Numerical Solutions to Stochastic Control Problems: When Monte Carlo Simulation Meets Nonparametric Regression

by

Zhiyi Shen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Actuarial Science

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:
Daniel Bauer
Associate Professor
Department of Risk and Insurance
Wisconsin School of Business
University of Wisconsin–Madison

Supervisor(s):
Chengguo Weng
Associate Professor
Department of Statistics and Actuarial Science
University of Waterloo

Internal Member:
Alexander Schied
Professor
Department of Statistics and Actuarial Science
University of Waterloo

Pengfei Li
Associate Professor
Department of Statistics and Actuarial Science
University of Waterloo

Internal-External Member:
Yuying Li
Professor
School of Computer Science
University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The theme of this thesis is to develop theoretically sound as well as numerically efficient Least Squares Monte Carlo (LSMC) methods for solving discrete-time stochastic control problems motivated by insurance and finance problems.

Despite its popularity in solving optimal stopping problems, the application of the LSMC method to stochastic control problems is hampered by several challenges. Firstly, the simulation of the state process is intricate in the absence of the optimal control policy in prior. Secondly, numerical methods only warrant the approximation accuracy of the value function over a bounded domain, which is incompatible with the unbounded set the state variable dwells in. Thirdly, given a considerable number of simulated paths, regression methods are computationally challenging. This thesis responds to the above problems.

Chapter 2 develops a novel LSMC algorithm to solve discrete-time stochastic optimal control problems, referred to as the Backward Simulation and Backward Updating (BSBU) algorithm. The BSBU algorithm has three pillars: a construction of auxiliary stochastic control model, an artificial simulation of the post-action value of state process, and a shape-preserving sieve estimation method which equip the algorithm with a number of merits including obviating forward simulation and control randomization, evading extrapolating the value function, and alleviating computational burden of the tuning parameter selection.

Chapter 3 proposes an alternative LSMC algorithm which directly approximates the optimal value function at each time step instead of the continuation function. This brings the benefits of faster convergence rate and closed-form expressions of the value function compared with the previously developed BSBU algorithm. We also develop a general argument for constructing an auxiliary stochastic control problem which inherits the continuity, monotonicity, and concavity of the original problem. This argument renders the LSMC algorithm circumvent extrapolating the value function in the backward recursion and can well adapt to other numerical methods.

Chapter 4 studies a complicated stochastic control problem: the no-arbitrage pricing of the "Polaris Choice IV" variable annuities issued by the American International Group. The Polaris allows the income base to lock in the high-water-mark of the investment account over a certain monitoring period which is related to the timing of the policyholders first withdrawal. By prudently introducing certain auxiliary state and control variables, we formulate the pricing problem into a Markovian stochastic optimal control framework. With a slight modification on the fee structure, we prove the existence of a bang-bang solution to the stochastic control problem: the policyholder's optimal withdrawal strategy is limited to a few choices. Accordingly, the price of the modified contract can be solved by the BSBU algorithm. Finally, we prove that the price of the modified contract is an upper bound for that of the Polaris with the real fee structure. Numerical experiments show that this bound is fairly tight.

Acknowledgements

Foremost, I would like to express my deep gratitude to my supervisor Professor Chengguo Weng. Without his selfless support, it would be impossible for me to complete the Ph.D. program in three years. We have known each other for more than six years and he has always been a considerate friend with whom I can talk heart to heart.

I would like to thank all the committee members: Professors Alexander Schied, Pengfei Li, Yuying Li, and Daniel Bauer (University of Wisconsin–Madison) for their time in reviewing my thesis and providing valuable comments.

This thesis has also been inspired by discussions with a number of people. I would like to thank Professor Zhaoxing Gao (Lehigh University) and Professor Degui Li (University of York) for stimulating discussions on nonparametric regression. I appreciate Professor Jan Palczewski (University of Leeds) for sending his research works. I am grateful to Doctor Parsiad Azimzadeh for his inspiring thesis work on computational finance. I thank Professor Yue Kuen Kwok (The Hong Kong University of Science and Technology) and Professor Pingping Zeng (Southern University of Science and Technology) for insightful discussions on option pricing. I appreciate Doctor Yao Tung Huang for introducing me to the field of variable annuities. Professor Yukun Liu (East China Normal University) brought me to the field of statistics and I have always been intrigued by the field; my thanks should also go to him.

I am indebted to Professors Yi Shen, Ruodu Wang, Alexander Schied, Phelim Boyle, and Ken Seng Tan for endless encouragement during my Ph.D. life and writing reference letters for me.

During my Ph.D. journey, a lot of friends have helped me in different ways and it is hard to imagine whether I could arrive at the destination without them. I would like to thank Wenju Jiang for giving me constructive career advice, improving my market sense, and stimulating me to stay positive in tough times. I am grateful to Yumin Wang for encouragement, patience, and inspiration. I thank Doctor Danqiao Guo and Professor Jinggong Zhang for the memorable companion at Waterloo. I thank Feiyu Zhu and Yi Shan for the discussions on C++ programming. I appreciate Bijun Sun, Doctor Lichen Chen, Doctor Jianfa Cong, Shenjia Xi, and Professor Zhenyu Cui for their generous help in my job seeking. I appreciate Yun Lin, Xiaobai Zhu, and Professor Mario Ghossoub for their time and roles in my mock interview. Special thanks go to Jie Zhang for her advice on the title of this thesis. I had a lot of good memories with the fellows in Waterloo: Lv Chen, Junhan Fang, Xiyue Han, Huameng Jia, Danping Li, Wenyuan Li, Hongcan Lin, Professor Peijun Sang, Jie Shen, Doctor Wei Wei, Jiayue Zhang, Kairuo Zhou, Ying Zuo, and those I forgot to mention. They are also greatly appreciated.

Special thanks go to Doctor Antti Pihlaja for his inspiring interview questions and constructive career advice. It was the interview experience with him that precipitated my determination of becoming a front-office quant.

Finally, I am indebted to my family for their unconditional love throughout life.

Dedication

To the memory of my grandmother.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

A variety of problems in insurance, finance, and operations research involve stochastic optimization. Notable examples include pricing financial derivative products with early-exercise features (see, e.g., [25], [58], and [81]), optimal inventory management (see, e.g., [80] and [68]), dynamic order execution (see, e.g., [66], [3], and [82]), high-frequency market making (see, e.g., [2] and [10]), and pricing variable annuities (see, e.g., [40] and [47]), among others. Such optimization problems are often formulated within a discrete-time Markovian stochastic optimal control framework where one decision-maker (DM) strives to optimize a certain *objective* by taking *action* over a lattice of time points. For example, in the problem of pricing variable annuity, a policyholder strives to optimize the monetary value of her cash inflows by taking withdrawals periodically from the investment account; see Figure 1.1 for a graphical illustration. By exploiting the prevailing Dynamic Programming Principle, solving a stochastic optimal control problem boils down to the investigation of the solution of a backward recursion equation which is known as the Bellman equation. The Bellman equations of many stochastic control models are not analytically tractable, and thus numerical solutions are naturally advocated. Over the past two decades, Monte Carlo simulation-based algorithms have been proposed in the literature to approach optimal solutions of stochastic control problems.

Since the ground-breaking works of [25], [58], and [81], the Least-Squares Monte Carlo (LSMC) algorithm, also known as the regression-based Monte Carlo method, has gained enduring popularity in solving optimal stopping problems, a special type of stochastic control problems. The associated convergence analysis has also been extensively studied in the literature, see, e.g., [30], [78], [44], [35], [36], [86], [87], and [88]. The LSMC has two building blocks: 1) a forward simulation of the state process and 2) an application of nonparametric regression to approximating

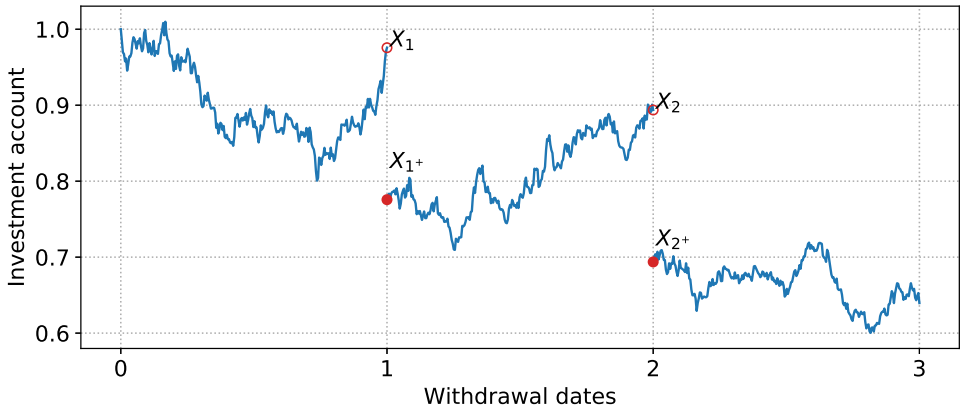


Figure 1.1: The evolution of the policyholder's investment account. Withdrawals cause instantaneous jumps of the account value across withdrawal dates.

the continuation value or the value function. In a statistical context, a nonparametric regression method aims to estimate certain conditional mean function in the following form:

$$g(\cdot) = \mathbb{E}\Big[\underbrace{Y}_{\text{Response}} \Big| \underbrace{X}_{\text{Covariate}} = \cdot \Big]. \tag{1.1}$$

In view of this, there is no surprise that it makes a perfect fit for approximating the continuation value and value function in an optimal stopping problem both of which can be expressed into the above conditional expectation form; see Section 3.2 of the subsequent Chapter 3.

Despite its popularity in solving optimal stopping problems, the application of the LSMC to a general stochastic control setting is much more arduous due to the challenges arising from several aspects.

1.1.1 Simulation

As pointed out previously, the simulation of the state process is an integral part of the LSMC algorithm. In stark contrast to optimal stopping problems, in a stochastic optimal control setting, the state variable at the present time step depends on the DM's action at the previous time step. Accordingly, the forward simulation of the whole trajectory of the state process calls for the specification of the DM's action at each time step. For example, in the context of the variable annuity, the policyholder's investment account value depends on how much she withdraws from the account at each withdrawal date; see Figure 1.1. This contravenes the fact that the DM's optimal action should be determined by solving the Bellman equation in a backward recursion

manner.

One popular way to circumvent this problem is to generate the DM's action either by a heuristic guess or a simulation from a certain probability distribution. Given the generated DM's action, the state process can be simulated smoothly in a forward manner as in the optimal stopping setting. Once the sample paths of the state process are simulated, a nonparametric regression method is employed to approximate either the continuation value or the value function and the optimal policy is updated by solving the local optimization associated with the Bellman equation in a backward recursion way. This forward simulation and backward updating argument is often referred to as the control randomization method ([51]) in the literature and has been used in the LSMC algorithm to solve a few stochastic control problems, see, e.g., [31] and [89] for dynamic portfolio optimization and [47] for pricing variable annuities. It is worth stressing that a similar idea appears earlier in [16] than [51] which develops an LSMC algorithm to solve a wide class of stochastic control problems by simulating the state process under a "reference measure". Although such expediency is intuitively appealing, as pointed out by [72], in some cases, a poor guess of the DM's action might lead the LSMC algorithm to miss the optimal solution because the sample distribution of the simulated state process highly relies on such an initial guess. To get a fleeting glimpse of this problem, let us revisit the instance of variable annuities: suppose the initial guess of the policyholder's withdrawal strategy is to deplete the investment account at a certain withdrawal date, then the investment account will get exhausted forever. Accordingly, the sample paths of this state variable do not evenly distribute over its feasible set but concentrate on a single point, which, in turn, skews the regression estimate for the continuation value.

1.1.2 Localization

In addition to the annoying issue of simulating the state process, the application of the LSMC algorithm to stochastic control problems is also bothered by the fact that the state variable dwells in an unbounded set. To be specific, on one hand, the accuracy of a numerical estimate for the value function is only ensured over a compact set say $\mathcal{D}$ because the regression estimate is merely a legitimate approximation of the continuation value over a compact set; see, e.g., [65] and [30]. On the other hand, solving the local optimization problem in the Bellman equation at time step t generally calls for the knowledge of the value function at time step $t+1$ over a set that is larger than $\mathcal{D}$.

It is notable that this dilemma also applies to other numerical methods, for example, the finite difference method and the Fourier transform method, as one needs to construct numerical grids on a compact set. Certain extrapolation techniques are often used as expediency for this problem; see, e.g., [4] and [48] in the context of using the Fourier transform method to solve the stochastic control problems accompanying variable annuities. Without a prudent and rigorous treatment on this intricate issue, the accuracy of the numerical estimate for the value function at

time zero is left in doubt as one has no idea about how the extrapolation error would propagate in the backward recursion procedure.

1.1.3 Regression

Different from usual statistical problems, the approximation of the continuation value of a stochastic control problem gives rise to new challenges. In the context of the LSMC, the sample size of the regression problem corresponds to the number of simulated paths of the state process which clearly is considerable. This observation renders most nonparametric regression methods computationally expensive or even prohibitive.

To be specific, the computational cost of prevailing nonparametric regression methods arise from three aspects: 1) computational complexity of model prediction[1], 2) memory cost of the regression estimate and 3) tuning parameter selection. For example, despite its faster convergence rate, the local polynomial regression method (see, e.g., [38]) requires storing all sample points in the memory and predicting the regression function at new point calls for rerunning a least-squares regression, which clearly is computationally luxury. This issue disturbs all local regression methods such as the kernel methods (see, e.g., [64] and [85]) and the isotonic regression method (see, e.g., [69]) which is a popular resolution to estimating monotone functions. The smoothing spline method does not need to store the sample points but the evaluation of the regression estimate is still time-consuming; see, e.g., [76]. The reproducing kernel Hilbert space (or reproducing kernel in short) method requires storing the training sample and the evaluating the regression estimate at a particular point involves M operations with M being the sample size; see e.g. [37, Eq. (4.2)]. It is also worth noting that almost all nonparametric regression methods require a tuning parameter selection procedure such as cross-validation which is required to avoid undesirable overfitting or underfitting. This ramps up the computational burden of fitting a regression model. For instance, the regression spline method (see, e.g., [71]) costs less computational time in the model prediction than the smoothing spline method but necessitate a prior selection of the number and position of the knots. The linear sieve estimation method (see, e.g., [65]) demands to optimize the number of basis functions in fitting the regression model. In Table 1.1, we summarize the pros (✓) and cons (×) of some popular nonparametric regression methods in terms of the computational issues mentioned above.

In addition to the issue of computational cost, it is notable that the regression estimate, in general, does not preserve some shape properties such as the concavity and monotonicity of the continuation value function and the optimal value function in an LSMC algorithm; see, e.g., [48, pp. 825–826] and [72]. Two undesirable issues arise accordingly. Firstly, the LSMC estimates for

[1]The model prediction here means computing the regression estimate at a particular point that is different from the sample points. This is intensively involved in an LSMC algorithm as one may see from the later chapters of this thesis.

Table 1.1: Pros (✓) and cons (×) of common nonparametric regression methods from a computational perspective.

Regression Method	Tuning Parameter	Computational Cost	
		Model prediction	Memory cost
Local polynomial	Bandwidth	×	×
Kernel regression	Kernel function, Bandwidth	×	×
Isotonic regression	No tuning parameter	×	×
Smoothing spline	Smoothing parameter	×	✓
Regression spline	Position of knots	✓	✓
Reproducing kernel	Kernel	×	×
Linear sieve	Number of basis functions	✓	✓
Neural network	Number of hidden layers, neurons in each layer	×	×

the continuation value and value function lose economic interpretation. Secondly, the concavity of the local optimization problem in the Bellman equation might be impaired, which makes the search for the optimizer fairly cumbersome; see Chapter 3 for a discussion. Shape-restricted regression methods have also been widely studied in statistical literature but they are also annoyed by the computational issues mentioned previously; see, e.g., [60].

1.2 Contributions and Road Map

In view of the aforementioned obstacles, the thrust of this thesis is to explore possible answers to the following research questions.

(Q1) Is the forward simulation imperative in an LSMC algorithm?

(Q2) Does the control randomization method always ensure the convergence of an LSMC algorithm to the optimal solution?

(Q3) How to develop a general truncation argument for discrete-time stochastic control problems?

(Q4) Is it possible to improve the efficiency of the regression methods by exploiting certain shape information of the continuation value and the optimal value function?

The subsequent chapters of this thesis respond to the above questions. Chapter 2 develops a novel regression-based Monte Carlo algorithm referred to as the Backward Simulation and Backward Updating (BSBU) algorithm. To be specific, this chapter shows that the answer to **(Q1)** is negative: it is not necessary to simulate the state process in a forward manner. In particular, Chapter 2 proposes to direct simulate the post-action value of the state process, which, in turn, eliminates the need of control randomization. This further brings several benefits in terms of memory cost and computational complexity. Furthermore, Chapter 2 gives one resolution to the problem of **(Q3)**. An auxiliary stochastic control model is constructed such that the accompanying state process is restrained in a compact domain. The corresponding optimal value function is shown to be a legitimate proxy for the original one, which enables one to switch the attention to the auxiliary problem for a solution. This bypasses undesirable extrapolation and the paves the way to applying the nonparametric regression technique of [65]. Moreover, in response to **(Q4)**, Chapter 2 proposes to employ a shape-preserving sieve estimation method to approximate the continuation value of the Bellman equation. The resulting regression estimate is less volatile and preserves certain economic interpretations as confirmed by numerical experiments, which provides an affirmative answer to **(Q4)**. Finally, the convergence result of the BSBU algorithm is established by resorting to the theory of nonparametric sieve estimation.

Chapter 3 develops an alternative regression-based Monte Carlo algorithm, referred to as the Regression-later Least Squares Monte Carlo (RL-LSMC) algorithm, by exploiting the regression-later technique of [43]. Specifically, the RL-LSMC algorithm aims to use a regression method to directly approximate the optimal value function instead of the continuation value. To this end, Chapter 3 gives a second resolution to **(Q3)** that is different with the one proposed in Chapter 2. By an elaborate construction of an auxiliary stochastic control model, the optimal value function of the auxiliary problem manages to inherit certain regularities of the original one such as the continuity, convexity/concavity, and monotonicity, which is not achievable by the counterpart in Chapter 2. This, in turn, brings several benefits to the implementation of the RL-LSMC algorithm. Chapter 3 further establishes the convergence result of the RL-LSMC algorithm which sheds lights on the choice of the sampling distribution and casts insights to the questions **(Q2)**. In particular, Chapter 3 shows that the continuity of the optimal value function plays an indispensable role in guaranteeing the legitimacy of direct simulation of the state process from an artificial probability distribution. Finally, Chapter 3 applies the RL-LSMC algorithm to the problem of hedging equity-linked insurance products, which shows the advantages of the RL-LSMC algorithm over the BSBU peer developed in Chapter 2 and pinpoints the primary motivation for using the regression-later technique.

Chapter 4 studies a convoluted stochastic optimal control problem: the no-arbitrage pricing of a new type of variable annuity contract, the Polaris variable annuities, issued by the American International Group (AIG). Specifically, by a slight modification of the fee scheme, a bang-bang solution to the stochastic control problem accompanying the pricing problem is proved to exist, that is, the optimal action at each time step is limited to a few explicit choices. This dramatically

reduces the complexity of the optimization problem involved in the Bellman equation. Consequently, the no-arbitrage price of the modified contract can be numerically approached by the BSBU algorithm developed in Chapter 2. We further show that the price function of the modified contract dominates that of the real policy and therefore gives a guide on the super-hedging cost of the policy writer. One crucial implication of this particular stochastic control problem is that the control randomization does not necessarily lead an LSMC algorithm to find the optimal solution and highlights the merits of the BSBU algorithm. This gives a negative answer to **(Q2)**. Chapter 4 also provides a new idea of solving the pricing problems of those variable annuity products in the absence of bang-bang solutions: one might first speculatively modify the contract provisions to construct a modified contract whose pricing problem is less formidable; and then show the subtle relation between the modified and original contracts.

Finally, Chapter 5 concludes the thesis and points out some future research directions.

Chapter 2

A Backward Simulation Method for Stochastic Optimal Control Problems

2.1 Introduction

The stochastic optimal control model is a prevalent paradigm for solving optimal decision problems with uncertainty in a variety of fields, particularly, financial engineering. In solving discrete-time stochastic optimal control problems, the Dynamic Programming Principle (DPP) is a prevailing tool which characterizes the optimal value function as the solution to a backward recursive equation system, often known as the *Bellman equation*. This dismantles the stochastic optimization problem into two separate problems: 1) solving a sequence of deterministic optimization problems and 2) evaluating the conditional expectation terms in the Bellman equation. In spite of the theoretical appealingness of the DPP, there generally does not exist a closed-form solution to the Bellman equation, which hampers the application of stochastic optimal control models to many intricate real-world problems. Recently, a number of numerical methods have been proposed in the literature to approach the optimal or suboptimal solutions to various stochastic optimal control problems by combining Monte Carlo simulation method with nonparametric regression methods.

In a statistical setting, the typical goal of nonparametric regression methods is to estimate the functional form of the expectation of a response variable conditioning on some covariate variables. This naturally motivates one to use certain nonparametric regression methods to evaluate the conditional expectation (also known as the continuation value in the context of pricing Bermudan option) involved in the Bellman equation where the optimal value function at the next time step and the state variable at the current time step are taken as the response and covariate variables, respectively. Such a ground-breaking idea was incubated in a series of papers including [25], [58],

and [81], and the corresponding numerical algorithms are often referred to as the Least-Squares Monte Carlo (LSMC) algorithms. Since then, the LSMC algorithm has witnessed remarkable popularity in solving optimal stopping problems, a special class of stochastic control problems; see, e.g., [30], [78], [44], [43], [35], [36], [86], [17], [15], [87], and the references therein.

Solving general stochastic control problems by resorting to the LSMC algorithm is considerably more involved. To understand the crux, let us note that the LSMC method has two building blocks: 1) a forward simulation of the state process and 2) a backward updating procedure which employs the nonparametric regression to estimate the continuation value. In an optimal stopping problem, the evolution of the state process is independent of decision maker's (DM's) action and therefore, the forward simulation of the sample paths of the state process is relatively straightforward. In stark contrast to this, in a general stochastic optimal control setting, the state process is influenced by the DM's action and accordingly, its simulation is unattainable without specifying the DM's action. Ideally, one may expect to simulate the state process driven by the optimal action of the DM. However, the optimal action should be determined by solving the Bellman equation in a backward recursion manner, which is incongruous with the need of forward simulation in an LSMC algorithm. To bypass this hurdle, [51] proposes to first draw the DM's action from a random distribution and then simulate the sample paths of the state process based on the initialized action. This method has been referred to as the control randomization method in the literature and applied in the LSMC algorithm to solve many specific stochastic control problems; see, e.g., [31], [89], and [47], among others. Despite the wide usage of the control randomization method, the accuracy of the numerical estimate is impaired over the region with sparse sample points ([89, Section 3.3]) and the spread of the sample paths is sensitive to the specific way of initializing the action. In an extreme case, the LSMC algorithm might even miss the optimal solution under a dismal choice of random distribution from which the DM's action are drawn; see [72], for instance. It is also notable that most literature bind together the control randomization and the forward simulation of the state process. However, this chapter will show that the forward simulation is not imperative in an LSMC algorithm and the merits of abjuring the forward simulation are extant in several aspects. The limitations of the control randomization and the consequential forward simulation will be elaborated in the subsequent Section 2.2.2.

Besides the simulation of the state process, the approximation of the conditional expectation term in a Bellman equation is also an arduous task for several reasons. Firstly, the prevalent regression methods only warrant the accuracy of the regression estimate over a compact support, see, e.g., [65], [78], and [87], whereas the state variable generally takes value in an unbounded set. Some literature compromise to first truncate the domain of the continuation function and then use extrapolation techniques when the knowledge of the function outside the truncated region is required. It is worth noting that this problem is not acute in the context of optimal stopping problem but is severe in a general stochastic control setting. This is because, in the latter case, one has to traverse through all admissible actions, which calls for the values of the continuation function over a domain that is wider than spreading range of sample paths; see

Figure 2.2 for a graphical illustration. Secondly, in order to avoid overfitting or underfitting, most nonparametric regression methods thirst for an appropriate choice of the tuning parameter, e.g., the number of basis functions in a linear sieve estimation method (see the subsequent Section 2.4.3). This is often resolved by computationally expensive cross-validation methods, see, e.g., [56]. However, in view of the extraordinarily large number of simulated paths, such a tuning parameter selection procedure is computationally intensive in implementing the LSMC algorithm. The aforementioned challenges will be investigated in details in the subsequent section.

The contribution of this chapter is summarized as follows. Firstly, we propose to restrain the value set of the state process into a compact set, which evades the undesirable extrapolating the optimal value function estimate during the backward recursion of the LSMC algorithm. The value function accompanying the truncated state process is shown to be a legitimate approximation for the original value function under a suitable choice of the truncation parameter. Secondly, we generalize the idea of [72] to simulate the post-action value of the state process from an artificial probability distribution. This eliminates the need for the forward simulation and is consistent with the backward induction nature of the Bellman equation. The memory as well as time costs of the artificial simulation method are considerably less than those of the control-randomization-based forward simulation method. Thirdly, a shape-preserving sieve estimation method is introduced to approximate the conditional expectation term involved in the Bellman equation. By exploiting certain shape information of the continuation function, the sieve estimate is insensitive to the tuning parameter and accordingly reduces the computational cost of the tuning parameter selection. We refer to the proposed LSMC algorithm as the Backward Simulation and Backward Updating (BSBU) algorithm. Finally, we establish the convergence result of BSBU algorithm which sheds light on how the numerical error propagates over the backward recursion procedure.

This chapter is organized as follows. Section 2.2 gives a tour through the LSMC algorithm and puts forward the motivations of the chapter. Section 2.3 constructs the auxiliary stochastic optimal control model. Section 2.4 develops the BSBU algorithm and establishes the associated convergence result. Section 2.5 applies the BSBU algorithm to the pricing problem of an equity-linked insurance product and Section 2.6 conducts the corresponding numerical experiments. Finally, Section 2.7 concludes the chapter.

2.2 Basic Framework and Motivations

2.2.1 Stochastic Optimal Control Model

We restrict our attention to a collection of consecutive time points labeled by $\mathcal{T} := \{0, 1, \ldots, T\}$ on which a decision maker (DM) may take action. The uncertainty faced by the DM is formulated by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in\mathcal{T}}$. The DM's action is described by a discrete-time stochastic process $\mathsf{a} = \{a_t\}_{t\in\mathcal{T}_0}$ valued in $\mathbb{A} \subseteq \mathbb{R}^p$ with $\mathcal{T}_0 = \mathcal{T}\backslash\{T\}$

and $p \in \mathbb{N}$. Let $X = \{X_t\}_{t\in\mathcal{T}}$ be a certain state process valued in $\mathcal{X} \subseteq \mathbb{R}^d$ with $d \in \mathbb{N}$. Starting from an initial state $X_0 = x_0 \in \mathcal{X}$, it evolves recursively according to the following transition equation:

$$X_{t+1} = S\left(X_t, a_t, \varepsilon_{t+1}\right), \quad \text{for} \quad t = 0, 1, \ldots, T-1, \tag{2.1}$$

where $\varepsilon := \{\varepsilon_{t+1}\}_{t\in\mathcal{T}_0}$ is a sequence of independent random variables valued in $\mathbb{D} \subseteq \mathbb{R}^q$ with $q \in \mathbb{N}$. ε_{t+1} reflects the uncertainty faced by the DM at time step t and is referred to as *random innovation* in what follows. For brevity of notation, in what follows, the dependency of the state process on the action is compressed and the readers should always bear in mind that X_t implicitly depends on the DM's action up to time $t-1$.

Let us restrict our attention to the following admissible set of the DM's action:

$$\mathcal{A} = \left\{ \mathsf{a} = \{a_t\}_{t\in\mathcal{T}_0} \;\middle|\; a_t \text{ is } \mathcal{F}_t\text{-measurable and } a_t \in A_t\left(X_t\right) \text{ for } t \in \mathcal{T}_0 \right\},$$

where $A_t(\cdot)$ is a set valued map (referred to as the *correspondence* in this thesis), that is, $A_t(x)$ is a subset of $\mathbb{A}$ for each $x \in \mathcal{X}$. $A_t(X_t)$ corresponds to a state-dependent constraint imposed on the DM's action at time step t.

Now consider a discrete-time stochastic optimal control problem in the following form:

$$V_0(x_0) = \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t f_t(X_t, a_t) + \varphi^T f_T\left(X_T\right)\right], \tag{2.2}$$

where $\varphi \in (0,1)$ is a certain constant discounting factor, $f_t(\cdot,\cdot)$ and $f_T(\cdot)$ are the intermediate and terminal reward functions, respectively. In order to ensure the well-posedness of the stochastic control problem (2.2), we impose the following assumption which is conventional in literature, see [70] and [16] for instance.

Assumption 2.1.

$$\sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left|f_t(X_t, a_t)\right|\right] < \infty, \quad \text{and} \quad \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\left|f_T(X_T)\right|\right] < \infty.$$

The Dynamic Programming Principle states that the value function $V_0(\cdot)$ can be solved recursively:

$$\begin{cases} V_T(x) &= f_T(x), \\ V_t(x) &= \sup\limits_{a\in A_t(x)} \left[f_t(x,a) + \varphi \bar{C}_t(x,a)\right], \quad \text{for} \quad t = 0, 1, \ldots, T-1, \end{cases} \tag{2.3}$$

where

$$\bar{C}_t(x,a) = \mathbb{E}\left[V_{t+1}\left(X_{t+1}\right)\Big|X_t = x, a_t = a\right]. \tag{2.4}$$

We proceed by rewriting the transition equation (2.1) into the following form:

$$S(X_t, a_t, \varepsilon_{t+1}) = H\big(K(X_t, a_t), \varepsilon_{t+1}\big), \tag{2.5}$$

where $H(\cdot,\cdot) : \mathbb{R}^{r+q} \longrightarrow \mathbb{R}^d$ and $K(\cdot,\cdot) : \mathbb{R}^{d+p} \longrightarrow \mathbb{R}^r$ are some measurable functions with $r \in \mathbb{N}$. It is worth stressing that any transition function $S(\cdot,\cdot,\cdot)$ can be rewritten into the above form since one may choose $K(\cdot,\cdot)$ as identity function (i.e., $K(x,a) = (x,a)$) and the above equation holds trivially. Nevertheless, it is instructive to introduce the function $K(\cdot,\cdot)$ as it brings the benefit of dimension reduction, which we will explain with more details in the sequel. Combing Eqs. (2.1), (2.4) and (2.5) gives

$$\bar{C}_t\left(X_t, a_t\right) = \mathbb{E}\left[V_{t+1}\big(H\left(X_{t^+}, \varepsilon_{t+1}\right)\big)\Big|X_{t^+} = K\left(X_t, a_t\right)\right].$$

Hereafter, X_{t^+} is referred to as the *post-action* value of the state process X_t at time t. It constitutes an essential component in the LSMC algorithm proposed in Section 2.4. Define function

$$C_t(k) := \mathbb{E}\Big[V_{t+1}\left(X_{t+1}\right)\Big|X_{t^+} = k\Big] = \mathbb{E}\Big[V_{t+1}\big(H\left(k, \varepsilon_{t+1}\right)\big)\Big]. \tag{2.6}$$

We observe the following relationship between $\bar{C}_t(\cdot,\cdot)$ and $C_t(\cdot)$:

$$\bar{C}_t(x,a) = C_t\big(K(x,a)\big). \tag{2.7}$$

The crucial implication of the above relation is that it suffices to recover the functional form of $C_t(\cdot)$ in order to evaluate $\bar{C}_t(\cdot,\cdot)$ since $K(\cdot,\cdot)$ is known at the first hand. The motivation of rewriting the transition equation into Eq. (2.5) is now clear: $K(\cdot,\cdot)$ maps a $(d+p)$-dimensional vector into a r-dimensional vector which compresses the dimension if $r < d+p$, and it is more efficient to recover the function $C_t(\cdot)$ than $\bar{C}_t(\cdot,\cdot)$ due to such a dimension reduction. It is also worth noting that $C_t(\cdot)$ is solely determined by the probability distribution of ε_{t+1} according to Eq. (2.6), and *it is not necessary to know the exact distribution of X_{t^+} in the evaluation of the function $C_t(\cdot)$*. In view of the relation (2.7), the Bellman equation (2.3) can be equivalently written as

$$\begin{cases} V_T(x) &= f_T(x), \\ V_t(x) &= \sup\limits_{a\in A_t(x)}\Big[f_t(x,a) + \varphi C_t\big(K(x,a)\big)\Big], \quad \text{for} \quad t = 0, 1, \ldots, T-1. \end{cases} \tag{2.8}$$

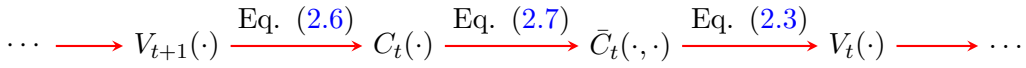


Figure 2.1: A diagram for backward information propagation in solving the Bellman equation.

The above equation system states that, given the value function at time step $t+1$, one may first evaluate continuation function according to Eqs. (2.6) and (2.7) and then obtain the value function at time step t via solving an optimization problem in the second line of Eq. (2.8). The information propagation behind the above recursive procedure is illustrated in Figure 2.1.

2.2.2 A Tour Through LSMC Algorithm

We proceed by briefly reviewing the Least-squares Monte Carlo (LSMC) algorithm. We will show its limitations in several aspects which motivate the algorithm we will propose in the subsequent sections.

"Forward simulation and backward updating" (FSBU) algorithm

There has been voluminous literature on the LSMC for optimal stopping problem, while the literature on the LSMC for general stochastic optimal control problem is thin. Most literature addresses the LSMC for the stochastic control problems arising in some specific applications, see, e.g., [24], [11], [47], [72], [31], and [89], among others. An LSMC algorithm for a class of stochastic control problem is developed in [16].

For most variants of the LSMC algorithm, they can be decomposed into two pillars: (i) a forward simulation of the state process and (ii) a backward updating of control policies. We review these algorithms in a unified paradigm as follows.

1. **Initiation:** Set $V_T^{\mathrm{E}}(x) = f_T(x)$. For $t = T-1, T-2, \ldots, 0$, do the two steps below.
2. **Forward Simulation:**
   2.1 **Control randomization** Generate a random sample of the DM's action up to time step t:
   $$\mathsf{a}_{0:t}^M := \left\{ \left(a_0^{(m)}, \ldots, a_t^{(m)} \right),\ m = 1, 2, \ldots, M \right\}$$
   generated by a certain heuristic rule.

2.2 **Simulation of state process** Simulate a random sample of the random innovations:

$$\left\{\left(\varepsilon_1^{(m)},\ldots,\varepsilon_{t+1}^{(m)}\right),\ m=1,2,\ldots,M\right\}.$$

The sample of the state process up to time step $t+1$ is given by

$$\mathbf{X}_{1:t+1}^{M}:=\left\{X_{1:t+1}^{(m)}:=\left(X_1^{(m)},\ldots,X_{t+1}^{(m)}\right),\ m=1,2,\ldots,M\right\},$$

where $X_n^{(m)}=S\left(X_{n-1}^{(m)},a_{n-1}^{(m)},\varepsilon_n^{(m)}\right)$ for $n=1,2,\ldots,t+1$.

3. **Backward Updating:**

3.1 **Regression** Given a numerical estimate of value function at time step $t+1$, denoted by $V_{t+1}^{\mathrm{E}}(\cdot)$, construct the random sample

$$\mathbf{Y}_{t+1}^{M}:=\left\{V_{t+1}^{\mathrm{E}}\left(X_{t+1}^{(m)}\right),m=1,2,\ldots,M\right\}.$$

Further construct a random sample of post-action value of the state process as follows:

$$\mathbf{X}_{t^+}^{M}:=\left\{X_{t^+}^{(m)}:=K\left(X_t^{(m)},a_t^{(m)}\right),m=1,2,\ldots,M\right\}. \tag{2.9}$$

Take $\mathbf{Y}_{t+1}^{M}$ and $\mathbf{X}_{t^+}^{M}$ as the samples of response variable and regressor, respectively, and employ a certain non-parametric regression to obtain a regression estimate $C_t^{\mathrm{E}}(\cdot)$ for $C_t(\cdot)$.

3.2 **Optimization** An estimate for the value function at time step t is given by

$$V_t^{\mathrm{E}}(x)=\sup_{a\in A_t(x)}\Big[f_t(x,a)+\varphi C_t^{\mathrm{E}}\big(K(x,a)\big)\Big]. \tag{2.10}$$

We henceforth call the above algorithm as the Forward Simulation and Backward Updating (FSBU) algorithm since the simulation of $\mathbf{X}_{t^+}^{M}$ involves a forward procedure.

Remark 2.1 (Randomness of $V_t^{\mathrm{E}}(\cdot)$ and $C_t^{\mathrm{E}}(\cdot)$)**.** *The superscript* E *in* $V_t^{\mathrm{E}}(\cdot)$ *and* $C_t^{\mathrm{E}}(\cdot)$ *stresses that they are numerical estimates of the true value function and continuation function, respectively. Since a certain regression technique is employed to get such numerical estimates, they essentially depend on the random samples* $\mathbf{Y}_{t+1}^{M}$ *and* $\mathbf{X}_{t^+}^{M}$ *and hence on all previously generated random samples going from step* $T-1$ *down to step* t*, i.e.,* $\mathbf{Y}_{n+1}^{M}$ *and* $\mathbf{X}_{n^+}^{M}$ *for* $n=t,\ldots,T-1$*. Such dependency is suppressed in notation for brevity, but the readers should keep in mind that both* $V_t^{\mathrm{E}}(\cdot)$ *and* $C_t^{\mathrm{E}}(\cdot)$ *are random functions.*

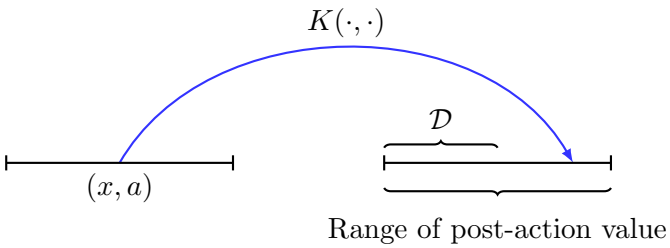


Figure 2.2: A diagram illustrating the map $K(\cdot,\cdot)$ relating pre-action value to the post-action value.

Challenges

There are several challenges in implementing the FSBU algorithm to solve a stochastic control problem. Some comments are made on the challenges from three aspects.

(I) Limitation of control randomization As the DM's optimal action is not tractable priori but should be solved in the backward updating stage of the algorithm, Step 2.1 randomly generates a feasible action, which is referred to as *control randomization* method; see [51]. For some selected action $\mathsf{a}_{0:n}^M$, the accuracy of the regression estimate $C_t^{\mathrm{E}}(\cdot)$ can be warranted only over the support of the resulting sampling points $\mathbf{X}_{t^+}^M$, say $\mathcal{D}$, which might be smaller than those for other actions. This is more perceivable from Figure 2.3: since most sample points distribute over the interval $[0, 0.8]$, the regression estimates agree with each other over this region but vary substantially over $[0.8, 1.0]$. In view of this, the accuracy of the regression estimates over $[0.8, 1.0]$ is left in doubt. In fact, the true regression function in this experiment is a polynomial with the maximal degree of 3. On the other hand, in order to solve the optimization problem in Step 3.2 (see Eq. (2.10)), one requires the knowledge of $C_t^{\mathrm{E}}(\cdot)$ over the range of post-action value $X_{t^+} = K(X_t, a_t)$ for all feasible actions a_t because all possible values of the action should be invoked and taken as the input of the function $C_t^{\mathrm{E}}(K(x,a))$ in evaluating $V_t^{\mathrm{E}}(x)$; see Figure 2.2 for a graphical illustration. As a compromise, one may use certain extrapolation methods to infer the value of $C_t^{\mathrm{E}}(\cdot)$ outside the region $\mathcal{D}$, which incurs extra error and is hard to justify its legitimacy.

(II) Cost of forward simulation It is notable that, at time step t of the above FSBU algorithm, a new random sample of the state process that is *independent* of the sample at time step $t+1$ is simulated; see Figure 2.4 for a graphical illustration. This is required in order to apply the nonparametric regression theory to establish the convergence result; see, e.g., [87, p. 511] and [16]. On the contrary, using a single sample causes in-sample bias because the numerical estimate of value function obtained at time step $t+1$ $V_{t+1}^{\mathrm{E}}(\cdot)$ is correlated with $\mathbf{X}_{t^+}^M$; see e.g. [29, Section 3.1] and the earlier Remark 2.1. The total

time cost in a forward simulation procedure of the above LSMC algorithm is of $O(T^2)$[1]. Simulating the whole path of state process can be time-consuming especially when one uses some approximation schemes to simulate general stochastic differential equations[2]. Besides the issue of time cost, the memory cost in a single simulation is of $O(dT)$ with T and d being the number of time steps and dimensionality of the state process, respectively, which is sizable for a large T.

(III) Choice of regression technique Despite the voluminous literature on nonparametric regression, the choice of the nonparametric regression method in Step 4.1 should be meticulous. In the above FSBU algorithm, the sample size in the regression problem corresponds to the number of simulated paths and is generally recommended in the literature be chosen larger than one hundred thousand, which makes most regression methods computationally *prohibitive*. Specifically, the local methods such as local-polynomial regression are clearly not wise choices as they require running a regression at each sample point. It is worthy to point out that even computing a *single* point in $\mathbf{Y}_t^M$ is fairly time-consuming as it involves a local optimization problem (see Eq. (2.10)). Furthermore, the nuisance of high memory cost also burdens most nonparametric regression methods. For example, the kernel regression and isotonic regression methods require storing all sample points in order to recover the functional form of the regression function over some support, and the memory cost is *extraordinarily* large accordingly. The above two thorny issues escalate by noting that almost all nonparametric regression techniques involve a computationally-intensive cross-validation procedure to determine the tuning parameter (e.g., the bandwidth in local regression methods and the number of basis functions in global regression methods) in order to avoid overfitting or underfitting.

Motivations

In view of the previous items (I)–(III), the thrust behind this chapter is to explore possible answers to the following questions:

(Q1) *How to avoid the theoretically shaky extrapolation step?*

(Q2) *Is it possible to bypass the forward simulation in an LSMC algorithm?*

(Q3) *Is there a regression method that is insensitive to tuning parameter?*

[1]Suppose the time cost of simulating a path over each time interval $[t, t+1]$ is $\mathcal{C}$. Then the time cost of simulating a whole path up to time step n is about $n \times \mathcal{C}$, and the forward simulation in the whole LSMC algorithm has approximate time cost of $\mathcal{C}(1+2+\cdots+T) = T(T+1)\mathcal{C}/2$, accordingly.

[2]It is the authors' experience that a single simulation of 10^5 paths of the Heston model over a 10-year period takes 365 seconds by using the R package "`msde`" on a MacBook Pro (2.8 GHz Intel Core i7).

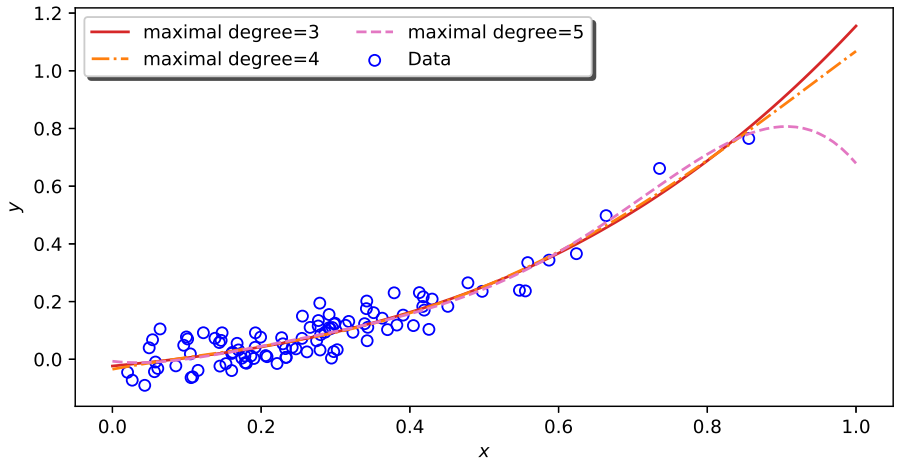


Figure 2.3: Regression estimates of polynomial regression method with different maximal degrees.

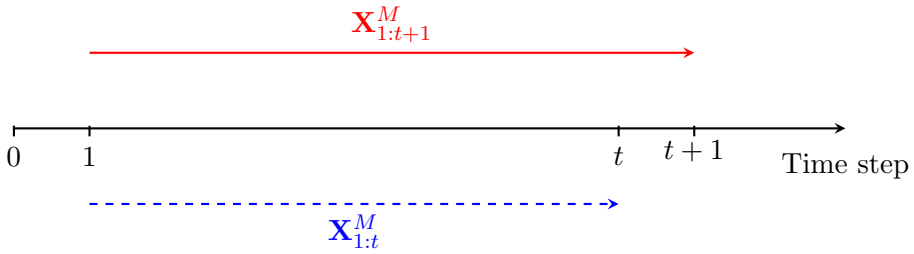


Figure 2.4: A diagram for illustrating the forward simulation of the state process. The solid line corresponds to a simulated sample of state process at time step t of the LSMC algorithm. The dashed line corresponds to a new *independent* simulated sample of the state process at time step $t-1$.

Now we present the road map of the remainder of the chapter. In terms of **(Q1)**, in the subsequent section, we will construct an auxiliary stochastic control problem where the accompanying state process only takes values in a bounded set. This enables us to sidestep extrapolating the regression function outside the region where the sample distributes. In response to **(Q2)**, We will propose to directly simulate the post-action value of state process in Section 2.4. For **(Q3)**, we will introduce a shape-preserving sieve estimation method to infer the continuation function. The resulting sieve estimate, on one hand, is insensitive to the tuning parameter, and on the other hand, preserves certain shape properties of the continuation function.

2.3 An Auxiliary Stochastic Control Problem

As commented in the item **(I)** "Limitation of control randomization" in the previous section, it is necessary to know the value of the continuation function over the whole range of the post-action value of the state process which is wider than the set where the regression sample suffuse. It is notable that the range of post-action value is unbounded if the state process takes value in an unbounded set, which is particularly the case in many finance applications. Therefore, it is generally inevitable to infer the continuation function outside the support of the sample and the error incurred by extrapolating the regression estimate is hard to quantify. Furthermore, as one will see in Section 2.4.3, the convergence of the regression method that is used to approximate the continuation function is only ensured over a compact domain.

In view of the above discussion, the thrust of this section is in dual-fold: first, we target to find a certain way to circumvent the unsound extrapolation in the implementation of an LSMC algorithm; second, we aim at confining the domain of the continuation function into a compact set. To realize this goal, we first construct an auxiliary stochastic optimal control problem where the accompanying state process only takes values in a bounded set and then show the discrepancy between the auxiliary problem and the original one is quantifiable and is marginal under certain conditions. It is worth stressing that the aforementioned challenges also burden other numerical algorithms for solving discrete-time stochastic control problems besides the LSMC such as the finite difference method. Therefore, the construction of the auxiliary problem can adapt to such numerical methods and is of independent interest.

2.3.1 Construction

Let $\mathcal{X}_R$ be a bounded subset of the set $\mathcal{X}$ where the subscript R denotes a certain truncation parameter. We may choose $\mathcal{X}_R$ such that its closure, $\mathrm{cl}\,(\mathcal{X}_R)$, is strictly convex. Further denote $\mathring{\mathcal{X}}_R$ (resp. $\partial\mathcal{X}_R$) as the interior (resp. boundary) of $\mathcal{X}_R$. Given the initial state $x_0 \in \mathring{\mathcal{X}}_R$, define

the following stopping time:

$$\tau^R := \inf\left\{t \in \mathcal{T} \;\middle|\; X_t \notin \mathring{\mathcal{X}}_R\right\}, \tag{2.11}$$

with the convention: $\tau^R = \infty$ if $X_t \in \mathring{\mathcal{X}}_R$ for all $t \in \mathcal{T}$.

We recursively define an auxiliary state process $\tilde{X} := \left\{\tilde{X}_t\right\}_{t\in\mathcal{T}}$ as follows:

$$\begin{cases} \tilde{X}_0 & = x_0, \\ \tilde{X}_t & = X_t \mathbb{1}_{\{\tau^R > t\}} + \mathcal{Q}\left(X_{\tau^R \wedge t}\right) \mathbb{1}_{\{\tau^R \leq t\}}, \quad \text{for} \;\; t = 1, 2, \ldots, T, \end{cases} \tag{2.12}$$

where $\mathcal{Q}(x) = \arg\inf_{y\in\mathrm{cl}(\mathcal{X}_R)} |y - x|$ with $|\cdot|$ denoting the Euclidean ℓ_2-norm.[3] Since $\mathrm{cl}\left(\mathcal{X}_R\right)$ is a compact and strictly convex set, $\mathcal{Q}(x)$ is unique and lies on the boundary set $\partial\mathcal{X}_R$ for $x \notin \mathring{\mathcal{X}}_R$.

Below we give some interpretations regarding the auxiliary state process defined in the above Eq. (2.12). The original state process X coincides with the auxiliary state process $\tilde{X}$ until the stopping time τ^R. Once the original state process passes through the interior of the truncated domain, the auxiliary state process freezes at a certain point in the boundary set $\partial\mathcal{X}_R$ thereafter. The evolution mechanisms of the original and auxiliary state processes are illustrated in Figure 2.5. The following proposition gives the transition equation of $\tilde{X}$.

Proposition 2.1. *The auxiliary state process $\tilde{X}$ defined by Eq. (2.12) admits the following transition equation across each time point: $\tilde{X}_0 = X_0$ and*

$$\tilde{X}_{t+1} = \tilde{X}_t \mathbb{1}_{\left\{\tilde{X}_t \in \partial\mathcal{X}_R\right\}} + \tilde{H}\Big(K\left(\tilde{X}_t, a_t\right), \varepsilon_{t+1}\Big) \mathbb{1}_{\left\{\tilde{X}_t \in \mathring{\mathcal{X}}_R\right\}}, \tag{2.13}$$

for $t = 0, 1, \ldots, T-1$, where

$$\tilde{H}(k, e) = \mathcal{Q}(H(k, e)), \tag{2.14}$$

and $K(\cdot,\cdot)$ is the transition equation relating the pre-action and post-action values of the original state process defined in Eq. (2.5).

The proof of the above proposition is relegated to Appendix A.3.1. The preceding Eq. (2.13) essentially states that $\tilde{X}$ is a Markov chain by itself, and accordingly, it is the sole state process of the auxiliary stochastic control model defined in the sequel.

Let $\tilde{\mathcal{A}}$ be the set of all admissible actions for the auxiliary state process which is defined as:

$$\tilde{\mathcal{A}} := \left\{\mathsf{a} = \{a_t\}_{t\in\mathcal{T}_0} \;\middle|\; a_t \text{ is } \mathcal{F}_t - \text{measurable}, \; a_t \in A_t\left(\tilde{X}_t\right), \; \text{for } t \in \mathcal{T}_0\right\}.$$

[3] For the brevity of notation, this chapter does not distinguish the notations for the Euclidean ℓ_2-norm of a scalar and a vector.

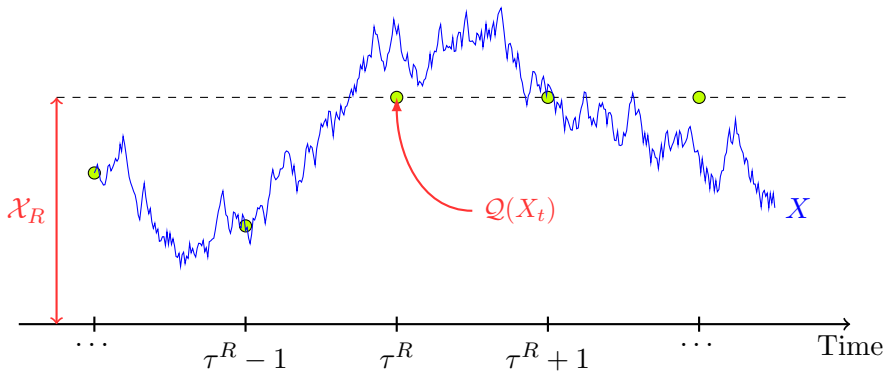


Figure 2.5: Graphical illustration of the evolution mechanisms of X and $\tilde{X}$. It is notable that X might evolve continuously between two discrete time points t and $t+1$. The stopping time τ^R corresponds to the first time point upon which X_t stays outside of $\mathring{\mathcal{X}}_R$ among all discrete time points $\{0, 1, \ldots, T\}$. The circles correspond to a path of $\tilde{X}$.

In parallel to the original stochastic optimal control problem (2.2), we consider the following auxiliary problem:

$$\tilde{V}_0(x_0) = \sup_{\mathsf{a} \in \tilde{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t f_t\left(\tilde{X}_t, a_t\right) + \varphi^T f_T\left(\tilde{X}_T\right)\right], \tag{2.15}$$

where $\tilde{X} = \left\{\tilde{X}_t\right\}_{t \in \mathcal{T}}$ is defined recursively by Eq. (2.13) for any given action a.

Remark 2.2. *It is worth stressing that in the above auxiliary problem (2.15) the DM's action is taken from the admissible set $\tilde{\mathcal{A}}$ that is different with the preceding set $\mathcal{A}$ accompanying the original stochastic control problem (2.2). This is resulted from the discrepancy between the state processes X and $\tilde{X}$. Subtle as the difference is, it will bring a technical difficulty in characterizing the gap between $V_0(x_0)$ and $\tilde{V}_0(x_0)$; see the proof of the subsequent Theorem 2.1 in Appendix A.3.2.*

Since the state process $\tilde{X}$ freezes once it reaches the boundary set $\partial\mathcal{X}_R$, the value function in Eq. (2.15) is given by

$$\tilde{V}_t(x) = \sum_{n=t}^{T-1} \varphi^{n-t} f_n\big(x; a_n^*(x)\big) + \varphi^{T-t} f_T(x), \quad \text{for} \quad x \in \partial\mathcal{X}_R,\ t \in \mathcal{T}, \tag{2.16}$$

with $a_n^*(x) \in \arg\sup_{a \in A_n(x)} f_n(x; a)$. Over the interior of the truncated domain, the above value

function $\tilde{V}_0(\cdot)$ can be solved in a similar backward recursion way as $V_0(\cdot)$ does, that is,

$$\begin{cases} \tilde{V}_T(x) & = f_T(x), \\ \tilde{V}_t(x) & = \sup\limits_{a\in A_t(x)} \Big[f_t(x,a) + \varphi \tilde{C}_t\big(K(x,a)\big) \Big], \quad \text{for } x \in \mathring{\mathcal{X}}_R,\ t = 0,1,\ldots,T-1, \end{cases} \tag{2.17}$$

where $\tilde{C}_t(\cdot)$ is defined in line with Eq. (2.6) with $H(\cdot,\cdot)$ replaced by $\tilde{H}(\cdot,\cdot)$. It is worth noting that, in evaluating $\tilde{C}_t\big(K(x,a)\big)$, the knowledge of $\tilde{V}_{t+1}(\cdot)$ over $\partial\mathcal{X}_R$ might be in need, and in such a situation, Eq. (2.16) is invoked.

We make some comparisons between the above Eq. (2.17) and the original Bellman equation (2.8). Firstly, in both equations, the state constraint $A_t(\cdot)$, the transition equation between pre-action and post-action values $K(\cdot,\cdot)$, and the reward functions are exactly the same. Secondly, the value function $\tilde{V}_t(\cdot)$ is solely defined on a bounded set $\mathrm{cl}\,(\mathcal{X}_R)$, whilst $V_t(\cdot)$ is defined on the set $\mathcal{X}$ which might be unbounded in many financial applications as the original state process X may correspond to a certain risky asset valued on the whole positive real line.

2.3.2 Error Analysis

In the following, we will characterize the discrepancy between $\tilde{V}_0(x_0)$ and $V_0(x_0)$. To this end, it is necessary to impose some assumptions on the state process and the reward functions.

Assumption 2.2. *Let $\tilde{X}_0 = X_0 \in \mathring{\mathcal{X}}_R$. There exists a measurable function $\mathcal{E}_T(\cdot,\cdot) : \mathring{\mathcal{X}}_R \times \mathbb{R}_{>0} \longrightarrow [0,1]$ satisfying*

$$\inf_{\mathsf{a}\in\mathcal{A}} \mathbb{P}\Big[X_t = \tilde{X}_t \ \text{ for all } \ t \in \mathcal{T} \Big] \geq 1 - \mathcal{E}_T(x_0, R). \tag{2.18}$$

$\mathcal{E}_T(x_0,R)$ in Eq. (2.18) gives an upper bound for the probability that the auxiliary state process disagrees with the original one at some time before maturity regardless of the DM's action. Since the primary difference between the auxiliary and original value functions stems from the disparity between the associated state processes, it is not surprising that the above inequality (2.18) plays an important role in characterizing the approximation error of $\tilde{V}_t(\cdot)$ as one will see later in the proof of Theorem 2.1. The expression of $\mathcal{E}_T(x_0,R)$ should be specified for each specific application; see Appendix A.2.1 for such an expression under the example of Section 2.5.

Assumption 2.3. (i) *There exists a measurable function $B(\cdot) : \mathbb{R}^d \longrightarrow \mathbb{R}_{>0}$ and a constant ζ such that*

$$\big|f_T(x)\big|^2 \leq B(x), \quad \sup_{a\in A_t(x)} \big|f_t(x,a)\big|^2 \leq B(x), \quad \text{and} \quad \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[B(X_{t+1})\right] \leq \zeta \tag{2.19}$$

for all $t \in \mathcal{T}_0$.

(ii) *There exists a measurable function* $\xi(\cdot) : \mathbb{R}_{>0} \longrightarrow \mathbb{R}_{>0}$ *such that*

$$\sup_{x \in \mathrm{cl}(\mathcal{X}_R)} \left| f_T(x) \right|^2 \leq \xi(R) \quad \text{and} \quad \sup_{x \in \mathrm{cl}(\mathcal{X}_R)} \left(\sup_{a \in A_t(x)} \left| f_t(x,a) \right|^2 \right) \leq \xi(R), \quad \text{for all} \quad t \in \mathcal{T}_0.$$

The existence of functions $B(\cdot)$ and $\xi(\cdot)$ is not hard to expect if (i) for each $x \in \mathcal{X}$, $A_t(x)$ is a compact set, and (ii) the reward function $f_t(x,a)$ in continuous in a. Their expressions are easy to specify in concrete applications; see Appendix A.2.1 for the verification of the above assumption in the context of Section 2.5.

The following theorem derives the legitimacy of using the auxiliary problem (2.15) as a proxy for the original problem (2.2). Its proof is relegated to Appendix A.3.2.

Theorem 2.1 (Truncation Error Bound)**.** *Suppose Assumptions 2.1, 2.2, and 2.3 hold. Then*

$$\left| V_0(x_0) - \tilde{V}_0(x_0) \right| \leq (T+1) \sqrt{2\big(\xi(R) + \zeta\big) \mathcal{E}_T(x_0, R)}. \tag{2.20}$$

Remark 2.3. *The primary difficulty of proving the above theorem stems from the nuance between the admissible sets associated with* $V_0(x_0)$ *and* $\tilde{V}_0(x_0)$*; see Eqs. (2.2) and (2.15) and the earlier Remark 2.2. This thorny issue is bypassed by showing that there is no loss in replacing the admissible set* $\tilde{\mathcal{A}}$ *of the auxiliary stochastic control problem (2.15) by* $\mathcal{A}$*; see Corollary A.1 of Appendix A.3.2. This is achieved by the elaborate construction of the auxiliary state process* $\tilde{X}$*. In particular,* $\tilde{X}$ *freezes at the truncation boundary once* X *exists the truncation region. It is worth stressing that the conclusion of Corollary A.1 does not hold in general.*

The error bound in the above inequality (2.20) can be understood as follows. The term $\big(\xi(R) + \zeta\big)\mathcal{E}_T(x_0, R)$ corresponds to an upper bound for the discrepancy between the reward functions of the two stochastic control models (2.2) and (2.15) at each time step. Since such a difference primarily stems from replacing the original state process X by $\tilde{X}$, it is not surprising that the term $\mathcal{E}_T(x_0, R)$ appears in the error estimate. Furthermore, the two terms $\xi(R)$ and ζ correspond to certain upper bounds of the magnitudes of the reward terms $f_t^2\left(\tilde{X}_t, a_t\right)$ and $f_t^2\left(X_t, a_t\right)$, respectively, and therefore a square root arises in the inequality (2.20). Finally, the discrepancy between the two value functions is amplified as the time progresses. This is reflected by the existence of a factor T in the above error estimate. The above interpretation is more perceivable from the proof of Theorem 2.2; see Appendix A.3.3. We further give a remark on how to determine an appropriate value of the truncation parameter R.

Remark 2.4 (Choice of R)**.** *The preceding Theorem 2.1 gives guide on how to choose the truncation parameter* R*. Note that* $\mathcal{E}_T(x_0, R)$ *is decreasing in* R *and whereas* $\xi(R)$ *is monotone with*

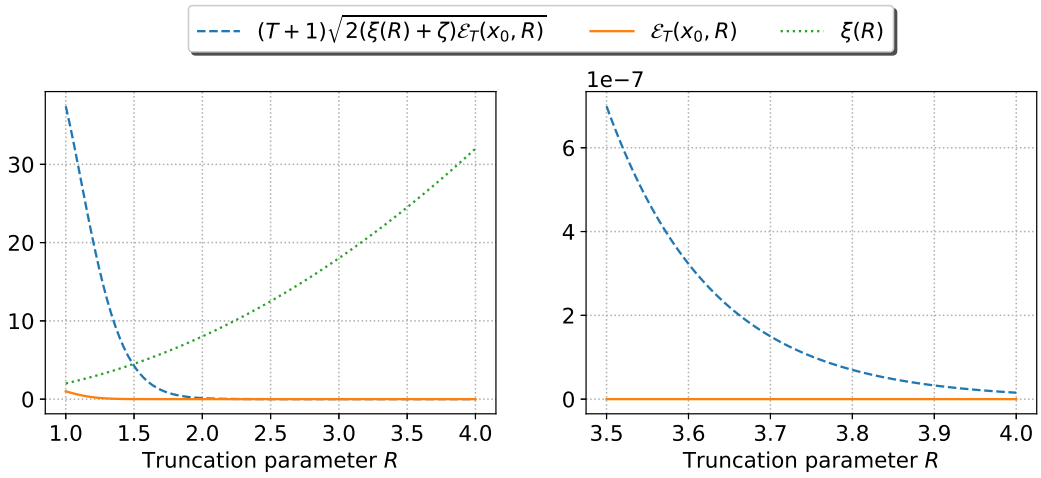


Figure 2.6: Plots of $\mathcal{E}_T(x_0, R)$, $\xi(R)$, and the error bound in (2.20) as functions of truncation parameter R. The left and right panels depict these functions over the intervals $[1, 4]$ and $[3.5, 4]$, respectively.

respect to R. This means that in order to ensure the R.H.S. of the inequality (2.20) is marginal $\mathcal{E}_T(x_0, R)$ should decay much faster than the growth speed of $\xi(R)$. Appendix A.2 derives an explicit expression for the error estimate in (2.20) in the concrete example of Section 2.5. Figure 2.6 depicts $\mathcal{E}_T(x_0, R)$ and $\xi(R)$ as functions of R under the parameter setting of Section 2.6 (see Table 2.1). From the left panel of Figure 2.6, one may clearly see that $\mathcal{E}_T(x_0, R)$ (solid line) dominates $\xi(R)$ (dotted line). The dashed line in Figure 2.6 shows that choosing $R = 4$ incurs a truncation error less than 10^{-7}.

