NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Galina Besstremyannaya, Sergei Golovan

INSTRUMENTAL VARIABLE QUANTILE REGRESSION FOR CLUSTERED DATA

BASIC RESEARCH PROGRAM
WORKING PAPERS

SERIES: ECONOMICS
WP BRP 255/EC/2022

Instrumental variable quantile regression for clustered data

Galina Besstremyannaya* Sergei Golovan**

June 13, 2022

Abstract

The purpose of the paper is to enable inference in case of quantile regression with endogenous covariates and clustered data. We prove that the instrumental variable quantile regression estimator is consistent where there is correlation of errors within clusters. We derive an asymptotic distribution for the estimator, which may be used for inference for a given τ. As regards inference based on the entire instrumental variable quantile regression process, we prove that cluster-based bootstrapping of a statistic of a certain class offers a computationally tractable approach for implementing asymptotic tests. Our theoretical results concerning the asymptotic properties of the instrumental variable quantile regression estimator for clustered data are supported by simulation analysis. The empirical part of the paper applies the technique to estimation of the earning equations of US men and women where female labor supply is endogenous and subject to the shock of World War II.

JEL Classification Codes: C21, C23, C26, D12

Keywords: quantile regression, endogeneity, clustered data, instrumental variables

*National Research University Higher School of Economics, Moscow, Pokrovsky blvd. 11, gbesstremyannya@hse.ru; The paper was prepared in the framework of the HSE University's Basic Research Program.

**New Economic School, Moscow, Skolkovskoe shosse, 45; sgolovan@nes.ru

1 Introduction

Empirical research often deals with datasets where observations come from a number of clusters, so that observations can be considered independent across clusters but correlation is assumed within each cluster. Clusters may be defined by geographic location (countries, regions, or municipalities), economic activity or peer groups (industries, establishments, classes), or kinship (families).

The assumption of independence of observations does not hold for clustered data. So inference requires generalization of estimators into the versions with cluster-robust standard errors, and this has been implemented for a range of least-squares methods (Cameron and Miller, 2011). As regards the quantile regression approach, to the best of our knowledge, cluster-robust standard errors were introduced only for the estimator in the model with exogenous covariates (Parente and Santos Silva, 2016, Hagemann, 2017). However, quantile regression is often applied to study heterogeneous response of the dependent variable to endogenous covariates. The instrumental variable quantile regression model of Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2006) has proved to be a widely used and computationally convenient econometric technique for this purpose.[1] But, the asymptotic properties of the estimator are developed in Chernozhukov and Hansen (2006) only under the assumption of i.i.d. observations.

The purpose of our paper is to extend the results of Chernozhukov and Hansen (2006) and enable inference in case of quantile regression with endogenous covariates and clustered data. We prove that the instrumental variable quantile regression estimator is consistent where there is correlation of errors within clusters, and we derive the asymptotic distribution for the estimator. As regards inference based on the instrumental variable quantile regression process as a whole, we extend the methodology of Chernozhukov and Hansen (2006) which uses bootstrap to compute critical values of the test statistics. We propose resampling by clusters and prove that it offers an approach to this computation and, hence, to the implementation of asymptotic tests. Our theoretical results concerning the asymptotic properties of the instrumental variable quantile regression estimator for clustered data are supported by the simulation analysis. The estimator for clustered data is given empirical application to gauge the impact of female labor supply on the wages of men and women in the US in 1940–1950.

The present paper is structured as follows. Section 2 sets up the model with the quantile regression with endogenous covariates and clustered data, and derives the asymptotic distribution for the estimator in the model. Section 3 develops inference for the entire instrumental variable quantile process for clustered data. The results of the simulations are shown in Section 4. Section 5 presents the results of the empirical study. Proofs are given in the Appendix.

2 The model

2.1. Setup

The instrumental variable quantile regression model of Chernozhukov and Hansen (2005) is applied to the τth structural quantile of the outcome variable Y as a function of the observed values x of exogenous covariates X and values d of the endogenous variables D, conditional on X and an instrumental variable Z. The conditional structural quantile $Q_\tau(Y, X, D|X, Z)$ is a

[1] As of May 2022, combined citations of the seminal paper of Chernozhukov and Hansen (2005) and of related work on instrumental variable quantile regression (Chernozhukov and Hansen, 2004, 2006, 2008) are close to 2,500.

function $q(d, x, \tau)$ which is assumed to be linear in covariates:

$$q(d, x, \tau) = d'\alpha(\tau) + x'\beta(\tau). \tag{1}$$

Let the following assumption hold:

Assumption 1 (Chernozhukov and Hansen (2005), P. 248. Main conditions of the instrumental variable quantile regression model)**.** *Given a common probability space (Ω, F, P) for P-almost every value of X, Z, the following conditions A1–A5 hold jointly:*

A1 *(Potential outcomes) Given $X = x$, for each d, $Y_d = q(d, x, U_d)$, where $U_d \sim U(0, 1)$ and $q(d, x, \tau)$ is strictly increasing in τ.*

A2 *(Independence) Given $X = x$, $\{U_d\}$ is independent of Z.*

A3 *(Selection) Given $X = x$, $Z = z$, for unknown function δ and random vector ν, $D \equiv \delta(z, x, \nu)$.*

A4 *(Rank invariance or rank similarity) For each d and d', given (ν, X, Z), either (a) $U_d = U'_d$ or (b) $U_d \sim U'_d$.*

A5 *Observed variables consist of $Y \equiv q(D, X, U_D)$; $D \equiv \delta(Z, X, \nu)$, X, Z.*

Under these regularity conditions, a set of moment equations expresses the quantiles of the outcome variable Y, conditional on exogenous covariates X and a vector of instruments Z, as a linear function of $q(D, X, \tau)$ (Theorem 1 in Chernozhukov and Hansen (2005)). The solution to these structural equations is the population-level estimator in the instrumental variable quantile regression model. The finite-sample analogue of the population-level estimator in Chernozhukov and Hansen (2005) and the theory for the general inference in Chernozhukov and Hansen (2006) are developed under the assumption that (Y_i, D_i, X_i, Z_i) are i.i.d. for all observations i in the sample.

In our model we keep the aforementioned setup by Chernozhukov and Hansen (2005)[2] but relax the i.i.d. assumption and consider that observations are sampled from data generating process with clusters:[3]

$$Y_t = D'_t\alpha(U_{Dt}) + X'_t\beta(U_{Dt}), \quad D_t = \delta(Z_t, X_t, \nu_t), \quad t = 1, \ldots, T,$$

where t is the index within a cluster, ν_t is a random variable, and $U_{D_t} \sim U[0, 1]$ conditionally on Z_t, X_t.

2.2. Identification of the finite-sample estimator with clustered standard errors

The set of conditions for identification of the estimator for clustered data modifies the corresponding assumptions of Chernozhukov and Hansen (2006)[4] to account for the fact that observations are sampled in clusters. In clustered data, condition R1 assumes the existence of intra-cluster correlation of observations but considers that observations are i.i.d. across clusters. Conditions of full rank and continuity of the Jacobian matrices in the moment conditions (R3), and uniformity and smoothness of instruments and weights (R4) are formulated for the data-generating process with clusters. The condition for compactness and convexity of the space for the vector of coefficients (R2) is taken directly from Chernozhukov and Hansen (2006). Index i denotes clusters in the assumptions below.

[2]For the reader's convenience, we keep the original notations from Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2006) throughout this paper.

[3]Without loss of generality the size of all clusters is considered fixed, see Parente and Santos Silva (2016).

[4]We keep the names of the original assumptions.

Assumption 2. R1 *(Sampling)* $\{(Y_{i1},\ldots,Y_{iT}),(D_{i1},\ldots,D_{iT}),(X_{i1},\ldots,X_{iT}),(Z_{i1},\ldots,Z_{iT})\}$ *are iid defined on the probability space* (Ω,F,P) *and take values in a compact set.*

R2 *(Compactness and convexity) For all* τ, $(\alpha(\tau),\beta(\tau))\in \operatorname{int}\mathcal{A}\times\mathcal{B}$, $\mathcal{A}\times\mathcal{B}$ *are compact and convex.*

R3 *(Full rank and continuity) Data generating process* Y_t *has bounded conditional density, a.s.* $\sup_{y\in\mathbb{R}} f_{Y_t|X_t,D_t,Z_t}(y)<K$, *and for* $\pi\equiv(\alpha,\beta,\gamma)$, $\theta\equiv(\alpha,\beta)$, *and*

$$\Pi_t(\pi,\tau)\equiv E((\tau-1(Y_t<D_t'\alpha+X_t'\beta+\Phi_t(\tau)'\gamma))\Psi_t(\tau)),$$
$$\Pi_t(\theta,\tau)\equiv E((\tau-1(Y_t<D_t'\alpha+X_t'\beta))\Psi_t(\tau)),\qquad \Psi_t(\tau)\equiv V_t(\tau)\cdot[\Phi_t(\tau)',X_t']',$$

where $\Phi_t(\tau)=\Phi_t(\tau,Z_t,X_t)$ *are transformations of instruments,* $V_t(\tau)=V_t(\tau,Z_t,X_t)$ *are weights, Jacobian matrices* $\frac{\partial}{\partial(\alpha,\beta)}\Pi_t(\theta,\tau)$ *and* $\frac{\partial}{\partial(\beta,\gamma)}\Pi_t(\pi,\tau)$ *are continuous and have full rank, uniformly over* $\mathcal{A}\times\mathcal{B}\times\mathcal{G}\times\mathcal{T}$ *and the image of* $\mathcal{A}\times\mathcal{B}$ *under the mapping* $(\alpha,\beta)\mapsto\Pi_t(\theta,\tau)$ *is simply connected for all* $t=1,\ldots,T$.

R4 *(Estimated instruments and weights)* $Wp\to 1$, *the functions* $\hat\Phi_t(\tau,z,x)$, $\hat V_t(\tau,z,x)\in\mathcal{F}$ *and* $\hat\Phi_t(\tau,z,x)\to_p\Phi_t(\tau,z,x)$, $\hat V_t(\tau,z,x)\to_p V_t(\tau,z,x)$ *uniformly in* (τ,z,x) *over compact sets, where* $\Phi_t(\tau,z,x)$ *and* $V_t(\tau,z,x)\in\mathcal{F}$, *the functions* $f_t(\tau,z,x)\in\mathcal{F}$ *are uniformly smooth functions in* (z,x) *with the uniform smoothness order* $\eta>\dim(d,z,x)/2$, *and* $\|f_t(\tau',z,x)-f_t(\tau,z,x)\|<C|\tau'-\tau|^a$, $C>0$, $a>0$, *for all* (z,x,τ,τ') *and* $t=1,\ldots,T$.

2.3. The estimator for clustered data and its asymptotic properties

Conditions R1–R4 from Assumption 2 and the Chernozhukov and Hansen (2005) set of regularity conditions for the population-level instrumental variable quantile regression make it possible to find a unique solution to the finite-sample analogue of conditional moment equations with clustered data.

Theorem 1. *Under Assumption 1 and Assumption 2,* $(\alpha',\beta')=(\alpha(\tau)',\beta(\tau)')$ *uniquely solves the system of equations* $E(\tau-I(Y_t<D_t'\alpha+X_t'\beta)\Psi_t(\tau))=0$ *over* $\mathcal{A}\times\mathcal{B}$ *for all* $t=1,\ldots,T$.

The estimation follows the two-step procedure of Chernozhukov and Hansen (2004) which minimizes the weighted objective function in quantile regression:

$$Q_N(\tau,\alpha,\beta,\gamma)=\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_\tau(Y_{it}-D_{it}'\alpha-X_{it}'\beta-\hat\Phi_{it}(\tau)'\gamma)\hat V_{it}(\tau),\tag{2}$$

where $\rho_\tau(u)=u(\tau-I(u\le 0))$ is the Koenker and Bassett (1978) loss function, $\hat\Phi_{it}(\tau)=\hat\Phi_t(\tau,X_{it},Z_{it})$ and $\hat V_{it}(\tau)=\hat V_t(\tau,X_{it},Z_{it})$ are weights.

Note that owing to clustered data, the weighted sums of the values of the loss function are taken over clusters t and observations i within each cluster.

The first step requires the estimate

$$\big(\hat\beta(\alpha,\tau),\hat\gamma(\alpha,\tau)\big)=\underset{\beta,\gamma}{\operatorname{argmin}}\,Q_N(\tau,\alpha,\beta,\gamma).\tag{3}$$

At the second step, the value of α that minimizes the norm of $\hat\gamma(\alpha,\tau)$ is found:

$$\hat\alpha(\tau)=\underset{\alpha\in\mathcal{A}}{\operatorname{argmin}}\,\|\hat\gamma(\alpha,\tau)\|^2_{A(\tau)},\ \text{where } \|\hat\gamma(\alpha,\tau)\|^2_{A(\tau)}=\hat\gamma(\alpha,\tau)'A(\tau)\hat\gamma(\alpha,\tau),\tag{4}$$

where $A(\tau)$ is a uniformly positive definite matrix in the set $\mathcal{T}$. Finally, $\hat{\beta}(\tau) = \hat{\beta}(\hat{\alpha}(\tau), \tau)$.[5]

Next, we derive the asymptotic distribution for the instrumental variable quantile regression estimator for clustered data.

