Chen, Heng

Working Paper

Sheep in wolf's clothing: Using the least squares criterion for quantile estimation

Bank of Canada Working Paper, No. 2014-24

Provided in Cooperation with:
Bank of Canada, Ottawa

Suggested Citation: Chen, Heng (2014) : Sheep in wolf's clothing: Using the least squares criterion for quantile estimation, Bank of Canada Working Paper, No. 2014-24, Bank of Canada, Ottawa, https://doi.org/10.34989/swp-2014-24

This Version is available at:
https://hdl.handle.net/10419/103028

BANK OF CANADA
BANQUE DU CANADA

Working Paper/Document de travail
2014-24

Sheep in Wolf's Clothing: Using the Least Squares Criterion for Quantile Estimation

by Heng Chen

www.bank-banque-canada.ca

Bank of Canada Working Paper 2014-24

June 2014

Sheep in Wolf's Clothing: Using the Least Squares Criterion for Quantile Estimation

by

Heng Chen

Currency Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
chhe@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I would like to thank participants at 2012 MEG, 2012 Microeconometrics Symposium at Carleton University, Bank of Canada's Brown Bag Seminar, 2013 MAESG, and also Jushan Bai, Ba M. Chu, Jean-Marie Dufour, Ben Fung, Yanqin Fan, Chris Henry, Kim P. Huynh, David T. Jacho-Chavez, Shakeeb Khan, Arthur Lewbel, Jack Porter, John Rust, Gregor Smith and Marcel Voia for their helpful suggestions. I would also like to thank Maren Hansen for her editorial assistance.

Abstract

Estimation of the quantile model, especially with a large data set, can be computationally burdensome. This paper proposes using the Gaussian approximation, also known as quantile coupling, to estimate a quantile model. The intuition of quantile coupling is to divide the original observations into bins with an equal number of observations, and then compute order statistics within these bins. The quantile coupling allows one to apply the standard Gaussian-based estimation and inference to the transformed data set. The resulting estimator is asymptotically normal with a parametric convergence rate. A key advantage of this method is that it is faster than the conventional check function approach, when handling a sizable data set.

JEL codes: C13; C14; C21
Bank classification: Econometric and statistical methods

Résumé

L'estimation d'un modèle de régression quantile peut nécessiter des calculs considérables, tout particulièrement si ce modèle s'appuie sur un grand ensemble de données. Une approximation gaussienne est ici proposée (couplage quantile) comme méthode d'estimation. Le principe du couplage quantile consiste à diviser les observations initiales en sous-groupes composés d'un nombre égal d'observations, puis à calculer des statistiques d'ordre pour chacun de ces sous-groupes. Ce couplage permet d'appliquer aux données transformées une méthode d'estimation et d'inférence gaussienne standard. L'estimateur obtenu suit une loi asymptotiquement normale et son taux de convergence est paramétrique. L'un des principaux avantages de cette méthode réside dans le fait qu'elle assure un traitement plus rapide des grands jeux de données que l'approche classique par fonction de perte.

Classification JEL : C13; C14; C21
Classification de la Banque : Méthodes économétriques et statistiques

1 Introduction

Estimation of the quantile model, especially with a large data set, can be computationally burdensome. This paper proposes using the Gaussian approximation, also known as quantile coupling, to estimate a quantile model. A key ingredient of the estimator is to compute local order statistics, where the original N observations are first divided into J bins with an equal number of observations, and then the order statistics are computed in these bins. The quantile coupling allows one to apply a standard Gaussian-based estimation and inference to the local order statistics. The Gaussian-based estimates applied to the local order statistics are consistent and have an asymptotically normal distribution with the $\sqrt{N}$ convergence rate. In addition to reducing the effective sample size to J, the proposed estimator has a closed-form expression, thus it is computationally efficient to handle a sizable data set. In a simulation study, this paper demonstrates that this superior speed occurs for sample sizes N larger than $20,000$.

Using the existing Gaussian methods applied to the local order statistics, the new method offers the following advantages: (1) estimates are from the differentiable least squares criterion and are thus computationally faster to calculate than the least absolute deviation criterion based on the original data; (2) consistency and asymptotic normality of the estimators are easy to derive when the bin size m is properly chosen; and (3) inference of the estimators can be obtained directly from the OLS (ordinary least squares) software package based on the transformed data, without estimating the quantile-density function, as in Koenker and Xiao (2002).

In addition, the paper generalizes the results in Brown et al. (2008) by allowing for (1) multivariate random design, (2) the quantile coupling for the arbitrary order statistics and (3) conditional heteroskedasticity of the error term. Notice that Brown et al. (2008) focus on the univariate fixed-design median models with homoscedastic errors.

The paper is organized as follows. Section 2 explains how the local order statistics for multivariate random design are computed, and proposes the new method. Section 3 discusses Monte Carlo simulation results and Section 4 concludes. All proofs are included in the appendix.

2 Proposed Estimator

We first introduce the classical check function approach for estimating the quantile model. Then we discuss the computation of local order statistics, and construct the estimator from these transformed data as if they were generated by a Gaussian data-generating process.

