1 \). To this end, we first produce the Bahadur representation of the QR estimator.
Moreover, as in Chernozhukov and Hong (2002), we consider $\delta_{NT}$ as a parameter sequence. Define

\[ S_{IT}^{(1)}(\alpha_i^1, \beta^1; \theta^0, \delta) = \frac{1}{T} \sum_{t=1}^{T} \{ \tau - 1(y_{it} \leq \alpha_i^1 + x_{it}^\top \beta^1) \} 1(X_{it}^\top \theta^0 > C_{it} + \delta) \]

\[ S_{IT}^{(1)}(\alpha_i^1, \beta^1; \theta^0, \delta) = E[S_{IT}^{(1)}(\alpha_i^1, \beta^1; \theta^0, \delta)] \]

\[ = E \{ [\tau - F_i(\alpha_i^1 - \alpha_{io} + x_{it}^\top (\beta^1 - \beta_0)|x_{it}] 1(X_{it}^\top \theta^0 > C_{it} + \delta) \} \]

\[ S_{NT}^{(2)}(\alpha^1, \beta^1; \theta^0, \delta) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \{ \tau - 1(y_{it} \leq \alpha_i^1 + x_{it}^\top \beta^1) \} x_{it}1(X_{it}^\top \theta^0 > C_{it} + \delta) \]

\[ S_{NT}^{(2)}(\alpha^1, \beta^1; \theta^0, \delta) = ES_{NT}^{(2)}(\alpha^1, \beta^1; \theta^0, \delta) \]

\[ = \frac{1}{N} \sum_{i=1}^{N} E[\{ \tau - F_i(\alpha_i^1 - \alpha_{io} + x_{it}^\top (\beta^1 - \beta_0)|x_{it}) \} x_{it}1(X_{it}^\top \theta^0 > C_{it} + \delta)] \]

where $\alpha^1 = (\alpha_1^1, ..., \alpha_k^1)^\top$.

Expanding $S_{IT}^{(1)}(\hat{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT})$ and $S_{NT}^{(2)}(\hat{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT})$ around $(\alpha_{io}, \beta_0)$, and $(\alpha_0, \beta_0)$, respectively, we have

\[ S_{IT}^{(1)}(\hat{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) = -a_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})(\hat{\alpha}^1 - \alpha_{io}) - A_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})(\hat{\beta}^1 - \beta_0) \]

\[ + O_p((\hat{\alpha}^1 - \alpha_{io})^2 \vee ||\hat{\beta}^1 - \beta_0||^2) \quad (A.3) \]

\[ S_{NT}^{(2)}(\hat{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) = -\frac{1}{N} \sum_{i=1}^{N} A_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})(\hat{\alpha}^1 - \alpha_{io}) - \frac{1}{N} \sum_{i=1}^{N} B_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})(\hat{\beta}^1 - \beta_0) \]

\[ + O_p((\hat{\alpha}^1 - \alpha_{io})^2 \vee ||\hat{\beta}^1 - \beta_0||^2) \quad (A.4) \]

where

\[ a_{S}^{\theta}(\theta, \delta) := E[f_i(0|X_{it})1(X_{it}^\top \theta > C_{it} + \delta)], \]

\[ A_{S}^{\theta}(\theta, \delta) := E[f_i(0|X_{it})x_{it}1(X_{it}^\top \theta > C_{it} + \delta)], \]

\[ B_{S}^{\theta}(\theta, \delta) := E[f_i(0|X_{it})x_{it}x_{it}^\top 1(X_{it}^\top \theta > C_{it} + \delta)]. \]

