STOCHASTIC APPROXIMATION SCHEMES FOR ECONOMIC CAPITAL AND RISK MARGIN COMPUTATIONS *

David Barrera[1], Stéphane Crépey[2], Babacar Diallo[2,3], Gersende Fort[4], Emmanuel Gobet[1] and Uladzislau Stazhynski[1]

Abstract. We consider the problem of the numerical computation of its economic capital by an insurance or a bank, in the form of a value-at-risk or expected shortfall of its loss over a given time horizon. This loss includes the appreciation of the mark-to-model of the liabilities of the firm, which we account for by nested Monte Carlo à la Gordy and Juneja [17] or by regression à la Broadie, Du, and Moallemi [10]. Using a stochastic approximation point of view on value-at-risk and expected shortfall, we establish the convergence of the resulting economic capital simulation schemes, under mild assumptions that only bear on the theoretical limiting problem at hand, as opposed to assumptions on the approximating problems in [17] and [10]. Our economic capital estimates can then be made conditional in a Markov framework and integrated in an outer Monte Carlo simulation to yield the risk margin of the firm, corresponding to a market value margin (MVM) in insurance or to a capital valuation adjustment (KVA) in banking parlance. This is illustrated numerically by a KVA case study implemented on GPUs.

Résumé. Nous considérons le problème pour une assurance, ou une banque, du calcul numérique de son capital économique sous forme d'une "value-at-risk", ou d'une "expected shortfall", de sa perte sur un horizon de temps donné. Cette perte inclut l'appréciation du modèle d'évaluation du passif de l'entité, que nous estimons par un Monte Carlo imbriquée à la Gordy et Juneja [17] ou par une régression à la Broadie, Du, et Moallemi [10]. En utilisant une approche d'approximation stochastique sur la value-at-risk, ou l'expected shortfall nous établissons la convergence des schémas résultants de la simulation du capital économique, sous des hypothèses modérées ne portant que sur le problème de limite théorique, par opposition aux hypothèses sur les problèmes d'approximation dans [17] et [10]. Nos estimations de capital économique peuvent ensuite être conditionnées dans un cadre de Markov et intégrées dans une simulation de Monte Carlo externe afin de générer la marge de risque de l'entité, correspondant à une "market value margin" (MVM) en assurance ou à un "capital valuation adjustment" (KVA) dans le langage bancaire. Ceci est illustré numériquement par une étude de cas KVA mise en oeuvre sur des GPU.

[1] Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique and CNRS, Université Paris-Saclay, Route de Saclay, 91128 Palaiseau Cedex, France;e-mail: `david.barrera@polytechnique.edu` & `emmanuel.gobet@polytechnique.edu` & `uladzislau.stazhynski@polytechnique.edu`

[2] LaMME, Univ. Evry, CNRS, Université Paris-Saclay, 91037, Evry, France; e-mail: `stephane.crepey@univ-evry.fr`

[3] Quantitative Research GMD/GMT Crédit Agricole CIB, 92160 Montrouge, France; e-mail: `babacar.diallo@ca-cib.com`

[4] CNRS, Institut de Mathématiques de Toulouse (IMT), 31062 Toulouse Cedex 9, France; e-mail: `Gersende.Fort@math.univ-toulouse.fr`

© EDP Sciences, SMAI 2019

Introduction

The current financial and insurance regulatory trends incentivize investment banks and insurance companies to charge to their clients, on top of a risk-neutral expectation of contractual cash flows, a suitable risk margin (see [11], [25]), meant to be gradually released to shareholders as return for their capital at risk in the future. This risk margin, sometimes called market value margin (MVM) in insurance and corresponding in banking to a capital valuation adjustment (KVA, see [2]), can be modeled as an expectation of the future economic capital of the firm. Future economic capital is modeled in our paper as the conditional expected shortfall (ES)[1] of the losses of the firm over a one-year horizon. These losses are assessed on a mark-to-model basis, which includes, at any future time point where the conditional expected shortfall is computed, the valuation one year later of the liabilities of the firm, such as variable annuities (VA) in the insurance case or a credit valuation adjustment (CVA) in the banking case, i.e. an expectation, conditionally on the information available one year later, of the future cash flows that the liability is pricing.

As such complex liabilities are typically intractable analytically and because the losses of the firm are specified through dynamic models of the underlying risk factors, in principle, the computation of the risk margin involves a nested and, in fact, a doubly nested simulation, whereby an outer Monte Carlo simulation gathers inner estimates of conditional expected shortfalls at future time points, themselves calling for recursive valuation one year later of the embedded liability. This makes it a challenging problem, both from a practical and from a convergence analysis point of view. In particular, on realistically heavy applications, such computations can only be implemented in parallel, with GPUs as a current hardware paradigm, which poses nontrivial programming optimization issues.

The assumptions made in Gordy and Juneja [17] for establishing the convergence of the simulation-and-sort value-at-risk and expected shortfall nested Monte Carlo estimates are hard to check (and might actually be violated) in practice, especially when considered dynamically in the context of risk margin computations. As the value-at-risk and expected shortfall of a given loss random variable can jointly be represented as zeros of suitable functions that can be written as expectations, an alternative is stochastic approximation (SA). In the base case without embedded liability of the firm, the convergence of the value-at-risk and expected shortfall SA estimates is established in Bardou, Frikha, and Pagès [7, 8]. In the present paper this convergence is extended to the case of dependent noise, corresponding to the presence of the nested future liability of the firm in our loss variable. This is then applied to risk margin computations by embedding the resulting inner conditional ES estimates into an outer sample mean.

Moreover we analyze a variant of this approach where the future liabilities are regressed as in Broadie, Du, and Moallemi [10], rather than re-simulated in a nested fashion, resulting in a simply nested procedure for the overall risk margin computation.

The different variants of the method are tested numerically, using GPU programming so that the inner conditional risk measures can be computed in parallel and then averaged out for yielding the outer risk margin estimate.

Beyond the extension of the base result of [7,8] to dependent noise and its economical capital and risk margin application, we refer the reader to the concluding section of the paper regarding the technical contributions of our approach with respect to [17] and [10].

The paper is organized as follows. Section 1 presents our stochastic approximation value-at-risk and expected shortfall algorithms in the presence of dependent noise, with nested Monte Carlo versus regression estimates of the latter in the respective cases of Algorithms 1 and 2 (whereas Algorithm 0 corresponds to the base case without dependent noise). Sections 2 and 3 deal with the convergence analyses of the respective Algorithms 1 and 2. Section 4 casts such estimates in a dynamic setup, integrating out the estimated conditional economic capital in the context of an outer simulation for the corresponding risk margin; this is then illustrated numerically in the context of a KVA case study. Section 5 concludes.

---

[1]In the context of this paper where we are considering conditional ES, we avoid the alternative terminology of conditional value-at-risk for (unconditional) ES.

Remark 0.1. In the motivating discussion above and in our application Section 4, for concreteness, we focus on economic capital, modeled as expected shorfall, and on the ensuing risk margin. However, the results of Sections 2 and 3 cover both expected shorfall and value-at-risk (establishing convergence for the latter is in fact a prerequisite for the former). Hence, our results also cover the cases of value-at-risk, conditional value-at-risks, and integration of the latter in the context of an outer expectation. Again this can be relevant together for bank and for insurance, noting that:

- In the insurance case, Solvency capital is determined as the 99.5%-value-at-risk of the one year loss of the firm for Solvency II (see [11]), and as the 99%-expected shortfall for the Swiss Solvency Test (see [25]);
- In the banking case, Basel II Pillar II defines economic capital as the 99% value-at-risk of the depletion over a one-year period of core equity tier I capital (CET1) (where the latter corresponds the one year trading loss of the bank as detailed in [2, Section A.2]); But the FRTB required a shift from 99% value-at-risk to 97.5% expected shortfall as the reference risk measure in capital calculations. Moreover, value-at-risk is relevant to banks for the computation of their initial margin (with a time horizon of one or two weeks, as opposed to one year conventionally in the paper) and, in turn, of their dynamic (conditional) initial margin (see [3]) in the context of the computation of their margin valuation adjustment (MVA).

1. Stochastic Algorithms for Economic Capital Calculations

On some probability space $(\Omega, \mathcal{A}, \mathbb{P})$, our financial loss L is defined as a real valued random variable of the form

$$L = \phi + \beta \, \mathbb{E}_0 \left[\psi | Z_1 \right] - \mathbb{E}_0 \left[\psi' \right], \tag{1.1}$$

where $(\beta, \phi, \psi, \psi')$ are four real valued random variables and (Z_0, Z_1) are two $\mathbb{R}^q$ valued random variables such that under $\mathbb{P}_0$, the conditional probability measure $\mathbb{P}$ given Z_0 (with related expectation and variance denoted by $\mathbb{E}_0$ and $\mathbb{V}\text{ar}_0$):

- i.i.d. samples from (ϕ, β, Z_1) given Z_0 are available;
- i.i.d. samples from the conditional distribution of ψ given Z_1, denoted by $\Pi(Z_1, \cdot)$, are available;
- i.i.d. samples from the conditional distribution of ψ' given $\{Z_0 = z\}$, denoted by $\Pi'(z, \cdot)$, are available;
- the discount factor β is bounded: there exists a positive constant c_β such that $|\beta| \leq c_\beta$.

We denote by $\text{P}(z, \cdot)$ and $\text{Q}(z, \cdot)$ the distributions of Z_1 and L conditionally on $Z_0 = z$. We also write

$$\Psi(Z_1) := \mathbb{E}_0[\psi | Z_1] \text{ and } \Psi'(z) := \mathbb{E}_0[\psi'], \text{ so that } L = \phi + \beta \Psi(Z_1) - \Psi'(z). \tag{1.2}$$

In the financial application the second and third terms in (1.1) will be used for modeling the future (conventionally taken as 1, i.e. one year) and present (time 0) liability valuations, whereas the first term corresponds to the realized loss of the firm on the time interval $[0, 1]$. The above-listed assumptions allow recovering $\mathbb{E}_0 \left[\psi | Z_1 \right]$ by nested Monte Carlo simulation restarting from time 1 (which is the approach in [17]) or by empirical regression at time 1 (which is the approach in [10]), whereas $\mathbb{E}_0 \left[\psi' \right]$ can be obtained by a standard Monte Carlo simulation rooted at $(0, z)$.

Let

$$H_1(\xi, x) := 1 - \frac{1}{1 - \alpha} \mathbb{1}_{x > \xi}, \ H_2(\xi, \chi, x) := \chi - \xi - \frac{1}{1 - \alpha} (x - \xi)^+. \tag{1.3}$$

A value-at-risk $\xi_\star$ at level α of the random variable (loss) L solves the equation

$$1 - \frac{1}{1 - \alpha} \mathbb{P}_0 \left(L > \xi_\star \right) = \mathbb{E}_0 \left[H_1(\xi_\star, L) \right] = 0; \tag{1.4}$$

it is uniquely defined if L has an increasing $\mathbb{P}_0$ c.d.f F e.g. if it has a nonvanishing $\mathbb{P}_0$ density f. Given a solution $\xi_\star$ to (1.4), the expected shortfall $\chi_\star$ at level α solves the equation

$$\chi_\star - \xi_\star - \frac{1}{1-\alpha}\mathbb{E}_0\left[(L-\xi_\star)^+\right] = \mathbb{E}_0\left[H_2(\xi_\star,\chi_\star,L)\right] = 0 \tag{1.5}$$

(noting that any solution $\xi_\star$ to (1.4) yields the same equation (1.5) for $\chi_\star$; see e.g. Lemma A.1 in Appendix A.3). Equivalently, $\xi_\star$ and $\chi_\star$ satisfy

$$\int_{\xi_\star}^{+\infty} Q(z,dx) = 1 - \alpha, \qquad \chi_\star = \xi_\star + \frac{1}{1-\alpha}\int (x - \xi_\star)^+ Q(z,dx). \tag{1.6}$$

We model economic capital (EC) at time 0 (known in the insurance regulation as the Solvency capital requirement, SCR) as the expected shortfall of level $\alpha \in \left(\frac{1}{2}, 1\right)$ of the distribution of L given $Z_0 = z$, i.e.

$$\mathrm{ES}(z) := (1-\alpha)^{-1}\int_\alpha^1 \mathbb{V}\mathrm{a}\mathbb{R}_0^a[L]\,\mathrm{d}a. \tag{1.7}$$

In (1.7), $\mathbb{V}\mathrm{a}\mathbb{R}_0^a[L]$ is a corresponding value-at-risk at level a. Throughout the paper, α is fixed, so the dependence of $\mathrm{ES}(z)$ upon α is omitted. Likewise we introduce the notation $\mathrm{VaR}(z)$ for the value-at-risk at the (fixed) level α of L.

1.1. Stochastic Approximation (SA) With Dependent Noise

We propose two approaches for computing $\mathrm{ES}(z)$. Both estimates $\widehat{\mathrm{ES}(z)}$ are defined as the output of a stochastic approximation (SA) algorithm with K iterations. However, in the applications targeted in this paper, the expectations in (1.4) and (1.5) are not known analytically, so that the quantities $(\xi_\star, \chi_\star)$ are roots of intractable functions. SA algorithms provide a numerical solution to (1.4)-(1.5) (see e.g. [9, 20]): given a deterministic stepsize sequence $\{\gamma_k, k \geq 1\}$ and a sequence $\{L_k, k \geq 1\}$ of random variables i.i.d. with distribution $Q(z, \cdot)$, we define iteratively, starting from (ξ_0, χ_0),

$$\begin{cases} \xi_{k+1} &= \xi_k - \gamma_{k+1}H_1(\xi_k, L_{k+1}) \\ \chi_{k+1} &= \chi_k - \gamma_{k+1}H_2(\xi_k, \chi_k, L_{k+1}). \end{cases} \tag{1.8}$$

Remark 1.1. In the case where $\gamma_k = \frac{1}{k}$, the first line in (1.8), for the value-at-risk specification of H_1 in (1.3), is equivalent to

$$\frac{1}{k}\sum_{l=1}^k \frac{\xi_{l-1}}{\xi_k}\mathbb{1}_{L_l > \xi_{l-1}} = 1 - \alpha, \ k \geq 1,$$

to be compared with the following empirical quantile ξ_k' specification:

$$\frac{1}{k}\sum_{l=1}^k \mathbb{1}_{L_l > \xi_k'} \approx 1 - \alpha.$$

The (almost-sure) limit $(\xi_\infty, \chi_\infty)$ of any convergent sequence $\{(\xi_k, \chi_k, k \geq 0\}$ is a solution to

$$(\xi, \chi) \mapsto \begin{cases} \mathbb{E}_0\left[H_1(\xi, L)\right] = 0, \\ \mathbb{E}_0\left[H_2(\xi, \chi, L)\right] = 0. \end{cases}$$

Therefore, any limit is a pair of solutions to (1.4)–(1.5). In particular, $\chi_\infty = \mathrm{ES}(z)$.

