ISSN 1066-5307, Mathematical Methods of Statistics, 2015, Vol. 24, No. 2, pp. 147–155. © Allerton Press, Inc., 2015.

The Cumulative Quantile Regression Function with Censored and Truncated Response

S. M. Tse[1]*

[1] *Dept. Appl. Math., National Dong Hwa University, Hualien, Taiwan*

Received May 23, 2014; in final form, March 3, 2015

Abstract—In this paper, we propose a nonparametric estimator of the cumulative quantile regression (CQR) function when the response is subjected to random truncation and censorship. The observed responses are weighted by the increments of the product-limit estimator for the underlying response distribution. Strong Gaussian approximations of the associated weighted partial sum process and the CQR process are established under appropriate assumptions. A functional law of the iterated logarithm for the CQR process is also derived. The construct provides a foundation for the asymptotic theory of functional statistics based on these processes.

Keywords: quantile regression function, strong Gaussian approximation, induced order statistics, product-limit estimator, weighted partial sum process, law of the iterated logarithm.

2000 Mathematics Subject Classification: 60F15, 60F17, 62G05, 62G30.

DOI: 10.3103/S1066530715020052

1. INTRODUCTION

Let (X, Y) be a bivariate positive random vector with marginal distribution functions (df's) of X and Y denoted by F and K respectively. Assume that Y is integrable. The regression function of Y on X is $m(x) = E[\,Y \,|\, X = x\,]$. Let $Q(u) := \inf\{\, x\colon F(x) > u\,\}$, $0 < u < 1$, denote the right-continuous quantile function (qf) associated with $F(x)$. With $U = F(X)$ denoting the rank variable, Rao & Zhao [13] defined the quantile regression (QR) function of Y on X as

$$r(u) = \mathbf{E}\,[\,Y \,|\, U = u\,] = m \circ Q(u), \qquad 0 \le u \le 1. \tag{1.1}$$

The cumulative quantile regression function (CQR) and its standardized form are

$$M(u) := \int_0^u m \circ Q(t)\, dt = \int_0^u r(t)\, dt, \qquad \tilde{M}(u) := \frac{1}{\mu}\, M(u), \tag{1.2}$$

for $u \in [0, 1]$, where $\mu = M(1) = \mathbf{E}\,[\,Y\,]$. We shall refer to this as the full model. In econometrics, with (X, Y) representing income and tax respectively, $\tilde{M}(u)$ is the fraction of the total tax contributed by the lowest uth fraction of income holders. In insurance, $M(1) - M(u)$ and its standardized version are more relevant since large claims are of more interest than small claims (Furman & Zitikis [5, 6]). More generally, $\tilde{M}(u)$ is the counterpart of the Lorenz curve in the presence of a covariate. In the special case $Y = X$, m reduces to the identity function and $\tilde{M}$ is the usual Lorenz curve (Gastwirth [7, 8]). By expressing the X values in terms of quantiles, the numbers are put into their proper context. This is particularly useful when comparing population groups with significant difference in the ranges of X values (see Schechtman *et al.* [15] for a review).

The empirical estimates of M and $\tilde{M}$ based on the partial sums of the Y's and the empirical df $F_n(x)$ of the X's have been studied by several authors. Rao & Zhao [13, 14] established the uniform strong consistency of M_n and $\tilde{M}_n$, the weak convergence of the associated CQR process (respectively the standardized form) in $D[0, 1]$ under Skorohod J_1 topology and a law of the iterated logarithm (LIL) for

*E-mail: tse@server.am.ndhu.edu.tw

the empirical CQR. Davydov and Egorov [4] arrived at the same results in a more general setting via entropy technique. Tse [19] elevated the weak convergence results to strong Gaussian approximation at good rates, thus greatly enhancing the efficiency in application.

However, data are often imperfect. Tse [20] studied the model, where the covariate X is subjected to random truncation and censorship (model I). Another common occurrence is when the response Y, instead of the covariate X, is subjected to random truncation and censorship. How should M and $\tilde{M}$ be estimated in this case? And what are the properties of these estimators? It is the objective of this paper to give a definite answer to these questions. In particular, we propose the empirical estimators of M and $\tilde{M}$, denoted by M_n and $\tilde{M}_n$, based on the empirical process of the covariate X and a weighted partial sum process of the observed responses (subjected to truncation and censorhsip), the weights being the increments of the product-limit (PL) estimator of the df of Y. It is the natural sequel to Tse [19, 20], where the full model and model I were investigated. As before, we continue to work in the unified framework of Csörgő [2] and Csörgő *et al.* [3]. We shall prove the strong uniform consistency of M_n and $\tilde{M}_n$ and construct strong Gaussian approximations of the associated empirical processes. Almost sure statements like the LIL follow as easy consequences. The results broaden the scope of applications in econometics, insurance, medical and health sciences among others. Along the way, we also establish the strong Gaussian approximation of the weighted partial sum process mentioned above. The result is of independent interest and finds applications in other areas. Ultimately, the basis of our constructions are the strong approximation results of Komlós *et al.* [11, 12] for the partial sum and empirical processes of i.i.d. random variables. The other essential ingredient is the LIL of the PL-estimator from Tse [17].

