J. Math. Anal. Appl. 451 (2017) 906–923

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

The least squares estimator of random variables under sublinear expectations

Sun Chuanfeng [a,1], Ji Shaolin [b,*,2]

[a] *School of Mathematical Sciences, University of Jinan, Jinan, Shandong 250022, PR China*
[b] *Institute for Financial Studies and Institute of Mathematics, Shandong University, Jinan, Shandong 250100, PR China*

A R T I C L E I N F O

Article history:
Received 27 May 2016
Available online 22 February 2017
Submitted by V. Pozdnyakov

Keywords:
Least squares estimator
Conditional expectation
Sublinear expectation
Coherent risk measure
g-Expectation

A B S T R A C T

In this paper, we study the least squares estimator for sublinear expectations. Under some mild assumptions, we prove the existence and uniqueness of the least squares estimator. The relationship between the least squares estimator and the conditional coherent risk measure (resp. the conditional g-expectation) is also explored. Then some characterizations of the least squares estimator are given.

1. Introduction

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let ξ be a random variable in $L^2_{\mathcal{F}}(P)$ where

$$L^2_{\mathcal{F}}(P) := \{\xi : \Omega \to \mathbb{R};\ \xi \in L^2(P) \text{ and } \xi \text{ is } \mathcal{F} - \text{measurable}\}.$$

If $\mathcal{C} \subset \mathcal{F}$ is a σ-algebra, then the conditional expectation of ξ, denoted by $E_P[\xi|\mathcal{C}]$, is just the solution of the following problem: find a $\hat{\eta} \in L^2_{\mathcal{C}}(P)$ such that

$$E_P[(\xi - \hat{\eta})^2] = \inf_{\eta \in L^2_{\mathcal{C}}(P)} E_P[(\xi - \eta)^2].$$

* Corresponding author. Fax: +86 0531 88564100.
E-mail addresses: sms_suncf@ujn.edu.cn (C. Sun), jsl@sdu.edu.cn (S. Ji).
[1] This work was supported by the Fund of Doctoral Program Research of University of Jinan (No. 160100119).
[2] This work was supported by National Natural Science Foundation of China (No. 11571203); Supported by the Programme of Introducing Talents of Discipline to Universities of China (No. B12023).

http://dx.doi.org/10.1016/j.jmaa.2017.02.020
0022-247X/

$\hat{\eta}$ is called the least squares estimator of ξ. The property that the conditional expectation coincides with the least squares estimator is the basis for the filtering theory (see for example Bensoussan [3], Davis [4] or Kallianpur [9]). From another viewpoint, the least squares estimator can also be used as an alternative definition of the conditional expectation.

In recent decades, nonlinear expectations have been proposed and developed rapidly. Various definitions of conditional nonlinear expectations are introduced. For example, Peng studied the g-expectation in [10] and defined the conditional g-expectation $\mathcal{E}_g[\xi|\mathcal{F}_t]$ as the solution of a backward stochastic differential equation at time t. The g-expectation has many good properties including time consistency, i.e., $\forall 0 \leq s \leq t \leq T$, $\mathcal{E}_g[\mathcal{E}_g[\xi|\mathcal{F}_t]|\mathcal{F}_s] = \mathcal{E}_g[\xi|\mathcal{F}_s]$. Artzner et al. [1] studied the coherent risk measures which can be seen as one kind of nonlinear expectations (see [6] and [12]). The conditional coherent risk measure in [2] is defined as

$$\bar{\Phi}_t[\xi] := \operatorname*{ess\,sup}_{P\in\mathcal{P}} E_P[\xi|\mathcal{F}_t],$$

where $\xi \in \mathcal{F}_T$, $\mathcal{F}_t$ is a sub-σ-algebra of $\mathcal{F}_T$ and $\mathcal{P}$ is a family of probability measures. They have proved that if $\mathcal{P}$ is 'm-stable', then the conditional coherent risk measure defined above is time consistent.

For the nonlinear expectation case, we do not know whether conditional nonlinear expectations still coincide with the least squares estimators. So it is interesting to explore the relationship between the least squares estimators and the nonlinear conditional expectations. In this paper, we focus on the sublinear expectation case. The least squares estimator under sublinear expectations is investigated.

For a sublinear expectation ρ, since ρ can be represented as a supremum of a family of linear expectations, then our least squares estimate problem can be seen as a minimax problem. The key tool to solve these problems is the minimax theorem. We assume ρ is continuous from above in order that the minimax theorem holds. Then the existence and uniqueness of the least squares estimator are obtained. If we denote $\rho(\xi|\mathcal{C})$ as the least squares estimator, note that for any real number λ, $\rho(\lambda\xi|\mathcal{C})) = \lambda\rho(\xi|\mathcal{C})$. This result induces, generally speaking, both of Artzner et al.'s and Peng's nonlinear conditional expectations fail to be least squares estimators.

The problem comes naturally: when least squares estimators will coincide with Artzner et al.'s and Peng's conditional expectations. To give the answer, we investigate the relationship between the least squares estimators and the nonlinear conditional expectations. A sufficient and necessary condition for the coincidence is given.

In the rest part, through variational methods, we show to get the least squares estimator is equivalent to obtain the solution of an associated nonlinear equation.

The paper is organized as follows. In section 2, we formulate our problem. Under some mild assumptions, we prove the existence and uniqueness of the least squares estimator in section 3. In section 4, we first give the basic properties of the least squares estimator. Then we explore the relationship between the least squares estimator and the conditional coherent risk measure and conditional g-expectation. In the last section, we obtain several characterizations of the least squares estimator.

2. Problem formulation

2.1. Preliminary

For a given measurable space $(\Omega, \mathcal{F})$, we denote $\mathbb{F}$ as the set of bounded $\mathcal{F}$-measurable functions, $\mathbb{N}$ as the set of natural numbers and $\mathbb{R}$ as the set of real numbers.

Definition 2.1. A sublinear expectation ρ is a functional $\rho : \mathbb{F} \mapsto \mathbb{R}$ satisfying

(i) Monotonicity: $\rho(\xi_1) \geq \rho(\xi_2)$ if $\xi_1 \geq \xi_2$;
(ii) Constant preserving: $\rho(c) = c$ for $c \in \mathbb{R}$;

(iii) Sub-additivity: For each $\xi_1, \xi_2 \in \mathbb{F}$, $\rho(\xi_1 + \xi_2) \leq \rho(\xi_1) + \rho(\xi_2)$;
(iv) Positive homogeneity: $\rho(\lambda\xi) = \lambda\rho(\xi)$ for $\lambda \geq 0$.

Note that $\mathbb{F}$ is a Banach space endowed with the supremum norm. Denote the dual space of $\mathbb{F}$ by $\mathbb{F}^*$. Since there is a one-to-one correspondence between $\mathbb{F}^*$ and the class of additive set functions, we denote the element in $\mathbb{F}^*$ by E_P where P is an additive set function. Sometimes we also use P instead of E_P.

Lemma 2.2. *Suppose that the sublinear expectation ρ can be represented by a family of probability measures $\mathcal{P}$, i.e., $\rho(\xi) = \sup_{P\in\mathcal{P}} E_P[\xi]$. For a sequence $\{\xi_n\}_{n\in\mathbb{N}}$, if there exists an $M \in \mathbb{R}$ such that $\xi_n \geq M$ for all n, then we have*

$$\rho(\liminf_n \xi_n) \leq \liminf_n \rho(\xi_n).$$

Proof. Let $\zeta_n = \inf_{k\geq n} \xi_k$. $\zeta_n \leq \xi_n$ and $\{\zeta_n\}_{n\in\mathbb{N}}$ is an increasing sequence. It is easy to see that

$$\rho(\liminf_n \xi_n) = \rho(\lim_n \zeta_n) = \lim_n \rho(\zeta_n) \leq \liminf_n \rho(\xi_n). \quad \square$$

Definition 2.3. We say a sublinear expectation ρ continuous from above on $\mathcal{F}$ if for any sequence $\{\xi_n\}_{n\in\mathbb{N}} \subset \mathbb{F}$ satisfying $\xi_n \downarrow 0$, we have

$$\rho(\xi_n) \downarrow 0.$$

Lemma 2.4. *If a sublinear expectation ρ is continuous from above on $\mathcal{F}$, then for any linear expectation E_P dominated by ρ, P is a probability measure.*

Proof. For any $A_n \downarrow \phi$, we have $\rho(I_{A_n}) \downarrow 0$. If a linear expectation E_P is dominated by ρ, then $P(A_n) \downarrow 0$. It is easy to see that $P(\Omega) = 1$. Thus P is a probability measure. $\square$

Proposition 2.5. *A sublinear expectation ρ is continuous from above on $\mathcal{F}$ if and only if there exists a probability measure P_0 and a family of probability measures $\mathcal{P}$ such that*

i) $\rho(X) = \sup_{P\in\mathcal{P}} E_P[X]$ for all $X \in \mathbb{F}$;
ii) any element in $\mathcal{P}$ is absolutely continuous with respect to P_0;
iii) the set $\{\frac{dP}{dP_0}; P \in \mathcal{P}\}$ is $\sigma(L^1(P_0), L^\infty(P_0))$-compact,

where $\sigma(L^1(P_0), L^\infty(P_0))$ denotes the weak topology defined on $L^1(P_0)$.

