Pure and Applied Functional Analysis
Volume 9, Number 3, 2024, 741–762

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

EXPECTILES IN RISK AVERSE STOCHASTIC PROGRAMMING AND DYNAMIC OPTIMIZATION

RAJMADAN LAKSHMANAN AND ALOIS PICHLER

Abstract. This paper features expectiles in dynamic and stochastic optimization. Expectiles are a family of risk functionals characterized as minimizers of optimization problems. For this reason, they enjoy various unique stability properties, which can be exploited in risk averse management, in stochastic optimization and in optimal control.

The paper provides tight relates of expectiles to other risk functionals and addresses their properties in regression. Further, we extend expectiles to a dynamic framework. As such, they allow incorporating a risk averse aspect in continuous-time dynamic optimization and a risk averse variant of the Hamilton–Jacobi–Bellman equations.

1. Introduction

Classical dynamic programming problems involve the expectation in the objective. The expectation is a risk neutral assessment of random outcomes. In many situations, specifically in economic environments, a risk averse assessment or risk management is much more favorable and desirable. For this reason there have been attempts to develop risk averse dynamic programming principles and risk averse Hamilton–Jacobi–Bellman equations.

Non-linear expectations (g-expectations, cf. Pardoux and Peng [20], Coquet et al. [8], Peng [21, 22, 23]) have been considered, e.g., to incorporate the aspect of risk to dynamic equations. A seemingly simpler approach involves risk measures (or risk functionals) instead of non-linear expectations, as risk measures are able to assess the risk associated with a random outcome (cf. Ruszczyński and Yao [34, 35]). By construction, risk measures are defined on random variables. For dynamic programming, they need to be extended to stochastic processes. The increments of stochastic processes are random variables so that composing risk measures over time and accumulating the corresponding risk is a promising approach to extend risk functionals from random variables to stochastic processes.

Specifically, this paper addresses expectiles in stochastic and dynamic optimization. Expectiles constitute a family of risk measure with unique properties. We

---

2020 *Mathematics Subject Classification.* 90C08, 90C15, 60G07.

Key words and phrases. Expectiles, multistage stochastic optimization, dynamic optimization, stochastic processes.

Funding: DFG, German Research Foundation – Project-ID 416228727 – SFB 1410.

demonstrate how they can be employed to incorporate risk aversion in dynamic programming and to develop risk averse Hamilton–Jacobi–Bellman equations.

Cont et al. [7] point out the importance of estimating risk measures in a robust way. In this context, Gneiting [13] proves that the Average Value-at-Risk, the most important risk measure in theory and practice, is not elicitable, that is, it is not possible to describe the risk measure as minimizer. More generally, Ziegel [39] proves that the only elicitable spectral risk measure is the (trivial) expectation. Bellini et al. [4] finally provide a proof that only expectiles constitute elicitable risk measures.

Expectiles have been introduced earlier in Newey and Powell [18] as

$$e_\alpha(X) \coloneqq \operatorname*{arg\,min}_{x\in\mathbb{R}} \mathbb{E}\,\ell_\alpha(X-x), \tag{1.1}$$

where X is a $\mathbb{R}$-valued random variable, $\alpha \in (0,1)$ and the scoring function (loss function) is[1]

$$\ell_\alpha(x) \coloneqq \alpha\cdot x_+^2 + (1-\alpha)(-x)_+^2 = \begin{cases} \alpha\cdot x^2 & \text{if } x\geq 0,\\ (1-\alpha)\,x^2 & \text{if } x\leq 0.\end{cases} \tag{1.2}$$

The characterization as a minimizer in the definition (1.1) applies for $X \in L^2$. The first order condition (cf. (1.3) below) is an equivalent characterization of the expectile, which applies – more generally – for $X \in L^1 \supset L^2$.

Definition 1.1 (Expectiles, cf. [18])**.** For $X \in L^1$ and a risk level $\alpha \in [0,1]$, the expectiles of a random variable X is the unique solution of the equation

$$\alpha\,\mathbb{E}(X-x)_+ = (1-\alpha)\,\mathbb{E}(x-X)_+, \tag{1.3}$$

where $x \in \mathbb{R}$.

Remark 1.2. In an alternative way, replacing the objective in (1.1) by $\mathbb{E}\big(\ell_\alpha(X-x)-\ell_\alpha(X-x_0)\big)$ for some fixed $x_0\in\mathbb{R}$ extends the definition to $X\in L^1$ as well, so that expectiles are well-defined for $X \in L^1$, even as minimizers.

For $\alpha = 1/2$, the expectile is the expectation, $e_{1/2}(X) = \mathbb{E}\,X$. It follows from symmetry of the loss function ℓ_α (i.e., $\ell_\alpha(x) = \ell_{1-\alpha}(-x)$) that

$$e_\alpha(X) = -e_{1-\alpha}(-X), \tag{1.4}$$

so that the expectile involves both tails, the lower and the upper tail of the distribution of the random variable X. For $X\in L^\infty$, the expectile approaches the essential supremum for increasing risk level, $e_\alpha(X)\to \operatorname{ess\,sup} X$ as $\alpha\to 1$.[2] More generally, we have the monotone behavior

$$\mathbb{E}\,X \leq e_\alpha(X) \leq e_{\alpha'}(X) \leq \operatorname{ess\,sup} X \tag{1.5}$$

for $1/2 \leq \alpha\leq \alpha' \leq 1$.

[1] $x_+ \coloneqq \max(0,x)$

[2] The essential supremum of X is the smallest *number* $c\in\mathbb{R}$ so that $X\leq c$ a.s.

Outline of the paper. In the following Section 2 we elaborate that expectiles constitute a risk measure, and we provide tight relations to other risk measures. Next, we introduce conditional risk functionals in Section 3. These are important for risk management in discrete and in continuous time. In continuous time (Section 4), we consider the risk-averse generator, which turns out to be a non-linear differential operator. We finally employ expectiles for dynamic optimization problems in Section 5 and conclude in Section 6.

2. Elicitable risk measures

The expectile $e_\alpha(\cdot)$ is a risk measure as introduced in Artzner et al. [3]. That is, the mapping $X \mapsto e_\alpha(X)$, provided that $\alpha \geq 1/2$, satisfies the following four axioms formulated for (convex) risk measures $\mathcal{R}\colon \mathcal{Y} \to \mathbb{R}$, where $\mathcal{Y}$ is an appropriate linear space of $\mathbb{R}$-valued random variables (for example $\mathcal{Y} = L^1(P)$) on the probability space $(\Omega, \mathcal{F}, P)$:

(i) $\mathcal{R}(X) \leq \mathcal{R}(Y)$ for all $X \leq Y$ almost everywhere,
(ii) $\mathcal{R}(X+Y) \leq \mathcal{R}(X) + \mathcal{R}(Y)$ for all X, $Y \in \mathcal{Y}$,
(iii) $\mathcal{R}(\lambda X) = \lambda\, \mathcal{R}(X)$ for all $\lambda > 0$, and
(iv) $\mathcal{R}(c + X) = c + \mathcal{R}(X)$ for all $c \in \mathbb{R}$.

The expectile is a risk functional satisfying the Axioms (i)–(iv) above (Appendix 7 presents a brief proof for the subadditivity (ii), while the other assertions are evident). Further, the expectile $e_\alpha(\cdot)$ is the only risk measure which can be expressed as a minimizer – as in (1.1) – in addition. We will elaborate below that the expectile is not a spectral risk measure. The natural space (cf. Pichler [27]) of expectiles is $\mathcal{Y} = L^1$, cf. also the discussion in Section 1 above. In what follows – unless stated differently – we will always assume that $\mathcal{Y} = L^1$.

Explicit expressions for the expectiles are available only in exceptional cases. For the uniform distribution in the interval $[0,1]$, $U \sim \mathcal{U}[0,1]$, e.g., the expectile is $e_\alpha(U) = \frac{\alpha - \sqrt{\alpha(1-\alpha)}}{2\alpha - 1}$.

To extend expectiles to a risk measure in continuous time employing the Wiener process (Brownian motion), we shall frequently need the expectile of the normal distribution, for which at least the following series expansion is available.