Solving for $(\hat{\alpha}^1 - \alpha_{io})$ from equation (A.3)

\[ \hat{\alpha}^1 - \alpha_{io} = -a_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})^{-1} S_{IT}^{(1)}(\hat{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) - a_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})^{-1} A_{S}^{\theta}(\hat{\theta}^0, \delta_{NT})^\top (\hat{\beta}^1 - \beta_0) \]

\[ + O_p((\hat{\alpha}^1 - \alpha_{io})^2 \vee ||\hat{\beta}^1 - \beta_0||^2), \]
and plugging in equation (A.4), we have

\[
S_{NT}^{(2)}(\boldsymbol{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) = \frac{1}{N} \sum_{i=1}^{N} A_i^S(\hat{\theta}^0, \delta_{NT})a_i^s(\hat{\theta}^0, \delta_{NT})^{-1} S_{iT}^{(1)}(\hat{\alpha}_i^1, \hat{\beta}_i^1; \hat{\theta}^0, \delta_{NT})
\]

\[
+ \frac{1}{N} \sum_{i=1}^{n} a_i^S(\hat{\theta}^0, \delta_{NT})^{-1} A_i^S(\hat{\theta}^0, \delta_{NT})A_i^S(\hat{\theta}^0, \delta_{NT})\top (\hat{\beta}_i^1 - \beta_0)
\]

\[
- \frac{1}{N} \sum_{i=1}^{N} B_i^S(\hat{\theta}^0, \delta_{NT})(\hat{\beta}_i^1 - \beta_0) + O_p(\max_{1 \leq i \leq N} |\hat{\alpha}_i^1 - \alpha_{i0}|^2 \lor ||\hat{\beta}^1 - \beta_0||^2).
\]

Solving for \((\hat{\beta}^1 - \beta_0)\), we obtain

\[
\hat{\beta}^1 - \beta_0 + o_p(||\hat{\beta}^1 - \beta_0||)
\]

\[
= \left( \frac{1}{N} \sum_{i=1}^{N} (a_i^S(\hat{\theta}^0, \delta_{NT})^{-1} A_i^S(\hat{\theta}^0, \delta_{NT})A_i^S(\hat{\theta}^0, \delta_{NT})\top - B_i^S(\hat{\theta}^0, \delta_{NT})) \right)^{-1}
\]

\[
\left[ S_{NT}^{(2)}(\boldsymbol{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT})
\right]
\]

\[
- \frac{1}{N} \sum_{i=1}^{N} A_i^S(\hat{\theta}^0, \delta_{NT})a_i^s(\hat{\theta}^0, \delta_{NT})^{-1} S_{iT}^{(1)}(\hat{\alpha}_i^1, \hat{\beta}_i^1; \hat{\theta}^0, \delta_{NT}) + O_p(\max_{1 \leq i \leq N} |\hat{\alpha}_i^1 - \alpha_{i0}|^2).
\]

Since \(\hat{\theta}^0 \xrightarrow{p} \theta_0\) and \(\delta_{NT} \rightarrow 0\), by the continuous mapping theorem,

\[
\frac{1}{N} \sum_{i=1}^{N} [a_i^S(\theta_0, 0)^{-1} A_i^S(\theta_0, 0)A_i^S(\theta_0, 0)\top - B_i^S(\theta_0, 0)] \xrightarrow{p}
\]

\[
\frac{1}{N} \sum_{i=1}^{N} [a_i^S(\theta_0, 0)^{-1} A_i^S(\theta_0, 0)A_i^S(\theta_0, 0)\top - B_i^S(\theta_0, 0)] := \Lambda_N.
\]

In addition, we can rewrite

\[
S_{NT}^{(2)}(\boldsymbol{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) = S_{NT}^{(2)}(\alpha_0, \beta_0; \theta_0, \delta_{NT}) + [S_{NT}^{(2)}(\boldsymbol{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) - S_{NT}^{(2)}(\alpha_0, \beta_0; \theta_0, \delta_{NT})
\]

\[
+ S_{NT}^{(2)}(\alpha^1, \hat{\beta}_1; \theta_0, \delta_{NT}) - S_{NT}^{(2)}(\alpha_0, \beta_0; \theta_0, \delta_{NT})] + O_p(1/T).
\]

The equality follows from the computational property of the QR estimator \(||S_{NT}^{(2)}(\boldsymbol{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT})|| = O_p(T^{-1}).\)