Theorem 2. *Given Assumption 1 and Assumption 2, for $\varepsilon_{it}(\tau) = Y_{it} - D_{it}'\alpha(\tau) + X_{it}'\beta(\tau)$ and $l_{it}(\tau, \theta(\tau)) = \tau - I(\varepsilon_{it}(\tau) < 0)$:*

$$\sqrt{N}(\hat{\theta}(\cdot) - \theta(\cdot)) = -J(\cdot)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T} l_{it}(\cdot, \theta(\cdot))\Psi_{it}(\cdot) + o_p(1) \Rightarrow b(\cdot) \tag{5}$$

for $N \to \infty$, where $b(\cdot)$ is a mean zero Gaussian process with covariance function $E(b(\tau)b(\tau')') = J(\tau)^{-1}S(\tau, \tau')J(\tau')^{-1}$,

$$J(\tau) = E\left(\sum_{t=1}^{T} f_{\varepsilon_t(\tau)}(0|X_t, D_t, Z_t)\Psi_t(\tau)[D_t', X_t']\right),$$

$$S(\tau, \tau') = E\left(\sum_{t=1}^{T}\sum_{s=1}^{T} l_t(\tau, \theta(\tau))l_s(\tau', \theta(\tau'))\Psi_t(\tau)\Psi_s(\tau')'\right).$$

Following Powell (1986), the estimator of $J(\tau)$ is

$$\hat{J}(\tau) = \frac{1}{2Nh_N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T} I(|\hat{\varepsilon}_{it}(\tau)| \leq h_N)\hat{\Psi}_{it}(\tau)[D_{it}', X_{it}']\right),$$

where $\hat{\varepsilon}_{it} = Y_{it} - D_{it}'\hat{\alpha}(\tau) - X_{it}'\hat{\beta}(\tau)$ and bandwidth h_N is chosen so that $h_N \to 0$ and $Nh_N^2 \to \infty$ (see Parente and Santos Silva (2016)).

The matrix $S(\tau, \tau')$ is estimated by its sample analogue:

$$\hat{S}(\tau, \tau') = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\sum_{s=1}^{T} l_{it}(\tau, \theta(\tau))l_{is}(\tau', \theta(\tau'))\hat{\Psi}_{it}(\tau)\hat{\Psi}_{is}(\tau')'\right).$$

It should be noted that Theorem 2 makes it possible to derive asymptotics of the instrumental variable quantile regression estimator for clustered data for each quantile index τ:

$$\sqrt{N}(\hat{\theta}(\tau) - \theta(\tau)) \xrightarrow{d} \mathcal{N}(0, J(\tau)^{-1}S(\tau, \tau)(J(\tau)^{-1})'). \tag{6}$$

It also gives the joint limiting distribution of the estimator for several quantiles $\{\tau \in J\}$, where J is a finite subset of $(0, 1)$:

$$\{\sqrt{N}(\hat{\theta}(\tau) - \theta(\tau))\}_{\tau \in J} \xrightarrow{d} \mathcal{N}(0, \{J(\tau)^{-1}S(\tau, \tau')(J(\tau')^{-1})'\}_{\tau, \tau' \in J}). \tag{7}$$

[5]Note that both the instrumental variable quantile regression estimator and the instrumental variable quantile regression estimator for clustered data depend on $A(\tau)$. Although any positive definite matrix can be employed in this context, Chernozhukov and Hansen (2006) recommend to use the asymptotic variance-covariance matrix of $\hat{\gamma}(\alpha(\tau), \tau)$ as $A(\tau)$. But the asymptotic variance-covariance matrix of $\hat{\gamma}(\alpha(\tau), \tau)$ in quantile regression without clusters is different from this matrix in presence of clusters. Hence, the estimator for clustered data may differ numerically from the Chernozhukov and Hansen (2006) instrumental variable quantile regression estimator. However, we focus on the asymptotic properties of the estimator for clustered data and these properties do not depend on $A(\tau)$.

2.4. General inference

Consider uniform inference for a set of quantiles $\tau \in \mathcal{T}$. We use the general form of the null hypotheses in the notations of Chernozhukov and Hansen (2006):

$$R(\tau)(\theta(\tau) - r(\tau)) = 0, \quad \text{for each } \tau \in \mathcal{T}, \tag{8}$$

where $R(\tau)$ is a given $q \times p$ matrix of rank q, $q \leq p = \dim \theta(\tau)$, and $r(\tau) \in \mathbb{R}^p$, $\theta(\tau)$ and $r(\tau)$ are functions to be estimated. The tests are based on the inference process: $v_N(\cdot) = R(\cdot)(\hat{\theta}(\cdot) - \hat{r}(\cdot))$.

Let the following assumption hold:

Assumption 3 (Chernozhukov and Hansen (2006), P. 501. Conditions for inference)**.**

I1. $R(\cdot)(\theta(\cdot) - r(\cdot)) = g(\cdot)$, *where the functions* $g(\tau)$, $R(\tau)$, $r(\tau)$ *are continuous and either (a)* $g(\tau) = 0$*) for all* τ *(the null hypothesis) or (b)* $g(\tau) \neq 0$ *for some* τ *(the alternative hypothesis).*

I2. $\sqrt{N}(\hat{\theta}(\cdot) - \theta(\cdot)) \Rightarrow b(\cdot)$ *and* $\sqrt{N}(\hat{r}(\cdot) - r(\cdot)) \Rightarrow d(\cdot)$ *jointly in* $\ell^\infty(\mathcal{T})$, *where* $b(\cdot)$ *and* $d(\cdot)$ *are jointly zero mean Gaussian functions that may have different laws under the null and the alternative.*

Assume that along with conditions I1 and I2 from Assumption 3, the estimates in quantile regression for clustered data admit the linear representations below.

Assumption 4. I3. *Linear representations:*

$$\sqrt{N}(\hat{\theta}(\cdot) - \theta(\cdot)) = -J(\cdot)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} l_{it}(\cdot, \theta(\cdot)) \Psi_{it}(\cdot) + o_p(1)$$

and

$$\sqrt{N}(\hat{r}(\cdot) - r(\cdot)) = -H(\cdot)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}(\cdot, r(\cdot)) \Upsilon_{it}(\cdot) + o_p(1)$$

in $\ell^\infty(\mathcal{T})$, *where* $J(\cdot)$ *and* $H(\cdot)$ *are constant invertible matrices, and vectors* $(l_{i1}(\cdot, \theta(\cdot))\Psi_{i1}(\cdot), \ldots, l_{iT}(\cdot, \theta(\cdot))\Psi_{iT}(\cdot))$ *and* $(d_{i1}(\cdot, r(\cdot))\Upsilon_{i1}(\cdot), \ldots, d_{iT}(\cdot, r(\cdot))\Upsilon_{iT}(\cdot))$ *are i.i.d. mean zero for each* τ.

I4. *(a) The estimates* $l_{it}(\cdot, \hat{\theta}(\cdot))\hat{\Psi}_{it}(\cdot)$ *and* $d_{it}(\cdot, \hat{r}(\cdot))\hat{\Upsilon}_{it}(\cdot)$ *take realizations in a Donsker class of functions with a constant envelope and are uniformly consistent in* τ *in the* $L_2(P)$ *norm. (b)* $Wp \to 1$, $E(l_{it}(\tau, \theta(\tau)\hat{\Psi}_{it}(\tau))) = 0$ *and* $E(d_{it}(\tau, r(\tau)\hat{\Upsilon}_{it}(\tau))) = 0$ *for each* i, t. *(c)* $E\|l_{it}(\tau, \theta) - l_{it}(\tau, \theta')\| < C\|\theta - \theta'\|$, $E\|d_{it}(\tau, r) - d_{it}(\tau, r')\| < C\|r - r'\|$, *uniformly in* $\tau \in \mathcal{T}$ *and in* (θ, θ', r, r') *over compact sets.*

Then the inference process in instrumental variable quantile regression for clustered data admits a linear representation.

Proposition 1. *Under Assumptions 1, 3, and 4*

$$\sqrt{N}(v_N(\cdot) - g(\cdot)) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} z_{it}(\cdot) \right) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{T}),$$

where $z_{it}(\cdot) = R(\cdot)(J(\cdot)^{-1} l_{it}(\cdot, \theta(\cdot))\Psi_{it}(\cdot) - H(\cdot)^{-1} d_{it}(\cdot, r(\cdot))\Upsilon_{it}(\cdot))$.

The estimate of $z_{it}(\cdot)$ becomes $\hat{z}_{it}(\cdot) = R(\cdot)(\hat{J}(\cdot)^{-1}l_{it}(\cdot,\hat{\theta}(\cdot))\hat{\Psi}_{it}(\cdot) - \hat{H}(\cdot)^{-1}d_{it}(\cdot,\hat{r}(\cdot))\hat{\Upsilon}_{it}(\cdot))$, where $\hat{J}(\cdot)$ and $\hat{H}(\cdot)$ are any uniformly consistent estimates of $J(\cdot)$ and $H(\cdot)$.

An approach for implementing the asymptotic tests with the help of the Kolmogorov–Smirnov (KS) and the Cramér–von Mises (CM statistic) is formulated in Theorem 4 in Chernozhukov and Hansen (2006): it develops the asymptotic theory for computation of critical values for test statistics. The null hypothesis is rejected when the test statistic exceeds its critical value.

In case of clustered data, we use the aforementioned approach but propose resampling by clusters in order to obtain the critical values for test statistics for the inference process.[6]

The inference is then conducted through the following algorithm.

Step 1. Resampling procedure with clustered data: To resample from $\{\hat{z}_{it}(\tau), i = 1,\ldots,N, t = 1,\ldots,T, \tau \in \mathcal{T}\}$, randomly select B_N subsets of $1,\ldots,N$ of size b without replacement, each of these subsets is denoted I_i, $i = 1,\ldots,B_N$. Define the inference process for the jth subset of data I_j as $v_{j,b,N}(\tau) \equiv 1/b\sum_{i\in I_j}\sum_{t=1}^{T}\hat{z}_{it}(\tau)$. Denote $\hat{S}_{j,b,N} \equiv f(\sqrt{b}v_{j,b,N}(\cdot))$ as

$$\hat{S}_{j,b,N} \equiv \sup_{\tau\in\mathcal{T}} \sqrt{b}\|v_{j,b,N}(\tau)\|_{\hat{\Lambda}(\tau)} \quad \text{or} \quad \hat{S}_{j,b,N} \equiv \int_{\mathcal{T}} \sqrt{b}\|v_{j,b,N}(\tau)\|^2_{\hat{\Lambda}(\tau)}d\tau,$$

where S_N is, respectively, the KS or CM statistic. Note that resampling is conducted by clusters i and not by individual observations.

Step 2. Computation of the critical value of the test statistic based on the resampling procedure with clustered data: The step fully follows Chernozhukov and Hansen (2006). Specifically, for each statistic $S = f(v(\cdot))$, it defines $\Gamma(x) \equiv P(S \leq x)$ and suggests the estimate of $\Gamma(x)$ as $\hat{\Gamma}_{b,N}(x) = 1/B_N\sum_{j=1}^{b_N} I(S_{j,b,N} \leq x)$. The critical value for the test is $c_{b,N}(1-\alpha) = \inf\{c\colon \hat{\Gamma}_{b,N}(c) \geq 1-\alpha\}$, i.e. $(1-\alpha)$th quantile of $\hat{\Gamma}_{b,N}(x)$. The null hypothesis is rejected by the test of level α when $S_N > c_{b,N}(1-\alpha)$.

Theorem 3 justifies the above described procedure.

Theorem 3 (Score subsampling inference for clustered data)**.** *Suppose Assumptions 1, 3, and 4 hold, and that $\hat{J}(\tau) = J(\tau) + o_p(1)$ and $\hat{H}(\tau) = H(\tau) + o_p(1)$ uniformly in τ over $\mathcal{T}$. Then as $B_N \to \infty$, $b \to \infty$, $N \to \infty$:*

(1) *Under the null hypothesis, if Γ is continuous at $\Gamma^{-1}(1-\alpha)$: $c_{b,N}(1-\alpha) \xrightarrow{p} \Gamma^{-1}(1-\alpha)$, $P(S_N > c_{b,N}(1-\alpha)) \to \alpha$;*

(2) *Under the alternative hypothesis, $S_N \xrightarrow{d} \infty$, $c_{b,N}(1-\alpha) = O_p(1)$, $P(S_N > c_{b,N}(1-\alpha)) \to 1$;*

(3) *$\Gamma(x)$ is absolutely continuous at $x > 0$ when the covariance function of v is nondegenerate a.e. in τ.*

The application of the approach is demonstrated below using examples of hypotheses from Chernozhukov and Hansen (2006), for which we formulate scores in case of instrumental variable quantile regression for clustered data and $\dim\alpha \equiv 1$.

