7-31-2016

# Essays on Computational Problems in Insurance

Hongjun Ha

# PERMISSION TO BORROW

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or publish this dissertation may be granted by the author or, in his/her absence, the professor under whose direction it was written or, in his absence, by the Dean of the Robinson College of Business. Such quoting, copying, or publishing must be solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential gain will not be allowed without written permission of the author.

HONGJUN HA

# NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:

Hongjun Ha
Department of Risk Management and Insurance
Georgia State University
35 Broad Street NW, 11th Floor, Atlanta, GA 30303

The director of this dissertation is:

Daniel Bauer
Department of Risk Management and Insurance
Georgia State University
35 Broad Street NW, 11th Floor, Atlanta, GA 30303

Essays on Computational Problems in Insurance BY

Hongjun Ha

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY

ROBINSON COLLEGE OF BUSINESS

2016

# ACCEPTANCE

This dissertation was prepared under the direction of the Hongjun Ha Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctoral of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Daniel Bauer

Ajay Subramanian

Enrico Biffis

Baozhong Yang

# ABSTRACT

Essays on Computational Problems in Insurance BY

Hongjun Ha

July 18th, 2016

Committee Chair: Daniel Bauer

Major Academic Unit: Department of Risk Management and Insurance

This dissertation consists of two chapters. The first chapter establishes an algorithm for calculating capital requirements. The calculation of capital requirements for financial institutions usually entails a reevaluation of the company's assets and liabilities at some future point in time for a (large) number of stochastic forecasts of economic and firm-specific variables. The complexity of this *nested* valuation problem leads many companies to struggle with the implementation. The current chapter proposes and analyzes a novel approach to this computational problem based on least-squares regression and Monte Carlo simulations. Our approach is motivated by a well-known method for pricing non-European derivatives. We study convergence of the algorithm and analyze the resulting estimate for practically important risk measures. Moreover, we address the problem of how to choose the regressors, and show that an *optimal* choice is given by the left singular functions of the corresponding valuation operator. Our numerical examples demonstrate that the algorithm can produce accurate results at relatively low computational costs, particularly when relying on the optimal basis functions.

The second chapter discusses another application of regression-based methods, in the context of pricing variable annuities. Advanced life insurance products with exercise-dependent financial guarantees present challenging problems in view of pricing and risk management. In particular, due to the complexity of the guarantees and since practical valuation frameworks include a variety of stochastic risk factors, conventional methods that are based on the discretization of the underlying (Markov) state space may not be feasible. As a practical alternative, this chapter explores the applicability of Least-Squares Monte Carlo (LSM) methods familiar from American option pricing in this context. Unlike previous literature we consider optionality beyond surrendering the contract, where we focus on popular withdrawal benefits – so-called GMWBs – within Variable Annuities. We introduce different LSM variants, particularly the regression-now and regression-later approaches, and explore their viability and potential pitfalls. We commence our numerical analysis in a basic Black-Scholes framework, where we compare the LSM results to those from a discretization approach. We then extend the model to include various relevant risk factors and compare the results to those from the basic framework.

# ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents for always supporting me. When I told them about my plan to study abroad, they helped me financially as well as emotionally. In particular, my father advised me frequently and encouraged me in difficult time. He is definitely my teacher in life. My mother shows her love by trusting me. When I changed my major to Risk Management and Insurance, she supported my decision without hesitation. My wife, Juhee Han, deserves a special thank you. Even though she had a good career in the Republic of Korea, she quit her job in order to support my studies at GSU. It is my turn now to support her in finding a "new" occupation in a "new" place. Thanks also to my sister, Jiyoen. I am sorry she has to deal with our family business in my stead. I am indebted to her.

I was lucky enough to meet fantastic people at GSU. Dr. Daniel Bauer, my advisor, was the biggest reason why I pursued my doctoral degree. I still remember his class where he widened my intellectual horizon about Actuarial Science and Financial Engineering. There, I dreamed of being a researcher just like him. Moreover, he was very patient with my slow progress and clearly pointed out what I should have done. I cannot list here what I have learned from him. Dr. Ajay Subramanian, who gave me the chance to study at the doctoral level at GSU, also taught me many advanced subjects. He encouraged me to do research of high quality by giving valuable feedback. I appreciate Dr. Enrico Biffis' supporting me as a committee member. He gave good advice on my research before and after his moving to GSU. Dr. Baozhong Yang spent a lot of time reading my papers and giving comments. His sharp insight into my researches was a great help in deriving meaningful results. I am also very grateful for financial support through the Hellen C. Leith, the Leyton Hunter and the Huebner Foundation, respectively.

Finally, thanks to everyone who has provided support. Especially, I want to thank Dr. Soon-ho Kim, the officiator of my wedding and my mentor. He always helped

me in succeeding in many important things in my life. I cannot forget his truthful advice and support for the last seven years. I was very happy to study, research, and live with my friends, Phillipe, Yiling, and Dongshin, who were all in the same boat as me. I am certain you will be good teachers and researchers. I should acknowledge my friends in Duluth: Yongseok, Seokbum, Minjae, Sunghak and Soonwoo.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# A Least-Squares Monte Carlo Approach to the Calculation of Capital Requirements[1]

## 1.1 Introduction

Many risk management applications within financial institutions entail a reevaluation of the company's assets and liabilities at some time horizon $\tau$ (usually called a *risk horizon*) for a large number of realizations of economic and firm-specific (state) variables. The resulting empirical loss distribution is then applied to derive risk measures such as the Value-at-Risk (VaR) or the Expected Shortfall (ES), which serve as the basis for capital requirements within several regulatory frameworks such as Basel III for banks and Solvency II for insurance companies. However, the high complexity of this *nested* computation structure leads firms to struggle with the implementation

[1]This chapter extends an earlier working paper Bauer et al. (2009), where the approach considered here was originally proposed. We thank Giuseppe Benedetti, Enrico Biffis, Matthias Fahrenwaldt, Andreas Reuss, Daniela Singer, Ajay Subramanian, Baozhong Yang, and seminar participants at the Bachelier Congress 2014, the World Risk and Insurance Economics Congress 2015, Georgia State University, Michigan State University, St. Joseph's University, Université de Montréal, and Barrie & Hibbert for helpful comments.

(Bauer et al., 2012).[2]

The present chapter proposes an alternative approach based on least-squares regression and Monte Carlo simulation akin to the well-known Least-Squares Monte Carlo method (LSM) for pricing non-European derivatives introduced by Longstaff and Schwartz (2001). Akin to the LSM pricing method, this approach relies on two approximations (Clément et al., 2002): On the one hand, the capital random variable, which can be represented as a conditional risk-neutral expected value at the time horizon $\tau$, is replaced by a finite linear combination of functions of the state variables, so-called *basis functions*. As the second approximation, Monte Carlo simulations and least-squares regression are employed to estimate this linear combination. Hence, for each realization of the state variables, the resulting linear combination presents an approximate realization of the capital at $\tau$, and the resulting sample can be used for estimating relevant risk measures.

Although this approach is increasingly popular in practice for calculating economic capital particularly in the insurance industry (Barrie and Hibbert, 2011; Milliman, 2013; DAV, 2015) and has been used in several applied research contributions (Floryszczak et al., 2011; Pelsser and Schweizer, 2015), these papers do not provide a detailed analysis of the properties of this algorithm and the choice of the basis functions. Our work closes this gap in literature.

We begin our analysis by introducing our setting and the algorithm. As an important innovation, we frame the estimation problem via a *valuation operator* that maps future payoffs (as functionals of the state variables) to the conditional expected value at the risk horizon. In particular, we base our definition on a hybrid probability measure that overcomes structural difficulties with the probability space – arising from the fact that simulations for risk estimation before the risk horizon are carried out under the physical measure whereas simulations for valuation after the risk horizon

[2] As a consequence, many companies rely on approximations within so-called *standard models* or *standardized approaches*, which are usually not able to accurately reflect an company's risk situation and may lead to deficient outcomes (Liebwein, 2006; Pfeifer and Strassburger, 2008)

are carried out under a risk-neutral measure.

We formally establish convergence of the algorithm for the risk distribution (in probability) and for families of risk measures under general conditions when taking limits sequentially in the first and second approximation. In addition, by relying on results from Newey (1997) on the convergence of series estimators, we present conditions for the joint convergence of the two approximations in the general case and more explicit results for the practically relevant case of orthonormal polynomials.[3] We then analyze in more detail the properties of the estimator for the important special case of VaR, which serves as the risk measure for regulatory frameworks such as Basel III or Solvency II. In particular, the conditions for joint convergence imply that the number of simulations has to increase faster than the cube of the number of basis functions when estimating VaR via the LSM algorithm based on polynomial basis functions. Moreover, by building on ideas from Gordy and Juneja (2010), we show that for a fixed number of basis functions, the least-squares estimation of the regression approximation, while unbiased when viewed as an estimator for the individual loss, carries a positive bias term for this tail risk measure. It is important to note, however, that this result only pertains to the regression approximation but not the approximation of the actual loss variables via the linear combination of the basis functions – which is the crux of the algorithm. In particular, the adequacy of the estimate crucially depends on the choice of basis functions.

This is where the operator formulation becomes especially useful. By expressing the valuation operator via its singular value decomposition (SVD), we show that under certain conditions, the (left) singular functions present an optimal choice for the basis functions. More precisely, we demonstrate that these singular functions approximate the valuation operator – and, thus, the distributions of relevant capital levels – in an optimal manner. The intuition is that similarly to an SVD for a matrix, the singular functions provide the most important dimensions in spanning the image space of the

[3]We thank Giuseppe Benedetti for pointing us to this issue of joint convergence.

operator.

We comment on the joint convergence of the LSM algorithm under this choice and also the calculation of the singular functions. While in general the decomposition has to be carried out numerically, for certain classes of models it is possible to derive analytic expressions. As an important example class for applications, we discuss the calculation of the SVD – and, thus, the derivation of optimal basis functions – for models with Gaussian transition density. In this case, (i) it is straightforward to show that the underlying assumptions are satisfied. And (ii), by following ideas from Khare and Zhou (2009), it is possible to derive the singular functions, which take the form of products of Hermite polynomials of linearly transformed states, by solving a related eigenvalue problem.

We illustrate our theoretical results considering two examples from life insurance in the context of annuitization options. We first we consider a simple Guaranteed Annuity Option (GAO) within a pure endowment insurance contract in the Vasicek (1977) stochastic interest rate model (Boyle and Hardy, 2003; Pelsser, 2003). Following Boyle and Hardy (2003), we obtain a closed form solution for the valuation problem at the risk horizon so that we can conveniently compare the approximated realizations of the loss distribution with the exact ones. Our results demonstrate that the algorithm can produce accurate results at relatively low computational costs, although the interplay of the sample variance and the functional approximation is finical. We find that optimal basis functions improve the performance of the algorithm when compared to alternative basis functions with a different span.

As a second example, we consider popular annuitization guarantees within Variable Annuity contracts, so-called Guaranteed Minimum Income Benefits (GMIBs).[4] In a setting with three stochastic risk factors (investment fund, interest, and mortality), we demonstrate that the algorithm still delivers reliable results when relying on

[4]Between 2011 and 2013, roughly 15% of the more than $150 billion worth of Variable Annuities sold in the US contained a GMIB. Source: Fact Sheets by the Life Insurance and Market Research Association (LIMRA).

sufficiently many basis functions and simulations. Here we emphasize that the optimal choice given by the singular functions not only determines the functional class – which are Hermite polynomials in this case, although of course different classes of univariate polynomials will generate the same span. But they also specify the most important combinations of stochastic factors, an indeed in our setting it turns out that higher-order combinations of certain risk factors are more important than lower-order combinations of others.[5] This latter aspect in particular is very relevant in practical settings with high-dimensional state vectors, so that our results provide immediate guidance for these pressing problems.

**Related Literature and Organization of the Chapter**

Our approach is inspired by the LSM approach for derivative pricing and relies on corresponding results (Carriere, 1996; Tsitsiklis and Van Roy, 2001; Longstaff and Schwartz, 2001; Clément et al., 2002). A similar regression-based algorithm for risk estimation is independently studied in Broadie et al. (2015) (their paper postdates early versions of this work; Bauer et al. (2009)). Their results are similar to our sequential convergence results in 1.3.1, and the authors additionally introduce a *weighted* version of their regression algorithm. Moreover, Benedetti (2016) provides alternative joint convergence results to ours in Section 1.3.1 under a different (weaker) set of conditions. However, these authors do not contemplate how to optimally choose the basis functions – although they emphasize the importance of this choice – which is a key contribution of the current chapter.

As already indicated, the LSM approach enjoys popularity in the context of calculating risk capital for life insurance liabilities in practice and applied research, so that providing a theoretical foundation and guidance for its application are key motivating factors for this chapter. A number of recent contributions discuss the so-called replicating portfolio approach as an alternative that enjoys certain advantages (Beutner et

[5]We thank Baozhong Yang for pointing us in this direction.

al., 2016; Natolski and Werner, 2016; Cambou and Filipović, 2016), and Pelsser and Schweizer (2015) point our that the difference between the LSM versus the replicating portfolio calculation aligns with the so-called *regression-now* versus the so-called *regression-later* algorithm, respectively, for non-European option pricing (Glasserman and Yu, 2002). While a detailed comparison is beyond the scope of this chapter, we note that although indeed in simple settings the performance of regress-later approaches appears superior (Beutner et al., 2013), the application comes with several caveats regarding the choice of the basis function and other complications in high-dimensional settings (Pelsser and Schweizer, 2015; Ha and Bauer, 2016).

The remainder of the chapter is structured as follows: Section 1.2 lays out the simulation framework and the algorithm; Section 1.3 addresses convergence of the algorithm and analyzes the estimator in special cases; Section 1.4 discusses optimal basis functions and derives them in models with Gaussian transition densities; Section 1.5 provides the numerical examples; and, finally, Section 1.6 concludes the chapter. All proofs are relegated to the Appendix.

## 1.2 The LSM Approach

### 1.2.1 Simulation Framework

We assume that investors can trade continuously in a frictionless financial market with time finite horizon $T$ corresponding to the longest-term liability of the company in view. Let $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ be a complete filtered probability space on which all relevant quantities exist, where $\mathbb{P}$ denotes the physical measure. We assume that all random variables in what follows are square-integrable (in $L^2(\Omega, \mathcal{F}, \mathbb{P})$). The sigma algebra $\mathcal{F}_t$ represents all information about the market up to time $t$, and the filtration $\mathbf{F}$ is assumed to satisfy the usual conditions.

The uncertainty with respect to the company's future assets and liabilities arises from the uncertain development of a number of influencing factors, such as equity re-

turns, interest rates, demographic or loss indices, etc. We introduce the $d$-dimensional, sufficiently regular Markov process $Y = (Y_t)_{t\in[0,T]} = (Y_{t,1}, \ldots, Y_{t,d})_{t\in[0,T]}$, $d \in \mathbb{N}$, the so-called *state process*, to model this uncertainty. We assume that all financial assets in the market can be expressed in terms of $Y$. Non-financial risk factors can also be incorporated (see e.g. Bauer et al. (2010) or Zhu and Bauer (2011) for settings specific to life insurance that include demographic risk). In this market, we take for granted the existence of a risk-neutral probability measure (equivalent martingale measure) $\mathbb{Q}$ equivalent to $\mathbb{P}$ under which payment streams can be valued as expected discounted cash flows with respect to a given numéraire process $(N_t)_{t\in[0,T]}$.[6]

In financial risk management, we are now concerned with the company's financial situation at a certain (future) point in time $\tau$, $0 < \tau < T$, which we refer to as the *risk horizon*. More specifically, based on realizations of the state process $Y$ over the time period $[0, \tau]$ that are generated under the physical measure $\mathbb{P}$, we need to assess the *available capital* $C_\tau$, at time $\tau$ calculated as the market value of assets minus liabilities. This amount can serve as a buffer against risks and absorb financial losses. The *capital requirement* is then defined via a risk-measure $\rho$ applied to the capital random variable. For instance, if the capital requirement is cast based on Value-at-Risk (VaR), the capitalization at time $\tau$ should be sufficient to cover the net liabilities at least with a probability $\alpha$, i.e. the additionally required capital is

$$\mathrm{VaR}_\alpha(-C_\tau) = \inf\left\{ x \in \mathbb{R} \middle| \, \mathbb{P}\left(x + C_\tau \geq 0\right) \geq \alpha \right\}. \tag{1.1}$$

The capital at the risk horizon, for each realization of the state process $Y$, is derived from a market-consistent valuation approach. While the market value of traded instruments is usually readily available from the model ("mark-to-market"), the valuation of complex financial positions on the firm's asset side such as portfolios of derivatives and/or the valuation of complex liabilities such as insurance contracts containing em-

[6] According to the *Fundamental Theorem of Asset Pricing*, this assumption is essentially equivalent to the absence of arbitrage. We refer to Schachermayer (2009) for details.

bedded options typically requires numerical approaches. This is the main source of complexity associated with this task, since the valuation needs to be carried out for each realization of the process $Y$ at time $\tau$, i.e. we face a *nested* calculation problem.

Formally, the available capital is derived as a (risk-neutral) conditional expected value of discounted cash flows $X_t$, where for simplicity and to be closer to modeling practice, we assume that cash flows only occur at the discrete times $t = 1, 2, \ldots, T$ and that $\tau \in \{1, 2, \ldots, T\}$ :

$$C_\tau = \mathbb{E}^{\mathbb{Q}} \left[ \sum_{k=\tau}^{T} \frac{N_\tau}{N_k} X_k \middle| (Y_s)_{0 \le s \le \tau} \right]. \tag{1.2}$$

Note that within this formulation, interim asset and liability cash flows in $[0, \tau]$ may be aggregated in the $\sigma(Y_s, 0 \le s \le \tau)$-measurable position $X_\tau$. Moreover, in contrast to e.g. Gordy and Juneja (2010), we consider aggregate asset and liability cash flows at times $k \ge \tau$ rather than cash flows corresponding to individual asset and liability positions. Aside from notational simplicity, the reason for this formulation is that we particularly focus on situations where an independent evaluation of many different positions is not advisable or feasible as it is for instance the case within economic capital modeling in life insurance (Bauer et al., 2012).

In addition to current interest rates, security prices, etc., the value of the asset and liability positions may also depend on path-dependent quantities. For instance, Asian options depend on the average of a certain price index over a fixed time interval, lookback options depend on the running maximum, and liability values in insurance with profit sharing mechanisms depend on entries in the insurer's bookkeeping system. In what follows, we assume that – if necessary – the state process $Y$ is augmented so that it contains all quantities relevant for the evaluation of the available capital and

still satisfies the Markov property (Whitt, 1986). Thus, we can write:

$$C_\tau = \mathbb{E}^{\mathbb{Q}}\left[\sum_{k=\tau}^{T} \frac{N_\tau}{N_k} X_k \middle| Y_\tau\right].$$

We refer to the state process $Y$ as our *model framework*. Within this *framework*, the asset-liability projection *model* of the company is given by cash flow projections of the asset-liability positions, i.e. functionals $x_k$ that derive the cash flows $X_k$ based on the current state $Y_k$:[7]

$$\frac{N_\tau}{N_k} X_k = x_k\left(Y_k\right),\ \tau \leq k \leq T.$$

Hence, each *model* within our *model framework* can be identified with an element in a suitable function space, $\mathbf{x} = (x_\tau, x_{\tau+1}, ..., x_T)$ . More specifically, we can represent:

$$C_\tau\left(Y_\tau\right) = \sum_{j=\tau}^{T} \mathbb{E}^{\mathbb{Q}}\left[x_j\left(Y_j\right) \middle| Y_\tau\right].$$

We now introduce the probability measure $\tilde{\mathbb{P}}$ via its Radon-Nikodym derivative:

$$\frac{\partial \tilde{\mathbb{P}}}{\partial \mathbb{P}} = \frac{\frac{\partial \mathbb{Q}}{\partial \mathbb{P}}}{\mathbb{E}^{\mathbb{P}}\left[\frac{\partial \mathbb{Q}}{\partial \mathbb{P}} \middle| \mathcal{F}_\tau\right]}.$$

**Lemma 1.2.1.** *We have:*

1. $\tilde{\mathbb{P}}(A) = \mathbb{P}(A),\ A \in \mathcal{F}_t,\ 0 \leq t \leq \tau.$
2. $\mathbb{E}^{\tilde{\mathbb{P}}}\left[X \middle| \mathcal{F}_\tau\right] = \mathbb{E}^{\mathbb{Q}}\left[X \middle| \mathcal{F}_\tau\right]$ *for every random variable* $X \in \mathcal{F}$.

Lemma 1.2.1 implies that we have

$$C_\tau\left(Y_\tau\right) \quad = \quad \sum_{j=\tau}^{T} \mathbb{E}^{\tilde{\mathbb{P}}}\left[x_j\left(Y_j\right) \middle| Y_\tau\right] = L\,\mathbf{x}\left(Y_\tau\right), \tag{1.3}$$

[7]Similarly to Section 8.1 in Glasserman (2004), without loss of generality, by possibly augmenting the state space or by changing the numéraire process (see Section 1.5), we assume that the discount factor can be expressed as a function of the state variables.

where the operator

$$L : \mathcal{H} = \bigoplus_{j=\tau}^{T} L^2\left(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_j}\right) \to L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right) \tag{1.4}$$

is mapping a model to capital. We call $L$ in (1.4) the *valuation operator*. For our applications later in the text, it is important to note the following:

**Lemma 1.2.2.** *$L$ is continuous linear operator.*

Moreover, for our results on the optimality of basis functions, we require compactness of the operator $L$. The following lemma provides a sufficient condition for $L$ to be compact in terms of the transition densities of the driving Markov process.

**Lemma 1.2.3.** *Assume there exists a joint density $\pi_{Y_\tau, Y_j}(y, x)$, $j = \tau, \tau+1, ..., T$, for $Y_\tau$ and $Y_j$. Moreover:*

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \pi_{Y_j|Y_\tau}(x|y)\, \pi_{Y_\tau|Y_j}(y|x)\, dx\, dy < \infty,$$

*where $\pi_{Y_j|Y_\tau}(x|y)$ and $\pi_{Y_\tau|Y_j}(y|x)$ denote the transition density and the reverse transition density, respectively. Then the operator $L$ is compact.*

The definition of $L$ implies that a model can be identified with an element of the Hilbert space $\mathcal{H}$ whereas the capital $C_\tau$ can be (state-wise) identified with an element of $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$. The task at hand is now to evaluate this element for a given model $\mathbf{x} = (x_\tau, \ldots, x_T)$ and to then determine the capital requirement via a (monetary) risk measure $\rho : L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}) \to \mathbb{R}$ as $\rho(L\mathbf{x})$, although the *model* may change between applications as the exposures may change (e.g. from one year to the next or when evaluating allocations).

One possibility to carry out this computational problem is to rely on *nested simulations*, i.e. to simulate a large number of scenarios for $Y_\tau$ under $\mathbb{P}$ and then, for each of these realizations, to determine the available capital using another simulation step

under $\mathbb{Q}$. The resulting (empirical) distribution can then be employed to calculate risk measures (Lee, 1998; Gordy and Juneja, 2010). However, this approach is computationally burdensome and, for some relevant applications, may requires a very large number of simulations to obtain results in a reliable range (Bauer et al., 2012). Hence, in the following, we propose and develop an alternative approach for such situations.

### 1.2.2 Least-Squares Monte-Carlo (LSM) Algorithm

As indicated in the previous section, the task at hand is to determine the distribution of $C_\tau$ given by Equation (1.3). Here, the conditional expectation causes the primary difficulty for developing a suitable Monte Carlo technique. This is akin to the pricing of Bermudan or American options, where "*the conditional expectations involved in the iterations of dynamic programming cause the main difficulty for the development of Monte-Carlo techniques*" (Clément et al., 2002). A solution to this problem was proposed by Carriere (1996), Tsitsiklis and Van Roy (2001), and Longstaff and Schwartz (2001), who use least-squares regression on a suitable finite set of functions in order to approximate the conditional expectation. In what follows, we exploit this analogy by transferring their ideas to our problem.

As pointed out by Clément et al. (2002), their approach consists of two different types of approximations. Proceeding analogously, as the first approximation, we replace the conditional expectation, $C_\tau$, by a finite combination of linear independent basis functions $e_k(Y_\tau) \in L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right)$:

$$C_\tau \approx \widehat{C}_\tau^{(M)}(Y_\tau) = \sum_{k=1}^{M} \alpha_k \cdot e_k(Y_\tau). \tag{1.5}$$

We then determine approximate $\mathbb{P}$-realizations of $C_\tau$ using Monte Carlo simulations. We generate $N$ independent paths $(Y_t^{(1)})_{0\le t\le T}$, $(Y_t^{(2)})_{0\le t\le T}$,..., $(Y_t^{(N)})_{0\le t\le T}$, where we generate the Markovian increments under the physical measure for $t \in (0, \tau]$

and under the risk-neutral measure for $t \in (\tau, T]$.[8] Based on these paths, we calculate the realized cumulative discounted cash flows

$$V_\tau^{(i)} = \sum_{j=\tau}^{T} x_j \left( Y_j^{(i)} \right), \, 1 \leq i \leq N.$$

We use these realizations in order to determine the coefficients $\alpha = (\alpha_1, \ldots, \alpha_M)$ in the approximation (1.5) by least-squares regression:

$$\hat{\alpha}^{(N)} = \text{argmin}_{\alpha \in \mathbb{R}^M} \left\{ \sum_{i=1}^{N} \left[ V_\tau^{(i)} - \sum_{k=1}^{M} \alpha_k \cdot e_k \left( Y_\tau^{(i)} \right) \right]^2 \right\}.$$

Replacing $\alpha$ by $\hat{\alpha}^{(N)}$, we obtain the second approximation:

$$C_\tau \approx \widehat{C}_\tau^{(M)}(Y_\tau) \approx \widehat{C}_\tau^{(M,N)}(Y_\tau) = \sum_{k=1}^{M} \hat{\alpha}_k^{(N)} \cdot e_k(Y_\tau), \tag{1.6}$$

based on which we may then determine $\rho\left(L\,\mathbf{x}\right) \approx \rho(\widehat{C}_\tau^{(M,N)})$.

In case the distribution of $Y_\tau$, $\mathbb{P}_{Y_\tau}$, is not directly accessible, we can calculate realizations of $\widehat{C}_\tau^{(M,N)}$ resorting to the previously generated paths $(Y_t^{(i)})_{0 \leq t \leq T}$, $i = 1, \ldots, N$, or, more precisely, to the sub-paths for $t \in [0, \tau]$. Based on these realizations, we may then determine the corresponding empirical distribution function and, consequently, an estimate for $\rho(\widehat{C}_\tau^{(M,N)})$. For the analysis of potential errors when approximating the risk measure based on the empirical distribution function, we refer to Weber (2007).

[8] Note that it is possible to allow for multiple *inner* simulations under the risk-neutral measure per *outer* simulation under $\mathbb{P}$ as in the algorithm proposed by Broadie et al. (2015). However, as shown in their paper, a single inner scenario as within our version will be the optimal choice when allocating a finite computational budget. The intuition is that the inner noise diversifies in the regression approach whereas additional outer scenarios add to the information regarding the relevant distribution.

## 1.3 Analysis of the Algorithm

### 1.3.1 Convergence

The following proposition establishes convergence of the algorithm described in Section 1.2.2 when taking limits sequentially:

**Proposition 1.3.1.** *$\widehat{C}_\tau^{(M)} \to C_\tau$ in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$, $M \to \infty$, and $\widehat{C}_\tau^{(M,N)} \to \widehat{C}_\tau^{(M)}$, $N \to \infty$, $\tilde{\mathbb{P}}$-almost surely. Furthermore, $Z^{(N)} = \sqrt{N}\left[\widehat{C}_\tau^{(M)} - \widehat{C}_\tau^{(M,N)}\right] \longrightarrow$ Normal $(0, \xi^{(M)})$, where $\xi^{(M)}$ is provided in Equation (15) in the Appendix.*

We note that the proof of this convergence result is related to and simpler than the corresponding result for the Bermudan option pricing algorithm in Clément et al. (2002) since we do not have to take the recursive nature into account. However, in contrast to their setting, we deal with a structurally more complex probability space due to the intermittent measure change and we show the adequacy of "any" linearly independent collection of basis functions rather then postulating certain properties.

The primary point of Proposition 1.3.1 is the convergence in probability – and, hence, in distribution – of $\widehat{C}_\tau^{(M,N)} \to C_\tau$ implying that the resulting distribution function of $\widehat{C}_\tau^{(M,N)}$ presents a valid approximation of the distribution of $C_\tau$ for large $M$ and $N$. The question of whether $\rho(\widehat{C}_\tau^{(M,N)})$ presents a valid approximation of $\rho(C_\tau)$ depends on the regularity of the risk measure. In general, we require continuity in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$ as well as point-wise continuity with respect to almost sure convergence (see Kaina and Rüschendorf (2009) for a corresponding discussion in the context of convex risk measures). In the special case of orthogonal basis functions, we are able to present a more concrete result:

**Corollary 1.3.1.** *If $\{e_k,\, k = 1, \ldots, M\}$ are orthonormal, then $\widehat{C}_\tau^{(M,N)} \to C_\tau$, $N \to \infty$, $M \to \infty$ in $L^1(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$. In particular, if $\rho$ is a finite convex risk measure on $L^1(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$, we have $\rho(\widehat{C}_\tau^{(M,N)}) \to \rho\left(C_\tau\right)$, $N \to \infty$, $M \to \infty$.*

Thus, at least for certain classes of risk measures $\rho$, the algorithm produces a

consistent estimate, i.e. if $N$ and $M$ are chosen *large enough*, $\rho(\hat{C}_\tau^{(M,N)})$ presents a viable approximation. In the next part, we make more precise what *large enough* means and, particularly, how large $N$ needs to be chosen relative to $M$.

### 1.3.2 Joint Convergence and Convergence Rate

The LSM algorithm approximates the capital level – which is given by the conditional expectation of the aggregated future cash flows $V_\tau = \sum_{j=1}^{T} x_j(Y_j^{(i)})$ – by its linear projection on the subspace spanned by the basis functions $e^{(M)}(Y_\tau) = (e_1(Y_\tau), \ldots, e_M(Y_\tau))'$ :

$$\mathbb{E}^{\tilde{\mathbb{P}}}\left[V_\tau | Y_\tau\right] \approx e^{(M)}(Y_\tau)'\,\hat{\alpha}^{(N)}.$$

Thus, the approximation takes the form of a *series estimator* for the conditional expectation. General conditions for the *joint* convergence of such estimators are provided in Newey (1997). Convergence of the risk measure then follows as in the previous subsection. We immediately obtain:[9]

**Proposition 1.3.2** (Newey (1997))**.** *Assume* $Var(V_\tau|Y_\tau)$ *is bounded and that for every* $M$*, there is a non-singular constant matrix* $B$ *such that for* $\tilde{e}^{(M)} = B\,e^{(M)}$ *we have:*

- *The smallest eigenvalue of* $\mathbb{E}^{\mathbb{P}}\left[\tilde{e}^{(M)}(Y_\tau)\,\tilde{e}^{(M)}(Y_\tau)'\right]$ *is bounded away from zero uniformly in* $K$*; and*
- *there is a sequence of constants* $\xi_0(M)$ *satisfying* $\sup_{y\in\mathcal{Y}} \|\tilde{e}^{(M)}(y)\| \leq \xi_0(M)$ *and* $M = M(N)$ *such that* $\xi_0(M)^2\,M/N \to 0$ *as* $N \to \infty$*, where* $\mathcal{Y}$ *is the support of* $Y_\tau$*.*

*Moreover, assume there exist* $\psi > 0$ *and* $\alpha_M \in \mathbb{R}^M$ *such that* $\sup_{y\in\mathcal{Y}} |C_\tau(y) - e^{(M)}(y)'\,\alpha_M| = O(M^{-\psi})$ *as* $M \to \infty$*.*

*Then:*

$$\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(C_\tau - \hat{C}_\tau^{(M,N)}\right)^2\right] = O(M/N + M^{-2\,\psi}),$$

[9] Newey (1997) also provides conditions for uniform convergence and for asymptotic normality of series estimators. We refer to his paper for details.