In light of the result delivered in the preceding theorem, one may henceforth turn our attention to the optimal value function of the auxiliary problem (2.15), that is, $\tilde{V}_0(x_0)$. In the subsequent section, we will develop an LSMC algorithm to approach $\tilde{V}_0(x_0)$.

2.4 A Backward Simulation Monte Carlo Algorithm

2.4.1 Simulation of post-action value

This subsection proposes an LSMC algorithm which simulates the state process without referring to the optimal action. Recall from Step 3.1 of the FSBU algorithm in Section 2.2.2 that the ultimate goal of simulating the state process is generating a random sample of the post-action value

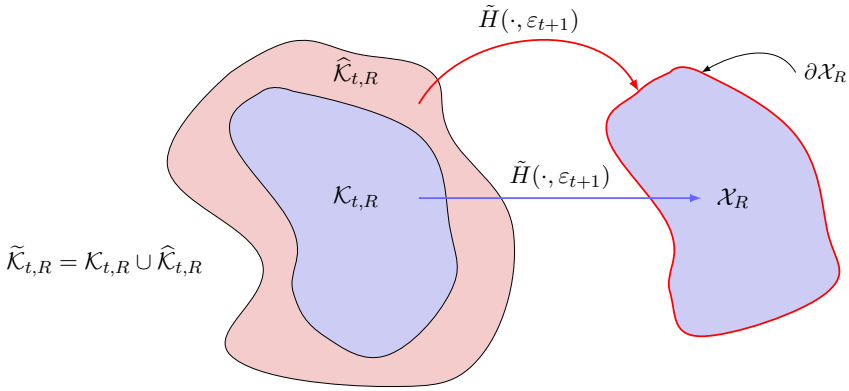


Figure 2.7: A graphical illustration for the relationships between $\mathcal{K}_{t,R}, \widehat{\mathcal{K}}_{t,R}$, and $\mathcal{X}_R$.

of the state process which acts as a crucial input for the regression step. This naturally inspires us to directly simulate the post-action value X_{t^+} from an artificial probability distribution. The term "artificial" stresses the fact that such a distribution might not coincide with the distribution of X_{t^+} under the optimal action process.

Since the value function $\tilde{V}_t(\cdot)$ is explicitly given by Eq. (2.16) over $\partial\mathcal{X}_R$, the primary goal of our proposed LSMC algorithm is to get a numerical estimate for the value function over the open set $\mathring{\mathcal{X}}_R$. In view of this, one may circumscribe the support of the artificial probability distribution that the post-action values are simulated from. First note that the range of post-action value of the auxiliary state process denoted by $\widetilde{\mathcal{K}}_{t,R}$ is given by

$$\widetilde{\mathcal{K}}_{t,R} := \bigcup_{x \in \mathring{\mathcal{X}}_R} \left(\bigcup_{a \in A_t(x)} \{K(x,a)\} \right), \quad \text{for} \quad t \in \mathcal{T}_0.$$

Consider the following subset:

$$\widehat{\mathcal{K}}_{t,R} := \left\{ k \in \widetilde{\mathcal{K}}_{t,R} \;\middle|\; \tilde{H}(k, e_1) = \tilde{H}(k, e_2) \in \partial\mathcal{X}_R, \ \forall e_1 \text{ and } e_2 \in \mathbb{D} \right\}, \tag{2.21}$$

for $t \in \mathcal{T}_0$, where one should recall that $\mathbb{D}$ is the set of all values the random innovation ε_{t+1} might take and $\tilde{H}(\cdot,\cdot)$ is the transition equation relating the post-action value at time step t to the state variable at time step $t+1$ which is given in Eq. (2.14).

The preceding equation states that X_{t+1} will stop at a certain point in the boundary set $\partial\mathcal{X}_R$ if $\tilde{X}_{t^+} := K\left(\tilde{X}_t, a_t\right)$ lies in the set $\widehat{\mathcal{K}}_{t,R}$; see Figure 2.7 for a graphical illustration. To make the matter more concrete, let us consider the example of pricing variable annuities (see, e.g. [47] and [72]) where $\tilde{X}_{t^+}$ corresponds the post-withdrawal value of the investment account. If the

investment account is depleted after the policyholder's withdrawal (i.e., $\tilde{X}_{t^+} = 0$), it remains exhausted forever (i.e., $X_n^R = 0$ for $n = t+1, \ldots, T$). In such an example, $\widehat{\mathcal{K}}_{t,R}$ is a singleton $\{0\}$. In view of the above discussion and Eq. (2.16), for any $k \in \widehat{\mathcal{K}}_{t,R}$,

$$\tilde{C}_t(k) = \mathbb{E}\left[\tilde{V}_{t+1}\big(\tilde{H}(k, \varepsilon_{t+1})\big)\right] = \tilde{V}_{t+1}\big(\tilde{H}(k, e)\big) \tag{2.22}$$

which has a value independent of e and is given by Eq. (2.16). Therefore, at time step t, it suffices to get a regression estimate for the continuation function $\tilde{C}_t(\cdot)$ on the set $\mathcal{K}_{t,R} := \widehat{\mathcal{K}}_{t,R} \backslash \widehat{\mathcal{K}}_{t,R}$.

2.4.2 The algorithm

Now the Backward Simulation and Backward Updating (BSBU) algorithm is presented as follows.

1. **Initiation:** Set $\tilde{V}_T^{\mathrm{E}}(x) = f_T(x)$ for $x \in \mathrm{cl}\,(\mathcal{X}_R)$. For $t = T-1, T-2, \ldots, 0$, do the two steps below.

2. **Backward Simulation:**

   2.1 **Simulation of post-action value** Generate a sample of the post-action values denoted by

   $$\mathbf{X}_{t^+}^M := \left\{X_{t^+}^{(m)},\ m = 1, 2, \ldots, M\right\}$$

   from a probability distribution $\mathsf{Q}_{t,R}$ with support $\mathcal{K}_{t,R}$.

   2.2 **Simulation of the state process** Construct the sample of the state process at time step $n+1$ according to

   $$\mathbf{X}_{t+1}^M := \left\{X_{t+1}^{(m)} = \tilde{H}\left(X_{t^+}^{(m)}, \varepsilon_{t+1}^{(m)}\right),\ m = 1, 2, \ldots, M\right\}. \tag{2.23}$$

   with $\left\{\varepsilon_{t+1}^{(m)}, m = 1, 2, \ldots, M\right\}$ being a sample of the random innovations.

3. **Backward Updating:**

   3.1 **Data preparation** Given a numerical estimate of value function at time step $t+1$, denoted by $\tilde{V}_{t+1}^{\mathrm{E}}(\cdot)$, construct the sample

   $$\mathbf{Y}_{t+1}^M := \left\{\tilde{V}_{t+1}^{\mathrm{E}}\left(X_{t+1}^{(m)}\right), m = 1, 2, \ldots, M\right\}. \tag{2.24}$$

   3.2 **Regression** Take $\mathbf{Y}_{t+1}^M$ and $\mathbf{X}_{t^+}^M$ as the samples of response variable and regressor, respectively, and employ a certain non-parametric regression to obtain a regression

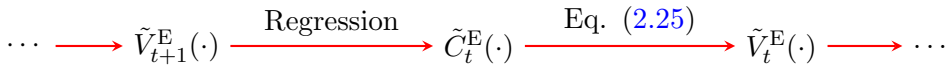


Figure 2.8: A diagram for backward information propagation in the BSBU algorithm.

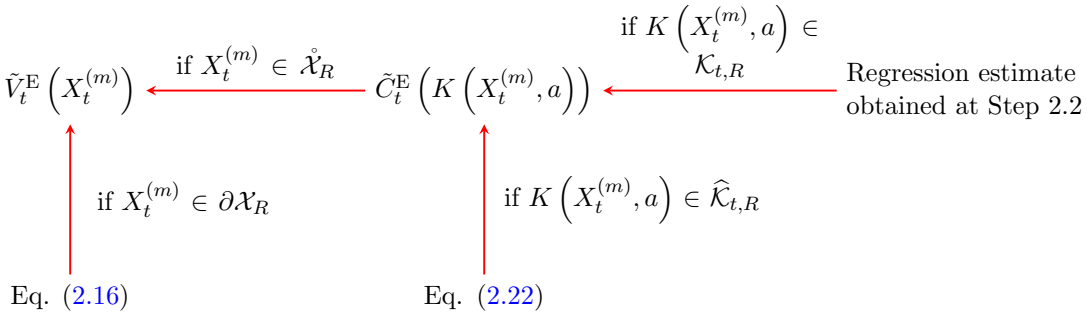


Figure 2.9: A diagram for the information propagation in evaluating $\tilde{V}_t^{\mathrm{E}}\left(X_t^{(m)}\right)$.

estimate $\tilde{C}_t^{\mathrm{E}}(\cdot)$ over the set $\mathcal{K}_{t,R}$. For $k \in \widehat{\mathcal{K}}_{t,R}$, set $\tilde{C}_t^{\mathrm{E}}(k) = \tilde{C}_t(k)$ with $\tilde{C}_t(\cdot)$ given by Eq. (2.22).

3.3 **Optimization** An estimate for the value function at time step t is given by:

$$\tilde{V}_t^{\mathrm{E}}(x) = \sup_{a \in A_t(x)} \left[f_t(x,a) + \varphi \tilde{C}_t^{\mathrm{E}}\big(K(x,a)\big) \right], \quad \text{for} \quad x \in \mathring{\mathcal{X}}_R. \tag{2.25}$$

For $x \in \partial\mathcal{X}_R$, set $\tilde{V}_t^{\mathrm{E}}(x) = \tilde{V}_t(x)$ with $\tilde{V}_t(\cdot)$ given by Eq. (2.16).

Step 3.2 prescribes $\tilde{C}_t(k)$ for the value of $\tilde{C}_t^{\mathrm{E}}(k)$ when $k \in \widehat{\mathcal{K}}_{t,R}$ because $K\left(X_t^{(m)}, a\right)$ might fall in the set $\widehat{\mathcal{K}}_{t,R}$. Similarly, in Step 3.3, Eq. (2.16) is invoked to evaluate $\tilde{V}_t^{\mathrm{E}}(x)$ for $x \in \partial\mathcal{X}_R$ as $X_t^{(m)}$ generated by Eq. (2.23) may lie on $\partial\mathcal{X}_R$, the boundary set of the truncated domain. The backward information propagation in the above BSBU algorithm is illustrated in Figure 2.8.

Comparing the above BSBU algorithm and the FSBU counterpart in Section 2.2.2, we have the following observations.

1. Firstly, the primary difference of the two algorithms lies in how to generate the post-action values of state process, i.e., $\mathbf{X}_{t^+}^M$. The FSBU algorithm is a forward simulation scheme while the BSBU algorithm directly generates post-action value from a certain prior distribution. Indeed, the FSBU algorithm can be viewed as a special BSBU algorithm if $\mathsf{Q}_{t,R}$ is chosen as the probability distribution of the post-action value from a control randomization procedure.

In general, both methods do not yield the distribution of X_{t^+} driven by the optimal action, and thus, there is no loss to directly generate $\mathbf{X}_{t^+}^M$ from a prior distribution $\mathsf{Q}_{t,R}$.

2. Secondly, the BSBU method has the advantage of reducing memory and time costs. On one hand, one does not need to store the sample of whole trajectories at each time step in the BSBU algorithm. On the other hand, the total time cost of simulating the state process is of $O(T)$ in the BSBU algorithm, while it is of $O\left(T^2\right)$ in the FSBU counterpart; see the item **(II)** "Cost of forward simulation" in Section 2.2.2.

3. Thirdly, the BSBU algorithm circumvents extrapolating the numerical estimates of the continuation function and the value function. It is notable that $\tilde{C}_t^{\mathrm{E}}(\cdot)$ and $\tilde{V}_t^{\mathrm{E}}(\cdot)$ are obtained over the sets $\widetilde{\mathcal{K}}_{t,R}$ and $\mathrm{cl}\left(\mathcal{X}_R\right)$, respectively, at time step t; see Steps 2.2-2.3 of the above BSBU algorithm. At the time step $t-1$, the BSBU algorithm does not require the knowledge of the value function (resp. the continuation function) outside $\mathrm{cl}\left(\mathcal{X}_R\right)$ (resp. $\widetilde{\mathcal{K}}_{t,R}$) in obtaining $\mathbf{Y}_t^M$; see Figure 2.9 for a graphical illustration. This nice property inherits from the construction of the auxiliary state process $\tilde{X}$ whose values are confined to a bounded set. In the FSBU algorithm, however, the state process is not restrained and $K\left(X_t^{(m)}, a\right)$ might lie outside the regression domain for $\tilde{C}_t^{\mathrm{E}}(\cdot)$. In such a situation, extrapolating the numerical solution causes extra error which is hard to quantify.

2.4.3 Sieve Estimation Method

In this subsection, we propose a linear sieve estimation method to estimate the continuation function in our BSBU algorithm.

Selection criteria for regression method

In Section 2.2.2 we have discussed potential issues associated with a regression method in estimating the continuation function of a stochastic control problem; see the item **(III)** "Choice of regression method". Based on the discussion, the following criteria for the choice of regression method in estimating continuation function are proposed.

(C1) Small memory cost The regression problem embedded in an LSMC algorithm usually exhibits extraordinarily large sample size. Thus, an appropriate regression method should have small memory requirement. This criterion excludes the kernel method ([64] and [85]), local-polynomial regression method ([38]), and isotonic regression method ([69]) which require storing all sample points in the memory in order to compute the regression function at any point in the domain.

(C2) Computationally cheap In almost all nonparametric regression methods, a certain parameter (referred to as *tuning* parameter in statistics literature) is used to avoid undesirable overfitting or underfitting of the regression model. Determining the optimal value of such a tuning parameter is usually computationally intensive. Therefore, an ideal regression method should be insensitive to the tuning parameter.

In view of the above two criteria, there are a limited number of suitable choices despite the voluminous nonparametric regression methods in the literature. In the following, we discuss a class of regression methods which is often referred to as the *sieve estimation method* in the literature.

Shape-preserving sieve estimation

Below, we give a brief introduction to the sieve estimation method; refer to [27] for a comprehensive review. Suppose one has a sample of independent and identically distributed (i.i.d.) random pairs $\left\{\left(U^{(m)}, Z^{(m)}\right)\right\}_{m=1}^{M}$ where $Z^{(m)}$ is a multivariate random vector with compact support $\mathcal{Z}$ and $U^{(m)}$ is a univariate random variable. Define the function $g(\cdot) : \mathcal{Z} \longrightarrow \mathbb{R}$ as

$$g\left(z\right) = \mathbb{E}\left[U^{(m)} \middle| Z^{(m)} = z\right] \tag{2.26}$$

which is independent of m. In the context of our BSBU algorithm, $U^{(m)}$ and $Z^{(m)}$ correspond to $\tilde{V}_{t+1}\left(X_{t+1}^{(m)}\right)$ and $X_{t^+}^{(m)}$, respectively, and the parallel function $g(\cdot)$ is the continuation function $\tilde{C}_t(\cdot)$.

The sieve estimation method strives to estimate the functional form of $g(\cdot)$ by solving the following optimization problem:

$$\hat{g}(\cdot) := \arg \min_{h(\cdot) \in \mathcal{H}_J} \frac{1}{M} \sum_{m=1}^{M} \left[U^{(m)} - h\left(Z^{(m)}\right)\right]^2, \tag{2.27}$$

where $\mathcal{H}_J$ is a finite-dimensional functional space depending on a certain parameter J and is called as *sieve space*. Intuitively, the ampler the sieve space is, the smaller the "gap" between the $\mathcal{H}_J$ and the function $g(\cdot)$ would be. The price to pay is that larger estimation error is incurred for a richer sieve space due to limited sample size M. Therefore, one has to balance such a trade-off by controlling the complexity of the sieve space and this is achieved by tuning the parameter J. To make the matter more concrete, we consider two examples of the sieve space in the sequel.

Example 2.1 (Linear Sieve Space)**.** *Let* $\{\phi_j(\cdot) : \mathcal{Z} \longrightarrow \mathbb{R}\}_{j \in \mathbb{N}}$ *be a sequence of basis functions.*

Consider the sieve space

$$\mathcal{H}_J = \left\{ h(\cdot): \quad h(z) = \sum_{j=0}^{J} \beta_j \phi_j(z), \quad \beta_j \in \mathbb{R} \right\}. \tag{2.28}$$

The above set $\mathcal{H}_J$ is essentially a linear span of finitely many basis functions and is referred to as *linear sieve space* in the statistics literature.

In the present context, the regression function $g(\cdot)$ corresponds to the continuation function and it exhibits some shape properties such as monotonicity in many applications; see [34] for pricing American option and [6] for valuing equity-linked insurance product, among others. In view of this, it is natural to expect the element in the sieve space satisfies such shape constraints, which makes the numerical result more economically sensible. This can be achieved by considering a special linear sieve space in the following example.

Example 2.2 (Shape-Preserving Sieve Space)**.** *Let* $\{\phi_j(\cdot): \mathcal{Z} \longrightarrow \mathbb{R}\}_{j\in\mathbb{N}}$ *be a sequence of basis functions. Denote* $\boldsymbol{\beta}_J = (\beta_0, \ldots, \beta_J)^\intercal$ *with* $\beta_j \in \mathbb{R},\ j = 0, 1, \ldots, J$*. Consider the sieve space*

$$\mathcal{H}_J = \left\{ h(\cdot): \quad h(z) = \sum_{j=0}^{J} \beta_j \phi_j(z), \quad \mathbf{A}_J \boldsymbol{\beta}_J \geq \mathbf{0}_{c(J)} \right\}, \tag{2.29}$$

where $c(\cdot): \mathbb{N} \longrightarrow \mathbb{N}$ *is some integer-valued function,* $\mathbf{A}_J$ *is a* $c(J)$*-by-*$(J+1)$ *matrix, and* $\mathbf{0}_{c(J)}$ *is a* $c(J)$*-by-1 null vector.*

[83, 84] show that each element in the sieve space in Eq. (2.29) is a convex, concave, or monotone function (with respect to each coordinate) with a special choice of the matrix $\mathbf{A}_J$ given that $\phi_j(\cdot),\ j = 0, 1, \ldots, J$, are Bernstein polynomials. See Appendix A.1.1 for the expressions of the Bernstein polynomials and matrix $\mathbf{A}_J$.

For a linear sieve space $\mathcal{H}_J$ defined either in Eq. (2.28) or Eq. (2.29), the solution of the preceding optimization problem (2.27) is given by the following form:

$$\hat{g}(z) = \hat{\boldsymbol{\beta}}^\intercal \boldsymbol{\phi}(z), \quad \text{for} \quad z \in \mathcal{Z}, \tag{2.30}$$

where $\boldsymbol{\phi}(z) := (\phi_1(z), \ldots, \phi_J(z))^\intercal$ and $\hat{\boldsymbol{\beta}}$ is the optimizer of the following optimization problem:

$$\min_{\boldsymbol{\beta}} \frac{1}{M} \sum_{m=1}^{M} \left[U^{(m)} - \boldsymbol{\beta}^\intercal \boldsymbol{\phi}\left(Z^{(m)}\right) \right]^2, \quad \text{subject to} \quad \boldsymbol{\beta}^\intercal \boldsymbol{\phi}(\cdot) \in \mathcal{H}_J. \tag{2.31}$$

The dependency of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\phi}(\cdot)$ on J is suppressed for brevity. In general, one has to solve a constrained quadratic programming problem to obtain $\hat{\boldsymbol{\beta}}$.

Discussions

One clear merit of the above linear sieve estimation method is that one only needs to store the vector $\hat{\boldsymbol{\beta}}$ for future evaluation of the regression function $\hat{g}(\cdot)$ at any point in the domain because basis functions $\boldsymbol{\phi}(\cdot)$ are explicitly known at the first hand. This makes the linear sieve estimation method tailored to our present problem in terms of the criterion **(C1)**.

For the criterion **(C2)**, it is documented in statistics literature that when the true regression function $g(\cdot)$ satisfies certain shape constraints, the shape-preserving estimate $\hat{g}(\cdot)$ obtained by (2.31) with $\mathcal{H}_J$ given by Eq. (2.29) is *insensitive* to the tuning parameter J; see, e.g., [61] and [83, 84]. When there is no prior shape information of $g(\cdot)$, one has to use the sieve space (2.28) and the regression estimate might be sensitive to the choice of J. Under such a situation, J can be determined in a data-driven manner. Appendix A.1.2 presents some common methods of selecting J discussed in the literature.

Finally, the convergence of the sieve estimate $\hat{g}(\cdot)$ to the conditional mean function $g(\cdot)$ is ensured under some technical conditions. These conditions are summarized in Assumption A.1 which is relegated to Appendix A.1.3 for the clarity of presentation.

2.4.4 Convergence Analysis

Now it is ready to conduct convergence analysis of the BSBU algorithm proposed in Section 2.4. For the regression method employed in the algorithm, we restrict our attention to the linear sieve estimator given by Eqs. (2.30) and (2.31) in the previous subsection.

A complete convergence analysis of the BSBU algorithm should take account of three types of errors:

(E0) Truncation Error The truncation error is caused by taking $\tilde{V}_0(x_0)$ as a proxy for $V_0(x_0)$.

(E1) Sieve Estimation Error At each step of the BSBU algorithm, the sieve estimation method is employed to get an estimate for the continuation function. The associated sieve estimation error stems from two resources: (a) the bias caused by using a finite-dimensional sieve space $\mathcal{H}_J$ to approximate continuation function; and (b) the statistical error in estimating coefficients of basis functions under a limited sample size of M.

(E2) Accumulation Error Recall that the nonparametric regression is used to approximate the function $\tilde{C}_t(\cdot) = \mathbb{E}\left[\tilde{V}_{t+1}(X_{t+1})\middle|X_{t^+} = \cdot\right]$. Thus, in principle, one should generate a random sample

$$\left\{\left(\tilde{V}_{t+1}\left(X_{t+1}^{(m)}\right), X_{t^+}^{(m)}\right)\right\}_{m=1}^{M}$$

based on which the sieve estimation method can be employed to get a regression estimate. However, $\tilde{V}_{t+1}(\cdot)$ is unknown and is replaced by its numerical estimate $\tilde{V}^{\mathrm{E}}_{t+1}(\cdot)$ in Step 2.2 of the BSBU algorithm. Such a compromise triggers a new type of error in addition to **(E0)** and **(E1)**.

(E0) has been investigated in Theorem 2.1. The discrepancy between $\tilde{V}_0(x_0)$ and $\tilde{V}^{\mathrm{E}}_0(X_0)$ is contributed by **(E1)** and **(E2)**. Distinguishing these two types of error plays a crucial role in our convergence analysis and this is inspired by [16]. Our main convergence result is summarized in the following theorem.

Theorem 2.2 (BSBU Algorithm Error)**.** *Suppose that*

(i) *Assumptions 2.1–2.3 and Assumption A.2 in Appendix A.3.3 hold;*

(ii) *Assumption A.1 in Appendix A.1 holds for* $U^{(m)} = V_{t+1}\left(X^{(m)}_{t+1}\right)$ *and* $Z^{(m)} = X^{(m)}_{t^+}$ *uniformly in* $t \in \mathcal{T}_0$*, where* $X^{(m)}_{t^+}$ *and* $X^{(m)}_{t+1}$ *are given in Steps 2.1 and 2.2 of the BSBU algorithm.*

Then, there exists a constant ψ *such that*

$$\left|\tilde{V}_0(x_0) - \tilde{V}^{\mathrm{E}}_0(x_0)\right| = O_{\mathbb{P}}\left(\sqrt{\psi^{T-1}\left(J/M + \rho_J^2\right)}\right), \quad \text{as} \quad M \longrightarrow \infty, \tag{2.32}$$

where ρ_J *is some sequence such that* $\rho_J \to 0$ *as* $M \to \infty$ *(see Assumption A.1) and the "Big O p" notation* $O_{\mathbb{P}}(\cdot)$ *is defined in Definition A.1 of Appendix A.3.3.*

The above theorem basically states that the numerical solution $\tilde{V}^{\mathrm{E}}_0(x_0)$ converges to $\tilde{V}_0(x_0)$ in probability as both the number of basis functions J and number of simulated paths M approach infinity at the rate specified by Condition (v) in Assumption A.1. Since Theorem 2.1 shows that the discrepancy between $\tilde{V}_0(x_0)$ and $V_0(x_0)$ is marginal for a sizable R, the numerical estimate $\tilde{V}^{\mathrm{E}}_0(x_0)$ is a legitimate approximation for $V_0(x_0)$ when R, J, and M are considerable. The R.H.S. of Eq. (2.32) reveals that the overall BSBU algorithm error arises from the two resources discussed in the previous item **(E1)**, which are indicated by the terms ρ_J and J/M, respectively. Furthermore, Eq. (2.32) also shows that such a regression error is magnified by a factor ψ at each time step, which reflects the error accumulation from time step $T-1$ down to time step 0. This is in line with the earlier discussion in the item **(E2)**.

2.5 Application: Pricing Equity-linked Insurance Products

This section applies the BSBU algorithm to the pricing of equity-linked insurance products. This pricing problem is an appropriate example to show the limitations of the FSBU algorithm

commented in Section 2.2.2. For the convenience of illustration, the contract studied here is a simplified version of variable annuities (VAs); for more complex products, see [6], [47], [48], and [72], among others.

2.5.1 Contract Description

VAs are prevailing equity-linked insurance products in North America. At inception of the contract, the policyholder (PH) pays a lump sum w_0 to the insurer which is invested into the equity market. The PH is entitled to withdraw any portion of the investment before maturity. She also enjoys certain guaranteed payments provided by the insurer regardless of the performance of the investment account. This exposes the insurer to downside risk of the equity market that is not diversifiable. As compensation, the insurer deducts insurance fees from the PH's investment account and trades available securities to hedge his risk exposure. Thus, no-arbitrage pricing has been the dominating paradigm for pricing VAs in the literature. The primary challenge of this pricing problem stems from the uncertainty of the PH's withdrawal behavior. This is conventionally resolved by studying the optimal withdrawal strategy of the PH in analogy to pricing American-style options, which naturally leads to a stochastic control problem; see [33], [28], [47], and many others.

2.5.2 Model Setup

We exemplify the model setup of Section 2.2 in the present pricing problem. The lattice $\mathcal{T}$ now labels the collection of all possible withdrawal dates. The first decision variable τ_t represents the PH's decision to initialize the withdrawal or not by taking values 1 and 0, respectively. As one may see later, the payoff functions depend on the timing of the first withdrawal of the PH. Therefore, a state variable $\{I_t\}_{t\in\mathcal{T}}$ is introduced to record the first-withdrawal-time, and its evolution mechanism is prescribed as follows: $I_0 = 0$, and

$$I_{t+1} = S_t^I(I_t, \tau_t) := \begin{cases} t, & \text{if } I_t = 0 \text{ and } \tau_t = 1, \\ I_t, & \text{otherwise,} \end{cases} \tag{2.33}$$

for $t \in \mathcal{T}_0$. The feasible set of τ_t is a singleton $\{1\}$ if the withdrawal has been initialized, i.e., $I_t > 0$; otherwise, it is $\{0, 1\}$.

Denote $(a)^+ := \max\{a, 0\}$ and $a \vee b := \max\{a, b\}$. The second state variable corresponds to the investment account and evolves according to

$$\begin{cases} W_0 &= w_0, \\ W_{t+1} &= \underbrace{\big(W_t - \gamma_t\big)^+}_{\text{post-withdrawal value}} \cdot \varepsilon_{t+1}, \quad \gamma_t \in \big[0, W_t \vee g_t(I_t) w_0\big], \quad t \in \mathcal{T}_0, \end{cases} \tag{2.34}$$

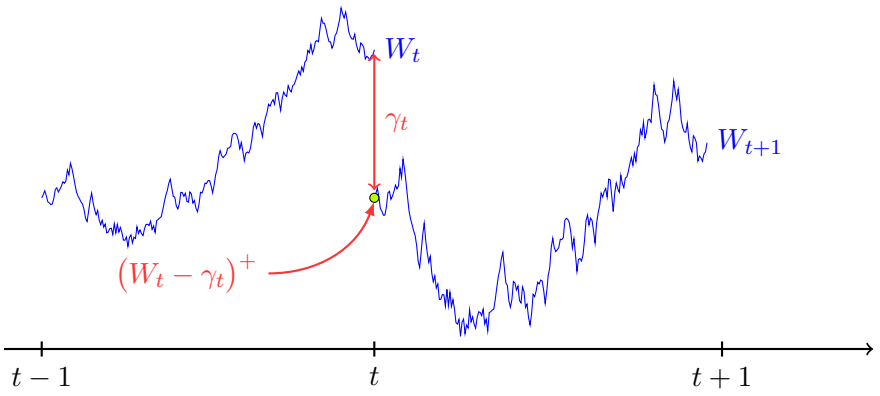


Figure 2.10: Jump mechanism of the investment account across a withdrawal date.

where γ_t is the withdrawal amount of the PH at time t, $\log \varepsilon_{t+1}$ corresponds to the log-return of the underlying asset over each time interval, and $g_t(I_t)$ is a certain percentage number depending on I_t. The above equation implies that the PH can withdraw up to the amount of $g_t(I_t)w_0$ even if the investment account is depleted, i.e., $W_t = 0$. The jump mechanism of the investment account across each withdrawal date is illustrated in Figure 2.10.

We assume that ε_{t+1} follows from a log-normal distribution with $\mathbb{E}[\log \varepsilon_{t+1}] = \big(r - q - \sigma^2/2\big)\delta := \mu\delta$ and $\mathbb{V}\mathrm{ar}[\log \varepsilon_{t+1}] = \sigma\delta$. In other words, the underlying asset evolves according to a geometric Brownian motion with drift and volatility rates $r - q$ and σ, respectively. r and q correspond to the risk-free rate and the insurance fee rate, respectively. Throughout this section, the expectation $\mathbb{E}[\cdot]$ is taken under a certain martingale pricing measure.

Now, the state process and the DM's action are $X = \{X_t = (W_t, I_t)\}_{t\in\mathcal{T}}$ and $\mathsf{a} = \{a_t = (\gamma_t, \tau_t)\}_{t\in\mathcal{T}}$, respectively. In accordance with Eqs. (2.33) and (2.34), the accompanying transition equation is $X_{t+1} = H\big(K(X_t, a_t), \varepsilon_{t+1}\big)$, where

$$K(X_t, a_t) = \Big(\big(W_t - \gamma_t\big)^+,\ S_t^I(I_t, \tau_t)\Big), \quad H\big(k, \varepsilon_{t+1}\big) = \big(k_1\varepsilon_{t+1}, k_2\big) \tag{2.35}$$

with $k = (k_1, k_2) \in [0, \infty) \times \mathcal{T}_0$. The dependency of $K(\cdot, \cdot)$ on t is suppressed for notational brevity.

Next, we discuss the feasible set of the PH's action. In principle, the withdrawal amount γ_t takes values in a continum $\big[0, W_t \vee g_t(I_t)w_0\big]$; see Eq. (2.34). However, it can be shown that the optimal withdrawal amount is limited to three choices: 1) $\gamma_t = 0$, 2) $\gamma_t = g_t(I_t)w_0$, and 3) $\gamma_t = W_t$ under certain contract specifications; see [6], [47], [48], and [72]. Via a similar argument adopted by the above references, one may show that this conclusion still holds for the contract considered

here. Therefore, one may restrict the feasible set of action a_t into the following discrete set:

$$A_t(X_t) = \begin{cases} \{(0,0), (g_t(I_t)w_0, 1), (W_t, 1)\}, & \text{if } I_t = 0, \quad \text{(withdrawal has not been initialized)} \\ \{(0,1), (g_t(I_t)w_0, 1), (W_t, 1)\}, & \text{if } I_t > 0, \quad \text{(withdrawal has been initialized)} \end{cases} \tag{2.36}$$

for $t = 1, 2, \ldots, T-1$. As a convention, the PH is not allowed to withdraw at inception, and thus $A_0(X_0) = \{(0,0)\}$.

We proceed by specifying the reward functions which correspond to the policy payoffs in the present context. Before maturity, the cash inflow of the PH is her withdrawal amount subject to some penalty:

$$f_t(X_t, a_t) = \gamma_t - \kappa\big(\gamma_t - g_t(I_t)w_0\big)^+, \quad \gamma_t \in \big[0, W_t \vee g_t(I_t)w_0\big], \quad \text{for} \quad t \in \mathcal{T}_0,$$

with $\kappa \in [0,1]$ being the penalty rate. In other words, the withdrawal amount in excess of the guaranteed amount is subject to a proportional penalty. At maturity, the policy payoff is the remaining value of the investment account, i.e., $f_T(X_T) = W_T$.

Finally, we give the interpretation of value function in the present context. $V_t(x) = V_t(W, I)$ with $I > 0$ (resp., $I = 0$) corresponds to the no-arbitrage price of the contract at withdrawal date t given that the investment account has a value of W and the first withdrawal is triggered at I-th withdrawal date (resp., no withdrawal has been taken).

2.5.3 A BSBU Algorithm for the Pricing Problem

The state process X generally takes value in the unbounded set $\mathcal{X} = [0, \infty) \times \mathcal{T}_0$. We consider a truncated domain: $\mathcal{X}_R = [0, R) \times \mathcal{T}_0$ with $R > 0$. Consequently, one may define the auxiliary state process $\tilde{X}$ as in Eq. (2.13). The range of the post-action value is given by $\widetilde{\mathcal{K}}_{t,R} = \widehat{\mathcal{K}}_{t,R} \cup \mathcal{K}_{t,R}$, where $\widehat{\mathcal{K}}_{t,R} = \{0, R\} \times \{0, 1, \ldots, t\}$ and $\mathcal{K}_{t,R} = (0, R) \times \{0, 1, \ldots, t\}$, respectively. This is in line with Eq. (2.21). Appendix A.2.1 verifies the preceding Assumptions 2.2 and 2.3 in the present context and derive an explicit expression for the truncation error estimate delivered in the preceding Theorem 2.1.

It is worth noting that a discrete state variable I_t appears in the present context and the continuation function, in general, is not continuous with respect to the post-action value accompanying this state variable, i.e., k_2; see Eq. (2.35). Consequently, Condition (iii) of Assumption A.1 might not hold here; see Appendix A.1.3. However, for each given value of k_2, the continuation function is still continuous with respect to k_1, the post-action value associated with the investment account value. Therefore, one may repeat Step 3.2 of the BSBU algorithm for every distinct value of k_2. It is easy to see the convergence of the resulting BSBU algorithm is not influenced by this modification.

Finally, it remains to specify the sampling distribution of the post-action value in order to pave the way to implementing the BSBU algorithm. In the subsequent section, we will address this issue in details and, in particular, we will compare the control randomization method with our artificial simulation method.

2.6 Numerical Experiments

This section devotes to conducting numerical experiments to show the merits of the BSBU algorithm in the context of pricing the variable annuity product addressed in the previous section.

2.6.1 Parameter Setting

We first present the parameter setting for our numerical experiments. We consider $T = 12$ time steps and the time interval between two consecutive withdrawal date is assumed to be $\delta = 1/12$. This corresponds to a contract with one-year maturity and monthly withdrawal frequency. The discounting rate is given by $\varphi = e^{-r\delta}$ with risk-free rate $r = 0.03$. The PH's initial investment is assumed to be one unit, i.e., $w_0 = 1$. The guaranteed payment percentage $g_t(I_t)$ is prescribed as $g_t(I_t) = G(t)\mathbb{1}_{\{I_t=0\}} + G(I_t)\mathbb{1}_{\{I_t>0\}}$ with $G(I) = 0.03\mathbb{1}_{\{1\le I\le 5\}} + 0.07\mathbb{1}_{\{6\le I\le 11\}}$. In other words, the PH enjoys a larger amount of guaranteed payment if she postpones the initiation of the withdrawal. In all subsequent numerical experiments, the truncation parameter R is fixed as 4, which causes a truncation error less than 10^{-7}; see the earlier Remark 2.4 and Figure 2.6. All the parameters used in our numerical experiments are summarized in Table 2.1.

2.6.2 Forward Simulation v.s. Artificial Simulation

Next, we would like to show the limitations of the forward simulation based on control randomization. Recall that $A_0(x)$ is a singleton $\{(0,0)\}$, so we may initialize $a_0^{(m)} = (0,0)$. Below, we consider some control randomization methods for generating $a_t^{(m)}$ for $t > 0$.

(CR0) The PH's action $a_t^{(m)}$ is simulated from a degenerated distribution with one single point mass at $(G(1)w_0, 1)$.

(CR1) Given $X_t^{(m)} = \left(W_t^{(m)}, I_t^{(m)}\right)$, the DM's action $a_t^{(m)}$ is simulated from a discrete uniform distribution with support set $A_t\left(X_t^{(m)}\right)$; see Eq. (2.36).

(CR2) Given $X_t^{(m)} = \left(W_t^{(m)}, I_t^{(m)}\right)$, the DM's action $a_t^{(m)}$ is simulated from a discrete uniform distribution with support set $A_t\left(X_t^{(m)}\right) \setminus \left\{\left(W_t^{(m)}, 1\right)\right\}$.

Table 2.1: Parameters used for numerical experiments.

Parameter	Value
Volatility rate σ	0.15
Risk-free rate r	0.03
Insurance fee rate q	0.01
Number of time steps T	12
Length of time interval δ	1/12
Discounting factor $\varphi = e^{-r\delta}$	0.9975
Initial purchase payment w_0	1
Withdrawal penalty κ	0.8
Guaranteed withdrawal percentage $G(I)$	$1 \leq I \leq 5 : 3\%,\ 6 \leq I \leq 11 : 7\%$
Truncation parameter R	4

Given the above rules of generating the PH's action, one may simulate the state process in a forward manner in accordance with Steps 2.1 and 2.2 of the FSBU algorithm; see Section 2.2.2.

(CR0) is first proposed by [47] in the context of pricing Guaranteed Lifelong Withdrawal Benefit, a particular type of variable annuity policy. It initializes the withdrawal at $t = 1$ and the resulting simulated state variable $I_t^{(m)}$ (resp., its accompanying post-action value $S_t^I\left(I_t^{(m)}, a_t^{(m)}\right)$) equals a fixed value for all $t = 1, 2, \ldots, T-1$ although I_t (resp., $S_t^I\left(I_t, a_t\right)$), in principle, can take any value in $\{0, 1, \ldots, t-1\}$ (resp., $\{0, 1, \ldots, t\}$). A consequential annoying issue is that the obtained estimate for the value function/continuation function is invariant to the first-withdrawal-time I_t. This is problematic because the later the PH initializes the withdrawal the larger guaranteed amount $g_t(I_t)$ she could enjoy in the remaining contract life.

(CR1) uniformly simulates the PH's action from its feasible set. By virtue of this, there always exist some paths with $I_t^{(m)} = 0$ which correspond to the scenario that the withdrawal has not been initialized. This in turn guarantees that, in principle, $I_t^{(m)}$ (resp., $S_t^I\left(I_t^{(m)}, a_t^{(m)}\right)$) can take any value in $\{0, 1, \ldots, t-1\}$ (resp., $\{0, 1, \ldots, t\}$). However, this strategy is also not satisfactory: an overwhelming portion of paths are absorbed by the state $W_t = 0$, i.e., the depletion of investment account, and very sparse sample points of the investment account are positive. This is graphically illustrated in the top panel of Figure 2.11 where 1000 sample paths are plotted for the clarity of presentation. So it is not hard to expect that the accuracy of the regression estimate is severely impaired over $\mathcal{K}_{t,R}$.

To alleviate the serious problem mentioned above, **(CR2)** discards the strategy of depleting the investment account, i.e., $(W_t, 1)$, in simulating the PH's action. Therefore, the simulated investment account value $W_t^{(m)}$ can spread over a wider range than that accompanying **(CR1)**; see the bottom panel of Figure 2.11. This phenomenon is more palpable from the histograms of $W_{T-1}^{(m)}$ collected by Figure 2.12. Nevertheless, **(CR2)**'s performance in simulating the $I_t^{(m)}$ is undesirable: Figure 2.13 shows that a substantial portion of sample points of the first-withdrawal-time $I_t^{(m)}$ are concentrated in first few values that I_t can take. To understand the crux, we note that at the first possible withdrawal date, one-half of sample paths exhibit the initiation of the withdrawal; among the remaining paths, one-half of them witness the withdrawal in the consecutive withdrawal date. Therefore, the portion of positive $I_t^{(m)}$ declines at an exponential rate as t increases, which is in line with Figure 2.13. In view of this, it can be expected that the consequential numerical estimate for the value function sustains significant error at state $x = (W, I)$ with a large I.

Overall, none of the above rules **(CR0)**–**(CR2)** gives satisfactory performance. It is hard to figure out an ideal way to randomize the PH's action which can circumvent the thorny issues mentioned above. This shows one drawback of binding together control randomization and forward simulation in addition to the issue of computational cost; see also the item **(I)** "Limitation of control randomization" of Section 2.2.2.

To circumvent the annoying problems mentioned above, in the subsequent numerical experiments, the post-action value of the state process at each time step is simulated as follows: $X_{t^+}^{(m)} := \left(W_{t^+}^{(m)}, I_{t^+}^{(m)}\right)$ where $W_{t^+}^{(m)}$ and $I_{t^+}^{(m)}$ are simulated from two independent uniform distributions with support sets $(0, R)$ and $\{0, 1, \ldots, t\}$, respectively. This ensures the post-action value evenly distributed over $\widetilde{\mathcal{K}}_{t,R} \backslash \widehat{\mathcal{K}}_{t,R}$.

2.6.3 Raw Sieve Estimation v.s. Shape-Preserving Sieve Estimation

In the sequel, we conduct several numerical experiments to compare the regression estimates for the continuation function produced by the following two sieve estimation methods.

- **(RSE)** The first sieve estimation method considered here is the raw sieve estimation (RSE) method which approximates a conditional mean function by Eq. (2.31) with the sieve space $\mathcal{H}_J$ given by (2.28). The basis function $\phi_j(z)$ is chosen as a univariate Bernstein polynomial with degree j; see Appendix A.1 for its expression. The RSE method is essentially the same as the least-squares regression method commonly adopted in the literature; see, for instance, [58].

- **(SPSE)** The second method is the shape-preserving sieve estimation method developed by [83, 84]. Analogous to the RSE method, it approximates a conditional mean function

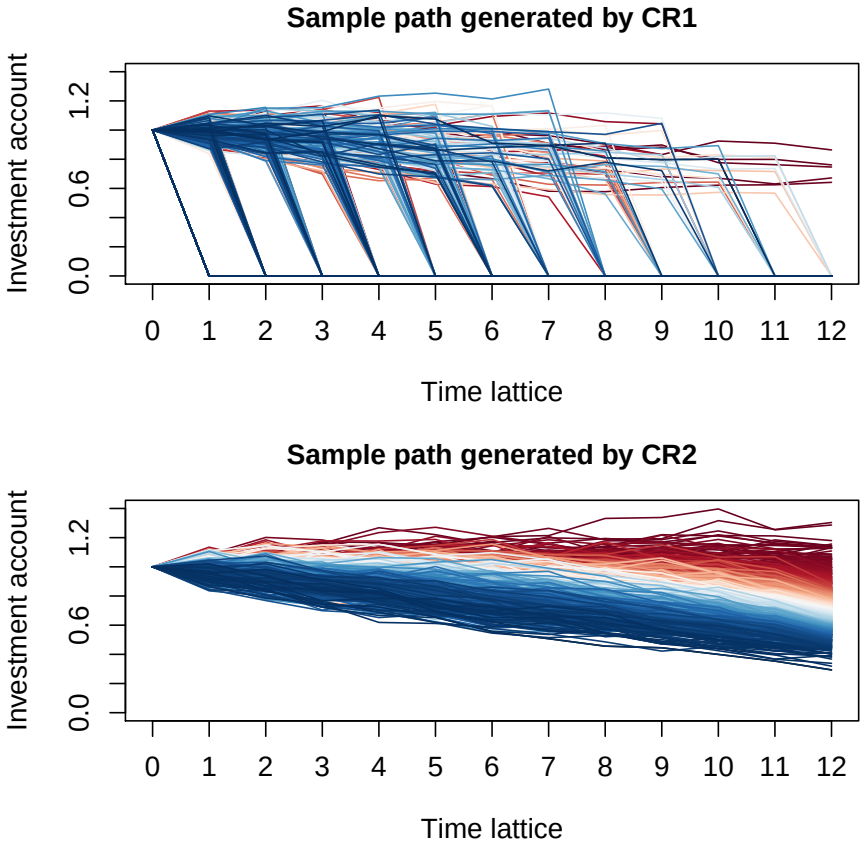


Figure 2.11: Sample paths of the investment account generated by control randomization methods **(CR1)** and **(CR2)**.

by Eq. (2.31) except that the sieve space $\mathcal{H}_J$ is chosen as the shape-preserving sieve space given in Eq. (2.29). In the context of Section 2.5, it is easy to show that $k_1 \mapsto \tilde{C}_t^{\mathrm{E}}(k_1, k_2)$ is a monotone function. So, we choose the shape constraint matrix $\mathbf{A}_J$ in Eq. (2.29) such that the resulting sieve estimate is a monotone function; see Appendix A.1 for the expression of such a matrix $\mathbf{A}_J$.

The first numerical experiment compares the regressione estimates of the SPSE and RSE for the continuation function at each time step. For fairness of comparison, for both methods, $\boldsymbol{\phi}(\cdot)$ is taken as a vector of univariate Bernstein polynomials up to order $J = 20$. Figure 2.14 collects the plots of regression estimates as a function of k_1 at odd time steps with $k_2 = 0$. To better show the subtle difference between the estimates produced by SPSE and RSE, the plots are restricted on the interval $[0, 1]$. From Figure 2.14, one may see that the discrepancy

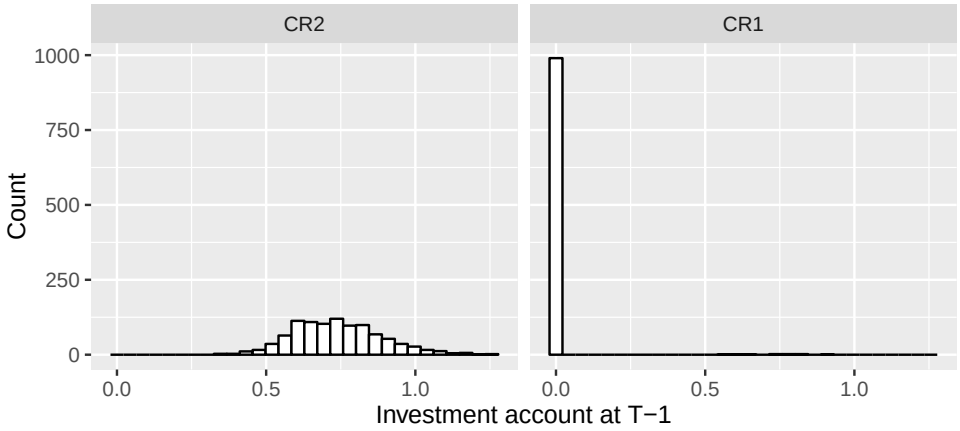


Figure 2.12: Histograms of $W_{T-1}^{(m)}$ generated by control randomization methods **(CR1)** and **(CR2)**.

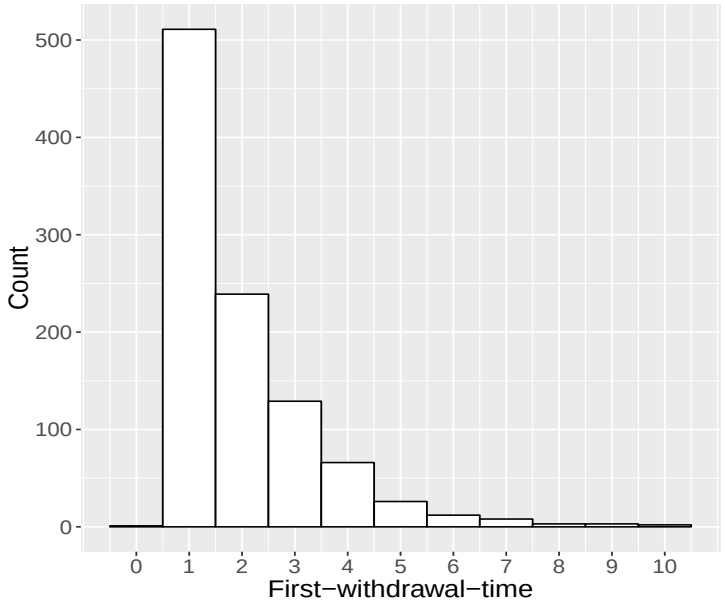


Figure 2.13: Histogram of $I_{11}^{(m)}$ generated by control randomization method **(CR2)**.

between the regression estimates accompanying SPSE and RSE is not conspicuous at large time step but becomes more significant as the time step goes down. Despite the continuation function $\tilde{C}_t(k_1, k_2)$, in principle, is monotone in k_1, the solid lines in Figure 2.14 show that its regression

estimate produced by the RSE does not inherit this monotonicity and loses certain economic interpretations, accordingly. This issue is more serious at smaller time steps as one may see from the bottom panel of Figure 2.14. This is not surprising because once the monotonicity is lost at a certain time step, the regression estimate obtained in the consecutive time step will be influenced, which in turn exaggerates the violation. In contrast, as delineated by the dashed lines in Figure 2.14, the SPSE method always preserves the monotonicity of the continuation function and thus the corresponding regression estimates are economically sensible. This shows the first advantage of the SPSE method in terms of preserving the monotonicity of the continuation function.

Next, we investigate the value function estimate at the initial state, that is, $\tilde{V}_0^{\mathrm{E}}(x_0)$ with $x_0 = (1, 0)$. This quantity approximates the no-arbitrage price of the VA policy at inception and thus is of most interest in the present context; see the last paragraph of Section 2.5.2. It is worth noting that $\tilde{V}_0^{\mathrm{E}}(x_0)$ is essentially a random variable due to the randomness of the simulated sample; see also Remark 2.1. In view of this, the BSBU algorithm is repeated 40 times in order to study the stability of $\tilde{V}_0^{\mathrm{E}}(x_0)$ under a finite sample size. Table 2.2 summarizes the mean and standard deviation of $\tilde{V}_0^{\mathrm{E}}(x_0)$ under different pairs of M and J. The "S.d." column of the table discloses that the standard deviation accompanying the SPSE is nearly one half of that associated with the RSE under all numerical settings. This is more perceivable from Figure 2.15 which depicts the density plots of $\tilde{V}_0^{\mathrm{E}}(x_0)$ under the numerical settings 1, 3, and 4 of Table 2.2: the numerical estimates accompanying SPSE are less volatile as reflected by the more spiked shape of the associated density plots. Overall, the SPSE surpasses the RSE in terms of faster convergence speed.

Table 2.2: Sample mean and standard deviation of $\tilde{V}_0^{\mathrm{E}}(x_0)$. The results are obtained by repeating the BSBU algorithm 40 times.

Setting	(M, J)	SPSE		RSE	
		Mean	S.d.	Mean	S.d.
0	$(1 \times 10^5, 15)$	1.0014	0.0037	1.0051	0.0078
1	$(1 \times 10^5, 20)$	0.9975	0.0036	1.0045	0.0069
2	$(1 \times 10^5, 25)$	0.9984	0.0035	1.0053	0.0061
3	$(2 \times 10^5, 20)$	0.9971	0.0023	0.9996	0.0057
4	$(8 \times 10^5, 20)$	0.9970	0.0015	0.9960	0.0022

From the settings 0-2 of Table 2.2, one may observe that for both methods the sample mean of the accompanying numerical estimate change little as the number of basis functions J hikes from 15 to 25. In the settings 1, 3 and 4 of Table 2.2, we fix $J = 20$ and increase the number of simulated paths M from 10^5 to 8×10^5. We witness that standard deviation decreases as M

climbs. This descending trend is also confirmed by the box plots depicted in Figure 2.16: the height of the box declines as the number of simulated paths hikes. Figure 2.16 also shows that the wedge between the price estimates produced by RSE and SPSE methods tends to shrink as one increases M. All of these show the convergence of the BSBU algorithm which is in line with the convergence result established in Theorem 2.2.

In view of the above observations, we summarize the advantages of the SPSE over the RSE counterpart in two-fold. Firstly, the SPSE produces economically sensible regression estimates by inheriting certain shape properties of the true continuation function. Secondly, the consequential estimate for the optimal value function accompanying the SPSE method is less volatile than that produced by the RSE method under a finite number of simulated sample paths.

2.6.4 Initiation Strategy of the Policyholder

This subsection is devoted to studying the PH's optimal initiation strategy. In the subsequent experiment, we choose $J = 20$ and $M = 2 \times 10^5$ and employ the SPSE method in the BSBU algorithm. This corresponds to the numerical setting 3 of Table 2.2. Suppose that the withdrawal has not been initiated up to time t, which implies $X_t = (w, 0)$ for some w corresponding to the investment account value. Now the PH faces three choices: (i) delaying the withdrawal ($a_t = (0, 0)$), (ii) withdrawing at guaranteed withdrawal amount ($a_t = (g_t(0)w_0, 1)$), and (iii) depleting the investment account ($a_t = (w, 1)$) according to Eq. (2.36). For a realized value of the state process $x = (w, 0)$ and a feasible decision a, one may define the PH's *contract value* at t-th possible withdrawal date as

$$J_t(w, a) = f_t\big(x, a\big) + \tilde{C}_t^{\mathrm{E}}\Big(K\big(x, a\big)\Big).$$

According to the Bellman equation (2.25), the PH aims to maximize the above contract value by choosing a decision among the above three strategies (i)–(iii).

Figure 2.17 depicts the PH's contract value as a function of the investment account w under the three withdrawal strategies, respectively. Two major observations are made from Figure 2.17. Firstly, depleting the investment account is generally suboptimal for the PH as highlighted by the dotted lines in Figure 2.17. Recall from Table 2.1 that a large penalty $\kappa = 0.8$ is imposed on any withdrawal that exceeds the guaranteed amount. Accordingly, depleting the investment account might lead the PH to lose a substantial portion of the contract value and is clearly not a wise choice. This observation implies that the control randomization method **(CR1)** is far from the PH's optimal withdrawal strategy despite that it is intuitively appealing.

The second observation from Figure 2.17 is that delaying the withdrawal is more favorable to the PH in the early phase of the contract life while is not appealing in the later phase. This is clearly reflected by Figure 2.17: the solid lines dominates the dashed lines when $t < 6$ and

this relationship is reversed for $t > 6$. This observation is not a big surprise. Indeed, we recall from Section 2.6.1 that the PH enjoys a guaranteed withdrawal percentage of 7% if she postpones the initiation of the withdrawal until sixth withdrawal date; otherwise, a smaller guaranteed withdrawal percentage of 3% applies. In view of this, the control randomization methods **(CR0)** and **(CR2)** largely deviate from the optimal strategy as they place substantial weights on small values of the first-withdrawal-time I_t in simulating the state process; see the earlier discussions in Section 2.6.2 and Figure 2.13.

2.7 Conclusion

This chapter developed a novel LSMC algorithm, referred to as Backward Simulation and Backward Updating (BSBU) algorithm, to approach numerical solutions to discrete-time stochastic optimal control problems. We first introduced an auxiliary stochastic control problem where the state process only takes value in a compact set. This enables the BSBU algorithm to successfully bypass extrapolating value function estimate. We further showed the optimal value function of the auxiliary problem is a legitimate approximation for that of the original problem with an appropriate choice of the truncation parameter. To circumvent the drawbacks of forward simulation, we proposed to directly simulate the post-action value of the state process from an artificial probability distribution. The pivotal idea behind this artificial simulation method is that the continuation function is solely determined by the distribution of random innovation term. Moreover, motivated by the shape information of the continuation function, We introduced a shape-preserving sieve estimation technique to alleviate the computational burden of tuning parameter selection involved in the regression step of an LSMC algorithm. Furthermore, convergence result of the BSBU algorithm was established by resorting to the theory of nonparametric sieve estimation. Finally, the merits of the BSBU algorithm are confirmed through an application to pricing equity-linked insurance products and the corresponding numerical experiments.

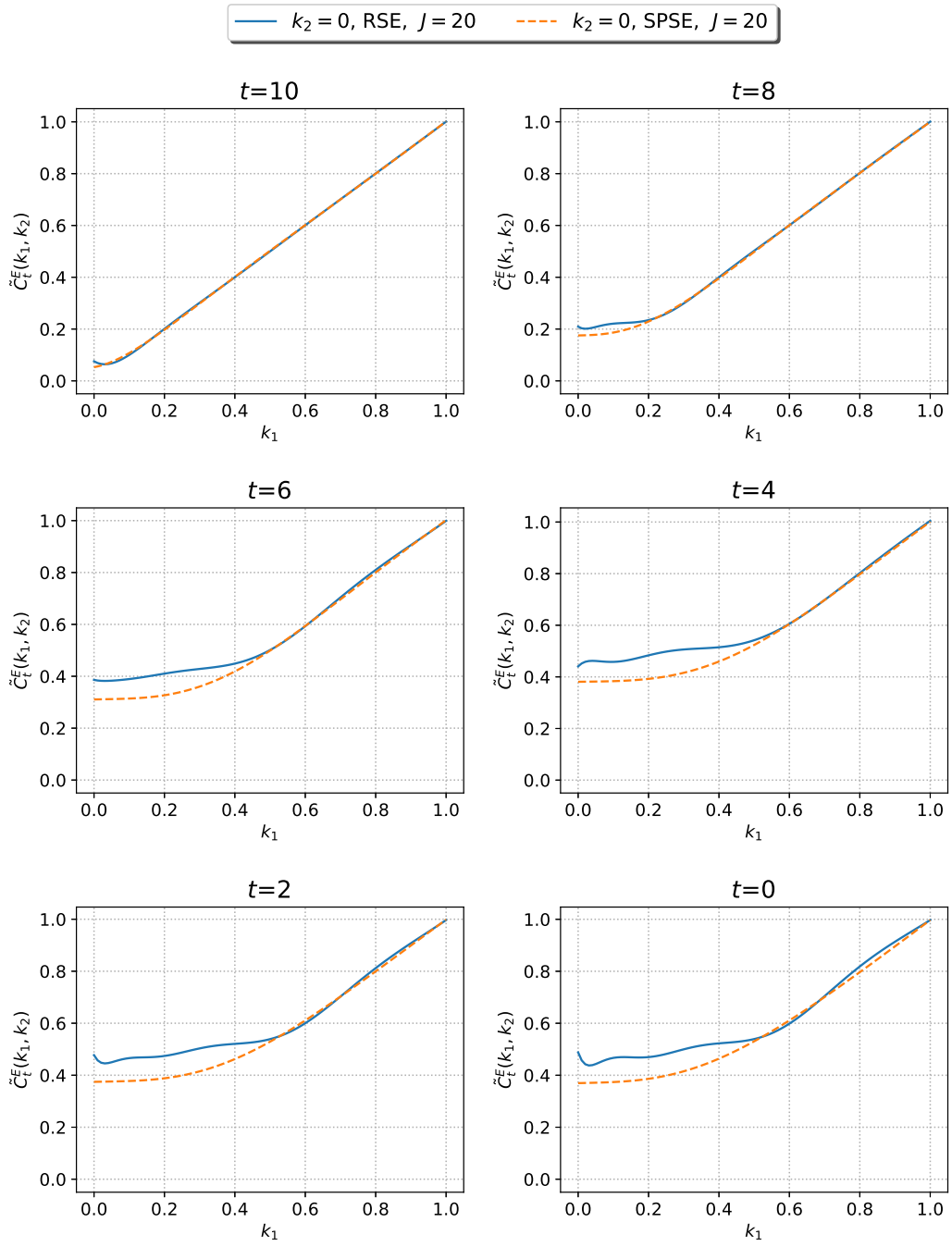


Figure 2.14: Plot of $k_1 \longmapsto \tilde{C}_t^{\mathrm{E}}(k_1, 0)$ with $J = 20$ and $M = 2 \times 10^4$.