Example 2.1. An explicit expression for the expectile of normally distributed random variables, $X \sim \mathcal{N}(\mu, \sigma^2)$, is not available. It holds that

$$e_\alpha(X) = \mu + \sigma\sqrt{\frac{8}{\pi}}\Big(\alpha - \frac{1}{2}\Big) + \sigma\frac{8\sqrt{2}}{\sqrt{\pi}^3}\Big(\alpha - \frac{1}{2}\Big)^3 + \mathcal{O}\Big(\alpha - \frac{1}{2}\Big)^5. \tag{2.1}$$

Proof. The general assertion derives from the standard normal distribution. Denoting the density of the standard normal distribution by $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ and by $\Phi(x) = \int_{-\infty}^x \varphi(t)\, dt$ its antiderivative, it holds that

$$\mathbb{E}(X - t)_+ = \int_t^\infty (x - t)\varphi(x)\, dx = \varphi(t) - t\big(1 - \Phi(t)\big)$$

and

$$\mathbb{E}(t-X)_+ = \int_{-\infty}^{t} (t-x)\varphi(x)\,dx = t\,\Phi(t)+\varphi(t),$$

which follows readily by employing the identity $\varphi'(x) = -x\,\varphi(x)$. Based on (1.3) define now

$$\begin{aligned} f(\alpha,e) &\coloneqq \alpha\cdot\bigl(\varphi(e)-e\bigl(1-\Phi(e)\bigr)\bigr)-(1-\alpha)\cdot\bigl(e\,\Phi(e)+\varphi(e)\bigr) \\ &= (2\alpha-1)\bigl(\varphi(e)+e\,\Phi(e)\bigr)-\alpha\,e \end{aligned} \tag{2.2}$$

so that the expectile e_α of the normally distributed random variable X satisfies $f(\alpha, e_\alpha) = 0$ for every $\alpha \in (0,1)$. We now apply the implicit function theorem.

As $e_{1/2}(X) = \mathbb{E}\,X = 0$ for the normal distribution it holds that $f(1/2, e_{1/2}) = 0$. Further, the partial derivatives of f at (α, e) are $f_\alpha(1/2, 0) = \sqrt{\frac{2}{\pi}}$ and $f_e(\alpha, e) = -\frac{1}{2}$ so that the first term in assertion (2.1) follows with the implicit function theorem. The coefficient for the next term $\bigl(\alpha - 1/2\bigr)^2$ is zero, because the function (2.2) is odd with respect to the center $1/2$, as the normal distribution is symmetric, cf. (1.4). The remaining coefficient is found by differentiating the function (2.2) further. We omit the rather technical computations here, as our further results build on the first two terms only. □

Example 2.2. For a log-normal random variable X with $\log X \sim \mathcal{N}(\mu, \sigma^2)$, the expectiles are

$$e_\alpha(X) = e^{\mu+\frac{\sigma^2}{2}} + \Bigl(e^{\sigma^2}-1\Bigr)\,e^{2\mu+\sigma^2}\Bigl(\alpha-\frac{1}{2}\Bigr)\,4\sqrt{e}\bigl(2\Phi(1/2)-1\bigr)+\mathcal{O}\Bigl(\alpha-\frac{1}{2}\Bigr)^2.$$

Proof. As above, the proof again relies on explicitly available expressions

$$\mathbb{E}(X-t)_+ = \int_{\log t}^{\infty} (e^x - t)\varphi(x)\,dx = \sqrt{e}\,\Phi(1-\log t) - t\,\Phi(-\log t)$$

and

$$\mathbb{E}(t-X)_+ = \int_{-\infty}^{\log t} \bigl(t-e^x\bigr)\varphi(x)\,dx = \sqrt{e}\,\Phi(1-\log t)-\sqrt{e}+t\,\Phi(\log t).$$

The statement follows again by the implicit function theorem. □

2.1. **Tight comparison with important risk measures.** In what follows, we shall compare expectiles with important risk measures and give the tightest-possible estimates and the smallest spectral risk measure enveloping the expectiles.

The Average Value-at-Risk is the smallest convex envelope of the Value-at-Risk (cf. Föllmer and Schied [12]). The Average Value-at-Risk can be stated in the equivalent forms (cf. Pflug [24])

$$\begin{aligned} \mathsf{AV@R}_\alpha(X) &\coloneqq \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(\alpha)\mathrm{d}\alpha \\ &= \min\left\{q+\frac{1}{1-\alpha}\,\mathbb{E}(X-q)_+ \colon q\in\mathbb{R}\right\}, \end{aligned} \tag{2.3}$$

where

$$\mathsf{V@R}_\alpha(X) \coloneqq F_X^{-1}(\alpha) \coloneqq \inf\bigl\{x\colon P(X\le x)\ge\alpha\bigr\} \tag{2.4}$$

is the Value-at-risk.

The Average Value-at-risk is the fundamental building block in the Kusoka representation (cf. Kusuoka [16]) and the most important risk functional in actuarial practice. Notice as well that the Average Value-at-Risk is the *minimum objective* of an optimization problem (problem (2.3)), while the expectile in (1.1) is the *minimizer* of an optimization problem.

Remark 2.3 (Quantiles)**.** Similarly to the expectile, the Value-at-Risk defined in (2.4) is a minimizer of an optimization problem, specifically the problem

$$\min_{q\in\mathbb{R}} \mathbb{E}\,\tilde{\ell}_\alpha(X-q)$$

with scoring function

$$\tilde{\ell}_\alpha(x) := \begin{cases} -(1-\alpha)\,x & \text{if } x\le 0,\\ \alpha\cdot x & \text{if } x\ge 0\end{cases} = \Big(\alpha-\frac{1}{2}\Big)x+\frac{1}{2}\,|x|\,,$$

well-known from quantile regression. Indeed, the first order condition is $0 = \frac{\partial}{\partial q}\mathbb{E}\,\ell_\alpha(X-q) = \alpha\,\mathbb{E}\,\mathbb{1}_{\{X>q\}} - (1-\alpha)\,\mathbb{E}\,\mathbb{1}_{\{X\le q\}} = \alpha - P(X\le q)$ and hence the assertion. However, by violating (ii) above, the Value-at-Risk is *not* a convex risk functional.

Definition 2.4 (Spectral risk measure, cf. Acerbi and Simonetti [2], Acerbi [1])**.** Let $\sigma\colon [0,1)\to\mathbb{R}_{\ge 0}$ be a non-negative, non-decreasing function with $\int_0^1 \sigma(u)\,\mathrm{d}u = 1$. Then

$$\mathcal{R}_\sigma(X) = \int_0^1 F_X^{-1}(\alpha)\sigma(\alpha)\,\mathrm{d}\alpha, \qquad X\in\mathcal{Y},$$

is a risk measure. $\mathcal{R}_\sigma$ is called a the *spectral risk measure* and the function σ is called the *spectrum* of $\mathcal{R}_\sigma$.

The expectiles are not a spectral risk measure themselves. But for every expectile, there is a smallest spectral risk measure.

Proposition 2.5 (Enveloping risk measure)**.** *If $\mathcal{R}_\sigma(X)$ is any spectral risk measure with*

$$e_\alpha(X) \le \mathcal{R}_\sigma(X) \tag{2.5}$$

for every random variable $X\in\mathcal{Y}$, then $e_\alpha(X)\le s_\alpha(X)\le\mathcal{R}_\sigma(X)$ for all X, where

$$s_\alpha(X) := \int_0^1 F_X^{-1}(u)\frac{\alpha(1-\alpha)}{\big(\alpha-u(2\alpha-1)\big)^2}\,\mathrm{d}u; \tag{2.6}$$

that is, s_α is the smallest spectral risk measure larger than e_α.

Proof. Above all, $s_\alpha(\cdot)$ is a spectral risk functional, as $u\mapsto \frac{\alpha(1-\alpha)}{\big(\alpha-u(2\alpha-1)\big)^2}$ is a non-negative, increasing function and $\int_0^1 \frac{\alpha(1-\alpha)}{\big(\alpha-u(2\alpha-1)\big)^2}\,\mathrm{d}u = 1$.

