Growth-Optimal Portfolio Selection under CVaR Constraints

October 13, 2017

1 Helping Lemmas

Before proving the main theorems of the paper, we state one known lemma and state and prove two lemmas that will be used repeatedly in this proofs. The first lemma is known as Breiman's generalized ergodic theorem. The second and the third lemmas concern the continuity of the saddle point w.r.t. the probability distribution.

Lemma 1 (Ergodicity, [3])**.** *Let $\mathbf{X} = \{X_i\}_{-\infty}^{\infty}$ be a stationary and ergodic process. For each positive integer i, let T_i denote the operator that shifts any sequence by i places to the left. Let $f_1, f_2, \dots$ be a sequence of real-valued functions such that $\lim_{n\to\infty} f_n(\mathbf{X}) = f(\mathbf{X})$ almost surely, for some function f. Assume that $\mathbb{E}\sup_n |f_n(\mathbf{X})| < \infty$. Then,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i(T^i\mathbf{X}) = \mathbb{E}f(\mathbf{X})$$

almost surely.

We denote by $\mathcal{X} \triangleq [1-B, 1+B]^n$.

Lemma 2 (Continuity and Minimax)**.** *Let $\mathcal{B}, \Lambda, \mathcal{X}$ be compact real spaces. $l:\ \mathcal{B}\times\Lambda\times\mathcal{X} \to \mathbb{R}$ be a continuous function. Denote by $\mathbb{P}(\mathcal{X})$ the space of all probability measures on $\mathcal{X}$ (equipped with the topology of weak-convergence). Then the following function $L^* : \mathbb{P}(\mathcal{X}) \to \mathbb{R}$ is continuous*

$$L^*(\mathbb{Q}) = \inf_{(\mathbf{b},c)\in\mathcal{B}} \sup_{\lambda\in\Lambda} \mathbb{E}_{\mathbb{Q}}\left[l(\mathbf{b}, c, \lambda, x)\right]. \qquad (1)$$

Moreover, for any $\mathbb{Q} \in \mathbb{P}(\mathcal{X})$,

$$\inf_{(\mathbf{b},c)\in\mathcal{B}} \sup_{\lambda\in\Lambda} \mathbb{E}_{\mathbb{Q}}\left[l(\mathbf{b}, c, \lambda, x)\right] = \sup_{\lambda\in\Lambda} \inf_{(\mathbf{b},c)\in\mathcal{B}} \mathbb{E}_{\mathbb{Q}}\left[l(\mathbf{b}, c, \lambda, x)\right].$$

Proof. $\mathcal{B}, \Lambda, \mathcal{X}$ are compact, implying that the function $l(\mathbf{b}, c, \lambda, x)$ is bounded. Therefore, the function $L:\ \mathcal{B}\times\Lambda\times\mathbb{P}(\mathcal{X}) \to \mathbb{R}$, defined as

$$L(\mathbf{b}, c, \lambda, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left[l(\mathbf{b}, c, \lambda, x)\right], \qquad (2)$$

is continuous. By applying Proposition 7.32 from [2], we have that $\sup_{\lambda\in\Lambda}\mathbb{E}_{\mathbb{Q}}\left[l(\mathbf{b}, c, \lambda, X)\right]$ is continuous in $\mathbb{Q}\times\mathcal{B}$. Again applying the same proposition, we get the desired result. The last part of the lemma follows directly from Fan's minimax theorem [4]. □

Lemma 3 (Continuity of the optimal selection)**.** *Let $\mathcal{B}, \Lambda, \mathcal{X}$ be compact real spaces, and let L be as defined in Equation (2). Then, there exist two measurable selection functions $h^{(\mathbf{b},c)}$, h^{λ} such that*

$$h^{(\mathbf{b},c)}(\mathbb{Q}) \in \arg\min_{(\mathbf{b},c)\in\mathcal{B}} \left(\max_{\lambda\in\Lambda} L(\mathbf{b}, c, \lambda, \mathbb{Q})\right),$$
$$h^{\lambda}(\mathbb{Q}) \in \arg\max_{\lambda\in\Lambda} \left(\min_{(\mathbf{b},c)\in\mathcal{B}} L(\mathbf{b}, c, \lambda, \mathbb{Q})\right)$$

for any $\mathbb{Q} \in \mathbb{P}(\mathcal{X})$. Moreover, let L^* be as defined in Equation (1). Then, the set

$$Gr(L^*) \triangleq \{(u^*, v^*, \mathbb{Q}) \mid u^* \in h^{(\mathbf{b},c)}(\mathbb{Q}), v^* \in h^{\lambda}(\mathbb{Q}), \mathbb{Q} \in \mathbb{P}(\mathcal{X})\},$$

is closed in $\mathcal{B}\times\Lambda\times\mathbb{P}(\mathcal{X})$.

Proof. The first part of the proof follows immediately from the minimax measurable theorem of [8] due to the compactness of $\mathcal{B}, \Lambda, \mathcal{X}$ and the properties of the loss function L. The proof of the second part is similar to the one presented in Theorem 3 of [1]. In order to show that $Gr(L^*)$ is closed, it is enough to show that if (i) $\mathbb{Q}_n \to \mathbb{Q}_\infty$ in $\mathbb{P}(\mathcal{X})$; (ii) $u_n \to u_\infty$ in $\mathcal{B}$; (iii) $v_n \to v_\infty$ in Λ and (iv) $u_n \in h^{(\mathbf{b},c)}(\mathbb{Q}_n), v_n \in h^{\lambda}(\mathbb{Q}_n)$ for all n, then,

$$u_\infty \in h^{(\mathbf{b},c)}(\mathbb{Q}_\infty), v_\infty \in h^{\lambda}(\mathbb{Q}_\infty).$$

The function $L(\mathbf{b}, c, \lambda, \mathbb{Q})$, as defined in Equation (2), is continuous. Therefore,

$$\lim_{n\to\infty} L(u_n, v_n, \mathbb{Q}_n) = L(u_\infty, v_\infty, \mathbb{Q}_\infty).$$