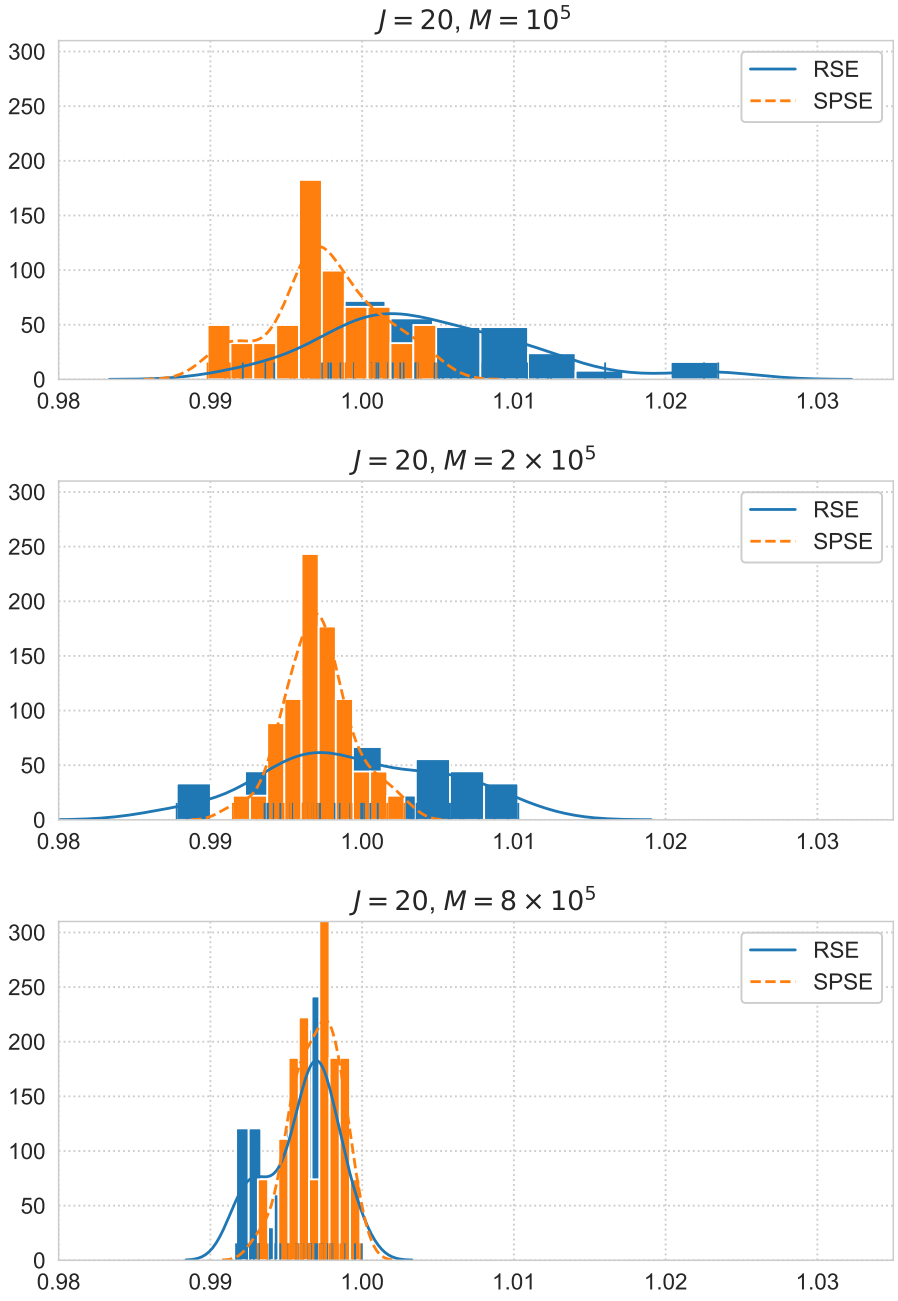


Figure 2.15: Density plots and histograms of $\tilde{V}_0^{\mathrm{E}}(x_0)$ with 40 repeats of the BSBU algorithm.

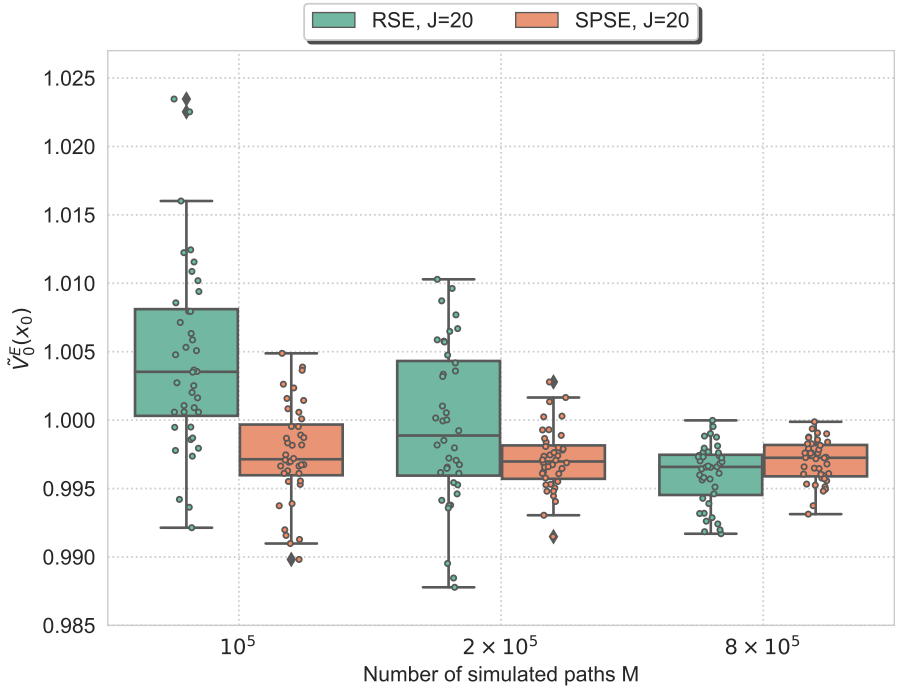


Figure 2.16: Box plots of $\tilde{V}_0^{\mathrm{E}}(x_0)$ with 40 repeats of the BSBU algorithm.

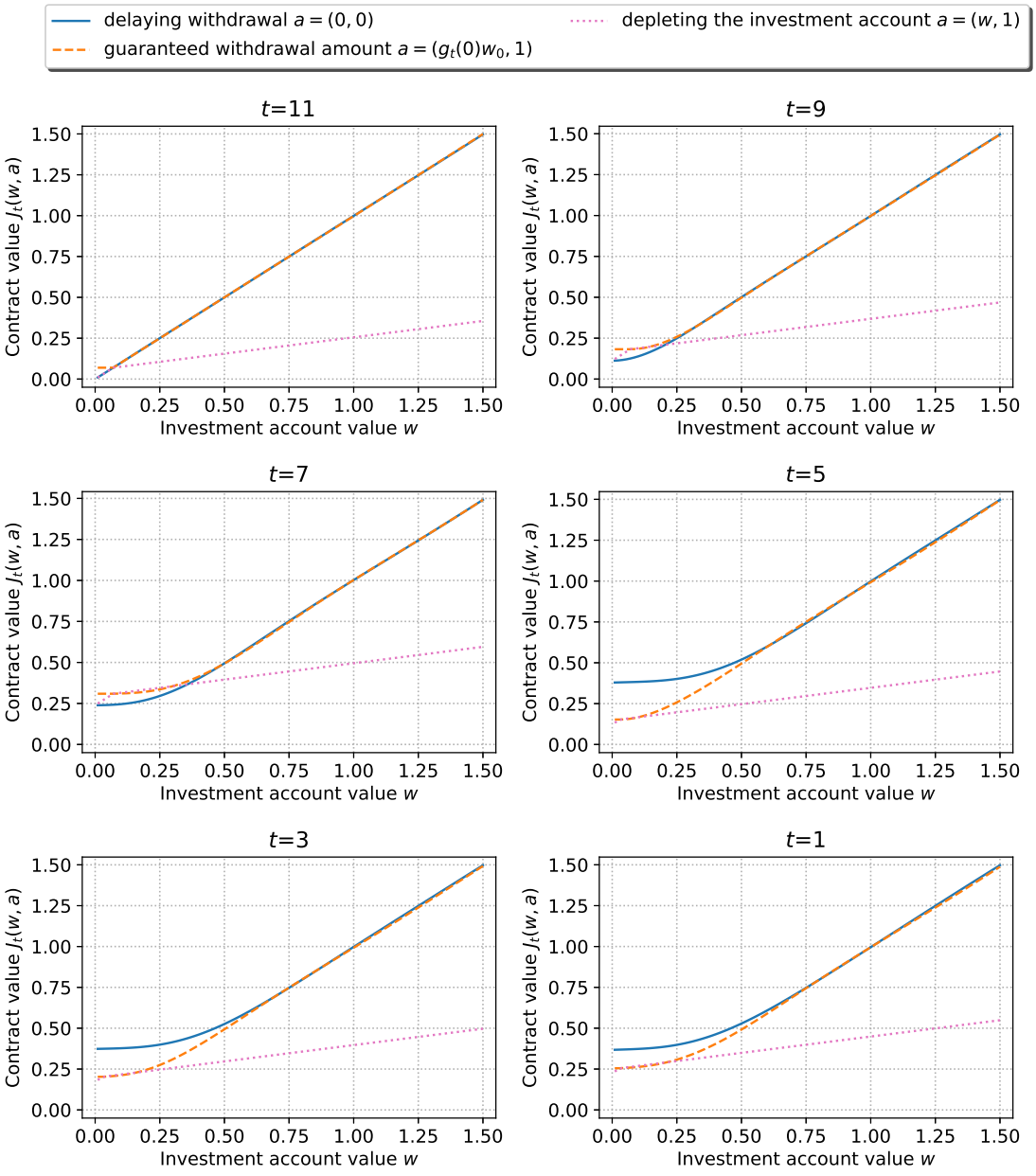


Figure 2.17: Policyholder's contract value under different withdrawal strategies.

Chapter 3

Regression-later Monte Carlo Method for Stochastic Optimal Control Problems

3.1 Introduction

Chapter 2 develops a Least Squares Monte Carlo (LSMC) algorithm, referred to as the Backward Simulation and Backward Updating (BSBU) algorithm, to solve general discrete-time stochastic optimal control problems. As we have seen from Section 2.5, one important motivation of the BSBU algorithm is the problem of pricing equity-linked insurance products which can be formulated as a discrete-time stochastic control problem. It is worth noting that determining the no-arbitrage pricing is only the starting point for managing the financial risk of these products. In reality, insurance companies are more concerned with how to trade available securities to hedge their risk exposures. This entails calculating the sensitivities of the price function with respect to the prices of hedging instruments. However, we recall from Chapter 2 that the numerical estimate for the optimal value function is obtained by solving an optimization problem (see Eq. (2.25)) and therefore has no closed-form expression in general. This observation renders the calculation of hedging ratio fairly cumbersome. The aim of this chapter is to circumvent this difficulty by developing a new LSMC algorithm for general discrete-time stochastic control problems.

The variants of the LSMC algorithm can be roughly categorized into two groups: regression-now and regression-later algorithms. The methods in the first category use regression methods to approximate the continuation function (see the subsequent Eq. (3.5)) involved in the Bellman equation, see, e.g., [25], [81], [44], [53], [24], [16], [87], [47], and the references therein. The BSBU algorithm developed in Chapter 2 is also in a similar spirit and fits into this category. The reason

behind the popularity of the regression-now algorithms stems from two aspects. On one hand, the convergence rate of a nonparametric regression estimate usually depends on the smoothness of the conditional mean function to be approximated; see, e.g., [65]. On the other hand, the smoothness of the continuation function is not hard to expect in many finance applications; see, e.g., [87, Remark 3.9]. In view of these, it is not surprising that the regression-now LSMCs have been widely used to solve various stochastic control problems. Despite the convenience commented above, the regression-now methods have one clear limitation: the optimal value function should be determined by an optimization problem and thus loses its analytical tractability as commented earlier.

The second category of LSMCs directly approximates the optimal value function by a certain regression method. For example, by using the linear sieve estimation method (see Section 2.4.3), the numerical estimate for the value function is solely a linear combination of basis functions which have explicit expression in prior. This idea was first proposed by [43] and the accompanying LSMC algorithm was often referred as the regression-later LSMC (RL-LSMC) algorithm. [43] also discloses the connection between the RL-LSMC and the stochastic mesh method of [22]. The analytical tractability of the value function from the RL-LSMC algorithms comes at a cost. Firstly, the value function of a stochastic control problem is not necessarily a smooth function or even continuous function; see, e.g., [79, pp. 50]. The smoothness assumption on the value function is more restrictive than that on the continuation function which is required by the regression-now algorithms. Secondly, the success of the RL-LSMC algorithm relies on the fast evaluation of the continuation function as it is no longer approximated by a regression method. This imposes more restrictions on the dynamics of the state variables and the type of basis functions. In view of these, one should balance the trade-off between the analytical tractability of the value function and the compromise in the flexibility. Notably that [19] argues that the RL-LSMC algorithm enjoys a faster convergence rate than the regression-now counterpart. This chapter will show that this conclusion holds only when the value function exhibits certain smoothness comparable to what required on the continuation function for the regression-now algorithms; see the subsequent Remark 3.6 for a discussion.

To generalize the idea of regression-later algorithms to solve stochastic control problems, one has to respond to several challenges as below.

(C1) In order to prepare for the ground of the nonparametric regression, one needs to render the value function into a compact domain. A direct application of the auxiliary stochastic control problem constructed by Chapter 2 is impeded by the fact that such a construction might impair the continuity of the value function while the continuity is crucial to applying the RL-LSMC algorithm; see subsequent Section 3.2.2 for a discussion. Thus, one has to propose a new way to construct an auxiliary stochastic control problem in order to inherit the continuity of the original value function.

(C2) As commented in the subsequent Remark 3.1, the continuity of the value function plays

an important role in the convergence of an RL-LSMC algorithm. Thus, it is necessary to investigate the conditions that warrant the continuity of the optimal value function.

(C3) Last but not least, as commented in Chapter 2, the implementation of an LSMC algorithm demands the simulation of the state process. In the context of the BSBU algorithm, we proposed to simulate the post-action value which paves the way for approximating the continuation function. In order to directly estimate the value function by a regression method, the state variable should be directly simulated in a certain way. In particular, one has to examine how to choose the sampling distribution such that the convergence of the RL-LSMC algorithm is ensured.

A recent work [9] develops several RL-LSMC algorithms to solve a class of stochastic optimal control problems. Specifically, in the spirit of the regression-later technique, they directly approximate the optimal value function by a linear combination of basis functions. In terms of the aforementioned challenge **(C3)**, they propose to simulate the state variable at each time step from a "training" measure, which circumvents the annoying problem of simulating the state process in the absence of the optimal control policy. The work of this chapter differs from preceding literature from several dimensions. Firstly, [9] restrict their attention to the stochastic control problems where the transition densities of the state process are bounded or defined on a compact set; see their Assumption 1. This assumption rules out many interesting finance applications as the state process usually has unbounded feasible set. As commented by [9, Remark 1], certain truncation argument is necessary to drop this restriction and this is one thrust of this chapter. Secondly, the convergence result established by [9] is in an L^2 sense that relates to the training measure. However, it might be the case that the point/region where the training measure places small odds exhibits large probability under the real measure induced by the optimal policy. The convergence result of this chapter is stated in a point-wise sense and it will be *independent* of the training measure if the value function is continuous and the sampling distribution satisfies some extra conditions; see the subsequent Remark 3.1 and Assumption 3.9.

The contribution of this chapter is summarized as follows.

1. This chapter gives a construction of an auxiliary stochastic optimal control problem which is different from that of Chapter 2. This auxiliary problem allows us to confine the domain of the state process into a compact set, which in turn paves the way for applying the theory of nonparametric sieve estimation method to prove the convergence of the subsequently developed Monte Carlo algorithm. The value function of this auxiliary problem is shown to be a legitimate proxy for the original value function by giving explicit truncation error bound in the subsequent Theorem 3.1. Mild conditions are established to ensure that the auxiliary value function inherits the continuity, monotonicity, and concavity of the original one; see the subsequent Propositions 3.1, 3.2, and 3.3. This inheritance property brings benefit in solving the local optimization problem in an LSMC algorithm; see the

subsequent Proposition 3.4 for a discussion. It is worth noting that besides the LSMC method, other numerical algorithms for solving stochastic control problems also call for a certain truncation argument to confine the domain of the value function, see, e.g., [40] and [8] for the finite difference method. In view of this, the auxiliary problem also lays the foundation for applying such numerical methods and therefore is of independent interest.

2. This chapter develops an RL-LSMC algorithm which directly approximates the optimal value function by a linear sieve estimation method. Convergence rate of the RL-LSMC algorithm is shown to be faster than the regression-now counterparts under some conditions. This is summarized in the subsequent Theorem 3.2. This generalizes the result of [19] from one-step regression problem to a multi-period stochastic optimal control setting. Such an improvement on the convergence speed of LSMC algorithms comes at the price of more restrictive structure of the stochastic control problem at hand; see Section 3.4 for a dedicated discussion.

3. The developed RL-LSMC algorithm is applied to solve the problem of hedging variable annuity, a prevailing type of equity-linked insurance products, which calls for an efficient calculation of the sensitivity of the value function with respect to the state variable. By the nature of the regression-later algorithm, the value function estimate has an explicit expression. As a result, computing the sensitivities is relatively straightforward and an explicit expression of the delta of the hedging portfolio can be derived; see Appendix B.1 of the chapter. This enriches the relatively thin literature on hedging dynamic withdrawal benefits in variable annuities.

This chapter proceeds as follows. Section 3.2 presents a basic Markovian stochastic optimal control framework and a brief introduction to the linear sieve estimation method. Section 3.3 constructs the auxiliary stochastic control problem and studies the properties of its optimal value function. Section 3.4 presents the RL-LSMC algorithm and Section 3.5 applies it to the problem of delta-hedging of variable annuities. Finally, Section 3.6 concludes the chapter.

3.2 Basic Framework and Preliminaries

As we mentioned in the introduction section, while the RL-LSMC algorithm we will propose brings some additional benefits compared with the regression-now counterparts, it is more restrictive for applications. In other words, we need to impose relatively stronger assumptions for the convergence of the RL-LSMC algorithm. To clarify the motivations of these conditions, it is helpful to go through the stochastic optimal control framework and the nonparametric sieve estimation method once again while the readers may find this section largely overlaps with some sections in Chapter 2.

3.2.1 Stochastic Control Framework

We start by considering a discrete-time stochastic control problem over a collection of consecutive time points labeled by $\mathcal{T} := \{0, 1, \ldots, T\}$. Let $X = \{X_t\}_{t\in\mathcal{T}}$ be a state process valued in $\mathcal{X} \subseteq \mathbb{R}^d$ with $d \in \mathbb{N}$. Starting from a state $X_0 = x_0 \in \mathbb{R}^d$, it evolves recursively according to the following transition equation:

$$X_{t+1} = S\left(X_t, a_t, \varepsilon_{t+1}\right), \quad t \in \mathcal{T}_0 := \mathcal{T}\backslash\{T\}, \tag{3.1}$$

where $\mathsf{a} := \{a_t\}_{t\in\mathcal{T}_0}$ is the action taken by the DM and $\varepsilon := \{\varepsilon_{t+1}\}_{t\in\mathcal{T}_0}$ is a sequence of independent random variables valued in $\mathbb{D} \subseteq \mathbb{R}^q$ with $q \in \mathbb{N}$. For notational brevity, the dependency of X on the action a is suppressed. Throughout the chapter, all the random elements are defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t\in\mathcal{T}}$ generated by ε where $\mathcal{F}_0$ is the trivial σ-algebra.

Further define the set of all admissible actions as follows:

$$\mathcal{A} = \left\{\mathsf{a} = \{a_t\}_{t\in\mathcal{T}_0} \;\middle|\; a_t \text{ is } \mathcal{F}_t\text{-measurable and } a_t \in A_t\left(X_t\right) \text{ for } t \in \mathcal{T}_0\right\}, \tag{3.2}$$

where $A_t : \mathcal{X} \rightrightarrows \mathbb{A}$ is a correspondence (see also Section 2.2), that is, $A_t(x)$ gives a subset of $\mathbb{A}$ for each $x \in \mathcal{X}$. $A_t(X_t)$ corresponds to a certain state constraint subjected by the DM's action at time step t given the state process X_t.

Next we revisit the discrete-time stochastic optimal control problem that we have come across in Chapter 2 as follows:

$$V_0(x_0) = \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t f_t(X_t, a_t) + \varphi^T f_T\left(X_T\right)\right], \tag{3.3}$$

where $\varphi \in (0, 1)$ is a certain constant discounting factor, $f_t : \mathbb{R}^d \times \mathbb{A} \to \mathbb{R}$ and $f_T : \mathbb{R}^d \to \mathbb{R}$ are two measurable functions which correspond to the DM's intermediate and terminal rewards, respectively. The following assumption is same as Assumption 2.1 and to make this chapter self-contained we restate it here:

Assumption 3.1.

$$\sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T} \left|f_t(X_t, a_t)\right|\right] < \infty, \quad \text{and} \quad \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\left|f_T(X_T)\right|\right] < \infty.$$

By exploiting the Dynamic Programming Principle, the optimal value function can be ob-

tained by the following backward recursion equation:

$$\begin{cases} V_T(x) & = f_T(x), \\ V_t(x) & = \sup\limits_{a \in A_t(x)} \Big[f_t(x,a) + \varphi \bar{C}_t(x,a) \Big], \ x \in \mathcal{X}, \quad \text{for} \ \ t = 0, 1, \ldots, T-1, \end{cases} \tag{3.4}$$

where

$$\bar{C}_t(x,a) := \mathbb{E}\left[V_{t+1}(X_{t+1}) \middle| X_t = x, a_t = a \right]. \tag{3.5}$$

3.2.2 Preliminaries on Nonparametric Regression

To prepare for the ground of the LSMC, we give a tour through the nonparametric regression and investigate a particular type of regression method, the *linear sieve estimation* method, in this subsection.

Linear Sieve Estimation Method

Suppose we have a sequence of i.i.d. random observations $\left\{ \left(U^{(m)}, Z^{(m)} \right) \right\}_{m=1}^{M}$, where $U^{(m)}$ and $Z^{(m)}$ are the response and covariate variables, respectively, in a statistical context, with the former taking values in $\mathbb{R}$ and the latter in $\mathbb{R}^r$. The nonparametric regression aims to estimate the conditional mean (regression) function $g(\cdot) := \mathbb{E}\left[U^{(m)} \middle| Z^{(m)} = \cdot \ \right]$ from the random sample. Here $g(\cdot)$ is assumed to be a continuous function over a compact set $\mathcal{Z}$. In the linear sieve estimation method, this goal is achieved by studying the following optimization problem:

$$\hat{g}(\cdot) := \arg \inf_{h(\cdot) \in \mathcal{H}_J} \frac{1}{M} \sum_{m=1}^{M} \left[U^{(m)} - h\left(Z^{(m)} \right) \right]^2, \tag{3.6}$$

where $\mathcal{H}_J$ is a finite-dimensional functional space depending on a tuning parameter J and is called as *sieve space*.

A typical choice of $\mathcal{H}_J$ is the linear span generated by a sequence of basis functions, for instance, polynomials with different degrees. Consider the following sieve space:

$$\mathcal{H}_J = \left\{ h(\cdot) : \quad h(z) = \sum_{j=0}^{J} \beta_j \phi_j(z), \quad \mathbf{A}_J \boldsymbol{\beta}_J \geq \mathbf{0}_{b(J)} \right\}, \tag{3.7}$$

where $\boldsymbol{\beta}_J := (\beta_0, \ldots, \beta_J)^\intercal$, $b(\cdot) : \mathbb{N} \longrightarrow \mathbb{N}$ is some integer-valued function, $\mathbf{A}_J$ is a $b(J)$-by-$(J+1)$ matrix, and $\mathbf{0}_{b(J)}$ is a $b(J)$-by-1 null vector. When the matrix $\mathbf{A}_J$ is a null matrix,

the sieve estimation considered here degenerates to the least-squares linear regression which is widely used in the literature of LSMC; see, e.g., [58]. Generally, the linear constraints $\mathbf{A}_J \boldsymbol{\beta}_J \geq \mathbf{0}_{b(J)}$ are imposed to ensure the regression estimate $\hat{g}(\cdot)$ inherits certain shape properties such as nonnegativity, monotonicity, and convexity from the true regression function $g(\cdot)$; see [83] and [84]. Some specific forms of $\mathbf{A}_J$ are relegated to Appendix A.1.1.

The core idea of the sieve estimation can be summarized as a two-stage approximation. In the first stage, the true regression function $g(\cdot)$ is approximated by $g_J^{\mathrm{o}}(\cdot)$ (referred to as the *oracle* in the literature, see e.g. [57, pp. 25]) as given by

$$g_J^{\mathrm{o}}(\cdot) = \arg \inf_{h \in \mathcal{H}_J} \|h - g\|, \tag{3.8}$$

with $\|\cdot\|$ denoting the supremum norm. In the second stage, for a given J, the oracle $g_J^{\mathrm{o}}(\cdot)$ is approximated by the sieve estimate $\hat{g}(\cdot)$ given by the preceding Eq. (3.6) and the approximation error (in the sense of the supremum norm) decays as one increases the sample size M.

Conditions for the Convergence

Below, we present a technical assumption associated with the validity of the preceding two-stage approximation procedure which is proposed by [65]. This assumption is also commonly referred to for the convergence of an LSMC algorithm in the literature when the sieve estimation method is involved in the regression step of the algorithm; see, e.g., [78] and [13]. We will frequently refer to the conditions of the subsequent Assumption 3.2 when we approach some critical questions arising from the RL-LSMC algorithm in the subsequent subsection in order to clarify the motivations of our approach.

Assumption 3.2. **(i)** $\left\{\left(U^{(m)}, Z^{(m)}\right)\right\}_{m=1}^{M}$ *are i.i.d. and* $Z^{(m)}$ *has compact support* $\mathcal{Z}$*. Furthermore,* $\mathbb{V}\mathrm{ar}\left[U^{(m)} \middle| Z^{(m)} = \cdot\right]$ *is bounded over* $\mathcal{Z}$*.*

(ii) *There exists a sequence* $\Upsilon(J)$ *such that* $\|\boldsymbol{\phi}\| \leq \Upsilon(J)$ *with* $\|\cdot\|$ *denoting the supremum norm of a continuous function over* $\mathcal{Z}$*.*

(iii) *For the sieve space* $\mathcal{H}_J$ *defined in Eq. (3.7), there exists a* $(J+1)$*-by-1 vector* $\boldsymbol{\beta}^{\star}$ *and a sequence* $\hat{\rho}_J$ *such that* $\hat{\rho}_J \longrightarrow 0$ *as* $J \longrightarrow \infty$*, and*

$$\|g_J^{\mathrm{o}} - g\| = \|\boldsymbol{\phi}^{\intercal}\boldsymbol{\beta}^{\star} - g\| = O\left(\hat{\rho}_J\right),$$

with $g_J^{\mathrm{o}}(\cdot)$ *defined in Eq. (3.8).*

(iv) *Let* $\Phi := \mathbb{E}\left[\boldsymbol{\phi}\left(Z^{(m)}\right)\boldsymbol{\phi}^{\intercal}\left(Z^{(m)}\right)\right]$*. There exists a positive constant* $\underline{c}_{\Phi}$ *independent of* J *such that* $0 < \underline{c}_{\Phi} \leq \lambda_{\min}(\Phi) \leq \lambda_{\max}(\Phi) \leq \bar{c}_{\Phi} < \infty$*, with* $\lambda_{\min}(\Phi)$ *and* $\lambda_{\max}(\Phi)$ *denoting the smallest and largest eigenvalues of* Φ*, respectively.*

(v) *As* $M \longrightarrow \infty$, $J \longrightarrow \infty$, *and* $\Upsilon^2(J)J/M \longrightarrow 0$.

In the above assumption, Part (i) requires the compactness of the support $\mathcal{Z}$, which makes the direct application of the sieve estimation method to the stochastic control problem theoretically shaky; see item "**(Q1)**" in the subsequent subsection for more discussions. The expression of $\Upsilon(J)$ in Part (ii) depends on the choice of the basis function $\boldsymbol{\phi}$; see [65] for a detailed discussion. Part (iii) requires the continuity of the conditional mean function g so that g_J^{o} is well-defined and thus restricts the application of RL-LSMC algorithms to some stochastic control problems; see item "**(Q2)**" in the sequel for more discussions. Part (v) specifies the growth rate of J and M.

Part (iv) might seem opaque but is mild. Indeed, [65, pp. 156] shows that for power series, this condition can be replaced by a more transparent condition as stated in Assumption 3.3 below. See also [5, pp. 320–327] for discussions under other possible basis functions.

Assumption 3.3. *$\mathcal{Z}$ is a Cartesian product of compact connected intervals. Furthermore, the sampling distribution of $Z^{(m)}$ has an absolutely continuous density that is bounded away from zero.*

Assumption 3.3 guides the choice of the sampling distribution in the subsequently developed RL-LSMC algorithm; see the subsequent Section 3.4.3.

Before ending this subsection, we make a remark on an important implication of the continuity of the conditional mean function $g(\cdot)$.

Remark 3.1. *The above discussions focus on the case where the function $g(\cdot)$ is continuous. If $g(\cdot)$ is not continuous, one may sill get a sieve estimate $\hat{g}(\cdot)$ from Eq. (3.6) but $\hat{g}(\cdot)$ converges to $g_{\mu,J}^{\mathrm{o}}(\cdot)$ (in an L^2 sense) as defined by*

$$g_{\mu,J}^{\mathrm{o}}(\cdot) = \arg \inf_{h \in \mathcal{H}_J} \int_{\mathcal{Z}} |g(z) - h(z)|^2 \mu(\mathrm{d}z), \tag{3.9}$$

with μ denoting the sampling measure of $Z^{(m)}$; see e.g. [27]. $g_{\mu,J}^{\mathrm{o}}(\cdot)$ defined in the above certainly depends on μ. That is, in such a situation, the sieve estimate generally depends on the sampling measure and different choices of μ might yield different limit functions.

*The implication of this observation in the context of the LSMC algorithm is that if the sampling measure does not agree with the measure driven by the optimal action (which is indeed the general case), the LSMC algorithm might fail to converge to the optimal solution to the stochastic control problem. The above discussion pinpoints the subtle role played by the continuity of the regression function in an LSMC algorithm; see also the item "***(Q2)***" in the sequel.*

3.2.3 LSMC for Stochastic Control Problems

Recall from Eq. (3.5) that the continuation function $\bar{C}_t(\cdot,\cdot)$ is defined as the conditional mean function:

$$\bar{C}_t(\cdot,\cdot) := \mathbb{E}\Big[\underbrace{V_{t+1}(X_{t+1})}_{\text{response variable}} \,\Big|\, \underbrace{(X_t, a_t)}_{\text{covariate variable}} = (\cdot,\cdot)\Big].$$

This observation naturally motivates one to take $V_{t+1}(X_{t+1})$ and (X_t, a_t) as response and covariate variables, respectively, and employ a certain nonparametric regression method to get a numerical estimate for the continuation function, say $\bar{C}_t^{\mathrm{E}}(\cdot,\cdot)$. The numerical estimate for the optimal value function is accordingly given by Eq. (3.4) with $\bar{C}_t(\cdot,\cdot)$ replaced by $\bar{C}_t^{\mathrm{E}}(\cdot,\cdot)$. This is one of the key ideas behind the LSMC algorithm and was incubated in a series of seminal papers including [25], [58], and [81].

On the other hand, we can think of the optimal value function $V_t(X_t)$ as an expectation of itself conditioning on the state variable X_t although the associated conditional distribution is degenerated. In view of this, one may alternatively take $V_t(X_t)$ and X_t as response and covariate variables, respectively, and exploit the nonparametric regression to directly recover the value function. The above methods of approximating the continuation function $\bar{C}_t(\cdot,\cdot)$ and value function $V_t(\cdot)$ by nonparametric regression are referred to as the *regression-now* and *regression-later* methods, respectively, in the literature; see, e.g., [44].

To sum up, the key idea of the LSMC is to either approximate the continuation function (regression-now) or the value function (regression-later) by some nonparametric regression method. To apply the LSMC algorithm to solve stochastic control problems, we need to approach the following questions.

(Q1) *How to confine the domain of continuation/value function into a compact set?*
Recall that Part (i) of the preceding Assumption 3.2 requires the regression function is solely defined on a compact domain. This is crucial to the convergence result of the linear sieve estimation method (or general nonparametric regression methods) and conventionally assumed in the statistics literature. However, in the current context, this requirement is incongruous with the fact that the value function or continuation function of a stochastic control model is commonly defined on an unbounded set $\mathcal{X}$ (see Eq. (3.4)) which is the range where the accompanying state process might take value. Therefore, a direct application of nonparametric regression to stochastic control problems might be problematic. This thorny issue is circumvented by studying an auxiliary stochastic control problem as we will see in the subsequent Section 3.3.

(Q2) *How to ensure the continuity of continuation/value function?*
As one may see from preceding Assumption 3.2 and Remark 3.1, the continuity of the

regression function plays a pivotal role in establishing the convergence result. Back to the stochastic control framework, the regression function corresponds to the continuation function or the optimal value function. The continuity of the continuation function is not hard to warrant since it is by definition a certain conditional expectation that can be taken as a smoothing operator. Even if $V_{t+1}(\cdot)$ is not continuous, $\bar{C}_t(\cdot)$ defined through Eq. (3.5) can be a smooth function in many finance applications; see [87, Remark 3.9] for a discussion. However, it is more restrictive to require the value function to be continuous and more dedicated investigation should be conducted.

(Q3) *How to generate the random sample of response and covariate variables?*
In typical statistical contexts, the sample of response and covariate variables are observed at the first hand. In contrast, in the context of stochastic control problems, the random sample of this pair is generated by Monte Carlo simulation which requires the simulation of the state process X. This brings a new challenge since the evolution of the state process is driven by the DM's optimal action which is not tractable in prior. This problem is often resolved by the control randomization technique; see, e.g., [51]. Chapter 2 proposes to simulate the post-action value of the state process, which eliminates the needs of control randomization and forward simulation. In a similar spirit of [9] and the artificial simulation method of Chapter 2, this chapter proposes to directly simulate X_t instead of its accompanying post-action value. This will bring several benefits which will be addressed in details in the remainder of the chapter.

Before ending this section, we present the road map of this chapter. In response to **(Q1)**, The subsequent section will construct an auxiliary stochastic control problem whose optimal value function is defined on a compact set. This value function will be shown to be a legitimate proxy for that of the original problem (3.3) in Theorem 3.1 in the sequel. The continuity of the value function is further established under some regular conditions see the subsequent Proposition 3.1, which responds to **(Q2)**. These pave the way for applying the nonparametric sieve estimation method to approximating the value function. In terms of **(Q3)**, Section 3.4 proposes to directly simulate the state variable at each time step from a certain probability distribution. The accompanying advantages over the simulation of post-action value proposed in Chapter 2 will be also be addressed.

3.3 A Truncation Argument

The aim of this section is to construct an auxiliary stochastic control problem where the accompanying optimal value function satisfies that (i) it is solely defined on a bounded domain, and that (ii) it is a continuous function. The thrust behind this is in two-fold. Firstly, the convergence of the LSMC algorithm developed in the subsequent Section 3.4 hinges on these two properties

as a result of the application of the sieve estimation theory introduced in the previous section. Secondly, this construction of auxiliary problem bypasses any theoretically unsound extrapolation in the backward recursion process of an LSMC algorithm; see Chapter 2 and Remark 3.4 in the sequel for more comments.

3.3.1 Construction

Below we construct an auxiliary state process. The difference between this construction and the one used in Section 2.3 of Chapter 2 is subtle and will be discussed in the subsequent Section 3.3.3.

We construct a new state process $\hat{X} := \left\{\hat{X}_t\right\}_{t\in\mathcal{T}}$ as follows:

$$\begin{cases} \hat{X}_0 &= x_0 \in \mathcal{X}_R \subseteq \mathbb{R}^d, \\ \hat{X}_t &= \mathcal{Q}\left(S\left(\hat{X}_t, \hat{a}_t, \varepsilon_{t+1}\right)\right) =: \hat{S}\left(\hat{X}_t, \hat{a}_t, \varepsilon_{t+1}\right), \quad \text{for} \quad t = 1, 2, \ldots, T, \end{cases} \tag{3.10}$$

where the function $\mathcal{Q} : \mathcal{X} \to \mathrm{cl}\left(\mathcal{X}_R\right)$ is defined in Eq. (2.12), $\hat{\mathsf{a}} := \{\hat{a}_t\}_{t\in\mathcal{T}_0} \in \hat{\mathcal{A}}$ and

$$\hat{\mathcal{A}} = \left\{\hat{\mathsf{a}} = \{\hat{a}_t\}_{t\in\mathcal{T}_0} \;\middle|\; \hat{a}_t \text{ is } \mathcal{F}_t\text{-measurable and } \hat{a}_t \in A_t\left(\hat{X}_t\right) \text{ for } t \in \mathcal{T}_0\right\}. \tag{3.11}$$

In relative to the original stochastic control problem (3.3), we consider the following auxiliary problem:

$$\hat{V}_0(x_0) = \sup_{\hat{\mathsf{a}}\in\hat{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t f_t\left(\hat{X}_t, \hat{a}_t\right) + \varphi^T f_T\left(\hat{X}_T\right)\right]. \tag{3.12}$$

By comparing the above equation with Eq. (3.3), we observe that the reward functions of these two stochastic control problems are the same. The essential difference stems from the discrepancy between the state processes X and $\hat{X}$. Due to the presence of state constraint $A_t(\cdot)$, the admissible sets of the PH's actions in these two problems also differ with each other and are given by $\mathcal{A}$ and $\hat{\mathcal{A}}$, respectively. This subtle difference makes the investigation of the gap between $V_0(x_0)$ and $\hat{V}_0(x_0)$ nontrivial; see the subsequent subsection for a discussion.

For the auxiliary stochastic control problem (3.12), the corresponding Bellman equation is given by

$$\begin{cases} \hat{V}_T(x) &= f_T(x), \\ \hat{V}_t(x) &= \sup\limits_{a\in A_t(x)} \left[f_t(x,a) + \varphi \hat{C}_t(x,a)\right], \; x \in \mathrm{cl}\left(\mathcal{X}_R\right), \quad \text{for} \quad t = 0, 1, \ldots, T-1, \end{cases} \tag{3.13}$$

where

$$\hat{C}_t(x,a) := \mathbb{E}\left[\hat{V}_{t+1}(X_{t+1})\middle| X_t = x, a_t = a\right]. \tag{3.14}$$

By comparing $\hat{V}_t(\cdot)$ with $V_t(\cdot)$, we have one important observation. On one hand, $V_t(\cdot)$ is defined on the set $\mathcal{X}$ which might be unbounded in many finance applications because financial assets prices usually take values in the positive real line. On the other hand, $\hat{V}_t(\cdot)$ is solely defined on the compact set $\mathrm{cl}\,(\mathcal{X}_R)$. This observation lays the foundation for applying the linear sieve estimation method introduced in Section 3.2.2 to recover the optimal value function $\hat{V}_t(\cdot)$ over $\mathrm{cl}\,(\mathcal{X}_R)$ as Part (i) of Assumption 3.2 is automatically satisfied. It remains to show that $\hat{V}_0(x_0)$ is a valid approximation for $V_0(x_0)$ and this will be addressed in the sequel.

3.3.2 Properties of Value Function

This subsection establishes the connection between the optimal value functions of the stochastic control problems (3.3) and (3.12).

Approximation error of $\hat{V}_0(x_0)$

Recall that these two stochastic control problems exhibit different admissible sets for the DM's action as given by Eqs. (3.2) and (3.11), respectively. This brings technical challenges in investigating the gap between $V_0(x_0)$ and $\hat{V}_0(x_0)$. To circumvent this, we put these two optimization problems into a common admissible set:

$$\bar{\mathcal{A}} := \left\{ \bar{\mathsf{a}} = \{\bar{a}_t\}_{t\in\mathcal{T}_0} \;\middle|\; \bar{a}_t \text{ is } \mathcal{F}_t\text{-measurable and } \bar{a}_t \in \mathbb{A} \text{ for } t \in \mathcal{T}_0 \right\}. \tag{3.15}$$

It is easy to see that the above set $\bar{\mathcal{A}}$ is richer than the previous sets $\mathcal{A}$ and $\hat{\mathcal{A}}$ because the state constraint $A_t(\cdot)$ is dropped in $\bar{\mathcal{A}}$.

We further impose the following technical assumptions.

Assumption 3.4. *Let $X_0 = x_0 \in \mathcal{X}_R$. There exists a measurable function $\mathcal{E}_{x_0,T}(\cdot) : \mathbb{R}_{>0} \longrightarrow [0,1]$ such that*

$$\inf_{\mathsf{a}\in\bar{\mathcal{A}}} \mathbb{P}\Big[X_t \in \mathcal{X}_R, \quad \text{for all} \quad 0 \le t \le T\Big] \ge 1 - \mathcal{E}_{x_0,T}(R). \tag{3.16}$$

$\mathcal{E}_{x_0,T}(R)$ in the above assumption is an upper bound for the probability that the auxiliary state process $\hat{X}$ disagrees with X at some time point before maturity T. Since the discrepancy between problems (3.3) and (3.12) essentially stems from different constructions of the accompanying state

processes, there is no surprise that $\mathcal{E}_{x_0,T}(R)$ plays a crucial role in controlling such a disagreement. To get a conservative estimate for this probability bound, one may construct a new stochastic process that is independent of the DM's action and always dominates X. Accordingly, it boils down to estimating the probability that this new process exits the region $\mathcal{X}_R$; see Appendix B.1 for a detailed discussion.

Assumption 3.5. **(i)** *For each $t \in \mathcal{T}_0$, there exists a measurable function $h_t(\cdot) : \mathcal{X} \longrightarrow \mathbb{R}$ such that $f_t(x,a) \geq h_t(x)$ for all $a \in \mathbb{A}$.*

(ii) *There exists a measurable function $\bar{B}(\cdot) : \mathcal{X} \longrightarrow \mathbb{R}_{>0}$ such that*

$$\left|f_T(x)\right|^2 \leq \bar{B}(x) \quad \text{and} \quad \sup_{a\in\mathbb{A}} \left|\bar{f}_t(x,a)\right|^2 \leq \bar{B}(x), \quad \text{for all} \quad t \in \mathcal{T}_0,$$

where $\bar{f}_t(x,a) := f_t(x,a)\mathbb{1}_{\{a\in A_t(x)\}} + h_t(x)\mathbb{1}_{\{a\notin A_t(x)\}}$.

(iii) *There exists a measurable function $\bar{\xi}(\cdot) : \mathbb{R}_{>0} \longrightarrow \mathbb{R}_{>0}$ and a constant ζ_X such that*

$$\sup_{x\in\mathcal{X}_R} \bar{B}(x) \leq \bar{\xi}(R) \quad \text{and} \quad \sup_{\mathsf{a}\in\bar{\mathcal{A}}} \mathbb{E}\left[\bar{B}(X_t)\right] \leq \zeta_X, \quad \text{for all} \quad t \in \mathcal{T}_0.$$

Remark 3.2. *Cautious readers might tell the nuanced difference between the above assumption and Assumption 2.3 of Chapter 2. In particular, the above Part (i) is new and one sufficient condition to ensure this is that the intermediate reward functions are positive. In this case, we can choose $h_t(x) \equiv 0$ in Assumption 3.5. The introduction of Part (i) is for the sake of relaxing the state constraint and paving the way for proving the subsequent Theorem 3.1; see also the subsequent Remark 3.3[1]. Although this condition is mild for many real applications, it is not required in Chapter 2. This is the price one has to pay for the advantages got from the new truncation argument developed in this chapter over that in Chapter 2; see the subsequent Section 3.3.3 for a discussion.*

Given Part (i), Part (ii) is fulfilled if Assumption 2.3 holds because one may choose $\bar{B}(x) = B(x) \vee \max_{t\in\mathcal{T}_0} h_t(x)$ with $B(x)$ given in Assumption 2.3. $\bar{B}(\cdot)$ usually has a polynomial expression. Accordingly, the constant ζ_X controls the moment of the state process X of a certain degree, say $d \in \mathbb{N}$. To get such a constant, we can construct another stochastic process, say Y_t, which is independent of the DM's action a and always dominates X_t. Then, we can choose ζ_X as $\max_{t\in\mathcal{T}} \mathbb{E}\left[Y_t^d\right]$. The above discussion will be more perceivable in the specific example presented in the subsequent Section 3.5. Overall, it is generally not hard to determine $\bar{B}(\cdot)$, $\bar{\xi}(\cdot)$, and ζ_X; see Appendix B.1 for the verification of the above assumptions for the example of Section 3.5.

The legitimacy of using $\hat{V}_0(x_0)$ as a proxy for $V_0(x_0)$ is derived from the following theorem.

[1]We are grateful to Professor Alexander Schied for inspiring discussions on this technical issue.

Theorem 3.1. *Suppose Assumptions 3.1, 3.4, and 3.5 hold. In addition, assume that for each* $(x, \bar{a}) \in \mathcal{X} \times \mathbb{A}$, *there exists* $a \in A_t(x)$ *such that* $S(x, \bar{a}, e) = S(x, a, e)$ *for* $t \in \mathcal{T}_0$ *and* $e \in \mathbb{D}$. *Then,*

$$\left|V_0(x_0) - \hat{V}_0(x_0)\right| \leq \sqrt{2(T+1)\frac{1-\varphi^{2(T+1)}}{1-\varphi^2}\mathcal{E}_{x_0,T}(R)\left(\bar{\xi}(R) + \zeta_X\right)}. \tag{3.17}$$

The proof of the above theorem is relegated to Appendix B.2. Below we make some comments on the additional assumption other than Assumptions 3.1, 3.4, and 3.5 in Theorem 3.1.

Remark 3.3. *The additional assumption naturally holds in the degenerated case where* $A_t(x) \equiv \mathbb{A}$ *for all* $x \in \mathcal{X}$, *that is, the feasible set of the DM's is state-independent. In the example considered in Section 3.5, one may encounter the general case where the feasible set is state-dependent.*

This additional requirement is used to place the stochastic control problems (3.3) *and* (3.12) *in the same admissible set* $\bar{\mathcal{A}}$ *(see Corollary B.2 of Appendix B.2), which brings much convenience in investigating the gap between* $V_0(x_0)$ *and* $\hat{V}_0(x_0)$, *as one may see from the proof of the Theorem 3.1. It is also interesting to compare Corollary B.2 with Corollary A.1 accompanying Chapter 2, which shows the subtle difference between the ideas of proving the above theorem and Theorem 2.1 of Chapter 2; also see Remark 2.3.*

The result of Theorem 3.1 can be understood as follows. The inequality (3.17) gives an explicit upper bound for the discrepancy between $\hat{V}_0(x_0)$ and $V_0(x_0)$. Firstly, recall that the discrepancy of problems (3.3) and (3.12) primarily stems from the difference between the accompanying state processes. Therefore, $\mathcal{E}_{x_0,T}(R)$ appears in the above error bound which characterizes the probability of disagreement between X and $\hat{X}$; see the subsequent paragraph below Assumption 3.4. Furthermore, it can be expected that the gap between $\hat{V}_0(x_0)$ and $V_0(x_0)$ widens as the time horizon is prolonged because the probability of the original state process X leaving the truncated region $\mathcal{X}_R$ increases as time progresses. This intuition is confirmed by the dependency of function $\mathcal{E}_{x_0,T}(\cdot)$ on T. It is also perceivable from Assumption 3.5 that the term $\sqrt{\bar{\xi}(R) + \zeta_X}$ corresponds to an upper bound for the difference between the reward functions of the two stochastic control problems (3.3) and (3.12) at each time step. Since there are $T+1$ reward functions in total, it is not surprising that the factor $T+1$ appears in the preceding inequality (3.17). Finally, the factor $\frac{1-\varphi^{2(T+1)}}{1-\varphi^2}$ appears due to the presence of a discounting multiplier in front of each reward function.

The crucial implication of the inequality (3.17) is that $\hat{V}_0(x_0)$ is a legitimate approximation for $V_0(x_0)$ as long as the error bound in inequality (3.17) is marginal. This is often achieved when (i) the truncation parameter R is sizable and (ii) the initial state x_0 is deep inside the region $\mathcal{X}_R$. This will be more perceivable in the concrete example in Section 3.5 and the associated Appendix B.1. Figure 3.1 depicts $\mathcal{E}_{x_0,T}(R)$ and the error bound in (3.17) as functions of truncation parameter R under the parameter setting in the subsequent Section 3.5.4. We can clearly see from Figure 3.1

that the error bound (dashed line) shrinks fairly fast as one increases R, which is not surprising as $\mathcal{E}_{x_0,T}(R)$ (solid line) decays faster than the growth rate of $\bar{\xi}(R)$.

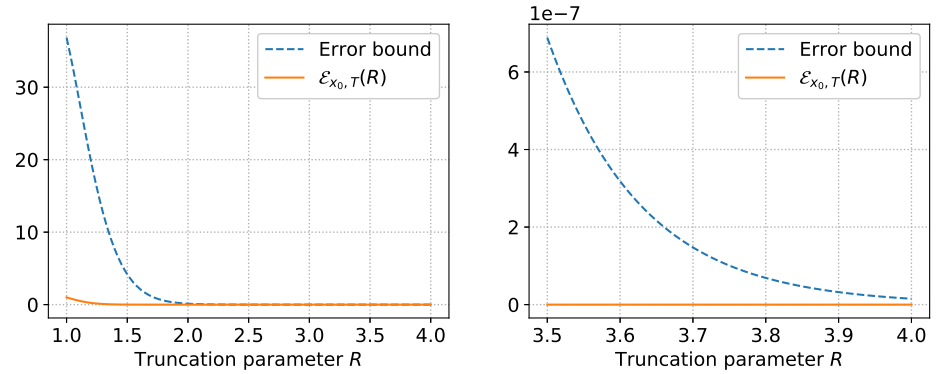


Figure 3.1: Plots of $\mathcal{E}_{x_0,T}(R)$ and the error bound in (3.17) as functions of truncation parameter R. The left and right panels depict the plots over the intervals $[1, 4]$ and $[3.5, 4]$, respectively.

Continuity of $\hat{V}_t(\cdot)$

In the following, we will show the continuity of $\hat{V}_t(\cdot)$ under some conditions.

Assumption 3.6. **(i)** *For each $t \in \mathcal{T}_0$, $A_t : \mathcal{X} \rightrightarrows \mathbb{A}$ is a compact valued and continuous correspondence (see Appendix B.2.2 for a definition).*

(ii) *For each $t \in \mathcal{T}_0$, $f_t : \mathcal{X} \times \mathbb{A} \to \mathbb{R}$ is continuous. In addition, $f_T : \mathcal{X} \to \mathbb{R}$ and $S : \mathcal{X} \times \mathbb{A} \times \mathbb{D} \to \mathcal{X}$ are continuous functions.*

If $\mathbb{A} \subseteq \mathbb{R}$, a sufficient condition for Part (i) of the above assumption to hold is $A_t(x) = [\theta_{0,t}(x), \theta_{1,t}(x)]$ where $\theta_{j,t} : \mathcal{X} \to \mathbb{R}$ are two continuous functions with $\theta_{0,t}(x) \leq \theta_{1,t}(x)$ for all $x \in \mathcal{X}$. The continuity of reward functions and transition equation required by Part (ii) in the above is used to ensure the application of the Berge's Maximum Theorem (see Lemma B.3 of Appendix B.2.2) to proving Proposition 3.1 in the sequel.

Proposition 3.1. *Suppose Assumptions 3.1 and 3.6 hold. Then $V_t : \mathcal{X} \longrightarrow \mathbb{R}$ is a continuous function. Under the additional assumption that $\mathcal{Q} : \mathcal{X} \longrightarrow \mathrm{cl}(\mathcal{X}_R)$ is continuous, $\hat{V}_t : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$ is also a continuous function.*

The proof of the above proposition is relegated to Appendix B.2.2.

Monotonicity and concavity of $\hat{V}_t(\cdot)$

Next, we establish some shape properties of the optimal value functions $V_t(\cdot)$ and $\hat{V}_t(\cdot)$.

Assumption 3.7. **(i)** *$f_T : \mathcal{X} \to \mathbb{R}$ is monotone (see Definition B.6 of Appendix B.2.3).*

(ii) *For each $x, x' \in \mathcal{X}$ with $x \leq_{\mathcal{X}} x'$, $t \in \mathcal{T}_0$, and $a \in A_t(x)$, there exists $a' \in A_t(x')$ such that*

$$f_t(x,a) \leq f_t(x',a') \quad \text{and} \quad S(x,a,e) \leq_{\mathcal{X}} S(x',a',e), \quad \text{for} \quad e \in \mathbb{D},$$

with $\leq_{\mathcal{X}}$ denoting a certain partial order equipped by $\mathcal{X}$; see Definition B.5 of Appendix B.2.3.

The definition of the monotonicity stated in the above assumption depends on how one defines the partial order equipped by the state space $\mathcal{X}$; see Appendix B.2.3 for a discussion. The following theorem establishes the monotonicity of the value functions V_t and $\hat{V}_t$. The proof is relegated to Appendix B.2.

Proposition 3.2. *Suppose Assumptions 3.1 and 3.7 hold. Then $V_t : \mathcal{X} \longrightarrow \mathbb{R}$ is a monotone function. Under the additional assumption that $\mathcal{Q} : \mathcal{X} \longrightarrow \mathrm{cl}(\mathcal{X}_R)$ is monotone, $\hat{V}_t : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$ is also a monotone function.*

The preceding Propositions 3.1 and 3.2 disclose that the optimal value function of the auxiliary stochastic control problem (3.12) inherits the continuity (resp., monotonicity) of its counterpart of the original problem (3.3) under a mild condition on $\mathcal{Q} : \mathcal{X} \to \mathrm{cl}(\mathcal{X}_R)$. Specifically, when $d = 1$, $\mathcal{X} = [0, \infty)$, and $\mathcal{X}_R = [0, R]$ which is the common case in many finance applications, we get $\mathcal{Q}(x) = \min[x, R]$ which satisfies the condition of the preceding two propositions. The continuity (resp., monotonicity) of $\hat{V} : \mathrm{cl}(\mathcal{X}_R) \to \mathbb{R}$ lays the foundation for employing the raw (resp., shape-preserving) sieve estimation method to approximate this value function.

By imposing more conditions, the concavity of the value functions can be established.

Assumption 3.8. **(i)** *For each $e \in \mathbb{D}$, $S(x,a,e)$ is a concave function of (x,a).*

(ii) *$f_T : \mathcal{X} \to \mathbb{R}$ and $f_t : \mathcal{X} \times \mathbb{A} \to \mathbb{R}$ are concave functions for $t \in \mathcal{T}_0$. In addition, there exists a constant ζ_f such that*

$$\sup_{x \in \mathcal{X}} \left(\sup_{a \in A_t(x)} |f_t(x,a)| \right) \leq \zeta_f, \quad \text{and} \quad \sup_{x \in \mathcal{X}} |f_T(x)| \leq \zeta_f, \quad \text{for} \quad t \in \mathcal{T}_0.$$

(iii) *For each $\lambda \in (0,1)$ and any two pairs (x', a') and (x'', a'') such that $a' \in A_t(x')$ and $a'' \in A_t(x'')$, $\lambda a' + (1-\lambda)a'' \in A_t(\lambda x' + (1-\lambda)x'')$ for $t \in \mathcal{T}_0$.*

The concavity of a multivariate function is defined in Definition B.8 of Appendix B.2.3.

Proposition 3.3. *Suppose Assumptions 3.1, 3.7, and 3.8 hold. Then $V_t : \mathcal{X} \longrightarrow \mathbb{R}$ is a concave and monotone function. In addition, if $\mathcal{Q} : \mathcal{X} \longrightarrow \mathrm{cl}\,(\mathcal{X}_R)$ is a concave and monotone function, then so is $\hat{V}_t : \mathrm{cl}\,(\mathcal{X}_R) \longrightarrow \mathbb{R}$.*

The proof of the above proposition is relegated to Appendix B.2. As we will see later, the regularities of $\hat{V}_t$ established by the preceding Propositions 3.1, 3.2, and 3.3 not only warrant the legitimacy of applying the shape-preserving sieve estimation method but also bring benefits to the consequential RL-LSMC algorithm as disclosed by the Proposition 3.4 in the sequel.

3.3.3 Discussions

Before ending this section, we make some comparisons between the truncation method developed in the present chapter with some alternatives in the literature and the one given in Chapter 2.

Comparison with the results of [49]

It is worth noting that a similar truncation argument has been proposed by [49] to confine the domain of the optimal value function into a compact set. Below, we point out the difference between the results established in this section and those of [49], in particular, their Proposition A.1.

(i) Firstly, [49] establishes the relationship between the optimal value functions of the truncated problem and the original one in certain $\mathbb{L}^1(\mu_t)$ sense:

$$\int \left| V_t(x) - \hat{V}_t(x) \right| \mu_t(\mathrm{d}x) \longrightarrow 0, \quad \text{as} \quad R \longrightarrow \infty, \tag{3.18}$$

where $\mu_t(\cdot)$ is the sampling distribution of the state process at time step t ([49, pp. 15]). The limitation of such an $\mathbb{L}^1$ loss criterion is that the discrepancy between the two value functions at a certain point/region might be considerable but with a small probability measured by the distribution function $\mu_t(\cdot)$. However, it is worth stressing that $\mu_t(\cdot)$ is not the *true* distribution of the state process driven by the DM's *optimal* action say $\mu_t^\star(\cdot)$ but some distribution of the sample path commonly generated by forward simulation and control randomization; see, e.g., [51]. It might be the case that the point/region where $\mu_t(\cdot)$ places marginal probability exhibits substantial odds under the distribution $\mu_t^\star(\cdot)$. In view of this, $\mathbb{L}^1(\mu_t)$ is not a good measure to quantify the gap between the optimal value functions of the original and the truncated problems (3.3) and (3.12).

In contrast, the preceding Theorem 3.1 characterizes the difference between two value functions in a point-wise sense (see inequality (3.17)) which is of more interest in reality; for instance, in the context of pricing variable annuity, $V_0(x_0)$ corresponds to the no-arbitrage price of the policy with initial purchase payment x_0 from the policyholder; see, e.g., [40], [6], [47], and [48].

(ii) Secondly, the convergence rate of the integral term in the preceding statement (3.18) is not explicitly given by [49] as their Assumption (HLoc) is less transparent. This brings inconvenience in choosing the truncation parameter R. However, the error bound established in Theorem 3.1 is more transparent as the terms $\bar{B}(\cdot)$, $\bar{\xi}(R)$, $\mathcal{E}_{x_0,T}(R)$, and ζ_X are not hard to specify in a specific application. Theorem 3.1 also sheds light on the question that under which conditions the approximation error of $\hat{V}_t(x_0)$ would vanish: $\mathcal{E}_{x_0,T}(R)$ should decay faster than the increasing speed of $\bar{\xi}(R)$ as the parameter R climbs as shown by Figure 3.1. This gives a guideline of choosing an appropriate size of R.

Comparison with the results of Chapter 2

It is notable that Chapter 2 uses a different way to construct an auxiliary state process $\tilde{X}$. Below we elaborate the difference between $\hat{X}$ and $\tilde{X}$ and the merit of the truncation argument developed in this chapter.

Before the original state process X leaves the truncated region $\mathcal{X}_R$ for the first time, $\hat{X}$, $\tilde{X}$, and X agree with each other. Once the original state process X exits $\mathcal{X}_R$, $\tilde{X}$ gets absorbed by the boundary of $\mathcal{X}_R$ forever according to the construction in Section 2.3; see Figure 2.5 for a graphical illustration. However, one should recall from Eq. (3.10) that $\hat{X}$ does not necessarily sojourn at the boundary of $\mathcal{X}_R$.

One benefit we can get from the way of constructing $\hat{X}$ in this Chapter is that the accompanying value function $\hat{V}_t$ inherits the continuity, monotonicity and concavity of the original value function V_t; see Propositions 3.1, 3.2, and 3.3. This is crucial to the application of sieve estimation method to approximating $\hat{V}_t$ as we will see in the subsequent section. On the other hand, the value function accompanying $\tilde{X}$ does not necessarily have such nice regularities as the construction of $\tilde{X}$ might impair the continuity of the transition equation. This is not a big problem in the context of Chapter 2 because it focuses on approximating the continuation function instead and accordingly it does not require the continuity of the value function.

It is worth stressing that the price we pay for such a benefit is that more conditions need to be imposed in Theorem 3.1 and Assumption 3.5. These conditions, however, are not required in Chapter 2; see Remarks 3.2 and 3.3.

3.4 The Regression-later Monte Carlo Algorithm

In what follows, let us turn our attention to the optimal value function $\hat{V}_t : \mathrm{cl}\left(\mathcal{X}_R\right) \to \mathbb{R}$ which is a valid proxy for the optimal value function of the original problem (3.3) according to Theorem 3.1. This section aims to develop an LSMC algorithm to estimate $\hat{V}_t$ for $t \in \mathcal{T}$.

3.4.1 The Algorithm

The regression-later Least Squares Monte Carlo (RL-LSMC) algorithm is presented as follows.

1. **Initiation:**

   1.1 Simulate a sample of independent and identically distributed (i.i.d.) random variables

   $$\left\{\hat{X}_T^{(m)}, m = 1, 2, \ldots, M\right\}$$

   from a probability distribution $\mathsf{Q}(\cdot)$ with support $\mathrm{cl}\left(\mathcal{X}_R\right)$.

   1.2 Let

   $$\left(U^{(m)}, Z^{(m)}\right) = \left(f_T\left(\hat{X}_T^{(m)}\right), \hat{X}_T^{(m)}\right), \quad \text{with} \quad m = 1, 2, \ldots, M.$$

   Conduct the linear sieve estimation given in (3.6) and the resulting regression estimate is denoted by $\hat{V}_T^{\mathrm{E}}(x) := \boldsymbol{\phi}^{\intercal}(x)\check{\boldsymbol{\beta}}_T$ for $x \in \mathrm{cl}\left(\mathcal{X}_R\right)$.

   For $t = T-1, \ldots, 0$, do the step below.

2. **Backward Simulation and Backward Updating:**

   3.1 **Data preparation** Simulate a sample of i.i.d. random variables

   $$\mathbf{X}_t^M := \left\{\hat{X}_t^{(m)},\ m = 1, 2, \ldots, M\right\},$$

   from $\mathsf{Q}(\cdot)$. Further construct the random sample $\left\{\hat{V}_t^*\left(\hat{X}_t^{(m)}\right), m = 1, 2, \ldots, M\right\}$, where

   $$\hat{V}_t^*(x) = \sup_{a \in A_t(x)} \left\{f_t(x, a) + \varphi \mathbb{E}^{\varepsilon}\left[\hat{V}_{t+1}^{\mathrm{E}}\left(\hat{S}\left(x, a, \varepsilon_{t+1}\right)\right)\right]\right\}, \quad \text{for} \quad x \in \mathbf{X}_t^M, \tag{3.19}$$

   $\mathbb{E}^{\varepsilon}[\cdot]$ denotes the expectation taken over the distribution of ε_{t+1}[2], and recall that

[2]The reader should be cautious that $\hat{V}_{t+1}^{\mathrm{E}}$ is essentially a random function as it implicitly depends on the random samples simulated from time step T down to time step $t+1$; see Remark 2.1 of Chapter 2 for a detailed discussion.

$\hat{S} : \mathcal{X}_R \times \mathbb{A} \times \mathbb{D} \to \mathcal{X}_R$ is the transition equation of the auxiliary state process $\hat{X}$; see Section 3.3.1.