In Section 2, we formulate the model and define the appropriate estimators. Section 3 states the main results of this paper. Auxiliary results and proofs are relegated to Section 4.

2. THE MODEL

Let (T, S) be independent positive random variables with continuous df's G and L respectively. Assume that (T, S) are independent of (X, Y) defined in the Introduction. Let $Z := \min(Y, S)$ and $\delta := I\{Y \le S\}$. If $Z \ge T$, one observes (X, Z, T, δ). If $Z < T$, only X is observed, the quadruplet is incomplete. In other words, the response Y is subjected to left truncation, T, and right censorship, S, mechanisms, while δ indicates whether or not the observed Z is censored. We shall call this model II, to distinguish it from the full model and model I, where the truncation and censorship apply to the covariate X rather than the response Y. In the absence of the covariate, (Z, Y, δ) reduces to the usual left truncation and right censorship (LTRC) model, and the df of Y is estimated by the product-limit estimator. In the presence of the covariate, the existence of a truncated response can now be detected. This marks a deviation from the plain LTRC model. Denote the df of Z by J. By the independence assumption, we have $1 - J = (1 - K)(1 - L)$. For any df H, let $a_H := \inf\{z\colon H(z) > 0\}$ and $b_H := \sup\{z\colon H(z) < 1\}$ denote the left and right endpoints of its support. We assume $a_G = a_J = 0$ and $b_G \le b_J$ throughout.

Let $(\tilde{X}_k, \tilde{Y}_k, \tilde{T}_k, \tilde{S}_k)$, $k = 1, \ldots, n_0$, be independent and identically distributed as (X, Y, T, S). Define $\tilde{Z}_k := \min(\tilde{Y}_k, \tilde{S}_k)$ and $\tilde{\delta}_k := I\{\tilde{Y}_k \le \tilde{S}_k\}$, $k = 1, \ldots, n_0$. Following the rule above, the observed data are $(\tilde{X}_{k_i}, \tilde{Z}_{k_i}, \tilde{T}_{k_i}, \tilde{\delta}_{k_i})$, $i = 1, \ldots, n$, where n is the number of observed complete quadruplets among the n_0. To avoid the complication of double layers of subscripts, we denote these observed quadruplets by $(X_i, Z_i, T_i, \delta_i)$, $i = 1, \ldots, n$. By the truncation mechanism, n is a $\mathrm{Bin}(n_0, \alpha)$ random variable with $\alpha := P(T \le Z)$. The ratio n/n_0 provides a natural estimator for α. Other than that, the incomplete quadruplets have little role to play in our model. The product-limit (PL) estimator of K, the distribution of Y, proposed by Tsai *et al.* [16] is

$$1 - K_n(y) = \prod_{i\colon Z_i \le y} \left[1 - \frac{1}{n\, C_n(Z_i)} \right]^{\delta_i} \tag{2.1}$$

assuming no ties in the data, where $n\, C_n(z) = \sum_{i=1}^{n} I\{T_i \le z \le Z_i\}$ counts the number of items among the n that are actually at risk at the point z. Then $C_n(z)$ is a consistent estimator of

$$C(z) := \frac{1}{\alpha} P(\, T \le z \le S\,)\,[\,1 - K(z-)\,] = \frac{1}{\alpha} G(z)\,[\,1 - L(z)\,]\,[\,1 - K(z)\,], \tag{2.2}$$

since L and K are assumed to be continuous.

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of $X_1, \ldots, X_n$ and denote the Z associated with $X_{(i)}$ by $Z_{(i)}$. The $Z_{(i)}$'s are called induced order statistics or concomitants (see Greselin *et al.* [10] for a review). Let $\Delta K_n(y) = K_n(y) - K_n(y-)$ be the jump of the PL estimator K_n at y and $Q_n(u) := \inf\{\, x\colon F_n(x) > u \,\}$ be the empirical qf for the covariate X. The empirical CQR for our model II is defined for $u \in [0, 1]$ by

$$M_n(u) = \sum_{i=1}^{n} Z_i \, \Delta K_n(Z_i) \, I\{\, X_i \le Q_n(u) \,\} = \sum_{i=1}^{s(u)} Z_{(i)} \, \delta_{(i)} \, \frac{1 - K_n(Z_{(i)}-)}{n \, C_n(Z_{(i)})}, \tag{2.3}$$

