Online Supplement for "Portfolio Optimization with Nonparametric Value-at-Risk: A Block Coordinate Descent Method"

Xueting Cui[a], Xiaoling Sun[b], Shushang Zhu[c], Rujun Jiang[d] and Duan Li[e]

[a]School of Mathematics, Shanghai University of Finance and Economics, Shanghai 200433, P. R. China, cui.xueting@shufe.edu.cn

[b]Department of Management Science, School of Management, Fudan University, Shanghai 200433, P. R. China

[c]Department of Finance and Investment, Sun Yat-Sen Business School, Sun Yat-Sen University, Guangzhou 510275, P. R. China, zhuss@mail.sysu.edu.cn

[d]School of Data Science, Fudan University, Shanghai 200433, P. R. China, rjjiang@fudan.edu.cn

[e]Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, N. T., Hong Kong, dli@se.cuhk.edu.hk

1 Proof of Lemma 1

Proof. For given s, let $r = [sN]$. Following Yang (1985), we have $E[T_N(\lambda)] - Q(\lambda) = I_1 + I_2$, where

$$I_1 = \int_0^1 E\left[X_{r,N} - Q\left(\frac{r}{N+1}\right)\right] \frac{1}{h} K\left(\frac{s-\lambda}{h}\right) ds$$

and

$$I_2 = \int_0^1 \left[Q\left(\frac{r}{N+1}\right) - Q(\lambda)\right] \frac{1}{h} K\left(\frac{s-\lambda}{h}\right) ds.$$

Next, we prove $I_1 = o\left(1/\sqrt{N}\right)$ and $I_2 = O(1/N) + O\left(h^2\right)$ uniformly for sufficiently large N and small h.

Recall that, under Assumptions 1, 2 and 3, Theorem 2.2(b) in Bickel (1967) showed that, for any $0 < a < 0.5$, $E[X_{r,N} - Q(\frac{r}{N+1})] = o(1/\sqrt{N})$ holds uniformly for sufficiently large N and r satisfying $aN \leq r \leq (1-a)N$, and this result is independent of the distribution F.

Choose a satisfying $0 < a < \min\{\lambda, 1-\lambda\} \leq 0.5$ and $a > 2\delta$ at the same time, where δ are defined in Assumption 4. Then we have $a - \lambda < 0$ and $1 - a - \lambda > 0$. Set h such that $\frac{a-\lambda}{h} < -c$ and $\frac{1-a-\lambda}{h} > c$, which can be done as $h \to 0$. More specifically, setting h small enough such that the support set $S = [-c, c] \subset [\frac{a-\lambda}{h}, \frac{1-a-\lambda}{h}]$ guarantees $K(\frac{s-\lambda}{h}) = 0$ (and thus

$E[X_{r,N} - Q(\frac{r}{N+1})]\frac{1}{h}K(\frac{s-\lambda}{h}) = 0)$ for $s \le a$ or $s \ge 1-a-\epsilon$ (ϵ is a sufficiently small number). Then we have

$$\begin{aligned} I_1 &= \int_0^1 E\left[X_{r,N} - Q\left(\frac{r}{N+1}\right)\right]\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds \\ &= \int_a^{(1-a-\epsilon)} E\left[X_{r,N} - Q\left(\frac{r}{N+1}\right)\right]K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right) \\ &= o(1/\sqrt{N}) \end{aligned}$$

uniformly, since $K(\cdot)$ is bounded and $E[X_{r,N} - Q(\frac{r}{N+1})] = o(1/\sqrt{N})$ holds uniformly for sufficiently large N and r satisfying $aN \le r = [Ns] \le (1-a-\epsilon)N+1 \le (1-a)N$, and, again, this result is independent of the distribution F.