2.1 Parametric quantile model

In this paper, we focus on the location-scale λth quantile model studied by Koenker and Bassett (1982) and Koenker and Xiao (2002). This model allows for data generated from the same distributional family, but the parameters of the distributions, such as the location and scale, vary with the observation index.

For $i = 1, \cdots, N$, Y is a scalar random variable, X is a $d_X \times 1$ vector of explanatory variables and e is a disturbance error, such that

$$Y_i = \alpha + X_i^T \beta + e_i, \qquad \text{where } e_i \equiv \sigma(X_i) \cdot U_i. \tag{1}$$

In order to estimate α and β, we minimize the check function as follows:

$$\arg\min_{\alpha,\beta} \sum_{i=1}^{N} \left[\lambda - I\left\{Y_i - \alpha - X_i^T\beta < 0\right\}\right]\left(Y_i - \alpha - X_i^T\beta\right). \tag{2}$$

The finite and large sample properties of this method are summarized in Koenker (2005), and numerical implementations are discussed in Portnoy and Koenker (1997). However, its limiting distribution needs to work on the non-differentiability of the criteria function, and the rounding and approximation have to be done in the computation in order to find the unique numerical solution. In addition to these issues, the inference based on the check function approach needs to non-parametrically estimate the density of U, which is sensitive to the choice of smoothing parameter in finite samples. The next subsection introduces the new method applied to local order statistics and shows that the method is easy to compute, avoiding the estimated density function for the inference.[1]

2.2 Local order statistics

To find local order statistics, we first divide the N observations into J groups (bins) with m observations in each bin, and then compute the order statistics value of the dependent variable Y in each bin. We denote $\lceil \cdot \rceil$ to be the rounded-up integer. Given the J bins, the transformed data set is generated as

$$\mathcal{Y}_{j,\lambda} \equiv \text{the } \lceil \lambda m \rceil \text{ th smallest value of } Y \text{ in the } j\text{th bin}, \qquad 1 \le j \le J.$$

We set

$$\begin{aligned} \mathcal{E}_{j,\lambda} &\equiv \text{the } \lceil \lambda m \rceil \text{ th smallest value of } e \text{ in the } j\text{th bin}, \qquad 1 \le j \le J, \\ \theta_j(\alpha, \beta) &\equiv \mathcal{Y}_{j,\lambda} - \mathcal{E}_{j,\lambda}, \end{aligned}$$

then $\mathcal{Y}_{j,\lambda}$ can be written as

$$\mathcal{Y}_{j,\lambda} = \theta_j(\alpha, \beta) + \mathcal{E}_{j,\lambda}, \qquad 1 \le j \le J. \tag{3}$$

By working on the local order statistics, the error e_i is transformed to $\mathcal{E}_{j,\lambda}$. Lemma 2.1 shows that $\mathcal{E}_{j,\lambda}$ can be closely approximated using Gaussian random variables. Then, according to Lemma 2.2, $\theta_j(\alpha, \beta)$ is approximated by $\alpha + \mathcal{X}_{[j]}^T \beta$, in which the vector $\mathcal{X}_{[j]}^T$ is the coordinates of endpoints in the jth bin (which will be precisely defined later).

Assumption 1: $\{U_i\}_{i=1}^n$ is a sequence of i.i.d. errors that is also independent of X_i, and U_i has an unknown density function f_U, such that $\int_{-\infty}^{0} f_U(u)du = \lambda$ and $|f_U(u) - f_U(0)| \le Cu^2$ in an open neighborhood of 0 for some constant C.

[1] Another "direct" inference approach for the quantile model is by Fan and Liu (2012).

Assumption 2: The scale function $\sigma : \mathcal{R}^{d_x} \to \mathcal{R}$ is $\delta-$Holder continuous with $\delta \geq 1$.

Assumption 3: $\int |u|^{\varepsilon} f_U(u)du < \infty$ for some $\varepsilon > 0$.

Assumption 4: $\max_{1\leq j\leq J} \left\| \mathcal{X}_{[j]}^T - \mathcal{X}_{[j-1]}^T \right\|_1 = O_p(J^{-1})$, where $\mathcal{X}_{[j]}$ is the coordinates of endpoints in the jth bin.

Assumptions 1-3 are similar to the ones made for estimating the location-scale model by Equation (2). The $|f_U(u) - f_U(0)| \leq Cu^2$ in Assumption 1 is satisfied, for example, by the Cauchy distribution, the Laplace distribution and the t distribution. Assumption 2 states the necessary smooth condition for the scale function $\sigma(\cdot)$ in Equation (1), so that the heteroskedastic error e can be treated as homoscedastic within each bin. Assumption 3 guarantees the existence of moments of the order statistics, as in Cramer et al. (2002). Assumption 4 imposes the conditions on the spacings of $\mathcal{X}_{[j]}^T$, which can be substituted by low-level conditions on the density function of X, as in Gasser and Müller (1979).

