Inference on linear quantile regression with dyadic data

Hongqi Chen*

May 2022

Abstract

In this paper, we study a robust inference procedure for the linear quantile regression estimator with a dyadic data structure. We derive asymptotic distribution for quantile regression estimator when dependence exists between any pair of dyads with common nodes in a network. We provide the consistency results for the covariance matrix estimator and show asymptotic normality for the corresponding $t-$statistic. Numerical simulations are provided to illustrate the excellent performance of our $t-$statistic in the inference of quantile regression with dyadic data.

Keywords: Dyadic data; Quantile regression; Robust inference procedure; Robust $t-$test

*Department of Economics, University of Illinois, Urbana-Champaign, 214 David Kinley Hall, 1407 West Gregory Drive, Urbana, IL, 61801, USA, Email:hongqic2@illinois.edu

1 Introduction

Dyadic data is a common type of data structure in the modern network or matched data analysis. In the field of economics, dyadic data have many applications. For example, the trade volume between any pair of countries in the world is a dyadic dataset. In finance, the dependence measurements across different financial institutions or assets also form a dyadic dataset. An essential feature of a dyadic dataset is the highly correlated cross-sectional dependence structure, in that any pairs of dyads are dependent if they share a common node in the network. For example, the trade volume between US and China is correlated with the trade volume between US and EU. This correlation structure is much more complicated than traditional heteroscedasticity data or clustered data frameworks. Therefore, a more careful robust inference procedure is needed for empirical researchers.

This paper considers a robust inference procedure on linear quantile regression estimator under a dyadic data framework. Quantile regression (QR), proposed by Koenker and Bassett (1978), is a valuable tool for researchers to estimate the conditional distribution of a response variable at different quantile levels. It would characterize a whole distribution to consider heterogeneity rather than focusing on the mean. A good reference for quantile regression is Koenker (2005). Past literature considers inference issues with different data structure types, including heteroscedastic observations and clustered data (see literature review below). However, with dyadic data, inference on the QR estimator is generally unknown. Our study contributes to the asymptotic distribution for the QR estimator with dyadic data structure, provides the consistency for its dyadic-robust covariance matrix estimation, and shows the asymptotic standard normality for the corresponding $t-$statistic. Our paper provides a theoretical foundation and guidance to applied researchers for further inference on the QR estimator.

Our asymptotic framework borrows from Tabord-Meehan (2019), who considers the robust inference issue for the least square estimator with dyadic data. Since dyadic data usually come from dependency graphs, we utilize a central limit theorem from Janson (1988). To make the theoretical framework reasonable, we keep our assumptions keep in line with Tabord-Meehan (2019) and other quantile regression literature. Then we establish the asymptotic distribution of the QR estimator and show a consistent estimator for the covariance matrix. Based on our theoretical results, inference on QR with dyadic data is straightforward.

Relevant literature. Our paper relates to two parts of existing literature. The first part is inference with dyadic data. For the least square estimation, Cameron and Miller (2014) propose robust inference for paired or dyadic data and mention that the traditional two-way clustered robust estimator is not suitable for dyadic data. Aronow et al. (2015) show the consistency of the covariance matrix estimation of parameters with dyadic data and extend their results into a panel data framework. Both papers do not consider the inference issue, such as the asymptotic distribution for the least square estimator. Niu (2021) considers density estimation and nonparametric regression with dyadic data. Tabord-Meehan (2019) first proposes a formal inference for linear regression with dyadic data based on limit theorems from Janson (1988). This paper inspires us to follow the same theoretical framework.

The second part of the literature is related to robust inference for the QR estimator. Under traditional non-i.i.d data structure, a typical "sandwich" form covariance matrix estimation follows classical literature by Hendricks and Koenker (1992) as well as Powell et al. (1991), who provide an alternative form using kernel estimation. For a more complex clustered data framework, Parente and Silva (2016) first propose an analytical form for a consistent estimator of the covariance matrix estimation and provides asymptotic distribution for the QR estimator. Hagemann (2017) provides a wild gradient bootstrap approach for the cluster-robust inference of the QR estimator and shows good performance when the with-in cluster correlation level is high. Our setting with dyadic data complements the literature on the robust inference for QR estimator.

The rest of this paper is organized as follows. The remaining part of this section mentions our notations throughout this paper. Section 2 introduces our model setup and relevant asymptotic framework. In section 3, we present our theoretical results on asymptotic distributions of the QR estimator, consistent covariance matrix estimation, and asymptotic normality for $t-$statistic. Relevant proofs are in the appendix. Finite-sample Monte Carlo simulations are provided in section 4. The last section concludes.

Notations. In this paper, for any vector a, $|a|$ denotes its L_1-norm while $\|a\|$ denotes its Euclidean norm. For any $\tau \in (0,1)$, $\rho_\tau(u) = u(\tau - 1(u<0))$ acts as the traditional quantile loss function, as well as its subgradient $\psi_\tau(u) = \tau - 1(u<0)$. Knight's identity means $\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v 1(u \leq s) - 1(u \leq 0)\, ds$.

2 Quantile Regression with Dyadic Data

In this section, we describe the linear QR model with dyadic data and show two asymptotic frameworks. Our asymptotic frameworks borrow from Tabord-Meehan (2019).

2.1 Model setup

We start to introduce some notations in our model setup formally. Considering an undirected networks with $i = 1, \ldots, n$ nodes, we denote the index set $\mathcal{I} = \left\{ \{g, h\} : \{g, h\} \in \{1, \ldots, n\}^2, g \neq h \right\}$ as the set of all dyads between distinct nodes. We denote $Y_{gh} \in \mathbb{R}$ as a response variable in $\{g, h\}$ and $X_{gh} \in \mathbb{R}^p$ as the corresponding covariates. For each observations, we use y_{gh} and x_{gh} respectively. Since we consider an undirected network, we have $y_{gh} = y_{hg}$ as well as $x_{gh} = x_{hg}$. We index our dyads by $d_{\{g,h\}} \in \{1, \ldots, D\}$. If all dyads exist in this network, we can express the cardinality of all different dyads as $D = \frac{n!}{(n-2)!2!} = \frac{n(n-1)}{2}$. For general situation, we do not need $D = \frac{n(n-1)}{2}$. In addition, we define the inverse correspondence $\phi : \{1, \ldots, D\} \to \mathcal{I}$ such that $\phi\left(d_{\{g,h\}}\right) = \{g, h\}$. So we can rewrite (y_{gh}, x_{gh}) as (y_d, x_d) for simplicity. We consider the following linear quantile regression model throughout the context

$$y_d = x_d^T \beta + u_d$$

where $\beta \in \mathbb{R}^p$ is our parameter of interest. u_d is the quantile error term satisfying $P\left(u_d < 0 | x_d\right) = \tau$. Our objective is to investigate a robust inference procedure for β. To achieve our goal, we should consider two steps. First, we must show the limiting distributions on our estimate $\hat{\beta} - \beta$. Second, we should provide a dyadic robust standard error estimator. In the end, the robust test statistic follows.

Since our data structure is dyadic, we should allow arbitrary dependence structure across any two dyads if they share a common node. Otherwise, if two dyads do not have a node in common, they are assumed to be independent. In the least square literature, e.g. Cameron and Miller (2014); Aronow et al. (2015); Tabord-Meehan (2019), a typical characterization is that $E\left(\epsilon_d \epsilon_{d'} | x_d, x_{d'}\right) = 0$ unless $\phi(d) \cap \phi\left(d'\right) \neq \emptyset$, where ϵ is the error term in mean regression. Here, we state a quantile analog form for the QR estimator.

$$E\left(\psi_\tau\left(u_d\right) \psi_\tau\left(u_{d'}\right) | x_d, x_{d'}\right) = 0$$

unless $\phi(d) \cap \phi\left(d'\right) \neq \emptyset$. This quantile-type analogy is useful in our formal analysis, and our first assumption below indicates this condition naturally.

We can now state our main assumptions. Note that assumption 4 are different in convergence rates with respect to asymptotic frameworks, which we will introduce in the following subsection.

Assumption 1. *(i) For all $d = 1, \ldots, D$, every sample $\left(y_d, x_d^T\right)^T$ are drawn from the same joint distribution $P(\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{p+1}$ where $\mathcal{X}$ and $\mathcal{Y}$* are the supports of random variables X and Y. *(ii) The random vectors $\left(y_d, x_d^T\right)^T$ and $\left(y_{d'}, x_{d'}^T\right)^T$ are independent if $\phi(d) \cap \phi\left(d'\right) = \emptyset$*

for any $d, d^{'} = 1, \ldots, D$.