Proof. $\Rightarrow$ By Theorem A.1 in Appendix A, ρ can be represented by the family of linear expectations dominated by ρ. We denote $\mathcal{P}$ as all the linear expectations dominated by ρ. Since ρ is continuous from above on $\mathcal{F}$, by Lemma 2.4, each element in $\mathcal{P}$ is a probability measure. By Theorem A.2, $\mathcal{P}$ is $\sigma(\mathbb{F}^*, \mathbb{F})$-compact, where $\sigma(\mathbb{F}^*, \mathbb{F})$ denotes *weak** topology defined on $\mathbb{F}^*$. By Theorem A.3, there exists a $P_0 \in \mathbb{F}^*_c$ such that all the elements in $\mathcal{P}$ are absolutely continuous with respect to P_0, where $\mathbb{F}^*_c$ is the set of linear expectations generated by countably additive measures. Then we can use $L^\infty(P_0)$ to replace $\mathbb{F}$. Note that $L^1(P_0)$ is the space of integrable random variables and $L^\infty(P_0)$ is the space of all equivalence classes of bounded real valued random variables. Since ρ is continuous from above on $\mathcal{F}$ and the dual space of $L^1(P_0)$ is $L^\infty(P_0)$, by Corollary 4.35 in [7], the set $\{\frac{dP}{dP_0}; P \in \mathcal{P}\}$ is $\sigma(L^1(P_0), L^\infty(P_0))$-compact.

$\Leftarrow$ The result is directly deduced by Dini's theorem. $\square$

Remark 2.6. It is worth to point out that the proof of Proposition 2.5 follows directly from Theorem 3.6 in [5].

In the following, we will denote $\mathcal{P}$ as linear expectations dominated by ρ and P_0 the probability measure as in Proposition 2.5.

Definition 2.7. We say a sublinear expectation ρ proper if all the elements in $\mathcal{P}$ are equivalent to P_0.

2.2. Least squares estimator of a random variable

Let $\mathcal{C}$ be a sub σ-algebra of $\mathcal{F}$ and $\mathbb{C}$ be the set of all the bounded $\mathcal{C}$-measurable functions on $(\Omega, \mathcal{F})$. For a given random variable $\xi \in \mathbb{F}$, our problem is to obtain its least squares estimator under the sublinear expectation ρ when we only know "the information" $\mathcal{C}$. In more details, we want to solve the following optimization problem.

Problem 2.8. For a given $\xi \in \mathbb{F}$, find a $\hat{\eta} \in \mathbb{C}$ such that

$$\rho(\xi - \hat{\eta})^2 = \inf_{\eta \in \mathbb{C}} \rho(\xi - \eta)^2. \tag{2.1}$$

The optimal solution $\hat{\eta}$ of (2.1) is called the least square estimator.

If ρ degenerates to a linear expectation, then $\mathcal{P}$ contains only one probability measure P and ρ is the mathematical expectation under P. In this case, it is well known that the least squares estimator $\hat{\eta}$ is just the conditional expectation $E_P[\xi \mid \mathcal{C}]$.

3. Existence and uniqueness

In this section, we study the existence and uniqueness of the least squares estimator.

For any $\xi \in \mathbb{F}$, there exists a positive constant M such that $\sup |\xi| \leq M$ and we will denote by $\mathbb{G}$ all the $\mathcal{C}$-measurable functions bounded by M.

3.1. Existence

Lemma 3.1. *Suppose that $\xi \in \mathbb{F}$. Then we have*

$$\inf_{\eta \in \mathbb{C}} \rho(\xi - \eta)^2 = \inf_{\eta \in \mathbb{G}} \rho(\xi - \eta)^2.$$

Proof. For any $\eta \in \mathbb{C}$, let

$$\bar{\eta} := \eta \mathbf{I}_{\{-M \leq \eta \leq M\}} + M \mathbf{I}_{\{\eta > M\}} - M \mathbf{I}_{\{\eta < -M\}}.$$

Then $\bar{\eta} \in \mathbb{G}$. For any $P \in \mathcal{P}$,

$$\begin{aligned} E_P[(\xi - \eta)^2] - E_P[(\xi - \bar{\eta})^2] &= E_P[(\bar{\eta} - \eta)(2\xi - \eta - \bar{\eta})] \\ &\geq E_P[(M - \eta)(2\xi - 2M)\mathbf{I}_{\{\eta > M\}}] + E_P[(-M - \eta)(2\xi + 2M)\mathbf{I}_{\{\eta < -M\}}]. \end{aligned}$$

Since $-M \leq \xi \leq M$, we have

$$E_P[(M - \eta)(2\xi - 2M)\mathbf{I}_{\{\eta > M\}}] \geq 0 \quad \text{and} \quad E_P[(-M - \eta)(2\xi + 2M)\mathbf{I}_{\{\eta < -M\}}] \geq 0.$$

Thus for any $P \in \mathcal{P}$,

$$E_P[(\xi-\eta)^2] \geq E_P[(\xi-\bar{\eta})^2],$$

which yields

$$\rho(\xi-\eta)^2 \geq \rho(\xi-\bar{\eta})^2$$

and

$$\inf_{\eta\in\mathbb{C}} \rho(\xi-\eta)^2 \geq \inf_{\eta\in\mathbb{G}} \rho(\xi-\eta)^2.$$

On the other hand, since $\mathbb{G} \subset \mathbb{C}$, the inverse inequality is obviously true. □

Theorem 3.2. *If the sublinear expectation ρ is continuous from above on $\mathcal{F}$, then there exists an optimal solution $\hat{\eta} \in \mathbb{G}$ for Problem 2.8.*

Proof. By Lemma 3.1, there exists a sequence $\{\eta_n\}_{n\geq 1} \subset \mathbb{G}$ such that

$$\rho(\xi-\eta_n)^2 < \alpha + \frac{1}{2^n},$$

where $\alpha := \inf_{\eta\in\mathbb{C}} \rho(\xi-\eta)^2$. By Komlós Theorem (see Theorem A.4), there exists a subsequence $\{\eta_{n_i}\}_{i\geq 1}$ and a random variable $\hat{\eta}$ such that

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \eta_{n_i} = \hat{\eta}, \quad P_0 - a.s.$$

Since $\{\eta_n\}_{n\geq 1}$ is bounded by M, then $\hat{\eta} \in \mathbb{G}$. By Lemma 2.2, we have

$$\rho(\xi-\hat{\eta})^2 \leq \liminf_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \rho(\xi-\eta_{n_i})^2 \leq \lim_{k\to\infty}(\alpha + \frac{1}{k}) = \alpha.$$

Thus $\hat{\eta}$ is an optimal solution for Problem 2.8. □

Remark 3.3. If we only assume there exists a probability measure P_0 such that the ρ is a sublinear expectation generated by a family of probability measures which are all absolutely continuous with respect to P_0, Theorem 3.2 still holds.

3.2. Uniqueness

In this subsection, if ρ is continuous from above on $\mathcal{F}$ and proper, we prove the optimal solution of Problem 2.8 is unique under P_0-a.s. sense.

