International Journal of Theoretical and Applied Finance
Vol. 24, No. 1 (2021) 2150003 (28 pages)
© World Scientific Publishing Company
DOI: 10.1142/S0219024921500035

World Scientific
www.worldscientific.com

TIME-INCONSISTENT MARKOVIAN CONTROL PROBLEMS UNDER MODEL UNCERTAINTY WITH APPLICATION TO THE MEAN-VARIANCE PORTFOLIO SELECTION

TOMASZ R. BIELECKI
Department of Applied Mathematics, Illinois Institute of Technology
10 W 32nd Street, Building RE, Room 220, Chicago, IL 60616, USA
tbielecki@iit.edu

TAO CHEN*
Department of Mathematics, University of Michigan
530 Church Street, East Hall, Room 2859, Ann Arbor, MI 48109, USA
chenta@umich.edu

IGOR CIALENCO
Department of Applied Mathematics, Illinois Institute of Technology
10 W 32nd Street, Building RE, Room 220, Chicago, IL 60616, USA
cialenco@iit.edu

Received 8 February 2020
Revised 8 September 2020
Accepted 23 December 2020
Published 3 March 2021

In this paper, we study a class of time-inconsistent terminal Markovian control problems in discrete time subject to model uncertainty. We combine the concept of the sub-game perfect strategies with the adaptive robust stochastic control method to tackle the theoretical aspects of the considered stochastic control problem. Consequently, as an important application of the theoretical results and by applying a machine learning algorithm we solve numerically the mean-variance portfolio selection problem under the model uncertainty.

Keywords: Adaptive robust control; model uncertainty; stochastic control; adaptive robust dynamic programming; recursive confidence regions; time-inconsistent Markovian control problem; optimal portfolio allocation; mean-variance portfolio selection; terminal criteria; machine learning; Gaussian surrogate processes; regression Monte Carlo.

We dedicate this paper to the memory of Tomas Björk.

1. Introduction

The main goal of this study is to develop a methodology to solve efficiently some *time-inconsistent* Markovian control problems subject to *model uncertainty*

*Corresponding author.

in a discrete time setup. The proposed approach hinges on the following main building concepts: first, incorporating model uncertainty through the *adaptive robust* paradigm introduced in Bielecki *et al.* (2019); second, dealing with time-inconsistency of the stochastic control problem at hand by exploiting the concept of *sub-game perfect* strategies as studied in Björk & Murgoci (2014); third, developing efficient numerical solutions for the obtained Bellman equations by adopting the *machine learning* techniques proposed in Chen & Ludkovski (2019).

There exists a significant body of work on incorporating model uncertainty (or model misspecification) in stochastic control problems, and among some of the well-known and prominent methods we would mention the robust control approach (Gilboa & Schmeidler 1989, Hansen *et al.* 2006, Hansen & Sargent 2008), adaptive control (Chen & Guo 1991, Duncan *et al.* 2001, 2006, Kumar & Varaiya 2015), and Bayesian adaptive control (Kumar & Varaiya 2015). We refer the reader to Bielecki *et al.* (2019) for a relatively comprehensive literature review on this subject and their connection to the adaptive robust methodology used in this paper and originally introduced in Bielecki *et al.* (2019). The adaptive robust methodology of Bielecki *et al.* (2019) is an approach that solves (time-consistent) Markovian control problems in discrete time subject to model uncertainty. The core of this methodology was to combine a recursive learning mechanism about the unknown model, with the Markovian dynamics of that model and with the time-consistent nature of the control problem studied therein, which allowed to derive an adequate system of recursive dynamic programming equations, which where dubbed the adaptive robust Bellman equations that gave a solution to the original control problem.

In all the above-mentioned methods, inherently the stochastic problems are (strongly) time-consistent in the sense that the dynamic programming principle holds true. For an overview of the time consistency in decision making, cf. Bielecki *et al.* (2018, 2017b). While lack of (strong) time consistency in decision making is not necessarily an unacceptable feature, from stochastic control point of view it may lead to undesirable properties that, in particular, may lack adequate numerically tractable solutions. A good body of literature have been dedicated to time-inconsistent stochastic control problems have been emerged in the recent years, primarily for continuous time setup. We refer the reader to He (2018, Chap. 2) and Shi & Cui (2017) for a comprehensive literature review of time-inconsistent stochastic control problems in discrete time. Broadly speaking, there are three avenues that researchers followed in dealing with time-inconsistent Markovian control problems (primarily in discrete time) when the underlying model is fully known:

(1) The *pre-commitment approach* emphasizes the global optimality, namely the controller optimizes the expected objective functional at the initial time point and sticks to the resulting strategy through the whole time period. In general, such strategy will be time-inconsistent if it is not revised by the controller afterwards. In the context of optimal portfolio selection, the authors of Li & Ng

(2000) and Li & Zhou (2000) introduced an embedding technique to obtain a pre-committed solution of the dynamic mean-variance optimization problem.

(2) The *sub-game perfect* approach, sometimes also called consistent planning approach, is based on ideas rooted in game theory, where the time-consistent strategy is derived by assuming that the investor is playing an optimization game with future-self. This approach, originated in Strotz (1955), Goldman (1980), and systemically studied in Ekeland & Lazrak (2006), Ekeland & Pirvu (2008), Ekeland & Lazrak (2010), Hu *et al.* (2012), Hernández & Possamai (2020) (in continuous time) leads to a specific notion of optimality (sub-game perfection), which can be characterized in terms of respective dynamic programming equations. In this regard, Basak & Chabakauri (2010), Björk & Murgoci (2014) were the first to apply and extend this approach to the mean-variance problem. Some more different examples are investigated in Björk *et al.* (2014), Bannister *et al.* (2016).

(3) By *modifying the criteria* as time evolves such that the dynamic programming principle holds that has been studied, in various forms in Bouchard *et al.* (2010), Cui *et al.* (2012), Karnam *et al.* (2017), Feinstein & Rudloff (2019), Kováčová & Rudloff (2019).

It should be mentioned that only a selected number of publications on time-inconsistency of stochastic control problems in continuous time setup were mentioned here. Generally speaking, there is no one right method in addressing the time-inconsistency, and each of these three approaches has its advantages and drawbacks.

In this work, we are focusing on the sub-game perfect approach that is appropriately formulated for the Markovian control problem with model uncertainty. As an important application of the proposed general theory, we consider the mean-variance portfolio selection problem under model uncertainty. Besides being an important contribution of our paper, arguably, the classical mean-variance portfolio optimization methodology is one of the most popular portfolio selection methodology among managers of financial portfolios. Notably, majority of the above cited literature is devoted to the mean-variance optimization problem.

It is well-documented that solving numerically stochastic control problems subject to model uncertainty is a challenging task, and classical numerical methods cannot be successfully applied even to the simplest problems. In Chen & Ludkovski (2019) the authors introduced a method, rooted in the machine learning methodology, to deal with such problems in the context of an adaptive robust, time-consistent, stochastic control problem. In the present work, we apply a similar computational approach for solving the aforementioned mean-variance problem.

The paper is organized as follows. In Sec. 2, we formulate the time-inconsistent Markovian control problem subject to model uncertainty, as well as, the corresponding time-inconsistent adaptive robust control problem (see Sec. 2.1). The main theoretical developments of this work are presented in Sec. 3. Specifically, in this section, we propose and analyze the time consistent sub-game perfect approach to deal with the adaptive robust control problem of Sec. 2.1. We derive the Bellman equations for the sub-game perfect strategies; see Theorem 3.1. In Sec. 3.1, we study the existence of sub-game perfect strategies. An illustrative example of our theoretical results that is rooted in the classical Markowitz's mean-variance portfolio theory, is presented in Sec. 4. Using machine learning methods, in Sec. 5, we provide numerical solutions of the example presented in Sec. 4. Finally, in Sec. 6, we outline some possible research directions and open problems.

2. Time-Inconsistent Markovian Control Problem with Model Uncertainty

In this section, we state the underlying time-inconsistent stochastic control problem. Let $(\Omega, \mathscr{F})$ be a measurable space, $T \in \mathbb{N}$ be a fixed time horizon, and let us denote by $\mathcal{T} := \{0, 1, 2, \dots, T\}$ and $\mathcal{T}' := \{0, 1, 2, \dots, T-1\}$. In what follows, we implicitly assume that all considered probabilities are defined on $(\Omega, \mathscr{F})$, and as usual $\mathbb{E}_{\mathbb{P}}$ will denote the expectation under a probability measure $\mathbb{P}$. We let $\mathbf{\Theta} \subset \mathbb{R}^d$ be a nonempty Borel set,[a] which will play the role of the known parameter space. We consider a random process $Z = \{Z_t,\ t \in \mathcal{T}\}$ on $(\Omega, \mathscr{F})$ taking values in $\mathbb{R}^m$, and we denote by $\mathbb{F} = (\mathscr{F}_t, t \in \mathcal{T})$ its natural filtration. We postulate that this process is observed by the controller, but the true law of Z is unknown to the controller and assumed to be generated by a probability measure belonging to a (known) parameterized family of probabilities $\mathbf{P}(\mathbf{\Theta}) = \{\mathbb{P}_\theta, \theta \in \mathbf{\Theta}\}$. For simplicity, we will write $\mathbb{E}_\theta$ instead of $\mathbb{E}_{\mathbb{P}_\theta}$. We denote by $\mathbb{P}_{\theta^*}$ the measure generating the true law of Z, and thus $\theta^* \in \mathbf{\Theta}$ is the unknown true parameter. We will assume that $\mathbf{\Theta} \neq \{\theta^*\}$, namely we consider the case of a nontrivial model uncertainty.

We let $A \subset \mathbb{R}^k$ to denote the set of control values. For technical reasons, such as the existence of measurable selectors, we assume that A is finite, although tentatively all stated results can be extended to A being a compact. An admissible control process φ is an $\mathbb{F}$-adapted process, taking values in A and we will denote by $\mathcal{A}$ the set of all admissible control processes.

We consider a discrete time controlled dynamical system with the state process X taking values in $\mathbb{R}^n$ and having the dynamics

$$X_{t+1} = S(X_t, \varphi_t, Z_{t+1}), \quad t \in \mathcal{T}', \quad X_0 = x_0 \in \mathbb{R}^n, \tag{2.1}$$

with $S : \mathbb{R}^n \times A \times \mathbb{R}^m \to \mathbb{R}^n$ a measurable mapping, and φ a control process. We limit ourselves to the class of Markovian strategies only.

[a] In general, the parameter space may be infinite-dimensional, consisting for example of dynamic factors, such as deterministic functions of time or hidden Markov chains. In this study, for simplicity, we chose the parameter space to be a subset of $\mathbb{R}^d$. In most applications, in order to avoid problems with constrained estimation, the parameter space is taken to be equal to the maximal relevant subset of $\mathbb{R}^d$.

The underlying uncertain control problem is

$$\sup_{\varphi \in \mathcal{A}} (\mathbb{E}_{\theta^*}[F(X_T)] + G(\mathbb{E}_{\theta^*}[X_T])), \tag{2.2}$$

where $F, G : \mathbb{R}^n \to \mathbb{R}$ are some given Borel measurable functions. Throughout, we will assume that all expectations are well-defined.

Remark 2.1. Generally speaking, all results obtained in this paper can be extended to a more general control problem of the form

$$\sup_{\varphi \in \mathcal{A}} (\mathbb{E}_{\theta^*}[F(X_T)] + G(\mathbb{E}_{\theta^*}[H(X_T)])),$$

where $H : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R} \to \mathbb{R}$. For example, if H is a bijection, then putting $Y_t = H(X_t)$ reduces the problem to the case (2.2). Otherwise, one can increase the dimension of the state process and replace X_t with $(X_t, H(X_t))$. With slight loss of generality, and gain in readability, we opted to focus on (2.2).

