ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 728 June 2013

Quantile regression with clustered data

Paulo M.D.C. Parente & J.M.C. Santos Silva

Quantile regression with clustered data*

Paulo M.D.C. Parente[†] J.M.C. Santos Silva[‡]

June 20, 2013

Abstract

We show that the quantile regression estimator is consistent and asymptotically normal when the error terms are correlated within clusters but independent across clusters. A consistent estimator of the covariance matrix of the asymptotic distribution is provided and we propose a specification test capable of detecting the presence of intra-cluster correlation. A small simulation study illustrates the finite sample performance of the test and of the covariance matrix estimator.

JEL classification code: C12, C21, C23.

Key words: Clustered standard errors, Moulton Problem, Panel data, Specification testing.

1. INTRODUCTION

In many applications inference is performed using micro data sampled from a number of groups or clusters; typically it is assumed that observations from different groups are conditionally independent but intra-cluster correlation is not ruled out. In this context valid inference can be performed by using a consistent estimator of the covariance matrix

*Santos Silva gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

[†]Department of Economics, University of Exeter and CReMic. Email: P.M.Parente@exeter.ac.uk.

[‡]Department of Economics, University of Essex and CEMAPRE. E-mail: jmcss@essex.ac.uk.

of the asymptotic distribution of the estimator when intra-cluster correlation is allowed for. This was the motivation for the work of Liang and Zeger (1986) and Arellano (1987), who extended the results in White (1984) to derive covariance matrix estimators that are valid when there is heteroskedasticity and intra-cluster correlation.

Although these methods were initially developed with panel data in mind they are also useful with dyadic data and even in cross-sections, where the clusters can be defined for example by regions or industries. A well-known example where it is important to allow for intra-cluster correlation occurs when cross-sectional regressions using micro data contain some regressors observed only at a more aggregate level; see Moulton (1986, 1990). The ubiquitous use of the so-called clustered standard errors in applied econometrics shows how prevalent this kind of situation is (see Cameron and Miller, 2011, for a recent survey on inference with this kind of data).

The asymptotic distribution of maximum likelihood and least squares estimators allowing for intra-cluster correlation have been widely studied, and popular software packages now implement covariance matrix estimators that are valid in this case (see, e.g., Rogers, 1993). However, it appears that so far the case of quantile regression has not been considered. This is unfortunate because quantile regression can suffer from the "Moulton problem" and because pooled quantile regression and correlated random effects quantile regression are gaining popularity in applied panel-data econometrics (see, e.g., the influential paper by Abrevaya and Dahl, 2008). In these cases practitioners perform inference by using bootstrap procedures but we are not aware of any formal proof that these estimators are consistent in this context. Moreover, bootstrapping quantile regression is somewhat impractical when the problem involves very large samples and many regressors because in this case the computation of the bootstrap covariance matrix using a reasonable number of bootstraps is still very time consuming.

In this paper we extend the results of Kim and White (2003) and show that the traditional quantile regression estimator (Koenker and Bassett, 1978) is consistent and asymptotically normal when there is within-cluster correlation of the error terms. Additionally we present a consistent estimator for the covariance matrix of the asymptotic distribution of the quantile regression estimator with intra-cluster correlation and propose a specification test capable of detecting the presence of this kind of correlation. A small simulation study is used to illustrate the finite sample performance of the proposed methods. An Appendix provides the proofs of all theorems.

2. QUANTILES WITH CLUSTERS

2.1. Set-up and asymptotic properties

Consider the case in which the researcher is interested in estimating the θ-th quantile of the conditional distribution of y given x, denoted $Q_\theta(y|x)$, and assume that

$$Q_\theta(y|x) = x'\beta_0,$$

where x and β_0 are $k \times 1$ vectors and for simplicity we omit that the vector of parameters is indexed by θ. We are interested in the case where estimation is to be performed using a sample $\{(y_{gi}, x_{gi}),\ \ g = 1, \ldots, G,\ i = 1, \ldots, n_g\}$, where g indexes a set of G predefined groups or clusters, each with n_g elements. That is, we are interested in estimating

$$\begin{aligned} y_{gi} &= x'_{gi}\beta_0 + u_{gi}, && (1) \\ \Pr(u_{gi} &\leq 0|x_{gi}) = \theta. && (2) \end{aligned}$$

In what follows we will consider the properties of the estimator of β_0, with n_g fixed and $G \to \infty$, for the case in which the disturbances u_{gi} are assumed to be uncorrelated across clusters but are permitted to be correlated within clusters. For simplicity, and without loss of generality, we consider only the case where $n_g = n$.

The quantile regression estimator for clustered data is defined by

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^k} \frac{1}{G} \sum\nolimits_{g=1}^{G} \sum\nolimits_{i=1}^{n} \rho_\theta\left(y_{gi} - x'_{gi}\beta\right),$$

where $\rho_\theta(a) = a\left(\theta - \mathrm{I}[a < 0]\right)$ is known as the check function and $\mathrm{I}[e]$ is the indicator function of the event e.

The consistency of $\hat{\beta}$ can be proved under the following assumption:

Assumption 1 ***(a)*** *Let* $x_g = (x'_{g1}, \ldots, x'_{gn})'$ *and* $y_g = (y_{g1}, \ldots, y_{gn})'$*; the data* $\left\{(y_g, x'_g)'\right\}_{g=1}^{G}$ *are independent and identically distributed across* g*;* ***(b)*** $\mathrm{E}[\|x_{gi}\|] < \infty$*;* ***(c)*** *The conditional distribution of* u_{gi} *given* x_{gi}*,* $\mathrm{F}(u_{gi}|x_{gi})$*, has a unique* θ*-th conditional quantile at* $u_{gi} = 0$*.*

Assumption 1 (a) is made for simplicity but can be relaxed to allow some dependence across g, although some of the remaining regularity conditions would have to be strengthened. Assumption 1 (b) and (c) are standard (see Theorem 2.11 of Newey and McFadden, 1994, p. 2140) but can be relaxed (see Koenker, 2005, p. 118).

We are now able to establish consistency.

Theorem 1 *Under Assumption 1,* $\hat{\beta} \xrightarrow{p} \beta_0$*.*

To prove asymptotic normality we need the following additional assumption:

Assumption 2 ***(a)*** $\mathrm{F}(u_{gi}|x_{gi})$ *is absolutely continuous with continuous density* $f(u|x_{gi})$ *that satisfies* $f(0|x_{gi}) < f_1 < \infty$ *and* $f(0|x_{gi}) > 0$ *for all* x_{gi} *and for a positive constant* f_1*;* ***(b)*** $\mathrm{E}[\|x_{gi}\|^3] < \infty$ *for all* i *and* g*;* ***(c)*** *The matrix* $\mathrm{A} = \mathrm{E}[(\sum_{i=1}^{n}\sum_{j=1}^{n} x_{gi}x'_{gj}\psi_\theta(u_{gi})\psi_\theta(u_{gj}))]$*, where* $\psi_\theta(a) = \theta - \mathrm{I}\left[a < 0\right]$*, is positive definite;* ***(d)*** *The matrix* $\mathrm{B} = \sum_{i=1}^{n} \mathrm{E}[x_{gi}x'_{gi}f(0|x_{gi})]$ *is positive definite.*

Assumption 2 (a), (c), and (d) are standard (see Koenker, 2005, p. 120). Assumption 2 (b) is stronger than that considered by Koenker (2005, p. 120) in the standard i.i.d. setting but coincides with that required by Powell (1984) and Kim and White (2003).