Lemma 3.4. *If the sublinear expectation ρ is continuous from above on $\mathcal{F}$, for a given $\xi \in \mathbb{F}$, we have*

$$\sup_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{G}} E_P[(\xi-\eta)^2] = \max_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{G}} E_P[(\xi-\eta)^2].$$

Proof. By Proposition 2.5, there exists a probability measure P_0 such that $P \ll P_0$ for all $P \in \mathcal{P}$. Let $f_P := \frac{dP}{dP_0}$ and

$$\beta := \sup_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{G}} E_P[(\xi-\eta)^2] = \sup_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{G}} E_{P_0}[f_P(\xi-\eta)^2].$$

Take a sequence $\{f_{P_n}; P_n \in \mathcal{P}\}_{n\geq 1}$ such that

$$\inf_{\eta\in\mathbb{G}} E_{P_0}[f_{P_n}(\xi-\eta)^2] \geq \beta - \frac{1}{2^n}.$$

By Theorem A.4, there exists a subsequence $\{f_{P_{n_i}}\}_{i\geq 1}$ of $\{f_{P_n}\}_{n\geq 1}$ and a random variable $f_{\hat{P}} \in L^1(P_0)$ such that

$$\lim_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} f_{P_{n_i}} = f_{\hat{P}} \quad P_0 - \text{a.s.}$$

Let $g_k := \frac{1}{k}\sum_{i=1}^{k} f_{P_{n_i}}$. $g_k \in \{f_P; P \in \mathcal{P}\}$. By Proposition 2.5, $\{f_P; P \in \mathcal{P}\}$ is $\sigma(L^1(P_0), L^\infty(P_0))$-compact. By Dunford–Pettis theorem, it is uniformly integrable. Thus $\{g_k\}_{k\geq 1}$ is also uniformly integrable and $||g_k - f_{\hat{P}}||_{L^1(P_0)} \to 0$, which shows $\hat{P} \in \mathcal{P}$.

On the other hand, for any $\eta \in \mathbb{G}$ and $k \in \mathbb{N}$, we have

$$E_{P_0}[g_k(\xi-\eta)^2] \geq \inf_{\tilde{\eta}\in\mathbb{G}} E_{P_0}[g_k(\xi-\tilde{\eta})^2].$$

Then for any $\eta \in \mathbb{G}$, we have

$$\lim_{k\to\infty} E_{P_0}[g_k(\xi-\eta)^2] \geq \limsup_{k\to\infty} \inf_{\tilde{\eta}\in\mathbb{G}} E_{P_0}[g_k(\xi-\tilde{\eta})^2].$$

Thus

$$\inf_{\eta\in\mathbb{G}} \lim_{k\to\infty} E_{P_0}[g_k(\xi-\eta)^2] \geq \limsup_{k\to\infty} \inf_{\tilde{\eta}\in\mathbb{G}} E_{P_0}[g_k(\xi-\tilde{\eta})^2].$$

Since $\{g_k\}_{k\geq 1}$ is uniformly integral and $||(\xi-\eta)^2||_{L^\infty} \leq 4M^2$, we have

$$\inf_{\eta\in\mathbb{G}} E_{P_0}[f_{\hat{P}}(\xi-\eta)^2] = \inf_{\eta\in\mathbb{G}} E_{P_0}[\lim_{k\to\infty} g_k(\xi-\eta)^2] = \inf_{\eta\in\mathbb{G}} \lim_{k\to\infty} E_{P_0}[g_k(\xi-\eta)^2].$$

Thus

$$\inf_{\eta\in\mathbb{G}} E_{P_0}[f_{\hat{P}}(\xi-\eta)^2] \geq \limsup_{k\to\infty} \frac{1}{k}\sum_{i=1}^{k} \inf_{\eta\in\mathbb{G}} E_{P_0}[f_{P_{n_i}}(\xi-\eta)^2] \geq \beta.$$

Since $\hat{P} \in \mathcal{P}$, we have

$$\inf_{\eta\in\mathbb{G}} E_{\hat{P}}[(\xi-\eta)^2] = \sup_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{G}} E_P[(\xi-\eta)^2]. \quad \square$$

Corollary 3.5. *If the sublinear expectation ρ is continuous from above on $\mathcal{F}$, then for a given $\xi \in \mathbb{F}$, we have*

$$\sup_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2] = \max_{P\in\mathcal{P}} \inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2].$$

Proof. Choose $\hat{P}$ as in Lemma 3.4. By Lemma 3.1 and Lemma 3.4, we have

$$\sup_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2] \le \sup_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{G}} E_P[(\xi-\eta)^2] = \inf_{\eta\in\mathbb{G}} E_{\hat{P}}[(\xi-\eta)^2] = \inf_{\eta\in\mathbb{C}} E_{\hat{P}}[(\xi-\eta)^2].$$

On the other hand, the inverse inequality is obvious. Then

$$\inf_{\eta\in\mathbb{C}} E_{\hat{P}}[(\xi-\eta)^2] = \sup_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2].$$

Since $\hat{P}\in\mathcal{P}$, we have

$$\sup_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2] = \max_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2]. \quad \square$$

Theorem 3.6. *If the sublinear expectation ρ is continuous from above on $\mathcal{F}$ and proper, then for any given $\xi\in\mathbb{F}$, there exists a unique optimal solution of Problem 2.8.*

Proof. The existence result is proved in Theorem 3.2. Now we prove the uniqueness.

Since $\mathcal{P}$ is $\sigma(L^1(P_0), L^\infty(P_0))$-compact, by Theorem A.5,

$$\inf_{\eta\in\mathbb{C}}\sup_{P\in\mathcal{P}} E_P[(\xi-\eta)^2] = \sup_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2].$$

Since the optimal solution exists, by Corollary 3.5,

$$\min_{\eta\in\mathbb{C}}\sup_{P\in\mathcal{P}} E_P[(\xi-\eta)^2] = \max_{P\in\mathcal{P}}\inf_{\eta\in\mathbb{C}} E_P[(\xi-\eta)^2].$$

Let $\hat{\eta}$ be an optimal solution and $\hat{P}$ as in Corollary 3.5. By Theorem A.7, $(\hat{\eta},\hat{P})$ is an saddle point, i.e.,

$$E_P[(\xi-\hat{\eta})^2] \le E_{\hat{P}}[(\xi-\hat{\eta})^2] \le E_{\hat{P}}[(\xi-\eta)^2], \quad \forall P\in\mathcal{P}, \eta\in\mathbb{C}.$$

This result shows that if $\hat{\eta}$ is an optimal solution, then there exists a $\hat{P}\in\mathcal{P}$ such that $\hat{\eta} = E_{\hat{P}}[\xi|\mathcal{C}]$.

Suppose that there exist two optimal solutions $\hat{\eta}_1$ and $\hat{\eta}_2$. Their accompanying probabilities are denoted by $\hat{P}_1$ and $\hat{P}_2$ respectively. Thus $\hat{\eta}_1 = E_{\hat{P}_1}[\xi|\mathcal{C}]$ and $\hat{\eta}_2 = E_{\hat{P}_2}[\xi|\mathcal{C}]$. Let $P^\lambda := \lambda\hat{P}_1 + (1-\lambda)\hat{P}_2$, where $\lambda\in(0,1)$. Denote $\lambda E_{P^\lambda}[\frac{d\hat{P}_1}{dP^\lambda}|\mathcal{C}]$ by $\lambda_{\hat{P}_1}$ and $(1-\lambda)E_{P^\lambda}[\frac{d\hat{P}_2}{dP^\lambda}|\mathcal{C}]$ by $\lambda_{\hat{P}_2}$. It is easy to see that $\lambda_{\hat{P}_1} + \lambda_{\hat{P}_2} = 1$. Since

$$\hat{\eta}_1 = E_{\hat{P}_1}[\xi|\mathcal{C}] = \frac{E_{P^\lambda}[\xi\frac{d\hat{P}_1}{dP^\lambda}|\mathcal{C}]}{E_{P^\lambda}[\frac{d\hat{P}_1}{dP^\lambda}|\mathcal{C}]}$$

and

$$\hat{\eta}_2 = E_{\hat{P}_2}[\xi|\mathcal{C}] = \frac{E_{P^\lambda}[\xi\frac{d\hat{P}_2}{dP^\lambda}|\mathcal{C}]}{E_{P^\lambda}[\frac{d\hat{P}_2}{dP^\lambda}|\mathcal{C}]},$$

then $\lambda_{P_1}\hat{\eta}_1 + \lambda_{P_2}\hat{\eta}_2 = E_{P^\lambda}[\xi|\mathcal{C}]$. And