3.2 **Regression** Let

$$\left(U^{(m)}, Z^{(m)}\right) = \left(\hat{V}_t^* \left(\hat{X}_t^{(m)}\right), \hat{X}_t^{(m)}\right), \quad m = 1, 2, \ldots, M.$$

Employ the linear sieve estimation method to obtain a regression estimate

$$\hat{V}_t^{\mathrm{E}}(x) = \boldsymbol{\phi}^{\intercal}(x)\check{\boldsymbol{\beta}}_t, \quad \text{for} \quad x \in \mathrm{cl}\left(\mathcal{X}_R\right). \tag{3.20}$$

Remark 3.4. *By virtue of the construction of the auxiliary state process $\hat{X}$ and the accompanying transition equation $\hat{S}$, the evaluation of Eq. (3.19) only requires the information of value function estimate $\hat{V}_{t+1}^{\mathrm{E}}$ over $\mathrm{cl}\left(\mathcal{X}_R\right)$. This convenience however cannot be achieved if one directly applies an LSMC algorithm to the original state process X; see Section 2.4 of Chapter 2 for a discussion.*

Remark 3.5. *It is worth stressing the subtle difference between $\hat{V}_t^{\mathrm{E}}$ and $\hat{V}_t^*$ On one hand, $\hat{V}_t^*(x)$ is the optimal value of the optimization problem (3.19) and accordingly, its expression is not explicitly given. On the other hand, $\hat{V}_t^{\mathrm{E}}(x)$ is a sort of approximation for $\hat{V}_t^*(x)$ by a linear combination of basis functions which is explicitly given by $\boldsymbol{\phi}^{\intercal}(x)\check{\boldsymbol{\beta}}_t$. This analytical tractability reduces the difficulty of evaluating the expectation term $\mathbb{E}^{\varepsilon}\left[\hat{V}_{t+1}^{\mathrm{E}}\left(\hat{S}\left(x, a, \varepsilon_{t+1}\right)\right)\right]$ in Eq. (3.19); more discussions will follow later.*

The success of implementing the above RL-LSMC algorithm rests on (i) the efficient evaluation of the expectation term in Eq. (3.19), and (ii) finding the global extremum of the optimization problem (3.19). These two issues are discussed respectively in the sequel.

Evaluation of continuation value

In view of Eqs. (3.19) and (3.20), we get

$$\mathbb{E}^{\varepsilon}\left[\hat{V}_{t+1}^{\mathrm{E}}\left(\hat{S}\left(x, a, \varepsilon_{t+1}\right)\right)\right] = \check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\varepsilon}\left[\boldsymbol{\phi}\left(\hat{S}\left(x, a, \varepsilon_{t+1}\right)\right)\right]. \tag{3.21}$$

The R.H.S. of the above equation is easy to calculate in the presence of explicit expressions of $\boldsymbol{\phi}$, $\hat{S}$, and the distribution function of ε_{t+1}. To make the matter more concrete, let us consider a specific case where $\hat{S}(x, a, e) = \min\left[(x-a)^+ e, R\right]$, $\boldsymbol{\phi}(x) = \left(1, x, \ldots, x^J\right)^{\intercal}$, and ε_{t+1} follows a log-normal distribution. This case will be studied in the subsequent Section 3.5. Evaluating the above expectation term boils down to the calculation of the partial expectation of a log-normal distribution which is straightforward; see Appendix B.1 for details. Such a convenience is resulted by the simple expression of $\hat{V}_{t+1}^{\mathrm{E}}$ and this highlights our motivation to use $\hat{V}_{t+1}^{\mathrm{E}}$ as a proxy for $\hat{V}_{t+1}$ instead of $\hat{V}_{t+1}^*$.

$$\cdots \Longrightarrow \hat{V}_{t+1}^{\mathrm{E}} \text{ is monotone/concave} \Rightarrow \hat{V}_t^* \text{ is monotone/concave} \Rightarrow \hat{V}_t^{\mathrm{E}} \text{ is monotone/concave} \Longrightarrow \cdots$$

Figure 3.2: A diagram for the propagation of monotonicity in the RL-LSMC algorithm.

Concavity preservation and local optimization

Solving the local optimization problem involved in (3.19) is formidable in general. Common numerical optimization algorithms cannot always guarantee the convergence to a global optimizer. As a compromise, in the absence of sufficient prior knowledge of the objective function of the optimization problem, one may first discretize the feasible set of the decision variable, i.e., $A_t(x)$, and then resort to a linear search method ([6]).

For a tame dynamic programming problem, we usually require the local optimization problem in the Bellman equation (3.13) is tractable in the sense that it can be solved with a light computational cost. This is attainable if the objective function of the local optimization problem i.e., $\mathcal{V}_t(x,a) := f_t(x,a) + \hat{C}_t(x,a)$, is concave in the decision variable a. However, one cannot ignore the fact that such concavity might be impaired after replacing the continuation/value function by its numerical estimate in an LSMC algorithm. Specifically, the objective function of the optimization problem one encounters in Eq. (3.19) is

$$\mathcal{V}_t^{\mathrm{E}}(x,a) := f_t(x,a) + \varphi \mathbb{E}^{\varepsilon}\left[\hat{V}_{t+1}^{\mathrm{E}}\left(\hat{S}\big(x,a,\varepsilon_{t+1}\big)\right)\right] \tag{3.22}$$

instead. In the sequel, we will show that this dilemma can be circumvented by the shape-preserving sieve estimation method.

Proposition 3.4. **(i)** *Suppose the conditions of Proposition 3.2 are satisfied. For each $t \in \mathcal{T}_0$, if $\hat{V}_{t+1}^{\mathrm{E}} : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$ is monotone, then so is $\hat{V}_t^* : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$.*

(ii) *If the conditions of Proposition 3.3 all hold, then for each $t \in \mathcal{T}_0$, $\hat{V}_t^* : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$ is a monotone and concave function as long as $\hat{V}_{t+1}^{\mathrm{E}} : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$ is monotone and concave.*

(iii) *Under the conditions of Proposition 3.3, $\mathcal{V}_t^{\mathrm{E}}(x,a)$ defined in Eq. (3.22) is concave in a if $\hat{V}_{t+1}^{\mathrm{E}} : \mathrm{cl}(\mathcal{X}_R) \longrightarrow \mathbb{R}$ is monotone and concave.*

The above proposition can be proved in the same manner as we did for Propositions 3.2 and 3.3, and thus the proof is omitted.

The implications of this proposition are given as follows. On one hand, Part (i) (resp., Part (ii)) states that the monotonicity (resp., monotonicity and concavity) of the value function estimate obtained at time step $t+1$ warrants the monotonicity (resp., monotonicity and concavity) of the function $\hat{V}_t^*$. On the other hand, recall that $\hat{V}_t^*$ is taken as the regression function in the

preceding RL-LSMC algorithm and $\hat{V}_t^{\mathrm{E}}$ is the corresponding regression estimate produced by the sieve estimation method. Combing these two observations together, an induction argument implies that $\hat{V}_t$, $\hat{V}_t^*$, and $\hat{V}_t^{\mathrm{E}}$ are monotone (resp., monotone and concave) functions for all $t \in \mathcal{T}$ if the conditions of the preceding Proposition 3.4 hold and the regression estimate may inherit the monotonicity (resp., monotonicity and concavity); see Figure 3.2 for an illustration diagram.

Furthermore, Part (iii) of Proposition 3.3 discloses that the concavity of the local optimization problem at time step t rests on the monotonicity and concavity of the value function estimate at time step $t+1$. In view of these, employing the shape-preserving sieve estimation method brings benefits at least in two-fold.

(B1) The shape properties such as monotonicity and concavity of the optimal value function accompanying the original stochastic control problem (3.3) (i.e., V_t) can be inherited by $\hat{V}_t$ and its regression estimate $\hat{V}_t^{\mathrm{E}}$. This renders $\hat{V}_t^{\mathrm{E}}$ economically sensible.

(B2) The concavity of the optimization problem involved in the RL-LSMC algorithm is warranted, which in turn makes the evaluation of Eq. (3.19) relatively more tame than the case without concavity.

3.4.2 Convergence Analysis

This subsection devotes to establishing the convergence result associated with the preceding RL-LSMC algorithm. To this end, a technical assumption is imposed.

Assumption 3.9. *For each $x \in \mathrm{cl}\,(\mathcal{X}_R)$ and each $a \in A_t(x)$, let $F_{\hat{S}}(\cdot; x, a)$ be the distribution function of $\hat{X}_{t+1}$ given that $\hat{X}_t = x$ and the DM's action at time step t is a. Assume that $F_{\hat{S}}(\cdot; x, a)$ is absolutely continuous with respect to $\mathsf{Q}(\cdot)$ with density $w(\cdot; x, a)$; and that there exists a constant ζ_w such that*

$$\sup_{a \in A_t(x)} |w(y; x, a)| \leq \zeta_w \quad \text{for all} \quad (x, y)^\intercal \in \mathrm{cl}\,(\mathcal{X}_R) \times \mathrm{cl}\,(\mathcal{X}_R)\,.$$

The following theorem summarizes the convergence result of the RL-LSMC algorithm. The associated proof is relegated to Appendix B.2.

Theorem 3.2. *Suppose Assumptions 3.1 and 3.9 hold. In addition, assume that the conditions of Assumption 3.2 holds for $g(\cdot) = V_t(\cdot)$, $U^{(m)} = V_t\left(\hat{X}_t^{(m)}\right)$, and $Z^{(m)} = \hat{X}_t^{(m)}$ uniformly in t. Then,*

$$\left|\hat{V}_0(x_0) - \hat{V}_0^{\mathrm{E}}(x_0)\right| = O_{\mathbb{P}}\left(\zeta^{T-1}\hat{\rho}_J\right), \quad \text{as} \quad M \longrightarrow \infty, \tag{3.23}$$

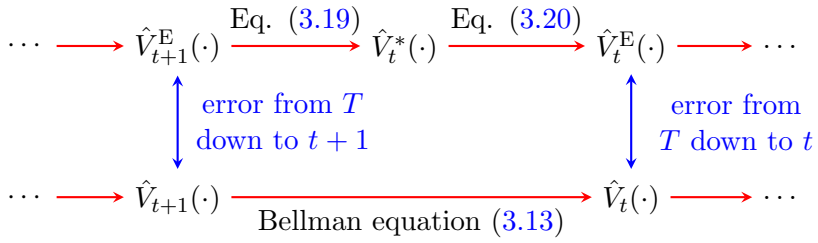


Figure 3.3: A diagram for the error propagation in the RL-LSMC algorithm.

where ζ is some constant independent of t, M, and J, $\hat{\rho}_J$ is some sequence such that $\hat{\rho}_J \to 0$ as $M \to \infty$ as given in Assumption 3.2, and the notation $O_{\mathbb{P}}(\cdot)$ is defined in Definition B.9 of Appendix B.2.

The gap between $V_0(x_0)$ and $\hat{V}_0^{\mathrm{E}}(x_0)$ is can be divided into two parts:

$$\left|V_0(x_0) - \hat{V}_0^{\mathrm{E}}(x_0)\right| \leq \underbrace{\left|V_0(x_0) - \hat{V}_0(x_0)\right|}_{\text{truncation error}} + \underbrace{\left|\hat{V}_0(x_0) - \hat{V}_0^{\mathrm{E}}(x_0)\right|}_{\text{LSMC error}}.$$

The first term on the R.H.S. of the above inequality stands for the truncation error caused by using the value function of the auxiliary problem as a proxy for the original value function. This has been characterized by the preceding Theorem 3.1. The second term corresponds to an error arising from the preceding RL-LSMC algorithm which is further contributed by two resources.

(E1) In each step of the RL-LSMC algorithm, the value function is approximated by a linear combination of basis functions. The resulting approximation error is of the order $O(\hat{\rho}_J)$, which clearly follows from Part (iii) of the preceding Assumption 3.2. Since a random sample with a finite size is used to estimate the coefficients of the basis function, extra statistical errors are incurred and this is reflected by the subscript $\mathbb{P}$ in $O_{\mathbb{P}}(\hat{\rho}_J)$.

(E2) Recall from Section 3.2.3 that the regression-later method takes $V_t(X_t)$ as the response variable and thus in principle one should generate a random sample $\left\{V_t\left(\hat{X}_t^{(m)}\right)\right\}_{m=1}^{M}$. However, one could only use $\left\{\hat{V}_t^*\left(\hat{X}_t^{(m)}\right)\right\}_{m=1}^{M}$ as a proxy which involves the estimation error accumulated from preceding time steps of the algorithm; see Step 3.1 of the RL-LSMC algorithm. In view of this, it is not surprising that the error of the algorithm is amplified as time step goes from T down to 0, which is reflected by a factor ζ^{T-1} in Eq. (3.23). Figure 3.3 depicts a diagram for the error propagation mechanism of the RL-LSMC algorithm.

3.4.3 When Does Forward Simulation Work?

This subsection discusses the choice of the sampling measure in the RL-LSMC algorithm and sheds light on the following question:

- *When is it legitimate to simulate the state process forward in time based on a control randomization technique?*

The preceding Assumptions 3.3 and 3.9 implicitly impose restrictions on the sampling distribution $\mathsf{Q}(\cdot)$. It is notable that $F_{\hat{S}}(\cdot; x, a)$ might not have a Lebesgue density over $\mathrm{cl}\,(\mathcal{X}_R)$ due to the construction of the auxiliary state process $\tilde{X}$. In particular, $\tilde{X}_t$ might exhibit nonzero probability over the truncation boundary. This observation casts shadow to using the uniform distribution with support set $\mathrm{cl}\,(\mathcal{X}_R)$ as the sampling distribution $\mathsf{Q}(\cdot)$.

Furthermore, the preceding Assumption 3.3 requires $\mathsf{Q}(\cdot)$ to have density bounded away from zero. [65, pp. 157] also points out that this requirement can be relaxed by allowing probability masses over some points, which is not hard to expect as the regression function of interest is assumed to be a continuous function.

Suppose that $F_{\hat{S}}(\cdot; x, a)$ has an absolutely continuous density function with respect to the Lebesgue measure restricted on the interior of $\mathrm{cl}\,(\mathcal{X}_R)$ and meanwhile exhibits probability masses over the boundary. In this case, it is easy to see that one may choose $\mathsf{Q}(\cdot)$ as a mixture of the uniform distribution on the interior and a degenerated distribution on the boundary such that the above two assumptions hold.

The above discussion casts insights to the legitimacy of using forward simulation to generate the regression data: if the resulting distribution of the state variable satisfies the conditions imposed by Assumptions 3.3 and 3.9, then the convergence of the RL-LSMC algorithm is ensured. However, it is hard to expect this can be achieved based on an arbitrary control randomization method especially when the state process has certain absorbing states, as we have seen from the numerical example in the previous Chapter 2.

3.4.4 Discussions

This subsection makes some comparisons between the proposed RL-LSMC algorithm and the BSBU counterpart developed in Chapter 2. The RL-LSMC exhibits the following advantages.

(A1) By comparing Eq. (3.23) with the convergence result delivered in Theorem 2.2 of Chapter 2, we notice that one additional term $\sqrt{J/M}$ appears in the converge rate in Theorem 2.2. This additional term arises because in Chapter 2 the sieve estimation method is used to estimate the continuation function instead of the value function. This shows the first

advantage of the RL-LSMC algorithm over the regression-now counterparts in terms of reducing regression error. The observation that the error of a regression-later Monte Carlo algorithm is solely contributed by the term $\hat{\rho}_J$ has also been revealed by [19, Theorem 3.1] in the context of one single period regression problem.

Remark 3.6. *It is worth stressing that the decay rate of $\hat{\rho}_J$ depends on the smoothness of the regression function (i.e., value function in the present context); see, e.g., [65]. Therefore the expression of $\hat{\rho}_J$ can be substantially different from that of ρ_J in Theorem 2.2. In view of this, the RL-LSMC has more appealing convergence rate than the BSBU counterpart if the value function has the same order of differentiability as the continuation function. Otherwise, it is unfair to compare the convergence speeds of the RL-LSMC and BSBU peers due to the difference between $\hat{\rho}_J$ and ρ_J.*

(A2) Rather than simulating the post-action value of the state process as in Chapter 2, the RL-LSMC algorithm proposed in the present chapter samples $\hat{X}_t^{(m)}$ over its feasible set $\mathrm{cl}\,(\mathcal{X}_R)$. Since $\mathrm{cl}\,(\mathcal{X}_R)$ is independent of t, one may choose one common sampling distribution Q for all time steps. However, the feasible set of the post-action value is generally time-dependent and accordingly, the sampling distribution in the BSBU algorithm is time-dependent; see Section 2.4 of Chapter 2.

(A3) As previously commented in Remark 3.5, the estimate of the value function $\hat{V}_t^{\mathrm{E}}$ in the RL-LSMC algorithm is a simple linear combination of the basis functions. However, the parallel value function estimate is obtained as the optimal value of an optimization problem in regression-later algorithms because in such algorithms the continuation function is instead approximated by basis functions. An explicit expression of $\hat{V}_t^{\mathrm{E}}$ brings much convenience to calculating its derivatives which is useful for finance applications. We will demonstrate such convenience with a specific application in Section 3.5.

On the other hand, the merits of the RL-LSMC mentioned above come at a price.

(P1) As stressed previously, the success of the RL-LSMC is placed on the ground that the expectation term in Eq. (3.21) can be evaluated in a facile way. This poses restrictions on the function $\hat{S}$ and the distribution of the disturbance term ε_{t+1}.

(P2) As commented in the item "**(Q2)**" in Section 3.2.3, the legitimacy of applying the RL-LSMC relies on the continuity of the value function which however is not required in the regression-now algorithms. This continuity requirement narrows the application range of the RL-LSMC.

To sum up, compared with regression-now LSMC algorithms, the benefits of the RL-LSMC algorithm come at the expense of extra restrictions on the stochastic control problem at hand.

3.5 Application: Delta-hedging of Variable Annuities

This section applies the RL-LSMC algorithm to the dynamic hedging of variable annuities.

3.5.1 Model Setup

Variable annuities (VAs) are equity-linked products issued by insurance companies which provide policyholders (PHs) protection against the downside risk of the equity market. Pricing variable annuities leads to a discrete-time stochastic optimal control problem; see [40], [6], [47], and [48], among others. For the convenience of illustration, this section considers a simplified VA contract.

In the VA policy, the PH's initial payment, say x_0, is invested into a certain risky asset (referred to as policy fund) and the PH is allowed to take periodical withdrawals from the investment account. The collection of withdrawal dates is labeled by $\mathcal{T}$ and the length between two consecutive withdrawal dates is assumed to be δ. Furthermore, X_t corresponds to the value of the PH's investment account right before the withdrawal at withdrawal date t. It evolves recursively as follows:

$$\begin{cases} X_0 & = x_0, \\ X_{t+1} & = (X_t - a_t)^+ \frac{S_{(t+1)\delta}}{S_{t\delta}}, \quad a_t \in \big[0, G \vee X_t\big], \quad t \in \mathcal{T}, \end{cases} \tag{3.24}$$

where S_u is the time-u price of the policy fund, and a_t corresponds to the PH's withdrawal amount. It is worth noting that the PH can withdraw up to the amount of G even if the investment account value falls below this level. This is the guarantee provided by the insurance company for a potential market decline.

Assume the price process of the policy fund follows a geometric Brownian motion (GBM) under a martingale pricing measure $\mathbb{Q}$, i.e.,

$$S_{(t+1)\delta} = S_{t\delta} \exp\Big(\big(r - q - \sigma^2/2\big)\delta + \sigma\sqrt{\delta} Z_{t+1}\Big) =: S_{t\delta} \cdot \varepsilon_{t+1}, \quad t \in \mathcal{T}, \tag{3.25}$$

where r, q, σ, and $\{Z_{t+1}\}_{t\in\mathcal{T}_0}$ denote the risk-free rate, insurance fee rate, volatility rate, and a sequence of independent standard normal random variables, respectively. $\log \varepsilon_{t+1}$ can be interpreted as the log-return of the underlying asset over $(t\delta, (t+1)\delta]$.

In view of Eqs. (3.24) and (3.25), one may write down the transition equation for the state process as follows:

$$X_{t+1} = S(X_t, a_t, \varepsilon_{t+1}) = \big(X_t - a_t\big)^+ \varepsilon_{t+1}, \quad t \in \mathcal{T}. \tag{3.26}$$

The reward functions in the present context correspond to the policy payoffs which are given

by

$$\begin{cases} f_T(X_T) & = X_T, \\ f_t(X_t, a_t) & = a_t - \kappa_t(a_t - G)^+, \quad t \in \mathcal{T}_0 = \mathcal{T}\backslash\{0\}. \end{cases} \tag{3.27}$$

In other words, the remaining value of the investment account is returned to the PH at maturity. Before the maturity, the payoff equals the PH's withdrawal amount subject to a certain proportional penalty κ_t.

Now the stochastic control problem (3.3) can be interpreted such that the PH strives to maximize the financial value of her cash inflows. $V_0(x_0)$ corresponds to the no-arbitrage price of the VA policy.

3.5.2 The RL-LSMC Algorithm

In the present context, $\mathcal{X} = [0, \infty)$ as the investment account always has nonnegative value. Let $\mathcal{X}_R = [0, R]$ and accordingly, $\mathcal{Q}(x) = \min[x, R]$. In the present case, the transition equation of the auxiliary state process (3.10) becomes

$$\hat{X}_{t+1} = \hat{S}\left(\hat{X}_t, a_t, \varepsilon_{t+1}\right) = \min\left[\left(X_t - a_t\right)^+ \varepsilon_{t+1}, R\right], \quad t \in \mathcal{T}_0, \tag{3.28}$$

given $\hat{X}_0 = x_0$.

In Appendix B.1, the preceding technical assumptions are verified in the context of the VA policy considered here, which warrants the legitimacy of applying the RL-LSMC algorithm. Similarly to [6], [47], and [72], one may show that the optimal decision at each time step is attained in the subset $A_t^\star(x) = \{0, G, x\}$ of $A_t(x)$. So, we can restrict our attention to the lattice $A_t^\star(x)$ in solving the optimization problem involved in Eq. (3.19).

It remains to address how to evaluate Eq. (3.19). Appendix B.1 presents explicit expressions of expectation term (3.21) involved in Eq. (3.19) under two particular types of basis functions, power function and Bernstein polynomial, respectively. Now it is ready to implement the RL-LSMC algorithm of Section 3.4 to get the estimate for the no-arbitrage price of the policy. In view of Theorems 3.1 and 3.2, we henceforth turn our attention to $\hat{V}_t^{\mathrm{E}}$ which is a good approximation for the value of the hedging portfolio.

3.5.3 The Delta-hedging

We consider the dynamic hedging of the VA policy which is an indispensable risk management tool for the insurance company. It is worth stressing that the policy fund is typically a basket

of mutual funds and there generally does not exist a vanilla options market over such an asset. Due to such a reason, the delta-hedging is the common practice of the insurance companies in offsetting their risk exposures which means the underlying asset is the sole hedging instrument. This amounts to calculating the first-order sensitivity of the replication portfolio[3] with respect to the policy fund price.

It is notable that the hedging frequency might be more intensive than the withdrawal frequency. For instance, suppose $\delta = 1/12$, that is, the PH takes monthly withdrawals, however, the hedger may dynamically rebalance the hedging portfolio on a daily basis. Thus, it is necessary to define the value of the replication portfolio at a particular time point between two consecutive withdrawal dates.

Suppose $(X_t - a_t)/S_{t\delta} = x$ which is observable to the hedger at time $u \in \big(t\delta, (t+1)\delta\big]$. Define the time-$u$ value of the hedging portfolio as:

$$P_{t,x}(u,y) := \varphi^{(t+1)\delta-u}\mathbb{E}^{\mathbb{Q}}\left[\hat{V}^{\mathrm{E}}_{t+1}\left(\hat{X}_{t+1}\right)\middle| S_u = y,\ \frac{X_t - a_t}{S_{t\delta}} = x\right],\quad u \in (t, t+1],\ t \in \mathcal{T}_0,$$

where φ is the discounting factor and $\mathbb{E}^{\mathbb{Q}}[\cdot]$ is taken under the martingale pricing measure $\mathbb{Q}$.

Remark 3.7. *The rationale behind the above definition is that the policy fund S_u is traded by the insurance company to replicate the payoff $\hat{V}^{\mathrm{E}}_{t+1}\left(\hat{X}_{t+1}\right)$. The ratio $(X_t - a_t)/S_{t\delta}$ corresponds to the number of shares of policy fund held by the PH immediately after her t-th withdrawal. In view of this, $P_{t,x}(u,y)$ is thought of as the time-t value of the replication portfolio given that time-to-expiry is $(t+1)\delta - u$, the PH holds x shares of policy fund, and the policy fund's time-u price is y.*

It follows from Eqs. (3.24) and (3.28) that

$$P_{t,x}(u,y) = \varphi^{(t+1)\delta-u}\mathbb{E}^{\mathbb{Q}}\left[\hat{V}^{\mathrm{E}}_{t+1}\Big(\min\left[xS_{(t+1)\delta}, R\right]\Big)\middle| S_u = y\right].$$

Our goal is to calculate the partial derivative $\partial P_{t,x}(u,y)/\partial y$ which corresponds to the number of shares of policy fund held in the hedging portfolio at time u. Recall from Eq. (3.20) that $\hat{V}^{\mathrm{E}}_{t+1}$ is a linear combination of basis functions that are explicitly given. This observation implies that the evaluation of $\partial P_{t,x}(u,y)/\partial y$ is essentially the same as calculating the delta of a European option. This is generally easy to handle. Under the GBM assumption (3.25), one may even get

[3] In this chapter, the hedging/replication portfolio corresponds to the portfolio that can replicate the cash inflows of the PH and the associated delta should be interpreted accordingly. As a matter of fact, the insurer's payout should be the part of the PH's cash inflow in excess of the balance of the investment account because the investment account value is fully covered by the mutual funds. Therefore, the number of shares of the hedging instrument held by the insurer should be adjusted from the delta calculated in this section; see, e.g., [74, Section 6.5].

an analytical expression for $\partial P_{t,x}(u,y)/\partial y$ whose derivation is relegated to Appendix B.1. For more general dynamics, one may resort to the path-wise method to calculate the delta; see [21] for instance.

To sum up, the replication of variable annuities boils down to replicating a sequence of European options and therefore demands to calculate the associated deltas. This is relatively easy to tackle since the RL-LSMC algorithm makes the $\hat{V}_{t+1}^{\mathrm{E}}$ exhibit a simple expression as given in Eq. (3.20).

3.5.4 Numerical Experiments

This subsection conducts several numerical experiments to show the merits of the proposed RL-LSMC algorithm.

Consider a contract with maturity $T\delta = 1$ (year) which allows the PH to take monthly withdrawals, i.e., $\delta = 1/12$. Appendix B.1 derives an explicit expression for the error bound in the preceding Theorem 3.1. This allows us to estimate the truncation error for any chosen truncation parameter R. In the subsequent numerical experiments, the truncation parameter is chosen as $R = 4$, which causes a truncation error less than 10^{-7}; see the dashed line depicted in Figure 3.1. All parameters are summarized in Table 3.1.

In the subsequent experiments, two sieve estimation methods are employed.

- **(RSE)** The first method chooses basis function $\boldsymbol{\phi}(x) = \big(1, x, \ldots, x^J\big)^\intercal$ and set $\mathbf{A}_J$ as the null matrix. In other words, this sieve estimation method uses a linear combination of power series to approximate the optimal value function. Henceforth it is referred to as the raw sieve estimation (RSE) method.

- **(SPSE)** The second method is the shape-preserving sieve estimation (SPSE) method of [83] which chooses $\boldsymbol{\phi}(x)$ as a vector of Bernstein polynomials. It is easy to verify that the conditions of the preceding Proposition 3.2 are satisfied in the stochastic control problem accompanying the VA contract considered here. In view of this, the optimal value function is a monotone function and the constraint $\mathbf{A}_J$ in Eq. (3.7) is chosen such that this monotonicity is inherited by the sieve estimate. The expressions of $\boldsymbol{\phi}(x)$ and $\mathbf{A}_J$ are relegated to Appendix A.1.1 for the clarity of presentation.

We first conduct a convergence test for the RL-LSMC algorithm. Table 3.2 collects the estimate for the no-arbitrage price of the VA contract, $\hat{V}_0^{\mathrm{E}}(x_0)$, under different settings of number of simulated paths M and maximal degree of the basis functions J. Recall that $\hat{V}_0^{\mathrm{E}}(x_0)$ is random due to the randomness of the sample generated throughout the algorithm. In view of this, for each pair (M, J), the algorithm is repeated 30 times and the associated sample mean and standard

Table 3.1: Parameters used for numerical experiments.

Parameter	Value
Volatility rate σ	0.15
Risk-free rate r	0.03
Insurance fee rate q	0.01
Number of time steps T	12
Length of time interval δ	1/12
Discounting factor $\varphi = e^{-r\delta}$	0.9975
Initial purchase payment x_0	1
Withdrawal penalty κ_t	0.8
Guaranteed amount G	0.05
Truncation parameter R	4

deviation as reported in the "Mean" and "S.d." columns, respectively. From Settings 0 and 1 of Table 3.2, we observe that the sample mean of the price estimate changes little as M increases from 10^3 to 2×10^3, whereas, the standard deviation is nearly halved. This observation is more perceivable from the empirical density plots of the price estimate delineated in Figure 3.4: a larger M brings more spiked empirical density of $\hat{V}_0^{\mathrm{E}}(x_0)$. One may conclude that simulating two thousand sample points in the RL-LSMC already produces a stable numerical result with sample standard deviation around 5×10^{-4}. This is in contrast to the case of regression-now algorithms where the number of simulated paths should be generally larger than hundred thousand in order to ensure the stability of the numerical estimate; see the numerical experiments of Chapter 2 for comparison. This discrepancy is not surprising if we note that the error rate of the RL-LSMC is smaller than that of the regression-now counterparts as disclosed by a comparison between Theorems 2.2 and 3.2. In the numerical settings 0, 2, and 3, the sample size is fixed as $M = 1 \times 10^3$ and the parameter J increases from 15 to 17. One can observe that as J hikes, the sample mean of the price estimate changes accordingly but the trend is very creeping. This signals the convergence of the RL-LSMC algorithm.

Next, we investigate the performance of the SPSE method in terms of preserving the monotonicity of the value function. We fix $J = 15$ and $M = 2 \times 10^3$. Figure 3.5 delineates the estimates of the value function, $\hat{V}_t^{\mathrm{E}}$, produced by the RSE and SPSE methods, respectively. To highlight the subtle difference between the estimates resulted by the two regression methods, the plots are restricted over the interval $[0, 1]$. Observe that the RSE method does not necessarily preserve the

Table 3.2: Mean and standard deviation (S.d.) of $\hat{V}_0^{\mathrm{E}}(x_0)$ produced by SPSE and RSE methods. The results are obtained by repeating the RL-LSMC algorithm 30 times.

Setting	(M, J)	SPSE		RSE	
		Mean	S.d.	Mean	S.d.
0	$(1 \times 10^3, 15)$	0.9839	0.0009	0.9950	0.0012
1	$(2 \times 10^3, 15)$	0.9840	0.0005	0.9952	0.0005
2	$(1 \times 10^3, 16)$	0.9838	0.0008	0.9963	0.0009
3	$(1 \times 10^3, 17)$	0.9848	0.0006	0.9965	0.0007

monotonicity of the value function and the lost of monotonicity is propagated from time step 10 down to 0; see solid lines depicted in Figure 3.5. Specifically, the monotonicity is clearly broken by the RSE method at time steps 10, 4, and 2. The violation of monotonicity is more perceivable from Figure 3.6 which depicts a local plot of $\hat{V}_4^{\mathrm{E}}(x)$ over the sub-interval $[0, 0.6]$. On the contrary, the SPSE method warrants the monotonicity of the regression estimate throughout all time steps as reflected by dashed lines in Figure 3.5. These observations are consistent with the preceding Proposition 3.4 and show the major merit of the SPSE method. Also, note that the difference between the estimates produced by these two regression methods is small over the whole domain of the value function $\hat{V}_t(x)$.

Finally, we study the delta of the hedging portfolio accompanying the VA policy. Recall that hedging the VA policy is essentially replicating a sequence of European payoffs sequentially. For the convenience of comparing the deltas of the hedging portfolio over different time period, define the quantity

$$\Delta_{t,x}(\tau, y) := \frac{\partial P_{t,x}(u, y)}{\partial y}\bigg|_{u=(t+\tau)\delta}, \quad \text{for} \quad \tau \in (0, 1),$$

with $\frac{\partial P_{t,x}(u,y)}{\partial y}$ given by Eq. (B.12) or (B.13) depending on the choice of the basis function $\boldsymbol{\phi}$. In view of the previous Remark 3.7, this quantity corresponds to the number of shares of the policy fund held in the hedging portfolio at time $(t+\tau)\delta$ given that the PH holds x shares of the policy fund over the time period $\big(t\delta, (t+1)\delta\big]$. For a fixed $\tau = 0.5$, $\Delta_{t,1}(\tau, y)$ as a function of policy fund price y is ploted in Figure 3.7. Two major observations can be made from Figure 3.7. First, the disagreement between the deltas produced by the RSE and SPSE methods exaggerates as fund price y declines. The delta accompanying the RSE method exhibits fluctuations around small fund price as reflected by the solid lines at time steps 6, 4, 2, and 0, whereas, the delta produced by the SPSE method has a smooth shape, see the dashed lines depicted in Figure 3.7. Second, the

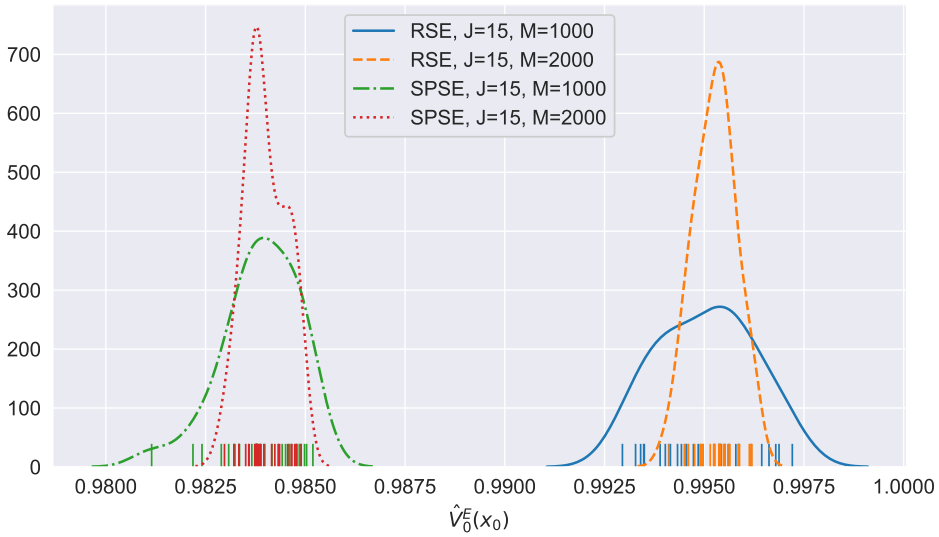


Figure 3.4: Density and rug plots of price estimate $\hat{V}_0^{\mathrm{E}}(x_0)$ obtained by repeating the RL-LSMC algorithm 30 times.

delta of the hedging portfolio is bounded above from 1. This is consistent with the observation of [47, pp. 922] in the context of the Guaranteed Lifelong Withdrawal Benefit, a particular type of VA policy; see their Figure 6.

The above two observations are interpreted as follows. On one hand, one recalls from Remark 3.7 that $\Delta_{t,1}(\tau, y)$ corresponds to the delta of the hedging portfolio that aims to replicate a European payoff $\hat{V}_{t+1}^{\mathrm{E}}\left(\min\left[S_{(t+1)\delta}, R\right]\right)$. On the other hand, Figure 3.5 shows that $\hat{V}_{t+1}$ has a shape similar to the payoff function of a call option plus a risk-free bond. Combing these together, it is not surprising that $\Delta_{t,1}(\tau, y)$ is bounded between 0 and 1. Moreover, since the value function estimate accompanying the RSE method exhibits perturbation around the origin $x = 0$, it is not hard to expect that the associated delta of the hedging portfolio vibrates at small fund price y.

Figure 3.8 plots the surface plot of the bivariate function $(\tau, y) \mapsto \Delta_{t,1}(\tau, y)$ which shows that the function is insensitive to τ. This means the delta of the hedging portfolio changes little as time progresses, which might be attributed to the small magnitude of the time interval δ.

To sum up, the RL-LSMC algorithm produces a stable numerical estimate for the optimal value function at a much smaller sample size of 2000 compared with the regression-now counterparts. This is in line with the result of Theorem 3.2 and the subsequent discussion in item "**(A1)**". Moreover, the SPSE method produces more economically sensible value function esti-

mate (resp., the delta of the hedging portfolio) when the investment account value x (resp., fund price y) is small. The numerical estimates accompanying these two regression methods agree with each other fairly well over the whole domain.

3.6 Conclusion

This chapter developed a Monte Carlo algorithm to solve stochastic optimal control problems. This algorithm approximates the optimal value function by a nonparametric sieve estimation method. The algorithm is referred to as the Regression-later Least Squares Monte Carlo (RL-LSMC) algorithm. An auxiliary stochastic control problem was constructed which confines the feasible set of the associated state process into a compact set. By virtue of the elaborate construction, the optimal value function of the auxiliary problem is shown to inherit the regularities such as continuity, monotonicity, and concavity of the value function accompanying the original problem. This paves the way for the application of the nonparametric sieve estimation theory to establishing the convergence result of the RL-LSMC algorithm. We then disclosed that the error rate of the proposed algorithm is better than those Monte Carlo algorithms that use nonparametric regression to estimate the continuation function. To further show the merits of the RL-LSMC, we studied a concrete application in hedging equity-linked insurance products. The consequential numerical experiments confirm the advantages of the RL-LSMC algorithm.

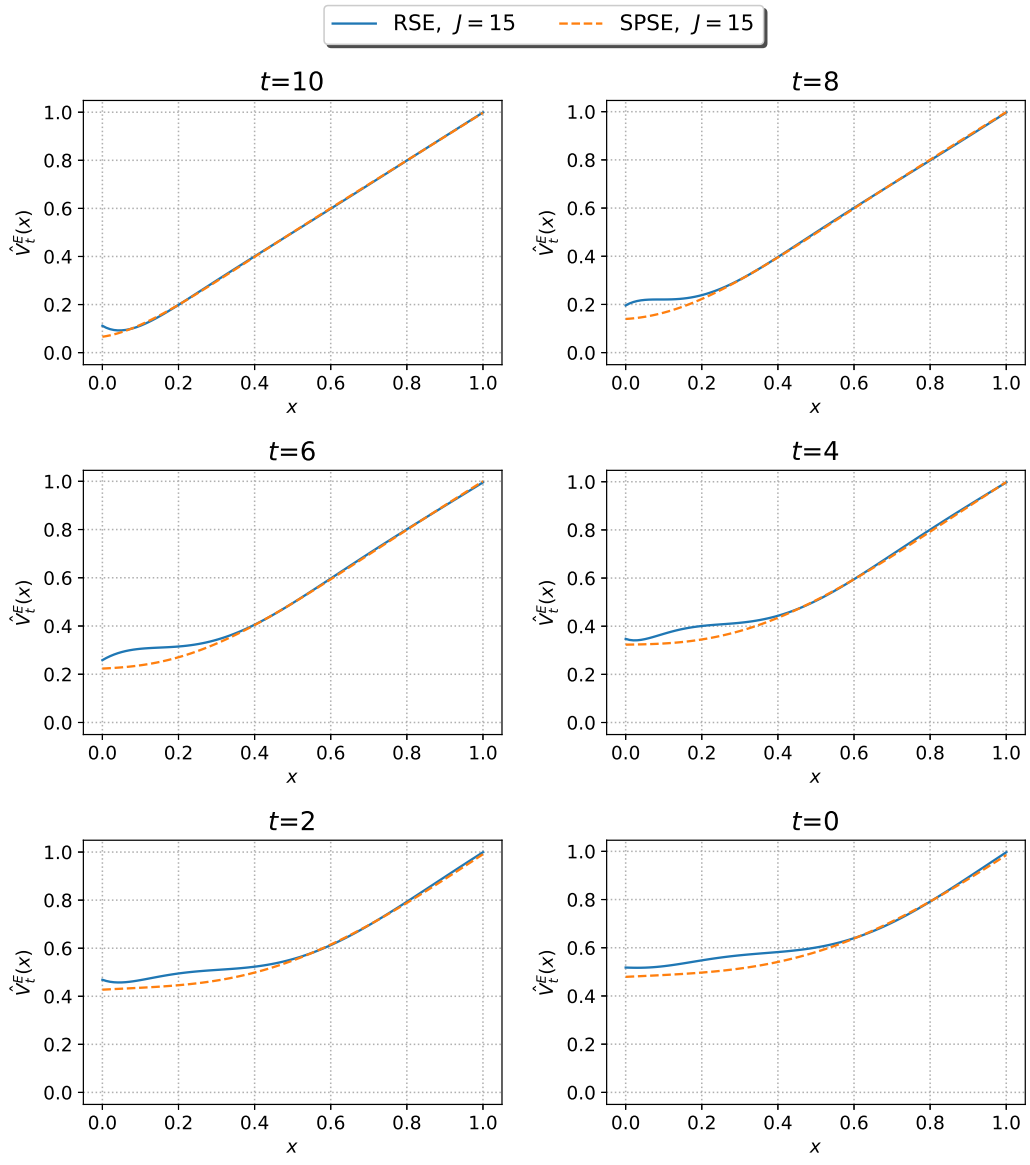


Figure 3.5: Plots of $\hat{V}_t^{\mathrm{E}}(x)$ produced by the RSE and SPSE methods. $J = 15$ and $M = 2000$.

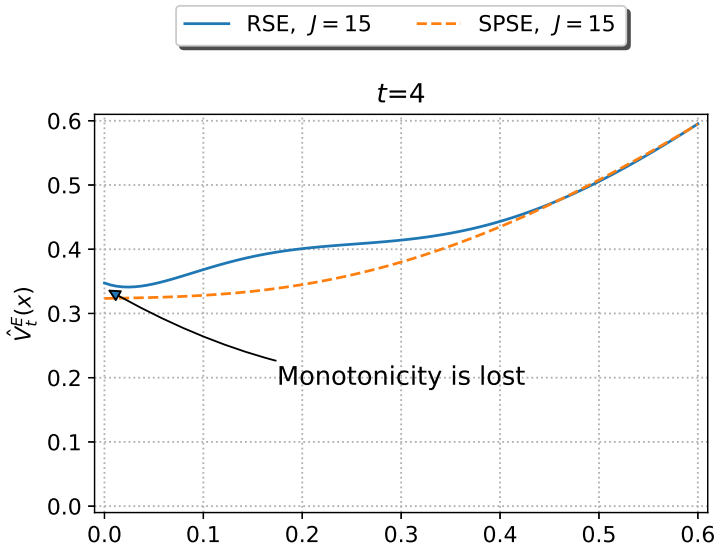


Figure 3.6: Loss of monotonicity of $\hat{V}_t^{\mathrm{E}}(x)$ at time step $t = 4$. $J = 15$ and $M = 2000$.

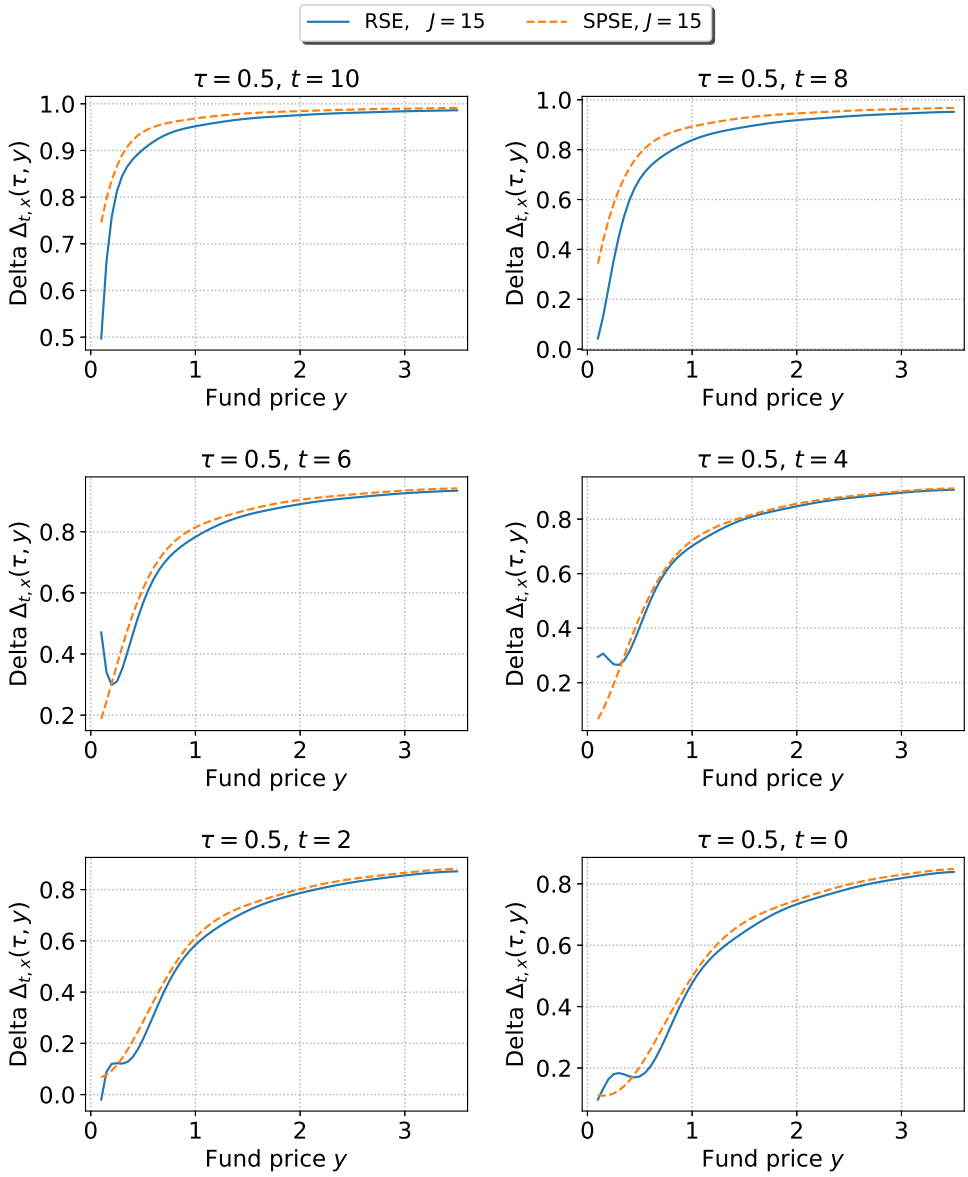
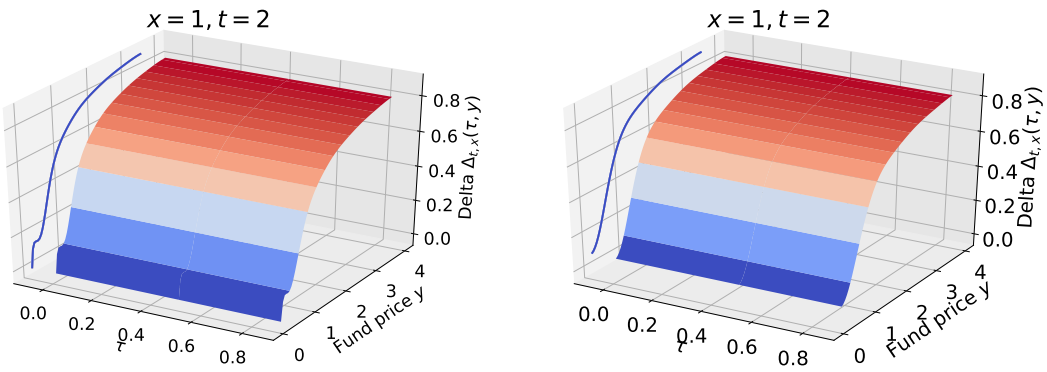


Figure 3.7: Plots of $y \mapsto \Delta_{t,1}(0.5, y)$ produced by the RSE and SPSE methods. $J = 15$ and $M = 2000$.



(a) Delta produced by RSE method.

(b) Delta produced by SPSE method.

Figure 3.8: Surface plot of $(\tau, y) \mapsto \Delta_{t,x}(\tau, y)$ with $x = 1$ and $t = 2$.

Chapter 4

Pricing Bounds and Bang-bang Analysis of the Polaris Variable Annuities

4.1 Introduction

The numerical experiments of Chapters 2 and 3 are conducted in the context of pricing and hedging equity-linked insurance products. The products considered in those chapters are simplified versions of the variable annuity which is a prevailing equity-linked product in North America. This chapter studies a more complicated product whose pricing problem is challenging from both theoretical and numerical sides.

The pace of innovation in the market of variable annuities (VAs) has been fairly remarkable in the past two decades. By the second quarter of 2017, the net amount of U.S. assets invested in VAs has been approximately 1.98 trillion dollars; see "Second-Quarter 2017 Annuity Sales Report" issued by Insured Retirement Institute ([50]). Roughly speaking, VAs are long-term equity-linked insurance products embedded with various guarantees which are also known as *riders*. These riders provide policyholders with the flexibility of dynamic withdrawals, additional purchases, longevity protection, and guaranteed minimum income payments even after the investment account is depleted which all contribute to the popularity of VAs. Therefore, it comes as no surprise that pricing and hedging various kinds of VA riders have raised great interest from both academics and practitioners.

This chapter studies the pricing problem of a new rider –"Polaris Income Plus Daily" income benefit– structured in the "Polaris Choice IV" VAs recently issued by the American International

Group.[1] Similar to most popular VA riders, the writer of the Polaris initially receives an up-front payment from the policyholder which is then invested into a basket of mutual funds. The policyholder is allowed to take dynamic withdrawals during the contract life and enjoys certain downside protection against the adverse performance of the underlying fund. As an important sweetener to policyholders, the income base in the Polaris can step up to a certain level (referred to as *step-up value*), periodically. The step-up value is prescribed as the high water mark (i.e., running maximum) of the investment account since the inception of the contract if the policyholder has not taken any withdrawal; otherwise, it only locks in the running maximum between two consecutive withdrawal dates. In other words, the income base steps up to the running maximum of the policy fund over a certain monitoring period while the length of the monitoring period depends on the withdrawal behaviors of the policyholder. In addition, both the *Protected Income Payment* applied after the underlying fund is depleted and the *Maximal Annual Withdrawal Amount* that is free of withdrawal charge are dependent on the policyholder's age at the first withdrawal. These complex behavior-dependent payoff features distinguish the Polaris from the other withdrawal benefits in VAs, such as Guaranteed Minimum Withdrawal Benefits (GMWB) and Guaranteed Lifelong Withdrawal Benefits (GLWB), which have been extensively studied in the literature.

Since the seminal work of [62] and [33], the no-arbitrage approach has been the dominating paradigm for pricing dynamic withdrawal benefits in VAs akin to pricing financial derivatives; see, e.g., [14], [28], [47], [48], [74], and the references therein. This prevailingness is attributed to the fact that the major risk undertaken by the insurer in VAs is the financial risk which cannot be diversified in the way like traditional insurance products by pooling but might be hedged (or at least mitigated) by trading available securities in the financial market; see, e.g., [45]. In recent years, there has also been an upsurge of studies in building lifetime utility models to value the withdrawal benefits from the perspective of policyholders, see, e.g., [41, 42], [77], and [46], among others. For an overview of this strand of research, see [12].

This chapter abides the no-arbitrage pricing framework to study the valuation problem of the Polaris because this paradigm guides the insurer's hedging strategy and is robust to the policyholder's actual withdrawal behavior; see [12]. Under such a framework, the pricing problem poses many challenges from both theoretical and numerical sides. In particular, it leads to a stochastic optimal control problem due to the uncertainty of the withdrawal behaviors of policyholders. By resorting to the Dynamic Programming Principle, the price function is characterized as a solution to a backward recursion equation system, often known as the Bellman equation. The analytical intractability of the solution to the Bellman equation renders the valuation problem fairly cumbersome unless some unrealistic simplifications are conducted, e.g., assuming static withdrawal behavior and discarding the stochastic control formulation. As a remedy, the existence of a bang-bang solution might be established for a stochastic control problem which means the optimal

[1] Specifically, the Polaris VAs are sold by American General Life Insurance Company beyond New York City. In New York, they are sold by The United States Life Insurance Company; see the cover page of the prospectus [1].

decision at each time step is confined to a finite number of choices. In such a case, one can drastically reduce the searching space for the optimization problem in the Bellman equation. The study on the existence of the bang-bang solution of the stochastic optimal control models accompanying VAs was pioneered by [6] which shows that the optimal withdrawal strategy[2] is among three choices: zero withdrawal, withdrawal at contractual withdrawal amount, and complete surrender. A similar result is established by [47] for the GLWB contract when the underlying risky asset follows 3/2 stochastic volatility model. Recently, the newest variant of GLWB which exhibits different contract features in income and accumulation phases has also been shown to exhibit a bang-bang solution, see [48]. In spite of those results for the GLWB, the bang-bang solution does not exist for general VA products and its analysis should be carried on a case by case base according to the specific contract provisions and model assumptions. Indeed, [6] also shows that the pricing model accompanying the GMWB does not exhibit a bang-bang solution. This chapter will show that the stochastic control model associated with the Polaris also exhibits a bang-bang solution upon a modification of the fee scheme. This modified contract is referred to as the *synthetic contract* which serves an important benchmark for the real contract. In particular, we will show that the no-arbitrage price of this synthetic contract is an upper bound for that of the real contract.

In terms of numerical implementation of the pricing models associated with VAs, the key challenge lies in the efficient calculation of conditional expectation involved in the Bellman equation and searching for the optimal withdrawal strategy at each time step. There are three prevalent numerical methods for solving stochastic optimal control problems arising from pricing VAs: finite difference method (lattice tree method), numerical integration approach, and Least Squares Monte Carlo (LSMC) method. The finite difference method rests on a certain partial differential equation (PDE) characterization of the conditional expectation. Since the withdrawal strategy changes the initial condition of the PDE at each time step, one should solve a sequence of PDEs under discrete-time withdrawal models; see, e.g., [33], [7], [40], [6], among others. The numerical integration method lays its ground on a certain integral representation of the conditional expectation. Specifically, the Gaussian-Quadrature approach and Fourier transform approach can be employed as long as the transition density and the characteristic function of the pricing process are known at first hand, respectively; see, e.g., [59], [74], [48], [4], and the references therein.

As a Monte-Carlo-simulation-based method, the LSMC brings large flexibility in choosing price dynamics and has gained enduring popularity since its emergence. As highlighted in the previous chapters, an LSMC typically requires a forward simulation of the state variable. In the context of pricing variable annuities, [47] proposes to simulate the policyholder's investment account value based on the strategy of withdrawing at a constant rate. As we have seen from the

[2]Throughout the chapter, the "optimality" of the withdrawal strategy should be understood as the strategy that maximizes the insurer's liability. Despite that the policyholder might not behave in accordance with such a strategy, it is instructive to study the pricing problem under such a stochastic optimal control framework, which in turn guides the insurer's hedging strategy; see [40] and [7] for detailed discussions.

numerical experiments of Chapter 2, such a forward simulation method might render the LSMC algorithm miss the optimal solution when the contract payoff depends on the first withdrawal time of the policyholder. This is particularly the case of the Polaris; see also the subsequent Remark 4.1. To circumvent this annoying issue, in this chapter we will implement the BSBU algorithm developed in Chapter 2 with a slight modification to get a numerical solution to the pricing problem of the Polaris.

The contributions of this chapter are in three-fold. Firstly, a comprehensive no-arbitrage pricing framework is established for the "Polaris Income Plus Daily". Despite the complex behavior-dependent features mentioned previously, we manage to formulate the pricing problem under a Markovian stochastic optimal control framework by prudently introducing certain auxiliary state and decision variables. By virtue of the nice Markov property, the no-arbitrage price of the Polaris can be obtained by solving the Bellman equation in a backward recursion manner. Secondly, the existence of a bang-bang solution is established for the accompanying stochastic control model after a modification on the fee structure of the contract. Specifically, this chapter will prove that the optimal withdrawal strategy at each withdrawal date is among three explicit choices for this synthetic contract. Thirdly, the no-arbitrage price of the synthetic contract is further shown to dominate that of the real contract. The argument used to establish this result also sheds lights on how to construct an upper bound for the optimal value function of a discrete-time stochastic control problem.

This chapter proceeds as follows. Certain notations and abbreviations are provided in the rest of this section. Section 4.2 provides a brief description of the contract and the model setup. Section 4.3 gives the mathematical formulation of the pricing model, some theoretical results related to the existence of bang-bang solution for the synthetic contract, and a study on the pricing bounds of the Polaris. Section 4.4 addresses the numerical approach. Section 4.5 provides numerical studies, and Section 4.6 concludes the chapter. Proofs and some technical results are relegated to Appendix C.

Notations and Abbreviations.

$\mathcal{I}_0$	$\mathcal{I}_0 := \{0, 1, \ldots, N-1\}$
$\mathcal{I}$	$\mathcal{I} := \{1, 2, \ldots, N-1\}$
P_0	upfront purchase payment of the policyholder
$W(t)$	time t value of investment account
$A(t)$	time t value of income base
$Z(t)$	time t value of step-up value
$B(t)$	time t value of adjusted purchase payment
I_n	an auxiliary state variable indicating first withdrawal time
X_n	state process $X_n = \big(W(t_n), Z(t_n), A(t_n), B(t_n), I_n\big)$ accompanying the synthetic contract
x	$x = (W, Z, A, B, I)$, a realized value of X_n
η	the annual insurance fee rate

MAWA	maximal annual withdrawal amount
PIP	protected income payment
$G(\xi)$	MAWA/PIP percentage given the withdrawal is initiated at t_ξ
$\xi_n(I)$	timing of the initiation of the withdrawal given $I_n = I$
$\tilde{G}_n(I)$	$\tilde{G}_n(I) = G(\xi_n(I))$, MAWA/PIP percentage given $I_n = I$
γ_n	policyholder's withdrawal amount at time t_n
τ_n	auxiliary decision variable indicating whether to take the first withdrawal at time t_n
π_n	$\pi_n := (\gamma_n, \tau_n)$
X_{n^+}	$X_{n^+} := \big(W(t_n^+), Z(t_n^+), A(t_n^+), B(t_n^+), I_{n+1}\big)$
$K_n(\cdot,\cdot)$	transition function of the state process of the synthetic contract across t_n
$\bar{K}_n(\cdot,\cdot)$	transition function of the state process of the real contract across t_n
$D_n(x)$	feasible set of π_n at $X_n = x$
$\Gamma_{i,n}$	$\Gamma_{1,n} := \{0\}$, and $\Gamma_{2,n}$ corresponds to the feasible set of γ_n given $I_n > 0$.
${}_kp_0$	probability that a t_0-age policyholder survives over the time interval $(t_0, t_k]$
q_{k-1}	probability that a t_{k-1}-age policyholder dies during the time interval $(t_{k-1}, t_k]$
$\widehat{X}_{n^+}$	$\widehat{X}_{n^+} := (W(t_n^+), Z(t_n^+), 1, B(t_{n+1}), I_{n+1})$
$\widehat{X}_{n+1}$	$\widehat{X}_{n+1} := (W(t_{n+1}), Z(t_{n+1}), 1, B(t_{n+1}), I_{n+1})$
$\widehat{X}_{n^+}^{(I,m)}$	$\widehat{X}_{n^+}^{(I,m)} := \Big(W_{n^+}^{(m)}, 1, 1, B_{n+1}^{(m)}, I\Big)$
$\widehat{X}_{n+1}^{(I,m)}$	$\widehat{X}_{n+1}^{(I,m)} := \Big(W_{n^+}^{(m)} e^{L_n^{(m)}}, 1 \vee W_{n^+}^{(m)} e^{\bar{L}_n^{(m)}}, 1, B_{n+1}^{(m)}, I\Big)$
$\hat{x}^I$	$\hat{x}^I := (w, z, 1, b, I)$, a realized value of $\widehat{X}_{n+1}^{(I,1)}$

4.2 Model Setup and Contract Descriptions

In this section, we present the model setup and a brief contract description of the "Polaris Income Plus Daily" rider. For the detailed contract provisions, we refer to the contract prospectus [1].

In reality, contract events, such as withdrawals and deduction of insurance fees, only happen at a collection of deterministic time points: $\{t_1, t_2, \ldots, t_{N-1}\}$ with each t_n referred to as *withdrawal time* or *event time*. Further denote t_0 and $t_N := T$ as the attained age of the policyholder and the contract maturity[3], respectively. The time interval between two consecutive event times $\Delta t := t_n - t_{n-1}$ is assumed to be one year without loss of generality (w.l.o.g.); see, e.g., [47] and [48]. Let $\mathcal{I} := \{1, 2, \ldots, N-1\}$ label the collection of all event times and denote $\mathcal{I}_0 := \mathcal{I} \cup \{0\}$. Assume that all the random elements involved are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$. Also assume that there exists an equivalent martingale measure $\mathbb{Q}$ such that the discounted underlying asset price is a martingale.

[3]Here the maturity refers to the *Latest Annuity Date* when the contract is converted into a fixed annuity automatically, see the discussion in "Terminal payoff" below.

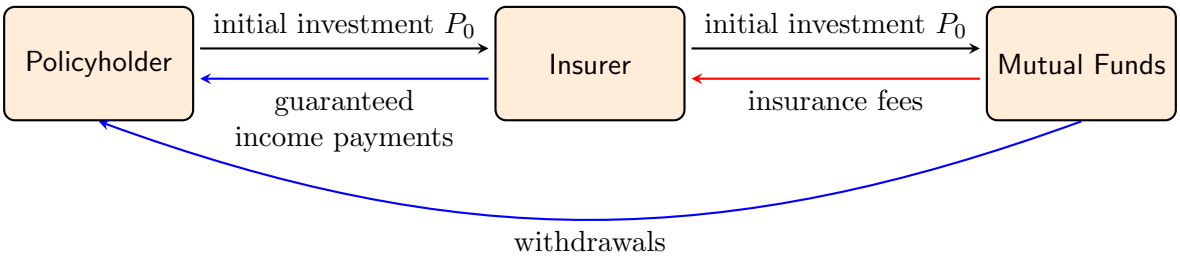


Figure 4.1: A diagram illustrating the mechanism of a variable annuity policy.

At the inception of the contract, the policyholder (PH) pays a lump sum to the writer, say P_0, which is invested in a basket of funds. An investment account is set to record the market value of the underlying funds and its time t value is denoted by $W(t)$. A shadow account, referred to as *income base* in the contract prospectus [1], is used to determine the guaranteed income payments whose time t value is denoted by $A(t)$. Throughout the contract life, the PH is entitled to withdraw any portion of the market value of the investment account. The insurer provides guaranteed income payments to the PH regardless of the performance of the underlying funds. In this sense, the insurer provides certain protection for the PH against the potential market decline. As compensation, the insurer deducts insurance fees from the investment account, periodically. The above mechanism is illustrated in Figure 4.1.