In the Appendix we prove the following theorem:

Theorem 2 *Under Assumptions 1 and 2 we have*

$$\sqrt{G}\left(\hat{\beta} - \beta_0\right) \xrightarrow{D} \mathcal{N}\left(0, \Omega\right),$$

with $\Omega = \mathrm{B}^{-1}\mathrm{A}\mathrm{B}^{-1}$*.*

2.2. Consistent covariance matrix estimation

For the estimator $\hat{\beta}$ to be useful it is necessary to have a consistent estimator of Ω. As mentioned before, practitioners often use bootstrap procedures to estimate Ω (see, e.g., Abrevaya and Dahl, 2008) but it is not clear that this estimator is valid in the context considered here. More importantly, bootstrap methods are still somewhat impractical in realistic applications, especially in models for which quantile regression takes many iterations to converge. In what follows we provide consistent estimators of A and B that can be used to obtain a consistent estimator of Ω.

A consistent estimator of

$$\mathrm{A} = \mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} x_{gi}x'_{gj}\psi_\theta(u_{gi})\psi_\theta(u_{gj})\right]$$

is given by

$$\hat{\mathrm{A}} = \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\sum_{j=1}^{n} x_{gi}x'_{gj}\psi_\theta(\hat{u}_{gi})\psi_\theta(\hat{u}_{gj}),$$

where $\hat{u}_{gi} = y_{gi} - x'_{gi}\hat{\beta}$. Given that $\hat{\beta} \xrightarrow{p} \beta_0$, Assumptions 1 and 2, Loève's c_r inequality (Davidson, 1994, p. 140), and a uniform weak law of large numbers, imply that $\hat{\mathrm{A}} \xrightarrow{p} \mathrm{A}$.

A more challenging task is to obtain a consistent estimator of B. Following Powell (1984, 1986) and Kim and White (2003) we consider the estimator

$$\hat{\mathrm{B}} = \frac{1}{2\hat{c}_G G}\sum_{g=1}^{G}\sum_{i=1}^{n} \mathrm{I}\left[|\hat{u}_{gi}| \leq \hat{c}_G\right] x_{gi}x'_{gi},$$

where the bandwidth $\hat{c}_G$ may be a function of the data. To establish the consistency of $\hat{\mathrm{B}}$ we require the following additional assumption which was also considered by Powell (1984) and Kim and White (2003):

Assumption 3 ***(a)*** *$f(a|x_{gi}) < f_1 < \infty$ for all a and x_g and for a positive constant f_1;* ***(b)*** *There is a stochastic sequence $\hat{c}_G$ and a non-stochastic sequence c_G such that $\hat{c}_G/c_G \xrightarrow{p} 1$, $c_G = o(1)$ and $c_G^{-1} = o(\sqrt{G})$.*

The following theorem gives the desired result:

Theorem 3 *Under assumptions 1, 2, and 3,* $\hat{\mathrm{B}} \xrightarrow{p} \mathrm{B}$.

In order to implement this estimator of B it is necessary to define a practical method of choosing the bandwidth $\hat{c}_G$. The solution used in the simulations presented in Section 3 is based on the method described by Koenker (2005, p. 81). In particular, we define

$$\hat{c}_G = \kappa \left[\Phi^{-1}\left(\theta + h_{nG}\right) - \Phi^{-1}\left(\theta - h_{nG}\right)\right],$$

where h_{nG} is (see Koenker, 2005, p. 140 or Koenker and Machado, 1999, p. 1301)

$$h_{nG} = (nG)^{-1/3}\left(\Phi^{-1}\left(1 - \frac{0.05}{2}\right)\right)^{2/3}\left(\frac{1.5\left(\phi\left(\Phi^{-1}\left(\theta\right)\right)\right)^2}{2\left(\Phi^{-1}\left(\theta\right)\right)^2 + 1}\right)^{1/3},$$

and κ is a robust estimate of scale. After some experimentation, we decided to define κ as the MAD (median absolute deviation) of the θ-th quantile regression residuals.[1]

3. A SPECIFICATION TEST

In the spirit of White (1980) and Kim and White (2003), in this section we propose a simple test to check whether the use of the covariance matrix estimator obtained in Subsection 2.2 is necessary. In particular, we derive a test based on the moment condition

$$\mathrm{E}\left[\sum\nolimits_{i=1}^{n}\left(\sum\nolimits_{j=1}^{n} z_{gi}z_{gj}\psi_\theta(u_{gi})\psi_\theta(u_{gj}) - \psi_\theta(u_{gi})^2 z_{gi}^2\right)\right] = 0, \tag{3}$$

where $z_{gi} = g(x_{gi})$ and $g(\cdot)$ is a scalar function. When (3) holds for z_{gi} defined as an arbitrary element of x_{gi}, the matrix Ω reduces to the covariance matrix obtained by Chamberlain (1994) and Kim and White (2003) for the case where the errors u_{gi} and u_{gj} are uncorrelated but may be heteroskedastic.

Further insights into the nature of (3) can be gained by noting that it is implied by the following twin sets of moment conditions:

$$\mathrm{E}[z_{gi}z_{gj}\left(\theta - \mathrm{I}\left[u_{gi} < 0\right]\right)] = 0, \tag{4}$$

$$\mathrm{E}[z_{gi}z_{gj}\left(\theta^2 - \mathrm{I}\left[u_{gi} < 0\right]\mathrm{I}\left[u_{gj} < 0\right]\right)] = 0. \tag{5}$$

[1] It is customary to multiply MAD by 1.4826 because in the normal distribution 1.4826MAD is approximately equal to the standard deviation. However, some preliminary simulations revealed that the results were substantially better when the scaling factor was not used.

The moment conditions in (4) are closely related to the first order conditions of the estimator and can be used to test the correct specification of (1) and (2) by checking for the omission of the variables $z_{gi}z_{gj}$. The set of moment conditions in (5) will hold if u_{gi} and u_{gj} are independent and are of particular interest in the context we are considering. Therefore, a test based on (3) is both a test of the validity of (2) and a test of the independence between u_{gi} and u_{gj}.