Example 2.1 (Mean-variance). A typical, and also practically important, example is the mean-variance (MV) control problem, where X is a scalar valued process, and $F(x) = x - \gamma x^2$, $G(x) = \gamma x^2$, for some fixed weight $\gamma > 0$. Sometimes γ is refereed to as risk aversion parameter. In this case, the stochastic control problem at hand becomes

$$\sup_{\varphi \in \mathcal{A}} (\mathbb{E}_{\theta^*}(X_T) - \gamma \operatorname{Var}_{\theta^*}(X_T)). \tag{2.3}$$

In Sec. 4, we will return to this example in the context of portfolio selection problem.

Problem (2.2), in general, is time-inconsistent in the sense that the dynamic programming principle fails; see, for instance, the discussion in Shi & Cui (2017) and Björk & Murgoci (2014). Additionally, the parameter θ^* is not known to the controller, and thus problem (2.2) cannot be solved as is. In the rest of the paper, we address these two issues via *time consistent adaptive robust sub-game perfect approach*. To achieve this, we first formally formulate an adaptive robust control problem corresponding to (2.2) (see Sec. 2.1). The solution to this problem, by analogy, can be called adaptive robust pre-commitment strategy. Hence, by similar arguments to the case without model uncertainty this problem is time-inconsistent. To overcome this, we proceed with time consistent adaptive robust sub-game approach (see Sec. 3). Finally, the adaptive robust method requires handling a double (sup-inf) optimization problem at each time instance, yielding Bellman equations that are intrinsically multi-dimensional. For these reasons, computing the solutions of the corresponding control problem is a nontrivial task, and we address this issue using machine-learning technique for the MV problem (see Sec. 5).

2.1. *Time-inconsistent adaptive robust control problem*

We mainly follow here the developments presented in Bielecki *et al.* (2019), with the key difference that herein the Markovian controls problems are inherently time-inconsistent, and for this reason we only use Markovian strategies.

A central building block in the adaptive robust approach is the recursive construction of confidence regions for the unknown model parameter θ^*. We refer to Bielecki *et al.* (2017a) for a general study of recursive constructions of (approximate) confidence regions for time homogeneous Markov chains, while in Sec. 4 we provide a specific such recursive construction corresponding to the optimal portfolio selection problem under uncertainty. Here, we just postulate that the recursive algorithm for building confidence regions uses an $\mathbb{R}^d$-valued and observed process, say $(C_t,\ t \in \mathcal{T}')$, satisfying the following abstract dynamics:

$$C_{t+1} = R(t, C_t, Z_{t+1}), \quad t \in \mathcal{T}',\ \ C_0 = c_0, \tag{2.4}$$

where $R\colon \mathcal{T}' \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^d$ is a deterministic measurable function, and $c_0 \in \mathbf{\Theta}$. Note that, given our assumptions about process Z, the process C is $\mathbb{F}$-adapted. Usually C_t is taken to be a consistent estimator of θ^*.

Now, we fix a confidence level $\alpha \in (0,1)$, and for each time $t \in \mathcal{T}'$, we assume that an $(1-\alpha)$-confidence region, say $\Theta_t \subset \mathbb{R}^d$, for θ^*, can be represented as

$$\Theta_t = \tau(t, C_t), \tag{2.5}$$

where, for each $t \in \mathcal{T}'$, $\tau(t,\cdot)\colon \mathbb{R}^d \to 2^{\Theta}$ is a deterministic set valued function, with 2^{Θ} denoting the set of all subsets of Θ. Note that in view of (2.4) the construction of confidence regions given in (2.5) is indeed recursive. In our construction of confidence regions, the mapping $\tau(t,\cdot)$ will be a measurable set valued function, with compact values. The important property of the recursive confidence regions constructed in Sec. 4 is that $\lim_{t\to\infty} \Theta_t = \{\theta^*\}$, where the convergence is understood $\mathbb{P}_{\theta^*}$ almost surely, and the limit is in the Hausdorff metric. This is not always the case, and in Bielecki *et al.* (2017a) it is shown that under some general assumptions the convergence holds in probability. The sequence $\Theta_t,\ t \in \mathcal{T}'$ represents learning about θ^* based on the observed history up to time $t \in \mathcal{T}$. We introduce the augmented state process $Y_t = (X_t, C_t),\ t \in \mathcal{T}$, and the augmented state space

$$E_Y = \mathbb{R}^n \times \mathbb{R}^d,$$

and we denote by $\mathcal{E}_Y$ the collection of Borel measurable sets in E_Y. In view of the above, the process Y has the following dynamics:

$$Y_{t+1} = \mathbf{G}(t, Y_t, \varphi_t, Z_{t+1}), \quad t \in \mathcal{T}',$$

where $\mathbf{G}$ is the mapping $\mathbf{G}\colon \mathcal{T}' \times E_Y \times A \times \mathbb{R}^m \to E_Y$ defined as

$$\mathbf{G}(t, y, a, z) = (S(x, a, z), R(t, c, z)), \tag{2.6}$$

where $y = (x, c) \in E_Y$.

A control process $\varphi = (\varphi_t,\ t \in \mathcal{T}')$ is called *Markovian control process* if (with a slight abuse of notation)

$$\varphi_t = \varphi_t(Y_t),$$

where (on the right-hand side) $\varphi_t\colon E_Y \to A$, is a measurable mapping.

From now on, we constrain the set $\mathcal{A}$ of admissible control processes to the set of Markovian control processes. For any admissible control process φ and for any $t \in \mathcal{T}'$, we denote by $\varphi^t = (\varphi_k,\ k = t, \dots, T-1)$ the "t-tail" of φ; in particular, $\varphi^0 = \varphi$. Accordingly, we denote by $\mathcal{A}^t$ the collection of t-tails φ^t; thus, $\mathcal{A}^0 = \mathcal{A}$.

Let $\breve{\psi}_t : E_Y \to \Theta$ be a measurable mapping (Knightian selector), and let us denote by $\breve{\psi} = (\breve{\psi}_t,\ t \in \mathcal{T}')$ the sequence of such mappings, and by $\breve{\psi}^t = (\breve{\psi}_s,\ s = t, \dots, T-1)$ the t-tail of the sequence ψ. The set of all sequences $\breve{\psi}$, and, respectively, $\breve{\psi}^t$, will be denoted by $\breve{\Psi}$ and $\breve{\Psi}^t$, respectively.

Similarly, we consider the measurable mappings $\psi_t : E_Y \to \Theta$, such that $\psi_t(x, c) \in \tau(t, c)$. This, in particular, implies that $\psi_t(X_t, C_t) \in \Theta_t$. Correspondingly, we define the set of all such selectors as Ψ_t, the set of all sequences of such mappings by $\Psi = \Psi_0 \times \cdots \times \Psi_{T-1}$, and the set of t-tails by Ψ^t. Clearly, $\Psi \subset \breve{\Psi}$. Moreover, $\psi^t \in \Psi^t$ if and only if $\psi^t \in \breve{\Psi}^t$ and $\psi_s(y_s) \in \tau(s, c_s),\ s = t, \dots, T-1$.

Next, for each $(t, y, a, \theta) \in \mathcal{T}' \times E_Y \times A \times \Theta$, we define a probability measure on $\mathcal{E}_Y$, given by

$$\begin{aligned} Q(B \mid t, y, a, \theta) &= \mathbb{P}_\theta(Z_{t+1} \in \{z : \mathbf{G}(t, y, a, z) \in B\}) \\ &= \mathbb{P}_\theta(\mathbf{G}(t, y, a, Z_{t+1}) \in B), \end{aligned} \tag{2.7}$$

for any $B \in \mathcal{E}_Y$. Throughout we assume that:

(A1) For every $t \in \mathcal{T}$ and every $a \in A$, the measure $Q(dy' \mid t, y, a, \theta)$ is a Borel measurable stochastic kernel with respect to (y, θ).

This assumption will be satisfied in the context of the mean-variance problem discussed in Sec. 4.

Using Ionescu–Tulcea theorem (cf. Bäuerle & Rieder (2011, Appendix B)), for every $t \in \mathcal{T}'$, control process $\varphi \in \mathcal{A}^t$, $\psi \in \mathbf{\Psi}^t$, time t state $y_t \in E_Y$, we define probability measure $\mathbb{Q}^{\varphi,\psi}_{y_t,t}$ on the concatenated canonical space $\mathsf{X}_{s=t+1}^{T} E_Y$ as follows:

$$\begin{aligned} &\mathbb{Q}^{\varphi,\psi}_{y_t,t}(B_{t+1} \times \cdots \times B_T) \\ &\qquad = \int_{B_{t+1}} \cdots \int_{B_T} \prod_{u=t+1}^{T} Q(dy_u \mid u-1, y_{u-1}, \varphi_{u-1}(y_{u-1}), \psi_{u-1}(y_{u-1})). \end{aligned} \tag{2.8}$$

Correspondingly, we define the family of probability measures $\mathcal{Q}^{\varphi}_{y_t,t} = \{\mathbb{Q}^{\varphi,\psi}_{y_t,t},\ \psi \in \Psi^t\}$. Here, and everywhere below, to simplify the notations, we simply write $\varphi \in \mathcal{A}^t$, instead of $\varphi^t \in \mathcal{A}^t$, and implicitly assume that the processes have the correct tail dimension.

Analogously we define the set $\breve{\mathcal{Q}}^{\varphi}_{y_0,0} = \{\mathbb{Q}^{\varphi,\psi}_{y_0,0},\ \psi \in \breve{\Psi}\}$. In Remark 2.2 we provide a game oriented interpretation of mappings ψ and $\breve{\psi}$, as strategies played by the nature seen as a Knightian adversary of the controller.

The *time inconsistent strong robust control* problem is then given as

$$\sup_{\varphi\in\mathcal{A}}\ \inf_{\mathbb{Q}\in\mathcal{Q}^{\varphi,\breve{\Psi}}_{y_0,0}}(\mathbb{E}_{\mathbb{Q}}(F(X_T))+G(\mathbb{E}_{\mathbb{Q}}(X_T))). \tag{2.9}$$

The corresponding *time inconsistent adaptive robust control* problem is

$$\sup_{\varphi\in\mathcal{A}}\ \inf_{\mathbb{Q}\in\mathcal{Q}^{\varphi,\Psi}_{y_0,0}}(\mathbb{E}_{\mathbb{Q}}(F(X_T))+G(\mathbb{E}_{\mathbb{Q}}(X_T))). \tag{2.10}$$

Remark 2.2. The strong robust control problem is essentially a game problem between the controller and his/her Knightian adversary — the nature, who may keep changing the dynamics of the underlying stochastic system over time. In this game, the nature is not restricted in its choices of model dynamics, except for the requirement that $\breve{\psi}_t(Y_t)\in\Theta$, and each choice is potentially based on the value of Y_t. On the other hand, the adaptive robust control problem is a game problem between the controller and the nature, who, as in the case of strong robust control problem, may keep changing the dynamics of the underlying stochastic system over time. However, in the latter game, the nature is restricted in its choices of model dynamics to the effect that $\psi_t(Y_t)\in\tau(t,C_t)$.

3. Time Consistent Adaptive Robust Sub-game Approach

In general, the dynamic programming principle proved in Bielecki *et al.* (2019, Sec. 2.2.1), does not apply to problem (2.10), which is the nature of the time inconsistency of this problem. In particular, the backward induction procedure (or dynamic programming principle) from Bielecki *et al.* (2019, Sec. 2.2.1), cannot be used, in general, to solve problem (2.10). Consequently, with no dynamic principle at hand, practically speaking such problems cannot be solved numerically, especially when the number of steps is large. Thus, instead of dealing with problem (2.10) as is, we adopt the concept of sub-game perfect controls of Björk & Murgoci (2014) to our setup, which we will transform (2.10) into a time consistent problem that can be solved by using backward induction.

For convenience, for $\varphi\in\mathcal{A}^t$, $y=(x,c)$, and for $\psi\in\mathbf{\Psi}^{t+1}$ we define

$$\mathcal{Q}^{\varphi,\psi+}_{y,t}:=\{\mathbb{Q}^{\varphi^t,(\theta;\psi^{t+1})}_{y,t},\ \theta\in\tau(t,c)\},$$

where $(\theta;\psi^{t+1}):=(\theta,\psi_{t+1},\dots,\psi_T)$. In what follows, we will use similar notation $(a;b)$ for concatenation of vectors a,b.