Now we verify that \(S_{NT}^{(2)}(\alpha, \beta; \theta, \delta) - S_{NT}^{(2)}(\alpha_0, \beta_0; \theta_0, \delta_{NT})\) is stochastically equicontinuous. To this end, we verify that \(\{[\tau - 1(y - X^T \theta)]1(p(X^T \gamma) > 1 - \tau + c) \in \Theta \times \Gamma\}\) is a VC class, which is presented previously in the proof of the consistency of \(\hat{\beta}_1\). Therefore,

\[
S_{NT}^{(2)}(\boldsymbol{\alpha}^1, \hat{\beta}^1; \hat{\theta}^0, \delta_{NT}) = S_{NT}^{(2)}(\alpha_0, \beta_0; \theta_0, 0) + o_p(1/\sqrt{T}).
\]
Similarly, we have that

$$\frac{1}{N} \sum_{i=1}^{N} A_i^S(\hat{\theta}, \delta_{NT}) a_i^S(\hat{\theta}_0, \delta_{NT})^{-1} S_i^{(1)}(\alpha, \beta; \hat{\theta}_0, \delta_{NT})$$

$$- \frac{1}{N} \sum_{i=1}^{N} A_i^S(\theta_0, 0) a_i^S(\theta_0, 0)^{-1} S_i^{(1)}(\alpha, \beta; \theta_0, 0) + o_p(1/\sqrt{T}).$$

Therefore,

$$\hat{\beta}^1 - \beta_0 = \Lambda^{-1}_N \left[ S_{NT}^{(2)}(\alpha_0, \beta_0; \theta_0, 0) - \frac{1}{N} \sum_{i=1}^{N} A_i^S(\theta_0, 0) a_i^S(\theta_0, 0)^{-1} S_i^{(1)}(\alpha, \beta; \theta_0, 0) \right] + o_p(1/\sqrt{T})$$

$$= \Lambda^{-1}_N \left[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \tau - 1(y_{it} \leq \alpha + x_{it}^T \beta_0) \right\} x_{it} 1(X_{it}^T \theta_0 > C_{it}) \right] - \frac{1}{N} \sum_{i=1}^{N} A_i^S(\theta_0, 0) a_i^S(\theta_0, 0)^{-1} \frac{1}{T} \sum_{t=1}^{T} \left\{ \tau - 1(y_{it} \leq \alpha + x_{it}^T \beta_0) \right\} 1(X_{it}^T \theta_0 > C_{it}) \right] + o_p(1/\sqrt{T})$$

$$= \Lambda^{-1}_N \left[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \tau - 1(y_{it} \leq \alpha + x_{it}^T \beta_0) \right\} 1(X_{it}^T \theta_0 > C_{it})(x_{it} - A_i^S(\theta_0, 0) a_i^S(\theta_0, 0)^{-1}) + o_p(1/\sqrt{T}) \right]$$

Thus, as $T \to \infty$,

$$\sqrt{T}(\hat{\beta}^1 - \beta_0) \overset{d}{\to} N(0, \Gamma^{-1}_N V_N \Gamma^{-1}_N / N)$$

where

$$V_N := \frac{1}{N} \sum_{i=1}^{N} \tau(1 - \tau) E\{ (x_{it} - A_i^S(\theta_0, 0) a_i^S(\theta_0, 0)^{-1})(x_{it} - A_i^S(\theta_0, 0) a_i^S(\theta_0, 0)^{-1})^T 1(X_{it}^T \theta_0 > C_{it}) \},$$

$$\Lambda_N := \frac{1}{N} \sum_{i=1}^{N} [a_i^S(\theta_0, 0)^{-1} A_i^S(\theta_0, 0) A_i^S(\theta_0, 0)^T - B_i^S(\theta_0, 0)].$$

Finally, as $N \to \infty$ and by condition A5 – noticing that $a_i^S(\theta_0, 0) = a_i$, $A_i^S(\theta_0, 0) = A_i$, and $B_i^S(\theta_0, 0) = B_i$ – we have

$$\sqrt{NT}(\hat{\beta}^1 - \beta_0) \overset{d}{\to} N(0, \Lambda^{-1} V \Lambda^{-1}),$$

where $V_N \to V$ and $\Lambda_N \to \Lambda$ as $N \to \infty$.  