However, in our case, i.i.d. samples from the law of L are not available, because of the quantities $\mathbb{E}_1[\psi]$ and $\mathbb{E}_0[\psi']$ in L, which are not explicit. Therefore, we propose to replace exact sampling of L by approximate sampling. Toward this aim, we introduce two strategies.

Introducing i.i.d. $\{(\phi^k, \beta^k, Z_1^k), k \geq 1\}$ with the same distribution as (ϕ, β, Z_1) conditionally on $Z_0 = z$, the first strategy consists in replacing the draws $\{L_k, k \geq 1\}$ in (1.8) by

$$\phi_k + \frac{\beta_k}{M_k} \sum_{m=1}^{M_k} \psi_{m,k} - \frac{1}{M_k'} \sum_{m=1}^{M_k'} \psi_m', \tag{1.9}$$

where, conditionally on Z_1^k, $\{\psi_{m,k}, m \geq 1\}$ are i.i.d. with distribution $\Pi(Z_1^k, \cdot)$; conditionally on $Z_0 = z$, $\{\psi_m', m \geq 1\}$ are i.i.d. with distribution $\Pi'(z, \cdot)$; M_k, M_k' are positive integers.

Of course, conditionally on $Z_0 = z$, the second average in (1.9) can be updated at each step k using only the corresponding partial sum at step $k-1$ and the samples $\{\psi_m', M_{k-1}' < m \leq M_k'\}$.

The second strategy consists in replacing the draws $\{L_k, k \geq 1\}$ in (1.8) by

$$\phi_k + \beta_k \widehat{\Psi}(Z_1^k) - \frac{1}{M_k'} \sum_{m=1}^{M_k'} \psi_m',$$

where the first and last terms are as before and where $\widehat{\Psi}(\cdot)$ is a regression-based estimator, computed prior and independently from the Z_1^k, of the function $\Psi(\cdot)$, such that

$$\Psi(Z_1) = \mathbb{E}_0[\psi | Z_1], \qquad \mathrm{P}(z, \cdot)\text{-a.s} \tag{1.10}$$

(recall that $\mathrm{P}(z, \cdot)$ denotes the conditional distribution of Z_1 given $Z_0 = z$).

The advantage of the first approach is that, under sufficiently good convergence hypotheses for the nested averages (see the assumptions of Theorem 2.2), the approximation of $\mathrm{ES}(z)$ can be made asymptotically as good as desired. On the other side, the approach based on the regression requires a previous knowledge of the global behavior of Ψ (as an element of a certain function space) in order to give approximations with small bias (see Theorem 3.3), which is essential to have good asymptotics in our error analysis. Nevertheless, the second strategy has a small computational cost compared with the first one (at least for large values of the M_k in (1.9)). This can be a significant advantage if we indeed know which function space can serve to build a good predictor of the function Ψ.

Algorithmic summaries of these two strategies are given in the respective Sections 1.3 and 1.4. In Section 1.2, for pedagogical purposes, we start by recalling essentially known results in the base case where $\psi = \psi' = 0$.

1.2. **Base-case Without Present and Future Liabilities**

---

1 Input: A positive sequence $\{\gamma_k, k \geq 1\}$, $K \in \mathbb{N}^*$, $\xi_0 \in \mathbb{R}$, $\chi_0 \in \mathbb{R}$, and $z \in \mathbb{R}^q$.
2 for $k = 1$ to K, do
3   /* Sampling step   */
4   **Sample ϕ^k with the same distribution as ϕ conditionally on $Z_0 = z$, independently from the past draws ;**
5   **Set $L_k := \phi^k$;**
6   /* Update the conditional VaR and ES estimates   */
7 $\xi_k = \xi_{k-1} - \gamma_k\, H_1(\xi_{k-1}, L_k)$;
8 $\chi_k = \chi_{k-1} - \gamma_k\, H_2(\xi_{k-1}, \chi_{k-1}, L_k)$.
9 Return(*The sequences $\{\chi_k, 1 \leq k \leq K\}$ and $\{\xi_k, 1 \leq k \leq K\}$* **)**

Algorithm 0: Estimates of VaR(z) and ES(z) in the base case without present and future liabilities ($\psi = \psi' = 0$).

Note that, when $\psi \equiv \psi' = 0$, the random variables $\{L_k, k \geq 1\}$ are i.i.d. with distribution Q($z, \cdot$). Therefore, sufficient conditions on this distribution and on the sequence $\{\gamma_k, k \geq 1\}$ for the almost-sure convergence of $\{\xi_k, k \geq 0\}$ to VaR(z) and $\{\chi_k, k \geq 0\}$ to ES(z) can be proven by application of standard results for stochastic approximation algorithms: By application of Theorem A.3 and Lemma A.2 in Appendix A.2, we prove in Theorem 1.2 that the algorithm produces a sequence $\{(\xi_k, \chi_k), k \geq 1\}$ converging to a pair solution of (1.4)-(1.5) where $L \sim$ Q($z, \cdot$). Hence, ξ_K is a strongly consistent estimator of a value-at-risk of level α of the distribution Q($z, \cdot$), while χ_K is a (strongly) consistent estimator of the associated expected shortfall.

More precisely, these convergences are established under the following assumptions.

H1. $\{\gamma_k, k \geq 1\}$ *is a $(0,1)$-valued deterministic sequence, such that for some $\kappa \in (0,1]$,*

$$\sum_k \gamma_k = +\infty, \qquad \sum_k \gamma_k^{1+\kappa} < +\infty.$$

H2. *a) Under $\mathbb{P}_0$, $L = \phi \sim$ Q($Z_0 = z, \cdot$) has a continuous cumulative distribution function.*
b) $\mathbb{E}_0\left[L^2\right] = \int x^2 \mathrm{Q}(z, \mathrm{d}x) < +\infty$.

H1 is standard in stochastic approximation, and is satisfied for example with $\gamma_n \sim \gamma_\star/n^c$ and $c \in (1/2, 1]$. The condition H2 essentially allows to characterize the set of the limiting points of the algorithm and to prove that the stochastic approximation algorithm is a perturbation of a discretized ODE with a controlled noise.

Theorem 1.2. *Let $\{(\xi_k, \chi_k), k \geq 1\}$ be the output of Algorithm 0. Assume H1 and H2. Then there exist a bounded random variable ξ_∞ and a real number χ_∞ satisfying (1.6) $\mathbb{P}_0$-a.s. and such that, for any $p \in (0,2)$,*

$$\mathbb{P}_0\left(\lim_{k \to \infty}(\xi_k, \chi_k) = (\xi_\infty, \chi_\infty)\right) = 1, \qquad \lim_{k \to \infty} \mathbb{E}_0\left[|\xi_k - \xi_\infty|^p\right] = 0.$$

The proof of this result, which is very close to [7, Theorem 1], is detailed in Appendix A.3. The proof consists in first proving the almost-sure convergence of the sequence $\{\xi_k, k \geq 0\}$ toward the set of solutions of (1.4) by applying classical results on the convergence of stochastic approximation scheme; for the sake of completeness these results are stated and proved as Theorem A.3 in Appendix A.2. We then deduce the convergence of the sequence $\{\chi_k, k \geq 0\}$ by using the fact that χ_k can be written as a weighted sum of the samples $\{L_j, \xi_j, 0 \leq j \leq k\}$ (see Lemma A.2 in Appendix A.1).

Remember that although the set of solutions to the equation $\xi : 1 - \alpha = \int_\xi^\infty \mathrm{Q}(z, \mathrm{d}x)$ might not be a singleton (when VaR(z) is not unique), ES(z) is unique - see Lemma A.1 in Appendix A.3.

1.3. With Future Liability Estimated by Nested Monte Carlo

---

1 **Input: A positive sequence $\{\gamma_k, k \geq 1\}$, $\mathbb{N}^*$-valued sequences $\{M_k, M_k', k \geq 1\}$, $\xi_0 \in \mathbb{R}$, $\chi_0 \in \mathbb{R}$,**
  $S_0' = 0$, $M_0' = 0$ **and** $z \in \mathbb{R}^q$.
2 **for** $k = 1$ **to** K, **do**
3   /* Sampling step   */
4   **Sample (ϕ^k, β^k, Z_1^k) with the same distribution as (ϕ, β, Z_1) conditionally on $Z_0 = z$,**
      independently from the past draws ;
5   **Sample $(M_k' - M_{k-1}')$ independent copies $\{\psi_m', M_{k-1}' < m \leq M_k'\}$ with the distribution**
      $\Pi'(z, \cdot)$**, independently from the past draws;**
6   **Given Z_1^k, sample M_k independent copies $\{\psi_{m,k}, 1 \leq m \leq M_k\}$ with the distribution $\Pi(Z_1^k, \cdot)$;**
7   **Compute**
8 $S_k' := S_{k-1}' + \sum_{m=M_{k-1}'+1}^{M_k'} \psi_m'$;
9 $L_k := \phi^k + \beta^k \frac{1}{M_k} \sum_{m=1}^{M_k} \psi_{m,k} - \frac{1}{M_k'} S_k'$;
10   /* Update the conditional VaR and ES estimates   */
11 $\xi_k = \xi_{k-1} - \gamma_k H_1(\xi_{k-1}, L_k)$;
12 $\chi_k = \chi_{k-1} - \gamma_k H_2(\xi_{k-1}, \chi_{k-1}, L_k)$.
13 **Return(***The sequences $\{\chi_k, 1 \leq k \leq K\}$ and $\{\xi_k, 1 \leq k \leq K\}$***)**

Algorithm 1: Estimates of VaR(z) and ES(z) with future liability estimated by nested Monte Carlo.

Note that the random variables $\{L_k, k \geq 1\}$ have the same distribution, but this distribution is not Q($z, \cdot$), the distribution of L given by (1.1): there is a bias which, roughly speaking, can be made as small as possible by choosing M_k, M_k' large enough.

We provide in Section 2.1 sufficient conditions on Q($z, \cdot$) and on the sequences $\{\gamma_k, k \geq 1\}$, $\{M_k, k \geq 1\}$, $\{M_k', k \geq 1\}$ for the $\mathbb{P}_0$-a.s. convergence of $\{\xi_k, k \geq 0\}$ to VaR(z) and $\{\chi_k, k \geq 0\}$ to ES(z). We also provide convergence rates in Section 2.2 and show the benefit of considering the averaged outputs $K^{-1} \sum_{k=1}^K \xi_k$ and $K^{-1} \sum_{k=1}^K \chi_k$ as estimators of VaR(z) and ES(z).

1.4. With Future Liability Estimated by Regression

The regression approach relies on the following observation: The function Ψ in (1.10) satisfies

$$\Psi = \operatorname{argmin}_{h \in \mathcal{H}} \int_{\mathbb{R}^q} \int_{\mathbb{R}} (w - h(z_1))^2 \, \Pi(z_1, \mathrm{d}w) \mathrm{P}(z, \mathrm{d}z_1), \tag{1.11}$$

where $\mathcal{H}$ denotes the set of Borel measurable, P($z, \cdot$)-square integrable functions from $\mathbb{R}^q$ to $\mathbb{R}$. Since the integral in (1.11) is not explicit but sampling from the conditional distribution of (ψ, Z_1) given $Z_0 = z$ is possible, we define the estimate $\widehat{\Psi}(\cdot)$ as the solution of the empirical criterion associated with (1.11), replacing this integral by a Monte Carlo sum with i.i.d. samples.

If furthermore we replace the 'complex' functional space $\mathcal{H}$ by a space $\widehat{\mathcal{H}}$ suitable to least squares estimation, typically a finite-dimensional vector space of functions (but not necessarily, possibly also e.g. a neural network), we obtain a version of (1.11) in which Ψ is approximated by the solution to a least squares regression problem. The best choice for $\widehat{\mathcal{H}}$ will depend on the specific problem at hand, typically on regularity assumptions regarding Ψ.

In order to make use of the distribution-free theory of non-parametric regression (as explained, for instance, in [18]), it is better to deal with bounded random variables to get nice statistical error estimates (through appropriate measure concentration inequalities). For this reason we consider the projection of the real-valued

random variable ψ on the interval $[-B, B]$:

$$\psi^B := \psi \mathbb{1}_{[|\psi| \leq B]} + B \, \text{sign}(\psi) \, \mathbb{1}_{[|\psi| > B]}, \tag{1.12}$$

where B is a large threshold assumed to be known by the user, and we write $\Psi^B(Z_1) := \mathbb{E}_0[\psi^B | Z_1]$.

This gives rise to the following Algorithm 2 for the estimation of $\text{ES}(z)$, using the embedded regression Algorithm 3.

---

1 Input: $B > 0$, $M, M' \in \mathbb{N}^*$, $K \in \mathbb{N}^*$, **a finite dimensional vector space** $\widehat{\mathcal{H}}$ **of Borel measurable functions from** $\mathbb{R}^q$ **to** $\mathbb{R}$, **a positive sequence** $\{\gamma_k, k \geq 1\}$, $\xi_0 \in \mathbb{R}$, $\chi_0 \in \mathbb{R}$ **and** $z \in \mathbb{R}^q$.

2 /* Regression step   */

3 Compute an approximation $\widehat{\Psi^B}(\cdot)$ **of** $\Psi(\cdot)$ **by Algorithm 3 with inputs** B, M **and** $\widehat{\mathcal{H}}$.

4 /* stochastic approximation step   */

5 Sample M' **independent copies** $\{\psi'_m, 1 \leq m \leq M'\}$ **with the distribution** $\Pi'(z, \cdot)$;

6 for $k = 1$ **to** K, **do**

7   **Sample** (ϕ^k, β^k, Z_1^k) **from the conditional distribution of** (ϕ, β, Z_1) **given** $Z_0 = z$, **independently from the past draws;**

8   **Compute** $L_k := \phi^k + \beta^k \, \widehat{\Psi^B}(Z_1^k) - \frac{1}{M'} \sum_{m=1}^{M'} \psi'_m$;

9   /* Update the conditional VaR and ES estimates   */

10 $\xi_k = \xi_{k-1} - \gamma_k H_1(\xi_{k-1}, L_k)$;

11 $\chi_k = \chi_{k-1} - \gamma_k H_2(\xi_{k-1}, \chi_{k-1}, L_k)$.