Bellini et al. [4, Proposition 9] provide the Kusuoka representation

$$e_\alpha(X) = \max_{\gamma\in[1/\beta,1]} \gamma\,\mathbb{E}\,X + (1-\gamma)\,\mathsf{AV@R}_{\frac{\beta-\frac{1}{\gamma}}{\beta-1}}(X) \tag{2.7}$$

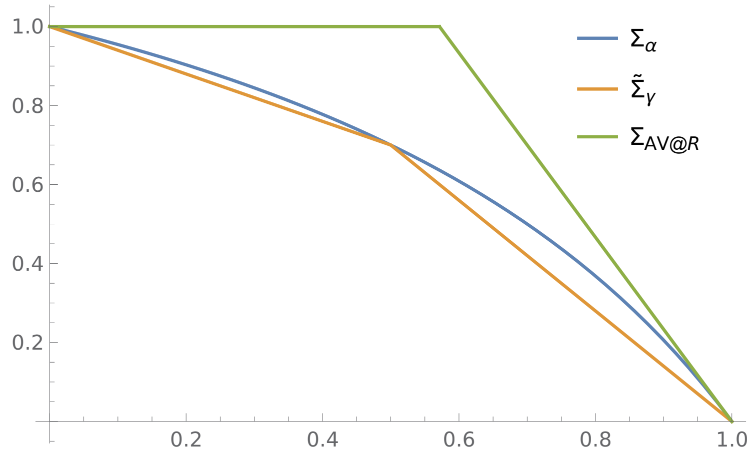

FIGURE 1. The function Σ_α and $\tilde{\Sigma}_\gamma$, exemplified for $\alpha = 70\,\%$ and $\gamma = 60\,\%$

for expectiles, where $\beta = \frac{\alpha}{1-\alpha}$. Define the functions $\Sigma_\gamma(u) := \gamma(1-u) + (1-\gamma)\min\left(1, \frac{1-u}{1-\frac{\beta-\frac{1}{\gamma}}{\beta-1}}\right)$ and $\Sigma(u) := \frac{\alpha(1-u)}{\alpha-u(2\alpha-1)}$. Both functions coincide at $u=0$, $u=1$ and $u = \frac{\alpha(1+\gamma)-1}{(2\alpha-1)\gamma}$; indeed $\Sigma_\gamma(0) = \Sigma(0) = 1$, $\Sigma_\gamma(1) = \Sigma(1) = 0$ and

$$\Sigma\left(\frac{\alpha(1+\gamma)-1}{(2\alpha-1)\gamma}\right) = \Sigma_\gamma\left(\frac{\alpha(1+\gamma)-1}{(2\alpha-1)\gamma}\right) = \frac{\alpha(1-\gamma)}{2\alpha-1}. \tag{2.8}$$

As Σ_γ is piece wise linear and Σ concave, it follows that $\Sigma_\gamma(u) \le \Sigma(u)$ for all $u \in [0,1]$. With integration by parts it follows further that

$$\begin{aligned} \gamma\,\mathbb{E}\,X + (1-\gamma)\,\mathsf{AV@R}_{\frac{\beta-\frac{1}{\gamma}}{\beta-1}}(X) &= -\int_0^1 F_X^{-1}(u)\,\mathrm{d}\Sigma_\gamma(u) \\ &= -\,F_X^{-1}(u)\Sigma_\gamma(u)\big|_{u=0}^1 + \int_0^1 \Sigma_\gamma(u)\,\mathrm{d}F_X^{-1}(u) \\ &\le -\,F_X^{-1}(u)\Sigma(u)\big|_{u=0}^1 + \int_0^1 \Sigma(u)\,\mathrm{d}F_X^{-1}(u) \\ &= -\int_0^1 F_X^{-1}(u)\,\mathrm{d}\Sigma(u) \\ &= s_\alpha(X) \end{aligned} \tag{2.9}$$

and thus $e_\alpha \le s_\alpha$. The assertion follows, as for every $u \in (0,1)$ there is $\gamma \in \left(\frac{1-\alpha}{\alpha}, 1\right)$ ($\gamma = \frac{1-\alpha}{u(1-2\alpha)+\alpha}$) so that $\Sigma_\gamma(u) = \Sigma(u)$ by (2.8) above (cf. Figure 1 for illustration). □

We have the following comparison with the Average Value-at-Risk. The comparison is sharp in the sense that the risk rates cannot be improved.

Corollary 2.6. *For every random variable $X \in L^1$ it holds that*

$$e_{\frac{1}{2-\alpha}}(X) \le \mathsf{AV@R}_\alpha(X), \qquad \alpha \in [0,1], \tag{2.10}$$

and

$$\frac{\alpha}{3\alpha-1}\,\mathbb{E}\,X + \frac{2\alpha-1}{3\alpha-1}\,\mathsf{AV@R}_{2-\frac{1}{\alpha}}(X) \le e_\alpha(X) \le \mathsf{AV@R}_{2-\frac{1}{\alpha}}(X) \tag{2.11}$$

for every $\alpha \in [1/2, 1]$.

For non-negative random variables ($X \ge 0$ a.s.) we further have

$$\mathsf{AV@R}_\alpha(X) \le \frac{1}{1-\alpha} e_{\frac{1}{2-\alpha}}(X) \tag{2.12}$$

and

$$e_\alpha(X) \le \frac{\alpha}{1-\alpha}\,\mathbb{E}\,X. \tag{2.13}$$

The risk rates in the preceding equations (2.10)–(2.13) *are optimal, they cannot be improved.*

Remark 2.7. The preceding corollary might give the impression that e_α is 'weak' in the sense that it attains smaller values than the average value at risk and is comparable to the risk neutral expectation. However, it holds that $e_\alpha(X) \to 1$ for $\alpha \to 1$, as follows readily from (1.3). Further, we have that the Average Value-at-Risk is a lower bound for the expectiles in view of (2.12), so that expectiles are at least as 'strong' as the Average Value-at-Risk.

Proof of Corollary 2.6. Employing the notation of the proof of Proposition 2.5 and $\Sigma_\alpha(u) := \min\left(1, \frac{1-u}{\frac{1}{\alpha}-1}\right)$, we have that $\Sigma(u) \le \Sigma_\alpha(u)$. As in the proof above we conclude that $\mathcal{R}_\alpha(X) \le \mathsf{AV@R}_\alpha(X)$ and with (2.5) that (2.10). The inequality (2.10) is tight, as $\Sigma'_\gamma(1) \xrightarrow[\gamma\to 1]{} \Sigma'(1)$.

As for the remaining inequality choose $\gamma = \frac{\alpha}{3\alpha-1}$ in (2.7), and replace α by $\frac{1}{2-\alpha}$ in (2.11) to obtain (2.10).

The inequality $\min\left(1, \frac{1-u}{1-\alpha}\right) \le \frac{1}{1-\alpha}\Sigma_\gamma(u)$ is evident for every $u \in [0,1]$, and the remaining assertion (2.12) follows by the same reasoning as above. However, for inequality (2.9) to hold true it is essential that $X \ge 0$ a.s.

Hölder's inequality, applied to (2.6), gives

$$\mathbb{E}\,X \le s_\alpha(X) \le \int_0^1 F_X^{-1}(u)\,\mathrm{d}u \cdot \max_{u\in[0,1]} \frac{\alpha(1-\alpha)}{\big(\alpha - u(2\alpha-1)\big)^2} = \mathbb{E}\,X \cdot \frac{\alpha}{1-\alpha}$$

and thus (2.13). □

3. Conditional and Dynamic Risk Measure

Risk functionals – as discussed above – are employed to assess the risk of a random outcome. For this reason, they have the economic interpretation of an insurance premium, while the random outcome is the random insurance benefit (the random variable). While the premium is known beforehand, the insurance benefit (the random outcome) is not, it is revealed later.

Conditional risk measures are employed in risk management over time, they address stochastic processes instead of random variables. Nested risk measures, which are compositions of risk functionals over time, enjoy the economic interpretation of risk premiums for insurance on a rolling horizon basis. For a discussion of nested risk functionals we may refer to Cheridito and Kupper [6], Riedel [31], Shapiro [37], Ruszczyński and Shapiro [33] and Pichler and Schlotter [28].

3.1. **The conditional expectile.** Definition 1.1 allows extending the expectile to conditional expectiles, which are conditioned on some σ-algebra. This constitutes a major building block to extend the definition of expectiles from random variables to stochastic processes.

Definition 3.1 (Conditional expectiles)**.** Let $X \in L^1$ be a random variable and $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$, $\mathcal{G} \subset \mathcal{F}$ and α a $\mathcal{G}$-measureable variable with values in $[0,1]$. The $\mathcal{G}$-measureable random variable Z satisfying

$$\alpha \cdot \mathbb{E}\big((X-Z)_+ \mid \mathcal{G}\big) = (1-\alpha)\cdot \mathbb{E}\big((Z-X)_+ \mid \mathcal{G}\big) \qquad \text{a.s.} \tag{3.1}$$

is called the *conditional expectile* (i.e., the conditional version of (1.3)) and denoted $Z = e_\alpha(X \mid \mathcal{G})$. As usual for the conditional expectation, we shall also write $e^{\mathcal{G}}(X) := e(X \mid \mathcal{G})$ and $e^{Y=y}(X) := e(X \mid Y=y)$ for the conditional expectile and its versions.

The solution of the problem (3.1) exists and is unique for the same reasons as for the usual expectile, and $e_\alpha(X \mid \mathcal{G}) \in L^1$, as $\big(e_\alpha(X \mid \mathcal{G}) - X\big)_+$ and $\mathbb{E}\big((X - e_\alpha(X \mid \mathcal{G}))_+ \mid \mathcal{G}\big)$ exist in (3.1).

Remark 3.2. Based on the properties of the conditional expectation (cf. Section 2), we have the following properties of the conditional expectile.