45
A.2 Nonparametric 2-step

Let \( A_T = \{ \alpha : ||\alpha - \alpha_0|| \leq \epsilon_T \} \), \( B_T = \{ \beta : ||\beta - \beta_0|| \leq \epsilon_T \} \), \( \Theta_T = A_T \times B_T \), and \( \mathcal{H}_T = \{ \pi : ||\pi - \pi_0|| \leq \epsilon_T \} \). We use \( a \lesssim b \) to signify that there is a constant \( C \) such that \( a \leq Cb \). We usually suppress arguments of the functions \( \psi = \tau - 1(y_{it} \leq \alpha_i + x_{it}^\top \beta) \), \( f_i(X_{it}^\top \theta_0 | X_{it}) \), and \( \pi(\alpha_i, x_{it}, C_{it}) \) for notational simplicity. Therefore, \( \psi(\alpha_i, \beta; y_{it}, x_{it}) \equiv \psi(\alpha_i, \beta; y_{it}, x_{it}) \equiv \psi(\theta_i; y_{it}, x_{it}) \equiv f_i(X_{it}^\top \theta_0 | X_{it}) \equiv f_i(\theta_0) \pi(\alpha_i, x_{it}, C_{it}) \equiv \pi(\alpha_i; \cdot) \) and \( \pi_0(\alpha_i, x_{it}, C_{it}) \equiv \pi_0(\alpha_i; \cdot) \).

Define

\[
\mathbb{H}^{(1)}_{it}(\alpha_i, \beta, \pi) = \frac{1}{T} \sum_{t=1}^{T} \psi(\alpha_i, \beta; \cdot)1(\pi(\alpha_i; \cdot) > 1 - \tau + c_T) = \frac{1}{T} \sum_{t=1}^{T} m_{it}^{(1)}(\alpha_i, \beta, \pi, c_T)
\]

\[
H^{(1)}_{it}(\alpha_i, \beta, \pi) = E[\mathbb{H}^{(1)}_{it}(\alpha_i, \beta)]
\]

\[
H^{(1)}_{it}(\alpha_i, \beta, \pi) = E\left[ \left\{ \left\{ \tau - P(y_{it}^* < \alpha_i + x_{it}^\top \beta | x_{it})1\{X_{it}^\top \theta_i > C_{it}\}\right\}1(\pi(\alpha_i; \cdot) > 1 - \tau + c_T) \right\} \right]
\]

\[
\mathbb{H}^{(2)}_{NT}(\alpha, \beta, \pi) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it} \psi(\alpha_i, \beta; \cdot)1(\pi(\alpha_i; \cdot) > 1 - \tau + c_T) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} m_{it}^{(2)}(\alpha_i, \beta, \pi, c_T)
\]

\[
H^{(2)}_{NT}(\alpha, \beta, \pi) = E[\mathbb{H}^{(2)}_{NT}(\alpha_i, \beta)]
\]

\[
H^{(2)}_{NT}(\alpha, \beta, \pi) = E\mathbb{H}^{(2)}_{NT}(\alpha_i, \beta)
\]

\[
H^{(2)}_{NT}(\alpha, \beta, \pi) = \left( \mathbb{H}^{(1)}_{1N}(\alpha_1, \beta, \pi), \cdots, \mathbb{H}^{(1)}_{NT}(\alpha_N, \beta, \pi), \mathbb{H}^{(2)}_{NT}(\alpha, \beta, \pi) \right)^\top
\]

\[
H_{NT}(\alpha, \beta, \pi) = E[H_{NT}(\alpha, \beta, \pi)].
\]

The next lemma helps in the derivation of the results.

**Lemma 1.** Under Assumptions B1–B5, for all positive \( \epsilon_T = o(1) \), we have for fixed \( N \)

\[
\sup_{(\theta, \pi) \in \Theta_T \times \mathcal{H}_T} ||H_{NT}(\alpha, \beta, \pi) - H_{NT}(\alpha, \beta, \pi_0) - H_{NT}(\alpha_0, \beta_0, \pi_0)|| = o_p(T^{-1/2})
\]