where $s(u) = \max\{\, i\colon X_i \le Q_n(u)\}$, applying (2.1) in the second equality. We shall work with the first representation in the following analysis. The second representation is more convenient for data analysis. As u goes from zero to one, $n \, M_n(u)$ gives the weighted partial sums of the induced ordered statistics $Z_{(i)}$'s. In the absence of truncation and censorship, $\delta_i = 1$, $Z_i = Y_i$, and then K_n is replaced by the usual empirical df, so that $\Delta K_n(Z_i)$ becomes $\frac{1}{n}$ for all $i = 1, \ldots, n$, $M_n(u)$ reduces to the empirical CQR for the full model. However, in model II, the presence of K_n and C_n couples the summands in $M_n(u)$. We are no longer dealing with the sum of i.i.d. random variables. Similarly, the standardized counterpart of $M_n(u)$ is

$$\tilde{M}_n(u) = \frac{1}{M_n(1)} \sum_{i=1}^{n} Z_i \, \Delta K_n(Z_i) \, I\{\, X_i \le Q_n(u) \,\}, \qquad 0 \le u \le 1, \tag{2.4}$$

where $M_n(1)$ is the empirical counterpart of $\mu = E\,Y$. The associated CQR process and the standardized version are defined respectively as

$$\zeta_n(u) := \sqrt{n} \, [\, M_n(u) - M(u) \,] \qquad \text{and} \qquad \tilde{\zeta}_n(u) := \sqrt{n} \, [\, \tilde{M}_n(u) - \tilde{M}(u) \,]. \tag{2.5}$$

The analysis thus involves the weighted partial sum process for the Z_i's, the PL-estimator of K and the empirical process for the X_i's simultaneously.

3. BASIC AUXILIARY RESULTS AND MAIN THEOREMS

This section presents the main theorems of this paper. Theorem 3.1 gives the strong uniform consistency of M_n and $\tilde{M}_n$ in the form of a LIL. Theorems 3.2 and 3.3 are the strong approximation theorems for $\zeta_n(u)$ and $\tilde{\zeta}_n(u)$. Theorem 3.4 derives a functional LIL for the CQR process. Theorems 3.2 and 3.4 contain the strong Gaussian approximation of a weighted partial sum process and the associated LIL. The constructions in Theorems 3.2 and 3.3 are based upon the strong approximation results of Komlós *et al.* [11, 12] for the partial sum and empirical processes for i.i.d. random variables in terms of Wiener and Kiefer processes. They are stated as Theorems A and B in Tse [19]. We shall refer to them by the same names throughout this paper. We shall build our strong approximation constructions of ζ_n and $\tilde{\zeta}_n$ on the probability spaces of Theorems A and B.

Here, we simply recall that if $\{\, W(t),\ 0 \le t < \infty \,\}$ is a standard Wiener process, then for $n = 1, 2, \ldots$ and $0 \le t \le 1$,

$$W_n(t) = \frac{1}{\sqrt{n}} W(nt), \tag{3.1}$$

is a sequence of standard Wiener processes with well-defined covariance structure. Also, the two-parameter Kiefer process $\{\, K(u, y)\colon 0 \le u \le 1, 1 \le y \le 1 \,\}$ (see, e.g., Csörgő & Révész [1] for its definition) has covariance

$$\mathbf{E} \, K(u_1, y_1) \, K(u_2, y_2) = (u_1 \wedge u_2 - u_1 \, u_2) \, (y_1 \wedge y_2),$$

and for any positive number n,

$$B_n(u) = \frac{K(u, n)}{\sqrt{n}} \qquad \text{for} \quad 0 \le u \le 1, \tag{3.2}$$

is a Brownian Bridge.

Returning to our model II, for $j = 1, 2, \dots, n$, let

$$D_j \;:=\; Z_j\,\delta_j\,\frac{1-K(Z_j)}{C(Z_j)}. \tag{3.3}$$

Since $Z_j = Y_j$ when $\delta_j = 1$, D_j have the same distribution as

$$D(Y_j) \;:=\; Y_j\,\delta_j\,\frac{1-K(Y_j)}{C(Y_j)}. \tag{3.4}$$

Think of the D_j's as the surrogates of the Y_j's when the observed Z_j's are weighted by the increments of K_n. Next, we introduce the following modified conditional variance of $D(Y)$:

$$\tilde V(x) = \mathbf{E}\left[\,(\,D(Y) - m(X))^2 \,|\, X = x\,\right]. \tag{3.5}$$

We shall see that this modified conditional variance marks the price that we have to pay when the response variable is subjected to truncation and censorship. In the absence of these mechanisms, $\delta = 1$ and $C(z) = 1 - K(z)$, so $D(Y)$ reduces to Y and $\tilde V$ to the true conditional variance of Y:

$$V(x) = \mathbf{E}\left[\,(Y - m(X))^2 \,|\, X = x\,\right]. \tag{3.6}$$

We shall assume that the following smoothness conditions hold for the distribution of (X, Y, T, S) in our theorems:

A1. $E\,|D_j|^3 < \infty$, $j = 1, \dots, n$.