[6] An alternative to the below modified approach of Chernozhukov and Hansen (2006) is the modification of the bootstrap procedure, proposed in Hagemann (2017) for quantile regression with clustered data under exogeneity. Indeed, Hagemann (2017) introduces the methodology for general inference (the null hypothesis is for instance, $\alpha(\tau) = 0$ under all $\tau \in \mathcal{T}$) and uses the statistics similar to those of Kolmogorov-Smirnov. But there are differences in the implementation of the procedures, proposed in Hagemann (2017) and Chernozhukov and Hansen (2006). Specifically, Chernozhukov and Hansen (2006) estimate quantile regression once, compute the scores for each observation and then bootstrap these scores. At the same time, Hagemann (2017) uses a conventional bootstrap approach and estimates a large number of quantile regressions: a regression for each sample. Such approach implies computational burden in case of instrumental variable quantile regression, and hence is not employed in the present paper.

1. *No effect of the endogenous variable.*

   H_0: $\alpha(\tau) = 0$ for all $\tau \in \mathcal{T}$, $R(\cdot) \equiv R = [1, 0, \dots, 0]$, $r(\cdot) \equiv 0$.

   In this case, $\hat{z}_{it}(\tau) = R(\tau)[\hat{J}(\tau)^{-1} l_{it}(\tau, \hat{\theta}(\tau))\hat{\Psi}_{it}(\tau)]$, where $l_{it}(\tau, \hat{\theta}(\tau)) = (\tau - I(Y_{it} < D_{it}'\hat{\alpha}(\tau) + X_i'\hat{\beta}(\tau)))$, $\hat{\Psi}_{it}(\tau) = \hat{V}_{it}[\hat{\Phi}_{it}(\tau), X_{it}']'$.

2. *Constant effect of the endogenous variable across quantiles.*

   H_0: $\alpha(\tau) \equiv \alpha = \text{const}$, for all $\tau \in \mathcal{T}$, $R(\cdot) \equiv R = [1, 0, \dots, 0]$ and $r(\cdot) = [\alpha, 0, \dots, 0]'$.

   In this case, $[\hat{\alpha}(1/2), 0, \dots, 0]$ can be taken for $\hat{r}(\cdot)$, and $\hat{z}_{it}(\tau) = R(\tau)[\hat{J}(\tau)^{-1} l_{it}(\tau, \hat{\theta}(\tau))\hat{\Psi}_{it}(\tau) - \hat{J}(1/2)^{-1} l_{it}(1/2, \hat{\theta}(1/2))\hat{\Psi}_{it}(1/2)]$, for $l_{it(\cdot, \hat{\theta}(\cdot))}$ defined in the example with the hypothesis of no effect.

3. *Exogeneity hypothesis.*

   H_0: The coefficient for the endogenous variable in instrumental variable quantile regression equals the coefficient for this variable in the quantile regression under exogeneity, $R(\cdot) \equiv R = [1, 0, \dots, 0]$, $r(\cdot) = \vartheta(\cdot)$, where $r(\cdot)$ is estimated using quantile regression under exogeneity.

   Then, the score is given by $\hat{z}_{it}(\tau) = R(\tau)[\hat{J}(\tau)^{-1} l_{it}(\tau, \hat{\theta}(\tau))\hat{\Psi}_{it}(\tau) - \hat{H}(\tau)^{-1} d_{it}(\tau, \hat{\vartheta}(\tau))]$, where $d_{it}(\tau, \hat{\vartheta}(\tau)) = (\tau - I(Y_{it} < \tilde{X}_{it}'\hat{\vartheta}(\tau)))\tilde{X}_{it}$, $\tilde{X}_{it} = [D_{it}', X_{it}']'$, and following Parente and Santos Silva (2016), $\hat{H}(\tau) = 1/(2Nh_N)\sum_{i=1}^{N}\left(\sum_{t=1}^{T} I(|\hat{\epsilon}_{it}| \le h_N)\tilde{X}_{it}\tilde{X}_{it}'\right)$.

2.5. Simulations

The present section shows the results of simulation analysis of performance of the estimator of the covariance matrix in the instrumental variable quantile regression for clustered data. The data-generating process is

$$
\begin{aligned}
Y_{it} &= D_{it} \cdot \alpha \cdot U_{it} + \beta_0 \cdot U_{it} + X_{it} \cdot \beta_1 \cdot U_{it}, \\
i &= 1, \dots, N, t = 1, \dots, T,
\end{aligned}
$$

where i is the index for cluster, t is the index for observation within a cluster, α, β_0, $\beta_1 \in \mathbb{R}$, so $\alpha(\tau) = \alpha \cdot \tau$, $\beta_0(\tau) = \beta \cdot \tau$, $\beta_1(\tau) = \beta \cdot \tau$.

The intra-cluster correlation of errors is introduced by adding random variable ξ_{it}, which varies across clusters and observations, and ζ_i, which denotes individual effect of clusters. For this purpose, we draw variables ξ_{it} and ζ_i from Gamma distribution:

$$
\begin{aligned}
\xi_{kit} &\sim \Gamma(1, 1), k = 1, \dots, 5, \\
\zeta_{ki} &\sim \Gamma(2, 1), k = 1, \dots, 5.
\end{aligned}
$$

The covariates D, X, the excluded instrument Z and the error term U are then constructed as follows:

$$
\begin{aligned}
D_{it} &= d \cdot (\xi_{1it} + \zeta_{1i}) + \xi_{2it} + \xi_{3it} + \zeta_{2i} + \zeta_{3i} \sim \Gamma(9, 1), \\
Z_{it} &= \xi_{3it} + \xi_{4it} + \zeta_{3i} + \zeta_{4i} \sim \Gamma(6, 1), \\
X_{it} &= \xi_{4it} + \xi_{5it} + \zeta_{4i} + \zeta_{5i} \sim \Gamma(6, 1), \\
U_{it} &= F_{\Gamma(3,1)}(\xi_{1it} + \zeta_{1i}) \sim U(0, 1).
\end{aligned}
$$

The number of clusters $N \in \{100, 200, 500, 1000, 2000\}$ and the size of cluster $T \in \{2, 5, 10\}$. The values of $d = 1$ and $d = 0$ are used to model endogenous and exogenous D, respectively.

The performance of the instrumental variable quantile regression estimator with clustered standard errors is evaluated for three quantile indices: $\tau \in \{0.25, 0.50, 0.75\}$.

For each τ we estimate the conditional τth quantile regression of Y_{it} on D_{it}, X_{it} and a constant, using Z_{it} as an instrument for D_{it}. Then, we focus on the basic inference process and test whether the slope coefficients are equal to their true values: namely, whether $\alpha(\tau) = \alpha \cdot \tau$ and $\beta_1(\tau) = \beta_1 \cdot \tau$. The performance of the Chernozhukov and Hansen (2006) estimator is contrasted with the performance of the instrumental variable quantile regression estimator for clustered data.

Next, we conduct inference based on the entire instrumental variable quantile regression process (approximating it by a set of three quantiles $\mathcal{T} = \{0.25, 0.5, 0.75\}$). Specifically, we test each of the three hypotheses: $\alpha \equiv 0$, $\alpha \equiv \text{const}$, D is exogenous. For each hypothesis we contrast the results of the test based on the Chernozhukov and Hansen (2006) bootstrap (i.e. resampling observations) and the bootstrap which resamples clusters. We simulated 500 samples, and used 10,000 bootstrap draws in each case.

The estimations are performed using custom code in Python.

Table 1 shows the true sizes of the basic inference concerning the coefficient for the endogenous variable α at the 10% level. The standard errors for the Chernozhukov and Hansen (2006) instrumental variable quantile regression estimator yield excessive rejection of the null hypothesis about equality of α to its true value: the rejection rates exceed 0.2 for $T = 5$ and 0.3 for $T = 10$. However, the rejection rate is close to 0.1 in case of the instrumental variable quantile regression estimator with clustered standard errors. Note that for each τ and each combination of the values of N and T, the rejection rate is higher for the estimator with standard errors based on i.i.d assumption than for the estimator with clustered standard errors. This fact points to underestimation of standard errors under application of Chernozhukov and Hansen (2006) resampling to the data generating-process with positive correlation of errors within clusters. Note that the asymptotic theory for the instrumental variable quantile regression estimator for clustered data requires a fairly large number of clusters. The results in Table 1 are in line with this requirement: the rejection rates of the null hypothesis are slightly overstated at 0.13–0.15 when N is 100 or 200.

Table 2 gives results concerning the basic inference about the vector of coefficients for exogenous covariates. Similarly to basic inference about α, the rejection rate of the null hypothesis of equality of the vector β_1 to its true value is excessively high with standard errors for the Chernozhukov and Hansen (2006) instrumental variable quantile regression estimator. But the rejection rate is close to 0.1 in case of the estimator with clustered standard errors.

The performance of the i.i.d. resampling procedure for implementation of the general inference is contrasted with resampling of clusters in Table 3. The results point to wrong conclusions that may be associated with using the Chernozhukov and Hansen (2006) resampling procedure where there is within-cluster correlation of errors. For each of the process tests the rejection rate of the null hypothesis is much higher in case of bootstrap under the i.i.d. assumption than in case of bootstrap based on clustering. While the i.i.d. resampling yields rejection rates which well exceed the significance level of the test, the resampling of clusters leads to better results, which confirms the implications of our Theorem 3.

The results of the process tests with clustered bootstrap show that rejection rates of null hypotheses are close to 0.1 (Table 3).

The probabilities of rejecting the null hypothesis under the alternative at the 10% level of significance of the test are shown in Table 4. The rejection rates exceed 0.1, rise with increase in the size of sample NT, and become close to 1 with large N.

Table 1: Simulated true sizes for basic inference tests at 10% level (H_0: $\alpha(\tau) = 1 \cdot \tau$, where $1 \cdot \tau$ is the true value of $\alpha(\tau)$)

		standard errors under i.i.d.			clustered standard errors		
T	N	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
2	100	0.082	0.128	0.118	0.066	0.096	0.096
	200	0.118	0.110	0.160	0.080	0.080	0.118
	500	0.150	0.116	0.154	0.112	0.086	0.112
	1000	0.150	0.132	0.136	0.106	0.088	0.098
	2000	0.148	0.160	0.138	0.108	0.102	0.100
5	100	0.230	0.294	0.258	0.084	0.122	0.128
	200	0.280	0.308	0.296	0.102	0.102	0.138
	500	0.282	0.248	0.268	0.098	0.092	0.110
	1000	0.238	0.278	0.268	0.088	0.092	0.102
	2000	0.244	0.274	0.276	0.104	0.102	0.116
10	100	0.376	0.400	0.436	0.108	0.124	0.154
	200	0.402	0.436	0.434	0.132	0.108	0.138
	500	0.426	0.446	0.388	0.110	0.098	0.098
	1000	0.358	0.412	0.396	0.098	0.094	0.102
	2000	0.410	0.418	0.384	0.110	0.096	0.110

Table 2: Simulated true sizes for basic inference tests at 10% level (H_0: $\beta_1(\tau) = 2 \cdot \tau$, where $2 \cdot \tau$ is the true value of $\beta_1(\tau)$)

		standard errors under i.i.d.			clustered standard errors		
T	N	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
2	100	0.094	0.104	0.128	0.074	0.070	0.082
	200	0.118	0.120	0.112	0.070	0.078	0.080
	500	0.128	0.116	0.126	0.080	0.066	0.082
	1000	0.134	0.144	0.134	0.086	0.084	0.084
	2000	0.142	0.170	0.154	0.100	0.110	0.098
5	100	0.216	0.246	0.246	0.094	0.094	0.094
	200	0.272	0.254	0.268	0.100	0.088	0.092
	500	0.234	0.252	0.248	0.090	0.088	0.084
	1000	0.222	0.272	0.272	0.094	0.096	0.100
	2000	0.254	0.280	0.240	0.094	0.100	0.076
10	100	0.360	0.394	0.370	0.092	0.094	0.082
	200	0.356	0.386	0.396	0.088	0.092	0.072
	500	0.368	0.390	0.370	0.098	0.082	0.108
	1000	0.354	0.402	0.412	0.096	0.096	0.116
	2000	0.376	0.408	0.366	0.090	0.084	0.076

Table 3: Simulated true sizes for process inference tests for $\tau \in \{0.25, 0.5, 0.75\}$ at 10% level under null hypotheses

		i.i.d. bootstrap			clustered bootstrap		
T	N	$\alpha(\tau) \equiv 0$	$\alpha(\tau) \equiv$ const	D is exog.	$\alpha(\tau) \equiv 0$	$\alpha(\tau) \equiv$ const	D is exog.
	100	0.144	0.058	0.110	0.124	0.054	0.090
	200	0.158	0.088	0.106	0.120	0.076	0.084
2	500	0.170	0.110	0.100	0.112	0.080	0.070
	1000	0.174	0.140	0.088	0.096	0.104	0.054
	2000	0.208	0.152	0.100	0.116	0.122	0.042
	100	0.324	0.188	0.280	0.170	0.122	0.112
	200	0.376	0.236	0.250	0.164	0.132	0.096
5	500	0.370	0.216	0.242	0.124	0.106	0.048
	1000	0.348	0.242	0.218	0.122	0.112	0.052
	2000	0.338	0.222	0.220	0.136	0.114	0.058
	100	0.550	0.322	0.422	0.188	0.156	0.138
	200	0.584	0.374	0.472	0.182	0.182	0.102
10	500	0.570	0.382	0.384	0.138	0.138	0.068
	1000	0.526	0.374	0.374	0.118	0.122	0.048
	2000	0.558	0.358	0.382	0.136	0.132	0.062

Notes: No effect and constant effect hypotheses are tested in the model with $\alpha(\tau) \equiv 0$ and endogeneous D. Exogeneity of D is tested in a model with $\alpha(\tau) = 1 \cdot \tau$ and exogeneous D.