**Proof.** We verify condition (3.2) of Theorem 3 of Chen, Linton, and van Keilegom (2003). Let

\[
m_{it}(\theta, \pi, c_T) = (m_{it}^{(1)}(\alpha_1, \beta, \pi, c_T), \cdots, m_{it}^{(1)}(\alpha_N, \beta, \pi, c_T), m_{it}^{(2)}(\alpha, \beta, \pi, c_T)).
\]
Note that
\[
||m_t(\theta, \pi, c_T) - m_t(\theta', \pi', c_T)||^2 \\
\leq \left( \psi(\theta_i; \cdot) \{ \pi(\alpha_i; \cdot) > 1 - \tau + c_T \} - \psi(\theta_i'; \cdot) \{ \pi'(\alpha_i; \cdot) > 1 - \tau + c_T \} \right)^2 + \cdots \\
+ \left( \psi(\theta_N; \cdot) \{ \pi(\alpha_N; \cdot) > 1 - \tau + c_T \} - \psi(\theta_N'; \cdot) \{ \pi'(\alpha_N; \cdot) > 1 - \tau + c_T \} \right)^2 \\
+ \frac{1}{N^2} \left( \frac{1}{N} \sum_{i=1}^{N} ||\psi(\theta_i; \cdot) \{ \pi(\alpha_i; \cdot) > 1 - \tau + c_T \} - \psi(\theta_i'; \cdot) \{ \pi'(\alpha_i; \cdot) > 1 - \tau + c_T \}|| \right)^2 M^2 \\
:= B_{1}^2 + B_{2}^2 + B_{3}^2 + \cdots + B_{N}^2 + \frac{M^2}{N^2} \left( \sum_{i=1}^{N} B_{1i} + B_{2i} + B_{3i} \right)^2,
\]
where \( B_{1i}^2 = (\tau^2 + 2) \{ \pi(\alpha_i; \cdot) > 1 - \tau + c_T \} - 1 \{ \pi'(\alpha_i; \cdot) > 1 - \tau + c_T \} \), \( B_{2i}^2 = |1 \{ y_{it}^* < C_{it}, X_{it}^\top \theta_i > C_{it} \} - 1 \{ y_{it}^* < C_{it}, X_{it}^\top \theta_i' > C_{it} \} | \), and \( B_{3i}^2 = |1 \{ y_{it}^* > C_{it}, y_{it}^* - X_{it}^\top \theta_i < C_{it} \} - 1 \{ y_{it}^* > C_{it}, y_{it}^* - X_{it}^\top \theta_i' < C_{it} \} | \), all of which are \( O(e_T) \) as are \( B_{1}, B_{2}, \) and \( B_{3} \) of Tang, Wang, He, and Zhu (2012), respectively. Thus, condition (3.2) of Chen, Linton, and van Keilegom (2003) holds with \( r = 2 \) and \( s_j = 1/2 \).

**A.2.1 Consistency**

Proof. Fixing \( N \), we need to show that \( (\hat{\alpha}, \hat{\beta}) \overset{p}{\to} (\alpha_0, \beta_0) \) as \( T \to \infty \). Note that under model (2.1) and B5.1, \( P(y_{it}^* < X_{it}^\top \theta_{i0} | x_{it}) = \tau \) and \( \pi_0(\alpha_i; \cdot) > 1 - \tau + c_T \) implies \( X_{it}^\top \theta_{i0} > C_{it} \). Therefore \( H_{it}^{(1)}(\alpha_0, \beta_0, \pi_0) = 0 \) and \( H_{NT}^{(2)}(\alpha_0, \beta_0, \pi_0) = 0 \).

Now we verify conditions (1.1)–(1.3) and (1.5') of Chen, Linton, and van Keilegom (2003) [CLK], as (1.4) follows from Lemma 1 above.

First, notice that (1.1) in CLK holds by the computational property of quantile regression estimators. Second, we verify condition (1.2) in CLK. By assumptions B3 and B5.1, \((\alpha_0, \beta_0)\) is the minimizer of
\[
\frac{1}{N} \sum_{i=1}^{N} E[\rho_T(y_{it} - X_{it}^\top \theta_{i0})1(\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T)],
\]
which is strictly convex on \((\alpha_0, \beta_0)\). Therefore \((\alpha_0, \beta_0)\) is the unique minimizer of \( \frac{1}{N} \sum_{i=1}^{N} E[\rho_T(y_{it} - X_{it}^\top \theta_{i0})1(\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T)] \).