We define the time-t time inconsistent criterion as

$$\begin{gathered}J_t(y,\varphi^t,\psi^t):=\mathbb{E}_{\mathbb{Q}^{\varphi,\psi}_{y,t}}(F(X_T))+G(\mathbb{E}_{\mathbb{Q}^{\varphi,\psi}_{y,t}}(X_T)),\\ y=(x,c)\in E_Y,\quad t\in\mathcal{T}',\end{gathered} \tag{3.1}$$

and let

$$J_T(y)=F(x)+G(x),\quad y=(x,c)\in E_Y.$$

Definition 3.1. The pair of strategies $(\widetilde{\varphi}, \widetilde{\psi})$ is called *sub-game perfect* if

$$\max_{a\in A} \inf_{\theta\in\tau(t,c)} J_t(y,(a;\widetilde{\varphi}^{t+1}),(\theta;\widetilde{\psi}^{t+1})) = J_t(y,\widetilde{\varphi}^t,\widetilde{\psi}^t), \tag{3.2}$$

for any $y=(x,c)\in E_Y$, and $t\in\mathcal{T}'$, with the convention that $(a;\widetilde{\varphi}^T)=a$, and $(\theta;\widetilde{\psi}^T)=\theta$.

Remark 3.1. Similarly, one can define the sub-game perfect strategies for the strong robust case, by replacing set $\tau(t,c)$ in Definition 3.1 with Θ. Due to the imposed model assumptions, in particular Assumption (A2) below, all obtained results hold true by similarity in the strong robust case.

Please note that Definition 3.1 is not a definition of equilibrium strategies in any game-theoretic sense, as we do not study any sort of game-theoretic equilibrium per se, even though we may think of the controller and the nature as two players. This definition is inspired by the concept of the sub-game perfect Nash equilibrium that was applied in the context of time-inconsistent control problems, as presented for example in Björk & Murgoci (2014, Definition 2.2). However, sub-game perfect Nash equilibrium is not really the same as the classical game-theoretic concept of Nash equilibrium, and should not be interpreted as such.

The idea of Definition 3.1 is to view the problem in the embedded sequential optimization terms: at each point of time there are two decision makers who chose decisions impacting evolution of the system: the controller and the nature. The two decision makers acting at time $t\in\mathcal{T}'$ know that all the pairs of decision makers coming after them will use the control $(\widetilde{\varphi}^{t+1},\widetilde{\psi}^{t+1})$. Given such knowledge, the two time-t decision makers optimize over A and $\tau(t,c)$, respectively, so generate their decisions. Following up on item (2) from the Introduction, we might interpret this as the game played by the two players, the controller and the nature, at time-t, with their future selves.

Throughout, we will use the notation

$$V_t(y) := J_t(y,\widetilde{\varphi}^t,\widetilde{\psi}^t),$$

for $y=(x,c)\in E_Y,\ t\in\mathcal{T}'$ and some $(\widetilde{\varphi}^t,\widetilde{\psi}^t)$ pair of sub-game perfect strategy. We remark that due to the Definition 3.1 and Theorem 3.1 proved below, the value of $V_t(y)$ does not depend on the choice of the sub-game perfect strategy. We will also show that there exists a sub-game perfect pair of strategies to the original stochastic control problem (2.10), and thus $V_t(y)$ is well-defined.

Note that, for any $\varphi^t\in\mathcal{A}^t,\ \psi^{t+1}\in\mathbf{\Psi}^{t+1}$ we have that

$$\inf_{\theta\in\tau(t,c)} J_t(y,\varphi^t,(\theta;\psi^{t+1})) = \inf_{\mathbb{Q}\in\mathcal{Q}_{y,t}^{\varphi,\psi+}} (\mathbb{E}_{\mathbb{Q}}(F(X_T))+G(\mathbb{E}_{\mathbb{Q}}(X_T))).$$

For $y=(x,c)\in E_Y$, $t\in\mathcal{T}'$, $\varphi\in A^t$, and $\psi\in\mathbf{\Psi}^t$ we denote

$$f_t^{\varphi,\psi}(y) := \mathbb{E}_{\mathbb{Q}_{y,t}^{\varphi,\psi}} F(X_T), \quad g_t^{\varphi,\psi}(y) := \mathbb{E}_{\mathbb{Q}_{y,t}^{\varphi,\psi}}(X_T),$$

$$f_T^{\varphi,\psi}(y) := F(x), \quad g_T^{\varphi,\psi}(y) := x,$$

and for $y=(x,c)\in E_Y$, and $t\in\mathcal{T}$, we put $\widetilde{g}_t(y)=g_t^{\widetilde{\varphi},\widetilde{\psi}}(y)$. In addition, we define the following integral operator:

$$Q^{a,\theta}_{y,t}f := \int_{E_Y} f(y')Q(dy'|t,y,a,\theta),$$

where Q is given by (2.7).

Clearly,

$$f^{\varphi,\psi}_t(y)=Q^{\varphi_t,\psi_t}_{y,t}f^{\varphi,\psi}_{t+1},\quad g^{\varphi,\psi}_t(y)=Q^{\varphi_t,\psi_t}_{y,t}g^{\varphi,\psi}_{t+1}.$$

With this at hand, we have the following counterpart of Björk & Murgoci (2014, Lemma 3.2).

Lemma 3.1. *For $y=(x,c)\in E_Y,\ a\in A,\theta\in\tau(t,c)$, and $t\in\mathcal{T}'$, the following identity holds*:

$$\begin{aligned}J_t(y,(a;\varphi^{t+1}),(\theta;\psi^{t+1})) &= Q^{a,\theta}_{y,t}J_{t+1}(\cdot,\varphi^{t+1},\psi^{t+1})\\ &\quad -[Q^{a,\theta}_{y,t}(G\circ g^{\varphi,\psi}_{t+1})-G(Q^{a,\theta}_{y,t}g^{\varphi,\psi}_{t+1})],\end{aligned}$$

with the convention that

$$J_T(y,\varphi^T,\psi^T)=J_T(y)=F(x)+G(x),\quad y=(x,c)\in E_Y.$$

Proof. We have

$$Q^{a,\theta}_{y,t}J_{t+1}(\cdot,\varphi,\psi)=Q^{a,\theta}_{y,t}f^{\varphi,\psi}_{t+1}+Q^{a,\theta}_{y,t}(G\circ g^{\varphi,\psi}_{t+1})$$

and

$$\begin{aligned}J_t(y,(a;\varphi^{t+1}),(\theta;\psi^{t+1})) &= f_t^{(a;\varphi^{t+1}),(\theta;\psi^{t+1})}(y)+G(g_t^{(a;\varphi^{t+1}),(\theta;\psi^{t+1})}(y))\\ &= Q^{a,\theta}_{y,t}f^{\varphi,\psi}_{t+1}+G(Q^{a,\theta}_{y,t}g^{\varphi,\psi}_{t+1}).\end{aligned}$$

This proves the result. □

Now we are in the position to prove the first main result about the backward recursion for V_t in terms of the corresponding Bellman equations.

Theorem 3.1. *A pair $(\widetilde{\varphi},\widetilde{\psi})$ of Markovian strategies is a pair of sub-game perfect strategies if and only if*

$$V_t(y)=\max_{a\in A}\inf_{\theta\in\tau(t,c)}(Q^{a,\theta}_{y,t}V_{t+1}-[Q^{a,\theta}_{y,t}(G\circ\widetilde{g}_{t+1})-G(Q^{a,\theta}_{y,t}\widetilde{g}_{t+1})]), \tag{3.3}$$

$$\widetilde{g}_t(y)=Q^{\widetilde{\varphi}_t,\widetilde{\psi}_t}_{y,t}\widetilde{g}_{t+1}, \tag{3.4}$$

$$V_T(y)=F(x)+G(x), \tag{3.5}$$

for any $y=(x,c)\in E_Y$, and $t\in\mathcal{T}'$.

Proof. ($\Rightarrow$) In view of Lemma 3.1, we have that

$$\begin{aligned}J_t(y,(a;\widetilde{\varphi}^{t+1}),(\theta;\widetilde{\psi}^{t+1})) &= Q^{a,\theta}_{y,t}J_{t+1}(\,\cdot\,,\widetilde{\varphi}^{t+1},\widetilde{\psi}^{t+1}) - [Q^{a,\theta}_{y,t}(G\circ g^{\widetilde{\varphi},\widetilde{\psi}}_{t+1})\\ &\quad - G(Q^{a,\theta}_{y,t}g^{\widetilde{\varphi},\widetilde{\psi}}_{t+1})]\\ &= Q^{a,\theta}_{y,t}V_{t+1} - [Q^{a,\theta}_{y,t}(G\circ\widetilde{g}_{t+1}) - G(Q^{a,\theta}_{y,t}\widetilde{g}_{t+1})],\\ &\quad y=(x,c)\in E_Y,\quad a\in A,\ \ \theta\in\tau(t,c),\ \ t\in\mathcal{T}'.\end{aligned}$$

Thus,

$$\begin{aligned}V_t(y) &= \max_{a\in A}\inf_{\theta\in\tau(t,c)} J_t(y,(a;\widetilde{\varphi}^{t+1}),(\theta;\widetilde{\psi}^{t+1}))\\ &= \max_{a\in A}\inf_{\theta\in\tau(t,c)}(Q^{a,\theta}_{y,t}V_{t+1} - [Q^{a,\theta}_{y,t}(G\circ\widetilde{g}_{t+1}) - G(Q^{a,\theta}_{y,t}\widetilde{g}_{t+1})]),\\ &\quad y=(x,c)\in E_Y,\quad a\in A,\ \ \theta\in\tau(t,c),\ \ t\in\mathcal{T}'.\end{aligned}$$

($\Leftarrow$) We start with $t=T-1$. Note that $V_T\equiv J_T$. Thus, for $y=(x,c)\in E_Y$, we have

$$\begin{aligned}V_{T-1}(y) &= \max_{a\in A}\inf_{\theta\in\tau(T-1,c)}(Q^{a,\theta}_{y,T-1}V_T - [Q^{a,\theta}_{y,T-1}(G\circ\check{g}_T) - G(Q^{a,\theta}_{y,T-1}\check{g}_T)])\\ &= Q^{\check{\varphi}_{T-1}(y),\check{\psi}_{T-1}(y)}_{y,T-1}V_T - [Q^{\check{\varphi}_{T-1}(y),\check{\psi}_{T-1}(y)}_{y,T-1}(G\circ\check{g}_T)\\ &\quad - G(Q^{\check{\varphi}_{T-1}(y),\check{\psi}_{T-1}(y)}_{y,T-1}\check{g}_T)].\end{aligned}$$

Using Lemma 3.1, we deduce

$$V_{T-1}(y) = \max_{a\in A}\inf_{\theta\in\tau(T-1,c)} J_{T-1}(y,a,\theta) \tag{3.6}$$

$$= J_{T-1}(y,\check{\varphi}_{T-1},\check{\psi}_{T-1}) = J_{T-1}(y,\check{\varphi}^{T-1},\check{\psi}^{T-1}), \tag{3.7}$$

which verifies (3.2) for $t=T-1$ and $\widetilde{\varphi}_{T-1}\equiv\check{\varphi}_{T-1}$ and $\widetilde{\psi}_{T-1}\equiv\check{\psi}_{T-1}$.