Table 4: Simulated true sizes for process inference tests for $\tau \in \{0.25, 0.5, 0.75\}$ at 10% level under alternative hypotheses

		clustered bootstrap		
T	N	$\alpha(\tau) \equiv 0$	$\alpha(\tau) \equiv$ const	D is exog.
	100	0.458	0.132	0.908
	200	0.530	0.190	0.996
2	500	0.770	0.354	1.000
	1000	0.938	0.538	1.000
	2000	0.998	0.842	1.000
	100	0.554	0.236	0.998
	200	0.658	0.326	1.000
5	500	0.890	0.530	1.000
	1000	0.984	0.792	1.000
	2000	1.000	0.966	1.000
	100	0.584	0.286	0.998
	200	0.702	0.394	1.000
10	500	0.920	0.662	1.000
	1000	0.996	0.876	1.000
	2000	1.000	0.996	1.000

Notes: All three hypotheses are tested in the model with $\alpha(\tau) = 1 \cdot \tau$ and endogeneous D.

3 The impact of female labor participation on the wages of men and women in the United States in 1940–1950

Female participation in the labor market has attracted the interest of labor economists since the 1930s and early reviews on the subject appeared in the 1970s–1980s (Mincer, 1962, Heckman, 1978, Killingsworth and Heckman, 1986, Psacharopoulos and Tzannatos, 1989). The amount of female labor was shown to affect the wages of both men and women (Juhn and Kim, 1999, Cain and Dooley, 1976). In particular, increasing presence of married women in the labor force caused a decline, on average, in the women's wages (Cain and Dooley, 1976).

It should be noted that the amount of labor is likely to be endogenous in the wage equation. Endogeneity occurs because both the amount of labor and wages are found as a solution of the system of demand and supply equations. A pioneering work Acemoglu and Autor (2011) uses an instrumental variable approach to account for endogeneity in the female labor force in time of war, estimating least-squares models and using mobilization rate in World War II to capture state-level exogenous variation in female labor supply in the US. The same instrument was employed in subsequent papers which focused on groups of women by their fertility, ethnicity, marital status (see review in (Rose, 2018)) or studied the effect in the longer run (e.g. employment by women in 1960, assessed in Fernández et al. (2004)). Note that a related approach is used in Boehnke and Gay (2020) for World War I where the instrument for female labor force is the military fatality rate.

In the present paper we extend the analysis for the case when the effect of labor supply on earnings is heterogeneous. Indeed, it is plausible to assume that the effect differs across high-wage and low-wage workers. Specifically, shortage of highly-skilled labor causes increase of wages in that segment, which do not decline despite subsequent increase of supply of highly skilled labor (see evidence for France, Germany, Austria, the US and the UK in respectively, (Abowd et al., 1999, Andrews et al., 2012, Borovičková and Shimer, 2017, van Reenen, 2011)).

We employ a quantile regression approach with endogeneity to estimate the analogue of the least-squares wage equation of Acemoglu et al. (2004). Quantile regression is a technique, widely used by labor economists to capture heterogeneous impact of the explanatory variables on the tails of conditional distribution of the dependent variable. Early applications of quantile regression models under exogeneity in the analysis of wages include Abadie (1997) and Buchinsky (1998).

We use the data of Acemoglu et al. (2004) available at MIT Economics: David Autor's Data Archive `http://economics.mit.edu/faculty/dautor/data/autacemly06`.[7] The data consist of one-percent random draws from the 1940 and 1950 censuses. For each census, the Acemoglu et al. (2004) analysis of the wage equation uses samples with white individuals aged 14–64, who were not self-employed or employed in farming, did not reside in prisons or barracks, and received wages and salaries (the range of hourly earnings in the previous year is 0.5–250 in 1990 US dollars). Total number of observations in the pooled data with samples from the 1940 and 1950 censuses numbered 198,385 men and 69,335 women. Census sampling weights are used in all estimations.

Following the logic of the two-stage least squares models for men and women, given in equation (10) of Acemoglu et al. (2004), we estimate their quantile regression analogues as

[7] Specifically, we use the data and variables for Table 9 on pp.534–535 in Acemoglu et al. (2004): `http://economics.mit.edu/~dautor/www/table5-9-10-11-12-a1_dta.zip` and `http://economics.mit.edu/~dautor/www/table9_do_log.zip`

follows:

$$\ln w_{ist} = D'_{st}\alpha(u_{ist}) + X'_{ist}\beta(u_{ist}), \tag{9}$$

$$D'_{st} = \delta(X_{ist}, Z_{st}, \nu_{ist}), \tag{10}$$

$$\tau \mapsto D_{st}\alpha(\tau) + X'_{ist}\beta(\tau) \quad \text{is monotonically increasing,} \tag{11}$$

where τ denotes the value of a given quantile for conditional distribution of the log of weekly earnings – $\ln w_{ist}$ – for individual i at state s in period t (1940 or 1950), the endogenous variable D_{st} is female labor supply (average weeks worked per woman in state s in year t), $X_{ist} = [X_{1,ist}, X_{2,ist}, X_{3,st}, X_{4,st}]$ is a vector of exogenous variables: $X_{1,ist}$ includes state of residence of the individual, years of education, marital status, WWII veteran dummy (for men), a quartic in potential experience; $X_{2,ist}$ is state/country of birth, $X_{3,st}$ is state-level female age structure, $X_{4,st}$ includes share of farmers and nonwhites, and average schooling structure in the state in 1940, all exogenous variables are interacted with the 1950 year dummy to account for the pooled structure of data, Z_{st} – an interaction between the 1950 dummy and mobilization rate in state s – is an instrument for female labor supply, ν_{ist} is statistically dependent on u_{ist}, $u_{ist} \perp (X_{ist}, Z_{st}) \sim U[0, 1]$.

The wage equation in Acemoglu et al. (2004) is estimated for each individual but the endogenous variable – female labor supply – is measured as the state average in a given year. So the data become clustered at the state-year level and the two-stage least squares models account for clustered standard errors.

In our quantile regression analysis we similarly account for clustered data. Specifically, we contrast the standard errors for the coefficient for female labor supply estimated in (9)–(11) under the assumption of standard errors based om i.i.d. assumption and under standard errors, clustered at state-year level.

We use 19 values of $\tau \in [0.05, 0.95]$, starting with $\tau = 0.05$ at the 0.05 step. For the purpose of comparison we present the coefficient for female labor supply, estimated in conventional quantile regression, along with the findings in quantile regressions under endogeneity. It should be noted that the coefficient under the latter approach is mainly insignificant.

The first set of our results describes the impact of female labor supply on women's earnings. As is shown in Figures 1–2, the coefficient for the amount of weeks worked per woman is negative in explaining the log of earnings by women. With an increase in quantile index, the coefficient becomes smaller in absolute terms. So the effect is weaker for higher-wage workers. The standard errors for the coefficient estimated under the Chernozhukov and Hansen (2006) approach are 2–3 times smaller than the standard errors of the estimator for clustered data (Tables C.1–C.2 of the Appendix).[8] This fact implies a positive correlation of errors within clusters.

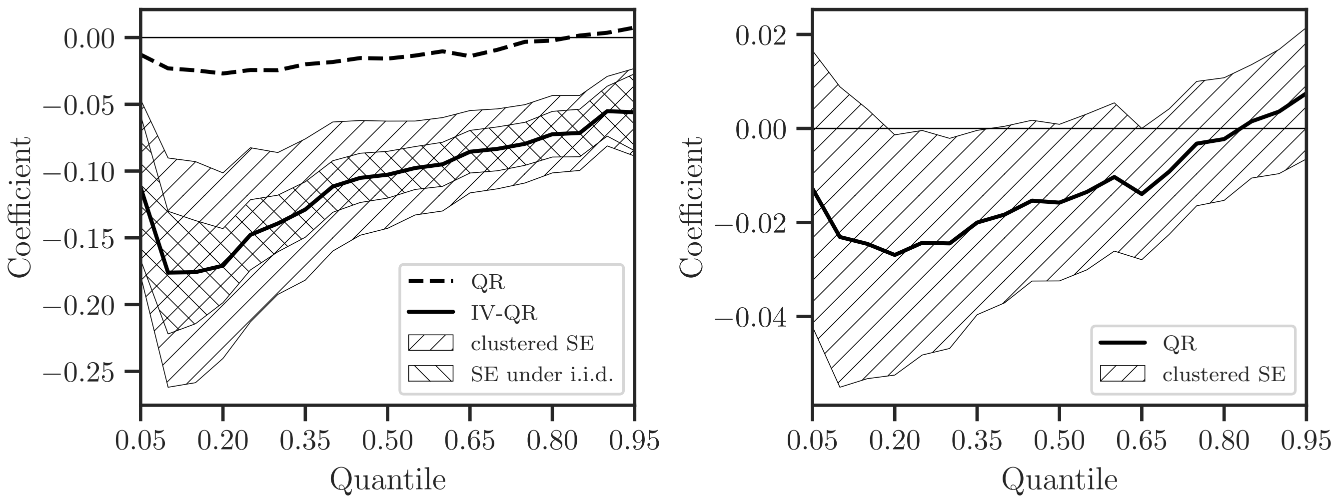
Impact of female labor on earnings is significant in the specification with $X_{1,ist}$, $X_{2,ist}$ and $X_{3,st}$ as controls both in case of the standard errors based on i.i.d. assumption and under the standard errors clustered at the state-year level (Figure 1). However, when additional state-level controls $X_{4,st}$ are included in the list of explanatory variables, the estimators with and without clusters yield different results related to basic inference. The Chernozhukov and Hansen (2006) estimator implies that the effect of female labor is significant in regressions with quantile indices from 0.05 to 0.85, while the estimator for clustered data gives a much smaller range of quantile indices with significant effect: from 0.05 to 0.55 (Figure 2). In other words, failure to account for clustered standard errors leads to wrong conclusions for high-wage workers.

The second set of results deals with the effect of female labor supply on men's earnings. The findings are similar to those for female earnings: the coefficient for female labor supply is

[8] For the sake of brevity, Tables C.1–C.4 show the estimated coefficients and their standard errors for $\tau \in [0.1, 0.9]$, with the 0.1–step.

negative in explaining men's earnings, and is inversely related to quantile index (Figures 3–4). In other words, the negative effect is smaller in absolute terms for higher-wage male workers. The standard errors for the coefficient are downwards biased under the Chernozhukov and Hansen (2006) approach in comparison with the estimator for clustered standard errors (Tables C.3–C.4 of the Appendix). Underestimation of standard errors owing to neglect of clusters does not affect basic inference in specification with fewer state-level controls (Figure 3), but becomes important when more controls are added to the list of covariates (Figure 4).

Finally, we carry out inference on the entire instrumental variable quantile regression process in the models with the richest list of exogenous controls (X_1 through X_4). We use resampling of test statistics based on clusters. The results, which are reported in Table 5, show that female labor supply affects the earnings of both men and women (the hypothesis of no effect is rejected at the 5% level). Female labor supply is endogenous for both sexes and the effect of the variable differs across quantiles (each of the hypotheses of exogeneity and of constant effect is rejected at the 5% level).



