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

Daniel Bauer[*]    &    Hongjun Ha

Department of Risk Management and Insurance. Georgia State University

35 Broad Street. Atlanta, GA 30303. USA

April 2015

## Abstract

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.

Relying on a well-known method for pricing non-European derivatives, the current paper proposes and analyzes a novel approach to this computational problem based on least-squares regression and Monte Carlo simulations. We show convergence of the algorithm, we analyze the resulting estimate for practically important risk measures, and we derive *optimal basis functions* based on spectral methods. Our numerical examples demonstrate that the algorithm can produce accurate results at relatively low computational costs, particularly when relying on the optimal basis functions.

*Keywords:* Loss distribution, least-square Monte Carlo, Value-at-Risk, spectral theory, variable annuities.

---

[*]Corresponding author. Phone: +1-(404)-413-7490. Fax: +1-(405)-413-7499. E-mail addresses: dbauer@gsu.edu (D. Bauer); hha6@gsu.edu (H. Ha).

# 1 Introduction

Many risk management applications within financial institutions entail a reevaluation of the company's assets and liabilities at some time horizon $T_H$ (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 II for banks, and Solvency II or the Swiss Solvency Test for insurance companies. However, the high complexity of this *nested* computation structure leads firms to struggle with their implementation. As a consequence, many companies rely on second-best 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 (see e.g. Liebwein (2006) or Pfeifer and Strassburger (2008)).[1]

The present paper 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 (Clement et al., 2002): On the one hand, the loss random variable, which can be represented as a conditional risk-neutral expected value at the time horizon $T_H$, 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 loss at $T_H$, 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), thus far there exists no detailed analysis of the properties of this algorithm or of how to choose the basis functions. This paper closes this gap in literature. We start our analysis by introducing our setting and the algorithm. As an important innovation, we frame the estimation problem via a *loss 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 analyses on a hybrid probability measure that easily 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 are carried out under the risk-neutral measure.

We formally establish convergence of the algorithm for the risk distribution (in probability) and for families of risk measures. Moreover, by building on ideas from Gordy and Juneja (2010), we analyze in more detail the properties of the estimator for two important special cases: the probability of a large loss and Value-at-Risk, which serves as the risk measure for important regular frameworks such as Basel II or Solvency II. More precisely,

---

[1] In a recent paper, Gordy and Juneja (2010) argue that a *nested simulation approach* can deliver reliable results even if the number of trials in the inner simulation step is relatively small. However, their results to a great extent depend on the existence of a large number of independent positions within the company's (asset) portfolio. If the positions are calculated based on the same scenarios or if there is only a single position as for instance within the *direct* and *indirect* approach for determining the liabilities of an insurance company, respectively, the results are less encouraging (Bauer et al., 2012).

we show that the least-squares estimation of the regression approximation, while unbiased when viewed as an estimator for the individual loss, typically yields a positive bias for these tail risk measures. 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.

This is where the operator formulation bears full fruit. More precisely, by expressing the loss operator via its spectral representation (see Linetsky (2008) for details on eigenvalue expansions/ spectral representations of valuation operators), we show that under certain conditions, the eigenfunctions present an optimal choice for the basis functions. More precisely, we demonstrate that these eigenfunctions – which are closely to related to the eigenfunctions of the infinitesimal generator of the underlying Markov process – approximate the loss operator in an optimal manner. The intuition is that similarly to a principal component analysis for finite-dimensional operators, the eigenfunctions provide the most important dimensions in spanning the image space of the operator.

We implement the proposed approach within the constant elasticity of variance (CEV) model for a relevant example from life insurance: a variable annuity with a guaranteed minimum death benefit (GMDB) and a guaranteed minimum accumulation benefit (GMAB). This example displays several advantages. On the one hand, we are able to derive a closed-form solution for the valuation problem at the risk horizon so that we can conveniently compare the approximated empirical loss distribution with the "exact" one. On the other hand, this example fits our setting for the spectral decomposition of the risk operator, which allows comparing the optimal basis functions to other basis functions commonly applied in LSM approaches such as Hermite, Chebyshev, or Legendre polynomials. Our results demonstrate that the algorithm can produce accurate results at relatively low computational costs. We show numerically that there exists a positive bias due to the regression estimation. However, we also find that the bias due to the choice of the basis functions is less transparent, although optimal basis functions considerably improve the performance of the algorithm. In particular, the approximation of the distribution as measured by various statistical distance measures is far superior for the eigenfunctions documenting the importance of the choice of basis function.

The remainder of the paper is structured as follows: Section 2 lays out the simulation framework and the algorithm; Section 3 addresses convergence of algorithm and analyzes the estimator in special cases; Section 4 discusses optimal basis functions based on spectral methods; Section 6 provides the numerical example of applications in the context of VA; and, finally, Section 7 concludes the paper and discusses extensions left for future research. All proofs are relegated to the Appendix.

## 2  The LSM Approach

### 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 $\mathbb{P}$ is the physical measure. 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 returns, 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 risky assets in the market can be expressed in terms of $Y$. Futhermore, we suppose the existence of a locally risk-free process $(B_t)_{t \in [0,T]}$ (the bank account) with $B_t = \exp\{\int_0^t r_u \, du\}$, where $r_t = r(Y_t)$ is the instantaneous risk-free interest rate at time $t$. 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 the numéraire process $(B_t)_{t \in [0,T]}$.[2] In particular, for the time $t$ price of a zero-coupon bond with maturity $k \geq t$, we have $p(t,k) = p(t,k,Y_t) = \mathbb{E}^{\mathbb{Q}}\left[ e^{-\int_t^k r_s \, ds} \middle| Y_t \right].$

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

$$\inf \left\{ x \in \mathbb{R} \middle| \mathbb{P}\left(-\mathrm{AC}_{T_H} \leq x\right) \geq \alpha \right\}. \tag{1}$$

The available 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 typically 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 embedded options may require numerical approaches. That is the main source of the complexity associated with this task, since this valuation needs to be carried out for each realization of the $Y$ at time $T_H$, 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 $T_H \in \{1, 2, \ldots, T\}$ :

$$\mathrm{AC}_{T_H} = \mathbb{E}^{\mathbb{Q}}\left[ \sum_{k=T_H}^{T} e^{-\int_{T_H}^{k} r_s \, ds} X_k \middle| (Y_s)_{0 \leq s \leq T_H} \right]. \tag{2}$$

Note that within this formulation, interim asset and liability cash-flows in $[0, T_H]$ maybe be aggregated in the $\sigma(Y_s, 0 \leq s \leq T_H)$-measurable position $X_{T_H}$. Moreover, in contrast to

---

[2]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.

e.g. Gordy and Juneja (2010), we consider aggregate asset and liability cash-flows at times $k \geq T_H$ 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 potentially not even possible as it is for instance the case within economic capital modeling in insurance (Bauer et al., 2012).

In addition to the 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 insurers bookkeeping system (see Bauer et al. (2010) for a detailed discussion in the context of life insurance). 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:

$$\text{AC}_{T_H} = \mathbb{E}^{\mathbb{Q}} \left[ \left. \sum_{k=T_H}^{T} \frac{B_{T_H}}{B_k} X_k \right| Y_{T_H} \right].$$

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 $Y_k$,[3]

$$\frac{B_{T_H}}{B_k} X_k = x_k (Y_k), \; T_H \leq k \leq T.$$

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

$$\text{AC}_{T_H} (Y_{T_H}) = \sum_{j=T_H}^{T} \mathbb{E}^{\mathbb{Q}} \left[ x_j (Y_j) | Y_{T_H} \right].$$

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

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

**Lemma 2.1.** *We have :*

  *1.* $\tilde{\mathbb{P}}(A) = \mathbb{P}(A), \; A \in \mathcal{F}_t, \; 0 \leq t \leq T_H.$

  *2.* $\mathbb{E}^{\tilde{\mathbb{P}}} [X | \mathcal{F}_{T_H}] = \mathbb{E}^{\mathbb{Q}} [X | \mathcal{F}_{T_H}]$ *for every random variable* $X \in \mathcal{F}.$

  Lemma 2.1 implies that we have

$$\begin{aligned} \text{AC}_{T_H} (Y_{T_H}) &= \sum_{j=T_H}^{T} \mathbb{E}^{\tilde{\mathbb{P}}} [x_j (Y_j) | Y_{T_H}] & (3) \\ &= T_L (\mathbf{x}) (Y_{T_H}), \end{aligned}$$

---

[3]Similarly to Section 8.1 in Glassermann (2004), without loss of generality by possibly augmenting the state space, we assume that the discount factor can be expressed as a function of the state variables.

where the operator

$$T_L : \mathcal{H} = \bigoplus_{j=T_H}^{T} L^2\left(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_j}\right) \to L^2\left(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_{T_H}}\right) \tag{4}$$

is mapping a model to capital. We call $T_L$ in (4) the *loss operator*. For our applications later in the text, it is important to note the following:

**Lemma 2.2.** *$T_L$ is a continuous linear operator.*

In particular, the above implies that a model can be identified with an element of the Hilbert space $\mathcal{H}$ whereas the capital $AC_{T_H}$ can be (state-wise) identified with an element of $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}}\right)$. The task at hand is now to evaluate this element for a given model $\mathbf{x} = (x_{T_H}, \ldots, x_T)$ and to then determine the capital requirement via a (monetary) risk measure $\rho : L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}}\right) \to \mathbb{R}$ as $\rho\left(T_L\left(\mathbf{x}\right)\right)$.

One possibility to carry out this computational problem is to rely on *nested simulations*, i.e. to simulate a large number of scenarios for $Y_{T_H}$ 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 applications, may requires a very large number of simulations to obtain results in a reliable range (Bauer et al., 2012). Whence, in the following, we propose and develop an alternative approach for such situations.