Now we turn to demonstrate that $I_2 = O(1/N) + O\left(h^2\right)$ holds uniformly for sufficiently large N and small h. Notice that $I_2 = \theta_1 + \theta_2$, where

$$\theta_1 = \int_0^1 \left[Q\left(\frac{r}{N+1}\right) - Q(s)\right]\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds \text{ and } \theta_2 = \int_0^1 [Q(s)-Q(\lambda)]\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds.$$

By the mean value theorem, we have

$$Q\left(\frac{r}{N+1}\right) - Q(s) = Q'(r_s)\left(\frac{r}{N+1} - s\right),$$

where $r_s \in [r/(N+1), s]$ (if $r/(N+1) < s$) or $r_s \in [s, r/(N+1)]$ (if $r/(N+1) \ge s$). Notice that $Ns-1 < r = [Ns] < Ns+1$ (implying $(-1-s)/(N+1) < r/(N+1) - s < (1-s)/(N+1)$) and further that $|Q'(r_s)| < M_1$ because $r_s \in (\delta, 1-\delta)$ for $N \ge 1$. We then have

$$\begin{aligned} |\theta_1| &= \left|\int_0^1 Q'(r_s)\left(\frac{r}{N+1} - s\right)K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right)\right| \\ &= \left|\int_a^{1-a} Q'(r_s)\left(\frac{r}{N+1} - s\right)K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right)\right| \\ &\le \int_a^{1-a} \left|Q'(r_s)\left(\frac{r}{N+1} - s\right)\right|K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right) \\ &\le \frac{M_1}{N+1}\int_0^1 (1+s)\,K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right) \\ &= O(1/N), \end{aligned}$$

where the last equality is due to that $K(\cdot)$ is bounded. Thus, we have $\theta_1 = O(1/N)$.

For θ_2, using Taylor's expansion, we have

$$Q(s) - Q(\lambda) = Q'(\lambda)(s-\lambda) + \frac{1}{2}Q''(s_\lambda)(s-\lambda)^2,$$

where s_λ is an interior point of the interval $[s, \lambda]$ (if $\lambda > s$) or $[\lambda, s]$ (if $s \geq \lambda$). Thus

$$\theta_2 = \int_0^1 Q'(\lambda)(s-\lambda)\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds + \frac{1}{2}\int_0^1 Q''(s_\lambda)\,(s-\lambda)^2\,\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds.$$

Due to $(s-\lambda) < [\min\{\lambda,(1-\lambda)\}]^{-1}(s-\lambda)^2$ for all $s \notin [0,1]$, following the proof of Theorem 1 in Yang (1985), for the first integral of θ_2, we have

$$\begin{aligned}
&\left|\int_0^1 Q'(\lambda)(s-\lambda)\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds\right| \\
&= \left|\int_{s\notin[0,1]} Q'(\lambda)(s-\lambda)\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds\right| \\
&\leq h^2\left|Q'(\lambda)\right|[\min\{\lambda,(1-\lambda)\}]^{-1}\int_{s\notin[0,1]}\left(\frac{s-\lambda}{h}\right)^2 K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right) \\
&= h^2|Q'(\lambda)|[\min\{\lambda,(1-\lambda)\}]^{-1}\int_{u\notin[-\frac{\lambda}{h},\frac{1-\lambda}{h}]} u^2K(u)du,
\end{aligned}$$

where the first equality is due to Assumption 7. Since $Q'(\lambda)$ is assumed to be bounded for all distributions and the boundedness and finite support of $K(\cdot)$ imply $\int_{u\notin[-\frac{\lambda}{h},\frac{1-\lambda}{h}]} u^2K(u)du < M_3$ for some constant $M_3 > 0$, we have

$$\int_0^1 \frac{1}{h}K\left(\frac{s-\lambda}{h}\right)Q'(\lambda)(s-\lambda)ds = O(h^2)$$

