Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Instrumental variable quantile regression for clustered data

Galina Besstremyannaya[a,*], Sergei Golovan[b]

[a] National Research University Higher School of Economics, Pokrovsky blvd. 11, Moscow, Russia
[b] New Economic School, Skolkovskoe shosse, 45, Moscow, Russia

ARTICLE INFO

Article history:
Received 11 October 2022
Revised 18 June 2023
Accepted 18 June 2023
Available online xxx

JEL classification:
C21
C23
C26
D12

Keywords:
Quantile regression
Endogeneity
Clustered data
Instrumental variables

ABSTRACT

The purpose is to enable inference in case of quantile regression with endogenous covariates and clustered data. It is proven that the instrumental variable quantile regression estimator is consistent where there is correlation of errors within clusters, and an asymptotic distribution for the estimator, which may be used for inference for a given quantile τ, is derived. As regards inference based on the entire instrumental variable quantile regression process, it is proven that cluster-based resampling of a statistic of a certain class offers a computationally tractable approach for implementing asymptotic tests. The theoretical results concerning the asymptotic properties of the instrumental variable quantile regression estimator for clustered data are supported by simulation analysis. An empirical illustration shows the use of the proposed technique in order to estimate the earning equations of US men and women where female labor supply is endogenous and subject to the shock of World War II.

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 dependence 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, which has been implemented for a wide range of least-squares based methods, through correction of the asymptotic variance matrix to account for intra-cluster correlation (Cameron and Miller, 2011). As regards quantile regression models, to the best of our knowledge, such an approach with cluster-robust standard errors is only available for the model with exogenous covariates. The correction of the asymptptic variance matrix of the estimator was suggested by Wooldridge (2007) and a formal proof of the asymptotic properties of the estimator can be found in Parente and Santos Silva (2016) and Hagemann (2017). However, quantile regression is often applied to study heterogeneous response of the dependent variable to endogenous covariates. There are a number of ways of dealing with endogeneity in quantile regression models (Abadie, 2002; Harding and Lamarche, 2014; Chernozhukov and Hansen, 2013; Harding and Lamarche, 2017; Chetverikov et al., 2016; Huber and Wüthrich, 2019), but the instrumental-variable quantile regression model of Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2006) is a widely used and computationally convenient econometric technique for

* Corresponding author.
E-mail addresses: gbesstremyannaya@hse.ru (G. Besstremyannaya), sgolovan@nes.ru (S. Golovan).

https://doi.org/10.1016/j.ecosta.2023.06.005
2452-3062/

Please cite this article as: G. Besstremyannaya and S. Golovan, Instrumental variable quantile regression for clustered data, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2023.06.005

this purpose (Chernozhukov et al., 2017). As of June 2023, 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 3,000. However, the asymptotic properties of the estimator are developed in Chernozhukov and Hansen (2006) only under the assumption of i.i.d. observations.

Our purpose here 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 resampling 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.

Although correction of the asymptotic variance matrix is the approach most commonly used in the general model of quantile regression, there are other ways of dealing with restricted versions of clustered data. In case of exogeneous covariates Geraci (2019) applies a semiparametric estimator to deal with intra-cluster correlation caused by random effects, while Zhang et al. (2019) propose a two-step procedure for the model where coefficients differ across groups of clusters. As for quantile regression under endogeneity, Chetverikov et al. (2016) offer an estimator for clustered data in the model with group-level endogeneous variables.

Section 2 below sets up the model with the quantile regression with endogenous covariates and clustered data. Section 3 derives the asymptotic distribution for the estimator in the model. Section 4 develops inference for the entire instrumental variable quantile process for clustered data. The results of the simulations are shown in Section 5. Section 6 presents the results of the empirical study. Proofs are given in the Appendix.

2. The model setup and identification

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

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

We suppose that regularity conditions A1–A5 for the instrumental variable quantile regression from (Chernozhukov and Hansen, 2005, p.248) hold. 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), p.249). 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 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 above-mentioned setup by Chernozhukov and Hansen (2005) but relax the i.i.d. assumption and consider that observations are sampled from a data generating process with clusters:

$$\{(Y_{i1}, \dots, Y_{iK}), (D_{i1}, \dots, D_{iK}), (X_{i1}, \dots, X_{iK}), (Z_{i1}, \dots, Z_{iK})\},$$

and

$$Y_{ik} = D'_{ik}\alpha(U_{D_{ik}}) + X'_{ik}\beta(U_{D_{ik}}), \quad D_{ik} = \delta(Z_{ik}, X_{ik}, v_{ik}), \quad i = 1, \dots, N,\ k = 1, \dots, K,$$

where i is the number of a cluster, k is the index within a cluster, v_{ik} is a random variable, and $U_{D_{ik}} \sim U[0, 1]$ conditionally on Z_{ik}, X_{ik}.

Similar to Parente and Santos Silva (2016) the size of all clusters is considered fixed.

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) 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.

Assumption 1.

R1 (Sampling) $\{(Y_{i1},\ldots,Y_{iK}),(D_{i1},\ldots,D_{iK}),(X_{i1},\ldots,X_{iK}),(Z_{i1},\ldots,Z_{iK})\}$ are i.i.d. 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_k has uniformly bounded conditional density, a.s. $\sup_{y\in\mathbb{R}} f_{Y_k|X_k,D_k,Z_k}(y)<M$, and for $\pi\equiv(\alpha,\beta,\gamma)$, $\theta\equiv(\alpha,\beta)$, and

$$\Pi_k(\pi,\tau)\equiv E((\tau-1(Y_k<D_k'\alpha+X_k'\beta+\Phi_k(\tau)'\gamma))\Psi_k(\tau)),$$
$$\Pi_k(\theta,\tau)\equiv E((\tau-1(Y_k<D_k'\alpha+X_k'\beta))\Psi_k(\tau)),\quad \Psi_k(\tau)\equiv V_k(\tau)\cdot[\Phi_k(\tau)',X_k']',$$

where $\Phi_k(\tau)=\Phi_k(\tau,Z_k,X_k)$ are transformations of instruments, $V_k(\tau)=V_k(\tau,Z_k,X_k)$ are weights, Jacobian matrices $\frac{\partial}{\partial(\alpha,\beta)}\Pi_k(\theta,\tau)$ and $\frac{\partial}{\partial(\beta,\gamma)}\Pi_k(\pi,\tau)$ 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 $(\alpha,\beta)\mapsto\Pi_k(\theta,\tau)$ is simply connected for all $k=1,\ldots,K$.