(i) $e_\alpha^{\mathcal{G}}(X) \le e_\alpha^{\mathcal{G}}(Y)$ a.e. for all $X \le Y$ almost everywhere,
(ii) $e_\alpha^{\mathcal{G}}(X+Y) \le e_\alpha^{\mathcal{G}}(X) + e_\alpha^{\mathcal{G}}(Y)$ a.e.,
(iii) $e_\alpha^{\mathcal{G}}(\lambda X) = \lambda\, e_\alpha^{\mathcal{G}}(X)$ for all $\lambda > 0$ and λ which is $\mathcal{G}$-measurable,
(iv) $e_\alpha^{\mathcal{G}}(c+X) = c + e_\alpha^{\mathcal{G}}(X)$ for all $\mathbb{R}$-valued c measurable with respect to $\mathcal{G}$.

In what follows, we shall consider the conditional expectile for a single σ-algebra first and discuss regression. Next, we consider filtrations $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$, typically generated by a stochastic process $X = (X_t)_{t\in\mathcal{T}}$.

3.2. **Conditional expectiles in stochastic optimization and regression.** Stochastic optimization and most typical problems in machine learning (as the training of neural networks) as well as specific problems in inverse problems (cf. Lu and Pereverzev [17]) consider the problem

$$\begin{aligned} &\text{minimize } f_0(x) := \mathbb{E}\, f(x,\xi) \\ &\text{subject to } x \in \mathcal{X}, \end{aligned} \tag{3.2}$$

where the objective is a risk neutral expectation, $f\colon \mathcal{X}\times\mathbb{R}^m \to \mathbb{R}$ is a function, $\mathcal{X} \subset \mathbb{R}^d$ is closed and ξ is a random variable with values in $\mathbb{R}^m$. Sample average approximation builds on independent realizations ξ_i of identically distributed random variable ξ, $i = 1, \dots,$ to solve (3.2) in real world applications. To this end, the empirical version

$$\hat{f}_n(x) \coloneqq \frac{1}{n}\sum_{i=1}^{n} f(x,\xi_i)$$

is considered instead of the expectation $\mathbb{E}\, f(x,\xi)$ in (3.2) for varying $x \in \mathcal{X}$.

We consider the measure points (observations) $X \in \mathcal{X}$ to be random (with measure P) as well and intend to 'learn' the function f_0 based on observations

$$\big(X_i, f(X_i,\xi_i)\big), \quad i = 1,\dots,n, \tag{3.3}$$

where (X_i,ξ_i) are revealed jointly (cf. Dentcheva and Lin [9] for further motivation in stochastic optimization and an alternative approach); even more generally, we consider the iid observations

$$(X_i, f_i), \quad i = 1,\dots n, \tag{3.4}$$

which is (3.3) with $f_i \coloneqq f(X_i,\xi_i)$.

To model (3.4), let ρ be the probability measure of the joint distribution (X,f) and denote the marginal measure by $P(A) \coloneqq \rho(A\times\mathbb{R})$. Then there exists a regular conditional probability kernel (cf. Kallenberg [14]) so that

$$\rho(A\times B) = \int_A \rho(f\in B|\, x)\, P(dx). \tag{3.5}$$

The bivariate measure ρ in (3.5) is not an artifact. Indeed, denote the conditional measures of f given X by the Markov kernel $\rho\colon \mathcal{X}\times\mathcal{B}(\mathbb{R})\to[0,1]$, that is, $\rho(f\in A \mid X = x) = \rho(x,A)$, then (X,f) jointly follow the composed measure (3.5),

$$(X,f)\sim\rho,$$

and hence both approaches are equivalent.

For a random vector $(X,f)\in\mathbb{R}^d\times\mathbb{R}$ with law ρ set

$$f_0(x) \coloneqq \mathbb{E}(f \mid X = x); \tag{3.6}$$

this definition notably corresponds to

$$f_0(x) = \mathbb{E}(f(X,\xi) \mid X = x)$$

in the setting (3.3) above. For this reason, the stochastic optimization problem (3.2) is equivalent to[3]

$$\operatorname*{ess\,inf}_{x\in\mathcal{X}} \mathbb{E}(f \mid X = x), \tag{3.7}$$

where (X,f) is a random variable with law ρ, provided that $\operatorname{supp} P = \mathcal{X}$, where

$$\operatorname{supp} P \coloneqq \bigcap \big\{A\colon A \text{ is closed and } P(A) = 1\big\}$$

is the support.[4]

[3]The essential infimumess $\inf(f \mid X)$ is the largest random variable g, measurable with respect to $\sigma(X)$ (the σ-algebra generated by X), so that $g\le f$, cf. Föllmer and Schied [12, Definition A.34]. Measurability is the crucial difference in comparison to the (unconditional) essential supremum in Footnote 2.

[4]Cf. Rüschendorf [32] for the support of the marginal measure P.

Note, however, that not every random vector (X, f) can be recast as in (3.3) for a function f and a random ξ. For this reason, the problem formulation (3.7) is more general than the genuine problem (3.2).

3.3. **Risk assessment with conditional expectiles.** To incorporate risk in the assessment, consider the conditional expectation (3.6) and define

$$f_\alpha(x) \coloneqq e_\alpha(f \mid X = x),$$

where $e_\alpha^{\sigma(X)}$ is the conditional expectile introduced in Section 3.1 above. Based on (1.5), we have that

$$f_\alpha(x) \geq f_0(x), \qquad \text{for } \alpha \geq 1/2,\ x \in \mathcal{X}.$$

The function f_α intentionally *overestimates* (overrates) the risk-free assessment f_0 and the surplus $f_\alpha - f_0$ is the amount attributed to risk aversion.

To solve the risk averse version of the stochastic optimization problem (3.7),

$$\begin{aligned} &\text{minimize } e_\alpha(f \mid X = x) \\ &\text{subject to } x \in \mathcal{X}, \end{aligned}$$

just find an estimator for $\hat{e}_\alpha$ for e_α first and then solve

$$\begin{aligned} &\text{minimize } \hat{e}_\alpha(x) \\ &\text{subject to } x \in \mathcal{X}. \end{aligned}$$

The substitute $\hat{e}_\alpha(\cdot)$ is chosen in an adequate space of functions. Dentcheva and Lin [9] consider the Nadaraya–Watson kernel estimator to solve the problem. Here, we exploit the problem by using reproducing kernel Hilbert spaces (RKHS) with kernel function k, where we may refer to Berlinet and Thomas-Agnan [5] for details.

Definition 3.3. For a kernel function $k\colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, the RKHS space $\mathcal{H}_k$ is the completion of the functions $f(x) = \sum_{i=1}^{\ell} w_i\, k(x, x_i)$ with respect to the inner product

$$\langle k(\cdot, x_i) \mid k(\cdot, x_j)\rangle = k(x_i, x_j), \qquad i, j = 1, \dots, \ell,$$

where x_i and $x_j \in \mathcal{X}$.

The regularized problem is

$$\text{minimize } \frac{1}{n}\sum_{i=1}^{n} \ell_\alpha\big(\hat{e}_\alpha(X_i) - f_i\big) + \lambda \|\hat{e}_\alpha\|_k^2, \tag{3.8}$$

where $\hat{e}_\alpha(\cdot) \in \mathcal{H}_k$. It follows from the generalized representer theorem (cf. Schölkopf et al. [36]), that the function e_α is given by

$$e_\alpha(\cdot) = \frac{1}{n}\sum_{i=1}^{n} w_i\, k(\cdot, X_i),$$

that is, the supporting points are exactly the points X_i, $i = 1, \dots, n$, where measurements f_i, $i = 1, \dots, n$, are available. It might be convenient in some situations

Algorithm 1. Newton-like iteration to solve (3.8).
Input: Measurements (f_i, X_i), $i = 1, \dots n$, and support points $\tilde{x}_j$, $j = 1, \dots, \tilde{n}$.
Output: The weights w_j, $j = 1, \dots, \tilde{n}$, of the function

$$\hat{e}_\alpha(\cdot) = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} w_j k(\cdot, \tilde{x}_j) \tag{3.9}$$

minimizing (3.8).
Set

$$K_{ij} := k(X_i, \tilde{x}_j)$$

for $i = 1, \dots n$ and $j = 1, \dots \tilde{n}$, and

$$\tilde{K}_{ij} := k(\tilde{x}_i, \tilde{x}_j)$$

for $i, j = 1, \dots, \tilde{n}$.
while *change of the weights w encountered* **do**
  for $i = 1$ **to** n **do**
    update $A_{ii} \leftarrow \begin{cases} \alpha & \text{if } f_i \le \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} w_j\, k(X_i, \tilde{x}_j), \\ 1-\alpha & \text{else} \end{cases}$
  end
  update

$$w \leftarrow w - \Big(\frac{\lambda}{\tilde{n}^2}\tilde{K} + \frac{1}{n^2\tilde{n}} K^\top A K\Big)^{-1} \cdot \Big(\frac{\lambda}{\tilde{n}^2}\tilde{K}w + \frac{1}{n^2\tilde{n}} K^\top A K w - \frac{1}{n\tilde{n}} K^\top A f\Big) \tag{3.10}$$

end
Result: The best approximating function (3.9).

to find the best approximation located at the points $\tilde{x}_j$, $j = 1, \dots, \tilde{n}$, that is, the function

$$\hat{e}_\alpha(\cdot) = \frac{1}{\tilde{n}} \sum_{j=1}^{\tilde{n}} w_j k(\cdot, \tilde{x}_j),$$

for fewer or special design points $\tilde{x}_j$, $j = 1, \dots, \tilde{n}$. We describe the equations for this generalized problem.