*i.e. we have joint convergence in* $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$.

In this result, we clearly see the influence of the two approximations: The functional approximation is reflected in the second part of the expression for the convergence rate. Here, it is worth noting that the speed $\psi$ will depend on the choice of the basis functions, emphasizing the importance of this aspect. The first part of the expression corresponds to the regression approximation, and in line with the second part of Proposition 1.3.1 it goes to zero linearly in $N$. However, it is important to note that to ensure convergence in the first place, the conditions require that $\xi_0(M)^2\, M/N \to 0$ – and not only $M/N \to 0$ as it appears in the convergence rate – where $\xi_0$ again depends on the choice of the basis functions and the underlying stochastic model.

The result provides general conditions that can be checked for any selection of basis functions, although ascertaining them for each underlying stochastic model may be cumbersome. Newey also provides explicit conditions for the practically relevant case of power series. In our notation, they read as follows:

**Proposition 1.3.3** (Newey (1997))**.** *Assume* $Var(V_\tau|Y_\tau)$ *is bounded and that the basis functions* $e^{(M)}(Y_\tau)$ *consist of orthonormal polynomials, that* $\mathcal{Y}$ *is a Cartesian product of compact connected intervals, and that a sub-vector of* $Y_\tau$ *has a density that is bounded away from zero. Moreover, assume that* $C_\tau(y)$ *is continuously differentiable of order* $s$.

*Then, if* $M^3/N \to 0$, *we have:*

$$\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(C_\tau - \hat{C}_\tau^{(M,N)}\right)^2\right] = O(M/N + M^{-2s/d}),$$

*i.e. we have joint convergence in* $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$.

Hence, for orthonormal polynomials, the conditions entail $M^3/N \to 0$, i.e. the number of simulations has to increase faster than the cube of the number of basis functions. In particular, to ascertain convergence for a large set of basis functions,

a very large number of simulations is required. Moreover, the smoothness of the conditional expectation is important. First-order differentiability is required ($s \geq 1$), and if $s = 1$, the convergence of the functional approximation will only be of order $M^{-2/d}$, where $d$ is the dimension of the underlying model.

For common financial models, particularly for diffusion models, smoothness is satisfied so the latter part of the assumptions seem innocuous. On the other hand, frequently the support of the stochastic variables is unbounded. However, here convergence in probability still may be established via the Markov inequality since we can limit the consideration to products of compact intervals (see also Andrews and Whang (1990) for related results on series estimators under a weaker condition).

Regarding the properties of the estimator beyond convergence, much rides on the first (functional) approximation that we discuss in more detail in the following section. With regards to the second approximation, it is well-known that as the OLS estimate, $\widehat{C}_\tau^{(M,N)}$ is unbiased – though not necessarily efficient – for $\widehat{C}_\tau^{(M)}$ under mild conditions (see e.g. Sec. 6 in Amemiya (1985)).[10] However, this clearly does not imply that $\rho(\widehat{C}_\tau^{(M,N)})$ is unbiased for $\rho(\widehat{C}_\tau^{(M,N)})$. Proceeding similarly to Gordy and Juneja (2010) for the nested simulation estimator, in the next subsection we analyze this question in more detail for VaR.

### 1.3.3 LSM Estimate for Value-at-Risk

An important special case is VaR, which is the risk measure applied in regulatory frameworks such as Basel III and Solvency II. VaR does not fall in the class of convex risk measures so that Corollary 1.3.1 does not apply. However, convergence immediately follows from Proposition 1.3.1-1.3.3:

[10] Note that, in financial applications, typically the residuals are not homoscedastic. Nevertheless, on relies on a simple OLS rather than a GLS estimate since the covariance matrix is usually not known and its estimation would yet again increase the complexity of the algorithm.

**Corollary 1.3.2.** *We have:*

$$F_{\widehat{C}_\tau^{(M,N)}}(l) = \mathbb{P}(\widehat{C}_\tau^{(M,N)} \leq l) \to \mathbb{P}(C_\tau \leq l) = F_{C_\tau}(l),\ N \to \infty,\ M \to \infty,\ l \in \mathbb{R},$$

*and*

$$F^{-1}_{\widehat{C}_\tau^{(M,N)}}(\alpha) \to F^{-1}_{C_\tau}(\alpha),\ N \to \infty,\ M \to \infty,$$

*for all continuity points $\alpha \in (0,1)$ of $F^{-1}_{C_\tau}$. Moreover, under the conditions of Propositions 1.3.2 and 1.3.3, we have joint convergence.*

Gordy and Juneja (2010) show that the nested simulations estimator for VaR carries a positive bias in the order of the number of simulations in the inner step. They derive their results by considering the joint density of the exact distribution of the capital at time $\tau$ and the error when relying on a finite number of inner simulations scaled by the square-root of the number of inner simulations. The following proposition establishes that their results carry over to our setting in view of the second approximation:

**Proposition 1.3.4** (Gordy and Juneja (2010))**.** *Let $g_N(\cdot,\cdot)$ denote the joint probability density function of $(-\hat{C}_\tau^{(M)}, Z^{(N)})$, and assume that it satisfies the regularity conditions from Gordy and Juneja (2010) collected in the Appendix. Then:*

$$\mathbb{E}\left[\widehat{VaR}_\alpha\left[-\widehat{C}_\tau^{(M,N)}\right]\right] = VaR_\alpha\left[-\widehat{C}_\tau^{(M)}\right] + \frac{\theta_\alpha}{N\bar{f}\left(VaR_\alpha\left(-\widehat{C}_\tau^{(M)}\right)\right)} + o_N(N^{-1}),$$

*where $\widehat{VaR}_\alpha\left[-\widehat{C}_\tau^{(M,N)}\right]$ denotes the $\lceil (1-\alpha)N \rceil$ order statistic of $V_\tau^{(i)}$, $1 \leq i \leq N$ (the sample quantile), $\theta_\alpha = -\frac{1}{2}\frac{d}{d\mu}\left[\bar{f}(\mu)\mathbb{E}\left[\sigma^2_{Z^{(N)}} | -\widehat{C}_\tau^{(M)} = \mu\right]\right]_{\mu = VaR_\alpha\left[-\widehat{C}_\tau^{(M)}\right]}$, $\sigma^2_{Z^{(N)}} = \mathbb{E}\left[\left(Z^{(N)}\right)^2 | Y_\tau\right]$, and $\bar{f}$ is the marginal density of $-\widehat{C}_\tau^{(M)}$.*

The key point of the proposition is that – similarly to the nested simulations estimator – the LSM estimator for VaR is biased. In particular, for large losses or a large value of $\alpha$, the derivative of the density in the tail is negative resulting in a positive bias. That is, ceteris paribus, on average the LSM estimator will err on the "conservative" side (see also Bauer et al. (2012)). However, note that here we ignore the variance due to estimating the risk measure from the finite sample, which

may well trump the inaccuracy due to the bias. Indeed, as is clear from Proposition 1.3.1, the convergence of the variance is of order $N$ and thus dominates the mean-square error for relatively large values of $N$ (in which the bias will enter as $O(N^{-2})$). Moreover, of course the result only pertains to the regression approximation but not the approximation of the capital variable via the linear combination of basis functions, which is at the core of the proposed algorithm.

## 1.4 Choice of Basis Functions

As demonstrated in Section 1.3.1, any set of independent functions will lead the LSM algorithm to converge. In fact, for the LSM method for pricing non-European derivatives, frequent choices of basis functions include Hermite polynomials, Legendre polynomials, Chebyshev polynomials, Fourier series, and even simple polynomials. Based on various numerical tests, Moreno and Navas (2003) conclude that the approach is robust to the choice of basis functions (see also the original paper by Longstaff and Schwartz (2001)). A key difference between the LSM pricing method and the approach here, however, is that it is necessary to approximate the distribution over its entire domain rather than the expected value only. Furthermore, the state space for estimating a company's capital can be high-dimensional and considerably more complex than that of a derivative security. Therefore, the choice of basis functions is not only potentially more complex but also more crucial in the present context.

### 1.4.1 Optimal Basis Functions for a Model Framework

As illustrated in Section 2.1, we can identify the capital – as a function of the state vector at the risk horizon $Y_\tau$ – for a cash flow *model* $\mathbf{x}$ within a certain *model framework* $Y$ with the output of the linear operator $L$ applied to $\mathbf{x}$: $C_\tau(Y_\tau) = L\mathbf{x}(Y_\tau)$ (cf. Eq. (1.3)). As discussed in Section 1.3.2, the LSM algorithm, in turn, approximates $C_\tau$ by its linear projection on the subspace spanned by the basis functions $e^{(M)}(Y_\tau)$, $P\,C_\tau(Y_\tau)$,

where $P$ is the projection operator.

For simplicity, in what follows, we assume that the basis functions are orthonormal in $L^2(\mathbb{R}, \mathcal{B}, \mathbb{P}_{Y_\tau})$. Then we can represent $P$ as:

$$P\cdot = \sum_{k=1}^{M} \langle \cdot, e_k(Y_\tau) \rangle_{L^2(\mathbb{P}_{Y_\tau})} \, e_k.$$

Therefore, the LSM approximation can be represented via the *finite rank* operator $L_F = P\,L$, where we have:

$$\begin{aligned} L_F\mathbf{x} &= P\,L\,\mathbf{x} = \sum_{k=1}^{M} \langle L\mathbf{x}, e_k(Y_\tau) \rangle_{L^2(\mathbb{P}_{Y_\tau})} \, e_k \\ &= \sum_{k=1}^{M} \mathbb{E}^{\mathbb{P}} \left[ e_k(Y_\tau) \sum_{j=\tau}^{T} \mathbb{E}^{\tilde{\mathbb{P}}} \left[ x_j(Y_j) | Y_\tau \right] \right] e_k = \sum_{k=1}^{M} \mathbb{E}^{\mathbb{P}} \left[ e_k(Y_\tau) \underbrace{\sum_{j=\tau}^{T} x_j(Y_j)}_{=V_\tau} \right] e_k \\ &= \sum_{k=1}^{M} \underbrace{\mathbb{E}^{\tilde{\mathbb{P}}} \left[ e_k(Y_\tau)\, V_\tau \right]}_{\alpha_k} e_k, \end{aligned} \tag{1.7}$$

where the fourth equality follows by the tower property of conditional expectations.

It is important to note that under this representation, ignoring the uncertainty arising from the regression estimate, the operator $L_F$ gives the LSM approximation for *each* model $\mathbf{x}$ within the *model framework*. That is, the choice of the basis function *precedes* fixing a particular cash flow model (payoff). Thus, we can define *optimal basis functions* as a system that minimizes the distance between $L$ and $L_F$, so that the approximation is optimal with regards to all possible cash flow models within the framework:

**Definition 1.4.1.** *We call the set of basis functions* $\{e_1^*, e_2^*, ..., e_M^*\}$ *optimal in* $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$ *if:*

$$\{e_1^*, e_2^*, ..., e_M^*\} = argmin_{\{e_1, e_2, ..., e_M\}} \|L - L_F\| = argmin_{\{e_1, e_2, ..., e_M\}} \sup_{\|\mathbf{x}\|=1} \|L\mathbf{x} - L_F\mathbf{x}\|.$$

This notion of *optimality* has various advantages in the context of calculating risk capital. Unlike pricing a specific derivative security with a well-determined payoff, capital may need to be calculated for subportfolios or only certain lines of business for the purposes of capital allocation. Moreover, a company's portfolio will change from one calculation date to the next, so that the relevant cash flow model is in flux. The underlying model framework, on the other hand, is usually common to all subportfolios since the purpose of a capital framework is exactly the enterprise-wide determination of diversification opportunities and systematic risk factors. Moreover, it is typically not frequently revised. Hence, it is expedient here to connect the optimality of basis functions to the framework rather than a particular model (payoff).

### 1.4.2 Optimal Basis Functions for a Compact Valuation Operator

In order to derive optimal basis functions, it is sufficient to determine the finite-rank operator $L_F$ that presents the best approximation to the infinite-dimensional operator $L$. If $L$ is a compact operator, this operator is immediately given by the *singular value decomposition* (SVD) of $L$ (for convenience, details on the SVD of a compact operator are collected in the Appendix). More precisely, we can then represent $L : \mathcal{H} \to L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$ as:

$$L\,\mathbf{x} = \sum_{k=1}^{\infty} \omega_k \,\langle \mathbf{x}, s_k \rangle\, \varphi_k, \tag{1.8}$$

where $\{\omega_k\}$ with $\omega_1 \geq \omega_2 \geq \ldots$ are the singular values of $L$, $\{s_k\}$ are the right singular functions of $L$, and $\{\varphi\}_k$ are the left singular functions of $L$ – which are exactly the eigenfunctions of $L\,L^*$. As demonstrated by the following proposition, the optimal basis functions are given by the left singular functions of $L$.

**Proposition 1.4.1.** *Assume the operator $L$ is compact. Then for each $M$, the left singular functions of $L$ $\{\varphi_1, \varphi_2, \ldots, \varphi_M\} \in L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$ are optimal basis functions in the sense of Definition 1.4.1. For a fixed cash flow model, we obtain $\alpha_k = \omega_k \,\langle \mathbf{x}, s_k \rangle$.*

The result that the left singular functions provide an optimal approximation may not be surprising given related results in finite dimensions. In particular, our proof is similar to the Eckart-Young-Mirsky Theorem on low-rank approximations of an arbitrary matrix. A sufficient condition for the compactness of the operator $L$ is provided in Lemma 1.2.3.

To appraise the impact of the two approximations simultaneously, we can analyze the joint convergence properties in $M$ and $N$ for the case of optimal basis functions. Here, in general, we have to check the conditions from Proposition 1.3.2. We observe that the convergence rate associated with the first approximation depends on the quantity $\psi$, which in the present context depends on the speed of convergence of the singular value decomposition:

$$\begin{aligned} O(M^{-\psi}) = \inf_{\alpha_M} \sup_{y \in \mathcal{Y}} |C_\tau(y) - e^{(M)}(y)' \, \alpha_M| &\leq \sup_{y \in \mathcal{Y}} |\, L\,\mathbf{x}\,(y) - L_F\,\mathbf{x}\,(y)| \\ &= \sup_{y \in \mathcal{Y}} \left| \sum_{k=M+1}^{\infty} \omega_k \left\langle \mathbf{x}, s_k \right\rangle \varphi_k(y) \right|. \quad (1.9) \end{aligned}$$

In particular, we are able to provide an explicit result in the case of bounded singular functions:

**Proposition 1.4.2.** *Assume $Var(V_\tau|Y_\tau)$ is bounded and that the singular functions, $\{\varphi_k\}_{k=1}^{\infty}$, are uniformly bounded on the support of $Y_\tau$. Then, if $M^2/N \to 0$, we have:*

$$\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(C_\tau - \hat{C}_\tau^{(M,N)}\right)^2\right] = O(M/N + \omega_M^2),$$

*i.e. we have joint convergence in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$.*

In the general (unbounded) case, according to Equation (1.9), the convergence will depend on the properties of the singular functions as well as the speed of convergence of the singular values. Here, similarly to Proposition 1.3.3 for orthonormal polynomials, the latter convergence depends on the smoothness of the kernel $k(x, y)$ (see

Birman and Solomyak (1977) for a survey on the convergence of singular values of integral operators). However, Equation (1.9) again illustrates the intuition behind the optimality criterion: To choose a basis function that minimizes the distance between the operators for all $\mathbf{x}$, although in the Definition we consider the $L^2$ norm rather than the supremum.

The derivation of the SVD of the valuation operator of course depends on the specific model framework. In some cases, it is possible to carry out the calculations and derive analytical expressions for the singular values. In the next subsection, we determine the SVD – and, thus, optimal basis functions – in the practically highly relevant case of Gaussian transition densities. Here, the optimal basis functions correspond to Hermite polynomials of suitably transformed state variables (Proposition 1.4.3).

### 1.4.3 Optimal Basis Functions for Gaussian Transition Densities

In what follows, consider a single cash flow at time $T$ only and let $(Y_t)$ be a $\mathbb{R}^d$-dimensional Markov process such that $(Y_\tau, Y_T)$ are jointly Gaussian distributed. We denote the distribution under $\tilde{\mathbb{P}}$ of this random vector via:

$$\begin{pmatrix} Y_\tau \\ Y_T \end{pmatrix} \sim N\left( \begin{pmatrix} \mu_\tau \\ \mu_T \end{pmatrix}, \quad \begin{pmatrix} \Sigma_\tau & \Gamma \\ \Gamma' & \Sigma_T \end{pmatrix} \right), \tag{1.10}$$

where $\mu_\tau$, $\mu_T$, $\Sigma_\tau$, and $\Sigma_T$ are the mean vectors and variance-covariance matrices of $Y_\tau$ and $Y_T$, respectively, and $\Gamma$ is the corresponding (auto) covariance matrix – which we assume to be non-singular.[11] Note that the specific form of these parameters depends on the choice of a numéraire $N_t$, $\tau \leq t \leq T$.

Denoting by $g(x; \mu, \Sigma)$ the joint normal probability density function at $x$ with mean vector $\mu$ and covariance matrix $\Sigma$, the marginal densities of $Y_\tau$ and $Y_T$ are $\pi_{Y_\tau}(x) = g(x; \mu_\tau, \Sigma_\tau)$ and $\pi_{Y_T}(y) = g(y; \mu_T, \Sigma_T)$, respectively. Mapping these assumption to the

[11] The distribution in (1.10) is the unconditional distribution with known $Y_0$.

previous notation yields $\mathbf{x} = x_T$, $L : \mathcal{H} = L^2(\mathbb{R}^d, \mathcal{B}, \pi_{Y_T}) \to L^2(\mathbb{R}^d, \mathcal{B}, \pi_{Y_\tau})$, and

$$C_\tau(Y_\tau) = L\mathbf{x}(Y_\tau) = \int_{\mathbb{R}^d} x_T(y)\, \pi_{Y_T|Y_\tau}(y|Y_\tau)\, dy,$$

where $\pi_{Y_T|Y_\tau}(y|x)$ denotes the transition density. In order to obtain optimal basis functions, the objective is to derive the SVD of $L$.

**Lemma 1.4.1.** *We have for the conditional distributions:*

$$Y_T|Y_\tau \sim N\left(\mu_{T|x}, \Sigma_{T|\tau}\right) \text{ and } Y_\tau|Y_T \sim N\left(\mu_{\tau|y}, \Sigma_{\tau|T}\right)$$

*with transition density and reverse transition density:*

$$\pi_{Y_T|Y_\tau}(y|x) = g(y; \mu_{T|\tau}(x), \Sigma_{T|\tau}) \text{ and } \pi_{Y_\tau|Y_T}(x|y) = g(x; \mu_{\tau|T}(y), \Sigma_{\tau|T}),$$

*respectively, where* $\mu_{T|\tau}(x) = \mu_T + \Gamma'\Sigma_\tau^{-1}(x - \mu_\tau)$, $\Sigma_{T|\tau} = \Sigma_T - \Gamma'\Sigma_\tau^{-1}\Gamma$, $\mu_{\tau|T}(y) = \mu_\tau + \Gamma\Sigma_T^{-1}(y - \mu_T)$, *and* $\Sigma_{\tau|T} = \Sigma_\tau - \Gamma\Sigma_T^{-1}\Gamma'$. *Moreover, $L$ is compact in this setting.*

Per Proposition 1.4.1, the optimal basis functions are given by the left singular functions, which are in turn the eigenfunctions of $L\,L^*$. We obtain:

**Lemma 1.4.2.** *The operator $LL^*$ and $L^*L$ are integral operators:*

$$LL^* f(\cdot) = \int_{\mathbb{R}^d} K_A(\cdot, y)\, f(y)\, dy \text{ and } L^*Lf(\cdot) = \int_{\mathbb{R}^d} K_B(\cdot, x)\, f(x)\, dx,$$

*where the kernels are given by Gaussian densities:*

$$K_A(x, y) = g(y; \mu_A(x), \Sigma_A) \text{ and } K_B(y, x) = g(x; \mu_B(y), \Sigma_B)$$

*with*

- $\mu_A(x) = \mu_\tau + A(x - \mu_\tau)$, $A = \Gamma\Sigma_T^{-1}\Gamma'\Sigma_\tau^{-1}$, *and* $\Sigma_A = \Sigma_\tau - A\Sigma_\tau A'$;
- $\mu_B(y) = \mu_T + B(y - \mu_T)$, $B = \Gamma'\Sigma_\tau^{-1}\Gamma\Sigma_T^{-1}$, *and* $\Sigma_B = \Sigma_T - B\Sigma_T B'$.

*We denote by* $\mathbb{E}_{K_A}[\cdot|x]$ *and* $\mathbb{E}_{K_B}[\cdot|y]$ *the expectation operators under the Gaussian densities* $K_A(x,\cdot)$ *and* $K_B(y,\cdot)$*, respectively.*

The problem of finding the singular values and the left singular functions therefore amounts to solving the eigen-equations:

$$\mathbb{E}_{K_A}\left[f(Y)|x\right] = \omega^2\, f(x).$$

We exploit analogies to the eigenvalue problem of the Markov operator of a first-order multivariate normal autoregressive (MAR(1)) process studied in Khare and Zhou (2009) to obtain the following:

**Lemma 1.4.3.** *Denote by* $P\Lambda P'$ *be the eigenvalue decomposition of*

$$\Sigma_\tau^{-1/2}\, A\, \Sigma_\tau^{1/2} = \Sigma_\tau^{-1/2}\, \Gamma \Sigma_T^{-1} \Gamma' \Sigma_\tau^{-1/2},$$

*where* $PP' = I$ *and* $\Lambda$ *is the diagonal matrix whose diagonal entries are the eigenvalues* $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_d|$ *of* $A$*. For* $y \in \mathbb{R}^d$*, define the transformation:*

$$z^P(y) = P'\Sigma_\tau^{-1/2}(y - \mu_\tau). \tag{1.11}$$

*Then:*

$$\mathbb{E}_{K_A}\left[z^P(Y)|x\right] = \Lambda\, z^P(x).$$

*Moreover,* $Var_{K_A}\left[z^P(Y)|x\right] = I - \Lambda^2$*,* $\mathbb{E}_{\pi_{Y_\tau}}\left[z^P(Y_\tau)\right] = 0$*, and* $Var_{\pi_{Y_\tau}}\left[z^P(Y_\tau)\right] = I$*.*

*Similarly, denote the diagonalization* $\Sigma_T^{-1/2} B \Sigma_T^{1/2} = Q\Lambda Q'$*, where* $Q'Q = I$ *and define the transformation:*

$$z^Q(x) = Q'\Sigma_T^{-1/2}(x - \mu_T). \tag{1.12}$$

*Then for* $X \sim K_B(y, \cdot)$, *we have:*

$$\mathbb{E}_{K_B}\left[z^Q(X)|y\right] = \Lambda\, z^A(y),$$

$Var_{K_B}\left[z^Q(X)|y\right] = I - \Lambda^2$, $\mathbb{E}_{\pi_{Y_T}}\left[z^Q(Y_T)\right] = 0$, *and* $Var_{\pi_{Y_T}}\left[z^Q(Y_T)\right] = I$.

Therefore, for a random vector $Y|x$ in $\mathbb{R}^d$ that is distributed according to $K(x, \cdot)$, the components $z_i^P(Y)$ of $z^P(Y)$ are independently distributed with $z_i^P(Y) \sim N(\lambda_i\, z_i^P(x), 1 - \lambda_i^2)$, where $z_i^P(x)$ is the $i$-th component of $z^P(x)$. Since eigenfunctions of standard Gaussian distributed random variables are given by Hermite polynomials, the SVD follows immediately from Lemma 1.4.3:

**Proposition 1.4.3.** *Denote the Hermite polynomial of degree* $j$ *by* $h_j(x)$, *that is:*[12]

$$h_0(x) = 1,\ h_1(x) = x,\ h_j(x) = \frac{1}{\sqrt{j}}\left(x\, h_{j-1}(x) - \sqrt{j-1}h_{j-2}(x)\right), \quad j = 2, 3, ...$$

*The singular values of* $L$ *in the current (Gaussian) setting are given by:*

$$\omega_{|n|} = \Pi_{i=1}^d \lambda_i^{n_i/2}, \quad n = (n_1, ..., n_d) \in \mathbb{N}_0^d, \tag{1.13}$$

*where* $\mathbb{N}_0^d$ *is the set of d-dimensional non-negative integers,* $|n| = \sum_{i=1}^d n_i$, *and the corresponding right and left singular functions are:*

$$s_{|n|}(x) = \Pi_{i=1}^d h_{n_i}(z_i^Q(x)) \text{ and } \varphi_{|n|}(y) = \Pi_{i=1}^d h_{n_i}(z_i^P(y)),$$

*respectively.*

We know from Proposition 1.4.1 that the left singular functions $\varphi$ will present optimal choices for the basis functions in the LSM algorithm. Note that in the univariate case ($d = 1$), $A = \lambda_1$ is simply the square of the correlation coefficient between $Y_\tau$ and

[12] See Kollo and Rosen (2006) for real and vector valued Hermite polynomials and the normalization employed here.

$Y_T$ – so that the singular values are simply powers of this correlation. Thus, the SVD takes the form

$$L\,\mathbf{x}(Y_\tau) = \sum_{k=1}^{\infty} \left(\mathrm{Corr}(Y_\tau, Y_T)\right)^{k-1} \left\langle x_T, h_{k-1}\left(\frac{Y_T - \mu_T}{\Sigma_T}\right)\right\rangle_{\pi_{Y_T}} h_{k-1}\left(\frac{Y_\tau - \mu_\tau}{\Sigma_\tau}\right).$$

In particular, the optimal basis functions are simply given by Hermite polynomials of the normalized Markov state – although other choices of polynomial bases will generate the same span so that the results will coincide.

In the general multivariate case, it is clear from Proposition 1.4.3 that the singular values of $L$ are directly related to eigenvalues of the matrix $A$ (or, equivalently, $B$). In particular, the largest eigenvalue gives the most important dimension for the basis function – which according to (1.11) is a linear transformation of the normalized state vector. However, it is important to note that the subscripts in Proposition 1.4.3 do not correspond to those in the SVD (1.8). There are $d$ vectors of indices $n$ such that $|n| = 1$, $d^2$ vectors of indices such that $|n| = 2$, etc. in Equation (1.13), whereas in the SVD (1.8) the functions are ordered according to the singular values. For instance, while for $1 > \lambda_1 > \lambda_2$ clearly $\sqrt{\lambda_1} > \sqrt{\lambda_1}^2 = \lambda_1$ and similarly for $\lambda_2$, it is not clear whether $\lambda_1^2 > \lambda_2$ or vice versa – and the order will determine which combination of basis functions is optimal. Thus in the multi-dimensional case – and particularly in high-dimensional settings that are very relevant for practical applications – is where the analysis here provides immediate guidance. Even if a user chooses the same function class (Hermite polynomials) or function classes with the same span (other polynomial families), it is unlikely that a naïve choice will pick the suitable combinations – and this choice becomes less trivial and more material as the number of dimensions increases.

From Proposition 1.3.1, we obtain sequential convergence. As polynomials with full support, the left singular functions do not satisfy the uniformly boundedness assumptions of Proposition 1.4.2 and due to the unbounded domain formally also the requirements of Propositions 1.3.2 and 1.3.3 are not satisfied. However, following the

discussion after Proposition 1.3.3, we have joint convergence in probability as long as ${}^{M^3}/_N \to 0$.

In models with non-Gaussian transitions, while an analytical derivation may not be possible, we can rely on numerical methods to determine approximations of the optimal basis functions. For instance, Huang (2012) explains how to solve the associated integral equation by discretization method, which allows to determine the singular function numerically. Alternatively, Serdyukov et al. (2014) apply the truncated SVD to solve inverse problems numerically.

## 1.5 Applications

To illustrate the LSM algorithm and its properties, we consider two examples from life insurance: A Guaranteed Annuity Option (GAO) within a conventional pure endowment policy and a Guaranteed Minimum Income Benefit (GMIB) within a Variable Annuity contract. As indicated in the Introduction, the LSM algorithm is particularly relevant in insurance, especially in light of the dawning Solvency II regulation that comes into effect in 2016. Here, the so-called *Solvency Capital Requirement* within an *internal model* takes the form of a 99.5% VaR of the available capital at the risk horizon $\tau = 1$ (see Bauer et al. (2012) for details).

### 1.5.1 Application to GAO

GAOs are common in many markets and, as described Boyle and Hardy (2003), these options were a major factor in the demise of Equitable Life, the world's oldest life insurance company, in 2000. We consider the valuation of a GAO attached to a basic pure endowment policy under the Vasicek (1977) interest rate model. This framework has two advantages. First, following Boyle and Hardy (2003) and Pelsser (2003), it is possible to derive a closed form valuation formula. Hence, we can exactly simulate the capital level at the risk horizon and derive a closed form for the VaR. This allows us to

appraise the performance of the LSM algorithm by comparing numerical results to the "exact" quantities that are not subject to the functional approximation. Moreover, since the Vasicek model is driven by a simple Ornstein-Uhlenbeck (OU) process, it falls in the class of models considered in Section 1.4.3 and we can rely on the corresponding results to obtain optimal basis functions.

**Payoff of the GAO and Valuation Formula**

We consider a large portfolio of pure endowment policies with a GAO. In particular, we abstract from mortality risk (aggregate systematic risk as well as small sample risk), and to ease notation we derive all expressions for a single policyholder aged $x$ at time zero. Following standard actuarial notation we denote the $k$-year survival probability by ${}_kp_x$.

Under a plain pure endowment policy, the policyholder receives a fixed payment $P$ upon survival until the maturity date $T$ and nothing in case death occurs before time $T$. Thus, the time-$t$ value of the basic contract – if the policyholder is alive – is $P\,p(t,T)\,{}_{T-t}p_{x+t}$, where $p(t,T)$ is the value at time $t$ of a zero-coupon bond with maturity $T$. The benefit can be taken out as a fixed payment or can be converted into a life annuity under the concurrent market annuity payout rate, $m_{x+T}(T)$. In the latter case, the policyholders will receive a payment of $P\,m_{x+T}(T)$ each year upon survival past year $T$.