The remainder of this section is organized as follows. Section 4.2.1 introduces some auxiliary decision and state variables to render our mathematical discussions under a Markovian setting. The evolution mechanisms of all state variables, in particular, the investment account and the income base, are presented in Section 4.2.2 which prepare the ground of our pricing model in the next section.

4.2.1 Auxiliary Decision and State Variables

As commented in the introduction section, the payoffs of the Polaris depend on the timing of the first withdrawal of the PH. Therefore, it is necessary to introduce an auxiliary state variable to record the time of the first withdrawal, which paves the way to formulating the pricing problem under a Markovian stochastic optimal control framework.

First withdrawal time

First introduce a decision variable τ_n to model the PH's decision to initialize the withdrawal or not by taking values 1 and 0, respectively. Moreover, we use a new state variable I_n to record

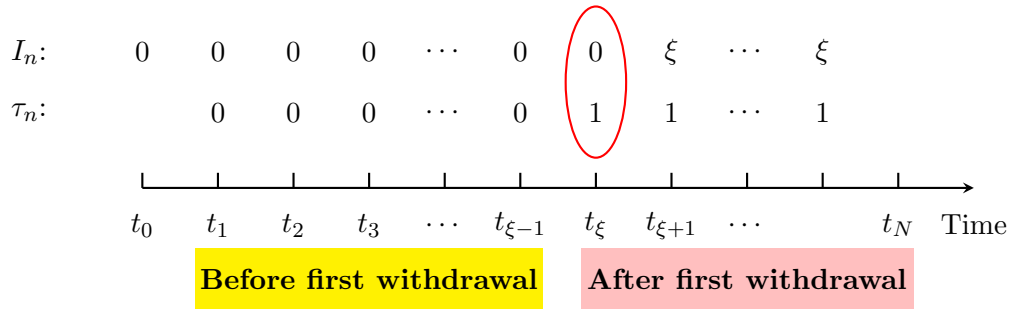


Figure 4.2: Evolution scheme of I_n and τ_n. t_ξ is the first withdrawal time.

the first withdrawal time which is recursively defined as follows: $I_0 = 0$ and

$$I_n = \begin{cases} n-1, & \text{if } I_{n-1} = 0 \text{ and } \tau_{n-1} = 1, \\ I_{n-1}, & \text{otherwise.} \end{cases}$$

The above definition is interpreted as follows. $I_n = 0$ corresponds to the case that the PH has not initialized the withdrawal by time t_n. Once the first withdrawal is taken at t_ξ for some $\xi \in \mathcal{I}$, I_n freezes at ξ forever. Therefore, I_n is a jump process with jump size equal to the index of the first withdrawal time.

Further prescribe that the feasible set of τ_n is a singleton $\{1\}$ once the withdrawal is initialized. Before the first withdrawal, the feasible set of decision variable τ_n is $\{0, 1\}$ instead. Therefore, the feasible set of τ_n is state-dependent.

Given I_n and $\tau_n = 1$, the index of the first withdrawal time is given by

$$\xi_n(I_n) = I_n \mathbb{1}_{\{I_n > 0\}} + n \mathbb{1}_{\{I_n = 0\}}, \tag{4.1}$$

where $\mathbb{1}_{\{E\}} = 1$ if event E happens; otherwise, $\mathbb{1}_{\{E\}} = 0$. The evolution mechanisms of I_n and τ_n are depicted by the diagram in Figure 4.2.

Remark 4.1. *From Figure 4.2, we can get a clear picture of the drawback of the forward simulation method proposed by [47] which simulates the state process by activating the withdrawal at a predetermined time step; see the first paragraph of their Section 3.3.2.*[4] *Suppose one initializes the withdrawal at t_ξ, then all simulated values of the above state variable I_n equal ξ for all $n \geq \xi + 1$. However, the feasible set of I_n is $\{0, 1, \ldots, n-1\}$ for $n \in \mathcal{I}$, and therefore the simulated values do not evenly spread over the feasible set and the regression estimate of the continuation function*

[4] To be more precise, in simulating the trajectories of the state process, [47] assumes the PH withdraws at a fixed amount throughout the contract life, or equivalently, the PH initiates the withdrawal at t_1 and maintains a constant withdrawal amount thereafter.

is severely skewed.

Guaranteed Income Payment

The necessity of introducing the state variable I_n stems from two important contract provisions of the Polaris. Firstly, any withdrawal amount smaller than a certain pre-specified amount does not reduce the income base value. This cap is called as *Maximal Annual Withdrawal Amount* (MAWA) and is calculated as a certain percentage of the income base. Specifically, if the first withdrawal happens at an age between 45 and 59, the MAWA is determined by 4.0% of the income base; and this MAWA percentage climbs to 5.0% if the age at the first withdrawal is between 60 and 64.

Secondly, even after the investment account is depleted, the PH is still entitled to withdraw a certain amount at each withdrawal time, which is referred to as the *Protected Income Payment* (PIP). Analogously, the PIP is computed as a certain percentage of the income base value and the PIP percentage steps up if the PH defers the withdrawal. In what follows, let us focus on the case that the MAWA percentage coincides with the PIP percentage and denote $G(\xi)$ as the MAWA/PIP percentage given the first withdrawal is taken at t_ξ. This is in accordance with the "Income Option 3" of the Polaris contract.[5]

In accordance with the above discussions, given the state variable I_n, the MAWA/PIP percentage is given by

$$\tilde{G}_n(I_n) := G(\xi_n(I_n)). \tag{4.2}$$

Feasible set of decision variables

Let γ_n denote the PH's withdrawal amount at event time t_n. Next, we discuss the feasible set of the bivariate decision variable $\pi_n := (\gamma_n, \tau_n)$. Define the feasible set of π_n, denoted by D_n, as follows:

$$D_n = \begin{cases} D_{n,1} \bigcup D_{n,2}, & \text{if } I_n = 0 \\ D_{n,2}, & \text{if } I_n > 0 \end{cases} \tag{4.3}$$

where

$$D_{n,1} := \Gamma_{n,1} \times \{0\} := \{0\} \times \{0\}, \tag{4.4}$$

$$D_{n,2} := \Gamma_{n,2} \times \{1\} = \Big[0, W(t_n) \vee \tilde{G}_n(I_n) A(t_n)\Big] \times \{1\} \tag{4.5}$$

[5] See the "Supplement to the variable annuity prospectus" of the prospectus [1] for details.

with $a \vee b := \max\{a, b\}$. The above equation system can be understood as follows. If $I_n = 0$, no withdrawal has been taken by time t_n and the PH has the option to initialize the withdrawal or not, which corresponds to a feasible set for π_n either being $D_{n,1}$ or $D_{n,2}$. In particular, if the PH chooses to postpone withdrawal ($\tau_n = 0$), the feasible set of withdrawal amount γ_n is clearly a singleton $\{0\}$. In the case $I_n > 0$, withdrawal was initialized before t_n and accordingly, the feasible set for the withdrawal amount is a continuum as given by $\Gamma_{n,2} = \left[0, W(t_n) \vee \tilde{G}_n(I_n)A(t_n)\right]$. It is notable that the PH can always enjoy a withdrawal amount up to the greater between MAWA/PIP and the investment account value, which is in accordance with the discussion in the last item "Guaranteed Income Payment".

It is worth stressing that the feasible set of the decision variable π_n depends on the state variables. Such dependency is sometimes highlighted by writing $D_n(X_n)$ with X_n being the state process defined in the next subsection.

4.2.2 Evolution Mechanisms of State Variables

Evolution of investment account

The investment account is reduced by withdrawals and insurance fees (also referred to as *rider charge*) periodically. We model the log-return of the underlying fund by a Lévy process and denote its increment over $(t_n, t]$ by $L_n(t)$, for $t > t_n$ and $n \in \mathcal{I}_0$. Hereafter, we use $W(t_n^+)$ to denote the value of investment account right after the deduction of the PH's withdrawal and insurance fee at an event time t_n. All the other quantities affixed with "(t_n^+)" in this chapter should be interpreted in the same manner.

The evolution mechanism of the investment account is inductively prescribed as follows.

- At the inception, the investment account value equals the initial investment, i.e., $W(t_0) = P_0$.
- Across the withdrawal time t_n, the investment account value is reduced by withdrawal and insurance fee charged by the insurer. We consider two different fee schemes in the sequel.
  - **Real contract.** In the real contract specification, the insurance fee charged at each event time is calculated as a certain percentage of the income base [1, pp. C-1], denoted by $\eta \in (0, 1)$.[6] Accordingly, investment account exhibits the following jump

[6]Here the rider charge percentage is assumed to be deterministic. In real contract specification, the value of η is tied to the VIX and thus is not a constant. Our framework can easily accommodate this feature by introducing an extra state variable as it is done in [32]. The key point is that the evolution of VIX is not influenced by the PH's decision and thus introducing this state variable would not ruin the argument for proving the existence of a bang-bang solution.

mechanism across an event time:

$$W(t_n^+) = \max\Big[W(t_n) - \underbrace{\gamma_n}_{\text{withdrawal}} - \underbrace{\eta A\big(t_n^+\big)}_{\text{insurance fee}}, 0\Big], \ n \in \mathcal{I}. \tag{4.6}$$

Remark 4.2. *One important observation can be made from the above equation: the insurance fees deducted by the insurer increases as the income base steps up. Therefore, a larger income base value might not be desirable to the PH as that means more fees levied by the insurer. This observation is in stark contrast with that of the GLWB contract studied by [6] and [47], as stated in Remark 4.16 of [6] that a larger benefit base*[7] *brings more value to the PH. In a formal mathematical language, the price function of the Polaris considered here is not necessarily monotone with respect to the state variable. This breaks the argument for proving the existence of the bang-bang solution where the monotonicity of the value function plays a pivotal role.*

– **Synthetic contract.** As alluded by Remark 4.2, the special fee scheme of the Polaris ruins the monotonicity of the price function which in turn brings a technical challenge to the pricing model discussed later. To circumvent this difficulty, we modify the fee scheme and consider the following transition equation of the investment account across an event time:

$$W(t_n^+) = (1-\eta)\big(W(t_n) - \gamma_n\big)^+, \quad n \in \mathcal{I}, \tag{4.7}$$

with $(a)^+ := \max[a, 0]$. The above equation assumes that the insurance fee is proportional to the investment account and therefore the amount $\eta\big(W(t_n)-\gamma_n\big)^+$ is deducted. This is commonly assumed in the literature of variable annuities, see, e.g., [6], [47], [74], among others. Under such an assumption, the monotonicity of the price function might hold and the accompanying pricing problem is easier to solve as we will see in the subsequent section. In the subsequent discussion, we will first concentrate on the jump mechanism (4.7) for our pricing model and call this modified contract as the "synthetic contract". We will show that the no-arbitrage price of this synthetic contract can be used as an upper bound for that of the "real contract" corresponding to the fee scheme (4.6).

- Between two consecutive event times, the investment account evolves according to:

$$W(t) = W(t_n^+)e^{L_n(t)}, \quad t \in (t_n, t_{n+1}], \quad n \in \mathcal{I}_0. \tag{4.8}$$

Since the PH is not allowed to withdraw at t_0, we have $W(t_0) = W(t_0^+)$. This convention

[7] The benefit base is the parallel income base in a GLWB contract. It basically plays the role of determining the guaranteed payment of the insurer.

applies to the other state variables.

Evolution of step-up value

Prior to PH's first withdrawal, the income base locks in the high water mark of the investment account since the inception ("global" running maximum); see the first case of system (4.9) below. This level is referred to as *step-up value*. After the initiation of the withdrawal, the income base value steps up to the running maximum of the investment account between two consecutive event times ("local" running maximum).

Introduce step-up value as an additional state variable and denote its time t value as $Z(t)$. In accordance with the preceding discussion, its evolution mechanism is inductively given as follows.

- At initiation, $Z(t_0) = P_0$.
- For $n \in \mathcal{I}$, the jump mechanism of step-up value across the withdrawal time t_n is given by

$$Z(t_n^+) = \begin{cases} Z(t_n), & \text{if } \tau_n = 0, \\ W(t_n^+), & \text{otherwise.} \end{cases} \tag{4.9}$$

- Between two consecutive withdrawal times,

$$Z(t) = \max \left[\sup_{s \in (t_n, t]} W(s), Z(t_n^+) \right], \quad t \in (t_n, t_{n+1}], \quad \text{for} \quad n \in \mathcal{I}_0. \tag{4.10}$$

Some interpretations regarding the above state variable $Z(t)$ are given as follows. $Z(t_n)$ gives the global running maximum of investment account over $[0, t_n]$ if no withdrawal has been made before t_n; see the first case of the above system (4.9). If a withdrawal is taken at t_n, i.e., $\tau_n = 1$, then the step-up value is reset to be the post-withdrawal value of the investment account, which corresponds to the second case of Eq. (4.9). Accordingly, the state variable $Z(t_{n+1})$ records the running maximum of the underlying fund over $(t_n, t_{n+1}]$.

Evolution of income base

As commented previously, the income base is a shadow account which plays a crucial role in determining the MAWA/PIP. Let us discuss the jump mechanism of the income base in several cases. If the first withdrawal has not been taken, the income base locks in the high water mark of the investment account since inception; see the first case of Eq. (4.11) below. If the first withdrawal is less than MAWA and is taken at t_n, the income base only steps up to the investment

account value after withdrawal, see the second case of Eq. (4.11) below. If the first withdrawal has been taken but the annual withdrawal is less than the MAWA, the income base steps up to the running maximum of the investment account between two consecutive anniversaries; see the third case of Eq. (4.11) below. If the annual withdrawal amount exceeds the MAWA which is referred to as *excess withdrawal*, the income base is first reduced by the factor $\frac{W(t_n)-\gamma_n}{W(t_n)-\tilde{G}_n(I)A(t_n)}$ and then steps up to the greater between the income base and the investment account, see the last case of Eq. (4.11) below. The detailed jump mechanism of the income base is inductively summarized as follows.

- At initiation, $A(t_0) = P_0$.
- For $n \in \mathcal{I}$, the jump mechanism of the income base across withdrawal time t_n is given by

$$A(t_n^+) = \begin{cases} A(t_n) \vee Z(t_n), & \text{if } \tau_n = 0, \\ A(t_n) \vee (W(t_n) - \gamma_n), & \text{if } I_n = 0, 0 < \gamma_n \leq \tilde{G}_n(I_n)A(t_n), \\ A(t_n) \vee Z(t_n), & \text{if } I_n > 0, 0 \leq \gamma_n \leq \tilde{G}_n(I_n)A(t_n), \\ \left(\frac{W(t_n)-\gamma_n}{W(t_n)-\tilde{G}_n(I_n)}A(t_n)\right) \vee (W(t_n) - \gamma_n), & \text{if } \tilde{G}_n(I_n)A(t_n) < \gamma_n \leq W(t_n), \end{cases} \tag{4.11}$$

  where one should recall from Eq. (4.2) that $\tilde{G}(I_n) = G(\xi_n(I_n))$ is the MAWA/PIP percentage.
- The income base remains constant over time interval $(t_n, t_{n+1}]$: $A(t) \equiv A(t_n^+)$, $t \in (t_n, t_{n+1}]$, for $n \in \mathcal{I}_0$.

Evolution of adjusted payment

If the PH passes away during the contract life, death benefits are paid to the beneficiary. The amount of death benefits is determined by the greater of the investment account value and the initial purchase payment of the PH adjusted by the withdrawals and this quantity is referred to as *adjusted payment* in what follows. Accordingly, it is necessary to introduce a state variable $B(t)$ to record the adjusted payment so as to further determine the death payment payable at each event time. The amount of withdrawal-adjusted payment is initially set as P_0 since there is no withdrawal at inception. If the annual withdrawal amount is less than MAWA, the adjusted payment is simply reduced by the withdrawal amount; otherwise, the adjusted payment is first reduced by MAWA and then scaled by the factor $\frac{W(t_n)-\gamma_n}{W(t_n)-\tilde{G}_n(I_n)A(t_n)}$. Therefore, the evolution mechanism of $B(t)$ is given as follows.

- At initiation, $B(t_0) = P_0$.

- The jump mechanism across withdrawal time t_n is given by

$$B(t_n^+) = \begin{cases} \big(B(t_n) - \gamma_n\big)^+, & \text{if } 0 \le \gamma_n \le \tilde{G}_n(I_n)A(t_n) \\ \frac{W(t_n)-\gamma_n}{W(t_n)-\tilde{G}_n(I_n)A(t_n)}\big(B(t_n) - \tilde{G}_n(I_n)A(t_n)\big)^+, & \text{if } \tilde{G}_n(I_n)A(t_n) < \gamma_n \le W(t_n), \end{cases} \tag{4.12}$$

  for $n \in \mathcal{I}$.

- The adjusted payment changes its value only at each withdrawal time and therefore $B(t) \equiv B(t_n^+)$ for $t \in (t_n, t_{n+1}]$ and $n \in \mathcal{I}_0$.

Transition of state process across withdrawal time

So far defined five state variables have been introduced and they together compose a multivariate controlled Markov process, denoted as $X_n := \big(W(t_n), Z(t_n), A(t_n), B(t_n), I_n\big)$. Its dependency on the PH's decision is suppressed for notational simplicity. The readers should bear in mind that X_n implicitly depends on the action taken by the PH up to time t_{n-1}, i.e., $\{\pi_i\}_{i=1}^{n-1}$. Define the transition equation of X_n across an event time t_n as

$$K_n(X_n, \pi_n) = \big(W(t_n^+), Z(t_n^+), A(t_n^+), B(t_n^+), I_{n+1}\big) =: X_{n^+}. \tag{4.13}$$

The transition function $K_n(\cdot,\cdot)$ is determined in accordance with Eqs. (4.7, 4.9, 4.11, 4.12) and its expression is relegated to Appendix C.1 for the clarity of presentation. Henceforth we will call X_{n^+} as the *post-withdrawal* value of the state process.

The expression of $K_n(\cdot,\cdot)$ is based on Eq. (4.7) for the jump mechanism of the investment account, whereas, as commented in Remark 4.2, the transition of the investment account across an event time is governed by Eq. (4.6) in the real contract specification of the Polaris. In such a case, the accompanying transition function, denoted by $\bar{K}_n(\cdot,\cdot)$, differs from $K_n(\cdot,\cdot)$ only in the first component which corresponds to the transition function of the investment account; see Appendix C.1 for the specific expression of $\bar{K}_n(\cdot,\cdot)$. The state process obeying this transition mechanism can be defined in a similar manner and is denoted by $\bar{X}_n$. Accordingly,

$$\bar{X}_{n^+} := \bar{K}_n\left(\bar{X}_n, \pi_n\right). \tag{4.14}$$

It is worth noting that while the transition mechanism across a withdrawal date differs between $\bar{X}_n$ and X_n, their evolution schemes between two consecutive withdrawal dates are exactly the same.

4.3 Pricing Model

4.3.1 Stochastic Control Formulation

This subsection formulates the pricing model of the Polaris variable annuity under a stochastic optimal control framework.

Withdrawal charge

Let $g_n(X_n, \pi_n)$ be the dollar amount of the PH's cash inflow resulted from a decision π_n at time t_n. It is given by

$$g_n(X_n, \pi_n) = \gamma_n - \kappa_n \left(\gamma_n - \tilde{G}_n(I_n)A(t_n)\right)^+.$$

In other words, only the excess withdrawal part $\gamma_n - \tilde{G}_n(I_n)A(t_n)$ is subject to a proportional penalty κ_n which is time-dependent [1, pp. 26–27].

Terminal payoff

Starting from the inception, the contract stays in the *Accumulation Phase* during which the preceding contract provisions apply. After the *Latest Annuity Date* t_N, the Accumulation Phase terminates and the *Income Phase* is automatically initiated, that is, the contract is converted into a fixed annuity with time-t_N value equal to $W(t_N)$. Therefore, w.l.o.g., we assume that the terminal payoff is a lump sum payment of $W(t_N)$[8].

Mortality risk

The literature of variable annuity conventionally assumes the mortality risk faced by the insurer can be diversified via issuing a large number of similar contracts; for a theoretical foundation behind this, see [39]. We adopt actuarial notation to denote ${}_kp_0$ as the probability that a t_0-age PH survives over the period $(t_0, t_k]$ and q_{k-1} as the probability that a t_{k-1}-age PH passes away during the time interval $(t_{k-1}, t_k]$.

[8]The notions of *income phase* and *accumulation phase* in the Polaris are different from those in the GLWB policy studied by [48].

Dynamic programming equation

Let r be a certain risk-free rate and denote $\varphi := e^{-r}$. Given an initial state X_0 and a PH's decision $\pi := (\pi_k)_{k=1}^{N-1}$, the expected present value (EPV) of all future policy payoffs is given by

$$\mathcal{J}(X_0, \pi) = \mathbb{E}^{\mathbb{Q}}\left[\sum_{k=1}^{N-1} \varphi^k f_k(X_k, \pi_k) + \varphi^N {}_{N-1}p_0 W(t_N)\right],$$

where

$$f_k(X_k, \pi_k) = {}_kp_0 g_k(X_k, \pi_k) + {}_{k-1}p_0 q_{k-1}\Big[B(t_k) \vee W(t_k)\Big], \tag{4.15}$$

and the superscript in $\mathbb{E}^{\mathbb{Q}}$ stresses the expectation is taken under the martingale measure $\mathbb{Q}$.

From the insurer's perspective, the first term ${}_kp_0 g_k(X_k, \pi_k)$ represents the cash inflows of living PHs at time t_k. The second term in Eq. (4.15) corresponds to the death benefits paid at time t_k for those PHs who passed away during the time period $[t_{k-1}, t_k]$. Therefore, $f_k(X_k, \pi_k)$ gives the total payoff of the policy at time t_k. Finally, the term ${}_{N-1}p_0 W(t_N)$ is the value of converted fixed annuity of living PHs at the maturity.

The no-arbitrage price of the synthetic contract at inception is characterized as the optimal value function of the following stochastic optimal control problem:

$$V_0(X_0) := \sup_{\pi \in \mathbf{\Pi}} \mathcal{J}(X_0, \pi),$$

where $\mathbf{\Pi}$ is the set of all admissible decisions:

$$\mathbf{\Pi} := \left\{(\pi_k)_{k=1}^{N-1} \;\middle|\; \pi_k \text{ is } \mathcal{F}_k - \text{measurable},\ \pi_k \in D_k\left(X_k\right), k = 1, \ldots, N-1\right\}, \tag{4.16}$$

and $X_0 = (P_0, P_0, P_0, P_0, 0)$ with P_0 being the intitial purchase amount of the policyholder.

Let $x = (W, Z, A, B, I)$ be a realized value of the state process. Via exploiting the Bellman's principle of optimality, the value function $V_0(\cdot)$ can be solved through the following backward recursion procedure:

1. For $n = N$, $V_N(x) = {}_{N-1}p_0 W$;
2. For $n = N-1, \ldots, 1$,

$$V_n(x) = \sup_{\pi_n \in D_n(x)} \Big[f_n(x, \pi_n) + \varphi C_n\big(K_n(x, \pi_n)\big)\Big], \tag{4.17}$$

where

$$C_n(\cdot) := \mathbb{E}^{\mathbb{Q}}\left[V_{n+1}(X_{n+1})\middle| X_{n^+} = \cdot\right]; \tag{4.18}$$

3. Finally, $V_0(x) = \varphi\mathbb{E}^{\mathbb{Q}}\left[V_1(X_1)\middle| X_0 = x\right]$.

Note that the transition equation (4.13) is adopted for calculating the conditional expectation $C_n\big(K_n(x,\pi_n)\big)$, and accordingly, $V_0(x_0)$ returns the fair value of the synthetic contract with the initial state $x_0 = \big(P_0, P_0, P_0, P_0, 0\big)$. The value function accompanying the real contract, denoted by $\bar{V}_n(\cdot)$, can be similarly solved via the preceding dynamic programming equation except that $K_n(\cdot,\cdot)$ should be replaced by $\bar{K}_n(\cdot,\cdot)$ as given by Eq. (4.14).

Two key issues arise in implementing the above backward recursion procedure. The first issue is the evaluation of the continuation function $C_n(\cdot)$. This can be numerically resolved in the spirit of the BSBU algorithm developed in Chapter 2. The second issue is the possibility of not being able to find the global optimizer of the optimization problem at each time step. This is due to the fact that the objective function of the optimization problem is in general neither convex nor concave and there does not exist a generic optimization algorithm to find its global optimizer in a computationally efficient way. One possible idea to circumvent this dilemma is considering the optimization problem over a finite subset of the original feasible set D_n, denoted by $\widehat{D}_n$, and then showing the equivalence of these two optimization problems. If this is achievable, the stochastic control problem is said to exhibit a *bang-bang solution*. In the subsequent subsection, we will prove the existence of the bang-bang solution for the stochastic optimization problem associated with the synthetic contract, and such a result is critical to the efficacy of the numerical approach we will develop in Section 4.4 for the computation of the fair value of the synthetic contract. We will further prove that $V_0(x_0) \geq \bar{V}_0(x_0)$ in Section 4.3.3, that is, the fair value for the real contract is bounded from above by that of the synthetic contract, which is the primary driving force behind introducing the synthetic contract.

4.3.2 Bang-bang Analysis

In the following, we will see that the stochastic optimal control problem associated with the synthetic contract exhibits the bang-bang solution, that is, the optimal withdrawal strategy at each event time is restricted into a few choices.

We will first establish the scaling property (positive homogeneity) of the optimal value function which is beneficial in two-fold: first, it simplifies the proof of the existence of the bang-bang solution; second, it reduces the dimension of the state variable by one (see Eq. (4.23) in the sequel) and accordingly alleviates the computational burden of the numerical algorithm proposed in Section 4.4 in the sequel.

Proposition 4.1 (Positive Homogeneity). *For $c > 0$, let $x^c := (c \cdot W, c \cdot Z, c \cdot A, c \cdot B, I)$ and $x := (W, Z, A, B, I)$ be two realized values of the state variable X_n. Then, the value functions of the synthetic contract and real contract exhibit the following scaling property:*

$$V_n(x^c) = c \cdot V_n(x) \quad \textit{and} \quad \bar{V}_n(x^c) = c \cdot \bar{V}_n(x), \quad \text{for} \quad n = 0, 1, \ldots, N.$$

The proof of the preceding proposition is relegated to Appendix C.4. The statements delivered in the above proposition are not trivial since the feasible set of the PH's decision depends on whether the withdrawals begin or not. A dedicated investigation should be carried out. It is worth stressing that the positive homogeneity does not hold for general VA products, for example, the GMWB; see [7] for a discussion.

The following theorem gives a full characterization of the optimizer of the optimization problem in the Bellman equation (4.17).

Theorem 4.1 (Bang-bang Solution). *For any withdrawal time t_n, let $x = (W, Z, A, B, I)$ be a realized value of X_n. Suppose that $\xi \longmapsto G(\xi)$ for any $\xi \in \mathcal{I}$ is monotone. Then the synthetic contract exhibits the following optimal withdrawal strategies:*

(i) *If the withdrawal has been initialized, i.e., $I > 0$, the optimal withdrawal amount is either (i) $\hat{\gamma}_n = 0$, (ii) $\hat{\gamma}_n = G(I)A$, or (iii) $\hat{\gamma}_n = W$.*

(ii) *If the withdrawal has not been initialized, i.e., $I = 0$, the optimal strategy is either to postpone the withdrawal, i.e. $\tau_n = 0$, or to activate the withdrawal, i.e. $\tau_n = 1$, and the corresponding optimal withdrawal amount is either (i) $\hat{\gamma}_n = G(n)A$, or (ii) $\hat{\gamma}_n = W$.*

The proof of Theorem 4.1 is relegated to Appendix C.4. Here we give some intuitions behind the existence of the bang-bang solution. In view of the Bellman equation (4.17), one has to solve a deterministic optimization problem at each event time. If the accompanying objective function is convex and the feasible set of the decision variable is a convex set, then the supremum of the objective is attained among the collection of extreme points. In order to guarantee the convexity of the optimization problem can be propagated from step T down to step 0, the monotonicity of value function plays an indispensable role. As commented in Remark 4.2, when the insurance fee is proportional to the income base (see Eq. (4.6)), it is not clear whether the value function is monotone w.r.t. the income base or not because a higher income base means more insurance fees charged by the insurance company. Accordingly, it is generally difficult to show the existence of a bang-bang solution for the pricing model corresponding to the real contract. We conjecture that the statements delivered in the above theorem do not hold for the real contract.

The conclusions in Theorem 4.1 can be equivalently stated as:

$$V_n(x) = \max_{\pi_n \in \widehat{D}_n} \Big[f_n(x, \pi_n) + \varphi C_n\big(K_n(x, \pi_n)\big) \Big] =: \max_{\pi_n \in \widehat{D}_n} \mathcal{V}_n(x, \pi_n), \tag{4.19}$$

for $n \in \mathcal{I}$, where

$$\widehat{D}_n = \begin{cases} \{(0,0), \big(G(n)A, 1\big), \big(W, 1\big)\}, & \text{if } I = 0, \\ \{(0,1), \big(G(I)A, 1\big), \big(W, 1\big)\}, & \text{otherwise,} \end{cases} \tag{4.20}$$

and its dependency on state $x = (W, Z, A, B, I)$ is suppressed for notational brevity.

The sequel proposition gives a closed-form expression for the value function given the investment account is depleted.

Proposition 4.2 (Explicit Solution)**.** *Let $X_n = (W, Z, A, B, I)$ with $I > 0$. Suppose $W = 0$ and $I > 0$. Then the no-arbitrage prices of the real contract and the synthetic contract at time t_n are exactly the same and exhibit the following analytical expression:*

$$\begin{aligned} V_n^{\mathrm{S}}(0, Z, A, B, I) &= {}_np_0 G(I)A + {}_{n-1}p_0 q_{n-1} B + \sum_{j=n+1}^{N-1} {}_jp_0 \varphi^{j-n} G(I)\big(A \vee Z\big) \\ &\quad + \sum_{j=n+1}^{N-1} {}_{j-1}p_0 q_{j-1} \varphi^{j-n} \Big(\widetilde{B} - (j-n)G(I)\big(A \vee Z\big)\Big)^+, \end{aligned} \tag{4.21}$$

with $\widetilde{B} := \big(B - G(I)A\big)^+$ and $n \in \mathcal{I}$.

The proof of the preceding proposition is straightforward. We give the intuition behind the above equation as follows. On one hand, it is easy to see that the optimal withdrawal strategy after the investment account is depleted is always withdrawing at the amount $G(I)A$. On the other hand, once the investment account is depleted, no future insurance fee will be deducted and accordingly, the impact of the difference in fee schemes on the contract value vanishes. In view of this, it is not surprising that the above Eq. (4.21) holds for both synthetic and real contracts. The first summation term corresponds to the present value of future withdrawals and the second summation gives the present value of future death benefits.

4.3.3 Pricing Bounds

The following theorem discloses the relationship between the no-arbitrage prices of the synthetic contract and the real contract.

Theorem 4.2 (Pricing Upper Bound)**.** *Let P_0 be initial investment amount of the policyholder. Denote $x_0 := \big(P_0, P_0, P_0, P_0, 0\big)$. The no-arbitrage price of the real contract is dominated by that of the synthetic contract, that is, $\bar{V}_0(x_0) \leq V_0(x_0)$.*

Remark 4.3. *The above theorem casts new insights on the design of the fee structure. In most earlier guaranteed withdrawal benefits, the rider charge is prescribed to be proportional to the investment account and the prevailing market fee rate is observed to be less than the theoretical fee rate obtained from the no-arbitrage pricing framework; see, e.g., [62], [28], [63], and [52]. In other words, the insurance fees received by the insurer are insufficient to finance the hedging portfolio. The above theorem discloses that the insurer's risk exposure can be reduced by charging the fees against the income base. This shows the advantage of the fee structure of the Polaris over those in the GMWB and GLWB studied in the literature, see, e.g., [6] and [47].*

Pricing lower bound A lower bound for $\bar{V}_0(x_0)$ is easy to obtain. One may first restrict the feasible set of the decision variable at each time step, D_n, into the finite subset $\widehat{D}_n$ given in Eq. (4.20) and then solve a similar stochastic optimization problem as (4.19) except that $K_n(\cdot,\cdot)$ is replaced by $\bar{K}_n(\cdot,\cdot)$. Since the original feasible set is reduced into a subset, this will produce a sub-optimal solution, or equivalently, a lower bound of the no-arbitrage price of the real contract.

4.4 Numerical Approach

This section develops a Least Squares Monte Carlo (LSMC) algorithm to compute the no-arbitrage price of the synthetic contract. It is worth noting that our proposed approach is also applicable to computing the lower bound of the no-arbitrage price of the real contract by the procedure described at the end of the last section.

The road map of this section is as follows: firstly, we will discuss the approximation of the continuation function $C_n(\cdot)$ by exploiting some dimension reduction tricks and a nonparametric sieve estimation method; secondly, we propose to directly simulate the post-withdrawal value of the state process and present the whole LSMC algorithm; finally, the convergence analysis of the algorithm and some discussions are given.

4.4.1 Nonparametric Sieve Estimation for Continuation Function

Dimension reduction

To enhance the efficiency of approximating $C_n(\cdot)$ by the regression method proposed in the sequel, we would like to first reduce the dimensionality of the state variable. Introduce the following normalized transition function:

$$\hat{K}_n(x,\pi_n) = \left(\hat{K}_{n,1}(x,\pi_n), \hat{K}_{n,2}(x,\pi_n), 1, \hat{K}_{n,4}(x,\pi_n), K_{n,5}(x,\pi_n)\right), \tag{4.22}$$

where $\hat{K}_{n,i}(x,\pi_n) = K_{n,i}(x,\pi_n)/K_{n,3}(x,\pi_n)$, $i = 1,2,4$, for $K_{n,3}(x,\pi_n) > 0$.

By exploiting the scaling property delivered in Proposition 4.1, $\mathcal{V}_n(x,\pi_n)$ (see Eq. (4.19)) can be equivalently written as

$$\mathcal{V}_n(x,\pi_n) = \begin{cases} f_n(x,\pi_n) + \varphi K_{n,3}(x,\pi_n)\cdot C_n\left(\hat{K}_n(x,\pi_n)\right), & \text{if } K_{n,3}(x,\pi_n) > 0, \\ {}_np_0\Big[(1-\kappa_n)W + \kappa_n\tilde{G}_n(I)A\Big] + {}_{n-1}p_0 q_{n-1}\big(B\vee W\big), & \text{otherwise,} \end{cases} \tag{4.23}$$

where the second case of the above system holds because $K_{n,3}(x,\pi_n) = 0$ entails $\gamma_n = W$.

Observe that all terms in the above equation are exact except for the continuation function $C_n(\cdot)$ which is essentially a four-variate function since the third component of $\hat{K}_n(x,\pi_n)$ is fixed as 1; see Eq. (4.22). Hence the dimension of the state process has been reduced by one. Compressing the dimension of the state process by exploiting the positive homogeneity property of the value function is a common technique in the literature of variable annuity; see, e.g., [7] and [47].

The following proposition discloses that the dimensionality of the problem can be further reduced by one if one restricts her attention of the continuation function into some bounded set.

Proposition 4.3 (Shape Constraints)**.** *The continuation function $C_n(\cdot)$ defined in Eq. (4.18) exhibits the following properties:*

(i) $k_2 \longmapsto C_n(k_1,k_2,1,k_4,k_5)$ *is constant over* $[0,1]$*;*

(ii) $k_1 \longmapsto C_n(k_1,k_2,1,k_4,k_5)$ *and* $k_4 \longmapsto C_n(k_1,k_2,1,k_4,k_5)$ *are convex and monotone.*

The proof of the above proposition is relegated to Appendix C.4.

Property (i) of Proposition 4.3 states that $C_n(k_1,k_2,1,k_4,k_5)$ is invariant to k_2 over certain region and therefore one may view it as a three-variate function. This property stems from the unique structure of the Polaris instead of the homogeneity result in Proposition 4.1 which has already been exploited by us to reduce the dimensionality by one in deriving Eq. (4.23). Such a special structure does not apply to the GMWB and the GLWB.

Property (ii) of the above proposition gives extra shape information of the continuation function which further implies the value function is convex and monotone (C.M.) according to the proof of Theorem 4.1; see Appendix C.4. The C.M. property is commonly observed in various financial products. For instance, it can be shown that the price function of an American call option is C.M. w.r.t. the spot price under some mild conditions, see, e.g., [34]. The C.M. property of the price function of the GLWB has been studied in [6] and [47]. In light of this, it is natural to expect a good numerical estimate for the continuation function can inherit the C.M. property which is meaningful in at least two aspects: first, the C.M. properties of the continuation function and price function have sensible economic interpretations; second, as we have seen in Chapter

2, the shape-preserving sieve estimation method makes the regression estimate insensitive to the tuning parameter and thus reduces the computational cost. Unfortunately, for most LSMC algorithms proposed in the literature, the regression estimate often violates the C.M. shape constraint because a linear combination of the basis functions might not be a C.M. function, see, e.g., [34] and [48, pp. 825–826]. This dilemma not only applies to the LSMC but also is suffered by many other numerical approaches such as Fourier transform approach ([48]). In the sequel, we will employ the shape-preserving sieve estimation technique of [83, 84] to approximate the continuation function as in Chapter 2.

Artificial simulation of the state process

Let us restrict our attention of the continuation function $C_n(\cdot)$ to the following bounded set:

$$\mathcal{K}_n := [0,1] \times \{1\} \times \{1\} \times [0,1] \times \{0,1,\ldots,n\}.$$

The motivation behind approximating $C_n(\cdot)$ over the above set $\mathcal{K}_n$ will become clear after the presentation of the whole LSMC algorithm; see the subsequent Remark 4.5. Inductively assume that there already exists an estimate for the value function at time step $n+1$, denoted by $V_{n+1}^{\mathrm{E}}(\cdot)$, because the value function V_{N-1} will be shown explicitly known.

Simulation of post-withdrawal value at t_n^+ Let

$$\widehat{X}_{n^+} := \left(W(t_n^+), Z(t_n^+), 1, B(t_{n+1}), I_{n+1}\right),$$

and

$$\widehat{X}_{n+1} := \left(W(t_{n+1}), Z(t_{n+1}), 1, B(t_{n+1}), I_{n+1}\right).$$

It is notable that the third component of $\widehat{X}_{n^+}$ (resp., $\widehat{X}_{n+1}$) is normalized to be one in comparison with X_{n^+} (resp., X_{n+1}) defined previously. For each $I_{n+1} = I \in \{0,1,\ldots,n\}$, denote the m-th simulated value of $\widehat{X}_{n^+}$ by

$$\widehat{X}_{n^+}^{(I,m)} = \left(W_{n^+}^{(m)}, 1, 1, B_{n+1}^{(m)}, I\right). \tag{4.24}$$

Note that the simulated values of $Z(t_n^+)$ and $A(t_n^+)$ are fixed as 1 because one is solely interested in the value of continuation function over $\mathcal{K}_n$. The superscript in $\widehat{X}_{n^+}^{(I,m)}$ stresses that the simulated value of I_{n+1} is fixed as I.

The pair $\left(W_{n^+}^{(m)}, B_{n+1}^{(m)}\right)$ is simulated from some exogenously given bivariate distribution with support $[0,1]^2$ and cumulative distribution function $\mu(\cdot,\cdot)$. For each m and n, we simulate the

pair $\left(W_{n^+}^{(m)}, B_{n+1}^{(m)}\right)$ as follows:

1. Simulate two independent random variables from exponential distribution with mean 1, denoted as Y_1 and Y_2, respectively.
2. Set $\left(W_{n^+}^{(m)}, B_{n+1}^{(m)}\right) = \left(Y_1 \wedge 1, Y_2 \wedge 1\right)$ with $a \wedge b := \min\{a, b\}$.

Simulation of the state process at t_{n+1} Given $\widehat{X}_{n^+}^{(I,m)}$, generate the random sample of the state process at t_{n+1} as follows. Let $\left(L_n^{(m)}, \bar{L}_n^{(m)}\right)$ be the m-th simulated value of the increment and running maximum of Lévy process over time interval $[t_n, t_{n+1}]$ which are independent of $\widehat{X}_{n^+}^{(I,m)}$. The simulated value of $\widehat{X}_{n+1}$ is given by

$$\widehat{X}_{n+1}^{(I,m)} := \left(W_{n^+}^{(m)} e^{L_n^{(m)}}, 1 \vee W_{n^+}^{(m)} e^{\bar{L}_n^{(m)}}, 1, B_{n+1}^{(m)}, I\right) \tag{4.25}$$

which is in accordance with the evolution mechanisms of the investment account and step-up value; see Eqs. (4.8) and (4.10), respectively.

Shape-preserving sieve estimation In the following, we employ the shape-preserving sieve estimation method of [83, 84] in the following. Different from the case in Chapter 2, the covariate variable/regressor has dimension two in the present regression problem. Let $\left\{b_{\ell,k}^J(x_1, x_2)\right\}_{0 \le \ell, k \le J}$ be a set of bivariate Bernstein polynomials, that is,

$$b_{\ell,k}^J(x_1, x_2) = \binom{J}{\ell}\binom{J}{k} x_1^\ell (1 - x_1)^{J-\ell} x_2^k (1 - x_2)^{J-k}, \tag{4.26}$$

for $x_1, x_2 \in [0, 1]$, $\ell, k = 0, 1, \ldots, J$.

Remark 4.4. *The notation J here stands for the maximal degree of the Bernstein polynomials. While cautious readers might note that $J + 1$ corresponds to the total number of basis functions used in the sieve estimation method in the previous Chapters 2 and 3. We abuse this notation a little bit as long as it causes no confusion in the present chapter.*

The shape-preserving sieve estimation method approximates the continuation function $C_n(\cdot)$ by a linear combination of the above Bernstein polynomials, that is,

$$C_n^{\mathrm{E}}\big(k_1, 1, 1, k_4, I\big) = \mathbf{b}^\intercal(k_1, k_4) \cdot \boldsymbol{\beta}_{n,I}^{\mathrm{E}}, \quad \text{for} \quad \big(k_1, 1, 1, k_4, I\big) \in \mathcal{K}_n, \tag{4.27}$$

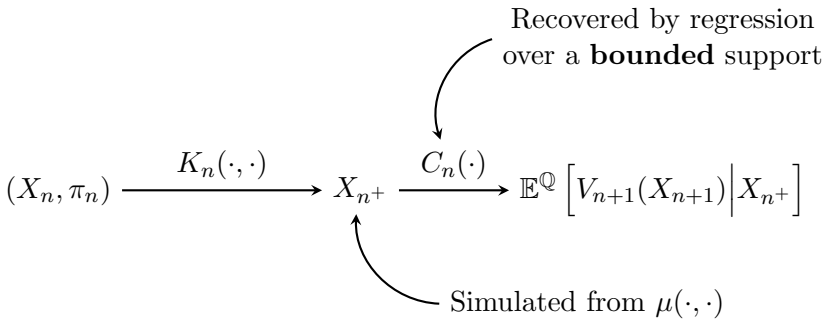


Figure 4.3: A diagram for estimating the continuation function. It is worth noting that the regression is conducted once to recover $C_n(\cdot)$ per time step and $C_n\big(K_n(X_n,\pi_n)\big)$ can be computed for various pairs of (X_n,π_n).

where $\mathbf{b}\,(k_1,k_4)$ is a $(J+1)^2$-by-1 vector-valued function with each element given by $b^J_{\ell,k}(k_1,k_4)$ and $\boldsymbol{\beta}^{\mathrm{E}}_{n,I}$ solves:

$$\begin{cases} \min\limits_{\boldsymbol{\beta}} \frac{1}{M}\sum_{m=1}^{M}\left[V^{\mathrm{E}}_{n+1}\left(\widehat{X}^{(I,m)}_{n+1}\right)-\mathbf{b}^{\intercal}\left(\widehat{X}^{(I,m)}_{n^+}\right)\boldsymbol{\beta}\right]^2 \\ \text{subject to}\quad \mathbf{A}\boldsymbol{\beta}\geq\mathbf{0}, \end{cases} \tag{4.28}$$

with $\widehat{X}^{(I,m)}_{n^+}$ and $\widehat{X}^{(I,m)}_{n+1}$ given in Eqs. (4.24) and (4.25), respectively, $\boldsymbol{\beta}$ being a $(J+1)^2$-by-1 vector, and $\mathbf{0}$ denoting a $2(J^2+J)$-by-1 null vector. Under a particular choice of $\mathbf{A}$ (see Appendix C.3), the linear constraint $\mathbf{A}\boldsymbol{\beta}\geq\mathbf{0}$ in the above optimization problem ensures $k_j\mapsto C^{\mathrm{E}}_n\big(k_1,1,1,k_4,I\big)$ is C.M. for $j=1$ and 4 ([84]).

It is worth noting that $\boldsymbol{\beta}^{\mathrm{E}}_{n,I}$ depends on I due to the dependency of the random sample $V^{\mathrm{E}}_{n+1}\left(\widehat{X}^{(I,m)}_{n+1}\right)$ on I. The superscript E emphasizes that $\boldsymbol{\beta}^{\mathrm{E}}_{n,I}$ is a statistical estimate (random vector). Figure 4.3 is an illustrative diagram for the artificial simulation method and the estimation of continuation function.

Estimate for the value function

Now it is ready to give a numerical estimate for the value function. Recall that an analytical expression of the value function is available when the investment account is depleted (Eq. (4.21)), and therefore, there is no need to use a numerical estimate under such a situation. In view of

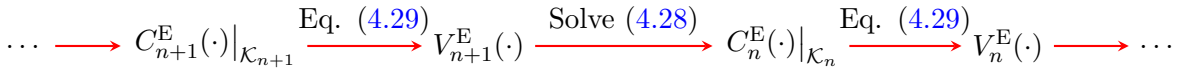


Figure 4.4: A diagram illustrating the propagation of information. It is worth noting that in the evaluation of $V^{\mathrm{E}}_n(\cdot)$, the information of $C^{\mathrm{E}}_{n+1}(\cdot)\big|_{\mathcal{K}_{n+1}}$ is implicitly required.

this and Eqs. (4.19, 4.23), the numerical estimate for the value function is given by

$$V^{\mathrm{E}}_n(x) = \begin{cases} \max\limits_{\pi_n \in \widehat{D}_n} \left[f_n(x,\pi_n) + \varphi K_{n,3}(x,\pi_n) C^{\mathrm{E}}_n\left(\hat{K}_n(x,\pi_n)\right)\right], & \text{if } K_{n,i}(x,\pi_n) > 0 \text{ for } i = 1,3, \\ \max\limits_{\pi_n \in \widehat{D}_n} \left[f_n(x,\pi_n) + \varphi K_{n,3}(x,\pi_n) V^{\mathrm{S}}_{n+1}\left(\hat{K}_n(x,\pi_n)\right)\right], & \text{if } K_{n,1}(x,\pi_n) = 0 \text{ and } K_{n,3}(x,\pi_n) > 0, \\ {}_np_0\Big[(1-\kappa_n)W + \kappa_n GA\Big] + {}_{n-1}p_0 q_{n-1} B, & \text{otherwise.} \end{cases} \tag{4.29}$$

Recall that $\widehat{D}_n$ is a lattice (see Eq. (4.20)), and therefore, the optimization problem in the above equation can be solved by a linear search.

Before closing this subsection, we make a remark on the set $\mathcal{K}_n$.

Remark 4.5. *It is easy to see from the proof of Proposition 4.3 in Appendix C.4 that only the knowledge of $C_{n+1}(\cdot)$ restricted on the set $\mathcal{K}_{n+1}$ is required in calculating $V^{\mathrm{E}}_n(x)$; see Figure 4.4 for an illustrative diagram. Therefore, although the domain of the value function is unbounded (as the policy fund value may take values in the positive real line), it is only necessary to acquire the continuation function over some bounded support. This observation implies that different from the previous Chapters 2 and 3 there is no need to construct an auxiliary stochastic control problem. Accordingly, the truncation error (see Section 2.4 of Chapter 2) does not contribute to the overall error of the subsequent LSMC algorithm.*

For general VA products, one cannot expect a similar property holds as Part (i) of Proposition 4.3 and thus a truncation argument is indispensable before the implementation of an LSMC algorithm.

4.4.2 The LSMC Algorithm

Before presenting the details of the numerical algorithm, it is convenient to derive a closed-form expression of the value function at the penultimate year. According to the derivations in Appendix C.2, one has

$$V_{N-1}(x) = {}_{N-2}p_0 q_{N-2}(B \vee W) + {}_{N-1}p_0 \left(\tilde{G}_{N-1}(I)A \vee W\right). \tag{4.30}$$

The LSMC algorithm used for pricing the Polaris is given as follows.

1 Initiation: Set $V_{N-1}^{\mathrm{E}}(\cdot) = V_{N-1}(\cdot)$ which is given by the above Eq. (4.30). For $n = N-2, N-3, \ldots, 1$, do the two steps below.

2 Artificial Simulation and Backward Updating:

2.1 For a given I, generate an independent and identically distributed (i.i.d.) random sample $\left\{\widehat{X}_{n+1}^{(I,m)}\right\}_{m=1}^{M}$ according to (4.25).

2.2 Solve the quadratic programming problem (4.28) and obtain the corresponding numerical estimates of the continuation function and the value function according to Eqs. (4.27) and (4.29), respectively.

2.3 *Repeat Steps 2.1 and 2.2 for* $I = 0, 1, \ldots, n$.

3 Price Estimate:

3.1 Generate a random sample:

$$X_1^{(m)} := \left(P_0 e^{L_0^{(m)}}, P_0 e^{\bar{L}_0^{(m)}}, P_0, P_0, 0\right), \ m = 1, 2, \ldots, M, \tag{4.31}$$

where $\left\{\left(L_0^{(m)}, \bar{L}_0^{(m)}\right)\right\}_{m=1}^{M}$ are M i.i.d. simulated values of the log-return of the policy fund and the associated running maximum over $[t_0, t_1]$.

3.2 The no-arbitrage price of the contract with initial purchase payment P_0 is estimated by

$$V_0^{\mathrm{E}}(x_0) := \frac{1}{M} \sum_{m=1}^{M} \varphi V_1^{\mathrm{E}}\left(X_1^{(m)}\right),$$

where $x_0 = (P_0, P_0, P_0, P_0, 0)$ and $\varphi := e^{-r\Delta t}$.

Remark 4.6. *In the real contract specification, the MAWA/PIP can only take a few distinct values. For example, suppose the PH's first withdrawal time is* t_ξ*, the MAWA/PIP percentage in the "Income Option 3" of the contract [1] are prescribed as follows:*

$$G(\xi) = \begin{cases} 5\%, & \text{if } \xi = 1, 2, \ldots, 6, \\ 5.25\%, & \text{otherwise.} \end{cases} \tag{4.32}$$

Since the state variable I_n *is mainly used to determine the MAWA/PIP percentage, Step 2.3 in the above algorithm is only necessarily repeated for very few distinct values of* I*, e.g.,* $0, 1,$ *and* 7 *under the above specification.*

4.5 Numerical Experiments

4.5.1 Contract and Model Parameters

In all subsequent numerical experiments, the dynamics of the underlying policy fund is modelled as a Geometric Brownian motion, i.e., $L_n(t) = r(t - t_n) + \sigma\mathcal{B}_{t-t_n}, t \in (t_n, t_{n+1}], n \in \mathcal{I}_0$, where $\mathcal{B}_t$ is the standard Brownian motion under the risk-neutral measure. Since the joint distribution of the Brownian motion and its running maximum has explicit expression ([55, Eq. (4.1.24)]), the simulation of $\big(L_n(t), \bar{L}_n(t)\big)$ is straightforward. For the simulation of the running maximum of a wide class of Lévy processes, see [54]. In terms of the financial market parameters, the risk-free interest rate r is chosen as 0.04 and the annualized volatility σ is set to be 0.19 which approximately equals the historical volatility of the S&P 500 index between 1989 and 2008 ([63]).[9] It is worth noting that the policy fund, in general, is not the equity index and might involve alternative investments and fixed-income securities depending on the choice of the PH. Accordingly, the annualized volatility rate might be smaller than 0.19 in reality as the diversification lowers the uncertainty of the investment.

The survival and death probabilities are determined according to the DAV 2004R mortality table for a 65-year-old male from [67]. The maturity date is chosen as the Latest Annuity Date which is the 95th birthday of the PH [1, pp. 4]. Accordingly, there are $N = 30$ time steps in the backward recursion procedure. As mentioned previously, the annual insurance fee rate η is tied to the VIX and further truncated to lie between the minimal and maximal annual fee rates, with 60 bps and 220 bps, respectively [1, Appendix C]. Here, we set $\eta = 220$ bps to investigate whether the maximal fee rate charged by the insurer is sufficient to hedge the policy. The initial purchase amount of the PH, P_0, is normalized to be 1 unit. The withdrawal penalty is only charged during the first four contract years, i.e., $\kappa_n = 0\%$ for $n > 4$[1, footnote 1, pp. 7]. The MAWA/PIP percentage is set to be consistent with the "Income Option 3" of the "Polaris Income Plus Daily" rider; see the earlier Remark 4.6. Table 4.2 summarizes the parameters used in sequel numerical experiments.

4.5.2 Impact of the Maximal Degree of the Basis Function

A numerical experiment is conducted to study the sensitivity of regression estimate $C_n^{\mathrm{E}}(\cdot)$ with respect to the degree of Bernstein polynomials. We take the continuation function at time step $N - 2$ as the illustrating example due to its analytical tractability (see Eq. (C.1) in Appendix

[9] The use of historical volatility might be debatable since the no-arbitrage pricing framework requires calibrating the model to the market. However, the lack of long-term (more than ten years or even lifetime) and highly liquid derivatives renders the usual model calibration procedure–as it is adopted in the pricing of financial derivatives–also questionable.

Table 4.2: Parameters used for numerical examples.

Parameter	Value
Financial market parameters	
Volatility σ	0.19
Interest rate r	0.04
Policyholder & contract specifications	
Latest Annuity Date t_N	95th birthday of the policyholder
Attained age t_0	65th birthday of the policyholder
Mortality	DAV 2004R (65-year-old male)
Withdrawal times	Yearly
Initial investment P_0	1
Time periods N	30
Insurance fee rate η	220 bps
Withdrawal penalty κ_n	$n = 1 : 8\%,\ n = 2 : 7\%,\ n = 3 : 6\%,$ $n = 4 : 5\%,\ n > 4 : 0\%$
MAWA/PIP percentage $G(\xi)$	$1 \leq \xi \leq 6 : 5\%,\ \xi > 6 : 5.25\%$

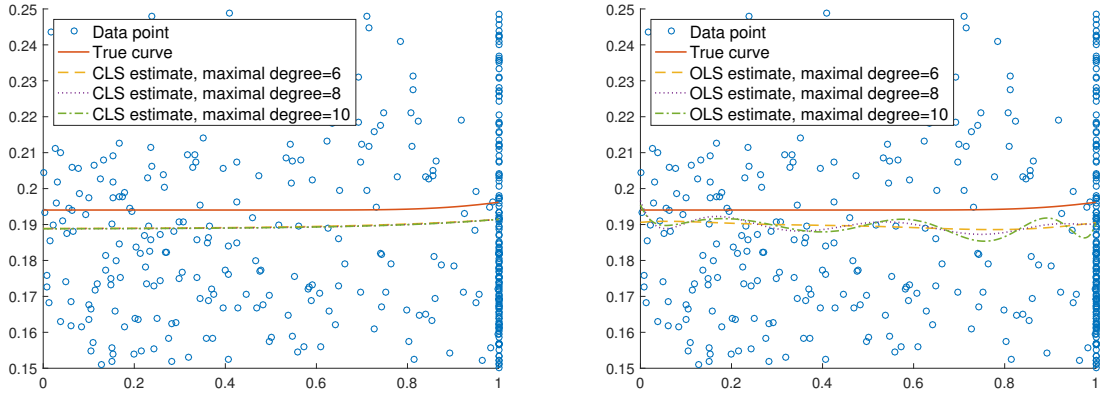
C.2). Let us focus on estimating the marginal function $k_4 \longmapsto C_{N-2}(1, 1, 1, k_4, 0)$ for the clarity of the presentation.

Specifically, simulate an i.i.d. random sample $\left\{k_4^{(m)}\right\}_{m=1}^{M}$ from the exponential distribution with mean 1 and further truncate them with the cap 1. We also generate a random sample $\left\{\left(L_{N-2}^{(m)}, \bar{L}_{N-2}^{(m)}\right)\right\}_{m=1}^{M}$ as discussed in Section 4.4. Let

$$\widetilde{X}_{N-1}^{(m)} = \left(e^{L_{N-2}^{(m)}}, 1 \vee e^{\bar{L}_{N-2}^{(m)}}, 1, k_4^{(m)}, 0\right).$$

Regress $\left\{V_{N-1}\left(\widetilde{X}_{N-1}^{(m)}\right)\right\}_{m=1}^{M}$ against $\left\{\mathbf{b}\left(1, k_4^{(m)}\right)\right\}_{m=1}^{M}$ according to the CLS regression in (4.28). We also conduct an Ordinary Least Squares (OLS) estimation by relaxing the linear constraints in (4.28) in order to compare the resulting regression estimate with that of the CLS regression. The number of simulated paths M is set to be 10^4 and the maximal degree of Bernstein polynomials is varied from 6 to 10 with an increment of 2.

Figure 4.5 depicts the regression estimates produced by these two regression methods. It is palpable that the OLS method is sensitive to J and gives economically insensible estimates which are neither convex nor monotone. Furthermore, the behaviors of these regression estimates are erratic near the boundaries of the support set. In contrast, the fitted curves of the CLS counterpart preserve the C.M. property and are insensitive to the choice of maximal degree. This confirms the advantages of the CLS method in terms of insensitivity with respect to the parameter J and mitigating overfitting/underfitting.



(a) Fitted curves using CLS method.

(b) Fitted curves using OLS method.

Figure 4.5: Fitted curves of $k_4 \longmapsto C_{N-2}\left((1,1,1,k_4,0)\right)$. The curves are fitted by using 10^4 simulated data points but only 10^3 data points are plotted in the figure for the clarity. The range of vertical axis is limited to $[0.15, 0.25]$.

4.5.3 Performance of Pricing Bounds

This subsection aims to validate the convergence of the numerical algorithm and demonstrate the performance of the pricing bounds. As discussed previously, increasing the maximal degree J brings marginal change to the numerical result. So, we set $J = 6$ and consider 4 different values (ranging from 1×10^4 to 5×10^5) for M. The "Upper bound" column of Table 4.3 gives the no-arbitrage price of the synthetic contract, i.e., $V_0(x_0)$, and the "Lower bound" column reports the time-0 objective value of the stochastic control problem with a finite feasible set $\widehat{D}_n$ as described in the last paragraph of Section 4.3. As there is no prior information about the C.M. property of the continuation function of the stochastic control problem corresponding to the real contract, the lower bound is obtained by the LSMC algorithm where the OLS regression

is conducted instead. Each result in columns "Lower bound" and "Upper bound" is obtained by repeating the LSMC algorithm 40 times. The associated sample mean and standard deviation (s.d.) are also reported.

The upper and lower pricing bounds become increasingly stable as the number of simulated paths climbs. Such a trend is more perceivable by the boxplots depicted in Figure 4.6. It is clear that the height of the boxes shrinks as the refinement level increases. Furthermore, the "Upper bound", obtained via the CLS regression, exhibits smaller standard deviation than that of the "Lower bound" which is produced by the OLS regression. This shows better finite sample performance of the CLS estimate. Finally and most importantly, the gap between pricing bounds is generally less than 3%. This indicates that the fair value of the synthetic contract can serve as a sharp upper bound of the no-arbitrage price of the real contract. It is also notable that the lower bound is generally larger than the initial purchase payment (1 unit). This reveals that the policy is underpriced in the sense that the fees received by the insurer are insufficient to hedge his financial risk exposure. Like most prevailing equity derivatives, the pricing results of Polaris are sensitive to the volatility level. The sequel subsection will investigates how sensitive the price bounds are to various model parameters including the volatility of the underlying fund.

Table 4.3: Results from validation test, data in Table 4.2. The mean and standard deviation are obtained by running the algorithm 40 times. The percentage difference is calculated as the difference between upper and lower bounds divided by the lower bound.

Refinement	Number of simulated paths	Lower bound		Upper bound		Percentage
		Mean	S.d.	Mean	S.d.	difference (%)
1	1×10^4	1.0278	0.0143	1.0409	0.0050	1.2746
2	5×10^4	1.0184	0.0051	1.0411	0.0024	2.2290
3	1.5×10^5	1.0169	0.0039	1.0408	0.0014	2.3503
4	5×10^5	1.0159	0.0021	1.0407	0.0007	2.4412

4.5.4 Sensitivity Analysis

This subsection aims to investigate the sensitivity of the pricing bounds with respect to various model parameters. In each subsequent numerical experiment, 5×10^5 sample paths are simulated. The pricing bounds are obtained via implementing the algorithm for one single time.

Let us start by exploring the effect of financial market parameters. Table 4.4 shows that the pricing bounds are rather sensitive to the perturbation of these parameters, which is in agreement with many other studies for VA products of other types, see, e.g., [40]. In general, a lower risk-free

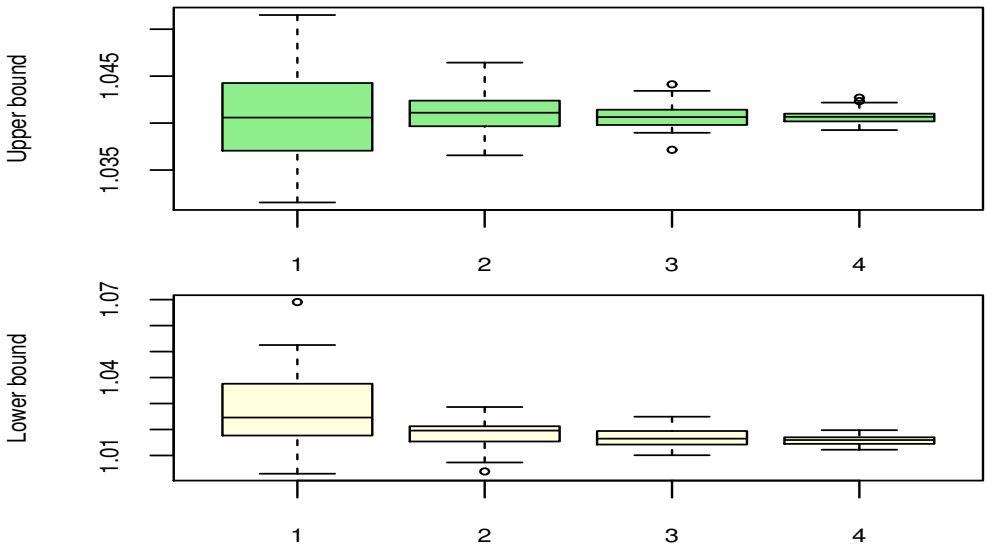


Figure 4.6: Boxplot of pricing bounds in Table 4.2. The x-axis represents the different refinement level in terms of the number of simulated paths ranging from 1×10^4 to 5×10^5; see Table 4.3 for details. The height of each box gives the discrepancy between 75th and 25th percentiles.

rate yields a higher price, whereas, the policy becomes more valuable to the PH with increasing market volatility. One possible interpretation is that a lower interest rate raises the EPV of the cash inflows of the PH and therefore increases the value of the policy. Moreover, a surge of the market volatility causes a wider spread of the investment account. Accordingly, the income base might step up to a higher level but never decrease as long as the PH does not take any excess withdrawal. This asymmetric evolution pattern of the income base exposes the insurer to larger financial risk, or equivalently, a larger value of the policy from the PH's point of view. It is also notable that the larger volatility, the wider wedge between the pricing bounds. Recall that the difference between the synthetic and real contracts arises from the discrepancy between the income base and the investment account. Intuitively, one would expect that this discrepancy is exaggerated when the policy fund price has large swings as the income base has limited downside risk while the policy fund might plunge.