Lemma 2.1 *When Assumptions 1-4 hold, we have*

$$\mathcal{E}_{j,\lambda} = \frac{1}{\sqrt{m}} \left[\frac{\sigma_j \sqrt{\lambda(1-\lambda)}}{f_U(0)} Z_j + \zeta_j \right],$$

where $\{Z_j\}_{j=1}^J$ *is i.i.d. standard Gaussian errors and* $\sigma_j \equiv \sigma\left(\mathcal{X}_{[j]}\right)$*. In addition, for all* $l > 0$*,*

$$E|\zeta_j|^l = O\left[\left(\frac{\sqrt{m}}{J^{\delta}} \right)^l + \left(\frac{1}{m} \right)^l \right].$$

Lemma 2.2 *When Assumption 4 holds, we have*

$$\theta_j(\alpha, \beta) - \alpha - \mathcal{X}_{[j]}^T \beta = O_p\left(\frac{1}{J} \right).$$

It is essential to have J bins. For the univariate fixed-design case, Brown et al. (2008) simply divide the data along the real line at equal interval lengths. However, when X is a multivariate random design, this is not possible. In particular, it is more difficult to bin the data in order to have an equal number of observations in each bin (see the following remark). Thus we propose using conditional (sequential) ordering, conducted on one of the marginal sets of the covariates, conditional on ordering within other marginal sets of observations. Heuristically speaking, the bins in the multivariate case are constructed by slicing according to the chosen ordered values of the first covariate, then internally slicing according to the ordered values of the second covariate within that slice, and so on.

Remark: Although bins with an unequal number of observations are allowed (for example, the data are binned with equal lengths in each dimension of X), the current bins with equal numbers of observations provide a homoscedastic Gaussian approximation for $\mathcal{E}_{j,\lambda}$, where the variance of this Gaussian random variable depends on m, the number of observations in the bin.

2.3 Bivariate case

We illustrate the conditional ordering with a bivariate case, $X^T \equiv \{X_1, X_2\}$, where $J = J_1 \times J_2$. The random design with more than the bivariate could be easily extended from Wald (1943).

Figure 1 provides an illustrative example of $\lambda = 1/4$. In Step 1, the 250 scattered observations are divided into 5 slices along the X_1 axis, ensuring an equal 50 observations in each slice. In Step 2, each slice is divided into another 5 slices along the X_2 axis, so that 10 observations are in each bin. This results in a total of 25 bins. Step 3 computes the quartile of Y in each bin, and its corresponding coordinates of $\{X_1, X_2\}$ are marked as red solid points.

The algorithm for computing local order statistics under the bivariate case is as follows:

Step 1 (Order X_1): We reorder the original sample $\{X_{1,i}, X_{2,i}, Y_i\}_{i=1}^N$ based on the order statistics of $X_{1,i}$, where $X_{[1,1]} \leq X_{[1,2]} \leq \cdots \leq X_{[1,N]}$. We define $S_{j_1} \equiv X_{[1,r_{1j}]}$ and $M_{j_1} \equiv X_{[1,s_{1j}]}$, where r_{1j} and s_{1j} denote some positive integers, $r_{1j} < s_{1j} < N$, and the number of observations between S_{j_1} and M_{j_1} are N/J_1, where $j_1 = 1, 2, \cdots, J_1$. We consider only those sample points $\{X_{1,a}, X_{2,a}, Y_a\}$ for which $X_{[1,r_{1j}]} < X_{1,a} < X_{[1,s_{1j}]}$, i.e., the sample points $\left\{X_{[1,r_{1j}+1]}, X_{(2,r_{1j}+1)}, Y_{(r_{1j}+1)}\right\}, \cdots, \{X_{[1,s_{1j}-1]}, X_{(2,s_{1j}-1)}, Y_{(s_{1j}-1)}\}$, which are ordered by $X_{[1,r_{1j}+1]}, \cdots, X_{[1,s_{1j}-1]}$ associated with the induced order statistics from $\left\{X_{(2,r_{1j}+1)}, Y_{(r_{1j}+1)}\right\}$ to $\left\{X_{(2,s_{1j}-1)}, Y_{(s_{1j}-1)}\right\}$.

Step 2 (Order X_2 conditional on the ordered X_1): We denote by $X^*_{[2,r_{1j}+1]}, \cdots, X^*_{[2,s_{1j}-1]}$ the values $X_{(2,r_{1j}+1)}, \cdots, X_{(2,s_{1j}-1)}$ arranged in order of increasing magnitude. We define $S_{j_2} \equiv X^*_{[2,r_{2j}]}$ and $M_{j_2} \equiv X^*_{[2,s_{2j}]}$, where r_{2j} and s_{2j} denote some positive integers for which $r_{2j} < s_{2j} < s_{1j} - r_{1j} - 1$, and the number of observations between S_{j_2} and M_{j_2} are $m = N/(J_1 \times J_2)$, where $j_2 = 1, 2, \cdots, J_2$.