In contrast, when the policy is equipped with a GAO, upon survival the policyholder has the right to choose at maturity between (i) a fixed payment of $P$, (ii) a life annuity at the market rate $P\,m_{x+T}(T)$, or (iii) a life annuity with a guaranteed payout rate $g$ fixed at the policy's inception. Clearly, (i) and (ii) will result in the same (market) value, so that the time $T$ payoff for the pure endowment plus GAO is

given by the maximum of options (ii) and (iii):[13]

$$P \max\{g, m_{x+T}(T)\} \underbrace{\sum_{k=1}^{\infty} {}_k p_{x+T}\, p(T, T+k)}_{=a_{x+T}(T)},$$

where $a_{x+T}(T)$ denotes the time $T$-value of an immediate annuity on an $(x+T)$-year old policyholder. We clearly have $m_{x+T}(T) = {}^1/_{a_{x+T}(T)}$, so that:

$$P \max\{g, m_{x+T}(T)\}\, a_{x+T}(T) = P + \underbrace{P \max\{g\, a_{x+T}(T) - 1, 0\}}_{=C(T)}.$$

Here, the bond prices within the annuity present value depend on the concurrent (time $T$) interest rate $r_T$, so that $C(T)$ takes the form of an interest rate derivative. For its valuation, we follow Vasicek (1977) and assume the interest rate evolves according to unidimensional OU process:

$$dr_t = \alpha(\gamma - r_t)\, dt + \sigma\, dW_t, \tag{1.14}$$

under the physical measure $\mathbb{P}$, whereas the dynamics under the risk-neutral measure $\mathbb{Q}$ are given by:

$$dr_t = \alpha(\bar{\gamma} - r_t)\, dt + \sigma\, dZ_t. \tag{1.15}$$

Here $\alpha$ is the speed of mean reversion, $\gamma$ is the mean reversion level, $\sigma$ is the volatility, $\bar{\gamma} = \gamma - \lambda\sigma/\alpha$ where $\lambda$ is market price of risk, and $(W_t)$ and $(Z_t)$ are standard Brownian motions under the physical measure and risk-neutral measure, respectively. Following Boyle and Hardy (2003), who rely on the approach by Jamshidian (1989) for pricing

[13] Clearly, this entails the strong assumption on the policyholder's behavior that she chooses the value-maximizing option. While this may not be the case in a realistic setting with financial frictions, incomplete markets, or behavioral biases (Bauer et al., 2015), we accept it here for illustrative purposes.

options on a coupon bond, we obtain for the value of the GAO:

$$\begin{aligned} c(t) &= \mathbb{E}^{\mathbb{Q}}\left[ {}_{T-t}p_{x+t}\, e^{-\int_t^T r_s\, ds}\, C(T)\,|r_t \right] && (1.16)\\ &= g\, {}_{T-t}p_{x+t} \sum_{k=1}^{\infty} {}_kp_{x+T}\, \left[ p(t,T)\,\Phi(h) - K_k\, p(t,T+k)\,\Phi(h-\tilde{\sigma}) \right]. && (1.17) \end{aligned}$$

Here $\Phi(\cdot)$ denotes the standard Normal cumulative distribution function,

$$\tilde{\sigma} = \sigma \sqrt{\frac{1-e^{-2\alpha(T-t)}}{2\alpha}}\, \frac{1-\exp(-\alpha\, k)}{\alpha}, \quad h = \frac{1}{\tilde{\sigma}} \log\left( \frac{p(t,T+k)}{p(t,T)K_k} \right) + \frac{\tilde{\sigma}}{2},$$

and the strike price $K_k$ is given by $p^*(T,T+k)$, where $r_T^*$ is the interest rate such that

$$\sum_{k=1}^{\infty} {}_kp_{x+T}\, p^*(T,T+k) = 1/g$$

and $p^*(T,T+k)$ is the price of zero coupon bond priced at rate $r_T^*$. Thus, the price of the pure endowment plus GAO policy is:

$$v(t) = P\left( c(t) + p(t,T)\, {}_{T-t}p_{x+t} \right). \tag{1.18}$$

**Capital Requirement for the GAO**

The (available) capital at the risk horizon $\tau$ is given by the present value of assets $A_\tau$ minus liabilities $L_\tau$. For the single pure endowment plus GAO policy considered here, we obtain:

$$\begin{aligned} C_\tau = A_\tau - L_\tau &= A_\tau - \underbrace{P\, \left( {}_{T-\tau}p_{x+\tau}\, p(\tau,T) + c(\tau) \right)}_{=v(\tau)} \\ &= A_\tau - P\, {}_{T-\tau}p_{x+\tau}\, p(\tau,T)\, \mathbb{E}^{\mathbb{Q}_T}\left[ 1 + C(T)|r_\tau \right], && (1.19) \end{aligned}$$

where $\mathbb{Q}_T$ denotes the $T$-forward measure, i.e. the risk-neutral measure when choosing $(p(t,T))$ as the numéraire process. For the dynamics of the risk-free rate, we have:

$$dr_t = \alpha(\bar{\gamma} - {}^{\sigma^2}/_{\alpha^2}\,(1 - e^{-\alpha(T-t)}) - r_t)\,dt + \sigma\,dZ_t^T,$$

where $\left(Z_t^T\right)$ is a Brownian motion under $\mathbb{Q}_T$. The capital requirement can then be determined by a risk measure $\rho$ applied to $-C_\tau$: $\rho(-C_\tau)$ (see e.g. Eq. (1.1) in the case of VaR).

For simplicity, we ignore asset risk in what follows and simply set $A_\tau = 0$, so that we can express the capital requirement as $\rho\left(v(\tau)\right)$. Since the distribution of the risk-free rate under the physical measure is Normal, $r_\tau \sim N(\mu_\tau, \sigma_\tau^2)$ (see the proof of Lemma 1.5.1 in the Appendix for the corresponding expressions in terms of the parameters) and since $v(t)$ is decreasing in $r_t$, we can determine the capital in closed form for various risk measures. For instance, in the case of VaR, we obtain

$$\mathrm{VaR}_\alpha = v(\tau, r_\tau = \mu_{r_\tau} - \Phi^{-1}(\alpha)\,\sigma_{r_\tau}). \tag{1.20}$$

For calculating the capital requirement via the LSM algorithm, we map the notation from the previous sections to the current setting. From Equation (1.19), it is clear that the relevant state process $Y_t = r_t$ is of dimension $d = 1$. Moreover, the cash flow functional $\mathbf{x} = x_T$, where

$$x_T(r_T) = -v(T, r_T) = -P\,{}_{T-\tau}p_{x+\tau}\,p(\tau, T)\,[1 + C(T, r_T)]$$

and

$$C_\tau = L\mathbf{x}\,(r_\tau) = \mathbb{E}^{\mathbb{Q}_T}\left[x_T(r_T)|r_\tau\right].$$

To apply Proposition 1.4.3 to the current problem, we require the joint distribution of the state variables $r_\tau$ and $r_T$:

**Lemma 1.5.1.** *The joint distribution of $r_\tau$ and $r_T$ under $\tilde{\mathbb{P}}$ is:*

$$\begin{bmatrix} r_\tau \\ r_T \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_\tau \\ \mu_T \end{bmatrix}, \begin{bmatrix} \sigma_\tau^2, & e^{-\alpha(T-\tau)}\sigma_\tau^2 \\ e^{-\alpha(T-\tau)}\sigma_\tau^2, & \sigma_T^2 \end{bmatrix} \right),$$

*where we refer to the proof in the Appendix for explicit expressions of $\mu_\tau$, $\sigma_\tau$, etc. in terms of the parameters.*

From Proposition 1.4.3, we then have:

$$L\mathbf{x}(r_\tau) = \sum_{k=1}^{\infty} \rho^{k-1} \langle \mathbf{x}, h_{k-1} \rangle h_{k-1}(z(r_\tau)),$$

where $\rho = e^{-\alpha(T-\tau)}\sigma_\tau/\sigma_T$ and $z(r_\tau) = (r_\tau - \mu_\tau)/\sigma_\tau$. Importantly, since the first $n$ Hermite polynomials are spanned by other families of orthogonal polynomials and even simply monomials, other polynomial families will lead to equivalent results (ignoring possible numerical issues in the calculation of the regression coefficients). However, we can compare this family to other basis functions with a different functional form; following Proposition 1.3.1, we will have (sequential) convergence for any (square-integrable) choice of basis functions.

**Numerical Results**

We parametrize the model by using representative values. We set the initial interest rate $r_0 = 5\%$, and for the interest rate parameters we assume $\alpha = 15\%$ (speed of mean reversion), $\gamma = 5\%$ (mean reversion level), $\sigma = 1\%$ (interest rate volatility), $\lambda = 3\%$ (market price of risk), and $x = 55$ (age of the policyholder). For the mortality rates, for illustrative purposes, we use a simple De Moivre model with terminal age $\omega = 110$, so that ${}_kp_x = {}^{\omega-x-k}/_{\omega-x}$. For the insurance contract, we let the face value $P = 100$, the maturity $T = 10$, and the guaranteed annuity rate $g = {}^1/_9$. This rate corresponds to a (flat) interest rate of a little over 6%, so that the option will frequently be in-the-money. Finally, we set the risk horizon $\tau = 1$ as it is typical in insurance.

We start by analyzing the LSM approximation to the capital variable as we vary the number of basis functions. In Figure 1.1, we display the empirical density functions based on $N = 60,000$ Monte Carlo simulations for *exact* realizations according to Equation (1.17) and approximate realizations calculated via the LSM algorithm for different numbers of basis functions $M$. Here we rely on the optimal basis functions from Proposition 1.4.3 (Hermite polynomials). As is evident from the figure, the approximation becomes closer as $M$ increases, although already for low values of $M$ the LSM algorithm seems to capture the basic shape of the density. Hence, this first analysis seems encouraging that the LSM algorithm can provide viable results at relatively low computational costs.

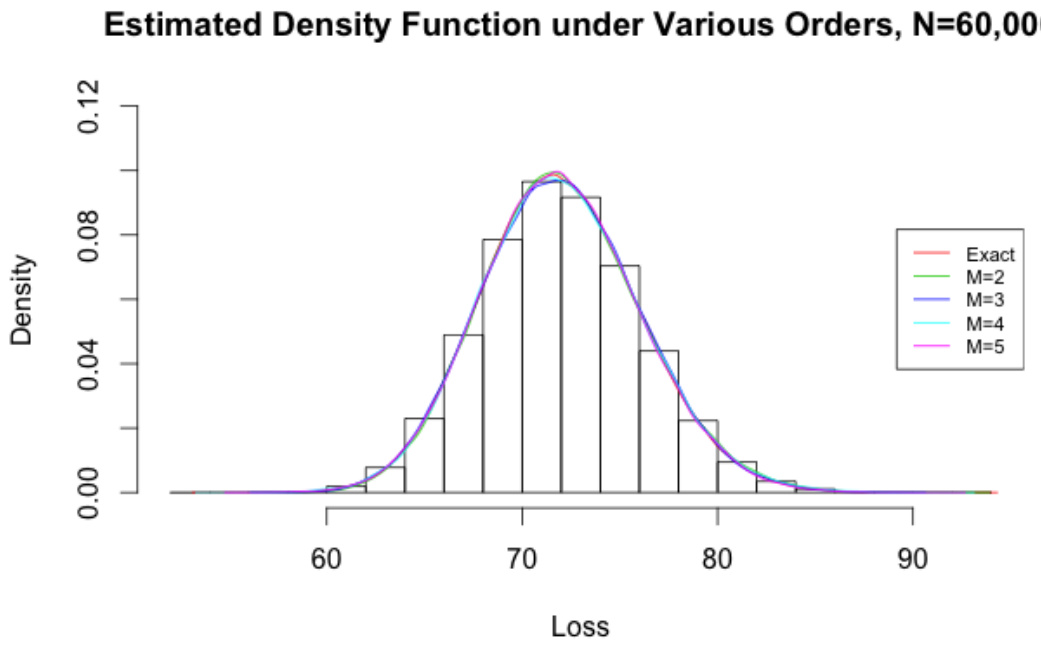
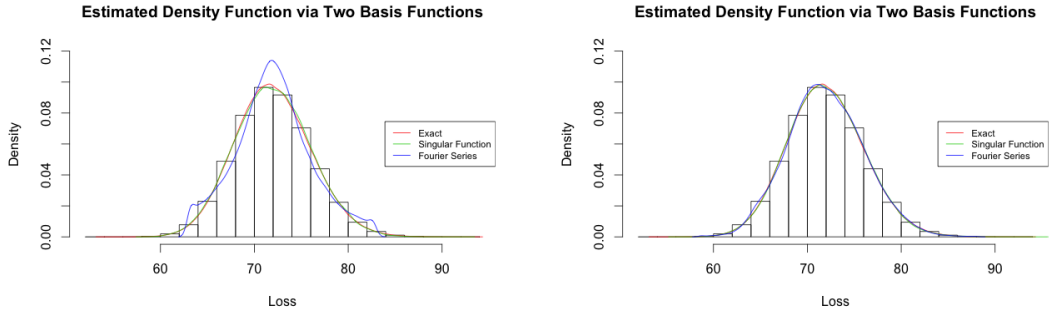



Figure 1.1: Empirical density functions of $v(\tau)$ based on $N = 60,000$ Monte Carlo realizations; exact and using the LSM algorithm with $M$ singular functions in the approximation.

To appraise the influence of the choice of basis functions, in Figure 1.2 we compare the LSM approximation based on the singular functions as used in Figure 1.1 to a different choice of basis functions, namely the first $M$ elements of the Fourier basis. We

observe that the approximation based on the (non-optimal) Fourier series is noticeably worse. In particular, from the upper panel (1.2a) with $M = 4$, we find that the Fourier basis is not able to accurately reflect the shape of the density function. As the number of basis functions increases, of course the approximation becomes better as is evident from lower panel (1.2b) with $M = 10$. However, still the optimal basis functions provide a considerably better fit.




(a) $M = 4$, $N = 60,000$ (b) $M = 10$, $N = 500,000$

Figure 1.2: Empirical density functions of $v(\tau)$ based on $N$ Monte Carlo realizations; exact and using the LSM algorithm with different basis functions ($M$ terms)

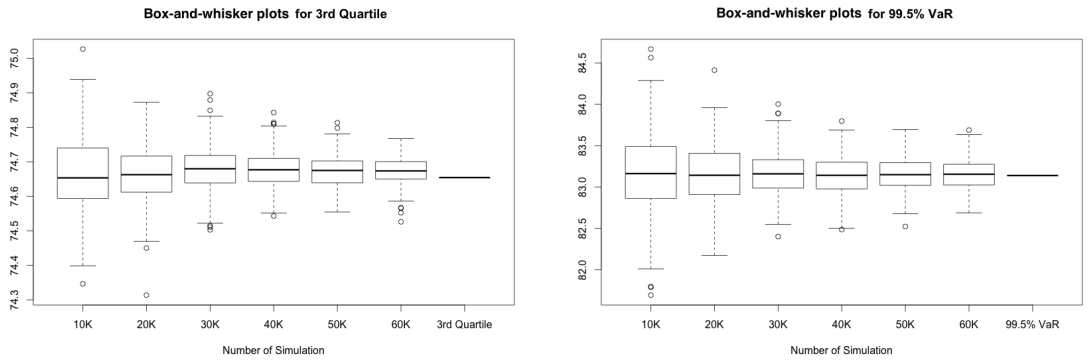
Table 1.1 reinforces this insight. Here, we show statistical differences between the empirical density functions based on $N = 700,000$ realizations (we report the mean of two-hundred runs) using, on the one hand, the exact realizations of the capital and, on the other hand, an LSM approximation. We compare differences for various choices of basis functions, both in view of the number of function terms $M$ and the function class (singular functions / polynomials vs. Fourier basis). For each combination, the table reports three common statistical distance measures: the Kolmogorov-Smirnov statistic (KS), the Kullback-Leibler divergence (KL), and the Jensen-Shannon divergence (JS). There are two key observations. First, the statistical distances are considerably smaller for the optimal choice of singular functions relative to the Fourier series. This holds for all combinations and distance measures, and, depending on the metric, the discrepancy is quite large. Second, the statistical difference increases for the singular functions as we add additional basis functions, i.e. as $M$ increases. The reason becomes clear

when recalling our results on joint convergence: When increasing $M$, the error due to the regression approximation increases (the second approximation from Section 1.2.2 corresponding to the first term in the convergence order from Proposition 1.3.2). For the Fourier basis, on the other hand, adding a basis term sometimes decreases and sometimes increases the distance. Here, both aspects in the convergence rate are at work – as $M$ increases and with fixed $N$, the regression approximation worsens but the functional approximation improves – with either of them dominating in some cases.

| Order | | Singular Functions | Fourier Series |
|---|---|---|---|
| $M = 3$ | KS | $2.218 \times 10^{-3}$ | $6.601 \times 10^{-2}$ |
| | KL | $1.413 \times 10^{-8}$ | $5.226 \times 10^{-5}$ |
| | JS | $5.465 \times 10^{-5}$ | $3.594 \times 10^{-3}$ |
| $M = 4$ | KS | $2.291 \times 10^{-3}$ | $6.570 \times 10^{-2}$ |
| | KL | $1.896 \times 10^{-8}$ | $9.582 \times 10^{-6}$ |
| | JS | $6.555 \times 10^{-5}$ | $1.507 \times 10^{-3}$ |
| $M = 5$ | KS | $2.423 \times 10^{-3}$ | $6.208 \times 10^{-2}$ |
| | KL | $2.421 \times 10^{-8}$ | $9.324 \times 10^{-6}$ |
| | JS | $7.435 \times 10^{-5}$ | $1.483 \times 10^{-3}$ |

Table 1.1: Statistical Distances between the empirical density function based on the exact realizations and the LSM approximation using different basis functions; mean of two-hundred realizations of $N = 700,000$.

The key application for the LSM algorithm in practice is calculating a company's capital requirement (economic capital), which is cast via a risk measure applied to the simulated distribution. Figure 1.3 shows results for the third quartile ($\text{VaR}_{75\%}$) and the 99.5% VaR ($\text{VaR}_{99.5\%}$). In both cases, we show results for different numbers of simulations $N$ used in the LSM algorithm on the $x$-axis. We use the first $M = 3$ (fixed) singular functions as basis functions. For each combination of risk measure and $N$, we run the LSM algorithm 300 times and determine the risk measure for each run. Figure 1.3 provides box plots of the outcomes (the box presents the area between the first and third quartile, with the inner line at the median; the whisker line spans samples that are located closer than 150% of the interquartile range to the upper and lower quartiles, respectively (Tukey boxplot)).



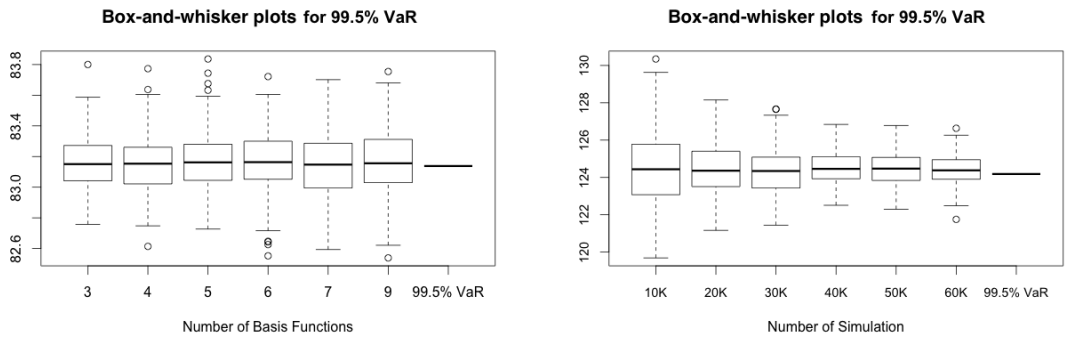

(a) Third quartile

(b) 99.5% VaR

Figure 1.3: Box-and-whisker diagrams for different risk measures (mean, third quartile, and 99.5% VaR) calculated using the LSM algorithm with different number of simulations $N$; the number of basis functions is fixed at $M = 3$.

The VaR formula from Equation (1.20) yields 74.65 and 83.14 for the third quartile and the 99.5% VaR, respectively. From Figure 1.3, it appears that the LSM algorithm produces viable results even with a relatively small number of simulations, e.g. ranging between about 74.5 to 74.8 for $\text{VaR}_{75\%}$ when using 20,000 simulations. However, this range becomes wider as we move towards the tail of the distribution, with the corresponding estimates for $\text{VaR}_{99.5\%}$ ranging between roughly 82 to 84.5. We observe a slight downward trend in the mean of the $\text{VaR}_{99.5\%}$ when increasing $N$ in line with the positive bias from Proposition 1.3.4. However, as also indicated in the discussion after the proposition, the bias is overshadowed by the sample variance resulting from the Monte Carlo estimation of the quantiles.

Increasing the number of simulations of course yields a more accurate estimation of the quantiles. In Figure 1.4a, we plot the distributions of 99.5% VaR for $N = 700,000$ and different choices for the number of basis functions $M$ (again box plots based on 300 runs). We find that the dispersion of the distribution becomes larger as the number of basis functions increases under fixed number of simulation. Again, this emphasizes the importance of the joint convergence rate: When increasing $M$, to ascertain the approximation improves, it is necessary to simultaneously increase $N$.




(a) $N = 700,000$ (b) $M = 3$

Figure 1.4: Box-and-whisker diagrams for 99.5% VaR calculated using the LSM algorithm with different number of basis functions and $N$ fixed (a); and different number of simulations $N$ and $M$ fixed under an increased volatility parameter.

The results are sensitive to changes in the parameters. For instance, in Figure 1.4b, we increase the volatility parameter ($\sigma$) from 1% to 2.5%. The VaR formula from Equation (1.20) yields 124.18 and we find that the range for $\text{VaR}_{99.5\%}$ at $\sigma = 2.5\%$ widens substantially relative to Figure 1.3. Thus, the required computational budget to obtain viable results may increase as the parameters change. Moreover, the positive bias arising from the VaR estimation is more evident in this case.

### 1.5.2 Application to GMIB

Within a Variable Annuity (VA) plus GMIB, at maturity $T$ the policyholder has the right to choose between a lump sum payment amounting to the current account value or a guaranteed annuity payment $b$ determined as a guaranteed rate applied to a guaranteed amount. GMIBs are popular riders for VA contracts: Between 2011 and 2013, roughly 15% of the more than $150 billion worth of Variable Annuities sold in the US contained a GMIB (LIMRA). Importantly, GMIBs are subject to a variety of risk factors, including fund (investment) risk, mortality risk, and – as long term contracts – interest rate risk. Consequently, we consider its risk and valuation in a multivariate Markov setting for these three risk factors.

**Model and Payoff of the GMIB**

As in the previous section, we consider a large portfolio of GMIBs with policyholder age $x$, policy maturity $T$, and a fixed guaranteed amount – so that the guaranteed annuity payment $b$ is fixed at time zero.[14] The payoff of the VA plus GMIB at $T$ in case of survival is given by:

$$\max\left\{S_T,\ b\,a_{x+T}(T)\right\}, \tag{1.21}$$

where $S_T$ is the underlying account value which evolves according to a reference asset net various fees (which we ignore for simplicity).

We consider a three-dimensional state process $Y_t$ governing financial and biometric risks:

$$Y_t = (q_t,\, r_t,\, \mu_{x+t})',$$

where $q_t$ denotes the log-price of the risky asset at time $t$, $r_t$ is the short rate, and $\mu_{x+t}$ is the force of mortality of an $(x+t)$-aged person at time $t$. We assume $Y_t$ satisfies the following stochastic differential equations under $\mathbb{P}$:

$$dq_t = \left(m - \frac{1}{2}\sigma_S^2\right)\,dt + \sigma_S\,dW_t^S, \tag{1.22}$$

$$dr_t = \alpha(\gamma - r_t)\,dt + \sigma_r\,dW_t^r, \tag{1.23}$$

$$d\mu_{x+t} = \kappa\mu_{x+t}\,dt + \psi\,dW_t^\mu, \tag{1.24}$$

where $m$ is the instantaneous rate of return of the risk asset, $\sigma_S$ is the asset volatility, $\kappa$ is an instantaneous rate of increment of mortality (Gompertz exponent), $\psi$ is the volatility of mortality, and $W_t^S$, $W_t^r$, and $W_t^\mu$ are standard Brownian motions under $\mathbb{P}$ with $dW_t^S\,dW_t^r = \rho_{12}\,dt$, $dW_t^S\,dW_t^\mu = dW_t^r\,dW_t^\mu = 0$, i.e. we assume independence of financial and biometric risks. Note that the solutions to the above stochastic dif-

[14] Some contract variants include path-dependent features such as ratchet guarantees (Bauer et al., 2008).

ferential equations at time $t$ are Normal distributed so that we can derive the optimal basis function using the approach in Section 1.4.3.

The dynamics of $Y_t$ under the risk-neutral measure $\mathbb{Q}$ are given by:

$$\begin{aligned} dq_t &= \left(r_t - \frac{1}{2}\sigma_S^2\right) dt + \sigma_S \, d\tilde{W}_t^S, \\ dr_t &= \alpha(\bar{\gamma} - r_t) \, dt + \sigma_r \, d\tilde{W}_t^r, \\ d\mu_{x+t} &= \kappa \mu_{x+t} \, dt + \psi \, d\tilde{W}_t^\mu, \end{aligned}$$

where $\tilde{W}_t^S$, $\tilde{W}_t^r$ and $\tilde{W}_t^\mu$ are standard Brownian motions under $\mathbb{Q}$ with the same correlation coefficients. Here, for simplicity and without loss of generality, we assume that there is no risk premium for mortality risk. Since the force of mortality is stochastic, the $k$-year survival probability ${}_kp_{x+t}$ is given by:

$${}_kp_{x+t} = \mathbb{E}^{\mathbb{Q}} \left[ e^{-\int_0^k \mu_{x+t+s} \, ds} | Y_t \right],$$

and the at time-$t$-value of the VA plus GMIB contract is:

$$V(t) = \mathbb{E}^{\mathbb{Q}} \left[ e^{-\int_t^T r_s \, ds} e^{-\int_0^{T-t} \mu_{x+t+y} \, dy} \max\left\{ e^{q_T}, \; b \, a_{x+T}(T) \right\} | Y_t \right]. \tag{1.25}$$

Since it is not possible to obtain an analytical expression for the GMIB, particularly when considering additional features such as step ups or ratchets, it is necessary to rely on numerical methods for valuation and estimating risk capital. To directly apply our LSM framework, we adjust the presentation by changing the numéraire to a pure endowment with maturity $T$ and maturity value one. The price of GMIB at time $k$ using the pure endowment as the numéraire is:

$$V(t) = {}_{T-t}E_{x+t} \, \mathbb{E}^{\mathbb{Q}_E} \left[ \max\left\{ e^{q_T}, \; b \, a_{x+T}(T) \right\} | Y_t \right], \tag{1.26}$$

where $\tau \leq t \leq T$, ${}_{T-t}E_{x+t}$ is the price of the pure endowment contract at time $t$, and

$\mathbb{Q}_E$ is the risk-neutral measure using the pure endowment contract as the numéraire.

Under our assumption of independence between financial and biometric risk, we obtain:

$$\begin{aligned} {}_{T-t}E_{x+t} &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_s\,ds}e^{-\int_0^{T-t}\mu_{x+t+y}\,dy}|Y_t\right] = p(t,T)\ \times\ {}_{T-t}p_{x+t} \\ &= A_r(t,T)\exp\left(-r_t\,B_r(t,T)\right)\ A_\mu(t,T)\exp\left(-\mu_{x+t}\,B_\mu(t,T)\right) \end{aligned}$$

since $(r_t)$ and $(\mu_t)$ are affine with

$$\begin{aligned} B_r(t,T) &= \frac{1-e^{-\alpha(T-t)}}{\alpha}, \quad A_r(t,T) = \exp\left\{\left(\bar{\gamma}-\frac{\sigma_r^2}{2\alpha^2}\right)\left(B_r(t,T)-T+t\right)-\frac{\sigma_r^2}{4\alpha}B_r^2(t,T)\right\}, \\ B_\mu(t,T) &= \frac{1-e^{\kappa(T-t)}}{\kappa}, \quad A_\mu(t,T) = \exp\left\{\frac{\psi^2}{2\kappa^2}\left(B_m u(t,T)+T-t\right)+\frac{\psi^2}{4\kappa}B_\mu(t,T)^2\right\}. \end{aligned}$$

Thus, applying Itô's formula, the dynamics of the pure endowment price are:

$$\begin{aligned} d_{T-t}E_{x+t} &= dp\times {}_{T-t}p_{x+t} + p\times d({}_{T-t}p_{x+t}) \\ &= {}_{T-t}E_{x+t}\left[(r_t+\mu_{x+t})dt-\sigma_r B_r(t,T)d\tilde{W}_t^t-\psi B_\mu(t,T)d\tilde{W}_t^\mu\right], \end{aligned}$$

and from Brigo and Mercurio (2006), the new dynamics of $Y_t$ under $\mathbb{Q}_E$ for $\tau \le t \le T$ become:

$$dq_t = \left(r_t-\frac{1}{2}\sigma_S^2-\rho_{12}\,\sigma_s\,\sigma_r B_r(t,T)\right)\,dt+\sigma_S\,dZ_t^S, \tag{1.27}$$

$$dr_t = \alpha(\bar{\gamma}-\sigma_r^2 B_r(t,T)/\alpha-r_t)\,dt+\sigma_r\,dZ_t^r, \tag{1.28}$$

$$d\mu_{x+t} = \kappa(\mu_{x+t}-\psi^2 B_\mu(t,T)/\kappa)\,dt+\psi\,dZ_t^\mu, \tag{1.29}$$

where $Z_t^S$, $Z_t^r$, and $Z_t^\mu$ are standard Brownian motions under $\mathbb{Q}_E$ with $dZ_t^S dZ_t^r = \rho_{12}dt$, $dZ_t^S dZ_t^\mu = 0$, and $dZ_t^r dZ_t^\mu = 0$.

Again proceeding similarly to the previous section, we ignore the asset side in the calculation of the risk capital for the VA plus GMIB contract, and estimate the risk

measure $\rho(V(\tau, Y_\tau))$ via the LSM algorithm. In particular, the cash flow functional in the current setting is $\mathbf{x} = x_T$ with

$$x_T(Y_T) = -V(T) = -\max\{e^{q_T}, ba_{x+T}(T)\}$$

and

$$C_\tau = L\,\mathbf{x}(Y_\tau) = {}_{T-t}E_{x+\tau}\,\mathbb{E}^{\mathbb{Q}_E}\left[x_T(Y_T)|Y_\tau\right].$$

To apply our results on optimal basis functions, we require the joint distribution of $Y_\tau$ and $Y_T$:

**Lemma 1.5.2.** *From (1.22)–(1.24) and (1.27)–(1.29), the joint (unconditional) distribution of $Y_\tau$ and $Y_T$ under $\tilde{\mathbb{P}}$ is:*

$$\begin{bmatrix} Y_\tau \\ Y_T \end{bmatrix} \sim N\left[\begin{bmatrix} \mu_\tau \\ \mu_T \end{bmatrix}, \quad \begin{bmatrix} \Sigma_\tau & \Gamma \\ \Gamma' & \Sigma_T \end{bmatrix}\right],$$

*where we refer to the proof in the Appendix for explicit expressions of $\mu_\tau$, $\mu_T$, $\Sigma_\tau$ etc. in terms of the parameters.*

Thus we can apply the results from Proposition 1.4.3 to derive optimal basis functions. More precisely, for any non negative integer vector $l = (l_1, l_2, l_3)$, $\omega_{|l|} = \lambda_1^{l_1}\lambda_2^{l_2}\lambda_3^{l_3}$ is the squared singular value of $L$ and the corresponding left singular functions is:

$$\varphi_{|l|}(x) = h_{l_1}(z_1^P(x))h_{l_2}(z_2^P(x))h_{l_3}(z_3^P(x)).$$

Thus, in order to find the set of optimal basis functions for the LSM algorithm consisting of $M = K + 1$ functions, we need to calculate $\omega_{|m|}$ for $m = (m_1, m_2, m_3)$ such that $|m| \leq M$, order them, and then determine the associated functions.

**Numerical Results**

As in the previous application, we set the model parameters using representative values. The initial price of the risky asset is one hundred – so $q_0 = 4.605$ – and for the risky asset parameters we assume $m = 0.05$ (instantaneous rate of return) and $\sigma_S = 20\%$ (asset volatility). The initial interest rate is assumed to be $r_0 = 2\%$, $\alpha = 20\%$ (speed of mean reversion), $\gamma = 2.5\%$ (mean reversion level), $\sigma_r = 1\%$ (interest rate volatility), $\lambda = 2\%$ (market price of risk), and $\rho_{12} = -30\%$ (correlation between asset and interest rate). For the mortality rate, $x = 55$ (age of the policyholder), $\mu_{55} = 1\%$ (initial value of mortality), $\kappa = 10\%$ (instantaneous rate of increment), and $\psi = 0.03\%$ (mortality volatility) are assumed. For the insurance contract, we let the maturity $T = 15$, and the guaranteed annuity payout $b = 30$ per year. We set the risk horizon $\tau = 1$ as in the previous application.

With the above parameters, the eigenvalues of $A$ are $\lambda_1 = 0.1908$, $\lambda_2 = 0.0669$, and $\lambda_3 = 0.0012$. The first singular value is one and its corresponding left singular function is $\varphi_1(x) = 1$. The second singular value of the valuation operator is $\sqrt{\lambda_1}$ and the corresponding left singular function is $\varphi_2(x) = z_1^P(x)$. The next three singular values are given by $\sqrt{\lambda_2}$, $\lambda_1$, and $\sqrt{\lambda_1\lambda_2}$, and corresponding left singular functions are $\varphi_3(x) = z_2^P(x)$, $\varphi_4(x) = \frac{1}{\sqrt{2}}\left(\left(z_1^P(x)\right)^2 - 1\right)$, and $\varphi_5(x) = z_1^P(x)z_2^P(x)$. In contrast, a naïve choice of five monomials may result in the sequence $(1, q_\tau, r_\tau, \mu_{x+\tau}, q_\tau^2)$ or another arbitrary arrangement.