Assumption 2. *For any fixed* $\tau \in T$ *where* T *is a closed subset of* $(0,1)$, *the* τ*th conditional quantile of* y_d *satisfies a linear form in the sense that* $Q_\tau\left(y_d|x_d\right) = x_d^T\beta$. *We assume* $\beta \in B$ *where* $B \subset \mathbb{R}^p$ *is a compact and convex set.*

Assumption 3. *(i)* $\{y_d, x_d\}$ *has uniform bounded supports* $\{\mathcal{Y}, \mathcal{X}\}$ *for any* $d = 1, \ldots, D$. *(ii)* $\frac{1}{D}\sum_{d=1}^{D} X_d X_d^T$ *is positive definite uniformly in* d. *(iii) The conditional density of* u_d, $f_{u_d}\left(u|x_d\right)$, *and its derivative ,* $f_{u_d}^{'}\left(u|x_d\right)$, *are bounded above uniformly in* u *and* x_d *for any* $d = 1, \ldots, D$. *We denote* $\bar{f}$ *as the upper bound for* $f_{u_d}\left(u|x_d\right)$ *in the sense that* $f_{u_d}\left(u|x_d\right) < \bar{f} < \infty$. *(iv) the conditional density* $f_{u_d}\left(u|x_d\right)$ *is bounded below away from* 0 *uniformly in* $\beta \in B$ *and* x_d *for any* $d = 1, \ldots, D$.

Assumption 4. *(i) The Jacobian*

$$J_d = \frac{1}{D}\sum_{d=1}^{D} E f_{u_d}\left(0|x_d\right) x_d x_d^T \to J = E f_{u_d}\left(0|x_d\right) x_d x_d^T$$

(ii) Under the first asymptotic framework, the covariance matrix

$$\Sigma_d = \frac{1}{D}\sum_{d=1}^{D}\sum_{d'=1}^{D} E\psi_\tau\left(y_d - x_d^T\beta\right)\psi_\tau\left(y_{d'} - x_{d'}^T\beta\right) x_d x_{d'}^T \to \Sigma$$

where $\Sigma = E\psi_\tau\left(y_d - x_d^T\beta\right)\psi_\tau\left(y_{d'} - x_{d'}^T\beta\right) x_d x_{d'}^T$.

(iii) Under the second asymptotic framework, the covariance matrix

$$\Sigma_d = \frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D} E\psi_\tau\left(y_d - x_d^T\beta\right)\psi_\tau\left(y_{d'} - x_{d'}^T\beta\right) x_d x_{d'}^T \to \Sigma$$

for some $r \in (0,1]$.

Remark 2.1. For assumption 1, this assumption ensures that samples on dyads with no common nodes are independent with each other, and we allow dyads with common nodes have arbitrary correlation. This assumption implies $E\left(\psi_\tau\left(u_d\right)\psi_\tau\left(u_{d'}\right)|x_d, x_{d'}\right) = 0$ unless $\phi\left(d\right) \cap \phi\left(d^{'}\right) \neq \emptyset$. Moreover, we assume observations follow the same distribution regardless of their dependency level. As for assumption 2, we assume the linear quantile regression structure of our model. We rule out the extremal quantile situation. In assumption 3, these are regular assumptions in QR literature. Those assumptions guarantee good behavior on conditional densities. We impose bounded supports on our data for simplicity, while they can be relaxed with more general moment conditions, as long as asymptotic theorems in Janson

(1988) hold. Assumption 4 is a technical assumption on the convergence of sample Jacobian and covariance matrix of $\psi_\tau x$. These assumptions are in line with Koenker (2005) and Hagemann (2017) but with a dyadic data structure. As for the limit of the covariance matrix, since we consider two asymptotic frameworks below, we denote two limiting assumptions here. The D and DN^r in denominators are related to our asymptotic frameworks, which we will introduce soon. Besides those regular assumptions above, it is worth mentioning that although our analysis is based on undirected networks, it is straightforward to extend our results into a directed network.

2.2 Asymptotic frameworks and a key central limit theorem

In order to consider the asymptotic properties of the QR estimator with dyadic data, we utilize the asymptotic frameworks in Tabord-Meehan (2019). In this subsection, we first introduce this framework and then mention a fundamental central limit theorem for the dependency graph from Janson (1988).

In a given undirected network, we define M_n be the number of dyads for every node $n = 1, \dots, N$ with $M^H = \max_n M_n$ and $M^L = \min_n M_n$. By the definition of a dyad, we know $M^H \leq N - 1$ and $\frac{M^L N}{2} \leq D \leq \frac{M^L N}{2}$.

We consider two asymptotic frameworks from Tabord-Meehan (2019) in assumption 5 and 6, respectively. These two frameworks consider the relative level of M^H and M^L with respect to D, therefore they can control the density of the network. We employ the same rate conditions on M^H and M^l as (AF1) and (AF2) assumptions in Tabord-Meehan (2019) to distinguish the two frameworks.

Assumption 5. *(First asymptotic framework) $M^H < C$ for some constant C as $D \to \infty$.*

Assumption 6. *(Second asymptotic framework) $M^L \geq cN$ for some positive constant c.*

Remark 2.2. For assumption 5, it restricts the number of dyads for all nodes; hence the network is relatively sparse under this setup. Although this assumption implies the dependency level within the network is limited as D diverges to infinity, it seems somewhat inappropriate if the dyads are dense. In real-world applications, the dependence structure is quite complex in a network. An intuitive extension would allow M_d to grow with the number of nodes N. In assumption 6, we allow M^L and therefore M^H to grow at a certain rate cN to infinity as N diverges. There is another implication that we also need M^L growing with M^H: If we only allow M^H to grow but M^L is fixed, then the variation of dyads in the dependency graph would be too high to apply Janson (1988)'s central limit theorem (see more discussion in Tabord-Meehan (2019)).

Given our asymptotic frameworks, we restate the key central limit theorem and its conditions from Janson (1988) by our notations below. Our results heavily rely on this theorem.

Theorem 2.1. *(Theorem 2 in Janson (1988)) Suppose that for each D, $\{X_d\}_{d=1}^{D}$ is a family of bounded random variables, $|X_d| \leq C$. Suppose further Γ_D is a dependency graph for this family and let M_{max} be the maximal degree of Γ_D. Let $S_D = \sum_{d=1}^{D} X_d$ and $\sigma_D^2 = var\left(S_D\right)$. If there exists an integer $l \geq 3$ such that*

$$\frac{\left(\frac{D}{M_{max}}\right)^{\frac{1}{l}} M_{max} C}{\sigma_D} \to 0$$

as $D \to \infty$, then

$$\frac{S_D - ES_D}{\sigma_D^2} \xrightarrow{d} N\left(0,1\right)$$

Remark 2.3. Given this central limit theorem, we can work on the asymptotic distribution of random variables defined on dyads. In our model, the family of dyadic random variables forms a dependency graph with maximal degree $M_{max} = 2\left(M^H - 1\right)$ at most. Throughout the paper, we need to verify $\frac{\left(\frac{D}{M_{max}}\right)^{\frac{1}{l}} M_{max} C}{\sigma_D} \to 0$ for those dyadic random variables and asymptotic distribution follows. In addition, note that this central limit theorem only considers scalar random variables. Therefore, in the formal proof, we first apply this central limit theorem with a single random variable and combine results with the Cramer-Wold device to obtain final asymptotic properties.

Based on the previous central limit theorem, we shall note that we use $\sigma_D^2 = var\left(\sum_{d=1}^{D} x_d \psi_\tau\left(y_d - x_d^T \beta\right)\right)$ throughout the paper. Similar to Tabord-Meehan (2019), we need to control the growth rate of σ_D^2. In the first asymptotic framework which σ_D^2 increase at rate D, we consider assumption 4 (ii) to control. For the second framework, we need to impose assumption 4 (iii) in order to achieve rate control on σ_D^2.

3 Asymptotic properties

In this section, we show results about asymptotic properties under both asymptotic frameworks. We first show the asymptotic distribution of the QR estimator for any fixed quantile level τ. Next, we provide a consistent covariance estimator $\hat{\Omega}$. Under these two results, we can obtain the traditional $t-$test statistic is asymptotically standard normal.