uniformly. We also have the following result for the second integral of θ_2,

$$\begin{aligned}
&\left|\frac{1}{2}\int_0^1 Q''(s_\lambda)\,(s-\lambda)^2\,\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds\right| \\
&= \left|\frac{1}{2}\int_a^{1-a} Q''(s_\lambda)\,(s-\lambda)^2\,\frac{1}{h}K\left(\frac{s-\lambda}{h}\right)ds\right| \\
&\leq \frac{h^2}{2}\int_a^{1-a}|Q''(s_\lambda)|\left(\frac{s-\lambda}{h}\right)^2 K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right) \\
&\leq \frac{h^2}{2}M_2\int_0^1\left(\frac{s-\lambda}{h}\right)^2 K\left(\frac{s-\lambda}{h}\right)d\left(\frac{s-\lambda}{h}\right) \\
&\leq \frac{h^2}{2}M_2\int_{-\infty}^{\infty} u^2K(u)du \\
&= O(h^2),
\end{aligned}$$

where the boundedness and finite support of $K(\cdot)$ implies $\int_{-\infty}^{\infty} u^2K(u)du < M_4$ for some constant $M_4 > 0$. Thus $I_2 = \theta_1 + \theta_2 = O(1/N) + O(h^2)$ holds independently of the distribution F. Since $O(1/N) = o(1/\sqrt{N})$, we have $E[T_N(\lambda)] - Q(\lambda) = I_1 + I_2 = o(1/\sqrt{N}) + O(h^2)$. □

2 Proof of Proposition 1

Proof. By definition, VaR is exactly the α-quantile ($\mathrm{VaR}_\alpha = Q(\alpha)$). According to Lemma 1, we have $|E[\mathrm{VaR}_\alpha^k(\boldsymbol{x})] - \mathrm{VaR}_\alpha(\boldsymbol{x})| = o(1/\sqrt{N}) + O(h^2)$ uniformly for $\boldsymbol{x} \in \Omega$. Then

$$\begin{aligned} E[\mathrm{VaR}_\alpha^k(\boldsymbol{x}_N^*)] - \mathrm{VaR}_\alpha(\boldsymbol{x}^*) &= E[\min_{\boldsymbol{x}\in\Omega} \mathrm{VaR}_\alpha^k(\boldsymbol{x})] - \mathrm{VaR}_\alpha(\boldsymbol{x}^*) \\ &\leq E[\mathrm{VaR}_\alpha^k(\boldsymbol{x}^*)] - \mathrm{VaR}_\alpha(\boldsymbol{x}^*) \\ &= o(1/\sqrt{N}) + O(h^2). \end{aligned}$$

On the other hand,

$$\begin{aligned} E[\mathrm{VaR}_\alpha^k(\boldsymbol{x}_N^*)] - \mathrm{VaR}_\alpha(\boldsymbol{x}^*) &= E[\mathrm{VaR}_\alpha^k(\boldsymbol{x}_N^*)] - \min_{\boldsymbol{x}\in\Omega} \mathrm{VaR}_\alpha(\boldsymbol{x}) \\ &\geq E[\mathrm{VaR}_\alpha^k(\boldsymbol{x}_N^*) - \mathrm{VaR}_\alpha(\boldsymbol{x}_N^*)] \\ &= o(1/\sqrt{N}) + O(h^2). \end{aligned}$$

Thus, $|E[\mathrm{VaR}_\alpha^k(\boldsymbol{x}_N^*)] - \mathrm{VaR}_\alpha(\boldsymbol{x}^*)| = o(1/\sqrt{N}) + O(h^2)$. □

