Optimal Multivariate Quota-share Reinsurance: A Nonparametric Mean-CVaR Framework

Haoze Sun†, Chengguo Weng‡, Yi Zhang††

†Department of Mathematics
Zhejiang University, Hangzhou, Zhejiang, 310027, China

‡Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada

Abstract

In this paper, the Conditional Value-at-Risk (CVaR) is adopted to measure the total loss of multiple lines of insurance business and two nonparametric estimation methods are introduced to explore the optimal multivariate quota-share reinsurance under a mean-CVaR framework. The proposed optimal reinsurance models are directly formulated on empirical data and no explicit distributional assumption on the underlying risk vector is required. The resulting nonparametric reinsurance models are convex and computationally amenable, circumventing the difficulty of computing CVaR of the sum of a generally dependent random variable. Statistical consistency of the resulting estimators for the best CVaR is established for both nonparametric models under mild conditions, which allow empirical data generated from any stationary process satisfying strong mixing conditions. In the end of the paper, a simulation example is presented to demonstrate the finite sample performance of both nonparametric models.

Keywords: Multiple optimal reinsurance, Quota-share reinsurance, Mean-CVaR, Nonparametric model, Kernel estimation, α-mixing process

Corresponding author. Tel: +086(571)8795-3667.
Email: Sun(sunhaoze@zju.edu.cn), Weng(c2weng@uwaterloo.ca), Zhang(zhangyi63@zju.edu.cn)
1. Introduction

Reinsurance can be an effective way of managing risk by transferring risk from an insurer (also referred to as the cedent) to a second insurance carrier (referred to as the reinsurer). The study of optimal reinsurance design has attracted great attentions from both academicians an practitioners since the seminar work of Borch (1960). In the past half-century, a variety of optimal reinsurance designs have been devised by either minimizing certain risk measure of an insurer’s risk exposure or maximizing the expected utility of the final wealth of a risk-averse insurer; see, for example, Borch (1960), Arrow (1963), Raviv (1979), Humberman et al. (1983), Young (1999); Kaluszka (2001); Kaluszka and Okolewski (2008), and references therein. More recently, the Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) based optimal reinsurance designs have been extensively studied due to the prevalent use of the two risk measures in financial and insurance practice; see, for instance, Gajek and Zagrodny (2004), Huang (2006), Cai et al. (2008), Balbás et al. (2009), Cheung (2010), Tan et al. (2011), Asimit et al. (2013), Chi and Weng (2013), and Cheung et al. (2014a), Tan and Weng (2014), just to name a few.

Almost all the optimal reinsurance models in the afore-mentioned literature are for a single random variable, and the results are therefore applicable only to aggregate loss. In the insurance practice, however, an insurer with multiple lines of business often purchases reinsurance separately for each line of business. Therefore, from the perspective of enterprise risk management, it is prudent for the insurer to investigate the strategies to buy reinsurance on each individual line of business and attain the optimality in certain integrated sense. Due to the inherent dependence among the risks of individual lines, optimal multivariate reinsurance problems are usually difficult to be solved. Under certain special dependence structure, Denuit and Vermandele (1998) and Cai and Wei (2012) show that an excess-of-loss reinsurance is optimal for the expected value reinsurance premium principle. Cheung et al. (2014b) propose a minimax model and its solution is also in favour of excess-of-loss reinsurance. Zhu et al. (2014) study the multiple optimal reinsurance by minimizing the multivariate lower-orthant Value-at-Risk, and shows that a two-layer reinsurance for each line is optimal for various reinsurance premium principles.

Extensive research has been conducted to explore the general optimal reinsurance contract in the literature, and different shapes of optimal reinsurance treaties have been identified corresponding to various optimality criteria, being non-proportional contracts, such as stop-loss, excess-of-loss, layer reinsurance and so on, in a majority of cases.
Nevertheless, quota-share reinsurance, which is one of the main proportional reinsurance contracts, remains among the most popular forms of reinsurance in insurance practice, particularly for life insurance. This motivates us to study the optimal quota-share reinsurance for multiple lines of insurance businesses in the present paper.

Under a quota-share reinsurance treaty, the reinsurer takes a stated percentage share of each policy that an insurer underwrites. The reinsurer receives that stated percentage of the premiums and pays the stated percentage of claims. Therefore, the determination of the optimal multiple quota-share reinsurance boils down to deciding on the optimal percentage share (called quota-share coefficient) for each line of business. In the present paper, the optimal reinsurance treaties are studied under a mean-CVaR framework, where the Conditional Value-at-Risk (CVaR) is adopted to measure the total loss of multiple lines of insurance business and a constraint on the insurer’s expected net profit is imposed.

The proposed mean-CVaR optimal reinsurance model is theoretically appealing as it allows the insurer to balance its risk and reward in exploring optimal reinsurance purchase strategies. Nevertheless, such theoretical model is subject to two major issues. First, its formulation heavily relies on the distribution of the underlying risk vector, which in practice is uncertain and needs to be estimated from empirical data. Second, even though the distribution of the underlying risk vector is explicitly known, the analysis of optimal reinsurance is often intimidatingly challenging because it involves computing the CVaR of the aggregate risk from a generally dependent random vector.

In the present paper, we follow the idea of Weng (2009) (also see, Tan and Weng (2014)) and formulate optimal reinsurance models directly from empirical data, leading to fully nonparametric models. We propose two types of nonparametric models upon the theoretical mean-CVaR framework and a representation result of CVaR from Rockafellar and Uryasev (2002). The first one is directly formulated under the empirical measure and it leads to a linear programming problem, which we refer to as “linear programming model” throughout the paper. Second, we propose a kernel estimation for the risk measure CVaR and the resulting nonparametric model becomes a convex programming. Such a model is termed as “kernel-based model” throughout the paper.

Our proposed empirically data-based models possess many advantages. First, they are completely data-based and no explicit assumption is required on the distribution of the loss random vector. Second, the proposed nonparametric models bypass the technical obstacle of computing the CVaR of the aggregate loss from a generally dependent random vector, which can not be circumvented in a theoretical model. Third, the proposed
nonparametric models are computationally amenable and can be solved efficiently as either a linear programming or a convex programming. Fourth, statistical consistency results are established to provide theoretical support to our proposed nonparametric models. The best risk measure level solved from both nonparametric models is formally shown convergent to the theoretically best level for empirical data generated from any stationary process satisfying certain moment and strong mixing conditions. Our numerical simulation with Exponential and Pareto marginal distributions and Gaussian copula illustrate that both the linear programming model and kernel-based model work well for reasonably large sample size, with the later performed slightly better than the former.

The rest of the paper proceeds as follows. The mathematical formulation of multiple quota-share reinsurance is introduced in section 2, and the theoretical mean-CVaR reinsurance model is defined in section 3. The linear programming model and the kernel-based model are specified in section 4. Convergence results for the proposed nonparametric reinsurance models are shown in section 5. Section 7 is our numerical simulation study, and section 8 concludes the paper. Some technical lemmas are given in Appendix A, and some proofs are collected in Appendix B.

2. Multiple Quota-share Reinsurance

Let \( X = (X_1, X_2, \cdots, X_p) \) be the claim (aggregate) loss vector on \( p \) lines of an insurer’s business, where \( X_i \) denotes the aggregate nonnegative loss random variable (in the absence of reinsurance) the insurer is subject to in its \( i \)th line of business with cumulative distribution function \( F_i(x) = \mathbb{P}(X_i \leq x) \), survival function \( F_i(x) = 1 - F_i(x) = \mathbb{P}(X_i > x) \) and mean \( \mu_i = \mathbb{E}[X_i] < \infty, i = 1, \ldots, p. \)

The insurance premium is often calculated by the expected value principle in insurance practice, because the safety loading on the top of the net premium can be justified by estimating the distributional shape of the underwritten risks, which is usually a priori determined before any reinsurance contract is reached. Therefore, in the present paper we assume that the insurance premium \( \pi_i \) collected by the insurer in underwriting the risk \( X_i \) is computed according to the expected value principle with a loading factor \( \beta_i > 0 \), i.e.,

\[
\pi_i = (1 + \beta_i)\mathbb{E}[X_i] = (1 + \beta_i)\mu_i, \quad i = 1, \ldots, p.
\]  
Consequently, the net risk of the insurer in the \( i \)th line of business in absence of any reinsurance is given by \( X_i - \pi_i, i = 1, \ldots, p. \)
We assume that the cedent considers to purchase a quota-share reinsurance with a proportional coefficient $c_i \in [0, 1]$ for the $i$th line of business, $i = 1, \ldots, p$. Under such multiple quota-share reinsurance, the cedent retains a risk of $(1 - c_i)X_i$ and transfers the risk of $c_iX_i$ to a reinsurer, $i = 1, \ldots, p$. Further, we assume that the premium principle used to compute the reinsurance premium is also the expected principle with a positive loading factor $\theta_i$ for the $i$th line of business, $i = 1, \ldots, p$. Consequently, the reinsurance premium which the insurer is obligated to pay the reinsurer as the compensation of transferring the part of risk $c_iX_i$ to the reinsurer is given by

$$\Pi_i(c_iX_i) = (1 + \theta_i)\mathbb{E}[c_iX_i] = (1 + \theta_i)c_i\mathbb{E}[X_i] = (1 + \theta_i)c_i\mu_i, \quad i = 1, \ldots, p.$$ 

In general, we assume that the loading factor for the insurance premium is smaller than that for the reinsurance premium, i.e., $\beta_i < \theta_i$, $i = 1, \ldots, p$.

The expected value principle is often criticized when it is used in a general optimal reinsurance model, but such premium principle is a legitimate assumption for the design of optimal quota-share reinsurance as in our models, which can be justified as follows. The distributional properties of the risk absorbed by the reinsurer do not vary with the proportional coefficient in quota-share reinsurance except its scale, and therefore, an a priori safety loading factor which the reinsurer may use to charge the reinsurance premium can be expected by the insurer before the optimal proportional coefficient is determined in a reinsurance contract. Therefore, the assumption of expected value principle for reinsurance premium calculation is legitimate for the design of optimal quota-share reinsurance. Indeed, any premium principle $\Pi_i$ is equivalent to an expected value principle in computing the premium of a quota-share reinsurance as long as it processes the property of scale invariance, i.e., $\Pi_i(c_iX_i) = c_i\Pi_i(X_i)$ for any constant $c_i > 0$, because in this case, $\Pi_i(c_iX_i)$ can be written as

$$\Pi_i(c_iX_i) = (1 + \theta_i)\mathbb{E}[c_iX_i]$$

with $\theta_i = \frac{\Pi_i(X_i)}{\mathbb{E}[X_i]} - 1$, which is independent of the proportional coefficient $c_i$ and known in advance before the reinsurance design.

With the above specification of the multiple quota-share reinsurance, the risk exposure in the presence of reinsurance for the $i$th line of business is given by

$$T_i(c_i) = (1 - c_i)X_i + (1 + \theta_i)c_i\mu_i - \pi_i,$$

where $\pi_i$ is given in (2.1), $i = 1, \ldots, p$. Accordingly, the total risk exposure of the insurer
over the $p$ lines of business is given by
\[
T(c) = \sum_{i=1}^{p} [(1 - c_i) X_i + (1 + \theta_i) c_i \mu_i - \pi_i],
\] (2.2)
where $c = (c_1, \ldots, c_p)$ is the proportional coefficient vector. Define
\[
L(c, X) = \sum_{i=1}^{p} (1 - c_i) X_i, \quad \mu = (\mu_i, \ldots, \mu_p)' \quad \text{and} \quad \zeta = (\zeta_1, \ldots, \zeta_p)',
\]
with $\zeta_i = (1 + \theta_i) c_i - (1 + \beta_i), \; i = 1, \ldots, p$. Then, from (2.1) and (2.2), the total risk exposure of the insurer over the $p$ lines of business
\[
T(c) = L(c, X) + \zeta' \mu,
\] (2.3)
and its expected net profit
\[
\Psi(c) = \mathbb{E}[-T(c)] = \sum_{i=1}^{p} (\beta_i - \theta_i c_i) \mu_i.
\] (2.4)

3. Theoretical Mean-CVaR Reinsurance Model

The determination of optimal reinsurance in the present paper is analyzed from the insurer’s perspective, and the problem is tackled under a mean-CVaR framework, where CVaR refers to the risk measure Conditional Value-at-Risk (CVaR). The CVaR at a confidence level $\alpha$ (e.g. $\alpha = 0.95$) of the insurer’s total risk is defined by
\[
\text{CVaR}_\alpha(T(c)) = \frac{1}{1 - \alpha} \int_{\alpha}^{1} \text{VaR}_u(T(c)) du,
\] (3.5)
where $\text{VaR}_u(T(c))$ is the Value-at-Risk of $T(c)$ at a confidence level of $u$ and formally defined by
\[
\text{VaR}_u(T(c)) := \inf\{q \in \mathbb{R} | \mathbb{P}(T(c) \leq q) \geq \alpha\}.
\]
The mathematical model we exploit to analyze optimal multiple quota-share reinsurance is to minimize CVaR of the insurer’s total risk exposure $T(c)$, constrained on a given expected total profit level of $P$, as follows:
\[
\left\{ \begin{array}{ll}
\min_{c' \in [0,1]^p} & \text{CVaR}_\alpha(T(c)), \\
\text{s.t.} & \Psi(c) = P,
\end{array} \right. \] (3.6)
where \( c' \) is the transpose of \( c \).