## 2.2 Least-Squares-Algorithm

As was pointed out in the previous section, the task at hand is to determine the distribution of $AC_{T_H}$ given by Equation (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*" (Clement 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 Clement et al. (2002), their approach consists of two different types of approximations. Proceeding analogously, as the first approximation, we replace the conditional expectation, $AC_{T_H}$, by a finite combination of linear independent basis functions $e_k(Y_{T_H}) \in L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}}\right)$:

$$AC_{T_H} \approx \widehat{AC}_{T_H}^{(M)}(Y_{T_H}) = \sum_{k=1}^{M} \alpha_k \times e_k(Y_{T_H}). \tag{5}$$

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

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

$$\mathrm{PV}_{T_H}^{(i)} = \sum_{j=T_H}^{T} x_j \left( Y_j^{(i)} \right), \ 1 \le i \le N.$$

We use these realizations in order to determine the coefficients $\alpha = (\alpha_1, \ldots, \alpha_M)$ in the approximation (5) by least-squares regression:

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

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

$$\mathrm{AC}_{T_H} \approx \widehat{\mathrm{AC}}_{T_H}^{(M)}(Y_{T_H}) \approx \widehat{\mathrm{AC}}_{T_H}^{(M,N)}(Y_{T_H}) = \sum_{k=1}^{M} \hat{\alpha}_k^{(N)} \times e_k(Y_{T_H}), \tag{6}$$

based on which we may then determine $\rho\left(T_L\left(\mathbf{x}\right)\right) \approx \rho(\widehat{\mathrm{AC}}_{T_H}^{(M,N)})$.

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

## 3    Analysis of the Algorithm

### 3.1    Convergence

The following proposition establishes the convergence of the algorithm described in Section 2.2:

**Proposition 3.1.** $\widehat{AC}_{T_H}^{(M)} \to AC_{T_H}$ in $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}}\right)$, $M \to \infty$, and $\widehat{AC}_{T_H}^{(M,N)} \to \widehat{AC}_{T_H}^{(M)}$, $N \to \infty$, $\tilde{\mathbb{P}}$-almost surely.

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 Clement et al. (2002) since we do not have to take the recursive nature into account. However, in contrast to 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 3.1 is the convergence in probability – and, hence, in distribution – of $\widehat{\mathrm{AC}}_{T_H}^{(M,N)} \to AC_{T_H}$ implying that the resulting distribution function of $\widehat{\mathrm{AC}}_{T_H}^{(M,N)}$ presents a valid approximation of the distribution of $\mathrm{AC}_{T_H}$. The question of

whether $\rho(\widehat{\text{AC}}_{T_H}^{(M,N)})$ presents a valid approximation of $\rho(\text{AC}_{T_H})$ depends on the regularity of the risk measure. In general, we require continuity in $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}})$ 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 3.1.** *If $\{e_k, k = 1, \ldots, M\}$ are orthonormal and $(\sum_{k=T_H}^{T} x_k) \in L^2(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$, $1 \leq j \leq m$, then $\widehat{AC}_{T_H}^{(M,N)} \to AC_{T_H}$, $N \to \infty$, $M \to \infty$ in $L^1(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}})$. In particular, if $\rho$ is a finite convex risk measure on $L^1(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}})$, we have $\rho(-\widehat{AC}_{T_H}^{(M,N)}) \to \rho(-AC_{T_H})$, $N \to \infty$, $M \to \infty$.*

An important special case that does not fall in the class of convex risk measures is Value-at-Risk. More specifically, as pointed out above, a practical situation where the proposed algorithm may be particularly advantageous is the calculation of the solvency capital requirement for insurance companies, which is usually cast in the form of a Value-at-Risk (Bauer et al., 2012). In this case, and for the related problem of estimating the probability of a large loss, convergence immediately follows from Proposition 3.1:

**Corollary 3.2.** *We have:*

$$F_{\widehat{AC}_{T_H}^{(M,N)}}(l) = \mathbb{P}(\widehat{AC}_{T_H}^{(M,N)} \leq l) \to \mathbb{P}(AC_{T_H} \leq l) = F_{AC_{T_H}}(l), \ M, N \to \infty, \ l \in \mathbb{R},$$

*and*

$$F_{\widehat{AC}_{T_H}^{(M,N)}}^{-1}(\alpha) \to F_{AC_{T_H}}^{-1}(\alpha), \ M, N \to \infty,$$

*for all continuity points $\alpha \in (0,1)$ of $F_{AC_{T_H}}^{-1}$.*

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, $\widehat{\text{AC}}_{T_H}^{(M,N)}$ is unbiased – though not necessarily efficient – for $\widehat{\text{AC}}_{T_H}^{(M)}$ (see e.g. Sec. 6 in Amemiya (1985)).[4] However, this clearly does not imply that $\rho(-\widehat{\text{AC}}_{T_H}^{(M,N)})$ is unbiased for $\rho(-\widehat{\text{AC}}_{T_H}^{(M,N)})$. Proceeding similarly to Gordy and Juneja (2010) for the nested simulation estimator, in the next subsection we analyze this question in more detail for the important special cases of the probability of a large loss and Value-at-Risk.

## 3.2 Probability of a Large Loss and Value-at-Risk

In Gordy and Juneja (2010), the authors show that the nested simulations estimator for tail risk measures such as the probability of a large loss or Value-at-Risk carries a positive bias in the order of the numbers of simulations in the inner step. They derive their results by considering the joint density of the exact distribution of the capital at time $T_H$ and the error

---

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

when relying on a finite number of inner simulations scaled by the square-root of the number of inner simulations by the central limit theorem.

The following Lemma establishes that their results carry over to our setting in a straightforward manners:

**Lemma 3.1.** $Z^{(N)} = \sqrt{N} \left[ \widehat{AC}_{T_H}^{(M)} - \widehat{AC}_{T_H}^{(M,N)} \right] \longrightarrow Normal\,(0, \xi)$ *where* $\xi$ *is provided in (15) in the Appendix.*

Denote by $g_N(\cdot, \cdot)$ the joint density function of $(-\widehat{AC}_{T_H}^{(M)}, Z^{(N)})$, and assume it satisfies the regularity conditions from Gordy and Juneja (2010) collected in the Appendix. Then we obtain:

**Proposition 3.2.** [5] *Under the regularity conditions we have:*

1. $\mathbb{P}\left( \widehat{AC}_{T_H}^{(M,N)} < -\mu \right) = \mathbb{P}\left( \widehat{AC}_{T_H}^{(M)} < -\mu \right) + \frac{\theta_\mu}{N} + O_N(N^{-1.5})$,

   *where* $\theta_\mu = -\frac{1}{2} \frac{d}{d\mu} \bar{f}(\mu) \mathbb{E}\left[ \sigma_{Z^{(N)}}^2 \,|\, -\widehat{AC}_{T_H}^{(M)} = \mu \right]$, $\sigma_{Z^{(N)}}^2 = \mathbb{E}\left[ \left( Z^{(N)} \right)^2 | Y_{T_H} \right]$ *and* $\bar{f}$ *is the marginal density of* $-\widehat{AC}_{T_H}^{(M)}$.

2. $VaR_\alpha\left[ -\widehat{AC}_{T_H}^{(M,N)} \right] = VaR_\alpha\left[ -\widehat{AC}_{T_H}^{(M)} \right] + \frac{\theta_\alpha}{N\bar{f}\left( VaR_\alpha\left( -\widehat{AC}_{T_H}^{(M)} \right) \right)} + o_N(N^{-1})$,

   *where* $\theta_\mu = -\frac{1}{2} \frac{d}{d\mu} \left[ \bar{f}(\mu) \mathbb{E}\left[ \sigma_{Z^{(N)}}^2 \,|\, -\widehat{AC}_{T_H}^{(M)} = \mu \right] \right]_{\mu = VaR_\alpha\left[ -\widehat{AC}_{T_H}^{(M)} \right]}$

The key point of the proposition is that–similar to the nested simulations estimator – the LSM estimator for the probability of a large loss and the Value-at-Risk 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, on average the Value-at-Risk LSM estimator will be "too conservative". However, note that here we ignore the variance of the estimate due to estimating the risk measure from the finite sample, which may well trump the inaccuracy due to the bias. Indeed, when estimating the probability of a large loss via the empirical probability or the Value-at-Risk via the empirical quantile, the convergence of the variance is in the order of $N$, and thus dominates the mean-square error for relatively large values of $N$.

Moreover, of course the result only pertains to the regression approximation but not the approximation of the loss variables via the linear combination of basis functions which is at the core of the proposed algorithm. In particular, due to the linear complexity, increasing $N$ is cheap relative to the nested simulation algorithm, so that analyzing the functional approximation is more essential.

## 4 Choice of Basis Function

As indicated in Section 3, any set of independent functions will lead the algorithm to converge. In fact, for the Least Squares Monte Carlo method for pricing non-European derivatives, frequent choices include Hermite polynomials, Legendre polynomials, or Chebyshev

---

[5]c. Proposition 1 and 2 in Gordy and Juneja (2010).

polynomials and based on various numerical tests, Moreno and Navas (2003) conclude that the approach is robust to the choice of basis functions. A key difference between the Least Squares Monte Carlo 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 for more complex but also more crucial in the present context.

In what follows, we first discuss in more detail the nature of the choice of the basis functions. In particular, we introduce the notion of optimal basis functions for a given model framework. Subsequently, we demonstrate in a diffusion framework, that under certain conditions the spectral expansion of the valuation operator yields optimal basis functions in this sense.

## 4.1   Nature of Choice of Basis Functions

The space, $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}})$, is a separable Hilbert Space. Therefore, there exists at least one complete orthonormal sequence, $\{e_j\}_{j \geq 1}$[6], and we can represent any element $T_L(\mathbf{x}) \in L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}})$, as

$$T_L(\mathbf{x}) = \sum_{j=1}^{\infty} \langle T_L(\mathbf{x}), e_j \rangle \, e_j. \tag{7}$$

The LSM algorithm prescribes approximation of $T_L(\mathbf{x})$ via the first $M$ elements of this sequence, $\{e_1, ..., e_M\}$:

$$T_L(\mathbf{x}) \approx \widehat{T_L}(\mathbf{x}) := \mathbf{P}T_L(\mathbf{x}) = \sum_{j=1}^{M} \langle T_L(\mathbf{x}), e_j \rangle \, e_j \tag{8}$$

The projection operator $\mathbf{P}$ is defined by

$$\mathbf{P} = \sum_{j=1}^{M} < \cdot, e_j > e_j \tag{9}$$

Since we use the finite number of basis functions, one possible objective may be to choose the set of basis functions that minimizes the distance between $T_L(\mathbf{x})$ and $\mathbf{P}T_L(\mathbf{x})$ among all conceivable basis functions. However, this choice may not be robust to change in the *model*, and in fact the necessity to derive optimal choices for each specification could be problematic. Thus, here we take a different approach and consider optimal basis functions for all $\mathbf{x}$, that is optimal basis functions for a *model framework*:

**Definition 4.1.** *We call the set of basis functions $\{e_1^*, e_2^*, ..., e_M^*\}$ optimal in $L^2(\mathbb{R}^d, \mathcal{B}, m)$ if*

$$\{e_1^*, e_2^*, ..., e_M^*\} = arginf_{\{e_1, e_2, ... e_M\}} ||T_L - \mathbf{P}T_L|| = arginf_{\{e_1, e_2, ... e_M\}} \sup_{||f||=1} ||T_L f - \mathbf{P}T_L f||$$

Note that the definition is subject to a given Hilbert space, and we determine the optimal approximation of the capital random variable for any model within our framework.