Step 3 (Take order statistics): Choose $\mathcal{Y}_{j,\lambda}$ as the $\lceil \lambda m \rceil$th smallest value of Y in the range of $[S_{j_1}, M_{j_1}] \times [S_{j_2}, M_{j_2}]$. Here $\mathcal{X}^T_{[j]}$ is denoted as $\left\{X_{[1,s_{1j}-1]}, X^*_{[2,s_{2j}-1]}\right\}$, which are simply the coordinates of the northeast endpoints in the jth bin.

2.4 Asymptotic properties

Based on Steps 1-3, we regress $\mathcal{Y}_{j,\lambda}$ on $\left\{1, \mathcal{X}^T_{[j]}\right\}$ for $j = 1, \cdots, J$ in order to estimate $(\alpha, \beta)^T$. Proposition 2.3 justifies the above regression, which states the order of approximation errors between $\mathcal{Y}_{j,\lambda}$ and $\alpha + \mathcal{X}^T_{[j]}\beta + \sigma_j \sqrt{\lambda(1-\lambda)} Z_j / [\sqrt{m} f_U(0)]$.

Proposition 2.3 *When Assumptions 1-4 hold, then*

$$\sqrt{m}\mathcal{Y}_{j,\lambda} = \sqrt{m}\left[\alpha + \mathcal{X}^T_{[j]}\beta\right] + \frac{\sigma_j \sqrt{\lambda(1-\lambda)}}{f_U(0)} Z_j + \xi_j,$$

where $\{Z_j\}_{j=1}^J$ is i.i.d. standard Gaussian errors, and for all $l > 0$

$$E|\xi_j|^l = O\left[\left(\frac{\sqrt{m}}{J^\delta}\right)^l + \left(\frac{\sqrt{m}}{J}\right)^l + \left(\frac{1}{m}\right)^l\right].$$

This Proposition demonstrates three approximation-error terms between Equation (3) and the Gaussian model. The first term is at the order of magnitude $\sqrt{m}/J^\delta$, which results from approximating the heteroskedastic errors $\sigma(X_i) \cdot U_i$ in each bin by the i.i.d. $\sigma\left(\mathcal{X}_{[j]}\right) \cdot U_i$. The second is at the order of magnitude $\sqrt{m}/J$, resulting from using the constant $\alpha + \mathcal{X}^T_{[j]}\beta$ in each bin to approximate the variation of the functional

values within the bin. The smaller the bin size m, the smaller these two error terms are. The last error term is at the order of magnitude $1/m$, and is due to the approximation by the Gaussian random variable. We require the bin size m to be large so that the last error term can be negligible.

Based on Proposition 2.3, the approximating Gaussian model is

$$\mathcal{Y}_{j,\lambda} \approx \alpha + \mathcal{X}_{[j]}^T \beta + \frac{\sigma_j \sqrt{\lambda(1-\lambda)}}{\sqrt{m} f_U(0)} Z_j. \tag{4}$$

Many Gaussian-based estimators are applicable, for example, the maximum-likelihood estimator and OLS estimator. For illustration purposes, we define the OLS estimator based on the local order statistics

$$\left(\widehat{\alpha}, \widehat{\beta}\right)^T \equiv \arg\min_{(\alpha,\beta)} \sum_{j=1}^{J} \left[\mathcal{Y}_{j,\lambda} - \alpha - \mathcal{X}_{[j]}^T \beta\right]^2. \tag{5}$$

Theorem 2.4 *When Assumptions 1-4 hold, then*

(a) Conditional on the design $\{X_i\}_{i=1}^N$ *and* $m \to \infty$, $J \to \infty$, *we have*

$$\left(\widehat{\alpha}, \widehat{\beta}\right)^T \xrightarrow{p} (\alpha, \beta)^T;$$

(b) Conditional on the design $\{X_i\}_{i=1}^N$ *and* $C_1 N^{1/3+\varepsilon_1} \leq m \leq C_2 N^{1/2-\varepsilon_2}$, *where* C_1, C_2 *are generic constants, and* $\varepsilon_1, \varepsilon_2$ *are small positive constants, we have*

$$\sqrt{N}\left[\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right] \xrightarrow{d} N\left[\mathbf{0}, \frac{\lambda(1-\lambda)}{f_U^2(0)} \mathcal{W}\Omega^{-1}\mathcal{W}\right],$$

where

$$\begin{aligned} \mathcal{W} &\equiv \frac{\sum_{j=1}^{J}\left[1, \mathcal{X}_{[j]}^T\right]^T \left[1, \mathcal{X}_{[j]}^T\right]}{J}, \\ \Omega &\equiv \frac{\sum_{j=1}^{J}\sigma_j^2 \cdot \left[1, \mathcal{X}_{[j]}^T\right]^T \left[1, \mathcal{X}_{[j]}^T\right]}{J}. \end{aligned}$$

Although the effective sample size of the proposed estimator is J instead of N, the variance of the Gaussian term $\sigma_j \sqrt{\lambda(1-\lambda)} Z_j / [\sqrt{m} f_U(0)]$ is proportionally reduced to $1/m$. Thus the convergence rate of $\left(\widehat{\alpha}, \widehat{\beta}\right)^T$ will still be $\sqrt{N}$.