Next, we consider the effects of different contract parameters on the pricing bounds. Table 4.4 shows that the price hikes as the insurance fee rate slips. This is not surprising since the investment account may remain at a higher level with less reduction of insurance fees, which in turn yields a larger cash inflow of the PH. Also observe that an annual insurance fee rate of 300

bps might be more appropriate in comparison to the prevailing rate of 220 bps levied by the insurer. Moreover, we investigate the effect of MAWA/PIP percentage. As mentioned previously, one new contract feature provided by the Polaris is the MAWA/PIP percentage depends on the PH's age at the first withdrawal. Therefore, it is interesting to compare the no-arbitrage prices of contracts with fixed MAWA/PIP percentage and the floating scheme in Table 4.2. By comparing the "MAWA/PIP percentage" panel with the "Base" case in Table 4.4, it is clear that the floating scheme of MAWA/PIP percentage adds value to the contract since the PH can enjoy a higher free-of-charge withdrawal amount in the later phase of the contract life. The last row of Table 4.4 also shows that a 25 bps drop in the MAWA/PIP percentage causes nearly 1% decline in the fair value of the contract.

Finally, we observe that the percentage difference between the upper and lower pricing bounds is generally less than 3% in most cases.

Table 4.4: Effect of parameters on pricing bounds under the numerical setting in Table 4.2, except as noted. 5×10^5 sample paths are simulated in the LSMC algorithm and the results are obtained by running the algorithm one time.

Parameter	Lower bound	Upper bound	Percentage difference (%)
Base	1.0159	1.0407	2.4412
Risk-free rate			
$r = 0.03$	1.1098	1.1390	2.6311
$r = 0.05$	0.9467	0.9630	1.7218
Volatility			
$\sigma = 0.15$	0.9582	0.9741	1.6594
$\sigma = 0.23$	1.0922	1.1270	3.1862
Rider charge rate			
260 (bps)	1.0042	1.0285	2.4198
300 (bps)	0.9907	1.0166	2.6143
MAWA/PIP percentage			
$\xi \geq 1 : 5\%$	1.0041	1.0284	2.4201
$\xi \geq 1 : 4.75\%$	0.9924	1.0123	2.0052

4.6 Concluding Remarks

This chapter established a pricing framework for "Polaris Choice IV" variable annuities. Auxiliary state and decision variables were introduced to make the mathematical discussion under a Markovian stochastic optimal control setting, and accordingly, the Dynamic Programming Principle is applicable. We considered a contract, referred to as the synthetic contract, which charges insurance fees proportional to the investment account and proved the existence of a bang-bang solution to the associated stochastic control problem. The no-arbitrage price of this synthetic contract can be approximated by an LSMC algorithm which is in the same spirit of the BSBU algorithm developed in Chapter 2. We have also shown that the price of the synthetic contract dominates that of the real contract. This upper bound has been shown to be fairly sharp by consequential numerical studies.

One crucial implication obtained from studying the present pricing problem is that the control randomization method does not necessarily guarantee the convergence of an LSMC algorithm to the optimal solution, which discloses one limitation of the forward simulation commonly adopted in the literature.

Finally, the modification of the fee structure in this chapter is essentially a modification on the transition equation of the state process, and the construction of the upper bound for the price function results from such a modification. This observation provides an idea for the construction of the upper bound of the value function for a stochastic control problem which is different from the information relaxation method commonly used in the literature; see e.g. [23].

Chapter 5

Conclusion and Future Work

5.1 Summary of the Thesis Work

This thesis studied numerical solutions to discrete-time stochastic optimal control (DTSOC) problems by integrating Monte Carlo simulation with nonparametric regression techniques.

- We constructed two different auxiliary DTSOC models whose state processes are confined into a compact domain and the accompanying optimal value functions are legitimate approximations for those of the original DTSOC problems. This not only enables one to steer by extrapolating the numerical estimate but also paves the way for applying nonparametric regression methods to approximate the value function or the continuation value. Since most numerical algorithms can barely afford to recover the value function over an unbounded domain with sufficient accuracy, the constructed auxiliary models can be expected to have wide applications.
- We proposed two Least Squares Monte Carlo (LSMC) algorithms to approach numerical solutions to general DTSOC problems. We showed that the forward simulation of the state process is not indispensable in an LSMC algorithm. We further proposed an artificial simulation method to bypass the forward simulation and control randomization which are widely used in the literature. Certain shape information of the value function/continuation function was exploited to improve the efficiency of the regression method and enhance the interpretability of the numerical estimates. Convergence results were established to build theoretical foundations of these algorithms.
- We studied the pricing problem of a complicated equity-linked insurance product, the Polaris variable annuities. With a prudent modification on the fee structure of the product, We

proved the existence of a bang-bang solution to the accompanying DTSOC model. We also disclosed the connection between the no-arbitrage price of this modified product and that of the original one. This casts new insights to pricing the variable annuity product in the absence of a bang-bang solution: one might first construct a modified contract whose pricing problem is less convoluted and then show the relation between the modified and real contracts. As a byproduct, we proposed a new way to construct upper and lower bounds for the optimal value function of a DTSOC problem, which is different from the information relaxation or duality method widely adopted in the literature and is of independent interest.

- We have not addressed how to solve the local optimization problem involved in the Bellman equation when the objective function is not convex or concave and there is no bang-bang solution. In such a situation, solving the local optimization problem is generally formidable and it is hard to expect that there exists a generic algorithm that can find the global optimizer with reasonable computational costs. This thesis focuses on responding to the problems mentioned in the introduction chapter, however, admittedly, this challenge hampers the application of the LSMC algorithms developed in the thesis. One attempt to handle this thorny issue is to discretize the feasible set of the action and run a linear search to solve the optimization problem. By doing so, one can still employ an LSMC algorithm to get a numerical estimate for the optimal value function which however is only a suboptimal solution.

5.2 Future Research Avenues

There are several avenues of research that remain to be explored.

- **Curse of Dimensionality** On one hand, it is notable from Chapters 2 and 3 that the convergence of an LSMC algorithm is ensured only when the number of simulated paths M increase at a faster rate than the number of basis function J; see in particular Condition (v) of Assumption A.1. On the other hand, when there are multiple state variables, as one may see from Chapter 4, the multivariate basis function should be generated by the tensor-product of its univariate peers, which means J surges exponentially as the dimension hikes. Combing these two observations together implies that the number of simulated paths should grow at an exponential rate with respect to the dimension of the state process in order to guarantee the convergence of the algorithm, which is computationally formidable in reality. This casts shadows to a statement in some literature that the LSMC can handle high-dimensional stochastic control problem that renders the traditional lattice-based methods impotent. This dilemma has been well acknowledged by statistics literature and is named as the "*Curse of the Dimensionality*". In view of this, it is instructive to study how one can leverage some dimension reduction techniques developed by the statistical community to mitigate this thorny problem in the context of DTSOC problems.

- **Sensitivity Calculation** It will be fruitful to numerically evaluate the sensitivities of the optimal value function of a DTSOC model. This research direction is driven by the risk management problem of the variable annuities. Although Chapter 3 addressed the delta-hedging of the variable annuity, the insurer might also be concerned with the sensitivity of his hedging portfolio with respect to various parameters such as volatility and interest rate. This naturally calls for developing a numerical algorithm to calculate the sensitivities of the optimal value function. It is notable that the optimal value function is not necessarily differentiable and therefore dedicated investigation should be carried out. One possible avenue for future research is to extend the result of [26] from the optimal stopping problem to a general DTSOC setting.

References

[1] AIG. Polaris choice iv prospectus. 2016.

[2] Yacine Ait-Sahalia and Mehmet Sağlam. High frequency market making: Optimal quoting. *Available at SSRN 2331613*, 2017.

[3] Aurélien Alfonsi, Antje Fruth, and Alexander Schied. Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2):143–157, 2010.

[4] Jennifer Alonso-García, Oliver Wood, and Jonathan Ziveyi. Pricing and hedging guaranteed minimum withdrawal benefits under a general lévy framework using the cos method. *Quantitative Finance*, 18(6):1049–1075, 2018.

[5] Donald WK Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica: Journal of the Econometric Society*, pages 307–345, 1991.

[6] Parsiad Azimzadeh and Peter A Forsyth. The existence of optimal bang-bang controls for gmxb contracts. *SIAM Journal on Financial Mathematics*, 6(1):117–139, 2015.

[7] Parsiad Azimzadeh, Peter A Forsyth, and Kenneth R Vetzal. Hedging costs for variable annuities under regime-switching. In *Hidden Markov Models in Finance*, pages 133–166. Springer, 2014.

[8] Azimzadeh, Parsiad. Impulse control in finance: Numerical methods and viscosity solutions, 2017.

[9] Alessandro Balata and Jan Palczewski. Regress-later monte carlo for optimal control of markov processes. *arXiv preprint arXiv:1712.09705*, 2017.

[10] Nicolas Baradel, Bruno Bouchard, David Evangelista, and Othmane Mounjid. Optimal inventory management and order book modeling. *arXiv preprint arXiv:1802.08135*, 2018.

[11] Christophe Barrera-Esteve, Florent Bergeret, Charles Dossal, Emmanuel Gobet, Asma Meziou, Rémi Munos, and Damien Reboul-Salze. Numerical methods for the pricing of swing options: a stochastic control approach. *Methodology and Computing in Applied Probability*, 8(4):517–540, 2006.

[12] Daniel Bauer, Jin Gao, Thorsten Moenig, Eric R Ulm, and Nan Zhu. Policyholder exercise behavior in life insurance: The state of affairs. *North American Actuarial Journal*, pages 1–17, 2017.

[13] Daniel Bauer and Hongjun Ha. A least-squares monte carlo approach to the calculation of capital requirements. *Department of Risk Management and Insurance, Georgia State University*, 2013.

[14] Daniel Bauer, Alexander Kling, and Jochen Russ. A universal pricing framework for guaranteed minimum benefits in variable annuities. *Astin Bulletin*, 38(02):621–651, 2008.

[15] Denis Belomestny. Pricing bermudan options by nonparametric regression: optimal rates of convergence for lower estimates. *Finance and Stochastics*, 15(4):655–683, 2011.

[16] Denis Belomestny, Anastasia Kolodko, and John Schoenmakers. Regression methods for stochastic control problems and their convergence analysis. *SIAM Journal on Control and Optimization*, 48(5):3562–3588, 2010.

[17] Denis Belomestny, Grigori Milstein, and Vladimir Spokoiny. Regression methods in pricing american and bermudan options using consumption processes. *Quantitative Finance*, 9(3):315–327, 2009.

[18] Claude Berge. *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*. Courier Corporation, 1997.

[19] Eric Beutner, Antoon Pelsser, and Janina Schweizer. Fast convergence of regress-later estimates in least squares monte carlo. *Available at SSRN 2328709*, 2013.

[20] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[21] Mark Broadie and Paul Glasserman. Estimating security price derivatives using simulation. *Management science*, 42(2):269–285, 1996.

[22] Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional american options. *Journal of Computational Finance*, 7:35–72, 2004.

[23] David B Brown, James E Smith, and Peng Sun. Information relaxations and duality in stochastic dynamic programs. *Operations Research*, 58(4-part-1):785–801, 2010.

[24] René Carmona and Michael Ludkovski. Valuation of energy storage: An optimal switching approach. *Quantitative Finance*, 10(4):359–374, 2010.

[25] Jacques F Carriere. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19(1):19–30, 1996.

[26] Nan Chen and Yanchu Liu. American option sensitivities estimation via a generalized infinitesimal perturbation analysis approach. *Operations Research*, 62(3):616–632, 2014.

[27] Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, 6:5549–5632, 2007.

[28] Zhang Chen, Ken Vetzal, and Peter A Forsyth. The effect of modelling parameters on the value of gmwb guarantees. *Insurance: Mathematics and Economics*, 43(1):165–173, 2008.

[29] Jaehyuk Choi, Chenru Liu, and Jeechul Woo. An efficient approach for removing look-ahead bias in the least square monte carlo algorithm: Leave-one-out. *arXiv preprint arXiv:1810.02071*, 2018.

[30] Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for american option pricing. *Finance and Stochastics*, 6(4):449–471, 2002.

[31] Fei Cong and Cornelis W Oosterlee. Multi-period mean–variance portfolio optimization based on monte-carlo simulation. *Journal of Economic Dynamics and Control*, 64:23–38, 2016.

[32] Zhenyu Cui, Runhuan Feng, and Anne MacKay. Variable annuities with vix-linked fee structure under a heston-type stochastic volatility model. *North American Actuarial Journal*, 21(3):458–483, 2017.

[33] Min Dai, Yue Kuen Kwok, and Jianping Zong. Guaranteed minimum withdrawal benefit in variable annuities. *Mathematical Finance*, 18(4):595–611, 2008.

[34] Pierre Del Moral, Bruno Rémillard, and Sylvain Rubenthaler. Monte carlo approximations of american options that preserve monotonicity and convexity. In *Numerical Methods in Finance*, pages 115–143. Springer, 2012.

[35] Daniel Egloff. Monte carlo algorithms for optimal stopping and statistical learning. *The Annals of Applied Probability*, 15(2):1396–1432, 2005.

[36] Daniel Egloff, Michael Kohler, and Nebojsa Todorovic. A dynamic look-ahead monte carlo algorithm for pricing bermudan options. *The Annals of Applied Probability*, 17(4):1138–1171, 2007.

[37] Theodoros Evgeniou, Massimiliano Pontil, and Tomaso Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13(1):1, 2000.

[38] Jianqing Fan and Irene Gijbels. *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*, volume 66. CRC Press, 1996.

[39] Runhuan Feng and Yasutaka Shimizu. Applications of central limit theorems for equity-linked insurance. *Insurance: Mathematics and Economics*, 69:138–148, 2016.

[40] Peter Forsyth and Kenneth Vetzal. An optimal stochastic control framework for determining the cost of hedging of variable annuities. *Journal of Economic Dynamics and Control*, 44:29–53, 2014.

[41] Jin Gao and Eric R Ulm. Optimal consumption and allocation in variable annuities with guaranteed minimum death benefits. *Insurance: Mathematics and Economics*, 51(3):586–598, 2012.

[42] Jin Gao and Eric R Ulm. Optimal allocation and consumption with guaranteed minimum death benefits, external income and term life insurance. *Insurance: Mathematics and Economics*, 61:87–98, 2015.

[43] Paul Glasserman and Bin Yu. Number of paths versus number of basis functions in american option pricing. *The Annals of Applied Probability*, 14(4):2090–2119, 2004.

[44] Paul Glasserman and Bin Yu. Simulation for american options: Regression now or regression later? In *Monte Carlo and Quasi-Monte Carlo Methods 2002*, pages 213–226. Springer, 2004.

[45] Mary R Hardy. Hedging and reserving for single-premium segregated fund contracts. *North American Actuarial Journal*, 4(2):63–74, 2000.

[46] Vanya Horneff, Raimond Maurer, Olivia S Mitchell, and Ralph Rogalla. Optimal life cycle portfolio choice with variable annuities offering liquidity and investment downside protection. *Insurance: Mathematics and Economics*, 63:91–107, 2015.

[47] Yao Tung Huang and Yue Kuen Kwok. Regression-based monte carlo methods for stochastic control models: Variable annuities with lifelong guarantees. *Quantitative Finance*, 16(6):905–928, 2016.

[48] Yao Tung Huang, Pingping Zeng, and Yue Kuen Kwok. Optimal initiation of guaranteed lifelong withdrawal benefit with dynamic withdrawals. *SIAM Journal on Financial Mathematics*, 8(1):804–840, 2017.

[49] Côme Huré, Huyên Pham, Achref Bachouch, and Nicolas Langrené. Deep neural networks algorithms for stochastic control problems on finite horizon, part i: convergence analysis. *arXiv preprint arXiv:1812.04300*, 2018.

[50] IRI. Second-quarter 2017 annuity sales report. 2017.

[51] Idris Kharroubi, Nicolas Langrené, and Huyên Pham. A numerical algorithm for fully nonlinear hjb equations: an approach by control randomization. *Monte Carlo Methods and Applications*, 20(2):145–165, 2014.

[52] Christian Knoller, Gunther Kraut, and Pascal Schoenmaekers. On the propensity to surrender a variable annuity contract: an empirical analysis of dynamic policyholder behavior. *Journal of Risk and Insurance*, 83(4):979–1006, 2016.

[53] Michael Kohler, Adam Krzyżak, and Nebojsa Todorovic. Pricing of high-dimensional american options by neural networks. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 20(3):383–410, 2010.

[54] Alexey Kuznetsov, Andreas E Kyprianou, Juan C Pardo, and Kees van Schaik. A wiener–hopf monte carlo simulation technique for lévy processes. *The Annals of Applied Probability*, 21(6):2171–2190, 2011.

[55] Yue Kuen Kwok. *Mathematical models of financial derivatives*. Berlin: Springer, 2008.

[56] Ker-Chau Li. Asymptotic optimality for cp, cl, cross-validation and generalized cross-validation: discrete index set. *The Annals of Statistics*, pages 958–975, 1987.

[57] Han Liu. Nonparametric Learning in High Dimensions. 12 2010.

[58] Francis A Longstaff and Eduardo S Schwartz. Valuing american options by simulation: a simple least-squares approach. *Review of Financial studies*, 14(1):113–147, 2001.

[59] Xiaolin Luo and Pavel V Shevchenko. Valuation of variable annuities with guaranteed minimum withdrawal and death benefits via stochastic control optimization. *Insurance: Mathematics and Economics*, 62:5–15, 2015.

[60] Rahul Mazumder, Arkopal Choudhury, Garud Iyengar, and Bodhisattva Sen. A computational framework for multivariate convex regression and its variants. *Journal of the American Statistical Association*, 114(525):318–331, 2019.

[61] Mary C. Meyer. Inference using shape-restricted regression splines. *The Annals of Applied Statistics*, 2:1013–1033, 2008.

[62] Moshe A Milevsky and Thomas S Salisbury. Financial valuation of guaranteed minimum withdrawal benefits. *Insurance: Mathematics and Economics*, 38(1):21–38, 2006.

[63] Thorsten Moenig and Daniel Bauer. Revisiting the risk-neutral approach to optimal policyholder behavior: a study of withdrawal guarantees in variable annuities. *Review of Finance*, page rfv018, 2015.

[64] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

[65] Whitney K Newey. Convergence rates and asymptotic normality for series estimators. *Journal of econometrics*, 79(1):147–168, 1997.

[66] Anna A Obizhaeva and Jiang Wang. Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1):1–32, 2013.

[67] Ulrich Pasdika and Jürgen Wolff. Coping with longevity: The new german annuity valuation table dav 2004 r. *The Living to 100 and Beyond Symposium*, 2(4), 2005.

[68] Evan L Porteus. Stochastic inventory theory. *Handbooks in operations research and management science*, 2:605–652, 1990.

[69] Tim Robertson, F.T. Wright, and R.L. Dykstra. *Order restricted statistical inference*. John Wiley and Sons, 1988.

[70] LCG Rogers. Pathwise stochastic optimal control. *SIAM Journal on Control and Optimization*, 46(3):1116–1132, 2007.

[71] David Ruppert, Matt P Wand, and Raymond J Carroll. *Semiparametric regression*. Number 12. Cambridge University Press, 2003.

[72] Zhiyi Shen and Chengguo Weng. Pricing bounds and bang-bang analysis of polaris variable annuities. Working paper of University of Waterloo, Available at SSRN: https://ssrn.com/abstract=3056794, 2017.

[73] Zhiyi Shen and Chengguo Weng. A backward simulation method for stochastic optimal control problems. *arXiv preprint arXiv:1901.06715*, 2019.

[74] Pavel V Shevchenko and Xiaolin Luo. A unified pricing of variable annuity guarantees under the optimal stochastic control framework. *Risks*, 4(3):1–31, 2016.

[75] Steven E Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.

[76] Bernard W Silverman. Spline smoothing: the equivalent variable kernel method. *The Annals of Statistics*, 12(3):898–916, 1984.

[77] Petra Steinorth and Olivia S Mitchell. Valuing variable annuities with guaranteed minimum lifetime withdrawal benefits. *Insurance: Mathematics and Economics*, 64:246–258, 2015.

[78] Lars Stentoft. Convergence of the least squares monte carlo approach to american option valuation. *Management Science*, 50(9):1193–1203, 2004.

[79] Nizar Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29. Springer Science & Business Media, 2012.

[80] John N Tsitsiklis. Periodic review inventory systems with continuous demand and discrete order sizes. *Management science*, 30(10):1250–1254, 1984.

[81] John N Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex american-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.

[82] Gerry Tsoukalas, Jiang Wang, and Kay Giesecke. Dynamic portfolio execution. *Management Science*, 2017.

[83] Jiangdian Wang and Sujit K Ghosh. Shape restricted nonparametric regression based on multivariate bernstein polynomials. Technical report, North Carolina State University. Dept. of Statistics, 2012.

[84] Jiangdian Wang and Sujit K Ghosh. Shape restricted nonparametric regression with bernstein polynomials. *Computational Statistics and Data Analysis*, 56(9):2729–2741, 2012.

[85] Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.

[86] Daniel Z Zanger. Convergence of a least-squares monte carlo algorithm for bounded approximating sets. *Applied Mathematical Finance*, 16(2):123–150, 2009.

[87] Daniel Z Zanger. Quantitative error estimates for a least-squares monte carlo algorithm for american option pricing. *Finance and Stochastics*, 17(3):503–534, 2013.

[88] Daniel Z Zanger. Convergence of a least-squares monte carlo algorithm for american option pricing with dependent sample data. *Mathematical Finance*, 28(1):447–479, 2018.

[89] Rongju Zhang, Nicolas Langrené, Yu Tian, Zili Zhu, Fima Klebaner, and Kais Hamza. Dynamic portfolio optimization with liquidity cost and market impact: A simulation-and-regression approach. *arXiv preprint arXiv:1610.07694*, 2016.

APPENDICES

Appendix A

Appendix for Chapter 2

A.1 Supplements for Sieve Estimation Method

A.1.1 Forms of Matrix $\mathbf{A}_J$

Below we collect several forms of the constraint matrix $\mathbf{A}_J$ in (2.29) which ensures monotonicity, convexity, or concavity of the sieve estimate (2.30) if the basis function $\boldsymbol{\phi}(z)$ is a vector of Bernstein polynomials. For the simplicity of notation, we only consider the univariate case, i.e., $z \in \mathbb{R}$.

It is worth noting that the Bernstein polynomials are solely defined on the unit interval and thus one may first normalize z in defining $\boldsymbol{\phi}(z)$. To be specific, when the domain of the conditional mean function $g(\cdot)$ is $\mathcal{Z} = [0, R]$, we choose

$$\boldsymbol{\phi}(z) = \mathbf{b}\big(z/R\big) = \big(b_{J,0}\big(z/R\big), \ldots, b_{J,J}\big(z/R\big)\big)^{\intercal}, \quad z \in [0, R], \tag{A.1}$$

where

$$b_{J,j}(v) = \binom{J}{j} v^j (1-v)^{J-j} = \sum_{\ell=j}^{J} (-1)^{\ell-j} \binom{J}{\ell} \binom{\ell}{j} v^{\ell}, \quad v \in [0, 1]. \tag{A.2}$$

Monotonicity Suppose the conditional mean $g(\cdot)$ defined in Eq. (2.26) is monotone. Then the corresponding monotonicity-preserved sieve estimate $\hat{g}(\cdot)$ is obtained from (2.30) with

$\mathcal{H}_J$ and $\boldsymbol{\phi}(z)$ given by Eqs. (2.29) and (A.1), respectively, and

$$\mathbf{A}_J = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ & & \ddots & & \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}_{J\times(J+1)}.$$

Convexity/Concavity If $g(\cdot)$ is convex, we choose the matrix $\mathbf{A}_J$ as

$$\mathbf{A}_J = \begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ & & \ddots & & & \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}_{(J-1)\times(J+1)}.$$

Moreover, the matrix $\mathbf{A}_J$ accompanying a concave $g(\cdot)$ is obtained by taking negative of the above matrix.

Convexity and Monotonicity If $g(\cdot)$ is convex and monotone, the corresponding $\mathbf{A}_J$ is given by

$$\mathbf{A}_J = \begin{pmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ & & \ddots & & & \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}_{J\times(J+1)}.$$

A.1.2 A Data-driven Choice of J

Below we present some common methods of choosing the number of basis functions J in a sieve estimation method; see, e.g., [56].

Mallows's C_p For a discrete set $\mathcal{J} \subseteq \mathbb{N}$, J is determined by solving the following minimization problem:

$$\hat{J} = \arg\min_{J\in\mathcal{J}} \frac{1}{M}\sum_{m=1}^{M}\left[U^{(m)} - \hat{g}\left(Z^{(m)}\right)\right]^2 + 2\hat{\sigma}^2(J/M),$$

where $\hat{g}(\cdot)$ is given in (2.30) and $\hat{\sigma}^2 := M^{-1}\sum_{m=1}^{M}\left[U^{(m)} - \hat{g}\left(Z^{(m)}\right)\right]^2$ which is an estimate for the variance of residual term.

Generalized cross-validation J is determined by

$$\hat{J} = \arg\min_{J \in \mathcal{J}} \frac{M^{-1} \sum_{m=1}^{M} \left[U^{(m)} - \hat{g}\left(Z^{(m)} \right) \right]^2}{\left(1 - \left(J/M \right) \right)^2},$$

with $\hat{g}(\cdot)$ given in (2.30).

Leave-one-out cross-validation Select J to minimize

$$\mathrm{CV}(J) := \frac{1}{M} \sum_{m=1}^{M} \left[U^{(m)} - \hat{g}_{-m}\left(Z^{(m)} \right) \right]^2,$$

where $\hat{g}_{-m}(\cdot)$ is similarly obtained by Eq. (2.30) with the sample point $\left(U^{(m)}, Z^{(m)} \right)$ removed.

A.1.3 Technical Assumption of Sieve Estimation Method

We impose the following assumption accompanying the sieve estimation method discussed in Section 2.4.3 which follows from [65].

Assumption A.1. **(i)** $\left\{ \left(U^{(m)}, Z^{(m)} \right) \right\}_{m=1}^{M}$ *are i.i.d. and* $Z^{(m)}$ *has compact support* $\mathcal{Z}$*. Furthermore,* $\mathbb{V}\mathrm{ar}\left[U^{(m)} \middle| Z^{(m)} = \cdot \right]$ *is bounded over* $\mathcal{Z}$*.*

(ii) *There exists a sequence* $\Upsilon(J)$ *such that* $\|\boldsymbol{\phi}\| \leq \Upsilon(J)$ *with* $\|\cdot\|$ *denoting the supremum norm of a continuous function over* $\mathcal{Z}$*.*

(iii) *For the sieve space* $\mathcal{H}_J$ *defined either in Eq. (2.28) or Eq. (2.29), there exists a* $(J+1)$*-by-1 vector* $\tilde{\boldsymbol{\beta}}$ *and a sequence* ρ_J *such that* $\rho_J \longrightarrow 0$ *as* $J \longrightarrow \infty$*, and*

$$\inf_{h(\cdot) \in \mathcal{H}_J} \|h - g\| = \left\| \tilde{\boldsymbol{\beta}}^{\intercal} \boldsymbol{\phi} - g \right\| = O\left(\rho_J \right), \tag{A.3}$$

where one should recall that $g(\cdot) := \mathbb{E}\left[U^{(m)} \middle| Z^{(m)} = \cdot \right]$*.*

(iv) *Let* $\Phi := \mathbb{E}\left[\boldsymbol{\phi}\left(Z^{(m)} \right) \boldsymbol{\phi}^{\intercal}\left(Z^{(m)} \right) \right]$*. There exists a positive constant* $\underline{c}_{\Phi}$ *independent of* J *such that* $0 < \underline{c}_{\Phi} \leq \lambda_{\min}(\Phi) \leq \lambda_{\max}(\Phi) \leq \bar{c}_{\Phi} < \infty$*, with* $\lambda_{\min}(\Phi)$ *and* $\lambda_{\max}(\Phi)$ *denoting the smallest and largest eigenvalues of* Φ*, respectively.*

(v) *As* $M \longrightarrow \infty$*,* $J \longrightarrow \infty$*, and* $\Upsilon^2(J) J/M \longrightarrow 0$*.*

We give some comments on the above technical conditions.

1. The i.i.d. condition in Part (i) of the above assumption pinpoints the necessity of generating an independent sample at each time step in an LSMC algorithm; see also the discussion in the earlier item "Cost of forward simulation" of Section 2.2.2. Part (i) further requires $Z^{(m)}$ has a compact support, which is conventional in statistics literature see, e.g., [65] and [27]. In the context of BSBU algorithm, this shows that restraining the state process into a bounded domain is not only beneficial in eliminating undesirable extrapolation but also indispensable in guaranteeing the convergence of the regression estimate to the continuation function. This has also been pointed out in the literature of LSMC, see, e.g., [78] and [87].

2. Part (ii) specifies how the magnitude of $\boldsymbol{\phi}(\cdot)$ is amplified as the number of basis function functions grows up. In particular, [65] shows that $\Upsilon(J) = O\left(\sqrt{J}\right)$ for B-splines and $\Upsilon(J) = O(J)$ for power series; for the cases of other types of basis functions, we refer to [27].

3. Part (iii) states that there exists a function $\tilde{\boldsymbol{\beta}}^{\intercal}\boldsymbol{\phi}(\cdot)$ in the sieve space $\mathcal{H}_J$ that "best" approximates the conditional mean function $g(\cdot)$ under the supremum norm; see Figure A.1 for a graphical illustration. The existence of $\tilde{\boldsymbol{\beta}}$ (referred to as *oracle*) is guaranteed by the convexity of sieve space $\mathcal{H}_J$. For the sieve space (2.29), the existence of ρ_J relies on the convexity, concavity or monotonicity of the function $g(\cdot)$ which follows by the Property 3.2 of [84]. Figure A.1 depicts the relationship between $\tilde{\boldsymbol{\beta}}^{\intercal}\boldsymbol{\phi}(\cdot)$ and $g(\cdot)$: their discrepancy vanishes as J increases and, for a fixed J, the sieve estimate $\hat{\boldsymbol{\beta}}^{\intercal}\boldsymbol{\phi}(\cdot)$ converges to $\tilde{\boldsymbol{\beta}}^{\intercal}\boldsymbol{\phi}(\cdot)$ as the sample size M approaches infinity. Therefore, one may view the sieve estimation as a two-stage approximation for the conditional mean function $g(\cdot)$.

4. The condition in Part (iv) ensures the design matrix of the regression problem is nonsingular with a high probability and does not blow up as J approaches infinity. Finally, Part (v) prescribes the growth rates of J and M in order to avoid overfitting or underfitting.

A.2 Supplements for Section 2.5

A.2.1 Verification of Assumptions

In this subsection we verify the Assumptions 2.2 and 2.3 in the context of Section 2.5. In particular, we will exemplify the terms $\mathcal{E}_T(x_0, R)$, $\xi(R)$, and ζ in order to give an explicit expression for the error estimate in (2.20).

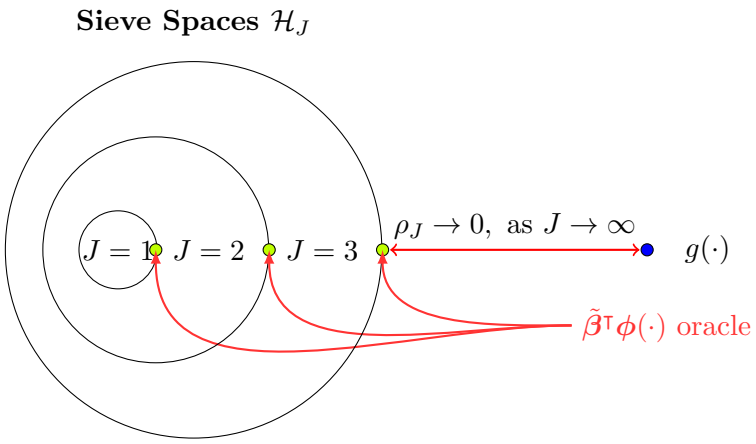


Figure A.1: A diagram illustrating the relationship between oracle $\tilde{\boldsymbol{\beta}}^\intercal \boldsymbol{\phi}(\cdot)$ and $g(\cdot)$.

Verification of Assumption 2.2

Recall from Section 2.5 that $\mathcal{X} = [0, \infty) \times \mathcal{T}_0$ and $\mathcal{X}_R = [0, R) \times \mathcal{T}_0$. It is also notable that X and $\tilde{X}$ has one common absorbing state $\{0\}$. In view of these, we observe that

$$\Big\{X_t = \tilde{X}_t \ \text{ for all } \ t \in \mathcal{T}\Big\}^c \subseteq \Big\{X_t \in \mathcal{X}_R \ \text{ for all } \ t \in \mathcal{T}\Big\}^c = \Big\{\max_{t \in \mathcal{T}} W_t \geq R\Big\}, \tag{A.4}$$

where one should recall that $X_t = (W_t, I_t)$.

We recall that the underlying asset price follows a Geometric Brownian Motion with drift and volatility rates $r - q$ and σ, respectively. Suppose the price process is driven by a standard Brownian Motion $\{\mathcal{B}_u\}_{u \in [0, T\delta]}$. We define a continuous time process $Y := \{Y_u\}_{u \in [0, T\delta]}$ as follows:

$$Y_u = w_0 e^{(r - q - \sigma^2/2)u + \sigma \mathcal{B}_u} = w_0 e^{\mu u + \sigma \mathcal{B}_u}. \tag{A.5}$$

In view of the above equation and the transition equation accompanying W_t (Eq. (2.34)), we get

$Y_{t\delta} \geq W_t$ for all $t \in \mathcal{T}$ regardless of the PH's withdrawal strategy. And therefore,

$$\begin{aligned}
\sup_{\mathsf{a}\in\mathcal{A}} \mathbb{P}\left[\max_{t\in\mathcal{T}} W_t \geq R\right] &\leq \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} Y_u > R\right] \\
&= \mathbb{P}\left[w_0 \exp\left(\sup_{u\in[0,\tilde{T}]} [\mu u + \sigma \mathcal{B}_u]\right) > R\right] \\
&= \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} \left[\mu u + \sigma \mathcal{B}_u\right] > \log\left(R/w_0\right)\right] \\
&= \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} \left[\alpha u + \mathcal{B}_u\right] > m\right],
\end{aligned}$$

with $\alpha := \mu/\sigma = (r - q - \sigma^2/2)/\sigma$, $\tilde{T} := T\delta$, and $m := \sigma^{-1}\log\left(R/w_0\right)$. Combing the above display with (A.4), we get

$$\begin{aligned}
\inf_{\mathsf{a}\in\mathcal{A}} \mathbb{P}\Big[X_t = \tilde{X}_t \ \text{ for all } \ t \in \mathcal{T}\Big] &\geq \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} \left[\alpha u + \mathcal{B}_u\right] > m\right] \\
&= 1 - \mathcal{N}\left(\frac{m - \alpha\tilde{T}}{\sqrt{\tilde{T}}}\right) + e^{2\alpha m}\mathcal{N}\left(\frac{-m - \alpha\tilde{T}}{\sqrt{\tilde{T}}}\right) \\
&:= \mathcal{E}_T(x_0, R), \qquad \text{(A.6)}
\end{aligned}$$

where $\mathcal{N}(\cdot)$ denotes the cdf of a standard normal random variable and the last equality follows by the cdf of the continuous running maximum of a Brownian Motion with nonzero drift; see [75, Corollary 7.2.2, pp. 297].

Verification of Assumption 2.3

Denote $x = (w, I)$ and $a = (\gamma, \tau)$. Recall that in the context of Section 2.5 we have $f_t(x, a) = \gamma - \kappa(\gamma - g_t(I)w_0)^+$. Further recall that $A_t(x) = [0, g_t(I)w_0 \vee w]$ and thus,

$$\sup_{a\in A_t(x)} |f_t(x, a)|^2 \leq (w \vee g_t(I)w_0)^2 \leq 2\left(w^2 + G_0^2\right) := B(x) \qquad \text{(A.7)}$$

with $G_0 := \max_{t\in\mathcal{T}_0}\left[\max_{0\le I\le t-1} g_t(I)\right]$. Consequently, we get

$$\sup_{x\in\mathrm{cl}(\mathcal{X}_R)}\left(\sup_{a\in A_t(x)}|f_t(x,a)|^2\right)\le B^2(R)=2\left(R^2+G_0^2\right):=\xi(R), \tag{A.8}$$

and

$$\sup_{\mathsf{a}\in\mathcal{A}}\mathbb{E}\left[B\left(X_t\right)\right]=2\left(G_0^2+\sup_{\mathsf{a}\in\mathcal{A}}\mathbb{E}\left[W_t^2\right]\right)\le 2\Big(G_0^2+\mathbb{E}\left[Y_{t\delta}^2\right]\Big)=2\left[G_0^2+e^{2\left(\mu+\sigma^2\right)\tilde{T}}\right]:=\zeta_X. \tag{A.9}$$

Combing Eqs. (A.6)–(A.9) together, one may get an explicit expression for the error bound in (2.20).

A.3 Proofs of Statements

A.3.1 Proof of Proposition 2.1

Preliminary

Lemma A.1. *Let*

$$\bar{\mathcal{A}}:=\left\{\mathsf{a}=\{\bar{a}_t\}_{t\in\mathcal{T}_0}\;\middle|\; a_t \text{ is } \mathcal{F}_t\text{-measurable and } a_t\in\mathbb{A} \text{ for } t\in\mathcal{T}_0\right\}.$$

For each $\mathsf{a}\in\bar{\mathcal{A}}$*, the following statements hold:*

(i) $\left\{\tau^R\le t\right\}=\left\{\tilde{X}_t\in\partial\mathcal{X}_R\right\}$ *for* $t=1,2,\ldots,T$*;*

(ii) $\left\{\tau^R=t+1\right\}=\left\{\tilde{X}_t\in\mathring{\mathcal{X}}_R,\ S\left(\tilde{X}_t,a_t,\varepsilon_{t+1}\right)\notin\mathring{\mathcal{X}}_R\right\}$ *for* $t=0,1,\ldots,T-1$*,*

where τ^R *and* $\tilde{X}_t$ *are defined in Eqs.* (2.11) *and* (2.12)*, respectively.*

Proof of Lemma A.1. **(i)** According to Eq. (2.12), we observe

$$\begin{aligned}\left\{\tilde{X}_t\in\partial\mathcal{X}_R\right\} &= \left\{X_t\in\partial\mathcal{X}_R,\tau^R>t\right\}\cup\left\{\mathcal{Q}\left(X_{\tau^R\wedge t}\right)\in\partial\mathcal{X}_R,\tau^R\le t\right\}\\ &= \left\{\mathcal{Q}\left(X_{\tau^R\wedge t}\right)\in\partial\mathcal{X}_R,\tau^R\le t\right\},\end{aligned}$$

where the second identity is by the definition of the stopping time τ^R and the fact that $\partial\mathcal{X}_R\cap\mathring{\mathcal{X}}_R=\emptyset$. To show the statement in Part (i) of Lemma A.1, it suffices to prove

$\{\tau^R \leq t\} \subseteq \{\mathcal{Q}(X_{\tau^R \wedge t}) \in \partial \mathcal{X}_R\}$. Indeed, $\tau^R \leq t$ implies $X_{\tau^R \wedge t} \notin \mathring{\mathcal{X}}_R$, and thus $\mathcal{Q}(X_{\tau^R \wedge t}) \in \partial \mathcal{X}_R$.

(ii) In view of Part (i) and Eq. (2.12), we obtain

$$\left\{\tilde{X}_t \in \mathring{\mathcal{X}}_R\right\} = \left\{\tilde{X}_t \in \partial \mathcal{X}_R\right\}^{\mathrm{c}} = \{\tau^R > t\} \subseteq \left\{\tilde{X}_t = X_t\right\}.$$

Therefore, we obtain

$$\begin{aligned}
\left\{\tilde{X}_t \in \mathring{\mathcal{X}}_R,\ S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right) \notin \mathring{\mathcal{X}}_R\right\} &= \left\{\tilde{X}_t \in \mathring{\mathcal{X}}_R,\ \tilde{X}_t = X_t,\ S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right) \notin \mathring{\mathcal{X}}_R,\ \tau^R > t\right\} \\
&= \left\{X_t \in \mathring{\mathcal{X}}_R,\ S\left(X_t, a_t, \varepsilon_{t+1}\right) \notin \mathring{\mathcal{X}}_R,\ \tau^R > t\right\} \\
&= \left\{X_t \in \mathring{\mathcal{X}}_R,\ X_{t+1} \notin \mathring{\mathcal{X}}_R,\ \tau^R > t\right\} = \{\tau^R = t+1\}.
\end{aligned}$$

This proves Part (ii) of Lemma A.1.

□

Proof of the Main Result

Proof of Proposition 2.1. By exploiting Lemma A.1 and Eq. (2.12), we get

$$\begin{aligned}
\tilde{X}_{t+1} &= X_{t+1}\mathbb{1}_{\{\tau^R > t+1\}} + \mathcal{Q}\left(X_{\tau^R \wedge (t+1)}\right)\mathbb{1}_{\{\tau^R \leq t+1\}} \\
&= X_{t+1}\mathbb{1}_{\{\tau^R > t+1\}} + \mathcal{Q}\left(X_{\tau^R \wedge t}\right)\mathbb{1}_{\{\tau^R \leq t\}} + \mathcal{Q}\left(X_{t+1}\right)\mathbb{1}_{\{\tau^R = t+1\}} \\
&= S\left(X_t, a_t, \varepsilon_{t+1}\right)\mathbb{1}_{\{\tau^R > t+1\}} + \mathcal{Q}\left(X_{\tau^R \wedge t}\right)\mathbb{1}_{\{\tau^R \leq t\}} \\
&\quad + \mathcal{Q}\left(S\left(X_t, a_t, \varepsilon_{t+1}\right)\right)\mathbb{1}_{\{\tau^R = t+1\}} \\
&= S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right)\mathbb{1}_{\{\tau^R > t+1\}} + \tilde{X}_t\mathbb{1}_{\{\tau^R \leq t\}} \\
&\quad + \mathcal{Q}\left(S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right)\right)\mathbb{1}_{\{\tau^R = t+1\}} \\
&= S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right)\mathbb{1}_{\{\tau^R > t+1\}} + \tilde{X}_t\mathbb{1}_{\{X_t \in \partial \mathcal{X}_R\}} \\
&\quad + \mathcal{Q}\left(S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right)\right)\mathbb{1}_{\left\{\tilde{X}_t \in \mathring{\mathcal{X}}_R,\ S\left(\tilde{X}_t, a_t, \varepsilon_{t+1}\right) \notin \mathring{\mathcal{X}}_R\right\}},
\end{aligned}$$

where the fourth equality follows from Eq. (2.12) and the last equality follows from Lemma A.1.

The above equation together with Eqs. (2.5) and (2.14) yields Eq. (2.13). This completes the proof. □

A.3.2 Proof of Theorem 2.1

Preliminary

Recall that X and $\tilde{X}$ implicitly depend on certain actions a; see Eqs. (2.1) and (2.13), respectively. In the sequel, we sometimes stress such dependency by writing $X_t(\mathsf{a})$ (resp. $\tilde{X}_t(\mathsf{a})$) and $X(\mathsf{a})$ (resp. $\tilde{X}(\mathsf{a})$).

Lemma A.2. *For the state process $\tilde{X}$ defined through Eq. (2.13), the following statements hold.*

(i) *For each $\mathsf{a} \in \mathcal{A}$, there exists $\tilde{\mathsf{a}} \in \tilde{\mathcal{A}}$ such that $\tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}})$ for all $t \in \mathcal{T}$ almost surely.*

(ii) *For each $\tilde{\mathsf{a}} \in \tilde{\mathcal{A}}$, there exists $\mathsf{a} \in \mathcal{A}$ such that $\tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}})$ for all $t \in \mathcal{T}$ almost surely.*

Proof of Lemma A.2. **(i)** Given $\mathsf{a} \in \mathcal{A}$ and $\tilde{X}(\mathsf{a})$, we construct $\tilde{\mathsf{a}}$ as follows: $\tilde{a}_0 = a_0$, and

$$\tilde{a}_t = a_t \mathbb{1}_{\left\{\tilde{X}_t(\mathsf{a}) \in \mathring{\mathcal{X}}_R\right\}} + a_t^*\left(\tilde{X}_t(\mathsf{a})\right) \mathbb{1}_{\left\{\tilde{X}_t(\mathsf{a}) \in \partial\mathcal{X}_R\right\}}, \tag{A.10}$$

for $t = 1, 2, \ldots, T-1$, where $a_t^*(\cdot)$ is any measurable function such that $a_t^*(x) \in A_t(x)$.

It is easy to see from the above construction that $\tilde{\mathsf{a}}$ is $\mathcal{F}$-adapted. It remains to show that

$$\tilde{a}_t \in A_t\left(\tilde{X}_t(\tilde{\mathsf{a}})\right) \quad \text{and} \quad \tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}), \quad \text{for} \quad t \in \mathcal{T}. \tag{A.11}$$

Firstly, we observe $\tilde{a}_0 = a_0 \in A_0(X_0)$ and $\tilde{X}_0(\tilde{\mathsf{a}}) = x_0$. As induction hypothesis, we assume the statement (A.11) holds for time step t. For time step $t+1$, we split the discussions into two cases.

1. If $\tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}) \in \partial\mathcal{X}_R$, then

$$\tilde{X}_{t+1}(\tilde{\mathsf{a}}) = \tilde{X}_t(\tilde{\mathsf{a}}) = \tilde{X}_t(\mathsf{a}) = \tilde{X}_{t+1}(\mathsf{a}),$$

   where the first and third equalities follow by Eq. (2.13) and the second equality is due to the induction hypothesis.

2. In the second case that $\tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}) \in \mathring{\mathcal{X}}_R$, we apply Eq. (2.13) to get

$$\tilde{X}_{t+1}(\tilde{\mathsf{a}}) = \tilde{H}\Big(K\left(\tilde{X}_t(\tilde{\mathsf{a}}), \tilde{a}_t\right), \varepsilon_{t+1}\Big) = \tilde{H}\Big(K\left(\tilde{X}_t(\mathsf{a}), a_t\right), \varepsilon_{t+1}\Big) = \tilde{X}_{t+1}(\mathsf{a}),$$

   where the second equality follows by Eq. (A.10) and the induction hypothesis (A.11).

Next, we will show that $\tilde{a}_{t+1} \in A_{t+1}\left(\tilde{X}_{t+1}(\tilde{\mathsf{a}})\right)$. In either of the above two cases, $\tilde{X}_{t+1}(\mathsf{a}) = \tilde{X}_{t+1}(\tilde{\mathsf{a}})$, which further implies that $X_{t+1}(\mathsf{a}) = \tilde{X}_{t+1}(\mathsf{a})$ if $\tilde{X}_{t+1}(\tilde{\mathsf{a}}) \in \mathring{\mathcal{X}}_R$ by Lemma A.1 and Eq. (2.12). In view of these and Eq. (A.10), we get

$$\tilde{a}_{t+1} = a_{t+1} \in A_{t+1}\left(X_{t+1}(\mathsf{a})\right) = A_{t+1}\left(\tilde{X}_{t+1}(\mathsf{a})\right) = A_{t+1}\left(\tilde{X}_{t+1}(\tilde{\mathsf{a}})\right),$$

if $\tilde{X}_{t+1}(\tilde{\mathsf{a}}) \in \mathring{\mathcal{X}}_R$. On the flip side, if $\tilde{X}_{t+1}(\mathsf{a}) = \tilde{X}_{t+1}(\tilde{\mathsf{a}}) \in \partial\mathcal{X}_R$, we get

$$\tilde{a}_{t+1} = a^*_{t+1}\left(\tilde{X}_{t+1}(\mathsf{a})\right) = a^*_{t+1}\left(\tilde{X}_{t+1}(\tilde{\mathsf{a}})\right) \in A_{t+1}\left(\tilde{X}_{t+1}(\tilde{\mathsf{a}})\right)$$

according to Eq. (A.10). This proves the statement (A.11) holds for time step $t+1$. The proof of Part (i) is complete.

(ii) Given $\tilde{\mathsf{a}} \in \tilde{\mathcal{A}}$ and $\tilde{X}(\tilde{\mathsf{a}})$, we construct a as follows: $a_0 = \tilde{a}_0$, and

$$a_t = \tilde{a}_t \mathbb{1}_{\left\{\tilde{X}_t(\tilde{\mathsf{a}}) \in \mathring{\mathcal{X}}_R\right\}} + \hat{a}_t\left(X_t(\mathsf{a})\right) \mathbb{1}_{\left\{\tilde{X}_t(\tilde{\mathsf{a}}) \in \partial\mathcal{X}_R\right\}}, \quad \text{for} \quad t = 1, 2, \ldots, T-1, \tag{A.12}$$

where $\hat{a}_t(\cdot)$ is any measurable function satisfying $\hat{a}_t(x) \in A_t(x)$ for $x \in \mathcal{X}$ and $t \in \mathcal{T}_0$.

It is easy to see that a is $\mathcal{F}$-adapted. Next, we use a forward induction argument to show that

$$a_t \in A_t\left(X_t(\mathsf{a})\right) \quad \text{and} \quad \tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}), \quad \text{for} \quad t \in \mathcal{T}. \tag{A.13}$$

The above statement holds trivially for $t = 0$. As induction hypothesis, we assume it holds for time step t. For time step $t+1$, we consider two separate cases.

1. If $\tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}) \in \partial\mathcal{X}_R$, Eq. (2.13) combined with (A.13) implies

$$\tilde{X}_{t+1}(\mathsf{a}) = \tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}) = \tilde{X}_{t+1}(\tilde{\mathsf{a}}).$$

2. In the second case that $\tilde{X}_t(\mathsf{a}) = \tilde{X}_t(\tilde{\mathsf{a}}) \in \mathring{\mathcal{X}}_R$, applying Eq. (2.13) gives

$$\tilde{X}_{t+1}(\mathsf{a}) = \tilde{H}\Big(K\left(\tilde{X}_t(\mathsf{a}), a_t\right), \varepsilon_{t+1}\Big) = \tilde{H}\Big(K\left(\tilde{X}_t(\tilde{\mathsf{a}}), \tilde{a}_t\right), \varepsilon_{t+1}\Big) = \tilde{X}_{t+1}(\tilde{\mathsf{a}}),$$

where the second equality follows by Eq. (A.12) and the induction hypothesis (A.13).

Overall, we always observe $\tilde{X}_{t+1}(\mathsf{a}) = \tilde{X}_{t+1}(\tilde{\mathsf{a}})$. To prove the statement (A.13) holds for time step $t+1$, it remains to show $a_{t+1} \in A_{t+1}\left(X_{t+1}(\mathsf{a})\right)$. We split the discussion into two separate cases.

1. Firstly, suppose $\tilde{X}_{t+1}(\mathsf{a}) = \tilde{X}_{t+1}(\tilde{\mathsf{a}}) \in \partial\mathcal{X}_R$. Eq. (A.12) implies

$$a_{t+1} = \hat{a}_{t+1}\left(X_{t+1}(\mathsf{a})\right) \in A_{t+1}\left(X_{t+1}(\mathsf{a})\right).$$

2. Secondly, suppose $\tilde{X}_{t+1}(\mathsf{a}) = \tilde{X}_{t+1}(\tilde{\mathsf{a}}) \in \mathring{\mathcal{X}}_R$. By Part (i) of Lemma 1, $\left\{\tilde{X}_{t+1}(\mathsf{a}) \in \mathring{\mathcal{X}}_R\right\} = \left\{\tau^R > t+1\right\}$ and thus, it follows from Eq. (2.12) that

$$\tilde{X}_{t+1}(\mathsf{a}) = X_{t+1}(\mathsf{a}), \quad \text{if} \quad \tilde{X}_{t+1}(\mathsf{a}) \in \mathring{\mathcal{X}}_R.$$

Consequently, we apply Eq. (A.12) to get

$$a_{t+1} = \tilde{a}_{t+1} \in A_{t+1}\left(\tilde{X}_{t+1}(\tilde{\mathsf{a}})\right) = A_{t+1}\left(\tilde{X}_{t+1}(\mathsf{a})\right) = A_{t+1}\left(X_{t+1}(\mathsf{a})\right).$$

The proof of Part (ii) is complete.

□

A direct consequence of the preceding lemma is the following corollary.

Corollary A.1. *The value function $\tilde{V}_0(x_0)$ defined in Eq. (2.15) exhibits:*

$$\tilde{V}_0(x_0) = \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t f_t\left(\tilde{X}_t, a_t\right) + \varphi^T f_T\left(\tilde{X}_T\right)\right]. \tag{A.14}$$

It is worth noting that the optimization problems in Eq. (A.14) and Eq. (2.15) are taken over the set $\mathcal{A}$ and $\tilde{\mathcal{A}}$, respectively. The above corollary states that the optimal values of these two optimization problems are exactly the same as given by $\tilde{V}_0(x_0)$.

Proof of the Main Result

Proof of Theorem 2.1. In view of Eqs. (2.2) and (A.14), we obtain

$$\begin{aligned}\left|\tilde{V}_0(x_0) - V_0(x_0)\right| \;\leq\; & \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left|f_t\left(\tilde{X}_t, a_t\right) - f_t(X_t, a_t)\right| \mathbb{1}_{\{\tilde{X}_t \neq X_t\}}\right] \\ & + \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\left|f_T\left(\tilde{X}_T\right) - f_T(X_T)\right| \mathbb{1}_{\{\tilde{X}_T \neq X_T\}}\right] := I_1 + I_2. \end{aligned} \tag{A.15}$$

Below we establish upper bounds for the I_1 and I_2 defined in the above display, respectively. Let $E := \left\{X_t = \tilde{X}_t \;\text{ for all }\; t \in \mathcal{T}\right\}$. Note that $E \subseteq \left\{X_t = \tilde{X}_t\right\} \Longrightarrow \left\{X_t \neq \tilde{X}_t\right\} \subseteq E^c$, and

accordingly, we get

$$\begin{aligned} I_1 &= \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left| f_t\left(\tilde{X}_t, a_t\right) - f_t(X_t, a_t)\right| \mathbb{1}_{\{X_t \neq \tilde{X}_t\}}\right] \\ &\leq \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\left| f_t\left(\tilde{X}_t, a_t\right)\right| + |f_t(X_t, a_t)|\right) \mathbb{1}_{\{X_t \neq \tilde{X}_t\}}\right] \\ &\leq \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\left| f_t\left(\tilde{X}_t, a_t\right)\right| + |f_t(X_t, a_t)|\right) \mathbb{1}_{E^c}\right] \\ &\leq \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\sum_{t=0}^{T-1} \left(\xi^{\frac{1}{2}}(R) + B^{\frac{1}{2}}(X_t)\right) \mathbb{1}_{E^c}\right] = \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[\left(\sum_{t=0}^{T-1} Y_t\right) \mathbb{1}_{E^c}\right], \end{aligned} \tag{A.16}$$

with $Y_t := \xi^{\frac{1}{2}}(R) + B(X_t)^{\frac{1}{2}}$, where the first inequality is by triangular inequality and Assumption 2.3 and the third inequality is due to Part (ii) of Assumption 2.3. Applying Cauchy–Schwarz inequality twice gives

$$\begin{aligned} I_1 &\leq \sup_{\mathsf{a}\in\mathcal{A}} \left\{ \mathbb{E}[\mathbb{1}_{E^c}] \cdot \mathbb{E}\left[\left(\sum_{t=0}^{T-1} Y_t\right)^2\right]\right\}^{\frac{1}{2}} \\ &\leq T^{\frac{1}{2}} \cdot \sup_{\mathsf{a}\in\mathcal{A}} \left\{ \mathbb{E}[\mathbb{1}_{E^c}] \cdot \mathbb{E}\left[\sum_{t=0}^{T-1} Y_t^2\right]\right\}^{\frac{1}{2}} \\ &\leq T^{\frac{1}{2}} \cdot \sup_{\mathsf{a}\in\mathcal{A}} \left\{ \mathbb{E}[\mathbb{1}_{E^c}] \cdot \mathbb{E}\left[2\sum_{t=0}^{T-1} \left(\xi(R) + B(X_t)\right)\right]\right\}^{\frac{1}{2}} \\ &\leq \sqrt{2}T^{\frac{1}{2}} \cdot \sup_{\mathsf{a}\in\mathcal{A}} \left\{ \mathbb{E}[\mathbb{1}_{E^c}] \cdot \sum_{t=0}^{T-1} \mathbb{E}\left[\xi(R) + B(X_t)\right]\right\}^{\frac{1}{2}}, \end{aligned}$$

where the third inequality follows because $(a+b)^2 \leq 2a^2 + 2b^2$ for two real numbers a and b. In view of Assumption 2.3, we get

$$\sum_{t=0}^{T-1} \mathbb{E}\left[\xi(R) + B(X_t)\right] \leq T\left(\xi(R) + \sup_{\mathsf{a}\in\mathcal{A}} \mathbb{E}\left[B(X_t)\right]\right) \leq T\big(\xi(R) + \zeta\big).$$

Combing the last two displays with Assumption 2.2 implies

$$\begin{aligned} I_1 &\leq \sqrt{2}T\big(\xi(R)+\zeta\big)^{\frac{1}{2}}\left(\sup_{\mathsf{a}\in\mathcal{A}}\mathbb{E}[\mathbb{1}_{E^c}]\right)^{\frac{1}{2}} \\ &= \sqrt{2}T\big(\xi(R)+\zeta\big)^{\frac{1}{2}}\left(1-\inf_{\mathsf{a}\in\mathcal{A}}\mathbb{E}[\mathbb{1}_{E}]\right)^{\frac{1}{2}} \\ &\leq T\sqrt{2\big(\xi(R)+\zeta\big)\mathcal{E}_T(x_0,R)}. \end{aligned} \tag{A.17}$$

A similar argument gives

$$I_2 \leq \sqrt{2\big(\xi(R)+\zeta\big)\mathcal{E}_T(x_0,R)}. \tag{A.18}$$

Combining (A.15), (A.17), and (A.18) together implies

$$\big|V_0(x_0)-\tilde{V}_0(x_0)\big| \leq (T+1)\sqrt{2\big(\xi(R)+\zeta\big)\mathcal{E}_T(x_0,R)}.$$

The proof is complete. □

A.3.3 Proof of Theorem 2.2

Preliminary lemmas

We first give the definitions of "Big O p" and "Small O p" notations which are commonplaces in statistics literature.

Definition A.1. **(i)** *For two sequences of random variables* $\{a_M\}_{M\in\mathbb{N}}$ *and* $\{b_M\}_{M\in\mathbb{N}}$*, we say* $a_M = O_{\mathbb{P}}(b_M)$ *if* $\lim_{k\to\infty}\limsup_{M\to\infty}\mathbb{P}\left(|a_M|>kb_M\right)=0$.

(ii) *Moreover, we say* $a_M = o_{\mathbb{P}}(b_M)$ *if* $\limsup_{M\to\infty}\mathbb{P}\left(|a_M|>kb_M\right)=0$ *for all* $k>0$.

Some Matrices Let $\mathbf{h}_t(x) = \left(\sup_{a\in A_t(x)}\phi_1\big(K(x,a)\big),\ldots,\sup_{a\in A_t(x)}\phi_J\big(K(x,a)\big)\right)^{\intercal}$, for $x\in\mathrm{cl}\left(\mathcal{X}_R\right)$, and we suppress its dependency on J. Define matrices

$$\Psi_t = \mathbb{E}\left[\mathbf{h}_t\left(X_t^{(m)}\right)\mathbf{h}_t^{\intercal}\left(X_t^{(m)}\right)\right] \quad\text{and}\quad \hat{\Psi}_t = \frac{1}{M}\sum_{m=1}^{M}\mathbf{h}_t\left(X_t^{(m)}\right)\mathbf{h}_t^{\intercal}\left(X_t^{(m)}\right)$$

for $t=1,2,\ldots,T-1$ with the superscript $\intercal$ denoting vector transpose. It is palpable that $\hat{\Psi}_t$ is a finite-sample estimate for Ψ_t. In the sequel, we denote $\lambda_{\max}(B)$ (resp. $\lambda_{\min}(B)$) as the largest

(resp. smallest) eigenvalue of a square matrix B. We impose the following Assumption on the eigenvalues of Ψ_t.

Assumption A.2. **(i)** *For each x and t, $A_t(x)$ is a compact set. Moreover, $a \longmapsto K(x,a)$ and $\phi_j(\cdot) : \mathbb{R}^r \longrightarrow \mathbb{R}$ are continuous functions for $1 \leq j \leq J$.*

(ii) *There exists a positive constant $\bar{c}_\Psi$ independent of t and J such that $\lambda_{\max}(\Psi_t) \leq \bar{c}_\Psi < \infty$.*

Part (i) of the preceding assumption guarantees that the function $\mathbf{h}_t(\cdot)$ is well-defined for $t \in \mathcal{T}_0$. The continuity requirement of $a \longmapsto K(x,a)$ can be removed if $A_t(x)$ is a lattice (discrete set), which is particularly the case when the stochastic optimal control problem exhibits the *bang-bang solution*, see, e.g., [6] and [47]. Part (ii) requires the largest eigenvalue of the matrix $\hat{\Psi}_t$ does not blow up as M and J approach infinity. This condition ensures the sample eigenvalue converges to the non-sample counterpart as M approaches infinity as shown in the subsequent Lemma A.3.