3.1 Asymptotic distribution for QR estimator

The QR estimator solves the following objective minimization for a given τ with dyadic data

$$\hat{\beta} = \arg\min_{\beta \in B} \frac{1}{D} \sum_{d=1}^{D} \rho_\tau \left(y_d - x_d^T \beta \right)$$

We emphasize that under different asymptotic frameworks, the rates of obtaining asymptotic normality are different. For the first asymptotic framework with assumption 5, we consider $\sqrt{D}\left(\hat{\beta} - \beta\right)$. But for the second framework with assumption 6, we consider $\sqrt{\frac{D}{N^r}}\left(\hat{\beta} - \beta\right)$. The differences in the scaling rates of asymptotic distributions are rooted in the distinctions of the growth rate of $var\left(\sum_{d=1}^{D} \psi_\tau \left(y_d - x_d^T \beta \right) x_d \right)$ under two frameworks as we explained before. Indeed, the factor r in assumption 4 (iii) controls this growth rate. Similar to Tabord-Meehan (2019), we rule out the case that $r = 0$, which means we have i.i.d data and traditional QR theory holds. In addition, although r is generally unknown, it does not impact the asymptotic distribution of test statistics. Therefore, knowing the true value of r is not necessary for our inference procedure.

We start to state our two asymptotic theorems on limiting distributions below.

Theorem 3.1. *Under the condition that assumptions 1-4 and 5 hold,*

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

where $\Omega = J^{-1} \Sigma J^{-1}$ *and* Σ *is defined in assumption 4 (ii).*

Theorem 3.2. *Under the condition that assumptions 1-4 and 6 hold,*

$$\sqrt{\frac{D}{N^r}}\left(\hat{\beta} - \beta\right) \xrightarrow{d} N\left(0, \Omega\right)$$

where $\Omega = J^{-1} \Sigma J^{-1}$ *and* Σ *is defined in assumption 4 (iii).*

3.2 Consistency of covariance matrix

To consider a suitable inference procedure for QR estimator with dyadic data, our straightforward idea is using traditional $t-$statistics, i.e. for any β_j in β

$$T_j = \frac{\hat{\beta}_j - \beta}{\sqrt{\hat{\Omega}_{jj}}}$$

Therefore, obtaining a consistency estimator for Ω is a top issue after we derive the limiting distribution of $\hat{\beta}_j - \beta$. A natural choice would be the "sandwich" formula with modifications on correlated dyads.

$$
\begin{aligned}
\hat{\Omega} &= \hat{J}^{-1}\hat{\Sigma}\hat{J}^{-1} \\
&= \left(\sum_{d=1}^{D} \hat{f}_{u_d}\left(0|x_d\right) x_d x_d^T\right)^{-1} \left(\sum_{d=1}^{D}\sum_{d'=1}^{D} 1_{d,d'}\psi_\tau\left(\hat{u}_d\right)\psi_\tau\left(\hat{u}_{d'}\right) x_d x_{d'}^T\right) \left(\sum_{d=1}^{D} \hat{f}_{u_d}\left(0|x_d\right) x_d x_d^T\right)^{-1}
\end{aligned}
$$

where $\hat{u}_d = y_d - x_d^T\hat{\beta}$ and $1_{d,d'}$ is the indicator function which has value 1 if $\phi\left(d\right) \cap \phi\left(d'\right) \neq \emptyset$ and 0 otherwise.

From previous asymptotic distributions of $\hat{\beta}$, the scaling factors of $\hat{\beta} - \beta$ are different for the two frameworks. For the first framework, we need to consider $D\hat{\Omega} \xrightarrow{p} \Omega$, while in the second, we should have $\frac{D}{N^r}\hat{\Omega} \xrightarrow{p} \Omega$ to ensure the standard normality of our $t-$statistics. In the expression of $\hat{\Omega}$, we note that $\hat{J}$ does not depend on the scaling factor for both frameworks, since it does not involve the summation of covariances on dyads. Therefore, we only need to consider the middle part by showing $\frac{1}{D}\hat{\Sigma} \xrightarrow{p} \Sigma$ for the first asymptotic framework and $\frac{1}{DN^r}\hat{\Sigma} \xrightarrow{p} \Sigma$ for the second. The following two lemmas show these results.

Lemma 3.1. *Under assumptions 1-4 and 5, we have*

$$\frac{1}{D}\hat{\Sigma} \xrightarrow{p} \Sigma$$

Lemma 3.2. *Under assumptions 1-4 and 6, for $r > \frac{1}{2}$, we have*

$$\frac{1}{DN^r}\hat{\Sigma} \xrightarrow{p} \Sigma$$

Remark 3.1. By the definition of $\hat{\Sigma}$, it is easy to find that two convergence results do not apply to i.i.d. data, since when all dyads are independent, $\hat{\Sigma} = 0_{p\times p}$. In that case, we do not have a proper estimate of $\hat{\Omega}$. Another important point in lemma 3.2 is that we require $r > \frac{1}{2}$ as Tabord-Meehan (2019). The reason to further restrict the growth rate of $var\left(\sum_{d=1}^{D}\psi_\tau\left(y_{d'} - x_{d'}^T\beta\right)x_d\right)$ is straightforward. When r is relatively small, it cannot control the divergence of $\hat{\Sigma}$.

Remark 3.2. In numerical implementations, we find that $\hat{\Sigma}$ may not always be positive definite in finite samples. Hence the estimated variance may not be positive. This problem also appears in the least square estimator. We follow the correction procedure in Cameron and Miller (2014) such that while $\hat{\Sigma}$ is not positive definite, we manually set those non-positive diagonal elements in $\hat{\Omega}$ to $\epsilon = 10^{-7}$ later.

Before we consider the consistency of $\hat{\Omega}$, we first impose a straightforward assumption on $\hat{J}$, which guarantees the convergence of J in covariance matrix estimation. This is a high-level assumption, and we will then relax this assumption by providing a consistent kernel estimator for J, like Powell et al. (1991).

Assumption 7. $\frac{1}{D}\hat{J} = \frac{1}{D}\sum_{d=1}^{D} \hat{f}_{u_d}(0|x_d)\, x_d x_d^T \xrightarrow{p} J$

Using this high-level assumption, we have the following consistent estimators for covariance matrix Ω under two asymptotic frameworks.

Theorem 3.3. *Under assumptions 1-4, 5 and 7, we have*

$$D\hat{\Omega} \xrightarrow{p} \Omega$$

Proof. This result follows from $\frac{1}{D}\hat{\Sigma} \xrightarrow{p} \Sigma$ and $\left(\frac{1}{D}\hat{J}\right)^{-1} \xrightarrow{p} J^{-1}$. □

Theorem 3.4. *Under assumptions 1-4 and 6-7, for $r > \frac{1}{2}$, we have*

$$\frac{D}{N^r}\hat{\Omega} \xrightarrow{p} \Omega$$

Proof. The result is from $\frac{1}{DN^r}\hat{\Sigma} \xrightarrow{p} \Sigma$ for $r > \frac{1}{2}$ and $\left(\frac{1}{D}\hat{J}\right)^{-1} \xrightarrow{p} J^{-1}$ directly □

As we can see, the convergence results of $\hat{\Omega}$ inherit conditions in the convergence of $\hat{\Sigma}$.

Now we relax our high-level assumption 7 by proposing a consistent kernel estimator by Powell et al. (1991). This estimator is also known as Powell's "sandwich" according to Koenker (2005).

$$\hat{J} = \frac{1}{2\hat{c}_D}\sum_{d=1}^{D} 1\left(|\hat{u}_d| < \hat{c}_D\right) x_d x_d^T$$

where $\hat{c}_D$ is a stochastic sequence acting as the bandwidth, and we shall show this covariance matrix estimator $\frac{1}{D}\hat{J} \xrightarrow{p} J$. Our proof procedure is modified from Kim and White (2003) and Parente and Silva (2016), and we need the following rate assumption on the bandwidth $\hat{c}_D$.

Assumption 8. *$\hat{c}_D$ is a stochastic sequence w.r.t D and c_D is a deterministic sequence satisfying $\frac{\hat{c}_D}{c_D} \xrightarrow{p} 1$, $c_D = o(1)$ and $\frac{1}{c_D} = o\left(\sqrt{\frac{D}{N^r}}\right)$ for some $r \in (0, 1]$.*

Remark 3.3. Our assumptions on the bandwidth $\hat{c}_D$ is common in kernel estimators. $\hat{c}_D$ can be a function on the data. We require $\frac{1}{c_D} = o\left(\sqrt{\frac{D}{N^r}}\right)$ for the second framework and this condition is also suitable for the first framework which requires $\frac{1}{c_D} = o\left(\sqrt{D}\right)$.