Conditional on the design $\{X_i\}_{i=1}^N$, the asymptotic variances of $\sqrt{n}\left(\widehat{\alpha} - \alpha, \widehat{\beta} - \beta\right)^T$ resemble the ones from the check function approach, that is, $\lambda(1-\lambda)\mathcal{V}\Theta^{-1}\mathcal{V}/f_U^2(0)$, where $\mathcal{V} \equiv \sum_{i=1}^N \left[1, X_i^T\right]^T \left[1, X_i^T\right]/N$ and $\Theta \equiv \sum_{i=1}^N \sigma^2(X_i) \cdot \left[1, X_i^T\right]^T \left[1, X_i^T\right]/N$. Moreover, inferences of $\left(\widehat{\alpha}, \widehat{\beta}\right)^T$ can be obtained directly from the OLS software package based on the transformed data, according to Equation (4). Hence both an estimation and inference of the proposed method can be easily obtained from the standard OLS software package, which is highly optimized and can handle a sizable data set quickly.

For the choice of the bin size m in Theorem 2.4, the term ξ_j in Proposition 2.3 is determined to have a negligible contribution to the first order of mean squared error (MSE), compared with the Gaussian term Z_j. At the same time, the bin size m is chosen so that $\left(\widehat{\alpha}, \widehat{\beta}\right)^T$ are asymptotically unbiased.

Remark: The above asymptotic arguments are conditional on the design $\{X_i\}_{i=1}^N$, because the proposed estimator is built upon the linear combination of the order statistics of induced order statistics, i.e., $\mathcal{Y}_{j,\lambda}$ is the $\lceil \lambda m \rceil$th order statistics in the jth bin of the induced order statistics in the original sample. Thus, extending to unconditional asymptotic theories is beyond the scope of this paper and is left for future research. See Delaigle and Hall (2012).[2]

3 Monte Carlo Simulations

In this section, we use a simulation study to evaluate the performance of the new estimator. We generate the data from

$$Y_i = X_{1i} + 2X_{2i} + e_i,$$

where the error e_i is $0.1 \cdot (X_{i1} + X_{2i} + X_{1i}X_{2i} + 0.1) \cdot U_i$ with U_i following a Cauchy distribution.

For the sample design, $\{X_1, X_2\}$ follow $Uniform\,[0,1]$ in Model A, while in Model B $\{X_1, X_2\}$ follow a standard bivariate normal distribution with a correlation coefficient of 0.01. The sample sizes are 1000 and 2500, with 5000 Monte Carlo replications.

The bias and MSE properties of the estimator are shown in Table 1 for the various bin sizes. The table shows that the new estimator $\widehat{\beta}$ performs well and has a small bias and low MSE, even when the bin size is 5. Notice that when m is equal to 1, the estimator simply runs the OLS on the original data so that the estimate is inconsistent and has a large bias. Next, when considering the different sample designs between Models A and B, we observe that compactly supported variables (Model A) have a smaller bias and lower MSE than the non-compactly supported ones (Model B). This result is because of the fact that we are using $\mathcal{X}_{[j]}$ to approximate the variation of the functional values within the bin, and we expect that an equal-spacing transformation, as in Hall et al. (1998), applied before the binning will have better adaptivity for different designs.

Moreover, we show the computational advantage of the new method on the sizable data set. The speed of the new estimator is compared with the check function estimator in Equation (2), where we use the **rq** command in the R **quantreg** package 5.05. The Monte Carlo experiment is based on Model A and we choose the bin size m as 0.1 percent of the sample size N. Note that the choice of m makes the results comparable with the check function estimator, in which the absolute Bias and MSE differences between the two methods are between 0.05 and 0.01. Figure 2 shows the incredible gains in the execution time of the new estimator relative to the check function approach. These gains seem to increase with the $\log_{10}$ sample size. All codes are available upon request.

[2]The only relevant paper is He and Nagaraja (2009), but it is for extreme order statistics of induced order statistics.

4 Conclusion

The key ingredient of this approach is to compute local order statistics from the original data. This saves a large amount of the computational cost, which makes it particularly appealing for analyzing a sizable data set: a Monte Carlo simulation indicates that the proposed estimator is quicker when the sample size N is larger than $20,000$.

References

[1] Brown, L. D., Cai, T. T., Zhou, H. H., 2008. Robust Nonparametric Estimation via Wavelet Median Regression. *Annals of Statistics*. 36, 2055–2084.

[2] Chen, H., 2013. Using Least Squares Loss Functions for Quantile Estimation. Technical Appendix.

[3] Cramer, E., Kamps, U., Rychlik, T., 2002. On the Existence of Moments of Generalized Order Statistics. *Statistics & Probability Letters*. 59, 397–404.

[4] Delaigle, A., Hall, P., 2012. Nonparametric Regression with Homogeneous Group Testing Data. *Annals of Statistics.* 40, 131–158.

[5] Fan, Y., Liu, R., 2012. A Direct Approach to Inference in Nonparametric and Semiparametric Quantile Regression Models. Working Paper.