12 Return(*The sequences* $\{\chi_k, 1 \leq k \leq K\}$ *and* $\{\xi_k, 1 \leq k \leq K\}$**)**

Algorithm 2: Estimates of $\text{VaR}(z)$ and $\text{ES}(z)$ with future liability estimated by regression.

---

1 Input: $B > 0$, $M \in \mathbb{N}^*$, **a function space** $\widehat{\mathcal{H}}$ **of Borel measurable functions from** $\mathbb{R}^q$ **to** $\mathbb{R}$,

2 Sample M **independent copies** $\mathcal{D} = \{(\psi_m, Z_1^m), m = 1, \cdots, M\}$ **from the conditional distribution of** (ψ, Z_1) **given** $Z_0 = z$;

3 Compute

$$\tilde{h} := \arg \min_{h \in \widehat{\mathcal{H}}} \frac{1}{M} \sum_{m=1}^{M} \left(\psi_m^B - h(Z_1^m) \right)^2$$

     and set

$$\widehat{\Psi^B}(\cdot) := \text{sign}(\tilde{h}(\cdot)) \left(|\tilde{h}(\cdot)| \wedge B \right);$$

Return(*The function* $\widehat{\Psi^B}(\cdot)$**)**

Algorithm 3: Approximation of Ψ in (1.10) by empirical regression.

The analysis of Algorithm 2 is established in Section 3. Theorem 3.3 gives a control of the deviation in the $\text{L}^1_{\mathbb{P}_0}$-norm of the respective limits $|\xi_\infty - \text{VaR}(z)|$ and $|\chi_\infty - \text{ES}(z)|$, where $(\xi_\infty, \chi_\infty)$ is the (almost-sure) limit of (ξ_K, χ_K) as K goes to infinity, from biases (or "deterministic errors") given, up to multiplicative constants, as the respective square and cube roots of

$$\inf_{h \in \widehat{\mathcal{H}}} \mathbb{E}_0 |h(Z_1) - \Psi^B(Z_1)|^2 + \mathbb{E}_0[((|\psi| - B)^+)^2]. \tag{1.13}$$

The control of (1.13) depends on analytic features of the problem at hand, typically on the regularity of Ψ for the choice of $\widehat{\mathcal{H}}$ and on the distribution of ψ for the choice of B (see [18, Chapter 10] for a general discussion).[2]

---

[2]Note also that the term $\mathbb{E}_0[((|\psi| - B)^+)^2]$ controls the truncation error $\mathbb{E}_0|\Psi(Z_1) - \Psi^B(Z_1)|^2$ by Jensen's inequality:
$$\mathbb{E}_0|\Psi(Z_1) - \Psi^B(Z_1)|^2 = \mathbb{E}_0|\mathbb{E}_0[(\psi - \psi^B)|Z_1]|^2 \leq \mathbb{E}_0|\psi - \psi^B|^2 = \mathbb{E}_0[((|\psi| - B)^+)^2].$$

Putting everything together, these results are telling us how we should choose the inputs $(\widehat{\mathcal{H}}, B, M, M')$ in order to make the limit $(\chi_\infty, \xi_\infty)$ of the (χ_K, ξ_K) as close as desired from the target values $(\chi_\star, \xi_\star)$.

2. Convergence Analysis of the Economic Capital SA Algorithm 1 (Future Liabilities Estimated by Nested Monte Carlo)

Section 2.1 deals with the almost-sure convergence of Algorithm 1. Section 2.2 addresses the rate of convergence of Algorithm 1 along a converging sequence: a central limit theorem is established as well as the rate of convergence when an averaging technique is applied to the output of Algorithm 1.

2.1. Almost-sure Convergence

The difference between Algorithm 0 and Algorithm 1 is that $\beta \mathbb{E}_0\left[\psi | Z_1\right] - \mathbb{E}_0\left[\psi' | Z_0\right]$ in the definition of L (see (1.1)) is non zero. The expectations are untractable and they are approximated by Monte Carlo sums. Hence, the random variables $\{L_k, k \geq 1\}$ in Algorithm 1 are no more i.i.d. under the distribution $Q(z, \cdot)$. Nevertheless, when the number of Monte Carlo points tends to infinity, the Monte Carlo error vanishes, and it is expected that Algorithm 1 inherits the same asymptotic behavior as the one of Algorithm 0, in which the L_k are i.i.d. with distribution $Q(z, \cdot)$. We provide sufficient conditions for this intuition to hold. H5 strenghtens H1 by showing how the stepsize γ_k and the number of Monte Carlo points M_k, M'_k have to be balanced; H3 is in echo to H2. H4 (see also H6) is introduced to control the bias between the distributions of the L_k and $Q(z, \cdot)$.

We assume

H3. *Under* $\mathbb{P}_0$, $L := \phi + \beta \mathbb{E}_0\left[\psi | Z_1\right] - \mathbb{E}_0\left[\psi' | Z_0 = z\right] \sim Q(z, \cdot)$ *and it has a density with respect to the Lebesgue measure on* $\mathbb{R}$, *bounded by* $C_0(z) > 0$. *In addition,*

$$\mathbb{E}_0\left[|L|^2\right] = \int x^2 Q(z, \mathrm{d}x) < +\infty.$$

H4. *There exists* $p_\star \geq 2$ *such that*

$$C_{p_\star}(z) := \mathbb{E}_0\left[\int \left| w - \int u \Pi(Z_1, \mathrm{d}u) \right|^{p_\star} \Pi(Z_1, \mathrm{d}w)\right] + \int \left| w - \int u \Pi'(z, \mathrm{d}u) \right|^{p_\star} \Pi'(z, \mathrm{d}w)$$

is finite.

H5. *The sequences* $\{M_k, k \geq 1\}$ *and* $\{M'_k, k \geq 1\}$ *are* $\mathbb{N}^\star$-*valued,* $\{\gamma_k, k \geq 1\}$ *is a* $(0,1)$-*valued sequence, and there exists* $\kappa \in (0,1]$ *such that*

$$\sum_k \gamma_k = +\infty, \qquad \sum_k \gamma_k^{1+\kappa} < +\infty, \tag{2.1}$$

$$\sum_{k \geq 1} \gamma_k^{1-\kappa} \left(M_k \wedge M'_k\right)^{-p_\star/(1+p_\star)} < +\infty, \qquad \sum_{k \geq 1} \gamma_k \left(M_k \wedge M'_k\right)^{-1/2} < +\infty. \tag{2.2}$$

Let us discuss the condition H5 in the case $\gamma_k \sim \gamma_\star k^{-c}(\ln k)^{-\bar{c}}$ and $(M_k \wedge M'_k) \sim m_\star k^\mu (\ln k)^{\bar{\mu}}$ when $k \to +\infty$ (for some $c, \mu \geq 0$). Then (2.1) in H5 implies that $c \in [1/2, 1]$ (the case $c = 1/2$ implies $\kappa = 1$ and $\bar{c} > 1$).

When $c = 1$, we have to choose $\bar{c} \leq 1$ and $\mu > 0$ (note that the last condition in (2.2) does not allow $\mu = 0$). Therefore, the number of Monte Carlo points has to increase, even slowly, along the iterations; this comes from the fact that the Monte Carlo bias has to vanish along iterations to force Algorithm 1 to have the same behavior as Algorithm 0.

When $c = 1/2$, the slowest rate for $M_k \wedge M'_k$ is $\mu = 1 + 1/p_\star$, and in that case, $\bar{\mu} > 1 + 1/p_\star$ and $\bar{c} > 1$. Therefore, the number of Monte Carlo points has to increase more than linearly with k.

When $c \in (1/2, 1)$, the slowest rate for $M_k \wedge M'_k$ is $\mu = 2(1-c)(1+1/p_\star)$, and in that case, $\bar{c} > 1/c$ and $\bar{\mu} > (1+1/p_\star)(1 - \bar{c}(2 - 1/c))$.

The above discussion makes it apparent that either we choose a rapidly decaying stepsize sequence, and we have the weakest Monte Carlo cost; or we choose a slowly decaying stepsize sequence, but the number of Monte Carlo points has to increase more than linearly. It is known that for implementation efficiency, a slow decaying rate for γ_k is better during the burn-in phase of the algorithm (while it has not reached its asymptotic convergence rate).

If H4 is strenghtened into

H6. *There exists $C_\infty(z) > 0$ such that for any $\delta > 0$ and any integer M,*

$$\mathbb{P}_0 \left(\left| \frac{1}{M} \sum_{m=1}^{M} \psi_m - \mathbb{E}_0 \left[\psi | Z_1 \right] \right| > \delta \right) \vee \mathbb{P}_0 \left(\left| \frac{1}{M} \sum_{m=1}^{M} \psi'_m - \mathbb{E}_0 \left[\psi' | Z_0 \right] \right| > \delta \right) \leq e^{-C_\infty(z) \, M \, \delta^2},$$

where conditionally to (Z_0, Z_1), $\{\psi_m, m \geq 1\}$ are i.i.d. with distribution $\Pi(Z_1, \mathrm{d}w)$, and $\{\psi'_m, m \geq 1\}$ are i.i.d. with distribution $\Pi'(z, \mathrm{d}w)$.

then the condition (2.2) in H5 is weakened into

$$\sum_{k \geq 1} \gamma_k^{1-\kappa} \frac{\ln(M_k \wedge M'_k)}{(M_k \wedge M'_k)} < +\infty, \qquad \sum_{k \geq 1} \gamma_k \, (M_k \wedge M'_k)^{-1/2} < +\infty.$$

The above discussion on the choice of (c, μ) is essentially modified as follows (the choice of the logarithmic terms $\bar{c}, \bar{\mu}$ is not detailed): either $c = 1$ and $\mu > 0$, or $c \in [1/2, 1)$ and $\mu = 2(1-c)$.

The following proposition is fundamental in the proof of Theorem 2.2. It allows to control the error induced by drawing samples L_k under a distribution approximating $\mathrm{Q}(z, \cdot)$ instead of sampling from $\mathrm{Q}(z, \cdot)$. Its proof is postponed to Appendix A.4.

Lemma 2.1. *Assume H3 and H4. Let L' be a random variable such that*

$$|L - L'| \leq c_\beta \left| \frac{1}{M} \sum_{m=1}^{M} \psi_m - \int w \Pi(z_1, \mathrm{d}w) \right| + \left| \frac{1}{M'} \sum_{m=1}^{M'} \psi'_m - \int w \Pi'(z, \mathrm{d}w) \right|,$$

where conditionally on (Z_1, Z_0), $\{\psi_m, m \geq 1\}$ (resp. $\{\psi'_m, m \geq 1\}$) are i.i.d. with distribution $\Pi(Z_1, \cdot)$ (resp. $\Pi'(z, \cdot)$). Then,

$$\sup_{\xi \in \mathbb{R}} \mathbb{E}_0 \left[|\mathbb{1}_{L > \xi} - \mathbb{1}_{L' > \xi}| \right] \leq (1 \vee c_\beta)^{p_\star} \frac{2^{p_\star} \left(C_0(z) + c_{p_\star} C_{p_\star}(z) \right)}{(M \wedge M')^{p_\star/(2(1+p_\star))}}, \tag{2.3}$$

$$\sup_{\xi \in \mathbb{R}} \mathbb{E}_0 \left[\left| (L - \xi)^+ - (L' - \xi)^+ \right|^{p_\star} \right] \leq (1 \vee c_\beta)^{p_\star} \frac{c_{p_\star} C_{p_\star}(z)}{(M \wedge M')^{p_\star/2}}, \tag{2.4}$$

where $c_{p_\star}$ is a universal constant depending only on $p_\star$. When H4 is replaced with H6, then for any $M, M' \geq 3$,

$$\sup_{\xi \in \mathbb{R}} \mathbb{E}_0 \left[|\mathbb{1}_{L > \xi} - \mathbb{1}_{L' > \xi}| \right] \leq 2 \left(1 + \frac{C_0(z)}{\sqrt{2 \, C_\infty(z)}} \right) \sqrt{\frac{\ln(M \wedge M')}{(M \wedge M')}}. \tag{2.5}$$

We can now prove that the output of Algorithm 1 provides strongly consistent estimators of VaR(z) and ES(z). The proof of the next theorem is postponed to Appendix A.4.

Theorem 2.2. *Let $\{(\xi_k, \chi_k), k \geq 1\}$ be the output of Algorithm 1. Assume H3, H4, and H5. Then there exists a bounded random variable ξ_∞ and a real χ_∞ satisfying $\mathbb{P}_0$-a.s. (1.6) and such that for any $p \in (0, 2)$*

$$\mathbb{P}_0 \left(\lim_{k \to \infty} (\xi_k, \chi_k) = (\xi_\infty, \chi_\infty) \right) = 1, \qquad \lim_{k \to \infty} \mathbb{E}_0 \left[|\xi_k - \xi_\infty|^p \right] = 0.$$

2.2. Rates of Convergence of Algorithm 1

We establish a rate of convergence in L^2 and a central limit theorem, along a sequence $\{(\xi_k, \chi_k), k \geq 1\}$ converging to $(\xi_\star, \chi_\star)$, where $(\xi_\star, \chi_\star)$ is a solution to (1.6); this solution is fixed throughout this section. These results are derived under the following conditions.

H7. *$(\xi_\star, \chi_\star)$ solves (1.6). H3 holds and is strenghtened as follows: under $\mathbb{P}_0$, the density of $L := \phi + \beta \mathbb{E}_0 [\psi | Z_1] - \mathbb{E}_0 [\psi' | Z_0 = z] \sim \mathrm{Q}(z, \cdot)$ w.r.t. the Lebesgue measure on $\mathbb{R}$, denoted by $f(z, \cdot)$, is continuously differentiable in a neighborhood of $\xi_\star$ and strictly positive at $\xi_\star$. In addition, there exists $\nu_\star > 0$ such that*

$$\mathbb{E}_0 \left[|L|^{2+\nu_\star} \right] = \int |x|^{2+\nu_\star} \mathrm{Q}(z, \cdot) < +\infty.$$

H4 is strenghtened as follows: there exists $p_\star > 2$ such that

$$C_{p_\star}(z) := \mathbb{E}_0 \left[\int \left| w - \int u \Pi(Z_1, \mathrm{d}u) \right|^{p_\star} \Pi(Z_1, \mathrm{d}w) \right] + \int \left| w - \int u \Pi'(z, \mathrm{d}u) \right|^{p_\star} \Pi'(z, \mathrm{d}w) \tag{2.6}$$

is finite.