We implement the LSM approximation to the capital variable as we vary the number of basis functions. In Figure 1.5, we provide empirical densities based on $N = 200,000$ and approximate realizations calculated via the LSM algorithm for different numbers of basis functions $M$. Here we rely on the optimal basis functions from Proposition 1.4.3 (Hermite polynomials). As is evident from the figure, the required number of basis function is relatively large comparing to the univariate case in the previous section.[15] The approximation becomes closer to the exact density as $M$

[15]Since it is impossible to obtain the exact loss distribution at the risk horizon, we consider the estimated

increases.

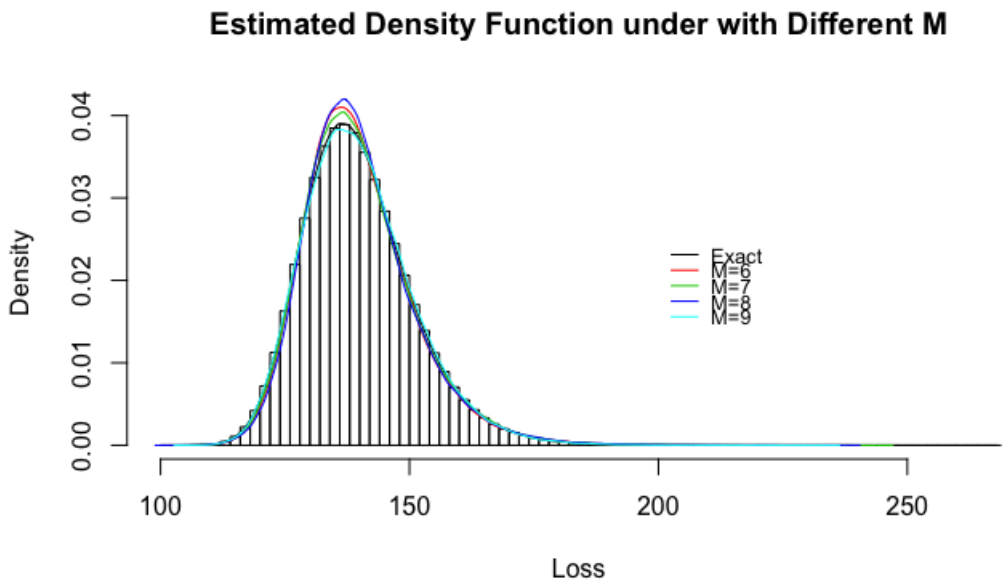



Figure 1.5: Empirical densities of $V(\tau)$ based on $N = 200,000$ Monte Carlo realizations; exact and using the LSM algorithm with $M$ singular functions in the approximation.

To assess the performance of optimal basis functions relative to naïve choices, in Table 1.2 we report statistical differences to the exact distribution according to various statistical distance measures for singular functions (left column) and simple monomials (right column).[16] We find that the optimal basis functions perform uniformly better than the simple polynomials. Furthermore, the table demonstrates that in this higher-dimensional setting, the functional approximation is more relevant than in the univariate setting in the previous section. More precisely, here we observe improvements in the statistical measures when using more basis functions even when keeping the number of simulations constant.

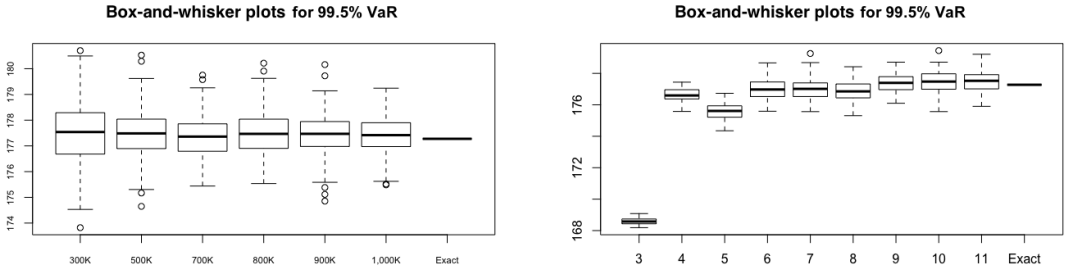
Moving to the calculation of the company's capital requirement, Figure 1.6 plots estimates for the VaR at 99.5% (a) using a fixed number of (optimal) basis functions

loss distribution obtained from the LSM algorithm with $M = 34$ monomials and $N = 25 \times 10^6$ simulations as the exact loss distribution to assess the performance of the LSM algorithm in this multivariate setting.

[16]Here, the set of monomial basis functions when $M = 5$ in Table 1.2 is $(1, q_\tau, r_\tau, \mu_{x+\tau}, q_\tau^2)$.

| Order | | Singular Functions | Simple Polynomials |
|---|---|---|---|
| | KS | $1.560 \times 10^{-2}$ | $3.700 \times 10^{-2}$ |
| $M = 4$ | KL | $8.227 \times 10^{-6}$ | $5.249 \times 10^{-5}$ |
| | JS | $1.428 \times 10^{-3}$ | $3.643 \times 10^{-3}$ |
| | KS | $7.693 \times 10^{-3}$ | $1.217 \times 10^{-2}$ |
| $M = 5$ | KL | $3.119 \times 10^{-6}$ | $3.945 \times 10^{-6}$ |
| | JS | $8.754 \times 10^{-4}$ | $9.885 \times 10^{-4}$ |
| | KS | $3.372 \times 10^{-3}$ | $5.754 \times 10^{-3}$ |
| $M = 10$ | KL | $5.139 \times 10^{-7}$ | $1.445 \times 10^{-6}$ |
| | JS | $3.325 \times 10^{-4}$ | $5.983 \times 10^{-4}$ |

Table 1.2: Statistical Distances between the empirical density function based on the exact realizations and the LSM approximation using different basis functions; mean of two-hundred realizations of $N = 800,000$.




(a) $M = 11$

(b) $N = 1,500,000$.

Figure 1.6: Box-and-whisker diagrams for 99.5% VaR calculated using the LSM algorithm with different number of simulations $N$ and a fixed number of basis functions (a); and with different number of basis functions $M$ and a fixed number of simulations (b).

and varying the number of simulations, and (b) using a fixed number of simulations and varying the number of basis functions (box plots based on 150 runs). Similarly to the previous section, Figure 1.6a displays that the dispersion of the distribution of VaR is decreasing as $N$ increases. However, for $N = 1,000,000$, the bulk of the estimates are located between 178.65 and 176.13, which cover 95% of VaR estimates and safely contains the correct estimate – illustrating the viability of the approach. We observe a slight downward trend in the mean of the $\text{VaR}_{99.5\%}$ in line with the positive bias from Proposition 1.3.4. In Figure 1.6b, we plot the distribution of 99.5% VaR for $N = 1,500,000$ and different choices for number of basis functions $M$. We

| | KS | KL | JS | $\text{VaR}_{99.5\%}$ |
|---|---|---|---|---|
| Singular | $1.097 \times 10^{-2}$ | $2.625 \times 10^{-6}$ | $7.990 \times 10^{-4}$ | 176.98 |
| Comb. 1 $(q_\tau^2, r_\tau^2)$ | $1.222 \times 10^{-2}$ | $3.370 \times 10^{-6}$ | $9.130 \times 10^{-4}$ | 177.03 |
| Comb. 2 $(q_\tau^2, \mu_{x+\tau}^2)$ | $1.402 \times 10^{-2}$ | $4.567 \times 10^{-6}$ | $1.064 \times 10^{-3}$ | 177.22 |
| Comb. 3 $(r_\tau^2, \mu_{x+\tau}^2)$ | $3.520 \times 10^{-2}$ | $5.079 \times 10^{-5}$ | $3.584 \times 10^{-3}$ | 169.46 |
| Comb. 4 $(q_\tau r_\tau, q_\tau \mu_{x+\tau})$ | $3.148 \times 10^{-2}$ | $4.123 \times 10^{-5}$ | $3.225 \times 10^{-3}$ | 170.15 |
| Comb. 5 $(q_\tau \mu_{x+\tau}, r_\tau \mu_{x+\tau})$ | $3.643 \times 10^{-2}$ | $5.417 \times 10^{-5}$ | $3.701 \times 10^{-3}$ | 169.26 |
| Comb. 6 $(q_\tau^2, r_\tau \mu_{x+\tau})$ | $1.387 \times 10^{-2}$ | $4.447 \times 10^{-6}$ | $1.049 \times 10^{-3}$ | 177.21 |

Table 1.3: Statistical Distances between the empirical density function based on the exact realizations and the LSM approximation using different combinations, and VaR at 99.5%; mean of two-hundred realizations of $N = 800,000$.

see that a small number of basis functions, e.g. $M = 3$ or $M = 5$, can lead to a severe misestimation. As we increase the number of basis functions, the estimated 99.5% VaRs converges to the exact 99.5% VaR, although the distribution becomes more dispersed. Again, this emphasizes the relevance of the joint behavior as $N$ and $M$ increase.

To analyze the viability of naïve choices, in Table 1.3, we compare the performance of six optimal basis functions to various combinations of six simple polynomial basis functions. In particular, we choose a constant term and first order terms in each variables, and we then consider six choices for the remaining two terms. Again, we observe that the singular functions provide a uniformly better fit than the polynomials. Furthermore, we notice that a poor choice in the basis function (Combinations 3-5) lead to a severe underestimation of the VaR at 99.5%, where it appears that omitting higher-order terms in $q$ is the key issue.

## 1.6 Conclusion

We propose a novel algorithm for estimating risk measures in "nested" settings, which delivers reliable results with a relatively small computational effort. The algorithm relies on functional approximations of conditional expected values and least-squares

regression. After establishing the algorithm, we analyze convergence of the approach and examine properties where estimating VaR. Moreover, we discuss the choice of basis functions in the functional approximation. Specifically, we show that, under certain conditions, the left singular functions of the valuation operator that maps cash flows to capital present *optimal* basis functions for a model framework. We derive optimal basis functions in settings where the underlying Markov state variable follows a Gaussian distribution, and we apply our ideas in two relevant examples from life insurance.

Our numerical illustrations document that the algorithm can provide viable results at relatively low computational costs. The algorithm therefore provides one potential solution to pressing practical problems such as the calculation of capital requirements in life insurance according to the forthcoming Solvency II directive. Two key insights emerge from our analyses in view of applying the LSM algorithm in practical settings. First, increasing the number of basis functions comes at a significant cost since it is necessary to simultaneously increase the number of simulations $N$. This is required to establish convergence in theory, since the number of simulations typically has to increase much faster; and also in our illustrations, the variance of the estimates increased markedly when adding in additional basis terms. Second, in multivariate settings, a key issue is not only choosing the functional class of basis functions – which appears less crucial in our exercises – but rather the combinations of basis functions that are important for spanning the payoff space in view of valuation. Even in the three-dimensional setting considered here, this is of critical importance as naïve choices may yield significantly worse results. We expect that the choice of basis functions will become even more important as the complexity and the dimensionality of the problem increase.

# Chapter 2

# A Least-Squares Monte Carlo Evaluation of Withdrawal Benefits in Variable Annuities[1]

## 2.1 Introduction

Within the life insurance industry, advanced savings products that combine capital market participation, financial guarantees, and insurance features play an increasingly important role. For instance, in the US since 1995, annual Variable Annuity (VA) sales have increased from roughly $30 billion to $150 billion (Morningstar Annuity Research Center[2]), whereas the annual growth rate in the US individual life insurance market was a mere 2% (LIMRA[3]). These products present challenging problems in view of pricing and risk management, particularly due to exercise-dependent features embedded in the contracts that render the resulting option valuation problem "non-European." The conventional approach is to rely on numerical methods that require a discretization of the (Markov) state space, such as finite difference schemes for the

[1] This chapater is co-authored with Daniel Bauer.
[2] Morningstar Annuity Research Center. `http://www.vards.com`.
[3] Life Insuance and Market Research Association. `http://www.limra.com`.

corresponding Black-Scholes partial differential equations (Bauer et al., 2008; Chen et al., 2008; Dai et al., 2008). However, due to the complexity of the guarantees and due to the multitude of relevant risk factors over the typically long horizons – e.g., interest, volatility, and longevity risk – the valuation problems are usually high-dimensional. Therefore, these conventional approaches may not be feasible for real-world applications.

In this context, a number of studies in the actuarial literature have proposed to rely on so-called Least-Square Monte Carlo (LSM) methods that are popular for pricing American options (Carriere, 1996; Longstaff and Schwartz, 2001; Clément et al., 2002). However, thus far the consideration was limited to optimal stopping problems as they arise when considering the option to surrender the contract (Nordahl, 2008; Bauer et al., 2010; Bacinello et al., 2010). The purpose of the current chapter is to explore the feasibility of the LSM approach for other (non-surrender) option features within life insurance contracts that depend on the policyholders exercise, particularly for popular withdrawal guarantees in VAs. Understanding the applicability – and potential pitfalls or limitations – of this method is highly relevant for providers faced with these valuation problems, particularly in light of the dawning Solvency II regulation that emphasizes market-consistent valuation.

Within a VA investment, the insurer invests an initial premium into a separate financial asset account according to the policyholder's choice and deducts a continuous fee – the so-called *option fee* – from the account. The option fee should be determined so as to provide for the benefits chosen by the policyholder. The problem of "pricing" an embedded option in this context then refers to the determination of this option fee. Within a Guaranteed Minimum Withdrawal Benefit (GMWB) rider, the policyholder has the option to make periodical withdrawals from her account, even in case the account value reaches zero. For pricing and managing the risk associated with a GMWB, of course the policyholder's behavior is a crucial factor.

A number of contributions comment on suitable assumptions for policyholder be-

havior. For instance, optimal withdrawals and fees when applying a conventional non-European option pricing approach to a basic VA plus GMWB do not square well with empirical patterns (Moenig and Bauer (2016) and references therein). Potential reasons include taxes and other frictions, market incompleteness, or behavioral biases. In this chapter, we follow the view in Bauer et al. (2015) that in the context of newly introduced guarantees and changing market environments, it is not sufficient to rely on past policyholder behavior but it is necessary to construct structural models. In particular, if one views policyholder behavior as actuarial risk, the conventional approach is to put up a reserve that is sufficient under any exercise pattern – which leads to the non-European option pricing problem (Bauer et al., 2010, 2015). However, also when considering the optimal decision in more advanced models that include taxes or risk aversion, it is typically necessary to solve a dynamic optimization problem so that the considerations here will apply, possibly after some modification.

There are many papers that determine optimal exercise behavior in insurance contracts according to American or Bermudan option pricing techniques (Steffensen, 2002; Tanskanen and Lukkarinen, 2003; Bauer et al., 2008, among many others). Here, the common approach is to rely on a discretization of the state space within the so-called grid algorithm (Judd, 1998). However, one of key problems when using this numerical method to approximate the value function is the so-called *curse of dimensionality*. That is, it is computationally demanding – and potentially prohibitive – to use this algorithm if the problem is high-dimensional. As indicated, this is typically the case for VA guarantees and for GMWBs in particular, especially within a practical model framework that includes relevant risk factors such as equity, interest, volatility, and longevity risks.

Recognizing this point, a number of studies have proposed to rely on an LSM approach, although the focus in actuarial studies has been on optimal stopping problems in the context of policy surrender. One exception is the recent paper by Huang and Kwok (2016), who consider a regression-based algorithm for pricing and heed-

ing related withdrawal guarantees, so-called Guaranteed Lifelong Withdrawal Benefits (GLWBs), in VAs. Within GLWBs, the withdrawal option does not elapse until the policyholder's time of death, whereas GMWBs feature a fixed maturity date and a guarantee account that gets depleted upon withdrawal. In particular, this means that within our setting, the guarantee account evolves as a state variable – rendering the problem different than usual LSM pricing. Moreover, in contrast to their paper, we consider variants of the LSM approach and discuss advantages and pitfalls of the various approaches. Further, we provide analyses in an advanced financial model with several risk factors.

After introducing the model framework and the GMWB pricing problem, we describe the conventional regression-based algorithm – which we refer to as *regression-now algorithm*. We formally establish the algorithm for pricing the GMWB. We first implement it in a basic Black-Scholes setting to analyze its performance and properties. In particular, we also implement a grid-based algorithm, which is feasible in this setting, and compare the results. This allows us to analyze the viability of the LSM algorithm in solving the dynamic optimization problem.

A number of insights emerge. First, for a relatively low number of simulation paths, the withdrawal rule from the LSM algorithm does not present a close approximation to the optimal withdrawal rule from the grid-based algorithm. While this "suboptimal" behavior should yield a *lower* value for the embedded option, we instead find that the LSM algorithm produces a *higher* price. This is due to a well-known bias in regression-based algorithms, since the algorithm emphasizes results that are high solely due to Monte Carlo variation (Glasserman, 2004). As the number of simulation paths increases, however, we observe that (i) the withdrawal pattern under the LSM algorithm becomes closer to that from the grid-based approach; and (ii) the positive bias decreases – so that for large choices, the LSM algorithm produces a viable approximation.

We then extend the model framework to include multiple stochastic factors:

Stochastic volatility, stochastic interest rates, and stochastic mortality. Here, the LSM algorithm allows us to solve the dynamic optimization problem associated with the GMWB valuation, whereas a conventional grid-based algorithm would require an at least five-dimensional grid – which complicates or even prohibits its application. We find that the option fee considerably increases when incorporating the various risk factors. Furthermore, optimal withdrawal patterns become relatively "rich" as they depend on the various risk factors in a non-trivial fashion. In view of the algorithm, more simulation paths and a high number of basis functions are required for the option fee to converge relative to the simple Black-Scholes setting – but these requirements are not prohibitive with reasonable computational resources.

Finally, we also analyze a different variant of the LSM algorithm, the so-called *regression-later* algorithm. This version relies on a functional approximation of the payoff profile with suitable basis functions that permit a closed-form evaluation of their expectation – rather than approximating the conditional expectations as within the conventional *regression-now* LSM variant. Our findings are mixed. In the basic Black-Scholes setting, it proves relatively straightforward to approximate the payoff profile based on few basis functions, and the regression-later algorithm then performs better (see Beutner et al. (2013) for similar findings). However, as we move to the multi-factor setting, (i) one faces the nontrivial problem of obtaining suitable basis functions for approximating the payoff that allow for closed-form solutions of their expected values; and (ii) the implementation becomes increasingly difficult and it is subject to other error sources, including the solution of equation systems associated with finding the moments of the basis functions. Due to these complications, we are not able to produce viable results.

The remainder of the chapter is organized as follows: Section 2.2 introduces model setting for GMWB and the related dynamic optimization problem. Section 2.3 presents the basic regression-now algorithm in the context of GMWB pricing. Section 2.4 provides our numerical results in the two model settings. Section 2.5 discusses the

regression-later variant and gives corresponding numerical illustrations. And, finally, Section 2.6 concludes.

## 2.2 Model Setting

This section sets up the general model framework for pricing the VA plus GMWB. In Section 2.2.1, we formalize the underlying assumptions and specify the quantities related to the GMWB. Section 2.2.2 then introduces the dynamic optimization problem to find the optimal withdrawal amount at each stage.

### 2.2.1 Framework

We assume that $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ is a complete filtered probability space on which all relevant quantities exist, where $\mathbb{P}$ is the physical measure. The sigma algebra $\mathcal{F}_t$ contains all information about the market and mortality up to time $t$, and the filtration $\mathbf{F}$ is assumed to satisfy the usual conditions. In addition, we take for granted the existence of a risk-neutral probability measure (equivalent martingale measure) $\mathbb{Q}$ equivalent to $\mathbb{P}$.

The uncertainty with respect to pricing and management of the VA plus GMWB arises from the uncertain development of a number of influencing factors, such as equity returns, interest rates, demographic indices, etc. We introduce the sufficiently regular $d$-dimensional Markov process $Y = (Y_t)_{t\in[0,T]} = (Y_{t,1}, ..., Y_{t,d})_{t\in[0,T]}$, $d \in \mathbb{N}$, to model the uncertainty. We assume that all market and mortality risk factors can be expressed in terms of $Y$. Moreover, we suppose the existence of a locally risk-free process $(B_t)_{t\in[0,T]}$ with $B_t = \exp\{\int_0^t r_u du\}$ where $(r_t)_{t\in\mathbb{R}_+}$ with $r_t = r(t, Y_t)$ is the risk-free interest rate process. In particular, for the time $t$ price of a zero-coupon bond with maturity $k \geq t$, we have $p(t,k) = p(t,k,Y_t) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^k r_s\, ds}\middle| Y_t\right]$. For considering mortality risk, we assume the existence of stochastic intensity $(\mu_{x+t})_{t\in\mathbb{R}_+}$, the force of mortality at age $x+t$. In particular, the probability that a policyholder whose age is

$x+t$ at time $t$ survives up to time $k > t$ is given by ${}_{k-t}p_{x+t} = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^{k-t}\mu_{x+t+s}ds}\middle| Y_t\right]$. The corresponding mortality probability is denoted by ${}_{k-t}q_{x+t} = 1 - {}_{k-t}p_{x+t}$.

Suppose that the policyholder at age $x$ enters into a VA plus GMWB contract. She pays an initial lump sum premium, $P_0$, which is invested into the reference asset whose price is given by $S_t = S_t(Y_t)$ at time $t$. A *personal account* is kept for the policyholder, and she has the right to make periodic withdrawals. More precisely, we assume that withdrawals are possible at $t_i$, $i = 1, 2, ..., n-1$, with:

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T,$$

where $T$ denotes the maturity of the contract.

Within a GMWB rider, the insurer maintains a (virtual) *guarantee account* which quantifies the remaining *guaranteed* total amount of withdrawals during the contract period, regardless of the performance of the reference asset. In particular, if the value of personal account is depleted due to poor market performance, the policyholder can still make withdrawals as long as the guaranteed account balance is positive.

We set up the law of motion for the personal account and guaranteed account of the GMWB following notations and the structure in Bauer et al. (2008) and Moenig and Bauer (2016). Let $X_{t_i}^-$ and $w_{t_i}$ denote the value of the personal account at time $t_i$ before making withdrawal and the amount of withdrawal at time $t_i$, respectively. After the withdrawal, the value of personal account is denoted by $X_{t_i}^+$. During the lifetime of the contract, the insurer deducts an (option) fee, $\phi$, continuously from the personal account to cover the cost for providing option features embedded in the contract. Then, the resulting law of motion of the personal account is given by:

$$\begin{aligned} X_{t_i}^- &= X_{t_{i-1}}^+ \frac{S_{t_i}}{S_{t_{i-1}}} e^{-\phi(t_i - t_{i-1})}, && i = 1, 2, ..., n, \\ X_{t_i}^+ &= \max\left(0, X_{t_i}^- - w_{t_i}\right), && i = 1, 2, ..., n-1, \end{aligned}$$

with $X_{t_0}^{+} = P_0$.

The possible withdrawal amount at time $t_i$ is:

$$0 \leq w_{t_i} \leq \max\left(X_{t_i}^{-}, \min\left(g_{t_i}, G_{t_i}\right)\right), \tag{2.1}$$

where $G_{t_i}$ is the guarantee account value at time $t_i$ and $g_{t_i}$ is the guaranteed contractual withdrawal amount. The constraint in (2.1) is a common form of the possible withdrawal amount and it differs slightly from some of the settings put forward in literature (Bacinello et al., 2013; Huang and Kwok, 2016). Note that we allow the policyholder to surrender the contract before the maturity by permitting $w_{t_i} = X_{t_i}^{-}$ if $X_{t_i}^{-} > G_{t_i}$ in (2.1). The guarantee account is updated based on the withdrawal amount according to the following transition equation:

$$G_{t_{i+1}} = \begin{cases} \max\left(0, G_{t_i} - w_{t_i}\right), & w_{t_i} \leq g_{t_i}. \\ \min\left(\max\left(0, G_{t_i} - w_{t_i}\right), \frac{X_{t_i}^{+}}{X_{t_i}^{-}} G_{t_i}\right), & w_{t_i} > g_{t_i}, \quad i = 1, 2, ..., n-1, \end{cases}$$

with $G_{t_1} = P_0$. It is also possible to include common ratchet or roll-up features of the guarantee account.

While the amount and timing of withdrawal are determined by the policyholder, it is common to impose penalties in the form of fees if the withdrawal amount exceeds some threshold or if the withdrawal occurs too early. Therefore, the cash amount going to the policyholder at time $t_i$, $C(t_i, w_{t_i})$, may be different from $w_{t_i}$. In line with practical contract designs, we assume that:

$$\begin{aligned} C(t_i, w_{t_i}) &= w_{t_i} - \text{fee}_{t_i}^{\text{I}} - \text{fee}_{t_i}^{\text{R}}, \\ \text{fee}_{t_i}^{\text{I}} &= ep_{t_i} \times \max(0, w_{t_i} - \min(g_{t_i}, G_{t_i})), \qquad (2.2) \\ \text{fee}_{t_i}^{\text{R}} &= pg_{t_i} \times (w_{t_i} - \text{fee}_{t_i}^{\text{I}}) \mathbb{1}_{\{x+t_i<59.5\}}, \qquad (2.3) \end{aligned}$$

where $\mathbb{1}_A$ is the indicator function; $ep_{t_i}$ and $pg_{t_i}$ are penalty percentages. In (2.2),

the fee is applied to the withdrawal and absorbed by the insurer if the withdrawal is greater than the threshold, $\min(g_{t_i}, G_{t_i})$, whereas the fee in (2.3) is collected by the regulator.

We define the death benefit, $D_{t_i}$, paid at $t_i$, $i = 1, 2, ..., n$, if death of the policyholder occurs during $(t_{i-1}, t_i]$. $D_{t_i}$ may be the account value $X_{t_i}^-$ or a guaranteed amount. Since most VA contracts also contain a Guaranteed Minimum Death Benefit (GMDB) rider free of charge, the death benefit in this chapter is assumed to be:

$$D_{t_i} = \max(X_{t_i}^-, G_{t_i}).$$

If the policyholder survives until maturity, on the other hand, the policyholder will receive a survival benefit, $V(t_n)$, which is given by

$$V(t_n) = \max\left(X_{t_n}^-,\ \min(g_{t_n}, G_{t_n})\right).$$

For the valuation, in this chapter, we assume that the policyholder determines the amount of withdrawal and the optimal surrender time, $\tau$, so as to maximize value of her contract. Suppose that $\mathcal{W} = (w_{t_1}, ..., w_{t_{n-1}})$ is an arbitrary withdrawal strategy and $\mathcal{A}$ denotes family of all conceivable (adapted) $\mathcal{W}$, and $\tau$ is a stopping time such that $w_{t_i} = 0$ for $t_i > \tau$ if $\tau \leq t_{n-1}$. Then the price/value $V(0)$ of the VA plus GMWB at $t = t_0$ is given by

$$V(0) = \sup_{\mathcal{W}\in\mathcal{A}} \mathbb{E}^{\mathbb{Q}}\Bigg[ \underbrace{\sum_{i=1}^{n-1\wedge\tau} e^{-\int_0^{t_i} r_s ds} e^{-\int_0^{t_i} \mu_x(s) ds} C(t_i, w_{t_i}) + e^{-\int_0^{T} r_s ds} e^{-\int_0^{T} \mu_x(s) ds} V(t_n) \mathbb{1}_{\{T\geq\tau\}}}_{\text{Survival Benefit}}$$
$$+ \underbrace{\sum_{j=1}^{n\wedge\tau} e^{-\int_0^{t_j} r_t dt} e^{-\int_0^{t_{j-1}} \mu_x(t) dt} \left(1 - e^{-\int_0^{t_j - t_{j-1}} \mu_{x+t_{j-1}}(t) dt}\right) D_{t_j}}_{\text{Death Benefit}} \Bigg| Y_0 \Bigg]. \quad (2.4)$$

To find $V(0)$ in (2.4), we need to determine $\mathcal{W}^*$ that maximizes the contract value.

Usually, it is not possible to determine $V(0)$ by finding $\mathcal{W}^*$ analytically. In the following section, we thus introduce a *dynamic optimization problem* to find the optimal withdrawal strategy.

From the insurer's point of view, the *fair* option fee, $\phi$, is then determined by setting $V(0) = P_0$ according to the equivalence principle.

### 2.2.2 Dynamic Optimization Problem

Finding the optimal withdrawal strategy in (2.4) is equivalent to solving the following *dynamic optimization problem* (2.5) recursively (Judd, 1998). At time $t_i, i = n-1, n-2, ..., 1$, the policyholder's problem reads:

$$
\begin{aligned}
&V_{t_i}(Y_{t_i}) = \max_{w_{t_i}} \quad C(t_i, w_{t_i}) + \\
&\mathbb{E}^{\mathbb{Q}}\left[ e^{-\int_{t_i}^{t_{i+1}} r_s ds} \left\{ e^{-\int_0^{t_{i+1}-t_i} \mu_{x+t_i}(s) ds} V_{t_{i+1}}(Y_{t_{i+1}}) + \left(1 - e^{-\int_0^{t_{i+1}-t_i} \mu_{x+t_i}(s) ds}\right) D_{t_{i+1}}(Y_{t_{i+1}}) \right\} \middle| Y_{t_i} \right], \\
&\text{subject to} \quad 0 \leq w_{t_i} \leq \max(X_{t_i}^-, \min(g_{t_i}, G_{t_i})), \\
&\qquad\qquad V_{t_n}(Y_{t_n}) = V_{t_n}(X_{t_n}^-, G_{t_n}) = \max(X_{t_n}^-, G_{t_n}), \\
&\qquad\qquad D_{t_{i+1}}(Y_{t_{i+1}}) = \max(X_{t_{i+1}}^-, G_{t_{i+1}}). \qquad (2.5)
\end{aligned}
$$

The challenging part in solving (2.5) is the calculation of the expected actuarially discounted benefits. Usually, it is not possible to obtain the expectation in closed form due to unknown form of $V_t(\cdot)$. Instead, it is common to rely on numerical techniques. The grid algorithm, which discretizes the underlying (Markov) state space, is the most common tool to get an approximation of the expectation. However, this method is effective and feasible only when the dimension of problem is low. For instance, the grid algorithm is easily implemented when the dimension of state variables is two (Bacinello et al., 2013). In case one wants to include multiple risk factors such as equity, interest rate and mortality risk, however, the previous method is subject to the so-called *curse of dimensionality*. As a potential solution, we adapt two *Least-Squares*

*Monte Carlo* (LSM) methods in the following sections – the so-called regression-now and regression-later algorithms.

## 2.3 Regression-Now Algorithm

The expectation operator in (2.5) may be regarded as the mapping:

$$\mathbb{E}^{\mathbb{Q}} : L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_{i+1}}\right) \to L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_i}\right), \tag{2.6}$$

where $\mathbb{Q}_t$ is the risk-neutral measure defined at time $t$ (according to the distribution of $Y_t$). Note that the considered two spaces in (2.6) are separable as $L^2$ spaces under a regular Borel measure. Thus, there exists a complete set of basis functions for the each space (Kreyszig, 1989). Accordingly, a conditional expectation in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_i})$ and a functional of random variables in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_{i+1}})$ can be approximated by linear combinations of the respective sets of complete basis functions.

Hence, there are two options for approximating the conditional expectation in (2.5): One is to approximate the conditional expectation using the basis for $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_i})$; the other is to approximate a functional of random variables using the basis for $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_{i+1}})$ and to then compute the expectation of the approximating functional. In this section, we discuss the first option to obtain the approximated value of the expected actuarially discounted benefits. The later option is explored in Section 2.5.

We label this method the *regression-now* algorithm. More precisely, the conditional expectation is approximated by a linear combination of basis functions for $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_i})$ *now* at time $t_i$. The corresponding coefficients are estimated via ordinary least-squares using information generated from a Monte Carlo simulation. As a Monte Carlo method, the regression-now algorithm does not suffer from the curse of dimensionality. The key ingredients are the number of simulations, the number of basis functions, and the form of basis functions.