Third, we check (1.3). By assumption B5.3, for any \( \epsilon_T \to 0, (\alpha, \beta) \in \mathcal{A}^N \times \mathcal{B} \) and \( ||\pi - \pi_0||_{\infty} \leq \epsilon_T \), we have
\[
||H_{it}^{(1)}(\alpha_i, \beta, \pi) - H_{it}^{(1)}(\alpha_i, \beta, \pi_0)|| \lesssim E[1(\pi(\alpha_i; \cdot) > 1 - \tau + c_T) - 1(\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T)] \lesssim \epsilon_T,
\]
Thus, $H_{NT}^{(1)}(\alpha, \beta, \pi)$ is uniformly continuous in $\pi$ at $\pi_0$. Similarly, $H_{NT}^{(2)}(\alpha, \beta, \pi)$ is uniformly continuous in $\pi$ at $\pi_0$.

Finally, to verify (1.5') in CLK note that $E[\|X_{it}1(y_{it} \leq X_{it}^\top \theta_0)1(\pi(\alpha_i; \cdot) > 1 - \tau + c_T)^2] \leq E\|X\|^2 \leq M_\tau^2$ by condition B1. Therefore, by Chebyshev’s inequality,

$$
\sup_{\theta_i \in \Theta_i, \|\pi - \pi_0\| \leq \epsilon} |H_{NT}^{(1)}(\alpha, \beta, \pi) - H^{(1)}_{NT}(\alpha, \beta, \pi)| = o_p(1),
$$

$$
\sup_{\theta_i \in \Theta_i, \|\pi - \pi_0\| \leq \epsilon} |H_{NT}^{(2)}(\alpha, \beta, \pi) - H^{(2)}_{NT}(\alpha, \beta, \pi)| = o_p(1).
$$

\[\square\]

A.2.2 Asymptotic normality

**Proof.** 4th-root consistency of $\hat{\theta}$

We rewrite $H_{NT}(\alpha, \beta, \pi) = b_1(\theta) + b_2(\pi) + b_3(\theta, \pi)$, where

$$
b_1(\theta) = \begin{pmatrix}
E[\psi(\theta_1; \cdot)]\{\pi_0(\alpha_1; \cdot) > 1 - \tau + c_T\} \\
\vdots \\
\frac{1}{N} \sum_{i=1}^{N} E[\psi(\theta_i; \cdot)]\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\} x_{it}
\end{pmatrix},
$$

$$
b_2(\pi) = \begin{pmatrix}
E[\psi(\theta_{N0}; \cdot)]\{\pi(\alpha_N; \cdot) > 1 - \tau + c_T\} - \{\pi_0(\alpha_N; \cdot) > 1 - \tau + c_T\} \\
\vdots \\
\frac{1}{N} \sum_{i=1}^{N} E[\psi(\theta_{0i}; \cdot)]\{\pi(\alpha_i; \cdot) > 1 - \tau + c_T\} - \{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\} x_{it}
\end{pmatrix},
$$

$$
b_3(\theta, \pi) = \begin{pmatrix}
E[\psi(\theta_1; \cdot) - \psi(\theta_{01}; \cdot)]\{\pi(\alpha_1; \cdot) > 1 - \tau + c_T\} - \{\pi_0(\alpha_1; \cdot) > 1 - \tau + c_T\} \\
\vdots \\
\frac{1}{N} \sum_{i=1}^{N} E[\psi(\theta_i; \cdot) - \psi(\theta_{0i}; \cdot)]\{\pi(\alpha_i; \cdot) > 1 - \tau + c_T\} - \{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\} x_{it}
\end{pmatrix}.
$$