It remains to show that $u_\infty \in h^{(\mathbf{b},c)}(\mathbb{Q}_\infty)$ and $v_\infty \in h^{\lambda}(\mathbb{Q}_\infty)$. From the optimality of u_n and v_n, we obtain

$$L(u_\infty, v_\infty, \mathbb{Q}_\infty) = \lim_{n\to\infty} L(u_n, v_n, \mathbb{Q}_n) = \lim_{n\to\infty} L^*(\mathbb{Q}_n). \qquad (3)$$

Finally, from the continuity of L^* (Lemma 2), we get

$$(3) = L^*(\lim_{n\to\infty} \mathbb{Q}_n) = L^*(\mathbb{Q}_\infty),$$

which gives the desired result. □

Corollary 1. *Under the conditions of Lemma 3. Define* $L_n(\mathbf{b}, c, \lambda, \mathbb{Q}) = L(\mathbf{b}, c, \lambda, \mathbb{Q}) + \frac{||(\mathbf{b},c)||^2 - ||\lambda||^2}{n}$ *and denote* $h^{(\mathbf{b},c)}_{L_n}(\mathbb{Q}_n), h^{\lambda}_{L_n}(\mathbb{Q}_n)$ *to be the measurable selection functions of* L_n*. If* $\mathbb{Q}_n \to \mathbb{Q}_\infty$ *weakly in* $\mathbb{P}(\mathcal{X})$ *and* $u_n \in h^{(\mathbf{b},c)}_{L_n}(\mathbb{Q}_n), v_n \in h^{\lambda}_{L_n}(\mathbb{Q}_n)$*, then*

$$L_n(u_n, v_n, \mathbb{Q}_n) \to L(u_\infty, v_\infty, \mathbb{Q}_\infty)$$

almost surely for $u_\infty \in h^{(\mathbf{b},c)}(\mathbb{Q}_\infty)$ *and* $v_\infty \in h^{\lambda}(\mathbb{Q}_\infty)$*.*

Proof. Denote $\hat{u}_n \in h^{(\mathbf{b},c)}(\mathbb{Q}_\infty)$ and $\hat{v}_n \in h^{\lambda}(\mathbb{Q}_\infty)$

$$\begin{aligned} &|L_n(u_n, v_n, \mathbb{Q}_n) - L(u_\infty, v_\infty, \mathbb{Q}_\infty)| \\ &\leq |L_n(u_n, v_n, \mathbb{Q}_n) - L(\hat{u}_n, \hat{v}_n, \mathbb{Q}_n)| \\ &+ |L(\hat{u}_n, \hat{v}_n, \mathbb{Q}_n) - L(u_\infty, v_\infty, \mathbb{Q}_\infty)|. \end{aligned} \tag{4}$$

Note that for every n and for constant $E > 0$,

$$\begin{aligned} &\min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} L(\mathbf{b}, c, \lambda, \mathbb{Q}) - \frac{||\lambda_{\max}||^2}{n} \\ &\leq \min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} L_n(\mathbf{b}, c, \lambda, \mathbb{Q}) \\ &= \min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} \left(\mathbb{E}_{\mathbb{Q}} \left[l(\mathbf{b}, c, \lambda, X) \right] + \frac{||(\mathbf{b},c)||^2 - ||\lambda||^2}{n} \right) \\ &\leq \min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} L(\mathbf{b}, c, \lambda, \mathbb{Q}) + \frac{E}{n}. \end{aligned}$$

Thus, for some constant C, $|L_n(u_n, v_n, \mathbb{Q}_n) - L(\hat{u}_n, \hat{v}_n, \mathbb{Q}_n)| < \frac{C}{n}$ and from Lemma 3, the last summand of Equation 4 also converges to 0 as n approaches ∞, we get the desired result, and clearly, if $h^{(\mathbf{b},c)}(\mathbb{Q}_\infty)$ and $h^{\lambda}(\mathbb{Q}_\infty)$ are singletons, then, the only accumulation point of $\{(v_n, u_n)\}_{n=1}^{\infty}$ is (v_∞, u_∞). □

The importance of Lemma 3 stems from the fact that it proves the continuity properties of the multi-valued correspondences $\mathbb{Q} \to h^{(\mathbf{b},c)}(\mathbb{Q})$ and $\mathbb{Q} \to h^{\lambda}(\mathbb{Q})$. This leads to the knowledge that if for the limiting distribution, $\mathbb{Q}_\infty$, the optimal set is a singleton, then $\mathbb{Q} \to h^{(\mathbf{b},c)}(\mathbb{Q})$ and $\mathbb{Q} \to h^{\lambda}(\mathbb{Q})$ are continuous in $\mathbb{Q}_\infty$.

2 Proof of Theorem 2

Theorem 2 (Optimality of $\mathcal{W}^*$)**.** *For any investment strategy* $\mathbf{S} \in \mathcal{S}_\gamma$*, whose portfolios are* $\mathbf{b}_1, \mathbf{b}_2, \ldots$ *the following holds a.s.*

$$\liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} f(\mathbf{b}_i, X_i) \geq \mathcal{W}^*.$$

Proof. For any given strategy $\mathcal{S} \in \mathcal{S}_\gamma$, we will look at the following sequence:

$$\frac{1}{T} \sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i). \tag{5}$$

where $\tilde{\lambda}_i^* \in h^{\lambda}(\mathbb{P}_{X_i|X_1^{i-1}})$, $\tilde{c}_i^* \in \arg\min_{c\in\mathbb{R}} \left(c + \frac{1}{1-\alpha} \mathbb{E}_{\mathbb{P}_{X_i|X_0^{i-1}}} \left[(-\log(\langle \mathbf{b}_i, \mathbf{X} \rangle) - c)^+ \right] \right)$.