Next, we let $t=T-2$. Then, using (3.6) we produce

$$\begin{aligned}V_{T-2}(y) &= \max_{a\in A}\inf_{\theta\in\tau(T-2,c)}(Q^{a,\theta}_{y,T-2}V_{T-1} - [Q^{a,\theta}_{y,T-2}(G\circ\check{g}_{T-1}) - G(Q^{a,\theta}_{y,T-2}\check{g}_{T-1})])\\ &= Q^{\check{\varphi}_{T-2}(y),\check{\psi}_{T-2}(y)}_{y,T-2}V_{T-1} - [Q^{\check{\varphi}_{T-2}(y),\check{\psi}_{T-2}(y)}_{y,T-2}(G\circ\check{g}_{T-1})\\ &\quad - G(Q^{\check{\varphi}_{T-2}(y),\check{\psi}_{T-2}(y)}_{y,T-2}\check{g}_{T-1})].\end{aligned}$$

Again, in view of Lemma 3.1 we obtain

$$V_{T-2}(y) = \max_{a\in A}\inf_{\theta\in\tau(T-2,c)} J_{T-2}(y,(a;\check{\varphi}_{T-1}),(\theta;\check{\psi}_{T-2})) \tag{3.8}$$

$$= J_{T-2}(y,(\check{\varphi}_{T-2};\check{\varphi}_{T-1}),(\check{\psi}_{T-2};\check{\psi}_{T-1})) = J_{T-1}(y,\check{\varphi}^{T-2},\check{\psi}^{T-2}), \tag{3.9}$$

which verifies (3.2) for $t=T-2$ and $\widetilde{\varphi}^{T-2}\equiv\check{\varphi}^{T-2}$ and $\widetilde{\psi}^{T-2}\equiv\check{\psi}^{T-2}$. Proceeding in the analogous way for $t=T-3,\dots,0$ we complete the proof. □

3.1. *Existence of sub-game perfect strategies*

In this section, we study the existence of a pair of sub-game perfect strategies. To this end, in addition the model assumptions from Sec. 2 and the Assumption (A1), we make the following standing assumptions:

(A2) The set $\boldsymbol{\Theta}$ is a compact subset of $\mathbb{R}^d$.
(A3) The probability measures in the family

$$\{Q(\,\cdot \mid t, y, a, \theta),\ t \in \mathcal{T},\ y \in E_Y, a \in A,\ \theta \in \boldsymbol{\Theta}\}$$

are equivalent.

We will show that these assumptions are satisfied in the example studied in Sec. 4. We note that Assumption (A3) could be alternatively formulated in terms of the probability measures generated by $Z^\theta,\ \theta \in \boldsymbol{\Theta}$.

Next we give the main result of this section.

Theorem 3.2. *The functions* V_t, $t \in \mathcal{T}$, *are lower semi-analytic* (*l.s.a.*), *and there exists a pair of sub-game perfect strategies.*

Proof. We will prove existence of sub-game perfect strategies by applying Theorem 3.1 and show, by backward induction, that for any $t \in \mathcal{T}'$, $y \in E_Y$, there exist universally measurable $\widetilde{\varphi}_t(y)$ and $\widetilde{\psi}_t(y, \widetilde{\varphi}_t(y))$ such that

$$V_t(y) = Q^{\widetilde{\varphi}_t, \widetilde{\psi}_t}_{y,t} V_{t+1} - Q^{\widetilde{\varphi}_t, \widetilde{\psi}_t}_{y,t}(G \circ \widetilde{g}_{t+1}) + G(Q^{\widetilde{\varphi}_t, \widetilde{\psi}_t}_{y,t} \widetilde{g}_{t+1}), \tag{3.10}$$

and $V_t(y)$ is l.s.a., with $\widetilde{g}_{t+1} = g^{\widetilde{\varphi}, \widetilde{\psi}}_{t+1}$, and where we recall that $V_t(y) := J_t(y, \bar{\varphi}^t, \bar{\psi}^t)$ for some pair of sub-perfect strategies.

In view of Lemma 3.1, we have that for $t = T-1$

$$V_{T-1}(y) = \max_{a \in A} \inf_{\theta \in \tau(T-1,c)} (Q^{a,\theta}_{y,T-1} V_T - Q^{a,\theta}_{y,T-1}(G \circ \widetilde{g}_T) + G(Q^{a,\theta}_{y,T-1} \widetilde{g}_T)),$$

where we put $\widetilde{g}_T(y) = x$ which is Borel measurable in y.

Hence, according to our assumptions, $G \circ \widetilde{g}_T$ and $V_T(y) = F(x) + G(x)$ are also Borel measurable. By Assumption (A1), and using Bertsekas & Shreve (1978, Proposition 7.29), the following functions:

$$Q^{a,\theta}_{y,T-1} V_T, \quad Q^{a,\theta}_{y,T-1}(G \circ \widetilde{g}_T), \quad G(Q^{a,\theta}_{y,T-1} \widetilde{g}_T)$$

are Borel measurable in (y, a, θ). Therefore, the function

$$\check{V}_{T-1}(y, a, \theta) := Q^{a,\theta}_{y,T-1} V_T - Q^{a,\theta}_{y,T-1}(G \circ \widetilde{g}_T) + G(Q^{a,\theta}_{y,T-1} \widetilde{g}_T)$$

is Borel measurable. Moreover, for any $b \in \mathbb{R}$, the set $\{(y, a, \theta) \in E_Y \times A \times \tau(T-1, c) : \check{V}_{T-1}(y, a, \theta) < b)\}$ is a Borel measurable subset of $E_Y \times A \times \tau(T-1, c)$. Since $E_Y \times A \times \tau(T-1, c)$ is a closed subset of $E_Y \times A \times \boldsymbol{\Theta}$, which is a Polish space (and thus a Borel space), then it is a Borel subspace. In turn, by Bertsekas & Shreve (1978, Proposition 7.36), the set $\{(y, a, \theta) \in E_Y \times A \times \tau(T-1, c) : \check{V}_{T-1}(y, a, \theta) < b)\}$ is analytic. Consequently, the function $\check{V}_{T-1}(y, a, \theta)$ is l.s.a.

By adopting the notations in Bertsekas & Shreve (1978, Proposition 7.50), we write[b]

$$
\begin{aligned}
&X = E_Y \times A = \mathbb{R}^n \times \mathbb{R}^d \times A, \quad x = (y, a),\\
&Y = \boldsymbol{\Theta}, \quad y = \theta,\\
&D = \bigcup_{(y,a)\in E_Y \times A} \{(y, a)\} \times \tau(T-1, c),\\
f(x, y) = {} & \check{V}_{T-1}(y, a, \theta).
\end{aligned}
$$

Recall that in view of the prior assumptions, X and Y are both Borel spaces. D is a closed set and therefore analytic, and the cross section $D_x = D_{(y,a)} = \{\theta \in \boldsymbol{\Theta} : (y, a, \theta) \in D\}$ is given by $D_{(y,a)} = \tau(T-1, c)$. Hence, by Bertsekas & Shreve (1978, Proposition 7.47), the function

$$
\check{V}^*_{T-1}(y, a) = \inf_{\theta\in\tau(T-1,c)} \check{V}_{T-1}(y, a, \theta), \quad (y, a) \in E_Y \times A,
$$

is l.s.a. Moreover, in view of Bertsekas & Shreve (1978, Proposition 7.50), for any $\varepsilon > 0$, there exists an analytically measurable function $\widetilde{\psi}^{\varepsilon}_{T-1}(y, a)$ such that

$$
\check{V}_{T-1}(y, a, \widetilde{\psi}^{\varepsilon}_{T-1}(y, a)) \leq \begin{cases} \check{V}^*_{T-1}(y, a) + \varepsilon & \text{if } \check{V}^*_{T-1}(y, a) > -\infty, \\ -\dfrac{1}{\varepsilon} & \text{if } \check{V}^*_{T-1}(y, a) = -\infty. \end{cases}
$$

Therefore, for any fixed (y, a), we obtain a sequence $\{\widetilde{\psi}^{\frac{1}{n}}_{T-1}(y, a),\ n \in \mathbb{N}\}$, such that

$$
\lim_{n\to\infty} \check{V}_{T-1}(y, a, \widetilde{\psi}^{\frac{1}{n}}_{T-1}(y, a)) = \check{V}^*_{T-1}(y, a).
$$

By Assumption (A2), there exists a convergent subsequence $\{\widetilde{\psi}^{\frac{1}{n_k}}_{T-1}(y, a), k \in \mathbb{N}\}$, such that its limit $\widetilde{\psi}^{*}_{T-1}(y, a)$ satisfies

$$
\check{V}_{T-1}(y, a, \widetilde{\psi}^{*}_{T-1}(y, a)) = \check{V}^*_{T-1}(y, a).
$$

Clearly, $V_{T-1}(y) = \max_{a\in A} \check{V}^*_{T-1}(y, a)$. Next, for every fixed $a \in A$, any $b \in \mathbb{R}$, we write the following complement of the upper level set as

$$
\begin{aligned}
\{y \in E_Y : \check{V}^*_{T-1}(y, a) < b\} &= \{y \in E_Y : (y, a) \in \check{V}^{*,-1}_{T-1}((-\infty, b))\}\\
&= \operatorname{proj}_{E_Y}\{\check{V}^{*,-1}_{T-1}((-\infty, b)) \cap (E_Y \times \{a\})\}.
\end{aligned}
$$

As a projection of an analytic set, such set is analytic, and moreover, $\check{V}^*_{T-1}(y, a)$ is l.s.a. in y for every $a \in A$. Thus, we get that $V_{T-1}(y)$ is l.s.a. as being the maximum of a finite collection of l.s.a. functions.

[b] The notation X and Y representing the relevant Borel spaces, should not be confused with the notation X and Y representing the relevant processes.

For every $y \in E_Y$, define $\widetilde{\varphi}^*_{T-1}(y) = \operatorname{argmax}_{a \in A}\{\check{V}^*_{T-1}(y,a) = V_{T-1}(y)\}$. Note that the set A is finite. Hence, we write

$$\{y \in E_Y : \check{V}^*_{T-1}(y,a) = V_{T-1}(y)\} = \{y \in E_Y : \check{V}^*_{T-1}(y,a) - V_{T-1}(y) = 0\}.$$

Since $\check{V}^*_{T-1}(y,a)$ is l.s.a. in (y,a), then it is analytically measurable and universally measurable in (y,a). Moreover, it is universally measurable in y for every a. Similarly, the function $V_{T-1}(y)$ is universally measurable in y as well. We get that the set $\{y \in E_Y : \check{V}^*_{T-1}(y,a)\}$ is universally measurable for every $a \in A$. Thus, the function $\widetilde{\varphi}^*_{T-1}(y)$ is universally measurable and it is the optimal selector. It is also straightforward to verify (3.10), and hence the proof for $t = T-1$ is complete.

Next, we note that the stochastic kernel $Q(dy'|T-1, y, \widetilde{\varphi}_{T-1}, \widetilde{\psi}_{T-1})$ is universally measurable as it is a composition of Borel measurable and universally measurable mappings (cf. Bertsekas & Shreve (1978, Proposition 7.44)) Hence, by Bertsekas & Shreve (1978, Proposition 7.46), we deduce that $\widetilde{g}_{T-1}(y) = Q^{\widetilde{\varphi}_{T-1},\widetilde{\psi}_{T-1}}_{y,T-1}\widetilde{g}_T$ is universally measurable.

For $0 < t \leq T-1$, assume that $V_t(y)$ is l.s.a. and $\widetilde{g}_t(y)$ is universally measurable. Then, by Bertsekas & Shreve (1978, Lemma 7.27), for any chosen $\theta \in \boldsymbol{\Theta}$, we have that there exists a Borel measurable function $\check{g}_t(y)$ such that $\check{g}_t(y) = \widetilde{g}_t(y)$ almost surely under the reference measure $\mathbb{P}$. Consequently, by Assumption (A3) we have $Q^{a,\theta}_{y,t}\check{f} = Q^{a,\theta}_{y,t}\widetilde{f}$ for any two integrable functions $\check{f}$ and $\tilde{f}$, such that $\check{f}$ is Borel measurable, $\widetilde{f}$ is universally measurable, and $\check{f} = \tilde{f}$ almost surely under $\mathbb{P}$. Thus, $Q^{a,\theta}_{y,t}\check{g}_t = Q^{a,\theta}_{y,t}\widetilde{g}_t$.