$$
\begin{aligned}
E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] &= E_{P^\lambda}[(\xi - \lambda_{\hat{P}_1}\hat{\eta}_1 - \lambda_{\hat{P}_2}\hat{\eta}_2)^2] \\
&= E_{P^\lambda}[(\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1) + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2))^2] \\
&= E_{P^\lambda}[\lambda^2_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 + \lambda^2_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 + 2\lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\xi - \hat{\eta}_1)(\xi - \hat{\eta}_2)] \\
&= E_{P^\lambda}[\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
&= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_2}(\xi - \hat{\eta}_1)^2 + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
&\quad + (1-\lambda)E_{\hat{P}_2}[\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_1}(\xi - \hat{\eta}_2)^2 + (\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
&= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] + \lambda E_{\hat{P}_1}[\lambda^2_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
&\quad + (1-\lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_2)^2] + (1-\lambda)E_{\hat{P}_2}[\lambda^2_{\hat{P}_1}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
&\geq \alpha,
\end{aligned}
$$

where $\alpha := \inf_{\eta\in\mathbb{C}} \rho(\xi - \eta)^2$. It yields $E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] = \alpha$ if and only if $\hat{\eta}_1 = \hat{\eta}_2$ P_0-a.s.

On the other hand, since $(\hat{\eta}_1, \hat{P}_1)$ is a saddle point,

$$
E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] \leq E_{P^\lambda}[(\xi - \hat{\eta}_1)^2] \leq E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] = \alpha.
$$

Then $E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] = \alpha$. Thus, $\hat{\eta}_1 = \hat{\eta}_2$ P_0-a.s. □

4. Properties of the least squares estimator

In this section, we will first give some basic properties of the least square estimator. Then we explore the relationship between the least square estimator and the conditional coherent risk measure and conditional g-expectation.

For a given $\xi \in \mathbb{F}$, we will use $\rho(\xi|\mathcal{C})$ to denote the least squares estimator with respect to $\mathcal{C}$ instead of $\hat{\eta}$. The following properties hold.

Proposition 4.1. *If the sublinear expectation ρ is continuous from above on $\mathcal{F}$ and proper, then for any $\xi \in \mathbb{F}$, we have:*

i) If $C_1 \leq \xi(\omega) \leq C_2$ for two constants C_1 and C_2, then $C_1 \leq \rho(\xi|\mathcal{C}) \leq C_2$.
ii) $\rho(\lambda\xi|\mathcal{C}) = \lambda\rho(\xi|\mathcal{C})$ for $\lambda \in \mathbb{R}$.
iii) For each $\eta_0 \in \mathbb{C}$, $\rho(\xi + \eta_0|\mathcal{C}) = \rho(\xi|\mathcal{C}) + \eta_0$.
iv) If under each $P \in \mathcal{P}$, ξ is independent of the sub σ-algebra $\mathcal{C}$, then $\rho(\xi|\mathcal{C})$ is a constant.

Proof. i) If $C_1 \leq \xi(\omega) \leq C_2$, then $C_1 \leq E_P[\xi|\mathcal{C}] \leq C_2$ for any $P \in \mathcal{P}$. Since $\rho(\xi|\mathcal{C}) \in \{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}$ (see the proof of Theorem 3.6), thus $\rho(\xi|\mathcal{C})$ lies in $[C_1, C_2]$.

ii) When $\lambda = 0$, the statement is obvious. When $\lambda \neq 0$, we have

$$
\lambda^2 \rho(\xi - \frac{\rho(\lambda\xi|\mathcal{C})}{\lambda})^2 = \rho(\lambda\xi - \rho(\lambda\xi|\mathcal{C}))^2 = \inf_{\eta\in\mathbb{C}} \rho(\lambda\xi - \eta)^2 = \lambda^2 \inf_{\eta\in\mathbb{C}} \rho(\xi - \eta)^2,
$$

which yields

$$
\rho(\xi - \frac{\rho(\lambda\xi|\mathcal{C})}{\lambda})^2 = \inf_{\eta\in\mathbb{C}} \rho(\xi - \eta)^2.
$$

Thus $\frac{\rho(\lambda\xi|\mathcal{C})}{\lambda} = \rho(\xi|\mathcal{C})$ due to the uniqueness.

iii) Note that

$$\rho(\xi+\eta_0-(\eta_0+\rho(\xi|\mathcal{C})))^2=\rho(\xi-\rho(\xi|\mathcal{C}))^2=\inf_{\eta\in\mathbb{C}}\rho(\xi-\eta)^2=\inf_{\eta\in\mathbb{C}}\rho(\xi+\eta_0-\eta)^2.$$

By the uniqueness of the least squares estimator, we have $\rho(\xi+\eta_0|\mathcal{C})=\eta_0+\rho(\xi|\mathcal{C})$.

iv) If under each $P\in\mathcal{P}$, ξ is independent of the sub σ-algebra $\mathcal{C}$, then $E_P[\xi|\mathcal{C}]$ is a constant for each $P\in\mathcal{P}$. Since $\rho(\xi|\mathcal{C})\in\{E_P[\xi|\mathcal{C}];P\in\mathcal{P}\}$, $\rho(\xi|\mathcal{C})$ is also a constant. $\square$

The coherent risk measure was introduced by Artzner et al. [1] and the g-expectation was introduced by Peng [10]. The conditional coherent risk measure and some special conditional g-expectations can be defined by $ess\sup_{P\in\mathcal{P}}E_P[\xi|\mathcal{C}]$. In next two examples, we will show the least squares estimator is different from the conditional coherent risk measure and the conditional g-expectation.

Example 4.2. Let $\Omega=\{\omega_1,\omega_2\}$, $\mathcal{F}=\{\phi,\{\omega_1\},\{\omega_2\},\Omega\}$ and $\mathcal{C}=\{\phi,\Omega\}$. Let $P_1=\frac{1}{4}\mathbf{I}_{\{\omega_1\}}+\frac{3}{4}\mathbf{I}_{\{\omega_2\}}$, $P_2=\frac{3}{4}\mathbf{I}_{\{\omega_1\}}+\frac{1}{4}\mathbf{I}_{\{\omega_2\}}$ and $\mathcal{P}=\{\lambda P_1+(1-\lambda)P_2;\lambda\in[0,1]\}$. For each $\xi\in\mathbb{F}$, define

$$\rho(\xi)=\sup_{P\in\mathcal{P}}E_P[\xi].$$

Take $\xi=2\mathbf{I}_{\{\omega_1\}}+8\mathbf{I}_{\{\omega_2\}}$. It is easy to check

$$\sup_{P\in\mathcal{P}}E_P[\xi]=6+\frac{1}{2}\quad\text{and}\quad\rho(\xi|\mathcal{C})=E_{\hat{P}}[\xi|\mathcal{C}]=5,$$

where $\hat{P}=\frac{1}{2}\mathbf{I}_{\{\omega_1\}}+\frac{1}{2}\mathbf{I}_{\{\omega_2\}}$.

Example 4.3. Let $W(\cdot)$ be a standard 1-dimensional Brownian motion defined on a complete probability space $(\Omega,\mathcal{F},P_0)$. The information structure is given by a filtration $\{\mathcal{F}_t\}_{0\le t\le T}$, which is generated by $W(\cdot)$ and augmented by all P_0-null sets. $M^2(0,T;\mathbb{R})$ denotes the space of all $\mathcal{F}_t$-progressively measurable processes y_t such that $E_P\int_0^T|y_t|^2dt<\infty$. Let us consider the g-expectation defined by the following BSDE:

$$y_t=\xi+\int_t^T|z_s|ds-\int_t^T z_s dW(s),\tag{4.1}$$

where ξ is a bounded $\mathcal{F}_T$-measurable function. Since there exists a unique adapted pair (y,z) solves (4.1), the solution y_t is defined as the g-expectation with respect to $\mathcal{F}_t$ and denote it by $\mathcal{E}_{|z|}(\xi|\mathcal{F}_t)$.