Observe that

$$\begin{aligned} (5) = &\frac{1}{T} \sum_{i=1}^{T} \mathbb{E} \left[l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i) \mid X_1^{i-1} \right] \\ &- \frac{1}{T} \sum_{i=1}^{T} (l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i) \\ &- \mathbb{E} \left[l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X) \mid X_1^{i-1} \right]). \end{aligned}$$

Since $A_i = l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i) - \mathbb{E} \left[l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i) \mid X_1^{i-1} \right]$ is a martingale difference sequence, the last summand converges to 0 a.s., by the strong law of large numbers (see, e.g., [9]). Therefore,

$$\begin{aligned} &\liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i) \\ &= \liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \mathbb{E} \left[l(\mathbf{b}_i, \tilde{c}_i^*, \tilde{\lambda}_i^*, X_i) \mid X_1^{i-1} \right] \\ &\geq \liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \min_{(\mathbf{b},c)\in\mathcal{B}()} \mathbb{E} \left[l(\mathbf{b}, c, \tilde{\lambda}_i^*, X_i) \mid X_1^{i-1} \right], \end{aligned} \tag{6}$$

where the minimum is taken w.r.t. all the $\sigma(X_1^{i-1})$-measurable functions. Because the process is stationary, we get for $\hat{\lambda}_i^* \in h^{\lambda}(\mathbb{P}_{X_0|X_{1-i}^{-1}})$,

$$\begin{aligned} (6) &= \liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} \min_{(\mathbf{b},c)\in\mathcal{B}()} \mathbb{E} \left[l(\mathbf{b}, c, \hat{\lambda}_i^*, X_0) \mid X_{1-i}^{-1} \right] \\ &= \liminf_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T} L^*(\mathbb{P}_{X_0|X_{1-i}^{-1}}). \end{aligned} \tag{7}$$

Using Levy's zero-one law, $\mathbb{P}_{X_0|X_{1-i}^{-1}} \to \mathbb{P}_\infty$ weakly as i approaches ∞ and from Lemma 2 we know that L^* is continuous. Therefore, we can apply Lemma 1 and get that a.s.

$$\begin{aligned} (7) &= \mathbb{E} \left[L^*(\mathbb{P}_\infty) \right] = \mathbb{E} \left[\mathbb{E}_{\mathbb{P}_\infty} \left[l\left(\mathbf{b}_\infty^*, c_\infty^*, \lambda_\infty^*, X_0 \right) \right] \right] \\ &= \mathbb{E} \left[\mathcal{L}\left(\mathbf{b}_\infty^*, c_\infty^*, \lambda_\infty^*, X_0 \right) \right]. \end{aligned} \tag{8}$$

Note also, that due to the complementary slackness condition of the optimal solution, i.e., $\lambda_\infty^* (\mathbb{E}_{\mathbb{P}_\infty} \left[c(\mathbf{b}_\infty^*, c_\infty^*, X_0) \right] - \gamma) = 0$, we get

$$(8) = \mathbb{E} \left[\mathbb{E}_{\mathbb{P}_\infty} \left[u\left(\mathbf{b}_\infty^*, c_\infty^*, X_0 \right) \right] \right] = \mathcal{W}^*.$$

From the uniqueness of λ_∞^*, and using Lemma 3 $\hat{\lambda}_i^* \to \lambda_\infty^*$ as i approaches ∞. Moreover, since l is continuous on a compact set, l is also uniformly continuous. Therefore, for any given $\epsilon > 0$, there exists $\delta > 0$, such that if $|\lambda' - \lambda| < \delta$, then

$$|l(\mathbf{b}, c, \lambda', x) - l(\mathbf{b}, c, \lambda, x)| < \epsilon$$

for any $(\mathbf{b}, c) \in \mathcal{B}$ and $x \in \mathcal{X}$. Therefore, there exists i_0 such that if $i > i_0$ then $|l(\mathbf{b}, c, \hat{\lambda}_i^*, x) - l(\mathbf{b}, c, \lambda_\infty^*, x)| < \epsilon$ for any $(\mathbf{b}, c) \in \mathcal{B}$ and $x \in \mathcal{X}$. Thus,

$$\begin{aligned}
&\liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \lambda_\infty^*, X_i) \\
&- \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \hat{\lambda}_i^*, X_i) \\
&= \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \lambda_\infty^*, X_i) \\
&+ \limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} -l(\mathbf{b}_i, \tilde{c_i^*}, \hat{\lambda}_i^*, X_i) \\
&\geq \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \lambda_\infty^*, X_i) - \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \hat{\lambda}_i^*, X_i) \\
&\geq -\epsilon \ \ a.s.,
\end{aligned}$$

and since ϵ is arbitrary,

$$\begin{aligned}
&\liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \lambda_\infty^*, X_i) \\
&\geq \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \hat{\lambda}_i^*, X_i).
\end{aligned}$$

Therefore we can conclude that

$$\liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, \tilde{c_i^*}, \lambda_\infty^*, X_i) \geq \mathcal{W}^* \ \ a.s.$$

We finish the proof by noticing that since $\mathcal{S} \in \mathcal{S}_\gamma$, then by definition

$$\limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} \min_{c\in\mathbb{R}} \left(c + \frac{1}{1-\alpha} \mathbb{E}_{\mathbb{P}_{X_i|X_0^{i-1}}} \left[(-\log(\langle \mathbf{b}_i, \mathbf{X}\rangle) - c)^+ \right] \right) \leq \gamma$$

a.s. and since λ_∞^* is non negative, we will get the desired result. □

3 Proof of Theorem 3

Before proving Theorem 3, we will prove the following two lemmas:

Lemma 4. *Assume that for any vector $w_1^k \in \mathbb{R}^k$ the random variable $||X_1^k - w_1^k||$ as a continuous distribution. Then, there exists a countable set of experts $\{H_{k,h}\}$ for which*

$$\lim_{k\to\infty} \lim_{h\to\infty} \lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_{k,h}^i, X_i) = \mathcal{W}^* \ \ a.s.,$$

where $(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_{k,h}^i)$ are the predictions made by expert $H_{k,h}$ at round i.