(i) Since $\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T$ implies $X_{it}^\top \theta_{0i} > C_u$, we have $P(y_{it} < X_{it}^\top \theta_{0i} | x_{it})1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\} = P(y_{it} < X_{it}^\top \theta_{0i} | x_{it})1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\} = \tau 1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\}$. Setting $\|\theta - \theta_0\| \leq \epsilon_T$,

$$
b_1(\theta) = -(K_1 + \tau K_2) \times \begin{pmatrix}
\alpha_1 - \alpha_{10} \\
\vdots \\
\alpha_N - \alpha_{N0} \\
\beta - \beta_0
\end{pmatrix} + o(\|\theta - \theta_0\|)
$$

48
where $K_1$ is
\[
\begin{pmatrix}
E_{f_i}(\theta_{10})1(X_{i1}^T\theta_{10} > C_{i1}), 0, \cdots, 0, E_{f_i}(\theta_{10})1(X_{i1}^T\theta_{10} > C_{i1})x_{i1}^T \\
\vdots \\
\frac{1}{N}E_{f_i}(\theta_{i0})1(X_{N_i}^T\theta_{i0} > C_{i1}), E_{f_i}(\theta_{i0})1(X_{N_i}^T\theta_{i0} > C_{i1})x_{N_i}^T
\end{pmatrix}
\]
and
\[
D_{11}, 0, \cdots, 0, D_{12} \\
\vdots \\
0, \cdots, 0, D_{N_1}, D_{N_2} \\
D_{12}, \cdots, D_{N_2}, D
\]
with $D_1 = E[1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\}1\{X_{i1}^T\theta \leq C_{u1}\}]$, $D_2 = E[x_{i1}1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\}1\{X_{i1}^T\theta \leq C_{u1}\}]$, and $D = E[x_{i1}x_{i1}^T1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T\}1\{X_{i1}^T\theta \leq C_{u1}\}]$.

(ii) Following Tang, Wang, He, and Zhu (2012), we have $b_2(\pi) = 0$ if $||\pi - \pi_0||_{\infty} = o_p(T^{-1/4})$.

(iii) In addition, similarly to Tang, Wang, He, and Zhu (2012) $||b_3(\theta, \pi)|| = o(T^{-1/4})$.

Therefore, by the above results and Lemma 1 we have
\[
\mathbb{H}_{NT}(\alpha, \beta, \pi) - \mathbb{H}_{NT}(\alpha_0, \beta_0, \pi_0)
= -[K_1 + \tau K_2](\hat{\theta} - \theta_0) + o_p(||\theta - \theta_0||) + b_3(\theta, \pi) + o_p(T^{-1/2})
\]
and
\[
-[K_1 + \tau K_2](\hat{\theta} - \theta_0) + o_p(||\hat{\theta} - \theta_0||) = -\mathbb{H}_{NT}(\alpha_0, \beta_0, \pi_0) + o_p(T^{-1/4}).
\]
By Central Limit Theorem, $\mathbb{H}_{NT}(\alpha_0, \beta_0, \pi_0) = O_p(T^{-1/2})$, we have $||\hat{\theta} - \theta_0|| = o_p(T^{-1/4})$.

**Root-N consistency and asymptotic distribution of $\hat{\beta}$**

By Assumption B5.1, $\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T$ implies $X_{i1}\theta_{i0} > C_{u1} + c_T$, where $T^{1/4}c_T > c^*$ for some $c^* > 0$. Since $||\hat{\theta} - \theta_0|| = o_p(T^{-1/4})$, we have $K_2(\hat{\theta} - \theta_0) = o_p(||\hat{\theta} - \theta_0||)$.

By Assumptions B1 and B5.1, $\pi_0(\alpha_i; \cdot) > 1 - \tau + c_T$ and $\hat{\pi}(\alpha_i; \cdot) > 1 - \tau + c_T$ imply $X_{i1}^T\theta_0 > C_{u1}$ and $X_{i1}^T\theta > C_{u1}$. Therefore,
\[
b_3(\hat{\theta}, \hat{\pi}) = o_p(||\hat{\theta} - \theta_0||)
\]
and by Lemma 1
\[
-\mathbb{H}_{NT}(\theta_0, \pi_0) = K_1(\hat{\theta} - \theta_0) + o_p(T^{-1/2}) + o_p(||\hat{\theta} - \theta_0||).
\]
By CLT, \( \mathbb{H}(\theta_0, \pi_0) = O_p(1) \). Therefore, \( \hat{\theta} - \theta_0 = O(T^{-1/2}) \). The previous equation is equivalent to