[6] Gasser, T., Muller, H. G., 1979. Kernel Estimation of Regression Functions. Smoothing Techniques for Curve Estimation: Lecture Notes in Mathematics 757, 23–68.

[7] Hall, P., Park, B.U., Turlach, B.A., 1998. A Note on Design Transformation and Binning in Nonparametric Curve Estimation. *Biometrika*. 85, 469–476.

[8] He, W., Nagaraja, H. N., 2009. Distribution of Extremal Order Statistics from Large Subsets of Concomitants. *Statistics and Probability Letters*. 80, 534–539.

[9] Koenker, R., 2005. *Quantile Regression.* Cambridge University Press.

[10] Koenker, R., Xiao, Z., 2002. Inference on the Quantile Regression Process. *Econometrica*. 70, 1583–1612.

[11] Koenker, R., Bassett, G., 1982. Robust Tests for Heteroskedasticity Based on Regression Quantiles. *Econometrica*. 50, 43–61.

[12] Mason, D., Zhou, H. H., 2012. Quantile Coupling Inequalities and Their Applications. *Probability Surveys*. 9, 439–479.

[13] Marohn, F., 2005. Exponential Inequalities for Tail Probabilities of Generalized Order Statistics. *Metrika*. 61, 251–260.

[14] Portnoy, S., Koenker, R., 1997. The Gaussian Hare and the Laplacian Tortoise. *Statistical Science*. 12, 279–300.

[15] Wald, A., 1943. An Extension of Wilk's Method for Setting Tolerance Limits. *Annals of Mathematics Statistics*. 14, 45–55.

Appendix

Proof for Lemma 2.1: Decompose ζ_j as $\zeta_j \equiv \zeta_j^A + \zeta_j^B$, where

$$\zeta_j^A = \sqrt{m}\mathcal{E}_{j,\lambda} - \sqrt{m}\sigma_j \cdot \mathcal{U}_{j,\lambda} \tag{6}$$

and

$$\zeta_j^B \equiv \sqrt{m}\sigma_j \cdot \mathcal{U}_{j,\lambda} - \frac{\sigma_j\sqrt{\lambda(1-\lambda)}}{f_U(0)} Z_j \tag{7}$$

and $\mathcal{U}_{j,\lambda}$ is the $\lceil \lambda m \rceil$th smallest value of U in the jth bin. Notice that U in the jth bin is still i.i.d..

According to Assumptions 2 and 4, we have $\mathcal{E}_{j,\lambda} \leq \left[CJ^{-d_\sigma} + \sigma_j\right] \cdot \mathcal{U}_{j,\lambda}$. Then, according to Assumption 3, we have $E\left|\zeta_j^A\right|^l = O\left[\left(\frac{\sqrt{m}}{J^{d_\sigma}}\right)^l\right]$.

In order to bound ζ_j^B, we need the following lemma, which states the quantile coupling results for the $\lceil \lambda m \rceil$th smallest value of U in the jth bin.

Lemma 4.1 *(a) Under Assumption 1, we have*

$$\left|\sqrt{\frac{m}{\lambda(1-\lambda)}} f_U(0) \cdot \mathcal{U}_{j,\lambda} - Z\right| \leq \frac{C}{m}\left(1 + |Z|^3\right)$$

for $|Z| \leq \kappa\sqrt{m}$, where constants $C, \kappa > 0$ do not depend on m.

(b) When Assumptions 1 and 3 hold and constants $C', C'' > 0$ do not depend on m, we have

$$\Pr\left[\sqrt{m}\left|\sqrt{\frac{m}{\lambda(1-\lambda)}} f_U(0) \cdot \mathcal{U}_{j,\lambda} - Z\right| > z\right] \leq C' \exp\left(-C''z\right) \text{ for all } z \geq 0.$$

For the complete proof, please refer to the technical appendix in Chen (2013). The outline of the main procedure is: (1) exponentialize the density of the $\lceil \lambda m \rceil$th smallest value of U in the jth bin; (2) exponentialize its distribution F_{U_λ}; (3) obtain the inequality $\Phi[\upsilon - u(\upsilon)] \leq F_{U_\lambda}(\upsilon-) \leq F_{U_\lambda}(\upsilon) \leq \Phi[\upsilon + u(\upsilon)]$ for some υ and $u(\upsilon) > 0$; (4) use Corollary 1 in Mason and Zhou (2012) and Section 3.2 in Marohn (2005) to prove the distributional bound.

The independence of Z_j is because it is constructed by $\Phi^{-1}[F_{U_\lambda}(\mathcal{U}_{j,\lambda})]$, where $\mathcal{U}_{j,\lambda}$ is independent for $1 \leq j \leq J$.

Proof for Lemma 2.2: Denote the set of indices in the jth bin as I_j, then $\mathcal{Y}_{j,\lambda} \leq \max_{k\in I_j}\left[\alpha + X_k^T\beta\right] + \mathcal{E}_{j,\lambda}$ and $\mathcal{Y}_{j,\lambda} \geq \min_{k\in I_j}\left[\alpha + X_k^T\beta\right] + \mathcal{E}_{j,\lambda}$, then according to Assumption 4, $\theta_j(\alpha,\beta) - \alpha - \mathcal{X}_{[j]}^T\beta = O_p\left(\frac{1}{J}\right)$.