---

[6]For simplicity, we restrict ourselves to orthonormal basis functions. We omit $Y_{T_H}$ in what follows for convenience.

## 4.2    Application of the Spectral Theory

In the following, we assume that the state process evolves according to a finite-dimensional diffusion:

$$dY_t = b(Y_t)dt + a(Y_t)dW_t, \qquad Y_0 \in \mathbb{R}^d \tag{10}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^{d \times d}$, and $W_t$ is an $d$-dimensional standard Brownian motion. Then the infinitesimal generator $(\mathcal{G})$ of $Y$, is given by

$$\mathcal{G}\mathbf{x}(y) := \lim_{t \downarrow 0} \frac{1}{t}(\mathcal{P}_t\mathbf{x} - \mathbf{x}) = \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(y)\frac{\partial^2}{\partial y_i \partial y_j}\mathbf{x}(y) + \sum_{i=1}^{d} b_i(y)\frac{\partial}{\partial y_i}\mathbf{x}(y), \quad \mathbf{x} \in \mathrm{Dom}(\mathcal{G})$$

$$\mathrm{Dom}(\mathcal{G}) = \left\{ \mathbf{x} \in \mathbf{B} : \lim_{t \downarrow 0}\frac{1}{t}(\mathcal{P}_t\mathbf{x} - \mathbf{x}) \text{ exists in } \mathbf{B} \right\}.$$

where the boundary condition is defined by the nature of problem and $t \in [0, T]$.

Note that

$$T_L(\mathbf{x}) = \sum_{j=T_H}^{T} \mathbb{E}^{\mathbb{P}}[x_j(Y_j)| \cdot ]$$

$$= \sum_{j=T_H}^{T} \mathcal{P}_{j-T_H}x_j$$

where $\mathcal{P}_t$ is the *pricing semigroup* (see Linetsky (2008) for a detailed introduction to operator methods and spectral theory in derivative pricing.) In particular, as a linear combination of pricing operators, in order to determine optimal basis function for $T_L$, it is sufficient to find optimal basis functions for $\mathcal{P}_t$. In what follows, we show that these are given by the spectral expansion of $\mathcal{P}_t$. Specifically, assume that $\mathcal{G}$ is self-adjoint in the Hilbert space $L^2(D, m)$ $D \subseteq \mathbb{R}^d$, and that its spectrum is simple and purely discrete with eigenfunctions $\{\varphi_j\}_{j=1}^{\infty}$ and eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$ ordered increasingly.[7] Then:

$$\mathcal{P}_t\mathbf{x} = \sum_{k=1}^{\infty} e^{-\lambda_k t} < \mathbf{x}, \varphi_k > \varphi_k \tag{11}$$

where $\lambda_k$ is the eigenvalue corresponding to $\varphi_k$ and we obtain :

**Proposition 4.1.** *Under these assumptions, the eigenfunctions of $-\mathcal{G}$, $\{\varphi_j\}_{j=1}^{\infty}$ are the optimal basis functions for $\mathcal{P}_t$ - and thus for $T_L$- in $L^2(D, m)$.*

Two remarks are in orders. On the one hand, we derive the optimal basis functions in $L^2(D, m)$ rather than $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}})$ or even a probability measure that is tailored to the risk measure $\rho$. This is required by the condition that the pricing operator be self-adjoint.

---

[7]This assumption is satisfied under certain assumptions on the model equation (10) and the payoff functions. For instance, in the cases of scalar diffusion, $\mathcal{G}$ will be self-adjoint under the so-called *speed measure*, and the spectrum will be discrete if there is no natural boundary for the payoff functions. See Davydov and Linetsky (2003) and Linetsky (2004, 2008).

On the other hand, certain conditions on the payoff functions must be met for the spectrum to be discrete. Nonetheless, if available, the spectral decomposition provides a theoretical foundation for the choice of basis functions, and provides a pathway to improve accuracy and control errors. For instance, frequently asymptotics are available for the eigenfunctions of model frameworks, which in turn allows to approve the joint convergence in $M$ and $N$.[8] We begin the detailed exploration of these issues in the next section.

## 5    Convergence Rate

In Section 3, we prove the convergence of the LSM estimator to true expectation. The convergence result is sequential. First, we show that $\widehat{AC}_{T_H}^{(M)}$ converges to $AC_{T_H}$ as $M \to \infty$, and $\widehat{AC}_{T_H}^{(M,N)}$ converges to $\widehat{AC}_{T_H}^{(M)}$ as $N \to \infty$.

In this section, we discuss general conditions for convergence and how fast the LSM estimator converges to the true expectation based on the choice of basis functions[9].

To assess convergence rates, we follow methods in Newey (1997). If one approximates $\mathbb{E}[y|x]$ nonparametrically, this conditional expectation may be represented by linear projection on the subspace spanned by $p_i(x)$, $i = 1, ..., M$, where $p_i(x)$ is some function. That is, an estimator, $\widehat{\mathbb{E}}[y|x]$, of $\mathbb{E}[y|x]$ is given by

$$\widehat{\mathbb{E}}[y|x] = \hat{\alpha}_1 p_1(x) + \cdots + \hat{\alpha}_M p_M(x)$$

where $\hat{\alpha}_i$, $i = 1, ..., M$ is an usual ordinary least-squares estimator. In the econometrical literature, this approximation is called "series estimator" for $\mathbb{E}[y|x]$ (e.g. see Newey (1995) , or Newey (1997)). Note that the series estimator has the same logic of the LSM algorithm for approximating $AC_{T_H}(Y_{T_H})$ excepting for data generating process. Hence, we utilize convergence rates of series estimators for convergence rates for the LSM estimators. We summarize general convergence rates of series estimators and its application to the case in which $p_i$ is orthonormal each other. In particular, we derive a convergence rate if we use eigenfunctions of the infinitesimal generator for the underlying process and show why the eigenfunctions are more optimal than any other orthonormal polynomials in the sense of speed of convergence, which is consistent result with the Proposition 4.1. In Newey (1997), the author considers two distance measures to quantify convergence rates : one is the integrated mean square error and $l_\infty$-metric. The convergence rates follow three assumptions. The assumptions and notations are provided in A.1 of the Appendix. Newey (1997) proves that,

**Proposition 5.1.** *If A.1 are satisfied with $v = 0$,*

$$\int [AC_{T_H}(y) - \widehat{AC}_{T_H}^{(M,N)}(y)]^2 dF_0(y) = O_p(M/N + M^{-2\psi})$$

*where $F_0(y)$ is the distribution function of $Y_{T_H}$ under $\tilde{\mathbb{P}}$.*

---

[8]For instance, Davydov and Linetsky (2003) show that for scalar diffusions, the eigenvalues grow as $n^2$ for large $n$, so that the eigenfunction expansion (11) converges very rapidly.

[9]We concentrate on only the expectation.

*Also, if A.1 is satisfied for some $v \geq 0$ then*

$$|\widehat{AC}_{T_H}^{(M,N)} - AC_{T_H}|_v = O_p(\zeta_v(M)[\sqrt{M}/\sqrt{N} + M^{-\psi}])$$

See Newey (1997) for a detailied proof of Theorem 1. Intuitively, it is eaty to think that the convergence rate may be decomposed into two parts : one is error from the least-squares estimation and the other error comes from functional approximation. As seen in the Proposition 5.1, the first term is related to regression estimation error and the second term is due to functional approximation. To have a convergence, the number of parameters should be smaller than the number of simulations. In particular, $\psi$ determines the speed of convergence of $\widehat{AC}_{T_H}^{(M)}$ to $AC_{T_H}$. The form of $\psi$ varys based on the form of basis functions. From now on, we are interested in when $v = 0$ only. In the following two subsections, we provide the convergence rate when using orthonormal polynomials and eigenfunctions of the infinitesimal generators.

## 5.1  Orthonormal Polynomials

In Newey (1997), the author provides the convergence rates when estimating the conditional expectation via orthonormal polynomials. To derive the result, the author makes two assumptions. The assumptions are provided in A.2 in the Appendix. If these assumptions are satisfied, the last two original assumptions of A.1 are consequences of A.2.

**Proposition 5.2.** *If the first assumption of A.1, and A.2 are satisfied with $v = 0$,*

$$\int [AC_{T_H}(y) - \widehat{AC}_{T_H}^{(M,N)}(y)]^2 dF_0(y) = O_p(M/N + M^{-2s/d})$$

$$|\widehat{AC}_{T_H}^{(M,N)} - AC_{T_H}|_0 = O_p(M[\sqrt{M}/\sqrt{N} + M^{-s/d}])$$

See Newey (1997) for a detailed proof of Theorem 4. In Proposition 5.2, we require that $s \geq d$ to have the convergence. Therefore, $s \geq d$ and $M^3/N \to 0$ are necessary conditions for the joint convergence when using ordinary orthonormal polynomials. In most financial application, the expectations are usually continuous and differentiable at least two times with respect to state variables. So these requirements are not too strong. However, the second error generating from the functional approximation diminishes with order of $s/d$ slowly.

## 5.2  Eigenfunctions

We derive the convergence rate when using the eigenfunctions of infinitesimal generator. To establish results, we check our framework satisfies assumptions of A.1. The first assumption is automatically satisfied due to the nature of Monte Carlo simulation. Moreover, the finite conditional variance is an usual assumption in financial application.

For the second assumption of A.1, we assume that

**Assumption 5.1.** *The eigenfunction, $\{\varphi'_j\}_{j=1}^{\infty}$, is bounded on the support of $Y_{T_H}$.*

If Assumption 5.1 is satisfied, the second assumption in A.1 is the consequence of this assumption and orthogonality of eigenfunctions. The Assumption 5.1 is similar to A.2. By

compactness of the supports in the finite Euclidean space, the support can be normalized closed unit ball. Then values of polynomials are bounded and there is a sequence such that $||P^M(y)|| \leq \zeta_0(M)$.

To have alternative to the final assumption of A.1, we prove the following Lemma.

