A Least-Squares Monte Carlo Approach to the Estimation of Enterprise Risk*

Hongjun Ha

Department of Mathematics; Saint Joseph's University

5600 City Avenue, Philadelphia, PA 19131. USA

Daniel Bauer†

Department of Risk and Insurance; University of Wisconsin-Madison

975 University Avenue, Madison, WI 53706. USA

First version: September 2013. This version: June 2019.

Abstract

The estimation of enterprise risk for financial institutions 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. Relying on well-known ideas for pricing non-European derivatives, the current paper discusses tackling this *nested* valuation problem based on Monte Carlo simulations and least-squares regression techniques. We formalize and analyze the algorithm in an operator setting. Importantly, we address the problem of how to choose the regressors ("basis functions"), and show that a *robust* choice is given by the left singular functions of the corresponding conditional expectation operator. Our numerical examples demonstrate that the algorithm can produce accurate results at relatively low computational costs, particularly when relying on robust basis functions.

Keywords: risk management, least-squares Monte Carlo, basis functions, Variable Annuity with GMB.

*This paper extends an earlier working paper Bauer et al. (2010). We thank Giuseppe Benedetti, Enrico Biffis, René Carmona, Matthias Fahrenwaldt, Jean-Pierre Fouque, Andreas Reuss, Daniela Singer, Ajay Subramanian, Baozhong Yang, and seminar participants at the Bachelier Congress 2014, the World Risk and Insurance Economics Congress 2015, the 2017 Conference Innovations in Insurance Risk and Asset Management at TU Munich, the 2017 Conference for the 10th Anniversary of the Center for Financial Mathematics and Actuarial Research at USCB, Georgia State University, Michigan State University, St. Joseph's University, Université de Montréal, and Barrie & Hibbert for helpful comments. The usual disclaimer applies.

†Corresponding author. Phone: +1-608-265-4119. E-mail addresses: hha@sju.edu (H. Ha); daniel.bauer@wisc.edu (D. Bauer).

1 Introduction

Many financial risk management applications entail a reevaluation of the company's assets and liabilities at some time horizon τ – sometimes called a *risk horizon* – for a large number of realizations of economic and firm-specific (state) variables. The resulting empirical distribution of firm economic capital is used for managing enterprise risk (Nocco and Stulz, 2006) and also for the derivation of enterprise risk measures such as Value-at-Risk (VaR) or Expected Shortfall (ES), which serve as the basis for capital requirements within several regulatory frameworks such as Solvency II for insurance companies and Basel III.5 for banks, respectively. However, the high complexity of this *nested* computation structure leads firms to struggle with the implementation.

This paper discusses an approach to this problem based on least-squares regression and Monte Carlo simulations akin to the well-known Least-Squares Monte Carlo method (LSM) for pricing non-European derivatives introduced by Longstaff and Schwartz (2001). Analogously 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 risk-neutral conditional expected value at the risk horizon τ, 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 the coefficients in this linear combination. Hence, for each realization of the state variables, the resulting linear combination presents an approximate realization of the capital at τ, and the resulting sample can be used for estimating 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; Nikolić et al., 2007) and has been used in several applied research contributions (Floryszczak et al., 2016; Pelsser and Schweizer, 2016; Krah et al., 2018, e.g.), these papers do not provide a detailed analysis or insights on how to choose 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. 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.

The joint convergence results illustrate the interplay between the two approximation steps. In particular, when increasing the number of basis functions in the functional approximation, it will be necessary to simultaneously increase the number of sample paths used in the regression approximation – which is the main source of complexity in high-dimensional enterprise risk models. Thus, choosing adequate basis functions is of crucial importance in such settings.

This is where our operator formulation becomes especially useful. By relying on representation results from functional analysis, we show that under certain conditions, the (left)

singular functions of the valuation operator present a *robust* choice for the basis functions. More precisely, we demonstrate that these singular functions yield the best approximation for the company's capital across all possible cash flow profiles given the company's risk model, in the minimax sense. The intuition is that akin to the singular value decomposition (SVD) for a matrix, the singular functions provide the most important dimensions in spanning the image space of the valuation operator.

The availability of robust basis functions is desirable in the risk management context for a variety of reasons. First, unlike pricing applications of LSM, enterprise risk models usually entail dozens of stochastic risk factors driving the asset and liability sides of the balance sheet, so that a systematic and parsimonious choice is crucial. Furthermore, again unlike pricing applications, the same firm-wide scenarios are used for evaluating the company's assets and liability portfolios, which contain a multitude of different cash flow profiles. Separate consideration of sub- or perturbed portfolios is common for calculating capital allocations, and positions will also adjust over time, whereas the underlying risk model typically remains in place. Thus tying the choice of the basis functions to the model framework rather than a particular payoff function is expedient.

To operationalize our ideas, we first discuss the calculation of the singular values of our valuation operator – and, thus, the derivation of robust basis functions – for models with a single payoff date and (multivariate) Gaussian transition densities. In this case, it is straightforward to show that the underlying assumptions are satisfied. And, 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. While these assumptions are rather restrictive, it is possible to rely on the analogous approach even in general situations with an arbitrary number of payment dates and arbitrary risk driver distributions, by solely considering the first and second moments. This is inspired by Discriminant Analysis for classification problems, where the class densities of the features are approximated by a multivariate Gaussian distribution focussing on the first and second moments (Hastie et al., 2009, Sec. 4.3). In particular, this approach circumvents potentially tedious calculations associated with a the numerical derivation of singular values in general models.

We illustrate our theoretical results considering popular annuitization guarantees within Variable Annuity contracts and three stochastic risk drivers (fund risk, interest rate risk, and mortality risk), both in a Gaussian and a general (non-Gaussian) setting. Thus, while the application does not match the complexities present in real-world enterprise risk applications, it serves to establish some relevant insights. We demonstrate that robust basis functions uniformly outperform the conventional choice of polynomials. In particular, two points deserve emphasis. First, it is important to note that each robust basis function is a linear combination of *all* stochastic states, and their order is determined uniquely by each functions relevance in spanning the valuation operator. No "intuitive" ad-hoc choices are necessary on which components are most relevant and how to sequence higher-order terms. This aspect is very relevant in practical settings with high-dimensional state vectors, so that our results provide immediate guidance for these pressing problems. Second, the derivation of approximations to robust basis functions in general settings is straightforward and only entails matrix calculations using first and second moments of the risk drivers, which can be easily estimated from the underlying simulations. Hence, there is little downside associated

with this approach relative to conventional choices of polynomial families.

Related Literature and Organization of the Paper

Our approach is inspired by the LSM approach for derivative pricing (Carriere, 1996; Tsitsiklis and Van Roy, 2001; Longstaff and Schwartz, 2001; Clément et al., 2002) and in principle our results also apply there, although – as argued above – we believe our results are particularly relevant in the risk management context. A similar regression-based algorithm for risk estimation is independently studied in Broadie et al. (2015). Their results are similar to our sequential convergence results in Section 3.1, and the authors additionally introduce a *weighted* version of their regression algorithm. Moreover, Benedetti (2017) provides joint convergence results under an alternative set of conditions. However, these authors do not contemplate how to choose the basis functions – although they emphasize the importance of this choice – which is a key contribution of our paper.

We refer to Makur and Zheng (2016) for the relevance of the SVD of conditional expectations in the information theory literature, which is driven by similar considerations. In particular, the authors derive the analogous SVD for the Gaussian setting in the univariate case (see also Abbe and Zheng (2012)). The relevance of Hermite polynomials in this context may not come as a surprise from a stochastic process perspective due to their relevance in the spectral analysis of the Ornstein-Uhlenbeck semigroup (Linetsky, 2004). However, as detailed in Makur and Zheng (2016, p. 636), we note that the setting here is distinct from Markov semigroup theory, where the relevant spaces are framed in terms of invariant measures and the time interval varies.

As already indicated, the LSM approach enjoys popularity in the context of calculating (life) insurance economic capital in practice and applied research, so that providing a theoretical foundation and guidance for these applications are key motivating factors for this paper. A number of competing algorithms have been proposed, including variants of the basic nested simulations approach (Gordy and Juneja, 2010, e.g.), the so-called replicating portfolio approach (Cambou and Filipović, 2018, e.g.), or non-parametric smoothing approaches in the nested simulations context (Liu and Staum, 2010; Hong et al., 2017; Risk and Ludkovski, 2018). While a detailed comparison is beyond the scope of this paper, some authors note deficiencies of some of these approaches in high-dimensional applications such as enterprise risk measurement, which is our focus (e.g., Bauer et al. (2012b) for nested simulations, Pelsser and Schweizer (2016) and Ha (2016) for replicating portfolios, and Hong et al. (2017) for non-parametric approaches).

The remainder of the paper is structured as follows: Section 2 lays out the simulation framework and the algorithm; Section 3 addresses convergence of the algorithm; Section 4 introduces the notion of robust basis functions and their derivation; Section 5 provides our numerical examples; and, finally, Section 6 concludes and points out avenues for future research. Proofs and technical details as well as some supplemental analyses are relegated to the Appendix.

2 The LSM Approach

2.1 Simulation Framework

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 T corresponds to the longest-term asset or liability of the company in view and $\mathbb{P}$ 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 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 financial assets in the market can be expressed in terms of Y. Non-financial risk factors can also be incorporated. In this market, we take for granted the existence of a risk-neutral probability measure (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 $(N_t)_{t\in[0,T]}$.

In financial risk management, we are now concerned with the company's financial situation at a certain (future) point in time τ, $0 < \tau < T$, 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_τ at time τ calculated as the market value of assets minus liabilities. This amount can serve as a buffer against risks and absorb financial losses, and thus describes the enteprise-wide risk situation. It also serves to define *capital requirements* via a risk-measure ρ. For instance, if the capital requirement is cast based on VaR, the capitalization at time τ should be sufficient to cover the net liabilities at least with a probability α, i.e., the additionally required capital is:

$$\mathrm{VaR}_\alpha(-C_\tau) = \inf\left\{ x \in \mathbb{R} | \, \mathbb{P}\left(x + C_\tau \geq 0\right) \geq \alpha \right\}. \tag{1}$$

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 embedded options typically require 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 τ, i.e., we face a *nested* valuation 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 discrete times $t = 1, 2, \ldots, T$ and that $\tau \in \{1, 2, \ldots, T\}$:

$$C_\tau = \mathbb{E}^{\mathbb{Q}}\left[\sum_{k=\tau}^{T} \frac{N_\tau}{N_k} X_k \middle| (Y_s)_{0\leq s\leq \tau} \right]. \tag{2}$$

Note that within this formulation, interim asset and liability cash flows in $[0, \tau]$ may be

aggregated in the $\sigma(Y_s, 0 \leq s \leq \tau)$-measurable position X_τ. Moreover, in contrast to, e.g., Gordy and Juneja (2010), we consider aggregate asset and liability cash flows at times $k \geq \tau$ 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., 2012b).

In this *risk measurement* setting, the relevant probability measure is the physical measure $\mathbb{P}$ until the risk horizon τ, and the risk-neutral measure $\mathbb{Q}$ after τ. In particular, a sample path of the state process Y will typically be generated using $\mathbb{P}$ over $[0, \tau]$ and $\mathbb{Q}$ over $(\tau, T]$. To streamline our discussion, we 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 2.1. *We have:*

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

Hence, $\tilde{\mathbb{P}}$ measures according to $\mathbb{P}$ over $[0, \tau]$ but coincides with $\mathbb{Q}$ when conditioning on the risk horizon, so that we can generally operate under $\tilde{\mathbb{P}}$ in what follows.

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 (see also Sec. 5.3). 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, with Lemma 2.1, we obtain:

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

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:[1]

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

[1]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 5), we assume that the discount factor can be expressed as a function of the state variables.

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:

$$C_\tau(Y_\tau) = \sum_{j=\tau}^{T} \mathbb{E}^{\tilde{\mathbb{P}}}[x_j(Y_j)|\, Y_\tau] = L\,\mathbf{x}(Y_\tau), \tag{3}$$

where the operator:

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

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

Lemma 2.2. *L is a continuous linear operator.*

Moreover, for our results on the robustness of basis functions, we require compactness of the operator L. The following lemma provides a version of the well-known Hilbert-Schmidt condition for L to be compact in terms of the transition densities (Breiman and Friedman, 1985):

Lemma 2.3. *Assume there exists a joint density $\pi_{Y_\tau,Y_j}(y,x)$, $j = \tau, \tau+1, ..., T$, for Y_τ and Y_j. Moreover, assume $\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \pi_{Y_j|Y_\tau}(y|x)\,\pi_{Y_\tau|Y_j}(x|y)\,dy\,dx < \infty$, where $\pi_{Y_j|Y_\tau}(y|x)$ and $\pi_{Y_\tau|Y_j}(x|y)$ 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 (state-dependent) capital C_τ can be 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 $\mathbf{x} = (x_\tau, \ldots, x_T)$, although the *model* may change between applications as the exposures may change (e.g., from one year to the next or when evaluating capital allocations). The resulting risk distribution can then be, e.g., applied to 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 $\rho(L\mathbf{x})$.

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_τ under $\mathbb{P}$ and then, for each of these realizations, to determine the available capital using another simulation step under $\mathbb{Q}$ (Lee, 1998; Gordy and Juneja, 2010). However, this approach is computationally burdensome and, for some relevant applications, may require a very large number of simulations to obtain results in a reliable range (Bauer et al., 2012b). Hence, in the following, we develop an alternative approach for such situations.

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_τ 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" (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_τ, by a finite combination of linearly independent basis functions $e_k(Y_\tau) \in L^2\left(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau}\right)$:

$$C_\tau \approx \widehat{C}_\tau^{(M)}(Y_\tau) = \sum_{k=1}^{M} \alpha_k \cdot e_k(Y_\tau). \tag{5}$$

We then determine approximate $\mathbb{P}$-realizations of C_τ 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 $t \in (\tau, T]$.[2] Based on these paths, we calculate the realized cumulative discounted cash flows:

$$V_\tau^{(i)} = \sum_{j=\tau}^{T} x_j\left(Y_j^{(i)}\right),\ 1 \le i \le N. \tag{6}$$

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

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

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

$$C_\tau \approx \widehat{C}_\tau^{(M)}(Y_\tau) \approx \widehat{C}_\tau^{(M,N)}(Y_\tau) = \sum_{k=1}^{M} \hat{\alpha}_k^{(N)} \cdot e_k(Y_\tau). \tag{7}$$

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

[2]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.