Proposition 3.1. *Under assumptions 1-4 above, we have* $\frac{1}{D}\hat{J} \xrightarrow{p} J$.

Remark 3.4. With the above proposition, we can get rid of our high-level assumption 7. We can apply the traditional kernel estimator to achieve the consistency of Jacobian $\hat{J}$. Note that the Jacobian does not involve covariance term; hence it has scaling factor D for both asymptotic frameworks.

3.3 Asymptotic distribution for test statistics

Following previous theorems, we can show the asymptotic distribution of the $t-$statistic for robust inference. The asymptotic standard normality for $t-$statistic holds under both frameworks.

Theorem 3.5. *Under assumptions 1-4 and one of the assumptions in 5 and 6 above, we have*

$$T_j = \frac{\hat{\beta}_j - \beta}{\sqrt{\hat{\Omega}_{jj}}} \xrightarrow{d} N(0,1)$$

for $j = 1, \ldots, p$.

Proof. Using Slutsky's theorem, asymptotic standard normality is directly from theorems 3.1 and 3.3 or 3.2 and 3.4. □

Remark 3.5. As we can see, it is obvious that the $t-$test statistic is asymptotically standard normal. We emphasize that this result implies the same limiting distribution for both asymptotic frameworks. Therefore, in practice, researchers do not need to know which asymptotic framework they are in and the specific value of r. They can still achieve the same asymptotic standard normality for the $t-$test.

4 Monte Carol Simulations

In this section, we show finite sample performance for robust inference of the QR estimator with dyadic data. We consider our dyadic robust $t-$statistic defined above and traditional QR inference $t-$statistic as a comparison. Our simulation design is modified from Hagemann (2017).

We consider a simple data generating process (DGP) with different dependence levels.

For any $d=1,\dots D$ with $\phi\left(d_{\{g,h\}}\right)=\{g,h\}$,

$$\begin{aligned} y_d &= u_d + x_d + x_d^2 u_d \\ x_d &= \sqrt{\rho} z_g + \sqrt{\rho} z_h + \sqrt{1-2\rho} z_d \\ u_d &= \sqrt{\rho} v_g + \sqrt{\rho} v_h + \sqrt{1-2\rho} v_d \end{aligned}$$

where z_g, z_h, and z_d follow independent $Unif\left(-\sqrt{3},\sqrt{3}\right)$ while v_g, v_h and, v_d follow independent $Unif\left(-1,1\right)$. Under this setup, we standardize the x_d with mean 0 and variance 1 and u_d with mean 0 and variance $\frac{1}{3}$. We choose $\rho \in [0, 0.5]$ to control the dependence level across dyads. When $\rho = 0$, we have pure i.i.d data for x_d and u_d. On the contrary, if $\rho = 0.5$, the dependence level is the highest among those dyads. Moreover, we note that any two observations x_d and $x_{d'}$ with $\phi\left(d\right) \cap \phi\left(d'\right) \neq \emptyset$ has correlation ρ. It is easy to show this DGP satisfies the following QR model

$$Q_\tau\left(y_d | x_d\right) = \beta_0 + \beta_1 x_d + \beta_2 x_d^2$$

with $\beta_0 = \frac{F_U^{-1}(\tau)}{\sqrt{3}}$, $\beta_1 = 1$, and $\beta_2 = \frac{F_U^{-1}(\tau)}{\sqrt{3}}$ where $F_U^{-1}\left(\tau\right)$ is the inverse cumulative distribution function of $Unif\left(-1,1\right)$.

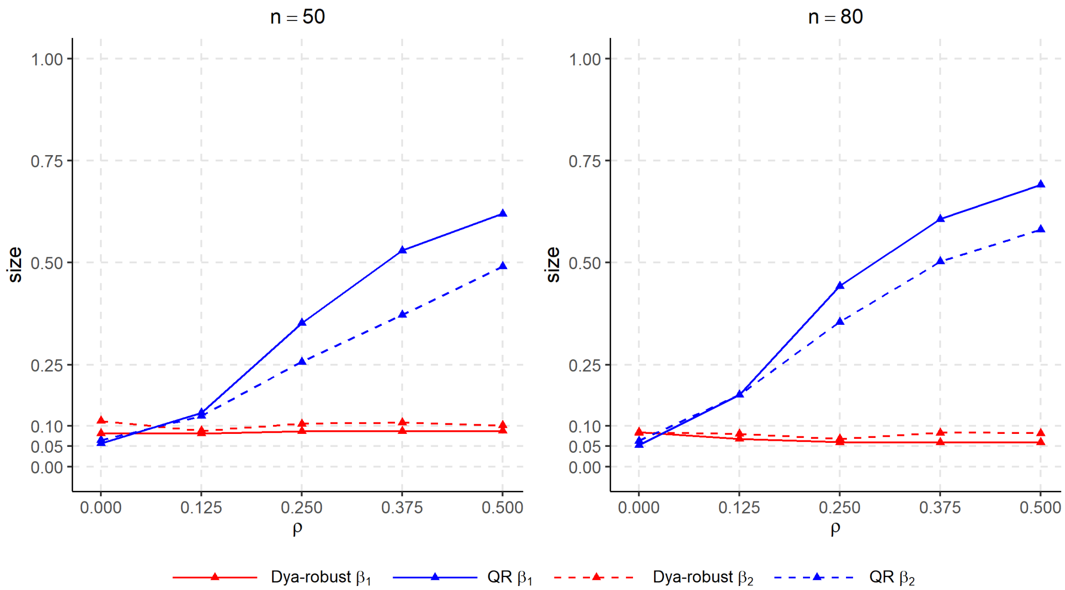
We consider a dense dyadic structure that dependence in all dyads between every node exists. For example, if we have $N = 50$ nodes, then there are $D = \frac{50\cdot 49}{2} = 1225$ dyads in total.

To show the inference performance of robust $t-$statistics, we obtain standard errors by estimating Ω above with $\hat{\Omega} = \hat{J}^{-1}\hat{\Sigma}\hat{J}^{-1}$. For $\hat{\Sigma}$, we follow our lemma 3.2 and use lemma 3.1 for $\hat{J}$. As a comparison, we also show the results with standard error of QR as if the data is independent but nonidentical distributed. For the standard error from traditional QR, we apply "quantreg" R package and follow the traditional Powell's sandwich estimator.

In the simulation, we consider different choices of number of nodes $n = 30, 50, 80$ and dependency level $\rho = 0, 0.25, 0.5$ at three quantile levels $\tau = 0.2, 0.5, 0.8$. We run our simulations 1000 times for each case. We mainly focus on comparing the empirical sizes of β_1 and β_2 for robust inference and traditional inference.

We first show the effect of different dependency levels on two inference procedures in figure 1 with $\tau = 0.5$ and $n = 50, 80$. It shows that the traditional QR has empirical sizes around 5% under the case $\rho = 0$ in which we have i.i.d data. The size performance deteriorates quickly as we impose more and more dependency across dyads. When the DGP reaches the highest dependency level, $\rho = 0.5$, the empirical size for traditional QR is more than 50%. As for our robust $t-$statistic, it controls the empirical size around 5% level well

Figure 1: Empirical size comparison with $n = 50, 80$, $\tau = 0.5$



as we increase ρ. These results show that our robust inference procedure is suitable even for the high dyadic dependence situation. Additional figures for $\tau = 0.2, 0.8$ are in the appendix.

In table 1, we summarize our simulation results under the dense model at $\tau = 0.5$. We also show empirical sizes as the sample size increases; hence the dyad size increases. It is easy to find that the performances of robust QR inference converge to 5% as we increase the sample size. Especially in the case with the largest sample size $n = 100$, the empirical size of β_1 almost coincides with the truth. A similar pattern can also be found for $\tau = 0.2, 0.8$ in the appendix, though their performances are slightly worse than the medium case.

5 Conclusion

In this paper, we consider a robust inference procedure for quantile regression with dyadic data. We show the necessity to estimate the robust covariance matrix with a dyadic data structure and propose a consistent covariance matrix estimation. In addition, we confirm that $t-$statistic is asymptotically normal under two different asymptotic frameworks. Numerical simulations show that our robust inference procedure can achieve correct empirical size while the non-robust traditional procedure suffers severe incorrect size problems under the dyadic dependence data structure. Our results suggest that consistent covariance matrix estimation and robust inference procedure are suitable for empirical practice.