Formally, we propose a test for the joint null hypothesis

$$H_0 : \begin{cases} \mathrm{F}_i\left(a|x_g\right) = \mathrm{F}\left(a|x_{gi}\right) & \text{for all } i, \\ \mathrm{F}_{i,j}\left(a,b|x_g\right) = \mathrm{F}_i\left(a|x_g\right) \times \mathrm{F}_j\left(b|x_g\right) & \text{for } i \neq j, \end{cases} \tag{6}$$

where $\mathrm{F}_i\left(a|x_g\right) = \Pr\left(u_{gi} \leq a|x_g\right)$ and $\mathrm{F}_{i,j}\left(a,b|x_g\right) = \Pr\left(u_{gi} \leq a, u_{gj} \leq b|x_g\right)$,[2] based on the following statistic, which is based on the sample analog of (3):

$$\mathcal{T} = \frac{1}{\sqrt{G}} \sum\nolimits_{g=1}^{G} \sum\nolimits_{i=1}^{n} \left(\sum\nolimits_{j=1}^{n} z_{gi}z_{gj}\psi_\theta(\hat{u}_{gi})\psi_\theta(\hat{u}_{gj}) - \psi_\theta(\hat{u}_{gi})^2 z_{gi}^2 \right).$$

In order to obtain the asymptotic distribution of $\mathcal{T}$ we define

$$\begin{aligned} D &= \mathrm{Var}\left(\sum\nolimits_{i=1}^{n} \left(\sum\nolimits_{j=1}^{n} z_{gi}z_{gj}\psi_\theta(u_{gi})\psi_\theta(u_{gj}) - \psi_\theta(u_{gi})^2 z_{gi}^2 \right) \right), \\ &= \mathrm{E}\left[\left(\sum\nolimits_{i=1}^{n} \left[\sum\nolimits_{j=1}^{n} z_{gi}z_{gj}\psi_\theta(u_{gi})\psi_\theta(u_{gj}) - \psi_\theta(u_{gi})^2 z_{gi}^2 \right] \right)^2 \right], \\ &= 2 \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1, i\neq j}^{n} \mathrm{E}\left[z_{gj}^2 z_{gi}^2 \theta^2 (1-\theta)^2 \right], \end{aligned}$$

and make the following additional assumption:

Assumption 4 ***(a)*** $\mathrm{E}[\|z_{gi}\|^{4+\eta}] < \infty$ *for some* $\eta > 0$*;* ***(b)*** D *is strictly positive.*

Theorem 4 *Under Assumptions 1, 2, and 4, and* H_0*,* $\mathcal{T} \xrightarrow{D} \mathcal{N}(0, D)$*.*

In practice D can be consistently estimated by

$$\hat{D} = \frac{2}{G} \sum\nolimits_{g=1}^{G} \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1, i\neq j}^{n} z_{gj}^2 z_{gi}^2 \theta^2 (1-\theta)^2,$$

[2] Notice that H_0 implies (3) but the reverse is not true; we will derive the distribution of the test statistic under H_0.

and therefore a test based on $\mathcal{T}$ is very easy to implement.[3] However, the test requires the choice of z_{gi}, which plays an important role in the interpretation of its outcome. In the simulation study to be presented below we focus on the case where $z_{gi} = 1$ and the model has an intercept. In this case the sample analog of (4) will necessarily hold because it is implied by the first order conditions of the estimator and consequently the test has non-trivial power only against intra-cluster correlation; i.e., only (5) is being tested. Therefore, this particular form of the test is the quantile regression analog of the heteroskedasticity and non-normality robust version of the Breusch and Pagan (1980) error components test introduced by Wooldridge (2002, p. 265).[4]

4. SIMULATION EVIDENCE

In this section we present the results of a small simulation study on the performance of the covariance matrix estimator proposed in Section 2 and of the test introduced in Section 3.

The simulated data were generated as

$$\begin{aligned} y_{gi} &= \gamma_0 + \gamma_1 x_{gi} + \left(x_{gi}^h\right) u_{gi}, \\ x_{gi} &= \xi_g + \varepsilon_{gi}, \\ u_{gi} &= \alpha_g + v_{gi}, \\ i &= 1, \ldots, n;\ g = 1, \ldots, G, \end{aligned}$$

where $\gamma_0 = \gamma_1 = 0$, $h \in \{0, 1\}$ is a parameter controlling the presence of heteroskedasticity, $n \in \{2, 5\}$, $G \in \{100, 1000, 10000\}$, $\xi_g \sim \chi^2_{(1)}$, $\varepsilon_{gi} \sim \chi^2_{(2)}$, $\alpha_g \sim \chi^2_{(d_\alpha)}$, $v_{gi} \sim \chi^2_{(d_v)}$, and ξ_g, ε_{gi}, α_g, and v_{gi} are independent. We considered cases with and without intra-cluster correlation in u_{gi}: in the first case we set $d_v = 2$ and $d_\alpha = 1$, and in the second case $d_v = 3$ and $d_\alpha = 0$. Therefore, in all cases $u_{gi} \sim \chi^2_{(3)}$ and $x_{gi} \sim \chi^2_{(3)}$. For each of the designs, ξ_g, ε_{gi},

[3] Note that when $n_g = n$ and $z_{gi} = 1$ we have that $\hat{D} = D = 2\theta^2(1-\theta)^2 n(n-1)$.

[4] Like Breusch and Pagan (1980) and Wooldridge (2002), we consider only the univariate case. However, following Kim and White (2003), it is also possible to develop a multivariate version of the test.

α_g, and v_{gi} were newly generated for each of the 10000 replications used in the experiment. The performance of the covariance estimator is evaluated for $\theta \in \{0.25, 0.50, 0.75\}$ by estimating the θ-th quantile regression of y_{gi} on x_{gi} and a constant and testing whether the slope parameter of the regression is equal to its true value $(\gamma_1 + hQ_\theta(u_{gi}))$. All the simulations where preformed in Stata 11 (StataCorp., 2009) using the command `qreg2` (Machado, Parente, and Santos Silva, 2013) that implements both the covariance matrix estimator and the test studied here.

Tables 1 and 2 give the rejection frequencies of the null hypothesis at the 5% level; we report the results obtained using both the covariance matrix estimator proposed in Section 2 and a covariance matrix estimator obtained using 100 cluster-bootstraps. In evaluating the results of these experiments we will follow Cochran (1952), who suggested that a test can be regarded as robust relative to a nominal level of 5% if its actual significance level is between 4% and 6%. Given the number of replicas used in these experiments, we will consider that estimated rejection frequencies within the range 3.62% to 6.47% provide evidence consistent with the robustness of the test.

In line with the findings for the case of independent observations reported by Buchinsky (1995), our results show that the bootstrap estimator performs well in most of the cases considered. As for the results based on $\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$, we see that there is some tendency to overreject the null when $n = 100$ and a slight tendency to under-reject for larger samples when the errors are heteroskedastic ($h = 1$) and there is no intra-cluster correlation ($d_v = 3$ and $d_\alpha = 0$). Crucially, the results obtained using $\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$ are quite reasonable when they are more interesting, i.e., when the samples are large and the errors actually have intra-cluster correlation ($d_v = 2$ and $d_\alpha = 1$).[5]

Overall, the results obtained when using $\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$ to estimate Ω are quite encouraging, suggesting that this estimator can be used in situations where the bootstrap is impractical.[6]

[5]We performed an additional set of experiments with a similar design but with normally distributed errors and the results for the tests based on $\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$ were slightly better than those reported here.

[6]Notice that computing the bootstrap results presented here is approximately 100 times slower than computing the results based on $\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$.