Proof for Proposition 2.3: Decompose ξ_j as

$$\begin{aligned}\xi_j &\equiv \sqrt{m}\mathcal{Y}_{j,\lambda} - \sqrt{m}\left[\alpha + \mathcal{X}_{[j]}^T\beta\right] - \frac{\sigma_j\sqrt{\lambda(1-\lambda)}}{f_U(0)} Z_j \\ &= \zeta_j^A + \zeta_j^B + \zeta_j^C\end{aligned}$$

where $\zeta_j^C = \sqrt{m}\mathcal{Y}_{j,\lambda} - \sqrt{m}\left[\alpha + \mathcal{X}_{[j]}^T\beta\right] - \sqrt{m}\mathcal{E}_{j,\lambda} = \sqrt{m}\theta_j(\alpha,\beta) - \sqrt{m}\left[\alpha + \mathcal{X}_{[j]}^T\beta\right]$, and ζ_j^A and ζ_j^B are the same as in Equations (6) and (7). Following Lemma 2.1 and Lemma 2.2, the proposition is proved.

Proof for Theorem 2.4:

(a) When $m \to \infty$ and $J \to \infty$,

$$\begin{aligned}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix} &= \left[\sum_{j=1}^J \left[1, \mathcal{X}_{[j]}^T\right]^T \left[1, \mathcal{X}_{[j]}^T\right]\right]^{-1} \left[\sum_{j=1}^J \left[1, \mathcal{X}_{[j]}^T\right]^T \left(\frac{\sigma_j\sqrt{\lambda(1-\lambda)}}{\sqrt{m}f_U(0)} Z_j + \frac{\xi_j}{\sqrt{m}}\right)\right] \\ &= O_p\left[\frac{1}{J^\delta} + \frac{1}{J} + \frac{1}{m^{3/2}}\right] \xrightarrow{p} \mathbf{0}.\end{aligned}$$

(b) Conditional on $\{X_i\}_{i=1}^N$,

$$Bias\left[\sqrt{n}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}\right] = O\left[\frac{\sqrt{n}}{m^{3/2}} + \frac{\sqrt{n}}{J}\right],$$

and

$$Var\left[\sqrt{n}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}\right] = Var\left[\frac{\lambda(1-\lambda)}{f_U^2(0)}\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[1, \mathcal{X}_{[j]}^T\right]^T Z_j\right] + O\left[\frac{m}{J^2} + \frac{1}{m^2}\right].$$

When $C_1 N^{1/3+\varepsilon_1} \leq m \leq C_2 N^{1/2-\varepsilon_2}$, where C_1 and C_2 are generic constants, and ε_1 and ε_2 are small positive constants, then

$$\begin{aligned} Bias\left[\sqrt{n}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}\right] &= o(1) \\ Var\left[\sqrt{n}\begin{pmatrix}\widehat{\alpha}\\ \widehat{\beta}\end{pmatrix}\right] &= Var\left[\frac{\lambda(1-\lambda)}{f_U^2(0)}\frac{1}{\sqrt{J}}\sum_{j=1}^{J}\left[1, \mathcal{X}_{[j]}^T\right]^T Z_j\right] + o(1). \end{aligned}$$

The asymptotic normality then follows.

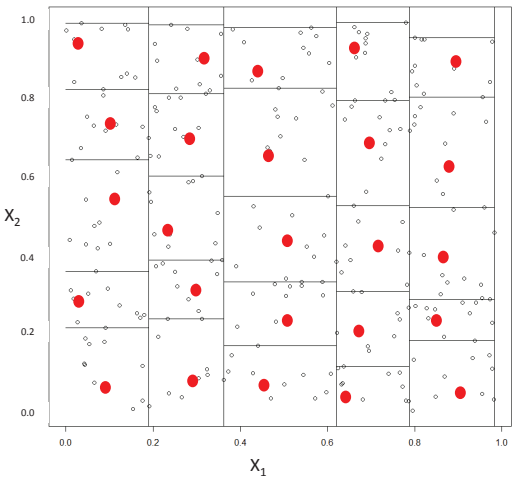
Table 1: Monte Carlo Simulation Results

Coefficient of X_1				Model A			
	m: bin size	1	5	10	20	25	50
$n = 1000$	Bias	−4.2007	−0.0048	−0.0084	−0.0185	−0.02345	−0.0519
	MSE	41323	0.0020	0.0016	0.0035	0.0042	0.0147
$n = 2500$	Bias	0.2553	−0.0004	−0.0025	−0.0067	−0.0076	−0.0179
	MSE	192.67	0.0008	0.0005	0.0007	0.0009	0.0027
				Model B			
	m: bin size	1	5	10	20	25	50
$n = 1000$	Bias	−0.1250	−0.0328	−0.0518	−0.0774	−0.0940	−0.1328
	MSE	1.7402	0.0026	0.0059	0.0148	0.0191	0.0475
$n = 2500$	Bias	−0.3235	−0.0182	−0.0304	−0.0507	−0.0574	−0.0808
	MSE	2.7170	0.0010	0.0017	0.0050	0.0063	0.0150