Proof. We start by defining a countable set of experts $\{H_{k,h}\}$ as follow: For each $h = 1, 2, \ldots$, we choose $p_h \in (0, 1)$ such that for the sequence $\{p_h\}_{h=1}^\infty$, $\lim_{h\to\infty} p_h = 0$. Moreover, set $\hat{h} = \lfloor np_h \rfloor$ For expert $H_{k,h}$, we define for a fixed $k \times n$-dimensional vector, denoted by w, the following set,

$$\begin{aligned}
B_{k,h}^{w,(1,t)} \triangleq \{x_i \mid k+1 \leq i \leq t, X_{i-k}^{i-1} &\text{ is among the } \hat{h} \\
\text{nearest neighbors of } \ w \ \text{ among } \ &X_1^k, \ldots, X_{t-k}^{t-1}\},
\end{aligned}$$

where $X_j^{j+k} \triangleq (\mathbf{X}_j, \ldots, \mathbf{X}_{j+k}) \in \mathbb{R}^{k\times n}$. where

$$X_j^{j+k} \triangleq (\mathbf{X}_j, \ldots, \mathbf{X}_{j+k}) \in \mathbb{R}^{k\times n}$$

.

Thus, expert $H_{k,h}$ has a window of length k and it looks for the $\hat{h}$ euclidean nearest-neighbors of w in the past. This results in a set of market vectors $B_{k,h}^{w,(1,t)}$. This set can also be seen as a conditional probability over the space of possible market vectors. Then, each expert recommends the actions $\mathbf{b}, c \in \mathcal{B}$ and $\lambda \in \Lambda$, which are the corresponding minimax solution. More formally, we define

$$\begin{aligned}
&h_{k,h}^{\mathbf{b}}, c(X_1^{t-1}, w) \triangleq \\
&\arg\min_{(\mathbf{b},c)\in\mathcal{B}} \left(\max_{\lambda\in\Lambda} \frac{1}{|B_{k,h}^{w,(1,t)}|} \sum_{x_i \in B_{k,h}^{w,(1,t)}} l_{k,l,t}(\mathbf{b}, c, \lambda, x_i) \right)
\end{aligned}$$

and

$$\begin{aligned}
&h_{k,h}^{\lambda}(X_1^{t-1}, w) \triangleq \\
&\arg\max_{\lambda\in\Lambda} \left(\min_{(\mathbf{b},c)\in\mathcal{B}} \frac{1}{|B_{k,h}^{w,(1,t)}|} \sum_{x_i \in B_{k,h}^{w,(1,t)}} l_{k,l,t}(\mathbf{b}, c, \lambda, x_i) \right)
\end{aligned}$$

for

$$\begin{aligned}
&l_{k,h,t}(\mathbf{b}, c, \lambda, x_i) \triangleq \\
&l(\mathbf{b}, c, \lambda, x_i) + \left(||(\mathbf{b}, c)||^2 - ||\lambda||^2\right) \left(\frac{1}{t} + \frac{1}{h} + \frac{1}{k}\right),
\end{aligned}$$

Using the above, we define the predictions of $H_{k,h}$ to be:

$$H^{\mathbf{b}}, c_{k,h}(X_1^{t-1}) = h^{\mathbf{b}}, c_{k,h}(X_1^{t-1}, X_{t-k}^{t-1}), \quad t = 1, 2, 3, \ldots \tag{9}$$

$$H_{k,h}^{\lambda}(X_1^{t-1}) = h_{k,h}^{\lambda}(X_1^{t-1}, X_{t-k}^{t-1}), \quad t = 1, 2, 3, \ldots \tag{10}$$

We will add two experts: $H_{0,0}$ whose predictions are always $(\mathbf{b}_0, c_0, \lambda_{\max})$ and $H_{-1,-1}$ whose predictions are always $(\mathbf{b}_0, c_0, 0)$.

Fixing $k, h > 0$ and w, we will define a (random) measure $\mathbb{P}_{j,w}^{(k,h)}$ that is the measure concentrated on the set $B_{k,h}^{w,(0,1-j)}$, defined by

$$\mathbb{P}_{j,w}^{(k,h)}(A) = \frac{\sum_{X_i \in B_{k,h}^{w,(0,1-j)}} 1_A(X_i)}{|B_{k,h}^{w,(0,1-j)}|},$$

where 1_A denotes the indicator function of the set $A \subset [-B, B]^n$.

In other words, $\mathbb{P}_{j,w}^{(k,h)}(A)$ is the relative frequency of the the vectors among $X_{1-j+k}, \ldots, X_0$ that fall in the set A. Let $r_{k,l}(w)$ be an arbitrary radius such that

$$\mathbb{P}(||X_{-k}^{-1} - w|| \leq r_{k,l}(w)) = p_l$$

Applying Lemma 1, and using similar arguments as in [5], it is straightforward to prove that for all w, w.p. 1

$$\mathbb{P}_{j,w}^{(k,h)} \to \mathbb{P}_{X_0| \ ||X_{-k}^{-1} - w|| \leq r_{k,l}(w)}$$

weakly as $j \to \infty$, where $\mathbb{P}_{X_0| \ ||X_{-k}^{-1} - w|| \leq r_{k,l}(w)}$ denotes the distribution of the vector X_0 conditioned on the event $||X_{-k}^{-1} - w|| \leq r_{k,l}(w)$.