**Lemma 5.1.** *There is an vector $\alpha_M = (\alpha_1, ..., \alpha_M)'$ such that*

$$\sup_{y \in Y} |AC_{T_H}(y) - \varphi(y)'\alpha_M| = O\left(\left(e^{T_H}\right)^{-\lambda_{M+1}}\right)$$

*where $\varphi(y) = (\varphi_1(y), ..., \varphi_M(y))'$ and $\varphi_j(y)$ is the eigenfunction associated to the jth eigenvalue of the infinitesimal generator.*

Then, we have the following theorem for the convergence rate;

**Proposition 5.3.** *If the first assumption of A.1 and Assumptions 5.1 is satisfied with $d = 0$,*

$$\int [AC_{T_H}(y) - AC_{T_H}^{(M,N)}(y)]^2 d\tilde{F}_0(x) = O_p(M/N + \left(e^{T_H}\right)^{-2\lambda_{M+1}})$$

$$|\widehat{AC}_{T_H}^{(M,N)} - AC_{T_H}|_0 = O_p(\sqrt{M}[\sqrt{M}/\sqrt{N} + \left(e^{T_H}\right)^{-\lambda_{M+1}}])$$

*where $\tilde{F}_0(y)$ is the distribution function of $Y_{T_H}$ under the speed measure.*

Usually, $\{\lambda_j\}_{j \geq 1}$ is a linear or quadratic function of $j$. The convergence rate when using eigenfuntions are exponential contrast to orthonormal polynomials which give the convergence with order $s/d$. Therefore we conclude that the convergence rate of the eigenfuctions is faster than usual orthonormal polynomials. On the uniform convergence, we see that the covergence rate of the eigenfunction is faster than orthonormal polynomials since the uniform convergence is obtainable as long as $M^2/N \to 0$.

Note that all Lemma and Proposition above are established under the speed measure. So we need to check the convergence rate is still preserverd under $\tilde{\mathbb{P}}$. Let $\tilde{\mathbb{S}}$ be the speed measure associated to the infinitesimal genertor. We assume that

**Assumption 5.2.** *The Radon-Nikodym derivative, $\frac{\tilde{\mathbb{S}}}{\mathbb{P}}$, is square integrable under $\tilde{\mathbb{P}}$, that is,*

$$\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left(\tilde{\mathbb{S}}/\tilde{\mathbb{P}}\right)^2\right] < \infty$$

Then we have the following Lemma,

**Lemma 5.2.** $\widehat{AC}_{T_H}^{(M,N)}$ *converges to $AC_{T_H}$ in $L^1$ under $\tilde{\mathbb{P}}$.*

Therefore, $\widehat{AC}_{T_H}^{(M,N)}$ converges to $AC_{T_H}$ in probability by Markov's inequality.

# 6  Application

In this section, we provide a numerical example for appraising the LSM algorithm. The example comes from the insurance setting: we analyze called a variable annuity (VA) within the constant elasticity variance (CEV) process setting. Most quantities (the distribution of loss at the time $T_H$, particularly) relevant to the VA product considered here can be obtained in closed form. Hence, we can measure the performance of the LSM algorithm by comparing various estimated quantities from the estimated distribution with the "true" values from the exact distribution of the future losses. First, we estimate the loss distribution via the LSM using the Hermite, Chebyshev, Legendre polynomials and eigenfunctions of the infinitesimal generator for the CEV process respectively as the basis functions. In particular, we estimate the mean, the 3rd quartile and VaR at 99% of the future losses via the proposed approach and show their convergence. Finally, we provide statistical distances to verify that the eigenfunctions are superior to other basis functions when approximating the distribution.

## 6.1  Variable Annuity with GMDB and GMAB

Variable Annuities (VAs) are savings product offered by insurance companies, and Guaranteed Minimum Benefits (GMBs) are long-term, mortality-dependent option features. In 2012, individual VA sales amounted to \$147 billion, increasing the combined net assets under management of VAs to \$1.64 trillion. The great majority of these VAs contain one or even multiple GMBs (Geneva Association, 2013).

For our application, we consider a VA with a simple Guaranteed Minimum Death Benefit (GMDB) and a Guaranteed Minimum Accumulated Benefit (GMAB). Under these contracts, the policyholder receives a death benefit with a guaranteed minimum amount at the end of the year of death or the survival benefit, also with a guaranteed minimum level, in case she survives until maturity. The VA is contracted with a net single premium (NSP), $P$, and the NSP is invested into the equity market where the current equity value at time zero is $S_0$. We ignore other expenses such as a loading premium. To make the problem simple, we assume that the insurer does not implement any hedging strategy and $P$ is equal to $S_0$.

In the following subsection, we analyze the payoff of the VA and provide a closed form solution for its value within the CEV framework.

## 6.2  Payoff and Pricing Formula of VA

We follow Bauer et al. (2008) to compute the relevant quantities of the VA. We denote $T(x)$ to be future life time of insured whose age is $x$. The discounted payoff of the VA for an insured under the constant risk free rate, $r$, is given by

$$\text{Payoff} = e^{-r(K(x)+1)} \max \left\{ S_{K(x)+1}, \ S_0(1+g_d)^{K(x)+1} \right\} \mathbb{1}_{\{K(x)<L\}}$$
$$+ e^{-rL} \max \left\{ S_L, \ S_0(1+g_i)^L \right\} \mathbb{1}_{\{K(x)\geq L\}}$$

where $K(x)$ is the greatest integer which is less than or equal to $T(x)$, $S_k$ is the account value at time $k$, $g_d$ is the minimum interest rate (roll up) under GMDB, $g_i$ is the minimum interest rate (roll up) under the GMAB, and $L$ is the maturity of the contract, and $\mathbb{1}_{\{A\}}$ is the indicator function with one if $A$ is true or zero otherwise.

In particular, we observe that the death benefit amount at the $(k+1)$th insurance year and the survival benefit are guaranteed. We can express these guarantees in forms of (protective) vanilla Put options as follows:

$$\text{Payoff} = e^{-r(K(x)+1)} \left[ \underbrace{\max\left\{0, \ S_0(1+g_d)^{k+1} - S_{k+1}\right\}}_{\text{Put Option Payoff for GMDB}} + S_{k+1} \right] \mathbb{1}_{\{K(x)<L\}}$$

$$+ e^{-rL} \left[ \underbrace{\max\left\{0, \ S_0(1+g_i)^L - S_L\right\}}_{\text{Put Option Payoff for GMAB}} + S_L \right] \mathbb{1}_{\{K(x)\geq L\}} \tag{12}$$

To compensate for these *embedded options*, a constant fee $\phi$ is continuously deducted from the account value $S_k$ and its fair "price" corresponds to the $\phi$ such that the expected value of the payoff equals the initial investment (Bauer et al., 2008).

There are various papers considering the valuation of VAs with GMBs in the Black-Scholes framework. Here, we consider the valuation in the Constant Elasticity of Variance (CEV) framework. More precisely, we assume the account value evolves according to:

$$dS_t = (\mu - \phi)S_t dt + \sigma S_t^{\beta/2} dW_t \tag{13}$$

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

$$dS_t = (r_t - \phi)S_t dt + \sigma S_t^{\beta/2} dZ_t, \tag{14}$$

where $(W_t)$ and $(Z_t)$ are standard Brownian motions under the physical measure and risk-neutral measure, respectively. This setting has the advantage that it satisfies the assumptions from Proposition 4.1, which allows us to consider "optimal basis functions". On the other hand, as within the Black-Scholes framework, explicit formulas are available for vanilla Put and Call option prices, which yields closed-form evaluation formulas of our Variable Annuity product. In particular, we obtain for the risk-neutral expected discounted value of the payoff:

$$\mathbb{E}^{\mathbb{B}\times\mathbb{Q}}[\text{Payoff}] = \sum_{k=0}^{L-1} {}_kp_x q_{x+k} \left[ \pi\left(S_0, S_0(1+g_d)^{k+1}, k+1\right) + S_0 e^{-\phi(k+1)} \right]$$

$$+ {}_Lp_x \left[ \pi\left(S_0, S_0(1+g_d)^L, L\right) + S_0 e^{-\phi L} \right]$$

where $\mathbb{B}$ is the probability measure associated to $K(x)$, $\mathbb{B} \times \mathbb{Q}$ is a product measure and ${}_kp_x$ is the $k$-year survival probability for an $x$-year old, $q_{x+k} = 1 - p_{x+k}$ is the one-year mortality probability.

For simplicity, we assume that mortality follows a simple De Moivre's law with maximal age $\omega$. More precisely, we assume that ${}_kp_x = 1 - k/(\omega - x)$, which allows for a straightforward

calculation of the survival and mortality probabilities.

$$\mathbb{E}^{\mathbb{B}\times\mathbb{Q}}[\text{Payoff}] = S_0 \left( \frac{1 - e^{-\phi L}}{(\omega - x)(e^\phi - 1)} + \frac{\omega - x - L}{\omega - x} e^{-\phi L} \right) + \frac{\omega - x - L}{\omega - x} \pi \left( S_0, S_0(1 + g_i)^L, L \right)$$
$$+ \sum_{k=0}^{L-1} \frac{1}{\omega - x} \pi \left( S_0, S_0(1 + g_d)^{k+1}, k + 1 \right)$$

where $\pi(S, E, T)$ is the European Put option price with current price of the underlying asset $S$, strike price $E$, and maturity $T$. Its closed form under the CEV process is given in Schroder (1989):

$$\pi(S, E, T) = Ee^{-rt}[1 - F(2x; 2/(2 - /\beta), 2y)] - Se^{-\phi t}F(2y; 2 + 2/(2 - \beta), 2x)$$
$$x = kS^{(2-\beta)}e^{(r-\phi)(2-\beta)t},$$
$$y = kE^{(2-\beta)},$$
$$k = \frac{2(r - \phi)}{\sigma^2(2 - \beta)[e^{(r-\phi)(2-\beta)t} - 1]},$$

where $F(x; \nu, n)$ is the non-central chi-square distribution function evaluated at $x$ with $\nu$ degrees of freedom and non-centrality parameter $n$.