R4 (Estimated instruments and weights) With probability going to 1, the functions $\hat\Phi_k(\tau,z,x)$, $\hat V_k(\tau,z,x)\in\mathcal{F}$ and $\hat\Phi_k(\tau,z,x)\to_p\Phi_k(\tau,z,x)$, $\hat V_k(\tau,z,x)\to_p V_k(\tau,z,x)$ uniformly in (τ,z,x) over compact sets, where $\Phi_k(\tau,z,x)$ and $V_k(\tau,z,x)\in\mathcal{F}$, the functions $f_k(\tau,z,x)\in\mathcal{F}$ are uniformly smooth functions in (z,x) with the uniform smoothness order $\eta>\dim(d,z,x)/2$, and $\|f_k(\tau',z,x)-f_k(\tau,z,x)\|<C|\tau'-\tau|^a$, $C>0$, $a>0$, for all (z,x,τ,τ') and $k=1,\ldots,K$.

Conditions R1–R4 from Assumption 1 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.

Note that condition R1 from Assumption 1 is fairly strict. It implies that all clusters are of equal size and that cluster size is fixed. Additionally, clusters are considered independent, and the interdependence of observations within the cluster does not differ across clusters. These are convenient requirements which make it possible to adapt most of the proofs of Chernozhukov and Hansen (2006) and to avoid technicalities.

There are several ways in which condition R1 may be violated:

- if cluster size grows indefinitely when sample size tends to infinity, one would expect the asymptotic properties of the instrumental variables quantile regression estimator to be completely different from what we are outlining here;
- if cluster sizes are different but bounded by a certain constant, or if observations in different clusters have only slightly different interdependence within the cluster, then the expression for the asymptotic covariance matrix of the estimator will change (see Section 3.1, which adapts the estimator to allow the latter extensions).

Theorem 1. *Under the regularity conditions from Chernozhukov and Hansen (2005) and Assumption 1,* $(\alpha',\beta')=(\alpha(\tau)',\beta(\tau)')$ *uniquely solves the system of equations* $E(\tau-I(Y_k<D_k'\alpha+X_k'\beta)\Psi_k(\tau))=0$ *over* $\mathcal{A}\times\mathcal{B}$ *for all* $k=1,\ldots,K$.

This theorem is analogous to the corresponding identification theorem in Chernozhukov and Hansen (2005).

3. The estimator for clustered data and its asymptotic properties

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

$$Q_N(\tau,\alpha,\beta,\gamma)=\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\rho_\tau(Y_{ik}-D_{ik}'\alpha-X_{ik}'\beta-\hat\Phi_{ik}(\tau)'\gamma)\hat V_{ik}(\tau),\tag{2}$$

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

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

The first step requires the estimate

$$\left(\hat\beta(\alpha,\tau),\hat\gamma(\alpha,\tau)\right)=\operatorname*{argmin}_{\beta,\gamma}Q_N(\tau,\alpha,\beta,\gamma).\tag{3}$$

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

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

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

The procedure for estimating the parameters $\theta(\tau)=(\alpha(\tau),\beta(\tau))$ exactly coincides with the corresponding procedure in Chernozhukov and Hansen (2006). Some differences can arise because 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 the 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 $A(\tau)$.

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

Theorem 2. *Given the regularity conditions from Chernozhukov and Hansen (2005) and Assumption 1, for* $\varepsilon_{ik}(\tau)=Y_{ik}-D_{ik}'\alpha(\tau)-X_{ik}'\beta(\tau)$ *and* $l_{ik}(\tau,\theta(\tau))=\tau-I(\varepsilon_{ik}(\tau)<0)$*:*

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

for $N\to\infty$, *where* $b(\cdot)$ *is a mean zero Gaussian process on (0,1) with covariance function* $E(b(\tau)b(\tau')')=J(\tau)^{-1}S(\tau,\tau')J(\tau')^{-1}$,

$$J(\tau)=E\left(\sum_{k=1}^{K}f_{\varepsilon_k(\tau)}(0|X_k,D_k,Z_k)\Psi_k(\tau)[D_k',X_k']\right),$$

$$S(\tau,\tau')=E\left(\sum_{k=1}^{K}\sum_{s=1}^{K}l_k(\tau,\theta(\tau))l_s(\tau',\theta(\tau'))\Psi_k(\tau)\Psi_s(\tau')'\right).$$

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

$$\hat{J}(\tau)=\frac{1}{2Nh_N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}I(|\hat{\varepsilon}_{ik}(\tau)|\le h_N)\hat{\Psi}_{ik}(\tau)[D_{ik}',X_{ik}']\right),$$

where $\hat{\varepsilon}_{ik}=Y_{ik}-D_{ik}'\hat{\alpha}(\tau)-X_{ik}'\hat{\beta}(\tau)$ 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:

$$\hat{S}(\tau,\tau')=\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K}\sum_{s=1}^{K}l_{ik}(\tau,\hat{\theta}(\tau))l_{is}(\tau',\hat{\theta}(\tau'))\hat{\Psi}_{ik}(\tau)\hat{\Psi}_{is}(\tau')'\right).$$

It should be noted that the expression for $J(\tau)$ and $S(\tau,\tau')$ in case of clustered data differs from the corresponding expression in Chernozhukov and Hansen (2006). This causes differences in the expressions for the estimators of $J(\tau)$ and $S(\tau,\tau')$.

Regarding $\hat{J}(\tau)$, the bandwidth in the model with clustered data depends only on N (the number of independent clusters). However, the corresponding version in the Chernozhukov and Hansen (2006) estimator would have bandwidth depending on NT (the total number of observations in the sample).

The expression for $\hat{S}(\tau,\tau')$ in the model with clustered data has a double sum over individual clusters in parentheses, and denominator N. However, the Chernozhukov and Hansen (2006) estimator would have the denominator NT and a single sum for all the observations. It should be noted that similar structural differences with respect to the estimators of $J(\tau)$ and $S(\tau,\tau')$ are observed in the model with clustered data under exogeneity (Parente and Santos Silva, 2016; Wooldridge, 2007).