Using the translation invariance property of CVaR, one may write

\[
\text{CVaR}_\alpha(T(c)) = \text{CVaR}_\alpha(L(c, X)) + \zeta' \mu,
\]

and consequently, applying the representation of CVaR from Rockafellar and Uryasev (2002), one gets

\[
\text{CVaR}_\alpha(T(c)) = \min_{q \in \mathbb{R}} H_\alpha(c, q) + \zeta' \mu, \tag{3.7}
\]

where

\[
H_\alpha(c, q) = q + \frac{1}{1 - \alpha} \mathbb{E}[(L(c, X) - q)^+], \tag{3.8}
\]

and \((x)^+ = \max(x, 0)\).

Let

\[
\mathcal{N}_P = \{ c' \in [0, 1]^p : \Psi(c) = P \} \tag{3.9}
\]

denote the feasible set of the problem (3.6). By the Theorem 14 in Rockafellar and Uryasev (2002), model (3.6) is equivalent to

\[
\min_{(c', q) \in \mathcal{N}_P \times \mathbb{R}} \{ H_\alpha(c, q) + \zeta' \mu \}, \tag{3.10}
\]

in the sense that

\[
\min_{(c', q) \in \mathcal{N}_P \times \mathbb{R}} \{ H_\alpha(c, q) + \zeta' \mu \} = \min_{c' \in \mathcal{N}_P} \{ \text{CVaR}_\alpha(L(c, X)) + \zeta' \mu \}
\]

and \((c^*, q^*)' = \arg \min_{(c', q) \in \mathcal{N}_P \times \mathbb{R}} \{ H_\alpha(c, q) + \zeta' \mu \}\) if and only if

\[
c^* = \min_{c' \in \mathcal{N}_P} \{ \text{CVaR}_\alpha(L(c, X)) + \zeta' \mu \} \quad \text{and} \quad q^* = \arg \min_{q \in \mathbb{R}} H_\alpha(c^*, q).
\]

Since \( L(c, X) \) is nonnegative for any \( c \in [0, 1]^p \), we must have \( q^* \geq 0 \), and therefore, we can restrict to the set of \( \mathcal{N}_P \times [0, \infty) \) for the solution of the problem (3.10), i.e., to consider problem

\[
\min_{(c', q) \in \mathcal{N}_P \times [0, \infty)} Q_\alpha(c, q) \tag{3.11}
\]

where

\[
Q_\alpha(c, q) = H_\alpha(c, q) + \zeta' \mu = q + \frac{1}{1 - \alpha} \mathbb{E}[(L(c, X) - q)^+] + \zeta' \mu.
\]
4. Nonparametric Mean-CVaR Reinsurance Model

The theoretical mean-CVaR optimal reinsurance model specified in the preceding section is theoretically appealing as it allows the insurer to strike balance between the risk retaining and expected profit gaining. Nevertheless, such theoretical model is subject to two major issues as we pointed out in the first section. First, the computation of the objective in the model calls for a full understanding on the joint distribution of the underlying risk vector, which is uncertain and needs to be estimated from empirical data in practice. Second, given that the distribution of the underlying risk vector is fully known, it is still intimidatingly challenging to compute the objective for a generally dependent random vector \( X \).

In this section, we will propose two nonparametric models which are directly formulated from empirical data. To proceed, let

\[
\{X_t = (X_{1,t}, X_{2,t}, \ldots, X_{p,t})', \ t = 1, 2, \ldots, T\}
\]

be a sample of the risk vector \( X = (X_1, X_2, \ldots, X_p)' \), and \( \{L_1, \ldots, L_T\} \) be the loss observations corresponding to a reinsurance proportional coefficient vector \( c \) so that \( L_t = L(c, X_t), \ t = 1, \ldots, T \).

4.1. Linear Programming Model

The first nonparametric model we propose is given as follows

\[
\min_{(c', q) \in \hat{\mathbb{N}}_P \times [0, \infty)} \tilde{Q}_\alpha(c, q), \quad (4.12)
\]

which is constructed by replacing the objective \( Q_\alpha(c, q) \) in (3.11) by its sample version

\[
\tilde{Q}_\alpha(c, q) = q + \frac{1}{(1 - \alpha)T} \sum_{t=1}^T (L(c, X_t) - q)^+ + \zeta^\top X, \quad (4.13)
\]

and substituting the feasible set by its sample counterpart

\[
\hat{\mathbb{N}}_P = \{c \in [0, 1]^p : \hat{\Psi}(c) = P\}, \quad (4.14)
\]

where

\[
X = (X_1, \ldots, X_p) = \sum_{t=1}^T X_t \quad (4.15)
\]
is the sample mean vector, and

\[ \hat{\Psi}(c) = \sum_{i=1}^{p} (\beta_i - \theta_i c_i) \overline{X}_i. \] (4.16)

is an estimator of \( \Psi(c) \). Here, \( \overline{X}_i \) is the sample mean of the losses from the \( i \)th line of business and used as an estimator of \( \mu_i = E[X_i], i = 1, \ldots, p. \)

The problem \( (4.12) \) is referred to as “linear programming model” throughout the paper, because it can be equivalently reformulated, by introducing auxiliary real variables \( \kappa_t \), for \( t = 1, \ldots, T \), as the following linear programming

\[
\begin{aligned}
& \min \left( c' q, \kappa' \right) \\
\text{s.t.} \quad & (c', q) \in \hat{\mathbb{R}}_P \times [0, \infty), \\
& \kappa_t \geq 0 \text{ and } \kappa_t \geq L(c, X_t) - q, \quad t = 1, \ldots, T
\end{aligned}
\] (4.17)

where \( \kappa = (\kappa_1, \ldots, \kappa_T)' \).

We have the following observations regarding \( (4.12) \) and \( (4.17) \). First, if \( (\hat{c}', \hat{q}) \in \hat{\mathbb{R}}_P \times [0, \infty) \) solves \( (4.12) \), \( (\hat{c}', \hat{q}, \hat{\kappa}') \) solves \( (4.17) \) with \( \hat{\kappa} = (\kappa_1, \ldots, \kappa_T) \) and \( \hat{\kappa}_t = (L(\hat{c}, X_t) - \hat{q})^+, \quad t = 1, \ldots, T. \) Second, if an element \( (\hat{c}', \hat{q}, \hat{\kappa}') \in \hat{\mathbb{R}}_P \times [0, \infty) \times \mathbb{R}^T \) solves \( (4.17) \), \( (\hat{c}', \hat{q}) \) must solve \( (4.12) \). Moreover, both optimization problems share the same optimal value.

### 4.2. Kernel-based Model

The second nonparametric model relies on kernel-based estimation for the objective \( Q_\alpha(c, q) \) in the theoretical model \( (3.11) \), and hence we refer such a model as “kernel-based model”. To proceed, let \( p(x; c) \) denote the probability density function of \( L_t(c, X) \), \( P(x; c) \) be its cumulative distribution function, and \( \overline{P}(x; c) = 1 - P(x; c) \) be the survival function of \( L_t(c, X) \). When it is clear from context, we suppress the dependence of these notations on vector \( c \) and simply write them by \( p(x) \), \( P(x) \) and \( \overline{P}(x) \), respectively.

In constructing the kernel-based model, we consider the kernel-based nonparametric estimation for the probability density function \( p(x; c) \) as follows

\[
\hat{p}(x; c) = \frac{1}{Th} \sum_{t=1}^{T} K \left( \frac{L_t - x}{h} \right),
\]

where \( K(\cdot) \) is a kernel function and \( h \) is a tuning parameter called bandwidth. It is well known that \( \hat{p}(x; c) \) is a consistent estimator of \( p(x; c) \) under some mild conditions; e.g.,
see Theorem 1.1 in Li and Racine (2007) for independent and identically distributed (i.i.d.) sequences and Theorems 5.2 and 5.3 in Fan and Yao (2003) for mixing processes.

A plug-in nonparametric estimator for $H_\alpha(c, q)$ in the objective of (3.11) is given by

$$
\hat{H}_\alpha(c, q) = q + \frac{1}{1 - \alpha} \int_0^{+\infty} (x - q)^+ \hat{p}(x; c) dx
$$

$$
= q + \frac{1}{1 - \alpha} \int_0^{+\infty} (x - q)^+ \frac{1}{T h} \sum_{t=1}^T K \left( \frac{L_t - x}{h} \right) dx
$$

$$
= q + \frac{1}{(1 - \alpha)T} \sum_{t=1}^T \left[ (L_t - q) G(X^{L_t}) - h H(X^{L_t}) \right], \quad (4.18)
$$

where

$$
X^{L_t} = \frac{L_t - q}{h}, \quad G(\nu) = \int_{-\infty}^{\nu} K(t) dt \quad \text{and} \quad H(\nu) = \int_{-\infty}^{\nu} tK(t) dt. \quad (4.19)
$$

We similarly apply the sample mean vector defined in (4.15) to estimate $\mu = \mathbb{E}[X]$, the true expectation of the random return $X$. This leads to a feasible set of $\hat{\mathbb{N}}_P$ as defined in (4.14). Plugging the estimators $\hat{H}_\alpha(c, q)$ and feasible set of $\hat{\mathbb{N}}_P$ into the theoretical model (3.11) gives the following nonparametric reinsurance model:

$$
\min_{(c', q) \in \hat{\mathbb{N}}_P \times [0, \infty)} \hat{Q}_\alpha(c, q), \quad (4.20)
$$

where

$$
\hat{Q}_\alpha(c, q) = \hat{H}_\alpha(c, q) + \zeta' \bar{X}
$$

$$
= q + \frac{1}{(1 - \alpha)T} \sum_{t=1}^T \left[ (L_t - q) G(X^{L_t}) - h H(X^{L_t}) \right] + \zeta' \bar{X}.
$$

5. Existence of Solutions and Convexity of Proposed Models

We first analyze the conditions for the feasible set $\mathbb{N}_P$ of the theoretical model (3.11) and the feasible set $\hat{\mathbb{N}}_P$ in both nonparametric models to be nonempty. Let $\eta = (\eta_1, \ldots, \eta_p)$ with $\eta_i = \theta_i \mu_i$, $i = 1, \ldots, p$, and define

$$
W = \sum_{i=1}^p \beta_i \mu_i, \quad \Delta_2 = W - P, \quad \text{and} \quad \Delta_1 = \sum_{i=1}^p \eta_i - \Delta_2 = \sum_{i=1}^p \eta_i - (W - P). \quad (5.21)
$$

Similarly define $\hat{\eta} = (\hat{\eta}_1, \ldots, \hat{\eta}_p)$ with $\hat{\eta}_i = \theta_i \bar{X}_i$, $i = 1, \ldots, p$, and

$$
\hat{W} = \sum_{i=1}^p \beta_i \bar{X}_i, \quad \hat{\Delta}_2 = \hat{W} - P, \quad \text{and} \quad \hat{\Delta}_1 = \sum_{i=1}^p \hat{\eta}_i - \hat{\Delta}_2 = \sum_{i=1}^p \hat{\eta}_i - (\hat{W} - P).
$$
Proposition 5.1. (a) \( \mathcal{N}_P \neq \emptyset \) if and only if \( \Delta_1 \geq 0 \) and \( \Delta_2 \geq 0 \). Moreover, \( \mathcal{N}_P = \{ c' : c_i = 0, \ i = 1, \ldots, p \} \) for \( \Delta_2 = 0 \), and \( \mathcal{N}_P = \{ c' : c_i = 1, \ i = 1, \ldots, p \} \) for \( \Delta_1 = 0 \).