Finally, we note that the stochastic kernel $Q(dy' \mid t, y, a, \theta)$ is Borel measurable in (y, a, θ). By Bertsekas & Shreve (1978, Proposition 7.48), it implies that $Q^{a,\theta}_{y,t-1}V_t$ is l.s.a. On the other hand, since $G \circ \tilde{g}_t$ is Borel measurable, we have that $-Q^{a,\theta}_{y,t}(G \circ \tilde{g}_t)$ and $G(Q^{a,\theta}_{y,t}\tilde{g}_t)$ are also Borel measurable. Thus, they are also l.s.a. The rest of the proof follows analogously. By induction, we conclude that (3.10) holds true for any $t \in \mathcal{T}'$, $y \in E_Y$, and an universally measurable pair sub-game perfect strategies exist. □

4. Uncertain Dynamic Mean-variance Portfolio Selection Problem

In this section, we will present an example that illustrates the results of Sec. 3. Namely, we consider a dynamic mean-variance portfolio selection problem, when an investor is deciding at time t on investing in a risky asset and a risk-free banking account in order to maximize the terminal weighted mean-variance criterion of the form (2.3), subject to market model uncertainty.

We take a risk-free asset B with a constant interest rate $r = (B_{t+1} - B_t)/B_t$, and a risky asset, say a stock, with the corresponding return from time t to $t+1$ denoted by r^s_{t+1}. We assume that the return process r^s, is observed. The dynamics of the wealth process, say W, produced by a self-financing trading strategy is given by

$$W_{t+1} = W_t(1 + r + \varphi_t(r^s_{t+1} - r)), \quad t \in \mathcal{T}', \tag{4.1}$$

with the initial wealth $W_0 = w_0 > 0$, and where φ_t denotes the proportion of the portfolio wealth invested in the risky asset from time t to $t+1$. We assume that the process φ takes finitely many values, say $a_i,\ i = 1, \dots, N$, where $a_i \in [0, 1]$.

We further assume that r_t^s+1, $t = 1, \dots, T-1$, is an i.i.d. sequence of log-normal distributed random variables, or saying differently we assume that the excess log-returns are normally distributed. Namely,

$$r_t^s = e^{Z_t} - 1,$$

where Z_t is an i.i.d. sequence of Gaussian random variables with mean μ and variance σ^2. Equivalently, we put $Z_t = \mu + \sigma\varepsilon_t$, where $\varepsilon_t,\ t \in \mathcal{T}'$ are i.i.d. standard Gaussian random variables. Note that under the above model assumptions, the wealth process remains positive a.s. The model uncertainty will come from the unknown parameters μ and/or σ. Using the notations from Sec. 2, here we have that $X_t = W_t$, and setting $x = w$ we get

$$S(w, a, z) = w(1 + r + a(e^z - 1 - r)), \quad A = \{a_i,\ i = 1, \dots, N\}.$$

Same as in Example 2.1, we take $F(w) = w - \gamma w^2$, and $G(w) = \gamma w^2$. Formally, the investor's adaptive robust mean variance problem is formulated as follows:

$$\sup_{\varphi\in\mathcal{A}} \inf_{\mathbb{Q}\in\mathcal{Q}^{\varphi,\Psi}_{y_0,0}} (\mathbb{E}_{\mathbb{Q}}(W_T) - \gamma \operatorname{Var}_{\mathbb{Q}}(W_T)), \tag{4.2}$$

where $\mathcal{A}$ is the set of Markovian trading strategies. We will find a pair of sub-game perfect strategy corresponding to (4.2).

We will discuss two cases: Case 1 — unknown mean μ and known standard deviation σ, and Case 2 — both μ and σ are unknown.

Case 1. Assume that σ is known, and the model ambiguity comes only from the parameter μ, whose true but unknown value is denoted by μ^*. Thus, using the notations from Sec. 2, we have that $\theta^* = \mu^*$, $\theta = \mu$, and we take $C_t = \widehat{\mu}_t$, $\mathbf{\Theta} = [\underline{\mu}, \overline{\mu}] \subset \mathbb{R}$, where $\widehat{\mu}$ is a point estimator of μ, given the observations of process Z, that takes values in $\mathbf{\Theta}$. The values of the boundaries $\underline{\mu}$ and $\overline{\mu}$ are fixed *a priori* by the observer. For the detailed discussion regarding the construction of such estimators we refer to Bielecki *et al.* (2017a). For this example, it is enough to take as $\widehat{\mu}$ the Maximum Likelihood Estimator (MLE), which is the sample mean in this case, projected appropriately on $\mathbf{\Theta}$. Formally, the recursion construction of $\widehat{\mu}$ is defined as follows:

$$\begin{aligned} \widetilde{\mu}_{t+1} &= \frac{t}{t+1}\widehat{\mu}_t + \frac{1}{t+1}Z_{t+1}, \\ \widehat{\mu}_{t+1} &= \pi(\widetilde{\mu}_{t+1}), \quad t \in \mathcal{T}', \end{aligned} \tag{4.3}$$

with $\widehat{\mu}_0 = c_0 \in \mathbf{\Theta}$, and where π is the projection to the closest point in $\mathbf{\Theta}$, namely $\pi(\mu) = \mu$ if $\mu \in [\underline{\mu}, \overline{\mu}]$, $\pi(\mu) = \underline{\mu}$ if $\mu < \underline{\mu}$, and $\pi(\mu) = \overline{\mu}$ if $\mu > \overline{\mu}$. We take as the initial guess c_0 any point in $\mathbf{\Theta}$.

Putting the above together we get that the function $\mathbf{G}$ defined in (2.6) is given here by

$$\mathbf{G}(t,w,c,a,z) = \left(w(1+r+a(e^z-1-r)), \pi\left(\frac{t}{t+1}c + \frac{1}{t+1}z\right)\right). \tag{4.4}$$

It can be easily verified that the kernel $Q(\,\cdot\,|t,y,a,\mu)$, defined in terms of function $\mathbf{G}$ given in (4.4), satisfies Assumption (A1), for example by using Bertsekas & Shreve (1978, Proposition 7.26). Obviously Assumption (A2) is satisfied.

As far as Assumption (A3), let $B \in \mathcal{E}_Y$ such that $Q(B \mid t,y,a,\mu) = 0$. In view of (2.7) we have that

$$\mathbb{P}_\mu(Z_{t+1} \in \{z : \mathbf{G}(t,y,a,z) \in B\}) = 0,$$

where $Z_{t+1} \sim N(\mu,\sigma^2)$. Due to the normality, it is clear that for any $\mu' \in \mathbf{\Theta}$, we also have

$$\mathbb{P}_{\mu'}(Z'_{t+1} \in \{z : \mathbf{G}(t,y,a,z) \in B\}) = 0$$

with $Z'_{t+1} \sim N(\mu',\sigma^2)$. Hence, $Q(B \mid t,y,a,\mu') = 0$, and thus the stochastic kernels $Q(\cdot \mid t,y,a,\mu)$ and $Q(\cdot \mid t,y,a,\mu')$ are equivalent and Assumption (A3) is fulfilled.

Now, we note that the $(1-\alpha)$-confidence region for μ^* at time t is given as

$$\Theta_t = \tau(t,\widehat{\mu}_t),$$

where

$$\tau(t,c) = \left[\max\left(c - \frac{\sigma}{\sqrt{t}}q_{\alpha/2}, \underline{\mu}\right), \min\left(c + \frac{\sigma}{\sqrt{t}}q_{\alpha/2}, \overline{\mu}\right)\right], \tag{4.5}$$

and where q_α denotes the α-quantile of a standard normal distribution. We take closed intervals in (4.5) to preserve compactness. With these at hand we construct the kernel Q according to (2.7), and the set of probability measures $\mathcal{Q}^{\varphi,\mathbf{\Psi}}_{y_0,0}$ on canonical space according to (2.8). We recall that in the present case $y_0 = (w_0, c_0)$.

The Bellman equations (3.3)–(3.5) take the form

$$V_t(y) = \max_{a\in A}\inf_{\theta\in\tau(t,c)}(Q^{a,\theta}_{y,t}V_{t+1} - [Q^{a,\theta}_{y,t}(\gamma\hat{g}^2_{t+1}(\cdot)) - \gamma(Q^{a,\theta}_{y,t}\hat{g}_{t+1})^2]), \tag{4.6}$$

$$\hat{g}_t(y) = Q^{\hat{\varphi}_t,\hat{\psi}_t}_{y,t}\hat{g}_{t+1}, \tag{4.7}$$

$$\begin{aligned} &V_T(y) = w,\\ &\quad y = (w,c) \in E_Y, \quad t \in \mathcal{T}', \end{aligned} \tag{4.8}$$

with $\tau(t,c)$ given in (4.5).

In view of Theorem 3.1 a pair $(\hat{\varphi},\hat{\psi})$ of Markov strategies satisfying (4.6)–(4.8) is a pair of sub-game perfect strategies for the adaptive robust mean-variance problem (4.2) with unknown μ.

In the next section, we will solve Eqs. (4.6)–(4.8) for a pair $(\hat{\varphi},\hat{\psi})$ using a machine learning based method. Note that although the dimension of the state space E_Y is two in the present case, which allows for efficient use of the traditional grid-based method to numerically solve the Bellman equations, our machine learning based method, originally proposed in Chen & Ludkovski (2019) can be applied to high dimensional problems where gridding is extremely inefficient. Generally speaking, this approach overcomes the challenges met in high dimensional (robust) stochastic control problems.

Case 2. Here we assume that both μ and σ are unknown, and thus, in the notations of Sec. 2, we have $\theta^* = (\mu^*, (\sigma^*)^2)$, $\theta = (\mu, \sigma^2)$, $\boldsymbol{\Theta} = [\underline{\mu}, \overline{\mu}] \times [\underline{\sigma}^2, \overline{\sigma}^2] \subset \mathbb{R} \times \mathbb{R}_+$, for some fixed $\underline{\mu}, \overline{\mu} \in \mathbb{R}$ and $\underline{\sigma}^2, \overline{\sigma}^2 \in \mathbb{R}_+$. Similar to the Case 1, we take as the point estimators for μ^* and $(\sigma^*)^2$ the corresponding MLEs, namely the sample mean and, respectively, the sample variance, projected appropriately to the rectangle $\boldsymbol{\Theta}$. It is shown in Bielecki *et al.* (2017a) that the following recursions hold true:

$$\widetilde{\mu}_{t+1} = \frac{t}{t+1}\widehat{\mu}_t + \frac{1}{t+1}Z_{t+1},$$

$$\widetilde{\sigma}^2_{t+1} = \frac{t}{t+1}\widehat{\sigma}^2_t + \frac{t}{(t+1)^2}(\widehat{\mu}_t - Z_{t+1})^2,$$

$$(\widehat{\mu}_{t+1}, \widehat{\sigma}^2_{t+1}) = \pi(\widetilde{\mu}_{t+1}, \widetilde{\sigma}^2_{t+1}), \quad t = 1, \ldots, T-1,$$

with some initial guess $\widehat{\mu}_0 = c'_0$, and $\widehat{\sigma}^2_0 = c''_0$, and where π is the projection[c] defined similarly as in (4.3). Consequently, we set $C_t = (C'_t, C''_t) = (\widehat{\mu}_t, \widehat{\sigma}^2_t), t \in \mathcal{T}$, and, respectively, we have

$$R(t, c, z) = \pi\left(\frac{t}{t+1}c' + \frac{1}{t+1}z, \frac{t}{t+1}c'' + \frac{t}{(t+1)^2}(c' - z)^2\right),$$

with $c = (c', c'')$. Thus, in this case, we have

$$\mathbf{G}(t, v, c, a, z) = \left(v(1 + r + a(e^z - 1 - r)), \pi\left(\frac{t}{t+1}c' + \frac{1}{t+1}z, \right.\right.$$
$$\left.\left.\frac{t}{t+1}c'' + \frac{t}{(t+1)^2}(c' - z)^2\right)\right). \tag{4.9}$$

Similarly as in Case 1 with regard to function $\mathbf{G}$ given in (4.4), it can be easily verified that the kernel $Q(\,\cdot \mid t, y, a, \mu)$, defined in terms of function $\mathbf{G}$ given in (4.9), satisfies Assumptions (A1) and (A3). The $(1-\alpha)$-confidence region for $(\mu^*, (\sigma^*)^2)$ at time t is the ellipsoid given by

$$\Theta_t = \tau(t, \widehat{\mu}_t, \widehat{\sigma}^2_t),$$

with

$$\tau(t, c) = \left\{(\mu, \sigma^2) \in \boldsymbol{\Theta} : \frac{t}{c''}(c' - \mu)^2 + \frac{t}{2(c'')^2}(c'' - \sigma^2)^2 \le \kappa\right\}, \tag{4.10}$$

where κ is the $(1-\alpha)$ quantile of the χ^2 distribution with two degrees of freedom. Accordingly, Eqs. (3.3)–(3.5) take the form analogous to (4.6)–(4.8) with, in

[c] We refer to Bielecki *et al.* (2017a) for precise definition of the projection π, but essentially it is defined as the closest point in the set $\boldsymbol{\Theta}$.

particular, $\tau(t,c)$ given in (4.10). In view of Theorem 3.1 a pair $(\hat{\varphi},\hat{\psi})$ of Markov strategies satisfying such equations is a pair of sub-game perfect strategies for the adaptive robust mean-variance problem (4.2) with unknown μ and σ. Note that the dimension of the state space in this case is three, and a grid-based method becomes extremely inefficient. Hence, developing a numerical solver with good scalability is crucial. As we mentioned earlier, and as described in next section, we will use the regression Monte Carlo idea and Gaussian process surrogates to compute the optimal pair $(\hat{\varphi},\hat{\psi})$ via backward recursion.