3 Proof of Theorem 1

Proof. Notice that $\boldsymbol{\lambda}^k = \boldsymbol{\lambda}$ for all k. At the k-th iteration, the first-order optimality conditions of subproblems

$$\begin{aligned} \boldsymbol{y}^{k+1} &= \arg\min \left\{ \mathcal{L}_{\sigma_k}(\boldsymbol{x}^k, \boldsymbol{y}, \boldsymbol{\lambda}^k) \mid \boldsymbol{y} \in \Re^N \right\}, \\ \boldsymbol{x}^{k+1} &= \arg\min \left\{ \mathcal{L}_{\sigma_k}(\boldsymbol{x}, \boldsymbol{y}^{k+1}, \boldsymbol{\lambda}^k) \mid \boldsymbol{x} \in \Omega \right\}. \end{aligned}$$

are

$$0 \in \partial\phi(\boldsymbol{y}^{k+1}) + \boldsymbol{\lambda} + \sigma_k(\boldsymbol{y}^{k+1} + R\boldsymbol{x}^k) \tag{1}$$

and

$$0 \in R^T\boldsymbol{\lambda} + \sigma_k R^T(\boldsymbol{y}^{k+1} + R\boldsymbol{x}^{k+1}) + N_\Omega(\boldsymbol{x}^{k+1}), \tag{2}$$

respectively.

As (2) can be rewritten as

$$0 \in R^T\boldsymbol{\lambda} + \sigma_k R^T(\boldsymbol{y}^{k+1} + R\boldsymbol{x}^k) + N_\Omega(\boldsymbol{x}^{k+1}) + \sigma_k R^T R(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k),$$

we have that

$$0 \in -R^T\partial\phi(\boldsymbol{y}^{k+1}) + N_\Omega(\boldsymbol{x}^{k+1}) + \sigma_k R^T R(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k).$$

If Algorithm 1 terminates at the k-th iteration, then we have $\|\boldsymbol{y}^{k+1} + R\boldsymbol{x}^{k+1}\| \le \epsilon_1$ and $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^k\| \le \epsilon_2$, which indicates $(\boldsymbol{x}^{k+1}, \boldsymbol{y}^{k+1})$ is an ϵ_1-feasible solution to (P_m) satisfying $\boldsymbol{\eta}$-near first-order stationary condition, where $\|\boldsymbol{\eta}\| = \|\sigma_k R^T R(\boldsymbol{x}^{k+1} - \boldsymbol{x}^k)\|$ is of the order $O(\sigma_k \epsilon_2)$.

If Algorithm 1 does not terminate in a finite number of iterations, then there exists a convergent subsequence $\{\boldsymbol{x}^{k_i}\}$ with an accumulation point $\bar{\boldsymbol{x}}$ since Ω is bounded. By Lemma 4, for sufficient large k_i, we have $\|\boldsymbol{y}^{k_i+1} + R\boldsymbol{x}^{k_i}\| \le \epsilon_1$, which also means that there exists a convergent subsequence of $\{\boldsymbol{y}^{k_i+1}\}$ with an accumulation point $\bar{\boldsymbol{y}}$. Furthermore, there exists a subsequence of $\{(\boldsymbol{x}^{k_i+1}, \boldsymbol{y}^{k_i+1})\}$ with an accumulation point $(\tilde{\boldsymbol{x}}, \bar{\boldsymbol{y}})$. Without loss of generality, we still denote this convergent subsequence as $\{(\boldsymbol{x}^{k_i+1}, \boldsymbol{y}^{k_i+1})\}$.

In the following, we show first that $\bar{\boldsymbol{x}} = \tilde{\boldsymbol{x}}$, and then we conclude that the second part of Theorem 1 holds. Since $\boldsymbol{x}^{k_i+1} = \arg\min\{\mathcal{L}_{\sigma_{k_i}}(\boldsymbol{x}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda}) \mid \boldsymbol{x} \in \Omega\}$ and $\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1}$ is a feasible direction of Ω, we must have

$$\nabla_x \mathcal{L}_{\sigma_{k_i}}(\boldsymbol{x}^{k_i+1}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda})^T(\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1}) \ge 0.$$