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:

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

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):

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

So Theorem 2 makes it possible to test hypotheses about coefficients for one quantile or several quantiles. For instance, the Wald statistic in the form

$$W=N(\hat{\theta}(\tau)-\theta_0)'(\hat{J}(\tau)^{-1}\hat{S}(\tau,\tau)(\hat{J}(\tau)^{-1})')^{-1}(\hat{\theta}(\tau)-\theta_0),$$

can be used to test hypothesis $H_0:\theta(\tau)=\theta_0$ against the alternative $H_1:\theta(\tau)\neq\theta_0$. Hypothesis $H_0:\theta(\tau)=\theta(\tau')$ can be tested against the alternative $H_1:\theta(\tau)\neq\theta(\tau')$ using the Wald statistic

$$W = N\left(\begin{pmatrix} I \\ -I \end{pmatrix}'\begin{pmatrix} \hat{\theta}(\tau) \\ \hat{\theta}(\tau') \end{pmatrix}\right)'\left(\begin{pmatrix} I \\ -I \end{pmatrix}'\begin{pmatrix} \hat{J}(\tau)^{-1}\hat{S}(\tau,\tau)(\hat{J}(\tau)^{-1})' & \hat{J}(\tau)^{-1}\hat{S}(\tau,\tau')(\hat{J}(\tau')^{-1})' \\ \hat{J}(\tau')^{-1}\hat{S}(\tau',\tau)(\hat{J}(\tau)^{-1})' & \hat{J}(\tau')^{-1}\hat{S}(\tau',\tau')(\hat{J}(\tau')^{-1})' \end{pmatrix}\begin{pmatrix} I \\ -I \end{pmatrix}\right)^{-1}$$
$$\left(\begin{pmatrix} I \\ -I \end{pmatrix}'\begin{pmatrix} \hat{\theta}(\tau) \\ \hat{\theta}(\tau') \end{pmatrix}\right).$$

3.1. Practical considerations

The analysis above is based on the three suppositions summarized in condition R1 of Assumption 1: firstly, that cluster size is fixed; secondly, that the clusters are of the same size; and thirdly, that the dependence within all clusters is identical.

In practice it is very unlikely that a researcher will be presented with a dataset that has equal-sized clusters which are identically distributed. Real datasets usually consist of clusters of different sizes. And even if clusters are of the same size, the correlation between individual observations within clusters may differ. Similarly to the approach in least squares regression with clusters of unequal size K_i, it seems reasonable to adjust the formulas for the parts of the covariance matrix $\hat{J}(\tau)$ and $\hat{S}(\tau,\tau')$:

$$\hat{J}(\tau) = \frac{1}{2Nh_N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K_i} I(|\hat{\varepsilon}_{ik}(\tau)| \le h_N)\hat{\Psi}_{ik}(\tau)[D'_{ik}, X'_{ik}]\right),$$
$$\hat{S}(\tau,\tau') = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K_i}\sum_{s=1}^{K_i} l_{ik}(\tau,\hat{\theta}(\tau))l_{is}(\tau',\hat{\theta}(\tau'))\hat{\Psi}_{ik}(\tau)\hat{\Psi}_{is}(\tau')'\right).$$

Our simulations, which are available upon request, suggest that the adjusted formulas work reasonably well in cases where

- clusters have approximately the same size, which does not grow indefinitely with increase of N;
- clusters have only slightly differing correlations between their individual observations.

However, if clusters are either very large and different in size, or have very different structure, clustered standard errors become biased. This is similar to the bias of standard errors under the i.i.d. assumption, but the magnitude of the bias in case of clustered standard errors is smaller.

4. General inference

While basic inference deals with the tests for coefficients for a given quantile or several quantiles, general inference evaluates null hypotheses formulated for a set of quantiles $\mathcal{T} \subset (0,1)$ (Koenker and Hallock, 2001; Koenker and Xiao, 2002; Koenker, 2005, Section 3.7, 3.8), and usually employs the Kolmogorov–Smirnov (KS) or Cramér–von Mises (CM) test statistic (Chernozhukov and Fernández-Val, 2005). Owing to difficulties in derivation of the asymptotic properties for these statistics, their critical values are usually computed through resampling (Chernozhukov and Fernández-Val, 2005; Koenker and Xiao, 2002; Abadie, 2002). The standard resampling procedure estimates the quantile regression model for each resampled sample He (2017). But it is computationally complex for the instrumental variables quantile regression. Therefore, Chernozhukov and Hansen (2006) use a linear approximation of the estimator $\hat{\theta}(\tau)$ to resample its summands. In case of our estimator with clustered data, observations are drawn from the data generating process by clusters. Accordingly, in this section we carry out resampling by clusters and prove validity of this approach. It should be noted that a similar approach is proposed in He (2017) for quantile regression under exogeneity.

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, $q \le p = \dim\theta(\tau)$, $r(\tau) \in \mathbb{R}^p$, and $\theta(\tau)$ and $r(\tau)$ are the 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 2 (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) $g(\tau) \ne 0$ 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 hypothesis and the alternative.

Assume that along with conditions I1 and I2 from Assumption 2, the estimates in quantile regression for clustered data admit the linear representations below.

Assumption 3.

I3. Linear representations:

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

and

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

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

I4. (a) The estimates $l_{ik}(\cdot,\hat{\theta}(\cdot))\hat{\Psi}_{ik}(\cdot)$ and $d_{ik}(\cdot,\hat{r}(\cdot))\hat{\Upsilon}_{ik}(\cdot)$ 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) With probability going to 1, $E(l_{ik}(\tau,\theta(\tau)\hat{\Psi}_{ik}(\tau)))=0$ and $E(d_{ik}(\tau,r(\tau)\hat{\Upsilon}_{ik}(\tau)))=0$ for each i, k. (c) $E\|l_{ik}(\tau,\theta)-l_{ik}(\tau,\theta')\|<C\|\theta-\theta'\|$, $E\|d_{ik}(\tau,r)-d_{ik}(\tau,r')\|<C\|r-r'\|$, uniformly in $\tau\in\mathcal{T}$ and in (θ,θ',r,r') over compact sets.