Table 1: Empirical sizes at 5% level with $\tau = 0.5$

	n	30			50		
	D	435			1225		
	ρ	0	0.25	0.50	0	0.25	0.50
β_1	QR	0.057	0.228	0.509	0.057	0.352	0.62
	Robust QR	0.14	0.087	0.105	0.081	0.087	0.088
β_2	QR	0.061	0.162	0.383	0.065	0.257	0.491
	Robust QR	0.149	0.145	0.17	0.112	0.105	0.101
	n	80			100		
	D	3160			4950		
	ρ	0	0.25	0.50	0	0.25	0.50
β_1	QR	0.053	0.443	0.691	0.054	0.472	0.674
	Robust QR	0.085	0.06	0.059	0.065	0.058	0.051
β_2	QR	0.063	0.355	0.581	0.048	0.391	0.624
	Robust QR	0.083	0.068	0.082	0.066	0.081	0.088

References

Andrews, D. W. (1989): "Asymptotics for semiparametric econometric models: II. Stochastic equicontinuity and nonparametric kernel estimation," .

Aronow, P. M., C. Samii, and V. A. Assenova (2015): "Cluster–robust variance estimation for dyadic data," *Political Analysis*, 23, 564–577.

Cameron, A. C. and D. L. Miller (2014): "Robust inference for dyadic data," *Unpublished manuscript, University of California-Davis.*

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

Hendricks, W. and R. Koenker (1992): "Hierarchical spline models for conditional quantiles and the demand for electricity," *Journal of the American statistical Association*, 87, 58–68.

Janson, S. (1988): "Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs," *The Annals of Probability*, 305–312.

Kim, T.-H. and H. White (2003): "Estimation, inference, and specification testing for possibly misspecified quantile regression," in *Maximum likelihood estimation of misspecified models: twenty years later*, Emerald Group Publishing Limited.

KNIGHT, K. (1998): "Limiting distributions for L1 regression estimators under general conditions," *Annals of statistics*, 755–770.

KOENKER, R. (2005): *Quantile Regression*, Cambridge University Press.

KOENKER, R. AND G. BASSETT (1978): "Regression quantiles," *Econometrica: journal of the Econometric Society*, 33–50.

NIU, F. (2021): "Essays on Econometrics of Dyadic Data," Ph.D. thesis, University of California, Berkeley.

PARENTE, P. M. AND J. M. S. SILVA (2016): "Quantile regression with clustered data," *Journal of Econometric Methods*, 5, 1–15.

POLLARD, D. (1991): "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, 7, 186–199.

POWELL, J. L. ET AL. (1991): "Estimation of monotonic regression models under quantile restrictions," *Nonparametric and semiparametric methods in Econometrics*, 357–384.

TABORD-MEEHAN, M. (2019): "Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic," *Journal of Business & Economic Statistics*, 37, 671–680.

A Proofs

In this section, we show detailed proofs of all results in section 3. Since we consider two different asymptotic frameworks and the first framework is similar but simpler than the second. So we will focus on theoretical proofs under the second asymptotic framework and sketch the first case.

A.1 Proof of theorem 3.1

The proof of theorem 3.1 is a simplified version of the proof in theorem 3.2. Under assumption 5, the argument is the same except the scaling factor is D instead of DN^r. Therefore, we provide the detailed proof for theorem 3.2 below and the proof of theorem 3.1 follows naturally.

A.2 Proof of theorem 3.2

Proof. To show the asymptotic distribution of QR estimator in theorem 3.2, we follow the Knight (1998) approach. We first consider x_d as a scalar random variable ($p = 1$) and we extend the univariate result into multivariate cases with Cramer-Wold device in the end.

We consider

$$Z(\delta) = \frac{1}{N^r}\left[\sum_{d=1}^{D} \rho_\tau\left(u_d - \frac{x_d\delta}{\sqrt{\frac{D}{N^r}}}\right) - \rho_\tau(u_d)\right]$$

where $u_d = y_d - x_d\beta$. It is obvious that $Z(\delta)$ is convex since ρ_τ is convex. Moreover, when $\hat{\delta} = \sqrt{\frac{D}{N^r}}\left(\hat{\beta} - \beta\right)$ minimize $Z(\delta)$ for any N. Using Knight (1998) identity, we have $Z(\delta) = Z_1(\delta) + Z_2(\delta)$, in which

$$Z_1(\delta) = -\sqrt{\frac{1}{DN^r}}\sum_{d=1}^{D} x_d\delta\psi_\tau(u_d)$$

$$Z_2(\delta) = \frac{1}{N^r}\sum_{d=1}^{D}\int_0^{\sqrt{\frac{N^r}{D}}x_d\delta} 1(u_d \le s) - 1(u_d \le 0)\, ds$$

For $Z_1(\delta)$, we consider applying Janson's central limit theorem w.r.t. $x_d\psi_\tau(u_d)$. It is natural to find $\{x_d\psi_\tau(u_d)\}_{d=1}^{D}$ forms a dependency graph with vertexes $\{x_d\psi_\tau(u_d)\}_{d=1}^{D}$ and edges $\left\{(x_d\psi_\tau(u_d), x_{d'}\psi_\tau(u_{d'}))_{\phi(d)\cap\phi(d')\neq\emptyset}\right\}$ since x_d and u_d are independent. By our definition, the maximal degree $\tilde{D}$ of this dependency graph is $2\left(M^H - 1\right)$. By assumption 3, we know $|x_d\psi_\tau(u_d)| \le |x_d| < C$ for some universal constant C. Define $\Sigma_D = var\left(\sum_{d=1}^{D} x_d\psi_\tau(u_d)\right)$, we then consider

$$\tilde{L}=\frac{\left(\frac{D}{2(M^H-1)}\right)^{\frac{1}{l}}2\left(M^H-1\right)C}{\sqrt{\Sigma_D}}=\frac{\left(\frac{D}{2(M^H-1)}\right)^{\frac{1}{l}}2\left(M^H-1\right)C}{\sqrt{DN^r}}\frac{1}{\sqrt{\frac{1}{DN^r}\Sigma_D}}$$

By definition of M^H, M^l and assumption 6, we know

$$\begin{aligned}\frac{\left(\frac{D}{2(M^H-1)}\right)^{\frac{1}{l}}2\left(M^H-1\right)C}{\sqrt{DN^r}}&\leq\frac{\left(\frac{M^HN}{4(M^H-1)}\right)^{\frac{1}{l}}2CN}{C\sqrt{\frac{N^2}{2}N^{\frac{r}{2}}}}\\&=\frac{O\left(N^{\frac{1}{l}}\right)O\left(N\right)}{O\left(N\right)N^{\frac{r}{2}}}\\&=O\left(N^{\frac{1}{l}-\frac{r}{2}}\right)\end{aligned}$$

Choose sufficiently large $l>\frac{2}{r}$, then we obtain $\frac{\left(\frac{D}{2(M^H-1)}\right)^{\frac{1}{l}}2\left(M^H-1\right)C}{\sqrt{DN^r}}\to 0$. In addition, by assumption 4, we have $\frac{1}{\sqrt{\frac{1}{DN^r}\Sigma_D}}\to\frac{1}{\sqrt{\Sigma}}$. Therefore, by Janson's central limit theorem, we obtain

$$\frac{\sum_{d=1}^{D}x_d\psi_\tau\left(u_d\right)}{\sqrt{\Sigma_D}}\xrightarrow{d}N\left(0,1\right)$$

and

$$\begin{aligned}Z_1\left(\delta\right)&=-\delta\sqrt{\frac{1}{DN^r}}\sum_{d=1}^{D}x_d\psi_\tau\left(u_d\right)\\&=-\delta\frac{\sum_{d=1}^{D}x_d\psi_\tau\left(u_d\right)}{\sqrt{\Sigma_D}}\left(\frac{\sqrt{\Sigma_D}}{\sqrt{DN^r}}\right)\\&\xrightarrow{d}-\delta W\end{aligned}$$

where $W\sim N\left(0,\Sigma\right)$.