3 Convergence of the Algorithm

3.1 Sequential Convergence

The following result establishes convergence of the LSM algorithm when taking limits sequentially:

Proposition 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 (26) 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. The primary point of Proposition 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_τ 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 orthonormal basis functions, we are able to present a more concrete result:

Corollary 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 ρ 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, the algorithm produces a consistent estimate of the available capital distribution and for capital requirements, at least for certain classes of risk measures ρ. That is, if N and M are chosen *large enough*, $\hat{C}_\tau^{(M,N)}$ and $\rho(\hat{C}_\tau^{(M,N)})$ presents viable approximations for C_τ and $\rho(C_\tau)$, respectively. In the next part, we make more precise what *large enough* means and, particularly, how large N needs to be chosen relative to M.

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 $V_\tau = \sum_{j=1}^T x_j(Y_j)$ – 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:[3]

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

Proposition 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 M;*
- *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_τ.*

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:

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

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 ψ 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 3.1 it goes to zero linearly in N.

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 highly relevant case of power series. In our notation, they read:

Proposition 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_τ 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 smoothness of the conditional expectation is important – which is not surprising given Jackson's inequality. 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 state dimension. Clearly, a more customized choice of basis functions may improve on this.

We note that although M/N enters the convergence rate, the general conditions require $\xi_0(M)^2\, M/N \to 0$ in general and $M^3/N \to 0$ for orthonormal polynomials, effectively to control for the influence of estimation errors in the empirical covariance matrix of the regressors. Moreover, for common financial models the assumption of a bounded conditional

variance or bounded support of the stochastic variables are not satisfied. Benedetti (2017) shows that if the distribution of the state process is known, convergence can still be ensured at a rate of $M^2 \log\{M\}/N \to 0$ under more modest – and in financial contexts appropriate – conditions. We refers to his paper for details.

An important risk measure that does not fall in the class of convex risk measures – so that Corollary 3.1 does not apply – is VaR, since it is the risk measure applied in regulatory frameworks, particularly Solvency II. However, convergence immediately follows from Propositions 3.1-3.3:

Corollary 3.2. *We have:*

$$F_{\widehat{C}_\tau^{(M,N)}}(l) = \mathbb{P}(\widehat{C}_\tau^{(M,N)} \leq l) \to \mathbb{P}(C_\tau \leq l) = F_{C_\tau}(l),\ N \to \infty,\ M \to \infty,\ l \in \mathbb{R},\ \textit{and}$$
$$F^{-1}_{\widehat{C}_\tau^{(M,N)}}(\alpha) \to F^{-1}_{C_\tau}(\alpha),\ N \to \infty,\ M \to \infty,$$

for all continuity points $\alpha \in (0,1)$ of $F^{-1}_{C_\tau}$. Moreover, under the conditions of Propositions 3.2 and 3.3, we have joint convergence.

Regarding the properties of the estimator beyond convergence, much rides on the quality of the two approximations. 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)). While this does not necessarily mean that $\rho(\widehat{C}_\tau^{(M,N)})$ is unbiased for $\rho(\widehat{C}_\tau^{(M)})$, typically the approximation error due to variance overshadows potential bias terms. For instance, following Gordy and Juneja (2010), it is straightforward to show based on Proposition 3.1 that a bias term will enter the mean squared error as $O(N^{-2})$, whereas the convergence of the variance is of order N. The quality of the first (functional) approximation, which is at the core of the proposed algorithm, directly relates to the choice of basis functions.

4 Choice of Basis Functions

As demonstrated in Section 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, the typical choice for basis functions are polynomial families. And while the choice is important for the pricing approximation (Glasserman, 2004, Sec. 8.6), several authors conclude based on numerical tests that the approach appears robust for typical problems when including a sufficiently large number of terms (see, e.g., Moreno and Navas (2003) and also the original paper by Longstaff and Schwartz (2001)). As we argue in the Introduction, a number of aspects differ in the risk measurement context considered here: enterprise risk models usually are high-dimensional, so that including a sufficiently large number of higher-ordered terms of all variables may not be feasible; also, the framework will be used for the evaluation of numerous different positions, so that tailoring the functional approximation to a particular payoff is not possible. Therefore, the choice of basis functions is not only potentially more complex but also more crucial in the present context.

4.1 Robust Basis Functions for a Model Framework

As illustrated in Section 2.1, we can identify capital – as a function of the state vector at the risk horizon Y_τ – for a cash flow *model* $\mathbf{x}$ within a certain *model framework* Y with the output of the linear operator L applied to $\mathbf{x}$: $C_\tau(Y_\tau) = L\mathbf{x}(Y_\tau)$ (Eq. (3)). As discussed in Section 3.2, the LSM algorithm, in turn, approximates C_τ 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:

$$\begin{aligned} L_F\mathbf{x} &= P\,L\,\mathbf{x} = \sum_{k=1}^{M} \langle L\mathbf{x}, e_k(Y_\tau)\rangle_{L^2(\mathbb{P}_{Y_\tau})}\, e_k \qquad (8)\\ &= \sum_{k=1}^{M} \mathbb{E}^{\mathbb{P}}\left[e_k(Y_\tau)\sum_{j=\tau}^{T}\mathbb{E}^{\tilde{\mathbb{P}}}\left[x_j(Y_j)|Y_\tau\right]\right]e_k = \sum_{k=1}^{M}\mathbb{E}^{\tilde{\mathbb{P}}}\left[e_k(Y_\tau)\underbrace{\sum_{j=\tau}^{T}x_j(Y_j)}_{=V_\tau}\right]e_k \\ &= \sum_{k=1}^{M}\underbrace{\mathbb{E}^{\tilde{\mathbb{P}}}\left[e_k(Y_\tau)\,V_\tau\right]}_{\alpha_k}e_k, \end{aligned}$$

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 *robust basis functions* as a system that minimizes the distance between L and L_F, so that the approximation is minimal with regards to all possible cash flow models within the framework (minimax robustness):

Definition 4.1. *We call the set of basis functions* $\{e_1^*, e_2^*, ..., e_M^*\}$ *robust in* $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$ *if:*

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

As outlined above, choosing *robust* basis functions is particularly expedient 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 purpose 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 an economic capital framework is exactly the enterprise-wide determination of diversification opportunities and systematic risk factors. Also, it is typically not frequently revised. Hence, it is expedient here to connect the choice of basis functions to the framework rather than a particular model (payoff).

4.2 Robust Basis Functions for a Compact Valuation Operator

In order to derive robust 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 approximation 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:

$$L\,\mathbf{x} = \sum_{k=1}^{\infty} \omega_k \,\langle \mathbf{x}, s_k \rangle\, \varphi_k, \tag{9}$$

where $\{\omega_k\}$ with $\omega_1 \geq \omega_2 \geq \dots$ 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 $L\,L^*$. The following proposition demonstrates that robust basis functions are given by the left singular functions of L.

Proposition 4.1. *Assume the operator L is compact. Then for each M, the left singular functions of L $\{\varphi_1, \varphi_2, \dots, \varphi_M\} \in L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$ are robust basis functions in the sense of Definition 4.1. For a fixed cash flow model, we obtain $\alpha_k = \omega_k \langle \mathbf{x}, s_k \rangle$.*

Our finding that the left singular functions provide an robust approximation is related to familiar 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 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 robust basis functions. Here, in general, we have to check the conditions from Newey's convergence result (Prop. 3.2). We observe that the convergence rate associated with the first (functional) approximation depends on the parameter ψ, which in the present context derives from the speed of convergence of the singular value decomposition:

$$\begin{aligned} O(M^{-\psi}) = \inf_{\alpha_M} \sup_{y \in \mathcal{Y}} |C_\tau(y) - e^{(M)}(y)'\,\alpha_M| &\leq \sup_{y \in \mathcal{Y}} |\,L\,\mathbf{x}\,(y) - L_F\,\mathbf{x}\,(y)| \\ &= \sup_{y \in \mathcal{Y}} \left| \sum_{k=M+1}^{\infty} \omega_k \,\langle \mathbf{x}, s_k \rangle\, \varphi_k(y) \right|. \end{aligned} \tag{10}$$

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

Proposition 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_τ. 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 $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$.

Comparing this convergence rate for singular functions to the general case from Proposition 3.2 and the orthonormal polynomial case from Proposition 3.3, we notice that the second term associated with the first (functional) approximation now is directly linked to the decay of the singular values. For integral operators, this rate 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). In any case, Equation (10) that directly enters Newey's convergence result illustrates the intuition behind the robustness 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 the next subsection, we determine the SVD – and, thus, robust basis functions – in the practically highly relevant case of Gaussian transition densities. Here, robust basis functions correspond to Hermite polynomials of suitably transformed state variables and the singular values decay exponentially for $d = 1$ (Proposition 4.3), demonstrating the merit of this choice.

4.3 Robust Basis Functions for Gaussian Transition Densities

In what follows, we consider a single cash flow at time T only, and we assume that (Y_τ, Y_T) are jointly Gaussian distributed. We denote the $\widetilde{\mathbb{P}}$-distribution of this random vector via:

$$\begin{pmatrix} Y_\tau \\ Y_T \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_\tau \\ \mu_T \end{pmatrix}, \quad \begin{pmatrix} \Sigma_\tau & \Gamma \\ \Gamma' & \Sigma_T \end{pmatrix} \right], \tag{11}$$

where μ_τ, μ_T, Σ_τ, and Σ_T are the mean vectors and variance-covariance matrices of Y_τ and Y_T, respectively, and Γ is the corresponding (auto) covariance matrix – which we assume to be non-singular.

Denoting by $g(x; \mu, \Sigma)$ the normal probability density function at x with mean vector μ and covariance matrix Σ, the marginal densities of Y_τ 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 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 robust basis functions, the objective is to derive the SVD of L.

Lemma 4.1. *We have for the conditional distributions:*

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

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 4.1, robust basis functions are given by the left singular functions, which are in turn the eigenfunctions of $L\,L^*$. We obtain:

Lemma 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 4.3. *Denote by $P\Lambda P'$ 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 Λ is the diagonal matrix whose 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:

$$z^P(y) = P'\Sigma_\tau^{-1/2}(y - \mu_\tau). \tag{12}$$

Then for $Y \sim K_A(x, \cdot)$, we have: $\mathbb{E}_{K_A}\left[z^P(Y)|x\right] = \Lambda\, z^P(x)$, $Var_{K_A}\left[z^P(Y)|x\right] = I - \Lambda^2$, $\mathbb{E}_{\pi_{Y_\tau}}\left[z^P(Y_\tau)\right] = 0$, 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:

$$z^Q(x) = Q'\Sigma_T^{-1/2}(x - \mu_T). \tag{13}$$

Then for $X \sim K_B(y, \cdot)$, we have: $\mathbb{E}_{K_B}\left[z^Q(X)|y\right] = \Lambda\, z^Q(y)$, $Var_{K_B}\left[z^Q(X)|y\right] = I - \Lambda^2$, $\mathbb{E}_{\pi_{Y_T}}\left[z^Q(Y_T)\right] = 0$, 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 $K_A(x, \cdot)$, 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 4.3:

Proposition 4.3. *Denote the Hermite polynomial of degree j by $h_j(x)$ (Kollo and Rosen, 2006):*

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

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

$$\omega_m = \Pi_{i=1}^d \lambda_i^{k_i/2}, \tag{14}$$

where $m = (k_1, ..., k_d) \in \mathbb{N}_0^d$ is a multi-index, $\mathbb{N}_0^d$ is the set of d-dimensional non-negative integers, and the corresponding right and left singular functions are:

$$s_m(x) = \Pi_{i=1}^d h_{k_i}(z_i^Q(x)) \text{ and } \varphi_m(y) = \Pi_{i=1}^d h_{k_i}(z_i^P(y)),$$

respectively.

Combining the insights from Proposition 4.1 and Proposition 4.3, we immediately obtain:

Corollary 4.1. *Let $(m_k)_{k\in\mathbb{N}}$ be a reordering of $\{m\} = \{(k_1, ..., k_d) \in \mathbb{N}_0^d\}$ such that: $\omega_{m_1} \geq \omega_{m_2} \geq \omega_{m_3} \geq \dots$ Then, in the current setting, a robust choice for the basis functions for the LSM algorithm in the sense of Definition 4.1 is given by:*

$$\varphi_k = \varphi_{m_k},\ k = 1, 2, 3, \dots$$

In the univariate case ($d = 1$), $A = \lambda_1$ is the square of the correlation coefficient between Y_τ and Y_T – so that the singular values are simply powers of this correlation, decaying exponentially. Thus, the SVD takes the form (Abbe and Zheng, 2012):

$$L\,\mathbf{x}(Y_\tau) = \sum_{k=1}^{\infty} \left(\mathrm{Corr}(Y_\tau, Y_T)\right)^{k-1} \left\langle x_T, h_{k-1}\left(\frac{Y_T - \mu_T}{\Sigma_T^{1/2}}\right)\right\rangle_{\pi_{Y_T}} h_{k-1}\left(\frac{Y_\tau - \mu_\tau}{\Sigma_\tau^{1/2}}\right).$$

In particular, robust basis functions are 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 4.3 that the singular values of L are directly related to eigenvalues of the matrix A (or, equivalently, B), and there are $\binom{d-1+l}{d-1}$ vectors of indices m such that $\sum_i k_i = l$ in Equation (14) (stars and bars problem). The order of these singular values will determine the order of the singular functions in the SVD (9). In particular, after $\varphi_1(x) = 1$ with coefficient equaling $\langle x_T, 1\rangle = \mathbb{E}[x_T]$, the first nontrivial basis function is given by the singular function associated with the largest singular value – which according to (12) is a component of the linearly transformed normalized state vector. The subsequent basis functions depend on the relative magnitudes of the different singular values. For instance, while for $1 > \lambda_1 > \lambda_2$ clearly $\sqrt{\lambda_1} > \sqrt{\lambda_1}^2 = \lambda_1$ and similarly for λ_2, it is not clear whether $\lambda_1 > \sqrt{\lambda_2}$ or vice versa – and this order will determine which combination of basis functions is robust. Thus, in the multi-dimensional case – and particularly in high-dimensional settings that are 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 (e.g., 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 3.1, we obtain sequential convergence. Joint convergence for (a class of) models $\mathbf{x}$ can be established by following Newey's approach from Propositions 3.2/3.3, or by relying on the results from Benedetti (2017) in case the parameters are known. While the Hermite polynomials do not satisfy the uniformly boundedness assumptions from Proposition 4.2, from Proposition 3.2 and the discussion following Proposition 4.2, it is clear that the convergence rate of the functional approximation is linked to the decay of the singular values ($O(\omega_M^2)$ in Prop. 4.2). In the current setting we have (Prop. 4.3):

$$\omega_M^2 = \omega_{m_M}^2 = \prod_{i=1}^{d} \lambda_i^{k_i} \leq \prod_{i=1}^{d} \max_{1\leq i\leq d}\{\lambda_i\}^{k_i} = \max_{1\leq i\leq d}\{\lambda_i\}^{\sum_i k_i},$$

where $\max_{1\leq i\leq d}\{\lambda_i\} < 1$ and there are $\binom{d-1+l}{d-1}$ vectors m such that $\sum_i k_i = l$. Thus, as in Proposition 3.3, the convergence is slowing down as the dimension d of the state process increases, although the relationship here is exponential rather than polynomial.