To make the assumptions simpler, we consider the case where the stepsize sequence $\{\gamma_k, k \geq 1\}$ is polynomially decreasing.

H8. *When $k \to \infty$, $\gamma_k \sim \gamma_\star k^{-c}$ where $c \in (1/2, 1]$ and $\gamma_\star > 0$; in the case $c = 1$, $2\gamma_\star > (1 \wedge (f(z, \xi_\star)/(1 - \alpha)))^{-1}$. In addition, c, M_k, M_k' satisfy*

$$\lim_k k^{c/2} \, (M_k \wedge M_k')^{-p_\star/(2(1+p_\star))} = 0, \tag{2.7}$$

where $p_\star$ is given by H7.

When $M_k \wedge M_k' \sim m_\star k^\mu$ when $k \to \infty$, the condition (2.7) is satisfied with $\mu > c(1 + 1/p_\star)$. In the case the condition (2.6) is replaced with H6, the condition (2.7) gets into

$$\lim_k k^{c/2} \sqrt{\frac{\ln(M_k \wedge M_k')}{M_k \wedge M_k'}} = 0,$$

which is satisfied with $\mu > c$.

Set $\theta_k := (\xi_k, \chi_k)$ and $\theta_\star := (\xi_\star, \chi_\star)$. Lemma 2.3 shows that $\theta_k - \theta_\star$ is bounded, in some sense, by $\sqrt{\gamma_k}$. Theorem 2.4 provides a central limit theorem, proving that, along converging paths, the normalized error $\gamma_k^{-1/2} (\theta_k - \theta_\star)$ behaves asymptotically as a Gaussian distribution.

Lemma 2.3. *Assume H7 and H8. Then, there exist positive random variables X_k, Y_k such that $\mathbb{P}_0(\sup_k |X_k| < \infty) = 1$, $\sup_k \mathbb{E}_0 [|Y_k|] < \infty$ and*

$$\gamma_k^{-1} |\theta_k - \theta_\star|^2 \mathbb{1}_{\lim_q \theta_q = \theta_\star} \leq X_k \, Y_k.$$

Set

$$\Gamma_c := \frac{1}{2} \begin{bmatrix} \alpha(1-\alpha)^2 \, (f(z, \xi_\star))^{-1} & 2\alpha \left(1 + \frac{f(z, \xi_\star)}{1-\alpha} \right)^{-1} \\ 2\alpha \left(1 + \frac{f(z, \xi_\star)}{1-\alpha} \right)^{-1} & \mathbb{V}\mathrm{ar}_0 \left[(L - \xi_\star)^+ \right] \end{bmatrix}$$

and

$$\Gamma_1 := \begin{bmatrix} (1-\alpha)\left(2\gamma_\star \frac{f(z,\xi_\star)}{1-\alpha} - 1\right)^{-1} & \mathbb{E}_0\left[(L-\xi_\star)^+\right]\left(\gamma_\star(1+\frac{f(z,\xi_\star)}{1-\alpha}) - 1\right)^{-1} \\ \mathbb{E}_0\left[(L-\xi_\star)^+\right]\left(\gamma_\star(1+\frac{f(z,\xi_\star)}{1-\alpha}) - 1\right)^{-1} & \mathbb{V}\mathrm{ar}_0\left[(L-\xi_\star)^+\right](2\gamma_\star - 1)^{-1}(\alpha\gamma_\star)^{-1} \end{bmatrix}$$

Theorem 2.4. *Assume H7 and H8. Let $\{\theta_k, k \geq 1\}$ be the output of Algorithm 1. Then, under the conditional probability $\mathbb{P}_0 \left(\cdot | \lim_q \theta_q = \theta_\star\right)$, the sequence $\{\gamma_k^{-1/2}(\theta_k - \theta_\star), k \geq 1\}$ converges in distribution to the centered bivariate normal distribution with covariance matrix $(1-\alpha)^{-2}\Gamma_c$ in the case $c \in (1/2, 1)$, and $(1-\alpha)^{-2}\alpha\gamma_\star\Gamma_1$ in the case $c = 1$ (where c is given by H8).*

The proof of Theorem 2.4 is postponed to Appendix A.6. Lemma 2.3 is a consequence of [16, Lemma 3.1.], applied to the same decomposition of $\theta_k - \theta_\star$ as in the proof of Theorem 2.4; details are omitted.

Theorem 2.4 shows that *(i)* the maximal rate of convergence is reached with a stepsize γ_k decaying at a rate $1/k$ as soon as $\gamma_\star$ is large enough (see H8 in the case $c = 1$); *(ii)* the limiting variance depends on $\gamma_\star$. In practice, the condition on $\gamma_\star$ is difficult to check since the quantity $f(z, \xi_\star)$ is unknown in many applications; in addition, it is known (see e.g. [9, Lemma 4, Chapter 3, Part I] or [16, Section 3]) that the optimal variance for an SA algorithm targeting the roots of the function

$$\theta = (\xi, \chi) \mapsto \begin{bmatrix} 1 - (1-\alpha)^{-1}\mathbb{P}_0(L > \xi) \\ \chi - \xi - (1-\alpha)^{-1}\mathbb{E}_0\left[(L-\xi)^+\right] \end{bmatrix}$$

is given by

$$\Gamma_\star := \begin{bmatrix} \frac{\alpha(1-\alpha)}{f^2(z,\xi_\star)} & \frac{\alpha}{1-\alpha}\frac{\mathbb{E}_0\left[(L-\xi_\star)^+\right]}{f(z,\xi_\star)} \\ \frac{\alpha}{1-\alpha}\frac{\mathbb{E}_0\left[(L-\xi_\star)^+\right]}{f(z,\xi_\star)} & \frac{\mathbb{V}\mathrm{ar}_0\left[(L-\xi_\star)^+\right]}{(1-\alpha)^2} \end{bmatrix}.$$

We prove in Theorem 2.5 that the optimal rate $O(1/k)$ and this optimal limiting variance $\Gamma_\star$ can be obtained by a simple post-processing of the output of Algorithm 1 run with $\gamma_k \sim \gamma_\star k^{-c}$ for some $c \in (1/2, 1)$. The proof of Theorem 2.5 is postponed to Appendix A.6. This post-processing technique is known in the literature as the Polyak-Ruppert averaging (see [21, 22]). Set

$$\bar{\theta}_k := \frac{1}{k}\sum_{\ell=1}^{k}\begin{bmatrix}\xi_\ell \\ \chi_\ell\end{bmatrix}$$

Theorem 2.5. *Let $\{\theta_k, k \geq 1\}$ be the output of Algorithm 1. Assume H7, $\gamma_k \sim \gamma_\star k^{-c}$ with $c \in (1/2, 1)$ and $\gamma_\star > 0$, and*

$$\lim_k k^c (M_k \wedge M_k')^{-p_\star/(2(1+p_\star))} = 0, \qquad \lim_k k^{-1/2}\sum_{l=1}^{k}(M_l \wedge M_l')^{-p_\star/(2(1+p_\star))} = 0. \qquad (2.8)$$

Then, under the conditional probability $\mathbb{P}_0 \left(\cdot | \lim_q \theta_q = \theta_\star\right)$, the sequence $\{k^{1/2}(\bar{\theta}_k - \theta_\star), k \geq 1\}$ converges in distribution to the centered bivariate normal distribution with covariance matrix $\Gamma_\star$.

When $M_k \wedge M_k' \sim m_\star k^\mu$, (2.8) is satisfied with $\mu > 2c(1 + 1/p_\star)$. In the case where the condition (2.6) is replaced with H6 in Theorem 2.5, then the condition (2.8) becomes

$$\lim_k k^c \frac{\ln(M_k \wedge M_k')}{(M_k \wedge M_k')^{1/2}} = 0, \qquad k^{-1/2}\sum_{l=1}^{k}\ln(M_l \wedge M_l')(M_l \wedge M_l')^{-1/2} = 0;$$

it is satisfied if $\mu > 2c$. Note that these conditions on μ are slightly more restrictive than what we obtained for the convergence of the sequence $\{\theta_k, k \geq 1\}$ in the case $c \in (1/2, 1)$.

3. Convergence Analysis of the Economic Capital SA Algorithm 2 (Future Liabilities Estimated by Regression)

In order to properly define $\widehat{\Psi^B}(Z_1)$ in $\widehat{L^B}$ as a random variable, we assume that the function space $\widehat{\mathcal{H}}$ is pointwise measurable.[3] We introduce the following object (cf. (3.1)):

$$\Psi^B := \mathbb{E}_0[\psi^B | Z_1], \qquad L^B := \phi + \beta \Psi^B(Z_1) - \Psi'(z). \tag{3.1}$$

For any fixed $g \in \widehat{\mathcal{H}}$, we define

$$L_{g^B} := \phi + \beta\, g^B(Z_1) - \frac{1}{M'} \sum_{m=1}^{M'} \psi'_m, \tag{3.2}$$

where $g^B : \mathbb{R}^q \to \mathbb{R}$ is the truncation of g by B: $g^B := \text{sign}(g)(|g| \wedge B)$. Last, for the approximation of Ψ^B obtained by regression (see Algorithm 3), we write

$$\widehat{L^B} := \phi + \beta\, \widehat{\Psi^B}(Z_1) - \frac{1}{M'} \sum_{m=1}^{M'} \psi'_m. \tag{3.3}$$

3.1. Existence of a Limit

H9. ϕ, β, Z_1 are independent from the regression sample $\mathcal{D}$ (defined in Algorithm 3) and $\{\psi'_m : 1 \le m \le M'\}$ are i.i.d. with distribution $\mathbb{P}_0$. They are independent from the ψ_m. In addition we have the square integrability conditions : $\mathbb{E}_0\left[|\phi|^2 + |\psi'|^2\right] < +\infty$.

Observe that the above assumption ensures in particular that, for any $g \in \widehat{\mathcal{H}}$,

$$\mathbb{E}_0\left[|L_{g^B}|^2\right] < +\infty.$$

We require an additional condition on L_{g^B}.

H10. For every $g \in \widehat{\mathcal{H}}$, L_{g^B} in (3.2) has a continuous cumulative distribution function under $\mathbb{P}_0$.

Lemma 3.1. Assume H1, H9, and H10. Let $\{(\xi_k, \chi_k), k \ge 1\}$ be the ouput of Algorithm 2. Then, conditionally on $\mathcal{D}$, there exist random variables $(\xi_\infty, \chi_\infty)$, finite a.s. which are a solution of (1.4)-(1.5) for $L = \widehat{L^B}$ under $\mathbb{P}_0(\,\cdot\mid\mathcal{D})$, and such that $\mathbb{P}_0(\lim_k(\xi_k, \chi_k) = (\xi_\infty, \chi_\infty)\mid\mathcal{D}) = 1$.

Proof. Given $g \in \widehat{\mathcal{H}}$, H1, H9 and H10 imply that the hypotheses of Theorem 1.2 are verified for every L_{g^B} as in (3.2) under the distribution $\mathbb{P}_0$. Hence, for fixed $\mathcal{D}$, the same is true for $\widehat{L^B}$ in (3.3) under the conditional distribution $\mathbb{P}_0(\,\cdot\mid\mathcal{D})$. The conclusion follows by application of Theorem 1.2. ∎

3.2. Error Analysis With a Given Approximate Model for the Regression Function

The next step is to bound the error between the initial model for L and the truncated and approximate model L_{g^B}, where we use the function g^B (for a given $g \in \widehat{\mathcal{H}}$) as a model of Ψ. For this we need Assumption 2 a) on the cumulative distribution function of L in (1.2) and its stronger version

H11. Assume Assumption H2 a). Denote by $(\xi_\star, \chi_\star)$ a solution to (1.4)-(1.5) with $L = \phi + \beta\Psi(Z_1) - \Psi'(z)$. The distribution of L, with $\mathbb{P}_0$ c.d.f. F, admits a density f under $\mathbb{P}_0$ bounded by C_f, this density is positive and continuous on a neighborhood of the interval

$$[\xi_\star - \zeta, \xi_\star + \zeta], \tag{3.4}$$

---

[3]I.e. there exists a countable subfamily of $\widehat{\mathcal{H}}$ with the property that every function in $\widehat{\mathcal{H}}$ is a pointwise limit of these functions. It includes finite dimensional vector spaces, neural networks with continuous activation function...

where

$$\zeta := 2^{1/3}(2C_f + 1)\left(\mathbb{E}_0|\beta g^B(Z_1) - \beta\Psi(Z_1)|^2 + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/3}.$$

Lemma 3.2. *Assume H9-H10 and H2 a), let $g \in \widehat{\mathcal{H}}$ be given and let $(\xi_{g^B,\star}, \chi_{g^B,\star})$ be a solution to (1.4)-(1.5) for L there defined by L_{g^B} in (3.2), then*

$$|\chi_{g^B,\star} - \chi_\star| \le \frac{1}{1-\alpha}\left(\mathbb{E}_0|\beta g^B(Z_1) - \beta\Psi(Z_1)| + \left(\frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/2}\right). \tag{3.5}$$

If the stronger condition H11 holds (for g), then

$$|\xi_{g^B,\star} - \xi_\star| \le 2^{1/3}(2C_f + 1)\sup_{x\in[\xi_\star-\zeta,\xi_\star+\zeta]}|f(F^{-1}(x))|^{-1}\left(\mathbb{E}_0|\beta g^B(Z_1) - \beta\Psi(Z_1)|^2 + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/3}. \tag{3.6}$$

Proof. We begin by proving (3.6), by an application of Corollary A.11. For this, we first estimate the Kolmogorov distance $d_{\mathrm{kol}}(L_{g^B}, L)$: actually Corollary A.13 with $p = 2$ gives

$$d_{\mathrm{kol}}(L_{g^B}, L) \le (2C_f + 1)(\mathbb{E}_0|L_{g^B} - L|^2)^{1/3}. \tag{3.7}$$

The difference in the expectation (3.7) is bounded as (see definitions (1.2) and (3.2))

$$|L_{g^B} - L|^2 \le 2\left|\beta g^B(Z_1) - \beta\Psi(Z_1)\right|^2 + 2\left(\frac{1}{M'}\sum_{m=1}^{M'}(\psi_m - \Psi'(z))\right)^2.$$

Therefore, we deduce

$$d_{\mathrm{kol}}(L_{g^B}, L) \le 2^{1/3}(2C_f + 1)\left(\mathbb{E}_0|\beta g^B(Z_1) - \beta\Psi(Z_1)|^2 + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/3} = \zeta. \tag{3.8}$$

Consequently, we can apply Corollary A.11 with $r = s = \zeta$, to get (3.6).