For $Z_2\left(\delta\right)=\sum_{d=1}^{D}\int_0^{\sqrt{\frac{N^r}{D}}x_d\delta}1\left(u_d\leq s\right)-1\left(u_d\leq 0\right)ds$, we can rewrite it as

$$Z_2\left(\delta\right)=\frac{1}{N^r}\sum_{d=1}^{D}E\left[EZ_{2d}\left(\delta|x_d\right)\right]+\frac{1}{N^r}\sum_{d=1}^{D}\left(Z_{2d}\left(\delta\right)-E\left[EZ_{2d}\left(\delta|x_d\right)\right]\right)$$

For the first term, we have

$$\begin{aligned}
\frac{1}{N^r}\sum_{d=1}^{D}E\left[EZ_{2d}\left(\delta|x_d\right)\right] &= \frac{1}{N^r}\sum_{d=1}^{D}E\int_0^{\sqrt{\frac{N^r}{D}}x_d\delta}F_{u_d}\left(s|x_d\right)-F_{u_d}\left(0|x_d\right)ds \\
&= \frac{1}{N^r}\sqrt{\frac{N^r}{D}}\sum_{d=1}^{D}E\int_0^{x_d\delta}F_{u_d}\left(\frac{t}{\sqrt{\frac{D}{N^r}}}|x_d\right)-F_{u_d}\left(0|x_d\right)dt \\
&= \frac{1}{N^r}\frac{N^r}{D}\sum_{d=1}^{D}E\int_0^{x_d\delta}\sqrt{\frac{D}{N^r}}\left(F_{u_d}\left(\frac{t}{\sqrt{\frac{D}{N^r}}}|x_d\right)-F_{u_d}\left(0|x_d\right)\right)dt \\
&= \frac{1}{D}\sum_{d=1}^{D}E\int_0^{x_d\delta}f_{u_d}\left(0|x_d\right)tdt+o\left(1\right) \\
&= \frac{1}{2D}\sum_{d=1}^{D}Ef_{u_d}\left(0|x_d\right)\delta x_d x_d\delta+o\left(1\right) \\
&\xrightarrow{p}\frac{1}{2}\delta J\delta
\end{aligned}$$

where the last convergence result follows assumption 4. In addition, we know that

$$\begin{aligned}
var\left(\frac{1}{N^r}Z_2\left(\delta\right)\right) &\leq \frac{1}{N^{2r}}E\left(\sum_{d=1}^{D}\int_0^{\sqrt{\frac{N^r}{D}}x_d\delta}1\left(u_d\leq s\right)-1\left(u_d\leq 0\right)ds\right)^2 \\
&\leq \frac{1}{N^r}D\sqrt{\frac{N^r}{D}}\max_d\left|x_d\delta\right|E\left(\sum_{d=1}^{D}\int_0^{\sqrt{\frac{N^r}{D}}x_d\delta}1\left(u_d\leq s\right)-1\left(u_d\leq 0\right)ds\right) \\
&\leq \sqrt{\frac{1}{N^rD}}\max_d\left|x_d\delta\right|\frac{1}{2}\sum_{d=1}^{D}Ef_{u_d}\left(0|x_d\right)\delta x_d x_d\delta+o\left(1\right) \\
&= o\left(1\right)
\end{aligned}$$

where the second inequality is based on Loève's c_r inequality with $r = 2$ and the third inequality is from the previous derivation on $\sum_{d=1}^{D}EZ_{2d}\left(\delta\right)$. By Chebyshev inequality, for any $\varepsilon > 0$, we have

$$P\left(\frac{1}{N^r}\left|\sum_{d=1}^{D}\left(Z_{2d}\left(\delta\right)-EZ_{2d}\left(\delta\right)\right)\right|\geq\varepsilon\right)\leq\frac{var\left(\frac{1}{N^r}Z_2\left(\delta\right)\right)}{\varepsilon}=o\left(1\right)$$

Therefore, $Z_2\left(\delta\right)\xrightarrow{p}\frac{1}{2}\delta J\delta$ and this implies $Z\left(\delta\right)\xrightarrow{d}Z_0\left(\delta\right)=-\delta W+\frac{1}{2}\delta J\delta$ is a convex limiting objective function. By the convexity lemma in Pollard (1991), we conclude that

$$\sqrt{\frac{D}{N^r}}\left(\hat{\beta}-\beta\right)=\hat{\delta}=\arg\min\left(\delta\right)\xrightarrow{d}\delta_0=\arg\min Z_0\left(\delta\right)$$

and then $\sqrt{\frac{D}{N^r}}\left(\hat{\beta}-\beta\right)\xrightarrow{d}N\left(0,J^{-1}\Sigma J^{-1}\right)$.

By Cramer Wold device, we can extend the above asymptotic distribution for the general case with $p\neq 1$. Hence the proof is complete. □

A.3 Proof of lemma 3.1

The procedure to show lemma 3.1 is a simplified version of the proof in lemma 3.2. The only difference is in the scaling factor. We refer to the next proof of lemma 3.2 for more details.

A.4 Proof of lemma 3.2

Proof. Our proof procedure is modified from Tabord-Meehan (2019). By definition and we express $\psi_\tau\left(\hat{u}_d\right)=\left(\psi_\tau\left(\hat{u}_d\right)-\psi_\tau\left(u_d\right)\right)+\psi_\tau\left(u_d\right)$ and $\psi_\tau\left(\hat{u}_{d'}\right)=\left(\psi_\tau\left(\hat{u}_{d'}\right)-\psi_\tau\left(u_{d'}\right)\right)+\psi_\tau\left(u_{d'}\right)$,

$$\begin{aligned}\frac{1}{DN^r}\hat{\Sigma}&=\frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\psi_\tau\left(\hat{u}_d\right)\psi_\tau\left(\hat{u}_{d'}\right)x_d x_{d'}^T\\&=\frac{1}{DN^r}\left(V_1+V_2+V_3+V_4\right)\end{aligned}$$

where

$$\begin{aligned}V_1&=\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\psi_\tau\left(u_d\right)\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T\\V_2&=\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\left(\psi_\tau\left(\hat{u}_d\right)-\psi_\tau\left(u_d\right)\right)\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T\\V_3&=\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\left(\psi_\tau\left(\hat{u}_{d'}\right)-\psi_\tau\left(u_{d'}\right)\right)\psi_\tau\left(u_d\right)x_d x_{d'}^T\\V_4&=\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\left(\psi_\tau\left(\hat{u}_d\right)-\psi_\tau\left(u_d\right)\right)\left(\psi_\tau\left(\hat{u}_{d'}\right)-\psi_\tau\left(u_{d'}\right)\right)x_d x_{d'}^T\end{aligned}$$

We consider to prove $\frac{1}{DN^r}V_1\xrightarrow{p}\Sigma$ and $\frac{1}{DN^r}V_i\xrightarrow{p}0$ for $i=2,3,4$. We first show $\frac{1}{DN^r}V_2\xrightarrow{p}0$

as an example , since V_3 and V_4 follow similar arguments. By the definition of ψ_τ, since $|1(a<b)-1(a<c)| \leq 1(|a-b|<|b-c|)$, for $\frac{1}{DN^r}V_2$, we have

$$\begin{aligned}
&\frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\left(\psi_\tau\left(\hat{u}_d\right)-\psi_\tau\left(u_d\right)\right)\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T\\
\leq&\frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\left|1\left(u_d<0\right)-1\left(\hat{u}_d<0\right)\right|\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T\\
=&\frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}\left|1\left(y_d<x_d^T\beta\right)-1\left(y_d<x_d^T\hat{\beta}\right)\right|\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T\\
\leq&\frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}1\left(\left|y_d-x_d^T\beta\right|<\left|x_d^T\beta-x_d^T\hat{\beta}\right|\right)\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T\\
\leq&\frac{1}{DN^r}\sum_{d=1}^{D}\sum_{d'=1}^{D}1_{d,d'}1\left(\frac{|u_d|}{\|x_d\|}<\left\|\beta-\hat{\beta}\right\|\right)\psi_\tau\left(u_{d'}\right)x_d x_{d'}^T
\end{aligned}$$

where the last inequality follows Cauchy-Schwarz inequality. From the previous asymptotic distribution theorem, we know $\left\|\beta-\hat{\beta}\right\|=o_p(1)$ as $D\to\infty$. Therefore, by assumption 3, we have $\frac{1}{DN^r}V_2\xrightarrow{p}0$.