### 2.3.1 Regression-now algorithm

Suppose that the policyholder is alive at time $t_{n-1}$. The policyholder's dynamic optimization problem then is:

$$V_{t_{n-1}}(Y_{t_{n-1}}) = \max_{w_{t_{n-1}}} \quad C(t_{n-1}, w_{t_{n-1}}) +$$
$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_{t_{n-1}}^{t_n} r_s ds}\left\{e^{-\int_0^{t_n-t_{n-1}} \mu_{x+t_{n-1}}(s)ds} V_{t_n}(Y_{t_n}) + \left(1 - e^{-\int_0^{t_n-t_{n-1}} \mu_{x+t_{n-1}}(s)ds}\right) D_{t_n}(Y_{t_n})\right\} \middle| Y_{t_{n-1}}\right],$$
$$\text{subject to} \quad 0 \le w_{t_{n-1}} \le \max(X^-_{t_{n-1}}, \min(g_{t_{n-1}}, G_{t_{n-1}})),$$
$$V_{t_n}(Y_{t_n}) = V_{t_n}(X^-_{t_n}, G_{t_n}) = \max(X^-_{t_n}, \min(g_{t_n}, G_{t_n})),$$
$$D_{t_n}(Y_{t_n}) = \max(X^-_{t_n}, G_{t_n}). \tag{2.7}$$

Suppose that $\{e_k\}_{k=1}^{\infty}$ is a set of complete basis functions for $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_{n-1}}\right)$. Then the expectation in (2.7) may be written as:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_{t_{n-1}}^{t_n} r_s ds}\left\{e^{-\int_0^{t_n-t_{n-1}} \mu_{x+t_{n-1}}(s)ds} V_{t_n}(Y_{t_n}) + \left(1 - e^{-\int_0^{t_n-t_{n-1}} \mu_{x+t_{n-1}}(s)ds}\right) D_{t_n}(Y_{t_n})\right\} \middle| Y_{t_{n-1}}\right]$$
$$= \sum_{k=1}^{\infty} \alpha_k^{t_{n-1}} e_k\left(Y_{t_{n-1}}\right), \tag{2.8}$$

where $\alpha_k^{t_{n-1}}$ is the corresponding coefficient of the $k$th basis function (Kreyszig, 1989).

The basic idea of the regression-now algorithm to approximate the (conditional) expectation consists of two steps:

1. (2.8) is replaced by finite linear combination of $M$ basis functions.

2. The corresponding coefficients are estimated via least-squares regression.

Common choices of basis functions for the first approximation are monomials, Hermite polynomials, or Legendre polynomials. After choosing the function class and the

number of basis functions, the first approximation is:

$$\mathbb{E}^{\mathbb{P}\times\mathbb{Q}}\left[e^{-\int_{t_{n-1}}^{t_n} r_s ds}\left\{e^{-\int_0^{t_n-t_{n-1}} \mu_{x+t_{n-1}}(s)ds} V_{t_n}(Y_{t_n}) + \left(1 - e^{-\int_0^{t_n-t_{n-1}} \mu_{x+t_{n-1}}(s)ds}\right) D_{t_n}(Y_{t_n})\right\} \middle| Y_{t_{n-1}}\right]$$
$$\approx \sum_{k=1}^{M} \alpha_k^{t_{n-1}} e_k\left(Y_{t_{n-1}}\right). \tag{2.9}$$

The associated coefficients in (2.9) are estimated via least-squares regression using information generated from a Monte Carlo simulation.[4] More precisely, we use $N$ samples of state variables generated at each $t = t_{n-1,j}$, $j = 1, ..., m$, such that:

$$t_{n-1} = t_{n-1,0} < t_{n-1,1} < \cdots < t_{n-1,m} = t_n.$$

At time $t_{n-1}$, using the $l$th sample path, we calculate:

$$y_{t_n}^l = df_{1,t_n}^l \left[df_{2,t_n}^l V_{t_n}(Y_{t_n}^l) + \left(1 - df_{2,t_n}^l\right) D_{t_n}(Y_{t_n}^l)\right], \tag{2.10}$$

where

$$\begin{aligned}
Y_{t_{n-1}}^l &= l\text{th realization of state variables,} \\
df_{1,t_n}^l &= \exp\left(-\sum_{j=0}^{m-1} r_{t_{n-1},j}^l \, \Delta_j^n\right), \\
df_{2,t_n}^l &= \exp\left(-\sum_{j=0}^{m-1} \mu_{x+t_{n-1}}^l(j) \, \Delta_j^n\right), \\
\Delta_j^n &= t_{n-1,j+1} - t_{n-1,j}.
\end{aligned}$$

Note that (2.10) is a (very) noisy estimator of the expectation in (2.7). We thus solve

[4] It may be (theoretically) possible to compute the $M$ coefficients by calculating inner products. However, since the structures of the space and the payoff functional are usually complicated, it is typically not feasible to calculate corresponding coefficients analytically in practice.

the following least-squares problem:

$$\hat{\alpha}^{t_{n-1}} = \text{argmin}_{\{\alpha^{t_{n-1}}\}} \sum_{k=1}^{N} \left[ e\left(Y_{t_{n-1}}^{l}\right) \cdot \alpha^{t_{n-1}} - y_{t_n}^{l} \right]^2,$$

where $e\left(Y_{n-1}^{l}\right) = \left(e_1\left(Y_{n-1}^{l}\right), ..., e_M\left(Y_{n-1}^{l}\right)\right)$, $\alpha^{t_{n-1}} = (\alpha_1^{t_{n-1}}, ..., \alpha_M^{t_{n-1}})'$ and $\cdot$ is the usual scalar product. Then, we replace $\alpha^{t_{n-1}}$ by $\hat{\alpha}^{t_{n-1}}$ to obtain the second approximation:

$$\sum_{k=1}^{M} \alpha_k^{t_{n-1}} e_k(Y_{t_{n-1}}) \approx \sum_{k=1}^{M} \hat{\alpha}_k^{t_{n-1}} e_k(Y_{t_{n-1}}). \tag{2.11}$$

At time $t_{n-1}$, therefore, we have the final approximated expectation after the two steps:

$$\mathbb{E}^{\mathbb{Q}}\left[ e^{-\int_{t_{n-1}}^{t_n} r_s ds} \left\{ e^{-\int_0^{t_n - t_{n-1}} \mu_{x+t_{n-1}}(s) ds} V_{t_n}(Y_{t_n}) + \left(1 - e^{-\int_0^{t_n - t_{n-1}} \mu_{x+t_{n-1}}(s) ds}\right) D_{t_n}(Y_{t_n}) \right\} \middle| Y_{t_{n-1}} \right]$$
$$\approx \sum_{k=1}^{M} \hat{\alpha}_k^{t_{n-1}} e_k\left(Y_{t_{n-1}}\right).$$

The policyholder thus solves the following (approximated) dynamic optimization problem:

$$V_{t_{n-1}}(Y_{t_{n-1}}) \approx \max_{w_{t_{n-1}}} \quad C(t_{n-1}, w_{t_{n-1}}) \; + \sum_{k=1}^{M} \hat{\alpha}_k^{t_{n-1}} e_k\left(Y_{t_{n-1}}\right),$$
$$\text{subject to} \quad w_{t_{n-1}} \in A_{t_{n-1}},$$

where $A_{t_{n-1}}$ is the set of discretized feasible solutions. Note that the continuous set of feasible solutions needs to be discretized as within the grid algorithm to solve the above approximated dynamic optimization problem. The discretized set of feasible solutions considered in this chapter is discussed in the next section. After solving the approximated dynamic optimization problem at time $t_{n-1}$, we follow similar approximation steps and solve the problem at each time $t_i$, $i = n-2, n-3, ..., 1$. More details

about the regression-now algorithm for pricing the GMWB are provided in Appendix A.2.1. Within the above steps, the values of the regressors are specified according to the current information. This is why the algorithm is called regression-*now* method.

It is relatively straightforward to anticipate that the algorithm converges to the true expectation as $M \to \infty$ and $N \to \infty$. If the number of simulation is increased, $\hat{\alpha}^{t_i} \to \alpha^{t_i}$ in probability (Amemiya, 1985). And as $M \to \infty$, the first approximation converges to (2.8) in $L^2$. Details about convergence of (2.11) can be found in Clément et al. (2002).

### 2.3.2 Discretized Feasible Solution

The continuous set of feasible solutions needs to be discretized to a finite set of feasible solutions for the implementation of approximation algorithms. For Guaranteed Minimum Life Benefits (GMLBs), recent contributions show that there are only three possible optimal solution: "do not withdraw," "withdraw the guaranteed contractual amount," and "surrender" (Azimzadeh and Forsyth, 2015; Huang and Kwok, 2016). It is difficult, however, to generalize the set of possible optimal solutions for the GMWB, although it is likely that a consideration of "corners" will suffice.

Here, we consider the following set of possible withdrawals, inspired by the literature on GMWB pricing and the corresponding results for GMLBs. However, we can extend the considered set to other sets that contains more feasible solutions based on the nature of problem, e.g. when incorporating tax features (Moenig and Bauer, 2016).

In our setting, when the policyholder makes decision at time $t_i$, she confronts the following six cases:

- If $X_{t_i}^- \leq g_{t_i} \leq G_{t_i}$, the possible withdrawal amount is $[0, g_{t_i}]$. The set of possible withdrawals is thus assumed to be:

$$A_{t_i} = \{0, g_{t_i}\}.$$

Note that there is no incentive for the policyholder to make a withdrawal of $X_{t_i}^-$.

- $X_{t_i}^- \leq G_{t_i} \leq g_{t_i}$, the possible withdrawal amount is $[0, G_{t_i}]$. The set of possible withdrawals is thus assumed to be:

$$A_{t_i} = \{0, X_{t_i}^-, G_{t_i}\}.$$

- $G_{t_i} \leq X_{t_i}^- \leq g_{t_i}$, the possible withdrawal amount is $[0, X_{t_i}^-]$. The set of possible withdrawals is thus assumed to be:

$$A_{t_i} = \{0, G_{t_i}, X_{t_i}^-\}.$$

- $G_t \leq g \leq X_t^-$, the possible withdrawal amount is $[0, X_{t_i}^-]$. The set of possible withdrawals is thus assumed to be:

$$A_{t_i} = \{0, G_{t_i}, X_{t_i}^-\}.$$

- $g_{t_i} \leq G_{t_i} \leq X_{t_i}^-$, the possible withdrawal amount is $[0, X_{t_i}^-]$. In this case, there are two sub-cases :

  - if $\sum_{s=t_i}^{t_n} g_s < G_{t_i}$

    In this sub-case, the policyholder has the motivation to rebalance her guarantee account value since she may not be able to fully enjoy her minimum guarantee feature if she only withdraws $g_{t_i}$ in the future. Thus, the set of possible withdrawals is assumed to be:

$$A_{t_i} = \left\{0, G_{t_i} - g, X_{t_i}^-\right\},$$

    where $g = G_{t_i} - \sum_{s=t_{i+1}}^{t_n} g_s$.

– if $\sum_{s=t_i}^{t_n} g_s \geq G_{t_i}$. The set of possible withdrawals is assumed to be:

$$A_{t_i} = \{0, g_{t_i}, X_{t_i}^-\}.$$

- $g \leq X_t^- \leq G_t$, the possible withdrawal amount is $[0, X_t^-]$. In this case, there are also two sub-cases :

  – If $\sum_{s=t_i}^{t_n} g_s < G_{t_i}$

  Similar to the previous case, the policyholder has the motivation to rebalance her guarantee account. The set of possible withdrawals is assumed that

  $$A_{t_i} = \left\{0, g, X_{t_i}^-\right\},$$

  where $g = X_{t_i}^- - X_{t_i}^-/G_{t_i}\left(\sum_{s=t_{i+1}}^{t_n} g_s\right)$.

  – If $\sum_{s=t_i}^{t_n} g_s \geq G_t$. The set of possible withdrawals is assumed to be:

  $$A_{t_i} = \{0, g_{t_i}, X_{t_i}^-\}.$$

After specifying $A_{t_i}$, we generate arbitrary withdrawals $w_{t_i}$ at each point in order to obtain variation in the personal account and the guarantee account states simultaneously. Note that, like in Monte Carlo simulations for conventional American-style derivatives, the random withdrawals causing surrender or situation rendering $G_{t_{i+1}} = 0$ should be ruled out.

## 2.4 Application of the Regression-Now Algorithm

We implement the regression-now algorithm to solve the dynamic optimization problem described in Section 2.2.2. First, in Section 2.4.1 we rely on a simple Black-Scholes framework with deterministic mortality. In this setting, the state vector is two dimensional and the grid algorithm is implementable for determining true optimal withdrawal

strategy and for pricing the GMWB. Thus, we are able to compare the results of the regression-now algorithm with the results of the grid method to appraise the viability of the LSM approach.

In Section 2.4.2, we introduce a more advanced model including stochastic interest rate, stochastic volatility, and stochastic mortality. We compare our results to the basic setting and discuss how the optimal withdrawal strategy is affected by state variables.

### 2.4.1 Application in the Black-Scholes framework

Under the Black-Scholes assumption, the price of the reference asset evolves according to the following stochastic differential equation under the measure $\mathbb{Q}$:

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where $r$ is the constant risk free rate, $\sigma$ is the asset volatility, and $W_t$ is a standard Brownian motion under the measure $\mathbb{Q}$. We assume that the force of mortality is a positive constant $\mu$ (exponential law of mortality). Under this assumption, the $k$-year survival probability of an $x$-year old policyholder is given by ${}_kp_x = e^{-\mu k}$. Note that the dimension of the state vector in this framework becomes two – $Y_t = (X_t^-, G_t)$.

The dynamic optimization problem at time $t_i$ $i = n-1, ..., 1$, reads:

$$\begin{aligned} V_{t_i}(Y_{t_i}) = &\max_{w_{t_i}} \; C(t_i, w_{t_i}) \\ &+ e^{-r(t_{i+1}-t_i)}\mathbb{E}^{\mathbb{Q}}\left[e^{-\mu(t_{i+1}-t_i)}V_{t_{i+1}}(Y_{t_{i+1}}) + \left(1 - e^{-\mu(t_{i+1}-t_i)}\right) D_{t_{i+1}}(Y_{t_{i+1}})\middle| Y_{t_i}\right], \\ \text{subject to} \quad & w_{t_i} \in A_{t_i}, \\ & V_{t_n}(Y_{t_n}) = \max(X_{t_n}^-, \min(g, G_{t_n})). \end{aligned} \tag{2.12}$$

We solve (2.12) via the grid method. Since the value of the personal account is log-normally distributed under the Black-Scholes framework, the expectation in (2.12)

is given by:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\mu(t_{i+1}-t_i)}V_{t_{i+1}}(Y_{t_{i+1}}) + \left(1-e^{-\mu(t_{i+1}-t_i)}\right)D_{t_{i+1}}(Y_{t_{i+1}})\middle| Y_{t_i}\right]$$
$$= \int_{-\infty}^{\infty}\left[e^{-\mu(t_{i+1}-t_i)}V_{t_{i+1}}(X_{t_i}^{+}e^{x}, G_{t_{i+1}}) + \left(1-e^{-\mu(t_{i+1}-t_i)}\right)\max(X_{t_i}^{+}e^{x}, G_{t_{i+1}})\right]g(x)dx$$

where

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2(t_{i+1}-t_i)}}e^{-\frac{\left(x-\left(r-\phi-\frac{1}{2}\sigma^2(t_{i+1}-t_i)\right)\right)}{2\sigma^2(t_{i+1}-t_i)}}, \quad -\infty < x < \infty.$$

The integral is evaluated via linear interpolation for $V_{t_{i+1}}$ and $D_{t_{i+1}}$, and the trapezoidal rule after discretizing $Y_t$.

In the LSM algorithm, we choose simple monomials as basis functions for the first approximation. After solving the least-squares problem, the approximated form of the expected value is given by:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{-\mu(t_{i+1}-t_i)}V_{t_{i+1}}(Y_{t_{i+1}}) + \left(1-e^{-\mu(t_{i+1}-t_i)}\right)D_{t_{i+1}}\middle| Y_{t_i}^{-}\right]$$
$$\approx \sum_{j=1}^{M}\hat{\alpha}_j^{t_i}\left(X_{t_i}^{+}\right)^{j_1}\left(G_{t_{i+1}}^{+}\right)^{j_2},$$

where $j_1, j_2 \in \mathbb{Z}_+$.

We use representative values to parameterize the model. The specification of the GMWB contract, financial market parameters, policyholder's age, etc. are provided in Table 2.1. The fair fee of 0.17% is calculated using grid-based algorithm. We first implement the regression-now algorithm using eight basis functions:

$$\mathbb{E}^{\mathbb{Q}}\left[\cdot|Y_t\right] = \alpha_0+\alpha_1 X_t^{+}+\alpha_2\left(X_t^{+}\right)^2+\alpha_3\left(X_t^{+}\right)^3+G_{t_{i+1}}+G_{t_{i+1}}\left(\beta_1 X_t^{+}+\beta_2\left(X_t^{+}\right)^2+\beta_3\left(X_t^{+}\right)^3\right).$$

For the grid-based algorithm, we consider equidistant grids for $X_t^{-}$ and $G_t$ with $201\times16$ gird points.

In Figure 2.1, we plot the expected present value of benefits at time zero $V(0)$

| **GMWB contract** | |
|---|---|
| Maturity | 15 |
| Number of withdrawal per year | 1 |
| Initial Premium ($P_0$) | 15 |
| Option fee ($\phi$) | 0.17% |
| $ep_{t_i}$ | $(8\%, 7\%, 6\%, 5\%, 4\%, 3\%, 2\%, 1\%, 0\%, \cdots, 0\%)$ |
| $pg_{t_i}$ | 10% |
| $g$ | 1 |
| **Policyholder** | |
| Age | 55 |
| $\mu$ | 0.02 |
| **Financial Market** | |
| Risk free rate | 4% |
| Volatility | 15% |

Table 2.1: Description of GMWB contract in the Black-Scholes framework

under the LSM algorithm for different numbers of simulation paths $N$. We compare it to the value under the grid-based algorithm, which is exactly the initial premium $P_0$ since we are relying on the *fair* fee that equates $P_0$ and $V(0)$ (as calculated via the grid-based algorithm). There are two immediate observations. First, the LSM valuation always exceeds the value from the grid algorithm. This is a familiar feature of regression-based algorithms (Glasserman, 2004) originating from an asymmetric influence of Monte Carlo errors: If an estimate is high, it is likely to be picked as the maximum in the Bellman equation, whereas it is likely that another option is favored if the estimate is low. Second, the LSM valuation converges to the grid-based valuation as the number of simulations increases – the bias vanishes and the optimal withdrawal strategy approximates the pattern from the grid-based approach.

To illustrate the latter aspect, Figure 2.2 plots the maximizing withdrawal rules resulting from the LSM algorithm in comparison to the corresponding patterns from the grid-based algorithm for $N = 30,000$ and $N = 6000,000$ simulations for different points in time on the left- and right-hand side, respectively. The grid-based withdrawal rules at $t = 14$ (red line in Panels 2.2a and 2.2b) indicates that it will be optimal to surrender if the value of the guaranteed account value is two and the value of personal

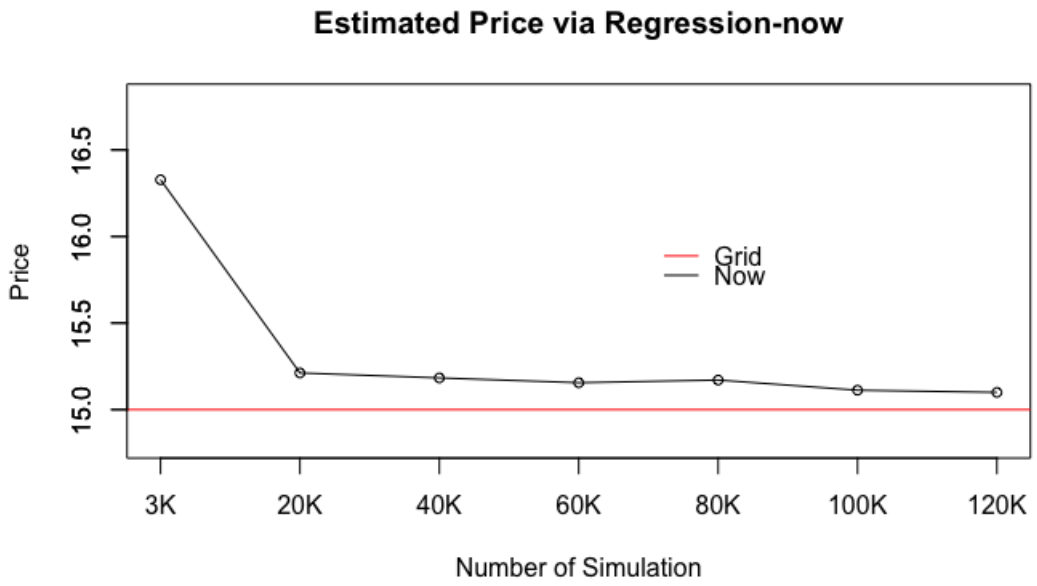



Figure 2.1: Convergence of Regression-Now Estimates: VA plus GMWB value for the grid-based algorithm (Grid) and the LSM estimates (Now) as a function of the number of simulations $N$, basic Black-Scholes model

account is larger than 2.7, whereas it is optimal to withdraw the guaranteed amount for smaller account values. Thus, 2.7 is the "critical value" dividing the region of withdrawing the guaranteed amount $g$ (to the left of it) and surrendering (to the right of it). Clearly, the LSM algorithm does not accurately reflect this pattern for $N = 30,000$ (Panel 2.2a), whereas for $N = 600,000$ (Panel 2.2b) the two graphs are similar. Indeed, the brief dip in the grid-based algorithm is arguably due to discretization errors – and we do not observe it for the withdrawal rule resulting from the LSM algorithm. The match for very large values of $X_t^-$ (beyond 29) is not very close, but note that $(G_t = 2, X_t^- > 29)$ correspond to unlikely combinations of the state variables – since it corresponds to withdrawals of 13 units prior to time $t = 14$ with still more than 29 units in the account value. In fact, a number of combinations illustrated in Figure 2.2 are relatively unlikely or even impossible ("off-equilibrium") given optimal withdrawals in prior years. As such, a poor fit in these regions may be a "feature" rather than a problem of the LSM algorithm. Observations for earlier dates

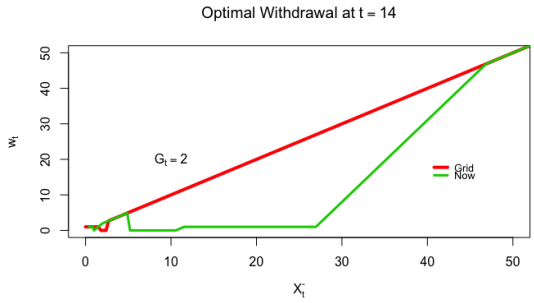
are similar, although there is a considerable region for relatively small account values where the grid and the LSM algorithm for $t = 11$ and $t = 12$. While these do not seem very material in view of the valuation as is evident from Figure 2.1, it is important to keep in mind these potential issues when interpreting withdrawal patterns resulting from the LSM algorithm.

### 2.4.2 Extended Model with Stochastic Volatility, Stochastic Interest Rates, and Stochastic Mortality

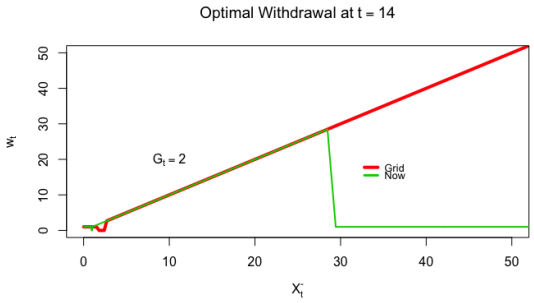
In this section, we consider the valuation of the VA plus GMWB contract in a more practical model including stochastic volatility, stochastic interest rate, and stochastic mortality. More precisely, the dynamics of $Y_t$ are assumed to follow a hybrid Heston-Cox-Ingersoll-Ross model (Heston, 1993; Cox et al., 1985) with the following stochastic differential equations under the risk-nuetral measure $\mathbb{Q}$:

$$\begin{aligned}
dS_t &= r_t S_t dt + \sqrt{\nu_t} S_t dW_t^{(1)} \\
dr_t &= a(b - r_t)dt + \sigma dW_t^{(2)}, \\
d\nu_t &= \kappa(\bar{\nu} - \nu_t)dt + \gamma\sqrt{\nu_t} dW_t^{(3)} \\
d\mu_x(t) &= \psi(\varphi - \mu_x(t))dt + \eta\sqrt{\mu_x(t)} dW_t^{(4)},
\end{aligned}$$

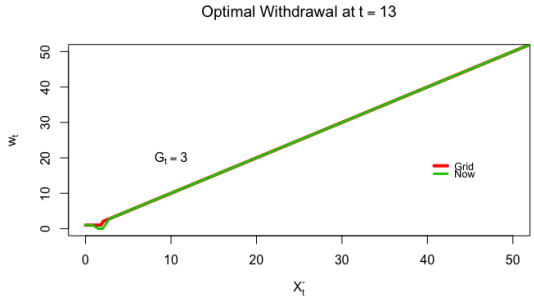
where $b$ is the long-term interest rate, $a$ is the speed of mean reversion for the risk-free interest rate, $\sigma$ is the volatility of the risk-free interest rate, $\nu_t$ is the (stochastic) variance of the risky asset, $\kappa$ is the speed of mean reversion for the variance process, $\bar{\nu}$ is the long-term mean of the variance, $\gamma$ is the volatility of volatility, $\psi$ is the speed of mean reversion for the force of mortality, $\varphi$ is the long term mean of the force of mortality, $\eta$ is the volatility of the force of mortality and $\mathbf{W}_t = (W_t^{(1)}, W_t^{(2)}, W_t^{(3)}, W_t^{(4)})'$ is a four-dimensional Brownian motion under the measure $\mathbb{Q}$ with the correlation matrix $\rho = [\rho_{i,j}]$ such that $dW_t^{(i)} dW_t^{(j)} = \rho_{i,j} dt$, $i, j = 1, 2, 3, 4$. Hence, the relevant state variable at time $t$ here is given by $Y_t = (X_t^-, G_t, r_t, \nu_t, \mu_x(t))$.



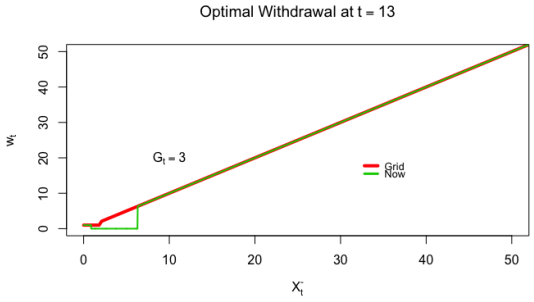

(a) $t = 14,\ N = 30,000$



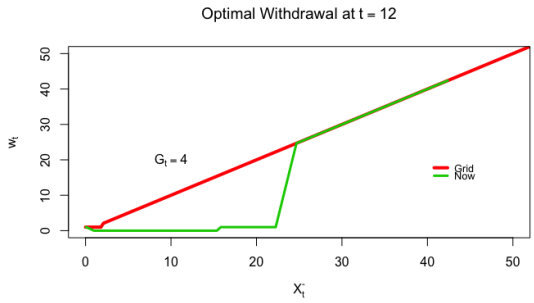

(b) $t = 14,\ N = 600,000$



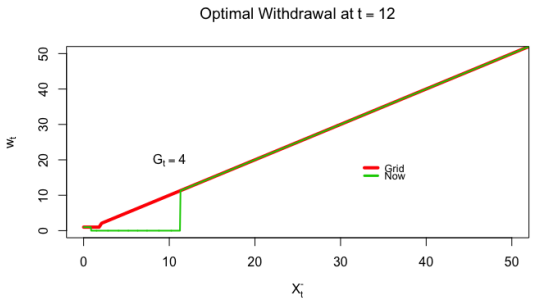

(c) $t = 13,\ N = 30,000$



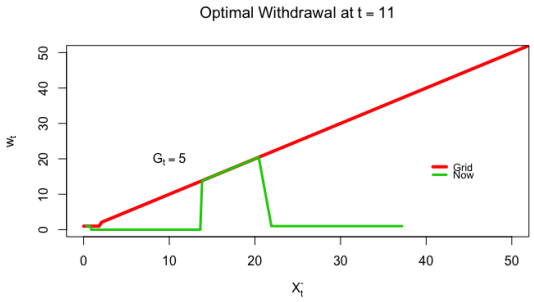

(d) $t = 13,\ N = 600,000$



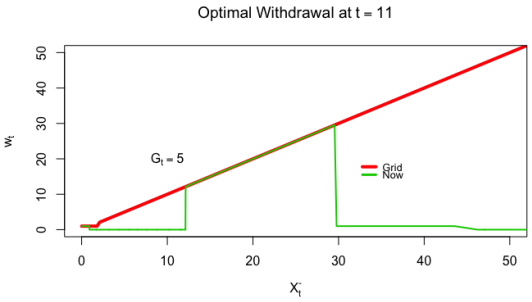

(e) $t = 12,\ N = 30,000$




(f) $t = 12,\ N = 600,000$




(g) $t = 11,\ N = 30,000$




(h) $t = 11,\ N = 600,000$

Figure 2.2: Estimated Optimal Withdrawals for the grid-based algorithm (Grid) and the regression-now algorithm (Now) for $N = 30,000$ (left-hand side) and $N = 600,000$ (right-hand side) at different times $t$

To generate sample paths of the stochastic variables, a simple Euler discretization scheme[5] with $m = 220$ is applied, and again monomials are used as the basis functions. We set the model parameters according to the representative values in Table 2.2. As evident from Table 2.2, we note the conventional assumption of independence between financial and mortality risk. At each time of withdrawal, we use sixty two ($M = 62$) basis functions of the following form:

$$\mathbb{E}^{\mathbb{Q}}\left[\cdot|Y_t\right] = f(Y_t, \alpha) + G_{t+1} + G_{t+1} f(Y_t, \beta),$$

where $f(Y_t, \alpha) = \sum_{j_1+j_2+j_3 \leq 3} \alpha_{j_1,j_2,j_3} \left(X_t^+\right)^{j_1} (r_t)^{j_2} (\sqrt{\nu_t})^{j_3}$.

We use the same specifications of the contract as in Table 2.1 except for the option fee, which again is chosen as the fair fee according to the LSM valuation with a large number of basis functions ($M = 62$) and simulations ($N = 35 \times 10^5$). This will be our reference value in what follows. The option fee is estimated to be 0.7%, which is over four times higher than the previous option fee. While there are obvious differences in the models (non-constant mortality, lower interest rate, etc.), higher number of stochastic factors may increase the overall riskiness – and, thus, the value of the protection provided by the GMWB.

In Figure 2.3, we again plot the expected present value of benefits at time zero $V(0)$ under the LSM algorithm for different numbers of simulation paths $N$ in this extended model framework. Notably more simulations are required than in the simple Black-Scholes framework with a single stochastic driver (Figure 2.1) – even for 400,000 simulations, we notice a significant deviation. However, for more than 600,000 simulations, the valuation slowly converges. This is a significant advantage of the Monte Carlo algorithms considered here: Implementing a grid-based algorithm in the current setting with five states presents a complex problem and requires considerable computational resources, whereas implementing the Monte Carlo algorithm here is only

[5]Here, we use the full truncation method to generate sample paths for a square root process.

| **Parameter Values** | |
|---|---|
| Interest rate | |
| $r_0$ | 2.2% |
| $a$ | 0.25 |
| $b$ | 2% |
| $\sigma$ | 1.5% |
| Volatility | |
| $\nu_0$ | 0.011 |
| $\kappa$ | 3.5 |
| $\bar{\nu}$ | 0.01 |
| $\gamma$ | 15% |
| Mortality | |
| $\mu_{55}$ | 1.2% |
| $\psi$ | 20% |
| $\varphi$ | 1.5% |
| $\eta$ | 1.2% |
| Correlation | |
| $\rho_{1,2}$ | 0.27 |
| $\rho_{1,3}$ | –0.30 |
| $\rho_{1,4}$ | 0.00 |
| $\rho_{2,3}$ | 0.20 |
| $\rho_{2,4}$ | 0.00 |
| $\rho_{3,4}$ | 0.00 |
| Option fee | |
| $\phi$ | 0.7% |

Table 2.2: Setting of Parameter Values and Option fee

slightly more complex than in the basic Black-Scholes setting and feasible even on a personal computer.