(b) \( \hat{\mathcal{N}}_P \neq \emptyset \) if and only if \( \hat{\Delta}_1 \geq 0 \) and \( \hat{\Delta}_2 \geq 0 \). Moreover, \( \hat{\mathcal{N}}_P = \{ c' : c_i = 0, \ i = 1, \ldots, p \} \) for \( \hat{\Delta}_2 = 0 \), and \( \hat{\mathcal{N}}_P = \{ c' : c_i = 1, \ i = 1, \ldots, p \} \) for \( \hat{\Delta}_1 = 0 \).

Proof. We only show the proof of part (a) because part (b) can be proved in complete parallel. From (2.4), \( \Psi(c) = \sum_{i=1}^{p} \beta_i \mu_i - \sum_{i=1}^{p} c_i \eta_i = W - c' \eta \), and therefore, it follows from (3.9) that \( \mathcal{N}_P = \{ c' \in [0, 1]^p : c' \eta = W - P \} \). Since \( \eta_i = \theta_i \mu_i > 0 \) for each \( i = 1, \ldots, p \), \( \mathcal{N}_P \neq \emptyset \) if and only if \( 0 \leq W - P \leq \sum_{i=1}^{p} \eta_i \), which is equivalent to \( \Delta_1 \geq 0 \) and \( \Delta_2 \geq 0 \). Again, due to the fact of \( \eta_i > 0 \) for \( i = 1, \ldots, p \), the only element \( c' \in [0, 1]^p \) satisfying \( c' \eta = 0 \) is \( (0, \ldots, 0) \). Similarly, the only element \( c' \in [0, 1]^p \) satisfying \( c' \eta = \sum_{i=1}^{p} \eta_i \) is \( (1, \ldots, 1) \).

Remark 5.1. In view of Proposition 5.1, in what follows we assume \( \Delta_1 > 0 \) and \( \Delta_2 > 0 \) to avoid the two trivial cases where \( \mathcal{N}_P \) is a singleton set. We will show in Proposition 5.2 in the sequel that, under quite general conditions, \( (\hat{W}, \hat{\Delta}_1, \hat{\Delta}_2) \) converge to \( (W, \Delta_1, \Delta_2) \) almost surely as the sample size \( T \) goes to infinity. This fact implies that, given \( \Delta_1 > 0 \) and \( \Delta_2 > 0 \), we must have \( \hat{\Delta}_1 > 0 \) and \( \hat{\Delta}_2 > 0 \) with probability tending to 1, i.e., the feasible set \( \hat{\mathcal{N}}_P \) of the empirical model (4.20) includes more than one element almost surely.

The existence of solutions to the problem (4.12) is obvious as it is equivalent to linear programming (4.17) as we justified in subsection 4.1. The convexity and existence of solutions for models (3.11) and (4.20) can be justified Proposition 5.2 below.

Proposition 5.2. Both (3.11) and (4.20) are convex problems, and have at least one solution.

Proof. We first show the convexity of both models. We note that the constraints in defining the feasible set \( \mathcal{N}_P \times [0, \infty) \) of the problem (3.11) are linear functions of decision variables \( (c', q) \). Therefore, in view of the fact that \( [(1 - c') \mathbf{x} - q]^{+} \) is a convex function of \( (c', q) \) for any given non-negative constant vector \( \mathbf{x} \in [0, \infty)^p \), one can concludes the convexity of the problem (3.11). Following the similar argument used in the proof of Theorem 1 in Yao et al. (2013), we can similarly show the convexity of the problem (4.20), and therefore omit the details.

To show the existence of solutions to each of both models, we first note that the objective functions in both models are continuous. Since \( 0 \leq L(c, \mathbf{X}) \leq \sum_{i=1}^{p} X_i \), a.s.,
for any $c \in \mathbb{R}$, $q^*$ in a minimizer (if it exists) of $Q_\alpha(c, q)$ must satisfy $0 \leq q^* \leq v_\alpha$, where $v_\alpha = \text{VaR}_\alpha \left( \sum_{i=1}^p X_i \right) < \infty$. This means that the solution to problem (3.11) must be within $[0,1]^p \times [0,v_\alpha]$, which is a compact set. Therefore, the problem (3.11) must admit at least one solution due to the well known fact that both the maximum and minimum values of a continuous function are attainable over a compact set. The existence of solutions to problem (4.20) can be proved similarly.

Remark 5.2. The convexity is one of the most appealing feature for an optimization problem as it implies that the optimal solutions can be obtained within polynomial time via various solvers (e.g., the built-in function `fmincon` in Matlab).

6. Convergence Properties of Nonparametric Models

The existence of solutions to each of the three problems (3.11), (4.12) and (4.20) was discussed in the previous section. For presentation convenience, we use $z^* \equiv (c^*, q^*)'$, $\tilde{z}_T \equiv (\tilde{c}_T, \tilde{q}_T)'$ and $\hat{z}_T \equiv (\hat{c}_T, \hat{q}_T)'$ to denote a solution to these three problems respectively. Plugging $z^*$ into the objective of problem (3.11), we obtain the theoretically best CVaR of the insurer’s total risk exposure under the profitability constraint of $\Psi(c) = P$ as follows

$$\text{CVaR}_\alpha(T(c^*)) = Q_\alpha(c^*, q^*).$$

We similarly plugging $\tilde{z}_T$ and $\hat{z}_T$ into the objective of problems (4.12) and (4.20) respectively to obtain the corresponding estimators for $\text{CVaR}_\alpha(T(c^*))$ as follows

$$\text{CVaR}_\alpha(T(c^*)) = \tilde{Q}_\alpha(\tilde{c}, \tilde{q}), \quad \text{and} \quad \text{CVaR}_\alpha(T(c^*)) = \hat{Q}_\alpha(\hat{c}, \hat{q}).$$

A natural question to ask is whether each of the above two estimators works well to estimate the theoretically best CVaR given by $Q_\alpha(c^*, q^*)$. We address such question by establishing the following convergence results:

$$Q_\alpha(\tilde{c}_T, \tilde{q}_T) \xrightarrow{p} Q_\alpha(c^*, q^*), \quad \tilde{Q}_\alpha(\tilde{c}_T, \tilde{q}_T) \xrightarrow{p} Q_\alpha(c^*, q^*), \quad \text{as} \ T \to \infty,$$

and

$$Q_\alpha(\hat{c}_T, \hat{q}_T) \xrightarrow{p} Q_\alpha(c^*, q^*), \quad \hat{Q}_\alpha(\hat{c}_T, \hat{q}_T) \xrightarrow{p} Q_\alpha(c^*, q^*), \quad \text{as} \ T \to \infty.$$

The first convergence result in (6.1) and (6.2) implies that an optimal reinsurance purchase strategy solved from a nonparametric model, either (4.12) or (4.20), leads to a risk,
measured in the risk measure of CVaR, asymptotically converging to the theoretically optimal one. The second convergence result in (6.1) and (6.2) implies that either of \( \tilde{Q}_\alpha(\tilde{c}, \tilde{q}) \) and \( \hat{Q}_\alpha(\hat{c}, \hat{q}) \) constructed from the nonparametric models is a consistent estimator of \( CVaR_\alpha(T(c^*)) = Q_\alpha(c^*, q^*) \).

6.1. Assumptions and Preliminaries

In the rest of the paper, for two functions \( a_T \) and \( b_T \) of the sample size \( T \), we use the notations \( a_T = o(b_T) \) and \( a_T = O(b_T) \) to mean \( \lim_{T \to \infty} a_T/b_T = 0 \) and \( \lim_{T \to \infty} a_T/b_T < \infty \), respectively. Similarly, if \( a_h \) and \( b_h \) are two functions of the bandwidth \( h \), \( a_h = o(b_h) \) and \( a_h = O(b_h) \) respectively denote \( \lim_{h \to 0^+} a_h/b_h \to 0 \) and \( \lim_{h \to 0^+} a_h/b_h < \infty \). Moreover, \( "Z_n \overset{p}{\rightarrow} Z" \) is used to denote the convergence in probability of a sequence \( \{Z_n, n = 1, 2, \ldots\} \) of random variables to a random variable \( Z \).

The convergence results (6.1) and (6.2) will be established for sample \( \{X_t, t = 1, \ldots, T\} \) to have a general dependence structure as follows.

**Definition 6.1.** The process \( \{X_t, t = 1, 2, \ldots\} \) is said to be \( \alpha \)-mixing if \( \alpha(k) \to 0 \) as \( k \to \infty \), where

\[
\alpha(k) = \sup_{A \in \mathcal{F}_1, B \in \mathcal{F}_{i+k}} |\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)|, \quad k = 1, 2, \ldots,
\]

and \( \mathcal{F}_i \) is the \( \sigma \)-algebra of events generated by \( \{X_t, i \leq t \leq j\} \).

The \( \alpha \)-mixing, also called strong mixing, is the weakest dependence among various mixing conditions in the literature (Doukhan [1994]). Intuitively, the mixing condition on a time series means that two elements in the series are almost independent when they are far enough away from each other in time. The \( \alpha \)-mixing time series includes many commonly used linear and nonlinear time series as special cases, such as AR, ARMA, GARCH and so on; see Masry and Tjostheim (1995), Pötscher and Prucha (1997), Cai (2002), Cai (2008), Carrasco and Chen (2002) and Chen and Tang (2005). It certainly includes the i.i.d. structure as one of the special cases as well.

The following Assumption [6.1] is required for the convergence results in (6.1).

**Assumption 6.1. (Process)**

(C1) \( \{X_t, t \geq 1\} \) is a strict stationary \( \alpha \)-mixing process such that \( \mathbb{E}\left[|X_{t,i}|^{r+\delta}\right] < \infty \) for some constants \( r > 2 \) and \( \delta > 0 \), \( t = 1, 2, \ldots\), and \( i = 1, \cdots, p \).
(C2) The mixing coefficients of \( \{X_t, t \geq 1\} \) satisfies \( \alpha(t) = O(t^{-\beta}) \), where \( \beta > \frac{r(\gamma + \delta)}{2\delta} \) with \( r, \delta \) given in condition (C1).

(C3) For any \( \mathbf{c}' \in \mathbb{N}_p, P(x; \mathbf{c}) = \alpha \) admits a unique solution denoted by \( q(\mathbf{c}, \alpha) \) and \( p(q; \mathbf{c}) = \partial P(x, \mathbf{c})/\partial x > 0 \) at \( q = q(\mathbf{c}; \alpha) \). The density function \( p(x; \mathbf{c}) \) has the first derivative \( p'(x; \mathbf{c}) \) which is continuous in the interior of the support of \( p(x; \mathbf{c}) \), and moreover, both \( p(x; \mathbf{c}) \) and \( p'(x; \mathbf{c}) \) are uniformly bounded by certain constant \( D > 0 \).

The moment conditions in (C1) and (C2) are used to apply Lemmas A.2 and A.3 for the development of our consequent results, and the smoothing condition on density function \( p(x; \mathbf{c}) \) in (C3) enables us to apply the Taylor expansion on the density function.

The linear programming model (4.12) does not involve any kernel function in its formulation, and therefore, the discussion on the resulting estimator \( \hat{\text{VaR}}_{\alpha}(T(c^*)) \) will not require any condition on the kernel function. Nevertheless, we need the following Assumption 6.2 for the study of the kernel-based estimator \( \hat{\text{VaR}}_{\alpha}(T(c^*)) \).

**Assumption 6.2. (Kernel and Bandwidth)**

(C4) \( K(\cdot) \) is a symmetric and bounded continuous function on \( \mathbb{R} \), such that \( \int_{-\infty}^{+\infty} K(x)dx = 1 \) and \( \int_{-\infty}^{+\infty} xK(x)dx = 0 \).

(C5) The bandwidth \( h \) is a positive function of the sample size \( T \) such that \( hT \to \infty \) and \( Th^4 \to 0 \) as \( T \to \infty \).

(C6) \( |K(u)| \leq C(1 + |u|)^{-1-\omega} \) for some constants \( \omega > 2 \) and \( C > 0 \).