The first order conditions of problem (3.8) for the weights w_j, $j = 1, \dots, \tilde{n}$, are

$$0 = \frac{1}{n}\sum_{i=1}^{n} 2 \cdot \left(\frac{1}{\tilde{n}}\sum_{j'=1}^{\tilde{n}} w_{j'} k(X_i, \tilde{x}_{j'}) - f_i\right) \cdot \left\{\begin{array}{cl} \alpha & \text{if } f_i \le \hat{e}_\alpha(X_i) \\ 1-\alpha & \text{if } f_i \ge \hat{e}_\alpha(X_i) \end{array}\right\} \cdot \frac{1}{\tilde{n}} k(\tilde{x}_j, X_i) + + 2\frac{\lambda}{\tilde{n}^2}\sum_{j=1}^{\tilde{n}} w_j k(\tilde{x}_i, \tilde{x}_j). \tag{3.11}$$

Define $\tilde{K} := \big(k(\tilde{x}_\ell, \tilde{x}_j)\big)_{\ell,j=1}^{n}$, $K := \big(k(X_i, \tilde{x}_j)\big)_{i=1,j=1}^{n,\tilde{n}}$ and

$$A(w) := \operatorname{diag}\big(a_i(w),\ i = 1, \dots, n\big)$$

with entries

$$a_i(w) = \begin{cases} \alpha & \text{if } f_i \le \frac{1}{\tilde{n}}\sum_{j=1}^{n} w_j k(X_i, \tilde{x}_j), \\ 1-\alpha & \text{if } f_i \ge \frac{1}{\tilde{n}}\sum_{j=1}^{n} w_j k(X_i, \tilde{x}_j) \end{cases}$$

on the diagonal. Then the equations (3.11) rewrite as

$$\left(\frac{\lambda}{\tilde{n}^2}\tilde{K} + \frac{1}{n^2\tilde{n}} K^\top A(w) K\right) w = \frac{1}{n\tilde{n}} K^\top A(w) f.$$

This equation is not linear in w, as $A(w)$ depends in a nonlinear way on w. However, the problem can be solved by inverting the matrix to obtain a fixed point equation. With that, the equation can be iterated, and the algorithm converges after finitely many iterations, cf. (3.10) in Algorithm 1. Figure 2 displays a typical result of expectile regression. Farooq and Steinwart [10] is a starting point in investigating convergence properties of the expectile regression problem.

Remark 3.4. Note that the inverted matrix in (3.10) is the derivative of the right-hand side with respect to w, as A is constant for small changes in w. For this reason, the iteration in Algorithm 1 is a Newton iteration in essence, although the function (1.2) is not differentiable. As $A(w)$ is constant for small variations of w, thus (3.10) vanishes locally.

4. Risk aversion in stochastic processes

The considerations on the expectile in the preceding sections are based on random variables. The conditional variant in the expectile regression is achieved with a single σ-algebra. In what follows, we generalize the expectile for stochastic processes – in a discrete time setting first, and then in continuous time.

4.1. Nested expectile in discrete time.

Consider a stochastic process $X = (X_{t_i})_{i=0}^{n}$ in discrete time, where $0 =: t_0 < t_1 < \dots < t_n = T$. For a dissection in time consider the increments

$$X_T = X_{t_0} + (X_{t_1} - X_{t_0}) + \dots + (X_{t_n} - X_{t_{n-1}}).$$

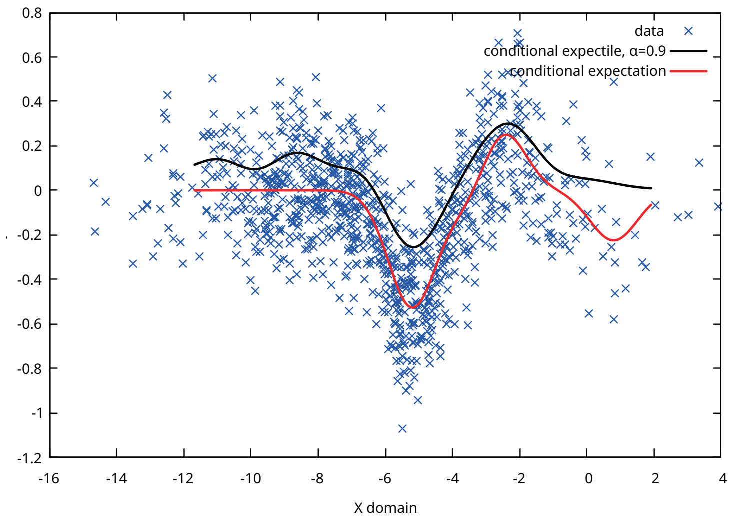

Figure 2. The expectile $\hat{e}_{90\,\%}(\cdot)$ based on $n = 1000$ observations overestimates the conditional expectation

The stochastic process X is adapted to the filtration $\mathcal{F}$, that is, X_t is measurable for every $\mathcal{F}_t$, $t \geq 0$, so most often we just may choose $\mathcal{F}_{t_i} \coloneqq \sigma(X_{t_j} : j \leq i)$. As well, we shall denote the sequence of σ-algebras by $\mathcal{F}_{t_0\,:\,t_n}$.

In what follows, we shall associate a certain risk for the time period $\Delta t \coloneqq t_{i+1} - t_i$ to come. For convenience in the presentation in what follows, we introduce the *rescaled* version of the expectile as

$$\tilde{e}_\beta(\cdot) \coloneqq e_{\frac{1+\sqrt{\beta}}{2}}(\cdot)$$

(i.e., $\alpha - \frac{1}{2} = \frac{\sqrt{\beta}}{2}$). The main reason for the rescaling is that $e_{1/2}(X) = \mathbb{E}\,X$, while $\mathsf{AV@R}_0(X) = \mathbb{E}\,X$, e.g. To ensure consistent parametrizations with other risk measures, we rescale the risk level so that $\tilde{e}_0(X) = \mathbb{E}\,X$ is associated with the risk-free assessment, while $\tilde{e}_1(X) = \operatorname{ess\,sup} X$ is the total risk averse assessment. The varying dynamic ($\sqrt{\beta}$ instead of β) turns out to be the natural choice in the continues time situation addressed below.

Definition 4.1 (Nested expectile)**.** Let $\big(\Omega, \mathcal{F} = (\mathcal{F}_{t_i})_{i=1}^n, P\big)$ be a filtered probability space and $\beta\colon \{t_0, \dots, t_n\} \to [0,1]$ be stochastic process adapted to the filtration $\mathcal{F} = (\mathcal{F}_{t_i})_{i=1}^n$. The nested expectile of the process with respect to the filtration $\mathcal{F}_{t_0:t_n}$, denoted $\tilde{e}_{\beta(\cdot)}^{\mathcal{F}_{t_0:t_n}}$, is

$$\tilde{e}_{\beta(\cdot)}^{\mathcal{F}_{t_0:t_n}}(X) \coloneqq X_0 + \tilde{e}^{\mathcal{F}_0}_{\beta(t_0, X_{t_0})\cdot(t_1 - t_0)}\Big(X_{t_1} - X_{t_0} + \dots \\ \dots + \tilde{e}^{\mathcal{F}_{t_{n-1}}}_{\beta(t_{n-1}, X_{t_{n-1}})\cdot(t_n - t_{n-1})}(X_T - X_{t_{n-1}})\Big),$$

or slightly more explicitly

$$\tilde{e}^{\mathcal{F}_{t_0:t_n}}_{\beta(\cdot)}(X) = X_0 + \tilde{e}^{\mathcal{F}_0}_{\beta(t_0,X_{t_0})\cdot(t_1-t_0)}\Big(X_{t_1} - X_{t_0} + \cdots + \tilde{e}^{\mathcal{F}_{t_{n-2}}}_{\beta(t_{n-2},X_{t_{n-2}})\cdot(t_{n-1}-t_{n-2})}\big(X_{t_{n-1}} - X_{t_{n-2}} + \tilde{e}^{\mathcal{F}_{t_{n-1}}}_{\beta(t_{n-1},X_{t_{n-1}})\cdot(t_n-t_{n-1})}(X_T - X_{t_{n-1}})\big)\Big).$$

Nested risk measures have been considered by Philpott et al. [26], Philpott and de Matos [25], e.g. In discrete time, fundamental properties of the Average Value-at-Risk have been elaborated by Xin and Shapiro [38], although for deterministic risk rates only and for random variables instead of stochastic processes. The definition above is dynamic, as the risk rate β is an adapted process itself. Note that the risk rate at time t may be chosen to reflect the history of observations up to t, it may depend on $\{t_i \leq t\colon i = 1, \ldots, n\}$.