Consider the following linear case:

$$y_t=\xi+\int_t^T\mu_s z_s ds-\int_t^T z_s dW(s),\tag{4.2}$$

where $|\mu_s|\le 1$ P_0-a.s. By Girsanov transform, there exists a probability P^μ such that the solution $\{y_t\}_{0\le t\le T}$ of (4.2) is a martingale under P^μ. Let $\mathcal{P}:=\{P^\mu\mid|\mu_s|\le 1\ P_0-\text{a.s.}\}$. By Theorem 2.1 in [8],

$$\mathcal{E}_{|z|}(\xi)=\sup_{P^\mu\in\mathcal{P}}E_{P^\mu}[\xi],\quad\forall\xi\in\mathcal{F}_T$$

and

$$\mathcal{E}_{|z|}(\xi|\mathcal{F}_t)=\operatorname*{ess\,sup}_{P^\mu\in\mathcal{P}}E_{P^\mu}[\xi|\mathcal{F}_t].$$

Since $\mathcal{E}_{|z|}(\cdot)$ is a sublinear expectation, if we denote the corresponding least squares estimator by $\rho_{|z|}(\xi|\mathcal{F}_t)$, then $\rho_{|z|}(\xi|\mathcal{F}_t)$ does not coincide with $\mathcal{E}_{|z|}(\xi|\mathcal{F}_t)$ for all bounded $\xi \in \mathcal{F}_T$. Otherwise, if for all bounded $\xi \in \mathcal{F}_T$, $\rho_{|z|}(\xi|\mathcal{F}_t) = \mathcal{E}_{|z|}(\xi|\mathcal{F}_t)$, then by the property ii) in Corollary 4.1, we have

$$\operatorname*{ess\,sup}_{P^\mu \in \mathcal{P}} E_{P^\mu}[\xi|\mathcal{F}_t] = \rho_{|z|}(\xi|\mathcal{F}_t) = -\rho_{|z|}(-\xi|\mathcal{F}_t) = \operatorname*{ess\,inf}_{P^\mu \in \mathcal{P}} E_{P^\mu}[\xi|\mathcal{F}_t].$$

Since the set $\mathcal{P}$ contains more than one probability measure, the above equation can not be true for all bounded $\xi \in \mathcal{F}_T$.

In the following, for simplicity, we denote $ess \sup\limits_{P \in \mathcal{P}} E_P[\xi|\mathcal{C}]$ by η_{ess}. We first prove that η_{ess} is the optimal solution of a constrained least squares optimization problem.

Definition 4.4. A sublinear expectation ρ is called strictly comparable if for any $\xi_1, \xi_2 \in \mathbb{F}$ satisfying $\xi_1 > \xi_2$ P_0-a.s., $\rho(\xi_1) > \rho(\xi_2)$.

The notion of the m-stable property was proposed in [1] (see Definition B.2). One result about this property is: If $\mathcal{P}$ is m-stable, then the sublinear expectation ρ is time consistent (see Proposition B.3).

Proposition 4.5. *Suppose that a sublinear expectation ρ is continuous from above, strictly comparable and the representation set $\mathcal{P}$ is m-stable. Then for $\xi \in \mathbb{F}$, η_{ess} is the unique solution of the following optimal problem:*

$$\inf_{\eta \in \mathbb{C}} \sup_{\tilde{\eta} \in \mathbb{C}^+} \rho[(\xi - \eta)^2 + \tilde{\eta}(\xi - \eta)], \tag{4.3}$$

where $\mathbb{C}^+$ denotes all nonnegative elements in $\mathbb{C}$.

Proof. We first show if the optimal solution of (4.3) exists, the optimal solution $\hat{\eta}$ of (4.3) lies in $\mathbb{B}$, where $\mathbb{B} := \{\eta \in \mathbb{C}; \eta \geq \eta_{ess} P_0 - a.s.\}$.

Since

$$\begin{aligned}
&\inf\nolimits_{\eta \in \mathbb{C}} \sup\nolimits_{\tilde{\eta} \in \mathbb{C}^+} \rho[(\xi - \eta)^2 + \tilde{\eta}(\xi - \eta)] \\
&\leq \sup\nolimits_{\tilde{\eta} \in \mathbb{C}^+} \rho[(\xi - \eta_{ess})^2 + \tilde{\eta}(\xi - \eta_{ess})] \\
&\leq \rho[(\xi - \eta_{ess})^2] + \sup\nolimits_{\tilde{\eta} \in \mathbb{C}^+} \rho[\tilde{\eta}(\xi - \eta_{ess})] \\
&= \rho[(\xi - \eta_{ess})^2] + \sup\nolimits_{\tilde{\eta} \in \mathbb{C}^+} \rho[ess \sup\nolimits_{P \in \mathcal{P}} E_P[\tilde{\eta}(\xi - \eta_{ess})]] \\
&= \rho[(\xi - \eta_{ess})^2] < \infty,
\end{aligned}$$

the value of our problem is finite. Then

$$\sup_{\tilde{\eta} \in \mathbb{C}^+} \rho[\tilde{\eta}(\xi - \hat{\eta})] \leq \sup_{\tilde{\eta} \in \mathbb{C}^+} \rho[(\xi - \hat{\eta})^2 + \tilde{\eta}(\xi - \hat{\eta})] < \infty.$$

On the other hand, since the representation set $\mathcal{P}$ is 'stable', we have

$$\sup_{\tilde{\eta} \in \mathbb{C}^+} \rho[\tilde{\eta}(\xi - \hat{\eta})] = \sup_{\tilde{\eta} \in \mathbb{C}^+} \rho[\tilde{\eta}(\eta_{ess} - \hat{\eta})].$$

If $A := \{\omega; \hat{\eta} < \eta_{ess}\}$ is not a P_0-null set, as ρ is strictly comparable, we can choose $\tilde{\eta}$ to let $\rho[\tilde{\eta}(\xi - \hat{\eta})]$ larger than any real number. Thus $P_0(A) = 0$ and $\hat{\eta} \geq \eta_{ess}$ P_0-a.s.

For any $\eta \in \mathbb{B}$ and $\tilde{\eta} \in \mathbb{C}^+$, we have

$$\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)]-\rho[(\xi-\eta)^2] \leq \rho[\tilde{\eta}(\xi-\eta)] \leq 0.$$

Thus for any $\eta \in \mathbb{B}$,

$$\sup_{\tilde{\eta}\in\mathbb{C}^+} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \leq \rho[(\xi-\eta)^2].$$

On the other hand,

$$\sup_{\tilde{\eta}\in\mathbb{C}^+} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \geq \rho[(\xi-\eta)^2+0(\xi-\eta)] = \rho[(\xi-\eta)^2].$$

Thus for any $\eta \in \mathbb{B}$,

$$\rho[(\xi-\eta)^2] = \sup_{\tilde{\eta}\in\mathbb{C}^+} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)]$$

and

$$\inf_{\eta\in\mathbb{B}} \sup_{\tilde{\eta}\in\mathbb{C}^+} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \inf_{\eta\in\mathbb{B}} \rho[(\xi-\eta)^2].$$

For any $\eta \in \mathbb{B}$, since

$$\begin{aligned}\rho[(\xi-\eta)^2] &= \rho[(\xi-\eta_{ess})^2+(\eta_{ess}-\eta)^2-2(\eta-\eta_{ess})(\xi-\eta_{ess})]\\ &\geq \rho[(\xi-\eta_{ess})^2+(\eta_{ess}-\eta)^2]-2\rho[(\eta-\eta_{ess})(\xi-\eta_{ess})] \geq \rho[(\xi-\eta_{ess})^2],\end{aligned}$$

then η_{ess} is the least squares estimator among $\mathbb{B}$.

On the other hand, for any $\eta \notin B$, $\sup_{\tilde{\eta}\in\mathbb{C}^+} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)]$ is equal to ∞. Thus η_{ess} is also the optimal solution of (4.3). The uniqueness comes from ρ is strictly comparable. □

Proposition 4.5 is just equivalent to say η_{ess} is the unique solution of the following problem:

$$\begin{gathered}\inf_{\eta\in\mathbb{C}} \rho[(\xi-\eta)^2]\\ \text{subject to} \quad \rho[(\eta_{ess}-\eta)^+]=0.\end{gathered}$$

A necessary and sufficient condition for η_{ess} being the least squares estimator is obtained.