Assumption 6.2 is commonly considered in the nonparametric statistical literature for kernel-based estimation (e.g., [Li and Racine 2007, Fan and Yao 2003]). Condition (C6) ensures that the second moment of the kernel density function exists, i.e., \( \kappa_2 = \int_{-\infty}^{\infty} u^2K(u)du < \infty \), and moreover,

\[
\int_{-\infty}^{\infty} K(u)du = o(h^2), \quad \text{and} \quad \int_{-\infty}^{\infty} uK(u)du = o(h).
\]

These implications will be useful in the development of our main results in the sequel.

### 6.2. Convergence Results

The main technical difficulty in establishing (6.1) and (6.2) lies in the stochastic nature of the empirical feasible set \( \hat{\mathbb{N}}_p \) as a result of its dependence on sample. The
standard procedure used for the consistency of M-estimators from statistics can not be
directly applied. Our proof relies on the uniform convergence results in Lemmas 6.1 and
6.2 and the technical Lemma 6.3 given in the sequel.

We denote
\[ d_T = T^{-\frac{1}{2}} \log^{\frac{1}{2}} T, \ T = 1, 2, \ldots, \]
and exploit the Euclidean norm \( \| x \| = \sqrt{\sum_{i=1}^{p} x_i^2} \) for any \( x = (x_1, \ldots, x_p) \in \mathbb{R}^p \).

**Lemma 6.1.** If the assumptions (C1)-(C3) are satisfied, then
\[
\| \overline{X} - \mu \| = o (d_T), \ a.s.,
\]
which further implies
\[
\sup_{c' \in [0,1]^p} \left| \tilde{\Psi}(c) - \Psi(c) \right| = o (d_T), \ a.s.
\]
and
\[
| \hat{W} - W | = o(d_T), \ | \hat{\Delta}_1 - \Delta | = o(d_T), \ | \hat{\Delta}_2 - \Delta_2 | = o(d_T), \ a.s.
\]

**Proof.** The proof follows from an application of Lemmas A.1 A.2 and A.3 with details
given in Appendix B. \( \square \)

**Remark 6.1.** From (6.3) in Proposition 6.1, the expected profit of the insurer resulted
from a strategy solved from either of the nonparametric models (4.12) and (4.20) con-
verges to the target value \( P \). Moreover, from the equation (6.5), we must have \( \hat{\Delta}_1 > 0 \)
and \( \hat{\Delta}_2 > 0 \) with probability tending to 1, given that \( \Delta_1 > 0 \) and \( \Delta_2 > 0 \). This observa-
tion is interesting because the feasible set \( \hat{\mathcal{F}}_P \) of the nonparametric models is nonempty
whenever \( \hat{\Delta}_1 > 0 \) and \( \hat{\Delta}_2 > 0 \) as we have shown in Proposition 5.1.

**Lemma 6.2.** Assume that the conditions (C1)-(C3) are satisfied.

(a) Then, for any constant \( M > 0 \),
\[
\sup_{(c',q) \in [0,1]^p \times [0,M]} \left| \tilde{Q}_\alpha(c, q) - Q_\alpha(c, q) \right| = o (d_T), \ a.s.,
\]
and
\[
\sup_{(c',q) \in [0,1]^p \times [0,\infty)} \left| \tilde{Q}_\alpha(c, q) - Q_\alpha(c, q) \right| = o (1), \ a.s.
\]
(b) If the conditions (C4)-(C6) are satisfied additionally, then for any constant \( M > 0 \),

\[
\sup_{(\hat{c}', q) \in [0, 1]^p \times [0, M]} | \hat{Q}_a(\hat{c}, q) - Q_a(\hat{c}, q) | = o(d_T), \quad a.s., \quad (6.8)
\]

and

\[
\sup_{(\hat{c}', q) \in [0, 1]^p \times [0, \infty)} | \hat{Q}_a(\hat{c}, q) - Q_a(\hat{c}, q) | = o(1), \quad a.s. \quad (6.9)
\]

**Proof.** We apply the Convex Lemma (see Lemma A.1) for the proof, with details given in Appendix B. \( \square \)

Define

\[
\Gamma_1 = \frac{\Delta_1}{\sum_{i=1}^{p} \eta_i} \quad \text{and} \quad \Gamma_2 = \frac{\Delta_2}{\sum_{i=1}^{p} \eta_i},
\]

where \( \Delta_1 \) and \( \Delta_2 \) are given in (5.21). According to Proposition 5.1, \( \mathcal{R}_P \neq \emptyset \) if and only if \( \Delta_1 \geq 0 \) and \( \Delta_2 \geq 0 \), which is equivalent to \( 0 \leq \Gamma_1 \leq 1 \) and \( 0 \leq \Gamma_2 \leq 1 \). The following Lemma 6.3 provides us with one way to approximate the value of objective in our nonparametric models at a decision \( \hat{c} \in \hat{\mathcal{R}}_P \) by its value at an element from the theoretical feasible set \( \mathcal{R}_P \).

**Lemma 6.3.** Assume that \( \Delta_1 > 0 \) and \( \Delta_2 > 0 \). For any constant \( \nu \geq 0 \), define event

\[
B_\nu = \{ \bar{X}_i \in [\mu_i - \nu, \mu_i + \nu], \quad i = 1, \ldots, p \}.
\]

Further denote \( \eta_{\min} = \min (\eta_1, \ldots, \eta_p) \), \( \eta_{\max} = \max (\eta_1, \ldots, \eta_p) \) and \( \theta_{\max} = \max (\theta_1, \ldots, \theta_p) \). Then, the following statements hold on events \( B_\nu \).

(a) Given any \( \hat{c} \in \hat{\mathcal{R}}_P \), there exists an element \( \tilde{c} \in \mathcal{R}_{P-\delta} \) such that \( ||\hat{c} - \tilde{c}|| \leq [\delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)] \sqrt{1 + 1/\eta_{\min}^2} \) for any constants \( \nu \) and \( \delta \) satisfying

\[
0 \leq \nu < \min \left( \frac{\Delta_1}{\sum_{i=1}^{p} (\beta_i + \theta_i)}, \frac{\sum_{i=1}^{p} \eta_i}{\sum_{i=1}^{p} \theta_i} \right), \quad \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \leq \delta, \quad (6.10)
\]

and

\[
\delta + \nu (1 + \eta_{\min}) \sum_{i=1}^{p} (\beta_i + \theta_i) \leq \eta_{\min} \Gamma_1. \quad (6.11)
\]
Proposition 6.1. \( \text{Assume that the conditions (C1)-(C3) are satisfied.} \)
See Appendix B.

Proof. We only show the proof of part (b) as part (a) can be obtained similarly. Fix \( \epsilon > 0 \), and define

\[
A_1 = \left\{ \tilde{Q}_a(\tilde{c}_T, \tilde{r}_T) < Q_a(\mathbf{c}^*, q^*) - \epsilon \right\}, \quad A_2 = \left\{ Q(\mathbf{c}, q) < \tilde{Q}_a(\tilde{c}_T, \tilde{r}_T) - \epsilon \right\}. \quad (6.14)
\]

To show \( \tilde{Q}_a(\tilde{z}_T) \overset{p}{\rightarrow} Q_a(\mathbf{z}^*) \), we need to show \( \mathbb{P}(A_1 \cup A_2) \rightarrow 0 \) as \( T \rightarrow \infty \). Further denote

\[
E_1 = \left\{ \left| \tilde{Q}_a(\tilde{c}_T, \tilde{r}_T) - Q_a(\tilde{c}_T, \tilde{r}_T) \right| \leq \epsilon/4 \right\} \quad (6.15)
\]

and

\[
E_2 = \left\{ \left| Q_a(\tilde{c}_T, \tilde{r}_T) - Q(\tilde{c}, \tilde{r}) \right| \leq \epsilon/4 \quad \text{and} \quad \left| Q_a(\tilde{c}, \tilde{r}_T) - Q(\mathbf{c}^*, \tilde{r}_T) \right| \leq \epsilon/4 \right\}
\]

for some \( \tilde{c} \in \mathbb{R}_{P-\delta} \) and \( \mathbf{c}^o \in \mathbb{R}_P \) \( \quad (6.16) \). On \( A_1 \cap E_1 \cap E_2 \),

\[
Q_a(\mathbf{c}^*, q^*) > \tilde{Q}_a(\tilde{c}_T, \tilde{r}_T) + \epsilon \geq Q_a(\tilde{c}_T, \tilde{r}_T) + \frac{3}{4} \epsilon \geq Q_a(\tilde{c}, \tilde{r}_T) + \frac{2}{4} \epsilon \geq Q_a(\mathbf{c}^o, \tilde{r}_T) + \frac{1}{4} \epsilon,
\]
which contradicts to the optimality of \((c^*, q^*)\) in minimizing \(Q_\alpha\) over \(\mathbb{K}_P \times [0, \infty)\). Thus, \(A_1 \cap E_1 \cap E_2 = \emptyset\), and

\[
P(A_1) \leq P(E^*_1) + P(E^*_2),
\]

(6.17)

where \(E^c\) for a set \(E\) denotes the complementary set of \(E\). We further denote

\[
E_3 = \{ |Q_\alpha(c^*, q^*) - Q(c, q^*)| < \epsilon/4 \text{ and } |Q_\alpha(c, q^*) - Q(\tilde{c}_{T}, \tilde{q}_T)| < \epsilon/4 \text{ for some } \tilde{c} \in \mathbb{K}_{P-\delta} \text{ and } \tilde{c}_{T} \in \hat{\mathbb{K}}_P \}
\]

(6.18)

Then, on \(E_3\), \(Q_\alpha(c^*, q^*) > Q(c, q^*) - \epsilon/4 > Q_\alpha(c_{T}, \hat{q}_T) - \epsilon/2\); thus, \(A_2 \cap E_3 = \emptyset\) so that

\[
P(A_2) \leq P(E^*_3).
\]

(6.19)

To proceed, note that \(Q_\alpha(c, q)\) as a function of \(c\) for any given \(q \in [0, \infty)\) is absolutely continuous over \([0, 1]^p\) as justified below: for any \(c_1, c_2 \in [0, 1]^p\),

\[
\begin{align*}
|Q_\alpha(c^{(1)}, q) - Q_\alpha(c^{(2)}, q)| &= \left| q + \frac{1}{1-\alpha} \mathbb{E}[(L(c^{(1)}, X) - q)^+] + \sum_{i=1}^{p}(1 + \theta_i)c^{(1)}_i \mu_i - s_\pi 
- \left( q + \frac{1}{1-\alpha} \mathbb{E}[(L(c^{(2)}, X) - q)^+] + \sum_{i=1}^{p}(1 + \theta_i)c^{(2)}_i \mu_i - s_\pi \right) \right|
\leq \frac{1}{1-\alpha} \mathbb{E} \left[ \left| (c^{(1)} - c^{(2)})' \mu \right| + (1 + \theta_{\max}) \left| (c^{(1)} - c^{(2)})' \mu \right| \right]
\leq \frac{1}{1-\alpha} (1 + \theta_{\max}) \| \mu \| \| c - c^* \|.
\end{align*}
\]

Therefore, for the given \(\epsilon\), there exists a constant \(\kappa > 0\) such that \(|Q_\alpha(c^{(1)}, q) - Q_\alpha(c^{(2)}, q)| \leq \epsilon/4\) whenever \(|c^{(1)} - c^{(2)}| \leq \kappa\) for any \(q \geq 0\). Consequently, for any \(\tilde{c}_{T} \in \hat{\mathbb{K}}_P\), taking both \(\nu\) and \(\delta\) small enough in part (a) of Lemma 6.3, we can find an element \(\tilde{c} \in \mathbb{K}_{P-\delta}\) such that \(\|\tilde{c}_{T} - \tilde{c}\| \leq \kappa\) on event \(B_{\nu}\), which is defined in Lemma 6.3 and given by \(B_{\nu} = \{ X \in [\mu_i - \nu, \mu_i + \nu], \ i = 1, \ldots, p \}\). Therefore,

\[
P \left( |Q_\alpha(\tilde{c}_{T}, \tilde{q}_T) - Q_\alpha(\tilde{c}, \hat{q}_T)| \geq \frac{\epsilon}{4} \forall \tilde{c} \in \mathbb{K}_{P-\delta} \right) \leq P(B^c_{\nu}) \leq \sum_{i=1}^{p} P (|\tilde{\mu}_i - \mu_i| > \nu).
\]