A2. F and K are continuous and strictly increasing.

A3. $m \circ Q$ and $\sqrt{\tilde V \circ Q}$ are functions of bounded variation on $[0, 1]$, so that we can write

$$m \circ Q \;=\; h_1 - h_2, \qquad \sqrt{\tilde V \circ Q} \;=\; h_3 - h_4,$$

where h_i, $i = 1, 2, 3, 4$, are nondecreasing differentiable functions on $[0, 1]$. Moreover, we assume that $\tilde V \circ Q(u) > 0$ for all $u \in (0, 1]$.

A4. The functions h_i, $i = 1, 2, 3, 4$, satisfy

$$\sup_{u\in(0,1)} u^{\alpha}\,(1-u)^{\beta}\,h_i'(u) < \infty \qquad \text{for some} \quad 0 < \alpha, \beta < 1,$$

and similarly for the function Q.

A5. With $J_1(z) := P(\,Z \le z,\ \delta = 1 \,|\, T \le Z\,)$, the following condition is satisfied:

$$\int_0^{\infty} \frac{dJ_1(u)}{C^3(u)} < \infty. \tag{3.7}$$

Also, the function

$$l(z) := \int_0^{z} \frac{d\,J_1(u)}{C^2(u)} \tag{3.8}$$

is strictly increasing and uniformly continuous.

Comparing A1–A4 to their counterparts in the full model (Tse [19]) reveals the adjustment needed for model II. In particular, the finite third moment of the Y's is replaced by that of the D's and the conditional variance of Y by that of the $D(Y)$. A5 is the assumption for the Gaussian approximation of the PL process $\gamma_n(z) := \sqrt{n}\,[\,K_n(z) - K(z)\,]$, which gives rise to an LIL for K_n that plays a key role in the following analysis. The other assumptions remain the same.

Theorem 3.1. *Suppose that conditions* A1–A5 *are satisfied. We have, almost surely,*

$$\begin{aligned} \sup_{u\in[0,1]} \big|\,M_n(u) - M(u)\,\big| \;&=\; O\bigg(\sqrt{\frac{\log\log n}{n}}\bigg), \\ \sup_{u\in[0,1]} \big|\,\tilde M_n(u) - \tilde M(u)\,\big| \;&=\; O\bigg(\sqrt{\frac{\log\log n}{n}}\bigg). \end{aligned} \tag{3.9}$$

The following theorems hold in the probability spaces of Theorems A and B.

Theorem 3.2. *Suppose that conditions* A1–A5 *are satisfied. On a sufficiently rich probability space, there exists a sequence of Gaussian processes*

$$\psi_n(u) = \Gamma_{n1}(u) + \int_0^u B_n(t)\, d\, m \circ Q(t) \tag{3.10}$$

on $[0,1]$*, where* Γ_{n1} *is a mean zero Gaussian process defined in terms of* W_n*'s in* (3.1) *with the covariance function*

$$\mathrm{Cov}\,[\,\Gamma_{n1}(u), \Gamma_{n2}(v)\,] = \int_0^{u\wedge v} \tilde{V} \circ Q(t)\, dt := J_1(u,v), \tag{3.11}$$

B_n *is the Brownian Bridge from a Kiefer process as in* (3.2)*, and the two are independent of each other, such that, for any* $\tau < 1/6$*, we have*

$$\sup_{u\in[0,1]} \big|\, \zeta_n(u) - \psi_n(u) \,\big| \;=\; O(\, n^{-\tau}\,) \tag{3.12}$$

almost surely.

We shall see in the next section that $\Gamma_{n1}(u)$ is the Gaussian approximation to a properly centered and weighted partial sum process of censored and truncated data. That result is of independent interest. The corresponding statement for the standardized version $\tilde{\zeta}_n$ is:

Theorem 3.3. *Suppose that conditions* A1–A5 *are satisfied. On a sufficiently rich probability space, there exists a sequence of Gaussian processes*

$$\tilde{\psi}_n(u) \;=\; \frac{1}{\mu}\,\big[\,\Gamma_{n1}(u) - \Gamma_{n1}(1)\,\tilde{M}(u)\,\big] + \frac{1}{\mu}\int_0^1 [\,I(t \le u) - \tilde{M}(u)\,]\, B_n(t)\, d\, m \circ Q(t)$$

on $[0,1]$*, where* V_n *and* B_n *are the same as in Theorem* 3.2*, such that, for any* $\tau < 1/6$*, we have*

$$\sup_{u\in[0,1]} \big|\, \tilde{\zeta}_n(u) - \tilde{\psi}_n(u) \,\big| \;=\; O(\, n^{-\tau}\,) \tag{3.13}$$

almost surely.