Moreover, we define matrices

$$\Phi_t = \mathbb{E}\left[\boldsymbol{\phi}\left(X_{t^+}^{(m)}\right)\boldsymbol{\phi}^\intercal\left(X_{t^+}^{(m)}\right)\right] \quad \text{and} \quad \hat{\Phi}_t = \frac{1}{M}\sum_{m=1}^{M}\boldsymbol{\phi}\left(X_{t^+}^{(m)}\right)\boldsymbol{\phi}^\intercal\left(X_{t^+}^{(m)}\right).$$

The following lemma relates the eigenvalues of $\hat{\Phi}_t$ and $\hat{\Psi}_t$ to those of Φ_t and Ψ_t.

Lemma A.3. **(i)** *Suppose Condition (ii) of Theorem 2.2 is satisfied. Then,*

$$\left|\lambda_{\max}(\Phi_t) - \lambda_{\max}\left(\hat{\Phi}_t\right)\right| = O_{\mathbb{P}}\left(\Upsilon(J)\sqrt{J/M}\right),$$

and

$$\left|\lambda_{\min}(\Phi_t) - \lambda_{\min}\left(\hat{\Phi}_t\right)\right| = O_{\mathbb{P}}\left(\Upsilon(J)\sqrt{J/M}\right),$$

for $t \in \mathcal{T}_0$.

(ii) *Suppose Assumption A.2 holds. In addition, Condition (v) of Assumption A.1 is satisfied. Then, $\lambda_{\max}\left(\hat{\Psi}_t\right) = O_{\mathbb{P}}(1)$ for $t = 1, 2, \ldots, T-1$.*

Proof of Lemma A.3. Lemma A.3 can be proved by a similar argument as that used in the proof of Eq. (A.1) in [65]. □

The above lemma shows the sample eigenvalues converge to the non-sample counterparts as M approaches infinity. In view of Condition (iv) of Assumption A.1, Lemma A.3 also implies the

largest (resp., smallest) eigenvalue of $\hat{\Phi}_t$ is bounded from above (resp., below) with probability approaching 1 as $M \longrightarrow \infty$. This fact is exploited in the proofs of subsequent Lemmas A.4 and A.5.

Pseudo Estimate, Oracle, and True Estimate Next, we introduce the concept of *pseudo estimate*. Let $\bar{\boldsymbol{\beta}}_t$ (resp. $\hat{\boldsymbol{\beta}}_t$) be the solution to the optimization problem in Eq. (2.31) with $U^{(m)} = \tilde{V}_{t+1}\left(X_{t+1}^{(m)}\right)$ (resp. $\tilde{V}_{t+1}^{\mathrm{E}}\left(X_{t+1}^{(m)}\right)$) and $Z^{(m)} = X_{t^+}^{(m)}$. Given $\bar{\boldsymbol{\beta}}_t$ and $\hat{\boldsymbol{\beta}}_t$, denote the associated regression estimates by $\tilde{C}_t^{\mathrm{PE}}(\cdot) = \bar{\boldsymbol{\beta}}_t^{\intercal}\boldsymbol{\phi}(\cdot)$ and $\tilde{C}_t^{\mathrm{E}}(\cdot) = \hat{\boldsymbol{\beta}}_t^{\intercal}\boldsymbol{\phi}(\cdot)$, respectively. $\tilde{C}_t^{\mathrm{PE}}(\cdot)$ is essentially the sieve estimate for the continuation function $\tilde{C}_t(\cdot)$ when the true value function $\tilde{V}_{t+1}(\cdot)$ is employed in the regression. We further define function $\tilde{V}_t^{\mathrm{PE}}(x)$ for $x \in \mathring{\mathcal{X}}_R$ by substituting $\tilde{C}_t^{\mathrm{E}}(\cdot)$ in Eq. (2.25) with $\tilde{C}_t^{\mathrm{PE}}(\cdot)$. For $x \in \partial\mathcal{X}_R$, we set $\tilde{V}_t^{\mathrm{PE}}(x) = \tilde{V}_t(x)$ with $\tilde{V}_t(\cdot)$ given by Eq. (2.16).

Admittedly, in the implementation of the BSBU algorithm, $\bar{\boldsymbol{\beta}}_t$ is not tractable because the true value function is unknown and should be replaced by the numerical estimate $\tilde{V}_{t+1}^{\mathrm{E}}(\cdot)$ obtained inductively. For this reason, following [16], we call $\bar{\boldsymbol{\beta}}_t$ the *pseudo estimate*. Despite this, the pseudo estimate plays an indispensable role in establishing the convergence result of Theorem 2.2. In addition to the two estimates $\bar{\boldsymbol{\beta}}_t$ and $\hat{\boldsymbol{\beta}}_t$ defined in the above, we further define the *oracle* $\boldsymbol{\beta}_t^{\mathrm{o}}$ as the solution to the optimization problem (A.3) with $g(\cdot)$ replaced by $\tilde{C}_t(\cdot)$.

The following lemma discloses that the gap between pseudo estimate and the oracle vanishes when both M and J increase at a certain rate.

Lemma A.4. *Suppose the conditions of Theorem 2.2 are satisfied. Then,*

$$\left|\bar{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right| = O_{\mathbb{P}}\left(\sqrt{J/M} + \rho_J\right), \quad \text{for} \quad t \in \mathcal{T}_0.$$

Proof of Lemma A.4. Recall that $\bar{\boldsymbol{\beta}}_t$ solves the optimization problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{J+1}} \frac{1}{M}\sum_{m=1}^{M}\left[\tilde{V}_{t+1}\left(X_{t+1}^{(m)}\right) - \boldsymbol{\beta}^{\intercal}\boldsymbol{\phi}\left(X_{t^+}^{(m)}\right)\right]^2, \quad \text{subject to} \quad \boldsymbol{\beta}^{\intercal}\boldsymbol{\phi}(\cdot) \in \mathcal{H}_J.$$

On the other hand, $\boldsymbol{\beta}_t^{\mathrm{o}}$ is a suboptimal solution to the above optimization problem. Therefore, we get

$$\left|\mathbf{V}_{t+1} - P\bar{\boldsymbol{\beta}}_t\right|^2 \leq \left|\mathbf{V}_{t+1} - P\boldsymbol{\beta}_t^{\mathrm{o}}\right|^2,$$

where P is a M-by-J matrix with m-th row being $\boldsymbol{\phi}^{\intercal}\left(X_{t^+}^{(m)}\right)$ and $\mathbf{V}_{t+1}$ is a M-by-1 vector with m-th element given by $\tilde{V}_{t+1}\left(X_{t+1}^{(m)}\right)$.

By adding and subtracting the term $P\boldsymbol{\beta}_t^{\mathrm{o}}$ in the L.H.S. of the above inequality, we get

$$\left|\tilde{\mathbf{U}} - P\bar{\boldsymbol{\delta}}\right|^2 \leq \left|\tilde{\mathbf{U}}\right|^2,$$

where we use the shorthand notations $\bar{\boldsymbol{\delta}} := \bar{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}$ and $\tilde{\mathbf{U}} := \mathbf{V}_{t+1} - P\boldsymbol{\beta}_t^{\mathrm{o}}$. Expanding both sides of the above inequality gives

$$\frac{\left|P\bar{\boldsymbol{\delta}}\right|^2}{2M} \leq \frac{\left|\tilde{\mathbf{U}}^\intercal P\bar{\boldsymbol{\delta}}\right|}{M} \leq \frac{\left|P^\intercal\tilde{\mathbf{U}}\right|\left|\bar{\boldsymbol{\delta}}\right|}{M},$$

where the second inequality is by Hölder's inequality. For the L.H.S. of the above inequality, it follows from the definition of the smallest eigenvalue that

$$\frac{\left|P\bar{\boldsymbol{\delta}}\right|^2}{2M} = \frac{\bar{\boldsymbol{\delta}}^\intercal P^\intercal P\bar{\boldsymbol{\delta}}}{2M} \geq \frac{\left|\bar{\boldsymbol{\delta}}\right|^2}{2}\lambda_{\min}\left(\hat{\Phi}_t\right).$$

Combing the last two inequalities together implies

$$\left|\bar{\boldsymbol{\delta}}\right|\lambda_{\min}\left(\hat{\Phi}_t\right) \leq \frac{2}{M}\left|P^\intercal\tilde{\mathbf{U}}\right|.$$

It follows from Lemma A.3 that the event $\left\{\underline{c}_\Phi/2 \leq \lambda_{\min}\left(\hat{\Phi}_t\right)\right\}$ holds with probability approaching 1 as $M \longrightarrow \infty$. And therefore,

$$\left|\bar{\boldsymbol{\delta}}\right| \leq \left(4/\underline{c}_\Phi\right)M^{-1}\left|P^\intercal\tilde{\mathbf{U}}\right| \tag{A.19}$$

holds with probability approaching 1 as $M \longrightarrow \infty$.

It follows as in Eq. (A.2) of [65, pp. 163] that

$$M^{-1}\left|P^\intercal\tilde{\mathbf{U}}\right| = O_{\mathbb{P}}\left(\sqrt{J/M} + \rho_J\right). \tag{A.20}$$

This in conjunction with the last display proves the desired result. The proof is complete. □

The next lemma relates the discrepancy between the pseudo estimate $\bar{\boldsymbol{\beta}}_t$ and the true estimate $\hat{\boldsymbol{\beta}}_t$ to the estimation error of the value function at time step $t+1$. This result is not hard to expect because the primary difference between the pseudo estimate and the true estimate stems from the estimation error of value function.

Lemma A.5. *Suppose the conditions of Theorem 2.2 are satisfied. Then, for $t \in \mathcal{T}_0$, there exists*

a constant $\psi > 0$ independent of t, R and J such that

$$\left|\bar{\boldsymbol{\beta}}_t - \hat{\boldsymbol{\beta}}_t\right| \leq \sqrt{\frac{\psi}{M}} \left|\mathbf{V}_{t+1} - \hat{\mathbf{V}}_{t+1}\right| + O_{\mathbb{P}}\left(\sqrt{J/M} + \rho_J\right)$$

holds with probability approaching 1 as $M \longrightarrow \infty$, where $\mathbf{V}_{t+1}$ and $\hat{\mathbf{V}}_{t+1}$ are two M-by-1 vectors with m-th element given by $\tilde{V}_{t+1}\left(X_{t+1}^{(m)}\right)$ and $\tilde{V}_{t+1}^{\mathrm{E}}\left(X_{t+1}^{(m)}\right)$, respectively.

Proof of Lemma A.5. Firstly, by triangle inequality, we get

$$\left|\bar{\boldsymbol{\beta}}_t - \hat{\boldsymbol{\beta}}_t\right| \leq \left|\bar{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right| + \left|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right| = O_{\mathbb{P}}\left(\sqrt{J/M} + \rho_J\right) + \left|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right|, \tag{A.21}$$

where the last equality is by Lemma A.4.

Next, we would like to establish an upper bound for $\left|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right|$. Using the argument as in the proof of inequality (A.19), we obtain

$$\left|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right| \leq \left(4/\underline{c}_\Phi\right) M^{-1} \left|P^\intercal \hat{\mathbf{U}}\right|,$$

holds with probability approaching 1 as $M \longrightarrow \infty$, where we adopt shorthand notation $\hat{\mathbf{U}} := \hat{\mathbf{V}}_{t+1} - P\boldsymbol{\beta}_t^{\mathrm{o}}$.

We also note that

$$\left|\hat{\mathbf{U}}\right| = \left|\hat{\mathbf{V}}_{t+1} - \mathbf{V}_{t+1} + \mathbf{V}_{t+1} - P\boldsymbol{\beta}_t^{\mathrm{o}}\right| \leq \left|\hat{\mathbf{V}}_{t+1} - \mathbf{V}_{t+1}\right| + \left|\mathbf{V}_{t+1} - P\boldsymbol{\beta}_t^{\mathrm{o}}\right|.$$

Combing the last two inequalities implies

$$\left|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right| \leq \left(4/\underline{c}_\Phi\right) M^{-1} \left(\left|P^\intercal \boldsymbol{\Delta}\right| + \left|P^\intercal \mathbf{U}\right|\right) = \left(4/\underline{c}_\Phi\right) M^{-1} \left|P^\intercal \boldsymbol{\Delta}\right| + O_{\mathbb{P}}\left(\sqrt{J/M} + \rho_J\right) \tag{A.22}$$

with shorthand notations $\boldsymbol{\Delta} := \hat{\mathbf{V}}_{t+1} - \mathbf{V}_{t+1}$ and $\tilde{\mathbf{U}} := \mathbf{V}_{t+1} - P\boldsymbol{\beta}_t^{\mathrm{o}}$, where the last equality follows by Eq. (A.20).

It follows from Lemma A.3 that

$$M^{-2} \left|P^\intercal \boldsymbol{\Delta}\right|^2 = M^{-1} \boldsymbol{\Delta}^\intercal \left(M^{-1} P P^\intercal\right) \boldsymbol{\Delta} \leq M^{-1} \lambda_{\max}\left(\hat{\Phi}_t\right) |\boldsymbol{\Delta}|^2 \leq M^{-1} 2\bar{c}_\Phi \left|\boldsymbol{\Delta}\right|^2$$

holds with probability approaching 1 as $M \longrightarrow \infty$. Combing the above inequality with (A.22)

implies

$$\left|\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{o}}\right| \leq \sqrt{\frac{32\bar{c}_\Phi}{\underline{c}_\Phi^2}\frac{1}{M}}\,|\boldsymbol{\Delta}| + O_{\mathbb{P}}\left(\sqrt{J/M} + \rho_J\right).$$

This in conjunction with (A.21) proves the desired result. □

The final lemma quantifies the discrepancy between the value function and its numerical estimate under the empirical L^2 norm.

Lemma A.6. *Let $F_t^X(\cdot)$ be the probability distribution function of $X_t^{(m)}$ for $t = 1, 2, \ldots, T-1$. Suppose the assumptions of Theorem 2.2 hold. Then*

$$M^{-1}\left|\mathbf{V}_t - \hat{\mathbf{V}}_t\right|^2 = O_{\mathbb{P}}\left(\psi^{T-t-1}\left(J/M + \rho_J^2\right)\right), \quad \text{for} \quad t = 1, 2, \ldots, T-1, \tag{A.23}$$

where $\mathbf{V}_t$ and $\hat{\mathbf{V}}_t$ are two M-by-1 vectors with m-th element being $\tilde{V}_t\left(X_t^{(m)}\right)$ and $\tilde{V}_t^{\mathrm{E}}\left(X_t^{(m)}\right)$, respectively.

Proof of Lemma A.6. We use a backward induction procedure to prove the statement of Lemma A.6. For $t = T-1$, we note that $\tilde{C}_{T-1}^{\mathrm{E}}(\cdot)$ is in agreement with $\tilde{C}_{T-1}^{\mathrm{PE}}(\cdot)$ because $\tilde{V}_T^{\mathrm{E}}(x) = \tilde{V}_T(x) = f_T(x)$ for $x \in \mathrm{cl}\left(\mathcal{X}_R\right)$. We get $\tilde{V}_{T-1}^{\mathrm{E}}(x) = \tilde{V}_{T-1}^{\mathrm{PE}}(x)$ for $x \in \mathrm{cl}\left(\mathcal{X}_R\right)$, accordingly. Furthermore, we observe that

$$\begin{aligned}
\left|\tilde{V}_{T-1}^{\mathrm{E}}(x) - \tilde{V}_{T-1}(x)\right| &= \left|\tilde{V}_{T-1}^{\mathrm{PE}}(x) - \tilde{V}_{T-1}(x)\right| \\
&\leq \sup_{a \in A_{T-1}(x)} \left|\tilde{C}_{T-1}^{\mathrm{PE}}\big(K(x,a)\big) - \tilde{C}_{T-1}\big(K(x,a)\big)\right| \\
&= \sup_{a \in A_{T-1}(x)} \left|\bar{\boldsymbol{\beta}}_{T-1}^{\intercal}\boldsymbol{\phi}\big(K(x,a)\big) - \tilde{C}_{T-1}\big(K(x,a)\big)\right| \\
&\leq \sup_{a \in A_{T-1}(x)} \left|\left(\bar{\boldsymbol{\beta}}_{T-1} - \boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right)^{\intercal}\boldsymbol{\phi}\big(K(x,a)\big)\right| \\
&\quad + \sup_{a \in A_{T-1}(x)} \left|\boldsymbol{\phi}^{\intercal}\big(K(x,a)\big)\boldsymbol{\beta}_{T-1}^{\mathrm{o}} - \tilde{C}_{T-1}\big(K(x,a)\big)\right| \\
&\leq \left|\left(\bar{\boldsymbol{\beta}}_{T-1} - \boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right)^{\intercal}\mathbf{h}_{T-1}(x)\right| + \left\|\boldsymbol{\phi}^{\intercal}\boldsymbol{\beta}_{T-1}^{\mathrm{o}} - \tilde{C}_{T-1}\right\| \\
&= \left|\left(\bar{\boldsymbol{\beta}}_{T-1} - \boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right)^{\intercal}\mathbf{h}_{T-1}(x)\right| + O(\rho_J),
\end{aligned} \tag{A.24}$$

where the third inequality is by the definition of function $\mathbf{h}_{T-1}(\cdot)$ and the last equality is guaranteed by Assumption A.1.

Consequently, we obtain

$$\begin{aligned} M^{-1}\left|\mathbf{V}_{T-1}-\hat{\mathbf{V}}_{T-1}\right|^2 &= \frac{1}{M}\sum_{m=1}^{M}\left|\tilde{V}_{T-1}^{\mathrm{E}}\left(X_{T-1}^{(m)}\right)-\tilde{V}_{T-1}\left(X_{T-1}^{(m)}\right)\right|^2 \\ &\leq \frac{1}{M}\sum_{m=1}^{M}\left|\left(\bar{\boldsymbol{\beta}}_{T-1}-\boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right)^{\intercal}\mathbf{h}_{T-1}\left(X_{T-1}^{(m)}\right)\right|^2+O\left(\rho_J^2\right) \\ &= \left(\bar{\boldsymbol{\beta}}_{T-1}-\boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right)^{\intercal}\hat{\Psi}_{T-1}\left(\bar{\boldsymbol{\beta}}_{T-1}-\boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right)+O\left(\rho_J^2\right) \\ &\leq 2\lambda_{\max}\left(\hat{\Psi}_{T-1}\right)\left|\bar{\boldsymbol{\beta}}_{T-1}-\boldsymbol{\beta}_{T-1}^{\mathrm{o}}\right|^2+O\left(\rho_J^2\right) \\ &= O_{\mathbb{P}}\left(J/M+\rho_J^2\right), \end{aligned} \tag{A.25}$$

where the second inequality follows from the definition of the largest eigenvalue of a matrix and the last equality is guaranteed by Lemma A.4 and Lemma A.3. In view of the above display, Eq. (A.23) holds for $t = T-1$.

As induction hypothesis, we assume (A.23) holds for $t+1$. Note that, for $x \in \mathring{\mathcal{X}}_R$,

$$\left|\tilde{V}_t(x)-\tilde{V}_t^{\mathrm{E}}(x)\right| \leq \left|\tilde{V}_t(x)-\tilde{V}_t^{\mathrm{PE}}(x)\right|+\left|\tilde{V}_t^{\mathrm{E}}(x)-\tilde{V}_t^{\mathrm{PE}}(x)\right|. \tag{A.26}$$

An argument similar to the one used in establishing (A.25) shows that

$$\frac{1}{M}\sum_{m=1}^{M}\left|\tilde{V}_t^{\mathrm{PE}}\left(X_t^{(m)}\right)-\tilde{V}_t\left(X_t^{(m)}\right)\right|^2 = O_{\mathbb{P}}\left(J/M+\rho_J^2\right). \tag{A.27}$$

Next, we investigate the term $\left|\tilde{V}_t^{\mathrm{E}}(x)-\tilde{V}_t^{\mathrm{PE}}(x)\right|$. Observe that

$$\begin{aligned} \left|\tilde{V}_t^{\mathrm{E}}(x)-\tilde{V}_t^{\mathrm{PE}}(x)\right| &\leq \sup_{a\in A_t(x)}\left|\tilde{C}_t^{\mathrm{E}}\big(K(x,a)\big)-\tilde{C}_t^{\mathrm{PE}}\big(K(x,a)\big)\right| \\ &= \sup_{a\in A_t(x)}\left|\left(\hat{\boldsymbol{\beta}}_t-\bar{\boldsymbol{\beta}}_t\right)^{\intercal}\boldsymbol{\phi}\big(K(x,a)\big)\right| \\ &\leq \left|\left(\hat{\boldsymbol{\beta}}_t-\bar{\boldsymbol{\beta}}_t\right)^{\intercal}\mathbf{h}_t(x)\right|. \end{aligned} \tag{A.28}$$

We adopt the same argument as in the proof of (A.25) to get

$$\frac{1}{M}\sum_{m=1}^{M}\left|\tilde{V}_t^{\mathrm{E}}\left(X_t^{(m)}\right)-\tilde{V}_t^{\mathrm{PE}}\left(X_t^{(m)}\right)\right|^2 \leq \lambda_{\max}\left(\hat{\Psi}_t\right)\left|\hat{\boldsymbol{\beta}}_t-\bar{\boldsymbol{\beta}}_t\right|^2.$$

Applying Lemma A.5 yields

$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M}\left|\tilde{V}_t^{\mathrm{E}}\left(X_t^{(m)}\right)-\tilde{V}_t^{\mathrm{PE}}\left(X_t^{(m)}\right)\right|^2 &\leq 2\lambda_{\max}\left(\hat{\Psi}_t\right)\left[\frac{\psi}{M}\left|\mathbf{V}_{t+1}-\hat{\mathbf{V}}_{t+1}\right|^2+O\left(\psi\rho_J^2\right)\right]\\ &= O_{\mathbb{P}}\left(\psi^{T-t-1}\left(J/M+\rho_J^2\right)\right),\end{aligned}$$

where the last equality is due to induction hypothesis (A.23) and $\lambda_{\max}\left(\hat{\Psi}_t\right)=O_{\mathbb{P}}(1)$ (see Lemma A.3). The above display in conjunction with (A.26) and (A.27) implies

$$\begin{aligned}M^{-1}\left|\mathbf{V}_t-\hat{\mathbf{V}}_t\right|^2 &= O_{\mathbb{P}}\left(J/M+\rho_J^2\right)+O_{\mathbb{P}}\left(\psi^{T-t-1}\left(J/M+\rho_J^2\right)\right)\\ &= O_{\mathbb{P}}\left(\psi^{T-t-1}\left(J/M+\rho_J^2\right)\right).\end{aligned}$$

This completes the proof. □

Proof of the Main Result

Proof of Theorem 2.2. Following the arguments used to prove (A.24), we get

$$\left|\tilde{V}_0^{\mathrm{PE}}(x_0)-\tilde{V}_0(x_0)\right|\leq\left|\left(\bar{\boldsymbol{\beta}}_0-\tilde{\boldsymbol{\beta}}_0\right)^{\intercal}\mathbf{h}_0(X_0)\right|+O(\rho_J)=O_{\mathbb{P}}\left(\sqrt{J/M}+\rho_J\right),$$

where the last equality is by Lemma A.4 and Part (ii) of Assumption A.2.

On the other hand, an argument similar to the one used in deriving (A.28) shows

$$\left|\tilde{V}_0^{\mathrm{PE}}(x_0)-\tilde{V}_0^{\mathrm{E}}(x_0)\right|\leq\left|\left(\hat{\boldsymbol{\beta}}_0-\bar{\boldsymbol{\beta}}_0\right)^{\intercal}\mathbf{h}_0(X_0)\right|\leq M^{-1/2}\left|\mathbf{V}_1-\hat{\mathbf{V}}_1\right|\left|\mathbf{h}_0(X_0)\right|.$$

The above two displays in conjunction with (A.23) implies

$$\begin{aligned}\left|\tilde{V}_0(x_0)-\tilde{V}_0^{\mathrm{E}}(x_0)\right| &\leq \left|\tilde{V}_0^{\mathrm{PE}}(x_0)-\tilde{V}_0(x_0)\right|+\left|\tilde{V}_0^{\mathrm{PE}}(x_0)-\tilde{V}_0^{\mathrm{E}}(x_0)\right|\\ &= O_{\mathbb{P}}\left(\sqrt{\psi^{T-1}\left(J/M+\rho_J^2\right)}\right).\end{aligned}$$

This shows (2.32) and completes the proof of Theorem 2.2. □

Appendix B

Appendix for Chapter 3

B.1 Supplements for Section 3.5

B.1.1 Verification of Assumptions

This subsection aims to verify the Assumptions 3.4 and 3.5 and the additional condition of Theorem 3.1 in the context of the variable annuity contract studied in Section 3.5. In particular, we will give concrete expressions of $\mathcal{E}_{x_0,T}(R)$, $\bar{B}(x)$, $\bar{\xi}(R)$, ζ_X, and the error bound in the R.H.S. of inequality (3.17). This allows us to choose an appropriate truncation parameter R in conducting the numerical experiments in Section 3.5.4.

First recall that in the context of Section 3.5 $A_t(x) = \big[0, G \vee x\big]$, $S(x,a,e) = \big(x-a\big)^+ e$, $f_t(x,a) = a - \kappa_t\big(a-G\big)^+$, and $f_T(x) = x$, with $x \in \mathcal{X} = [0,\infty)$, $a \in \mathbb{A} = [0,\infty)$, and $e \in \mathbb{D} = (0,\infty)$. Moreover, ε_{t+1} follows a log-normal distribution with $\mathbb{E}[\log \varepsilon_{t+1}] = \big(r - q - \sigma^2/2\big)\delta := \mu\delta$ and $\mathbb{V}\mathrm{ar}[\log \varepsilon_{t+1}] = \sigma\delta$.

Verification of Assumption 3.4

First recall that

$$\bar{\mathcal{A}} = \Big\{ \mathsf{a} = \{a_t\}_{t\in\mathcal{T}_0} \Big|\ a_t \text{ is } \mathcal{F}_t\text{-measurable and } a_t \geq 0 \Big\}.$$

Next define a continuous-time process $Y := \{Y_u\}_{u\in[0,T\delta]}$ as follows:

$$Y_u = x_0 \exp\Big(\big(r - q - \sigma^2/2\big)u + \sigma\mathcal{B}_u\Big) = x_0 \exp\left(\mu u + \sigma\mathcal{B}_u\right)$$

with $\{\mathcal{B}_u\}_{u\in[0,T\delta]}$ being a standard Brownian Motion under the pricing measure $\mathbb{Q}$. In view of the above display and the transition equation of the PH's investment account (Eq. (3.24)), one gets $Y_{\delta t} \geq X_t$ for all $t \in \mathcal{T}$ and $\mathsf{a} \in \bar{\mathcal{A}}$.

To proceed, note that the truncated region is $\mathcal{X}_R = [0, R]$ and accordingly,

$$\{\mathcal{X}_t \in \mathcal{X}_R, \quad \text{for all} \quad t \in \mathcal{T}_0\}^c \subseteq \left\{\max_{t\in\mathcal{T}_0} X_t \geq R\right\} \subseteq \left\{\sup_{u\in[0,T\delta]} Y_u \geq R\right\}.$$

And thus,

$$\begin{aligned}
\sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}} \mathbb{P}\Big[\{X_t \in \mathcal{X}_R, \quad \text{for all} \quad t \in \mathcal{T}_0\}^c\Big] &\leq \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} Y_u > R\right] \\
&= \mathbb{P}\left[x_0 \exp\left(\sup_{u\in[0,\tilde{T}]} [\mu u + \sigma\mathcal{B}_u]\right) > R\right] \\
&= \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} \left[\mu u + \sigma\mathcal{B}_u\right] > \log\left(R/x_0\right)\right] \\
&= \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} \left[\alpha u + \mathcal{B}_u\right] > m\right],
\end{aligned}$$

with $\alpha := \mu/\sigma = (r - q - \sigma^2/2)/\sigma$, $\tilde{T} := T\delta$, and $m := \sigma^{-1}\log\left(R/x_0\right)$. In view of the above inequality, one may choose

$$\begin{aligned}
\mathcal{E}_{x_0,T}(R) &= \mathbb{P}\left[\sup_{u\in[0,\tilde{T}]} [\alpha u + \mathcal{B}_u] > m\right] \\
&= 1 - \mathcal{N}\left(\frac{m - \alpha\tilde{T}}{\sqrt{\tilde{T}}}\right) + e^{2\alpha m}\mathcal{N}\left(\frac{-m - \alpha\tilde{T}}{\sqrt{\tilde{T}}}\right) \quad m \geq 0,
\end{aligned}$$

where $\mathcal{N}(\cdot)$ denotes the cdf of a standard normal random variable and the second equality follows by the distribution function of the continuous running maximum of a Brownian Motion with nonzero drift; see [75, Corollary 7.2.2, pp. 297]. The plot of function $\mathcal{E}_{x_0,T} : (0,\infty) \to [0,1]$ is depicted in Figure 3.1; see the solid line.

Verification of Assumption 3.5

For any $\bar{a} \in \mathbb{A} = [0, \infty)$ one may choose $a = \min\left[\bar{a}, G \vee x\right]$ such that $a \in A_t(x) = \left[0, G \vee x\right]$ and

$$S\left(x, \bar{a}, e\right) = \left(x - \bar{a}\right)^+ e = \left(x - \bar{a}\right) e \mathbb{1}_{\{\bar{a} \leq x\}} = \left(x - a\right) e \mathbb{1}_{\{a \leq x\}} = S\left(x, a, e\right),$$

where the second equality follows by $\{\bar{a} \leq G \vee x\} = \{a = \bar{a}\} \subseteq \{\bar{a} \leq x\}$. Thus, the additional condition of Theorem 3.1 is fulfilled under the above choice of a.

Next we would like to verify Assumption 3.5. Note that one may choose $h_t(x) \equiv 0$ such that Part (i) is satisfied and accordingly,

$$\left|\bar{f}_t(x, a)\right| \leq (G \vee x)^2 \leq (G + x)^2 \leq 2G^2 + 2x^2, \quad \text{for all} \quad a \in \mathbb{A} = [0, \infty).$$

In view of this, one may choose $\bar{B}(x) = 2G^2 + 2x^2$. Recall that $\mathcal{X}_R = [0, R]$ and that $\bar{\xi}(R)$ is an upper bound (uniformly in t) for $\bar{B}(x)$ over $\mathcal{X}_R$. Thus one may choose $\bar{\xi}(R) = 2\left(G^2 + R^2\right)$.

Recall that $Y_{\delta t} \geq X_t$ for all $t \in \mathcal{T}$ and $\mathsf{a} \in \bar{\mathcal{A}}$ almost surely. This implies

$$\sup_{\mathsf{a} \in \bar{\mathcal{A}}} \mathbb{E}\left[X_t^2\right] \leq \mathbb{E}\left[Y_{\delta t}^2\right] = e^{2\mu\delta t + 2\sigma^2 \delta t} = e^{2\left(\mu + \sigma^2\right)\tilde{T}}$$

with $\tilde{T} = \delta T$ for all $t \in \mathcal{T}$. And consequently,

$$\sup_{\mathsf{a} \in \bar{\mathcal{A}}} \mathbb{E}\left[\bar{B}\left(X_t\right)\right] = 2\left(G^2 + \sup_{\bar{\mathsf{a}} \in \bar{\mathcal{A}}} \mathbb{E}\left[X_t^2\right]\right) \leq 2\left[G^2 + e^{2\left(\mu + \sigma^2\right)\tilde{T}}\right] := \zeta_X.$$

To sum up, in the context of Section 3.5, one may choose $\bar{\xi}(R) = 2\left(G^2 + R^2\right)$, $\bar{B}(x) = 2\left(G^2 + x^2\right)$, and $\zeta_X = 2\left[G^2 + e^{2\left(\mu + \sigma^2\right)\tilde{T}}\right]$ such that Assumption 3.5 holds.

Combing the above results together, one can give an explicit expression of the error bound given in (3.17). Its plot as a function of truncation parameter R is depicted in Figure 3.1 under the parameter setting of Section 3.5.4.

B.1.2 Evaluation of Eq. (3.21)

Combing Eq. (3.19) with Eq. (3.21) gives

$$V_t^*(x) = \sup_{a \in A_t(x)} \left\{ f_t(x, a) + \varphi \check{\boldsymbol{\beta}}_{t+1}^{\intercal} \mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}\left(\hat{S}(x, a, \varepsilon_{t+1})\right)\right]\right\}, \quad t \in \mathcal{T}_0.$$

In the setting of Section 3.5, ε_{t+1} follows a log-normal distribution such that $\mathbb{E}^{\mathbb{Q}}[\log \varepsilon_{t+1}] = \big(r - q - \sigma^2/2\big)\delta$ and $\mathbb{V}\mathrm{ar}^{\mathbb{Q}}\left[\log \varepsilon_{t+1}\right] = \sigma^2\delta$.

In the sequel, we will evaluate the expectation term involved in the above equation under different choices of basis functions. To this end, we present a formula for the partial expectation of a log-normal random variable which will be frequently used. Let Z be a log-normal random variable with $\mathbb{E}[\log Z] = m$ and $\mathbb{V}\mathrm{ar}[\log Z] = v$. Then

$$\mathbb{E}\big[Z\mathbb{1}_{\{Z\le y\}}\big] = e^{m+v^2/2}\mathcal{N}\left(\frac{\log y - m - v^2}{v}\right), \quad \text{for} \quad y > 0, \tag{B.1}$$

where $\mathcal{N}(\cdot)$ denotes the cdf of a standard normal random variable.

Power series

Let $\mathbf{p}(x) = \big(1, x, \ldots, x^J\big)^\intercal$ and $\check{\boldsymbol{\beta}}_t = \big(\check{\beta}_{t,0}, \check{\beta}_{t,1}, \ldots, \check{\beta}_{t,J}\big)^\intercal$. By choosing $\boldsymbol{\phi}(x) = \mathbf{p}(x)$, Eq. (3.21) reads

$$\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\mathbb{Q}}\left[\mathbf{p}\left(\hat{S}(x,a,\varepsilon_{t+1})\right)\right] = \check{\beta}_{t+1,0} + \sum_{j=1}^{J}\check{\beta}_{t+1,j}\mathbb{E}^{\mathbb{Q}}\left[\hat{S}^j(x,a,\varepsilon_{t+1})\right]. \tag{B.2}$$

Since $\hat{S}(x,a,e) = \mathcal{Q}(S(x,a,e))$ with $S(x,a,e)$ given in Eq. (3.26) and $\mathcal{Q}(x) = \min\big[x, R\big]$,

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[\hat{S}^j(x,a,\varepsilon_{t+1})\right] &= \mathbb{E}^{\mathbb{Q}}\left[\min\left[\big((x-a)^+\varepsilon_{t+1}\big)^j, R^j\right]\right] \\ &= \begin{cases} 0, & \text{if} \quad x = a, \\ (x-a)^j\mathbb{E}^{\mathbb{Q}}\Big[\min\big[Z_j, Y_j\big]\Big], & \text{if} \quad x > a, \end{cases} \quad j = 1, 2, \ldots, J, \end{aligned} \tag{B.3}$$

where $Z_j := \varepsilon_{t+1}^j$ and $Y_j := \left(\frac{R}{x-a}\right)^j$, respectively.

Note that Z_j follows a log-normal distribution with

$$\mathbb{E}^{\mathbb{Q}}[\log Z_j] = \mathbb{E}^{\mathbb{Q}}\left[\log\left(\varepsilon_{t+1}^j\right)\right] = j(r - q - \sigma^2/2)\delta := m_j(\delta), \tag{B.4}$$

and

$$\mathbb{V}\mathrm{ar}^{\mathbb{Q}}[\log Z_j] = \mathbb{V}\mathrm{ar}^{\mathbb{Q}}\left[\log\left(\varepsilon_{t+1}^j\right)\right] = \left(j\sigma\sqrt{\delta}\right)^2 := v_j(\delta)^2. \tag{B.5}$$

By exploiting the partial expectation formula (B.1), some routine calculations give

$$\mathbb{E}^{\mathbb{Q}}\Big[\min\big[Z_j, Y_j\big]\Big] = e^{m_j(\delta)+v_j(\delta)^2/2}\mathcal{N}(d_{1,j}) + Y_j\mathcal{N}(d_{2,j}), \tag{B.6}$$

where

$$d_{1,j} = \frac{\log Y_j - m_j(\delta) - v_j(\delta)^2}{v_j(\delta)} = \frac{\log\left(\frac{R}{x-a}\right) - \left(r - q - \frac{\sigma^2}{2}\right)\delta - j\sigma^2\delta}{\sigma\sqrt{\delta}}, \tag{B.7}$$

$$d_{2,j} = -\frac{\log Y_j - m_j(\delta)}{v_j(\delta)} = \frac{\log\left(\frac{x-a}{R}\right) + \left(r - q - \frac{\sigma^2}{2}\right)\delta}{\sigma\sqrt{\delta}}, \tag{B.8}$$

for $x > a$ and $j = 1, 2, \ldots, J$.

Combing Eqs. (B.2), (B.3), and (B.6) together implies

$$\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\mathbb{Q}}\left[\mathbf{p}\left(\hat{S}(x, a, \varepsilon_{t+1})\right)\right] = \begin{cases} \check{\boldsymbol{\beta}}_{t+1}^{\intercal}\Big[\boldsymbol{\psi}_1(x-a)\circ\mathbf{E} + \boldsymbol{\psi}_2(x-a)\circ\mathbf{p}(R)\Big], & x > a, \\ \check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{e}, & x = a, \end{cases} \tag{B.9}$$

where

$$\begin{aligned} \mathbf{e} &= (1, 0, \ldots, 0)^{\intercal}, \\ \mathbf{E} &= \left(1, e^{m_1(\delta)+v_1(\delta)^2/2}, \ldots, e^{m_J(\delta)+v_J(\delta)^2/2}\right)^{\intercal}, \\ \boldsymbol{\psi}_1(x-a) &= \left(1, (x-a)\mathcal{N}(d_{1,1}), \ldots, (x-a)^J\mathcal{N}(d_{1,J})\right)^{\intercal}, \\ \boldsymbol{\psi}_2(x-a) &= (0, \mathcal{N}(d_{2,1}), \ldots, \mathcal{N}(d_{2,J}))^{\intercal} \end{aligned}$$

with $m_j(\delta)$, $v_j(\delta)$, and $d_{i,j}, i = 1, 2$ given in Eqs. (B.4), (B.5), (B.7), and (B.8), respectively, and $\circ$ denoting the element-wise product between two vectors.

It is worth noting that vectors $\mathbf{E}$ and $\mathbf{p}(R)$ are independent of $x - a$ and thus can be reused in computing Eq. (B.9) for different input pairs (x, a).

Bernstein polynomials

Consider $\boldsymbol{\phi}(x) = \mathbf{b}\big(x/R\big) = \big(b_{J,0}\big(x/R\big), b_{J,1}\big(x/R\big), \ldots, b_{J,J}\big(x/R\big)\big)^{\intercal}$ with $b_{J,j}(v)$ given in Eq. (A.2). We note from Eq. (A.1) that $\mathbf{b}(v) = \mathbf{B}\mathbf{p}(v)$, where $\mathbf{p}(v) = \big(1, v, \ldots, v^J\big)^{\intercal}$ and $\mathbf{B} = (B_{j,\ell})_{0\le j,\ell\le J}$ is a $(J+1)$-by-$(J+1)$ matrix with $B_{j,\ell} = (-1)^{\ell-j}\binom{J}{\ell}\binom{\ell}{j}\mathbb{1}_{\{j\le\ell\le J\}}$.

Under the above choice of $\boldsymbol{\phi}$, Eq. (3.21) reads

$$\begin{aligned}
\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}\left(\hat{S}(x,a,\varepsilon_{t+1})\right)\right] &= \check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{B}\mathbb{E}\left[\mathbf{p}\left(R^{-1}\hat{S}(x,a,\varepsilon_{t+1})\right)\right] \\
&= \check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{B}\left(\mathbf{p}\left(R^{-1}\right)\circ\mathbb{E}\left[\mathbf{p}\left(\hat{S}(x,a,\varepsilon_{t+1})\right)\right]\right) \\
&= \begin{cases}\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{B}\Big[\mathbf{p}\left(R^{-1}\right)\circ\boldsymbol{\psi}_1(x-a)\circ\mathbf{E}+\boldsymbol{\psi}_2(x-a)\Big], & x>a,\\ \check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{B}\mathbf{e}, & x=a,\end{cases}
\end{aligned}$$

where the last equality is by Eq. (B.9) and $\mathbf{e}\circ\mathbf{p}\left(R^{-1}\right)=\mathbf{e}$.

B.1.3 Delta Calculation

This subsection derives an explicit expression for $\partial P_{t,x}(u,y)/\partial y$ with P_t^x given in Eq. (3.29).

Recall from Eq. (3.29) that

$$\begin{aligned}
P_{t,x}(u,y) &= \varphi^{(t+1)\delta-u}\mathbb{E}^{\mathbb{Q}}\left[\hat{V}_{t+1}^{\mathrm{E}}\Big(\min\left[xS_{(t+1)\delta},R\right]\Big)\Big|S_u=y\right] \\
&= \varphi^{(t+1)\delta-u}\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}\Big(\min\left[xS_{(t+1)\delta},R\right]\Big)\Big|S_u=y\right],
\end{aligned}$$

where the last equality is by the definition of $\hat{V}_{t+1}^{\mathrm{E}}$ (Eq. (3.20)). Since S_u follows a Geometric Brownian Motion,

$$P_{t,x}(u,y)=\varphi^{(t+1)\delta-u}\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}\Big(\min\left[xye^{L},R\right]\Big)\right]$$

where $L:=(r-q-\sigma^2/2)\big((t+1)\delta-u\big)+\sigma\sqrt{(t+1)\delta-u}Z$ and Z stands for a standard normal random variable.

Taking derivatives on the both sides of the above equation gives

$$\begin{aligned}
\frac{\partial P_{t,x}(u,y)}{\partial y} &= \varphi^{(t+1)\delta-u}\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\frac{\partial}{\partial y}\mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}\Big(\min\left[xye^{L},R\right]\Big)\right] \\
&= \varphi^{(t+1)\delta-u}x\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}'\Big(xye^{L}\Big)e^{L}\mathbb{1}_{\{xye^{L}<R\}}\right]
\end{aligned}\tag{B.10}$$

with $\boldsymbol{\phi}'$ denoting the derivative of $\boldsymbol{\phi}$. Note that we interchange the derivative and expectation operators in deriving the last equality. By exploiting the dominated convergence theorem, this is legitimate if

$$\mathbb{E}^{\mathbb{Q}}\left[\boldsymbol{\phi}'\Big(xye^{L}\Big)e^{L}\right]<\infty,\quad\text{and}\quad\mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{xye^{L}=R\}}]=0.$$

In the following, we give explicit expressions of the above display under two particular choices of the basis function $\boldsymbol{\phi}$.

Power series

Let $\boldsymbol{\phi}(x) = \mathbf{p}(x) = \big(1, x, \ldots, x^J\big)^\intercal$ and accordingly, $\boldsymbol{\phi}'(x) = \big(0, 1, \ldots, x^{J-1}\big)^\intercal$. Substituting this expression into Eq. (B.10) gives

$$\begin{aligned}\frac{\partial P_{t,x}(u,y)}{\partial y} &= \varphi^{(t+1)\delta-u} x \sum_{j=1}^{J} \check{\beta}_{t+1,j} (xy)^{j-1} \mathbb{E}^{\mathbb{Q}}\left[e^{jL} \mathbb{1}_{\{xye^L<R\}}\right] \\ &= \varphi^{(t+1)\delta-u} y^{-1} \sum_{j=1}^{J} (xy)^{j} \check{\beta}_{t+1,j} \mathbb{E}^{\mathbb{Q}}\left[e^{jL} \mathbb{1}_{\{xye^L<R\}}\right].\end{aligned}$$

Further note that

$$\mathbb{E}^{\mathbb{Q}}\left[e^{jL} \mathbb{1}_{\{xye^L<R\}}\right] = \exp\left[m_j\big((t+1)\delta - u\big) + \frac{v_j^2\big((t+1)\delta-u\big)}{2}\right] \mathcal{N}\left(\tilde{d}_j^x(y)\right), \tag{B.11}$$

where

$$\tilde{d}_j^x(y) = \frac{j \log\left(\frac{R}{xy}\right) - m_j\big((t+1)\delta-u\big) - v_j^2\big((t+1)\delta-u\big)}{v_j\big((t+1)\delta-u\big)}$$

with $m_j(\cdot)$ and $v_j(\cdot)$ given in Eqs. (B.4) and (B.5), respectively.

Combing the last two displays together gives

$$\frac{\partial P_{t,x}(u,y)}{\partial y} = \left(\varphi^{(t+1)\delta-u} y^{-1}\right) \check{\boldsymbol{\beta}}_{t+1}^\intercal \Big[\mathbf{p}\,(xy) \circ \boldsymbol{\eta}_{t,x}(u,y)\Big], \tag{B.12}$$

where

$$\boldsymbol{\eta}_{t,x}(u,y) = \left(0, \mathbb{E}^{\mathbb{Q}}\left[e^{L} \mathbb{1}_{\{xye^L<R\}}\right], \ldots, \mathbb{E}^{\mathbb{Q}}\left[e^{JL} \mathbb{1}_{\{xye^L<R\}}\right]\right)^\intercal$$

is a $(J+1)$-by-1 vector with $\mathbb{E}^{\mathbb{Q}}\left[e^{jL} \mathbb{1}_{\{xye^L<R\}}\right]$ given by Eq. (B.11).

Bernstein polynomials

In the sequel we choose $\boldsymbol{\phi}(x) = \mathbf{b}(x/R)$ with $\mathbf{b}(v)$ being a vector of Bernstein polynomials up to degree J as given in Eq. (A.2). Accordingly, $\boldsymbol{\phi}(x) = \mathbf{B}\big[\mathbf{p}\left(R^{-1}\right) \circ \mathbf{p}(x)\big]$. This combined with Eq. (B.10) gives

$$\begin{aligned}\frac{\partial P_{t,x}(u,y)}{\partial y} &= \varphi^{(t+1)\delta-u}\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{B}\left(\mathbf{p}\left(R^{-1}\right) \circ \frac{\partial}{\partial y}\mathbb{E}^{\mathbb{Q}}\left[\mathbf{p}\Big(\min\left[xye^{L}, R\right]\Big)\right]\right) \\ &= \left(\varphi^{(t+1)\delta-u}y^{-1}\right)\check{\boldsymbol{\beta}}_{t+1}^{\intercal}\mathbf{B}\Big[\mathbf{p}\left(xyR^{-1}\right) \circ \boldsymbol{\eta}_{t,x}(u,y)\Big], \end{aligned} \tag{B.13}$$

where the last equality is by Eq. (B.12).

B.2 Proofs of Statements

B.2.1 Proof of Theorem 3.1

Preliminaries

Recall that the state variable X_t (resp., $\hat{X}_t$) implicitly depends on the DM's action $\mathsf{a} = \{a_t\}_{t\in\mathcal{T}_0}$. In what follows, we sometimes stress such a dependency by writing X_t^{a} (resp., $\hat{X}_t^{\mathsf{a}}$).

Lemma B.1. *For each* $\mathsf{a} \in \bar{\mathcal{A}}$ *defined by Eq.* (3.15),

$$\bigcup_{t=1}^{T}\left\{X_t^{\mathsf{a}} \neq \hat{X}_t^{\mathsf{a}}\right\} \subseteq \left\{X_t^{\mathsf{a}} \in \mathcal{X}_R, \quad \text{for all} \quad 0 \le t \le T\right\}^{c}.$$

Proof of Lemma B.1. Observe that

$$\bigcup_{t=1}^{T}\left\{X_t^{\mathsf{a}} \neq \hat{X}_t^{\mathsf{a}}\right\} = \left(\bigcap_{t=1}^{T}\left\{X_t^{\mathsf{a}} = \hat{X}_t^{\mathsf{a}}\right\}\right)^{c} = \left\{X_t^{\mathsf{a}} = \hat{X}_t^{\mathsf{a}}, \quad \text{for all} \quad 0 \le t \le T\right\}^{c}.$$

In view of the above display, the statement of Lemma B.1 follows if one proves that

$$\left\{X_t^{\mathsf{a}} \in \mathcal{X}_R, \quad \text{for all} \quad 0 \le t \le T\right\} \subseteq \left\{X_t^{\mathsf{a}} = \hat{X}_t^{\mathsf{a}}, \quad \text{for all} \quad 0 \le t \le T\right\},$$

or equivalently, $X_t^{\mathsf{a}} \in \mathcal{X}_R$, for all $0 \le t \le T$, implies

$$X_t^{\mathsf{a}} = \hat{X}_t^{\mathsf{a}}, \quad \text{for all} \quad 0 \le t \le T. \tag{B.14}$$

Next, we would like to prove the above display. Clearly, $X_0^{\mathsf{a}} = \hat{X}_0^{\mathsf{a}} = x_0$ according to Eq. (3.10). Suppose (B.14) holds for time step t. Eq. (3.1) implies

$$X_{t+1}^{\mathsf{a}} = S\left(X_t^{\mathsf{a}}, a_t, \varepsilon_{t+1}\right) = \mathcal{Q}\left(S\left(X_t^{\mathsf{a}}, a_t, \varepsilon_{t+1}\right)\right) = \mathcal{Q}\left(S\left(\hat{X}_t^{\mathsf{a}}, a_t, \varepsilon_{t+1}\right)\right) = \hat{X}_{t+1}^{\mathsf{a}},$$

where the second equality holds because $\mathcal{Q}(x) = x$ for $x \in \mathrm{cl}\left(\mathcal{X}_R\right)$, the third equality is by induction hypothesis that $X_t^{\mathsf{a}} = \hat{X}_t^{\mathsf{a}}$, and the last equality follows by Eq. (3.10). This proves the statement (B.14). The proof of Lemma B.1 is complete. □

A direct consequence of the preceding lemma is that one may control the probability that the state processes X and $\hat{X}$ disagrees with each other by $\mathcal{E}_{x_0,T}(R)$ given in Assumption 3.4.

Corollary B.1. *Under the conditions of Lemma B.1, then,*

$$\sup_{\mathsf{a}\in\bar{\mathcal{A}}} \mathbb{P}\left[\bigcup_{t=1}^{T}\left\{X_t^{\mathsf{a}} \neq \hat{X}_t^{\mathsf{a}}\right\}\right] \leq \mathcal{E}_{x_0,T}(R).$$

Proof of Corollary B.1.

$$\begin{aligned}
\sup_{\mathsf{a}\in\bar{\mathcal{A}}} \mathbb{P}\left[\bigcup_{t=1}^{T}\left\{X_t^{\mathsf{a}} \neq \hat{X}_t^{\mathsf{a}}\right\}\right] &\leq \sup_{\mathsf{a}\in\bar{\mathcal{A}}} \mathbb{P}\Big[\left\{X_t^{\mathsf{a}} \in \mathcal{X}_R, \quad \text{for all} \quad 0 \leq t \leq T\right\}^c\Big] \\
&= \sup_{\mathsf{a}\in\bar{\mathcal{A}}} \Big(1 - \mathbb{P}\Big[X_t^{\mathsf{a}} \in \mathcal{X}_R, \quad \text{for all} \quad 0 \leq t \leq T\Big]\Big) \\
&= 1 - \inf_{\mathsf{a}\in\bar{\mathcal{A}}} \mathbb{P}\Big[X_t^{\mathsf{a}} \in \mathcal{X}_R, \quad \text{for all} \quad 0 \leq t \leq T\Big] \leq \mathcal{E}_{x_0,T}(R),
\end{aligned}$$

where the above two inequalities follows by Lemma B.1 and Assumption 3.4, respectively. □

Recall that sets $\mathcal{A}$ and $\bar{\mathcal{A}}$ are defined in Eqs. (3.2) and (3.15), respectively.

Lemma B.2. *Suppose the conditions of Theorem 3.1 are satisfied. Then, for the state process X defined through Eq. (3.1), the following statements hold.*

(i) *For each $\mathsf{a} \in \mathcal{A}$, there exists $\bar{\mathsf{a}} \in \bar{\mathcal{A}}$ such that $f_t\left(X_t^{\mathsf{a}}, a_t\right) = \bar{f}_t\left(X_t^{\bar{\mathsf{a}}}, \bar{a}_t\right)$ and $f_T\left(X_T^{\mathsf{a}}\right) = f_T\left(X_T^{\bar{\mathsf{a}}}\right)$ for all $t \in \mathcal{T}_0$ almost surely.*

(ii) *For each $\bar{\mathsf{a}} \in \bar{\mathcal{A}}$, there exists $\mathsf{a} \in \mathcal{A}$ such that $f_t\left(X_t^{\mathsf{a}}, a_t\right) \geq \bar{f}_t\left(X_t^{\bar{\mathsf{a}}}, \bar{a}_t\right)$ and $f_T\left(X_T^{\mathsf{a}}\right) = f_T\left(X_T^{\bar{\mathsf{a}}}\right)$ for all $t \in \mathcal{T}_0$ almost surely.*

Proof of Lemma B.2. **(i)** First note that $\mathcal{A} \subset \bar{\mathcal{A}}$. Then Part (i) holds trivially by letting $\bar{\mathsf{a}} = \mathsf{a}$.

(ii) Let us divide the discussion into two cases. If $\bar{a}_0 \in A_0(x_0)$, then one may let $a_0 = \bar{a}_0$ and therefore,

$$f_0\left(X_0^{\mathsf{a}}, a_0\right) = f_0\left(x_0, a_0\right) = \bar{f}_0(x_0, \bar{a}_0) = \bar{f}_0\left(X_0^{\bar{\mathsf{a}}}, \bar{a}_0\right).$$

If $\bar{a}_0 \notin A_0(x_0)$, one may choose an $a_0 \in A_t(x_0)$ such that $S(x_0, a_0, e) = S\left(x_0, \bar{a}_0, e\right)$ for all $e \in \mathbb{D}$. In this case, Parts (i) and (ii) of Assumption 3.5 imply that

$$f_0\left(X_0^{\mathsf{a}}, a_0\right) = f_0\left(x_0, a_0\right) \geq h_0(x_0) = \bar{f}_0(x_0, \bar{a}_0) = \bar{f}_0\left(X_0^{\bar{\mathsf{a}}}, \bar{a}_0\right).$$

Combing the above cases together proves the first statement of Part (ii) for $t = 0$.

For the second claim, it is easy to see that in either of the above cases one gets

$$X_1^{\mathsf{a}} = S\left(X_0^{\mathsf{a}}, a_0, \varepsilon_1\right) = S(x_0, a_0, \varepsilon_1) = S\left(x_0, \bar{a}_0, \varepsilon_1\right) = S\left(X_0^{\bar{\mathsf{a}}}, \bar{a}_0, \varepsilon_1\right) = X_1^{\bar{\mathsf{a}}}.$$

Applying the above argument inductively proves the desired statement. The proof is complete.

□

Remark B.1. *From the above proof, one may see the statement of the preceding Lemma B.2 also holds for the state process $\hat{X}$ defined through Eq. (3.10).*

The preceding Lemma B.2 in conjunction with Remark B.1 implies the following corollary.

Corollary B.2. *Under the conditions of Theorem 3.1,*

$$V_0(x_0) = \sup_{\bar{\mathsf{a}} \in \bar{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t \bar{f}_t(X_t, \bar{a}_t) + \varphi^T f_T(X_T)\right], \tag{B.15}$$

$$\hat{V}_0(x_0) = \sup_{\bar{\mathsf{a}} \in \bar{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t \bar{f}_t\left(\hat{X}_t, \bar{a}_t\right) + \varphi^T f_T\left(\hat{X}_T\right)\right], \tag{B.16}$$

where $V_0(x_0)$ and $\hat{V}_0(x_0)$ are defined in Eqs. (3.3) and (3.12), respectively.

Proof of Corollary B.2. We only sketch the proof. By exploiting Part (i) of the preceding Lemma B.2 and the definition of $V_0(x_0)$ (Eq. (3.3)), it is straightforward to see that

$$V_0(x_0) \leq \sup_{\bar{\mathsf{a}} \in \bar{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t \bar{f}_t(X_t, \bar{a}_t) + \varphi^T f_T(X_T)\right].$$

On the other hand, Part (ii) of Lemma B.2 implies that the above inequality also holds reversely. This proves Eq. (B.15). Eq. (B.16) can be proved in parallel by exploiting Remark B.1. □

The above corollary states that to study the optimal value function $V_0(x_0)$ (resp., $\hat{V}_0(x_0)$), one may switch her attention from its primal definition Eq. (3.3) (resp., Eq. (3.12)) to the problem Eq. (B.15) (resp., Eq. (B.16)) where the state constraint is eliminated and a penalty term is imposed on the reward functions; see also the subsequent paragraph below Assumption 3.5.

Proof of main result

Proof of Theorem 3.1. Corollary B.2 implies that

$$\begin{aligned}
\left|V_0(x_0) - \hat{V}_0(x_0)\right| &\leq \sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t \left|\bar{f}_t\left(X_t, \bar{a}_t\right) - \bar{f}_t\left(\hat{X}_t, \bar{a}_t\right)\right| + \varphi^T \left|f_T\left(X_T\right) - f_T\left(\hat{X}_T\right)\right|\right] \\
&= \sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T-1} \varphi^t \left|\bar{f}_t\left(X_t, \bar{a}_t\right) - \bar{f}_t\left(\hat{X}_t, \bar{a}_t\right)\right| \mathbb{1}_{\{X_t \neq \hat{X}_t\}}\right. \\
&\qquad \left. + \varphi^T \left|f_T\left(X_T\right) - f_T\left(\hat{X}_T\right)\right| \mathbb{1}_{\{X_T \neq \hat{X}_T\}}\right] \\
&\leq \sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}} \mathbb{E}\left[\sum_{t=0}^{T} \varphi^t \left|\bar{B}^{\frac{1}{2}}(X_t) + \bar{\xi}^{\frac{1}{2}}(R)\right| \mathbb{1}_{\{X_t \neq \hat{X}_t\}}\right] := I,
\end{aligned}$$

where the last inequality follows from Assumption 3.5.

Next, we would like to establish an upper bound for the term I. Denote

$$E_0 := \bigcup_{t=1}^{T} \left\{X_t^{\mathsf{a}} \neq \hat{X}_t^{\mathsf{a}}\right\} \quad \text{and} \quad \bar{Y}_t := \bar{B}^{\frac{1}{2}}(X_t) + \bar{\xi}^{\frac{1}{2}}(R).$$

Note that $\left\{X_t \neq \hat{X}_t\right\} \subset E_0$ for $t \in \mathcal{T}$. This implies

$$I \leq \sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}} \mathbb{E}\left[\left(\sum_{t=0}^{T} \varphi^t \bar{Y}_t\right) \cdot \mathbb{1}_{E_0}\right].$$

An application of Cauchy-Schwarz inequality to the R.H.S. of the above display gives

$$\begin{aligned} I &\leq \sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}}\left\{\sqrt{\mathbb{P}\left[E_0\right]}\cdot\left(\mathbb{E}\left[\left(\sum_{t=0}^{T}\varphi^t\bar{Y}_t\right)^2\right]\right)^{\frac{1}{2}}\right\} \\ &\leq \left(\sum_{t=0}^{T}\varphi^{2t}\right)^{\frac{1}{2}}\cdot\sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}}\left\{\sqrt{\mathbb{P}\left[E_0\right]}\cdot\left(\mathbb{E}\left[\sum_{t=0}^{T}\bar{Y}_t^2\right]\right)^{\frac{1}{2}}\right\} \\ &\leq \left(\frac{1-\varphi^{2(T+1)}}{1-\varphi^2}\right)^{\frac{1}{2}}\cdot\sqrt{\mathcal{E}_{x_0,T}(R)}\cdot\sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}}\left\{\left(\mathbb{E}\left[\sum_{t=0}^{T}\bar{Y}_t^2\right]\right)^{\frac{1}{2}}\right\} \\ &= \left\{\frac{1-\varphi^{2(T+1)}}{1-\varphi^2}\cdot\mathcal{E}_{x_0,T}(R)\cdot\sup_{\bar{\mathsf{a}}\in\bar{\mathcal{A}}}\left(\mathbb{E}\left[\sum_{t=0}^{T}\bar{Y}_t^2\right]\right)\right\}^{\frac{1}{2}}, \end{aligned} \tag{B.17}$$

where the last inequality follows from Corollary B.1 and the last equality holds because taking square root preserves monotonicity.

Assumption 3.5 implies that

$$\mathbb{E}\left[\sum_{t=0}^{T}\bar{Y}_t^2\right] \leq 2\sum_{t=0}^{T}\mathbb{E}\left[\bar{\xi}(R)+\bar{B}(X_t)\right] \leq 2(T+1)\left(\bar{\xi}(R)+\zeta_X\right), \tag{B.18}$$

where the first inequality holds because $(a+b)^2 \leq 2a^2+2b^2$ for two real numbers a and b and the last inequality is by Assumption 3.5. Combing (B.17) with (B.18) implies

$$I \leq \sqrt{2(T+1)\frac{1-\varphi^{2(T+1)}}{1-\varphi^2}\mathcal{E}_{x_0,T}(R)\left(\bar{\xi}(R)+\zeta_X\right)}.$$

This proves the desired result. □

B.2.2 Proof of Proposition 3.1

Preliminaries

We first give the definition of the lower (resp., upper) semicontinuity of a correspondence.

Definition B.1 (Lower Semicontinuity)**.** *A correspondence $A:\mathcal{X}\rightrightarrows\mathbb{A}$ is lower semicontinuous at $x\in\mathcal{X}$ if $A(x)\neq\emptyset$ and if, for each sequence $x_n\to x$ and for each $y\in A(x)$, there exists $N\geq 1$ and a sequence $\{y_n\}_{n\geq N}$ such that $y_n\to y$ and $y_n\in A(x_n)$ for all $n\geq N$.*

Definition B.2 (Upper Semicontinuity). *A correspondence $A : \mathcal{X} \rightrightarrows \mathbb{A}$ is upper semicontinuous at $x \in \mathcal{X}$ if $A(x) \neq \emptyset$ and if, for each sequence $x_n \to x$ and for each sequence $\{y_n\}_{n\geq 1}$ such that $y_n \in A(x_n)$ for all $n \geq 1$, there exists a convergent sub-sequence of $\{y_n\}_{n\geq 1}$ whose limit point is in $A(x)$.*

Definition B.3 (Continuous Correspondence). *A correspondence $A : \mathcal{X} \rightrightarrows \mathbb{A}$ is continuous at $x \in \mathcal{X}$ if and only if it is both upper and lower semicontinuous at x. Furthermore, A is called a continuous correspondence if it is continuous at each $x \in \mathcal{X}$.*

Definition B.4 (Compact-valued Correspondence). *A correspondence $A : \mathcal{X} \rightrightarrows \mathbb{A}$ is compact-valued if and only if for each $x \in \mathcal{X}$, $A(x)$ is a compact set.*

Next, we present the Berge's Maximum Theorem; see [18].