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 robust basis functions. For instance, Huang (2012) explains how to solve the associated integral equation by discretization methods, which allows to determine the singular functions numerically, and Serdyukov et al. (2014) apply the truncated SVD to solve inverse problems numerically. As an alternative, in what follows we rely on the ideas from the Gaussian setting to derive approximations to robust basis functions in arbitrary settings.

4.4 Approximations of Robust Basis Functions in General Settings

In the Gaussian setting from the previous section, the robust basis functions are sorted according to the singular values of L, which with (14) are given by products of the eigenvalues of:

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

Here A appears in μ_A and Σ_A, which expresses how the current state affects the distribution of current states when going forward and backward in time ($\tau \to T \to \tau$). The leading eigenvectors of A thus provide the most relevant directions in the multi-dimensional state space for evaluating L – or, in other words, the most important directions in spanning the image space of the valuation operator.

The same intuition carries over to more general situations. Let:

$$\bar{Y} = \begin{pmatrix} Y_\tau \\ Y_{\tau+1} \\ \vdots \\ Y_T \end{pmatrix} \quad \text{with} \quad \mathbb{E}^{\tilde{\mathbb{P}}}[\bar{Y}] = \begin{pmatrix} \mu_\tau \\ \bar{\mu} \end{pmatrix} \in \mathbb{R}^{d+d(T-\tau)} \quad \text{and} \quad \mathrm{Cov}^{\tilde{\mathbb{P}}}[\bar{Y}] = \begin{pmatrix} \Sigma_\tau & \bar{\Gamma} \\ \bar{\Gamma}' & \bar{\Sigma} \end{pmatrix},$$

where $\Gamma \in \mathbb{R}^{d\times d(T-\tau)}$ and $\bar{\Sigma} \in \mathbb{R}^{d(T-\tau)\times d(T-\tau)}$. Then, proceeding analogously to the Gaussian case above, it is possible to derive approximations to robust basis functions by considering $\bar{A} = \bar{\Gamma}\bar{\Sigma}^{-1}\bar{\Gamma}'\Sigma_\tau^{-1}$ and sample values from cash flow simulations:

Algorithm 4.1. *For given N and M:*

- *Generate N independent paths of the state process, $(Y_t^{(i)})_{t=\tau,\tau+1,...,T}$, $1 \le i \le N$, under $\tilde{\mathbb{P}}$.*
- *Determine the sample mean and sample covariance matrix of $\bar{Y}^{(i)}$:*

$$\hat{\bar{\mu}} = \begin{pmatrix} \hat{\mu}_\tau \\ \hat{\bar{\mu}} \end{pmatrix}, \quad \widehat{Cov}(\bar{Y}) = \begin{pmatrix} \hat{\Sigma}_\tau & \hat{\bar{\Gamma}} \\ \hat{\bar{\Gamma}}' & \hat{\bar{\Sigma}} \end{pmatrix}.$$

- *Determine the eigenvalue decomposition $\hat{\bar{P}}\hat{\bar{\Lambda}}\hat{\bar{P}}$ of $\hat{\Sigma}_\tau^{-1/2}\hat{\bar{A}}\hat{\Sigma}_\tau^{1/2} = \hat{\Sigma}_\tau^{-1/2}\hat{\bar{\Gamma}}\hat{\bar{\Sigma}}^{-1}\hat{\bar{\Gamma}}'\hat{\Sigma}_\tau^{-1/2} \in \mathbb{R}^{d\times d}$.*
- *Set $\bar{\omega}_m = \Pi_{i=1}^d \hat{\lambda}_i^{k_i/2}$ for a multi-index $m = (k_1, ..., k_d) \in \mathbb{N}_0^d$, where $\hat{\lambda}_i$ is the i-th diagnonal elemement of $\hat{\bar{\Lambda}}$, and determine $m_1, ..., m_M$ as the M largest elements of $(\omega_m)_{m\in\mathbb{N}_0^d}$.*
- *Set:*

$$\varphi_j(y) = \varphi_{m_j}(y) = \Pi_{i=1}^d h_{k_i^{(m_j)}}\left(\left[\hat{\bar{P}}'\hat{\Sigma}_\tau^{-1/2}(y - \hat{\mu}_\tau)\right]_i\right),$$

where $k_i^{(m_j)}$ are the entries of m_j, $j = 1, ..., M$.

Of course, the generated sample paths $(Y_t^{(i)})_{t=\tau,\tau+1,...,T}$, $1 \le i \le N$ may also be used for determining the cumulative discounted cash flows in (6) in the LSM algorithm. The key idea behind Algorithm 4.1 is an approximation of the (joint) distribution of the state vector entries via a Gaussian distribution by focussing on the first and second moments, and then leveraging the results from the Gaussian case (Corollary 4.1). This is akin to Linear/Quadratic Discriminant Analysis for classification (Hastie et al., 2009, Sec. 4.3), where the class densities of the features are approximated by a multivariate Gaussian distribution to obtain approximations to the (optimal) Bayes classifier.

5 Application

To demonstrate the LSM algorithm and its properties, we consider an example from life insurance: 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 new Solvency II regulation that came into effect in 2016. Here, the so-called *Solvency Capital Requirement* takes the form of a 99.5% VaR of firm capital at the risk horizon $\tau = 1$. And while our application certainly does not feature the level of complexity of practical enterprise risk measurement frameworks, it serves to illustrate some of the key tradeoffs, particularly with regards to the choice of basis functions.

5.1 Variable Annuities with GMIBs

Within a Variable Annuity (VA) plus GMIB contract, at maturity the policyholder has the right to choose between a lump sum payment amounting to the current account value or a guaranteed annuity payment determined by 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.[4] Importantly, GMIBs are subject to a variety of risk factors, including fund (investment) risk, mortality risk, and – as long term contracts – interest rate risk.

We consider a large portfolio of different VA plus GMIB contracts indexed by j with policyholder (cohort) age x, policy maturities T_j, $\tau < T_1 < T_2 < \ldots$, and guaranteed amounts $G_{T_j}^{(j)}$. The payoff of VA j at time T_j in case of survival is given by:

$$\max\left\{S_{T_j},\ \frac{G_{T_j}^{(j)}}{a_{x+T_j}^*(0)}\, a_{x+T_j}(T_j)\right\}, \tag{15}$$

where S_{T_j} is the underlying account value which evolves according to a reference asset net various fees (which we ignore for simplicity), $a_{x+T_j}(T_j)$ denotes the time T_j value of an immediate annuity on an $(x+T_j)$-year old policyholder of $1 annually, and $a_{x+T_j}^*(0)$ gives the guaranteed annuity rate at time zero.

We rely on a model for the financial environment that features stochastic interest rates and a stochastic evolution of mortality. Specifically, we assume the reference asset S_t, the interest (short) rate r_t, and the force of mortality (hazard rate) μ_{x+t} for an $(x+t)$-aged individual at time t evolve according to the following stochastic differential equations (SDEs) under $\mathbb{P}$:

$$dS_t = S_t\left(m\,dt + \sigma_S\,dW_t^S\right), \tag{16}$$

$$dr_t = \alpha(\gamma - r_t)\,dt + \sigma_r\,dW_t^r, \tag{17}$$

$$d\mu_{x+t} = \kappa\mu_{x+t}\,dt + \psi\,dW_t^\mu, \tag{18}$$

where m is the instantaneous rate of return of the risky asset and σ_S is the asset volatility; α, γ, and σ_r are the speed of mean reversion, the mean reversion level, and the interest rate volatility in the Vasicek (1977) interest rate model, respectively; κ is an instantaneous rate of increment of mortality (Gompertz exponent) and ψ is the volatility of mortality; and W_t^S, W_t^r, and W_t^μ are standard Brownian motions under $\mathbb{P}$ with $dW_t^S\,dW_t^r = \rho_{12}\,dt$, $dW_t^S\,dW_t^\mu = \rho_{13}\,dt$, and $dW_t^r\,dW_t^\mu = \rho_{23}\,dt$.

For evaluating the present value of future benefits at the risk horizon τ as in Equation (2), we need to also contemplate the dynamics under the risk-neutral measure $\mathbb{Q}$. For simplicity, we assume a constant risk premium λ for interest rate risk, so that we obtain similar dynamics under $\mathbb{Q}$, with m being replaced by r_t in (16) and γ being replaced by $\bar{\gamma} = \gamma - \lambda\,\sigma_r/\alpha$ in (17). To directly apply our LSM setting from previous sections, we change the numéraire process to a pure endowment (survival benefit) with maturity T_1 and maturity value one on $[\tau, T_1)$, to a pure endowment with maturity T_2 on $[T_1, T_2)$, and so on. The value of VA j at

[4] Source: Fact Sheets by the Life Insurance and Market Research Association (LIMRA).

time $t \in [T_{j-1}, T_j)$, where we set $T_0 = \tau$, is then:

$$V(t) = {}_{T_j-t}E_{x+t}\, \mathbb{E}^{\mathbb{Q}_{E_j}} \left[\max \left\{ S_{T_j}, {}^{G^{(j)}_{T_j}} /_{a^*_{x+T_j}(0)} \times a_{x+T_j}(T_j) \right\} \middle| Y_t \right], \tag{19}$$

where ${}_{T_j-t}E_{x+t}$ is the value of the pure endowment contract with maturity T_j at time t and $\mathbb{Q}_{E_j}$ is the risk-neutral measure using the pure endowment contract with maturity T_j as the numéraire. By successively applying (19) and adding up the values of the different VAs, we can derive the portfolio value at the risk horizon τ. We refer to Section 5.3 for more details on this procedure.

Under our assumptions, since (r_t) and (μ_t) are affine, we obtain:

$${}_{T_j-t}E_{x+t} = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^{T_j} r_s + \mu_{x+s}\, ds} | Y_t \right] = A(t, T_j) \exp\left[-B_r(t, T_j) r_t - B_\mu(t, T_j) \mu_{x+t} \right], \tag{20}$$

where $B_r(\cdot,\cdot)$, $B_\mu(\cdot,\cdot)$, and $A(\cdot,\cdot)$ are given in the Appendix. Applying Itô's formula, the dynamics of the pure endowment price are (here, $\tilde{W}_t$ denote Brownian motions under $\mathbb{Q}$):

$$d_{T_j-t}E_{x+t} = {}_{T_j-t}E_{x+t} \left[(r_t + \mu_{x+t})dt - \sigma_r B_r(t, T_j) d\tilde{W}^r_t - \psi B_\mu(t, T_j) d\tilde{W}^\mu_t \right],$$

and from Brigo and Mercurio (2006), the new dynamics of Y_t under $\mathbb{Q}_{E_j}$ for $T_{j-1} \le t \le T_j$ become:

$$\begin{aligned}
dS_t &= S_t \left((r_t - \rho_{12}\sigma_S\sigma_r B_r(t, T_j) - \rho_{13}\sigma_S\psi B_\mu(t, T_j))\, dt + \sigma_S\, dZ^S_t \right), && (21)\\
dr_t &= \left(\alpha(\bar{\gamma} - r_t) - \sigma_r^2 B_r(t, T_j) - \rho_{23}\sigma_r\psi B_\mu(t, T_j) \right) dt + \sigma_r dZ^r_t, && (22)\\
d\mu_{x+t} &= \left(\kappa\mu_{x+t} - \rho_{23}\sigma_r\psi B_r(t, T_j) - \psi^2 B_\mu(t, T_j) \right) dt + \psi\, dZ^\mu_t, && (23)
\end{aligned}$$

where Z^S_t, Z^r_t, and Z^μ_t are standard Brownian motions under $\mathbb{Q}_{E_j}$ with $dZ^S_t dZ^r_t = \rho_{12}\, dt$, $dZ^S_t dZ^\mu_t = \rho_{13}\, dt$, and $dZ^r_t dZ^\mu_t = \rho_{23}\, dt$. Moreover, we have:

$$a_{x+T_j}(T_j) = \sum_{k=1}^{\infty} {}_kE_{x+T_j}, \text{ and } a^*_{x+T_j}(0) = \sum_{k=1}^{\infty} {}^{T_j+kE_x} /_{TE_x}$$

is calculated based on forward mortality rates (Bauer et al., 2012a).

For our numerical experiments, we set the model parameters using representative values. The initial price of the risky asset is one hundred and for the risky asset parameters we assume $m = 5\%$ (instantaneous rate of return) and $\sigma_S = 18\%$ (asset volatility). The initial interest rate is assumed to be $r_0 = 2.5\%$, $\alpha = 25\%$ (speed of mean reversion), $\gamma = 2\%$ (mean reversion level), $\sigma_r = 1\%$ (interest rate volatility), and $\lambda = 2\%$ (market price of risk). For the mortality rate, we set $x = 55$ (age of the policyholder), $\mu_{55} = 1\%$ (initial value of mortality), $\kappa = 7\%$ (instantaneous rate of increment), and $\psi = 0.12\%$ (mortality volatility). For correlations, we assume $\rho_{12} = -30\%$ (correlation between asset and interest rate), $\rho_{13} = 6\%$ (correlation between asset and mortality rate), and $\rho_{23} = -4\%$ (correlation between interest rate and mortality rate).

5.2 Gaussian Case: Single Roll-Up GMIB

We start by considering a single VA with a simple roll-up benefit that compounds the initial investment at a guaranteed rate m_g. Hence, we set $T = T_1$, $\mathbb{Q}_{E_1} = \mathbb{Q}_E$, and:

$$G_{T_1}^{(1)} = G_T = S_0 \times (1 + m_g)^T. \tag{24}$$

For our numerical experiments, we assume a guaranteed rate of return $m_g = 2\%$. This implies a probability that $S_T > b\, a_{x+T}(T)$ of approximately 40%. Also, we set the risk horizon $\tau = 1$ in line with the Solvency II regulation.