In the absence of truncation and censorship, Theorems 3.2 and 3.3 reduce properly to those of the full model in Tse [19], and Remarks 1 and 2 there apply. They also imply that ζ_n and $\tilde{\zeta}_n$ converge weakly to Gaussian processes with the distributions of ψ_n and $\tilde{\psi}_n$ respectively. More importantly, the strong construction renders inference in the strong sense feasible. Thus, we have the following LIL for ζ_n. Let $\mathcal{D}[0,1]$ be the space of left limit, right continuous functions over $[0,1]$. Let $b_n = \sqrt{2\,\log\log n}$. Define the following set of absolutely continuous functions that originates from Strassen:

$$\mathcal{K} = \Big\{\, k \,|\, k\colon [0,1] \to \mathbf{R},\ k(0) = 0,\ \int_0^1 \big(k'(t)\big)^2\, dt \;\le\; 1 \Big\},$$

$$\mathcal{H} = \Big\{\, h \,|\, h\colon [0,1] \to \mathbf{R},\ h(0) = h(1) = 0,\ \int_0^1 \big(h'(t)\big)^2\, dt \;\le\; 1 \Big\}.$$

Theorem 3.4. *Suppose that conditions* A1–A5 *are satisfied. Then the sequence* ζ_n/b_n *is almost surely relatively compact in the function space* $\mathcal{D}[0,1]$ *endowed with the Skorohod topology, and its set of limit points is the following set of absolutely continuous functions:*

$$\mathcal{G} = \Big\{\, g \,|\, g\colon [0,1] \to \mathbf{R},\ g(u) = \;\sigma\, k\big(\,\eta(u)/\sigma^2\,\big) - \int_0^u h(t)\, dr(t),\ k \in \mathcal{K},\ h \in \mathcal{H} \Big\}, \tag{3.14}$$

where

$$\eta(u) = \int_0^u \tilde{V} \circ Q(t)\, dt \qquad \textit{and} \qquad \eta(1) \equiv \sigma^2. \tag{3.15}$$

A similar result holds for the standardized form of the CQR process. These LILs are particularly useful aids in studying the asymptotic properties of functionals of the ζ_n and $\tilde{\zeta}_n$ processes.

4. AUXILIARY RESULTS AND PROOFS

Since F is continuous and strictly increasing, so is its inverse Q. The empirical process pertaining to F can be expressed in terms of the uniform empirical process. For $u \in [0,1]$, let $E_n(u)$ and $U_n(u)$ be the empirical df and the associated qf of i.i.d. unif$[0,1]$ random variables. Let

$$\beta_n(u) := \sqrt{n}\,\big[\,u - F_n \circ Q(u)\,\big] = \sqrt{n}\,\big[\,u - E_n(u)\,\big], \quad u \in [0,1]. \tag{4.1}$$

Also, for the empirical CQR function, we define the following processes over $[0,1]$:

$$\begin{aligned} \alpha_{n1}(u) &= \sqrt{n}\sum_{i=1}^{n}\Big[\,Z_i \Delta K_n(Z_i) - \frac{1}{n} m(X_i)\Big]\, I\big(\,X_i \le Q_n(u)\,\big), \\ \alpha_{n2}(u) &= \sqrt{n}\,\Big[\,\sum_{i=1}^{n}\frac{1}{n} m(X_i)\, I\big(\,X_i \le Q_n(u)\,\big) - M(u)\Big]. \end{aligned} \tag{4.2}$$

In terms of α_{n1} and α_{n2}, the CQR process is

$$\zeta_n(u) = \alpha_{n1}(u) + \alpha_{n2}(u), \quad 0 \le u \le 1. \tag{4.3}$$

By the observation at the beginning of this section, v_{n2} can be represented by an integration with respect to β_n process as in the full model. But α_{n1} is a very different process. Hence, the analysis of α_{n2} remains unchanged. Our consideration concentrates on α_{n1}.