Table 1: Rejection frequencies at the 5% level ($d_v = 3$ and $d_\alpha = 0$)

			$\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$			Bootstrap		
h	n	G	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$
0	2	100	0.0383	0.0675	0.1069	0.0508	0.0566	0.0585
		1000	0.0541	0.0580	0.0641	0.0570	0.0571	0.0567
		10000	0.0474	0.0515	0.0543	0.0515	0.0530	0.0578
	5	100	0.0548	0.0716	0.0859	0.0576	0.0599	0.0580
		1000	0.0511	0.0540	0.0517	0.0567	0.0576	0.0492
		10000	0.0504	0.0522	0.0502	0.0527	0.0554	0.0500
1	2	100	0.0621	0.0684	0.1055	0.0567	0.0609	0.0624
		1000	0.0340	0.0442	0.0551	0.0507	0.0590	0.0544
		10000	0.0338	0.0374	0.0461	0.0512	0.0535	0.0564
	5	100	0.0488	0.0546	0.0791	0.0574	0.0606	0.0603
		1000	0.0356	0.0330	0.0457	0.0558	0.0479	0.0481
		10000	0.0396	0.0370	0.0416	0.0561	0.0517	0.0493

Table 2: Rejection frequencies at the 5% level ($d_v = 2$ and $d_\alpha = 1$)

			$\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$			Bootstrap		
h	n	G	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$
0	2	100	0.0419	0.0677	0.1102	0.0545	0.0575	0.0604
		1000	0.0482	0.0557	0.0612	0.0528	0.0585	0.0550
		10000	0.0562	0.0554	0.0554	0.0615	0.0583	0.0566
	5	100	0.0598	0.0740	0.0920	0.0632	0.0610	0.0644
		1000	0.0520	0.0558	0.0616	0.0553	0.0567	0.0582
		10000	0.0496	0.0487	0.0565	0.0520	0.0539	0.0601
1	2	100	0.0717	0.0721	0.1103	0.0660	0.0608	0.0667
		1000	0.0332	0.0419	0.0582	0.0506	0.0539	00547
		10000	0.0366	0.0398	0.0445	0.0543	0.0544	0.0510
	5	100	0.0584	0.0630	0.0946	0.0627	0.0626	0.0663
		1000	0.0375	0.0400	0.0566	0.0540	0.0514	0.0561
		10000	0.0363	0.0364	0.0448	0.0481	0.0492	0.0501

Moreover, we note that although tests based on the bootstrap standard errors performed well in these experiments their use is not generally recommended (see, e.g., the comment in Davidson and MacKinnon, 2004, p. 208). Therefore, if bootstrap is at all feasible, it is perhaps better to use the processing time to obtain bootstrap confidence intervals, or to compute bootstrap p-values for the test statistics based on $\hat{\mathrm{B}}^{-1}\hat{\mathrm{A}}\hat{\mathrm{B}}^{-1}$. The study of the performance of these methods is, however, beyond the scope of the present paper.

The performance of the specification test is again evaluated by computing the rejection frequencies at the 5% level of the null hypothesis, which in this case is defined by (6). The test is based on the statistic

$$T = \sum_{g=1}^{G} \frac{1}{\sqrt{G}} \frac{\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \psi_\theta(\hat{u}_{gi})\psi_\theta(\hat{u}_{gj}) - \psi_\theta(\hat{u}_{gi})^2 \right)}{\sqrt{2\theta^2(1-\theta)^2 n(n-1)}},$$

and in these experiments we took $G \in \{100, 500, 1000\}$ because the power of the test increases quickly with the sample size.

Table 3 presents the rejection frequencies under the null ($d_v = 3$ and $d_\alpha = 0$) and Table 4 presents the results under the alternative ($d_v = 2$ and $d_\alpha = 1$). For comparison, the rejection frequencies obtained with the test suggested by Wooldridge (2002, p. 265) are also included in Tables 3 and 4.

Table 3: Rejection frequencies under the null (at 5%)

		$\theta = 0.25$		$\theta = 0.50$		$\theta = 0.75$		Wooldridge	
n	G	$h = 0$	$h = 1$	$h = 0$	$h = 1$	$h = 0$	$h = 1$	$h = 0$	$h = 1$
2	100	0.0494	0.0524	0.0566	0.0630	0.0496	0.0565	0.0513	0.0349
	500	0.0493	0.0513	0.0501	0.0547	0.0539	0.0543	0.0512	0.0425
	1000	0.0523	0.0483	0.0540	0.0516	0.0475	0.0456	0.0513	0.0449
5	100	0.0411	0.0401	0.0482	0.0498	0.0400	0.0399	0.0624	0.0506
	500	0.0487	0.0470	0.0509	0.0510	0.0506	0.0510	0.0533	0.0557
	1000	0.0522	0.0519	0.0570	0.0550	0.0513	0.0519	0.0540	0.0499

Table 4: Rejection frequencies under the alternative (at 5%)

		$\theta = 0.25$		$\theta = 0.50$		$\theta = 0.75$		Wooldridge	
n	G	$h = 0$	$h = 1$	$h = 0$	$h = 1$	$h = 0$	$h = 1$	$h = 0$	$h = 1$
2	100	0.4285	0.4385	0.6066	0.5998	0.6409	0.6427	0.7005	0.3343
	500	0.9709	0.9685	0.9995	0.9998	0.9992	0.9991	0.9999	0.9096
	1000	0.9997	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	0.9891
5	100	0.9962	0.9973	0.9998	0.9998	0.9963	0.9966	0.9493	0.7131
	500	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9867
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9974

Under the null all tests generally perform well and there is little to choose between them. Under the alternative the results depend on the value of h. In the heteroskedastic case ($h = 1$) the quantile-based tests clearly dominate Wooldridge's test, whereas in the homoskedastic case ($h = 1$) the situation is reversed. However, for realistic sample sizes there is little to choose between the tests because their power quickly approaches 1.

5. CONCLUDING REMARKS

We present the asymptotic results needed to perform inference with quantile regression when the data are obtained by sampling from different groups and it is assumed that observations from different groups are conditionally independent but intra-cluster correlation is not ruled out. We propose a consistent estimator of the covariance matrix of the asymptotic distribution of the estimator allowing for possible intra-cluster correlation and propose a simple test to check the presence of this type of correlation. The results of a small simulation study suggest that the proposed tools are likely to work reasonably well in practice.

APPENDIX

Throughout the Appendix c_r, CS, M, and T denote the c_r, Cauchy-Schwarz, Markov, and triangle inequalities respectively. LLN denotes the Khintchine's Weak Law of Large Numbers, UWL denotes a uniform weak law of large numbers such as Lemma 2.4 of Newey and McFadden (1994), and CLT is the Lindeberg-Lévy central limit theorem.

Proof of Theorem 1: We use Theorem 2.7 of Newey and McFadden (1994). Note that

$$\begin{aligned}\hat{\beta} &= \underset{\beta\in\mathbb{R}^k}{\arg\min} S_G(\beta) - S_G(\beta_0) \\ &= \underset{\beta\in\mathbb{R}^k}{\arg\min}\left[\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\rho_\theta(y_{gi}-x'_{gi}\beta) - \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\rho_\theta(u_{gi})\right].\end{aligned}$$

We have to show that $S_G(\beta) - S_G(\beta_0)$ converges uniformly to a function. In this case pointwise convergence suffices as pointwise convergence of convex functions implies uniform convergence on compact subsets. Note that

$$\begin{aligned}S_G(\beta) - S_G(\beta_0) &= \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}[\rho_\theta(y_{gi}-x'_{gi}\beta) - \rho_\theta(u_{gi})] \\ &= \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}[\rho_\theta(u_{gi}-x'_{gi}\delta) - \rho_\theta(u_{gi})],\end{aligned}$$

where $\delta = \beta - \beta_0$. Note that Knight's identity (Koenker, 2005, p. 121) tells us that

$$\rho_\theta(u-v) - \rho_\theta(v) = -v\psi_\theta(u) + \int_0^v \{\mathrm{I}[u \le s] - \mathrm{I}[u \le 0]\}\, ds,$$

where $\psi_\theta(u) = \theta - \mathrm{I}(u < 0)$. Thus

$$\begin{aligned}S_G(\beta) - S_G(\beta_0) &= \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n} -x'_{gi}\delta\psi_\theta(u_{gi}) \\ &\quad + \frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n}\int_0^{x'_{gi}\delta}\{I[u_{gi} \le s] - I[u_{gi} \le 0]\}\, ds.\end{aligned}$$