For the calculation of the risk capital, we ignore unsystematic mortality risk arising from finite samples and stochastic investments. Thus, it suffices to use the three-dimensional state process $Y_t = (q_t,\, r_t,\, \mu_{x+t})'$, where $q_t = \log\{S_t\}$ denotes the log-price of the risky asset. We estimate the distribution of C_τ via the LSM algorithm. In particular, the cash flow functional in the current setting is $\mathbf{x} = x_T$ with:

$$x_T(Y_T) = -V(T) = -\max\{e^{q_T},\, {}^{S_0\times(1+m_g)^T}/_{a^*_{x+T}(0)} \times a_{x+T}(T)\}$$

and

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

We have:

Lemma 5.1. *From (16)-(18) and (21)-(23), the joint (unconditional) distribution of Y_τ and Y_T is:*

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

where we refer to the proof in the Appendix for explicit expressions of μ_τ, μ_T, Σ_τ etc. in terms of the parameters.

Thus we can apply the results from Proposition 4.3 to derive robust basis functions.

LSM using Robust Basis Functions.

From Proposition 4.3, for any non negative integer vector $m = (k_1, k_2, k_3)$, $\omega_m = \lambda_1^{k_1/2}\lambda_2^{k_2/2}\lambda_3^{k_3/2}$ is a singular value of L, with corresponding left singular function:

$$\varphi_m(\cdot) = h_{k_1}(z_1^P(\cdot))h_{k_2}(z_2^P(\cdot))h_{k_3}(z_3^P(\cdot)),$$

where $h_j(\cdot)$ Hermite polynomial of degree j. Thus, following Corollary 4.1, in order to find the set of M robust basis functions for the LSM algorithm, we need to calculate ω_m for $|m| \le M$, order them, and then determine the associated functions. With the above parameters, from Lemma 4.3, we obtain $\lambda_1 = 0.1490$, $\lambda_2 = 0.0671$, and $\lambda_3 = 0.0004$ by solving for the eigenvalues A, and we have:

$$\begin{pmatrix} z_1^P(y_\tau) \\ z_2^P(y_\tau) \\ z_3^P(y_\tau) \end{pmatrix} = P'\Sigma_\tau^{-1/2}(y - \mu_\tau) = \begin{pmatrix} -0.0981 & (q_\tau - 4.6390) & -0.2969 & (r_\tau - 0.0239) & +804.9935 & (\mu_{x+\tau} - 0.0107) \\ 5.7992 & (q_\tau - 4.6390) & +23.6115 & (r_\tau - 0.0239) & -29.9839 & (\mu_{x+\tau} - 0.0107) \\ -0.5878 & (q_\tau - 4.6390) & -115.7881 & (r_\tau - 0.0239) & -27.0757 & (\mu_{x+\tau} - 0.0107) \end{pmatrix}$$

Two observations are immediate: First, each first-order robust basis function is a linear combination of the state variables. Hence, unlike when sequentially choosing polynomials in

each state variable, the question is not what component is most relevant but rather which direction in state variable space is most significant. Clearly, the latter question is more general and has advantages particularly when the state variable space is high-dimensional. Second, higher-order robust basis functions may enter before going through all first-order combinations. Indeed, $\omega_{(2,0,0)} = \lambda_1 > \sqrt{\lambda_3} = \omega_{(0,0,1)}$, so that the first four robust basis functions are $(1, z_1^P(\cdot), z_2^P(\cdot), \frac{1}{\sqrt{2}}((z_1^P(\cdot))^2 - 1))'$, as opposed to, e.g., $(1, q_\tau, r_\tau, \mu_{x+\tau})'$ when choosing $M = 4$ monomials.

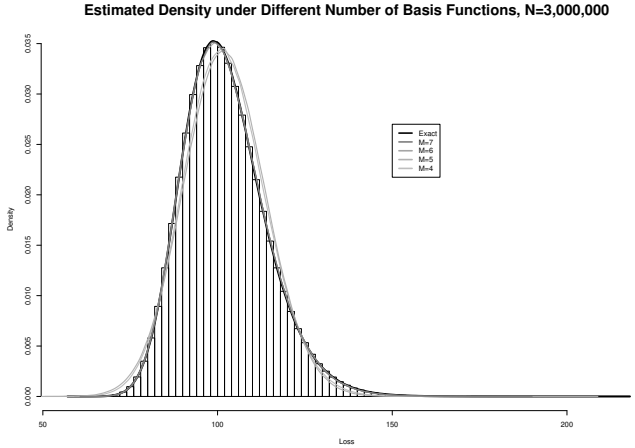


Figure 1: Empirical densities of $V(\tau)$ based on $N = 3,000,000$ Monte Carlo realizations; "exact" and using the LSM algorithm with M singular functions in the approximation.

We implement the LSM approximation to the capital variable and vary the number of robust basis functions. In Figure 1, we provide an empirical density based on $N = 3,000,000$ and approximate realizations calculated via the LSM algorithm for different numbers of basis functions M.[5] As is evident from the figure, a small number of basis function does not produce satisfactory results. However, the approximation becomes closer to the "exact" density as M increases, in line with our convergence results from Section 3.

Comparison to Conventional Basis Functions.

It is instructive to consider the functional approximation from Figure 1 based on robust basis functions for the case $M = 4$:

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}_E}\left[\max\{e^{q_T}, a_{x+T}(T)\}|Y_\tau\right] &\approx 173.7539 - 5.1805\, z_1^P(y_\tau) + 15.6972\, z_2^P(y_\tau) + 1.0281\,[\frac{1}{\sqrt{2}}(z_1^P(y_\tau)^2 - 1)] \\ &= -162.194 + 92.704\, q_\tau + 0.007\, q_\tau^2 + 375.7\, r_\tau + 0.042\, q_\tau r_\tau + 0.064\, r_\tau^2 \\ &\quad -14,204.8\, \mu_{x+\tau} - 114.768\, q_\tau \mu_{x+\tau} - 347.528\, r_\tau \mu_{x+\tau} + 471,074\, \mu_{x+\tau}^2.\end{aligned}$$

In contrast, when using $M = 4$ monomials in the LSM algorithm (again $N = 3,000,000$), we obtain:

$$\mathbb{E}^{\mathbb{Q}_E}\left[\max\{e^{q_T}, a_{x+T}(T)\}|Y_\tau\right] \approx -205.605 + 90.515\, q_\tau + 346.019\, r_\tau - 4,533.644\, \mu_{x+\tau}.$$

[5]Since it is impossible to obtain the exact loss distribution at the risk horizon, we consider the estimated loss distribution obtained from the LSM algorithm with $M = 37$ monomials and $N = 40 \times 10^6$ simulations as "exact."

		$N = 100,000$		$N = 3,000,000$	
	Div.	Singular	Monomials	Singular	Monomials
$M = 4$	KS	2.52×10^{-2}	2.86×10^{-2}	2.41×10^{-2}	2.77×10^{-2}
	KL	2.17×10^{-4}	2.32×10^{-4}	2.13×10^{-4}	2.28×10^{-4}
	JS	7.43×10^{-3}	7.68×10^{-3}	7.36×10^{-3}	7.62×10^{-3}
$M = 6$	KS	7.91×10^{-3}	9.60×10^{-3}	2.24×10^{-3}	5.79×10^{-3}
	KL	1.09×10^{-5}	4.93×10^{-5}	4.53×10^{-6}	4.31×10^{-5}
	JS	1.62×10^{-3}	3.52×10^{-3}	1.06×10^{-3}	3.29×10^{-3}
$M = 12$	KS	8.28×10^{-3}	8.26×10^{-3}	1.49×10^{-3}	1.53×10^{-3}
	KL	1.43×10^{-5}	1.55×10^{-5}	6.02×10^{-7}	1.74×10^{-6}
	JS	1.84×10^{-3}	1.93×10^{-3}	3.82×10^{-4}	6.58×10^{-4}

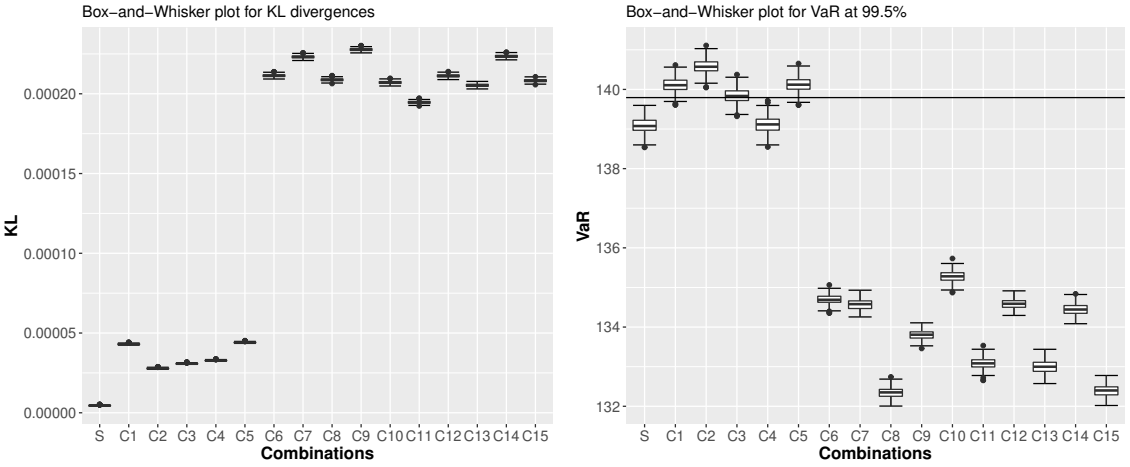
Table 1: Statistical divergence measures between the empirical density function based on the "exact" realizations and the LSM approximation using different basis functions; mean of 300 runs with $N = 3,000,000$ sample paths each.

We note that the latter regression result is not linked to the choice of monomials: Aside from numerical errors, we will obtain the same results when using orthonormal polynomial families with only first-order terms, since the span is the equivalent. Also, both approximations have the same degrees of freedom ($M = 4$). The reason the fit has a different form is that the first four robust basis functions include higher order terms, since their composition casts a broader net.

This results in a better functional approximation. To illustrate, in Table 1 we report statistical differences to the "exact" distribution according to various statistical divergence measures for singular functions and monomials for various numbers of simulations N and basis functions M. Here, the set of monomial basis functions when $M = 6$ is $(1, q_\tau, r_\tau, \mu_{x+\tau}, q_\tau^2, r_\tau^2)$, and for $M = 12$ we include all second-order terms and (q_τ^3, r_τ^3). For each combination, we report three common statistical divergence measures: the Kolmogorov-Smirnov statistic (KS), the Kullback-Leibler divergence (KL), and the Jensen-Shannon divergence (JS). We report the mean of three-hundred runs. We find that the singular functions significantly outperform the monomials, particularly for the larger number of sample paths ($N = 3,000,000$), with a relative difference up to an order of magnitude for the KL divergence. This result documents the importance of choosing appropriate basis functions, and the virtue of singular functions as the tangible and expedient choice.

This insight is even more relevant when considering that it is necessary to choose which higher-oder monomials to include – and indeed the choice in Table 1 turns out to be favorable. To illustrate, Figure 2a provides box-and-whisker plots for the KL divergence for 15 different possible combinations of six monomials, where in addition to the first-order terms for each of the state variables we include all possible combinations of two second-order terms.[6] The plots are based on three-hundred runs of $N = 3,000,000$ samples (corresponding KS and JS

[6]Here and in what follows, the box presents the area between the first and third quartile, with the inner line placed 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 box-and-whisker plot).



(a) KL (b) VaR at 99.5%

Figure 2: Box-and-whisker plots for the KL divergence and 99.5% VaR calculated using different basis functions with $M = 6$; based on 300 runs with $N = 3,000,000$ sample paths each.

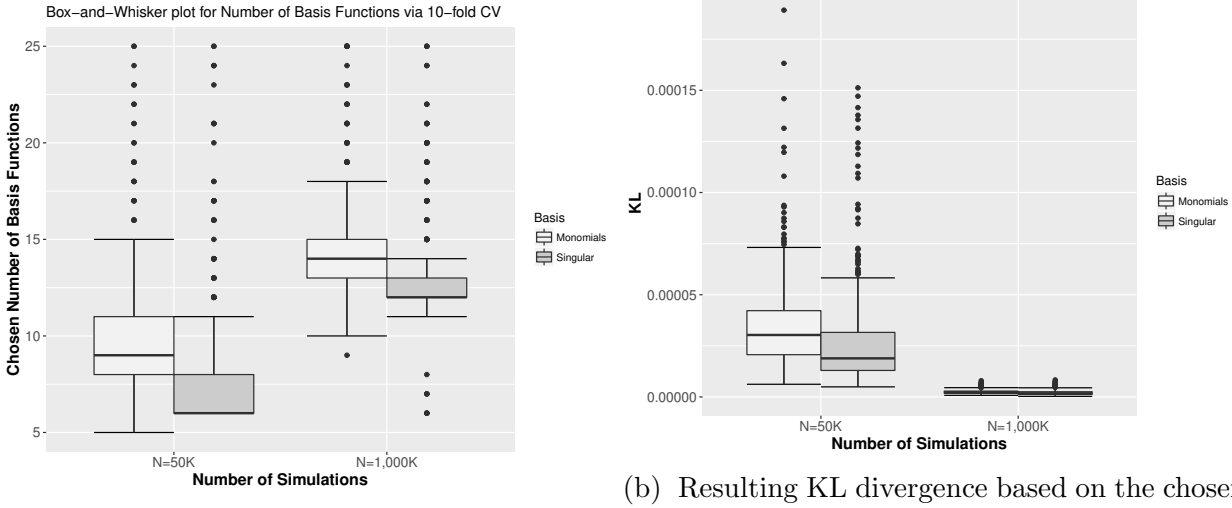
(Higher-oder terms: C1: q_τ^2, r_τ^2; C2: q_τ^2, $\mu_{x+\tau}^2$; C3: q_τ^2, $q_\tau r_\tau$; C4: q_τ^2, $q_\tau \mu_{x+\tau}$; C5: q_τ^2, $r_\tau \mu_{x+\tau}$; C6: r_τ^2, $\mu_{x+\tau}^2$; C7: r_τ^2, $q_\tau r_\tau$; C8: r_τ^2, $q_\tau \mu_{x+\tau}$; C9: r_τ^2, $r_\tau \mu_{x+\tau}$; C10: $\mu_{x+\tau}^2$, $q_\tau r_\tau$; C11: $\mu_{x+\tau}^2$, $q_\tau \mu_{x+\tau}$; C12: $\mu_{x+\tau}^2$, $r_\tau \mu_{x+\tau}$; C13: $q_\tau r_\tau$, $q_\tau \mu_{x+\tau}$; C14: $q_\tau r_\tau$, $r_\tau \mu_{x+\tau}$; C15: $q_\tau \mu_{x+\tau}$, $r_\tau \mu_{x+\tau}$).

box-plots are provided in Figure 6 in the Appendix). Again, we find that singular functions significantly outperform all combinations of monomials, and the differences can be drastic, particularly when not including the square of the log-price of the risky asset (combinations C6-C15). While some of the combinations may be unrealistic choices (e.g., those including combinations with the mortality rate), including a squared term if the interest rate and a combination of interest rate and the log-price of the risky asset (C7) may seem reasonable ex-ante. Importantly, it will be difficult to develop intuition on "reasonable" choices in high-dimensional enterprise risk measurement frameworks employed in practice.