By definition, $\left(h_{k,h}^{\mathbf{b},c}(X_{1-t}^{-1}, w), h_{k,h}^{\lambda}(X_{1-t}^{-1}, w)\right)$ is the minimax of $l_{t,k,h}$ w.r.t. $\mathbb{P}_{j,w}^{(k,h)}$. The sequence of functions $l_{t,k,h}$ converges uniformly as t approaches ∞ to

$$l_{k,h}(\mathbf{b}, c, \lambda, x) = l(\mathbf{b}, c, \lambda, x) + \left(||\mathbf{b}, c||^2 - ||\lambda||^2\right)\left(\frac{1}{h} + \frac{1}{k}\right).$$

Note also that for any fixed $\mathbb{Q}$, $L_{k,h}(\mathbf{b}, c, \lambda, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left[l_{k,h}(\mathbf{b}, c, \lambda, X)\right]$ is strictly convex in $(\mathbf{b}, c)$ and strictly concave in λ, and therefore, has a unique saddle-point (see, e.g., [7]). Therefore, since w is arbitrary, and following a Corollary 1 of Lemma 3, we get that a.s.

$$(\mathbf{b}_{k,h}^t, c_{k,h}^t) \to (\mathbf{b}_{k,h}^*, c_{k,h}^*) \qquad \lambda_{k,h}^t \to \lambda_{k,h}^*,$$

where $\left(\mathbf{b}_{k,h}^*, c_{k,h}^*, \lambda_{k,h}^*\right)$ is the minimax of $L_{k,h}$ w.r.t. $\mathbb{P}_{X_{-k}^{-1}}^{*(k,h)} \triangleq$. Thus, we can apply Lemma 1 and conclude that as T approaches ∞,

$$\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_{k,h}^i, X_i) \to \mathbb{E}\left[l(\mathbf{b}_{k,h}^*, c_{k,h}^*, \lambda_{k,h}^*, X_0)\right].$$

a.s.. We now evaluate

$$\lim_{h\to\infty} \mathbb{E}\left[l(\mathbf{b}_{k,h}^*, c_{k,h}^*, \lambda_{k,h}^*, X_0)\right].$$

Using the properties of the nearest neighbour estimates (see, e.g., [5]), we get that

$$\mathbb{P}_{X_{-k}^{-1}}^{*(k,h)} \to \mathbb{P}_{\left\{X_0|X_{-k}^{-1}\right\}}$$

weakly as $h \to \infty$. Moreover, the sequence of functions $l_{k,h}$ converges uniformly as h approaches ∞

$$l_k(\mathbf{b}, c, \lambda, x) = l(\mathbf{b}, c, \lambda, x) + \frac{||\mathbf{b}, c||^2 - ||\lambda||^2}{k}.$$

Note also, that for any fixed $\mathbb{Q}$, $L_k(\mathbf{b}, c, \lambda, \mathbb{Q}) = \mathbb{E}_{\mathbb{Q}}\left[l_k(\mathbf{b}, c, \lambda, X)\right]$ is strictly convex-concave, and therefore, has a unique saddle point. Accordingly, by applying Corollary 1 again, we get that a.s

$$(\mathbf{b}_{k,h}^*, c_{k,h}^*) \to (\mathbf{b}_k^*, c_k^*) \qquad \lambda_{k,h}^* \to \lambda_k^*,$$

where $(\mathbf{b}_k^*, c_k^*, \lambda_k^*)$ is the minimax of L_k w.r.t. $\mathbb{P}_{\left\{X_0|X_{-k}^{-1}\right\}}$. Therefore, as h approaches ∞,

$$l(\mathbf{b}_{k,h}^*, c_{k,h}^*, \lambda_{k,h}^*, X_0) \to l\left(\mathbf{b}_k^*, c_k^*, \lambda_k^*, X_0\right)$$

a.s.. Thus, by Lebesgue's dominated convergence,

$$\lim_{h\to\infty} \mathbb{E}\left[l(\mathbf{b}_{k,h}^*, c_{k,h}^*, \lambda_{k,h}^*, X_0)\right] = \mathbb{E}\left[l\left(\mathbf{b}_k^*, c_k^*, \lambda_k^*, X_0\right)\right].$$

Notice that for any $\mathbb{Q} \in \mathbb{P}(\mathcal{X})$, the distance between the saddle point of L_k w.r.t. $\mathbb{Q}$ and the the saddle point of L w.r.t. $\mathbb{Q}$ converges to 0 as k approaches ∞. To see this, notice that

$$\begin{aligned}
&\min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} L(\mathbf{b}, c, \lambda, \mathbb{Q}) - \frac{||\lambda_{\max}||^2}{k} \\
&\leq \min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} L_k(\mathbf{b}, c, \lambda, \mathbb{Q}) \\
&\leq \min_{(\mathbf{b},c)\in\mathcal{B}} \max_{\lambda\in\Lambda} L(\mathbf{b}, c, \lambda, \mathbb{Q}) + \frac{E}{k}
\end{aligned}$$

for some constant E, since $\mathcal{B}$ is bounded. The last part in our proof will be to show that if $(\hat{\mathbf{b}_k^*}, \hat{c_k^*}, \hat{\lambda_k^*})$ is the minimax of L w.r.t. $\mathbb{P}_{\left\{X_0|X_{-k}^{-1}\right\}}$, then as k approaches ∞, $\mathbb{E}\left[l\left(\hat{\mathbf{b}_k^*}, \hat{c_k^*}, \hat{\lambda_k^*}, X_0\right)\right]$ will converge a.s. to $\mathcal{W}^*$ and so $\mathbb{E}\left[l\left(\mathbf{b}_k^*, c_k^*, \lambda_k^*, X_0\right)\right]$.