5. Machine Learning Algorithm and Numerical Results

It is important to note that even though the market model of Sec. 4 is the same as the one considered in Bielecki *et al.* (2019, Sec. 4), the Bellman equations associated to the problem in Sec. 4 are more difficult to treat numerically than those from Bielecki *et al.* (2019, Sec. 4). In Bielecki *et al.* (2019) the authors used a (classical) nonmachine-learning based algorithm to solve numerically the Bellman equations, which cannot be used in the current work, for reasons outlined below.

The essence of the machine learning algorithm that we will use solving numerically the example from previous section is the same for both Cases 1 and 2. The algorithm begins with discretizing the relevant state space, for which we employ the regression Monte Carlo method to create a random (nongridded) mesh for the process $Y=(W,C)$. Note that the component W depends on the control process, hence at each time t we randomly select from the set A a value of φ_t, and we randomly generate a value of r^S_{t+1}, so to simulate the value of W_{t+1}. The resulting random mesh consists of a number of simulated paths of Y. Then, we solve Eqs. (4.6)–(4.8) in Case 1, and their counterparts in Case 2, and compute the optimal trading strategies at all mesh points.

The need for applying machine learning to solve our Bellman equations is twofold. On one hand, to approximate the integral operations such as $Q^{a,\theta}_{y,t}V_{t+1}$, we replace the integrals with weighted sums through Monte Carlo simulation or a Gaussian quadrature recipe. Accordingly, interpolation and/or extrapolation, via appropriate Gaussian Processes (GP) surrogates, will be used to evaluate the terms in the summations. Note that the state space used in the adaptive robust control method is E_Y, which is potentially highly dimensional, where traditional linear interpolation/extrapolation methods bring multiple limitations, and therefore GP surrogates are used to overcome these limitations. On the other hand, the computation procedure involving solving Eqs. (4.6)–(4.8) in Case 1, and their counterparts in Case 2, outputs approximate values of the optimal strategies for the mesh points on the sample paths only. Hence, to obtain the value of $\hat{\varphi}_t(y)$ for arbitrary $y\in E_Y$, an efficient regression model for $\hat{\varphi}$, such as a GP surrogate, is desirable.

5.1. *Description of the algorithm*

In view of the above mentioned computational challenges, we numerically tackle the adaptive robust stochastic control problem by following the novel method introduced in Chen & Ludkovski (2019). The key idea of this method is to utilize a nonparametric value function approximation strategy (cf. Powell (2007)) called Gaussian process surrogate (cf. Rasmussen (2006)). For the purpose of solving the Bellman equations (4.6)–(4.8) in Case 1, and their counterparts in Case 2, we build GP regression model for the value function $V_{t+1}(\,\cdot\,)$ and the operator $\hat{g}_{t+1}$ so that we can evaluate

$$Q^{a,\theta}_{y,t}V_{t+1} - \gamma Q^{a,\theta}_{y,t}\hat{g}^2_{t+1} + \gamma(Q^{a,\theta}_{y,t}\hat{g}_{t+1})^2. \tag{5.1}$$

We also construct GP regression model for the optimal control $\hat{\varphi}$. It permits us to apply the optimal strategy to out-of-sample paths without actual optimization, which allows for a significant reduction of the computational cost.

As the GP surrogate for the value function V_t we consider a regression model $\widetilde{V}_t(y)$ such that for any $y^1,\ldots,y^N \in E_Y$, with $y^i \neq y^j$ for $i \neq j$, the random variables $\widetilde{V}_t(y^1),\ldots,\widetilde{V}_t(y^N)$ are jointly normally distributed. Then, given training data $(y^i, V_t(y^i))$, $i=1,\ldots,N$, for any $y \in E_Y$, the predicted value $\widetilde{V}_t(y)$, providing an estimate (approximation) of $V_t(y)$, is given by

$$\widetilde{V}(y) = (k(y,y^1),\ldots,k(y,y^N))[\mathbf{K}+\epsilon^2\mathbf{I}]^{-1}(V_t(y^1),\ldots,V_t(y^N))^T,$$

where ϵ is a tuning parameter, $\mathbf{I}$ is the $N\times N$ identity matrix and the matrix $\mathbf{K}$ is defined as $\mathbf{K}_{i,j} = k(y^i,y^j)$, $i,\ j=1,\ldots,N$. The function $k(\,\cdot\,,\,\cdot\,)$ is the kernel function for the GP model, and in this work we choose the kernel as the Matern-5/2 (cf. Rasmussen (2006)). Fitting the GP surrogate $\widetilde{V}_t$ means to estimate the hyper-parameters inside $k(\,\cdot\,,\,\cdot\,)$ through training data $(y^i, V_t(y^i))$, $i=1,\ldots,N$. We note that since we do not have the closed form expression for $V_t(y)$, we numerically evaluate $V_t(y)$ instead. The GP surrogates for $\hat{g}_t$ and $\hat{\varphi}_t$ are obtained in an analogous way. We take $\epsilon = 10^{-5}$.

Given the mesh points $\{y^i_t,\ i=1,\ldots,N,\ t=0,\ldots,T-1\}$, the overall algorithm proceeds as follows:

Part A: Time backward recursion for $t = T-1,\ldots,0$.

(1) Assume that $V_{t+1}(y^i_{t+1})$, $\hat{g}_{t+1}(y^i_{t+1})$ and $\hat{\varphi}_{t+1}(y^i_{t+1})$, $i=1,\ldots,N$, are numerically approximated as $\overline{V}_{t+1}(y^i_{t+1})$, $\bar{\hat{g}}_{t+1}(y^i_{t+1})$ and $\bar{\hat{\varphi}}_{t+1}(y^i_{t+1})$, $i=1,\ldots,N$, respectively. Also suppose that the corresponding GP surrogates $\widetilde{V}_{t+1}$, $\tilde{\hat{g}}_{t+1}$, and $\tilde{\hat{\varphi}}_{t+1}$ are fitted through training data $(y^i_{t+1}, \overline{V}_{t+1}(y^i_{t+1}))$, $(y^i_{t+1}, \bar{\hat{g}}_{t+1}(y^i_{t+1}))$, and $(y^i_{t+1}, \bar{\hat{\varphi}}_{t+1}(y^i_{t+1}))$, $i=1,\ldots,N$, respectively.

(2) For time t, any $a \in A$, $\theta \in \tau(t,c)$ and each y_t^i, $i = 1, \ldots, N$, use one-step Monte Carlo simulation to estimate the quantities

$$Q_{y_t^i,t}^{a,\theta} V_{t+1} = \mathbb{E}_\theta[V_{t+1}(\boldsymbol{G}(t, y_t^i, a, Z_{t+1}))],$$

$$Q_{y_t^i,t}^{a,\theta} \hat{g}_{t+1}^2 = \mathbb{E}_\theta[\hat{g}_{t+1}^2(\boldsymbol{G}(t, y_t^i, a, Z_{t+1}))],$$

$$Q_{y_t^i,t}^{a,\theta} \hat{g}_{t+1} = \mathbb{E}_\theta[\hat{g}_{t+1}(\boldsymbol{G}(t, y_t^i, a, Z_{t+1}))].$$

For that, if $Z_{t+1}^1, \ldots, Z_{t+1}^M$ is a sample of Z_{t+1} drawn from the normal distribution corresponding to parameter θ, where $M > 0$ is a positive integer, then estimate the above expectations as

$$Q_{y_t^i,t}^{a,\theta} V_{t+1} \approx \widetilde{Q}_{y_t^i,t}^{a,\theta} \widetilde{V}_{t+1} := \frac{1}{M} \sum_{i=1}^{M} \widetilde{V}_{t+1}(\boldsymbol{G}(t, y_t^i, a, Z_{t+1}^i)),$$

$$Q_{y_t^i,t}^{a,\theta} \hat{g}_{t+1}^2 \approx \widetilde{Q}_{y_t^i,t}^{a,\theta} \widetilde{\hat{g}}_{t+1}^2 := \frac{1}{M} \sum_{i=1}^{M} \widetilde{\hat{g}}_{t+1}^2(\boldsymbol{G}(t, y_t^i, a, Z_{t+1}^i)),$$

$$Q_{y_t^i,t}^{a,\theta} \hat{g}_{t+1} \approx \widetilde{Q}_{y_t^i,t}^{a,\theta} \widetilde{\hat{g}}_{t+1} := \frac{1}{M} \sum_{i=1}^{M} \widetilde{\hat{g}}_{t+1}(\boldsymbol{G}(t, y_t^i, a, Z_{t+1}^i)).$$

Next, estimate the values of (5.1).

(3) For each y_t^i, $i = 1, \ldots, N$, and any $a \in A$, build a uniform grid over the set $\tau(t,c)$, and search for a grid point, say $\hat{\theta}(y_t^i, a)$, that minimizes

$$\widetilde{Q}_{y_t^i,t}^{a,\theta} \widetilde{V}_{t+1} - \gamma \widetilde{Q}_{y_t^i,t}^{a,\theta} \widetilde{\hat{g}}_{t+1}^2 + \gamma(\widetilde{Q}_{y_t^i,t}^{a,\theta} \widetilde{\hat{g}}_{t+1})^2.$$

(4) Compute

$$\overline{V}_t(y_t^i) = \max_{a \in A} \{\widetilde{Q}_{y_t^i,t}^{a,\hat{\theta}(y_t^i,a)} \widetilde{V}_{t+1} - \gamma \widetilde{Q}_{y_t^i,t}^{a,\hat{\theta}(y_t^i,a)} \widetilde{\hat{g}}_{t+1}^2 + \gamma(\widetilde{Q}_{y_t^i,t}^{a,\hat{\theta}(y_t^i,a)} \widetilde{\hat{g}}_{t+1})^2\},$$

and obtain a maximizer $\overline{\hat{\varphi}}_t(y_t^i)$, and corresponding $\overline{\hat{g}}_t(y_t^i) = \widetilde{Q}_{y_t^i,t}^{\hat{\varphi}_t(y_t^i),\hat{\theta}(y_t^i,\hat{\varphi}_t(y_t^i))}$ $\widetilde{\hat{g}}_{t+1}$, $i = 1, \ldots, N$.

(5) Fit GP regression models for $V_t(\cdot)$ and $\hat{g}_t(\cdot)$ using the results from Step 4 above. Fit a GP model for $\hat{a}_t(\cdot)$ as well; this is needed for obtaining values of the optimal strategies for out-of-sample paths in Part B of the algorithm.

(6) Goto (1): Start the next recursion for $t-1$.

Part B: Forward simulation to evaluate the performance of the GP surrogate $\widetilde{\hat{\varphi}}_t(\cdot)$, $t = 0, \ldots, T-1$, over the out-of-sample paths.

(1) Draw $K > 0$ samples of i.i.d. $Z_1^{*,i}, \ldots, Z_T^{*,i}$, $i = 1, \ldots, K$, from the normal distribution corresponding to the assumed true parameter θ^*.