The summands in α_{n1} are neither independent nor identically distributed. There are two sources for their dependence. The first one comes from the empirical qf Q_n of the covariate. This can be treated by observing that

$$I(\,X_i \le Q_n(u)\,) = I\big(\,F(X_i) \le (F \circ Q_n)(u)\,\big) = I\big(\,U_i \le U_n(u)\,\big), \tag{4.4}$$

where U_i's are i.i.d. unif$[0,1]$ random variables, and then approximating $U_n(u)$ by u within error bound. The second source of dependence among the summands in α_{n1} comes from the factor $\Delta\, K_n$, where the presence of K_n and C_n couples all the items. To deal with that dependence, recall that under A5, the strong Gaussian approximation of the PL-process $\gamma_n(z)$ leads to a LIL for K_n. For our purpose, the following simplified version suffices:

$$\limsup \sup_{z \in [0,b]} |\, K_n(z) - K(z)\,| = O\bigg(\sqrt{\frac{\log\log n}{n}}\bigg) \tag{4.5}$$

almost surely, where $b < b_J$. Under the same assumption, we also have

$$\limsup \sup_{z \in [0,b]} \bigg|\,\frac{1}{C_n(z)} - \frac{1}{C(z)}\,\bigg| = O\bigg(\sqrt{\frac{\log\log n}{n}}\bigg) \tag{4.6}$$

almost surely (Zhou & Yip [21], Tse [18]). Thus we can replace K_n and C_n by K and C within error bounds. Applying (4.4), (4.5) and (4.6) in α_{n1}, we get almost surely, within an error term of $O(n^{-\tau})$, $\tau < 1/2$,

$$\begin{aligned} \alpha_{n1}(u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\,Z_i\,\delta_i\,\frac{1-K(Z_i)}{C(Z_i)} - m(X_i)\Big]\, I\big(\,U_i \le U_n(u)\,\big) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big[\,D_i - m(X_i)\,\big]\, I\big(\,U_i \le U_n(u)\,\big). \end{aligned} \tag{4.7}$$

Note that when $u = 1$, all the indicator variables in (4.7) are equal to one, and we get the weighted partial sum process of the plain LTRC model

$$S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\,D_i - m(X_i)\,]. \tag{4.8}$$

Since a proper understanding of this process would facilitate our analysis and also the process is interesting in its own right, we make a detour to consider this process first.

Conditionally on n, the summands in (4.8) are now independent. To check the proper centering of the summands in (4.8), we note that an observed Z_i implies $T_i \le Z_i$. Smoothing then gives

$$E\left[Z_i\,\delta_i\,\frac{1-K(Z_i)}{C(Z_i)} \right] = E\left\{ E\left[Z_i\,\delta_i\,\frac{1-K(Z_i)}{C(Z_i)} \,\Big|\, Y_i,\, T_i \le Z_i \right] \right\}. \tag{4.9}$$

The conditional expectation in the right-hand side is non-vanishing only if $\delta_i = 1$, which occurs with conditional probability

$$P(\,\delta_i = 1\,|\,Y_i\,) \;=\; P(\,Y_i \le S_i\,) \;=\; 1 - L(Y_i-) \;=\; 1 - L(Y_i),$$

by the continuity assumption of L in the last step. But $\delta_i = 1$ implies $Z_i = Y_i$. So, recalling that the df of T is G, (4.9) becomes

$$\begin{aligned} &E\left\{ Y_i\,\frac{[1-L(Y_i)]\,[1-K(Y_i)]}{C(Y_i)}\,I_{\{T_i \le Y_i\}} \right\} \frac{1}{P(T \le Z)} \\ &\quad = \frac{1}{\alpha}\int_0^\infty y\,\frac{[1-L(y)]\,[1-K(y)]}{C(y)}\,G(y)\,dK(y) = \int_0^\infty y\,dK(y) \;=\; \mu, \end{aligned} \tag{4.10}$$

applying (2.2) in the last line. We have shown that the summands in S_n are centered properly.

The second moments of the summands can be computed in a similar manner. The result is

$$E\,D_i^2 \;=\; E\left[\left(Z_i\delta_i\frac{1-K(Z_i)}{C(Z_i)} \right)^2 \right] \;=\; \int_0^\infty y^2\,\frac{\alpha}{G(y)\,[1-L(y)]}\,dK(y),$$

which differs from the second moment of Y, but reduces to it properly in the absence of censorship and truncation ($\alpha = 1$, $G(t) = 1$ and $L(t) = 0$ for all $t \ge 0$). Thus, if we let

$$\sigma^2 \;=\; \mathrm{Var}(D_i) \;=\; \int_0^\infty y^2\,\frac{\alpha}{G(y)\,[1-L(y)]}\,dK(y) \;-\; \mu^2, \tag{4.11}$$

then σ is the scale parameter for the weighted summands in the LTRC model. It is also the modified conditional variance $\tilde{V}$ introduced in (3.5).

To summarize, we have reduced the weighted partial sum process for the LTRC model to a sum of i.i.d. random variables with proper mean μ but adjusted variance σ^2 defined in (4.11). We can now invoke Theorem A to obtain the strong Gaussian approximation of this process.