In Figure 2.4, we analyze optimal withdrawal rules in this extended model setting using $M = 62$ basis functions and $N = 35 \times 10^5$ simulations. More precisely, we display different exercise regions – no withdrawal (blue), surrender (green), and withdrawing at the guaranteed rate $g$ (red) – for different combinations of the financial state variables $(r_t,\ v_t,\ X_t^-)$ for $\mu_{55+t} = 1.5\%$ and at different times $t$. The results here come with the caveat from the previous section: While the valuation converges, the validity of the withdrawal rules for low-probability or "off-equilibrium" combinations of the state variables is questionable. The first take-away from the figure is that the optimal

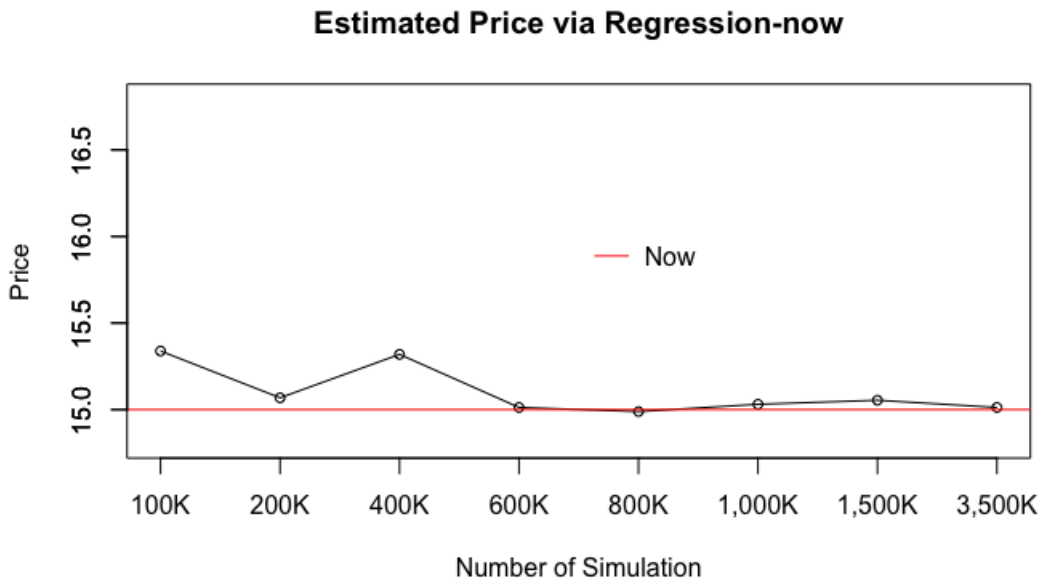
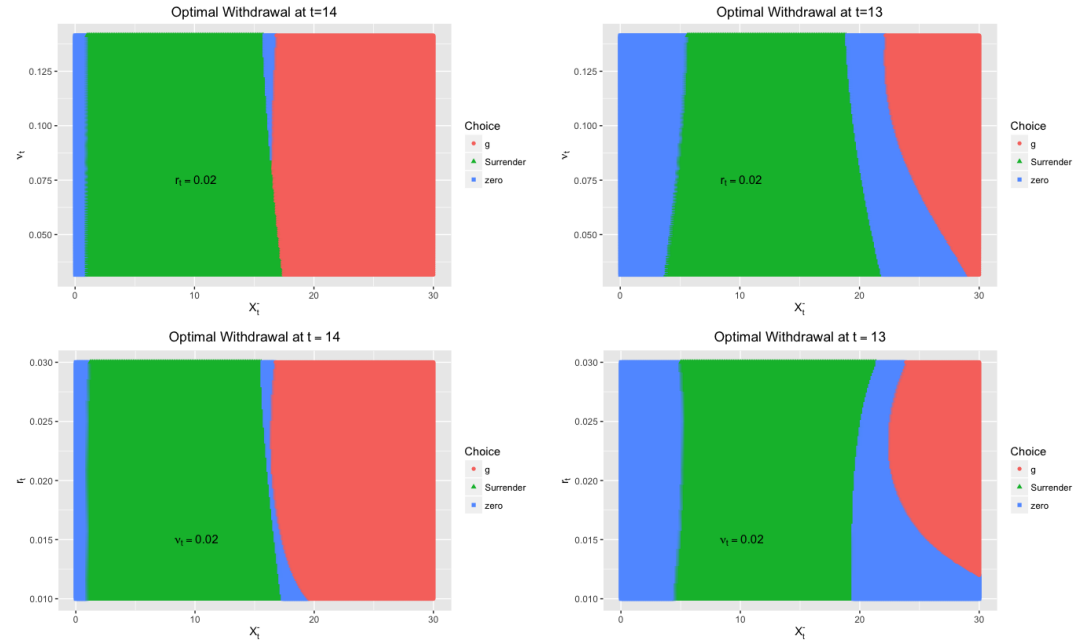



Figure 2.3: Convergence of Regression-Now Estimates: VA plus GMWB value for the LSM estimates (Now) as a function of the number of simulations $N$, four-factor model

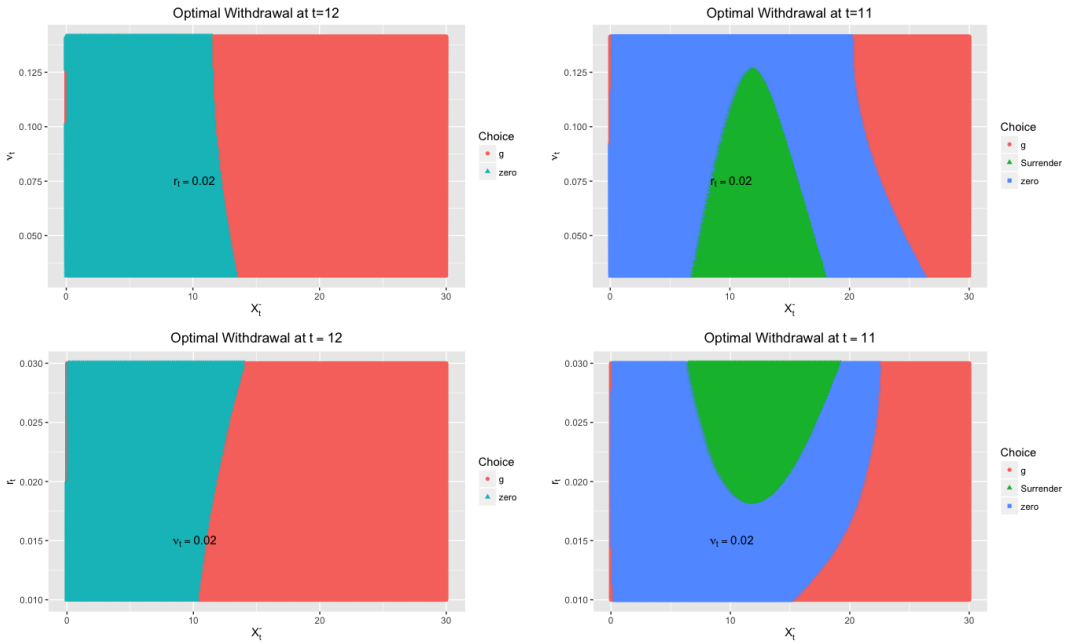
withdrawal strategy appears to be relatively "rich" in the sense that it depends on the state variables in a non-trivial – and sometimes even non-monotone – manner. Generally, it seems advisable to remain in the contract for relatively low account values whereas withdrawing/surrendering seems optimal in a moderate range of the account value.

## 2.5 Regression-later Algorithm

In some cases, it is possible to apply an alternative regression-based method to the valuation problem to speed up the convergence – the so-called *regression-later algorithm*. This variant approximates a functional of random variables in (2.5) via a linear combination of basis functions and least-squares regression, rather than the conditional expectation as within the regression-now algorithm. The conditional expectation of the approximation is then computed in the closed-form and serves as an approximation to the conditional expected value of interest. Since this approach uses information




(a) $t = 14$, $G_t = 2$ (b) $t = 13$, $G_t = 3$




(c) $t = 12$, $G_t = 4$ (d) $t = 11$, $G_t = 5$

Figure 2.4: Estimated Optimal Withdrawals based on the LSM algorithm (Now) for $N = 3,500,000$ at time $t$ with $\mu_{55+t} = 1.5\%$ at different times $t$

available at the end of the period in view to approximate the expectation, it is called the regression-*later* algorithm (Glasserman and Yu, 2002). Applications of regression-later can be found in Nadarajah et al. (2012) and Beutner et al. (2013).

In this section, after introducing the general approach we discuss the regression-later algorithm in the Black-Scholes framework from Section 2.4.1 with the same deterministic mortality assumption. Subsequently we comment on its application in the advanced model framework from Section 2.4.2.

We begin with the dynamic optimization problem at $t = t_{n-1}$. Under the regression-later algorithm, we approximate:

$$F_{t_n} = e^{-\mu(t_n - t_{n-1})} V_{t_n}(Y_{t_n}) + (1 - e^{-\mu(t_n - t_{n-1})}) D_{t_n}(Y_{t_n})$$

via the following steps:

1. The random variable $F_{t_n}$ is approximated using a linear combination of basis functions.

2. The associated coefficients are estimated by least-squares regression.

More precisely, suppose that $\{\varphi_k\}_{k=1}^{\infty}$ is a set of complete basis functions for $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{Q}_{t_n}\right)$. The random variable $F_{t_n}$ is then first approximated with the set of $M$ basis functions $\varphi = \{\varphi_1, ..., \varphi_M\}$:

$$F_{t_n} \approx \sum_{h=1}^{M} \beta_h^{t_n} \varphi_h(Y_{t_n}).$$

Note that we have realizations of $Y_{t_n}^l$ from the Monte Carlo simulations, $1 \leq l \leq N$. We rely on these to calculate:

$$F_{t_n}^l = e^{-(t_n - t_{n-1})\mu} V_{t_n}(Y_{t_n}^l) + (1 - e^{-(t_n - t_{n-1})\mu}) D_{t_n}(Y_{t_n}^l).$$

Using $\{F_{t_n}^l\}_{l=1}^N$, the coefficients $\beta_h^{t_n}$ are estimated by least-squares regression:

$$\hat{\beta}^{t_n} = \text{argmin}_{\{\beta^{t_n}\}} \sum_{l=1}^{N} \left[\varphi\left(Y_{t_n}^l\right) \cdot \beta^{t_n} - F_{t_n}^l\right]^2,$$

where $\varphi\left(Y_n^l\right) = \left(\varphi_1\left(Y_n^l\right), ..., \varphi_M\left(Y_n^l\right)\right)$ and $\beta^{t_n} = (\beta_1^{t_n}, ..., \beta_M^{t_n})'$. Then, the second approximation is obtained by replacing $\beta^{t_n}$ with $\hat{\beta}^{t_n}$:

$$F_{t_n} \approx \sum_{h=1}^{M} \hat{\beta}_h^{t_n} \varphi_h(Y_{t_n}). \tag{2.13}$$

It is important to note that (2.13) is not an expectation, but we approximate a random variable (payoff). For solving the dynamic optimization problem, however, we need to calculate expected value of (2.13). Thus, under the regression-later algorithm, it is assumed that a closed-form expression is available for calculating the following expectation:

$$\mathbb{E}^{\mathbb{Q}}\left[\varphi_h(Y_{t_n})|Y_{t_{n-1}}\right], \qquad h = 1, ..., M.$$

Therefore, the approximated dynamic problem at time $t_{n-1}$ becomes:

$$\begin{aligned}
V_{t_{n-1}}(Y_{t_{n-1}}) &= \max_{w_{t_{n-1}}} \; C(t_{n-1}, w_{t_{n-1}}) \\
&\quad + e^{-r(t_n - t_{n-1})} \mathbb{E}^{\mathbb{Q}}\left[e^{-\mu(t_n - t_{n-1})} V_{t_n}(Y_{t_n}) + (1 - e^{-\mu(t_n - t_{n-1})}) D_{t_n}(Y_{t_n})|Y_{t_{n-1}}\right], \\
&\approx \max_{w_{t_{n-1}}} \; C(t_{n-1}, w_{t_{n-1}}) + e^{-r(t_n - t_{n-1})} \sum_{h=1}^{M} \hat{\beta}_h^{t_n} \mathbb{E}^{\mathbb{Q}}\left[\varphi_h(Y_{t_n})|Y_{t_{n-1}}\right], \\
&\text{subject to} \qquad w_{t_i} \in A_{t_i}, \\
&\qquad\qquad V_{t_n}(Y_{t_n}) = \max(X_{t_n}^-, \min(g, G_{t_n})).
\end{aligned}$$

We assume that policyholder solves the above approximated dynamic optimization problem based on discretized feasible solutions.

At the previous withdrawal times, analogous approximations are applied and

approximated dynamic optimization problems are solved. More details about the regression-later algorithm for our valuation problem are provided in Appendix A.2.2.

Even though the final approximation of the regression-later algorithm in (2.13) seems to be similar to the regression-now algorithm, the two approaches are fundamentally different. In the regression-later algorithm, not the expectation is approximated, but the functional of random variable is via information available at time $t_{i+1}$. The primary advantage of the regression-later algorithm is that the algorithm approximates the $\mathcal{F}_{t_{i+1}}$-measurable functional accurately since the value function and death benefit are highly correlated with $Y_{t_{i+1}}$, whereas the $\mathcal{F}_{t_{i+1}}$-measurable functional is projected on the spanned space by $Y_{t_i}$ in the regression-now algorithm.

### 2.5.1 Regression-later Approach for GMWB Valuation

We implement the regression-later algorithm in the Black-Scholes framework introduced in Section 2.4.1 with the same assumptions. We again rely on monomial basis functions, so that we have the following estimator for functional of state variables:

$$\begin{aligned} & e^{-\mu(t_{i+1}-t_i)}V_{t_{i+1}}(Y_{t_{i+1}}) + (1-e^{-\mu(t_{i+1}-t_i)})D_{t_{i+1}}(Y_{t_{i+1}}) \\ & \approx \sum_{h=1}^{M} \hat{\beta}_h^{t_{i+1}} \left(X_{t_{i+1}}^-\right)^{h_1} \left(G_{t_{i+1}}\right)^{h_2}, \end{aligned} \tag{2.14}$$

where $h_1, h_2 \in \mathbb{Z}_+$. Note that $G_{t_{i+1}}$ is $\mathcal{F}_{t_i}$-measurable. Therefore, the expectation of (2.14) is:

$$\begin{aligned} & \mathbb{E}^{\mathbb{Q}}\left[\sum_{h=1}^{M} \hat{\beta}_h^{t_{i+1}} \left(X_{t_{i+1}}^-\right)^{h_1} \left(G_{t_{i+1}}\right)^{h_2} \middle| Y_{t_i}\right] = \sum_{h=1}^{M} \hat{\beta}_h^{t_{i+1}} \mathbb{E}^{\mathbb{Q}}\left[\left(X_{t_{i+1}}^-\right)^{h_1} \middle| Y_{t_i}\right] \left(G_{t_{i+1}}\right)^{h_2} \\ & = \sum_{h=1}^{M} \hat{\beta}_h^{t_{i+1}} \left(X_{t_i}^+\right)^{h_1} \exp\left(\left(r-\phi-\frac{1}{2}\sigma^2\right)(t_{i+1}-t_i)h_1 + \frac{1}{2}\sigma^2(t_{i+1}-t_i)^2 h_1^2\right) \left(G_{t_{i+1}}\right)^{h_2}. \end{aligned}$$

We use the same contract specifications and parameter values as in Section 2.4.1 (Table 2.1). We first implement the regression-later algorithm with $M = 8$ basis functions.[6] A key question is to check whether it converges faster to the exact price than the regression-now algorithm.

In Figure 2.5, we plot the expected present value of benefits at time zero $V(0)$ for different numbers of simulation paths $N$ under the three considered algorithms: The grid-based algorithm, the regression-now algorithm, and the regression-later algorithm. We notice the regression-later estimates provide a close approximation to the grid-based value already for relatively low choices of $N$. This is in line with results from the literature that emphasize the advantages of the regression-later approach since it uses informationoin $Y_{t_{i+1}}$ directly to approximate the $\mathcal{F}_{t_{i+1}}$-measurable function (Nadarajah et al., 2012; Beutner et al., 2013).

To obtain insights on the improved convergence properties of the regression-later approach, Figure 2.6 again displays the maximizing withdrawal rules from Figure 2.2 ($N = 30,000$ and $N = 6000,000$ simulations for different points in time on the left- and right-hand side, respectively), but we now overlay the optimal withdrawal rules from the regression-later algorithm. The key observation is that for the regression-later algorithm, the withdrawal pattern for $N = 30,000$ simulations presents a closer approximation to the grid-based withdrawal rule. While not perfect, the match is markedly better resulting in a closer valuation as observed from Figure 2.5.

Despite these advantages, the regression-later approach also comes with a variety of disadvantages that are exacerbated in high-dimensional settings. More specifically, while the application is straightforward in the simple Black-Scholes framework considered here, in multi-factor models (i) it is necessary to put structural constraints on the state variables to assure the existence of closed-forms for the expected values

[6] The regression equation considered is

$$F = \alpha_0 + \alpha_1 X_t^- + \alpha_2 \left(X_t^-\right)^2 + \alpha_3 \left(X_t^-\right)^3 + G_{t_{i+1}} + G_{t_{i+1}} \left(\beta_1 X_t^- + \beta_2 \left(X_t^-\right)^2 + \beta_3 \left(X_t^-\right)^3\right).$$

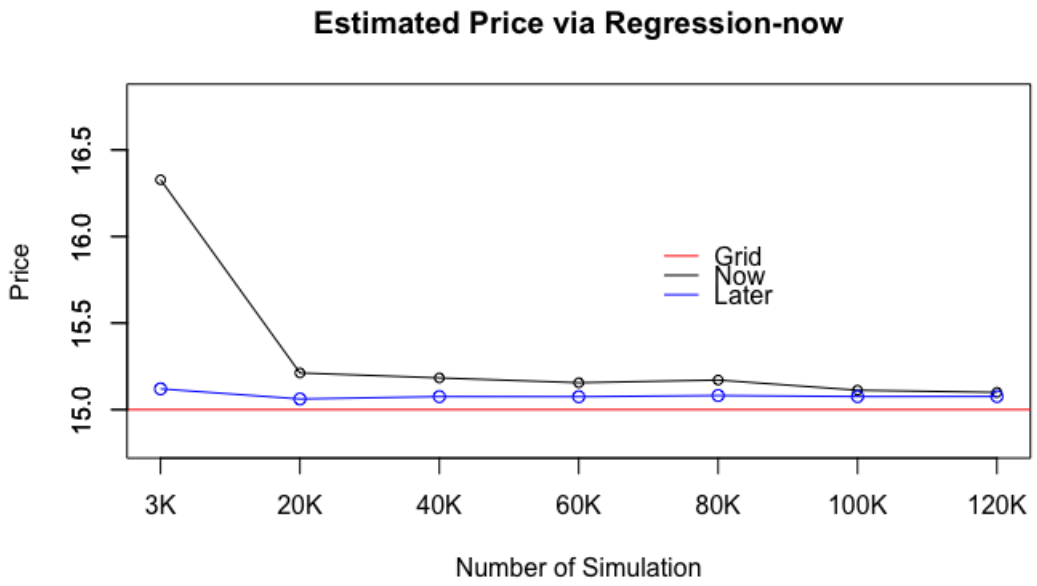
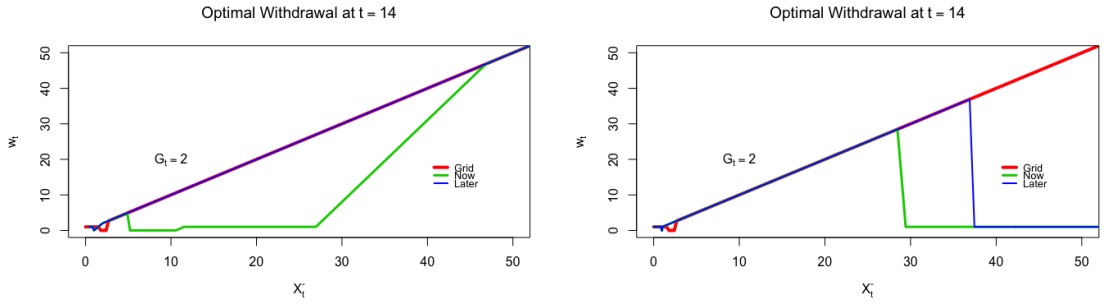



Figure 2.5: Convergence of Regression-Later Estimates: VA plus GMWB value for the grid-based algorithm (Grid), the regression-now estimates (Now), and the regression-later estimates (Later) as a function of the number of simulations $N$, basic Black-Scholes model

of the basis functions; and (ii) even then the choice of basis functions is constrained by the availability of closed-forms. Depending on the form of the random variables that need to be approximated, these constraints may be prohibitive for determining approximations. Indeed, although the model from Section 2.4.2 falls in the affine class when suitably transformed, our numerical experiments did not yield viable results.

## 2.6 Conclusion

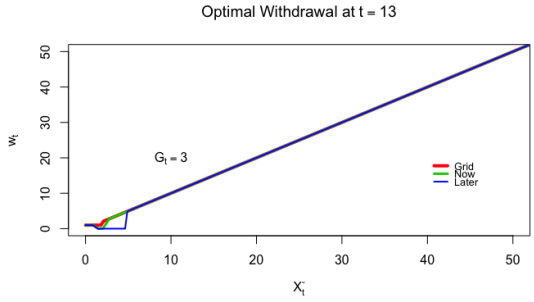
We establish and apply two LSM algorithms to price a VA plus GMWB contract. We find that the two approaches require sizable computational budgets to produce viable results, where the regression-later estimator converge faster than the regression-now estimators. However, the regression-now algorithm is more robust in the sense that it does not require structural assumptions on the form of the stochastic drivers or the form of the basis functions. In particular, with little modification relative to the basic



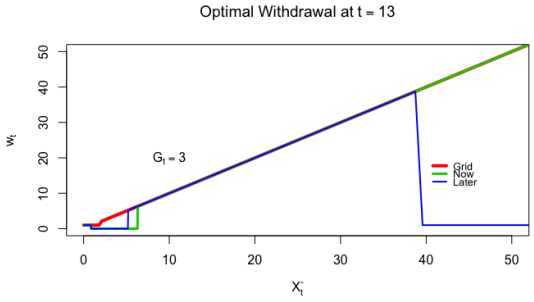

(a) $t = 14,\ N = 30,000$



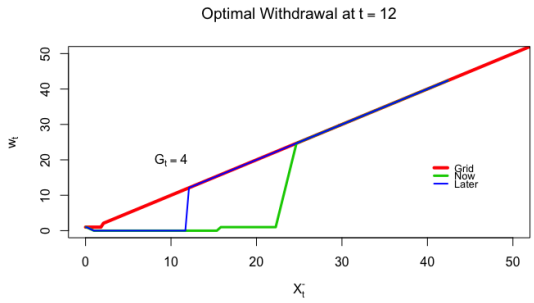

(b) $t = 14,\ N = 600,000$



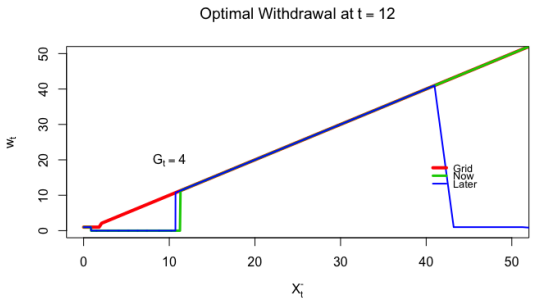

(c) $t = 13,\ N = 30,000$



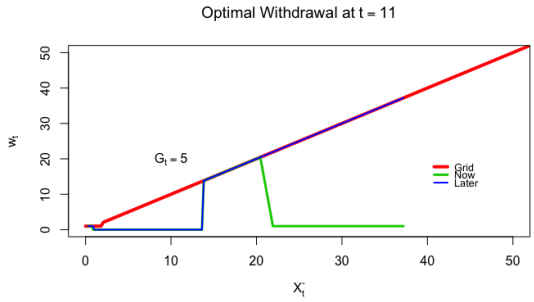

(d) $t = 13,\ N = 600,000$



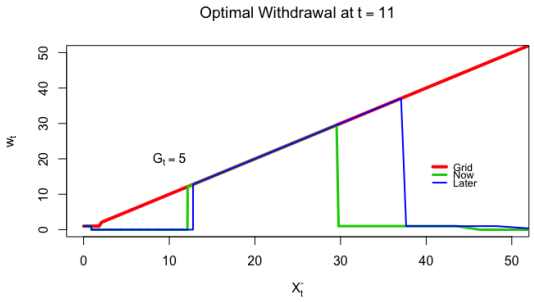

(e) $t = 12,\ N = 30,000$




(f) $t = 12,\ N = 600,000$




(g) $t = 11,\ N = 30,000$




(h) $t = 11,\ N = 600,000$

Figure 2.6: Estimated Optimal Withdrawals for the grid-based algorithm (Grid), the regression-now algorithm (Now), and the regression-later algorithm (Later) for $N = 30,000$ (left-hand side) and $N = 600,000$ (right-hand side) at different times $t$

Black-Scholes setup, it is possible to evaluate the VA plus GMWB contract using the regression-now algorithm in a more advanced model with several stochastic drivers.

# Appendices

## A.1 Proofs

***Proof of Lemma 1.2.1.*** 1. Let $A \in \mathcal{F}_t,\ 0 \leq t \leq \tau$. Then:

$$\begin{aligned}\tilde{\mathbb{P}}(A) &= \mathbb{E}^{\tilde{\mathbb{P}}}[\mathbf{1}_A] = \mathbb{E}^{\mathbb{P}}\left[\frac{\partial\tilde{\mathbb{P}}}{\partial\mathbb{P}}\mathbf{1}_A\right] = \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{P}}\left[\frac{\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}}{\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}\,|\mathcal{F}_\tau\right]}\mathbf{1}_A\middle|\,\mathcal{F}_\tau\right]\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[\frac{\mathbf{1}_A}{\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}\,|\mathcal{F}_\tau\right]}\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}\middle|\,\mathcal{F}_\tau\right]\right] = \mathbb{P}(A).\end{aligned}$$

2. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then:

$$\begin{aligned}\mathbb{E}^{\tilde{\mathbb{P}}}[X\,|\mathcal{F}_\tau] &= \underbrace{\frac{1}{\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\tilde{\mathbb{P}}}{\partial\mathbb{P}}\,|\mathcal{F}_\tau\right]}}_{=1}\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\tilde{\mathbb{P}}}{\partial\mathbb{P}}X\middle|\,\mathcal{F}_\tau\right] = \mathbb{E}^{\mathbb{P}}\left[\frac{X\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}}{\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}\,|\mathcal{F}_\tau\right]}\middle|\,\mathcal{F}_\tau\right] \\ &= \frac{1}{\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}\,|\mathcal{F}_\tau\right]}\mathbb{E}^{\mathbb{P}}\left[\frac{\partial\mathbb{Q}}{\partial\mathbb{P}}X\middle|\,\mathcal{F}_\tau\right] = \mathbb{E}^{\mathbb{Q}}\left[X|\,\mathcal{F}_\tau\right].\end{aligned}$$

□

***Proof of Lemma 1.2.2.*** Linearity is obvious. For the proof of continuity, consider a sequence $h^{(n)} \to h \in \mathcal{H}$. Then:

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[L\,h^{(n)} - L\,h\right]^2 \\ =\ &\mathbb{E}^{\mathbb{P}}\left[\left(\sum_{j=\tau}^{T}\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_j^{(n)} - h_j\right)(Y_j)\,|Y_\tau\right]\right)^2\right] \\ =\ &\mathbb{E}^{\mathbb{P}}\left[\sum_{j,k}\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_j^{(n)} - h_j\right)(Y_j)\,|Y_\tau\right]\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_k^{(n)} - h_k\right)(Y_k)\,|Y_\tau\right]\right] \\ \leq\ &\sum_{j,k}\sqrt{\mathbb{E}^{\mathbb{P}}\left[\left(\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_j^{(n)} - h_j\right)(Y_j)\,|Y_\tau\right]\right)^2\right]}\times\sqrt{\mathbb{E}^{\mathbb{P}}\left[\left(\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_k^{(n)} - h_k\right)(Y_k)\,|Y_\tau\right]\right)^2\right]} \\ \leq\ &\sum_{j,k}\sqrt{\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_j^{(n)} - h_j\right)^2(Y_j)\right]}\times\sqrt{\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(h_k^{(n)} - h_k\right)^2(Y_k)\right]} \to 0,\ n \to \infty,\end{aligned}$$

where we used the Cauchy-Schwarz inequality, the conditional Jensen inequality, and the tower property of conditional expectations. □

***Proof of Lemma 1.2.3.*** Consider an operator $L^{(j)}$, which maps from $L^2\left(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_j}\right)$ to $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right)$. Since $L^{(j)}$ is the (conditional) expectation under the assumption that there exists a joint density, it can be represented as an integral:

$$\begin{aligned} L^{(j)}\,\mathbf{x} &= \int_{\mathbb{R}^d} \mathbf{x}(y)\,\pi_{Y_\tau|Y_j}(y|x)\,dy = \int_{\mathbb{R}^d} \mathbf{x}(y)\frac{\pi_{Y_\tau,Y_j}(y,x)}{\pi_{Y_j}(x)}\,dy \\ &= \int_{\mathbb{R}^d} \mathbf{x}(y)\frac{\pi_{Y_\tau,Y_j}(y,x)}{\pi_{Y_j}(x)\pi_{Y_\tau}(y)}\pi_{Y_\tau}(y)\,dy = \int_{\mathbb{R}^d} \mathbf{x}(y)\,k(x,y)\,\pi_{Y_\tau}(y)\,dy, \end{aligned}$$

where $\mathbf{x}$ is an element of $L^2\left(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_j}\right)$, $\pi_{Y_\tau,Y_j}(y,x)$ is the joint density function of $Y_\tau$ and $Y_j$, $\pi_{Y_j}(y)$ and $\pi_{Y_\tau}(x)$ are marginal density functions for $Y_j$ and $Y_\tau$ in $L^2\left(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_j}\right)$ and $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right)$, respectively, and $k(x,y) = \frac{\pi_{Y_\tau,Y_j}(y,x)}{\pi_{Y_j}(x)\pi_{Y_\tau}(y)}$. Thus, $L^{(j)}$ is an integral operator with kernel $k(x,y)$. Moreover,

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |k(x,y)|^2\,\pi_{Y_j}(x)\pi_{Y_\tau}(y)\,dx\,dy = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \pi_{Y_j|Y_\tau}(x|y)\,\pi_{Y_\tau|Y_j}(y|x)\,dx\,dy < \infty.$$

Thus, according to p. 9 in Carrasco and Florens (2011), $L^{(j)}$ is a Hilbert-Schdmit operator and therefore compact. Finally, $L$ is the sum of $L^{(j)}$, $j = \tau, ..., T$, and therefore also compact. □

***Proof of Proposition 1.3.1.*** $\mathbb{P}_{Y_\tau}$ is a regular Borel measure as a finite Borel measure and hence $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right)$ is separable (see Proposition I.2.14 and p. 33 in Werner (2005)). Now if $\{e_k,\ k = 1, 2, \ldots, M\}$ are independent, by Gram-Schmidt we can find an orthonormal system $S = \{f_k,\ k = 1, 2, \ldots, M\}$ with $\text{lin}\{e_k,\ k = 1, 2, \ldots, M\} = \text{lin}\, S$.[7] For $S$, on the other hand, we can find an orthonormal basis $\{f_k,\ k \in \mathbb{N}\} = S' \supset S$. Hence:

$$\widehat{C}_\tau^{(M)} = \sum_{k=1}^{M} \alpha_k\, e_k = \sum_{k=1}^{M} \underbrace{\tilde{\alpha}_k}_{\langle C_\tau, f_k\rangle}\ f_k \to \sum_{k=1}^{\infty} \tilde{\alpha}_k\, f_k = C_\tau,\ M \to \infty,$$

[7] We denote by $\text{lin}\, S$ the (sub-)space spanned by the elements of $S$.

where

$$\left\|\widehat{C}_\tau^{(M)} - C_\tau\right\|^2 = \sum_{k=M+1}^{\infty} |\langle C_\tau, f_k\rangle|^2 \to 0,\ M \to \infty,$$

by Parseval's identity.