To show this, we will use the sub-martingale convergence theorem twice. First, we define Z_k as

$$Z_k \triangleq \min_{(\mathbf{b},c)\in\mathcal{B}()} \mathbb{E}\left[\max_{\lambda\in\Lambda()} \mathbb{E}\left[l\left(\mathbf{b}, c, \lambda, X_0\right) \mid X_{-\infty}^{-1}\right] \mid X_{-k}^{-1}\right],$$

where the minimum is taken w.r.t. all $\sigma(X_{-k}^{-1})$-measurable strategies and the maximum is taken w.r.t. all $\sigma(X_{-\infty}^{-1})$-measurable strategies. Notice that Z_k is a super-martingale. We can see this by using the tower property of conditional expectations,

$$\mathbb{E}[Z_{k+1} \mid X_{-k}^{-1}] = \mathbb{E}\left[\mathbb{E}\left[Z_{k+1} \mid X_{-k-1}^{-1}\right] \mid X_{-k}^{-1}\right]$$

and since Z_{k+1} is the optimal choice in $\mathcal{B}$ w.r.t. to X_{-k-1}^{-1},

$$\leq \mathbb{E}\left[\mathbb{E}[Z_k \mid X_{-k-1}^{-1}] \mid X_{-k}^{-1}\right] = \mathbb{E}[Z_k \mid X_{-k}^{-1}] = Z_k.$$

Note also that $\mathbb{E}[Z_k]$ is uniformly bounded. Therefore, we can apply the super-martingale convergence theorem and get that $Z_k \to Z_\infty$ a.s., where,

$$Z_\infty = \mathbb{E}\left[l(\mathbf{b}^*_\infty, c^*_\infty, \lambda^*_\infty, X_0) \mid X^{-1}_{-\infty}\right] = \mathcal{W}^*,$$

and by using Lebesgue's dominated convergence theorem, also $\mathbb{E}[Z_k] \to \mathbb{E}[Z_\infty] = \mathcal{W}^*$. Using the same arguments, Z'_k, defined as

$$Z'_k \triangleq \max_{\lambda\in\Lambda()} \mathbb{E}\left[\min_{(\mathbf{b},c)\in\mathcal{B}()} \mathbb{E}\left[l\left(\mathbf{b}, c, \lambda, X_0\right) \mid X^{-1}_{-\infty}\right] \mid X^{-1}_{-k}\right],$$

where the maximum is taken w.r.t. all $\sigma(X^{-1}_{-k})$-measurable strategies and the minimum is taken w.r.t. all $\sigma(X^{-1}_{-\infty})$-measurable strategies, is a sub-martingale that also converges a.s. to Z_∞ and thus $\mathbb{E}[Z'_k] \to \mathbb{E}[Z_\infty] = \mathcal{W}^*$.

We conclude the proof by noticing that the following relation holds for any k,

$$\begin{aligned}
&\mathbb{E}[Z'_k] \\
&= \mathbb{E}\left[\max_{\lambda\in\Lambda()} \mathbb{E}\left[\min_{(\mathbf{b},c)\in\mathcal{B}()} \mathbb{E}\left[l\left(\mathbf{b}, c, \lambda, X_0\right) \mid X^{-1}_{-\infty}\right] \mid X^{-1}_{-k}\right]\right] \\
&\le \mathbb{E}\left[\max_{\lambda\in\Lambda()} \mathbb{E}\left[\mathbb{E}\left[l\left(\hat{\mathbf{b}^*_k}, \hat{c^*_k}, \lambda, X_0\right) \mid X^{-1}_{-\infty}\right] \mid X^{-1}_{-k}\right]\right] \\
&= \mathbb{E}\left[\max_{\lambda\in\Lambda()} \mathbb{E}\left[l\left(\hat{\mathbf{b}^*_k}, \hat{c^*_k}, \lambda, X_0\right) \mid X^{-1}_{-k}\right]\right] \\
&= \mathbb{E}\left[l\left(\hat{\mathbf{b}^*_k}, \hat{c^*_k}, \hat{\lambda^*_k}, X_0\right)\right],
\end{aligned}$$

and using similar arguments we can show that also

$$\mathbb{E}\left[l\left(\mathbf{b}, \hat{c^*_k}, \hat{\lambda^*_k}, X_0\right)\right] \le \mathbb{E}[Z_k],$$

and since both $\mathbb{E}[Z_k]$ and $\mathbb{E}[Z'_k]$ converge to $\mathcal{W}^*$, we get the desired result.

□

Lemma 5. *For $\{H_{k,h}\}$ the following relation holds a.s.:*

$$\begin{aligned}
&\inf_{k,h} \limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l\left(\mathbf{b}^i_{k,h}, c^i_{k,h}, \lambda_i, X_i\right) \le \mathcal{W}^* \\
&\le \sup_{k,h} \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l\left(\mathbf{b}_i, c_i, \lambda^i_{k,h}, X_i\right),
\end{aligned}$$

where $(\mathbf{b}_i, c_i, \lambda_i)$ are the predictions of CANN when applied on $\{H_{k,h}\}$.

Proof. Set

$$f(\mathbf{b}, c, \mathbb{Q}) \triangleq \max_{\lambda\in\Lambda} \mathbb{E}_{\mathbb{Q}}\left[l\left(\mathbf{b}, c, \lambda, X_0\right)\right].$$

We will start from the LHS,

$$\inf_{k,h} \limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l\left(\mathbf{b}^i_{k,h}, c^i_{k,h}, \lambda_i, X_i\right), \tag{11}$$

and similarly to Lemma 2, by using the strong law of large numbers we can write

$$\begin{aligned}
(11) &= \inf_{k,h} \limsup_{n\to\infty} \frac{1}{T}\sum_{i=1}^{T} \mathbb{E}\left[l\left(\mathbf{b}^i_{k,h}, c^i_{k,h}, \lambda_i, X_0\right) \mid X^{-1}_{1-i}\right] \\
&\le \inf_{k,h} \limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} f(\mathbf{b}^i_{k,h}, c^i_{k,h}, \mathbb{P}_{X_0|X^{-1}_{1-i}}) \;\; a.s. \qquad (12)
\end{aligned}$$