Theorem 4.1. *Consider the data* (Z_i, T_i, δ_i), $i = 1, \ldots, n$, *in the LTRC model. Assume that* $E\,D_i^p < \infty$ *for some* $p > 2$, *where* D_i *defined in* (3.3) *have mean* $\mu = EY$ *and variance* σ^2 *defined in* (4.11). *There exists a Wiener process such that*

$$\sup_{t\in[0,T]} \left| n\sum_{i=1}^{[t]} Z_i\,\Delta K_n(Z_i) \;-\mu t \;-\; \sigma\,W(t) \right| \;=\; O(\,T^{1/p}\,) \tag{4.12}$$

almost surely.

This result generalizes the classical strong approximation result of KMT [11, 12] for regular partial sum process to summands subjected to random truncation and/or censorship mechanisms. In particular, it enables us to transplant many properties of the Wiener process to the weighted partial sum process with the minimum effort.

We now end our detour and return to $\alpha_{n1}(u)$ in our model II. Introduce the standardized form of the weighted observed responses:

$$\xi_i := \frac{1}{\sqrt{\tilde{V}(X_i)}}\left[Z_i\,\delta_i\,\frac{1-K(Z_i)}{C(Z_i)} - m(X_i) \right] = \frac{D_i - m(X_i)}{\sqrt{\tilde{V}(X_i)}} \tag{4.13}$$

and its associated partial sum process $\{ S_n(t)\colon 0 \le t \le 1 \}$ by

$$S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]+1} \xi_i, \qquad n = 1, 2, \ldots. \tag{4.14}$$

To deal with the last dependence among the summands of $\alpha_{n1}(u)$, we replace $U_n(u)$ in the summands of $\alpha_{n1}(u)$ by u, the difference thus incurred is:

$$\begin{aligned} \Delta_n(u) &= \alpha_{n1}(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\, D_i - m(X_i)\,]\, I(\, U_i \le u\,), \\ |\, \Delta_n(u)\,| &= \left| \alpha_{n1}(u) - \int_0^u \sqrt{\tilde{V} \circ Q(t)}\, dS_n(t) \right| \\ &= \left| \int_{u\wedge U_n(u)}^{u\vee U_n(u)} \sqrt{\tilde{V} \circ Q(t)}\, dS_n(t) \right| \\ &\le \left| \int_{u\wedge U_n(u)}^{u\vee U_n(u)} h_3(t)\, dS_n(t) \right| + \left| \int_{u\wedge U_n(u)}^{u\vee U_n(u)} h_4(t)\, dS_n(t) \right| \\ &=: \Delta_{n1}(u) + \Delta_{n2}(u), \end{aligned} \tag{4.15}$$

with the help of A3 in the next to last step. Luckily, it takes the same form as in the full model and the analysis there carries over. From the proof of Lemma 3.1 in Tse [19], we have

$$\sup_{0\le u\le 1} |\, \Delta_{n1}(u)\,| \;+\; \sup_{0\le u\le 1} |\, \Delta_{n2}(u)\,| \;=\; O(\, n^{-\tau}\,)$$

almost surely for any $\tau < 1/6$. Thus, we have succeeded in representing $\alpha_{n1}(u)$ by a scaled partial sum process within error bounds.

Now, we are ready for the strong Gaussian approximation for the CQR process defined in (4.3). For $0 \le u \le 1$, let

$$\Gamma_{n1}(u) = \int_0^u \sqrt{(\tilde{V} \circ Q)(t)}\, dW_n(t), \qquad \Gamma_{n2}(u) = \int_0^u B_n(t)\, d\, m \circ Q(t), \tag{4.16}$$

where $W_n(t)$ are the Wiener processes defined in (3.1) and $B_n(t)$ are the Brownian Bridges defined in terms of the Kiefer process in (3.2).

Proof of Theorem 3.2. Observe that $\psi_n(u) = \Gamma_{n1}(u) + \Gamma_{n2}(u)$. Hence, writing sup for $\sup_{u\in[0,1]}$, we have

$$\sup |\, \zeta_n(u) - \psi_n(u)\,| \;\le\; \sup |\, \Gamma_{n1}(u) - \alpha_{n1}(u)\,| + \sup |\, \Gamma_{n2}(u) - \alpha_{n2}(u)\,| \tag{4.17}$$

almost surely. The second term is exactly the same as in the full model. So Lemma 3.2 in Tse [19] gives $O(n^{-\tau})$ almost surely for any $\tau < 1/6$. The first term is also analogous to that of the full model except that $\tilde{V}$ replaces V. The analysis in Lemma 3.1 there carries over, with the role of Theorem A there taken by our Theorem 4.1 with $p = 3$. Again, we get the same order $O(n^{-\tau})$ almost surely for any $\tau < 1/6$. □

The proof of Theorem 3.3 is similar. Theorem 3.4 follows from Theorem 3.2 and the standard functional LIL for the Wiener and Kiefer processes (see Theorems 1.3.2 and 1.15.1 of [1]). Finally, noting that $M(u)$ is continuous, Theorem 3.1 follows easily from Theorem 3.4.