Using the same argument with part (c) of Lemma 6.3 and taking both \(\nu\) and \(\delta\) small enough, for any element \(\tilde{c} \in \mathbb{K}_{P-\delta}\), we can find an element \(\tilde{c}^o \in \mathbb{K}_P\) such that

\[
P (|Q_\alpha(\tilde{c}, \hat{q}_T) - Q_\alpha(\tilde{c}^o, \hat{q}_T)| \geq \epsilon/4 \forall \tilde{c}^o \in \mathbb{K}_P) \leq P(B^c_{\nu}) \leq \sum_{i=1}^{p} P (|\tilde{\mu}_i - \mu_i| > \nu).
\]
The last two displays, along with (6.16), imply
\[ P(E_2^c) \leq 2 \sum_{i=1}^{p} P(|\hat{\mu}_i - \mu_i| > \nu). \] (6.20)

Similarly, using parts (b) and (d) of Lemma 6.3, we can take small \( \delta \) and \( \nu \) such that
\[ P(E_3^c) \leq 2 \sum_{i=1}^{p} P(|\hat{\mu}_i - \mu_i| > \nu). \] (6.21)

Finally, combining (6.14), (6.17), (6.19), (6.20) and (6.21), we get
\[ P\left(\left|\hat{Q}_\alpha(\hat{c}_T, \hat{q}_T) - Q_\alpha(c^*, q^*)\right| > \epsilon\right) \]
\[ \leq P\left\{ \left|\hat{Q}_\alpha(\hat{c}_T, \hat{q}_T) - Q_\alpha(c^*, q^*)\right| > \epsilon/4 \right\} + 4 \sum_{i=1}^{p} P(|\hat{\mu}_i - \mu_i| > \nu) \]
\[ \to 0, \text{ as } T \to \infty, \]

where the last step follows from (6.3) and the uniform convergence results in (6.9). Furthermore,
\[ P\left(\left|Q_\alpha(\hat{c}_T, \hat{q}_T) - Q_\alpha(c^*, q^*)\right| > \epsilon\right) \leq P(A_1) + P(A_2) \]
\[ \leq P\left(\left|\hat{Q}_\alpha(\hat{c}_T, \hat{q}_T) - Q_\alpha(\hat{c}_T, \hat{q}_T)\right| > \epsilon/2 \right) + P\left(\left|\hat{Q}_\alpha(\hat{c}_T, \hat{q}_T) - Q_\alpha(c^*, q^*)\right| > \epsilon/2 \right) \]
\[ \leq P\left\{ \sup_{(c', q) \in [0,1]^p \times [0,\infty)} |\hat{Q}_\alpha(c, q) - Q_\alpha(c, q)| > \epsilon/4 \right\} \]
\[ + P\left(\left|\hat{Q}_\alpha(\hat{c}_T, \hat{q}_T) - Q_\alpha(c^*, q^*)\right| > \epsilon/2 \right) \]
\[ \to 0, \text{ as } T \to \infty, \]
by which the proof is complete. \( \square \)

7. Numerical Examples

In this section, we show some numerical examples to demonstrate the finite sample performance of our proposed nonparametric models. We consider a three-dimensional claim loss vector \( X = (X_1, X_2, X_3) \) in a numerical setting as follows.

A1. The confidence level \( \alpha = 0.95 \) in the risk measure CVaR.
A2. The insurer’s expected total profit $P \in \{10, 30, 50\}$.

A3. The insurance premium loading factor $\beta_i = 5\%, \ i = 1, 2, 3$.

A4. The reinsurance premium loading factor $\theta_i = 10\%, \ i = 1, 2, 3$.

A5. Either exponential distributions or Pareto distributions are considered as the marginal distributions:

A5.1. $X_i$ has an exponential distribution with mean $300 + 100i$, $i = 1, 2, 3$;

A5.2. $X_i$ has a Pareto distribution function $F_{X_i}(x) = 1 - \left(\frac{z_i}{x + z_i}\right)^\mu$, $x > 0$, $i = 1, 2, 3$, with parameters $\mu = 3$, $z_1 = 800$, $z_2 = 1000$ and $z_3 = 1200$, so that $\mathbb{E}[X_i] = 300 + 100i$, $i = 1, 2, 3$.

A6. The loss vector $X$ is governed by a Gaussian copula with two distinct correlation matrices investigated respectively:

A6.1. positive correlation $\Sigma^{(1)} = \begin{pmatrix} 1 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 1 \end{pmatrix}$, and

A6.2. negative correlation $\Sigma^{(2)} = \begin{pmatrix} 1 & 0 & -0.25 \\ 0 & 1 & -0.25 \\ -0.25 & -0.25 & 1 \end{pmatrix}$.

The above assumptions give us four scenarios resulted from combining different types of marginal distributions and different correlation matrices. The numerical simulation is conducted for each scenario with a sample size of $T \in \{50, 100, 200, 1000, 2000, 3000\}$ respectively. The computation regarding the model (4.12) is relatively straightforward, because it boils down to solving a linear programming problem after we plug a sample into the model. In contrast, we have to specify the kernel function and the bandwidth for the kernel-based model (4.20).

It is well known that the nonparametric kernel method is insensitive to the choice of the kernel function but sensitive to the choice of the bandwidth. A commonly used Rule of Thumb in the literature selects the bandwidth $h = 1.06 \cdot \sigma_L T^{-1/3}$, where $\sigma_L$ is the standard deviation of $L(c, X)$. Note that the rate $T^{-1/3}$ is the optimal rate in smooth distribution function estimation in terms of mean square errors; see Li and
The unknown quantity $\sigma_L$ is estimated by

$$\hat{\sigma}_L = \sqrt{\frac{1}{T-1} \sum_{i=1}^{T} (L_t - \overline{L})^2},$$

where $\overline{L} = \frac{1}{T} \sum_{t=1}^{T} L_t$. Let $k = ((1 - c)X_1, \cdots, (1 - c)X_T)$ and $M_0 = I - \frac{1}{T} \cdot 1^{T \times 1}$, where $1 = (1, \ldots, 1)'$ and $I$ is the $T \times T$ identity matrix. Then, $\hat{\sigma}_L$ can be further rewritten as

$$\hat{\sigma}_L = \sqrt{\frac{k'M_0k}{T-1}} = \sqrt{(1 - c)'\Omega(1 - c)},$$

where $\Omega = \frac{1}{T-1}(X_1, \cdots, X_T)M_0(X_1, \cdots, X_T)'$.

In the specific implementation, we apply the Gaussian kernel function $K(u) = (1/\sqrt{2\pi})e^{-u^2/2}$ and select a bandwidth of $h = 1.06 \cdot \hat{\sigma}_L T^{-1/3}$ where $c$ is replaced by $c_0 = 0.5 \times 1$ in computing the value of $\hat{\sigma}_L$.

The specific simulation procedure for each scenario and sample size goes as follows.

S1. Simulate a sample $\{X_t = (X_{1,t}, X_{2,t}, X_{3,t}), \ t = 1, \ldots, T\}$ of size $T$ from the joint distribution of $(X_1, X_2, X_3)$ defined by the given marginal distributions and Gaussian copula.

S2. Plug the simulated sample into model (4.17) and use the Matlab built-in function `linprog` to solve the resulting linear programming problem. Return the optimal solution $(\tilde{c}, \tilde{q})$ and the optimal value $\tilde{\text{CVaR}}_\alpha(T(c^*)) = \tilde{Q}_\alpha(\tilde{c}, \tilde{q})$.

S3. Plug the simulated sample into model (4.20) and use the Matlab built-in function `fmincon` to solve the the resulting constrained convex program. Return the optimal solution $(\hat{c}, \hat{q})$ and the optimal value $\hat{\text{CVaR}}_\alpha(T(c^*)) = \hat{Q}_\alpha(\hat{c}, \hat{q})$.

For each scenario (either Exponential or Pareto marginal distribution, either correlation matrix $\Sigma^{(1)}$ or $\Sigma^{(2)}$) and each sample size $T$, we replicate the above simulation procedure for $M = 2,000$ times independently to obtain $2,000$ independent estimates, denoted by $\hat{V} = \{\hat{V}_j, j = 1, \ldots, M\}$, for the minimum risk measure $\text{CVaR}_\alpha(c^*)$. Based on these $2,000$, we compute its mean $\overline{V} = \frac{1}{M} \sum_{j=1}^{M} \hat{V}_j$ and standard error $SE = \text{St.D}(\hat{V})/\sqrt{T}$, where $\text{St.D}(\hat{V})$ denotes the standard deviation of those $2,000$ simulated estimates $\hat{V}$.

The results are reported in Table 1 with some explanations given in its caption. Upon Table 1 an empirical 95% confidence interval for $E[\text{CVaR}_\alpha(T(c))]$ and $E[\hat{\text{CVaR}}_\alpha(T(\hat{c}))]$ can be constructed in the form of $[\overline{V} - 1.96 \times \text{SD}, \overline{V} + 1.96 \times \text{SD}]$. 

21
To demonstrate how the estimators $E \left[ \tilde{\text{CVaR}}_\alpha (T(\tilde{c})) \right]$ and $E \left[ \hat{\text{CVaR}}_\alpha (T(\hat{c})) \right]$ become more stable as the sample size increases, we build their boxplots based on those 2,000 simulated estimates for the case with an expected profit of 30, a positive correlation $\Sigma^{(1)}$ and exponential marginals. The plots are shown in Figure 1. The boxplots are studied for the other cases and similar trend is observed.

Upon the results reported in Table 1 and Figure 1 we have the following comments.

(a) For sample size as large as 1,000, the values of both $\tilde{\text{CVaR}}_\alpha (T(\tilde{c}))$ and $\hat{\text{CVaR}}_\alpha (T(\hat{c}))$ obtained from our nonparametric models become convergent, regardless of the assumptions on other components in the numerical setting. The boxplot in Figure 1 also indicates such trend. Moreover, the standard error decreases along with sample size, meaning that the estimates are more efficient for large sample size than small one. Combining such observation with Proposition 6.1 we can conclude that both $\tilde{\text{CVaR}}_\alpha (T(\tilde{c}))$ and $\hat{\text{CVaR}}_\alpha (T(\hat{c}))$ are good estimates for the theoretically optimal value of $\text{CVaR}_\alpha (c^*)$ when the sample size is reasonably large.

(b) A comparison in the values of $\tilde{\text{CVaR}}_\alpha (T(\tilde{c}))$ and $\hat{\text{CVaR}}_\alpha (T(\hat{c}))$ among different sample sizes indicates that these two estimates are increasing and eventually convergent along with sample size, and thus, in light of the remarks of part (a), we can conclude that $\tilde{\text{CVaR}}_\alpha (T(\tilde{c}))$ and $\hat{\text{CVaR}}_\alpha (T(\hat{c}))$ underestimate the theoretically optimal value of $\text{CVaR}_\alpha (c^*)$.

(c) If we compare the values of $\tilde{\text{CVaR}}_\alpha (T(\tilde{c}))$ and $\hat{\text{CVaR}}_\alpha (T(\hat{c}))$ between KBM (kernel-based model) and LPM (linear programming model) in Table 1, we can conclude that the two nonparametric models perform comparably well, with KMB performing slightly better that LPM across all the scenarios and sample sizes shown in Table 1.

(d) Table 1 also shows that a positive correlation leads to a higher risk measure than a negative correlation does. It is equally interesting to note that the heavy-tailed Pareto marginal distributions yield to higher risk measure than the light-tailed exponential marginal distributions.

8. Concluding Remarks

In the present paper, a mean-CVaR framework is proposed to determine the optimal multivariate quota-share reinsurance. The solution to the theoretical model depends on
Table 1: The Table reports the mean of the 2,000 empirical estimates of CVaR(\(\alpha\)) solved from either the linear programming model (4.12) or the kernel-based model (4.20) as described in section 7. The numbers reported in parenthesis are the standard errors computed as the standard deviation of those 2,000 estimates divided by the square root of sample size. The top panel reports the simulation results for exponential marginal distributions, while the bottom panel summarizes the results for Pareto marginal distributions. For each combination of the expected profit level \(P\), sample size of \(T\), and marginal distributions, the results for the correlation matrixes \(\Sigma(1)\) and \(\Sigma(2)\) are respectively indicated by "Posi. Corr." and "Nega. Corr." An indication of "KBM" is used for results obtained from the kernel-based model (4.20), whereas a label of "LPM" indicates results from the linear programming model (4.12).
Figure 1: Boxplots of $\hat{\text{CVaR}}_{\alpha}(T(c^*))$ and $\tilde{\text{CVaR}}_{\alpha}(T(c^*))$ for an expected profit of 30, loss vector of weekly positive correlation, and exponential marginals. The graphs indicate that the estimator becomes more stable as the sample size increases. The same pattern has been observed at the other expected profit levels.
the dependence structure which governs the underlying loss vector, and can hardly be obtained analytically. To circumvent such technical issue, we propose two nonparametric models to determine the optimal solutions. The first nonparametric model is developed by replacing the probability measure which governs the mean-CVaR model by its empirical measure and the resulting model becomes a linear programming problem. The second model is developed based on a kernel estimation method for the risk measure of CVaR, and the resulting nonparametric model becomes a convex programming problem.