For $\frac{1}{DN^r}V_1$, we know $\lim_{N\to\infty}E\frac{1}{DN^r}V_1=\Sigma$ by definition. Hence, it is enough to show

$$\lim_{N\to\infty}var\left(\frac{1}{DN^r}V_1\right)=0$$

We know that

$$var\left(\frac{1}{DN^r}V_1\right)=\frac{1}{D^2N^{2r}}\sum_{i=1}^{D}\sum_{j=1}^{D}\sum_{k=1}^{D}\sum_{l=1}^{D}cov\left(1_{i,j}\psi_\tau\left(u_i\right)\psi_\tau\left(u_j\right)x_i x_j^T,1_{k,l}\psi_\tau\left(u_k\right)\psi_\tau\left(u_l\right)x_k x_l^T\right)$$

To show this summation has limit 0, we should determine how many terms satisfy

$$\begin{aligned}
&cov\left(1_{i,j}\psi_\tau\left(u_i\right)\psi_\tau\left(u_j\right)x_i x_j^T,1_{k,l}\psi_\tau\left(u_k\right)\psi_\tau\left(u_l\right)x_k x_l^T\right)\\
=&E\left(1_{i,j}\psi_\tau\left(u_i\right)\psi_\tau\left(u_j\right)x_i x_j^T\cdot 1_{k,l}\psi_\tau\left(u_k\right)\psi_\tau\left(u_l\right)x_k x_l^T\right)\\
&-E\left(1_{i,j}\psi_\tau\left(u_i\right)\psi_\tau\left(u_j\right)x_i x_j^T\right)E\left(1_{k,l}\psi_\tau\left(u_k\right)\psi_\tau\left(u_l\right)x_k x_l^T\right)\\
\neq&0
\end{aligned}$$

There are two sufficient conditions need to be hold:

1. $\phi(i)\cap\phi(j)\neq\emptyset$ and $\phi(k)\cap\phi(l)\neq\emptyset$. Otherwise, the covariance is 0 by definition.

2. $\phi(i) \cap \phi(k) \neq \emptyset$ or $\phi(i) \cap \phi(l) \neq \emptyset$ or $\phi(j) \cap \phi(k) \neq \emptyset$ or $\phi(j) \cap \phi(l) \neq \emptyset$. Otherwise the second line above is 0 by our assumption 1 and $E\psi_\tau(u_d) = 0$.

Denote S be the set of tuples $(i, j, k, l) \in \{1, \dots, D\}^4$ with the above two sufficient conditions hold. We need to find an upper bound for the cardinality of S, $|S|$.

Following the same argument in Tabord-Meehan (2019). For any i, there are $O\left(M^H\right)$ indices j satisfying $\phi(i) \cap \phi(j) \neq \emptyset$. For those i and j, we first consider index k. There are $O\left(M^H\right)$ indices k with $\phi(i) \cap \phi(k) \neq \emptyset$ or $\phi(j) \cap \phi(k) \neq \emptyset$. For those i, j and k above, there are $O\left(M^H\right)$ indices l with $\phi(k) \cap \phi(l) \neq \emptyset$. We have a similar argument if we first consider index l. Therefore, we have

$$|S| = D \cdot O\left(M^H\right) = O\left(N^5\right)$$

where the last equality is from the definition of M^H and assumption 6.

Since the summands in V_1 are bounded by 3, we have

$$\begin{aligned} var\left(\frac{1}{DN^r}V_1\right) &\leq \frac{1}{D^2N^{2r}}O\left(N^5\right) \\ &\leq \frac{C}{N^{4+2r}}O\left(N^5\right) \\ &= o(1) \end{aligned}$$

for some positive constant C. The second inequality follows from 6 and the third equality holds for all $r > \frac{1}{2}$. Therefore, we obtain $\frac{1}{DN^r}V_1 \xrightarrow{p} \Sigma$.

Combining with results on V_2, V_3 and V_4, we show $\frac{1}{DN^r}\hat{\Sigma} \xrightarrow{p} \Sigma$ for $r > \frac{1}{2}$. $\square$

A.5 Proof of proposition 3.1

Proof. We first define $J_D = \frac{1}{2c_D D}\sum_{d=1}^{D} 1\left(|u_d| < c_D\right) x_d x_d^T$ and $\tilde{J} = \frac{1}{2c_D D}\sum_{d=1}^{D} 1\left(|\hat{u}_d| < \hat{c}_D\right) x_d x_d^T$ and we shall show (i) $|J_D - J| \xrightarrow{p} 0$, (ii)$\left|\tilde{J} - J_D\right| \xrightarrow{p} 0$ and (iii) $\left|\tilde{J} - \frac{1}{D}\hat{J}\right| \xrightarrow{p} 0$. Then the conclusion follows by triangular inequality.

For (i), we start with

$$\begin{aligned}
EJ_D &= E\frac{1}{2c_D D}\sum_{d=1}^{D} 1\left(|u_d| < c_D\right) x_d x_d^T \\
&= \frac{1}{D}\sum_{d=1}^{D} E\left(\frac{F_{u_d}\left(c_D|x_d\right) - F_{u_d}\left(-c_D|x_d\right)}{2c_D} x_d x_d^T\right) \\
&= E\left(\frac{1}{D}\sum_{d=1}^{D}\left(f_{u_d}\left(\xi_d|x_d\right) x_d x_d^T\right)\right) + o\left(1\right)
\end{aligned}$$

where $\xi_d \in [-c_D, c_D]$ and $\xi_d = o(1)$ by assumption 8. Since x_d has bounded support, by dominated convergence theorem, we have $|EJ_D - J| \to 0$. Following Kim and White (2003), by double array law of large number in Andrews (1989), we obtain $|J_D - J| \xrightarrow{p} 0$.

For (ii), by the facts that $\hat{u}_{di} = u_{di} - x_{di}^T\left(\hat{\beta} - \beta\right)$ and $|1(a < b) - 1(a < c)| \leq 1(|a - b| < |b - c|)$, the (i, j) −th element is given by

$$\begin{aligned}
&\left|\frac{1}{2c_D D}\sum_{d=1}^{D}\left(1\left(|\hat{u}_d| < \hat{c}_D\right) - 1\left(|u_d| < c_D\right)\right) x_{di} x_{dj}^T\right| \\
\leq &\frac{1}{2c_D D}\sum_{d=1}^{D} 1\left(|u_d + c_D| < a_D\right)|x_{di}|\,|x_{dj}| \\
&+ \frac{1}{2c_D D}\sum_{d=1}^{D} 1\left(|u_d - c_D| < a_D\right)|x_{di}|\,|x_{dj}| \\
= &U_{1D} + U_{2D}
\end{aligned}$$

where $a_D = |c_D - \hat{c}_D| + \left\|\hat{\beta} - \beta\right\| \|x_{Di}\|$.

Now we consider proving $U_{1D} \xrightarrow{p} 0$, while $U_{2D} \xrightarrow{p} 0$ follows exactly the same procedure. Denote three event sets $A_{1D} = \{U_{1D} > \eta\}$, $A_{2D} = \left\{c_D^{-1}\left\|\hat{\beta} - \beta\right\| \leq z\right\}$ and $A_{3D} = \left\{c_D^{-1}|c_D - c_D| \leq z\right\}$ for a positive constant z. We have

$$\begin{aligned}
P\left(A_{1D}\right) &= P\left(U_{1D} > \eta\right) \\
&\leq P\left(A_{1D} \cap A_{2D} \cap A_{3D}\right) + P\left(A_{2D}^c\right) + P\left(A_{3D}^c\right)
\end{aligned}$$

By theorem 3.2, we have $\sqrt{\frac{D}{N^r}}\left(\hat{\beta} - \beta\right) = O_p(1)$. Together with assumption 8 that $\frac{1}{c_D} = o\left(\sqrt{\frac{D}{N^r}}\right)$, we have $\lim_{D\to\infty} P\left(A_{2D}^c\right) = 0$. Moreover, we have $\lim_{D\to\infty} P\left(A_{3D}^c\right) = 0$ by assumption 8 as $\frac{\hat{c}_D}{c_D} \xrightarrow{p} 1$.