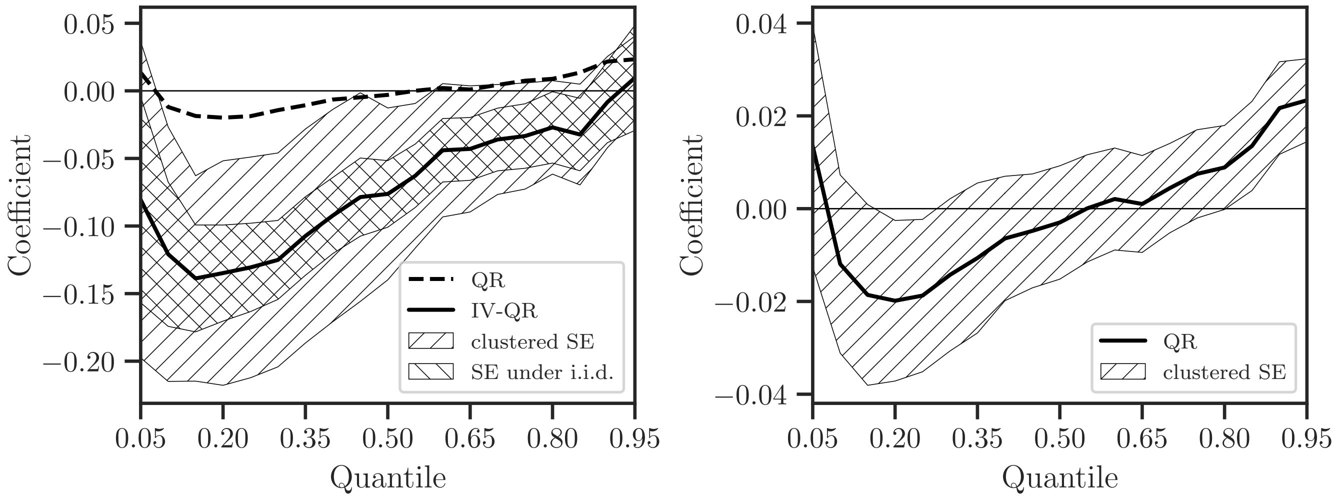
(a) 90% confidence intervals for IV-QR estimates.

(b) 90% confidence intervals for QR estimates

Figure 1: Impact of female labor supply on earnings of women, 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$ and $X_{3,st}$ as controls

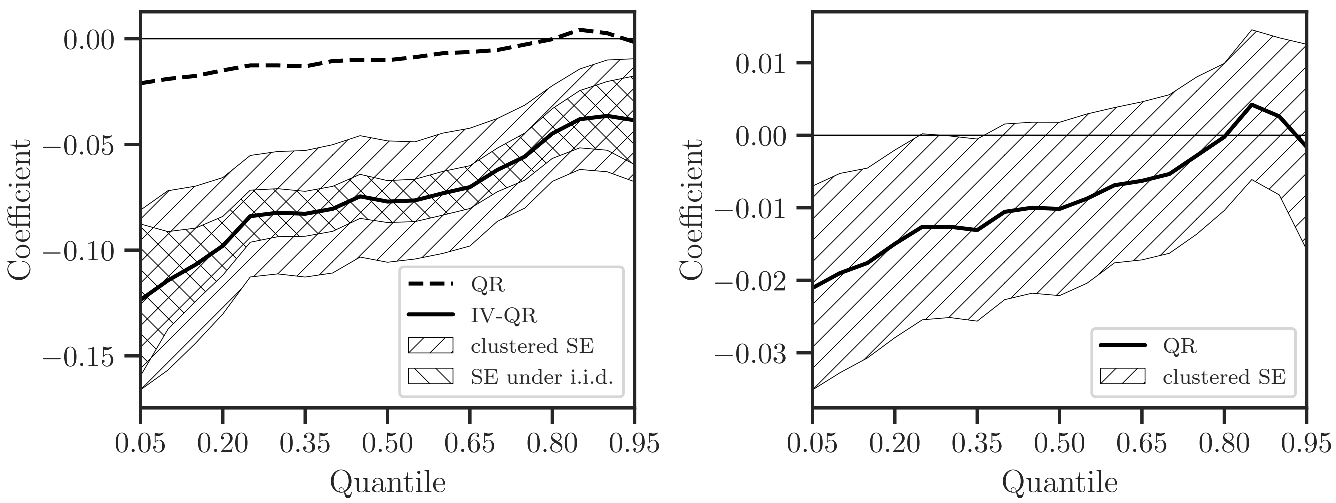
Table 5: Results of tests based on the instrumental variable quantile regression process (resampling by clusters)

	Impact of female labor supply on female earnings		Impact of female labor supply on male earnings	
Null hypothesis	KS statistic	P-value	KS statistic	P-value
No effect ($\alpha(\cdot) \equiv 0$)	7.047	0.002	3.953	0.026
Constant effect ($\alpha(\cdot) \equiv$ const)	3.562	0.041	3.139	0.001
Exogeneity (D is exogeneous)	6.292	0.000	3.429	0.011



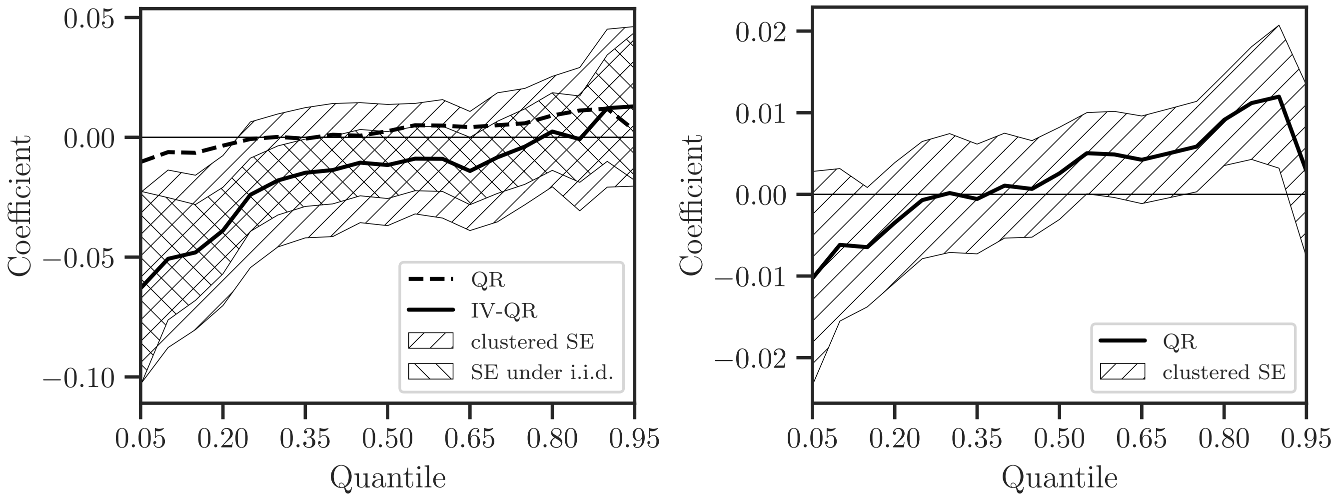
(a) 90% confidence intervals for IV-QR estimates (b) 90% confidence intervals for QR estimates

Figure 2: Impact of female labor supply on earnings of women, 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$, $X_{3,st}$ and $X_{4,st}$ as controls



(a) 90% confidence intervals for IV-QR estimates (b) 90% confidence intervals for QR estimates

Figure 3: Impact of female labor supply on earnings of men, 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$ and $X_{3,st}$ as controls



(a) 90% confidence intervals for IV-QR estimates

(b) 90% confidence intervals for QR estimates

Figure 4: Impact of female labor supply on earnings of men, 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$, $X_{3,st}$ and $X_{4,st}$ as controls

4 Conclusion

The present paper deals with robust inference in conditional quantile regression model with endogenous covariates and within-cluster correlation of error terms. We show that the widely used Chernozhukov and Hansen (2006) instrumental variable quantile regression estimator is consistent and asymptotically normal when applied to the data-generating process with clustered data. We derive the asymptotic distribution of the instrumental variable quantile regression estimator for clustered data, and the consistent estimator of the covariance matrix enables basic inference where there is intra-cluster correlation. As regards inference based on the entire instrumental variable quantile regression process, we extend the approach of Chernozhukov and Hansen (2006) and prove that resampling by clusters offers an approach to implementation of asymptotic tests.

Our theoretical results are supported by simulation analysis, where we compare the asymptotic behavior of the instrumental variable quantile regression estimator for clustered data and the behavior of the Chernozhukov and Hansen (2006) estimator. The empirical illustration of the instrumental variable quantile regression estimator under clustered standard errors uses the data from Acemoglu et al. (2004). We quantify the quantile regression analogues of the two-stage least squares models for wage equations for men and women, and data are clustered at state-year level. The results demonstrate that failure to incorporate the clustered structure of data leads to wrong conclusions about the effect of female labor supply on the earnings of high-wage male and female workers.

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the contents of the paper.

References

Abadie, A. (1997). Changes in Spanish labor income structure during the 1980's: A quantile regression approach. *Investigaciones Economicas*, 21(2):253–272.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. *Handbook of Labor Economics*, 4:1043–1171.

Acemoglu, D., Autor, D. H., and Lyle, D. (2004). Women, war, and wages: The effect of female labor supply on the wage structure at midcentury. *Journal of Political Economy*, 112(3):497–551.

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2012). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters*, 117(3):824–827.

Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. *Econometrica*, 74(2):539–563.

Boehnke, J. and Gay, V. (2020). The missing men World War I and female labor force participation. *Journal of Human Resources*, pages 0419–10151R1.

Borovičková, K. and Shimer, R. (2017). High wage workers work for high wage firms. Technical Report 24074, National Bureau of Economic Research.

Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, 13(1):1–30.

Cain, G. G. and Dooley, M. D. (1976). Estimation of a model of labor supply, fertility, and wages of married women. *Journal of Political Economy*, 84(4, Part 2):S179–S199.

Cameron, A. C. and Miller, D. L. (2011). Robust inference with clustered data. In Ullah, A. and Giles, D. E., editors, *Handbook of Empirical Economics and Finance*, pages 1–28. Chapman and Hall/CRC.

Chernozhukov, V. and Hansen, C. (2004). Instrumental variable quantile regression. Working Paper, available at http://www.mit.edu/∼vchern.

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

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.

Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398.

Chernozhukov, V. and Hong, H. (2002). Three-step censored quantile regression and extramarital affairs. *Journal of the American Statistical Association*, 97(459):872–882.

Fernández, R., Fogli, A., and Olivetti, C. (2004). Mothers and sons: Preference formation and female labor force dynamics. *The Quarterly Journal of Economics*, 119(4):1249–1299.

Hagemann, A. (2017). Cluster-robust bootstrap inference in quantile regression models. *Journal of the American Statistical Association*, 112(517):446–456.

Heckman, J. J. (1978). A partial survey of recent research on the labor supply of women. *The American Economic Review*, 68(2):200–207.

Juhn, C. and Kim, D. I. (1999). The effects of rising female labor supply on male wages. *Journal of Labor Economics*, 17(1):23–48.

Killingsworth, M. R. and Heckman, J. (1986). Female labor supply: A survey. In Ashenfelter, O. and Layard, R., editors, *Handbook of Labor Economics*, volume 1, pages 103–204. North Holland Amsterdam.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, 46:33–50.

Mincer, J. (1962). Labor force participation of married women: A study of labor supply. In *Aspects of labor economics*, pages 63–105. Princeton University Press.

Parente, P. M. and Santos Silva, J. (2016). Quantile regression with clustered data. *Journal of Econometric Methods*, 5(1):1–15.

Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32(1):143–155.

Psacharopoulos, G. and Tzannatos, Z. (1989). Female labor force participation: An international perspective. *The World Bank Research Observer*, 4(2):187–201.

Rose, E. K. (2018). The rise and fall of female labor force participation during World War II in the United States. *The Journal of Economic History*, 78(3):673–711.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag Press.

van Reenen, J. (2011). Wage inequality, technology and trade: 21st century evidence. *Labour Economics*, 18(6):730–741.

Appendix A Theorems from Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2006)

Theorem 1 in Chernozhukov and Hansen (2005) (P. 249. Main implication)**.** *Suppose conditions A1–A5 hold, then for any $\tau \in (0,1)$, a.s.*

$$P(Y \leq q(D,X,\tau)|X,Z) = P(Y < q(D,X,\tau)|X,Z) = \tau, \tag{A.1}$$

and U_D is independent of Z and X.