## 6.3 Capital Requirements for the Variable Annuity

For calculating the capital requirement for this variable annuity product, we map the previous notations to the current problem. The dimension of the state process for the VA problem considered here equals to one and its dynamics are given directly by the equity price. So, $Y_t = S_t$. The discounted future cash flow, $x_j(S_j)$ $(j = T_H, ..., L - 1)$, is the discounted death benefit and $x_L(S_L)$ is the sum of discounted death benefit and survival benefit.[10] Finally, the loss operator, $T_L$, maps the discounted future cash flows to the price at time $T_H$. At $T_H$, therefore, we have:

$$\text{AC}_{T_H} = T_L\mathbf{x}(S_{T_H}) = \sum_{j=T_H}^{L} \mathbb{E}^{\tilde{\mathbb{P}}}[x_j | S_{T_H}]$$
$$= S_{T_H} \left( \frac{1 - e^{-\phi(L-T_H)}}{(\omega - x - T_H)(e^\phi - 1)} + \frac{\omega - x - L}{\omega - x - T_H} e^{-\phi(L-T_H)} \right)$$
$$+ \sum_{j=0}^{L-T_H-1} \frac{1}{\omega - x - T_H} \pi \left( S_{T_H}, S_0(1 + g_d)^{j+T_H+1}, k + 1 \right)$$
$$+ \frac{\omega - x - L}{\omega - x - T_H} \pi \left( S_{T_H}, S_0(1 + g_i)^L, L - T_H \right).$$

---

[10]We have discounted values with regards to mortality, i.e. we do not consider unsystematic mortality risk for the calculation of capital requirements. See Zhu and Bauer (2011) for a setting that includes the mortality risk.

Note that the CEV process allows us to express the pricing semigroup - and thus the loss operator - in terms of the spectral representation since the infinitesimal generator for the CEV process is self-adjoint and its spectrum is simple and purely discrete. As in (11), the loss operator has the following spectral representation:

$$T_L\mathbf{x}(S_{T_H}) = \sum_{n=1}^{\infty} e^{-\lambda_n T_H} <\mathbf{x}, \varphi_n> \varphi_n(S_{T_H})$$

The forms of eigenvalues ($\{\lambda_n\}_{n\geq 1}$) and the associated normalized eigenfunctions($\{\varphi_n\}_{n\geq 1}$) are given in Linetsky (2008):

$$\lambda_n = (r - \phi)|\beta - 2|n, \ \varphi_n(S) = N_n S e^{-l(S)} L_{n-1}^{(2m)}(l(S)),$$

$$N_n = \sqrt{\frac{(n-1)!(r-\phi)}{\Gamma(2m+n)}} \left(\frac{(r-\phi)}{\sigma^2|(\beta-2)/2|}\right)^m,$$

$$l(S) = \frac{(r-\phi)}{\sigma^2|(\beta-2)/2|} S^{-(\beta-2)}, \ m = \frac{1}{8|\beta-2|},$$

where $L_{n-1}^{(\nu)}(x)$ are the generalized Lauguerre polynomial of $(n-1)$th order.[11] The convergence rate of the LSM via eigenfunctions is given by

$$\int [AC_{T_H}(x) - AC_{T_H}^{(M,N)}(x)]^2 dF_0(x) = O_p(M/N + (e^{T_H})^{-2M})$$

$$|\widehat{AC}_{T_H}^{(M,N)} - AC_{T_H}|_0 = O_p(\sqrt{M}[\sqrt{M}/\sqrt{N} + (e^{T_H})^{-M}])$$

since the eigenfunction is bounded.[12] and $||P^M(x)|| \leq C\sqrt{M}$.[13] As we expect, the convergence rate is faster than if we use other orthonormal polynomials as basis functions.

Note that we can derive the exact "realizations" of the capital at time $T_H$. That is, if we know $S_{T_H}$, we are able to obtain the capital requirement for the VA using the previous formula for $AC_{T_H}$. Alternatively, we can use the LSM algorithm introduced in previous sections. In particular, we are able to compare the results for the two approaches providing the opportunity to assess the performance of the LSM algorithm. In subsection 6.4, we provide results from the LSM algorithm using Hermite, Chebyshev, Legendre polynomials and eigenfunctions.

## 6.4  Results

If not mentioned otherwise, in the section we consider the following *base case parameters*: $S_0 = 10$, $g_d = 4\%$, $g_i = 5\%$, $r = 5\%$, $\omega = 100$, $x = 45$, $\mu = 10\%$, $\phi = 3.032\%$, $\sigma = 25\%$,

---

[11]The generalized Lauguerre polynomial, $L_{n-1}^{(\nu)}(x)$, is the solution to the differential equation, $xy'' + (\nu + 1 - x)y' + (n-1)y = 0$.

[12]Note that $\varphi_n(S) = N_n S e^{-l(S)} L_{n-1}^{(2m)}(l(S)) = N_n \left(\frac{\sigma^2|(\beta-2)/2|}{(r-\phi)} l(S)\right)^{-\frac{1}{(\beta-2)}} e^{-l(S)} L_{n-1}^{(2m)}(l(S))$. We see that $\varphi(S)$ is continuous and $\lim_{S\to\infty} \varphi(S) = 0$. Therefore, $\varphi_n(S)$ should be bounded.

[13]$\sup_{S\in(0,\infty)} ||P^M(S)|| \leq C\sqrt{\frac{1}{\Gamma(2m+1)} + \frac{1}{\Gamma(2m+2)} + \cdots + \frac{(M-1)!}{\Gamma(2m+M)}} \leq C'\sqrt{M}$, where $C = \max_{1\leq n\leq M} \sup_{S\in(0,\infty)} \left(S e^{-l(S)} L_{n-1}^{(2m)}(l(S))\right)^2$ and $C' = \frac{C}{\Gamma(2m+1)}$.
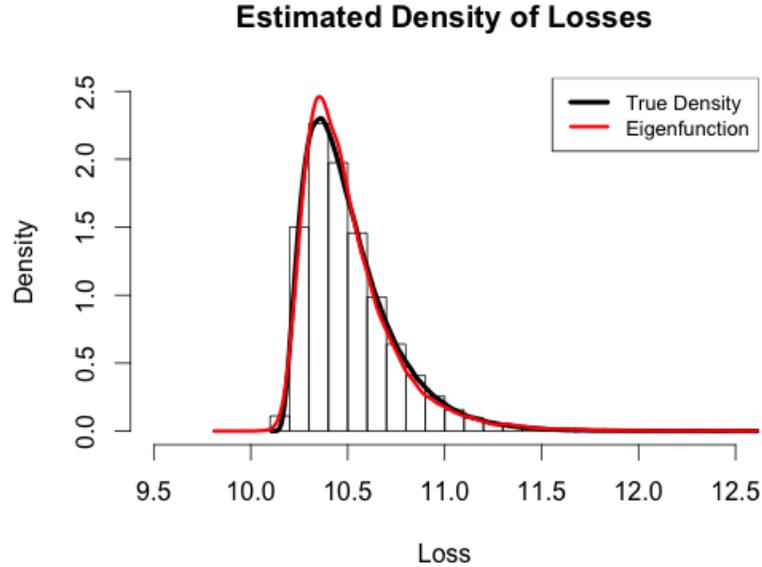
Figure 1: Estimated Density Function via Eigenfunctions

$\beta = 1.4$, $L = 15$, and $T_H = 1$. The one-year risk horizon is common in the insurance industry, the "fair" constant fee, $\phi$, is obtained by Newton's method. The other parameters are chosen arbitrary in a reasonable range.

We start by estimating the density of losses using Hermite, Chebyshev, Legendre polynomials– which are popular in applications (see e.g. Longstaff and Schwartz (2001) or Moreno and Navas (2003))– and eigenfunctions under $N = 50,000$ and $M = 5$. As seen in Figure 1, the LSM algorithm provides a good approximation of the distribution of losses. Though overall fit seems decent, we see that the estimated distribution is slightly more dispersed than the exact distribution.

From the 3.1, we know that the estimated distribution converges in $M$ and $N$. Figure 2 shows the approximation of the density for different $M$. We see that if we choose number of basis functions too small, the estimates are biased considerably, whereas five basis functions seem to provide a good result. This result generally holds for all basis function choices although we observe some differences as seen Figure 3. We come back to the comparison of the basis functions later in this section.

Next, we estimate risk measures. The risk measures considered here are the mean, the 3rd quartile, and the VaR at 99%. We implement the LSM algorithm two hundred times at $N = 50,000$, $N = 100,000$, $N = 300,000$, and $N = 600,000$ respectively. The number of basis functions is five for each $N$. We regard estimators obtained from $N = 600,000$ as $AC^{(5)}$. Based on the 200 samples, we generate the box-whisker plot to check the convergence and to appraise the estimators.

For the mean, Figure 4 shows that the estimate converges to its correct value. Clearly, the variance of the estimator decreases in $N$. Overall, the results are encouraging even for small number of $N$. In particular, no large bias observable. As pointed out in Section 3.1
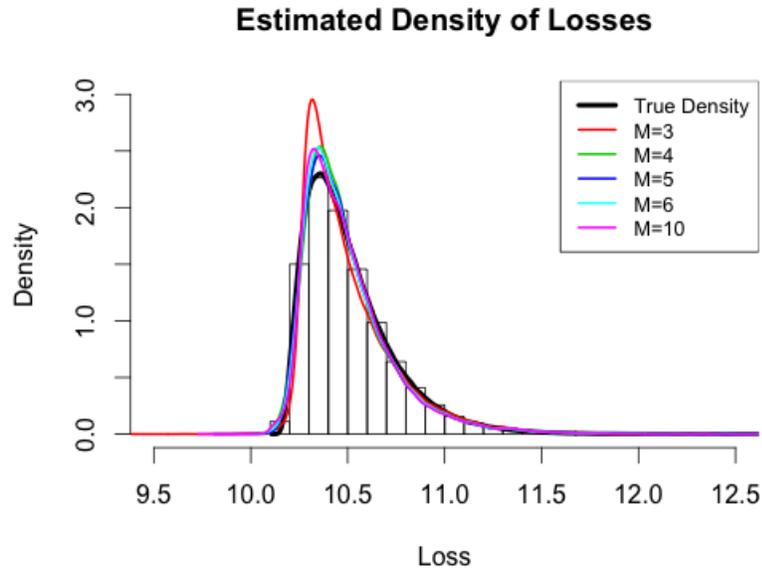
Figure 2: Estimated Density Function via Eigenfunctions under different number of basis functions
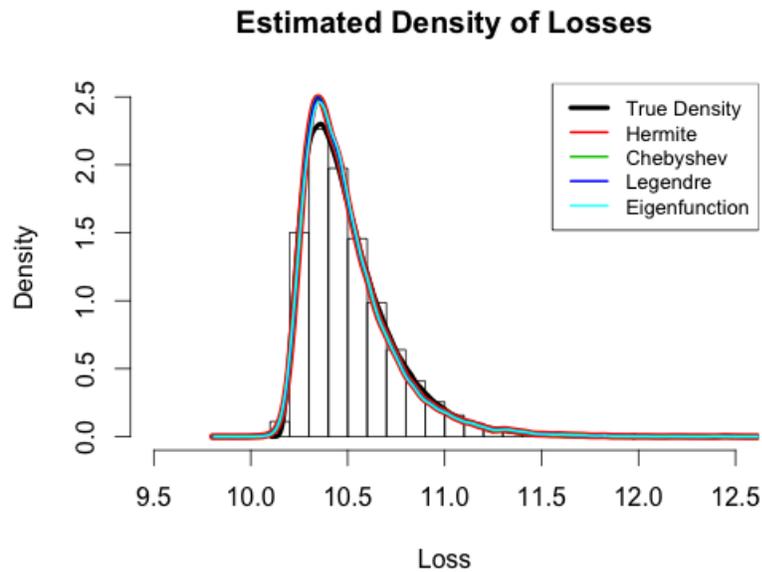


Figure 3: Estimated Density Function via various basis functions with $M = 5$

every estimated mean is expected to have a small deviation from the true mean though the deviation should decrease in $N$. Results for other basis functions are similar.