We consider the following example, which prepares for the Wiener process.

Example 4.2 (Random walk, cf. Pichler and Schlotter [30])**.** Consider a random walk process starting at X_0 with independent Markovian increments

$$X_{t_{i+1}} - X_{t_i} \sim \mathcal{N}(0, t_{i+1} - t_i) \tag{4.1}$$

and constant risk rate $\beta(t,x) = \beta$. With (4.1) and the asymptotic formula (4.1) for the normal distribution, we have that

$$\begin{aligned} X_{t_1} + \tilde{e}^{\mathcal{F}_{t_1}}_{\beta}(X_{t_2} - X_{t_1}) &= X_{t_1} + \sqrt{t_{i+1} - t_i}\sqrt{\frac{2}{\pi}}\sqrt{\beta(t_{i+1} - t_i)} + o\big(t_{t+1} - t_i\big) \\ &= X_{t_1} + \sqrt{\frac{2\beta}{\pi}}(t_{i+1} - t_i) + o\big(t_{t+1} - t_i\big). \end{aligned} \tag{4.2}$$

Nesting these expressions as in Definition 4.1 gives the explicit expression

$$\tilde{e}^{\mathcal{F}_{t_0:T}}_{\beta(\cdot)}(X) = X_0 + \sqrt{\frac{2\beta}{\pi}}T + o(T), \tag{4.3}$$

where T is the terminal time, while

$$\tilde{e}^{\mathcal{F}_{t_0:T}}_{0}(X) = X_0$$

for the risk rate $\beta = 0$. The amount attributed to the risk averse assessment in (4.3) thus accumulates linearly with time.

Remark 4.3 (Tower property)**.** We emphasize as well that Definition 4.1 explicitly involves time, the risk $\beta(t_i)\cdot(t_{i+1} - t_i)$ is associated to the time interval starting at t_i and ending at t_{i+1}. With a further point in between, $t_{i+1/2}$, the components of the risk functionals above are

$$\tilde{e}^{\mathcal{F}_0}_{\beta(t_i)\cdot(t_{i+1/2}-t_i)}\left(X_{t_{i+1/2}} - X_{t_i} + \tilde{e}^{\mathcal{F}_{t_{i+1/2}}}_{\beta(t_{i+1/2})\cdot(t_{i+1}-t_{i+1/2})}\left(X_{t_{i+1}} - X_{t_{i+1/2}}\right)\right)$$

instead of

$$\tilde{e}^{\mathcal{F}_0}_{\beta(t_i)\cdot(t_{i+1}-t_i)}\left(X_{t_{i+1}} - X_{t_i}\right).$$

With that, the risk rates accumulate over time: accumulated risk rates are $\beta(t_i)(t_{i+1/2} - t_i) + \beta(t_{i+1/2})(t_{i+1} - t_{i-1/2})$ in the first case. This amount indeed coincides with

$\beta(t_i)(t_{i+1} - t_i)$ (this is the risk rate in the second case), provided that $\beta(t_i) = \beta(t_{i+1/2})$, i.e., the risk assessment does not vary over time.

For the expectation, the corresponding property is the tower property, that is, $\mathbb{E}\big(\mathbb{E}\, X \mid \mathcal{G})\big) = \mathbb{E}\, X$.

4.2. **The nested expectile in continuous time.** In order to assign risk to a stochastic process in continuous time, we consider the nested formulation introduced above for decreasing time-steps.

Definition 4.4 (Nested expectile)**.** Let $X = (X_t)_{t\leq T}$ be a stochastic process adapted to $\mathcal{F} = (\mathcal{F}_t)_{t\leq T}$ and $\beta = (\beta_t)_{t\leq T}$ be càdlàg (i.e., right continuous, with left limits) and adapted. With the nested expectile defined in Definition 4.1, the nested expectile is

$$\tilde{e}_\beta^{\mathcal{F}}(X) = \lim_{\max \Delta t \to 0} \tilde{e}_{\beta_{t_0:t_n}}^{\mathcal{F}_{t_0:t_n}}(X), \tag{4.4}$$

provided that the limit with respect to decreasing mesh sizes $\max \Delta t \coloneqq \max_{i=1}^n t_{i+1} - t_i$ exists.

Example 4.5 (State independent risk rates)**.** Example 4.2 generalizes for a state independent, but time dependent Riemann integrable risk rate $\beta(x,t) = \beta(t)$. As above, we obtain that

$$X_{t_1} + \tilde{e}_\beta^{\mathcal{F}_{t_1}}(X_{t_2} - X_{t_1}) = X_{t_1} + \sqrt{\frac{2\beta(t_i)}{\pi}}(t_{i+1} - t_i) + o\big(t_{t+1} - t_i\big)$$

and thus

$$\tilde{e}_{\beta(\cdot)}^{\mathcal{F}_{t_0:T}}(X) = X_0 + \sqrt{\frac{2}{\pi}} \int_0^T \sqrt{\beta(t)}\,\mathrm{d}t$$

for $\Delta t \to 0$, as β is Riemann integrable. Again, this is an explicit expression for the total risk aversion of the entire random walk process with increments (4.1).

Definition 4.6 (Risk generator)**.** Let $(X_t)_{t\geq 0}$ be a stochastic process adapted to the filtration $\sigma(X)$ and $\beta(t,x)$ be a risk rate. The risk generator is

$$\mathcal{G}_\beta f(x,t) \coloneqq \lim_{h\to 0} \frac{\tilde{e}_{\beta(t,X_t)}^{\sigma(X)}\big(f(X_{t+h}) \,|\, X_t = x\big) - f(x)}{h},$$

provided that the limit exists.

Note, that $\mathcal{G}$ is an operator, which maps the (smooth) function f to $\mathcal{G}_\beta f$, which is a function again. In contrast to the risk-neutral generator, the risk generator $\mathcal{G}_\beta$ is possibly not linear, as we will see in what follows.

Proposition 4.7. *Let X_t follow the stochastic differential equation*

$$\mathrm{d}X_t = \mu(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t \tag{4.5}$$

with respect to the Wiener process (Brownian motion) $(W_t)_{t\geq 0}$ and the functions μ and σ be Lipschitz, i.e., $|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K|x - y|$ so that *strong solutions of* (4.5) *exist. For a smooth function f, the risk generator is*

$$\begin{aligned}\mathcal{G}_\beta f(t,x) &= \frac{\partial f(t,x)}{\partial t} + \mu(t,x)\cdot\frac{\partial f(t,x)}{\partial x} + \frac{1}{2}\sigma(t,x)^2\cdot\frac{\partial^2 f(t,x)}{\partial x^2} \\ &\quad + \sqrt{\frac{2}{\pi}\beta(t,x)}\cdot\left|\sigma(x,t)\cdot\frac{\partial f(t,x)}{\partial x}\right|. \end{aligned}\tag{4.6}$$

Proof. The proof follows Øksendal [19, Section 7.3] (another valuable reference is Karatzas and Shreve [15]).