Compared to the theoretical model, our proposed nonparametric models have the following advantages. First, they are data-driven models and thus distribution free. In formulating the models and exploring their solutions, we do not need to make any explicit assumptions about the distribution of the underlying loss vector. Second, the proposed nonparametric models are computationally friendly, involving solving either a linear programming or a convex programming with linear constraints. Third, convergence results are established under mild conditions, which allow general dependence among the underlying loss vector, and sample from an \( \alpha \)-mixing process. Moreover, our simulation results indicate that the optimal risk measure resulting from a strategy of a nonparametric model approximates the theoretically optimal one well for reasonably large sample size.

Acknowledgements

Weng acknowledges the financial support from the Natural Sciences and Engineering Research Council of Canada (No: 368474), and Society of Actuaries Centers of Actuarial Excellence Research Grant. Zhang thanks support from the Fundamental Research Funds for the Central Universities, the MOE Project of Humanities and Social Sciences (No:13YJA910005), and Zhejiang Provincial Natural Science Foundation (No: LY13A010001).

Appendix A: Technical Lemmas

The proof of our main results relies on several technical lemmas given in this appendix. Lemma A.1 is an extension of the Convexity Lemma in Pollard (1991). Lemma A.1 gives a strong uniform convergence result, whereas what in Pollard (1991) is a convergence in probability.
Lemma A.1. (An Extension of the Convexity Lemma) Let \( \{h_n(u) : u \in U\} \) be a sequence of random convex functions defined on a convex, open subset \( U \) of \( \mathbb{R}^d \). Suppose that \( h(u) \) is a real-valued function on \( U \) for which \( h_n(u) \to h(u) \) a.s. for each \( u \in U \). Then, for each compact subset \( K \) of \( U \),

\[
\sup_{u \in K} |h_n(u) - h(u)| \to 0 \quad \text{a.s.}
\]

The function \( h(\cdot) \) is necessarily convex on \( U \).

Proof. The proof is adapted from that of the Convexity Lemma of Pollard (1991). It is sufficient to show the convergence result for any compact set \( K \) which is a cube with edges all parallel to the coordinate directions, since every compact subset of \( U \) can be covered by finitely many such cubes.

For fixed constant \( \varepsilon > 0 \), since convexity implies continuity, there is a constant \( \delta > 0 \) such that \( h \) varies by less than \( \varepsilon \) over each cube of side 2\( \delta \) that intersects \( K \). For convenience, we may assume that the edge length of \( K \) is an integer multiple of \( \delta \). Partition \( K \) into a union of cubes of side \( \delta \), then expand \( K \) to a larger cube \( K^{\delta} \) by adding an extra layer of these \( \delta \)-cubes around each face. We may assume that \( \delta \) is small enough to ensure that \( K^{\delta} \) lies within \( U \). Write \( V \) for the finite set of all vertices of all the \( \delta \)-cubes that make up \( K \). Then, the assumption that \( h_n(u) - h(u) \overset{a.s.}{\to} 0 \) for each \( u \in U \) implies uniform convergence over \( V \):

\[
\lim_{n \to \infty} P \left( \max_{t \in V} \sup_{k \ge n} |h_k(t) - h(t)| > \varepsilon \right) = 0.
\]

Each \( u \) in \( K \) lies within a \( \delta \)-cube with vertices \( \{u_i\} \in V \), and it can be written as a convex combination \( \sum_i \alpha_i u_i \) of those vertices. Convexity of \( h_k \) gives

\[
h_k(u) \le \sum_i \alpha_i h_k(u_i) \\
\le \sum_i \alpha_i (h_k(u_i) - h(u_i) + h(u_i) - h(u) + h(u)) \\
\le \max_{t \in V} M_k(t) + \max_{i} |h(u_i) - h(u)| + h(u),
\]

where \( M_k(t) = |h_k(t) - h(t)| \). Thus, we have \( \lim_{n \to \infty} P \left( \sup_{u \in K} \sup_{k \ge n} \{h_k(u) - h(u)\} > 2\varepsilon \right) = 0 \). Following the similar steps in the proof of the Convexity Lemma in Pollard (1991), we can similarly show \( \lim_{n \to \infty} P \left( \sup_{u \in K} \sup_{k \ge n} \{h_k(u) - h(u)\} < -2\varepsilon \right) = 0 \), and thus, the desired result follows. \( \square \)

Lemma A.2. Let \( r > 2 \), \( \delta > 0 \) and \( \{X_i, 1 \le i \le T\} \) be an \( \alpha \)-mixing sequence of random variables with \( \mathbb{E}[X_i] = 0 \) and \( \mathbb{E} \left[ |X_i|^{r+\delta} \right] < \infty \). Suppose that \( \alpha(T) \le CT^{-\beta} \) for some
constants $C > 0$ and $\beta > r(r + \delta)/(2\delta)$. Then, for any $\varepsilon > 0$, there exists a positive constant $K = K(\varepsilon, r, \delta, \beta, C) < \infty$ such that

$$
\mathbb{E} \left[ \max_{1 \leq j \leq T} |S_j|^r \right] \leq K \left\{ T^r \sum_{i=1}^{T} \mathbb{E} \left[ |X_i|^r \right] + \left( \sum_{i=1}^{T} \left( \mathbb{E} \left[ |X_i|^r \right] \right) \right)^2 (r + \delta) \right\}^{r/2}.
$$


Lemma A.3. Let $\{b_1, b_2, \ldots\}$ be a nondecreasing unbounded sequence of positive numbers, and $\{a_1, a_2, \ldots\}$ be a sequence of nonnegative numbers. Let $r$ be a fixed positive number such that

$$
\mathbb{E} \left[ \max_{1 \leq j \leq T} |S_j|^r \right] \leq \sum_{j=1}^{T} a_j, \quad T = 1, 2, \ldots,
$$

for some sequence $\{S_j, j = 1, 2, \ldots\}$ of random variables. If $\sum_{j=1}^{T} (a_j/b_j^r) < \infty$, then

$$
\lim_{T \to \infty} \frac{S_T}{b_T} = 0, \text{ a.s.}
$$


Lemma A.4. Assume that the conditions (C1)-(C3) are satisfied. Then, the quantities $D_1$ and $D_2$ defined in (B.3) satisfy $D_1 = o(d_T)$ and $D_2 = o(d_T)$, a.s.

Proof. To show $D_1 = o(d_T)$, a.s., we write $D_1 = \frac{1}{T} \sum_{t=1}^{T} V_t$ with $V_t = (L(c, X_t) - q)^+ - \mathbb{E}[(L(c, X_t) - q)^+]$. Obviously, $\mathbb{E}[V_t] = 0$. Since $|V_t| \leq \sum_{i=1}^{T} X_t + q + \mathbb{E}[(L(c, X_t) - q)^+]$, we have $\mathbb{E}[|V_t|^{r+\delta}] < \infty$ by the condition (C1), for $t = 1, \ldots, T$. By Lemma A.2, for any constant $\varepsilon > 0$, there exists a constant $K > 0$ such that

$$
\mathbb{E} \left[ \max_{1 \leq j \leq T} \left| \sum_{i=1}^{j} V_i \right|^r \right] \leq K \left\{ T^r \sum_{i=1}^{T} \mathbb{E} \left[ |V_i|^r \right] + \left( \sum_{i=1}^{T} \left( \mathbb{E} \left[ |V_i|^r \right] \right) \right)^2 (r + \delta) \right\}^{r/2}
$$

$$
\leq C_1 T^{r/2} \leq C_2 \sum_{t=1}^{T} t^{r/2-1}
$$

for some positive constants $C_1$ and $C_2$. Let $a_t = t^{r/2-1}$ and $b_t = td_t = t^{1/2} \log^{1/2} t$. Then,

$$
\sum_{t=1}^{\infty} \frac{a_t}{b_t^r} = \sum_{t=1}^{\infty} \frac{1}{t \log^{r/2} t} < \infty,
$$

since $r > 2$. Therefore, applying Lemma A.3 yields $\lim_{T \to \infty} \frac{\sum_{t=1}^{T} V_t}{T d_T} = 0$, a.s. Thus, $D_1 = \frac{1}{T} \sum_{t=1}^{T} V_t = o(d_T)$, a.s.
To prove $D_2 = o(d_T)$, a.s., we write $D_2 = \zeta'(\overline{X} - \mu) = \frac{1}{T} \sum_{t=1}^{T} \zeta'(X_t - \mu)$. Note that $E[\zeta'(X_1 - \mu)] = 0$ and $|\zeta'(X_t - \mu)| \leq \sum_{i=1}^{p} (2 + \theta_i + \beta_i) (|X_{i,t}| + |\mu_i|)$, a.s., for $t = 1, \ldots, T$. Then $E[|\zeta'(X_t - \mu)|^{r+\delta}] < \infty$ by the condition (C1) for $t = 1, \ldots, T$. Following the same procedure as what we did for $D_1$, we can similarly obtain $D_2 = o(d_T)$, a.s. □

**Lemma A.5.** Assume that the conditions (C1)-(C6) are satisfied. Then, the quantities $J_1, J_2,$ and $J_3$, defined in (B.7) and (B.8), satisfy

$$J_1 = o(d_T), \quad J_2 = o(d_T) \quad \text{and} \quad J_3 = o(d_T), \quad a.s.$$ 

**Proof.** The strictly stationary property of $\{X_t, t = 1, 2, \ldots\}$ implies

$$E[J_1] = E\left[\frac{1}{(1 - \alpha)} (L_1 - q) G \left( \frac{L_t - q}{h} \right) \right] - \frac{1}{1 - \alpha} E(L - q)^+$$

$$= \frac{1}{(1 - \alpha)} \int_{-\infty}^{\infty} K(u) du (y) dy - \frac{1}{1 - \alpha} E(L - q)^+$$

$$= -\frac{1}{(1 - \alpha)} \int_{-\infty}^{\infty} K(u) (y - q) du (y) dy + \frac{1}{1 - \alpha} E(L - q)^+$$

$$= \frac{1}{(1 - \alpha)} \int_{-\infty}^{\infty} K(u) \left( \int_{-\infty}^{\infty} P(y) dy + uh\overline{P}(uh + q) \right) du - \frac{1}{1 - \alpha} \int_{-\infty}^{\infty} \overline{P}(y) dy.$$

Applying the Taylor expansion, we get for every $u \in \mathbb{R}$,

$$\left| uh\overline{P}(uh + q) - \int_{-\infty}^{uh+q} \overline{P}(y) dy \right|$$

$$= \left| uh \left[ \overline{P}(q) - p(q + \xi)uh \right] - \left( \overline{P}(q)uh - p(q + \xi') \left( \frac{uh}{2} \right)^2 \right) \right|$$

$$\leq 3D \cdot (uh)^2/2,$$

where $\xi$ and $\xi'$ (both depending on $u$) are two constants between 0 and $uh$, and $D$ is the bound for the density function $p(\cdot)$ given in condition (C3). Therefore, (A.1) implies

$$|E[J_1]| \leq \frac{3D}{2} \frac{1}{(1 - \alpha)} \int_{-\infty}^{\infty} K(u)(uh)^2 du = \frac{3D}{2} \frac{1}{(1 - \alpha)} \kappa_2 h^2,$$

i.e., $E[J_1] = O(h^2)$. Thus, the condition (C5) implies $E[J_1] = o(d_T)$.