This changes when we move away from the center of the distribution. Figure 5 provides

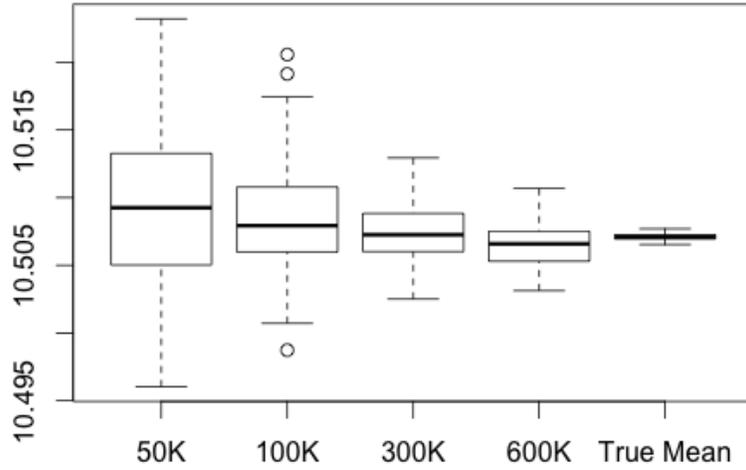**Box-Whisker Plot of mean via Eigenfunction**

Figure 4: Box Plot of Mean, Eigenfunctions, $M = 5$

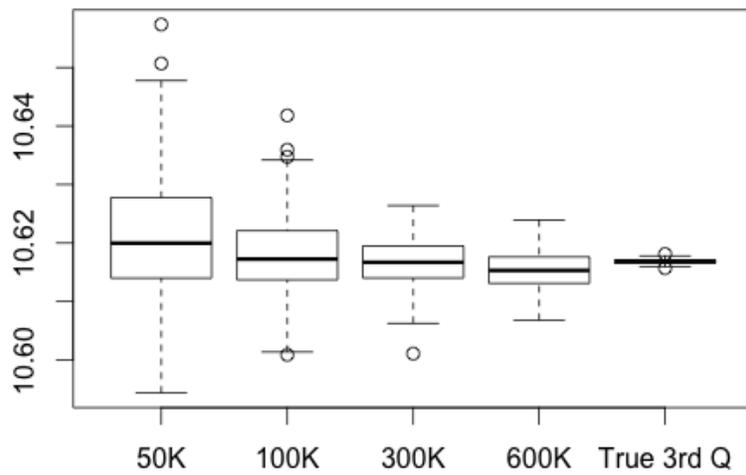**Box-Whisker Plot of 3rd Quartile via Eigenfunction**

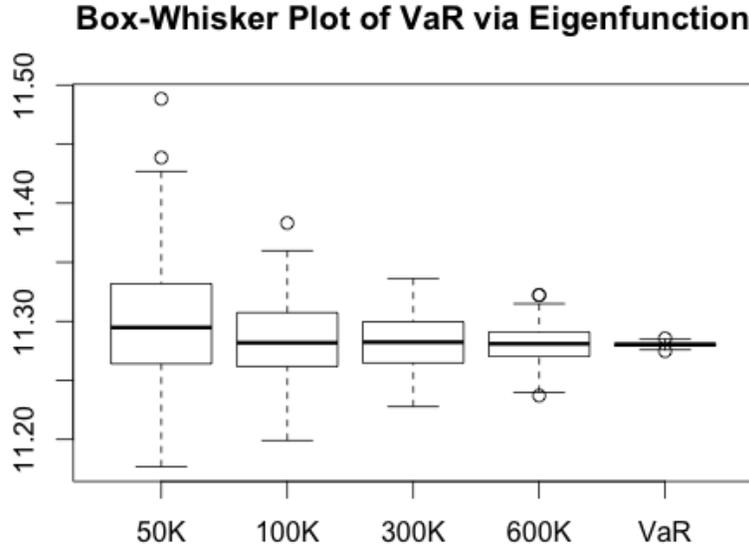Figure 5: Box Plot of 3rd Quartile, Eigenfunctions, $M = 5$

Figure 6: Box Plot of VaR at 99%, Eigenfunctions, $M = 5$

box-whisker plots for the 3rd quartile. Even though we can see that the estimate converges, there are systematic positive biases consistent with our results from Section 3.2. It seems that the bias is not disappearing fast. Obviously the estimator gets smaller bias as $N$ increases, but the speed of convergence is relatively slow compared to the case of mean and good relatively large amount of samples is necessary. However, the estimates are still relatively reliable since all estimates are in range of 0.01. It is important to note that the bias is not due to the number of basis functions used but due to the regression equation. In particular, we obtain similar results for other basis functions. One possible solution is to consider bias reduction methods such as the Jacknife method.

This observation comes forward as we move out in the tail, but in contrast to the previous situation, with a range of more than 0.05 it is about 5 times as wide as for the 3rd quartile. In Figure 6, we also see that the estimator has a systematic positive bias. The results from estimating 3rd quartile and VaR get worse because of difficulty of assessing the quantile based on finite samples. Nonetheless, for large $N$ we have still relatively reliable results. In particular, it is important to note that if we rely on the nested simulation, for instance, under $N = 600,000$ which is very restrictive computational budget, we may only have 600 outside simulations with 1,000 inner simulations. Then the estimated VaR will be considerably less reliable.

We test the convergence of VaR as $M$ increases. Figure 7 shows approximation for different number of basis functions based in the different classes. In Figure 7, we get pretty good results for all choices of basis functions where $M = 5$ or $M = 6$ appear to be the best choice overall. In particular, there seems to be a tradeoff between the accuracy of the functional approximation and the regression approximation. In particular, we see that the
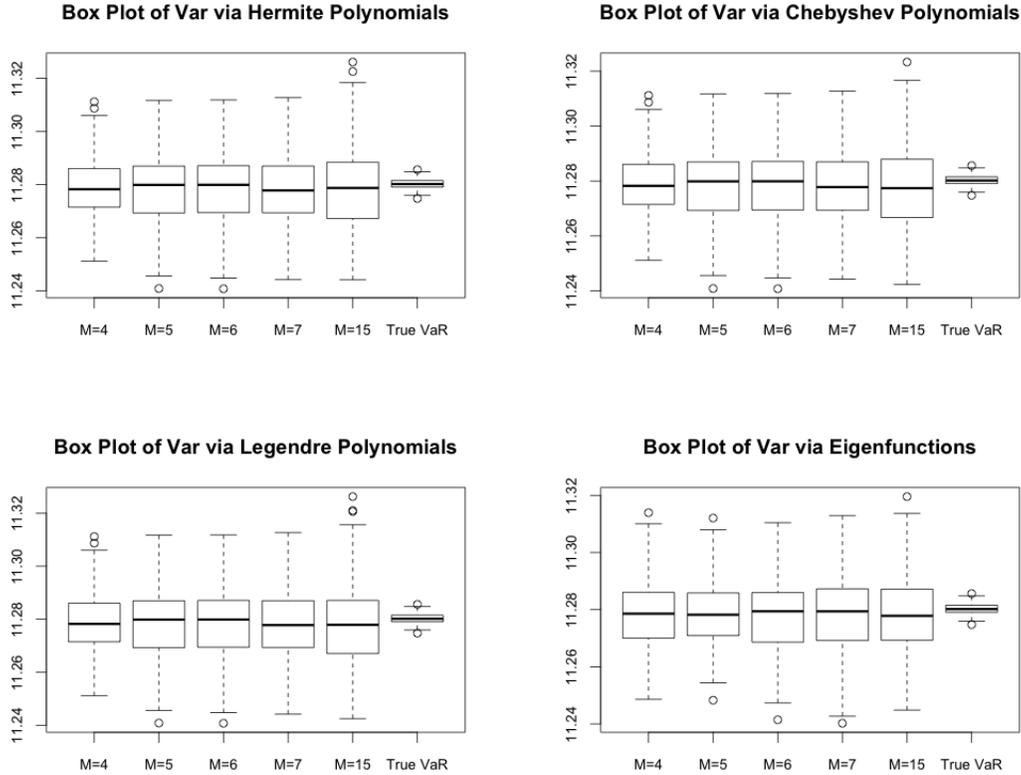
Figure 7: Box Plot for VaR at different Basis Functions, $N = 600,000$

variance of VaR is increased as increasing orders. However, at $M = 15$, the highest order, some estimates approach to the true value of VaR, which are relatively far from the true value of VaR at $M = 5$ or $M = 6$.

Finally, to compare the quality of the approximations based on the different basis functions, we consider the *statistical distances* here means measuring distance between two probability distribution. The first probability distribution is the "exact" distribution and the second distribution is the estimated one by the LSM. We implement the LSM two hundred times with $M = 5$, $M = 10$, and $M = 20$ with $N = 1,000,000$. Then we compute the mean of the distance. The statistical distance measure considered here are the Kullback-Liebler (KL) divergence,[14] the Jensen-Shannon (JS) divergence, and the Kolmogorov-Smirnov (KS) statistic.[15] When looking at the overall approximation of the overall distributions, we see

---

[14]The KL divergence is not symmetric, that is, $KL(A, B) \neq KL(B, A)$. We calculate $KL(A, B)$, where $A$ is the true distribution and $B$ is an estimated distributions, that is, $KL(A, B)$ measures the divergence from $A$ to $B$.