For the second part, we note that

$$(\hat{\alpha}_1^{(N)}, \ldots, \hat{\alpha}_M^{(N)})' = \hat{\alpha}^{(N)} = \left(A^{(M,N)}\right)^{-1} \frac{1}{N} \sum_{i=1}^{N} e\left(Y_\tau^{(i)}\right)\ V_\tau^{(i)},$$

where $e(\cdot) = (e_1(\cdot), \ldots, e_M(\cdot))'$ and $A^{(M,N)} = \left[\frac{1}{N}\sum_{i=1}^{N} e_k(Y_\tau^{(i)})\, e_l(Y_\tau^{(i)})\right]_{1\le k,l\le M}$ is invertible for large enough $N$ since we assumed that the basis functions are linearly independent. Hence,

$$\hat{\alpha}^{(N)} \to \alpha = (\alpha_1, \ldots, \alpha_M)' = \left(A^{(M)}\right)^{-1} \mathbb{E}^{\tilde{\mathbb{P}}}\left[e\left(Y_\tau\right)\ \left(\sum_{k=\tau}^{T} x_k\right)\right]\ \tilde{\mathbb{P}}\text{-a.s.},$$

by the law of large numbers, where $A^M = \left[\mathbb{E}^{\tilde{\mathbb{P}}}\left[e_k\left(Y_\tau\right)\ e_l\left(Y_\tau\right)\right]\right]_{1\le k,l\le M}$, so that

$$\widehat{C}_\tau^{(M,N)} = e'\,\hat{\alpha}^{(N)} \to e'\alpha = \widehat{C}_\tau^{(M)}\ \tilde{\mathbb{P}}\text{-a.s.}$$

Finally, for the third part, let

$$V_\tau^{(i)} = \sum_{k=\tau}^{T} x_k\left(Y_\tau^{(i)}\right) = \sum_{j=1}^{M} \alpha_j e_j\left(Y_\tau^{(i)}\right) + \epsilon_j,$$

where $\mathbb{E}\left[\epsilon_j | Y_\tau\right] = 0, \mathrm{Var}\left[\epsilon_j | Y_\tau\right] = \Sigma(Y_\tau)$, and $\mathrm{Cov}\left[\epsilon_i,\ \epsilon_j | Y_\tau\right] = 0$.

Thus (see e.g. Section 6.13 in Amemiya (1985)):

$$\sqrt{N}[\alpha - \hat{\alpha}^{(N)}] \longrightarrow \text{Normal}\left[0, \underbrace{\left(A^{(M)}\right)^{-1}\left[\mathbb{E}^{\mathbb{P}}\left[e_k(Y_\tau)e_l(Y_\tau)\Sigma(Y_\tau)\right]\right]_{1\le k,l\le M}\left(A^{(M)}\right)^{-1}}_{\tilde{\xi}}\right],$$

so that

$$\sqrt{N}\left[\widehat{C}_\tau^{(M)} - \widehat{C}_\tau^{(M,N)}\right] = e'[\alpha - \hat{\alpha}^{(N)}]\sqrt{N} \longrightarrow \text{Normal } (0, \xi^{(M)}),$$

where

$$\xi^{(M)} = e'\,\tilde{\xi}\,e. \tag{15}$$

□

***Proof of Corollary 1.3.1.*** Relying on the notation from the proof of Proposition 1.3.1, we now have (supposing square integrability)

$$\hat{\alpha}^{(N)} = \frac{1}{N}\sum_{i=1}^{N} e\left(Y_\tau^{(i)}\right)\, V_\tau^{(i)} \to \alpha,\ N \to \infty$$

in $L^2\left(\Omega, \mathcal{F}, \tilde{\mathbb{P}}\right)$ by the $L^2$-version of the weak law of large numbers (Durett, 1996). Thus,

$$\begin{aligned}
& \mathbb{E}^{\tilde{\mathbb{P}}}\left[\left|e(Y_\tau)'\,\hat{\alpha}^{(N)} - e(Y_\tau)'\,\alpha\right|\right] \\
\leq\ & \sum_{k=1}^{M} \mathbb{E}^{\tilde{\mathbb{P}}}\left[\left|e_k(Y_\tau)'\,\left(\hat{\alpha}_k^{(N)} - \alpha_k\right)\right|\right] \\
\leq\ & \sum_{k=1}^{M} \sqrt{\mathbb{E}^{\tilde{\mathbb{P}}}\left[e_k^2(Y_\tau)\right]}\sqrt{\mathbb{E}^{\tilde{\mathbb{P}}}\left[\hat{\alpha}_k^{(N)} - \alpha_k\right]^2} \to 0,\ N \to \infty.
\end{aligned}$$

The last assertion in the statement is a direct consequence of the Extended Namioka Theorem in Biagini and Fritelli (2009). □

***Proof of Corollary 1.3.2.*** The first assertion immediately follows from convergence in distribution as discussed in Section 1.3.1. For the quantiles, the convergence for all continuity points of $F_{C_\tau}^{-1}$ follows from Proposition 1.3.1 and the standard proof of Skorokhod's representation theorem (see e.g. Lemma 1.7 in Whitt (2002)). □

***Proof of Proposition 1.3.2.*** Since as Monte Carlo trials $(V_\tau^{(i)}, Y_\tau^{(i)})$ are i.i.d., the first part of Assumption 1 in Newey (1997) is automatically satisfied. The conditions

in the proposition are then exactly Assumptions 1 (part 2), 2, and 3 in his paper for $d = 0$. Thus, the claim follows by the first part of Theorem 1 in Newey (1997). □

***Proof of Proposition 1.3.3.*** Analogously to the proof or Proposition 3.2, the first part of Assumption 1 in Newey (1997) is automatically satisfied. The conditions in the proposition are taken from the second part of Assumption 1, Assumption 8, the discussion following Assumption 8, and Assumption 9 in his paper. Thus, the claim follows by the first part of Theorem 4 in Newey (1997). □

***Regularity Conditions in Proposition 1.3.4.*** *(Gordy and Juneja, 2010).* Regularity conditions on the joint probability function (pdf) $g$ of $(-\widehat{C}_\tau^{(M)}, Z^{(N)})$:

- The joint pdf $g_N(\cdot,\cdot)$, its partial derivatives $\frac{\partial}{\partial y} g_N(y,z)$ and $\frac{\partial^2}{\partial y^2} g_N(y,z)$ exist for each $N$ and for all $(y,z)$.
- For $N \geq 1$, there exist non-negative functions $p_{0,N}(\cdot)$, $p_{1,N}(\cdot)$ and $p_{2,N}(\cdot)$ such that:
  - $g_N(y,z) \leq p_{0,N}(z)$,
  - $\left|\frac{\partial}{\partial y} g_N(y,z)\right| \leq p_{1,N}(z)$,
  - $\left|\frac{\partial^2}{\partial y^2} g_N(y,z)\right| \leq p_{2,N}(z)$, and

  for all $y$ and $z$. In addition:

$$\sup_N \int_{-\infty}^{\infty} |z|^r p_{i,N}(z) dz < \infty$$

  for $i = 0, 1, 2$ and $0 \leq r \leq 4$.

□

***Singular Value Decomposition of a Compact Operator (Section 1.4.2).*** Suppose $A$ is a compact operator mapping from $\mathcal{H}_1$ to $\mathcal{H}_2$, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are

separable Hilbert spaces. Then, $A$ can be represented in the following form (see Section VI.3 in Werner (2005) or Huang (2012)):

$$A\mathbf{x} = \sum_{k=1}^{\infty} \lambda_k < \mathbf{x}, g_k >_{\mathcal{H}_1} f_k, \tag{16}$$

where:

- $< \cdot, \cdot >_{\mathcal{H}_1}$ denotes the inner product in $\mathcal{H}_1$;
- $\{\lambda_k^2\}$ are non-zero eigenvalues of $A^*A$ and $AA^*$ with $\lambda_1 \geq \lambda_2 \geq \cdots$, counted according to their multiplicity. Here, $\lambda_k$ is called the $k$-th singular value of $A$;
- $\{g_k\}$, called the (right) *singular functions*, are the orthonormal eigenfunctions of $A^*A$; and
- $\{f_k\}$, called the (left) *singular functions*, are the orthonormal eigenfunctions of $AA^*$ satisfying $A\, g_k = \lambda_k\, f_k$.

The representation (16) is called *singular value decomposition* (SVD) of $A$ and the triplet $(\lambda_k, g_k, f_k)$ is called *singular system* for $A$. The functional sequences, $\{g_k\}_{k\geq 1}$ and $\{f_k\}_{k\geq 1}$, form a complete orthonormal sequence for $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The singular values $\lambda_k$ are non-negative and the only possible accumulation point is zero. For more details about the SVD of a compact operator, we refer to Huang (2012). □

***Proof of Proposition 1.4.1.*** We consider the approximation of $L$ by an arbitrary rank-$M$ operator $L_F$, which can be represented as

$$L_F = \sum_{k=1}^{M} \alpha_k \, < \cdot, u_k > \, e_k,$$

where $\{\alpha_k\}_{k=1}^M \subseteq \mathbb{R}_+$, $\{u_k\}_{k=1}^M$ are orthonormal in $\mathcal{H}$, and $\{e_k\}_{k=1}^M$ are orthonormal in $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right)$. Denote by $L_F^*$ the operator when choosing $(\alpha_k, u_k, e_k) = (\omega_k, s_k, \varphi_k)$.

Then:

$$\begin{aligned}\inf_{\{L_F\}} ||L - L_F||^2 &\leq \sup_{||\mathbf{x}||=1} ||L\mathbf{x} - L_F^*\mathbf{x}||^2 \\ &= \sup_{||\mathbf{x}||=1} || \sum_{k=M+1}^{\infty} \omega_k < \mathbf{x}, s_k > \varphi_j||^2 \\ &= \sup_{||\mathbf{x}||=1} \sum_{k=M+1}^{\infty} \omega_k^2 < \mathbf{x}, s_k >^2 = \omega_{M+1}^2.\end{aligned}$$

On the other hand, consider any alternative system $(\alpha_k, u_k, e_k)$ for an arbitrary finite-rank operator $L_F$. Then choose a non-zero $\mathbf{x}_0$ such that $\mathbf{x}_0 \in \text{lin}\{s_1, ..., s_{M+1}\} \cap \text{lin}\{u_1, ..., u_M\}^{\perp} \neq \{0\}$. Note that $L - L_F$ is compact and bounded. Therefore:

$$\begin{aligned}||L - L_F||^2 \geq \frac{||L\mathbf{x}_0 - L_F\,\mathbf{x}_0||^2}{||\mathbf{x}_0||^2} &= \frac{||L\mathbf{x}_0||^2}{||\mathbf{x}_0||^2} \\ &= \frac{\sum_{k=1}^{M+1} \omega_k^2 | < \mathbf{x}_0, s_k > |^2}{\sum_{k=1}^{M+1} | < \mathbf{x}_0, s_k > |^2} \geq \omega_{M+1}^2.\end{aligned}$$

Hence:

$$\inf_{\{L_F\}} ||L - L_F||^2 = \omega_{M+1}^2 = ||L - L_F^*||.$$

Now since:

$$\inf_{\{L_F\}} ||L - L_F||^2 = \inf_{\{e_1,...,e_M\}} ||L - P(e_1, ..., e_M) \cdot\, L||^2,$$

where $P(e_1, ..., e_M)$ denotes the orthogonal projection on the subspace spanned by $(e_1, ..., e_M)$, the claim follows by Equation (1.7). □

***Proof of Proposition 1.4.2.*** Proceeding as in Equation (1.9) and with Equation

(1.7), we obtain:

$$\begin{aligned}
\inf_{\alpha_M} \sup_{y\in\mathcal{Y}} \left| C_\tau(y) - \sum_{k=1}^{M} \alpha_{M,k} e_k(y) \right| &\leq \sup_{y\in\mathcal{Y}} \left| C_\tau(y) - \widehat{C}_\tau^{(M)}(y) \right| \\
&= \sup_{y\in\mathcal{Y}} | \sum_{k=M+1}^{\infty} \omega_k \langle \mathbf{x}, s_k \rangle \varphi_k(y) | \\
&\leq \sum_{k=M+1}^{\infty} \omega_k |\langle \mathbf{x}, s_k \rangle| \sup_{y\in\mathcal{Y}} |\varphi_k(y)| \\
&\leq \sum_{k=M+1}^{\infty} \omega_k \|\mathbf{x}\| \|s_k\| \sup_{y\in\mathcal{Y}} |\varphi_k(y)| \\
&= \sum_{k=M+1}^{\infty} \omega_k \|\mathbf{x}\| \sup_{y\in\mathcal{Y}} |\varphi_k(y)| = O(\omega_M)
\end{aligned}$$

for a fixed $\mathbf{x}$ since the $\{\varphi_k\}$ are uniformly bounded, where the second and third inequality follow by the triangle and Cauchy-Schwarz inequalities, respectively.

Then, going through of the assumptions of Proposition 1.3.2 with $B = I$ and $e^{(M)} = (e_1, ..., e_M)'$, we obtain:

$$\mathbb{E}^{\tilde{\mathbb{P}}} \left[ \tilde{e}^{(M)}(Y_\tau) \tilde{e}^{(M)}(Y_\tau)' \right] = I$$

due to the orthonormality of the singular functions. Therefore, the smallest eigenvalues is bounded away from zero uniformly for every $M$. Moreover, for fixed $y \in \mathcal{Y}$, $||\tilde{e}^{(M)}(y)|| = \sqrt{\varphi_1(y)^2 + \cdots \varphi_M(y)^2}$, so that

$$\begin{aligned}
\sup_{y\in\mathcal{Y}} ||\tilde{e}^{(M)}(y)|| &= \sup_{y\in\mathcal{Y}} \sqrt{\varphi_1(y)^2 + \cdots \varphi_1(y)^2} \\
&\leq \sqrt{\sum_{k=1}^{M} \sup_{y\in\mathcal{Y}} \varphi_k(y)^2} \leq \sqrt{\max_{1\leq k\leq M} \sup_{y\in\mathcal{Y}} \varphi_k(y) \cdot M} = C\sqrt{M} = \zeta_0(M)
\end{aligned}$$

since the $\{\varphi_k\}$ are uniformly bounded. Thus, the claim follows by Proposition 1.3.2.

□

***Proof of Lemma 1.4.1.*** The assertions on the conditional distributions are standard. For showing that $L$ is compact, we check that the transition and the reverse transition density functions satisfy the condition in Lemma 1.2.3. Note that the transition density function can be written as:

$$\begin{aligned}
\pi_{Y_T|Y_\tau}(y|x) &= g(y; \mu_T + \Gamma'\Sigma_\tau^{-1}(x-\mu_\tau), \Sigma_{T|\tau}) \\
&= \frac{1}{(2\pi)^{d/2}|\Sigma_{T|\tau}|^{1/2}} \exp\left[-\frac{1}{2}(y-\mu_T-\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau))'\Sigma_{T|\tau}^{-1}(y-\mu_T-\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau))\right] \\
&= \frac{1}{(2\pi)^{d/2}|\Sigma_{T|\tau}|^{1/2}} \frac{|\Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau|^{1/2}}{|\Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau|^{1/2}} \\
&\quad \times \exp\left[-\frac{1}{2}(x-\mu_\tau-\Sigma_\tau(\Gamma')^{-1}(y-\mu_T))'\Sigma_\tau^{-1}\Gamma\Sigma_{T|\tau}^{-1}\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau-\Sigma_\tau(\Gamma')^{-1}(y-\mu_T))\right] \\
&= \frac{|\Sigma_\tau|}{|\Gamma|} g(x; \mu_\tau + \Sigma_\tau(\Gamma')^{-1}(y-\mu_T), \Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau).
\end{aligned}$$

We evaluate the following integral:

$$\begin{aligned}
&\int_{\mathbb{R}^d} \pi_{Y_T|Y_\tau}(y|x)\pi_{Y_\tau|Y_T}(x|y)dx \\
&= \frac{|\Sigma_\tau|}{|\Gamma|}\int_{\mathbb{R}^d} g(x; \mu_\tau + \Sigma_\tau(\Gamma')^{-1}(y-\mu_T), \Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau) \\
&\qquad\qquad \times g(x; \mu_\tau + \Gamma\Sigma_\tau^{-1}(y-\mu_T), \Sigma_{\tau|T})\, dx \\
&= \frac{|\Sigma_\tau|}{|\Gamma|(2\pi)^{d/2}} \frac{1}{|\Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau + \Sigma_{\tau|T}|^{1/2}} \\
&\quad \times \exp\Bigg[-\frac{1}{2}(\Sigma_\tau(\Gamma')^{-1}(y-\mu_T) - \Gamma\Sigma_T^{-1}(y-\mu_T))'(\Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau + \Sigma_{\tau|T})^{-1} \\
&\qquad\qquad \times (\Sigma_\tau(\Gamma')^{-1}(y-\mu_T) - \Gamma\Sigma_T^{-1}(y-\mu_T))\Bigg] \\
&= \frac{|\Sigma_\tau|}{|\Gamma|(2\pi)^{d/2}} \frac{1}{|\Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau + \Sigma_{\tau|T}|^{1/2}} \\
&\quad \times \exp\Bigg[-\frac{1}{2}(y-\mu_T)' \underbrace{(\Gamma^{-1}\Sigma_\tau - \Sigma_T^{-1}\Gamma')(\Sigma_\tau(\Gamma')^{-1}\Sigma_{T|\tau}\Gamma^{-1}\Sigma_\tau + \Sigma_{\tau|T})^{-1}(\Sigma_\tau(\Gamma')^{-1} - \Gamma\Sigma_T^{-1})}_{V^{-1}} \\
&\qquad\qquad \times (y-\mu_T)\Bigg] \\
&= C_1 \times g(y; \mu_T, V),
\end{aligned}$$

where we use results on the product of Gaussian densities (Vinga, 2004) and where

$C_1$ is an appropriate constant to obtain $g(y; \mu_T, V)$. Therefore:

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \pi_{Y_T|Y_\tau}(y|x) \pi_{Y_\tau|Y_T}(x|y)\, dx\, dy = \int_{\mathbb{R}^d} C_1 g(y; \mu_T, V)\, dy = C_1 < \infty.$$

□

***Proof of Lemma 1.4.2.*** $L^*$ can be found by:

$$\begin{aligned}
\langle Lh, m \rangle_{\pi_{Y_\tau}} &= \int_{\mathbb{R}^d} Lh(x)\, m(x)\, \pi_{Y_\tau}(x)\, dx = \int \left[ \int h(y) \pi_{Y_T|Y_\tau}(y|x)\, dy \right] m(x)\, \pi_{Y_\tau}(x)\, dx \\
&= \int_{\mathbb{R}^d} h(y) \left[ \int_{\mathbb{R}^d} m(x) \pi_{Y_\tau|Y_T}(x|y)\, dx \right] \pi_{Y_T}(y)\, dy = \langle h, L^* m \rangle,
\end{aligned}$$

where $L^* m(y) = \int_{\mathbb{R}^d} m(x) \pi_{Y_\tau|Y_T}(x|y)\, dx$. We calculate $LL^*$:

$$\begin{aligned}
LL^* \varphi(x) &= \int_{\mathbb{R}^d} L^* \varphi(s) \pi_{Y_T|Y_\tau}(s|x)\, ds \\
&= \int_{\mathbb{R}^d} \left[ \int \varphi(y) \pi_{Y_\tau|Y_T}(y|s)\, dy \right] \pi_{Y_T|Y_\tau}(s|x)\, ds \\
&= \int_{\mathbb{R}^d} \varphi(y) \underbrace{\int_{\mathbb{R}^d} \pi_{Y_\tau|Y_T}(y|s) \pi_{Y_T|Y_\tau}(s|x)\, ds}_{K_A(x,y)}\, dy.
\end{aligned}$$

It is useful to express the reverse density in the following form to find $K_A(x, y)$ as in the proof of Lemma 1.4.1:

$$g(y; \mu_{Y_\tau|s}, \Sigma_{\tau|T}) = \frac{|\Sigma_T|}{|\Gamma|} g(s; \mu_T + \Sigma_T \Gamma^{-1}(y - \mu_\tau),\ \Sigma_T \Gamma^{-1} \Sigma_{\tau|T} (\Gamma')^{-1} \Sigma_T).$$

Hence:

$$
\begin{aligned}
K_A(x,y) &= \int_{\mathbb{R}^d} \pi_{Y_\tau|Y_T}(y|s)\pi_{Y_T|Y_\tau}(s|x)\,ds \\
&= \frac{|\Sigma_T|}{|\Gamma|}\int_{\mathbb{R}^d} g(s;\mu_T+\Sigma_T\Gamma^{-1}(y-\mu_\tau),\ \Sigma_T\Gamma^{-1}\Sigma_{\tau|T}(\Gamma')^{-1}\Sigma_T)\times g(s;\mu_{T|x},\Sigma_{T|\tau})\,ds \\
&= \frac{|\Sigma_T|}{|\Gamma|}\times\frac{1}{(2\pi)^{d/2}|\Sigma_T\Gamma^{-1}\Sigma_{\tau|T}(\Gamma')^{-1}\Sigma_T+\Sigma_{T|\tau}|^{1/2}} \\
&\exp\Big(-\frac{1}{2}\left(\Sigma_T\Gamma^{-1}(y-\mu_\tau)-\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau)\right)' \\
&\qquad\qquad \times\left(\Sigma_T\Gamma^{-1}\Sigma_{\tau|T}(\Gamma')^{-1}\Sigma_T+\Sigma_{T|\tau}\right)^{-1}\left(\Sigma_T\Gamma^{-1}(y-\mu_\tau)-\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau)\right)\Big) \\
&= \frac{1}{(2\pi)^{d/2}|\Gamma\Sigma_T^{-1}\left(\Sigma_T\Gamma^{-1}\Sigma_{\tau|T}(\Gamma')^{-1}\Sigma_T+\Sigma_{T|\tau}\right)\Sigma_T^{-1}\Gamma'|^{1/2}} \\
&\exp\Big(-\frac{1}{2}\left(y-\mu_\tau-\Gamma\Sigma_T^{-1}\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau)\right)'(\Gamma^{-1})'\Sigma_T\left(\Sigma_T\Gamma^{-1}\Sigma_{\tau|T}(\Gamma')^{-1}\Sigma_T+\Sigma_{T|\tau}\right)^{-1} \\
&\qquad\qquad \times\Sigma_T\Gamma^{-1}\left(y-\mu_\tau-\Gamma\Sigma_T^{-1}\Gamma'\Sigma_\tau^{-1}(x-\mu_\tau)\right)\Big) \\
&= g(y;\mu_\tau+\underbrace{\Gamma\Sigma_T^{-1}\Gamma'\Sigma_\tau^{-1}}_{A}(x-\mu_\tau),\ \Sigma_\tau-\Gamma\Sigma_T^{-1}\Gamma'\Sigma_\tau^{-1}\Gamma\Sigma_T^{-1}\Gamma') \\
&= g(y;\underbrace{\mu_\tau+A(x-\mu_\tau)}_{\mu_A(x)},\ \underbrace{\Sigma_\tau-A\Sigma_\tau A'}_{\Sigma_A}) \\
&= g(y;\mu_A(x),\Sigma_A).
\end{aligned}
$$

$L^*L$ can be calculated by similar method. □

***Proof of Lemma 1.4.3.*** We start by recalling the considerations from Khare and Zhou (2009): Let $(X_t)$ on $\mathbb{R}^d$ be a MAR(1) process satisfying the following stochastic difference equation:

$$X_t = \Phi X_{t-1} + \eta_t, \quad t \geq 1, \tag{17}$$

where $\Phi \in \mathbb{R}^{d\times d}$ and $(\eta_t)_{t\geq 1}$ are independent and identically distributed, $\eta_1 \sim N(0,H)$. $(X_t)$ has a unique stationary distribution $N(0,\Sigma)$ if and only if $H = \Sigma - \Phi\Sigma\Phi'$, and the process is reversible if and only if $\Phi\Sigma = \Sigma\Phi'$. Khare and Zhou (2009) show that if these assumptions are satisfied, the transformed Markov operator for (17) has eigenvalues which are products of eigenvalues of $\Phi$ and the corresponding eigenfunctions

are products of Hermite polynomials.

Now note that under for a random variable $Y$ that is distributed according to $K_A(x\cdot)$, we can write:

$$Y - \mu_\tau = A(x - \mu_\tau) + \zeta_A, \tag{18}$$

where $\zeta \sim N(0, \Sigma_A)$. Since from Lemma 1.4.2 we have that $\Sigma_A = \Sigma_\tau - A\Sigma_\tau A'$ and

$$A\Sigma_\tau = \Gamma\Sigma_T^{-1}\Gamma' = \Sigma_\tau\, A',$$

the operator $L\,L^*$ has the same structure of the Markov operator for (17) that is reversible and stationary.

Following the approach by Khare and Zhou (2009), denote by $\Sigma_\tau^{1/2}$ the square root matrix of $\Sigma_\tau$. Then

$$\Sigma_\tau^{-1/2} A \Sigma_\tau^{1/2} = \Sigma_\tau^{-1/2}\,\Gamma\Sigma_T^{-1}\Gamma'\Sigma_\tau^{-1/2}$$

is symmetric and thus orthogonally diagonalizable:

$$\Sigma_\tau^{-1/2} A \Sigma_\tau^{1/2} = P\Lambda P' \;\Leftrightarrow\; A = (\Sigma_\tau^{1/2}\,P)\,\Lambda\,(P'\Sigma_\tau^{-1/2}).$$

In particular, the entries of the diagonal matrix $\Lambda$ are the eigenvalues of $A$.

Now for the transformation (1.11) of the random vector $Y$ from (18), $z^P(Y)$, we obtain:

$$\begin{aligned}
\mathbb{E}_{K_A}\left[z^P(Y)|x\right] &= P'\Sigma_\tau^{-1/2}A(x-\mu_\tau) \\
&= P'\Sigma_\tau^{-1/2}\Sigma_\tau^{1/2}P\Lambda P'\Sigma_\tau^{-1/2}(x-\mu_\tau) = \Lambda z^P(x),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}_{K_A}\left[z^P(Y)|x\right] &= P'\Sigma_\tau^{-1/2}\Sigma_A\Sigma_\tau^{-1/2}P \\
&= P'\Sigma_\tau^{-1/2}(\Sigma_\tau - A\Sigma_\tau A')\Sigma_\tau^{-1/2}P = I - \Lambda^2.
\end{aligned}$$

Moreover,

$$\mathbb{E}_{\pi_{Y_\tau}}\left[z^P(Y_\tau)\right] = P'\Sigma_\tau^{-1/2}\mathbb{E}_{\pi_{Y_\tau}}\left[Y_\tau - \mu_\tau\right] = 0$$

and

$$\mathrm{Var}_{\pi_{Y_\tau}}\left[z^P(Y_\tau)\right] = P'\Sigma_\tau^{-1/2}\Sigma_\tau\Sigma_\tau^{-1/2}P = I.$$

The second part follows analogously. □

***Proof of Proposition 1.4.3.*** For fixed $z_i^P(Y)$, we obtain from Carrasco and Florens (2011) that the univariate orthonormal Hermite polynomial of order $n_i$ is the eigenfunction under $K_A$ :

$$\mathbb{E}_{K_A}\left[h_{n_i}(z_i^P(Y))|x\right] = \lambda_i^{n_i}h_{n_i}(z_i^P(x)).$$

Moreover, the product of these orthonormal polynomials are also eigenfunction since:

$$\mathbb{E}_{K_A}\left[\Pi_{i=1}^d h_{n_i}(z_i^P(Y))|x\right] = \Pi_{i=1}^d\mathbb{E}_{K_A}\left[h_{n_i}(z_i^P(Y))|x\right] = \left(\Pi_{i=1}^d\lambda_i^{n_i}\right)\left(\Pi_{i=1}^d h_{n_i}(z_i^P(x))\right).$$

The orthogonality of the eigenfunctions is proved in Khare and Zhou (2009). Note that the product of normalized Hermite polynomials is already normalized since:

$$\mathbb{E}_{\pi_{Y_\tau}}\left[\left(\Pi_{i=1}^d h_{n_i}(z_i^P(Y))\right)^2\right] = \mathbb{E}_{\pi_{Y_\tau}}\left[\Pi_{i=1}^d h_{n_i}\left(z_i^P\right)^2(Y)\right] = \Pi_{i=1}^d\mathbb{E}_{\pi_{Y_\tau}}\left[h_{n_i}\left(z_i^P(Y)\right)^2\right] = 1.$$

Right singular functions are obtained similarly with fixed $z_i^Q(X)$.

□

***Proof of Lemma 1.5.1.*** Under $\mathbb{P}$, we have:

$$r_\tau = r_0e^{-\alpha\tau} + \gamma(1 - e^{-\alpha\tau}) + \sigma\int_0^\tau e^{-\alpha(\tau-t)}dW_t,$$

so that $r_\tau \sim N(\mu_\tau, \sigma_\tau^2)$ with $\mu_\tau = \gamma - (\gamma - r_0)e^{-\alpha\tau}$ and $\sigma_\tau^2 = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\tau})$.