Next we show $J_1 - E[J_1] = o(d_T)$, a.s. To this end, we write $J_1 - E[J_1] = \frac{1}{T} \sum_{t=1}^{T} Z_t$ with

$$Z_t = (L_t - q)G \left( \frac{L_t - q}{h} \right) - E \left[ (L_t - q)G \left( \frac{L_t - q}{h} \right) \right], \quad t \geq 1.$$
Since $|Z_t| \leq |L_t| + q + \mathbb{E}[(L_t - q)G(\frac{L_t - q}{h})]$, we have $\mathbb{E}|Z_t|^{r+\delta} < \infty$ in view of the conditions (C1). Following the same procedure in the proof of Lemma A.4 for the quantity $D_1$, we can similarly obtain $J_1 - \mathbb{E}[J_1] = \frac{1}{p} \sum_{t=1}^T Z_t = o(d_T)$, a.s., which combined with the fact of $\mathbb{E}[J_1] = o(d_T)$ leads to $J_1 = o(d_T)$, a.s.

Now we analyze $J_2 = \frac{h}{(1-\alpha)T} \sum_{t=1}^T \mathbb{E} \left[ H \left( \frac{L_t - q}{h} \right) \right]$.

$$
\mathbb{E} \left[ H \left( \frac{L_t - q}{h} \right) \right] = \int_0^\infty H \left( \frac{x - q}{h} \right) p(x) dx = \int_0^\infty \int_{-\infty}^{\frac{x-q}{h}} tK(t) dp(x) dx 
= \int_{-\infty}^{-\frac{q}{h}} \int_0^\infty p(x) dx tK(t) dt + \int_{-\frac{q}{h}}^{\infty} \int_{-\frac{q}{h}}^{\infty} p(x) dx tK(t) dt 
= \int_{-\infty}^{-\frac{q}{h}} tK(t) dt + \int_{-\frac{q}{h}}^{\infty} P(ht + q) tK(t) dt 
= \int_{-\infty}^{\infty} P(q)tK(t) dt + \int_{-\infty}^{-\frac{q}{h}} P(ht + q) tK(t) dt + \int_{-\frac{q}{h}}^{\infty} (P(ht + q) - P(q)) tK(t) dt 
\leq o(h) + I,
$$

and

$$
|I| = h \int_{-\frac{q}{h}}^{\infty} P(q) t^2 K(t) dt \leq h D \int_{-\frac{q}{h}}^{\infty} t^2 K(t) dt \leq h D \kappa = O(h),
$$

where $\xi$ (depending on $t$) is a constant between 0 and $ht$. The above two equations imply $\mathbb{E} \left[ H \left( \frac{L_t - q}{h} \right) \right] = O(h)$, and thus $\mathbb{E}[J_2] = O(h^2)$. Following similar procedure as for $J_1$ (consider $J_2 - \mathbb{E}[J_2]$) and apply Lemmas A.2 and A.3, we can obtain $J_2 = o(d_T)$.

Following the same procedure in the proof of (6.3) in Lemma 6.1, we can similarly prove $J_3 = o(d_T)$, a.s. \hfill \Box

**Appendix B: Proofs of Lemmas 6.2, 6.1 and 6.3**

**Proof of Lemma 6.1.** To show (6.3), it is sufficient to prove $|\bar{X}_i - \mu_i| = o(d_T)$, $i = 1, \ldots, p$. We write $\bar{X}_i - \mu_i = \frac{1}{T} \sum_{t=1}^T V_t$ with $V_t = X_{i,t} - \mu_i$, $t = 1, \ldots, T$. By Lemma A.2, for any constant $\epsilon > 0$, there exists a constant $K > 0$ such that

$$
\mathbb{E} \left( \max_{1 \leq j \leq T} \left| \sum_{t=1}^j V_t \right|^r \right) \leq K \left[ T^\epsilon \sum_{t=1}^T \mathbb{E}|V_t|^r + \left( \sum_{t=1}^T (\mathbb{E}|V_t|^{r+\delta})^{2/(r+\delta)} \right)^{r/2} \right] \leq C_1 T^{r/2} \leq C_2 \sum_{t=1}^T t^{r/2-1}
$$
for some positive constants $C_1$ and $C_2$. Let $a_t = \frac{t^{r/2}}{2} - 1$ and $b_t = t^{1/2} \log^{1/2} t$ so that

$$\sum_{t=1}^{\infty} \frac{a_t}{b_t^r} = \sum_{t=1}^{\infty} \frac{1}{t \log^{r/2} t} < \infty,$$

(B.1)
since $r > 2$. Consequently, applying Lemma A.3 yields $\lim_{T \to \infty} \sum_{t=1}^{T} \frac{V_t}{T d_t} = 0$, a.s. This means that (6.3) holds.

By Continuous Mapping Theorem, $T \Psi(c) - \Psi(c) = o(dT)$, a.s., for fixed $c \in [0,1]$, since $\Psi(c)$ is a continuous function of $X$ for given $c \in [0,1]$. Moreover, $\Psi(c)$, as a function of $c$, is linear (thus convex) over $[0,1]$. Therefore, by the Convexity Lemma (Lemma A.1, the uniform convergence result in (6.4) holds.

Finally, $\hat{W}$, $\hat{\Delta}_1$ and $\hat{\Delta}_2$ are all linear as functions of $X$. Thus, (6.3) and the Continuous Mapping Theorem together imply (6.5).

□

Proof of Lemma 6.2. (a) We first show (6.6). By the Convexity Lemma (Lemma A.1), it suffices to show

$$\left| \tilde{Q}_\alpha(c, q) - Q_\alpha(c, q) \right| = o(dT),$$

a.s., for each $(c', q) \in [0,1]^p \times [0, M]$, since $[0,1]^p \times [0, M]$ is a compact set. We write

$$\tilde{Q}_\alpha(c, q) - Q_\alpha(c, q) = \frac{1}{1 - \alpha} D_1 + D_2,$$

(B.2)

where

$$D_1 = \frac{1}{T} \sum_{t=1}^{T} \left( (L(c, X_t) - q)^+ - E[(L(c, X_t) - q)^+] \right), \text{ and } D_2 = \zeta'(X - \mu).$$

(B.3)

By Lemma A.4, $D_1 = o(dT)$ and $D_2 = o(dT)$, a.s. Thus, (6.6) holds.

Next, we show (6.7). Since $D_2$ is independent of $q$, by (B.2) it is sufficient to show

$$\sup_{(c', q) \in [0,1]^p \times [0, M]} |D_1| = o(1), \text{ a.s.}$$

(B.4)

for some constant $M$. To this end, we further write $D_1 = \tilde{\Gamma}_T(q; c) - \Gamma(q; c)$ with

$$\tilde{\Gamma}_T(q; c) = \frac{1}{T} \sum_{t=1}^{T} (L(c, X_t) - q)^+, \text{ and } \Gamma(q; c) = E[(L(c, X_t) - q)^+] \text{.}$$

Then, the fact of $D_1 = o(dT)$, a.s., along with the Convexity Lemma (Lemma A.1), implies

$$\sup_{(c', q) \in [0,1]^p \times [0, M]} \left| \tilde{\Gamma}_T(q, c) - \Gamma(q, c) \right| = o(dT), \text{ a.s.}$$

(B.5)
Due to the condition (C1), \( \sup_{c \in [0,1]^p} \Gamma(q; c) = E \left[ (\sum_{i=1}^p X_i - q)^+ \right] < \infty \) for any \( q \geq 0 \), and moreover, \( \sup_{c \in [0,1]^p} \Gamma(q; c) \) is decreasing to 0 as \( q \to \infty \). Therefore, given a constant \( \epsilon > 0 \), we can take sufficiently large \( M \) such that \( \sup_{c' \in [0,1]^p} \Gamma(q; c) \leq \epsilon/4 \) for any \( q \geq M \). Since \( \Gamma(q; c) \leq \sup_{c' \in \mathbb{N}_p} \Gamma(M, c) \) and \( \hat{\Gamma}_T(q; c) \leq \sup_{c' \in [0,1]^p} \hat{\Gamma}_T(M, c) \) for \( (c', q) \in [0,1]^p \times (M, \infty) \), we obtain

\[
\sup_{(c', q) \in [0,1]^p \times (M, \infty)} \left| \hat{\Gamma}_T(q; c) - \Gamma(q; c) \right| \leq \sup_{(c', q) \in [0,1]^p \times [M, \infty)} \hat{\Gamma}_T(q; c) + \sup_{(c', q) \in [0,1]^p \times [M, \infty)} \Gamma(q; c) 
\leq \sup_{c' \in [0,1]^p} \left| \hat{\Gamma}_T(M; c) - \Gamma(M; c) \right| + 2 \sup_{c' \in [0,1]^p} \Gamma(M; c) 
\leq \sup_{c' \in [0,1]^p} \left| \hat{\Gamma}_T(M; c) - \Gamma(M; c) \right| + \epsilon/2,
\]

which, combined with (B.5), yields

\[
\mathbb{P} \left( \sup_{c' \in [0,1]^p \times [M, \infty)} \sup_{k \geq T} \left| \hat{\Gamma}_k(q, c) - \Gamma(q, c) \right| > \epsilon \right) 
\leq \mathbb{P} \left( \sup_{c' \in [0,1]^p \times [M, \infty)} \sup_{k \geq T} \left| \hat{\Gamma}_k(M, c) - \Gamma(M, c) \right| > \epsilon/2 \right) \to 0, \quad \text{as } T \to \infty,
\]

i.e., (B.4) holds. This proves (6.7).

(b) We first show (6.8). By the Convexity Lemma (Lemma A.1), it suffices to show the result for fixed \( (c', q) \in \mathbb{N}_p \times [0, M] \). To this end, we write

\[
\hat{Q}_\alpha(c, q) - Q_\alpha(c, q) = J_1 - J_2 + J_3 \quad \text{(B.6)}
\]

with

\[
J_1 = \frac{1}{1 - \alpha} \left( \frac{1}{T} \sum_{t=1}^T (L_t - q)G \left( \frac{L_t - q}{h} \right) - E(L - q)^+ \right), \quad \text{(B.7)}
\]

\[
J_2 = \frac{h}{(1 - \alpha)T} \sum_{t=1}^T H \left( \frac{L_t - q}{h} \right), \quad \text{and} \quad J_3 = \zeta'(\bar{x} - r). \quad \text{(B.8)}
\]

According to Lemma A.5, \( J_1 = o(d_T) \), \( J_2 = o(d_T) \) and \( J_3 = o(d_T) \), a.s., which leads to (6.8). The proof of (6.9) is similar to that of (6.7) and thus omitted. \( \square \)

**Proof of Lemma 6.3.** Since \( \bar{x}_i \in [\mu_i - \nu, \mu_i + \nu], \ i = 1, \ldots, p, \ \hat{W} = \sum_{i=1}^p \beta_i \bar{x}_i \) and \( W = \sum_{i=1}^p \beta_i \mu_i \), we obtain

\[
W - \nu \sum_{i=1}^p \beta_i \leq \hat{W} \leq W + \nu \sum_{i=1}^p \beta_i. \quad \text{(B.9)}
\]
Furthermore, combining $\sum_{i=1}^{p} \eta_i = \sum_{i=1}^{p} \theta_i \mu_i$ and $\sum_{i=1}^{p} \eta_i = \sum_{i=1}^{p} \theta_i \mu_i$ yields
\[
\sum_{i=1}^{p} \theta_i \xi_i \leq \sum_{i=1}^{p} \xi_i \theta_i \leq \sum_{i=1}^{p} \theta_i \eta_i \leq \sum_{i=1}^{p} \theta_i \eta_i + \sum_{i=1}^{p} \theta_i \eta_i.
\] (B.10)

Equations (B.9) and (B.10) will be frequently used in the rest of the proof. Furthermore, we also recall that $0 < \Gamma_1 < 1$ and $0 < \Gamma_2 < 1$ whenever $\Delta_1 > 0$ and $\Delta_2 > 0$.