\[
- \frac{1}{T} \sum_{t=1}^{T} (\tau - 1\{y_{it} \leq X_{it}^\top \theta_{10}\})1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + \epsilon T\}
= E_{i_t}(\theta_{10})1(X_{it}^\top \theta_{10} > C_{it})(\hat{\alpha}_i - \alpha_{10}) + E_{i_t}(\theta_{10})1(X_{it}^\top \theta_{10} > C_{it})x_{it}(\hat{\beta} - \beta_0) + o_p(1)
\]

\[
- \frac{1}{T} \sum_{t=1}^{T} (\tau - 1\{y_{Ni} \leq X_{Ni}^\top \theta_{N0}\})1\{\pi_0(\alpha_i; \cdot) > 1 - \tau + \epsilon T\}
= E_{i_t}(\theta_{N0})1(X_{Ni}^\top \theta_{N0} > C_{it})(\hat{\alpha}_i - \alpha_{N0}) + E_{i_t}(\theta_{N0})1(X_{Ni}^\top \theta_{N0} > C_{it})x_{Ni}(\hat{\beta} - \beta_0) + o_p(1)
\]

\[
- \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tau - 1(y_{it} \leq \alpha_{i0} + x_{it}^\top \beta_0))\{x_{it}1(\pi_0(\alpha_i; \cdot) > 1 - \tau + \epsilon T)\}
= \frac{1}{N} \sum_{i=1}^{N} E_{i_t}(\theta_{10})1(X_{it}^\top \theta_{10} > C_{it})x_{it}(\hat{\alpha}_i - \alpha_{i0}) + \frac{1}{N} \sum_{i=1}^{N} E_{i_t}(\theta_{10})1(X_{it}^\top \theta_{10} > C_{it})x_{it}x_{it}^\top (\hat{\beta} - \beta_0).
\]

Let \( a_i := E_{i_t}(f_{i_t}(0|X_{it})1(X_{it}^\top \theta_{10} > C_{it})], \ A_i := E_{i_t}(f_{i_t}(0|X_{it})x_{it}1(X_{it}^\top \theta_{10} > C_{it})], \ B_i := E_{i_t}(f_{i_t}(0|X_{it})x_{it}x_{it}^\top 1(X_{it}^\top \theta_{10} > C_{it})]. \) Then, solving the above equations for \( (\hat{\beta} - \beta_0) \), we have

\[
\sqrt{T}(\hat{\beta} - \beta_0)
= \Lambda_N^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \{\tau - 1(y_{it} \leq \alpha_{i0} + x_{it}^\top \beta_0)\}(x_{it} - A_i a_i^{-1})1(\pi_0(\alpha_i; \cdot) > 1 - \tau + \epsilon T) + o_p(1)
\]

where \( \Lambda_N = \frac{1}{N} \sum_{i=1}^{N} [a_i^{-1} A_i A_i^\top - B_i] 1(\pi_0(\alpha_i; \cdot) > 1 - \tau). \) Thus, by Central Limit Theorem,

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} \{\tau - 1(y_{it} \leq \alpha_{i0} + x_{it}^\top \beta_0)\}(x_{it} - A_i a_i^{-1})1(\pi_0(\alpha_i; \cdot) > 1 - \tau + \epsilon T) \xrightarrow{d} N(0, V_N/N),
\]

where \( V_N = \tau(1 - \tau) \frac{1}{N} \sum_{i=1}^{N} E(x_{it} - A_i a_i^{-1})(x_{it} - A_i a_i^{-1})^\top 1(\pi_0(\alpha_i; \cdot) > 1 - \tau). \)

Thus,

\[
\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Lambda_N^{-1} V_N \Lambda_N^{-1}/N).
\]

Now, letting \( N \to \infty \), by B7 we have

\[
\sqrt{NT}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \Lambda^{-1} V \Lambda^{-1}),
\]

where \( \Lambda_N \to \Lambda \) and \( V_N \to V \) as \( N \to \infty. \)

\[\square\]