Proof of Theorem 3.1. Recalling that $h \in \mathcal{H}$ satisfies $|h(t)| \le t^{1/2}$ on $(0,1)$ (Lemma 1.7.1 in [1]), we have

$$\limsup_{n\to\infty} \sqrt{\frac{n}{\log\log n}} \sup_{u\in[0,1]} |\, M_n(u) - M(u)\,| \;\le\; v, \tag{4.18}$$

where

$$v^2 = \sigma^2 + \left| \int_0^1 dr(t) \right|^2 < \infty. \tag{4.19}$$

This gives the first statement of Theorem 3.1. The second statement follows easily. □

5. ACKNOWLEDGMENTS

.

The author would like to thank the editor and the referee for their suggestions and remarks. This work was partially supported by NSC Grant 101-2118-M-259-003, R.O.C.

REFERENCES

1. M. Csörgő and P. Révész, *Strong Approximations in Probability and Statistics*. (Academic Press, New York, 1981).
2. M. Csörgő, *Quantile Processes with Statistical Applications* (SIAM, Philadelphia, 1983).
3. M. Csörgő, S. Csörgő, and L. Horváth, *An Asymptotic Theory for Empirical Reliability and Concentration Processes*, in *Lecture Notes in Statistics* (Springer, New York, 1986), Vol. 33.
4. Yu. Davydov and V. Egorov, "Functional Limit Theorems for Induced Order Statistics", Math. Methods Statist. **9**, 297–313 (2000).
5. E. Furman and R. Zitikis, "Weighted Risk Capital Allocations", Insurance: Mathematics and Economics **43**, 263–269 (2008).
6. E. Furman and R. Zitikis, "A Monotonicity Property of the Composition of Regularized and Inverted-Regularized Gamma Functions with Applications", J. Math. Anal. and Appl. **348**, 971–976 (2008).
7. J. L. Gastwirth, "A General Definition of the Lorenz Curve", Econometrica **39**, 1037–1039 (1971).
8. J. L. Gastwirth, "The Estimation of the Lorenz Curve and Gini Index", Rev. Econometric Statist. **54**, 306–316 (1972).
9. C. M. Goldie, "Convergence Theorems for Empirical Lorenz Curve and Their Inverses", Adv. Appl. Probab. **9**, 765–791 (1977).
10. F. Greselin, M. L. Puri, and R. Zitikis, "*L*-functions, Processes, and Statistics in Measuring Economic Inequality and Acturial Risks", Statist. & Its Interface **2**, 227–245 (2009).
11. J. Komlós, P. Major, and G. Tusnády, "An Approximation of Partial Sums of Independent r.v.'s and the Sample d.f. I", Z. Wahrsch. verw. Gebiete **32**, 111–132 (1975).
12. J. Komlós, P. Major, and G. Tusnády, "An Approximation of Partial Sums of Independent r.v.'s and the Sample d.f. II", Z. Wahrsch. verw. Gebiete **34**, 33–58 (1976).
13. C. R. Rao and L. C. Zhao, "Convergence Theorems for Empirical Cumulative Quantile Regression Functions", Math. Methods Statist. **4**, 81–91 and 359 (1995).
14. C. R. Rao and L. C. Zhao, "Law of the Iterated Logarithm for Empirical Cumulative Quantile Regression Functions", Statistica Sinica **6**, 693–702 (1996).
15. E. Schechtman, A. Shelef, S. Yitzhaki, and R. Zitikis, "Testing Hypotheses about Absolute Concentration Curves and Marginal Conditional Stochastic Dominance", Econometric Theory **24**, 1044–1062 (2008).
16. W. Y. Tsai, N. P. Jewel, and M. C. Wang, "A Note on the Product-Limit Estimator under Right Censoring and Left Truncation", Biometrika **74**, 883–886 (1987).
17. S. M. Tse, "Strong Gaussian Approximations in the Left Truncated and Right Censored Model", Statistica Sinica **13**, 275–282 (2003).
18. S. M. Tse, "Quantile Processes for Left Truncated and Right Censored Data", Ann. Inst. Statist. Math. **57**, 61–69 (2005).
19. S. M. Tse, "On the Empirical Cumulative Quantile Regression Process", Math. Methods Statist. **18** (3), 270–279 (2009). Also available on SpringerLink DOI:10.3103/S1066530709030053.
20. S. M. Tse, "The Cumulative Quantile Regression Function with Censored and Truncated Covariate", Math. Methods Statist. **21** (3), 238–249 (2012). Also available on SpringerLink DOI:10.3103/S1066530712030040.
21. Y. Zhou and S. F. Yip, "A Strong Representation of the Product-Limit Estimator for Left Truncated and Right Censored Data", J. Multivar. Anal. **69**, 261–280 (1999).