Then, we consider the case that $c_D^{-1}\left\|\hat{\beta}-\beta\right\| \leq z$ and $c_D^{-1}\left|c_D-c_D\right| \leq z$, we have

$$\left|a_D\right| \leq c_D z\left(1+\left\|x_{Di}\right\|\right)$$

Therefore, by Markov inequality, we have

$$\begin{aligned}
& P\left(A_{1D} \cap A_{2D} \cap A_{3D}\right) \\
\leq & \frac{1}{\eta 2 c_D D} \sum_{d=1}^{D} E\left(\int_{-c_D z\left(1+\left\|x_{Di}\right\|\right)-c_D}^{c_D z\left(1+\left\|x_{Di}\right\|\right)-c_D} f_{u_d}\left(\xi \mid x_d\right) d\xi\left|x_{di}\right|\left|x_{dj}\right|\right) \\
\leq & \frac{1}{\eta 2 c_D D} \sum_{d=1}^{D} E\left(\int_{-c_D z\left(1+\left\|x_{Di}\right\|\right)-c_D}^{c_D z\left(1+\left\|x_{Di}\right\|\right)-c_D} \bar{f} d\xi\left|x_{di}\right|\left|x_{dj}\right|\right) \\
= & \frac{z \bar{f}}{\eta} E\left(\left(1+\left\|x_{Di}\right\|\right)\left|x_{di}\right|\left|x_{dj}\right|\right) \\
< & \infty
\end{aligned}$$

Since we can take $z>0$ arbitrarily small $z \rightarrow 0$, we conclude that $U_{1D} \xrightarrow{p} 0$. Hence $\left|\tilde{J}-J_D\right| \xrightarrow{p} 0$.

For (iii), since $\frac{1}{D}\hat{J}-\tilde{J}=\left(\frac{c_D}{\hat{c}_D}-1\right)\tilde{J}$ and $\tilde{J}=O_p(1)$, by assumption 8 that $\frac{\hat{c}_D}{c_D} \xrightarrow{p} 1$, we obtain $\left|\tilde{J}-\frac{1}{D}\hat{J}\right| \xrightarrow{p} 0$.

Using triangular inequality on (i), (ii) and (iii), our final result follows. □

B Additional simulation results

In this section, we report additional Monte Carlo simulations with $\tau=0.2, 0.8$. The performances are similar to the case $\tau=0.5$. For β_1, the empirical sizes are around 0.05 as we increase the number of dyads. For β_2, although there are some size distortions, their performance is still better than the traditional QR inference. The first two figures show the change in empirical sizes when the dependency level ρ increases. The last two tables show the performance of empirical sizes as we increase n.

Figure 2: Empirical size comparison with $n = 50, 80$, $\tau = 0.2$

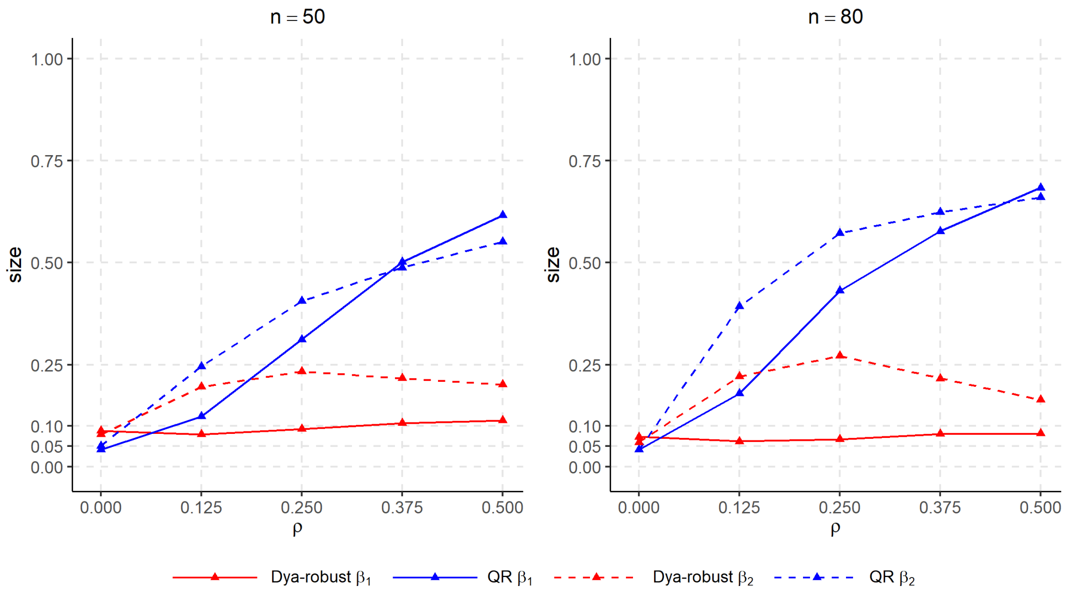


Figure 3: Empirical size comparison with $n = 50, 80$, $\tau = 0.8$

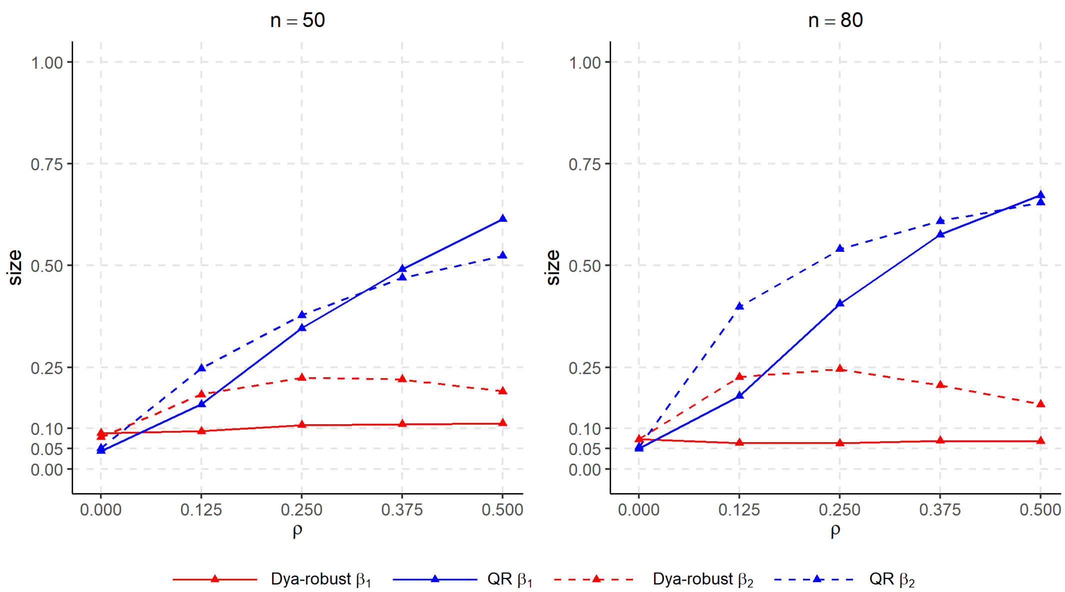


Table 2: Empirical sizes at 5% level with $\tau = 0.2$

	n	30			50		
	D	435			1225		
	ρ	0	0.25	0.50	0	0.25	0.50
β_1	QR	0.037	0.214	0.522	0.042	0.312	0.616
	Robust QR	0.103	0.126	0.151	0.088	0.092	0.113
β_2	QR	0.028	0.247	0.428	0.051	0.406	0.551
	Robust QR	0.107	0.235	0.278	0.079	0.233	0.201
	n	80			100		
	D	3160			4950		
	ρ	0	0.25	0.50	0	0.25	0.50
β_1	QR	0.042	0.431	0.684	0.048	0.451	0.692
	Robust QR	0.073	0.067	0.081	0.06	0.055	0.066
β_2	QR	0.042	0.572	0.66	0.055	0.622	0.672
	Robust QR	0.06	0.272	0.164	0.071	0.291	0.164

Table 3: Empirical sizes at 5% level with $\tau = 0.8$

	n	30			50		
	D	435			1225		
	ρ	0	0.25	0.50	0	0.25	0.50
β_1	QR	0.038	0.221	0.509	0.044	0.346	0.614
	Robust QR	0.098	0.111	0.138	0.088	0.108	0.112
β_2	QR	0.038	0.247	0.445	0.051	0.378	0.524
	Robust QR	0.122	0.223	0.29	0.078	0.224	0.191
	n	80			100		
	D	3160			4950		
	ρ	0	0.25	0.50	0	0.25	0.50
β_1	QR	0.05	0.406	0.673	0.044	0.442	0.694
	Robust QR	0.073	0.063	0.068	0.066	0.058	0.058
β_2	QR	0.054	0.54	0.655	0.061	0.657	0.701
	Robust QR	0.073	0.245	0.159	0.074	0.278	0.142