(a) Since $\nu < \Delta_1/\sum_{i=1}^{p} (\beta_i + \theta_i)$ and $\Delta_1 = \sum_{i=1}^{p} \eta_i - (W - P)$, we obtain
\[
W - P + \nu \sum_{i=1}^{p} \beta_i < \sum_{i=1}^{p} \eta_i - \nu \sum_{i=1}^{p} \theta_i.
\]

Therefore, it follows from (B.9) and (B.10) that
\[
\sum_{i=1}^{p} \tilde{\eta}_i > \sum_{i=1}^{p} \eta_i - \nu \sum_{i=1}^{p} \theta_i > W - P + \nu \sum_{i=1}^{p} \beta_i \geq \tilde{W} - W.
\]

Moreover, given $\tilde{c} \in \tilde{N}_P$, we obtain $\tilde{c}'\tilde{\eta} = \tilde{W} - P$ by the definition of $\tilde{N}_P$, and thus, there must exist an integer $i_0 \in \{1, \ldots, p\}$ such that $\tilde{c}_{i_0} \leq \frac{\tilde{W} - P}{\sum_{i=1}^{p} \tilde{\eta}_i} < 1$, whereby
\[
\tilde{m}_1 := 1 - \left(\frac{\tilde{W} - P}{\sum_{i=1}^{p} \tilde{\eta}_i}\right) > 0.
\] (B.11)

Denote
\[
\tilde{\delta}_1 := W - \tilde{W} + \tilde{c}' (\tilde{\eta} - \eta) + \delta = W - P - \tilde{c}' \eta + \delta.
\] (B.12)

In view of (B.9), (B.10) and the condition of $\nu \sum_{i=1}^{p} (\beta_i + \theta_i) \leq \delta$ given in (6.10), we get
\[
\tilde{\delta}_1 \in \left[\delta - \nu \sum_{i=1}^{p} (\beta_i + \theta_i), \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)\right] \subseteq \left[0, \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)\right].
\] (B.13)

Upon the given $\tilde{c} \in \tilde{N}_P$, we denote $\mathcal{N}_{i_0} = \{1, \ldots, p\} \setminus \{i_0\}$ and construct
\[
\left\{ \begin{array}{l}
\tilde{c}_i = \min \left\{ \tilde{c}_i + \tilde{\delta}_1 \frac{\eta_i}{\eta' \eta}, 1 \right\}, \quad i \in \mathcal{N}_{i_0}, \\
\tilde{c}_{i_0} = \tilde{c}_{i_0} + \tilde{\delta}_1 \frac{\eta_{i_0}}{\eta' \eta} + \sum_{i \neq i_0} \left( \tilde{c}_i + \tilde{\delta}_1 \frac{\eta_i}{\eta' \eta} - 1 \right) \frac{\eta_i}{\eta_{i_0}}.
\end{array} \right.
\] (B.14)

Since $\tilde{c}_i \in [0, 1]$ and $\tilde{\delta}_1 \in [0, \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)]$, we must have $\tilde{c}_i \in [0, 1]$ for all $i \in \mathcal{N}_{i_0}$. Moreover, denoting $\mathcal{I}_{i_0} = \left\{ i \in \mathcal{N}_{i_0} : \tilde{c}_i + \tilde{\delta}_1 \frac{\eta_i}{\eta' \eta} > 1 \right\}$, we get
\[
\tilde{c}_{i_0} = \tilde{c}_{i_0} + \tilde{\delta}_1 \frac{\eta_{i_0}}{\eta' \eta} + \sum_{i \in \mathcal{I}_{i_0}} \left( \tilde{c}_i + \tilde{\delta}_1 \frac{\eta_i}{\eta' \eta} - 1 \right) \frac{\eta_i}{\eta_{i_0}} \leq \tilde{c}_{i_0} + \tilde{\delta}_1 \frac{\eta_{i_0}^2}{\eta' \eta} + \tilde{\delta}_1 \frac{\eta_{i_0}}{\eta_{i_0}} \sum_{i \in \mathcal{I}_{i_0}} \frac{\eta_i^2}{\eta' \eta}
\]
\[
\leq \tilde{c}_{i_0} + \tilde{\delta}_1 \frac{\eta_{i_0}}{\eta_{i_0}} \leq \tilde{c}_{i_0} + \tilde{\delta}_1 / \eta_{\min},
\]
which is within $[0, 1]$ whenever $\hat{\delta}_1/\eta_{\min} \leq \hat{m}_1$. By equations (B.9), (B.10) and (B.11), we obtain

$$\hat{m}_1 = 1 - \frac{\hat{W} - P}{\sum_{i=1}^{p} \hat{\eta}_i} \geq 1 - \frac{W - P + \nu \sum_{i=1}^{p} \beta_i}{\sum_{i=1}^{p} \eta_i - \nu \sum_{i=1}^{p} \theta_i},$$

and in light of (B.13), we get a sufficient condition for $\hat{\delta}_1/\eta_{\min} \leq \hat{m}_1$ as follows:

$$1 - \frac{W - P + \nu \sum_{i=1}^{p} \beta_i}{\sum_{i=1}^{p} \eta_i - \nu \sum_{i=1}^{p} \theta_i} \geq 1 - \nu \sum_{i=1}^{p} \beta_i + \theta_i,$$

which, along with the condition of $\nu < \sum_{i=1}^{p} \eta_i / \sum_{i=1}^{p} \theta_i$, is equivalent to

$$\frac{\delta}{\eta_{\min}} \left( \sum_{i=1}^{p} \eta_i - \nu \sum_{i=1}^{p} \theta_i \right) + \nu \left[ (1 + \sum_{i=1}^{p} \eta_i) \sum_{i=1}^{p} (\beta_i + \theta_i) - \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \sum_{i=1}^{p} \theta_i \right] \leq \Delta_1.$$

Obviously, a sufficient condition for the above inequality to hold is

$$\frac{\delta}{\eta_{\min}} \sum_{i=1}^{p} \eta_i + \nu \left( 1 + \sum_{i=1}^{p} \eta_i \right) \sum_{i=1}^{p} (\beta_i + \theta_i) \leq \Delta_1,$$

which is equivalent to the second inequality in (6.11). Moreover, from (B.14), we obtain

$$\tilde{c}' \eta = \tilde{c}' + \sum_{i=1}^{p} \tilde{\delta}_1 \eta_i \eta_{i}' = \tilde{c}' + \tilde{\delta}_1 = W - (P - \delta). \quad (B.15)$$

The above analysis implies that $\tilde{c} \in \mathcal{A}_{P-\delta}$ whenever (6.10) and (6.11) are satisfied. Furthermore, since $\tilde{c}_i \leq \hat{c}_i + \tilde{\delta}_1 \eta_i / \eta_{\min}$ for $i \in N_0$ and $\tilde{c}_{i_0} \leq \tilde{c}_{i_0} + \tilde{\delta}_1 / \eta_{\min}$,

$$||\tilde{c} - \hat{c}|| \leq \left( \sum_{i \neq i_0} \tilde{\delta}_1^2 \eta_i^2 / \eta_{\min}^2 + \tilde{\delta}_1^2 / \eta_{\min}^2 \right)^{1/2} \leq \tilde{\delta}_1 \sqrt{1 + \frac{1}{\eta_{\min}^2}} \leq \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \sqrt{1 + \frac{1}{\eta_{\min}^2}}.$$

Thus, the proof is complete for part (a).

(b) For any $c^o \in \mathcal{A}_{P-\delta}$, $c^o \eta = W - P + \delta$. Therefore, there must exist $i_0 \in \{1, \ldots, p\}$ such that

$$c^o_{i_0} \geq \frac{W - P + \delta}{\sum_{i=1}^{p} \eta_i} \geq \frac{W - P}{\sum_{i=1}^{p} \eta_i} = \Gamma_2 > 0. \quad (B.16)$$

Denote

$$\tilde{\delta}_2 = c^o_{i_0} (\tilde{\eta} - \eta) + W - \tilde{W} + \delta. \quad (B.17)$$
In view of (B.9), (B.10) and the condition of \( \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \) given in (6.12), we get
\[
\hat{\delta}_2 \in \left[ \delta - \nu \sum_{i=1}^{p} (\beta_i + \theta_i), \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \right] \subseteq \left[ 0, \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \right].
\] (B.18)

Upon \( c^o \), we define
\[
\begin{aligned}
\overline{c}_i^o &= \left( c_i^o - \hat{\delta}_2 \frac{\hat{\eta}_i}{\eta} \right)^+, \quad i \in N_{i_0}, \\
\overline{c}_{i_0}^o &= c_{i_0}^o - \hat{\delta}_2 \frac{\hat{\eta}_{i_0}}{\eta} - \sum_{i \neq i_0} \left( \hat{\delta}_2 \frac{\hat{\eta}_i}{\eta} - c_i^o \right) \frac{\hat{\eta}_i}{\eta_{i_0}} \geq c_{i_0}^o - \frac{\hat{\delta}_2}{\eta_{i_0}}.
\end{aligned}
\] (B.19)

Since \( c_i^o \in [0, 1], \overline{c}_i^o \in [0, 1] \) for all \( i \in N_{i_0} \). Moreover, denoting
\[
J_{i_0} = \left\{ i \in N_{i_0} : \frac{\hat{\delta}_2 \hat{\eta}_i}{\eta} > c_i^o \right\},
\]
we obtain
\[
\overline{c}_{i_0}^o = c_{i_0}^o - \frac{\delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)}{\eta_{i_0}} \geq c_{i_0}^o - \frac{\delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)}{\eta_{i_0}}.
\] (B.20)

Note that \( \hat{\eta}_{i_0} = \theta_{i_0} \hat{\eta}_{i_0} \geq \theta_{i_0} (\mu_{i_0} - \nu) = \eta_{i_0} - \theta_{i_0} \nu \geq \eta_{\text{min}} - \theta_{\text{max}} \nu \) on the set \( B_{\nu} \). Therefore, it follows from (B.20) that
\[
\overline{c}_{i_0}^o \geq c_{i_0}^o - \frac{\delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)}{\eta_{\text{min}} - \theta_{\text{max}} \nu},
\]
and in view of (B.16), a sufficient condition for \( \overline{c}_{i_0}^o \in [0, 1] \) is
\[
\frac{\delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i)}{\eta_{\text{min}} - \theta_{\text{max}} \nu} \leq \Gamma_2,
\]
which is equivalent to the condition (6.13) because of the condition \( \nu < \eta_{\text{min}}/(2\theta_{\text{max}}) \) in (6.12). Moreover, from (B.19),
\[
(c^o)' \hat{\eta} = (c^o)' \hat{\eta} - \sum_{i=1}^{p} \hat{\delta}_2 \frac{\eta_i}{\eta} \hat{\eta}_i = (c^o)' \hat{\eta} - \hat{\delta}_2 = (c^o)' \hat{\eta} - \left[ (c^o)' \left( \hat{\eta} - \eta \right) + W - \hat{W} + \delta \right]
\]
\[
= \hat{W} - P.
\]

The above analysis implies that \( \overline{c}^o \in \hat{N}_P \) whenever conditions (6.12) and (6.13) are satisfied. Furthermore, from equations (B.18), (B.19), (B.20) and the condition \( \nu < \eta_{\text{min}}/(2\theta_{\text{max}}) \) in (6.12),
\[
\|\overline{c}^o - c^o\| \leq \left( \sum_{i \neq i_0} \hat{\delta}_2^2 \frac{\eta_i^2}{(\eta')^2} + \frac{\hat{\delta}_2^2}{\eta_{i_0}^2} \right)^{1/2} \leq \hat{\delta}_2 \sqrt{1 + \frac{1}{\eta_{i_0}^2}} \leq \left[ \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \right] \sqrt{1 + \frac{1}{\eta_{i_0}^2}},
\]
and \( \hat{\eta}_0 = \theta_i X_i = \eta_0 + \theta_i(\bar{X}_i - \mu_i) \geq \eta_{\min} - \theta_{\max} \nu \geq \eta_{\min}/2 \) on the set \( B_\nu \). Therefore, 
\[ \| \hat{c}^o - c^o \| \leq \left[ \delta + \nu \sum_{i=1}^{p} (\beta_i + \theta_i) \right] \sqrt{1 + 4/\eta_{\min}^2} \]
whenever (6.12) and (6.13) are satisfied.

(c) Let \( \nu = 0 \) in the result of part (b). Then, the condition (6.12) is automatically satisfied, and (6.11) becomes \( \delta \leq \eta_{\min} \Gamma_2 \). Moreover, in this case, \( \hat{\mu}_i = \mu_i, \quad i = 1, \ldots, p \), and therefore \( \hat{\mathbb{R}}_P = \mathbb{R}_P \). Applying the result of part (b), we get the desired result.

(d) The proof follows from the result of part (a) by setting \( \nu = 0 \). Indeed, in this case, the condition (6.10) is satisfied automatically, and the condition (6.11) is equivalent to \( \delta \leq \eta_{\min} \Gamma_1 \).

□

References