(2) All paths will start from the initial state y_0. The state along each path i is updated according to $\boldsymbol{G}(t, y_t^i, \widetilde{\hat{\varphi}}_t^i(y_t^i), Z_{t+1}^{*,i})$, where $\widetilde{\hat{\varphi}}_t$ is the GP surrogate fitted in Part A.

(3) Obtain the terminal wealth $\hat{W}_T^{*,i}$, generated by $\widetilde{\hat{\varphi}}$ along the path corresponding to the sample of $Z_1^{*,i}, \dots, Z_T^{*,i}$, $i = 1, \dots, K$, and compute

$$V^{\mathrm{ar}} := \underbrace{\frac{1}{K}\sum_{i=1}^{K} \widehat{W}_T^{\mathrm{ar},i}}_{\text{sample mean of } \widehat{W}_T^{\mathrm{ar}}} \underbrace{-\gamma\left(\frac{1}{K}\sum_{i=1}^{K}\left(\widehat{W}_T^{\mathrm{ar},i}\right)^2 - \left(\frac{1}{K}\sum_{i=1}^{K}\widehat{W}_T^{\mathrm{ar},i}\right)^2\right)}_{\text{sample variance of } \widehat{W}_T^{\mathrm{ar}}} \tag{5.2}$$

as an estimate of the performance of the optimal adaptive robust sub-game perfect strategy $\hat{\varphi}$.

We compare (5.2) to the performance of strategy generated by the strong robust sub-game perfect methodology (cf. Remark 3.1) on $K = 2000$ out-of-sample paths, where the latter performance is measured in terms of the mean-variance utility, say V^{sr}, which is computed in analogy to (5.2).

5.2. *Numerical results*

In this section, we apply the machine learning algorithm described above by taking a specific set of parameters. For both, Cases 1 and 2 we take: $T = 52$ with one period of time corresponding to one week; the annualized return on banking account being equal to 0.02 or equivalently $r = 0.0003846$; the initial wealth $W_0 = 100$; in Part A of our algorithm the number of Monte Carlo simulations is $N = 200$, and $M = 100$; the number of forward simulations in Part B is taken $K = 2000$; the confidence level $\alpha = 0.1$. For both cases, we analyze the performance of the control methods for $\gamma = 0.2$ and $\gamma = 0.9$. The assumed true parameter values, the initial guesses for the parameters, the bounds for the uncertainty set $\boldsymbol{\Theta}$, as well as the numerical results, are presented for each case separately.

In what follows we will abbreviate adaptive robust as AR, and strong robust as SR.

Case 1. Recall that in this case only the return μ is assumed to be unknown. The assumed true parameter value is denoted by μ^*, the initial guess is denoted by μ_0, the uncertainty set is the interval $\boldsymbol{\Theta} = [\underline{\mu}, \overline{\mu}]$. The relevant parameters are summarized in Table 1.

In Fig. 2 we display the histogram of out-of-sample terminal wealth W_T that corresponds to the two subcases (optimistic and pessimistic) and two stochastic

Table 1. Model parameters for mean-variance portfolio selection problem; Case 1.

$T = 52,\quad r = 0.0003846,\quad \gamma = 0.2,\quad W_0 = 100$
$\alpha = 0.1,\quad N = 200,\quad M = 100,\quad K = 2000$
$\mu^* = 0.00192,\quad \underline{\mu} = 0.000192,\quad \overline{\mu} = 0.0096,\quad \sigma = 0.0166$
$\mu_0 = 0.002308$ (optimistic), or $\mu_0 = 0.001538$ (pessimistic)

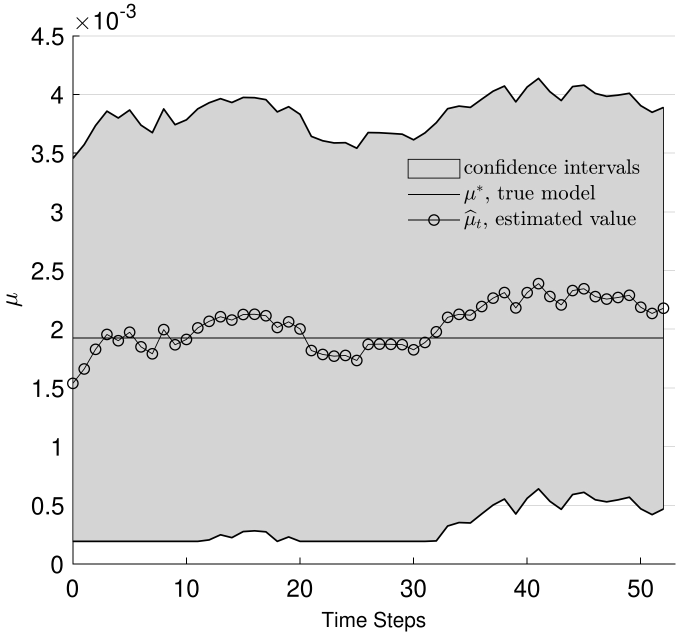

Fig. 1. Evolution of the confidence intervals $\tau(t,c)$ at the confidence level $\alpha = 10\%$; Case 1, pessimistic.

control approaches (AR and SR). The summary statistics are presented in Tables 2 and 3.

We start by presenting the evolution of the confidence intervals $\tau(t,c)$ for the unknown parameter μ; see Fig. 1, which represents the pessimistic initial guess (i.e. $\mu_0 = 0.001538 < \mu^* = 0.00192$). We recall that the SR methodology searches at each time for the worst-case model in $\mathbf{\Theta}$, while the AR searches over the confidence region $\tau(t,c)$, and then approximates the corresponding optimal strategies.

We observe that the performance of the AR and SR methods is comparable in Case 1. This indicates that in this case the uncertainty reduction is not very effective. We attribute this to the fact that the uncertainty regarding the mean return requires a longer time horizon. Nevertheless, the closer inspection of the results shows that in some situations (e.g. optimistic case and $\gamma = 0.9$) the performance of AR is better than performance of SR.

Case 2. We take the same set of parameters as in Case 1 (see Table 1), except that instead of the known and fixed σ, we now take

$$\sigma^* = 0.0416, \quad \underline{\sigma} = 0.0069, \quad \overline{\sigma} = 0.1109,$$

$$\sigma_0 = 0.0347 \quad \text{(optimistic)}, \quad \sigma_0 = 0.0485 \quad \text{(pessimistic)}.$$

With both μ and σ unknown, the model uncertainty set is the two-dimensional rectangle $\mathbf{\Theta} = [\underline{\mu}, \overline{\mu}] \times [\underline{\sigma}^2, \overline{\sigma}^2]$. The evolution of the projected confidence regions, which

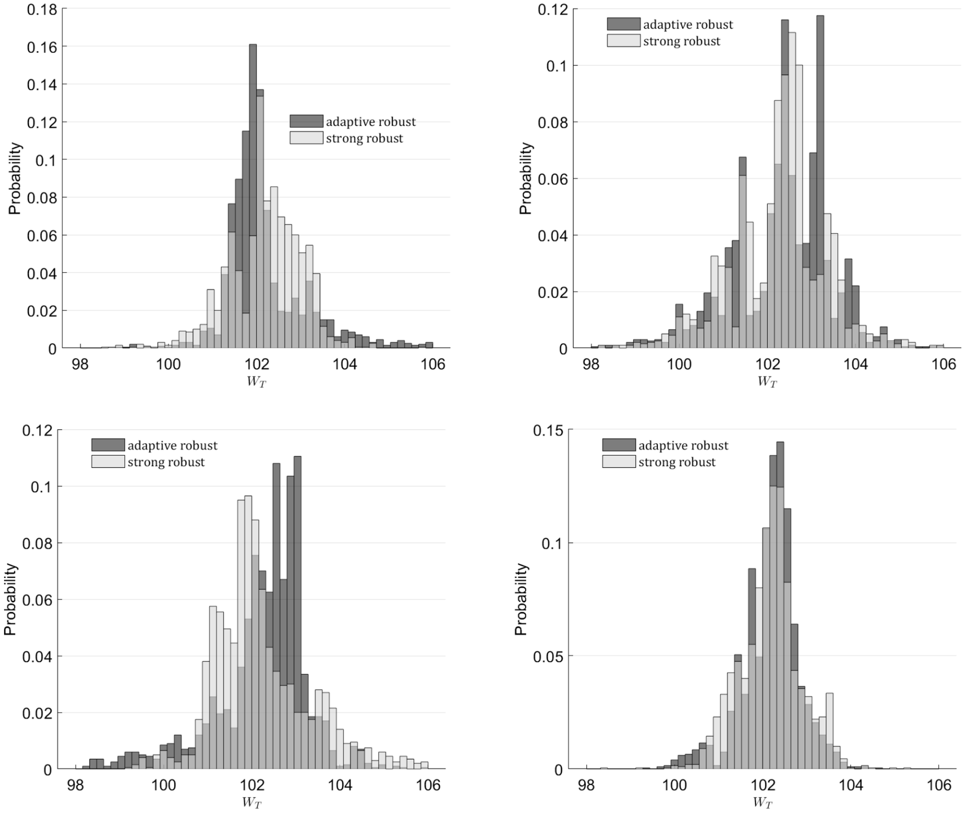

Fig. 2. Histogram of the out-of-sample terminal wealth W_T for Case 1, unknown μ and know σ. Risk averse coefficient $\gamma = 0.2$ — top row, and $\gamma = 0.9$ — bottom row; optimistic case — left panel, and pessimistic case — right panel.

Table 2. Mean, variance, 90%-quantile, maximum and minimum of the out-of-sample terminal wealth and mean-variance utility V for AR and SR for Case 1: optimistic.

	$\gamma = 0.2$		$\gamma = 0.9$	
	AR	SR	AR	SR
mean(W_T)	102.204	102.203	102.263	102.267
var(W_T)	0.887	0.673	1.151	1.473
$q_{0.90}(W_T)$	103.399	103.203	103.185	103.840
max(W_T)	107.664	105.747	106.504	108.971
min(W_T)	98.664	98.564	97.534	97.295
V	102.027	102.068	101.227	100.941

Table 3. Mean, variance, 90%-quantile, maximum and minimum of the out-of-sample terminal wealth and mean-variance utility V for AR and SR for Case 1: pessimistic.

	Pessimistic			
	$\gamma = 0.2$		$\gamma = 0.9$	
	AR	SR	AR	SR
mean(W_T)	102.338	102.262	102.189	102.190
var(W_T)	1.328	1.122	0.430	0.523
$q_{0.90}(W_T)$	103.673	103.481	102.912	103.107
max(W_T)	107.253	107.008	104.672	107.699
min(W_T)	97.537	97.187	99.3915	98.352
V	102.073	102.038	101.802	101.719

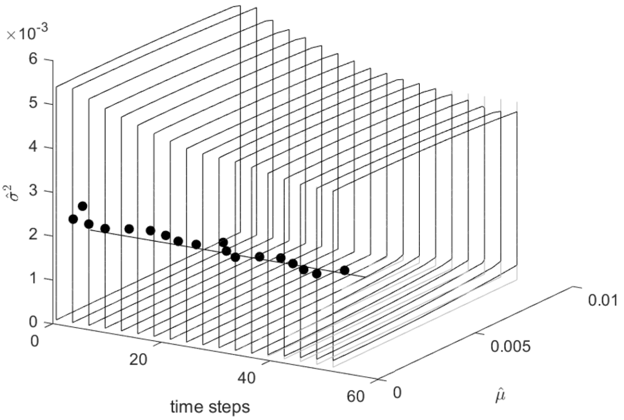

Fig. 3. Evolution of $\tau(t,c)$ at confidence level $\alpha = 10\%$ (ellipsoids), the true parameters value $(\mu^*, (\sigma^*)^2)$ (the solid straight line), and the MLE $(\widehat{\mu}, \widehat{\sigma}^2)$ (dotted line), for Case 2, pessimistic.

are derived from confidence ellipsoids in this case, along with the true parameter values $(\mu^*, (\sigma^*)^2)$ and the MLEs $(\widehat{\mu}, \widehat{\sigma}^2)$ are displayed in Fig. 3.