Theorem 4 in Chernozhukov and Hansen (2006) (P. 506. Inference)**.** *For f denoting the two- and one-sided KS or CM statistics:*

1. *Under Assumption 1, I1(a), and I2 $S_N \xrightarrow{d} S \equiv f(v(\cdot))$, where $v(\cdot) = R(\cdot)(b(\cdot) - d(\cdot))$. If $v(\cdot)$ has nondegenerate covariance kernel, then for $\alpha < 1/2$, $P(S_N > c(1-\alpha)) \to \alpha = P(f(v(\cdot)) > c(1-\alpha))$, where $c(1-\alpha)$ is chosen so that $P(f(v(\cdot)) > c(1-\alpha)) = \alpha$.*
2. *Under Assumption 1 and I1(b) $S_N \xrightarrow{d} \infty$, and $P(S_N > c(1-\alpha)) \to 1$.*

Appendix B Proofs

Define for $W_t = (Y_t, D_t, X_t, Z_t)$, $\vartheta \equiv (\beta, \gamma)$ and $\varphi_\tau(u) = I(u < 0) - \tau$

$$\hat{f}(W_t, \alpha, \vartheta, \tau) \equiv \varphi_\tau(Y_t - D'\alpha - X'\beta - \hat{\Phi}_t(\tau)'\gamma)\hat{\Psi}_t(\tau),$$
$$f(W_t, \alpha, \vartheta, \tau) \equiv \varphi_\tau(Y_t - D'\alpha - X'\beta - \Phi_t(\tau)'\gamma)\Psi_t(\tau),$$

$\Psi_t(\tau) \equiv V_t(\tau) \cdot [\Phi_t(\tau)', X_t']'$, $\Phi_t(\tau) = \Phi_t(\tau, X_t, Z_t)$, $V_t(\tau) = V_t(\tau, X_t, Z_t)$, $\hat{\Psi}_t(\tau) \equiv \hat{V}_t(\tau) \cdot [\hat{\Phi}_t(\tau)', X_t']'$, $\hat{\Phi}_t(\tau) = \hat{\Phi}_t(\tau, X_t, Z_t)$, $\hat{V}_t(\tau) = \hat{V}_t(\tau, X_t, Z_t)$; for $\rho_\tau(u) = (\tau - I(u<0))u = -\varphi_\tau(u)u$

$$\hat{g}(W_t, \alpha, \vartheta, \tau) \equiv \rho_\tau(Y_t - D'\alpha - X'\beta - \hat{\Phi}_t(\tau)'\gamma)\hat{V}_t(\tau),$$
$$g(W_t, \alpha, \vartheta, \tau) \equiv \rho_\tau(Y_t - D'\alpha - X'\beta - \Phi_t(\tau)'\gamma)V_t(\tau).$$

Denote $\mathbb{E}_N(\xi) \equiv \frac{1}{N}\sum_{i=1}^N \xi_i$, $\mathbb{G}_N(\xi) \equiv \sqrt{N}(\mathbb{E}_N(\xi) - E(\xi))$. Define

$$Q_N(\alpha, \vartheta, \tau) \equiv \mathbb{E}_N\left(\sum_{t=1}^T \hat{g}(W_t, \alpha, \vartheta, \tau)\right) = \frac{1}{N}\sum_{i=1}^N \sum_{t=1}^T \hat{g}(W_{it}, \alpha, \vartheta, \tau),$$
$$Q(\alpha, \vartheta, \tau) \equiv E\left(\sum_{t=1}^T g(W_t, \alpha, \vartheta, \tau)\right),$$

and

$$\hat{\vartheta}(\alpha, \tau) \equiv (\hat{\beta}(\alpha,\tau)', \hat{\gamma}(\alpha,\tau)') = \operatorname*{argmin}_{\vartheta \in \mathcal{B}\times\mathcal{G}} Q_N(\alpha, \vartheta, \tau),$$
$$\vartheta(\alpha, \tau) \equiv (\beta(\alpha,\tau)', \gamma(\alpha,\tau)') = \operatorname*{argmin}_{\vartheta \in \mathcal{B}\times\mathcal{G}} Q(\alpha, \vartheta, \tau),$$
$$\hat{\alpha}(\tau) = \operatorname*{argmin}_{\alpha \in \mathcal{A}} \|\hat{\gamma}(\alpha,\tau)\|^2_{A(\tau)},$$
$$\alpha(\tau) = \operatorname*{argmin}_{\alpha \in \mathcal{A}} \|\gamma(\alpha,\tau)\|^2_{A(\tau)},$$
$$\hat{\vartheta}(\tau) \equiv (\hat{\beta}(\tau)', \hat{\gamma}(\tau)') \equiv \hat{\vartheta}(\hat{\alpha}(\tau), \tau),$$
$$\vartheta(\tau) \equiv (\beta(\tau)', 0) \equiv \vartheta(\alpha(\tau), \tau).$$

Proof of Theorem 1. See proof of Theorem 2 in Chernozhukov and Hansen (2006), P.514. □

Proof of Theorem 2. Step 1 (Identification): See step 1 in the proof of Theorem 3 in Chernozhukov and Hansen (2006), P.514–516.

Step 2 (Consistency): Apply step 2 in the proof of Theorem 3 in Chernozhukov and Hansen (2006), P.516 to $Q_N(\alpha, \vartheta, \tau) \equiv \mathbb{E}_N \left(\sum_{t=1}^T \hat{g}(W_t, \alpha, \vartheta, \tau) \right) = \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T \hat{g}(W_{it}, \alpha, \vartheta, \tau)$.

Step3 (Asymptotics):

We adapt Step 3 in the proof of Theorem 3 in Chernozhukov and Hansen (2006), P.516–518. Consider a set of closed balls $B_{\delta_N}(\alpha(\tau))$ centered at $\alpha(\tau)$ for each τ. The radius of balls δ_N is independent of τ and $\delta_N \to 0$ slowly enough. Denote by $\alpha_N(\tau)$ any value inside $B_{\delta_N}(\alpha(\tau))$. Following the argument in Theorem 3.3 in Koenker and Bassett (1978),

$$O(1/\sqrt{N}) = \sqrt{N}\mathbb{E}_N \left(\sum_{t=1}^T \hat{f}(W_t, \alpha_N(\cdot), \hat{\vartheta}(\alpha_N(\cdot), \cdot), \cdot) \right).$$

By Lemma B.1, expand the right hand side for any $\sup_{\tau \in \mathcal{T}} \|\alpha_N(\tau) - \alpha(\tau)\| \xrightarrow{p} 0$:

$$\begin{aligned}
O(1/\sqrt{N}) &= \sqrt{N}\mathbb{E}_N \left(\sum_{t=1}^T \hat{f}(W_t, \alpha_N(\cdot), \hat{\vartheta}(\alpha_N(\cdot), \cdot), \cdot) \right) \\
&= \sqrt{N} E \left(\sum_{t=1}^T \hat{f}(W_t, \alpha_N(\cdot), \hat{\vartheta}(\alpha_N(\cdot), \cdot), \cdot) \right) + \mathbb{G}_N \left(\sum_{t=1}^T \hat{f}(W_t, \alpha_N(\cdot), \hat{\vartheta}(\alpha_N(\cdot), \cdot), \cdot) \right) \\
&= \sqrt{N} E \left(\sum_{t=1}^T \hat{f}(W_t, \alpha_N(\cdot), \hat{\vartheta}(\alpha_N(\cdot), \cdot), \cdot) \right) + \mathbb{G}_N \left(\sum_{t=1}^T f(W_t, \alpha(\cdot), \vartheta(\alpha(\cdot), \cdot), \cdot) \right) \\
&\quad + o_p(1) \qquad \text{in } \ell^\infty(\mathcal{T}).
\end{aligned}$$

Further expansion of the last line yields

$$\begin{aligned}
O(1/\sqrt{N}) &= (J_\vartheta(\cdot) + o_p(1))\sqrt{N}(\hat{\vartheta}(\alpha_N(\cdot), \cdot) - \vartheta(\cdot)) + (J_\alpha(\cdot) + o_p(1))\sqrt{N}(\alpha_N(\cdot) - \alpha(\cdot)) \\
&\quad + \mathbb{G}_N \left(\sum_{t=1}^T f(W_t, \alpha(\cdot), \vartheta(\alpha(\cdot), \cdot), \cdot) \right) + o_p(1) \qquad \text{in } \ell^\infty(\mathcal{T}),
\end{aligned} \tag{B.1}$$

where

$$\begin{aligned}
J_\alpha(\cdot) &= \frac{\partial}{\partial \alpha'} E \left(\sum_{t=1}^T \varphi_\cdot(Y_t - D_t'\alpha - X_t'\beta(\cdot))\Psi_t(\cdot) \right) \Bigg|_{\alpha=\alpha(\cdot)} = E \left(\sum_{t=1}^T f_{\varepsilon_t(\tau)}(0|X_t, D_t, Z_t)\Psi_t(\tau)D_t' \right), \\
J_\vartheta(\cdot) &= [J_\beta(\cdot), J_\gamma(\cdot)], \\
J_\beta(\cdot) &= \frac{\partial}{\partial \beta'} E \left(\sum_{t=1}^T \varphi_\cdot(Y_t - D_t'\alpha(\cdot) - X_t'\beta)\Psi_t(\cdot) \right) \Bigg|_{\beta=\beta(\cdot)} = E \left(\sum_{t=1}^T f_{\varepsilon_t(\tau)}(0|X_t, D_t, Z_t)\Psi_t(\tau)X_t' \right), \\
J_\gamma(\cdot) &= \frac{\partial}{\partial \gamma'} E \left(\sum_{t=1}^T \varphi_\cdot(Y_t - D_t'\alpha(\cdot) - X_t'\beta(\cdot) - \Phi_t(\cdot)'\gamma)\Psi_t(\cdot) \right) \Bigg|_{\gamma=0}.
\end{aligned}$$

So for any $\sup_{\tau \in \mathcal{T}} \|\alpha_N(\tau) - \alpha(\tau)\| \xrightarrow{p} 0$:

$$\begin{aligned}
\sqrt{N}(\hat{\vartheta}(\alpha_N(\cdot), \cdot) - \vartheta(\cdot)) &= -J_\vartheta^{-1}(\cdot)\mathbb{G}_N \left(\sum_{t=1}^T f(W_t, \alpha(\cdot), \vartheta(\alpha(\cdot), \cdot), \cdot) \right) \\
&\quad - J_\vartheta(\cdot)J_\alpha(\cdot)(1 + o_p(1))\sqrt{N}(\alpha_N(\cdot) - \alpha(\cdot)) + o_p(1) \qquad \text{in } \ell^\infty(\mathcal{T}),
\end{aligned}$$

and specifically

$$\sqrt{N}(\hat{\gamma}(\alpha_N(\cdot),\cdot)-0) = -\bar{J}_\gamma^{-1}(\cdot)\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right)$$
$$- \bar{J}_\gamma(\cdot)J_\alpha(\cdot)(1+o_p(1))\sqrt{N}(\alpha_N(\cdot)-\alpha(\cdot)) + o_p(1) \qquad \text{in } \ell^\infty(\mathcal{T}),$$

where $[\bar{J}_\beta(\cdot)', \bar{J}_\gamma(\cdot)']'$ is the conformable partition of $J_\vartheta^{-1}(\cdot)$.

According to Step 2 $wp \to 1$

$$\hat{\alpha}(\tau) = \operatorname*{arginf}_{\alpha_N(\tau)\in B_{\delta_N}(\alpha(\tau))} \|\hat{\gamma}(\alpha_N(\tau),\tau)\|^2_{A(\tau)} \qquad \text{for all } \tau \in \mathcal{T}.$$

By Lemma B.1, $\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) = O_p(1)$, and therefore

$$\sqrt{N}\|\hat{\gamma}(\alpha_N(\cdot),\cdot)\|_{A(\cdot)} = \|O_p(1) - \bar{J}_\gamma(\cdot)J_\alpha(\cdot)(1+o_p(1))\sqrt{N}(\alpha_N(\cdot)-\alpha(\cdot))\|_{A(\cdot)} \quad \text{in } \ell^\infty(\mathcal{T}).$$

By R3 $\bar{J}_\gamma(\tau)J_\alpha(\tau)$ and $A(\tau)$ have full rank uniformly in τ, so $\sqrt{N}(\hat{\alpha}(\cdot)-\alpha(\cdot)) = O_p(1)$ in $\ell^\infty(\mathcal{T})$. Expanding the argument used in the proof of Lemma D.1 in Chernozhukov and Hansen (2006), P.520,

$$\sqrt{N}(\hat{\alpha}(\cdot)-\alpha(\cdot)) = \operatorname*{arginf}_{\mu\in\mathbb{R}^l}\left\| -\bar{J}_\gamma(\cdot)\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) - \bar{J}_\gamma(\cdot)J_\alpha(\cdot)\mu\right\|^2_{A(\cdot)}$$
$$+ o_p(1) \quad \text{in } \ell^\infty(\mathcal{T}).$$

Accordingly, in $\ell^\infty(\mathcal{T})$ jointly

$$\sqrt{N}(\hat{\alpha}(\cdot)-\alpha(\cdot)) = -(J_\alpha(\cdot)'\bar{J}_\gamma(\cdot)'A(\cdot)\bar{J}_\gamma(\cdot)J_\alpha(\cdot))^{-1}J_\alpha(\cdot)'\bar{J}_\gamma(\cdot)'A(\cdot)\bar{J}_\gamma(\cdot)$$
$$\times\,\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) + o_p(1) = O_p(1),$$
$$\sqrt{N}(\hat{\vartheta}(\hat{\alpha}(\cdot),\cdot)-\vartheta(\cdot)) = -J_\vartheta^{-1}(\cdot)\big(I - J_\alpha(\cdot)(J_\alpha(\cdot)'\bar{J}_\gamma(\cdot)'A(\cdot)\bar{J}_\gamma(\cdot)J_\alpha(\cdot))^{-1}J_\alpha(\cdot)'\bar{J}_\gamma(\cdot)'A(\cdot)\bar{J}_\gamma(\cdot)\big)$$
$$\times\,\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) + o_p(1) = O_p(1).$$