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

Proposition 1. *Under the regularity conditions from Chernozhukov and Hansen (2005), and Assumptions 2, 3*

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

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

The estimate of $z_{ik}(\cdot)$ becomes $\hat{z}_{ik}(\cdot)=R(\cdot)(\hat{J}(\cdot)^{-1}l_{ik}(\cdot,\hat{\theta}(\cdot))\hat{\Psi}_{ik}(\cdot)-\hat{H}(\cdot)^{-1}d_{ik}(\cdot,\hat{r}(\cdot))\hat{\Upsilon}_{ik}(\cdot))$, 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, p.506): 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. An alternative to the modified approach of Chernozhukov and Hansen (2006), set out below, is modification of the resampling procedure, as 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 statistics similar to those of Kolmogorov–Smirnov. But there are differences in 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 resample these scores. By contrast, Hagemann (2017) uses a conventional resampling approach and estimates a large number of quantile regressions: a regression for each sample. Such an approach implies a high computational burden in case of instrumental variable quantile regression, and is therefore not employed here.

The inference is then carried out through the following algorithm.

Step 1. Resampling procedure with clustered data: To resample from $\{\hat{z}_{ik}(\tau), i=1,\ldots,N, k=1,\ldots,K, \tau\in\mathcal{T}\}$, 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 $\upsilon_{j,b,N}(\tau)\equiv 1/b\sum_{i\in I_j}\sum_{k=1}^{K}\hat{z}_{ik}(\tau)$. Denote $\hat{S}_{j,b,N}\equiv f(\sqrt{b}\upsilon_{j,b,N}(\cdot))$ as

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

where S_N is, respectively, the KS or CM statistic. Note that resampling is carried out 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 $S=f(\upsilon(\cdot))$, it defines $\Gamma(x)\equiv P(S\le x)$ and suggests the estimate of $\Gamma(x)$ as $\hat{\Gamma}_{b,N}(x)=1/B_N\sum_{j=1}^{b_N}I(S_{j,b,N}\le x)$. The critical value for the test is $c_{b,N}(1-\alpha)=\inf\{c:\hat{\Gamma}_{b,N}(c)\ge 1-\alpha\}$, i.e. the $(1-\alpha)$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 the regularity conditions from Chernozhukov and Hansen (2005), and Assumptions 2, 3 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,* $S_N \xrightarrow{p} \infty$, $c_{b,N}(1-\alpha) = O_p(1)$, $P(S_N > c_{b,N}(1-\alpha)) \to 1$;

(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 three examples of hypotheses from Chernozhukov and Hansen (2006), for which we formulate scores in case of instrumental variable quantile regression for clustered data and $\dim\alpha \equiv 1$. Lemma A.2 in the Appendix ensures that conditions I3 and I4 of Assumption 3 are satisfied under Assumption 1 for our implementation of these three tests.

1. *No effect of the endogenous variable.*
   H_0: $\alpha(\tau) = 0$ for all $\tau \in \mathcal{T}$, $R(\cdot) \equiv R = [1, 0, \ldots, 0]$, $r(\cdot) \equiv 0$.
   In this case, $\hat{z}_{ik}(\tau) = R(\tau)[\hat{J}(\tau)^{-1} l_{ik}(\tau, \hat{\theta}(\tau))\hat{\Psi}_{ik}(\tau)]$, where $l_{ik}(\tau, \hat{\theta}(\tau)) = (\tau - I(Y_{ik} < D'_{ik}\hat{\alpha}(\tau) + X'_{ik}\hat{\beta}(\tau)))$, $\hat{\Psi}_{ik}(\tau) = \hat{V}_{ik}[\hat{\Phi}_{ik}(\tau), X'_{ik}]'$.
2. *Constant effect of the endogenous variable across quantiles.*
   H_0: $\alpha(\tau) \equiv$ const, for all $\tau \in \mathcal{T}$, $R(\cdot) \equiv R = [1, 0, \ldots, 0]$ and $r(\cdot) = [\alpha(1/2), 0, \ldots, 0]'$.
   In this case, $[\hat{\alpha}(1/2), 0, \ldots, 0]$ can be taken for $\hat{r}(\cdot)$, and $\hat{z}_{ik}(\tau) = R(\tau)[\hat{J}(\tau)^{-1} l_{ik}(\tau, \hat{\theta}(\tau))\hat{\Psi}_{ik}(\tau) - \hat{J}(1/2)^{-1} l_{ik}(1/2, \hat{\theta}(1/2))\hat{\Psi}_{ik}(1/2)]$, for $l_{ik(\cdot,\hat{\theta}(\cdot))}$ defined in the example with the hypothesis of no effect.
3. *Exogeneity hypothesis.*
   H_0: The coefficient for the endogenous variable in the instrumental variable quantile regression equals the coefficient for this variable in the quantile regression under exogeneity, $R(\cdot) \equiv R = [1, 0, \ldots, 0]$, $r(\cdot) = \vartheta(\cdot)$, where $\vartheta(\cdot)$ is estimated using quantile regression under exogeneity.
   Then, the score is given by $\hat{z}_{ik}(\tau) = R(\tau)[\hat{J}(\tau)^{-1} l_{ik}(\tau, \hat{\theta}(\tau))\hat{\Psi}_{ik}(\tau) - \hat{H}(\tau)^{-1} d_{ik}(\tau, \hat{\vartheta}(\tau))]$, where $d_{ik}(\tau, \hat{\vartheta}(\tau)) = (\tau - I(Y_{ik} < \tilde{X}'_{ik}\hat{\vartheta}(\tau)))\tilde{X}_{ik}$, $\tilde{X}_{ik} = [D'_{ik}, X'_{ik}]'$, and following Parente and Santos Silva (2016), $\hat{H}(\tau) = \frac{1}{2Nh_N}\sum_{i=1}^{N}\left(\sum_{k=1}^{K} I(|\hat{\epsilon}_{ik}| \le h_N)\tilde{X}_{ik}\tilde{X}'_{ik}\right)$ for $\hat{\epsilon}_{ik}(\tau) = Y_{ik} - \tilde{X}'_{ik}\hat{\vartheta}(\tau)$.