Theorem 4.6. *Under assumptions in Proposition 4.5, for a given $\xi \in \mathbb{F}$, η_{ess} is the optimal solution of Problem 2.8 if and only if*

$$\inf_{\eta\in\mathbb{C}} \sup_{\tilde{\eta}\in\mathbb{C}^+} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \sup_{\tilde{\eta}\in\mathbb{C}^+} \inf_{\eta\in\mathbb{C}} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)].$$

Proof. If η_{ess} is the optimal solution of Problem 2.8, then

$$\sup_{\tilde{\eta}\in\mathbb{C}^+} \inf_{\eta\in\mathbb{C}} \rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \geq \inf_{\eta\in\mathbb{C}} \rho[(\xi-\eta)^2] = \rho[(\xi-\eta_{ess})^2].$$

On the other hand, by Proposition 4.5,

$$\begin{aligned}\inf_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] &= \sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta_{ess})^2+\tilde{\eta}(\xi-\eta_{ess})]\\ &\leq \rho[(\xi-\eta_{ess})^2]+\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[\tilde{\eta}(\xi-\eta_{ess})]\\ &= \rho[(\xi-\eta_{ess})^2].\end{aligned}$$

Then

$$\inf_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \leq \sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)].$$

Since

$$\inf_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \geq \sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)]$$

is obvious, thus

$$\inf_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)].$$

Conversely, for any $\tilde{\eta}\in\mathbb{C}^+$,

$$\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \inf_{\eta\in\mathbb{C}}\rho[(\xi+\frac{\tilde{\eta}}{2}-\eta)^2-\frac{\tilde{\eta}^2}{4}] \leq \inf_{\eta\in\mathbb{C}}\rho[(\xi+\frac{\tilde{\eta}}{2}-\eta)^2] = \inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2],$$

which implies

$$\sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \leq \inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2].$$

Since

$$\sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] \geq \inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+0(\xi-\eta)] = \inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2]$$

is obvious, thus

$$\sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2].$$

This shows $\sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)]$ attains its supremum when $\tilde{\eta}=0$. By Proposition 4.5, we also know $\inf_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)]$ attains its infimum when $\eta=\eta_{ess}$. Since

$$\inf_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \sup_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)],$$

then

$$\min_{\eta\in\mathbb{C}}\sup_{\tilde{\eta}\in\mathbb{C}^+}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)] = \max_{\tilde{\eta}\in\mathbb{C}^+}\inf_{\eta\in\mathbb{C}}\rho[(\xi-\eta)^2+\tilde{\eta}(\xi-\eta)].$$

By Theorem A.7, $(\eta_{ess},0)$ is the saddle point, i.e., for any $\eta\in\mathbb{C}$ and $\tilde{\eta}\in\mathbb{C}^+$, we have

$$\rho[(\xi-\eta_{ess})^2+\tilde{\eta}(\xi-\eta_{ess})] \leq \rho[(\xi-\eta_{ess})^2] \leq \rho[(\xi-\eta)^2].$$

The second inequality means η_{ess} is the optimal solution of Problem 2.8. $\square$

5. Characterizations of the least squares estimator

In this section, we obtain several characterizations of the least square estimator.

5.1. The orthogonal projection

If $\mathcal{P}$ contains only one probability P, by probability theory, the least squares estimator $\hat{\eta}$ is just the conditional expectation $E_P[\xi|\mathcal{C}]$. It is well known that a conditional expectation is an orthogonal projection. In more details, for any $\eta \in \mathbb{C}$,

$$E_P[(\xi - \hat{\eta})\eta] = E_P[(\xi - E_P[\xi|\mathcal{C}])\eta] = 0.$$

Does the above property still hold when ρ is a sublinear expectation? Note that for any $\eta \in \mathbb{C}$,

$$\begin{aligned} & \rho[(\xi - \hat{\eta})\eta] \\ = & \sup_{P\in\mathcal{P}} E_P[(\xi - \hat{\eta})\eta] \\ = & \sup_{P\in\mathcal{P}} E_P[(\xi - E_{\hat{P}}[\xi|\mathcal{C}])\eta] \\ \geq & \ 0, \end{aligned}$$

where $\hat{P}$ is the accompanying probability of $\hat{\eta}$ as in the proof of Theorem 3.6. But we notice that $\inf\limits_{\eta\in\mathbb{C}} \rho[(\xi - \hat{\eta})\eta] = 0$. This motivates us to introduce the following definition. For any given $\xi \in \mathbb{F}$, define f: $\mathbb{C} \mapsto \mathbb{R}$ by

$$f(\tilde{\eta}) = \inf_{\eta\in\mathbb{C}} \rho[(\xi - \tilde{\eta})\eta].$$

Denote the kernel of f by

$$\ker(f) := \{\tilde{\eta} \in \mathbb{C} \mid \ f(\tilde{\eta}) = 0\}.$$

We have proved that the least squares estimator is one element of the set $\{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}$. In the following, we show that this set can be described by the kernel of f.

Lemma 5.1. *If ρ is a sublinear expectation continuous from above on $\mathcal{F}$, for any given $\xi \in \mathbb{F}$,*

$$\ker(f) = \{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}.$$

Proof. For any $P \in \mathcal{P}$ and $\eta \in \mathbb{C}$, considering $E_P[\xi|\mathcal{C}]$, we have

$$\rho[(\xi - E_P[\xi|\mathcal{C}])\eta] \geq E_P[(\xi - E_P[\xi|\mathcal{C}])\eta] = 0.$$

Then

$$\inf_{\eta\in\mathbb{C}} \rho[(\xi - E_P[\xi|\mathcal{C}])\eta] \geq 0.$$

It is obvious that

$$\inf_{\eta\in\mathbb{C}} \rho[(\xi - E_P[\xi|\mathcal{C}])\eta] \leq \rho[(\xi - E_P[\xi|\mathcal{C}])0] = 0,$$

which leads to $\inf_{\eta\in\mathbb{C}} \rho[(\xi - E_P[\xi|\mathcal{C}])\eta] = 0$ for any $P \in \mathcal{P}$. Thus $\{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\} \subset \ker(f)$.

On the other hand, for any $\tilde{\eta} \in \ker(f)$, since $\mathbb{C}$ is a convex set and ρ is a sublinear expectation continuous from above, by Theorem A.6, there exists a probability $\tilde{P} \in \mathcal{P}$ such that

$$\inf_{\eta \in \mathbb{C}} E_{\tilde{P}}[(\xi - \tilde{\eta})\eta] = \inf_{\eta \in \mathbb{C}} \rho[(\xi - \tilde{\eta})\eta] = 0.$$

If $\tilde{\eta} \neq E_{\tilde{P}}[\xi|\mathcal{C}]$, then it is easy to find a $\eta' \in \mathbb{C}$ such that $E_{\tilde{P}}[(\xi - \tilde{\eta})\eta'] < 0$. Thus $\tilde{\eta} = E_{\tilde{P}}[\xi|\mathcal{C}]$, which deduces $\ker(f) \subset \{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}$. $\square$

If $\mathcal{P}$ satisfies the m-stable property, we obtain the following result.

Theorem 5.2. *If ρ is a sublinear expectation continuous from above on $\mathcal{F}$ and the corresponding $\mathcal{P}$ is m-stable, then for a given $\xi \in \mathbb{F}$, $\ker(f)$ is just the set*

$$\mathbb{B} := \{\tilde{\eta} \in \mathbb{C} \mid \operatorname*{ess\,inf}_{P \in \mathcal{P}} E_P[\xi|\mathcal{C}] \leq \tilde{\eta} \leq \operatorname*{ess\,sup}_{P \in \mathcal{P}} E_P[\xi|\mathcal{C}]\}.$$

Proof. By Lemma 5.1, $\ker(f)$ is a subset of $\mathbb{B}$. So we only need to prove $\mathbb{B} \subset \ker(f)$.