Consider the stochastic process $Y_t \coloneqq f(t,X_t)$. From Ito's rule we deduce that

$$\begin{aligned}Y_{t+\Delta t} = Y_t &+ \int_t^{t+\Delta t}\Big(\frac{\partial f}{\partial t}(s,X_s) + \mu(s,X_s)\frac{\partial f}{\partial x}(s,X_s) \\ &\qquad + \frac{1}{2}\sigma(s,X_s)^2\frac{\partial^2 f}{\partial x^2}(s,X_s)\Big)\mathrm{d}s \\ &+ \int_t^{t+\Delta t}\sigma(s,X_s)\frac{\partial f}{\partial x}(s,X_s)\mathrm{d}W_s,\end{aligned}$$

where the second part is a martingale with increments following the Wiener process. Following the proof of the Ito formula in Øksendal [19, p. 46ff], the functions μ and σ are approximated by the constants $\mu(s,X_s)\approx\mu(t,X_t)$ and $\sigma(s,X_s)\approx\sigma(t,X_t)$ for $s\in[t,t+\Delta t)$ so that

$$\begin{aligned}Y_{t+\Delta t} - Y_t = &\left(\frac{\partial f}{\partial t}(t,X_t) + \mu(t,X_t)\frac{\partial f}{\partial x}(t,X_t) + \frac{1}{2}\sigma(t,X_t)^2\frac{\partial^2 f}{\partial x^2}(t,X_t)\right)\Delta t \\ &+ \sigma(t,X_t)\frac{\partial f}{\partial x}(t,X_t)\cdot\big(W_{t+\Delta t}-W_t\big).\end{aligned}$$

$Y_{t+\Delta t} - Y_t$ is a normally distributed random variable with mean

$$Y_t + \left(\frac{\partial f}{\partial t}(t,X_t) + \mu(t,X_t)\frac{\partial f}{\partial x}(t,X_t) + \frac{1}{2}\sigma(t,X_t)^2\frac{\partial^2 f}{\partial x^2}(t,X_t)\right)\Delta t$$

and variance

$$\left(\sigma(t,X_t)\frac{\partial f}{\partial x}(t,X_t)\right)^2\Delta t.$$

We deduce from (2.1) that

$$\begin{aligned}\tilde{e}^{X_t}_{\beta\cdot\Delta t}(Y_{t+\Delta t}) - Y_t = &\left(\frac{\partial f}{\partial t}(t,X_t) + \mu(t,X_t)\frac{\partial f}{\partial x}(t,X_t)\right. \\ &\qquad\left. + \frac{1}{2}\sigma(t,X_t)^2\frac{\partial^2 f}{\partial x^2}(t,X_t)\right)\Delta t \\ &+ \left|\sigma(t,X_t)\frac{\partial f}{\partial x}(t,X_t)\right|\sqrt{\Delta t}\cdot\sqrt{\frac{8}{\pi}}\left(\frac{1+\sqrt{\beta(t,x)\,\Delta t}}{2} - \frac{1}{2}\right).\end{aligned}$$

Now, by the definition of the risk generator (4.4), we get the assertion. ☐

Remark 4.8. The drift (4.2) in Example 4.2 now turns out to be a specific case of the general relation revealed by (4.6), both reveal the same pattern: any risk averse assessment adds the additional drift term

$$\sqrt{\frac{2}{\pi}\beta(t,x)}\cdot\left|\sigma(x,t)\cdot\frac{\partial f(t,x)}{\partial x}\right|.$$

For the absolute value $|\cdot|$ in the expression, the additional drift term cannot be negative and always points in one direction, the direction of risk. This is in line with risk aversion, as deviations in the different directions are associated with profits and (for the other direction) losses. Further, the coefficient β models the amount of local risk aversion.

The behavior (4.6) has been found with other risk measures as well, for example for the Entropic Value-at-Risk, cf. Pichler and Schlotter [30]. For this reason, various results from the literature extend to the nested expectile.

5. The risk averse control problem

While the classical theory on dynamic optimization builds on the risk-neutral expectation (cf. Fleming and Soner [11]), we take risk into consideration to the optimal control problem and derive a risk averse variant of the Hamilton–Jacobi–Bellman equation. In what follows we derive the governing equations formally by adapting the presentation from Pichler and Schlotter [30] for expectiles.

Consider the stochastic differential equation

$$\mathrm{d}X_t^u=\mu\big(t,X_t^u,u(t,X_t^u)\big)\mathrm{d}t+\sigma\big(t,X_t^u,u(t,X_t^u)\big)\mathrm{d}W_t \tag{5.1}$$

driven by an adapted control policy $u(t,X_t)$, where u is a measurable function. It is the objective to minimize the risk-averse expectation of the accumulated costs,

$$\int_t^T c\big(s,X_s,u(s,X_s)\big)\mathrm{d}s+\Psi\big(X_T\big),$$

where $\Psi(\cdot)$ is a terminal cost. Recall that the nested expectiles accumulate costs and risk so that it is the objective to minimize the value function

$$V^u(t,x)\coloneqq\tilde{e}_{\beta(\cdot)}^{\sigma(X)}\left(\int_t^T\big(s,X_s^u,u(s,X_s^u)\big)\mathrm{d}s+\Psi\big(X_T^u\big)\middle|\,X_t^u=x\right)$$

among all policies $u\in\mathcal{U}$ chosen in a suitable set, where X_t^u solves the stochastic differential equation (5.1) for the policy u.

Proposition 5.1. *The value function*

$$V(t,x)\coloneqq\inf_{u(\cdot)\in\mathcal{U}}V^u(t,x),$$

solves the differential equation

$$\frac{\partial V}{\partial t}(t,x)=\mathcal{H}_\beta\left(t,x,\frac{\partial V}{\partial x},\frac{\partial^2 V}{\partial x^2}\right) \tag{5.2}$$

with terminal condition $V(T,x)=\Psi(x)$, where

$$\mathcal{H}_\beta(t,x,g,A) \coloneqq \sup_{u\in U}\Bigg\{-c(t,x,u)-g\cdot\mu(t,x,u) \\ -\frac{1}{2}A\,\sigma(t,x,u)^2-\sqrt{\frac{2}{\pi}\beta(t,x)}\cdot\big|g\cdot\sigma(t,x,u)\big|\Bigg\} \tag{5.3}$$

is the Hamiltonian, cf. Fleming and Soner [11, Section IV, (3.2)].

To accept the assertion recall that

$$\frac{1}{h}\tilde{e}^{\sigma(X)}_{\beta(\cdot)}\left(\int_t^{t+h} c\big(s,X_s^u,u(s,X_s^u)\big)\mathrm{d}s+V(t+h,X_{t+h})-V(t,x)\,\middle|\, X_t^u=x\right) \\ \xrightarrow[h\to 0]{} c(t,x,u)+\mathcal{G}_\beta V(t,x)$$

by the definition of the risk generator. While the left-hand side vanishes by the dynamic programming principle for the optimal policy, it follows for the right-hand side that

$$0=\inf_{u\in U} c(t,x,u)+\mathcal{G}_\beta V(t,x).$$

With Proposition 4.7, this leads to the equation (5.2) with Hamiltonian (5.3).

The fundamental equation (5.2) is the Hamilton–Jacobi–Bellman (HJB) partial differential equation. It is essential to observe that the HJB equation has the additional term

$$\sqrt{\frac{2}{\pi}\beta(t,x)}\cdot\left|\sigma(t,x,u)\frac{\partial V}{\partial x}\right|$$

involving the gradient; the total gradient in the Hamiltonian (5.2) thus comes with the coefficient

$$\mu(t,x,u)+\sqrt{\frac{2}{\pi}\beta(t,x)}\cdot\sigma(t,x,u)\cdot\operatorname{sign}\left(\sigma(t,x,u)\frac{\partial V}{\partial x}\right).$$

That is, risk aversion increases the trend μ by the amount $+\sqrt{\frac{2}{\pi}\beta(t,x)}\cdot\sigma(t,x,u)$, while letting the volatility σ of the process unaffected.

In typical situations, $\frac{\partial V}{\partial x}$ does not change its sign. For this reason, the classical theory on viscosity solutions on existence of solutions of (5.2) applies directly, without modifications. As well, explicit solutions of specific equations are known. In these situations, the explicit results can be adapted to the risk averse situation, cf. Pichler and Schlotter [29] for applications from financial mathematics.

6. Summary

This paper exploits the unique properties of expectiles in stochastic and in dynamic optimization. We start by giving tight comparisons with common risk measures first. Next, we define the conditional expectile. The conditional expectile can be nested to extend the scope of risk functionals (risk measures) to stochastic processes in discrete and in continuous time. For the random walk process or stochastic processes driven by a stochastic differential equation, explicit evaluations of the nested risk functional are available.

The risk generator is defined in analogy to the generator for stochastic processes. The risk generator involves an additional term which is caused by risk. With that, the risk generator is a non-linear differential operator. The aspect of risk augments the Hamiltonian via an additional term, which is responsible for risk only and the risk averse Hamilton–Jacobi–Bellman equations thus derive accordingly.

References

[1] C. Acerbi, *Spectral measures of risk: A coherent representation of subjective risk aversion*, Journal of Banking & Finance **26** (2002), 1505–1518.

[2] C. Acerbi and P. Simonetti, *Portfolio optimization with spectral measures of risk*, URL `http://arXiv.org/abs/cond-mat/0203607v1`, 2002.

[3] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, *Coherent Measures of Risk*, Mathematical Finance **9** (1999), 203–228.

[4] F. Bellini, B. Klar, A. Müller and E. Rosazza Gianin, *Generalized quantiles as risk measures*, Insurance: Mathematics and Economics **54** (2014), 41–48.