The inequality (3.5) follows in an easier way via (A.16) in Lemma A.10. ∎

3.3. **Error Analysis for the Randomly Optimal Regression Function**

Observe that by taking formally $g^B = \widehat{\Psi^B}$, we obtain, as a corollary of the previous proposition, a pathwise control between $(\xi_\infty, \chi_\infty)$ (associated to $\widehat{L^B}$) and $(\xi_\star, \chi_\star)$ (associated to L), for a given regression sample $\mathcal{D}$. By reintegrating over the learning sample $\mathcal{D}$, we shall obtain an estimate about the corresponding mean L_1 error. This strategy works nicely, in particular if we allow Assumption H11 to be valid with a

$$\zeta = 2^{1/3}(2C_f + 1)\left(\mathbb{E}_0|\beta g^B(Z_1) - \beta\Psi(Z_1)|^2 + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/3}$$

uniform in the learning sample $\mathcal{D}$. For this, set

$$\zeta_\infty := 2^{1/3}(2C_f + 1)\left(C_\beta^2\mathbb{E}_0\left[\max_{\varepsilon=\pm1}|\varepsilon B - \Psi(Z_1)|^2\right] + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/3} \tag{3.9}$$

which stands for a (rough) upper bound for ζ. This explains the following new assumption.

H12. *Assume Assumption H11 with $\zeta = \zeta_\infty$ defined in (3.9).*

Regarding the error analysis about the limits of Algorithm 2 (given by Lemma 3.1), our main result is now the following.

Theorem 3.3. *Assume H1, H9, H10, and H12. Let $B > 0$ and let $\widehat{\mathcal{H}}$ be a pointwise measurable function space with finite Vapnik-Chervonenkis dimension[4] $\mathbf{VC}_{\widehat{\mathcal{H}}}$. Set*

$$\mathcal{E}(\widehat{\mathcal{H}}, M, B) := C_\star B^2 \mathbf{VC}_{\widehat{\mathcal{H}}} \frac{(1 + \ln(M))}{M} + 4 \inf_{h \in \widehat{\mathcal{H}}} \mathbb{E}_0[|h(Z_1) - \Psi_B(Z_1)|^2] + 4\,\mathbb{E}_0[((|\psi| - B)^+)^2], \qquad (3.10)$$

where $C_\star$ is the constant that appears in (A.20). We have

$$\mathbb{E}_0|\chi_\infty - \chi_\star| \le \frac{1}{1 - \alpha} \left(C_\beta (\mathcal{E}(\widehat{\mathcal{H}}, M, B))^{1/2} + \left(\frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'}\right)^{1/2} \right) \qquad (3.11)$$

$$\mathbb{E}_0|\xi_\infty - \xi_\star| \le 2^{1/3}\,(2C_f + 1) \sup_{x \in [\xi_\star - \zeta_\infty, \xi_\star + \zeta_\infty]} |f(F^{-1}(x))|^{-1} \left(|\beta|_\infty^2 \mathcal{E}(\widehat{\mathcal{H}}, M, B) + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'} \right)^{1/3}. \qquad (3.12)$$

Theorem 3.3 gives a precise and useful guide for tuning the parameters all together. Namely, to make the (asymptotic) errors $\mathbb{E}_0|\chi_\infty - \chi_\star|$ and $\mathbb{E}_0|\xi_\infty - \xi_\star|$ less than some tolerance ϵ, we can choose $\widehat{\mathcal{H}}$ and B such that the "bias" given by the second line in (3.10) is sufficiently small; then one can choose M and M' large enough so that the right hand sides in (3.11) are less than ϵ. Unsurprisingly, when the complexity of $\widehat{\mathcal{H}}$ increases, the bias term ($\inf_{h \in \widehat{\mathcal{H}}} \dots$) goes to 0 and the variance term explodes ($\mathbf{VC}_{\widehat{\mathcal{H}}} \to +\infty$), hence one has to find a trade-off between those types of error. When one increases the threshold B, the bias decreases $\mathbb{E}_0[((|\psi| - B)^+)^2]$ but the variance increases (factor $C_\star B^2 \dots$).

Proof. First, by H1, H9, H10 and Lemma 3.1, the limits

$$\xi_\infty = \lim_k \xi_k, \quad \chi_\infty = \lim_k \chi_k$$

indeed exist for every fixed $\mathcal{D}$ and they correspond to solutions of (1.4)-(1.5) for $L = \widehat{L^B}$ (see (1.2)) under $\mathbb{P}_0(\cdot \mid \mathcal{D})$. Now apply Lemma 3.2, valid for any $\mathcal{D}$ since H2 a) holds for all g^B owing to the choice $\zeta = \zeta_\infty$. As β is bounded, we obtain

$$|\chi_\infty - \chi_\star| \le \frac{1}{1 - \alpha} \left(C_\beta \left(\mathbb{E}_0 \left(|\widehat{\Psi^B}(Z_1) - \Psi(Z_1)|^2 \mid \mathcal{D} \right) \right)^{1/2} + \left(\frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'} \right)^{1/2} \right), \qquad (3.13)$$

$$|\xi_\infty - \xi_\star| \le 2^{1/3}\,(2C_f + 1) \sup_{x \in [\xi_\star - \zeta_\infty, \xi_\star + \zeta_\infty]} |f(F^{-1}(x))|^{-1} \qquad (3.14)$$

$$\times \left(C_\beta^2 \mathbb{E}_0 \left(|\widehat{\Psi^B}(Z_1) - \Psi(Z_1)|^2 \mid \mathcal{D} \right) + \frac{\mathbb{V}\mathrm{ar}_0(\psi)}{M'} \right)^{1/3}.$$

Now, write

$$\mathbb{E}_0 \left(|\widehat{\Psi^B}(Z_1) - \Psi(Z_1)|^2 \mid \mathcal{D} \right) \le 2\mathbb{E}_0 \left(|\widehat{\Psi^B}(Z_1) - \Psi^B(Z_1)|^2 \mid \mathcal{D} \right) + 2\mathbb{E}_0 \left(|\Psi^B(Z_1) - \Psi(Z_1)|^2 \right).$$

Note that the first expectation of the right hand side is exactly controlled using Theorem A.14. For the second term, write

$$|\Psi^B(Z_1) - \Psi(Z_1)| \le \mathbb{E}_0(|-B \vee \psi \wedge B - \psi| \mid Z_1) \le \mathbb{E}_0((|\psi| - B)^+ \mid Z_1). \qquad (3.15)$$

---

[4]See [18, Section 9.4].

We now easily obtain the desired estimates by taking the expectation in (3.13)-(3.14), applying Theorem A.14 and using (3.15), together with $\mathbb{E}(|Z|^{1/p}) \leq (\mathbb{E}(|Z|))^{1/p}$ for any $p \geq 1$. ∎

4. RISK MARGIN

4.1. **Dynamization of the Setup**

Let there be given an $\mathbb{R}^q$ valued process $Z = \{Z_t, t \geq 0\}$, with $Z_0 = z$, non-homogeneous Markov in its own filtration on our probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The process Z plays the role of observable risk factors. Conditional probabilities, expectations, value-at-risks and expected shortfalls at a level $a \in (0, 1)$, given Z_t, are denoted by $\mathbb{P}_t$, $\mathbb{E}_t$, $\mathbb{V}a\mathbb{R}_t^a$, and $\mathbb{E}\mathbb{S}_t^a$. Other sources of randomness arising in $(\Omega, \mathcal{A}, \mathbb{P})$ may be unobservable factors (like hidden financial variables, private information). We assume that Z can be simulated exactly (in other words, we ignore for the sake of simplicity a vanishing time discretization bias regarding Z, which could be considered without major difficulty). We denote by $\bar{Z}_t = (t, Z_t)$ the time-homogenized Markov extension of Z. We write $Z_{[s,t]}$ and $\bar{Z}_{[s,t]}$ for the paths of Z and $\bar{Z}$ on the interval $[s, t]$. We define the discount factor

$$\beta(\bar{Z}_{[0,t]}) := e^{-\int_0^t r(\bar{Z}_s)\mathrm{d}s},$$

for some bounded from below, continuous interest rate function r (hence, in particular, a bounded discount factor). We may then consider the following specification of (1.1):

$$L := \phi(\bar{Z}_{[0,1]}) + \beta(\bar{Z}_{[0,1]})\mathbb{E}_1\left[\psi(\bar{Z}_{[1,T]})\right] - \mathbb{E}_0\left[\psi'(\bar{Z}_{[0,T]})\right], \tag{4.1}$$

where ϕ and ψ are real valued measurable functions.

Remark 4.1. The functions ϕ and ψ could depend on variables other than Z, it would not have any significant impact on the analysis.

For instance, we could consider a Euro Median Term Note (EMTN), issued by a bank, with a performance linked to the 1 year Euribor rate denoted by Z; then the cashflow for the bank may take the form $\varphi(Z_1)\mathbb{1}_{\tau \geq 1} = \psi'(\bar{Z}_{[0,1]}, \tau)$, where τ is the default time for the bank (assumed independent from Z for simplicity).

In the regression setup of Algorithm 2, this flexibility of using "Z smaller than an underlying high-dimensional factor process" allows embedding in our framework the common industry practice of "partial regressions" with respect to reduced sets of factors.

More broadly, let, for $t \geq 0$ (cf. (4.1) for $t = 0$),

$$L_{t+1}^t := \phi(\bar{Z}_{[t,t+1]}) + \beta(\bar{Z}_{[t,t+1]})\,\mathbb{E}_{t+1}[\psi(\bar{Z}_{[t+1,T]})] - \mathbb{E}_t[\psi(\bar{Z}_{[t,T]})]. \tag{4.2}$$

Let $\mathbb{V}a\mathbb{R}_t^a[L_{t+1}^t]$ denote a value-at-risk at level $a \in \left(\frac{1}{2}, 1\right)$ of L_{t+1}^t for the conditional distribution[5] of L_{t+1}^t given $\mathcal{F}_t$, i.e.

$$\mathbb{P}_t\left(L_{t+1}^t > \mathbb{V}a\mathbb{R}_t^a[L_{t+1}^t]\right) = 1 - a.$$

Let

$$\mathrm{ES}(\bar{Z}_t) := (1 - \alpha)^{-1} \int_\alpha^1 \mathbb{V}a\mathbb{R}_t^a[L_{t+1}^t]\,\mathrm{d}a \tag{4.3}$$

denote the corresponding expected shortfall of (fixed) level $\alpha \in \left(\frac{1}{2}, 1\right)$.

---

[5]Assumed atomless, cf. Assumptions H2 a) and H3 again in the basic, nested, and regressed setups of Algorithms 0, 1, and 2, respectively.

4.2. **Theoretical Risk Margin Estimate**

The risk margin RM (called KVA in banking parlance) estimates how much it would cost the firm (bank or insurance) to remunerate its shareholders at the hurdle rate $h > 0$ (e.g. 10%) for their capital at risk $\mathrm{ES}(\bar{Z}_t)$ at any future time t (see Section). Given the final maturity T of the portfolio, the corresponding formula in [2] reads as

$$\mathrm{RM} = h\mathbb{E}\left[\int_0^T e^{-ht}\beta_t\mathrm{ES}(\bar{Z}_t)dt\right] = \mathbb{E}\left[\beta_\zeta\mathrm{ES}(\bar{Z}_\zeta)\,\mathbb{1}_{\zeta \leq T}\right], \tag{4.4}$$

where the second equality follows by randomization of the integral with an independent exponential time ζ of parameter h.

Accordingly, we propose the risk margin estimator

$$\mathbb{E}\left[\beta_\zeta\mathrm{ES}(\bar{Z}_\zeta)\,\mathbb{1}_{\zeta \leq T}\right] \approx \frac{1}{N}\sum_{n=1}^N \beta_{\zeta^n}\widehat{\mathrm{ES}}(\bar{Z}_{\zeta^n}^n)\mathbb{1}_{\zeta^n \leq T}, \tag{4.5}$$

where $\{\bar{Z}_{\zeta^n}^n, n \geq 1\}$ are independent random variables with the same distribution as $\bar{Z}_\zeta$ and where $\widehat{\mathrm{ES}}(\cdot)$ is one of the estimators of $\mathrm{ES}(\cdot)$ considered in the previous sections, now made conditional on Z_t.

The convergence of the ensuing estimator to the risk margin obtained by sampling an outer expectation of inner conditional expected shortfall estimates could be established by taking an outer expectation of the errors for $\widehat{\mathrm{ES}}(\bar{Z}_{\zeta^n}^n)$ estimates of the $\mathrm{ES}(\bar{Z}_{\zeta^n}^n)$ in (4.5), errors obtained from the conditional version of the results of Sections 2 and 3 (or, more precisely, of the awaited but technical developments of these results in terms of convergence rates). By contrast, how to "make conditional" the convergence arguments of [17] or [10] and "aggregate them" to establish the convergence of an outer risk margin estimate is far from clear.

4.3. **KVA Case Study**

Our case study is based on the setup of Armenti and Crépey [5], Section 4 (see also Section 4.4 in [1]), which we recall as a starting point. We consider a clearing house (or central counterparty, CCP for short) with a finite number (≥ 2) of clearing members labeled by i. We denote by:

- T: an upper bound on the maturity of all claims in the CCP portfolio, also accounting for a constant time $\delta > 0$ of liquidating the positions of defaulting clearing members;
- D_t^i: The cumulative contractual cash flow process of the CCP portfolio of the member i, cash flows being counted positively when they flow from the clearing member to the CCP;
- $\mathrm{MtM}_t^i = \mathbb{E}_t[\int_t^T \beta_t^{-1}\beta_s dD_s^i]$: The mark-to-market of the CCP portfolio of the member i;
- τ_i, $\tau_i^\delta = \tau_i + \delta$ and $\boldsymbol{\delta}_{\tau_i^\delta}(dt)$: The default and liquidation times of the member i, a Dirac measure at time τ_i^δ;
- $\Delta_{\tau_i^\delta}^i = \int_{[\tau_i,\tau_i^\delta]} \beta_t^{-1}\beta_s dD_s^i$: The cumulative contractual cash flows of the member i, accrued at the OIS rate, over the liquidation period of the clearing member i;
- IM_t^i: The initial margin (IM) posted by the member i as a guarantee in case it defaults, given at time t as a conditional value-at-risk (at a given confidence level a_{ma}) of $\beta_t^{-1}\left(\beta_{t+\delta}(\mathrm{MtM}_{t+\delta}^i + \Delta_{t+\delta}^i) - \mathrm{MtM}_t^i\right)^+$.