Now by a LLN $\frac{1}{G}\sum_{g=1}^{G}\sum_{i=1}^{n} -x'_{gi}\delta\psi_\theta(u_{gi}) = o_p(1)$ and the second term of the rhs converges to

$$\overline{S}(\delta) = \sum_{i=1}^{n}\mathrm{E}[\int_0^{x'_{gi}\delta}\{\mathrm{F}[s|x_{gi}] - \theta\}\, ds].$$

Note that $\overline{S}(\delta) = 0$ if and only if $\delta = 0$ and $\overline{S}(\delta) > 0$ if $\delta \neq 0$. To see this note that if $x'_{gi}\delta > 0$ for some i $\mathrm{F}[s|x_{gi}] - \theta > 0$, thus $\int_0^{x'_{gi}\delta} \{\mathrm{F}[s|x_{gi}] - \theta\}\, ds > 0$. If $x'_{gi}\delta < 0$ for some i $\mathrm{F}[s|x_{gi}] - \theta < 0$, thus $\int_0^{x'_{gi}\delta} \{\mathrm{F}[s|x_{gi}] - \theta\}\, ds > 0$. Since $\delta = \beta - \beta_0 = 0$ is a unique local minimizer and the limiting function is convex, $\delta = \beta - \beta_0 = 0$ is also a global minimizer and the function is convex and consequently $\hat{\beta} = \beta_0 + o_p(1)$. ■

Proof of Theorem 2: We adapt the proof of Koenker (2005, p. 121). Consider the objective function

$$Z_G(\delta) = \sum_{g=1}^{G} \sum_{i=1}^{n} [\rho_\theta(u_{gi} - x'_{gi}\delta/\sqrt{G}) - \rho_\theta(u_{gi})].$$

This function is convex and minimized at $\hat{\delta}_G = \sqrt{G}(\hat{\beta} - \beta_0)$. Using Knight's identity we have

$$\begin{aligned} Z_G(\delta) &= Z_{1G}(\delta) + Z_{2G}(\delta) \\ Z_{1G}(\delta) &= \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \sum_{i=1}^{n} -x'_{gi}\delta\psi_\theta(u_{gi}) \\ Z_{2G}(\delta) &= \sum_{g=1}^{G} \sum_{i=1}^{n} \int_0^{G^{-1/2}x'_{gi}\delta} \{\mathrm{I}[u_{gi} \leq s] - \mathrm{I}[u_{gi} \leq 0]\}\, ds. \end{aligned}$$

Now $Z_{1G}(\delta) = -\delta' W$, where $W = G^{-1/2} \sum_{g=1}^{G} \sum_{i=1}^{n} x_{gi}\psi_\theta(u_{gi})$. Also, by a CLT, $W \xrightarrow{D} \mathcal{N}(0, C)$ where

$$\begin{aligned} C &= \mathrm{Var}(\sum_{i=1}^{n} x_{gi}\psi_\theta(u_{gi})) \\ &= \mathrm{E}[\sum_{i=1}^{n} x_{gi}\psi_\theta(u_{gi})(\sum_{j=1}^{n} x_{gj}\psi_\theta(u_{gj}))'] \\ &= \mathrm{E}[\sum_{i=1}^{n} \sum_{j=1}^{n} x_{gi}x'_{gj}\psi_\theta(u_{gi})\psi_\theta(u_{gj})]. \end{aligned}$$

Now write

$$Z_{2Ggi}(\delta) = \int_0^{G^{-1/2}x'_{gi}\delta} \{\mathrm{I}[u_{gi} \leq s] - \mathrm{I}[u_{gi} \leq 0]\}\, ds,$$

and $Z_{2G}(\delta) = \sum_{g=1}^{G} \sum_{i=1}^{n} Z_{2Ggi}(\delta)$. Note that

$$Z_{2G}(\delta) = \sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[Z_{2Ggi}(\delta)|x_{gi}] + R_G(\delta),$$

where

$$R_G(\delta) = \sum_{g=1}^{G} \sum_{i=1}^{n} \left\{ Z_{2Ggi}(\delta) - \mathrm{E}[Z_{2Ggi}(\delta)|x_{gi}] \right\}.$$

Note also that

$$\begin{aligned}
\sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[Z_{2Gg}(\delta)|x_g] &= \sum_{g=1}^{G} \sum_{i=1}^{n} \int_0^{G^{-1/2}x'_{gi}\delta} \left\{ \mathrm{F}[s|x_{gi}] - \theta \right\} ds \\
&= \frac{1}{\sqrt{G}} \sum_{g=1}^{G} \sum_{i=1}^{n} \int_0^{x'_{gi}\delta} \left\{ \mathrm{F}[\frac{t}{G^{1/2}}|x_{gi}] - \theta \right\} dt \\
&= \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n} \int_0^{x'_{gi}\delta} \left\{ \frac{\mathrm{F}[\frac{t}{G^{1/2}}|x_{gi}] - \theta}{t/\sqrt{G}} \right\} tdt \\
&= \frac{1}{G} \sum_{g=1}^{G} \sum_{i=1}^{n} \int_0^{x'_{gi}\delta} \left\{ f[0|x_{gi}] \right\} tdt + o_p(1) \\
&= \frac{1}{2G} \sum_{g=1}^{G} \sum_{i=1}^{n} \delta' f[0|x_{gi}] x_{gi} x'_{gi} \delta + o_p(1).
\end{aligned}$$

Now by CS

$$\begin{aligned}
Z_{2Ggi}(\delta) &= \int_0^{G^{-1/2}x'_{gi}\delta} \left\{ \mathrm{I}[u_{gi} \le s] - \mathrm{I}[u_{gi} \le 0] \right\} ds \\
&\le \frac{|x'_{gi}\delta|}{G^{1/2}} \le \frac{\|\delta\| \, \|x_{gi}\|}{G^{1/2}}.
\end{aligned} \tag{7}$$

Note that $\mathrm{E}[R_G(\delta)] = 0$ and that by c_r, (7), and CS we have

$$\begin{aligned}
\mathrm{Var}(R_G(\delta)) &= \sum_{g=1}^{G} \mathrm{E}[(\sum_{i=1}^{n} Z_{2Ggi}(\delta))^2] \\
&\le n \sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[Z_{2Ggi}(\delta)^2] \\
&\le n \frac{\|\delta\|}{G^{1/2}} \sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[Z_{2Ggi}(\delta) \, \|x_{gi}\|] \\
&= n \frac{\|\delta\|}{G^{1/2}} \sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[\mathrm{E}[Z_{2Ggi}(\delta)|x_{gi}] \, \|x_{gi}\|] \\
&= n \frac{\|\delta\|}{G^{3/2}} \sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[\|x_{gi}\| \int_0^{x'_{gi}\delta} \left\{ \frac{\mathrm{F}[\frac{t}{G^{1/2}}|x_{gi}] - \theta}{t/\sqrt{G}} \right\} tdt] \\
&\le n \frac{\|\delta\|}{2G^{3/2}} \sum_{g=1}^{G} \sum_{i=1}^{n} \mathrm{E}[\|x_{gi}\| \, \delta' f[0|x_{gi}] x_{gi} x'_{gi} \delta] + o(1) \\
&\le n \frac{\|\delta\|^3}{2G^{1/2}} f_1 \sum_{i=1}^{n} \mathrm{E}[\|x_{gi}\|^3] + o(1) \\
&= o(1).
\end{aligned}$$