5. Simulations

The next task is a 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

$$\begin{aligned} &Y_{ik} = D_{ik} \cdot \alpha \cdot U_{ik} + \beta_0 \cdot U_{ik} + X_{ik} \cdot \beta_1 \cdot U_{ik}, \\ &\quad i = 1, \ldots, N, \quad k = 1, \ldots, K, \end{aligned}$$

where i is the index for a cluster, k 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_{ik}, which varies across clusters and observations, and ζ_i, which denotes the individual effect of clusters. For this purpose, we draw variables ξ_{ik} and ζ_i from the Gamma distribution:

$$\begin{aligned} &\xi_{jik} \sim \Gamma(1, 1), j = 1, \ldots, 5, \\ &\zeta_{ji} \sim \Gamma(2, 1), j = 1, \ldots, 5. \end{aligned}$$

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

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

The number of clusters $N \in \{100, 200, 500, 1000, 2000\}$ and the size of cluster $K \in \{2, 5, 10\}$. 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_{ik} on D_{ik}, X_{ik} and a constant, using Z_{ik} as an instrument for D_{ik}. We then 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: $\alpha \equiv 0$, $\alpha \equiv$ const, D is exogenous. For each hypothesis we contrast the results of the test based on the Chernozhukov and Hansen (2006) procedure (i.e. resampling individual observations) and the procedure which resamples clusters. We simulated 500 samples, and used 10,000 draws in each case. The estimations are carried out using custom code in Python.

Table 1
Simulated true sizes for basic inference tests at the 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		
K	N	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
	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
2	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
	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
5	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
	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
10	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 the 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		
K	N	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
	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
2	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
	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
5	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
	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
10	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 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 of equality of α to its true value: the rejection rates exceed 0.2 for $K = 5$ and 0.3 for $K = 10$. However, the rejection rate is close to 0.1 for the instrumental variable quantile regression estimator with clustered standard errors. Note that for each τ and each combination of the values of N and K, 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 when Chernozhukov and Hansen (2006) resampling is applied 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 for the basic inference as to 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 for 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 possible wrong conclusions if the Chernozhukov and Hansen (2006) resampling procedure is used when there is within-cluster correlation of errors. For each of the process tests the rejection rate of the null hypothesis is much higher for resampling under the i.i.d. assumption than in case of resampling based on cluster-

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

		i.i.d. resampling			clustered resampling		
K	N	$\alpha(\tau)\equiv 0$	$\alpha(\tau)\equiv$ const	D is exog.	$\alpha(\tau)\equiv 0$	$\alpha(\tau)\equiv$ const	D is exog.
2	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
	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
5	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
	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
10	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
	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: Hypotheses of no effect and of constant effect are tested in the model with $\alpha(\tau)\equiv 0$ and endogenous D. Exogeneity of D is tested in a model with $\alpha(\tau) = 1\cdot\tau$ and exogenous D.

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

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

ing. 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. This confirms the implications of our Theorem 3.

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

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

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

The impact of women's 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 volume 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 of the average level of women's wages (Cain and Dooley, 1976).

It should be noted that the volume of labor is likely to be endogenous in the wage equation. Endogeneity occurs because both the volume of labor and wages are found as a solution of the system of demand and supply equations. A pioneering work Acemoglu et al. (2004) uses an instrumental variable approach to account for endogeneity in the female labor force in time of war, estimating least-squares models and using the 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 the female labor force is the military fatality rate.

We extend the analysis for the case when the effect of labor supply on earnings is heterogeneous. It is indeed 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 the 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 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 https://economics.mit.edu/people/faculty/david-h-autor/data-archive. Specifically, our empirical illustration is based on the data and variables for Table 9 on pp.534–535 in Acemoglu et al. (2004): https://www.dropbox.com/s/4shgtz5i0t0to6i/table5-9-10-11-12-a1_dta.zip?dl=0 and https://www.dropbox.com/s/sjjlgvihzuidz92/table9_do_log.zip?dl=0. 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). The total numbers of observations in the pooled data with samples from the 1940 and 1950 censuses were 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 follows:

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

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

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

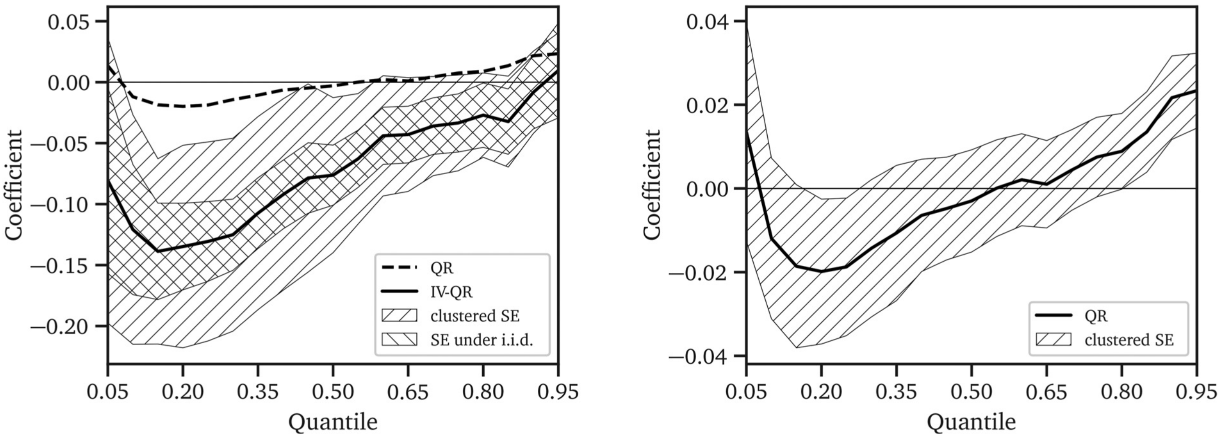
where τ denotes the value of a given quantile for conditional distribution of the log of weekly earnings ($\ln w_{ist}$) for individual i at state s in period t (1940 or 1950), the endogenous variable D_{st} is female labor supply (average weeks worked per woman in state s in year t) and $X_{ist} = [X_{1,ist}, X_{2,ist}, X_{3,st}, X_{4,st}]$ is a vector of exogenous variables. Specifically: $X_{1,ist}$ 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, v_{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 account for clustered data in a similar way. Specifically, we contrast standard errors for the coefficient for female labor supply estimated in (9)–(11) assuming standard errors based on the i.i.d. assumption with standard errors for the same coefficient at state-year level. The analysis uses 19 values of $\tau \in [0.05, 0.95]$, starting with $\tau = 0.05$ at the 0.05 step.

The first set of results describes the impact of female labor supply on women's earnings. As shown in Figure 1, the coefficient for the amount of weeks worked per woman is negative in explaining the log of women's earnings. With an increase in the 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 (Table B.1 of the Appendix, which, for the sake of brevity, shows the estimated coefficients and their standard errors for $\tau \in [0.1, 0.9]$, with the 0.1 step). This fact implies a positive correlation of errors within clusters.