Owing to invertibility of $\bar{J}_\gamma(\tau)J_\alpha(\tau)$

$$\sqrt{N}(\hat{\gamma}(\hat{\alpha}(\cdot),\cdot)-0) = -\bar{J}_\gamma(\cdot)\big(I - J_\alpha(\cdot)(\bar{J}_\gamma(\cdot)J_\alpha(\cdot))^{-1}\bar{J}_\gamma(\cdot)\big)$$
$$\times\,\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) + o_p(1) = 0\cdot O_p(1) + o_p(1) \quad \text{in } \ell^\infty(\mathcal{T}).$$

Substitute the above expression for $\sqrt{N}(\hat{\gamma}(\hat{\alpha}(\cdot),\cdot)-0)$ and the expression for $(\alpha_N(\cdot),\hat{\vartheta}(\alpha_N(\cdot),\cdot)) = (\hat{\alpha}(\cdot),\hat{\vartheta}(\cdot)) = (\hat{\alpha}(\cdot),\hat{\beta}(\cdot),o_p(1/\sqrt{N}))$ into expansion (B.1):

$$-\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) = J(\cdot)\sqrt{N}\begin{pmatrix}\hat{\alpha}(\cdot)-\alpha(\cdot)\\ \hat{\beta}(\cdot)-\beta(\cdot)\end{pmatrix} + o_p(1) \quad \text{in } \ell^\infty(\mathcal{T}),$$

where $J(\cdot) = [J_\alpha(\cdot)', J_\beta(\cdot)']'$. By Lemma B.1, $\mathbb{G}_N\left(\sum_{t=1}^{T} f(W_t,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right) \Rightarrow \mathbb{G}(\cdot)$ in $\ell^\infty(\mathcal{T})$, where $\mathbb{G}(\cdot)$ is the Gaussian process with covariance function $S(\tau,\tau')$ defined in Theorem 2. Therefore,

$$\sqrt{N}\begin{pmatrix}\hat{\alpha}(\cdot)-\alpha(\cdot)\\ \hat{\beta}(\cdot)-\beta(\cdot)\end{pmatrix} \Rightarrow J(\cdot)^{-1}\mathbb{G}(\cdot) \quad \text{in } \ell^\infty(\mathcal{T}).$$

$\square$

Proof of Proposition 1. The result follows immediately from assumptions. □

Proof of Theorem 3. Use the proof of Theorem 5 in Chernozhukov and Hansen (2006), P.518–520 for the functions and processes below:

- $\tau \mapsto \hat{z}(W_t, \tau)$, $t = 1, \ldots, T$,
- a Donsker set of functions $\{\xi(W_t, \tau), \tau \in \mathcal{T}, \xi \in \Xi, t = 1, \ldots, T\}$,
- the empirical process $(\tau, \xi) \mapsto \mathbb{G}_N(\xi(\tau)) \equiv 1/\sqrt{N} \sum_{i=1}^{N} \sum_{t=1}^{T} (\xi(W_{it}, \tau) - E\xi(W_{it}, \tau))$,
- its subsample realizations $(\tau, \xi) \mapsto \mathbb{G}_{j,b,N}(\xi(\tau)) \equiv 1/\sqrt{b} \sum_{i \in I_j} \sum_{t=1}^{T} (\xi(W_{it}, \tau) - E\xi(W_{it}, \tau))$, $j = 1, \ldots, B_N$. Let J_N denote the sampling (outer) law of $(\tau, \xi) \mapsto \mathbb{G}_N(\xi(\tau))$ in $\ell^\infty(\mathcal{T} \times \Xi)$,
- the subsampling law $L_{b,N}$ of $(\tau, \xi) \mapsto \mathbb{G}_{j,b,N}(\xi(\tau))$ in $\ell^\infty(\mathcal{T} \times \Xi)$.

□

Lemma B.1 (Stochastic expansion)**.** *Under Assumption 2, the following statements are true.*

I. $\sup_{(\alpha,\beta,\gamma,\tau) \in \mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T}} \left| \mathbb{E}_N \left(\sum_{t=1}^{T} \hat{g}(W_t, \alpha, \beta, \gamma, \tau) \right) - E \left(\sum_{t=1}^{T} g(W_t, \alpha, \beta, \gamma, \tau) \right) \right| \xrightarrow{p} 0.$

II. $\mathbb{G}_N f(W, \alpha(\cdot), \beta(\cdot), 0, \cdot) \Rightarrow \mathbb{G}(\cdot) \in \ell^\infty(\mathcal{T})$, *where* $\mathbb{G}$ *is a Gaussian process with covariance function* $S(\tau, \tau')$ *defined in Theorem 2. Furthermore, for any* $\hat{\alpha}(\tau)$, $\hat{\beta}(\tau)$, $\hat{\gamma}(\tau)$ *such that*

$$\sup_{\tau \in \mathcal{T}} \|(\hat{\alpha}(\tau), \hat{\beta}(\tau), \hat{\gamma}(\tau)) - (\alpha(\tau), \beta(\tau), 0)\| \xrightarrow{p} 0,$$

it is the case that

$$\sup_{\tau \in \mathcal{T}} \left\| \mathbb{G}_N \left(\sum_{t=1}^{T} \hat{f}(W_t, \hat{\alpha}(\tau), \hat{\beta}(\tau), \hat{\gamma}(\tau)) \right) - \mathbb{G}_N \left(\sum_{t=1}^{T} f(W_t, \alpha(\tau), \beta(\tau), 0) \right) \right\| \xrightarrow{p} 0.$$

Proof. We adapt the proof of Lemma B.2 in Chernozhukov and Hansen (2006), P.520–522.

Denote $\pi = (\alpha, \beta, \gamma)$ and $\Pi = \mathcal{A} \times \mathcal{B} \times \mathcal{G}$ where $\mathcal{G}$ is a closed ball at 0. The first step is to prove that the class of functions

$$\begin{aligned} \mathcal{H} = \{h = (\Phi_1, \ldots, \Phi_T, \Psi_1, \ldots, \Psi_T, \pi, \tau) \\ \mapsto \sum_{t=1}^{T} \varphi_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma)\Psi_t(X_t, Z_t), \\ \pi \in \Pi, \Phi_t \in \mathcal{F}, \Psi_t \in \mathcal{F}, t = 1, \ldots, T\} \end{aligned}$$

is Donsker, where $\mathcal{F}$ is defined in R4. By Corollary 2.7.4 in van der Vaart and Wellner (1996) the bracketing number of $\mathcal{F}$ satisfies

$$\log N_{[\cdot]}(\varepsilon, \mathcal{F}, L_2(P)) = O(\varepsilon^{-\dim(z,x)/\eta}) = O(\varepsilon^{-2-\delta'}),$$

for some $\delta' < 0$. So $\mathcal{F}$ is Donsker with a constant envelope. By Corollary 2.7.4 in van der Vaart and Wellner (1996) the bracketing number of

$$\mathcal{X} = \{(\Phi_1, \ldots, \Phi_T, \pi) \mapsto D_t'\alpha + X_t'\beta + \Phi_t(X_t, Z_t)'\gamma, t = 1, \ldots, T, \pi \in \Pi, \Phi_1, \ldots, \Phi_T \in \mathcal{F}\}$$

satisfies

$$\log N_{[\cdot]}(\varepsilon, \mathcal{X}, L_2(P)) = O(\varepsilon^{-\dim(z,d,x)/\eta}) = O(\varepsilon^{-2-\delta''}),$$

for some $\delta'' < 0$. By Remark 4 in Chernozhukov and Hong (2002), the bracketing number of

$$\mathcal{V} = \{(\Phi_1, \ldots, \Phi_T, \pi) \mapsto I(Y_t < D_t'\alpha + X_t'\beta + \Phi_t(X_t, Z_t)'\gamma), t = 1, \ldots, T, \\ \pi \in \Pi, \Phi_1, \ldots, \Phi_T \in \mathcal{F}\}$$

also satisfies

$$\log N_{[\cdot]}(\varepsilon, \mathcal{V}, L_2(P)) = O(\varepsilon^{-2-\delta''}).$$

$\mathcal{V}$ is Donsker since it has constant envelope by R1 and R4.

Class $\mathcal{H}$, defined as $\mathcal{H} \equiv \sum_{t=1}^{T} (\mathcal{T} - \mathcal{V}) \cdot \mathcal{F}$, is Lipschitz over $\mathcal{T} \times \mathcal{V} \times \mathcal{F}$. By Theorem 2.10.6 in van der Vaart and Wellner (1996) class $\mathcal{H}$ is Donsker, since it defined through products and sums of bounded Donsker classes $\mathcal{F}$, $\mathcal{V}$, and $\mathcal{T} \equiv \{\tau \mapsto \tau\}$.

The fact of $\mathcal{H}$ being Donsker is next used to prove claim II. The process

$$h = (\Phi_1, \ldots, \Phi_T, \Psi_1, \ldots, \Psi_T, \pi, \tau) \mapsto \mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_\tau (Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma) \Psi_t(X_t, Z_t) \right)$$

is Donsker in $\ell^\infty(\mathcal{H})$.

By R3 and R4, the process

$$\tau \mapsto (\tau, \alpha(\tau)', \beta(\tau)', \Phi_1(\tau, X_1, Z_1), \ldots, \Phi_T(\tau, X_T, Z_T), \Psi_1(\tau, X_1, Z_1), \ldots, \Psi_T(\tau, X_T, Z_T))$$

is uniformly Holder continuous in τ with respect to the supremum norm.

So the process

$$\tau \mapsto \mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_\tau (Y_t - D_t'\alpha(\tau) - X_t'\beta(\tau)) \Psi_t(\tau, X_t, Z_t) \right)$$

is also Donsker in $\ell^\infty(\mathcal{T})$.

Therefore,

$$\mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_\cdot (Y_t - D_t'\alpha(\cdot) - X_t'\beta(\cdot)) \Psi_t(\cdot, X_t, Z_t) \right) \Rightarrow \mathbb{G}(\cdot).$$

The process $\mathbb{G}(\cdot)$ has covariance function

$$\begin{aligned} S(\tau, \tau') &= E(\mathbb{G}(\tau)\mathbb{G}(\tau')') \\ &= E\left(\left(\sum_{t=1}^{T} \varphi_\tau (Y_t - D_t'\alpha(\tau) - X_t'\beta(\tau)) \Psi_t(\tau, X_t, Z_t) \right) \right. \\ &\qquad \left. \times \left(\sum_{s=1}^{T} \varphi_{\tau'} (Y_s - D_s'\alpha(\tau') - X_s'\beta(\tau')) \Psi_s(\tau', X_s, Z_s) \right)' \right) \\ &= E\left(\sum_{t=1}^{T} \sum_{s=1}^{T} l_t(\tau, \theta(\tau)) l_s(\tau', \theta(\tau')) \Psi_t(\tau) \Psi_s(\tau')' \right). \end{aligned}$$

where $l_t(\tau, \theta(\tau)) = \varphi_\tau(Y_t - D_t'\alpha(\tau) - X_t'\beta(\tau))$.

$\hat{\Psi}_t(\cdot) \xrightarrow{p} \Psi_t(\cdot)$ and $\hat{\Phi}_t(\cdot) \xrightarrow{p} \Phi_t(\cdot)$ uniformly over compact sets and $\hat{\pi}(\tau) - \pi(\tau) \xrightarrow{p} 0$ uniformly in τ. So by R3 and R4, $\delta_N \equiv \sup_{\tau \in \mathcal{T}} \rho(h'(\tau), h(\tau))|_{h'(\tau) = \hat{h}(\tau)} \xrightarrow{p} 0$, where ρ is the $L_2(P)$ pseudometric on $\mathcal{H}$:

$$\rho(h, \tilde{h}) \equiv \sqrt{ E\Big(\sum_{t=1}^{T} \| \varphi_\tau (Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma) \Psi_t(X_t, Z_t) \\ - \varphi_\tau (Y_t - D_t'\tilde{\alpha} - X_t'\tilde{\beta} - \tilde{\Phi}_t(X_t, Z_t)'\tilde{\gamma}) \tilde{\Psi}_t(X_t, Z_t) \|^2 \Big) }.$$

Accordingly, as $\delta_N \to 0$

$$
\begin{aligned}
\sup_{\tau\in\mathcal{T}} \Bigg\| \mathbb{G}_N & \left(\sum_{t=1}^{T} \varphi_\tau(Y_t - D_t'\hat{\alpha}(\tau) - X_t'\hat{\beta}(\tau) - \hat{\Phi}_t(\tau, X_t, Z_t)'\hat{\gamma}(\tau))\hat{\Psi}_t(\tau, X_t, Z_t) \right) \\
& - \mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_\tau(Y_t - D_t'\alpha(\tau) - X_t'\beta(\tau))\Psi_t(\tau, X_t, Z_t) \right) \Bigg\| \\
\leq & \sup_{\rho(\tilde{h},h)\leq\delta_N, \tilde{h}, h\in\mathcal{H}} \Bigg\| \mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_{\tilde{\tau}}(Y_t - D_t'\tilde{\alpha} - X_t'\tilde{\beta} - \tilde{\Phi}_t(X_t, Z_t)'\tilde{\gamma})\tilde{\Psi}_t(X_t, Z_t) \right) \\
& - \mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma)\Psi_t(X_t, Z_t) \right) \Bigg\| \xrightarrow{p} 0,
\end{aligned}
$$

by stochastic equicontinuity of

$$
h \mapsto \mathbb{G}_N \left(\sum_{t=1}^{T} \varphi_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma)\Psi_t(X_t, Z_t) \right),
$$

which proves claim II.