Coefficient of X_2				Model A			
	m: bin size	1	5	10	20	25	50
$n = 1000$	Bias	3.6691	−0.0113	−0.0212	−0.0377	−0.0371	−0.0613
	MSE	43447	0.0022	0.0021	0.0046	0.0051	0.0152
$n = 2500$	Bias	−0.5909	−0.0027	−0.0077	−0.0154	−0.0186	−0.0370
	MSE	382.77	0.0015	0.0005	0.0009	0.0012	0.0037
				Model B			
	m: bin size	1	5	10	20	25	50
$n = 1000$	Bias	0.4858	−0.0838	−0.1289	−0.1985	−0.1924	−0.2584
	MSE	3.2424	0.0100	0.0235	0.0564	0.0565	0.1176
$n = 2500$	Bias	0.6351	−0.0431	−0.0780	−0.1257	−0.1401	−0.1993
	MSE	2.3059	0.0029	0.0079	0.0209	0.0264	0.0568

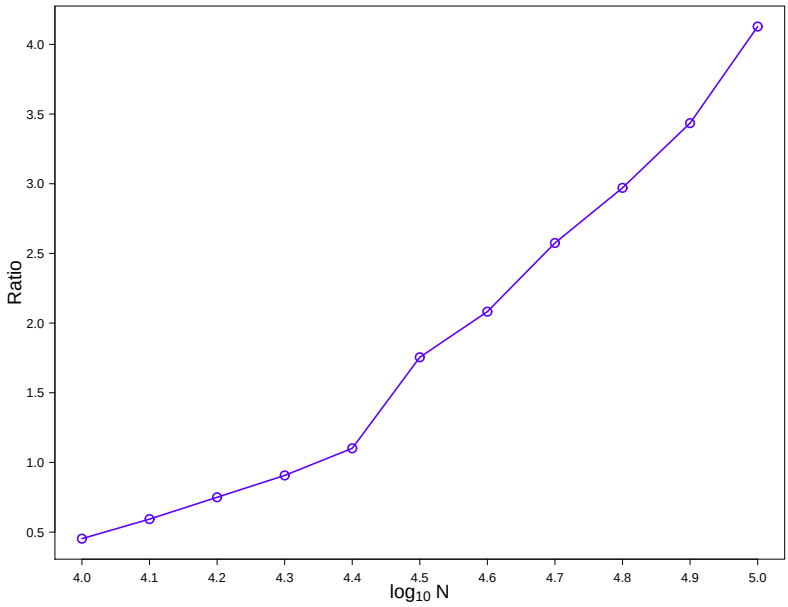
Notes: The Bias and MSE properties of the proposed estimators are listed. The model is $Y_i = X_{1i} + 2X_{2i} + e_i$, where e_i is $0.1 \cdot (X_{i1} + X_{2i} + X_{1i}X_{2i} + 0.1) \cdot U_i$ with U_i following a Cauchy distribution. Model A specifies $\{X_{1i}, X_{2i}\} \sim Uniform\,[0, 1]$, while in Model B $\{X_1, X_2\}$ follow a standard bivariate normal distribution with a correlation coefficient of 0.01. Here for $k = 1, 2$, Bias is defined as $\sum_{s=1}^{5000} \left(\widehat{\beta}_{s,k} - \beta_k\right) /5000$, where $\widehat{\beta}_{s,k}$ is the kth covariate estimated coefficient in the sth replication, and MSE is defined as $\left[\sum_{s=1}^{5000} \left(\widehat{\beta}_{s,k} - \beta_k\right) /5000\right]^2 + \sum_{s=1}^{5000} \left(\widehat{\beta}_{s,k} - \widehat{\beta}_{\cdot,k}\right)^2 /5000$, where $\widehat{\beta}_{\cdot,k}$ is the sample mean of $\widehat{\beta}_{s,k}$ for $s = 1, \cdots, 5000$. The m is the bin size.

Figure 1: Local order statistics of bivariate design



Note: The original $\{X_{1i}, X_{2i}, Y_i\}_{i=1}^{250}$ are first divided into 5 slices along the X_1 axis, ensuring an equal 50 observations in each slice. Each slice is then divided into another 5 slices along the X_2 axis, in order to have 10 observations in each bin. Next the quartile of Y is computed for each bin, and its corresponding coordinates of $\{X_1, X_2\}$ are marked as red solid points.

Figure 2: Execution time ratio



Note: The blue line represents the mean of the execution time ratio across 100 replications versus the $\log_{10}$ of the sample size N, where the ratio is defined as the execution time between the check function approach (Equation 2) and the new estimator (Equation 5). Note that the execution time of the new estimator includes both time from the local order statistics transformation and the ordinary least squares regression. We use Intel i7-2640M CPU@2.80GHz, running Windows 7-64bit using R x64 3.0.2.