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 1). In other words, failure to account for clustered standard errors leads to wrong conclusions for high-wage workers.



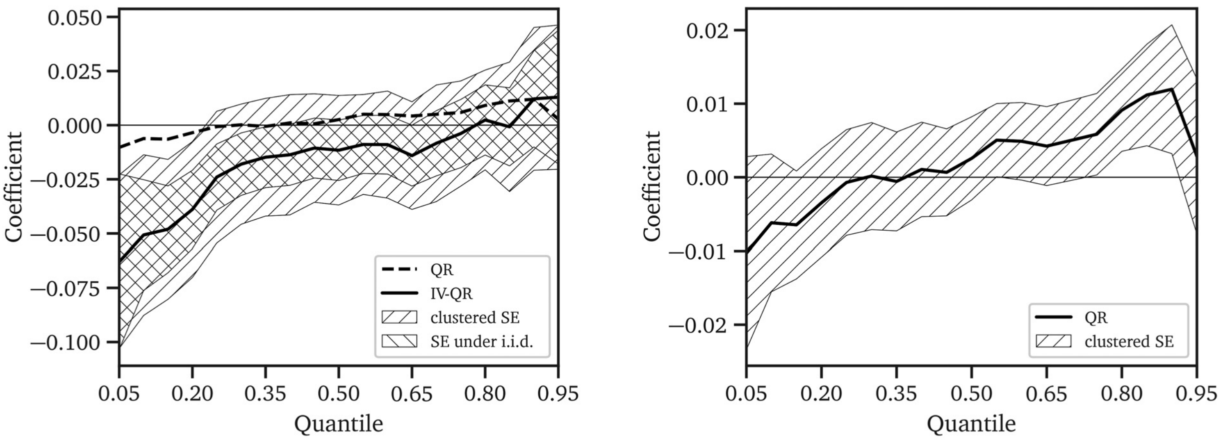
(a) 90% confidence intervals for IV-QR estimates

(b) 90% confidence intervals for QR estimates

Fig. 1. Impact of female labor supply on earnings of women, 194050, specification with $X_{1,ist}$, $X_{2,ist}$, $X_{3,st}$ and $X_{4,st}$ as controls

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

Null hypothesis	Impact of female labor supply on female earnings		Impact of female labor supply on male earnings	
	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

Fig. 2. Impact of female labor supply on earnings of men, 194050, specification with $X_{1,ist}$, $X_{2,ist}$, $X_{3,st}$ and $X_{4,st}$ as controls

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 negative in explaining men's earnings, and is inversely related to the quantile index (Figure 2). In other words, the negative effect is smaller in absolute terms for higher-wage male workers. The standard errors for the coefficient have downward bias under the Chernozhukov and Hansen (2006) approach in comparison with the estimator for clustered standard data (Table B.2 of the Appendix).

Finally, we carry out inference on the entire instrumental variable quantile regression process using 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).

7. Conclusion

We have examined robust inference in a 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 with the behavior of the Chernozhukov and Hansen (2006) estimator. Further, we give recommendations for practitioners who deal with data that have non-i.i.d. clusters. 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.

Declaration of Competing Interest

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

Acknowledgements

The paper was prepared in the framework of the HSE University's Basic Research Program.

Our thanks for helpful comments to the Editor, Erricos John Kontoghiorghes, to two anonymous referees, and participants of the North American Summer Meeting of the Econometric Society (2022), the Asian Meetings of the Econometric Society (China and Tokyo, 2022), the 2nd International Conference on Econometrics and Business Analytics (2022), and the International Conference on Modern Econometric Tools and Applications (2022).

We are grateful to an anonymous referee for suggesting that a discussion of the impact of violating condition R1 (by relaxing the assumptions that all clusters are of equal size and that cluster size is fixed) should be added to the text.

Appendix A. Proofs

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

Proof of Theorem 2. Define for $W_k = (Y_k, D_k, X_k, Z_k)$, $\vartheta \equiv (\beta, \gamma)$ and $\varphi_\tau(u) = I(u < 0) - \tau$

$$\hat{f}(W_k, \alpha, \vartheta, \tau) \equiv \varphi_\tau(Y_k - D_k'\alpha - X_k'\beta - \hat{\Phi}_k(\tau)'\gamma)\hat{\Psi}_k(\tau),$$
$$f(W_k, \alpha, \vartheta, \tau) \equiv \varphi_\tau(Y_k - D_k'\alpha - X_k'\beta - \Phi_k(\tau)'\gamma)\Psi_k(\tau),$$

$\Psi_k(\tau) \equiv V_k(\tau) \cdot [\Phi_k(\tau)', X_k']'$, $\Phi_k(\tau) = \Phi_k(\tau, X_k, Z_k)$, $V_k(\tau) = V_k(\tau, X_k, Z_k)$, $\hat{\Psi}_k(\tau) \equiv \hat{V}_k(\tau) \cdot [\hat{\Phi}_k(\tau)', X_k']'$, $\hat{\Phi}_k(\tau) = \hat{\Phi}_k(\tau, X_k, Z_k)$, $\hat{V}_k(\tau) = \hat{V}_k(\tau, X_k, Z_k)$; for $\rho_\tau(u) = (\tau - I(u < 0))u = -\varphi_\tau(u)u$

$$\hat{g}(W_k, \alpha, \vartheta, \tau) \equiv \rho_\tau(Y_k - D_k'\alpha - X_k'\beta - \hat{\Phi}_k(\tau)'\gamma)\hat{V}_k(\tau),$$
$$g(W_k, \alpha, \vartheta, \tau) \equiv \rho_\tau(Y_k - D_k'\alpha - X_k'\beta - \Phi_k(\tau)'\gamma)V_k(\tau).$$

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

$$Q_N(\alpha, \vartheta, \tau) \equiv \mathbb{E}_N\left(\sum_{k=1}^K \hat{g}(W_k, \alpha, \vartheta, \tau)\right) = \frac{1}{N}\sum_{i=1}^N\sum_{k=1}^K \hat{g}(W_{ik}, \alpha, \vartheta, \tau),$$
$$Q(\alpha, \vartheta, \tau) \equiv E\left(\sum_{k=1}^K g(W_k, \alpha, \vartheta, \tau)\right),$$

and

$$\hat{\vartheta}(\alpha, \tau) \equiv (\hat{\beta}(\alpha, \tau)', \hat{\gamma}(\alpha, \tau)') = \operatorname*{argmin}_{\vartheta \in \mathcal{B}\times\mathcal{G}} Q_N(\alpha, \vartheta, \tau),$$