[5] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Springer US, 2004.

[6] P. Cheridito and M. Kupper, *Composition of time-consistent dynamic monetary risk measures in discrete time*, International Journal of Theoretical and Applied Finance **14** (2011), 137–162.

[7] R. Cont, R. Deguest and G. Scandolo, *Robustness and sensitivity analysis of risk measurement procedures*, SSRN Electronic Journal, doi: 10.2139/ssrn.1086698, 2008.

[8] F. Coquet, Y. Hu and S. Peng, *Filtration-consistent nonlinear expectations and related g-expectations*, Probability Theory and Related Fields **123** (2002), 1–27.

[9] D. Dentcheva and Y. Lin, *Bias reduction in sample-based optimization*, SIAM Journal on Optimization **32** (2021), 130–151.

[10] M. Farooq and I. Steinwart, *Learning rates for kernel-based expectile regression*, Machine Learning **108** (2018), 203–227.

[11] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, Springer New York, NY, second edition, 2006.

[12] H. Föllmer and A. Schied, *Stochastic Finance: An Introduction in Discrete Time*, de Gruyter Studies in Mathematics 27, Berlin, Boston: De Gruyter, 2004.

[13] T. Gneiting, *Making and evaluating point forecasts*, Journal of the American Statistical Association **106** (2011), 746–762.

[14] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York, 2002. doi: 10.1007/b98838.

[15] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics. Springer-Verlag New York, 1991.

[16] S. Kusuoka, *On law invariant coherent risk measures*, in: Advances in mathematical economics, volume 3, chapter 4, Springer, 2001, pp. 83–95.

[17] S. Lu and S. V. Pereverzev, *Regularization Theory for Ill-posed Problems*, DE GRUYTER, 2013.

[18] W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55(4):819–847, 1987. doi: 10.2307/1911031.

[19] B. Øksendal, *Stochastic Differential Equations*, Springer, 6th edition, 2003.

[20] E. Pardoux and S. G. Peng, *Adapted solution of a backward stochastic differential equation*, Systems & Control Letters **14** (1990), 55–61.

[21] S. Peng, *A generalized dynamic programming principle and Hamilton-Jacobi-Bellman equation*, Stochastics and Stochastic Reports **38** (1992), 119–134.

[22] S. Peng, *Nonlinear expectations, nonlinear evaluations and risk measures*, in: Lecture Notes in Mathematics, Springer Berlin Heidelberg, 2004, pp. 165–253.

[23] S. Peng, *Nonlinear Expectations and Stochastic Calculus under Uncertainty*, Springer Berlin, Heidelberg, 2010.

[24] G. Ch. Pflug, *Some remarks on the Value-at-Risk and the Conditional Value-at-Risk*, in: Probabilistic Constrained Optimization, S. Uryasev (ed.), volume 49, chapter 15. Springer US, 2000, pp. 272–281.

[25] A. B. Philpott and V. L. de Matos, *Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion*, European Journal of Operational Research **218** (2012), 470–483.

[26] A. B. Philpott, V. L. de Matos and E. Finardi, *On solving multistage stochastic programs with coherent risk measures*, Operations Research **61** (2013), 957–970.

[27] A. Pichler, *The natural Banach space for version independent risk measures*, Insurance: Mathematics and Economics **53** (2013), 405–415.

[28] A. Pichler and R. Schlotter, *Martingale characterizations of risk-averse stochastic optimization problems*, Mathematical Programming **181** (2019), 377–403.

[29] A. Pichler and R. Schlotter, *Quantification of risk in classical models of finance*, Quantitative Finance **22** (2021), 31–45,.

[30] A. Pichler and R. Schlotter, *Risk-averse optimal control in continuous time by nesting risk measures*, *Mathematics of Operations Research*, 2022.

[31] F. Riedel, *Dynamic coherent risk measures*, Stochastic Processes and their Applications **112** (2004), 185–200.

[32] L. Rüschendorf, *Mathematische Statistik*, Springer Berlin Heidelberg, 2014.

[33] A. Ruszczyński and A. Shapiro, *Conditional risk mappings*, Mathematics of Operations Research **31** (2006), 544–561.

[34] A. Ruszczyński and J. Yao, *A Risk-Averse Analog of the Hamilton–Jacobi–Bellman Equation*, chapter 62, Society for Industrial & Applied Mathematics (SIAM), Philadelphia, PA, 2015, pp. 462–468.

[35] A. Ruszczyński and J. Yao, *A dual method for evaluation of dynamic risk in diffusion processes*, ESAIM Control Optimisation and Calculus of Variations **26** (2020): 96.

[36] B. Schölkopf, R. Herbrich and A. J. Smola, *A generalized representer theorem*, in: Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2001, pp. 416–426.

[37] A. Shapiro, *Time consistency of dynamic risk measures*, Operations Research Letters **40** (2012), 436–439.

[38] L. Xin and A. Shapiro, *Bounds for nested law invariant coherent risk measures*, Operations Research Letters **40** (2012), 431–435.

[39] J. F. Ziegel, *Coherence and elicitability*, Mathematical Finance **26** (2014), 901–918.

7. Appendix

The function $x \mapsto (1-\alpha)\,\mathbb{E}(x-X)_+ - \alpha\,\mathbb{E}(X-x)_+$ has slope

$$\begin{aligned}(1-\alpha)\,P(X\le x)+\alpha\,P(X\ge x) &= \alpha+(1-2\alpha)P(X\le x)\\ &\ge \alpha P(X\le x)+(1-2\alpha)P(X\le x)\\ &=(1-\alpha)P(X\le x)\\ &\ge 0\end{aligned}$$

and is therefore strictly increasing for every $\alpha\in(0,1)$ in the support of X so that the expectile is unique. Further, the slope is so that the function is convex for $\alpha \le 1/2$ and concave for $\alpha \ge 1/2$.

Denote by x_α (y_α, resp.) the expectile for X (Y, resp.), i.e.,

$$\begin{aligned}\alpha\,\mathbb{E}(X-x_\alpha)_+ &= (1-\alpha)\,\mathbb{E}(x_\alpha-X)_+ \text{ and}\\ \alpha\,\mathbb{E}(Y-y_\alpha)_+ &= (1-\alpha)\,\mathbb{E}(y_\alpha-Y)_+.\end{aligned}$$

With $x_+-(-x)_+=x$ we have further

$$\begin{aligned}\Big(\alpha-\frac12\Big)\,\mathbb{E}(X-x_\alpha)_+ - \Big(\frac12-\alpha\Big)\,\mathbb{E}(x_\alpha-X)_+ &= \frac12\,\mathbb{E}(X-x_\alpha) \text{ and}\\ \Big(\alpha-\frac12\Big)\,\mathbb{E}(Y-y_\alpha)_+ - \Big(\frac12-\alpha\Big)\,\mathbb{E}(y_\alpha-Y)_+ &= \frac12\,\mathbb{E}(Y-y_\alpha).\end{aligned}$$

For $\alpha\ge 1/2$ we obtain by convexity of the function $x\mapsto x_+$ that

$$\begin{aligned}\Big(\alpha-\frac12\Big)\,\mathbb{E}(X+Y-x_\alpha-y_\alpha)_+ &\le \Big(\alpha-\frac12\Big)\,\mathbb{E}(X-x_\alpha)+\Big(\alpha-\frac12\Big)\,\mathbb{E}(Y-y_\alpha)\\ &= \frac12\,\mathbb{E}(X-x_\alpha)+\Big(\frac12-\alpha\Big)\,\mathbb{E}(x_\alpha-X)_+\\ &\quad+\frac12\,\mathbb{E}(Y-y_\alpha)+\Big(\frac12-\alpha\Big)\,\mathbb{E}(y_\alpha-Y)_+\\ &\le \frac12\,\mathbb{E}(X-x_\alpha)+\frac12\,\mathbb{E}(Y-y_\alpha)\\ &\quad+\Big(\frac12-\alpha\Big)\,\mathbb{E}(x_\alpha+y_\alpha-X-Y)_+.\end{aligned}$$

It follows that

$$\alpha\,\mathbb{E}(X+Y-x_\alpha-y_\alpha)_+ \le (1-\alpha)\,\mathbb{E}(x_\alpha+y_\alpha-X-Y)_+.$$

The assertion follows by monotonicity again.

Manuscript received February 28 2023
revised March 7 2023

R. Lakshmanan
University of Technology, Chemnitz, Faculty of Mathematics, 09126 Chemnitz, Germany
E-mail address: rajmadan.lakshmanan@math.tu-chemnitz.de

A. Pichler
University of Technology, Chemnitz, Faculty of Mathematics, 09126 Chemnitz, Germany
E-mail address: alois.pichler@math.tu-chemnitz.de