The final step proves claim I. Functions

$$
\begin{aligned}
\mathcal{P} = \{ & h = (\Phi_1, \ldots, \Phi_T, V_1, \ldots, V_T, \alpha, \beta, \gamma, \tau) \\
& \mapsto \sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma)V_t(X_t, Z_t)\}
\end{aligned}
$$

are Donsker, since they are bounded by R1 and uniformly Lipschitz over $\mathcal{F}^T \times \mathcal{F}^T \times \mathcal{A} \times \mathcal{B} \times \mathcal{G} \times \mathcal{T}$ (Theorem 2.10.6 in van der Vaart and Wellner (1996)).

Donskerness leads to a uniform law of large numbers

$$
\begin{aligned}
\sup_{h\in\mathcal{H}} \Bigg| \mathbb{E}_N & \left(\sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma)V_t(X_t, Z_t) \right) \\
& - E \left(\sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(X_t, Z_t)'\gamma)V_t(X_t, Z_t) \right) \Bigg| \xrightarrow{p} 0,
\end{aligned}
$$

which further yields

$$
\begin{aligned}
\sup_{(\alpha,\beta,\gamma,\tau)\in\mathcal{A}\times\mathcal{B}\times\mathcal{G}\times\mathcal{T}} \Bigg| \mathbb{E}_N & \left(\sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \hat{\Phi}_t(\tau, X_t, Z_t)'\gamma)\hat{V}_t(\tau, X_t, Z_t) \right) \\
& - E \left(\sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \hat{\Phi}_t(\tau, X_t, Z_t)'\gamma)\hat{V}_t(\tau, X_t, Z_t) \right) \Bigg| \xrightarrow{p} 0.
\end{aligned}
$$

$\hat{\Phi}_t(\cdot)$ and $\hat{\Psi}_t(\cdot)$, $t = 1, \ldots, T$, are uniformly consistent. By R4,

$$
\begin{aligned}
\sup_{(\alpha,\beta,\gamma,\tau)\in\mathcal{A}\times\mathcal{B}\times\mathcal{G}\times\mathcal{T}} \Bigg| E & \left(\sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \hat{\Phi}_t(\tau, X_t, Z_t)'\gamma)\hat{V}_t(\tau, X_t, Z_t) \right) \\
& - E \left(\sum_{t=1}^{T} \rho_\tau(Y_t - D_t'\alpha - X_t'\beta - \Phi_t(\tau, X_t, Z_t)'\gamma)V_t(\tau, X_t, Z_t) \right) \Bigg| \xrightarrow{p} 0,
\end{aligned}
$$

which proves claim I. $\square$

Lemma B.2. *Conditions I3 and I4(a,b) hold for the proposed implementation in Examples 1–2 under conditions R1–R4. In Example 3 conditions I3 and I4 hold under conditions R1–R4 for the instrumental variable quantile regression estimator for clustered data and the standard regularity conditions for the conventional quantile regression estimator for clustered data, e.g. those in Angrist et al. (2006) and Parente and Santos Silva (2016).*

Proof. We adapt the proof of Lemma C.1 in Chernozhukov and Hansen (2006), P.523.

Consider Example 1. I3 holds for $\hat{\theta}(\cdot)$ by Theorem 2 in Chernozhukov and Hansen (2005). As $r = 0$, $z_{it}(\tau) = R(\tau)(J(\tau)^{-1} l_{it}(\tau, \theta(\tau))\Psi_{it}(\tau))$, where

$$l_{it}(\tau, \theta(\tau)) = (\tau - I(Y_{it} < D_{it}'\alpha(\tau) - X_{it}'\beta(\tau))), \quad \Psi_{it}(\tau) = V_{it}(\tau)[\Phi_{it}(\tau)', X_{it}']'. \tag{B.2}$$

The proof of the fact that condition I4(a) holds in Example 1 is similar to the proof of Lemma B.2 in Chernozhukov and Hansen (2006), P.521 for the class of functions $\mathcal{H}$.

Since Ψ_{it} is a function of only X_{it} and Z_{it}, condition I4(b) holds by Theorem 1 in Chernozhukov and Hansen (2005). Condition I4(c) holds by R3.

Next, consider Example 2. Without the loss of generality, use $\hat{r}(\cdot) = \hat{\theta}(1/2)$. For $l_{it}(\cdot)$ defined in (B.2), $z_{it}(\tau) = R(\tau)(J(\tau)^{-1} l_{it}(\tau, \theta(\tau))\Psi_{it}(\tau) - J(1/2)^{-1} l_{it}(1/2, \theta(1/2))\Psi_{it}(1/2))$. In other words, $d_{it}(\tau, r(\tau)) = J(1/2)^{-1} l_{it}(1/2, \theta(1/2))\Psi_{it}(1/2)$and I3—I4 hold by the argument used in Example 1.

Finally, consider Example 3. $\hat{\vartheta}(\tau)$ is the estimate of $\hat{r}(\tau)$ in conventional quantile regression of Y on D and X, under the assumption of clustered data. Owing to the regularity conditions in Angrist et al. (2006) and Parente and Santos Silva (2016),

$$\sqrt{N}(\hat{\vartheta}(\cdot) - \vartheta(\cdot)) = -H(\cdot)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} d_{it}(\cdot, \vartheta(\cdot)) + o_p(1),$$

where $d_{it}(\tau, \vartheta(\tau)) = (\tau - I(Y_{it} < \tilde{X}_{it}'\vartheta(\tau)))\tilde{X}_{it}$ for $\tilde{X}_{it} = [D_{it}', X_{it}']'$, $H(\tau) = E\left(\sum_{t=1}^{T} f_{Y_t|\tilde{X}_t}(\tilde{X}_t'\vartheta(\tau))\tilde{X}_t\tilde{X}_t'\right)$. Therefore, $z_{it} = R(\tau)(J(\tau)^{-1} l_{it}(\tau, \theta(\tau))\Psi_{it}(\tau) - H(\tau)^{-1} d_{it}(\tau, \vartheta(\tau))$.

The proof that conditions I3 and I4 hold for $l_{it}(\tau, \theta(\tau))\Psi_{it}(\tau)$ is given in Example 1.

To prove that I4(a) holds for $d_{it}(\tau, \vartheta(\tau))$, exploit the proof of Lemma B.1 by substituting $\tilde{X}_{it}$ for Ψ_{it} and setting $\gamma = 0$.

I4(b) holds since $E(d_{it}(\tau, \vartheta(\tau))) = 0$, and I4(c) holds by R3. □

Appendix C Results of the empirical analysis

Table C.1: Impact of female labor supply (weeks worked per female) on female earnings in 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$ and $X_{3,st}$ as controls

Quantile	Mean	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
IV-QR with clustered SE		-0.176^{***}	-0.171^{***}	-0.139^{***}	-0.112^{***}	-0.103^{***}	-0.095^{***}	-0.083^{***}	-0.072^{***}	-0.055^{***}
		(0.052)	(0.042)	(0.032)	(0.029)	(0.024)	(0.021)	(0.018)	(0.018)	(0.016)
IV-QR with i.i.d. SE		-0.176^{***}	-0.171^{***}	-0.139^{***}	-0.112^{***}	-0.103^{***}	-0.095^{***}	-0.083^{***}	-0.072^{***}	-0.055^{***}
		(0.028)	(0.017)	(0.013)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)	(0.011)
QR with clustered SE		-0.023	-0.027^{*}	-0.024^{*}	-0.018	-0.016	-0.010	-0.009	-0.002	0.004
		(0.019)	(0.016)	(0.014)	(0.011)	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)
TSLS with clustered SE	-0.108^{***}									
	(0.025)									

Notes: The model (9)–(11) is estimated using 69,335 observations. Dependent variable is log weekly earnings. Specification includes age structure and state of birth. Standard errors are in parentheses. *, **, *** means significance at 10%, 5%, 1% respectively.

Table C.2: Impact of female labor supply (weeks worked per female) on female earnings in 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$, $X_{3,st}$ and $X_{4,st}$ as controls

Quantile	Mean	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
IV-QR with clustered SE		-0.121^{**}	-0.135^{***}	-0.125^{***}	-0.092^{*}	-0.076^{**}	-0.044	-0.036	-0.027	-0.008
		(0.057)	(0.050)	(0.048)	(0.048)	(0.039)	(0.030)	(0.025)	(0.021)	(0.021)
IV-QR with i.i.d. SE		-0.121^{***}	-0.135^{***}	-0.125^{***}	-0.092^{***}	-0.076^{***}	-0.044^{***}	-0.036^{**}	-0.027^{*}	-0.008
		(0.032)	(0.022)	(0.018)	(0.017)	(0.015)	(0.014)	(0.014)	(0.016)	(0.018)
QR with clustered SE		-0.012	-0.020^{*}	-0.014	-0.006	-0.003	0.002	0.004	0.009	0.022^{***}
		(0.012)	(0.011)	(0.010)	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
TSLS with clustered SE	-0.073^{**}									
	(0.037)									

Notes: The model (9)–(11) is estimated using 69,335 observations. Dependent variable is log weekly earnings. Specification includes age structure, state of birth, share farmers, share nonwhite, and average education. Standard errors are in parentheses. *, **, *** means significance at 10%, 5%, 1% respectively.

Table C.3: Impact of female labor supply (weeks worked per female) on male earnings in 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$ and $X_{3,st}$ as controls

Quantile	Mean	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
IV-QR with clustered SE		-0.114^{***}	-0.098^{***}	-0.082^{***}	-0.080^{***}	-0.077^{***}	-0.073^{***}	-0.062^{***}	-0.045^{***}	-0.036^{**}
		(0.026)	(0.020)	(0.018)	(0.018)	(0.017)	(0.017)	(0.015)	(0.014)	(0.016)
IV-QR with i.i.d. SE		-0.114^{***}	-0.098^{***}	-0.082^{***}	-0.080^{***}	-0.077^{***}	-0.073^{***}	-0.062^{***}	-0.045^{***}	-0.036^{***}
		(0.014)	(0.008)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.010)
QR with clustered SE		-0.019^{**}	-0.015^{*}	-0.013^{*}	-0.011	-0.010	-0.007	-0.005	-0.000	0.003
		(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
TSLS with clustered SE	-0.070^{***}									
	(0.015)									

Notes: The model (9)–(11) is estimated using 198,385 observations. Dependent variable is log weekly earnings. Specification includes age structure and state of birth. Standard errors are in parentheses. *, **, *** means significance at 10%, 5%, 1% respectively.

Table C.4: Impact of female labor supply (weeks worked per female) on male earnings in 1940–50, specification with $X_{1,ist}$, $X_{2,ist}$, $X_{3,st}$ and $X_{4,st}$ as controls

Quantile	Mean	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
IV-QR with clustered SE		-0.051^{**}	-0.039^{**}	-0.018	-0.014	-0.012	-0.009	-0.008	0.002	0.012
		(0.023)	(0.019)	(0.017)	(0.017)	(0.015)	(0.015)	(0.016)	(0.014)	(0.020)
IV-QR with i.i.d. SE		-0.051^{***}	-0.039^{***}	-0.018^{**}	-0.014	-0.012	-0.009	-0.008	0.002	0.012
		(0.016)	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)	(0.013)
QR with clustered SE		-0.006	-0.003	0.000	0.001	0.003	0.005	0.005	0.009^{***}	0.012^{**}
		(0.006)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
TSLS with clustered SE	-0.021									
	(0.017)									

Notes: The model (9)–(11) is estimated using 198,385 observations. Dependent variable is log weekly earnings. Specification includes age structure, state of birth, share farmers, share nonwhite, and average education. Standard errors are in parentheses. *, **, *** means significance at 10%, 5%, 1% respectively.