[15]Let $A$ and $B$ be samples from distribution $\mathbb{P}$ and $\mathbb{Q}$, respectively, and we have $n$ data points for each sample. We assume that whole samples are ordered increasingly. The $i$th sample of $A$ and $B$ are defined by $a_i$ and $b_i$. Also, we normalize each sample point by $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Each statistical distance between $\mathbb{P}$ and $\mathbb{Q}$ is calculated by the following:

- KL divergence, $KLD(\mathbb{P}, \mathbb{Q}) = \sum_i a_i' \log \frac{a_i'}{b_i'}$;

| Order | | Hermite | Chebyshev | Legendre | Eigenfunction |
|---|---|---|---|---|---|
| | KL diver. | $6.402 \times 10^{-8}$ | $6.402 \times 10^{-8}$ | $6.402 \times 10^{-8}$ | $4.904 \times 10^{-8}$ |
| $M = 5$ | JS diver. | $1.169 \times 10^{-4}$ | $1.169 \times 10^{-4}$ | $1.169 \times 10^{-4}$ | $1.024 \times 10^{-4}$ |
| | KS stats. | $4.945 \times 10^{-3}$ | $4.945 \times 10^{-3}$ | $4.945 \times 10^{-3}$ | $5.087 \times 10^{-3}$ |
| | KL diver. | $1.253 \times 10^{-7}$ | $1.261 \times 10^{-7}$ | $1.255 \times 10^{-7}$ | $9.919 \times 10^{-8}$ |
| $M = 10$ | JS diver. | $1.663 \times 10^{-4}$ | $1.671 \times 10^{-4}$ | $1.667 \times 10^{-4}$ | $1.482 \times 10^{-4}$ |
| | KS stats. | $6.061 \times 10^{-3}$ | $6.121 \times 10^{-3}$ | $6.147 \times 10^{-3}$ | $6.246 \times 10^{-3}$ |
| | KL diver. | $1.781 \times 10^{-7}$ | $1.789 \times 10^{-7}$ | $1.785 \times 10^{-7}$ | $1.324 \times 10^{-7}$ |
| $M = 20$ | JS diver. | $2.021 \times 10^{-4}$ | $2.026 \times 10^{-4}$ | $2.021 \times 10^{-4}$ | $1.720 \times 10^{-4}$ |
| | KS stats. | $7.419 \times 10^{-3}$ | $7.400 \times 10^{-3}$ | $7.444 \times 10^{-3}$ | $6.930 \times 10^{-3}$ |

Table 1: Statistical Distances between the Exact Distribution and the Estimated Distribution.

that eigenfunctions perform best. In particular, we can see that the distances between the distribution obtained by eigenfunctions and the exact distribution is smaller than for other basis functions if we use extremely small number of basis functions. So we expect that the choice of eigenfunctions is optimal when we have a highly constraint computation budget or a very high-dimensional problem.

Since we know the basis funtions are optimal not only optimal for our particular model but for *model framework*, this result is not to be driven by the specifics of our example. All in all, we document theory can help identifying good basis functions.

## 7    Conclusion

We provide a novel algorithm for estimating risk measures in "nested" settings, which provides 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 discuss how to estimate risk measures of losses and analyze convergence of the approach. Specifically, we show that the eigenfunctions of the infinitesimal generator of the underlying Markov process are optimal basis functions that minimize the distance between estimated and "true" distribution.

Our numerical analyses document that the algorithm can provide accurate results based on relatively few samples. Moreover, we show that the eigenfunctions are superior to any other orthonormal basis functions though all perform relatively well. We expect that the optimality of the eigenfunctions become important as the complexity and the dimensionality of the problem increase.

For future research, we intend to study basis functions if the infinitesimal generator is not self-adjoint and/or if there is an essential spectrum of the self-adjoint infinitesimal generator since this situation frequently arises when the underlying process is multi-dimensional

---

- JSD divergence, $JSD(\mathbb{P}, \mathbb{Q}) = \frac{1}{2}KLD(A, m) + \frac{1}{2}KLD(B, m)$, where each point, $m_i$, of $m$ is calculated by $m_i = \frac{1}{2}(a_i' + b_i')$;
- KS statistics, $KS(\mathbb{P}, \mathbb{Q}) = \sup_x |F_{A,n}(x) - F_{B,n}(x)|$, where $F_n$ is an empirical distribution.

application. Moreover, we intend to analyze the joint convergence of the functional and the regression approximation.

# Appendix

## A Proofs

***Proof of Lemma 2.1.*** 1. Let $A \in \mathcal{F}_t$, $0 \le t \le T_H$. Then

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

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

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

$\square$

***Proof of Lemma 2.2.*** Linearity is obvious. For the proof of continuity, consider a sequence $h^{(n)} \to h \in \mathcal{H}$. Then,

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

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

***Proof of Proposition 3.1.*** $\mathbb{P}_{Y_{T_H}}$ is a regular Borel measure as a finite Borel measure and hence $L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_{T_H}}\right)$ 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 $\lin\{e_k, \ k = 1, 2, \ldots, M\} = \lin S$. For $S$, on the other hand, we can find an orthonormal basis $\{f_k, \ k \in \mathbb{N}\} = S' \supset S$. Whence,

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

where

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

by Parseval's identity.

For the second part, we note that

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

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

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

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

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

$\square$

***Proof of Corollary 3.1.*** Relying on the notation from the proof of Proposition 3.1, we now have (supposing square integrability)

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

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,

$$\mathbb{E}^{\tilde{\mathbb{P}}} \left[\left| e'(Y_{T_H})\, \hat{\alpha}^{(N)} - e'(Y_{T_H})\, \alpha \right|\right]$$

$$\le \sum_{k=1}^{M} \mathbb{E}^{\tilde{\mathbb{P}}} \left[|e_k(Y_{T_H})\, (\hat{\alpha}_k - \alpha_k)|\right]$$

$$\le \sum_{k=1}^{M} \sqrt{\mathbb{E}^{\tilde{\mathbb{P}}} \left[e_k^2(Y_{T_H})\right]} \sqrt{\mathbb{E}^{\tilde{\mathbb{P}}} \left[\hat{\alpha}_k - \alpha_k\right]^2} \to 0, \ N \to \infty.$$

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

***Proof of Corollary 3.2.*** The first assertion immediately follows from convergence in distribution as discussed in Section 3. For the quantiles, the convergence for all continuity points of $F_{\mathrm{AC}_{T_H}}^{-1}$ follows from Proposition 3.1 and the standard proof of Skorokhod'd representation theorem (see e.g. Lemma 1.7 in Whitt (2002)). □

***Proof of Lemma 3.1.*** Relying on the notation from the proof of Proposition 3.1, let

$$PV_{T_H}^{(i)} = \sum_{k=T_H}^{T} x_k \left(Y_{T_H}^{(i)}\right)$$

$$= \sum_{j=1}^{M} \alpha_j e_j \left(Y_{T_H}^{(i)}\right) + \epsilon_j,$$

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

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

$$\sqrt{N}[\alpha - \hat{\alpha}^{(N)}] \longrightarrow \mathrm{Normal}\left[0, \underbrace{\left(A^{(M)}\right)^{-1}\left[\mathbb{E}^{\mathbb{P}}\left[e_k(Y_{T_H})e_l(Y_{T_H})\Sigma(Y_{T_H})\right]\right]_{1\leq k,l\leq M}\left(A^{(M)}\right)^{-1}}_{S}\right]$$

so that

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

where

$$\xi = e'Se \tag{15}$$

□

***Regularity Conditions on*** $g_N(\cdot,\cdot)$ ***(cf. Gordy and Juneja (2010)).*** We collect regularity conditions.

- 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 $p_{0,N}(\cdot)$, $p_{1,N}(\cdot)$, $p_{2,N}(\cdot)$ such that

  - $g_N(y,z) \leq p_{0,N}(z)$
  - $\left|\frac{\partial}{\partial y}g_N(y,z)\right| \leq p_{1,N}(z)$
  - $\left|\frac{\partial^2}{\partial y^2}g_N(y,z)\right| \leq p_{2,N}(z)$

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 $0 \leq r \leq 4$.

$\square$

**Proof of Proposition 4.1.** Let $\mathbf{P}^* = \sum_{n=1}^{M} <\cdot, \varphi_j> \varphi_j$ be the projection operator when choosing the eigenfunctions as the basis function. Then:

$$\inf_{\{e_1,...,e_M\}} \sup_{||\mathbf{x}||=1} ||\mathcal{P}_t\mathbf{x} - \mathbf{P}\mathcal{P}_t\mathbf{x}||^2 \leq \sup_{||\mathbf{x}||=1} ||\mathcal{P}_t\mathbf{x} - \mathbf{P}^*\mathcal{P}_t\mathbf{x}||^2$$

$$= \sup_{||\mathbf{x}||=1} || \sum_{n=M+1}^{\infty} e^{-\lambda_n t} <\mathbf{x}, \varphi_j> \varphi_j ||^2$$

$$= \sup_{||\mathbf{x}||=1} \sum_{n=M+1}^{\infty} e^{-2\lambda_n t} <\mathbf{x}, \varphi_j>^2 = e^{-2\lambda_{M+1}t}$$

On the other hand, consider an alternative set of basis functions $\{\varphi_1, ..., \varphi_M\}$. Then there exists an $f_0 \in \text{span}\{\varphi_1, ..., \varphi_M, \varphi_{M+1}\}$ with $\mathcal{P}_t f_0 \in \text{span}\{\varphi_1, ..., \varphi_M\}^{\perp}$. To see this, consider $\Phi_{M+1} = \{\varphi_1, ..., \varphi_{M+1}\}$. Taking $\mathcal{P}_t$ on this set yields $\mathcal{P}_t\Phi_{M+1} = \{e^{-\lambda_1 t}\varphi_1, ..., e^{-\lambda_{M+1}t}\varphi_{M+1}\}$ which is an orthogonal set of $M + 1$ elements. Since $\text{span}\{\varphi_1, ..., \varphi_M\}$ is of dimension $M$, such an $f_0$ exists. Therefore,

$$||\mathcal{P}_t - \sum_{n=1}^{M} <\mathcal{P}_t\cdot, \varphi_n> \varphi_n||^2 \geq \frac{||\mathcal{P}_t f_0 - \sum_{n=1}^{M} <\mathcal{P}_t f_0, \varphi_n> \varphi_n||^2}{||f_0||^2} = \frac{||\mathcal{P}_t f_0||^2}{||f_0||^2}$$

$$= \frac{\sum_{n=1}^{M+1} e^{-2\lambda_n t} |<f_0, \varphi_n>|^2}{\sum_{n=1}^{M+1} |<f_0, \varphi_n>|^2} \geq e^{-2\lambda_{M+1}t}$$