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

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

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

Step3 (Asymptotics): We adapt Step 3 in the proof of Theorem 3 in Chernozhukov and Hansen (2006), pp.516–518 as follows:

- substitute $(W_1,\dots,W_K)$ for W;
- substitute $\sum_{k=1}^K \hat{g}(W_k,\alpha,\beta,\gamma,\tau)$ for $\hat{g}(W,\alpha,\beta,\gamma,\tau)$ and $\sum_{k=1}^K g(W_k,\alpha,\beta,\gamma,\tau)$ for $g(W,\alpha,\beta,\gamma,\tau)$;
- substitute $\sum_{k=1}^K \hat{f}(W_k,\alpha,\beta,\gamma,\tau)$ for $\hat{f}(W,\alpha,\beta,\gamma,\tau)$ and $\sum_{k=1}^K f(W_k,\alpha,\beta,\gamma,\tau)$ for $f(W,\alpha,\beta,\gamma,\tau)$;
- substitute $\sum_{k=1}^K \hat{\varphi}_\cdot(Y_k - D_k'\alpha(\cdot) - X_k'\beta - \Phi_k(\cdot)'\gamma)\Psi_k(\cdot)$ for $\sum_{k=1}^K \hat{\varphi}_\cdot(Y - D'\alpha(\cdot) - X'\beta - \Phi(\cdot)'\gamma)\Psi(\cdot)$ and $\sum_{k=1}^K \hat{\varphi}_\cdot(Y_k - D_k'\alpha - X_k'\beta(\cdot))\Psi_k(\cdot)$ for $\sum_{k=1}^K \hat{\varphi}_\cdot(Y - D'\alpha - X'\beta(\cdot))\Psi(\cdot)$.

Since the sample of vectors $(W_1,\dots,W_K)$ is i.i.d., it is possible to apply all steps of the proof of Theorem 3 in Chernozhukov and Hansen (2006). The last step, i.e. obtaining the asymptotic distribution of the Gaussian process $\mathbb{G}_N\left(\sum_{k=1}^K f(W_k,\alpha(\cdot),\vartheta(\alpha(\cdot),\cdot),\cdot)\right)$, is carried out using Lemma A.1. □

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

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

- $\tau \mapsto \hat{z}(W_k,\tau)$, $k = 1,\dots,K$,
- a Donsker set of functions $\{\xi(W_k,\tau), \tau\in\mathcal{T}, \xi\in\Xi, k=1,\dots,K\}$,
- the empirical process $(\tau,\xi)\mapsto \mathbb{G}_N(\xi(\tau)) \equiv 1/\sqrt{N}\sum_{i=1}^N\sum_{k=1}^K(\xi(W_{ik},\tau) - E\xi(W_{ik},\tau))$,
- its subsample realizations $(\tau,\xi)\mapsto \mathbb{G}_{j,b,N}(\xi(\tau)) \equiv 1/\sqrt{b}\sum_{i\in I_j}\sum_{k=1}^K(\xi(W_{ik},\tau) - E\xi(W_{ik},\tau))$, $j=1,\dots,B_N$. 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 A.1 (Stochastic expansion)**.** *Under Assumption 1, the following statements are true.*

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

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

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

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

- substitute $(W_1,\dots,W_K)$ for W;
- substitute $\sum_{k=1}^K \hat{g}(W_k,\alpha,\beta,\gamma,\tau)$ for $\hat{g}(W,\alpha,\beta,\gamma,\tau)$ and $\sum_{k=1}^K g(W_k,\alpha,\beta,\gamma,\tau)$ for $g(W,\alpha,\beta,\gamma,\tau)$;
- substitute $\sum_{k=1}^K \hat{f}(W_k,\alpha,\beta,\gamma,\tau)$ for $\hat{f}(W,\alpha,\beta,\gamma,\tau)$ and $\sum_{k=1}^K f(W_k,\alpha,\beta,\gamma,\tau)$ for $f(W,\alpha,\beta,\gamma,\tau)$.

Since the sample of vectors $(W_1,\dots,W_K)$ is i.i.d., it is possible to apply all the steps of the proof of Lemma B.2 in Chernozhukov and Hansen (2006). Note that the expression for the covariance function becomes

$$S(\tau,\tau') = E(\mathbb{G}(\tau)\mathbb{G}(\tau')') = E\left(\left(\sum_{k=1}^{K}\varphi_\tau(Y_k - D_k'\alpha(\tau) - X_k'\beta(\tau))\Psi_k(\tau,X_k,Z_k)\right)\times\left(\sum_{s=1}^{K}\varphi_{\tau'}(Y_s - D_s'\alpha(\tau') - X_s'\beta(\tau'))\Psi_s(\tau',X_s,Z_s)\right)'\right) = E\left(\sum_{k=1}^{K}\sum_{s=1}^{K} l_k(\tau,\theta(\tau))l_s(\tau',\theta(\tau'))\Psi_k(\tau)\Psi_s(\tau')'\right).$$

where $l_k(\tau,\theta(\tau)) = \varphi_\tau(Y_k - D_k'\alpha(\tau) - X_k'\beta(\tau))$. Unlike Chernozhukov and Hansen (2006), this expression cannot be further simplified. □

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

$$l_{ik}(\tau,\theta(\tau)) = (\tau - I(Y_{ik} < D_{ik}'\alpha(\tau) - X_{ik}'\beta(\tau))), \quad \Psi_{ik}(\tau) = V_{ik}(\tau)[\Phi_{ik}(\tau)', X_{ik}']'. \tag{A.1}$$