Beyond the first ring of defense provided by initial margin (and, of course, variation margin, which we assume equal to the process MtM_t^i stopped at time τ_i), a CCP maintains an additional resource, known as the default fund, against extreme and systemic risk. The current EMIR regulation sizes the default fund of a CCP by the Cover 2 rule, i.e. enough to cover the joint default of the two clearing members with the greatest CCP exposures, which purely relies on market risk. By contrast, we consider in the setup of this case study a broader risk-based specification, in the form of an economic capital of the CCP, which would be defined as a conditional expected shortfall, at some confidence level a_{df}, of its one-year ahead loss-and-profit if there was no default fund, as it results from the combination of the credit risk of the clearing members and of the market risk of

their portfolios. As developed in [5], such a specification can be used for allocating the default fund between the clearing members, after calibration of the quantile level a_{df} to the Cover 2 regulatory rule at time 0.

Specifically, we define the loss process of a CCP that would be in charge of dealing with member counterparty default losses through a CVA^{ccp} account (earning the risk-free rate r) as, for $t \in (0, T]$ (starting from some arbitrary initial value, since it is only the fluctuations of L^{ccp} that matter in what follows),

$$
\begin{aligned}
\beta_t dL_t^{ccp} = \sum_i &\left(\beta_{\tau_i^\delta}(\text{MtM}_{\tau_i^\delta}^i + \Delta_{\tau_i^\delta}^i) - \beta_{\tau_i}(\text{MtM}_{\tau_i}^i + \text{IM}_{\tau_i}^i) \right)^+ \boldsymbol{\delta}_{\tau_i^\delta}(dt) \\
&+ \beta_t (d\text{CVA}_t^{ccp} - r_t \text{CVA}_t^{ccp}) dt,
\end{aligned}
\tag{4.6}
$$

where the CVA of the CCP is given as

$$
\text{CVA}_t^{ccp} = \mathbb{E}_t \sum_{t < \tau_i^\delta < T} \beta_t^{-1} \left(\beta_{\tau_i^\delta}(\text{MtM}_{\tau_i^\delta}^i + \Delta_{\tau_i^\delta}^i) - \beta_{\tau_i}(\text{MtM}_{\tau_i}^i + \text{IM}_{\tau_i}^i) \right)^+, \ 0 \le t \le T
\tag{4.7}
$$

(in particular, L^{ccp} is constant from time T onward).

We define the corresponding economic capital process of the CCP as

$$
\text{EC}_t^{ccp} = \mathbb{ES}_t^{a_{df}} \left(\int_t^{t+1} \beta_t^{-1} \beta_s dL_s^{ccp} \right),
\tag{4.8}
$$

where, by (4.6),

$$
\begin{aligned}
\beta_t^{-1} \int_t^{t+1} \beta_s dL_s^{ccp} = \beta_t^{-1} \sum_{t < \tau_i^\delta \le t+1} &\left(\beta_{\tau_i^\delta}(\text{MtM}_{\tau_i^\delta}^i + \Delta_{\tau_i^\delta}^i) - \beta_{\tau_i}(\text{MtM}_{\tau_i}^i + \text{IM}_{\tau_i}^i) \right)^+ \\
&+ (\beta_t^{-1} \beta_{t+1} \text{CVA}_{t+1}^{ccp} - \text{CVA}_t^{ccp}).
\end{aligned}
\tag{4.9}
$$

The KVA (or risk margin) of the CCP estimates how much it would cost the CCP to remunerate all clearing members at some hurdle rate h for their capital at risk in the default fund from time 0 onward, namely, assuming a constant interest rate r (cf. (4.4), (4.8), and [5]):

$$
\text{KVA}^{ccp} = h\mathbb{E} \left[\int_0^T e^{-(r+h)s} \text{EC}_s^{ccp} ds \right].
\tag{4.10}
$$

For our numerics we consider the CCP toy model of Section 4 in [5] and Section 4.4 in [1], where nine members are clearing (interest rate or foreign exchange) swaps on a Black–Scholes underlying rate process X, with all the numerical parameters used there. The default times of the clearing members are defined through the common shock model or dynamic Marshall-Olkin copula (DMO) model of [12], Chapt. 8–10 and [13] (see also [14, 15]).

4.3.1. *Mapping with the General Setup*

This model, where defaults can happen simultaneously with positive probabilities, results in a Markovian pair $Z = (X, Y)$ made of, on the one hand, the underlying Black–Scholes rate X and, on the other hand, the vector Y of the default indicator processes of the clearing members. As a consequence, all conditional expectations, value-at-risks (embedded in the IM^i numbers), and expected shortfalls (embedded in the EC^{ccp} numbers) are

functions of the pair (t, Z), so that, with $Z = (X, Y)$ as above, we can identify

$$L_{t+1}^t \leftarrow \int_t^{t+1} \beta_t^{-1} \beta_s dL_s^{ccp},$$

$$\mathrm{ES}(t, z) \leftarrow \mathbb{ES}^{a_{df}} \left(\int_t^{t+1} \beta_t^{-1} \beta_s dL_s^{ccp} \,\Big|\, Z_t = z \right).$$

The ensuing KVA can be computed by Algorithms 0 (for validation purposes, building on the explicit CVA^{ccp} formulas that are available in our stylized setup, cf. [5, Section A]), 1, or 2 for the inner EC^{ccp} computations, which are then aggregated as explained above. However, for GPU optimization reasons developed in [1, Appendices A and B], we do not rely on the randomized version (given by the right-hand side formulation) of the risk margin in (4.4), i.e. we do not use the unbiased estimator (4.5), resorting instead on a Riemann sum approximation of step six months of the time integral that is visible in the left-hand side in (4.4).

Depending on the algorithm that is used, we can identify further $\beta(\bar{Z}_{[s,u]}) = e^{-r(u-s)}$ and:

- In the case of Algorithm 0:

$$\phi(\bar{Z}[t, t+1]) \leftarrow L_{t+1}^t, \qquad \psi \leftarrow 0;$$

- In the case of Algorithms 1 or 2 :

$$\phi(\bar{Z}[s, u]) = \psi(\bar{Z}_{[s,u]}) \leftarrow \beta_s^{-1} \sum_{s < \tau_i^\delta \leq u} \left(\beta_{\tau_i^\delta}(\mathrm{MtM}_{\tau_i^\delta}^i + \Delta_{\tau_i^\delta}^i) - \beta_{\tau_i}(\mathrm{MtM}_{\tau_i}^i + \mathrm{IM}_{\tau_i}^i) \right)^+. \tag{4.11}$$

With respect to the general setup of previous sections, the methodological assumptions, such as the ones on the sequences γ_k of the SA parameters or the requirement made in H9 of using a regression sample independent from the rest of the simulation in the context of Algorithm 2, can always be met at implementation stage.

Regarding now the abstract assumptions there, we only make a general comment that they should all hold in our lognormal model for X combined with randomized sampling at the times of defaults of the counterparties, which are all times with an intensity, recalling the corresponding modeling assumptions related to Algorithms 0 (SA scheme for the basic case without liabilities), 1 (SA scheme with nested simulation of future liabilities) and Algorithm 2 (SA scheme with regression of future liabilities), respectively:

- H2 a) [continuous cdf of the loss L], H2 b) [second moment of L],
- H3 [density and second moment of the loss] (the density part should follow from Malliavin calculus considerations), H4 [moments of order ≥ 2 of the (present and future) liabilities in L], H6 [concentration inequalities related to (and implying exponential moments of) the present and future liabilities in L], H7 [second moment of L, moments of order > 2 of the (present and future) liabilities in L];
- H2 a) again, H9 [square integrability of future loss components], H10 [continuous cdf of the loss where one replaces the future liability function Ψ by an arbitrary regression basis function $g \in \widehat{\mathcal{H}}$], H11 [bounded density of L, positive and continuous on an interval (specified further in H12) around $\xi_\star$].

Regarding the regression algorithm for CVA$_{t+1}^{ccp}$ that is required in the context of Algorithm 2, we apply to CVA$_{t+1}^{ccp} = \mathbb{E}_{t+1}\psi(\bar{Z}_{[t+1,T]})$ the approach that is used for computing the "CA process" in Section 4.4 of [1], using as a regression basis 1, X_{t+1}, X_{t+1}^2 (recall X is the underlying Black–Scholes rate) and the default indicator processes at time $(t+1)$ of the clearing members. In the present case of CVA$_{t+1}^{ccp}$ the situation is in fact a bit simpler as no time-stepping is required, i.e. we just need one regression for each time $(t+1)$ that occurs via the discretization times t of the integral visible in (4.4), because CVA$_{t+1}^{ccp}$ is a conditional expectation, as opposed to the above-mentioned CA process, which only solves a semi-linear BSDE.

For an SA scheme launched at time t of the outer KVA simulation, we use $\gamma_k = \frac{\gamma_0}{(100+k^{0.75})} \times \frac{(T-t)}{T}$, starting from the initial condition $\xi_0 = \chi_0 = 0$.

4.3.2. *Numerical Results*

All our simulations are run on a laptop that has an Intel i7-7700HQ CPU and a single GeForce GTX 1060 GPU programmed with the CUDA/C application programming interface (API).

Table 1 shows the time 0 (unconditional) expected shortfalls over the first year, obtained by four variants of the SA scheme and for three levels of the quantile a_{df}.

K	a_{df}	$ES^{(a)}$	$ES^{(b)}$	$ES^{(c)}$	$ES^{(d)}$
10^4	85%	311.23	248.05	253.11	259.13
	95.5%	924.72	924.72	924.72	924.72
	99%	2406.77	2406.77	2406.77	2406.77
10^5	85%	296.24	202.27	207.81	211.32
	95.5%	858.72	858.72	858.72	858.72
	99%	2347.83	2347.83	2347.83	2347.83
5×10^5	85%	287.85	200.12	206.05	209.37
	95.5%	849.12	849.12	849.12	849.12
	99%	2327.45	2327.45	2327.45	2327.45

TABLE 1.   Time 0 unconditional expected shortfalls, computed: [$ES^{(a)}$] by Algorithm 0 without the CVA terms, i.e. forgetting about the second line in (4.9), in order to assess, by comparison with the other results, the impact of these CVA terms on economical capital, depending on the confidence level a_{df}; [$ES^{(b)}$] by Algorithm 0 with the CVA terms computed by the explicit formulas that are available in the lognormal market model of this case study; [$ES^{(c)}$] by Algorithm 1 with the CVA_1 terms computed by nested Monte Carlo (and CVA_0 computed by outer Monte Carlo); [$ES^{(d)}$] by Algorithm 2 with the CVA_1 terms computed by regression against X_1 and the default indicators of the clearing members at time 1 (and CVA_0 computed by outer Monte Carlo).

In the case $a_{df} = 85\%$, Figure 1 shows the corresponding (time discretized) ES processes obtained after $K = 10^4$ and $K = 5 \times 10^5$ iterations of the SA schemes;

K	$KVA^{(a)}$	$KVA^{(b)}$	$KVA^{(c)}$	$KVA^{(d)}$
10^4	66.70	32.62	34.71	37.31
10^5	57.54	24.09	26. 85	29.11
5×10^5	54.89	23.57	25.08	28.76

TABLE 2. KVA for $a_{df} = 85\%$ and $N = 1024$ outer trajectories.

5. Conclusion and Perspectives

In this paper we propose convergent stochastic approximation estimators for the economic capital of a loss random variable L that entails a future liability (conditional expectation). The latter is estimated either by nested Monte Carlo as in Gordy and Juneja [17], or by regression as in Broadie, Du, and Moallemi [10]. Then we embed conditional versions of the above into outer risk margin (or KVA) computations.