We note that these findings with regards to the superior performance of the singular functions are not driven by the specific payoff function: As discussed in detail in Section 4, our notion of robustness is tied to the *model framework* rather than a specific cash flow *model*. In the Appendix, we include analyses based on modified payoffs, where we adjust guaranteed amount G_T. The results are analogous: Singular functions perform better than monomials, and the difference can be substantial.

M vs. N: The tradeoff between the approximations.

A second observation from Table 1 is that, in line with Figure 1, increasing the number of basis functions M generally yields closer approximations of the "true" distribution. However, an exception is the transition from $M = 6$ to $M = 12$ singular functions for $N = 100,000$ sample paths, where the approximation worsens. The reason is the interplay between the number of basis functions M and samples N that is emphasized in Section 3.2: The convergence rate in Proposition 3.3 is a function of M and N, and increasing M while keeping N at



(a) Number of Basis Functions M based on CV

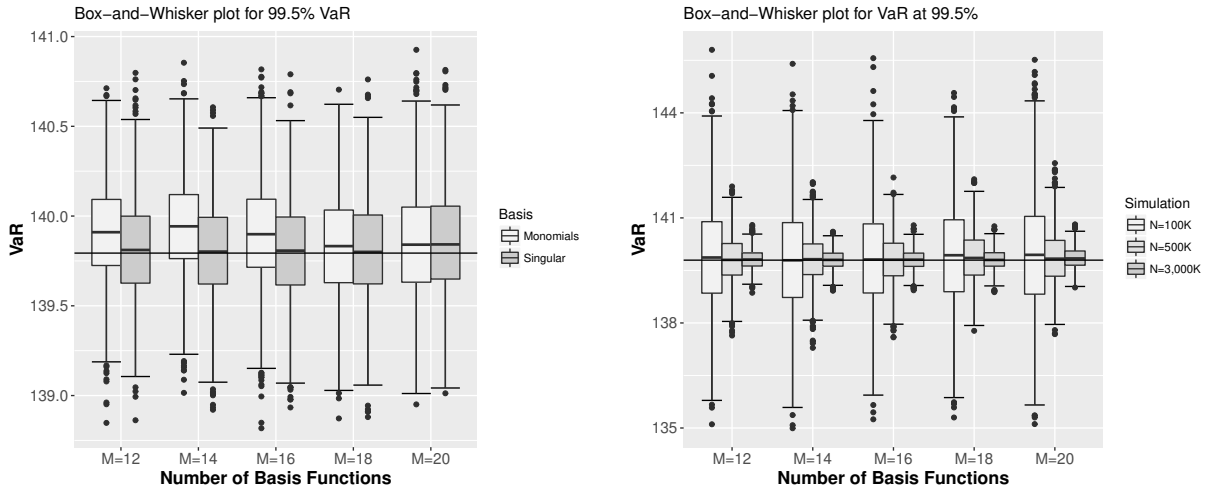
(b) Resulting KL divergence based on the chosen M

Figure 3: Box-and-whisker plots for the 10-fold cross-validated number of basis function M and the resulting KL divergence; based on 500 runs.

the same level, although yielding a closer functional approximation (second term), adversely affects the regression approximation. For the monomials, on the other hand, adding terms always decreases the divergence in Table 1, so that generally both aspects in the convergence rate are at work, with either of them dominating in some cases. This demonstrates that choosing the right number M given a numerical budget (N) is not trivial. Indeed, since N will have to grow as M increases, given a fixed computational budget, feeding the algorithm with a very large number of basis functions will be futile.

A common approach to navigate this tradeoff is to rely on cross-validation (CV) (Hastie et al., 2009, Chap. 7). In Figure 3a, we present results based on 10-fold CV for choosing the number of basis functions (monomials or singular functions) based on five hundred runs and $N = 50,000$ and $N = 1,000,000$. More precisely, for each run, the data is separated into ten folds and then the root-mean-square-errors (RMSE) of the predictions for the hold-out fold is calculated when fitting the regression using the remaining nine folds; the ten RMSEs are averaged to produce 10-fold CV RMSE. We report the M that leads to the smallest CV RMSE for each run. For the monomials, we go through the terms by order, where we sequence higher order terms by first considering q_τ, then r_τ, and then $\mu_{x+\tau}$. Figure 3b shows the KL divergence resulting from the cross-validated M in each run.

The cross-validated M for $N = 50,000$ are generally lower than for $N = 1,000,000$, which is not surprising given our observations on the M vs. N tradeoff. Moreover, the cross-validated M is typically lower for the singular functions, due to the improved functional approximation. As a consequence, the fit is less dispersed for singular functions than for monomials, which is evident from the distribution of the KL divergence in Figure 3b being more concentrated. However, even though the chosen M is usually lower for singular functions, the approximation remains closer: the KL divergence also is lower on average.



(a) Comparison between basis functions (b) Impact of N

Figure 4: Box-and-whisker plots for 99.5% VaR calculated using the LSM algorithm with different numbers of basis functions M; the left-hand panel is based on 1,500 runs with $N = 3,000,000$ sample paths each. The right-hand plot is also based on 1,500 runs and different choices for N and M.

Risk measure estimation.

One of the key take-aways from the foregoing analyses is that a significant number of basis functions M is necessary to obtain an accurate approximation to the capital distribution. For instance, panel (b) in Figure 2 shows that the "exact" VaR estimate lies outside the whiskers for the singular value functions as well as for most of the monomial combinations – and when it is inside the whiskers that appears to happen rather arbitrary. For larger choices for M, the situation improves. To illustrate, in Figure 4a, we show box-and-whisker plots of 99.5% VaR LSM estimates for singular functions as well as monomials, using $N = 3,000,000$ sample paths each. Again, the singular functions outperform the monomials, although the difference is less pronounced that when focussing on the entire distribution. The reason is that the LSM algorithm is – and the basis functions are – tailored to approximate the distribution, and not a specific risk measure. We will return to this point in our Conclusion, where we depict a potential avenue of how to improve the situation when targeting a specific risk measure.

From Figure 4a, we also notice that the central estimate does not markedly improve when going from $M = 12$ to $M = 20$ singular basis functions. It appears that for $M = 12$, the functional approximation is sufficiently accurate even in this tail region, and adding additional terms does not improve on the central estimate. Indeed, the key issue here is the dispersion of the estimates. For instance, for $M = 12$ singular functions, the whiskers based on the different runs span roughly 139.1 to 140.5, so the lengths is 1.4 or roughly 1% of the central estimate. Given the usual scale of capital requirements in the financial services industry, the cost for an additional percent of required capital is substantial. And, the whisker plots become wider as the number of basis functions increases (roughly $[139, 140.7]$ for $M = 20$). The dispersion becomes even larger when decreasing N, as is illustrated in

Panel 4b. For $N = 100,000$, the length of the whisker plot amounts to more than 5% of the central estimate and it increases as the number of basis functions M increases. This is again linked to the N vs. M tradeoff discussed above.

An alternative way of managing this tradeoff is to rely on regularized regression approaches (Hastie et al., 2009, Chap. 5), although their application in the present context is not without problem. To elaborate, regularized regression approaches generally introduce some bias for a reduction in variance. And since the regularization parameter is chosen so as to minimize the mean-squared prediction error across all realizations, it is not surprising that the predictions may worsen in a certain area of the distribution – which may cause issues when estimating tail risk measures. To illustrate, in the Appendix, we repeat our VaR estimations for $M = 16$ basis functions – using both, singular functions and monomials – where instead of OLS, we employ ridge regression to fit the approximation (see Figures 7a and 9 in part B). We find that relying on a regularized regression for estimating VaR can indeed be precarious, particularly for monomials. For the singular functions, the impact on the VaR estimate is limited, because overall there is very little change in the estimates relative to OLS – whereas for the monomial basis functions, the estimates are zooming in on the "wrong" target.

5.3 General Case: Multiple Complex GMIBs

We now consider two VAs with maturities T_1 and T_2, $T_2 > T_1$, with look-back (ratchet) and Asian guarantee features, respectively, issued to the same (cohort of) policyholders aged x. More precisely, we set the guaranteed amounts as:

$$G^{(1)}_{T_1} = \max_{0\le u\le T_1}\{S_u\} \text{ and } G^{(2)}_{T_2} = \frac{1}{T_2}\int_0^{T_2} S_u\,du.$$

Hence, it is necessary to consider the five-dimensional state variable $Y_t = (S_t, M_t, A_t, r_t, \mu_{x+t})'$, where $M_t = \max_{0\le u\le t}\{S_u\}$ and $A_t = \frac{1}{t}\int_0^t S_u\,du$. In particular, we no longer have Gaussian distributions, so it is not possible to rely on Section 4.3 in the present application. However, it is possible to rely on Algorithm 4.1 from Section 4.4 to obtain approximations for robust basis functions.

We proceed as follows: For each $i = 1, \ldots, N$:

- We simulate $Y^{(i)}_{t_k}$ on $[0, \tau]$ using an Euler discretization of (16)-(18), $t_k = \Delta \times k$ and a step size $\Delta = {}^1/_{240}$. We update M_{t_k} and A_{t_k} according to:

$$M_{t_k} = \max\{M_{t_{k-1}}, S_{t_k}\} \text{ and } A_{t_k} = \frac{(k-1)A_{t_{k-1}} + S_{t_k}}{k}. \tag{25}$$

- Starting with $Y^{(i)}_\tau$, we simulate $Y^{(i)}_{t_k}$ on $[\tau, T_1]$ using an Euler discretization of (21)-(23) setting $T_j = T_1$, where we update M_{t_k} and A_{t_k} according to (25).

- Starting with $Y^{(i)}_{T_1}$, we simulate $Y^{(i)}_{t_k}$ on $[T_1, T_2]$ using an Euler discretization of (21)-(23) setting $T_j = T_2$, where we update M_{t_k} and A_{t_k} according to (25).

- We determine the sample covariance matrix of $\bar{Y}^{(i)} = (Y_\tau^{(i)\prime}, Y_{T_1}^{(i)\prime}, Y_{T_2}^{(i)\prime})$, $i = 1, \ldots, N$, and determine approximations of robust basis functions $\varphi_1, \ldots, \varphi_M$ as in Algorithm 4.1.

- In each path, we determine the (discounted) payoff via:

$$V_\tau^{(i)} = {}_{T_1-\tau}E_{x+\tau}(Y_\tau^{(i)})\ \max\{S_{T_1}^{(i)}, G_{T_1}^{(1,i)}\} + {}_{T_1-\tau}E_{x+\tau}(Y_\tau^{(i)})\ {}_{T_2-T_1}E_{x+T_1}(Y_{T_1}^{(i)})\ \max\{S_{T_2}^{(i)}, G_{T_2}^{(2,i)}\}.$$

We can then obtain approximations for the distribution of C_τ by regressing $V_\tau^{(i)}$ on

$$(\varphi_1(Y_\tau^{(i)}), \ldots, \varphi_M(Y_\tau^{(i)})),\ i = 1, \ldots, N,$$

or another set of basis functions. We set $T_1 = 10$ and $T_2 = 15$ in our numerical experiments.

Approximations to Robust Basis Functions.

In this case, the (ordered) eigenvalues of $\hat{\hat{A}}$ are:

$$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4, \hat{\lambda}_5) = (0.3115, 0.1735, 0.0064, 0.0038, 0.0001).$$

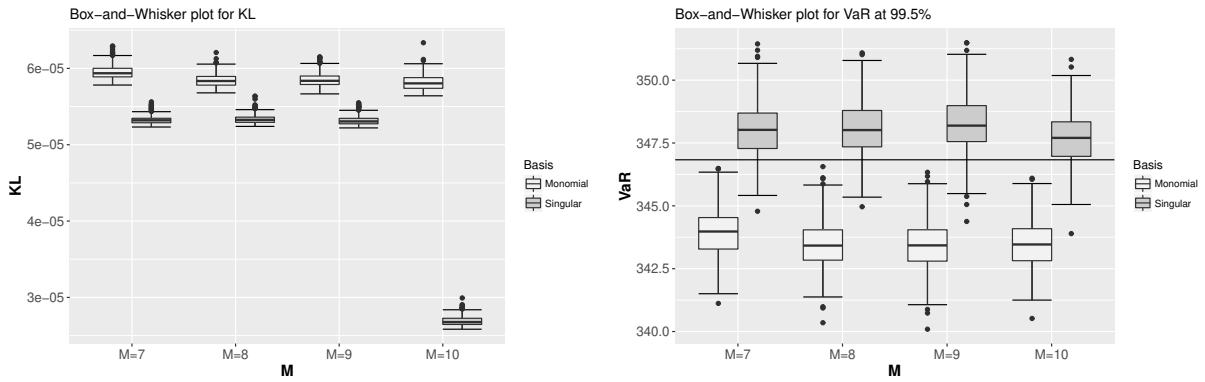
Thus, as in the previous application, functions including higher-order combinations of the first and second components of z^P occur before other first-order terms. For instance, the third (z_3^P) and fourth (z_4^P) first-order terms occur as robust basis function nine and eleven, respectively.

Comparison of LSM with Conventional Basis Functions.