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_{ik} is a function of only X_{ik} and Z_{ik}, 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 $l_{ik}(\cdot)$ defined in (Appendix A.1), $z_{ik}(\tau) = R(\tau)(J(\tau)^{-1}l_{ik}(\tau,\theta(\tau))\Psi_{ik}(\tau) - J(1/2)^{-1}l_{ik}(1/2,\theta(1/2))\Psi_{ik}(1/2))$. In other words, $d_{ik}(\tau, r(\tau)) = J(1/2)^{-1}l_{ik}(1/2,\theta(1/2))\Psi_{ik}(1/2)$ 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),

$$\sqrt{N}(\hat\vartheta(\cdot) - \vartheta(\cdot)) = -H(\cdot)^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{k=1}^{K} d_{ik}(\cdot,\vartheta(\cdot)) + o_p(1),$$

where $d_{ik}(\tau,\vartheta(\tau)) = (\tau - I(Y_{ik} < \tilde X_{ik}'\vartheta(\tau)))\tilde X_{ik}$ for $\tilde X_{ik} = [D_{ik}', X_{ik}']'$, $H(\tau) = E\left(\sum_{k=1}^{K} f_{Y_k|\tilde X_k}(\tilde X_k'\vartheta(\tau))\tilde X_k\tilde X_k'\right)$. Therefore, $z_{ik} = R(\tau)(J(\tau)^{-1}l_{ik}(\tau,\theta(\tau))\Psi_{ik}(\tau) - H(\tau)^{-1}d_{ik}(\tau,\vartheta(\tau))$.

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

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

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

Appendix B. Results of the empirical analysis

Table B.1
Impact of female labor supply (weeks worked per woman) 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. The dependent variable is log weekly earnings. The specification includes age structure, state of birth, share of farmers, share of nonwhites, and average education. Standard errors are in parentheses. *, **, *** mean significance at 10%, 5% and 1% respectively.

Table B.2
Impact of female labor supply (weeks worked per woman) 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. The dependent variable is log weekly earnings. The specification includes age structure, state of birth, share of farmers, share of nonwhites, and average education. Standard errors are in parentheses. *, **, *** mean significance at 10%, 5% and 1% respectively.

References

Abadie, A., 1997. Changes in Spanish labor income structure during the 1980's: A quantile regression approach. Investigaciones Economicas 21 (2), 253–272.
Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association 97 (457), 284–292.
Abowd, J.M., Kramarz, F., Margolis, D.N., 1999. High wage workers and high wage firms. Econometrica 67 (2), 251–333.
Acemoglu, D., Autor, D.H., 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., 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., Fernández-Val, I., 2006. Quantile regression under misspecification, with an application to the US wage structure. Econometrica 74 (2), 539–563.
Boehnke, J., Gay, V., 2020. The missing men World War I and female labor force participation. Journal of Human Resources 57 (4), 1209–1241.
Borovičková, K., Shimer, R., 2017. High wage workers work for high wage firms. Technical Report. 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., 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., Miller, D.L., 2011. Robust inference with clustered data. In: Ullah, A., Giles, D.E. (Eds.), Handbook of Empirical Economics and Finance. Chapman and Hall/CRC, pp. 1–28.
Chernozhukov, V., Fernández-Val, I., 2005. Subsampling inference on quantile regression processes. Sankhyā: The Indian Journal of Statistics 253–276.
Chernozhukov, V., Hansen, C., 2004. Instrumental variable quantile regression. Working Paper, available at http://www.mit.edu/~vchern.
Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. Econometrica 73 (1), 245–261.
Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effect models. Journal of Econometrics 132 (2), 491–525.
Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: A robust inference approach. Journal of Econometrics 142 (1), 379–398.
Chernozhukov, V., Hansen, C., 2013. Quantile models with endogeneity. Annual Reviews of Economics 5 (1), 57–81.
Chernozhukov, V., Hansen, C., Wüthrich, K., 2017. Instrumental variable quantile regression. In: Koenker, R., Chernozhukov, V., He, X., Peng, L. (Eds.), Handbook of Quantile Regression. Chapman and Hall/CRC, pp. 119–143.
Chetverikov, D., Larsen, B., Palmer, C., 2016. IV quantile regression for group-level treatments, with an application to the distributional effects of trade. Econometrica 84 (2), 809–833.
Fernández, R., Fogli, A., Olivetti, C., 2004. Mothers and sons: Preference formation and female labor force dynamics. The Quarterly Journal of Economics 119 (4), 1249–1299.
Geraci, M., 2019. Additive quantile regression for clustered data with an application to children's physical activity. Journal of the Royal Statistical Society. Series C, Applied Statistics 68 (4), 1071.
Hagemann, A., 2017. Cluster-robust bootstrap inference in quantile regression models. Journal of the American Statistical Association 112 (517), 446–456.
Harding, M., Lamarche, C., 2014. Estimating and testing a quantile regression model with interactive effects. Journal of Econometrics 178, 101–113.
Harding, M., Lamarche, C., 2017. Penalized quantile regression with semiparametric correlated effects: An application with heterogeneous preferences. Journal of Applied Econometrics 32 (2), 342–358.
He, X., 2017. Resampling methods. In: Koenker, R., Chernozhukov, V., He, X., Peng, L. (Eds.), Handbook of Quantile Regression. Chapman and Hall/CRC, pp. 7–19.
Heckman, J.J., 1978. A partial survey of recent research on the labor supply of women. The American Economic Review 68 (2), 200–207.
Huber, M., Wüthrich, K., 2019. Local average and quantile treatment effects under endogeneity: a review. Journal of Econometric Methods 8 (1), 20170007.
Juhn, C., 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., Heckman, J., 1986. Female labor supply: A survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics. North Holland Amsterdam, pp. 103–204.
Koenker, R., 2005. Quantile Regression. Cambridge University Press.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Koenker, R., Hallock, K., 2001. Quantile regression. Journal of Economic Perspectives 15 (4), 143–156.
Koenker, R., Xiao, Z., 2002. Inference on the quantile regression process. Econometrica 70 (4), 1583–1612.
Mincer, J., 1962. Labor force participation of married women: A study of labor supply. In: Aspects of Labor Economics. Princeton University Press, pp. 63–105.
Parente, P.M., 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., Tzannatos, Z., 1989. Female labor force participation: An international perspective. The World Bank Research Observer 4 (2), 187–201.
van Reenen, J., 2011. Wage inequality, technology and trade: 21st century evidence. Labour Economics 18 (6), 730–741.
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.
Wooldridge, J., 2007. Quantile methods. NBER Summer Institute, Lecture Notes 14, available at http://www.nber.org/WNE/lect_14_quantile.pdf.
Zhang, Y., Wang, H.J., Zhu, Z., 2019. Quantile-regression-based clustering for panel data. Journal of Econometrics 213 (1), 54–67.