Thus $R_G(\delta) = o_p(1)$. Hence by a LLN

$$\begin{aligned} Z_{2G}(\delta) &= \frac{1}{2G}\sum\nolimits_{g=1}^{G}\sum\nolimits_{i=1}^{n}\delta' f[0|x_{gi}]x_{gi}x_{gi}'\delta + o_p(1) \\ &= \delta' B\delta/2 + o_p(1). \end{aligned}$$

Therefore

$$Z_G(\delta) = -\delta' W + \delta' B\delta/2 + o_p(1).$$

The convexity of $-\delta' W + \delta' B\delta/2$ assures that the minimizer is unique and therefore

$$\sqrt{G}(\hat{\beta} - \beta_0) = \arg\min Z_G(\delta) \xrightarrow{D} \hat{\delta}_0 = \arg\min -\delta' W + \delta' B\delta/2.$$

Now note that $\hat{\delta}_0 = B^{-1}W$ (see Koenker, 2005, p. 122, and the references therein). ■

Proof of Theorem 3: The proof is similar to that of Lemma 5 of Kim and White (2003). Let $\mathrm{B}_G = (2c_G G)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{n}\mathrm{I}(|u_{gi}| \leq c_G)x_{gi}x_{gi}'$. Using the mean value theorem we have $\mathrm{E}[\mathrm{B}_G] = \mathrm{E}[\sum_{i=1}^{n} f(\tilde{c}_G|x_{gi})x_{gi}x_{gi}']$, where $|\tilde{c}_G| \leq c_G$ and therefore $\tilde{c}_G = o(1)$. Hence, by the Lebesgue dominated convergence theorem $\mathrm{E}[\mathrm{B}_G] = \mathrm{B}$. It follows from the law of large numbers for double arrays (Davidson, 1994, Corollary 19.9, p. 301, and Theorem 12.10, p. 190) that $\mathrm{B}_G \xrightarrow{p} \mathrm{B}$. We now show (i) $\left|\tilde{\mathrm{B}}_G - \mathrm{B}_G\right| \xrightarrow{p} 0$ where $\tilde{\mathrm{B}}_G = (2c_G G)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{n}\mathrm{I}(|\hat{u}_{gi}| \leq \hat{c}_G)x_{gi}x_{gi}'$, and (ii) $\left|\hat{\mathrm{B}} - \tilde{\mathrm{B}}_G\right| \xrightarrow{p} 0$. The conclusion follows from T.

To prove (i) consider the $(h,j)^{th}$ element of $\left|\tilde{\mathrm{B}}_G - \mathrm{B}_G\right|$, which is given by

$$\left|(2c_G G)^{-1}\sum\nolimits_{g=1}^{G}\sum\nolimits_{i=1}^{n}[\mathrm{I}(|\hat{u}_{gi}| \leq \hat{c}_G) - \mathrm{I}(|u_{gi}| \leq c_G)]x_{gih}x_{gij}'\right|.$$

Now using the facts that $\hat{u}_{gi} = u_{gi} - (\hat{\beta} - \beta_0)'x_{gi}$, $\mathrm{I}(|a| \leq b) = \mathrm{I}(a \leq b) - \mathrm{I}(a < -b)$, $|\mathrm{I}(x \leq 0) - \mathrm{I}(y \leq 0)| \leq \mathrm{I}(|x| \leq |x - y|)$, $|\mathrm{I}(x < 0) - \mathrm{I}(y < 0)| \leq \mathrm{I}(|x| \leq |x - y|)$, T, and CS we have

$$\begin{gathered} \left|(2c_G G)^{-1}\sum\nolimits_{g=1}^{G}\sum\nolimits_{i=1}^{n}[\mathrm{I}(|\hat{u}_{gi}| \leq \hat{c}_G) - \mathrm{I}(|u_{gi}| \leq c_G)]x_{gih}x_{gij}\right| \leq \mathcal{U}_{1G} + \mathcal{U}_{2G} \\ \mathcal{U}_{1G} = (2c_G G)^{-1}\sum\nolimits_{g=1}^{G}\sum\nolimits_{i=1}^{n}[\mathrm{I}(|u_{gi} - c_G| \leq d_G)]\,|x_{gih}|\,|x_{gij}| \\ \mathcal{U}_{2G} = (2c_G G)^{-1}\sum\nolimits_{g=1}^{G}\sum\nolimits_{i=1}^{n}[\mathrm{I}(|u_{gi} + c_G| \leq d_G)]\,|x_{gih}|\,|x_{gij}|, \end{gathered}$$

where $d_G = |c_G - \hat{c}_G| + \left\|\hat{\beta} - \beta_0\right\| \|x_{gi}\|$. We prove that $\mathcal{U}_{1G} \xrightarrow{p} 0$, the proof $\mathcal{U}_{2G} \xrightarrow{p} 0$ is similar.

Let $\mathcal{D}_{1G} = \{\mathcal{U}_{1G} > \eta\}$, $\mathcal{D}_{2G} = \left\{c_G^{-1} \left\|\hat{\beta} - \beta_0\right\| \leq \Delta\right\}$, and $\mathcal{D}_{3G} = \left\{c_G^{-1} |c_G - \hat{c}_G| \leq \Delta\right\}$ for a constant $\Delta > 0$. Thus

$$
\begin{aligned}
\Pr(\mathcal{U}_{1G} > \eta) &= \Pr(\mathcal{D}_{1G}) \\
&\leq \Pr(\mathcal{D}_{1G} \cap \mathcal{D}_{2G} \cap \mathcal{D}_{3G}) + \Pr(\mathcal{D}_{2G}^c) + \Pr(\mathcal{D}_{3G}^c).
\end{aligned}
$$