Since there now are five state variables, it is not trivial to consider polynomials basis functions with higher degrees. For instance, there are 20 monomials up to second degree, and 56 monomials up to third degree. In contrast, with the robust basis functions, we start to include higher order terms starting with the third basis function ($\varphi_3(Y_\tau) = \frac{1}{\sqrt{2}}(z_1^P(Y_\tau)^2 - 1), \varphi_5(Y_\tau) = \frac{1}{\sqrt{6}}(z_1^P(Y_\tau)^3 - 3z_1^P(Y_\tau))$), so they efficiently capture non-linear effects.

As before, this results in a superior functional approximation. Figure 5 shows box-and-whisker plots for the KL divergences as well as the 99.5% VaR under approximately robust basis functions and monomials (KS and JS results are provided in the Appendix, Figure 10).[7] Here, for monomials, we used the "best" combination of higher-order terms that, similarly to the previous application, first include higher order terms in stock price and interest rate. All implementations are based on $N = 1,000,000$, and distributions are obtained from five hundred runs. Again, we find that the approximately robust basis functions uniformly outperform the polynomials. As before, the difference is particularly pronounced in the KL case, which consider the entire distribution, although the superiority also manifests for VaR.

[7]The (assumed) exact density is obtained from $N = 20,000,000$ and $M = 56$ monomials including a constant function (which means that we use all monomials up to degree three). The "exact" VaR at 99.5% at the risk horizon ($\tau = 1$) is then 346.8318.



(a) KL (b) VaR at 99.5%

Figure 5: Box-and-whisker plots for the KL divergence and 99.5% VaR calculated using different basis functions and different M; based on 500 runs with $N = 1,000,000$ sample paths each.

6 Conclusion

We discuss a Least-Squares Monte Carlo (LSM) algorithm for estimating financial capital in *nested* valuation settings. The algorithm relies on functional approximations of conditional expected values and least-squares regression. As a key contribution, we discuss the choice of basis functions in the functional approximation. Specifically, we show that the left singular functions of the valuation operator that maps cash flows to capital present *robust* basis functions for a model framework. We discuss the derivation of robust basis functions. Our numerical illustrations document that the algorithm can provide viable results at relatively low computational costs.

Various extensions are possible. One aspect that is particularly interesting is tailoring the approach – and underlying basis functions – to the evaluation of a particular (tail) risk measure, to boost performance, e.g., when estimating VaR or ES. One possibility is using *Importance Sampling* for variance reduction, where the central idea is to select an alternative probability measure for evaluation of the risks at the risk horizon (Glasserman et al., 2000; Capriotti, 2008). Proceeding analogously to Section 4, it appears possible to derive robust basis functions when replacing $\mathbb{P}$ with the importance sampling measure. Another interesting extension is to combine our ideas with non-parametric smoothing approaches for capital estimation. This is also pointed out by Risk and Ludkovski (2018, Sec. 5.3). Integrating regression-based approaches with emulation techniques considered in their paper presents a promising avenue for future research.

References

Abbe, E., Zheng, L., 2012. A coordinate system for Gaussian networks. IEEE Transactions on Information Theory, 58: 721-733.

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

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

Bauer, D., Benth, F.E., Kiesel, R., 2012a. Modeling the forward surface of mortality. SIAM Journal on Financial Mathematics, 3: 639-666.

Bauer, D., Bergmann, D., Reuss, A., 2010. Solvency II and Nested Simulations – a Least-Squares Monte Carlo Approach. Paper presented at the International Congress of Actuaries 2010. Available at `http://www.africanagenda.com/ica2010/`.

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

Benedetti, G., 2017. On The Calculation Of Risk Measures Using Least-Squares Monte Carlo. International Journal of Theoretical and Applied Finance, 20: 1-14.

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

Breiman, L., Friedman, J.H., 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, 80: 580-598.

Brigo, D., Mercurio, F., 2006. Interest rate models – Theory and Practice with Smile, Inflation and Credit. Second Edition. Springer, Berlin.

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

Cambou, M., Filipović, D., 2018. Replicating Portfolio Approach to Capital Calculation. Finance and Stochastics, 22: 181-203.

Capriotti, L., 2008. Least-squares importance sampling for Monte Carlo security pricing. Quantitative Finance, 8: 485-497.

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

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.

Floryszczak A., Le Courtois O., Majri M., 2016. Inside the Solvency 2 Black Box: Net Asset Values and Solvency Capital Requirements with a Least-Squares Monte-Carlo Approach. Insurance: Mathematics and Economics, 71: 15-26.

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

Glasserman, P., Heidelberger, P., Shahabuddin, O., 2000. Variance reduction techniques for estimating value-at-risk. Management Science, 46: 1349-1364.

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

Ha, H., 2016. Essays on Computational Problems in Insurance. Dissertation, Georgia State University. https://scholarworks.gsu.edu/rmi_diss/40.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning – Data Mining, Inference, and Prediction, Second Edition. Springer, New York.

Hong, L.J., Juneja, S., Liu, G., 2017. Kernel smoothing for nested estimation with application to portfolio risk measurement. Operations Research, 65: 657-673.

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

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.

Krah, A. S., Nikolić, Z., Korn, R., 2018. A Least-Squares Monte Carlo Framework in Proxy Modeling of Life Insurance Companies. Risks, 6: 62.

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

Linetsky, V., 2004. The spectral decomposition of the option value. International Journal of Theoretical and Applied Finance, 7: 337-384.

Liu, M., Staum, J., 2010. Stochastic kriging for efficient nested simulation of expected shortfall. The Journal of Risk, 12: 3-27.

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

Makur, A., Zheng, L., 2016. Polynomial spectral decomposition of conditional expectation operators. 54th Annual Allerton Conference Communication, Control, and Computing (pp. 633-640). IEEE.

Milliman, 2013. "Least Squares Monte Carlo for fast and robust Capital projections" by C. Bettels, F. Ketterer, M. Hoerig, and M. Leitschkis. Available at 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.

Nikolić, Z., Jonen, C., Zhu, C., 2017. Robust regression technique in LSMC proxy modeling. Der Aktuar, 01.2017: 8-16.

Nocco, B.W., Stulz, R.M., 2006. Enterprise risk management: Theory and practice. Journal of applied corporate finance, 18: 8-20.

Pelsser, A., Schweizer, J., 2016. The Difference between LSMC and Replicating Portfolio in Insurance Liability Modeling. European Actuarial Journal, 6: 441-494.

Risk, J., Ludkovski, M., 2018. Sequential Design and Spatial Modeling for Portfolio Tail Risk Measurement. SIAM Journal on Financial Mathematics, 9: 1137-1174.

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: 208-218.

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.

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

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

Appendix

The first part of this Appendix collects all proofs and technical material. The second part B contains supplemental analyses to the application Section 5 of the main text.

A Proofs and Technical Material

Proof. Proof of Lemma 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:

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

□

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

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

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

Proof. Proof of Lemma 2.3. Consider the operator $L^{(j)}$ mapping from $L^2(\mathbb{R}^d, \mathcal{B}, \tilde{\mathbb{P}}_{Y_j})$ to $L^2(\mathbb{R}^d, \mathcal{B}, \mathbb{P}_{Y_\tau})$. Since $L^{(j)}$ is the (conditional) expectation under the assumption that there exists a joint density, it can be represented as:

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

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

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

Thus, $L^{(j)}$ is a Hilbert-Schmidt operator (e.g., Prop. VI.6.3 in Werner (2005)), and therefore compact. Finally, L is the sum of $L^{(j)}$, $j = \tau, ..., T$, and therefore also compact. □

Proof. Proof of Proposition 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 $\operatorname{lin}\{e_k,\ k = 1, 2, \ldots, M\} = \operatorname{lin} S$. For S, on the other hand, we can find an orthonormal basis $\{f_k,\ k \in \mathbb{N}\} = S' \supset S$. Hence:

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

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 assume 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 $A^M = \left[\mathbb{E}^{\tilde{\mathbb{P}}}\left[e_k\left(Y_\tau\right)\, e_l\left(Y_\tau\right)\right]\right]_{1\leq k,l\leq M}$, 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, write:

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

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)):

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

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:

$$\xi^{(M)} = e'\,\tilde{\xi}\,e. \tag{26}$$

□

Proof. Proof of Corollary 3.1. Relying on the notation from the proof of Proposition 3.1, we now have:

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

in $L^2(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ 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. Proof of Proposition 3.2. Since $(V_\tau^{(i)}, Y_\tau^{(i)})$ are i.i.d. as Monte Carlo trials, 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. Proof of Proposition 3.3. Analogously to the proof of 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).

□

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

Proof. Singular Value Decomposition of a Compact Operator (Section 4.2). Suppose the operator A mapping from $\mathcal{H}_1$ to $\mathcal{H}_2$ is compact, 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)):

$$A\mathbf{x} = \sum_{k=1}^{\infty} \lambda_k \langle \mathbf{x}, g_k \rangle_{\mathcal{H}_1} f_k, \tag{27}$$

where:

- $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ 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, λ_k is called the k-th singular value of A;
- $\{g_k\} \subset \mathcal{H}_1$, called the (right) *singular functions*, are the orthonormal eigenfunctions of A^*A; and
- $\{f_k\} \subset \mathcal{H}_2$, called the (left) *singular functions*, are the orthonormal eigenfunctions of AA^* satisfying $A\, g_k = \lambda_k\, f_k$.

The representation (27) is called *singular value decomposition* (SVD) of A and the triple (λ_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 complete orthonormal sequences of $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The singular values λ_k are non-negative and the only possible accumulation point is zero. □

Proof. Proof of Proposition 4.1. We consider the approximation of L by an arbitrary rank-M operator L_F, which can be represented as:

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

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:

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

On the other hand, consider any alternative system (α_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:

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

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 (8). □

Proof. Proof of Proposition 4.2. Proceeding as in Equation (10) and with Equation (8), 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 inequalities follow by the triangle and Cauchy-Schwarz inequalities, respectively.

Then, going through the assumptions of Proposition 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 3.2. □

Proof. Proof of Lemma 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 2.3. Note that the transition density function can be written as:

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

We evaluate the following integral:

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

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. Proof of Lemma 4.2. L^* can be found via:

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

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

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

It is useful to express the reverse density as in the proof of Lemma 4.1:

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

Hence:

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

where in the third equality we again rely on results on the products of Gaussian densities from Vinga (2004). L^*L can be derived analogously. □

Proof. Proof of Lemma 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:

$$X_t = \Phi X_{t-1}+\eta_t,\quad t\geq 1, \tag{28}$$

where $\Phi\in\mathbb{R}^{d\times d}$ and $(\eta_t)_{t\geq 1}$ are independent and identically distributed, $\eta_t\sim N(0,H)$. (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 (28) has eigenvalues which are products of eigenvalues of Φ and the corresponding eigenfunctions are products of Hermite polynomials.

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

$$Y-\mu_\tau = A(x-\mu_\tau)+\zeta_A, \tag{29}$$

where $\zeta\sim N(0,\Sigma_A)$. Since from Lemma 4.2 we have that $\Sigma_A=\Sigma_\tau-A\Sigma_\tau A'$ and:

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

for $\Sigma = \Sigma_\tau$ the operator $L\,L^*$ has the same structure of the Markov operator for (28) that is reversible and stationary.

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

Now for the transformation (12) of the random vector Y from (29), $z^P(Y)$, 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. Proof of Proposition 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 an 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 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 from $z_i^Q(X)$. □

Proof. Derivation of the Endowment Value in Equation (20) (Section 5.1).

Following Duffie et al. (2000), we obtain:

$$\begin{aligned} B_r(t,T) &= \frac{1-e^{-\alpha(T-t)}}{\alpha},\ B_\mu(t,T) = \frac{e^{\kappa(T-t)}-1}{\kappa}, \\ \text{and } A(t,T) &= \exp\left\{\bar{\gamma}\left(B_r(t,T)-T+t\right)+\frac{1}{2}\left\{\frac{\sigma_r^2}{\alpha^2}\left(T-t-2B_r(t,T)+\frac{1-e^{-2\alpha(T-t)}}{2\alpha}\right)\right.\right. \\ &\quad +\frac{\psi^2}{\kappa^2}\left(T-t-2B_\mu(t,T)+\frac{e^{2\kappa(T-t)}-1}{2\kappa}\right) \\ &\quad \left.\left.+\frac{2\rho_{23}\sigma_r\psi}{\alpha\kappa}\left(B_\mu(t,T)-T+t+B_r(t,T)-\frac{1-e^{-(\alpha-\kappa)(T-t)}}{\alpha-\kappa}\right)\right\}\right\}. \end{aligned}$$

□

Proof. Proof of Lemma 5.1. Under $\mathbb{P}$, the solutions of (16), (17), and (18) at time τ are:

$$\begin{aligned} q_\tau &= q_0+\left(m-\frac{1}{2}\sigma_S^2\right)\tau+\sigma_S\int_0^\tau dW_s^S, \\ r_\tau &= r_0e^{-\alpha\tau}+\gamma\left(1-e^{-\alpha\tau}\right)+\sigma_r\int_0^\tau e^{-\alpha(\tau-t)}dW_t^r, \\ \mu_{x+\tau} &= \mu_x e^{\kappa\tau}+\psi\int_0^\tau e^{\kappa(\tau-u)}dW_u^\mu. \end{aligned}$$

Thus, the joint Gaussian distribution of Y_τ is given by:

$$\begin{bmatrix} q_\tau \\ r_\tau \\ \mu_{x+\tau} \end{bmatrix} \sim N\left(\begin{bmatrix} q_0+\left(m-\frac{1}{2}\sigma_S^2\right)\tau \\ r_0e^{-\alpha\tau}+\gamma\left(1-e^{-\alpha\tau}\right) \\ \mu_x e^{\kappa\tau} \end{bmatrix}, \begin{bmatrix} \sigma_S^2\,\tau & \rho_{12}\sigma_S\sigma_r B_r(0,\tau) & \rho_{13}\sigma_S\psi B_\mu(0,\tau) \\ \rho_{12}\sigma_S\sigma_r B_r(0,\tau) & \sigma_r^2\frac{1-e^{-2\alpha\tau}}{2\alpha} & \rho_{23}\sigma_r\psi\frac{1-e^{-(\alpha-\kappa)\tau}}{\alpha-\kappa} \\ \rho_{13}\sigma_S\psi B_\mu(0,\tau) & \rho_{23}\sigma_r\psi\frac{1-e^{-(\alpha-\kappa)\tau}}{\alpha-\kappa} & \psi^2\frac{e^{2\kappa\tau}-1}{2\kappa} \end{bmatrix}\right), \tag{30}$$

so that μ_τ and Σ_τ are given by

$$\mu_\tau = \begin{bmatrix} q_0+\left(m-\frac{1}{2}\sigma_S^2\right)\tau \\ r_0e^{-\alpha\tau}+\gamma\left(1-e^{-\alpha\tau}\right) \\ \mu_x e^{\kappa\tau} \end{bmatrix}, \quad \Sigma_\tau = \begin{bmatrix} \sigma_S^2\,\tau & \rho_{12}\sigma_S\sigma_r B_r(0,\tau) & \rho_{13}\sigma_S\psi B_\mu(0,\tau) \\ \rho_{12}\sigma_S\sigma_r B_r(0,\tau) & \sigma_r^2\frac{1-e^{-2\alpha\tau}}{2\alpha} & \rho_{23}\sigma_r\psi\frac{1-e^{-(\alpha-\kappa)\tau}}{\alpha-\kappa} \\ \rho_{13}\sigma_S\psi B_\mu(0,\tau) & \rho_{23}\sigma_r\psi\frac{1-e^{-(\alpha-\kappa)\tau}}{\alpha-\kappa} & \psi^2\frac{e^{2\kappa\tau}-1}{2\kappa} \end{bmatrix}.$$