Hence,

$$\inf_{\{e_1,...,e_M\}} ||\mathcal{P}_t - \mathbf{P}\mathcal{P}_t||^2 = e^{-2\lambda_{M+1}t} = ||\mathcal{P}_t - \mathbf{P}^*\mathcal{P}_t||^2$$

$\square$

**Assumption A.1.** 
- $\left(PV_{T_H}^{(1)}, Y_{T_H}^{(1)}\right),..., \left(PV_{T_H}^{(N)}, Y_{T_H}^{(N)}\right)$ *are i.i.d. and* $Var(PV_{T_H}|Y_{T_H})$ *is bounded.*

- *Let* $p^M(Y_{T_H}) = (e_1(Y_{T_H}), ..., e_M(Y_{T_H}))'$ *denote a vector of approximating function. For every $M$ there is a nonsingular constant matrix $B$ such that for $P^M(Y_{T_H}) = Bp^M(Y_{T_H})$; (i) the smallest eigenvalue of $\mathbb{E}\left[P^M\left(Y_{T_H}^{(i)}\right) P^M\left(Y_{T_H}^{(i)}\right)'\right]$ is bounded away from zero uniformly in $M$ and; (ii) there is a sequence of constants $\zeta_0(M)$ satisfying $\sup_{y \in \mathcal{Y}} ||P^M(y)|| \leq \zeta_0(M)$ and $M = M(N)$ such that $\zeta_0(M)^2 M/N \rightarrow 0$ as $N \rightarrow \infty$ where $\mathcal{Y}$ is a support of $Y_{T_H}$, where $|| \cdot ||$ is an usual Euclidean matrix norm.*

*For further discussion, Newey introduces following notations. Let d be the dimension of $Y_{T_H}$ and $(\lambda_1, ..., \lambda_d)' = \lambda$ denote a vector of nonnegative integer, having the same dimension as $Y_{T_H}$. Also let $|\lambda| = \sum_{j=1}^d \lambda_j$ and v be any nonnegative integer. For a vector of function $h(x)$ define the vector of partial derivatives $\partial^\lambda h(x) = \partial^{|\lambda|} h(x)/\partial x_1^{\lambda_1} \cdots \partial x_d^{\lambda_d}$ and let $|g|_v = \max_{|\lambda| \leq v} \sup_{x \in \chi} |\partial^{|\lambda|} g(x)|$, where $\chi$ is a domain of x and the absence of an x argument for g denotes the entire function rather than its value at a point.*

- *For an integer $v \geq 0$ there are $\psi$ and $\alpha_k$, $k = 1, ..., M$ such that $|AC_{T_H} - \sum_{k=1}^M \alpha_k e_k|_v = O(M^{-\psi})$ as $M \to \infty$.*

*Also, Newey develop a new notation for the general convregence rate. The author defines,*

$$\zeta_v(M) = \max_{|\lambda| \leq v} \sup_{y \in \mathcal{Y}} ||P^M(y)||$$

**Assumption A.2.**    • *The support of $Y_{T_H}$ is a Cartesian product of compact connected intervals on which $Y_{T_H}$ has a probability density function that is bounded away from zero.*

- *$AC_{T_H}$ is continuously differentiable of order s on the support of $Y_{T_H}$.*

***Proof of Lemme 5.1.*** Note that the considered space is a separable Hilbert space. For this Hilbert space, there exists at least one complete orthonormal basis functions and the eigenfunctions forms such a basis system. Thefore, any element in the space can be written as in the following form,

$$f = \sum_{j=1}^\infty e^{-\lambda_j T_H} < f, \varphi_j > \varphi_j$$

and therefore, $\alpha_j$ is given by

$$\alpha_j = e^{-\lambda_j T_H} < f, \varphi_j >$$

We prove that $\left| AC_{T_H} - \widehat{AC}_{T_H}^{(M)} \right|_0 = O\left( \left( e^{T_H} \right)^{-\lambda_{M+1}} \right)$,

$$\left| AC_{T_H} - \widehat{AC}_{T_H}^{(M)} \right|_0 = \left| \sum_{j=1}^{\infty} e^{-\lambda_j T_H} < f, \varphi_j > \varphi_j - \sum_{j=1}^{M} e^{-\lambda_j T_H} < f, \varphi_j > \varphi_j \right|_0$$

$$= \left| \sum_{j=M+1}^{\infty} e^{-\lambda_j T_H} < f, \varphi_j > \varphi_j \right|_0$$

$$= \sup_{y \in \mathcal{Y}} \left| \sum_{j=M+1}^{\infty} e^{-\lambda_j T_H} < f, \varphi_j > \varphi_j(y) \right|$$

$$\leq \sum_{j=M+1}^{\infty} e^{-\lambda_j T_H} | < f, \varphi_j > | \sup_{y \in \mathcal{Y}} |\varphi_j(y)|$$

$$\leq \sum_{j=M+1}^{\infty} e^{-\lambda_j T_H} ||f|| \, ||\varphi_j|| \sup_{y \in \mathcal{Y}} |\varphi_j(y)| \quad \because \text{Cauchy Schwarz inequality}$$

$$= \sum_{j=M+1}^{\infty} e^{-\lambda_j T_H} ||f|| \sup_{y \in \mathcal{Y}} |\varphi_j(y)| \quad \because \varphi_j \text{ is normalized.}$$

$$= \sum_{j=M+1}^{\infty} e^{-\lambda_j T_H} C_1 \, C_{2,j}$$

$$= O\left( \left( e^{T_H} \right)^{-\lambda_{M+1}} \right)$$

where $C_1 = ||f||$ and $C_{2,j} = \sup_{y \in \mathcal{Y}} |\varphi_j(y)|$.

$\square$

***Proof of Proposition 5.3.*** The first assumption immediately follows from the LSM algorithm and $PV_{T_H}^{(i)} \in \mathcal{H}$.

Let $p^M(y) = (\varphi_1(y), ..., \varphi_M(y))'$. Setting $B = I$ yields

$$\mathbb{E}\left[ P^M\left( Y_{T_H}^{(i)} \right) P^M\left( Y_{T_H}^{(i)} \right)' \right] = \mathbb{E}\left[ p^M\left( Y_{T_H}^{(i)} \right) p^M\left( Y_{T_H}^{(i)} \right)' \right]$$

$$= I$$

due to the orthonormality of eigenfunctions. Therefore, the smallest eigenvalue of $\mathbb{E}\left[ P^M\left( Y_{T_H}^{(i)} \right) P^M\left( Y_{T_H}^{(i)} \right)' \right]$ is bounded away from zero uniformly for every $M$. Moreover, for fixed $y \in \mathcal{Y}$,

$$||P^M(y)|| = \sqrt{\varphi_1(y)^2 + \cdots + \varphi_M(y)^2}.$$

Then

$$\sup_{y\in\mathcal{Y}}||P^M(y)|| = \sup_x \sqrt{\varphi_1(y)^2 + \cdots + \varphi_M(y)^2}$$

$$\leq \sqrt{\sum_{j=1}^{M} \sup_{y\in\mathbb{Y}} \varphi_j(y)^2}$$

$$\leq \sqrt{\max_{1\leq j\leq M} sup_{y\in\mathcal{Y}}\varphi_j(y)^2 \times M}$$

$$= C\sqrt{M} = \zeta_0(M)$$

where $C = \sqrt{\max_{1\leq j\leq M} sup_{y\in\mathcal{Y}}\varphi_j(y)^2}$. We choose $N$ to have $M^2/N \to 0$ as $N \to \infty$. We prove that the framework satisfy the first and second assumptions of A.1. Also, error of functional approximation is given by $O\left(\left(e^{T_H}\right)^{-\lambda_{M+1}}\right)$ from Lemma 5.1. Therefore, from Newey (1997), we have Proposition 5.3. $\qquad\square$

**Proof of Proposition 5.2.**

$$\mathbb{E}^{\tilde{\mathbb{P}}}\left[\left|\mathrm{AC}_{T_H} - \widehat{\mathrm{AC}}_{T_H}(Y_{T_H})\right|\right] = \mathbb{E}^{\tilde{\mathbb{S}}}\left[\frac{\tilde{\mathbb{P}}}{\tilde{\mathbb{S}}}\left|\mathrm{AC}_{T_H} - \widehat{\mathrm{AC}}_{T_H}(Y_{T_H})\right|\right]$$

$$\leq \sqrt{\mathbb{E}^{\tilde{\mathbb{S}}}\left[\frac{\tilde{\mathbb{P}}^2}{\tilde{\mathbb{S}}^2}\right]}\sqrt{\mathbb{E}^{\tilde{\mathbb{S}}}\left[\left|\mathrm{AC}_{T_H} - \widehat{\mathrm{AC}}_{T_H}(Y_{T_H})\right|^2\right]} \quad \because \text{Cauchy Schwarz inequality}$$

$$\leq C \cdot O_p(M/N + \left(e^{T_H}\right)^{-2\lambda_{M+1}})$$

$$= O_p(M/N + \left(e^{T_H}\right)^{-2\lambda_{M+1}})$$

where $C = \sqrt{\mathbb{E}^{\tilde{\mathbb{S}}}\left[\frac{\tilde{\mathbb{P}}^2}{\tilde{\mathbb{S}}^2}\right]}$. $\qquad\square$

# References

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

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

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

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

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.

Brodie, M., Du, Y., Moallemi, C.C, 2012. Risk Estimation via Regression. Working Paper, Columbia University.

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

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

Davydov, D., Linetsky, V., 2003. Pricing Options on Scalar Diffusions: An Eigenfunction Expansion Approach. Operations Research, 51: 185-209.

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

Haefeli Ed., D., The Geneva Association, Geneva, 2013. Variable Annuities - An Analysis of Financial Stability. `https://www.genevaassociation.org/media/99582/GA2013-Variable_Annuities.pdf`.

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

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

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

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.

Linetsky, V., 2008. Spectral Methods in Derivatives Pricing. Handbooks in OR and MS, 15: Chapter 6, 223-299.

Linetsky, V., 2004. The Spectral Decomposition of the Option Value. International Journal of Theoretical ad Applied Finance, 7: 337-384.

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. `http://us.milliman.com`.

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.

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

Convergence rates for series estimators. In: Maddalla, G.S., Phillips, P.C.B., Srinavasan, T.N. Eds. Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao, Blackwell, Cambridge, USA: 254-275.

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.

Schroder, M., 1989. Computing the constant elasticity of variance option pricing formula. Journal of Finance, 40: 211-219.

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

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/`.

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