For fixed $k, h > 0$, from the proof of Theorem (4), $(\mathbf{b}^i_{k,h}, c^i_{k,h}) \to (\mathbf{b}^*_{k,h}, c^*_{k,h})$ a.s. as i approaches ∞, and from Levy's zero-one law also $\mathbb{P}_{X_0|X^{-1}_{1-i}} \to \mathbb{P}_\infty$ weakly. From Lemma 2 we know that f is continuous, therefore, we can apply Lemma 1 and get that

$$\begin{aligned}
(12) &= \inf_{k,h} \mathbb{E}\left[\mathbb{E}\left[f(\mathbf{b}^*_{k,h}, c^*_{k,h}, \mathbb{P}_\infty)\right]\right] \\
&\le \lim_{k\to\infty} \lim_{h\to\infty} \mathbb{E}\left[f((\mathbf{b}^*_{k,h}, c^*_{k,h}, \mathbb{P}_\infty)\right]. \qquad (13)
\end{aligned}$$

From the uniqueness of the saddle point and from the proof of Theorem (4), for fiked $k > 0$,

$$\lim_{h\to\infty} (\mathbf{b}^*_{k,h}, c^*_{k,h}) \to (\mathbf{b}^*_k, c^*_k)$$

a.s.. Thus, from the continuity of f we get that

$$\lim_{h\to\infty} f(\mathbf{b}^*_{k,h}, c^*_{k,h}, \mathbb{P}_\infty) \to f(\mathbf{b}^*_k, c^*_k, \mathbb{P}_\infty)$$

and again by Lebesgue's dominated convergence,

$$\begin{aligned}
(13) &= \lim_{k\to\infty} \mathbb{E}\left[f(\mathbf{b}^*_k, c^*_k, \mathbb{P}_\infty)\right] \\
&= \lim_{k\to\infty} \mathbb{E}\left[\max_{\lambda\in\Lambda} \mathbb{E}_{\mathbb{P}_\infty}\left[l\left(\mathbf{b}^*_k, c^*_k, \lambda, X_0\right)\right]\right]. \qquad (14)
\end{aligned}$$

Now, from Theorem 4 we know that every accumulation point of the sequence $\{(\mathbf{b}^*_k, c^*_k)\}$ is in the optimal set

$$\arg\min_{(\mathbf{b},c)\in\mathcal{B}} \left(\max_{\lambda\in\Lambda} \mathbb{E}_{\mathbb{P}_\infty}\left[l\left(\mathbf{b}, c, \lambda, X_0\right)\right]\right).$$

Therefore a.s.

$$\lim_{k\to\infty} \max_{\lambda\in\Lambda} \mathbb{E}_{\mathbb{P}_\infty}\left[l\left(\mathbf{b}^*_k, c^*_k, \lambda, X_0\right)\right] \to \mathbb{E}_{\mathbb{P}_\infty}\left[l\left(\mathbf{b}^*_\infty, c^*_\infty, \lambda^*_\infty, X_0\right)\right],$$

and using Lebesgue's dominated convergence,

$$(14) = \mathbb{E}\left[\mathbb{E}_{\mathbb{P}_\infty}\left[l\left(\mathbf{b}^*_\infty, c^*_\infty, \lambda^*_\infty, X_0\right)\right]\right] = \mathcal{W}^*.$$

Using similar arguments, we can show the second part of the lemma.

□

We are now ready to state and prove the optimality of CANN .

Theorem 3 (Optimality of CANN)**.** *Assume that for any vector $w_1^k \in \mathbb{R}^k$ the random variable $||X_1^k - w_1^k||$ as a continuous distribution. Then, for any $\gamma > 0$ and for any bounded process $\{X_i\}_{-\infty}^{\infty}$: CANN is a γ-bounded and γ-universal investment strategy.*

Proof. We first show that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, X_i) = \mathcal{W}^* \ \ a.s. \qquad (15)$$

Applying Lemma 5 in [6], we know that updates of CANN guarantee that for every expert $H_{k,h}$,

$$\frac{1}{T0}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i) \leq \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_i, x_i) + \frac{C_{k,h}}{\sqrt{T}} \qquad (16)$$

$$\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i) \geq \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_{k,h}^i, x_i) - \frac{C'_{k,h}}{\sqrt{T}}, \qquad (17)$$

where $C_{k,h}, C'_{k,h} > 0$ are some constants independent of T. In particular, using Equation (16),

$$\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i) \leq \inf_{k,h}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_i, x_i) + \frac{C_{k,h}}{\sqrt{T}}\right).$$

Therefore, we get

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i) \\
&\leq \limsup_{T\to\infty}\inf_{k,h}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_i, x_i) + \frac{C_{k,h}}{\sqrt{T}}\right) \\
&\leq \inf_{k,h}\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_i, x_i) + \frac{C_{k,h}}{\sqrt{T}}\right) \\
&\leq \inf_{k,h}\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_{k,h}^i, c_{k,h}^i, \lambda_i, x_i)\right), \qquad (18)
\end{aligned}$$

where in the last inequality we used the fact that $\limsup$ is sub-additive. Using Lemma (5), we get that

$$\begin{aligned}
&(18) \leq \mathcal{W}^* \\
&\leq \sup_{k,h}\liminf_{n\to\infty} \frac{1}{N}\sum_{i=1}^{N} l\left(\mathbf{b}_i, c_i, \lambda_{k,h}^i, X_i\right). \qquad (19)
\end{aligned}$$

Using similar arguments and using Equation (17) we can show that

$$(19) \leq \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i).$$

Summarizing, we have

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i) \leq \mathcal{W}^* \\
&\leq \liminf_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i).
\end{aligned}$$

Therefore, we can conclude that a.s.

$$\lim_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, X_i) = \mathcal{W}^*.$$

To show that MHA is indeed a γ-bounded strategy and to shorten the notation, we will denote

$$g(\mathbf{b}, c, \lambda, x) \triangleq \lambda\left(c + \frac{1}{1-\alpha}\left(\omega(\mathbf{b}, x) - c\right)^+ - \gamma\right).$$