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

$$\begin{aligned} r_s =& e^{-\alpha(s-\tau)}r_\tau+\left(\bar{\gamma}-\frac{\sigma_r^2}{\alpha^2}\right)\left(1-e^{-\alpha(s-\tau)}\right)+\frac{\sigma_r^2}{2\alpha^2}\left(e^{-\alpha(T-s)}-e^{-\alpha(T+s-2\tau)}\right) \\ &-\frac{\rho_{23}\sigma_r\psi}{\kappa}\left(\frac{e^{\kappa(T-s)}-e^{-\alpha(s-\tau)+\kappa(T-\tau)}}{\alpha-\kappa}-\frac{1-e^{-\alpha(s-\tau)}}{\alpha}\right)+\sigma_r\int_\tau^s e^{-\alpha(s-y)}dZ_y^r, \end{aligned}$$

so that the integral of $\int_\tau^T r_s\,ds$ can be evaluated using the stochastic Fubini theorem:

$$\begin{aligned}\int_\tau^T r_s ds =& \left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) r_\tau + \left(\bar\gamma - \frac{\sigma_r^2}{\alpha^2}\right)\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)\\ &+\frac{\sigma_r^2}{2\alpha^2}\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha}-\frac{e^{-\alpha(T-\tau)}-e^{-2\alpha(T-\tau)}}{\alpha}\right)\\ &-\frac{\rho_{23}\sigma_r\psi}{\kappa}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa(\alpha-\kappa)}-\frac{e^{\kappa(T-\tau)}-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha(\alpha-\kappa)}-\frac{1}{\alpha}\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)\right)\\ &+\sigma_r\int_\tau^T\frac{1-e^{-\alpha(T-y)}}{\alpha}dZ_y^r.\end{aligned}$$

Thus, under $\mathbb{Q}_E$ with known Y_τ, the solutions of (21), (22), and (23) are:

$$\begin{aligned}q_T =& q_\tau + \left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) r_\tau + \left(\bar\gamma - \frac{\sigma_r^2}{\alpha^2}\right)\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)\\ &+\frac{\sigma_r^2}{2\alpha^2}\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha}-\frac{e^{-\alpha(T-\tau)}-e^{-2\alpha(T-\tau)}}{\alpha}\right)\\ &-\frac{1}{2}\sigma_S^2(T-\tau)-\frac{\rho_{12}\sigma_S\sigma_r}{\alpha}\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)-\frac{\rho_{13}\sigma_S\psi}{\kappa}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa}-T+\tau\right)\\ &-\frac{\rho_{23}\sigma_r\psi}{\kappa}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa(\alpha-\kappa)}-\frac{e^{\kappa(T-\tau)}-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha(\alpha-\kappa)}-\frac{1}{\alpha}\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)\right)\\ &+\sigma_S\int_\tau^T dZ_s^S+\sigma_r\int_\tau^T\frac{1-e^{-\alpha(T-y)}}{\alpha}dZ_y^r,\\ r_T =& e^{-\alpha(T-\tau)}r_\tau+\left(\bar\gamma-\frac{\sigma_r^2}{\alpha^2}\right)\left(1-e^{-\alpha(T-\tau)}\right)+\frac{\sigma_r^2}{2\alpha^2}\left(1-e^{-2\alpha(T-\tau)}\right)\\ &-\frac{\rho_{23}\sigma_r\psi}{\kappa}\left(\frac{1-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha-\kappa}-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)+\sigma_r\int_\tau^T e^{-\alpha(T-y)}dZ_y^r,\\ \mu_{x+T} =& e^{\kappa(T-\tau)}\mu_{x+\tau}-\frac{\psi^2}{\kappa}\left(\frac{e^{2\kappa(T-\tau)}-1}{2\kappa}-\frac{e^{\kappa(T-\tau)}-1}{\kappa}\right)-\frac{\rho_{23}\sigma_r\psi}{\alpha}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa}-\frac{1-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha-\kappa}\right)\\ &+\psi\int_\tau^T e^{\kappa(T-t)}dZ_t^\mu,\end{aligned}$$

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

$$\begin{pmatrix}q_T\\ r_T\\ \mu_{x+T}\end{pmatrix}|Y_\tau \sim\ N\left(\begin{bmatrix}\mu_{q_T|q_\tau}\\ \mu_{r_T|r_\tau}\\ \mu_{\mu_{x+T}|\mu_{x+\tau}}\end{bmatrix},\ \underbrace{\begin{bmatrix}\sigma^2_{q_T|q_\tau} & \sigma_{q_T,r_T|q_\tau,r_\tau} & \sigma_{q_T,\mu_{x+T}|q_\tau,\mu_{x+\tau}}\\ \sigma_{q_T,r_T|q_\tau,r_\tau} & \sigma^2_{r_T|r_\tau} & \sigma_{r_T,\mu_{x+T}|r_\tau,\mu_{x+\tau}}\\ \sigma_{q_T,\mu_{x+T}|q_\tau,\mu_{x+\tau}} & \sigma_{r_T,\mu_{x+T}|r_\tau,\mu_{x+\tau}} & \sigma^2_{\mu_{x+T}|\mu_{x+\tau}}\end{bmatrix}}_{\Sigma_{T|\tau}}\right), \tag{31}$$

where:

$$\begin{aligned}
\mu_{q_T|q_\tau} =& q_\tau + B_r(\tau,T)r_\tau + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) \\
&+ \frac{\sigma_r^2}{2\alpha^2}\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha} - \frac{e^{-\alpha(T-\tau)}-e^{-2\alpha(T-\tau)}}{\alpha}\right) \\
&- \frac{\rho_{23}\sigma_r\psi}{\kappa}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa(\alpha-\kappa)} - \frac{e^{\kappa(T-\tau)}-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha(\alpha-\kappa)} - \frac{1}{\alpha}\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right)\right) \\
&- \frac{1}{2}\sigma_S^2(T-\tau) - \frac{\rho_{12}\sigma_S\sigma_r}{\alpha}\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) - \frac{\rho_{13}\sigma_S\psi}{\kappa}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa} - T + \tau\right),
\end{aligned}$$

$$\begin{aligned}
\mu_{r_T|r_\tau} =& e^{-\alpha(T-\tau)}r_\tau + \left(\bar{\gamma} - \frac{\sigma_r^2}{\alpha^2}\right)\left(1-e^{-\alpha(T-\tau)}\right) + \frac{\sigma_r^2}{2\alpha^2}\left(1-e^{-2\alpha(T-\tau)}\right) \\
&- \frac{\rho_{23}\sigma_r\psi}{\kappa}\left(\frac{1-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha-\kappa} - \frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right), \\
\mu_{\mu_{x+T}|\mu_{x+\tau}} =& \mu_{x+\tau}e^{\kappa(T-\tau)} - \frac{\rho_{23}\sigma_r\psi}{\alpha}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa} - \frac{1-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha-\kappa}\right) \\
&- \frac{\psi^2}{\kappa}\left(\frac{e^{2\kappa(T-\tau)}-1}{2\kappa} - \frac{e^{\kappa(T-\tau)}-1}{\kappa}\right), \\
\sigma_{q_T|q_\tau}^2 =& \sigma_S^2(T-\tau) + \frac{\sigma_r^2}{\alpha^2}\left(T-\tau-2\frac{1-e^{-\alpha(T-\tau)}}{\alpha} + \frac{1-e^{-2\alpha(T-\tau)}}{2\alpha}\right) \\
&+ \frac{2\rho_{12}\sigma_S\sigma_r}{\alpha}\left(T-\tau-\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right), \\
\sigma_{q_T,r_T|q_\tau,r_\tau} =& \rho_{12}\sigma_S\sigma_r\left(\frac{1-e^{-\alpha(T-\tau)}}{\alpha}\right) + \frac{\sigma_r^2}{\alpha}\left(\frac{1-2e^{-\alpha(T-\tau)}+e^{-2\alpha(T-\tau)}}{2\alpha}\right), \\
\sigma_{q_T,\mu_{x+T}|q_\tau,\mu_{x+\tau}} =& \rho_{13}\sigma_S\psi\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa}\right) + \frac{\rho_{23}\sigma_r\psi}{\alpha}\left(\frac{e^{\kappa(T-\tau)}-1}{\kappa} - \frac{1-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha-\kappa}\right), \\
\sigma_{r_T|r_\tau}^2 =& \sigma_r^2\left(\frac{1-e^{-2\alpha(T-\tau)}}{2\alpha}\right), \\
\sigma_{r_T,\mu_{x+T}|r_\tau,\mu_{x+\tau}} =& \rho_{23}\sigma_r\psi\left(\frac{1-e^{-(\alpha-\kappa)(T-\tau)}}{\alpha-\kappa}\right), \\
\sigma_{\mu_{x+T}|\mu_{x+\tau}}^2 =& \psi^2\left(\frac{e^{2\kappa(T-\tau)}-1}{2\kappa}\right).
\end{aligned}$$

It is possible to write the conditional mean of Y_T given Y_τ in the following affine 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_τ is a constant matrix defined by remaining terms of mean vector of $Y_T|Y_\tau$ after defining HY_τ. The unconditional distribution of Y_T under $\tilde{\mathbb{P}}$ is also Gaussian since Y_τ 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:

$$\begin{aligned}\mu_T = \mathbb{E}^{\tilde{\mathbb{P}}}[Y_T] &= \mathbb{E}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}_E}[Y_T|Y_\tau]\right] = \mathbb{E}^{\mathbb{P}}\left[HY_\tau + C_\tau\right] = H\mu_\tau + C_\tau, \\ \Sigma_T = \mathrm{Cov}^{\tilde{\mathbb{P}}}[Y_T] &= \mathrm{Cov}^{\mathbb{P}}\left[\mathbb{E}^{\mathbb{Q}_E}[Y_T|Y_\tau]\right] + \mathbb{E}^{\mathbb{P}}\left[\mathrm{Cov}^{\mathbb{Q}_E}[Y_T|Y_\tau]\right] \\ &= \mathrm{Cov}^{\mathbb{P}}\left[HY_\tau + C_\tau\right] + \mathbb{E}^{\mathbb{P}}\left[\Sigma_{T|\tau}\right] = H\Sigma_\tau H' + \Sigma_{T|\tau}.\end{aligned}$$

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

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

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

Therefore,

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

□

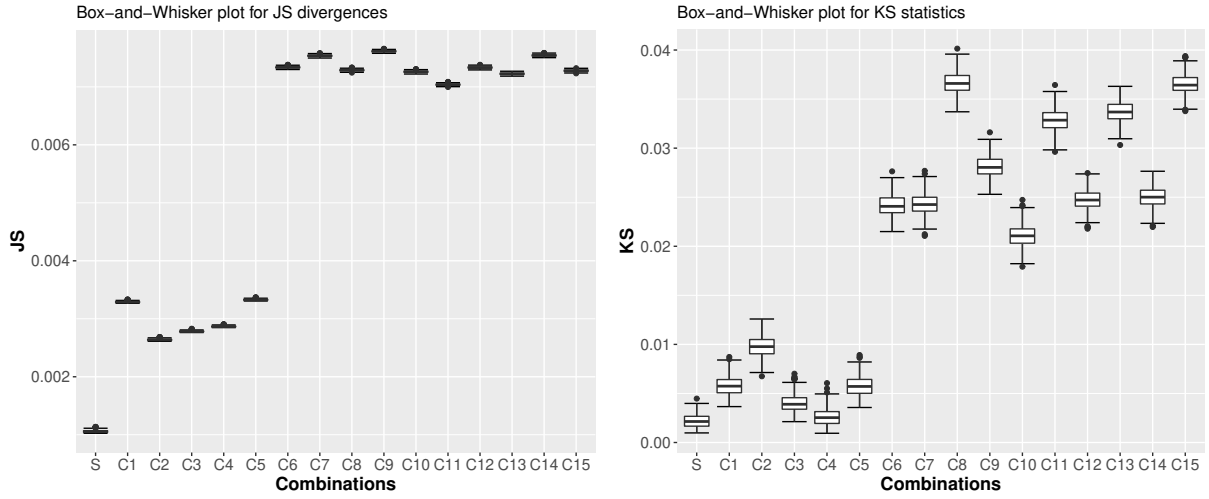
B Supplemental Analyses to Section 5

B.1 Additional Analyses in the Gaussian Case (Section 5.2)

Comparison to Conventional Basis Functions.

Figure 6 shows supplemental results for the analysis in Figure 2, particularly JS divergences and KS statistics. The conclusions are analogous to the KL and VaR results in Figure 2, respectively.

Table 2 shows results for different basis functions for different payoffs, where we modify the guaranteed annuity payment G_T in the payoff definition (24). More precisely, we show results for a less generous contract where we decrease the annuity by 5% ($0.95\,G_T$) and a



(a) JS (b) KS

Figure 6: Box-and-whisker plots for various statistical divergence measures and 99.5% VaR calculated using different basis functions with $M = 6$; based on 300 runs with $N = 3,000,000$ sample paths each.