On the other hand, since $\mathcal{L}_\sigma(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda})$ is a quadratic function of variable $\boldsymbol{x}$, we have from Taylor expansion that

$$\begin{aligned}
&\mathcal{L}_{\sigma_{k_i}}(\boldsymbol{x}^{k_i}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda}) - \mathcal{L}_{\sigma_{k_i}}(\boldsymbol{x}^{k_i+1}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda}) \\
&= \nabla_x \mathcal{L}_{\sigma_{k_i}}(\boldsymbol{x}^{k_i+1}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda})^T(\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1}) + \frac{\sigma_{k_i}}{2}(\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1})^T R^T R(\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1}) \\
&\ge \frac{\alpha}{2}\|\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1}\|^2,
\end{aligned}$$

where $\alpha = \frac{\sigma_{k_i}\lambda_{min}(R^T R)}{2}$ and $\lambda_{min}(R^T R)$ denotes the minimum eigenvalue of $R^T R$. Since $R^T R \succ 0$, we have $\alpha > 0$.

By Lemma 4, σ_{k_i} should be a constant for all the sufficient large k_i and we denote it as σ. Thus, for sufficient large k_i, we have

$$\mathcal{L}_\sigma(\boldsymbol{x}^{k_i}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda}) \ge \mathcal{L}_\sigma(\boldsymbol{x}^{k_i+1}, \boldsymbol{y}^{k_i+1}, \boldsymbol{\lambda}) \ge \mathcal{L}_\sigma(\boldsymbol{x}^{k_{(i+1)}}, \boldsymbol{y}^{k_{(i+1)}+1}, \boldsymbol{\lambda}),$$

which implies

$$\mathcal{L}_\sigma(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \boldsymbol{\lambda}) = \mathcal{L}_\sigma(\tilde{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \boldsymbol{\lambda}).$$

Then taking limits on both sides of (3) yields

$$\lim_{k_i \to +\infty} \frac{\alpha}{2}\|\boldsymbol{x}^{k_i} - \boldsymbol{x}^{k_i+1}\|^2 = \frac{\alpha}{2}\|\bar{\boldsymbol{x}} - \tilde{\boldsymbol{x}}\|^2 = 0.$$

Denote

$$\bar{\boldsymbol{\lambda}} = \boldsymbol{\lambda} + \sigma(\bar{\boldsymbol{y}} + R\bar{\boldsymbol{x}}) = \lim_{k_i \to +\infty} \boldsymbol{\lambda} + \sigma(\boldsymbol{y}^{k_i+1} + R\boldsymbol{x}^{k_i}) = \lim_{k_i \to +\infty} \boldsymbol{\lambda} + \sigma(\boldsymbol{y}^{k_i+1} + R\boldsymbol{x}^{k_i+1}). \quad (3)$$

According to Proposition 2.4.4 in Clarke (1983) and Lemma 3.6 in Balder (2008) on the continuous property on Clarke generalized gradient and the closedness property of the normal cone of a closed and bounded convex set, we have the following from the two conditions (1) and (2),

$$0 \in \partial\phi(\bar{\boldsymbol{y}}) + \bar{\boldsymbol{\lambda}} \text{ and } 0 \in R^T\bar{\boldsymbol{\lambda}} + N_\Omega(\bar{\boldsymbol{x}}).$$

Hence, $0 \in -R^T\partial\phi(\bar{\boldsymbol{y}}) + N_\Omega(\bar{\boldsymbol{x}})$, where $\|\bar{\boldsymbol{y}} + R\bar{\boldsymbol{x}}\| \leq \epsilon_1$. Therefore, $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$ is an ϵ_1-feasible solution to (P_m) satisfying the first-order stationary condition given by Proposition 2. □

References

Balder EJ (2008) On generalized gradients and optimization. Technical report, lNMB Ph.D. course "Convex Analysis for Optimization".

Bickel PJ (1967) Some contributions to the theory of order statistics. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (The Regents of the University of California).

Clarke FH (1983) *Optimization and Nonsmooth Analysis* (New York: Wiley).

Yang SS (1985) A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc.* 80:1004–1011.