Since $\mathcal{P}$ is ‘m-stable’, for any $\eta \in \mathbb{C}$ and $\tilde{\eta} \in \mathbb{B}$, we have

$$\begin{aligned} \rho[(\xi - \tilde{\eta})\eta] &= \rho[\operatorname*{ess\,sup}_{P \in \mathcal{P}} E_P[(\xi - \tilde{\eta})\eta|\mathcal{C}]] \\ &= \rho[\eta^+(\operatorname*{ess\,sup}_{P \in \mathcal{P}} E_P[\xi|\mathcal{C}] - \tilde{\eta}) - \eta^-(\operatorname*{ess\,inf}_{P \in \mathcal{P}} E_P[\xi|\mathcal{C}] - \tilde{\eta})], \end{aligned}$$

where $\eta^+ := \eta \bigvee 0$ and $\eta^- := -(\eta \bigwedge 0)$. It yields that for any $\tilde{\eta} \in \mathbb{B}$ and $\eta \in \mathbb{C}$,

$$\rho[(\xi - \tilde{\eta})\eta] \geq 0.$$

Thus for any $\tilde{\eta} \in \mathbb{B}$,

$$\inf_{\eta \in \mathbb{C}} \rho[(\xi - \tilde{\eta})\eta] \geq 0.$$

Since

$$\inf_{\eta \in \mathbb{C}} \rho[(\xi - \tilde{\eta})\eta] \leq \rho[(\xi - \tilde{\eta})0] = 0,$$

then $\inf_{\eta \in \mathbb{C}} \rho[(\xi - \tilde{\eta})\eta] = 0$ for any $\tilde{\eta} \in \mathbb{B}$, which implies that $\mathbb{B} \subset \ker(f)$. $\square$

5.2. A sufficient and necessary condition

We give a sufficient and necessary condition for the existence of the least squares estimator in this subsection. Especially, here we do not need to assume ρ is continuous from above on $\mathcal{F}$.

Lemma 5.3. *For a given $\xi \in \mathbb{F}$, if $\hat{\eta}$ is an optimal solution of Problem 2.8, then*

$$\rho[(\xi - \eta)(\xi - \hat{\eta})] \geq \rho(\xi - \hat{\eta})^2, \quad \forall \eta \in \mathbb{C}.$$

Proof. For any $\eta \in \mathbb{C}$, define f: $[0, 1] \mapsto \mathbb{R}$ by

$$f(\lambda) = \lambda^2 \rho(\xi - \eta)^2 + (1 - \lambda)^2 \rho(\xi - \hat{\eta})^2 + 2\lambda(1 - \lambda)\rho[(\xi - \eta)(\xi - \hat{\eta})].$$

It is easy to check that

$$f(\lambda) \geq \rho[\xi - (\lambda\eta + (1-\lambda)\hat{\eta})]^2 \geq \rho(\xi - \hat{\eta})^2,$$

which implies that $f(\lambda)$ attains the minimum on $[0,1]$ when $\lambda = 0$. Then for any $\eta \in \mathbb{C}$,

$$f'(\lambda)|_{0+} = -2\rho(\xi - \hat{\eta})^2 + 2\rho[(\xi - \eta)(\xi - \hat{\eta})] \geq 0.$$

Thus

$$\rho[(\xi - \eta)(\xi - \hat{\eta})] \geq \rho(\xi - \hat{\eta})^2, \quad \forall \eta \in \mathbb{C}. \quad \square$$

Lemma 5.4. *For any $\xi_1, \xi_2 \in \mathbb{F}$ such that $\rho(|\xi_1|^p) > 0$ and $\rho(|\xi_2|^q) > 0$, we have*

$$\rho(|\xi_1\xi_2|) \leq (\rho(|\xi_1|^p))^{\frac{1}{p}}(\rho(|\xi_2|^q))^{\frac{1}{q}},$$

where $1 < p$, $q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$.

Proof. Let

$$X = \frac{\xi_1}{(\rho(|\xi_1|^p))^{\frac{1}{p}}}, \quad Y = \frac{\xi_2}{(\rho(|\xi_2|^q))^{\frac{1}{q}}}.$$

Since $|XY| \leq \frac{|X|^p}{p} + \frac{|Y|^q}{q}$, then

$$\rho(|XY|) \leq \rho(\frac{|X|^p}{p} + \frac{|Y|^q}{q}) \leq \rho(\frac{|X|^p}{p}) + \rho(\frac{|Y|^q}{q}) = 1.$$

Thus

$$\rho(|\xi_1\xi_2|) \leq (\rho(|\xi_1|^p))^{\frac{1}{p}}(\rho(|\xi_2|^q))^{\frac{1}{q}}. \quad \square$$

Remark 5.5. If ρ can be represented by a family of probability measures, then the condition $\rho(|\xi_1|^p) > 0$ and $\rho(|\xi_2|^q) > 0$ can be abandoned.

Theorem 5.6. *Suppose $\inf_{\eta \in \mathbb{C}} \rho(\xi - \eta)^2 > 0$. For a given $\xi \in \mathbb{F}$, $\hat{\eta}$ is the optimal solution of Problem 2.8 if and only if it is the bounded $\mathcal{C}$-measurable solution of the following equation*

$$\inf_{\eta \in \mathbb{C}} \rho[(\xi - \hat{\eta})(\xi - \eta)] = \rho(\xi - \hat{\eta})^2. \tag{5.1}$$

Proof. $\Rightarrow$ Since $\hat{\eta}$ is the optimal solution of Problem 2.8, by Lemma 5.3,

$$\inf_{\eta \in \mathbb{C}} \rho[(\xi - \hat{\eta})(\xi - \eta)] \geq \rho(\xi - \hat{\eta})^2.$$

It is obvious that

$$\inf_{\eta \in \mathbb{C}} \rho[(\xi - \hat{\eta})(\xi - \eta)] \leq \rho(\xi - \hat{\eta})^2.$$

Then $\hat{\eta}$ is the solution of (5.1).

$\Leftarrow$ If $\hat{\eta} \in \mathbb{C}$ satisfying equation (5.1), by Lemma 5.4, we have

$$\begin{aligned}\rho(\xi-\hat{\eta})^2 &= \inf_{\eta\in\mathbb{C}}\rho[(\xi-\hat{\eta})(\xi-\eta)] \\ &\le \inf_{\eta\in\mathbb{C}}(\rho(\xi-\hat{\eta})^2)^{\frac{1}{2}}(\rho(\xi-\eta)^2)^{\frac{1}{2}} \\ &= (\rho(\xi-\hat{\eta})^2)^{\frac{1}{2}}[\inf_{\eta\in\mathbb{C}}\rho(\xi-\eta)^2]^{\frac{1}{2}}.\end{aligned}$$

Thus $\rho(\xi-\hat{\eta})^2 \le \inf_{\eta\in\mathbb{C}}\rho(\xi-\eta)^2$. □

Remark 5.7. If ρ is a linear expectation generated by probability measure P, then

$$E_P[E_P(\xi|\mathcal{C})\eta] = E_P[\xi\eta], \quad \forall \eta \in \mathbb{C}.$$

This means $E_P(\xi|\mathcal{C})$ not only satisfies (5.1) but also satisfies the following equation

$$\sup_{\eta\in\mathbb{C}} E_P[(\xi-\hat{\eta})(\xi-\eta)] = E_P(\xi-\hat{\eta})^2.$$

Remark 5.8. If ρ can be represented by a family of probability measures, then the condition $\inf_{\eta\in\mathbb{C}}\rho(\xi-\eta)^2 > 0$ in Theorem 5.6 can be abandoned since Lemma 5.4 still holds for either $\rho(|\xi_1|^p) = 0$ or $\rho(|\xi_2|^q) = 0$.

Appendix A. Some basic results

In this section, some results used in our paper are restated.

Theorem A.1. *If ρ is a sublinear expectation and $\mathcal{P}$ is the family of all linear expectations dominated by ρ, then*

$$\rho(\xi) = \max_{P\in\mathcal{P}} E_P[\xi], \forall \xi \in \mathbb{F}.$$

Proof. By Corollary 2.4 of Chapter I in [13], for any $\xi \in \mathbb{F}$, there exists a linear expectation L such that $L \le \rho$ and $L(\xi) = \rho(\xi)$. If we take all linear expectations dominated by ρ, then

$$\rho(\xi) = \max_{P\in\mathcal{P}} E_P[\xi], \forall \xi \in \mathbb{F}. \quad \square$$

Theorem A.2. *Let $\mathbb{F}$ be a normed space and ρ be a sublinear expectation from $\mathbb{F}$ to $\mathbb{R}$ dominated by some scalar multiple of the norm of $\mathbb{F}$. Then*

$$\{x^* \in \mathbb{F}^* : x^* \le \rho \quad on \quad \mathbb{F}\} \quad is \quad \sigma(\mathbb{F}^*, \mathbb{F}) - compact.$$

Proof. See Theorem 4.2 of Chapter 1 in [13]. □

We denote $\mathbb{F}^*_c$ as the set of linear expectations generated by countably additive measures.