First, from Equation (17) applied on the expert $H_{0,0}$, we get that:

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_{\max}, x_i) \\
&\leq \limsup_{T\to\infty} \frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_i, x_i). \qquad (20)
\end{aligned}$$

Moreover, since l is uniformly continuous, for any given $\epsilon > 0$, there exists $\delta > 0$, such that if $|\lambda' - \lambda| < \delta$, then

$$|l(\mathbf{b}, c, \lambda', x) - l(\mathbf{b}, c, \lambda, x)| < \epsilon$$

for any $(\mathbf{b}, c) \in \mathcal{B}$ and $x \in \mathcal{X}$. We also know that

$$\lim_{k\to\infty}\lim_{h\to\infty}\lim_{i\to\infty} \lambda_{k,h}^i = \lambda_\infty^*.$$

Therefore, there exist k_0, h_0, i_0 such that $|\lambda_{k_0,h_0}^i - \lambda_\infty^*| < \delta$ for any $i > i_0$. Since $\lim_{k\to\infty}\lambda_k^* = \lambda_\infty^*$ there exists k_0 such that $|\lambda_{k_0}^* - \lambda_\infty^*| < \frac{\delta}{3}$. Note that $\lim_{h\to\infty}\lambda_{k_0,h}^* = \lambda_{k_0}^*$, so there exists h_0 such that $|\lambda_{k_0,h_0}^* - \lambda_{k_0}^*| < \frac{\delta}{3}$. Finally, since $\lim_{i\to\infty}\lambda_{k_0,l_0}^i = \lambda_{k_0,l_0}^*$, there exists i_0 such that if $i > i_0$, then $|\lambda_{k_0,l_0}^i - \lambda_{k_0,l_0}^*| < \frac{\delta}{3}$. Combining all the above, we get that for k_0, h_0, i_0 if $i > i_0$, then

$$\begin{aligned}
&|\lambda_{k_0,h_0}^i - \lambda_\infty^*| < |\lambda_{k_0,h_0}^i - \lambda_{k_0,h_0}^*| \\
&+|\lambda_{k_0,h_0}^i - \lambda_{k_0}^*| + |\lambda_{k_0}^* - \lambda_\infty^*| < \delta.
\end{aligned}$$

Therefore,

$$\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) - \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i)\right) \le$$
$$\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) - \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_{k_0,h_0}^i, x_i)\right) +$$
$$\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_{k_0,h_0}^i, x_i) - \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i)\right) \tag{21}$$

From the uniform continuity we also learn that the first summand is bounded above by ϵ, and from Equation (17), we get that the last summand is bounded above by 0. Thus,

$$(21) \le \epsilon,$$

and since ϵ is arbitrary, we get that

$$\limsup_{T\to\infty}\left(\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) - \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i)\right) \le 0.$$

Thus,

$$\limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_\infty^*, X_i) \le \mathcal{W}^*,$$

and from Theorem 1 we can conclude that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_\infty^*, X_i) = \mathcal{W}^*.$$

Therefore, we can deduce that

$$\begin{aligned}
&\limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_i, x_i) \\
&- \limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) = \\
&\limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{N} g(\mathbf{b}_i, c_i, \lambda_i, x_i) \\
&+ \liminf_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} -g(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) \\
&\le \limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_i, x_i) - \frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) \\
&= \limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_i, x_i) - \frac{1}{T}\sum_{i=1}^{T} l(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i) \\
&= 0,
\end{aligned}$$

which results in

$$\begin{aligned}
&\limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_i, x_i) \\
&\le \limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i).
\end{aligned}$$

Combining the above with Equation (20), we get that

$$\begin{aligned}
&\limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_{\max}, x_i) \\
&\le \limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_\infty^*, x_i).
\end{aligned}$$

Since $0 \le \lambda_\infty^* < \lambda_{\max}$, we get that

$$\begin{aligned}
&\limsup_{T\to\infty}\frac{1}{T}\sum_{i=1}^{T} g(\mathbf{b}_i, c_i, \lambda_{\max}, x_i) \\
&\le \gamma.
\end{aligned}$$

Since the choices of c_i are not necessarily optimal w.r.t. $\mathbb{P}_{X_0|X_{-i-1}^{-1}}$ we get that CANN is γ-bounded. Using Equation (15), we get that CANN is also γ-universal.

□

References

[1] P.H. Algoet and T.M. Cover. Asymptotic optimality and asymptotic equipartition properties of log-optimum investment. *The Annals of Probability*, pages 876–898, 1988.

[2] D. Bertsekas and S. Shreve. *Stochastic optimal control: The discrete time case*, volume 23. Academic Press New York, 1978.

[3] L. Breiman. The individual ergodic theorem of information theory. *The Annals of Mathematical Statistics*, 28(3):809–811, 1957.

[4] K. Fan. Minimax theorems. *Proceedings of the National Academy of Sciences*, 39(1):42–47, 1953.

[5] L. Györfi, F. Udina, and H. Walk. Nonparametric nearest neighbor based empirical portfolio selection strategies. *Statistics & Decisions, International Mathematical Journal for Stochastic Methods and Models*, 26(2):145–157, 2008.

[6] Y. Kalnishkan and M. Vyugin. The weak aggregating algorithm and weak mixability. In *International Conference on Computational Learning Theory*, pages 188–203. Springer, 2005.

[7] Y. Lou, Y. Hong, L. Xie, G. Shi, and K. Johansson. Nash equilibrium computation in subnetwork zero-sum games with switching communications. *IEEE Transactions on Automatic Control*, 61(10):2920–2935, 2016.

[8] A. Nowak. Measurable selection theorems for minimax stochastic optimization problems. *SIAM Journal on Control and Optimization*, 23(3):466–476, 1985.

[9] W. Stout. Almost sure convergence, vol. 24 of probability and mathematical statistics, 1974.