From a practical point of view, an incremental SA scheme uses a limited amount of memory but, being a loop, is less easy to parallelize than a simulation-and-sort algorithm, on which several processors can fruitfully be used (see [1, Appendix C]). On the other hand SA schemes can be efficiently combined with importance sampling as studied in [7, 8], whereas [17] and [10] introduce respective jacknife and weighted regression acceleration procedures for the simulation-and-sort schemes.
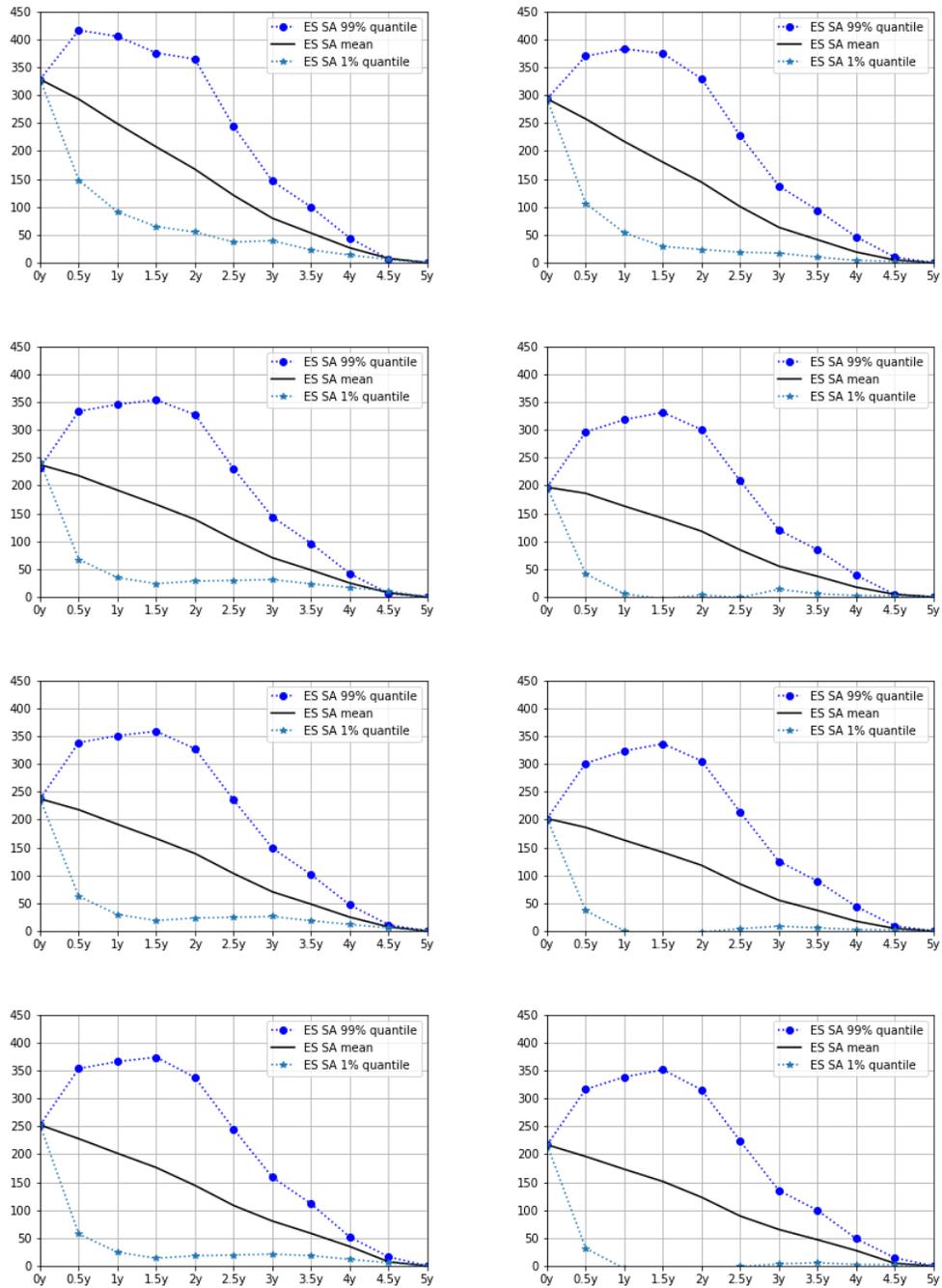
FIGURE 1. $\mathrm{ES}^{(a)}, \mathrm{ES}^{(b)}, \mathrm{ES}^{(c)}$, and $\mathrm{ES}^{(d)}$ processes (top to bottom) for $a_{df} = 85\%$ and $N = 1024$ outer trajectories. Left: $K = 10^4$. Right: $K = 5 \times 10^5$.
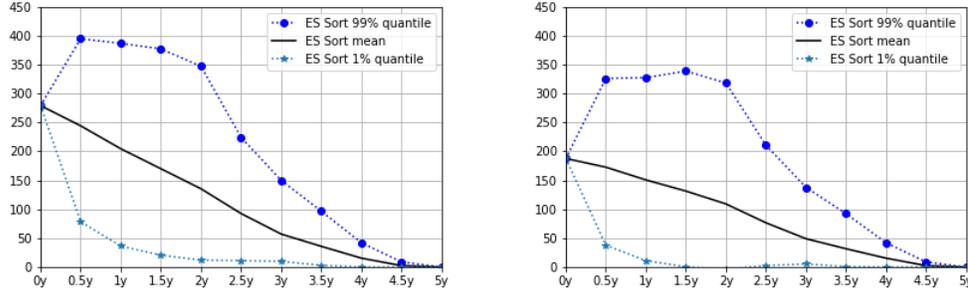
FIGURE 2. ES benchmark process obtain by tri method for $a_{df} = 85\%$, $N = 1024$ outer trajectories and $M_{ec} = 32 * 100$ inner trajectorires. Left: case without CVA terms. Right: case with CVA terms computed by the explicit formulas.

From a theoretical point of view, the stochastic approximation viewpoint leads to stronger convergence results under considerably smoother assumptions than together [17] and [10]. In particular, our assumptions (recalled in Section 4.3.1) only bear on the limiting problem, as opposed to unverifiable (not to say implausible) assumptions on the perturbed approximating problems in [17] and [10]:

- Assumptions on the density of the nested Monte Carlo surrogate of the loss in [17];
- Invertibility of the empirical covariance matrix of the regressors and an orthonormal basis of empirical regressors in [10]. By contrast, we do not even need to assume a vector space of theoretical regressors; for instance, our space of theoretical regressors could be given in the form of a neural network.

About now the results:

- [17] only shows mean square convergence, whereas we show almost sure convergence;
- [10] considers a very stylized proxy of expected shorfall in the form of $\mathbb{E}(L - \xi)^+$, for a known and fixed ξ, instead of the value-at-risk of L that needs to be estimated in the first place in a genuine expected shorfall perspective. Moreover, their study is asymptotic in the number of simulations for a fixed number of basis functions, they do not address the global convergence problem when the size of the regression basis and the number of simulations jointly go to infinity.

Last, regarding the comparison between the stochastic approximation schemes with nested versus regressed estimation of future liabilities, the assumptions that allow establishing the convergence of either approach are discussed and compared along the paper. In order to compare the fine convergence properties of each approach, it would be useful to push the computations to obtain the L_2 errors in both cases, which we leave for further research.

A. APPENDIX: TECHNICAL DEVELOPMENTS

We denote by $|x|$ the (Euclidean) norm of $x \in \mathbb{R}^d$ and by $\langle x, y \rangle$ the inner product of two vectors $x, y \in \mathbb{R}^d$. Vectors $x \in \mathbb{R}^d$ are column-vectors, and A^T denotes the transpose of a matrix A.

Some of the results are general (not specific to the setup of the main body of the paper) and therefore stated in terms of an abstract probability measure $\mathbb{Q}$, with related expectation denoted by $\mathbb{E}$.

A.1. **Two Identities**

Lemma A.1. *Let $\alpha \in (0,1)$ and μ be a probability distribution on $\mathbb{R}$ having a first order moment. If $\xi_\infty \neq \xi'_\infty$ are two solutions of* $\quad \xi : 1 - \alpha = \int_\xi^\infty \mu(\mathrm{d}x)$ *then*

$$\xi_\infty + \frac{1}{1-\alpha} \int (x - \xi_\infty)^+ \mu(\mathrm{d}x) = \xi'_\infty + \frac{1}{1-\alpha} \int (x - \xi'_\infty)^+ \mu(\mathrm{d}x).$$

Proof. We can assume that $\xi_\infty < \xi'_\infty$. Since

$$\int_{\xi_\infty}^{+\infty} \mu(\mathrm{d}x) = \int_{\xi'_\infty}^{+\infty} \mu(\mathrm{d}x) = 1 - \alpha,$$

then $\int_{\xi_\infty}^{\xi'_\infty} \mu(\mathrm{d}x) = 0$. Upon noting that

$$\xi_\infty + \frac{1}{1-\alpha} \int (x - \xi_\infty)^+ \mu(\mathrm{d}x) = \frac{1}{1-\alpha} \int_{\xi_\infty}^{+\infty} x \, \mu(\mathrm{d}x)$$

we obtain

$$\xi_\infty + \frac{1}{1-\alpha} \int (x - \xi_\infty)^+ \mu(\mathrm{d}x) = \xi'_\infty + \frac{1}{1-\alpha} \int (x - \xi'_\infty)^+ \mu(\mathrm{d}x). \qquad \blacksquare$$

Lemma A.2. *Let $\{\gamma_k, k \geq 1\}$ be a positive sequence such that $\sum_k \gamma_k = +\infty$ and $\lim_k \gamma_k = 0$. Given a $\mathbb{R}^d$-valued sequence $\{V_k, k \geq 1\}$ and $\theta_0 \in \mathbb{R}^d$, define the sequence $\{\theta_k, k \geq 0\}$ by*

$$\theta_{k+1} = (1 - \gamma_{k+1})\theta_k + \gamma_{k+1} V_{k+1}.$$

Set $S_0 = 1$ and $S_k := \prod_{j=1}^k (1 - \gamma_j)^{-1}$ for any $k \geq 1$. Then $\lim_k S_k = +\infty$ and when it exists,

$$\lim_k \theta_k = \lim_k \frac{1}{S_k} \sum_{j=1}^k S_j \gamma_j V_j.$$

Proof. The proof is adapted from [7]. Note that $S_{k+1}(1 - \gamma_{k+1}) = S_k$, hence, for all k large enough such that $\gamma_k \to 0$, $\ln S_{k+1} + \ln(1 - \gamma_{k+1}) = \ln S_k$, i.e.

$$\ln S_{k+1} - \ln S_k = - \ln(1 - \gamma_{k+1}) \geq \gamma_{k+1}.$$

Since $\sum_k \gamma_k = \infty$, then $\lim_k S_k = +\infty$.
   We write

$$S_{k+1}\theta_{k+1} = S_{k+1}(1 - \gamma_{k+1})\theta_k + S_{k+1}\gamma_{k+1} V_{k+1} = S_k \theta_k + S_{k+1}\gamma_{k+1} V_{k+1}.$$

Hence, by induction, $S_{k+1}\theta_{k+1} = S_0\theta_0 + \sum_{j=1}^{k+1} S_j \gamma_j V_j$, which implies the result and the conclusion follows from this. $\qquad \blacksquare$

A.2. **A General Convergence Result for Stochastic Approximation Algorithms**

Let $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ be a measurable function and let $\{\gamma_k, k \geq 1\}$ be a sequence of positive numbers. Let $\mathbb{R}^q$-valued random variables $\{V_k, k \geq 0\}$ and $\theta_0 \in \mathbb{R}^d$ be defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Theorem A.3

provides sufficient conditions for the almost-sure convergence and the L^p-convergence, $p \in (0,2)$, of the sequence $\{\theta_k, k \geq 0\}$ given by

$$\theta_{k+1} = \theta_k - \gamma_{k+1} \, H(\theta_k, V_{k+1}). \tag{A.1}$$

These conditions are general enough to cover the case when the r.v. $\{V_k, k \geq 1\}$ are not i.i.d. but have a distribution converging, in some sense, to the distribution of a r.v. $V_\star$.

We write

$$H(\theta_k, V_{k+1}) = h(\theta_k) + e_{k+1} + r_{k+1}, \tag{A.2}$$

where

$$h(\theta) := \mathbb{E}\left[H(\theta, V_\star)\right],$$
$$e_{k+1} := H(\theta_k, V_{k+1}) - \mathbb{E}\left[H(\theta_k, V_{k+1})|\mathcal{G}_k\right],$$
$$r_{k+1} := \mathbb{E}\left[H(\theta_k, V_{k+1})|\mathcal{G}_k\right] - h(\theta_k),$$

and where the filtration $\{\mathcal{G}_k, k \geq 1\}$ is defined by $\mathcal{G}_k := \sigma\{V_1, \cdots, V_k\}$.

Theorem A.3. *Suppose that*

(i) $\{\gamma_k, k \geq 1\}$ *is a deterministic positive sequence such that* $\sum_k \gamma_k = +\infty$ *and there exists* $\kappa \in (0,1]$ *such that* $\sum_{k \geq 1} \gamma_k^{1+\kappa} < \infty$,

(ii) $H : \mathbb{R}^d \times \mathbb{R}^q \to \mathbb{R}^d$ *is measurable and* $h : \mathbb{R}^d \to \mathbb{R}^d$ *is continuous,*

(iii) *the set* $\mathcal{L} := \{h = 0\}$ *is a non-empty compact subset of* $\mathbb{R}^d$ *and for any* $\theta^* \in \mathcal{L}$ *and* $\theta \notin \mathcal{L}$, *we have* $\langle \theta - \theta^*, h(\theta) \rangle > 0$.

Let $\{\theta_k, k \geq 0\}$ *be given by* (A.1) *where the r.v.* $\{V_k, k \geq 0\}$ *satifsy*

(iv) $\sum_{k \geq 1} \gamma_k^{1-\kappa}|r_k|^2 < +\infty$ $\mathbb{P}$-*a.s.*

(v) *There exist non-negative constants* $C_{H,1}, C_{H,2}$ *such that, for any* $\theta \in \mathbb{R}^d$,

$$\sup_{k \geq 1} \mathbb{E}\left[|H|^2(\theta, V_k)\right] \leq C_{H,1} + C_{H,2}|\theta|^2.$$

Then there exists a $\mathcal{L}$-*valued random variable* θ_∞ *such that* $\mathbb{P}(\lim_k \theta_k = \theta_\infty) = 1$. *If, in addition,*

(vi) $\sum_{k \geq 1} \gamma_k^{1-\kappa}\mathbb{E}[|r_k|^2] < +\infty$,

then $\sup_{k \geq 0} \mathbb{E}[|\theta_k - \theta_\infty|^2] < +\infty$ *and for any* $p \in (0,2)$, $\lim_k \mathbb{E}\left[|\theta_k - \theta_\infty|^p\right] = 0$.

Proof. **Step 1. Almost-sure boundedness and convergence.** Let $\theta^* \in \mathcal{L}$. We have, by (A.2),

$$\begin{aligned}
|\theta_{k+1} - \theta^*|^2 &= |\theta_k - \theta^* - \gamma_{k+1}\left(h(\theta_k) + e_{k+1} + r_{k+1}\right)|^2 \\
&= |\theta_k - \theta^*|^2 - 2\gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle \\
&\quad - 2\gamma_{k+1}\langle \theta_k - \theta^*, e_{k+1}\rangle - 2\gamma_{k+1}\langle \theta_k - \theta^*, r_{k+1}\rangle + \gamma_{k+1}^2|H|^2(\theta_k, V_{k+1}).
\end{aligned}$$

Since $\{e_k, k \geq 1\}$ is a martingale-increment w.r.t. the filtration $\{\mathcal{G}_k, k \geq 1\}$ and θ_k is $\mathcal{G}_k$-measurable, we have for any k,

$$\begin{aligned}
\mathbb{E}\left[|\theta_{k+1} - \theta^*|^2|\mathcal{G}_k\right] \leq |\theta_k - \theta^*|^2 &- 2\gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle + \gamma_{k+1}^{1+\kappa}|\theta_k - \theta^*|^2 + \gamma_{k+1}^{1-\kappa}|r_{k+1}|^2 \\
&+ \gamma_{k+1}^2 C_{H,1} + \gamma_{k+1}^2 C_{H,2}|\theta_k|^2,
\end{aligned}$$

where we used the inequality $-2\gamma\langle a, b\rangle \leq \gamma^{1+\kappa}|a|^2 + \gamma^{1-\kappa}|b|^2$, the equality $\mathbb{E}[r_{k+1}|\mathcal{G}_k] = r_{k+1}$ and the assumption (v). Hence, by using $|\theta_k|^2 \leq 2|\theta_k - \theta^*|^2 + 2|\theta^*|^2$,

$$\begin{aligned}
\mathbb{E}\left[|\theta_{k+1} - \theta^*|^2|\mathcal{G}_k\right] \leq \left(1 + 2\gamma_{k+1}^2 C_{H,2} + \gamma_{k+1}^{1+\kappa}\right)|\theta_k - \theta^*|^2 & \\
- 2\gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle + \gamma_{k+1}^{1-\kappa}|r_{k+1}|^2 + \gamma_{k+1}^2 C', &
\end{aligned} \tag{A.3}$$

where $C' := C_{H,1} + 2C_{H,2}|\theta^*|^2$. From the assumptions (i), (iii), and (iv), we have that, $\mathbb{P}$-a.s.,

$$\forall k \geq 0 \quad \gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle \geq 0, \qquad \sum_{k\geq 0}(\gamma_{k+1}^{1-\kappa}|r_{k+1}|^2 + \gamma_{k+1}^2 + \gamma_{k+1}^{1+\kappa}) < +\infty.$$

By the Robbins-Siegmund lemma (see [23]), $\mathbb{P}$-a.s. (for an almost-sure set depending upon θ^*)

$$\lim_k |\theta_k - \theta^*| \text{ exists}, \qquad \sum_{k\geq 0}\gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle < +\infty.$$

Since $\mathcal{L}$ is bounded and $\theta^* \in \mathcal{L}$, this implies that the sequence $\{\theta_k, k \geq 0\}$ is bounded with probability one. Using the separability of $\mathbb{R}^d$ and since $\theta' \mapsto \lim_k |\theta_k - \theta'|$ is continuous, we have $\mathbb{P}$-a.s.:

$$\forall \theta' \in \mathcal{L}, \quad \text{the limit } \lim_k |\theta_k - \theta'| \text{ exists}. \tag{A.4}$$

Set $\varsigma := \liminf_{k\to+\infty}\langle \theta_k - \theta^*, h(\theta_k)\rangle$. By (iii), $\varsigma \geq 0$. As $\sum_{k\geq 0}\gamma_k = +\infty$, we have

$$\{\varsigma > 0\} \subseteq \sum_{k\geq 0}\gamma_{k+1}\langle \theta_k - \theta^*, h(\theta_k)\rangle = +\infty;$$

the probability of the second event is zero. Hence $\mathbb{P}(\varsigma = 0) = 1$.