Theorem A.3. *If $\mathcal{P}$ is a subset of $\mathbb{F}^*_c$ which is $\sigma(\mathbb{F}^*, \mathbb{F})$ − compact, then there exists a nonnegative $P_0 \in \mathcal{P}$ such that the measures in $\mathcal{P}$ are absolutely P_0-continuous, i.e., if $P_0(A) = 0$, then $\sup_{P\in\mathcal{P}} P(A) = 0$.*

Proof. See Corollary 1.2 in [15]. □

Theorem A.4 *(Komlós theorem). Let $\{\xi_n\}_{n\geq 1}$ be a sequence of random variables bounded in $L^1(P_0)$. Then there exist a subsequence $\{\xi_{n_i}\}_{i\geq 1}$ and a random variable $\xi \in L^1(P_0)$ such that*

$$\lim_{n\to\infty} \frac{1}{k}\sum_{i=1}^{k} \xi_{n_i} = \xi \quad P_0 - a.s.$$

Proof. See Theorem A.3.4 in [11]. □

Theorem A.5. *Let X be a normed vector space, $A \subset X$ be a convex set which is compact for the weak topology $\sigma(X, X^*)$ and B be a nonempty convex set of a vector space Y. Let also $f: A \times B \mapsto \mathbb{R}$ be a function with the property that*

(i) $x \mapsto f(x, y)$ is concave and continuous on X for each $y \in B$;
(ii) $y \mapsto f(x, y)$ is convex on Y for each $x \in A$.

Then

$$\sup_{x\in A} \inf_{y\in B} f(x, y) = \inf_{y\in B} \sup_{x\in A} f(x, y).$$

Proof. See Theorem B.1.2 of Appendix B in [11]. □

Theorem A.6 *(Mazur–Orlicz theorem). Let $\mathcal{X}$ be a nonzero space, $\rho : \mathcal{X} \to \mathbb{R}$ be sublinear and $\mathcal{D}$ be a nonempty convex subset of $\mathcal{X}$. Then there exists a linear functional L on $\mathcal{X}$ such that L is dominated by ρ, i.e., $L(X) \leq \rho(X)$ for all $X \in \mathcal{X}$ and*

$$\inf_{X\in\mathcal{D}} L(X) = \inf_{X\in\mathcal{D}} \rho(X).$$

Proof. See Lemma 1.6 of Chapter 1 in [13]. □

Theorem A.7. *Let A and B be two nonempty sets and f from $A \times B$ to $\mathbb{R} \bigcup \{\infty\}$. Then f has saddle points, i.e., there exists $(\bar{x}, \bar{y}) \in A \times B$ such that*

$$\forall x \in A, \forall y \in B : \quad f(x, \bar{y}) \leq f(\bar{x}, \bar{y}) \leq f(\bar{x}, y)$$

if and only if

$$\inf_{y\in B} f(\bar{x}, y) = \max_{x\in A} \inf_{y\in B} f(x, y) = \min_{y\in B} \sup_{x\in A} f(x, y) = \sup_{x\in A} f(x, \bar{y}).$$

Proof. See Theorem 2.10.1 of Chapter 2 in [14]. □

Appendix B. Some results about coherent risk measure

In this section, we give some basic definitions and results about coherent risk measures. Readers can see [1,2,5,7] for more details. Note that in order to ensure our statements of the entire paper is consistent, the expectation we used in our paper is sublinear which is different from the coherent risk measure defined in [1,2] or [5], in which it is superlinear. Thus it is represented as $\sup_{P\in\mathcal{P}} E_P$ instead of $\inf_{P\in\mathcal{P}} E_P$ and the conditional expectation is taken as $\operatorname{ess\,sup}_{P\in\mathcal{P}} E_P[\cdot|\mathcal{C}]$ instead of $\operatorname{ess\,inf}_{P\in\mathcal{P}} E_P[\cdot|\mathcal{C}]$. Though the definition is different, the methods and results are not affected.

For a given probability set $(\Omega, \mathcal{F}, P_0)$, $\{\mathcal{F}_n\}_{n\geq 1}$ is the filtration satisfying $\mathcal{F} := \vee_{n=1}\mathcal{F}_n$.

Definition B.1. A map $\pi : L^\infty(P_0) \mapsto \mathbb{R}$ is called a coherent risk measure, if it satisfies the following properties:

i) Monotonicity: for all X and Y, if $X \geq Y$, $\pi(X) \geq \pi(Y)$.
ii) Translation invariance: if λ is a constant, for all X, $\pi(\lambda + X) = \lambda + \pi(X)$.
iii) Positive homogeneity: if $\lambda \geq 0$, for all X, $\pi(\lambda X) = \lambda\pi(X)$.
iv) Subadditivity: for all X and Y, $\pi(X + Y) \leq \pi(X) + \pi(Y)$.

Definition B.2 *(Stability).* We say that the set $\mathcal{P}$ is stable if for elements $Q^0, Q \in \mathcal{P}^e$ with associated martingales Z_n^0, Z_n and for each stopping time τ, the martingale L defined as $L_n = Z_n^0$ for $n \leq \tau$ and $L_n = Z_\tau^0 \frac{Z_n}{Z_\tau}$ for $n \geq \tau$ defines an element of $\mathcal{P}$, where $\mathcal{P}^e$ denotes the elements in $\mathcal{P}$ which is equivalent to P_0 and $Z_n^Q := E_{P_0}[\frac{dQ}{dP_0}|\mathcal{F}_n]$.

Proposition B.3. *Let $\Phi_\tau(\xi) = \operatorname{ess\,inf}_{Q\in\mathcal{P}^e} E_Q[\xi|\mathcal{F}_\tau]$. Then the following is equivalent:*

i) Stability of the set $\mathcal{P}$.
ii) Recursivity: for each bounded $\mathcal{F}_N$-random variable ξ, the family $\{\Phi_\nu(\xi)|\nu$ is a stopping time$\}$ satisfies: for every two stopping times $\sigma \leq \tau$, we have $\Phi_\sigma(\xi) = \Phi_\sigma(\Phi_\tau(\xi))$.

References

[1] P. Artzner, F. Delbaen, J.M. Eber, D. Heath, Coherent measures of risk, Math. Finance 9 (3) (1999) 203–228.
[2] P. Artzner, F. Delbaen, J.M. Eber, D. Heath, H. Ku, Coherent multiperiod risk adjusted values and Bellman's principle, Ann. Oper. Res. 152 (2007) 5–22.
[3] A. Bensoussan, Stochastic Control of Partially Observable Systems, Cambridge Univ. Press, 1992.
[4] M. Davis, Lectures on Stochastic Control and Nonlinear Filtering, Tata Institute of Fundamental Research, vol. 75, Springer-Verlag, 1984.
[5] F. Delbaen, Coherent risk measures on general probability spaces, in: Advances in Finance and Stochastics, Springer-Verlag, Berlin, Heidelberg, New York, 2002.
[6] F. Delbaen, S. Peng, E. Rosazza Gianin, Representation of the penalty term of dynamic concave utilities, Finance Stoch. 14 (2010) 449–472.
[7] H. Föllmer, A. Schied, Stochastic Finance. An Introduction in Discrete Time, Walter de Gruyter, Berlin, New York, 2002.
[8] L. Jiang, Z. Chen, A result on the probability measures dominated by g-expectation, Acta Math. Appl. Sin. Engl. Ser. 20 (2004) 507–512.
[9] G. Kallianpur, Stochastic Filtering Theory, Springer-Verlag, 1980.
[10] S. Peng, BSDE and related g-expectations, in: Backward Stochastic Differential Equations, vol. 364, Pitman, 1997, pp. 141–159.
[11] H. Pham, Continuous-Time Stochastic Control and Optimization with Financial Applications, Springer-Verlag, Berlin, Heidelberg, 2009.
[12] E. Rosazza Gianin, Risk measures via g-expectations, Insurance Math. Econom. 39 (2006) 19–34.
[13] S. Simons, From Hahn–Banach to Monotonicity, Springer-Verlag, Berlin, Heidelberg, 2008.
[14] C. Zălinescu, Convex Analysis in General Vector Spaces, World Scientific, River Edge, NJ, 2002.
[15] X. Zhang, On weak compactness in spaces of measures, J. Funct. Anal. 143 (1997) 1–9.