Lemma B.3 (Berge's Maximum Theorem). *Let $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathbb{A} \subseteq \mathbb{R}^p$ with $d, p \in \mathbb{N}$. Let $\mathcal{V} : \mathcal{X} \times A \to \mathbb{R}$ be a continuous function and $A : \mathcal{X} \rightrightarrows \mathbb{A}$ be a compact valued and continuous correspondence in accordance with Definitions B.4 and B.3. Then function $V(x) = \max_{a\in A(x)} \mathcal{V}(x, a)$ is well-defined and continuous.*

Proof of main results

Proof of Proposition 3.1. We would like to prove the proposition via a backward induction. First, notice that $V_T(\cdot) = f_T(\cdot)$ which is continuous by Assumption 3.6.

Now assume that $V_{t+1} : \mathcal{X} \to \mathbb{R}$ is continuous. The Bellman equation (3.4) implies that $V_t(x) = \sup_{a\in A_t(x)} \mathcal{V}_t(x, a)$ where

$$\mathcal{V}_t(x, a) := f_t(x, a) + \bar{C}_t(x, a). \tag{B.19}$$

Eq. (3.5) in conjunction with Eq. (3.1) implies that

$$\bar{C}_t(x, a) = \mathbb{E}\left[V_{t+1}(X_{t+1})\middle|X_t = x, a_t = a\right] = \mathbb{E}\Big[V_{t+1}\left(S\big(x, a, \varepsilon_{t+1}\big)\right)\Big] \tag{B.20}$$

which is continuous in $(x, a)^\intercal$ by Assumption 3.6. In view of the last two displays and Assumption 3.6, one gets $\mathcal{V}_t : \mathcal{X}\times\mathbb{A} \to \mathbb{R}$ is continuous. Applying Lemma B.3 implies $V_t : \mathcal{X} \to \mathbb{R}$ is continuous.

A similar argument will lead us to conclude that $\hat{V}_t : \mathcal{X} \to \mathbb{R}$ is continuous. The proof is complete. □

B.2.3 Proofs of Propositions 3.2 and 3.3

Preliminaries

To make the paper self-contained, we give some preliminary definitions.

Definition B.5 (Partial Order)**.** *A partial order is a binary relation $\leq$ over a set $\mathcal{X}$ satisfying the following axioms:*

(i) (Reflexivity) *For each $x \in \mathcal{X}$, $x \leq x$;*

(ii) (Antisymmetry) *For each x and $y \in \mathcal{X}$, if $x \leq y$ and $y \leq x$, then $x = y$;*

(iii) (Transitivity) *For each x, y and $z \in \mathcal{X}$, if $x \leq y$ and $y \leq z$, then $x \leq z$.*

Definition B.6 (Monotone Function)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets equipped with partial orders $\leq_{\mathcal{X}}$ and $\leq_{\mathcal{Y}}$, respectively. $f : \mathcal{X} \to \mathcal{Y}$ is monotone if for all $x, x' \in \mathcal{X}$, $x \leq_{\mathcal{X}} x'$ implies $f(x) \leq_{\mathcal{Y}} f(x')$.*

Definition B.7 (Convex Set)**.** *Let W be a vector space over $\mathbb{R}$. $\mathcal{X} \subset W$ is a convex set if for all $x', x'' \in \mathcal{X}$ and $\lambda \in (0, 1)$, $\lambda x' + (1 - \lambda)x'' \in \mathcal{X}$.*

Definition B.8 (Concave Function)**.** *Let $\mathcal{X}$ be a convex set and $\mathcal{Y}$ a vector space over $\mathbb{R}$ equipped with a partial order $\leq_{\mathcal{Y}}$. $f : \mathcal{X} \to \mathcal{Y}$ is a concave function if for all $x', x'' \in \mathcal{X}$ and $\lambda \in (0, 1)$,*

$$\lambda f\left(x'\right) + (1 - \lambda)\left(x''\right) \leq_{\mathcal{Y}} f\left(\lambda x' + (1 - \lambda)x''\right).$$

Next, we state a lemma from convex analysis whose proof is omitted; see e.g. [20].

Lemma B.4. *Let $\mathcal{X}$ be a convex set and let $\mathcal{B}$ and $\mathcal{C}$ be two vector spaces over $\mathbb{R}$ equipped with partial orders $\leq_{\mathcal{B}}$ and $\leq_{\mathcal{C}}$, respectively. If $f : \mathcal{X} \to \mathcal{B}$ and $h : \mathcal{B} \to \mathcal{C}$ are concave functions with h monotone, then $h \circ f$ is a concave function.*

The following lemma is used in the proof of Proposition 3.3.

Lemma B.5. *Suppose Assumptions 3.1 and 3.8 hold. Then $V_t : \mathcal{X} \to \mathbb{R}$ is bounded uniformly in t, that is, there exists a constant ζ_V such that $\sup_{x \in \mathcal{X}} |V_t(x)| \leq \zeta_V$ for all $t \in \mathcal{T}$.*

Proof. Firstly, by recalling that $V_T = f_T$, Part (ii) of Assumption 3.8 implies that $\sup_{x \in \mathcal{X}} |V_T(x)| \leq \zeta_f$. Inductively assume that $\sup_{x \in \mathcal{X}} |V_{t+1}(x)| \leq (T - t)\zeta_f$.

Recall from Eqs. (3.4) and (3.5) that

$$V_t(x) = \sup_{a \in A_t(x)} \left\{ f_t(x, a) + \mathbb{E}\left[V_{t+1}(X_{t+1}) \middle| X_t = x, a_t = a \right] \right\}.$$

Combing the above display with Part (ii) of Assumption 3.8 gives $V_t(x) \leq (T - t + 1)\zeta_f$ with the R.H.S. independent of x.

Thus, one may choose $\zeta_V = (T + 1)\zeta_f$ such that the assertion of Lemma B.5 holds. The proof is complete. □

Proof of main results

Proof of Proposition 3.2. First note that $V_T(x) = f_T(x)$ is monotone in x by Assumption 3.7. Inductively assume $V_{t+1} : \mathcal{X} \to \mathbb{R}$ is a monotone function.

Let $x, x' \in \mathcal{X}$ such that $x \leq_{\mathcal{X}} x'$. We observe that for each $a \in A_t(x)$,

$$\begin{aligned} \mathcal{V}_t(x,a) &:= f_t(x,a) + \bar{C}_t(x,a) \\ &= f_t(x,a) + \mathbb{E}\Big[V_{t+1}\left(S\big(x,a,\varepsilon_{t+1}\big)\right)\Big] \\ &\leq f_t\left(x',a'\right) + \mathbb{E}\Big[V_{t+1}\left(S\big(x',a',\varepsilon_{t+1}\big)\right)\Big] = \mathcal{V}_t\left(x',a'\right), \end{aligned}$$

where the second equality is by Eqs. (3.1) and (3.5) and the first inequality follows from Assumption 3.7 in conjunction with the monotonicity of $V_{t+1} : \mathcal{X} \to \mathbb{R}$. And thus, for each $a \in A_t(x)$,

$$\mathcal{V}_t(x,a) \leq \mathcal{V}_t\left(x',a'\right) \leq \sup_{a' \in A_t(x')} \mathcal{V}_t\left(x',a'\right) = V_t\left(x'\right),$$

where the last equality is by the Bellman equation (3.4). Since the R.H.S. of the above inequality is independent of $a \in A_t(x)$,

$$V_t(x) = \sup_{a \in A_t(x)} \mathcal{V}_t(x,a) \leq V_t\left(x'\right).$$

This proves the monotonicity of $V_t : \mathcal{X} \to \mathbb{R}$.

Analogously, one can prove that $\hat{V}_t : \mathcal{X} \to \mathbb{R}$ is a monotone function for all $t \in \mathcal{T}_0$. □

Proof of Proposition 3.3. We only prove the assertion for $V_t : \mathcal{X} \to \mathbb{R}$. The assertion for $\hat{V}_t : \mathcal{X} \to \mathbb{R}$ can be proved in parallel.

Firstly, Assumption 3.8 in conjunction with the Bellman equation (3.4) implies that $V_T = f_T$ is a concave and monotone function. As induction hypothesis, one may assume V_{t+1} is a concave and monotone function.

Let

$$\mathcal{V}_t(x,a) := f_t(x,a) + \bar{C}_t(x,a) = f_t(x,a) + \mathbb{E}\Big[V_{t+1}\left(S\big(x,a,\varepsilon_{t+1}\big)\right)\Big],$$

where the last equality is by Eq. (3.5). Recall that $V_{t+1} : \mathcal{X} \to \mathbb{R}$ is concave and monotone (induction hypothesis) and that for each e, $S(\cdot,\cdot,e) : \mathcal{X} \times \mathbb{A} \to \mathcal{X}$ is a concave function (Part (i) of Assumption 3.8). Then Lemma B.4 implies that $\bar{C}_t(x,a)$ is a concave function. This in conjunction with the concavity of $f_t : \mathcal{X} \times \mathbb{A} \to \mathbb{R}$ implies that $\mathcal{V}_t : \mathcal{X} \times \mathbb{A} \to \mathbb{R}$ is a concave function.

Recall from Eq. (3.4) that $V_t(x) = \sup_{a \in A_t(x)} \mathcal{V}_t(x,a)$. In the sequel, we would like to show that $\mathcal{V}_t : \mathcal{X} \to \mathbb{R}$ is a concave function. By Lemma B.5, and V_t is uniformly bounded. This combined with Assumption 3.8 implies that for any pair (x', x'') and any $\epsilon > 0$, there exists $a' \in A_t(x')$ and $a'' \in A_t(x'')$ such that $\lambda a' + (1-\lambda)a'' \in A_t(\lambda x' + (1-\lambda)x'')$,

$$\mathcal{V}_t\left(x', a'\right) + \epsilon > V_t\left(x'\right), \quad \text{and} \quad \mathcal{V}_t\left(x'', a''\right) + \epsilon > V_t\left(x''\right).$$

Consequently, for any $\lambda \in (0,1)$,

$$\begin{aligned} \lambda V_t\left(x'\right) + (1-\lambda) V_t\left(x''\right) &< \lambda \mathcal{V}_t\left(x', a'\right) + (1-\lambda)\mathcal{V}_t\left(x'', a''\right) + \epsilon \\ &\leq \mathcal{V}_t\Big(\lambda x' + (1-\lambda)x'', \lambda a' + (1-\lambda)a''\Big) + \epsilon \\ &\leq V_t\left(\lambda x' + (1-\lambda)x''\right) + \epsilon \end{aligned}$$

where the last second inequality is by the concavity of $\mathcal{V}_t$. Since ϵ can be arbitrarily small by choosing a' and a'', we get

$$\lambda V_t\left(x'\right) + (1-\lambda) V_t\left(x''\right) \leq V_t\left(\lambda x' + (1-\lambda)x''\right).$$

This proves the desired result. □

B.2.4 Proof of Theorem 3.2

Preliminaries

We first give the definitions of "Big O p" and "Small O p" notations.

Definition B.9. **(i)** *For two sequences of random variables* $\{a_M\}_{M\in\mathbb{N}}$ *and* $\{b_M\}_{M\in\mathbb{N}}$*, we say* $a_M = O_{\mathbb{P}}(b_M)$ *if* $\lim_{k\to\infty} \limsup_{M\to\infty} \mathbb{P}\left(|a_M| > k b_M\right) = 0$.

(ii) *Moreover, we say* $a_M = o_{\mathbb{P}}(b_M)$ *if* $\limsup_{M\to\infty} \mathbb{P}\left(|a_M| > k b_M\right) = 0$ *for all* $k > 0$.

Recall that $\check{\boldsymbol{\beta}}_t$ solves the following optimization problem:

$$\inf_{\boldsymbol{\beta}\in\mathbb{R}^{J+1}} \frac{1}{M} \sum_{m=1}^{M} \left[\hat{V}_t^*\left(\hat{X}_t^{(m)}\right) - \boldsymbol{\phi}\left(\hat{X}_t^{(m)}\right)\right]^2 \quad \text{s.t.} \quad \mathbf{A}_J \boldsymbol{\beta} \geq \mathbf{0}_{b(J)}, \tag{B.21}$$

where $\mathbf{A}_J$ is given in Eq. (3.7). Similarly define a *pseudo estimate* $\boldsymbol{\beta}_t^{\mathrm{PE}}$ as the optimizer to the above optimization problem with $\hat{V}_t^*\left(\hat{X}_t^{(m)}\right)$ replaced by $\hat{V}_t\left(\hat{X}_t^{(m)}\right)$. Consequently, define

$$\hat{V}_t^{\mathrm{PE}}(x) = \boldsymbol{\phi}^{\intercal}(x)\boldsymbol{\beta}_t^{\mathrm{PE}}, \quad \text{for} \quad x \in \mathrm{cl}\left(\mathcal{X}_R\right). \tag{B.22}$$

Further define the *oracle* $\boldsymbol{\beta}_t^\star$ as the optimizer of the following problem:

$$\inf_{\boldsymbol{\beta}\in\mathbb{R}^{J+1}} \left\|\boldsymbol{\phi}^\intercal\boldsymbol{\beta} - V_{t+1}\right\|, \quad \text{s.t.} \quad \mathbf{A}_J\boldsymbol{\beta} \geq \mathbf{0}_{b(J)}. \tag{B.23}$$

The first lemma discloses the discrepancy between the oracle and the pseudo estimate.

Lemma B.6. *Under the conditions of Theorem 3.2,*

$$\left|\boldsymbol{\beta}_t^\star - \boldsymbol{\beta}_t^{\mathrm{PE}}\right| = O_{\mathbb{P}}\left(\hat{\rho}_J\right) \quad \text{for} \quad t \in \mathcal{T}.$$

Proof. The proof is parallel to that of [73, Lemma 4] and thus is omitted. □

Remark B.2. *The conclusion of the above lemma is substantially different with that of [73, Lemma 4], although they can be proved by the same argument. In particular, the wedge between the pseudo estimate and the oracle is contributed by one additional term $\sqrt{J/M}$ in [73, Lemma 4], whereas this term does not appear in the above lemma. This is the benefit one may get from directly approximating the value function instead of the continuation function as it is done in [73].*

The subsequent lemma relates the gap between the regression coefficient estimate $\check{\boldsymbol{\beta}}_t$ and the pseudo estimate $\boldsymbol{\beta}_t^{\mathrm{PE}}$ at time step t to the estimation error of the optima value function at time step $t+1$.

Lemma B.7. *Under the conditions of Theorem 3.2, there exits a generic constant $\Gamma > 0$ such that*

$$\left|\check{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{PE}}\right| \leq \left(\Gamma \int \left|\hat{V}_{t+1}(y) - \hat{V}_{t+1}^{\mathrm{E}}(y)\right|^2 \mathrm{d}\mathsf{Q}(y)\right)^{\frac{1}{2}} + O_{\mathbb{P}}\left(\hat{\rho}_J\right)$$

holds with probability approaching 1 as $M \to \infty$ for $t \in \mathcal{T}$, where $\mathsf{Q}(\cdot)$ is the sampling distribution of $\hat{X}_t^{(m)}$ which is independent of m and t.

Proof. Via the argument used by [73] in proving their Lemma 5, one may get

$$\left|\check{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_t^{\mathrm{PE}}\right| \leq \sqrt{\frac{\chi}{M}}\left|\hat{\mathbf{V}}_t - \hat{\mathbf{V}}_t^*\right| + O_{\mathbb{P}}\left(\hat{\rho}_J\right), \quad \text{for some generic} \quad \chi > 0, \tag{B.24}$$

where $\hat{\mathbf{V}}_t$ and $\hat{\mathbf{V}}_t^*$ are two M-by-1 vectors with their m-th elements being $\hat{V}_t\left(\hat{X}_t^{(m)}\right)$ and $\hat{V}_t^*\left(\hat{X}_t^{(m)}\right)$, respectively.

Next it follows from Eq. (3.19) that for each $x \in \mathcal{X}_R$,

$$\begin{aligned}\left|\hat{V}_t(x) - \hat{V}_t^*(x)\right| &\leq \sup_{a \in A_t(x)} \int \left|\left(\hat{V}_{t+1}(y) - \hat{V}_{t+1}^{\mathrm{E}}(y)\right) w(y; x, a)\right| \mathrm{d}\mathsf{Q}(y) \\ &\leq \zeta_w \int \left|\hat{V}_{t+1}(y) - \hat{V}_{t+1}^{\mathrm{E}}(y)\right| \mathrm{d}\mathsf{Q}(y)\end{aligned}$$

where the second inequality is guaranteed by Assumption 3.9. It is worth stressing that the R.H.S. of the above inequality is independent of x and m.

The above display in conjunction with Jensen's inequality implies

$$\left|\hat{V}_t\left(\hat{X}_t^{(m)}\right) - \hat{V}_t^*\left(\hat{X}_t^{(m)}\right)\right|^2 \leq \zeta_w^2 \int \left|\hat{V}_{t+1}(y) - \hat{V}_{t+1}^{\mathrm{E}}(y)\right|^2 \mathrm{d}\mathsf{Q}(y),$$

and therefore,

$$\left|\hat{\mathbf{V}}_t - \hat{\mathbf{V}}_t^*\right| = \left(\sum_{m=1}^{M} \left|\hat{V}_t\left(\hat{X}_t^{(m)}\right) - \hat{V}_t^*\left(\hat{X}_t^{(m)}\right)\right|^2\right)^{\frac{1}{2}} \leq \zeta_w \sqrt{M} \left(\int \left|\hat{V}_{t+1}(y) - \hat{V}_{t+1}^{\mathrm{E}}(y)\right|^2 \mathrm{d}\mathsf{Q}(y)\right)^{\frac{1}{2}}.$$

Coming the above display with inequality (B.24) yields the desired result with $\Gamma = \chi\zeta_w$. □

We give some intuition behind the result delivered by the above lemma. Recall that the difference between $\check{\boldsymbol{\beta}}_t$ and $\boldsymbol{\beta}_t^{\mathrm{PE}}$ comes from the mismatch between $\hat{V}_t^*$ and $\hat{V}_t$. In view of the definitions of these two functions (see Eqs. (3.19) and (3.13), respectively), one may see that their discrepancy is essentially triggered by the estimation error of the value function at time step $t+1$, i.e., the disagreement between $\hat{V}_{t+1}^{\mathrm{E}}$ and $\hat{V}_{t+1}$. Thus, it is not surprising that the conclusion of Lemma B.7 holds.

Proof of main results

Proof of Theorem 3.2. First recall that $\hat{V}_T^*(x) = \hat{V}_T(x) = f_T(x)$. In view of this and the definitions of $\boldsymbol{\beta}_t^{\mathrm{PE}}$ and $\check{\boldsymbol{\beta}}_t$, one gets

$$\hat{V}_T^{\mathrm{E}}(x) = \boldsymbol{\phi}^\intercal(x)\check{\boldsymbol{\beta}}_T = \boldsymbol{\phi}^\intercal(x)\boldsymbol{\beta}_T^{\mathrm{PE}} = \hat{V}_T^{\mathrm{PE}}(x) \quad \text{for} \quad x \in \mathrm{cl}\left(\mathcal{X}_R\right),$$

and thus,

$$\begin{aligned}\left|\hat{V}_T(x)-\hat{V}_T^{\mathrm{E}}(x)\right| &= \left|\hat{V}_T(x)-\hat{V}_T^{\mathrm{PE}}(x)\right| \\ &\leq \left|\hat{V}_T(x)-\boldsymbol{\phi}^{\intercal}(x)\boldsymbol{\beta}_T^{\star}\right|+\left|\boldsymbol{\phi}^{\intercal}(x)\left(\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right)\right| \\ &\leq \left|\boldsymbol{\phi}^{\intercal}(x)\left(\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right)\right|+O\left(\hat{\rho}_J\right),\end{aligned}$$

where the last inequality follows because Condition (iii) of Assumption 3.2 holds for $g(\cdot)=\hat{V}_T(\cdot)$ and $\boldsymbol{\beta}^{\star}=\boldsymbol{\beta}_T^{\star}$. And therefore,

$$\begin{aligned}\int\left|\hat{V}_T(x)-\hat{V}_T^{\mathrm{E}}(x)\right|^2\mathrm{d}\mathsf{Q}(x) &\leq 2\left(\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right)^{\intercal}\int\boldsymbol{\phi}(x)\boldsymbol{\phi}^{\intercal}(x)\mathrm{d}\mathsf{Q}(x)\left(\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right)+2O\left(\hat{\rho}_J^2\right) \\ &= 2\left(\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right)^{\intercal}\mathbb{E}\left[\boldsymbol{\phi}\left(X_T^{(m)}\right)\boldsymbol{\phi}^{\intercal}\left(X_T^{(m)}\right)\right]\left(\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right)+O\left(\hat{\rho}_J^2\right) \\ &\leq 2\bar{c}_{\Phi}\left|\boldsymbol{\beta}_T^{\star}-\boldsymbol{\beta}_T^{\mathrm{PE}}\right|^2+O\left(\hat{\rho}_J^2\right) \qquad \text{(B.25)}\end{aligned}$$

where the first equality follows because $\mathsf{Q}(\cdot)$ is the sampling distribution of $X_T^{(m)}$ and the last inequality is by Condition (iv) of Assumption 3.2. The last display in conjunction with Lemma B.6 implies $\int\left|\hat{V}_T(x)-\hat{V}_T^{\mathrm{E}}(x)\right|^2\mathrm{d}\mathsf{Q}(x)=O_{\mathbb{P}}\left(\hat{\rho}_J^2\right)$.

Now assume that

$$\left(\int\left|\hat{V}_{t+1}(x)-\hat{V}_{t+1}^{\mathrm{E}}(x)\right|^2\mathrm{d}\mathsf{Q}(x)\right)^{\frac{1}{2}}=O_{\mathbb{P}}\left(\zeta^{T-t-1}\hat{\rho}_J\right)\quad\text{for some}\quad\zeta>0. \qquad \text{(B.26)}$$

Applying triangle inequality implies

$$\left|\hat{V}_t(x)-\hat{V}_t^{\mathrm{E}}(x)\right|\leq\left|\hat{V}_t(x)-\hat{V}_t^{\mathrm{PE}}(x)\right|+\left|\hat{V}_t^{\mathrm{PE}}(x)-\hat{V}_t^{\mathrm{E}}(x)\right|. \qquad \text{(B.27)}$$

An argument similar to that used in establishing (B.25) implies

$$\int\left|\hat{V}_t(x)-\hat{V}_t^{\mathrm{PE}}(x)\right|^2\mathrm{d}\mathsf{Q}(x)=O_{\mathbb{P}}\left(\hat{\rho}_J^2\right). \qquad \text{(B.28)}$$

Next, we would like to investigate the second term in the R.H.S. of (B.27). By exploiting Lemma

B.7, one gets

$$\begin{aligned}\int \left|\hat{V}_t^{\mathrm{PE}}(x) - \hat{V}_t^{\mathrm{E}}(x)\right|^2 \mathrm{d}\mathsf{Q}(x) &= \left(\boldsymbol{\beta}_t^{\mathrm{PE}} - \check{\boldsymbol{\beta}}_t\right)^{\intercal} \mathbb{E}\left[\boldsymbol{\phi}\left(\hat{X}_t^{(m)}\right)\boldsymbol{\phi}^{\intercal}\left(\hat{X}_t^{(m)}\right)\right]\left(\boldsymbol{\beta}_t^{\mathrm{PE}} - \check{\boldsymbol{\beta}}_t\right) \\ &\leq \bar{c}_{\Phi}\left|\boldsymbol{\beta}_t^{\mathrm{PE}} - \check{\boldsymbol{\beta}}_t\right|^2 \\ &\leq \bar{c}_{\Phi}\Gamma \int \left|\hat{V}_{t+1}(y) - \hat{V}_{t+1}^{\mathrm{E}}(y)\right|^2 \mathrm{d}\mathsf{Q}(y) + O_{\mathbb{P}}\left(\hat{\rho}_J\right) \\ &= \bar{c}_{\Phi}\Gamma O_{\mathbb{P}}\left(\zeta^{2(T-t-1)}\hat{\rho}_J^2\right) + O_{\mathbb{P}}\left(\hat{\rho}_J^2\right), \end{aligned} \tag{B.29}$$

where the last equality follows by the induction hypothesis (B.26). Combing inequalities (B.27), (B.28), and (B.29) yields:

$$\left(\int \left|\hat{V}_t(x) - \hat{V}_t^{\mathrm{E}}(x)\right|^2 \mathrm{d}\mathsf{Q}(x)\right)^{\frac{1}{2}} = O_{\mathbb{P}}\left(\zeta^{T-t}\hat{\rho}_J\right) \quad \text{for all} \quad t \in \mathcal{T}. \tag{B.30}$$

Now it is ready to establish a bound for $\left|\hat{V}_0(x_0) - \hat{V}_0^{\mathrm{E}}(x_0)\right|$. We observe that

$$\left|\hat{V}_0(x_0) - \hat{V}_0^{\mathrm{PE}}(x_0)\right| \leq \left|\boldsymbol{\phi}^{\intercal}(x_0)\left(\boldsymbol{\beta}_0^{\star} - \boldsymbol{\beta}_0^{\mathrm{PE}}\right)\right| + \left|\boldsymbol{\phi}^{\intercal}(x_0)\left(\boldsymbol{\beta}_0^{\star} - \boldsymbol{\beta}_0^{\mathrm{PE}}\right)\right| = O_{\mathbb{P}}\left(\hat{\rho}_J\right).$$

Also note that

$$\begin{aligned}\left|\hat{V}_0^{\mathrm{PE}}(x_0) - \hat{V}_0^{\mathrm{E}}(x_0)\right| &= \left|\boldsymbol{\phi}(x_0)\right|\left|\boldsymbol{\beta}_0^{\mathrm{PE}} - \check{\boldsymbol{\beta}}_0\right| \\ &\leq \left|\boldsymbol{\phi}(x_0)\right|\left[\left(\Gamma \int \left|\hat{V}_1(x) - \hat{V}_1^{\mathrm{E}}(x)\right|^2 \mathrm{d}\mathsf{Q}(x)\right)^{\frac{1}{2}} + O_{\mathbb{P}}\left(\hat{\rho}_J\right)\right] \\ &= O_{\mathbb{P}}\left(\zeta^{T-1}\hat{\rho}_J\right)\end{aligned}$$

where the last two inequalities follow by Lemma B.7 and (B.30), respectively. Combing the last two displays yields the desired result. The proof is complete. □

Appendix C

Appendix for Chapter 4

C.1 Expressions of Transition Functions

In this appendix, we give the expressions of transition functions defined in Section 4.2.2.

C.1.1 Transition function accompanying the synthetic contract

Let $x = (W, Z, A, B, I)$. The transition function accompanying the state process X_n across the event time t_n is given by

$$K_n(x, \pi_n) = \Big(K_{n,i}(x, \pi_n)\Big)_{1 \le i \le 5},$$

where

$$
\begin{aligned}
K_{n,1}(x,\pi_n) &= \big(W-\gamma_n\big)^+(1-\eta),\\
K_{n,2}(x,\pi_n) &= Z\mathbb{1}_{\{\tau_n=0\}} + K_{n,1}(x,\pi_n)\mathbb{1}_{\{\tau_n=1\}},
\end{aligned}
\tag{C.1}
$$

$$
K_{n,3}(x,\pi_n) = \begin{cases} A\vee Z, & \text{if } \tau_n = 0,\\ A\vee\big(W-\gamma_n\big), & \text{if } I=0, 0<\gamma_n\le \tilde{G}_n(I)A(t_n),\\ A\vee Z, & \text{if } I>0, 0\le\gamma_n\le \tilde{G}_n(I)A(t_n),\\ \frac{W-\gamma_n}{W-\tilde{G}_n(I)A}A\vee(W-\gamma_n), & \text{otherwise,}\end{cases}
\tag{C.2}
$$

$$
K_{n,4}(x,\pi_n) = \begin{cases} \big(B-\gamma_n\big)^+, & \text{if } 0\le\gamma_n\le\tilde{G}_n(I)A,\\ \frac{W-\gamma_n}{W-\tilde{G}_n(I)A}\big(B-\tilde{G}_n(I)A\big)^+, & \text{otherwise,}\end{cases}
\tag{C.3}
$$

$$
K_{n,5}(x,\pi_n) = \begin{cases} n, & \text{if } I=0 \text{ and } \tau_n=1,\\ I, & \text{otherwise.}\end{cases}
\tag{C.4}
$$

C.1.2 Transition function accompanying the real contract

The transition function of the real contract differs from that of the synthetic contract only in its first component. Specifically, it is given by $\bar{K}_n(x,\pi_n) = \Big(\bar{K}_{n,i}(x,\pi_n)\Big)_{0\le i\le 5}$, where

$$
\bar{K}_{n,i}(x,\pi_n) = K_{n,i}(x,\pi_n), \quad \text{for } \ 2\le i\le 5,
$$

and

$$
\bar{K}_{1,n}(x,\pi_n) = \max\Big[\,(W-\gamma_n)^+ - \eta K_{n,3}(x,\pi_n), 0\Big].
$$

C.2 Some Explicit Results

For the synthetic contract, this appendix presents some closed-form expressions of the value function and the continuation function. In what follows, we stipulate $\kappa_{N-2}=0$ for the synthetic contract because it is the case in the real contract specification in Table 4.2.

C.2.1 Expression of $V_{N-1}(\cdot)$

According to the dynamic programming equation Eq. (4.19), direct calculation gives

$$\begin{aligned} V_{N-1}(x) &= {}_{N-2}p_0 q_{N-2}\big(B \vee W\big) \\ &\quad + \max_{\gamma \in \Gamma_{N-1,2}} \Big\{ {}_{N-1}p_0 \gamma + {}_{N-1}p_0 \varphi \mathbb{E}^{\mathbb{Q}} \Big[W(t_N) \Big| W(t_{N-1}^+) = \big(W-\gamma\big)^+ (1-\eta) \Big] \Big\} \\ &= {}_{N-2}p_0 q_{N-2}\big(B \vee W\big) + \max_{\gamma \in \Gamma_{N-1,2}} \Big[{}_{N-1}p_0 \gamma + {}_{N-1}p_0 \varphi \big(W-\gamma\big)^+ (1-\eta) \Big] \\ &= {}_{N-2}p_0 q_{N-2}\big(B \vee W\big) + {}_{N-1}p_0 \Big(\tilde{G}_{N-1}(I) A \vee W \Big). \end{aligned}$$

Remark C.1. *The first equality of the above display states that it is not optimal to defer withdrawal at t_{N-1}. This is because the PH's investment is subject to insurance fee as reflected by the factor $(1-\eta)$ in the above display, whereby withdrawing the investment is free of withdrawal charge if $\kappa_{N-2} = 0$.*

C.2.2 Expression of $C_{N-2}(\cdot)$

This subsection aims to derive an analytical expression for $C_{N-2}(\cdot)$. This result is used in conducting the numerical experiment in Section 4.5.2 where one needs to compare the regression estimates with the true continuation function.

Let $k = (k_1, k_2, k_3, k_4, k_5)$. The continuation function at time t_{N-2} is given by

$$\begin{aligned} C_{N-2}(k) &= \mathbb{E}^{\mathbb{Q}} \Big[V_{N-1}(X_{N-1}) \Big| W(t_{N-2}^+) = k_1, A(t_{N-1}) = k_3, B(t_{N-1}) = k_4 \Big] \\ &= {}_{N-1}p_0 \mathbb{E}^{\mathbb{Q}} \Big[\tilde{G}_{N-1}(k_5) k_3 \vee k_1 e^{L_{N-2}} \Big] + {}_{N-2}p_0 q_{N-2} \mathbb{E}^{\mathbb{Q}} \Big[k_4 \vee k_1 e^{L_{N-2}} \Big] \\ &= {}_{N-1}p_0 \Big\{ \tilde{G}_{N-1}(k_5) k_3 + \mathbb{E}^{\mathbb{Q}} \Big[\Big(k_1 e^{L_{N-2}} - \tilde{G}_{N-1}(k_5) k_3 \Big)^+ \Big] \Big\} \\ &\quad + {}_{N-2}p_0 q_{N-2} \Big\{ k_4 + \mathbb{E}^{\mathbb{Q}} \Big[\big(k_1 e^{L_{N-2}} - k_4 \big)^+ \Big] \Big\}, \end{aligned}$$

where L_{N-2} denotes the log-return of the policy fund over time horizon $[t_{N-2}, t_{N-1}]$.

When the underlying fund follows a geometric Brownian motion, one has

$$\begin{aligned} C_{N-2}(k) &= {}_{N-1}p_0 \Big[\tilde{G}_{N-1}(k_5) k_3 + \varphi^{-1} \mathcal{BS}_C \Big(k_1, \tilde{G}_{N-1}(k_5) k_3 \Big) \Big] \\ &\quad + {}_{N-2}p_0 q_{N-2} \Big[k_4 + \varphi^{-1} \mathcal{BS}_C \big(k_1, k_4 \big) \Big], \end{aligned} \tag{C.1}$$

where one should recall that $\varphi = e^{-r}$ and $\mathcal{BS}_C(S, K)$ denotes the Black-Scholes formula for an European call option with spot price S, strike price K, and time to expiry 1.

C.3 Form of Matrix A in (4.28)

In this appendix, we give the specific form of matrix $\mathbf{A}$ in the quadratic programming problem (4.28).

Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \end{pmatrix},$$

with $\mathbf{A}^{(i)}, i = 1, 2$ being $(J^2 + J)$-by-$(J + 1)^2$ sub-matrices. Thus $\mathbf{A}$ is a $2(J^2 + J)$-by-$(J + 1)^2$ matrix.

The sub-matrix $\mathbf{A}^{(1)}$ is given by

$$\mathbf{A}^{(1)} = \begin{pmatrix} -\mathbf{I}_{J+1} & \mathbf{I}_{J+1} & & & \\ \mathbf{I}_{J+1} & -2\mathbf{I}_{J+1} & \mathbf{I}_{J+1} & & \\ & & \ddots & & \\ & & \mathbf{I}_{J+1} & -2\mathbf{I}_{J+1} & \mathbf{I}_{J+1} \end{pmatrix}_{(J^2+J)\times(J+1)^2},$$

where $\mathbf{I}_{J+1}$ is a $(J + 1)$-by-$(J + 1)$ identical matrix.

The sub-matrix $\mathbf{A}^{(2)}$ is given by

$$\mathbf{A}^{(2)} = \begin{pmatrix} \mathbf{B}_J & & & \\ & \mathbf{B}_J & & \\ & & \ddots & \\ & & & \mathbf{B}_J \end{pmatrix}_{(J^2+J)\times(J+1)^2},$$

with

$$\mathbf{B}_J = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & -2 & 1 \end{pmatrix}_{J\times(J+1)}.$$

C.4 Some Proofs

This appendix section collects the proofs of Theorems 4.1, 4.2 and Propositions 4.1, 4.3.

C.4.1 Preliminaries

Below we present some preliminary technical results which will be useful in the proofs of our main results. In what follows, denote $\mathbb{R}^n_+ := [0,\infty)^n$ for some positive integer n and $x = (W, Z, A, B, I)$, respectively. The definitions of partial order and the monotonicity of a multivariate function are given as follows.

Definition C.1 (Partial Order)**.** *Let m be some integer bigger than 1. For two vectors $(y_1, y_2, \ldots, y_m)$ and $(z_1, z_2, \ldots, z_m) \in \mathbb{R}^m_+$, we say $(y_1, y_2, \ldots, y_m) \leq (z_1, z_2, \ldots, z_m)$ if and only if $y_j \leq z_j$ for all $1 \leq j \leq m$.*

Definition C.2 (Monotonicity)**.** *We say a function $F : \mathbb{R}^m_+ \longrightarrow \mathbb{R}$ is monotone if and only if for any $y, z \in \mathbb{R}^m_+$, $y \leq z$ implies $F(y) \leq F(z)$.*

The following two technical lemmas come from convex analysis; see e.g. [20].

Lemma C.1. *Let $\mathcal{A} \subset \mathbb{R}$ be a convex set, and let $\mathcal{B}$ and $\mathcal{C}$ be vector spaces equipped with partial orders $\leq_{\mathcal{B}}$ and $\leq_{\mathcal{C}}$, respectively. If $h_1 : \mathcal{A} \longrightarrow \mathcal{B}$ and $h_2 : \mathcal{B} \longrightarrow \mathcal{C}$ are convex functions with h_2 monotone, then $h_2 \circ h_1$ is a convex function.*

Lemma C.2. *Suppose function $F(\cdot,\cdot) : \mathbb{R}^{n-1}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}$ for some integer $n > 1$ satisfies:*

(i) for each $z \in \mathbb{R}_+$, $y \in \mathbb{R}^{n-1}_+ \longmapsto F(y, z)$ is convex, and

(ii) for each $c > 0$, $y \in \mathbb{R}^{n-1}_+$ and $z \in \mathbb{R}_+$, $F(cy, cz) = cF(y, z)$.

Then, $F(y, z)$ is a convex function.

Lemma C.3. *For $\gamma_n = 0$ or $\gamma_n = \tilde{G}_n(I)A$, $\big(W, Z, A, B\big) \longmapsto K_n(x, \pi_n)$ and $\big(W, Z, A, B\big) \longmapsto f_n(x, \pi_n)$ are convex with $\tilde{G}_n(\cdot)$, $K_n(\cdot,\cdot)$, and $f_n(\cdot,\cdot)$ defined in Eq. (4.2), Appendix C.1.1 and Eq. (4.15), respectively.*

Proof. The proof follows from checking the definition of convexity by routine calculation. □

Lemma C.4. *For each $n \in \mathcal{I}$ and $I \in \{0, 1, \ldots, n-1\}$, $V_n\big(W, Z, A, B, I\big) \leq V_n\left(W', Z', A', B', I\right)$ if $W \leq W'$, $Z \leq Z'$, $A \leq A'$, and $B \leq B'$.*

Proof. This lemma can be proved in a manner similar to the one used in proving Theorem 4.2. The proof is thus omitted. □

To proceed, denote

$$\mathcal{R} := \left\{ \left(W, Z, A, B\right) \in \mathbb{R}_+^4 : \quad W \leq Z \right\}. \tag{C.2}$$

Lemma C.5. *Suppose for each I $(W, Z, A, B) \longmapsto V_{n+1}(W, Z, A, B, I)$ is C.M. on $\mathcal{R}$. Then, for each k_5, $(k_1, k_2, k_3, k_4) \longmapsto C_n(k)$ is C.M. on $\mathbb{R}_+^4$ with $k := (k_1, k_2, k_3, k_4, k_5)$ and $C_n(\cdot)$ defined in Eq. (4.18).*

Proof. We first prove the convexity of $(k_1, k_2, k_3, k_4) \longmapsto C_n(k)$. Let $k' = \left(k_1', k_2', k_3', k_4', k_5\right)$. For any $\lambda \in (0, 1)$, direct calculation gives

$$\begin{aligned}
C_n\Big(\lambda k + (1-\lambda)k'\Big) &= \mathbb{E}^{\mathbb{Q}}\left[V_{n+1}\left(\tilde{k}_1 e^{L_n}, \tilde{k}_2 \vee \tilde{k}_1 e^{\bar{L}_n}, \tilde{k}_3, \tilde{k}_4, k_5\right)\right] \\
&\leq \mathbb{E}^{\mathbb{Q}}\left\{V_{n+1}\Big(\tilde{k}_1 e^{L_n}, \lambda\left(k_2 \vee k_1 e^{\bar{L}_n}\right) + (1-\lambda)\left(k_2' \vee k_1' e^{\bar{L}_n}\right), \tilde{k}_3, \tilde{k}_4, k_5\Big)\right\} \\
&\leq \lambda \mathbb{E}^{\mathbb{Q}}\left[V_{n+1}\left(k_1 e^{L_n}, k_2 \vee k_1 e^{\bar{L}_n}, k_3, k_4, k_5\right)\right] \\
&\quad + (1-\lambda)\mathbb{E}^{\mathbb{Q}}\left[V_{n+1}\left(k_1' e^{L_n}, k_2' \vee k_1' e^{\bar{L}_n}, k_3', k_4', k_5\right)\right] \\
&= \lambda C_n\left(k\right) + (1-\lambda) C_n\left(k'\right),
\end{aligned}$$

where $\tilde{k}_i := \lambda k_i + (1-\lambda)k_i'$, $0 \leq i \leq 4$, and the last two inequalities follow by the convexity of $(x, y) \longmapsto x \vee y e^{L_n}$ and $(W, Z, A, B) \longmapsto V_{n+1}(x)$, respectively.

The monotonicity of $k \longmapsto C_n(k)$ can be proved in a similar way by exploiting Lemma C.4. This completes the proof. □

C.4.2 Proof of Proposition 4.1

Proof of Proposition 4.1. Only the statements for $V_n(\cdot)$ will be proved because the results for $\bar{V}_n(\cdot)$ follow in parallel. A backward induction strategy is adopted for the proof.

In the following, denote $x := (W, Z, A, B, I)$, $x^c := (c \cdot W, c \cdot Z, c \cdot A, c \cdot B, I)$ and $\pi_n^c := (c \cdot \gamma_n, \tau_n)$ for some $c > 0$. Also recall that $\pi_n = (\gamma_n, \tau_n)$ and that $V_N(x) = {}_{N-1}p_0 W$ which obviously exhibits the positive homogeneity. Inductively assume that $V_{n+1}(x)$ has the homogeneity and need to show so does $V_n(x)$. Firstly, routine calculation in conjunction with Eq. (4.15) yields

$$f_n\left(x^c, \pi_n^c\right) = c \cdot f_n\left(x, \pi_n\right). \tag{C.3}$$

Further, it is easy to see from the expression of $K_n(\cdot,\cdot)$ given in Appendix C.1.1 that $K_n\left(x^c,\pi_n^c\right) = c\cdot K_n\left(x,\pi_n\right)$. This combined with Eq. (4.18) implies

$$\begin{aligned} C_n\Big(K_n\big(x^c,\pi_n^c\big)\Big) &= \mathbb{E}\left[V_{n+1}(X_{n+1})\Big|X_{n^+} = c\cdot K_n\left(x,\pi_n\right)\right] \\ &= \mathbb{E}\left[V_{n+1}\left(X_{n+1}^c\right)\Big|X_{n^+} = K_n\left(x,\pi_n\right)\right] \\ &= \mathbb{E}\left[c\cdot V_{n+1}\left(X_{n+1}\right)\Big|X_{n^+} = K_n\left(x,\pi_n\right)\right] = c\cdot C_n\Big(K_n\big(x,\pi_n\big)\Big) \end{aligned}$$

with $X_{n+1}^c := (c\cdot W(t_{n+1}), c\cdot Z(t_{n+1}), c\cdot A(t_{n+1}), c\cdot B(t_{n+1}), I_{n+1})$, where the second equality holds because $A(t_{n+1}) = A(t_n^+)$, $B(t_{n+1}) = B(t_n^+)$, and the distribution of $W(t)/W(t_n^+)$ is independent of $W(t_n^+)$ for $t\in(t_n,t_{n+1}]$ and the third equality is by the induction hypothesis.

The above equation combined with Eq. (C.3) implies

$$\mathcal{V}_n\left(x^c,\pi_n^c\right) = c\cdot\mathcal{V}_n\left(x,\pi_n\right), \tag{C.4}$$

where one should recall from Eq. (4.19) that

$$\mathcal{V}_n(x,\pi_n) = f_n(x,\pi_n) + \varphi C_n\Big(K_n\left(x,\pi_n\right)\Big). \tag{C.5}$$

In view of Eq. (C.4), it suffices to show $(\hat\gamma_n^c,\hat\tau_n^c) = (c\cdot\hat\gamma_n,\hat\tau_n)$, where $(\hat\gamma_n^c,\hat\tau_n^c)$ and $(\hat\gamma_n,\hat\tau_n)$ denote the optimal policy at $X_n = x^c$ and $X_n = x$, respectively. Now divide the discussion into two cases with $I=0$ and $I>0$, respectively. The proofs for the two cases are similar so we focus on the first case.

For $I=0$, the feasible set of the decision variable is $D_{n,1}\cup D_{n,2}$ where one should recall that $D_{n,1}$ and $D_{n,2}$ are given in Eqs. (4.4) and (4.5), respectively. The supremum value of the objective function over $D_{n,2}$ is given by

$$\hat\gamma_{1,n}^c := \arg\sup_{\gamma_n\in\Gamma_{n,2}(x^c)}\mathcal{V}_n\Big(x^c,(\gamma_n,1)\Big) = c\cdot\arg\sup_{(\gamma_n/c)\in\Gamma_{n,2}(x)}\mathcal{V}_n\Big(x,(\gamma_n/c,1)\Big) := c\cdot\hat\gamma_{1,n}.$$

Comparing this value with the objective value over $D_{n,1}$, the optimal solution over whole feasible set D_n is obtained by

$$\begin{aligned} (\hat\gamma_n^c,\hat\tau_n^c) &= \begin{cases}(\hat\gamma_{1,n}^c,1), & \text{if } \mathcal{V}_n\left(x^c,\left(\hat\gamma_{1,n}^c,1\right)\right) > \mathcal{V}_n\left(x^c,(0,0)\right),\\ (0,0), & \text{otherwise,}\end{cases} \\ &= \begin{cases}(c\cdot\hat\gamma_{1,n},1), & \text{if } c\mathcal{V}_n\left(x,(\hat\gamma_{1,n},1)\right) > c\mathcal{V}_n\left(x,(0,0)\right),\\ (0,0), & \text{otherwise,}\end{cases} \\ &= (c\cdot\hat\gamma_n,\hat\tau_n), \end{aligned}$$

where the second equality follows from Eq. (C.4). The proof of Proposition 4.1 is complete. □

C.4.3 Proof of Theorem 4.1

Proof of Theorem 4.1. Let $x := \big(W, Z, A, B, I\big)$, $x' := (W', Z', A, B', I)$, and

$$\tilde{x} = \lambda x + (1-\lambda)x' := \Big(\tilde{W}, \tilde{Z}, A, \tilde{B}, I\Big).$$

Further denote $k := (k_1, k_2, k_3, k_4, k_5)$.

Assertion for $n = N-1$ The assertion of Theorem 4.1 for $n = N-1$ is proved in two steps.

(1) In the first step, we show that the optimal withdrawal amount at x is among three choices:

$$\hat{\gamma}_{N-1}(x) \in \big\{0, G_{N-1}(I)A, W\big\}. \tag{C.6}$$

In the sequel we split the discussion into two cases.

$I > 0$ If $I > 0$, one must have $\hat{\tau}_{N-1}(x) = 1$ and $\xi_{N-1}(I) = I$ according to Eqs. (4.5) and (4.1), respectively. One should also recall from Eq. (4.5) that the feasible set of γ_{N-1} at $X_{N-1} = x$ is given by

$$\Gamma_{N-1,2}(x) = \begin{cases} \big[0, G(I)A\big], & \text{if } 0 \le W < G(I)A, \\ [0, W], & \text{otherwise.} \end{cases}$$

It is straightforward to see $\gamma_{N-1} \longmapsto K_{N-1,1}\big(x, (\gamma_{N-1}, 1)\big)$ is convex on $\Gamma_{N-1,2}(x)$. Furthermore, Lemma C.5 implies,

$$k_1 \longmapsto C_{N-1}(k) = \mathbb{E}\left[{}_{N-1}p_0 W(t_N) \middle| W(t_{N-1}^+) = k_1\right]$$

is C.M.. Combining this with Lemma C.1 implies

$$\gamma_{N-1} \longmapsto C_{N-1}\Big(K_{N-1}\big(x, (\gamma_{N-1}, 1)\big)\Big)$$

is convex on $\Gamma_{N-1,2}(x)$.

It is easy to see $\gamma_{N-1} \longmapsto f_{N-1}\big(x, (\gamma_{N-1}, 1)\big)$ is convex on $\big[0, G(I)A\big]$ and $\big[G(I)A, W\big]$, respectively, if $W > G(I)A$. Accordingly, $\gamma_{N-1} \longmapsto \mathcal{V}_{N-1}\big(x, (\gamma_{N-1}, 1)\big)$ (see Eq. (C.5)) is convex on $\Gamma_{N-1,2}(x) = \big[0, G(I)A\big]$, if $W \le G(I)A$; otherwise it is convex on $\big[0, G(I)A\big]$ and

$\big[G(I)A, W\big]$, respectively. Thus its supremume over $\Gamma_{N-1,2}(x)$ is obtained at $0, G(I)A$, or W, as desired.

I=0 In the second case $I = 0$, the feasible set of τ_{N-1} is $\{0, 1\}$. Accordingly, the optimal decision at $X_n = x$ is either $(0, 0)$ or $(\tilde{\gamma}_{N-1}(x), 1)$, where

$$\tilde{\gamma}_{N-1}(x) = \arg \max_{\gamma \in \Gamma_{N-1,2}(x)} \mathcal{V}_{N-1}\left(x, (\gamma, 1)\right)$$

with $\mathcal{V}_n(\cdot, \cdot)$ given in Eq. (4.19). Note that the policy $(0, 0)$ is always superior to the policy $(0, 1)$ due to the monotonicity of $\xi \longmapsto G(\xi)$ and $k \longmapsto C_n(k)$. $\tilde{\gamma}_{N-1}(x)$ can be obtained via a similar argument used in Case 1.

(2) In the second step, we show that for each I, $\big(W, Z, A, B\big) \longmapsto V_{N-1}(x)$ is convex on $\mathcal{R}$.

Lemma C.2 implies that it suffices to show that for each I and A, $\big(W, Z, B\big) \longmapsto V_{N-1}(x)$ is a convex function. Equivalently, we would like to prove

$$V_{N-1}\left(\tilde{x}\right) \leq \lambda V_{N-1}(x) + (1-\lambda) V_{N-1}\left(x'\right). \tag{C.7}$$

In the sequel we only prove the above inequality for $I > 0$. The case for $I = 0$ can be proved in parallel. Below we divide the discussion into two cases.

(i) If the optimal withdrawal amount at $X_{N-1} = \tilde{x}$, denoted as $\hat{\gamma}_{N-1}\left(\tilde{x}\right)$, is bigger than $G(I)A$, then one must have $\hat{\gamma}_{N-1}\left(\tilde{x}\right) = \tilde{W}$ according to (C.6). Accordingly,

$$\begin{aligned} V_{N-1}\left(\tilde{x}\right) &= f_{N-1}\left(\tilde{x}, \left(\tilde{W}, 1\right)\right) \\ &\leq \lambda\Big[{}_{N-2}p_0 q_{N-2}\big(B \vee W\big) + {}_{N-1}p_0 g_{N-1}\left(x, \left(W, 1\right)\right)\Big] \\ &\quad +(1-\lambda)\Big[{}_{N-2}p_0 q_{N-2}\big(B' \vee W'\big) + {}_{N-1}p_0 g_{N-1}\left(x', \left(W', 1\right)\right)\Big] \\ &= \lambda f_{N-1}\left(x, \left(W, 1\right)\right) + (1-\lambda) f_{N-1}\left(x', \left(W', 1\right)\right) \\ &\leq \lambda V_{N-1}(x) + (1-\lambda) V_{N-1}\left(x'\right). \end{aligned}$$

(ii) If $\hat{\gamma}_{N-1}\left(\tilde{x}\right) \leq G(I)A$, then it equals either 0 or $G(I)A$ according to Eq. (C.6). Lemma C.3 implies

$$\begin{aligned} K_{N-1,1}\left(\tilde{x}, \left(\hat{\gamma}_{N-1}\left(\tilde{x}\right), 1\right)\right) \leq\ & \lambda K_{N-1,1}\left(x, \left(\hat{\gamma}_{N-1}\left(\tilde{x}\right), 1\right)\right) \\ & +(1-\lambda) K_{N-1,1}\left(x', \left(\hat{\gamma}_{N-1}\left(\tilde{x}\right), 1\right)\right). \end{aligned}$$

It is also notable that $\left(\hat{\gamma}_{N-1}\left(\tilde{x}\right), 1\right) \in D_{N-1,2}(x)$ and $\left(\hat{\gamma}_{N-1}\left(\tilde{x}\right), 1\right) \in D_{N-1,2}\left(x'\right)$. Then the C.M. property of $k_1 \longmapsto C_{N-1}(k_1)$, Lemma C.1, and the above inequality

together imply inequality (C.7).

Assertion for $n = N-2, N-3, \ldots, 1$ We only sketch the proof of the assertion for $n = N-2$. Firstly, by using Lemmas C.3 and C.5 one may show that the optimal withdrawal amount at x can be determined from the few desired choices. Secondly, to show the convexity of $\big(W, Z, A, B\big) \longmapsto V_{N-2}(x)$, the discussion is divided into two cases: the optimal withdrawal amount is greater than MAWA or not. Thirdly, the the monotonicity of $\big(W, Z, A, B\big) \longmapsto V_{N-2}(x)$ is guaranteed by Lemma C.4. The statements for $n = N-3, \ldots, 1$ can be proved by adopting the above arguments inductively. □

C.4.4 Proof of Theorem 4.2

Proof of Theorem 4.2. On one hand, the no-arbitrage price of the real contract is given by

$$\bar{V}_0(x_0) = \sup_{\pi \in \bar{\mathbf{\Pi}}} \mathbb{E}^{\mathbb{Q}} \left[\sum_{k=1}^{N-1} \varphi^k f_k \left(\bar{X}_k, \pi_k \right) + \varphi^N{}_{N-1} p_0 \bar{W}(t_N) \right], \tag{C.8}$$

where $\bar{\mathbf{\Pi}}$ is defined in a similar way as Eq. (4.16) expect that X_k is replaced by $\bar{X}_k$.

On the other hand, the no-arbitrage price of the synthetic contract is similarly given by

$$V_0(x_0) = \sup_{\pi \in \mathbf{\Pi}} \mathbb{E}^{\mathbb{Q}} \left[\sum_{k=1}^{N-1} \varphi^k f_k \left(X_k, \pi_k \right) + \varphi^N{}_{N-1} p_0 W(t_N) \right]. \tag{C.9}$$

By comparing the last two equations, one may observe that the discrepancy between $V_0(x_0)$ and $\bar{V}_0(x_0)$ stems from two sources: the admissible sets and the state processes. In view of this, to prove $\bar{V}_0(x_0) \leq V_0(x_0)$, it suffices to show:

(i) $\bar{\mathbf{\Pi}} \subseteq \mathbf{\Pi}$ and

(ii) for each $\pi \in \bar{\mathbf{\Pi}}$, the term in the bracket of (C.8) is dominated by that in (C.9) almost surely.

Recall from Eq. (4.3) that $D_n(x) \subseteq D_n(y)$ if $x \leq y$. Therefore, the preceding statements (i) and (ii) hold if one could show, for each $\pi \in \bar{\mathbf{\Pi}}$, $\bar{X}_k \leq X_k$ for all $k \in \mathcal{I}$, almost surely. From the transition functions accompanying $B(t)$ and I_t (see Eqs. (C.3) and (C.4), respectively), it is easy to see that for each $\pi \in \bar{\mathbf{\Pi}}$ $\bar{B}(t_k) = B(t_k)$ and $\bar{I}_k = I_k$ for all $k \in \mathcal{I}$ almost surely. It remains to show

$$\bar{W}(t_k) \leq W(t_k), \ \bar{Z}(t_k) \leq Z(t_k), \ \bar{A}(t_k) \leq A(t_k), \tag{C.10}$$

for all $k \in \mathcal{I}$ almost surely (a.s.).

Adopt a forward induction strategy to show (C.10). For $k = 1$, from the end of Section 4.2, one can tell that all the inequalities in (C.10) hold with equalities. Inductively assume that (C.10) holds for n. Recall that $\bar{A}(t)$ and $A(t)$ share the same transition function across t_n:

$$\bar{A}(t_{n+1}) \quad = \quad \begin{cases} \bar{A}(t_n) \vee \bar{Z}(t_n), & \text{if } \tau_n = 0, \\ \bar{A}(t_n) \vee \left(\bar{W}(t_n) - \gamma_n\right), & \text{if } I_n = 0, 0 < \gamma_n \le \tilde{G}_n(I_n)\bar{A}(t_n), \\ \bar{A}(t_n) \vee \bar{Z}(t_n), & \text{if } I_n > 0, 0 \le \gamma_n \le \tilde{G}_n(I_n)\bar{A}(t_n), \\ \frac{\bar{W}(t_n)-\gamma_n}{\bar{W}(t_n)-\tilde{G}_n(I_n)\bar{A}(t_n)}\bar{A}(t_n) \vee \left(\bar{W}(t_n) - \gamma_n\right), & \text{otherwise,} \end{cases}$$

and note $\bar{Z}(t_n) \ge \bar{W}(t_n)$. Therefore, one may conclude $\bar{A}(t_{n+1}) \ge \left(\bar{W}(t_n) - \gamma_n\right)^+$.

From Eq. (4.6), it is straightforward to see

$$\begin{aligned} \bar{W}(t_n^+) &= \max\left[\left(\bar{W}(t_n) - \gamma_n\right)^+ - \eta\bar{A}(t_{n+1}), 0\right] \\ &\le \left(\bar{W}(t_n) - \gamma_n\right)^+(1-\eta) \\ &\le \left(W(t_n) - \gamma_n\right)^+(1-\eta) = W(t_n^+) \quad \text{a.s.}, \end{aligned}$$

where the first inequality follows because $\bar{A}(t_{n+1}) \ge \left(\bar{W}(t_n) - \gamma_n\right)^+$, and the second inequality is by the induction hypothesis. The above inequality implies

$$\bar{W}(t_{n+1}) = \bar{W}(t_n^+)e^{L_n(t_{n+1})} \le W(t_n^+)e^{L_n(t_{n+1})} = W(t_{n+1}) \quad \text{a.s.}$$

with $L_n(t_{n+1})$ being the log-return of the policy fund over $[t_n, t_{n+1}]$.

One can obtain $\bar{Z}(t_{n+1}) \le Z(t_{n+1})$ in a similar way and thus (C.10) is true. This completes the proof. □

C.4.5 Proof of Proposition 4.3

Proof of Proposition 4.3. Only property (i) will be proved because property (ii) is clear from the proof of Theorem 4.1. Let L_n and $\bar{L}_n$ be the increment and the running maximum of the log-return process of the policy fund over $[t_n, t_{n+1}]$, respectively. Further denote $\hat{k} := \left(k_1, k_2, 1, k_4, k_5\right)$ and

$$x_{n+1} := \left(\widehat{W}_{n+1}, \widehat{Z}_{n+1}, 1, \widehat{B}_{n+1}, I\right) = \left(k_1 e^{L_n}, k_2 \vee k_1 e^{\bar{L}_n}, 1, k_4, k_5\right),$$

where the dependency of x_{n+1} on $\hat{k}$ is suppressed. In the remainder of the proof, the readers should always bear in mind that x_{n+1} depends on $\hat{k}$ and thus implicitly depends on k_2.

In view of (4.18) and $V_N(X_N) = {}_{N-1}p_0 W(t_N)$, property (i) obviously holds for $n = N-1$. Assume as the induction hypothesis that $k_2 \longmapsto C_{n+1}\left(\hat{k}\right)$ is constant for all $k_2 \in [0,1]$. Recall from Eq. (4.18) that

$$C_n\left(\hat{k}\right) = \mathbb{E}^{\mathbb{Q}}\left[V_{n+1}\left(k_1 e^{L_n}, k_2 \vee k_1 e^{\bar{L}_n}, 1, k_4, k_5\right)\right] = \mathbb{E}^{\mathbb{Q}}\left[V_{n+1}\left(x_{n+1}\right)\right].$$

Therefore, in order to prove $C_n\left(\hat{k}\right)$ is invariant with respect to k_2 for all $k_2 \in [0,1]$, it suffices to show that $k_2 \mapsto V_{n+1}\left(x_{n+1}\right)$ is constant over $[0,1]$.

In view of Eqs. (4.19) and (4.23), one gets

$$V_{n+1}\left(x_{n+1}\right) = \sup_{\pi_{n+1} \in \widehat{D}_{n+1}} \mathcal{V}_{n+1}\left(x_{n+1}, \pi_{n+1}\right),$$

where

$$\mathcal{V}_{n+1}\left(x_{n+1}, \pi_{n+1}\right) = f_{n+1}\left(x_{n+1}, \pi_{n+1}\right) + \varphi K_{n+1,3}\left(x_{n+1}, \pi_{n+1}\right) C_{n+1}\left(\hat{K}_{n+1}\left(x_{n+1}, \pi_{n+1}\right)\right).$$

It is easy to see $\widehat{D}_{n+1}$ and $f_{n+1}\left(x_{n+1}, \pi_{n+1}\right)$ are independent of k_2; see Eqs. (4.20) and (4.15), respectively. Further observe from Eqs. (C.2) and (C.1) that

$$K_{n+1,3}(x_{n+1}, \pi_{n+1}) = \begin{cases} 1 \vee \left(k_2 \vee k_1 e^{\bar{L}_n}\right), & \text{if } \tau_{n+1} = 0, \\ 1 \vee \left(k_1 e^{\bar{L}_n} - \gamma_{n+1}\right), & \text{if } k_5 = 0, 0 < \gamma_{n+1} \leq \tilde{G}_{n+1}(k_5), \\ 1 \vee \left(k_2 \vee k_1 e^{\bar{L}_n}\right), & \text{if } k_5 > 0, 0 \leq \gamma_{n+1} \leq \tilde{G}_{n+1}(k_5), \\ \frac{k_1 e^{L_n} - \gamma_{n+1}}{k_1 e^{L_n} - \tilde{G}_{n+1}(k_5)} \vee \left(k_1 e^{L_n} - \gamma_{n+1}\right)^+, & \text{otherwise}, \end{cases}$$

which is invariant with respect to k_2 for all $k_2 \in [0,1]$, and that

$$\hat{K}_{n+1,2}(x_{n+1}, \pi_{n+1}) = \begin{cases} \frac{\widehat{Z}_{n+1}}{1 \vee \widehat{Z}_{n+1}}, & \text{if } \tau_{n+1} = 0, \\ \frac{\left(\widehat{W}_{n+1} - \gamma_{n+1}\right)^+ (1-\eta)}{1 \vee \left(\widehat{W}_{n+1} - \gamma_{n+1}\right)^+}, & \text{if } \tau_{n+1} = 1, k_5 = 0, \\ \frac{\left(\widehat{W}_{n+1} - \gamma_{n+1}\right)^+ (1-\eta)}{1 \vee \widehat{Z}_{n+1}}, & \text{if } k_5 > 0, 0 \leq \gamma_{n+1} \leq \tilde{G}_{n+1}(k_5), \\ \left(\frac{1}{\widehat{W}_{n+1} - \tilde{G}_{n+1}(k_5)} \vee 1\right)^{-1}, & \text{otherwise}, \end{cases}$$

which takes value in $[0,1]$.

These, combined with the induction hypothesis, imply that $k_2 \mapsto V_{n+1}\left(x_{n+1}\right)$ is constant over $[0,1]$. The proof is complete. □