Now as $\sqrt{G}(\hat{\beta} - \beta_0) = O_p(1)$ and $c_G^{-1} = o(\sqrt{G})$ it follows that $\lim_{G\to\infty} \Pr(\mathcal{D}_{2G}^c) = 0$. Also as $\hat{c}_G/c_G \xrightarrow{p} 1$, we have $\lim_{G\to\infty} \Pr(\mathcal{D}_{3G}^c) = 0$. Additionally if $c_G^{-1} \left\|\hat{\beta} - \beta_0\right\| \leq \Delta$ and $c_G^{-1} |c_G - \hat{c}_G| \leq \Delta$ we have $|d_G| \leq c_G\Delta + c_G\Delta \|x_{gi}\|$. Hence by M

$$
\begin{aligned}
\Pr(\mathcal{D}_{1G} \cap \mathcal{D}_{2G} \cap \mathcal{D}_{3G}) &\leq (2\eta c_G G)^{-1} \sum\nolimits_{g=1}^{G} \sum\nolimits_{i=1}^{n} \mathrm{E}[\int_{-c_G\Delta - c_G\Delta\|x_{gi}\|}^{c_G\Delta + c_G\Delta\|x_{gi}\|} f(s|x_{gi}) ds\, |x_{gih}|\, |x_{gij}|] \\
&\leq (2\eta c_G G)^{-1} \sum\nolimits_{g=1}^{G} \sum\nolimits_{i=1}^{n} \mathrm{E}[\int_{-c_G\Delta - c_G\Delta\|x_{gi}\|}^{c_G\Delta + c_G\Delta\|x_{gi}\|} f_1 ds\, |x_{gih}|\, |x_{gij}|] \\
&= (\eta)^{-1}\Delta \sum\nolimits_{i=1}^{n} \mathrm{E}[(\|x_{gi}\| + 1)\, |x_{gih}|\, |x_{gij}|] < \infty
\end{aligned}
$$

under Assumptions 3. Now take Δ arbitrarily small and consequently $\mathcal{U}_{1G} \xrightarrow{p} 0$.

To prove (ii), note that $\hat{\mathrm{B}} - \tilde{\mathrm{B}}_G = \left(\frac{c_G}{\hat{c}_G} - 1\right) \tilde{\mathrm{B}}_G$. Note also that by (i) $\tilde{\mathrm{B}}_G = O_p(1)$ and since $\left(\frac{c_G}{\hat{c}_G} - 1\right) = o_p(1)$ by assumption, the result follows. ■

Proof of Theorem 4: For simplicity of notation we write $\sum_{gi} := \sum_{g=1}^{G} \sum_{i=1}^{n}$ and $\sum_j := \sum_{j=1}^{n}$. Note that

$$
\begin{aligned}
\mathcal{T} &= \frac{1}{\sqrt{G}} \sum\nolimits_{gi} \left(\sum\nolimits_{j} z_{gi} z_{gj} \psi_\theta(\hat{u}_{gi}) \psi_\theta(\hat{u}_{gj}) - \psi_\theta(\hat{u}_{gi})^2 z_{gi}^2\right) \\
&= \frac{1}{\sqrt{G}} \sum\nolimits_{gi} \sum\nolimits_{j, i\neq j} z_{gi} z_{gj} \psi_\theta(\hat{u}_{gi}) \psi_\theta(\hat{u}_{gj}) \\
&= \frac{1}{\sqrt{G}} \sum\nolimits_{gi} \left(\sum\nolimits_{j, i\neq j} z_{gi} z_{gj} \psi_\theta(u_{gi}) \psi_\theta(u_{gj})\right) + \mathcal{R}_G,
\end{aligned}
$$

where

$$
\mathcal{R}_G = \frac{1}{\sqrt{G}} \sum\nolimits_{gi} \left(\sum\nolimits_{j, i\neq j} z_{gi} z_{gj} [\psi_\theta(\hat{u}_{gi}) \psi_\theta(\hat{u}_{gj}) - \psi_\theta(u_{gi}) \psi_\theta(u_{gj})]\right).
$$

Now by Lemma 5

$$\mathcal{R}_G = \frac{1}{G^{1/2}} \sum\nolimits_{gi} \sum\nolimits_{j,i\neq j} z_{gi} z_{gj} h^*_{gij}(\hat{\beta}) + o_p(1),$$

where

$$\begin{aligned} h^*_{gij}(\hat{\beta}) &= \theta^2 - \theta \mathrm{F}(x'_{gj}(\hat{\beta} - \beta_0)|x_{gj}) - \theta \mathrm{F}(x'_{gi}(\hat{\beta} - \beta_0)|x_{gi}) \\ &\quad + \mathrm{F}(x'_{gi}(\hat{\beta} - \beta_0)|x_{gi}) \times \mathrm{F}(x'_{gj}(\hat{\beta} - \beta_0)|x_{gj}),\ i \neq j. \end{aligned}$$

By a Taylor expansion around β_0 we have

$$\mathcal{R}_G = \frac{1}{G} \sum\nolimits_{gi} \sum\nolimits_{j,\ i\neq j} z_{gi} z_{gj} H^*_{gij}(\tilde{\beta}) \sqrt{G}(\hat{\beta} - \beta_0) + o_p(1),$$

where $\tilde{\beta}$ is on the line segment joining $\hat{\beta}$ and β_0 and

$$\begin{aligned} H^*_{gij}(\tilde{\beta}) &= -\theta f(x'_{gj}(\tilde{\beta} - \beta_0)|x_{gj}) x'_{gj} - \theta f(x'_{gi}(\hat{\beta} - \beta_0)|x_{gi}) x'_{gi} \\ &\quad + f(x'_{gi}(\hat{\beta} - \beta_0)|x_{gi}) \times \mathrm{F}(x'_{gj}(\hat{\beta} - \beta_0)|x_{gj}) x'_{gi} \\ &\quad + f(x'_{gj}(\hat{\beta} - \beta_0)|x_{gj}) \mathrm{F}(x'_{gi}(\hat{\beta} - \beta_0)|x_{gi}) x'_{gj}, \end{aligned}$$

where $i \neq j$. Now notice that

$$\frac{1}{G} \sum\nolimits_{gi} \sum\nolimits_{j,i\neq j} z_{gi} z_{gj} H^*_{gij}(\tilde{\beta}) = o_p(1)$$

by a UWL. Since $\sqrt{G}(\hat{\beta} - \beta_0) = O_p(1)$ we have $\mathcal{R}_G = o_p(1)$. Thus

$$\mathcal{T} = \frac{1}{\sqrt{G}} \sum\nolimits_{gi} \left(\sum\nolimits_{j,i\neq j} z_{gi} z_{gj} \psi_\theta(u_{gi}) \psi_\theta(u_{gj}) \right) + o_p(1)$$

and consequently $\mathcal{T} \to \mathcal{N}(0, D)$ as $D > 0$ and

$$\begin{aligned} D &= \mathrm{E}[\Big(\sum\nolimits_{i=1}^n \Big[\sum\nolimits_{j=1,i\neq j}^n z_{gi} z_{gj} \psi_\theta(u_{gi}) \psi_\theta(u_{gj}) \Big] \Big)^2] \\ &\leq n^2 \sum\nolimits_{i=1}^n \sum\nolimits_{j=1,i\neq j}^n \mathrm{E}[z^4_{gi}]^{1/2} \mathrm{E}[z^4_{gj}]^{1/2} (\theta + 1)^2 < \infty \end{aligned}$$

by two applications of c_r and one of CS. ■

Let

$$m_G(\beta) = \frac{1}{G} \sum\nolimits_{gi} \left(\sum\nolimits_{j,i\neq j} z_{gi} z_{gj} [\theta - \mathrm{I}(u_{gi} \leq x'_{ig}(\beta - \beta_0))][\theta - \mathrm{I}(u_{gj} \leq x'_{gi}(\beta - \beta_0))] \right).$$