Similar to Case 1, we present the histograms of the out-of-sample terminal wealth W_T, Fig. 4, for AR and SR. The corresponding summary statistics are listed in Tables 4 and 5.

The results clearly indicate that overall the performance of AR is better than the performance of SR. Across the parameterizations, the value of the optimization criterion (i.e. V_T) is larger for AR than for SR. This is because AR produces much smaller variance of W_T than SR, while both methods produce comparable values of the mean of W_T. This, together with the values of other statistics indicates that SR is less risky than AR. We attribute this to better handling of model uncertainty by AR than it is done by SR.

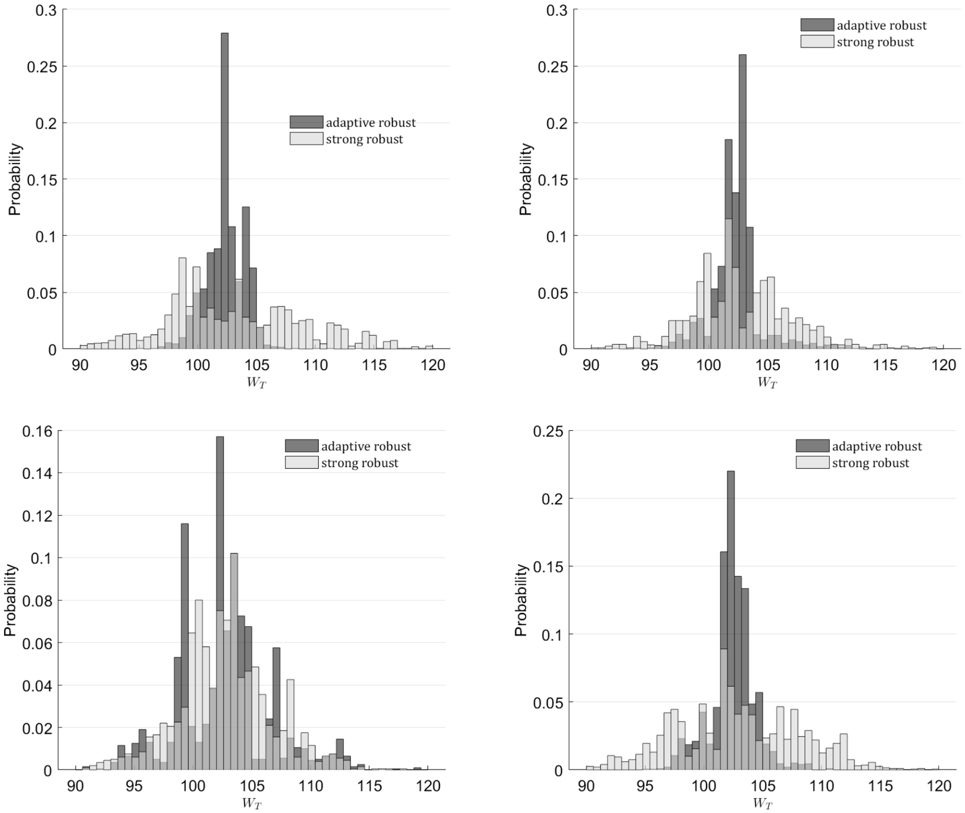

Fig. 4. Histogram of the out-of-sample terminal wealth W_T for Case 2, unknown μ and σ. Risk averse coefficient $\gamma = 0.2$ — top row, and $\gamma = 0.9$ — bottom row; optimistic case — left panel, and pessimistic case — right panel.

Table 4. Mean, variance, 90%-quantile, maximum and minimum of the out-of-sample terminal wealth and mean-variance utility V for the AR and SR methods; Case 2: optimistic.

	Optimistic			
	$\gamma = 0.2$		$\gamma = 0.9$	
	AR	SR	AR	SR
mean(W_T)	102.371	103.132	102.693	102.703
var(W_T)	2.653	34.654	15.983	16.149
$q_{0.90}(W_T)$	104.396	111.554	107.308	108.186
max(W_T)	113.741	126.470	123.094	121.139
min(W_T)	95.027	81.330	90.838	87.413
V	101.840	96.201	88.309	88.169

Table 5. Mean, variance, 90%-quantile, maximum and minimum of the out-of-sample terminal wealth and mean-variance utility V for the AR and SR methods; Case 2: pessimistic.

	Pessimistic			
	$\gamma = 0.2$		$\gamma = 0.9$	
	AR	SR	AR	SR
mean(W_T)	102.339	102.746	102.396	102.869
var(W_T)	4.653	18.911	3.349	29.698
$q_{0.90}(W_T)$	103.521	108.125	104.542	110.036
max(W_T)	118.592	121.682	110.559	128.299
min(W_T)	91.981	80.703	95.238	82.473
V	101.408	98.964	99.382	76.142

Finally, we want to mention that while we performed a similar analysis for various parameters sets and usually the obtained results are similar to the above, some cases may require a deeper analysis and understanding. For example, when the true parameter is close to the worst case one may expect that the strong robust strategy would outperform the adaptive robust strategy, which is not always the case. Also, in some examples, it may happen that the adaptive robust strategy might perform better than the strategy generated by knowing the true parameter. Understanding such phenomena will be part of the future work.

6. Concluding Remarks and Future Research

In this paper we have provided a methodology for dealing with a class of time-inconsistent Markovian decision problems in discrete time subject to model uncertainty, which is also known as Knigthian uncertainty. A version of the adaptive robust approach, that was originated in Bielecki *et al.* (2019), combined with sub-game perfect approach to time-inconsistent stochastic control problems as in Björk & Murgoci (2014), have been successfully used here. For simplicity we were assuming that the set of available actions is finite. This assumption, although quite fine from the numerical perspective, will be generalized to an appropriate compactness assumption in a follow up study. We only proved the existence of a pair of sub-game perfect strategies in Sec. 3.1. The study of the uniqueness of such strategies is deferred to a follow-up paper. Finally, we want to mention that while in this paper we only studied time-inconsistent Markovian decision problems with terminal cost, the generalizations to the case of terminal plus running cost will be addressed in future works.

Acknowledgments

Tomasz R. Bielecki and Igor Cialenco acknowledge support from National Science Foundation Grant DMS-1907568. The authors would like to thank the anonymous referees and the editors for their valuable comments and remarks, which helped us to improve the manuscript.

References

N. Bäuerle & U. Rieder (2011) *Markov Decision Processes with Applications to Finance*. Universitext. Heidelberg: Springer.

H. Bannister, B. Goldys, S. Penev & W. Wu (2016) Multiperiod mean-standard-deviation time consistent portfolio selection, *Automatica* **73**, 15–26.

S. Basak & G. Chabakauri (2010) Dynamic mean-variance asset allocation, *The Review of Financial Studies* **23**, 2970–3016.

D. P. Bertsekas & S. Shreve (1978) *Stochastic Optimal Control: The Discrete-Time Case*. San Diego, CA: Academic Press.

T. R. Bielecki, I. Cialenco & T. Chen (2017a) Recursive construction of confidence regions, *Electronic Journal of Statistics* **11** (2), 4674–4700.

T. R. Bielecki, I. Cialenco & M. Pitera (2017b) A survey of time consistency of dynamic risk measures and dynamic performance measures in discrete time: LM-measure perspective, *Probability, Uncertainty and Quantitative Risk* **2** (3), 1–52.

T. R. Bielecki, I. Cialenco & M. Pitera (2018) A unified approach to time consistency of dynamic risk measures and dynamic performance measures in discrete time, *Mathematics of Operations Research* **43** (1), 204–221.

T. R. Bielecki, I. Cialenco, T. Chen, A. Cousin & M. Jeanblanc (2019) Adaptive Robust Hedging Under Model Uncertainty, *SIAM Journal on Control and Optimization* **57** (2), 925–946.

T. Björk, A. Murgoci & X. Y. Zhou (2014) Mean-variance portfolio optimization with state-dependent risk aversion, *Mathematical Finance* **24**, 1–24.

T. Björk & A. Murgoci (2014) A theory of Markovian time-inconsistent stochastic control in discrete time, *Finance and Stochastics* **18** (3), 545–592.

B. Bouchard, R. Elie & N. Touzi (2010) Stochastic target problems with controlled loss, *SIAM Journal on Control and Optimization* **48** (5), 3123–3150.

H. F. Chen & L. Guo (1991) *Identification and Stochastic Adaptive Control*. Systems & Control: Foundations & Applications. Boston, MA: Birkhäuser Boston.

T. Chen & M. Ludkovski (2019) A machine learning approach to adaptive robust utility maximization and hedging, arXiv:1912.00244.

X. Cui, D. Li, S. Wang & S. Zhu (2012) Better than dynamic mean-variance: time inconsistency and free cash flow stream, *Mathematical Finance* **22** (2), 346–378.

T. E. Duncan, B. Pasik-Duncan & Ł . Stettner (2001) Risk sensitive adaptive control of discrete time Markov processes, *Theory of Probability and Mathematical Statistics* **21** (2), 493–512.

T. E. Duncan, B. Pasik-Duncan & Ł. Stettner (2006) Remarks on risk sensitive adaptive control of Markov processes. In: *Proc. 45th IEEE Conf. Decision and Control*, San Diego, CA, USA, pp. 2861–2865.

I. Ekeland & A. Lazrak (2006) Being serious about noncommitment: Subgame perfect equilibrium in continuous time, preprint.

I. Ekeland & A. Lazrak (2010) The golden rule when preferences are time inconsistent, *Mathematics and Financial Economics* **4** (1), 29–55.

I. Ekeland & T. Pirvu (2008) Investment and consumption without commitment. *Mathematics and Financial Economics* **2** (1), 57–86.

Z. Feinstein & B. Rudloff (2019) Time consistency for scalar multivariate risk measures, arXiv:1810.04978.

I. Gilboa & D. Schmeidler (1989) Maxmin expected utility with nonunique prior, *Journal of Mathematical Economics* **18** (2), 141–153.

S. Goldman (1980) Consistent plans, *The Review of Economic Studies* **47**, 533–537.

P. L. Hansen & T. J. Sargent (2008) *Robustness*. Princeton University Press.

L. P. Hansen, T. J. Sargent, G. Turmuhambetova & N. Williams (2006) Robust control and model misspecification, *Journal of Economic Theory* **128** (1), 45–90.

Z. He (2018) *Equilibrium Strategies for Mean-variance Problem*. PhD thesis, University of Leeds.

C. Hernández & D. Possamaï (2020) Me, myself and I: A general theory of non-markovian time-inconsistent stochastic control for sophisticated agents, arXiv:2002.12572.

Y. Hu, H. Jin & X. Y. Zhou (2012) Time-inconsistent stochastic linear–quadratic control, *SIAM Journal on Control and Optimization* **50** (3), 1548–1572.

C. Karnam, J. Ma & J. Zhang (2017) Dynamic approaches for some time-inconsistent optimization problems, *The Annals of Applied Probability* **27** (6), 3435–3477.

G. Kováčová & B. Rudloff (2019) Time consistency of the mean-risk problem, arXiv:1806.10981.

P. R. Kumar & P. Varaiya (2015) *Stochastic Systems: Estimation, Identification and Adaptive Control*, Classics in Applied Mathematics, Vol. 75. Philadelphia, PA: SIAM.

D. Li & W.-L. Ng (2000) Optimal dynamic portfolio selection: Multiperiod mean-variance formulation, *Mathematical Finance* **10** (3), 387–406.

D. Li & X. Y. Zhou (2000) Continuous-time mean-variance portfolio selection: A stochastic LQ framework, *Applied Mathematics and Optimization* **42**, 19–33.

W. Powell (2007) *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Oxford: Wiley-Blackwell.

C. E. Rasmussen (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press.

Y. Shi & X. Cui (2017) *Time Inconsistency and Self-Control Optimization Problems: Progress and Challenges*. 33–42. Berlin: Springer International Publishing.

R. H. Strotz (1955) Myopia and inconsistency in dynamic utility maximization, *The Review of Economic Studies* **3**, 165–180.