Under $\mathbb{Q}_T$, we have:

$$r_T = r_\tau e^{-\alpha(T-\tau)} + M(\tau, T) + \sigma \int_\tau^T e^{-\alpha(T-t)} dZ_t^T,$$

where $M(\tau,T) = (\bar{\gamma} - \frac{\sigma^2}{\alpha^2})(1 - e^{-\alpha(T-\tau)}) + \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(T-\tau)})$, so that $r_T|r_\tau \sim N(\mu_{r_T|r_\tau}, \sigma^2_{r_T|r_\tau})$ with $\mu_{r_T|r_\tau} = r_\tau e^{-\alpha(T-\tau)} + M(\tau,T)$ and $\sigma^2_{r_{r_T|r_\tau}} = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(T-\tau)})$. Note that this distribution specifies the transition density of $r_T$ given $r_\tau$. The unconditional mean and variance of $r_T$ is given by:

$$\begin{aligned}\mathbb{E}^{\tilde{\mathbb{P}}}[r_T] &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}_T}[r_T|r_\tau]\right] = \mathbb{E}^{\mathbb{P}}\left[r_\tau e^{-\alpha(T-\tau)} + M(\tau,T)\right] \\ &= \mu_{r_\tau} e^{-\alpha(T-\tau)} + M(\tau,T) = \mu_T\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}^{\tilde{\mathbb{P}}}[r_T] &= \mathbb{E}^{\mathbb{P}}\left[\mathrm{Var}^{\mathbb{Q}_T}[r_T|r_\tau]\right] + \mathrm{Var}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}_T}[r_T|r_\tau]\right] \\ &= \mathbb{E}^{\mathbb{P}}\left[\frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(T-\tau)})\right] + \mathrm{Var}^{\mathbb{P}}\left[r_\tau e^{-\alpha(T-\tau)} + M(\tau,T)\right] \\ &= \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha(T-\tau)}) + \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha\tau})e^{-2\alpha(T-\tau)} = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha T}) = \sigma_T^2,\end{aligned}$$

so that $r_T \sim N(\mu_T, \sigma_T^2)$. Moreover:

$$\begin{aligned}\mathrm{Cov}(r_\tau, r_T) &= \mathbb{E}^{\tilde{\mathbb{P}}}(r_T \cdot r_\tau) - \mu_\tau \mu_T = \mathbb{E}^{\tilde{\mathbb{P}}}\left[\mathbb{E}^{\tilde{\mathbb{P}}}[r_T \cdot r_\tau | r_\tau]\right] - \mu_\tau \mu_T \\ &= e^{-\alpha(T-\tau)}\sigma_\tau^2.\end{aligned}$$

Thus, we have for the joint distribution of $r_\tau$ and $r_T$:

$$\begin{bmatrix} r_\tau \\ r_T \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_\tau \\ \mu_T \end{bmatrix}, \begin{bmatrix} \sigma_\tau^2, & e^{-\alpha(T-\tau)}\sigma_\tau^2 \\ e^{-\alpha(T-\tau)}\sigma_\tau^2, & \sigma_T^2 \end{bmatrix}\right),$$

with $\rho = \rho_{r_\tau, r_T} = \mathrm{Corr}(r_\tau, r_T) = e^{-\alpha(T-\tau)}\frac{1-e^{-2\alpha\tau}}{1-e^{-2\alpha T}}$. □

***Proof of Lemma 1.5.2.*** Under $\mathbb{P}$, the solutions of (1.22), (1.23), and (1.24) at time $\tau$ are:

$$\begin{aligned} q_\tau &= q_0 + \left(m - \frac{1}{2}\sigma_S^2\right)\tau + \sigma_S \int_0^\tau dW_s^S, \\ r_\tau &= r_0 e^{-\alpha\tau} + \gamma\left(1 - e^{-\alpha\tau}\right) + \sigma_r \int_0^\tau e^{-\alpha(\tau-t)} dW_t^r, \\ \mu_{x+\tau} &= \mu_x e^{\kappa\tau} + \psi \int_0^\tau e^{\kappa(\tau-u)} dW_u^\mu. \end{aligned}$$

Thus, the joint Gaussian distribution of $Y_\tau$ is given by:

$$\begin{bmatrix} q_\tau \\ r_\tau \\ \mu_{x+\tau} \end{bmatrix} \sim N\left( \begin{bmatrix} q_0 + \left(m - \frac{1}{2}\sigma_S^2\right)\tau \\ r_0 e^{-\alpha\tau} + \gamma\left(1 - e^{-\alpha\tau}\right) \\ \mu_x e^{\kappa\tau} \end{bmatrix}, \begin{bmatrix} \sigma_S^2\,\tau, & \frac{\rho_{12}\sigma_S\sigma_r(1-e^{-\alpha\tau})}{\alpha} & 0 \\ \frac{\rho_{12}\sigma_S\sigma_r(1-e^{-\alpha\tau})}{\alpha} & \frac{\sigma_r^2(1-e^{-2\alpha\tau})}{2\alpha} & 0 \\ 0 & 0 & \frac{\psi^2(e^{2\kappa\tau}-1)}{2\kappa} \end{bmatrix} \right), \tag{19}$$

so that $\mu_\tau$ and $\Sigma_\tau$ are given by

$$\mu_\tau = \begin{bmatrix} q_0 + \left(m - \frac{1}{2}\sigma_S^2\right)\tau \\ r_0 e^{-\alpha\tau} + \gamma\left(1 - e^{-\alpha\tau}\right) \\ \mu_x e^{\kappa\tau} \end{bmatrix}, \quad \Sigma_\tau = \begin{bmatrix} \sigma_S^2\,\tau, & \frac{\rho_{12}\sigma_S\sigma_r(1-e^{-\alpha\tau})}{\alpha} & 0 \\ \frac{\rho_{12}\sigma_S\sigma_r(1-e^{-\alpha\tau})}{\alpha} & \frac{\sigma_r^2(1-e^{-2\alpha\tau})}{2\alpha} & 0 \\ 0 & 0 & \frac{\psi^2(e^{2\kappa\tau}-1)}{2\kappa} \end{bmatrix}.$$

To derive the distribution under $\mathbb{Q}_E$, first note that for $\tau \le s < T$,

$$r_s = e^{-\alpha(s-\tau)} r_\tau + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)(1 - e^{-\alpha(s-\tau)}) + \frac{\sigma_r^2}{2\alpha^2}\left(e^{-\alpha(T-s)} - e^{-\alpha(T+s-2\tau)}\right) + \sigma_r \int_\tau^s e^{-\alpha(s-y)} dW_y^r,$$

so that the integral of $\int_\tau^T r s\, ds$ can be evaluated using the stochastic Fubini's theorem:

$$\begin{aligned} \int_\tau^T r_s\, ds &= \frac{e^{\alpha\tau} r_\tau}{\alpha}(e^{-\alpha\tau} - e^{-\alpha T}) + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(T - \tau + e^{\alpha\tau}\frac{e^{-\alpha T} - e^{-\alpha\tau}}{\alpha}\right) \\ &\quad + \frac{\sigma_r^2}{2\alpha^2}\left(e^{-\alpha T}\frac{e^{\alpha T} - e^{\alpha\tau}}{\alpha} + e^{-\alpha T + 2\alpha\tau}\frac{e^{-\alpha T} - e^{-\alpha\tau}}{\alpha}\right) + \sigma_r \int_\tau^T \int_\tau^s e^{-\alpha(s-y)} dW_y^r\, ds \\ &= \frac{e^{\alpha\tau} r_\tau}{\alpha}(e^{-\alpha\tau} - e^{-\alpha T}) + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(T - \tau + e^{\alpha\tau}\frac{e^{-\alpha T} - e^{-\alpha\tau}}{\alpha}\right) \\ &\quad + \frac{\sigma_r^2}{2\alpha^2}\left(e^{-\alpha T}\frac{e^{\alpha T} - e^{\alpha\tau}}{\alpha} + e^{-\alpha T + 2\alpha\tau}\frac{e^{-\alpha T} - e^{-\alpha\tau}}{\alpha}\right) + \frac{\sigma_r}{\alpha}\int_\tau^T 1 - e^{-\alpha T + \alpha y} dW_y^r. \end{aligned}$$

Thus, under $\mathbb{Q}_E$ with known $Y_\tau$, the solutions of (1.27), (1.28), and (1.29) are:

$$
\begin{aligned}
q_T &= q_\tau + \frac{1-e^{-\alpha(T-\tau)}}{\alpha} r_\tau + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(T-\tau+\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) \\
&\quad + \frac{\sigma_r^2}{2\alpha^2}\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha} - \frac{e^{-\alpha(T-\tau)}-e^{-2\alpha(T-\tau)}}{\alpha}\right) + \frac{\sigma_r}{\alpha}\int_\tau^T 1-e^{-\alpha(T-t)} dZ_t^r \\
&\quad - \frac{\sigma_S^2}{2}(T-\tau) - \frac{\rho_{12}\sigma_S\sigma_r}{\alpha}\left[T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right] + \sigma_S\int_\tau^T dZ_s^S, \\
r_T &= e^{-\alpha(T-\tau)} r_\tau + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(1-e^{-\alpha(T-\tau)}\right) + \frac{\sigma_r^2}{2\alpha^2}\left(1-e^{-2\alpha(T-\tau)}\right) + \sigma_r\int_\tau^T e^{-\alpha(T-t)} dZ_t^r, \\
\mu_{x+T} &= e^{\kappa(T-\tau)}\mu_{x+\tau} - \frac{\psi^2}{\kappa^2}\left(1-e^{\kappa(T-\tau)}\right) + \frac{\psi^2}{2\kappa^2}\left(1-e^{2\kappa(T-\tau)}\right) + \psi\int_\tau^T e^{\kappa(T-u)} dZ_u^\mu,
\end{aligned}
$$

so that the (Gaussian) conditional distribution of $Y_T|Y_\tau$ is given by:

$$
\begin{bmatrix} q_\tau \\ r_\tau \\ \mu_{x+\tau} \end{bmatrix} |Y_\tau \sim \; N\left(\begin{bmatrix} \mu_{q_T}|q_\tau \\ \mu_{r_T}|r_\tau \\ \mu_{\mu_{x+T}}|\mu_{x+\tau} \end{bmatrix}, \underbrace{\begin{bmatrix} \sigma^2_{q_T|q_\tau}, & \sigma_{q_T,r_T|q_\tau,r_\tau} & 0 \\ \sigma_{q_T,r_T|q_\tau,r_\tau} & \sigma^2_{r_T|r_\tau} & 0 \\ 0 & 0 & \sigma^2_{\mu_{x+T}|\mu_{x+\tau}} \end{bmatrix}}_{G}\right), \tag{20}
$$

where

$$
\begin{aligned}
q_T|q_\tau &= q_\tau + r_\tau\frac{1-e^{-\alpha(T-\tau)}}{\alpha} + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(T-\tau+\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) \\
&\quad + \frac{\sigma_r^2}{2\alpha^2}\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha} - \frac{e^{-\alpha(T-\tau)}-e^{-2\alpha(T-\tau)}}{\alpha}\right) \\
&\quad - \frac{\sigma_S^2}{2}(T-\tau) - \frac{\rho_{12}\sigma_S\sigma_r}{\alpha}\left[T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right], \\
r_T|r_\tau &= e^{-\alpha(T-\tau)} r_\tau + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(1-e^{-\alpha(T-\tau)}\right) + \frac{\sigma_r^2}{2\alpha^2}\left(1-e^{-2\alpha(T-\tau)}\right), \\
\mu_{x+T}|\mu_{x+\tau} &= e^{\kappa(T-\tau)}\mu_{x+\tau} - \frac{\psi^2}{\kappa^2}\left(1-e^{\kappa(T-\tau)}\right) + \frac{\psi^2}{2\kappa^2}\left(1-e^{2\kappa(T-\tau)}\right),
\end{aligned}
$$

and

$$\begin{aligned}
\sigma^2_{q_T|q_\tau} &= \left(\frac{\sigma_r}{\alpha}\right)^2 \left[T-\tau-\frac{2\left(1-e^{-\alpha(T-\tau)}\right)}{\alpha}+\frac{1-e^{-2\alpha(T-\tau)}}{2\alpha}\right] \\
&\quad + \sigma_S^2(T-\tau)+\frac{2\sigma_r\sigma_S\rho_{12}}{\alpha}\left[T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right], \\
\sigma^2_{r_T|r_\tau} &= \frac{\sigma_r^2}{2\alpha}\left(1-e^{-2\alpha(T-\tau)}\right), \\
\sigma_{q_T,r_T|q_\tau,r_\tau} &= \left(\frac{\sigma_r^2}{\alpha}+\sigma_S\sigma_r\rho_{12}\right)\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)-\frac{\sigma_r^2}{\alpha}\left(\frac{1-e^{-2\alpha(T-\tau)}}{2\alpha}\right), \\
\sigma^2_{\mu_{x+T}|\mu_{x+\tau}} &= \frac{\psi^2}{2\kappa}\left(e^{2\kappa(T-\tau)}-1\right).
\end{aligned}$$

It is possible to write the conditional mean of $Y_T$ given $Y_\tau$ in the following linear form:

$$\begin{aligned}
\begin{bmatrix} \mu_{q_T}|q_\tau \\ \mu_{r_T}|r_\tau \\ \mu_{\mu_{x+T}}|\mu_{x+\tau} \end{bmatrix} &= \underbrace{\begin{bmatrix} 1 & \frac{1-e^{-\alpha(T-\tau)}}{\alpha} & 0 \\ 0 & e^{-\alpha(T-\tau)} & 0 \\ 0 & 0 & e^{\kappa(T-\tau)} \end{bmatrix}}_{H} \begin{bmatrix} q_\tau \\ r_\tau \\ \mu_{x+\tau} \end{bmatrix} + C_\tau \\
&= HY_\tau + C_\tau
\end{aligned}$$

where $C_\tau$ is a constant matrix defined by remaining terms of mean vector of $Y_T|Y_\tau$ after defining $HY_\tau$. The unconditional distribution of $Y_T$ under $\tilde{\mathbb{P}}$ is also Gaussian since $Y_\tau$ and $Y_T|Y_\tau$ follow Gaussian distributions. Thus, it suffices to specify a mean vector and a covariance matrix of $Y_T$ under $\tilde{\mathbb{P}}$ to specify its distribution:

$$\begin{aligned}
\mathbb{E}^{\tilde{\mathbb{P}}}[Y_T] &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}_E}[Y_T|Y_\tau]\right] = \mathbb{E}^{\mathbb{P}}\left[HY_\tau + C_\tau\right] = H\mu_\tau + C_\tau = \mu_T, \\
\mathrm{Var}^{\tilde{\mathbb{P}}}[Y_T] &= \mathrm{Var}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}_E}[Y_T|Y_\tau]\right] + \mathbb{E}^{\mathbb{P}}\left[\mathrm{Var}^{\mathbb{Q}_E}[Y_T|Y_\tau]\right] \\
&= \mathrm{Var}^{\mathbb{P}}\left[HY_\tau + C_{Y_\tau}\right] + \mathbb{E}^{\mathbb{P}}\left[G\right] = H\Sigma_\tau H' + G = \Sigma_T.
\end{aligned}$$

Hence, $Y_T \sim N(\mu_T,\ \Sigma_T)$.

The final step is to specify the joint distribution of $Y_\tau$ and $Y_T$ by finding $\mathrm{Cov}(Y_\tau, Y_T)$. Note that

$$\begin{aligned}
\mathrm{Cov}(Y_\tau, Y_T) &= \mathbb{E}^{\tilde{\mathbb{P}}}[Y_\tau Y_T'] - \mathbb{E}^{\tilde{\mathbb{P}}}[Y_\tau]\mathbb{E}^{\tilde{\mathbb{P}}}[Y_T'] \\
&= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{Q}_E}[Y_\tau Y_T'|Y_\tau]] - \mu_\tau \mu_T' \\
&= \mathbb{E}^{\mathbb{P}}\left[Y_\tau(Y_\tau' H' + C_\tau')\right] - \mu_\tau \mu_T' \\
&= \Sigma_\tau H' = \Gamma.
\end{aligned}$$

Therefore,

$$\begin{bmatrix} Y_\tau \\ Y_T \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_\tau \\ \mu_T \end{bmatrix}, \quad \begin{bmatrix} \Sigma_\tau & \Gamma \\ \Gamma' & \Sigma_T \end{bmatrix} \right).$$

□

## A.2 Simulation Scheme

In this section, we provide simulation algorithms of the regression-now and regression-later approaches. All notations are defined in the main text.

### A.2.1 Regression-now algorithm

- Initiate state variables, $X_{t_0}^{+}$, $G_{t_1}$, $r_{t_0}$, $\nu_{t_0}$, and $\mu_x(t_0)$.
- For $(t_{i-1}, t_i]$ , $i = 1, 2, ..., n$:
  1. Divide $(t_{i-1}, t_i]$ into $m$ sub-intervals such that $t_{i-1} = t_{i-1,0} < t_{i-1,2} < \cdots < t_{i-1,m} = t_i$.
  2. Generate $Y_{t_{i-1},j}^{l}$, $l = 1, 2, ..., N$, $j = 1, ..., m$ where the superscript denotes $l$th simulation and $N$ is the total number of simulations.
  3. Calculate $X_{t_{i-1},j}^{-,l}$, $l = 1, 2, ..., N$, $j = 1, 2, ..., m$.
  4. Generate random $w_{t_i}^{l}$ from $A_{t_i}^{l}$ not allowing for surrenders or setting $G_{t_{i+1}}^{l} = 0$.
  5. Update $X_{t_i}^{+,l}$ and $G_{t_{i+1}}^{l}$, $l = 1, 2, ..., N$.
  6. Drop $w_{t_n}^{l}$, $l = 1, 2, ..., N$.
- Set $\hat{V}_{t_n}^{l,now}\left(Y_{t_n}^{l}\right) = \max(X_{t_n}^{-,l}, \min(g, G_{t_n}^{l}))$ and $\hat{D}_{t_n}(Y_{t_n}^{l}) = \max(X_{t_n}^{-,l}, G_{t_n}^{l})$.
- For $i = n-1, ..., 1$:
  1. Calculate $y_{t_i}^{l} = df_{1,t_i}^{l}\left[df_{2,t_i}^{l} V_{t_{i+1}}(Y_{t_{i+1}}^{l}) + \left(1 - df_{2,t_i}^{l}\right)\hat{D}_{t_{i+1}}(Y_{t_{i+1}}^{l})\right]$, where

$$df_{1,t_i}^{l} = \exp\left(-\sum_{j=0}^{m-1} r_{t_i,j}^{l}\,\Delta_j^{i}\right),$$

$$df_{2,t_i}^{l} = \exp\left(-\sum_{j=0}^{m-1} \mu_{x+t_i,j}^{l}\,\Delta_j^{i}\right),$$

and

$$\Delta_j^i = t_{i,j+1} - t_{i,j}, l = 1, ..., N.$$

2. Solve

$$\hat{\alpha}^{t_i} = \text{argmin}_{\{\alpha^{t_i}\}} \sum_{l=1}^{N} \left[ e\left(Y_{t_i}^l\right) \cdot \alpha^{t_i} - y_{t_i}^l \right]^2.$$

3. Solve

$$\hat{V}_{t_i}^{l,now}\left(Y_{t_i}^l\right) = \max_{w_{t_i}^l \in \mathcal{A}_{t_i}^l} C(t_i, w_{t_i}) + \left(e\left(Y_{t_i}\right)\hat{\alpha}^{t_i}\right).$$

4. Set $\hat{D}_{t_i}^l(Y_{t_i}^l) = \max(X_{t_i}^{-,l}, G_{t_i}^l)$.

- The price of GMWB at $t = 0$ is:

$$\hat{V}_{now}(0) = \frac{1}{N} \sum_{l=1}^{N} df_{1,t_0}^l \left( df_{2,t_0}^l \hat{V}_{t_i}^{l,now}(Y_{t_1}^l) + (1 - df_{2,t_0}^l)\hat{D}_{t_i}^l(Y_{t_1}^l) \right).$$

### A.2.2 Regression-later algorithm in the Black-Scholes framework with deterministic survival probability

- Initiate state variables, $X_{t_0}^+$, $G_{t_1}$, $r_{t_0}$, $\nu_{t_0}$, and $\mu_x(t_0)$.

- For each $t_i$ , $i = 1, 2, ..., n$:

1. Generate $N$ independent standard normal random variable $Z_l$, $l = 1, ..., N$.
2. Calculate $X_{t_i}^{-,l} = X_{t_{i-1}}^{+,l} \exp\left(\left(r - \phi - \frac{1}{2}\sigma^2\right)(t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}} Z_l\right)$.
3. Generate random $w_{t_i}^l$ from $A_{t_i}^l$ not allowing for surrenders or setting $G_{t_{i+1}}^l = 0$.
4. Update $X_{t_i}^{+,l}$ and $G_{t_{i+1}}^l$, $l = 1, 2, ..., N$.
5. Drop $w_{t_n}^l$, $l = 1, 2, ..., N$.

- Set $\hat{V}_{t_n}^{l,later}\left(Y_{t_n}^l\right) = \max(X_{t_n}^{-,l}, \min(g, G_{t_n}^l))$ and $\hat{D}_{t_n}(Y_{t_n}^l) = \max(X_{t_n}^{-,l}, G_{t_n}^l)$.

- For $i = n-1, ..., 1$:

  1. Calculate $F^l_{t_{i+1}} = e^{-r(t_{i+1}-t_i)} \left[ e^{-\mu(t_{i+1}-t_i)} V_{t_{i+1}}(Y^l_{t_{i+1}}) + \left(1 - e^{-\mu(t_{i+1}-t_i)}\right) \hat{D}_{t_{i+1}}(Y^l_{t_{i+1}}) \right]$, $l = 1, ..., N$.

  2. Solve
  $$\hat{\beta}^{t_{i+1}} = \text{argmin}_{\{\beta^{t_{i+1}}\}} \sum_{l=1}^{N} \left[ \varphi\left(Y^l_{t_{i+1}}\right) \cdot \beta^{t_{i+1}} - F^l_{t_{i+1}} \right]^2.$$

  3. Solve
  $$\hat{V}^{l,later}_{t_i}\left(Y^l_{t_i}\right) = \max_{w^l_{t_i} \in \mathcal{A}^l_{t_i}} C(t_i, w_{t_i}) + \mathbb{E}^{\mathbb{Q}} \left[ \left( \varphi\left(Y_{t_{i+1}}\right) \hat{\beta}^{t_{i+1}} \right) | Y_{t_i} \right]. \tag{21}$$

  4. Set $\hat{D}^l_{t_i}(Y^l_{t_i}) = \max(X^{-,l}_{t_i}, G^l_{t_i})$

- The price of GMWB at $t = 0$ is:

$$\hat{V}_{later}(0) = \frac{1}{N} \sum_{l=1}^{N} e^{-r(t_1-t_0)} \left( e^{-\mu(t_1-t_0)} \hat{V}^{l,later}_{t_1}(Y^l_{t_1}) + \left(1 - e^{-\mu(t_1-t_0)}\right) \hat{D}^l_{t_1}(Y^l_{t_1}) \right).$$

# Bibliography

Amemiya, T., 1985. Advanced econometrics. Harvard university press, Cambridge.

Andersen, L.B.G., Piterbarg, V., 2007. Moment Explosions in Stochastic Volatility Models. Finance and Stochastics, 11: 29-50.

Andrews, D.W.K., Whang, Y.J., 1990. Additive interactive regression models: circumvention of the curse of dimensionality. Econometric Theory 6: 466-479.

Azimzadeh, P., Forsyth, P., 2015. The existence of optimal bang-bang controls of GMxB contracts. SIAM Journal on Financial Mathematics, 6: 117-139.

Bacinello, A. R., Biffis, E., Enrico B., Milossovich, P., 2010. Regression-based algorithms for life insurance contracts with surrender guarantees. Quantitative Finance, 10: 1077-1090.

Bacinello, A.R., Milossovich, P., Montealegre A., 2013. A Dynamic Programming Algorithm for the Valuation of Guaranteed Minimum Withdrawal Benefits in Variable Annuities. Research Paper Series, N. 4, 2013, Bruno de Finetti.

Bacinello, A.R., Milossovich, P., Olivieri, A., Pitacco, E., 2011. Variable Annuities: A Unifying Approach. Insurance: Mathematics and Economics, 49: 285-297.

Barrie and Hibbert, 2011. A Least Squares Monte Carlo Approach to Liability Proxy Modeling and Capital Calculation. Available at `http://www.barrhibb.com`.

Bauer, D., Bergmann, D., Kiesel, R., 2010. On the Risk-Neutral Valuation of Life Insurance Contracts with Numerical Methods in View. ASTIN Bulletin, 40: 31-61.

Bauer, D., Bergmann, D., Reuss, A., 2009. Solvency II and Nested Simulations – a Least-Squares Monte Carlo Approach. Working paper, Georgia State University and Ulm University.

Bauer, D., Gao, J., Moenig, T., Ulm, E.R., Zhu, N., 2015. Policyholder Exercise Behavior in Life Insurance: The State of Affairs. Working Paper, Georgia State University.

Bauer, D., Kling, A., Russ, A., 2008. A Universal Pricing Framework for Guaranteed Minimum Benefits in Variable Annuities. ASTIN Bulletin, 38: 621-651.

Bauer, D., Reuss, A., Singer, D., 2012. On the Calculation of the Solvency Capital Requirement based on Nested Simulations. ASTIN Bulletin, 42: 453-499.

Benedetti, G., 2016. A note on the calculation of risk measures using Least Squares Monte Carlo. Working Paper.

Beutner, E., Pelsser, A., Schweizer, J., 2013. Fast Convergence of Regress-Later Estimates in Least Squares Monte Carlo. Available at SSRN: http://ssrn.com/abstract=2328709 or http://dx.doi.org/10.2139/ssrn.2328709.

Beutner, E., Pelsser, A., Schweizer, J., 2016. Theory and validation of replicating portfolios in insurance risk management. Working Paper, Maastricht University.

Biagini, F., Frittelli, M., 2009. On the extension of the Namioka-Klee theorem and on the Fatou property of risk measures. In Optimality and risk: modern trends in mathematical finance – The Kebanov Festschrift (Eds. F. Delbaen, M. Rasonyi, and Ch. Stricker), 1-29.

Birman, M.S., Solomyak, M.Z., 1977. Estimates of singular numbers of integral operators. Russian Mathematical Surveys, 32: 15-89.

Björk, T., 2009. Arbitrage Theory in Continuous Time, Third Edition. Oxford.

Boyle, P., Hardy, M., 2003. Guaranteed Annuity Options. ASTIN Bulletin, 33: 125-152.

Bravo, J.,M., 2011. Pricing Longevity Bonds Using Affine-Jump Diffusion Models. Working Paper.

Brigo, D., Mercurio, F., 2006. Interest rate models – theory and practice. Springer.

Brodie, M., Du, Y., Moallemi, C.C, 2015. Risk Estimation via Regression. Operations Research, 63: 1077-1097.

Cairns, A.J.G., David, B., Kevin, D., 2008. Modeling and management of mortality risk: a review. Scandinavian Actuarial Journal, 2008: 79-113.

Cambou, M., Filipović, D., 2016. Replicating Portfolio Approach to Capital Calculation. Working Paper, EPFL Lausanne.

Carrasco M., Florens J., 2011. Spectral Method for Deconvolving a Density. Econometric Theory, 27: 546-581.

Carriere, J.F., 1996. Valuation of the early-exercise price for options using simulations and nonparametric regression. Insurance: Mathematics and Economics, 19: 19-30.

Chen, Z., Ken V., Peter A. F., 2008. The effect of modelling parameters on the value of GMWB guarantees. Insurance: Mathematics and Economics, 43: 165-173.

Clément, E., Lamberton, D., Protter, P., 2002. An analysis of a least squares regression method for american option pricing. Finance and Stochastics, 6: 449-471.

Cohn, D., 2013. Measure Theory. Birkhäuser.

Cox, J.C., Ingersoll J.E., Ross S.A. , 1985. A Theory of the Term Structure of Interest Rates. Econometrica, 53: 385-407.

Dai, M., Yue K. K., Jianping Z., 2008. Guaranteed minimum withdrawal benefit in variable annuities. Mathematical Finance, 18: 595-611.

DAV, 2015. Proxy-Modelle für die Risikokapitalberechnung. Technical Report, Ausschuss Investment der Deutschen Aktuarvereinigung, e.V. (DAV).

Durett, R., 1996. Probability: theory and examples, Second Edition. Duxbury Press, Belmont, CA.

Floryszczak A., Le Courtois O., Majri M., 2011. Inside the Solvency 2 Black Box: Net Asset Values and Solvency Capital Requirements with a Least-Squares Monte-Carlo Approach. Working Paper, EM Lyon.

Glasserman, P., 2004. Monte Carlo Methods in Financial Engineering. Springer, New York.

Glasserman, P., Yu, B., 2002. Simulation for American options: regression now or regression later? In In Monte Carlo and Quasi-Monte Carlo Methods, 213-226, Springer, Berlin.

Gordy, M.B., Juneja, D., 2010. Nested simulations in portfolio risk measurement. Management Science, 56: 1833-1848.

Ha, H., Bauer., D., 2016. A Least-Squares Monte Carlo Evaluation of Withdrawal Benefits in Variable Annuities. Working Paper, Georgia State University.

Heston, Steven L. , 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options Review of Financial Studies, 6: 327-343.

Huang, Q., 2012. Some Topics Concerning the Singular Value Decomposition and Generalized Singular Value Decomposition. Ph.D. thesis, Arizona State University

Huang Y.T., Kwok, Y.K., 2016. Regression-based Monte Carlo methods for stochastic control models: Variable annuities with lifelong guarantees. Quantitative Finance, 16: 905-928.

Jamshidian, F., 1989. An Exact Bond Option Formula. Journal of Finance, 44: 205-209

Judd, K., 1998. Numerical Methods in Economics. MIT press, Cambridge.

Kaina, M., Rüschendorf, L., 2009. On convex risk measures on $L^p$-spaces. Mathematical Methods of Operations Research, 69: 475-495.

Khare, K., Zhou, H., 2009. Rate of Convergence of Some Multivariate Markov Chains with Polynomial Eigenfunctions. The Annals of Applied Probability, 19: 737-777.

Kollo, T., Rosen, D.V., 2006. Advanced multivariate statistics with matrices. Springer, New York.

Kreyszig, E., 1989. Introductory Functional Analysis with Applications 1st Edition. Wiley, Chichester.

Lee, S.-H., 1998. Monte Carlo computation of conditional expectation quantities. Ph.D. thesis, Stanford University.

Liebwein, P., 2006. Risk Models for Capital Adequacy: Applications in the Context of Solvency II and Beyond. The Geneva Papers on Risk and Insurance: Issues and Practice, 31: 528-550.

Longstaff, F.A., Schwartz, E.S., 2001. Valuing American Options by Simulation: A Simple Least-Squares Approach. Review of Financial Studies, 14: 113-147.

Milliman, 2013. Least Squares Monte Carlo for fast and robust Capital projections. Available at `http://us.milliman.com`.

Moenig T., Bauer, D., 2016. Revisiting the risk-neutral approach to optimal policyholder behavior: a study of withdrawal guarantees in variable annuities. Review of Finance, 20: 759-794.

Moreno, M., Navas, J.F., 2003. On the Robustness of Least-Squares Monte Carlo (LSM) for Pricing American Derivatives. Review of Derivatives Research, 14: 113-147.

Nadarajah S., Margot, F., Secomandi, N., 2012. Valuation of Multiple Exercise Options with Energy Applications. Tepper Working Paper.

Natolski, J., Werner, R., 2016. Mathematical foundation of the replicating portfolio approach. Working Paper, University of Augsburg.

Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics, 79: 147-168.

Nordahl, H. A., 2008. Valuation of life insurance surrender and exchange options. Insurance: Mathematics and Economics 42.3 (2008): 909-919.

Pelsser, A., 2003. Pricing and Hedging Guaranteed Annuity Options via Static Option Replication. Insurance: Mathematics and Economics, 33: 283-296.

Pelsser, A., Schweizer, J., 2015. The Difference between LSMC and Replicating Portfolio in Insurance Liability Modeling. Working Paper, Maastricht University.

Pfeifer, D., Strassburger, D., 2008. Solvency II: stability problems with the SCR aggregation formula. Scandinavian Actuarial Journal, 2008/1: 61-77.

Schachermayer, W., 2009. The fundamental theorem of asset pricing. Encyclopedia of Quantitative Finance.

Serdyukov A.S., Patutin A.V., Shilova T.V., 2014. Numerical Evaluation of the Truncated Singular Value Decomposition Within the Seismic Traveltimes Tomography

Framework. Journal of Siberian Federal University. Mathematics and Physics, 7(2): 208-218.

Steffensen, M., 2002. Intervention options in life insurance. Insurance: Mathematics and Economics, 31: 71-85.

Tanskanen, A. J., Lukkarinen J., 2003. Fair valuation of path-dependent participating life insurance contracts. Insurance: Mathematics and Economics, 33: 595-609

Tsitsiklis, J.N., Van Roy, B., 2001. Regression methods for pricing complex American-style options. IEEE Transactions on Neural Networks, 12: 694-763.

Vasicek, O., 1997. An Equilibrium Characterization of the Term Structure. Journal of Financial Economics, 5: 177-188.

Vinga, S., 2004. Supplementary material to Vinga, S., Almeida, J.S., 2004. Rényi continuous entropy of DNA sequences. Journal of Theoretical Biology 231: 377-388.

Weber, S., 2007. Distribution-Invariant Risk Measures, Entropy, and Large Deviations. Journal of Applied Probability, 44: 16-40.

Werner, D., 2005. Funktionalanalysis, Fifth Edition. Springer, Berlin.

Whitt, W., 1986. Stochastic comparisons for non-Markov processes. Mathematics of Operations Research, 11: 608-618.

Whitt, W., 2002. Internet Supplement to "Stochastic-Process Limits" (Springer Series in Operations Research, 2002). Available at `http://www.columbia.edu/~ww2040/`.

Yan L., 2010. Guaranteed Minimum Withdrawal Benefits in Variable Annuities. Ph.D. Thesis, University of Waterloo.

Zhu, N., Bauer, D., 2011. Applications of Forward Mortality Factor Models in Life Insurance Practice. The Geneva Papers on Risk and Insurance, 36: 567-594