Lemma 5 *Suppose that Assumption 4 holds. Then, under H_0, for any $\delta_G = o(1)$ we have*

$$\sup_{\|\beta-\beta_0\|\leq\delta_G}\left|\sqrt{G}(m_G(\beta)-m_G(\beta_0))-\frac{1}{G^{1/2}}\sum\nolimits_{gi}\sum\nolimits_{j,i\neq j} z_{gi}z_{gj}h^*_{gij}(\beta)\right| = o_p(1),$$

where

$$\begin{aligned} h^*_{gij}(\beta) &= \theta^2 - \theta\mathrm{F}(x'_{gj}(\beta-\beta_0)|x_{gj}) - \theta\mathrm{F}(x'_{gi}(\beta-\beta_0)|x_{gi}) \\ &\quad +\mathrm{F}(x'_{gi}(\beta-\beta_0)|x_{gi}) \times \mathrm{F}(x'_{gj}(\beta-\beta_0)|x_{gj}). \end{aligned}$$

Proof: Note that

$$m_G(\beta) = \frac{1}{G}\sum\nolimits_{gi}\sum\nolimits_{j,i\neq j} z_{gi}z_{gj}h_{gij}(\beta),$$

where

$$\begin{aligned} h_{gij}(\beta) &= [\theta^2 - \theta\mathrm{I}(u_{gj} \leq x'_{gj}(\beta-\beta_0)) - \theta\mathrm{I}(u_{gi} \leq x'_{gi}(\beta-\beta_0)) \\ +\mathrm{I}(u_{gi} &\leq x'_{ig}(\beta-\beta_0))\mathrm{I}(u_{gj} \leq x'_{gj}(\beta-\beta_0))],\ i\neq j. \end{aligned}$$

Now taking the expected value of $h_{gij}(\beta)$ conditional on x_g we have

$$\begin{aligned} \mathrm{E}[h_{gij}(\beta)|x_g] &= \theta^2 - \theta\mathrm{F}_j(x'_{gj}(\beta-\beta_0)|x_g) + \theta\mathrm{F}_i(x'_{gi}(\beta-\beta_0)|x_g) \\ &\quad +\mathrm{F}_{i,j}(x'_{ig}(\beta-\beta_0), x'_{gj}(\beta-\beta_0)|x_g) \\ &= \theta^2 - \theta\mathrm{F}(x'_{gj}(\beta-\beta_0)|x_{gj}) + \theta\mathrm{F}(x'_{gi}(\beta-\beta_0)|x_{gi}) \\ &\quad +\mathrm{F}(x'_{gi}(\beta-\beta_0)|x_{gi}) \times \mathrm{F}(x'_{gj}(\beta-\beta_0)|x_{gj}), \end{aligned}$$

where the last line follows from H_0 and $i \neq j$.

Note now that

$$\sqrt{G}(m_G(\beta)-m_G(\beta_0)) - \frac{1}{G^{1/2}}\sum\nolimits_{gi}\sum\nolimits_{j,i\neq j} z_{gi}z_{gj}h^*_{gij}(\beta) =$$
$$\frac{1}{G^{1/2}}\sum\nolimits_{gi}\sum\nolimits_{j,i\neq j} z_{gi}z_{gj}[h_{gij}(\beta) - h_{gij}(\beta_0) - h^*_{gij}(\beta)].$$

Since the indicator functions $\mathrm{I}(u_{gi} \leq x'_{ig}(\beta-\beta_0))$ and $\mathrm{I}(u_{gj} \leq x'_{gi}(\beta-\beta_0))$ and the conditional distribution functions $\mathrm{F}(x'_{gi}(\beta-\beta_0)|x_{gi})$ and $\mathrm{F}(x'_{gj}(\beta-\beta_0)|x_{gj})$ are functions of bounded variation (and hence *type I class* of functions in the sense of Andrews, 1994) and as Assumptions 1 (a) and 4 (a) hold, it follows that

$$\frac{1}{G^{1/2}} \sum\nolimits_{gi} \sum\nolimits_{j,i \neq j} z_{gi} z_{gj} [h_{gij}(\beta) - h_{gij}(\beta_0) - h^*_{gij}(\beta)]$$

is stochastic equicontinuous by Theorems 1, 2 and 3 of Andrews (1994). ■

REFERENCES

Abrevaya, J. and Dahl, C.M. (2008). "The Effects of Birth Inputs on Birthweight," *Journal of Business & Economic Statistics*, 26, 379-397.

Andrews, D.W.K. (1994). "Empirical Process Methods in Econometrics," in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics,* Vol. 4, 2247-2294. New York(NY): North Holland.

Arellano, M. (1987). "Computing Robust Standard Errors for Within-Group Estimators," *Oxford Bulletin of Economics and Statistics*, 49, 431-434.

Breusch, T.S. and Pagan, A.R. (1980). "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47, 239-253.

Buchinsky, M. (1995), "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models a Monte Carlo Study," *Journal of Econometrics*, 68, 303-38.

Cameron, A.C. and Miller, D.L. (2011). "Robust Inference with Clustered Data," in A. Ullah and D.E. Giles, eds., *Handbook of Empirical Economics and Finance*, 1-28, Boca Raton (FL): Chapman & Hall/CRC Press.

Chamberlain, G. (1994). "Quantile Regression, Censoring and the Structure of Wages," in C.A. Sims, ed., *Advances in Econometrics*, 171–209. Cambridge: Cambridge University Press.

Cochran, W.G. (1952). "The χ^2 Test of Goodness of Fit," *Annals of Mathematical Statistics*, 23, 315-345.

Davidson, J.E.H. (1994). *Stochastic Limit Theory.* Oxford: Oxford University Press.

Davidson, R. and MacKinnon, J.G. (2004). *Econometric Theory and Methods.* New York, Oxford University Press.

Kim, T.H. and White, H. (2003). "Estimation, Inference, and Specification Testing for Possibly Misspecified Quantile Regressions," in T. Fomby and R.C. Hill, eds., *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*, 107-132. New York (NY): Elsevier.

Koenker, R. (2005). *Quantile Regression.* New York (NY): Cambridge University Press.

Koenker, R. and Bassett, G. (1978). "Regression Quantiles," *Econometrica*, 46, 33-50.

Koenker, R. and Machado, J.A.F. (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression", *Journal of the American Statistical Association*, 94, 1296-1310.

Liang, K.-Y. and Zeger, S.L. (1986). "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13-22.

Machado, J.A.F., Parente, P.M.D.C, and Santos Silva, J.M.C. (2013). `qreg2`: Stata module to perform quantile regression with robust and clustered standard errors, Statistical Software Components S457369, Boston College Department of Economics.

Moulton, B.R. (1986), "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385-97.

Moulton, B.R. (1990), "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72, 334-38.

Newey, W.K. and McFadden, D.L. (1994). "Large Sample Estimation and Hypothesis Testing", in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics,* Vol. 4, 2111-2245. New York (NY): North Holland.

Powell, J.L. (1984). "Least Absolute Deviation Estimation for Censored Regression Model", *Journal of Econometrics*, 25, 303-325.

Powell, J.L. (1986). "Censored Regression Quantiles", *Journal of Econometrics*, 32, 143-155.

Rogers, W.H. (1993), "Regression Standard Errors in Clustered Samples," *Stata Technical Bulletin*, 13, 19-23.

StataCorp. (2009). *Stata Release 11. Statistical Software.* College Station (TX): StataCorp LP.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

White, H. (1984), *Asymptotic Theory for Econometricians*, San Diego (CA): Academic Press.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Cambridge(MA): MIT Press.