(Higher-oder terms: C1: q_τ^2, r_τ^2; C2: q_τ^2, $\mu_{x+\tau}^2$; C3: q_τ^2, $q_\tau r_\tau$; C4: q_τ^2, $q_\tau \mu_{x+\tau}$; C5: q_τ^2, $r_\tau \mu_{x+\tau}$; C6: r_τ^2, $\mu_{x+\tau}^2$; C7: r_τ^2, $q_\tau r_\tau$; C8: r_τ^2, $q_\tau \mu_{x+\tau}$; C9: r_τ^2, $r_\tau \mu_{x+\tau}$; C10: $\mu_{x+\tau}^2$, $q_\tau r_\tau$; C11: $\mu_{x+\tau}^2$, $q_\tau \mu_{x+\tau}$; C12: $\mu_{x+\tau}^2$, $r_\tau \mu_{x+\tau}$; C13: $q_\tau r_\tau$, $q_\tau \mu_{x+\tau}$; C14: $q_\tau r_\tau$, $r_\tau \mu_{x+\tau}$; C15: $q_\tau \mu_{x+\tau}$, $r_\tau \mu_{x+\tau}$).

more generous contract where we increase the annuity by 5% ($1.05\, G_T$). In addition to the statistical divergences as in Table 1, we also report the first and third quartiles of 300 VaR estimates based on different sample paths. The estimated "exact" 99.5% VaR when using $0.95\, G_T$ is 137.89, and the "exact" 99.5% VaR for $1.05\, G_T$ is 141.79. The results are analogous to our analyses in the main text: Singular functions perform better than monomials, and the difference can be substantial for $M = 6$ and 12.

M vs. N: The tradeoff between the approximations.

Figure 7 shows supplemental results for the analysis in Figure 3, particularly JS divergences and VaR estimates. The conclusions for JS are analogous to KL, and the VaR results in line with Figure 4.

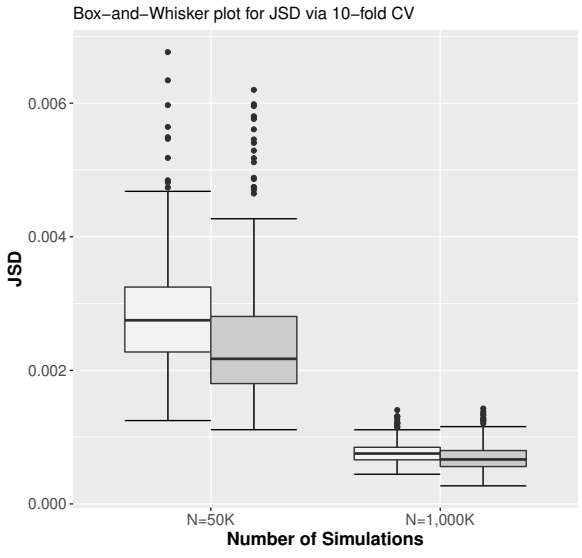
Risk measure estimation.

As outlined in the main text, we repeat our VaR estimations for $M = 16$ basis functions – using both, singular functions and monomials – where instead of OLS, we employ ridge regression to fit the approximation.[8] We choose the regularization parameter based on 10-fold cross-validation. Figure 8 displays box-and-whisker plots for the 99.5% VaR for $N = 50,000$

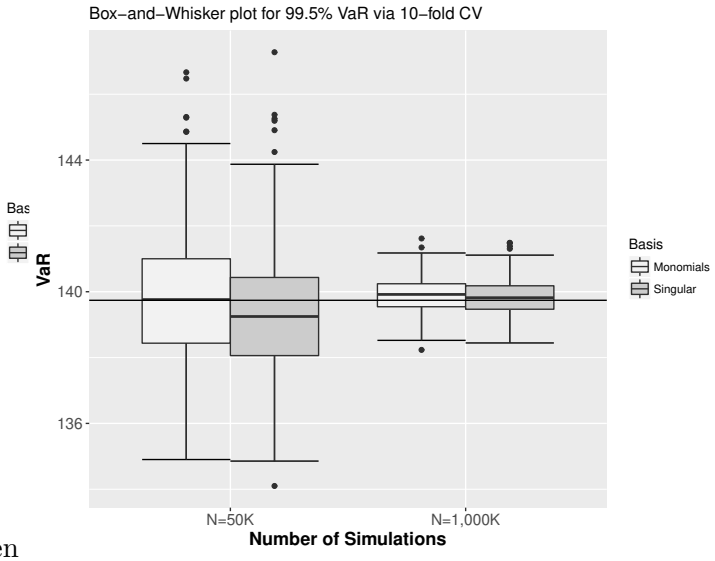
[8]The ridge regression coefficients, $\alpha_{ridge}^{(N)}$, minimize the sum of squares plus a regularization term $\eta \sum_{k=2}^{M} \alpha_k^2$, where η is a regularization parameter. We assume that the first basis function is always constant.

		$0.95\,G_T$		$1.05\,G_T$	
	Div.	Singular	Monomials	Singular	Monomials
$M=4$	KS	2.71×10^{-2}	2.98×10^{-2}	2.16×10^{-2}	2.59×10^{-2}
	KL	2.25×10^{-4}	2.39×10^{-4}	1.98×10^{-4}	2.14×10^{-4}
	JS	7.58×10^{-3}	7.80×10^{-3}	7.10×10^{-3}	7.38×10^{-3}
	$Q_{1,\text{VaR}}$	132.17	131.66	137.02	136.22
	$Q_{3,\text{VaR}}$	132.31	131.80	137.18	136.36
$M=6$	KS	2.48×10^{-3}	5.90×10^{-3}	1.99×10^{-3}	4.94×10^{-3}
	KL	4.61×10^{-6}	4.27×10^{-5}	4.43×10^{-6}	4.31×10^{-5}
	JS	1.07×10^{-3}	3.27×10^{-3}	1.05×10^{-3}	3.29×10^{-3}
	$Q_{1,\text{VaR}}$	137.09	138.10	140.99	141.99
	$Q_{3,\text{VaR}}$	137.34	138.35	141.26	142.25
$M=12$	KS	1.64×10^{-3}	1.70×10^{-3}	1.46×10^{-3}	1.58×10^{-3}
	KL	5.99×10^{-7}	1.56×10^{-6}	5.97×10^{-7}	1.57×10^{-6}
	JS	3.79×10^{-4}	6.22×10^{-4}	3.79×10^{-4}	6.26×10^{-4}
	$Q_{1,\text{VaR}}$	137.77	137.85	141.65	141.76
	$Q_{3,\text{VaR}}$	138.15	138.26	142.03	142.12

Table 2: Statistical divergence measures between the empirical density function based on the "exact" realizations and the LSM approximation using different basis functions; mean of 300 runs with $N = 3,000,000$ sample paths each. $Q_{1,\text{VaR}}$ and $Q_{3,\text{VaR}}$ are the first and third quartile of the distribution of the 99.5% VaR.

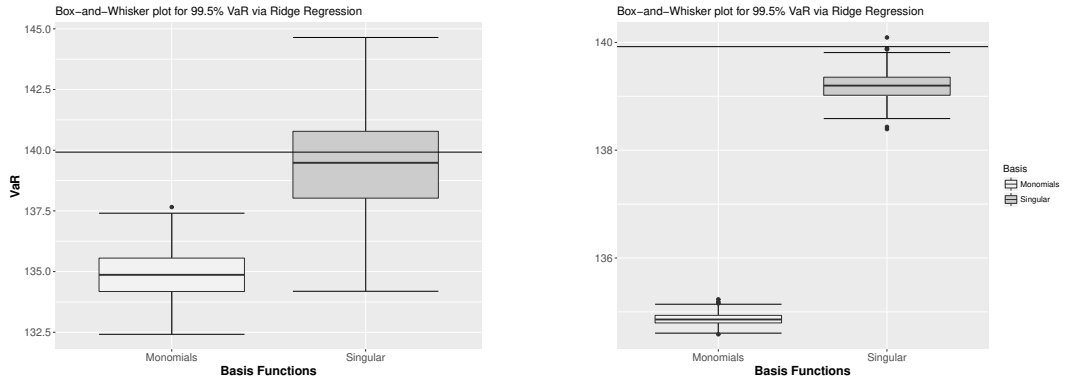


(a) Resulting JS divergence based on the chosen M



(b) Resulting VaR based on the chosen M

Figure 7: JS and VaR results based on cross-validated number of basis function M



(a) $N = 50,000$ (b) $N = 3,000,000$

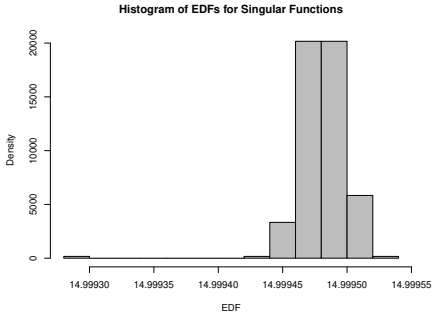
Figure 8: Box-and-whisker plots for VaR at 99.5% calculated using the LSM algorithm with ridge regression to fit parameters with different basis functions ($M = 16$) and different numbers of simulations N; based on 300 runs.

sample paths (left-hand panel (a)) and $N = 3,000,000$ sample paths (right-hand panel (b)), using both singular functions and monomials as the basis functions.

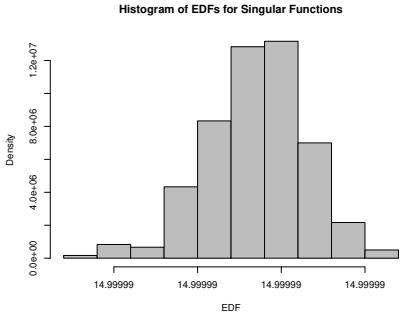
Comparing the results to the VaR estimates from Figure 4a, we find that relying on a regularized regression can be precarious. More precisely, while the box-and-whisker plot for the singular functions approximation in the case $N = 50,000$ seems to be roughly in line with the results from Figure 4, the plot when using monomials as basis functions – while notably tighter – now significantly undershoots and it no longer includes the "exact" VaR. The findings for $N = 3,000,000$ are similar, although now here the "exact" VaR is outside the whiskers also for the singular basis functions.

To provide intuition for these results, Figure 9 plots effective degrees of freedom (EDF) for the fits underlying Figure 8 (we refer to Hastie et al. (2009) for the definition of EDF in ridge regression). We notice that EDFs are very close to 15 for the singular functions, both for $N = 50,000$ and $N = 3,000,000$ (Panels (a) and (b)). Since 16 basis functions correspond exactly to 15 degrees of freedom (the constant term is always included), the shrinkage coming from the regularization is relatively minor – although it is clear from Figure 8b that the effect is still significant enough to move the VaR estimates downward. In contrast, for the monomial basis functions, the EDFs are around 10.9 and 13.25 for $N = 50,000$ and $N = 3,000,000$, respectively, so that here the shrinkage is substantial (Panels (c) and (d)) – explaining both the compression and the shifting of the box-and-whisker plots.

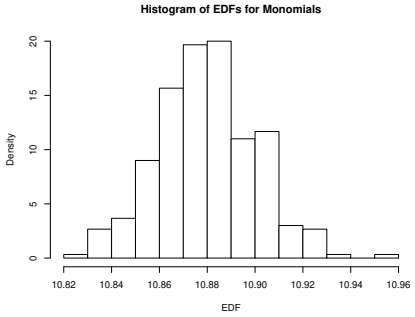
Since the regularization parameter is chosen so as to minimize the mean-squared prediction error across all realizations, it is not surprising that the predictions worsen in a certain area of the distribution. Indeed, it appears that the incurred bias for gains in terms of variance has an adverse effect for estimating VaR. For the singular functions, while the impact on the VaR estimate is limited, this is because overall there is very little change in the estimates relative to OLS – whereas for the monomial basis functions, the estimates are zooming in on the wrong target.



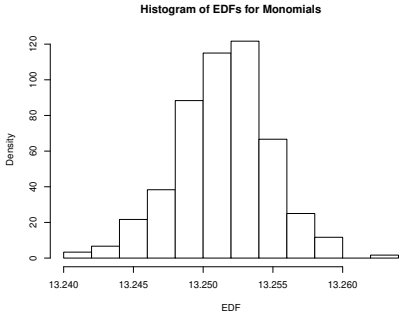
(a) $N = 50,000$, Singular Functions



(b) $N = 3,000,000$, Singular Functions

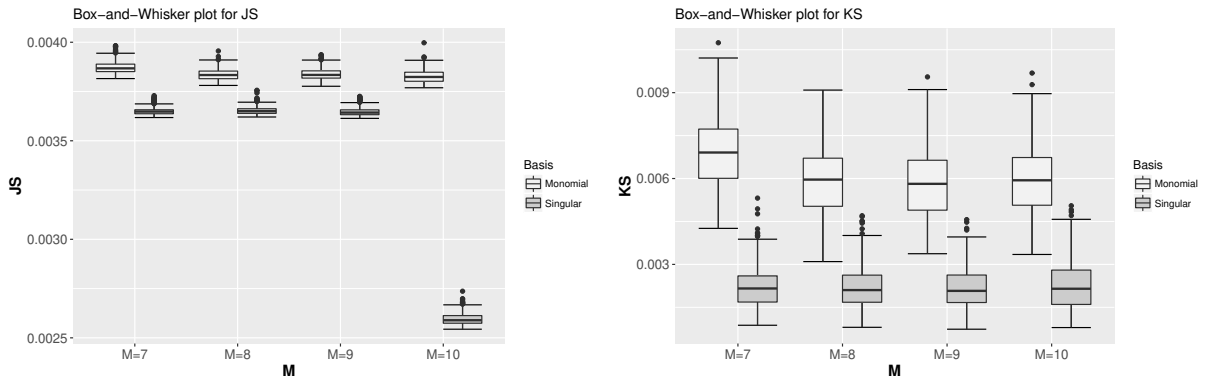


(c) $N = 50,000$, Monomials



(d) $N = 3,000,000$, Monomials

Figure 9: Histograms of EDFs for the ridge regression with singular functions and monomials using $M = 16$ basis functions and different numbers of simulations N; based on 300 runs.



(a) JS (b) KS

Figure 10: Box-and-whisker plots for the JS divergence and KS statistic calculated using different basis functions and different M; based on 500 runs with $N = 1,000,000$ sample paths each.

B.2 Additional Analyses in the General Case (Section 5.3)

Figure 10 shows supplemental results for the analysis in Figure 5, particularly JS divergences and KS statistics. The conclusions are analogous to the KL and VaR results in Figure 5, respectively.

References

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

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.

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

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica, 68: 1343-1376.

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

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

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

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

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

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.

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

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