Therefore, with probability one, there exists a subsequence $\{n_k, k \geq 1\}$ such that $\lim_k\langle \theta_{n_k} - \theta^*, h(\theta_{n_k})\rangle = 0$. Since the sequence $\{\theta_{n_k}, k \geq 1\}$ is bounded a.s., we can still assume (up to extraction of another subsequence) that $\{\theta_{n_k}, k \geq 1\}$ converges to some limit θ_∞. By assumption (ii), we have $\langle \theta_\infty - \theta^*, h(\theta_\infty)\rangle = 0$, and by assumption (iii), this implies that $\theta_\infty \in \mathcal{L}$. But using (A.4) we get $\lim_k |\theta_k - \theta_\infty| = \lim_k |\theta_{n_k} - \theta_\infty| = 0$. This implies that $\lim_k \theta_k = \theta_\infty$.

Step2. Uniform boundedness in L^2. Let a (deterministic) point $\theta^* \in \mathcal{L}$ be given. By taking expectation in (A.3), we have

$$\mathbb{E}\left[|\theta_{k+1} - \theta^*|^2\right] \leq \left(1 + 2\gamma_{k+1}^2 C_{H,2} + \gamma_{k+1}^{1+\kappa}\right)\mathbb{E}\left[|\theta_k - \theta^*|^2\right] + \gamma_{k+1}^{1-\kappa}\mathbb{E}[|r_{k+1}|^2] + C'\gamma_{k+1}^2.$$

Applying again the Robbins-Siegmund lemma with the assumptions (i) and (vi), we deduce that the sequence $\lim_k \mathbb{E}\left[|\theta_k - \theta^*|^2\right]$ exists and thus $\sup_k \mathbb{E}\left[|\theta_k|^2\right] < \infty$ since $\mathcal{L}$ is bounded. This implies $\sup_k \mathbb{E}\left[|\theta_k - \theta_\infty|^2\right] < +\infty$ for any $\mathcal{L}$-valued random variable θ_∞, using again that $\mathcal{L}$ is bounded.

Step 3. Convergence in L^p. Let $C > 0$ and $p \in (0,2)$. We write

$$\mathbb{E}\left[|\theta_k - \theta_\infty|^p\right] = \mathbb{E}\left[|\theta_k - \theta_\infty|^p \; \mathbb{1}_{|\theta_k-\theta_\infty|<C}\right] + \mathbb{E}\left[|\theta_k - \theta_\infty|^p \; \mathbb{1}_{|\theta_k-\theta_\infty|\geq C}\right].$$

The first term converges to zero by the dominated convergence theorem. For the second term, Hölder's and Markov's inequalities give that

$$\mathbb{E}[|\theta_k - \theta_\infty|^p\mathbb{1}_{|\theta_k-\theta_\infty|\geq C}] \leq \frac{\mathbb{E}[|\theta_k - \theta_\infty|^2]}{C^{2-p}} \leq \frac{\sup_{l\geq 0}\mathbb{E}[|\theta_l - \theta_\infty|^2]}{C^{2-p}};$$

which is lower than $\epsilon > 0$ for some C large enough. This holds true for any ϵ, thus concluding the proof. ∎

A.3. **Proof of Theorem 1.2**

For the study of the sequence $\{\xi_k, k \geq 1\}$, we check the assumptions of Theorem A.3, applied to $\theta_k \leftarrow \xi_k$ (so that $d = 1$), $\mathbb{P} \leftarrow \mathbb{P}_0$, $V_k \leftarrow L_k$, the distribution of $V_\star$ is $Q(z, \cdot)$ and

$$H(\theta, V) \leftarrow H_1(\xi, L) = 1 - \frac{1}{1-\alpha}\mathbb{1}_{L>\xi}$$

(cf. (1.3)). By H1, the condition (i) is satisfied. In addition, by H2a, the function

$$\xi \mapsto h(\xi) := 1 - \frac{1}{1-\alpha} \int_\xi^{+\infty} Q(z, \mathrm{d}x)$$

is continuous on $\mathbb{R}$, so that (ii) holds. The set $\mathcal{L} := \{h = 0\}$ is the set of the points $\xi_\star$ satisfying $\mathbb{P}_0(L > \xi_\star) = 1 - \alpha$: under H2a, this set is non empty and compact and for any $\xi < \xi_\star$ (resp. $\xi > \xi_\star$) such that $h(\xi) \neq 0$, $h(\xi) < 0$ (resp. $h(\xi) > 0$). Hence, (iii) is satisfied. In this algorithm, we have $r_k = 0$ since $h(\cdot) = \mathbb{E}_0[H(\cdot, L)]$ and L_{k+1} is independent of ξ_k. Hence, (iv) and (vi) hold. Finally, $\sup_{\xi \in R, L \in \mathbb{R}} |H(\xi, L)|^2 \leq C_{H,1} := (1-\alpha)^{-2}$ so that (v) holds with $C_{H,2} = 0$.

For the results on the sequence $\{\chi_k, k \geq 1\}$, we check the assumptions of Lemma A.2 with $\theta_k \leftarrow \chi_k$ (so that $d = 1$) and $V_{k+1} \leftarrow \xi_k + (1-\alpha)^{-1}(L_{k+1} - \xi_k)^+$. We write

$$V_{k+1} = \xi_k + (1-\alpha)^{-1} \int (x - \xi_k)^+ Q(z, \mathrm{d}x) + \tilde{e}_{k+1}, \qquad (A.5)$$

$$\tilde{e}_{k+1} := (1-\alpha)^{-1} \left((L_{k+1} - \xi_k)^+ - \int (x - \xi_k)^+ Q(z, \mathrm{d}x) \right).$$

Set $S_0 := 1$ and $S_k := \prod_{j=1}^k (1 - \gamma_j)^{-1}$ for any $k \geq 1$, so that $S_k(1 - \gamma_k) = S_{k-1}$ and $S_k - S_{k-1} = \gamma_k S_k$. By H1 and Lemma A.2, $\lim_k S_k = +\infty$ so that from the above almost-sure convergence on $\{\xi_k, k \geq 1\}$ and from the Cesaro lemma,

$$\lim_{k \to \infty} \frac{1}{S_k} \sum_{j=1}^k S_j \gamma_j \xi_j = \lim_{k \to \infty} \frac{1}{S_k} \sum_{j=1}^k (S_j - S_{j-1}) \xi_j = \xi_\infty \qquad \mathbb{P}_0 - \text{a.s.}$$

By H2a, the second term in the RHS of (A.5) is a continuous function of ξ_k. Therefore, by similar arguments,

$$\lim_{k \to \infty} \frac{1}{S_k} \sum_{j=1}^k S_j \gamma_j \int (x - \xi_j)^+ Q(z, \mathrm{d}x) = \int (x - \xi_\infty)^+ Q(z, \mathrm{d}x) \qquad \mathbb{P}_0 - \text{a.s.}$$

Finally, $\{\tilde{e}_k, k \geq 1\}$ is a $\mathcal{G}_k$-martingale increment; by using $|(a-c)^+ - (b-c)^+| \leq |a| + |b|$ and $(|a| + |b|)^2 \leq 2a^2 + 2b^2$, and since $\{L_k, k \geq 1\}$ are i.i.d. with distribution $Q(z, \cdot)$, we have

$$\mathbb{E}_0 \left[\left| (L_{j+1} - \xi_j)^+ - \int (x - \xi_j)^+ Q(z, \mathrm{d}x) \right|^2 \right] \leq 2 \int x^2 Q(z, \mathrm{d}x) + 2 \left(\int |x| Q(z, \mathrm{d}x) \right)^2;$$

by H2b, the RHS is finite. Therefore, H1, [19, Corollary 2.2.] and the Kronecker Lemma (see e.g. [19, Section 2.6, page 31]) imply that

$$\lim_{k \to +\infty} \frac{1}{S_k} \sum_{j=1}^k S_j \gamma_j \left((L_{j+1} - \xi_j)^+ - \int (x - \xi_j)^+ Q(z, \mathrm{d}x) \right) = 0, \qquad \mathbb{P}_0 - \text{a.s.}$$

By Lemma A.2, we obtain that $\mathbb{P}_0$-a.s., $\lim_k \chi_k$ exists and solves

$$\lim_k \chi_k = \xi_\infty + \frac{1}{1-\alpha} \int (x - \xi_\infty)^+ Q(z, \mathrm{d}x).$$

∎

A.4. **Proofs of the Results of Section 2.1**

We start the proof with two preliminary lemmas.

Lemma A.4. *Let $\{\varphi_m, m \geq 1\}$ be $\mathbb{R}$-valued random variable defined on $(\Omega, \mathcal{A}, \mathbb{P})$, i.i.d. with distribution μ. Assume that μ has a finite moment of order $p_\star > 1$ and set $C_{p_\star} := \int |w - \int w\mu(\mathrm{d}w)|^{p_\star}\,\mu(\mathrm{d}w)$. Then for any $M \geq 1$,*

$$\mathbb{E}\left[\left|\frac{1}{M}\sum_{m=1}^{M}\varphi_m - \int w\,\mu(\mathrm{d}w)\right|^{p_\star}\right] \leq \frac{c_{p_\star}\,C_{p_\star}}{M^{(p_\star/2)\wedge(p_\star-1)}},$$

where $c_p := (18p\sqrt{q})^p$ and q is the Hölder conjugate of p.

Proof. Set $\mu[1] := \int w\mu(\mathrm{d}w)$. The Burkholder inequality (see e.g. [19, Theorem 2.10]) applied to the sequence $\frac{1}{M}(\varphi^m - \mu[1]), m = 1, \ldots, M$, yields

$$\mathbb{E}\left[\left|\frac{1}{M}\sum_{m=1}^{M}\varphi_m - \mu[1]\right|^{p_\star}\right] \leq \frac{c_{p_\star}}{M^{p_\star}}\mathbb{E}\left[\left|\sum_{m=1}^{M}(\varphi_m - \mu[1])^2\right|^{p_\star/2}\right].$$

If $p_\star \geq 2$, we obtain by the Minkowsky inequality,

$$\mathbb{E}\left[\left|\sum_{m=1}^{M}(\varphi_m - \mu[1])^2\right|^{p_\star/2}\right] \leq \left(\sum_{m=1}^{M}\left(\mathbb{E}\left[|\varphi_m - \mu[1]|^{p_\star}\right]\right)^{2/p_\star}\right)^{p_\star/2} \leq C_{p_\star}\,M^{p_\star/2}.$$

If $p_\star < 2$, by using $(x+y)^{p_\star/2} \leq x^{p_\star/2} + y^{p_\star/2}$ for any $x, y \geq 0$, we have

$$\mathbb{E}\left[\left|\sum_{m=1}^{M}(\varphi_m - \mu[1])^2\right|^{p_\star/2}\right] \leq \mathbb{E}\left[\sum_{m=1}^{M}|\varphi_m - \mu[1]|^{p_\star}\right] = C_{p_\star}\,M.$$

This concludes the proof.   ∎

Lemma A.5. *Let $V, V', \{\varphi_m, m \geq 1\}$ and $\{\varphi'_m, m \geq 1\}$ be $\mathbb{R}$-valued random variables defined on $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{B} \subset \mathcal{A}$ be a σ-field on Ω such that conditionally on $\mathcal{B}$, $\{\varphi_m, m \geq 1\}$ are i.i.d. and $\{\varphi'_m, m \geq 1\})$ are i.i.d. Assume*

(i) *the distribution of V admits a density with respect to the Lebesgue measure on $\mathbb{R}$, which is upper bounded by $C_0 > 0$;*

(ii) *there exists $p_\star > 1$ such that φ_1 and φ'_1 have finite $p_\star$-moments; set*

$$C_{p_\star} := \mathbb{E}\left[|\varphi_1 - \mathbb{E}[\varphi_1 \mid \mathcal{B}]|^{p_\star}\right] \vee \mathbb{E}\left[|\varphi'_1 - \mathbb{E}[\varphi'_1 \mid \mathcal{B}]|^{p_\star}\right].$$

(iii) *there exists a constant $C > 0$ such that*

$$|V - V'| \leq \frac{C}{M}\left|\sum_{m=1}^{M}(\varphi_m - \mathbb{E}[\varphi_1|\mathcal{B}])\right| + \frac{1}{M'}\left|\sum_{m=1}^{M'}(\varphi'_m - \mathbb{E}[\varphi'_1|\mathcal{B}])\right|.$$

Then for any positive integer M,

$$\sup_{\xi \in \mathbb{R}} \mathbb{E}\left[|\mathbb{1}_{V > \xi} - \mathbb{1}_{V' > \xi}|\right] \leq 2^{p_\star}(1 \vee C)^{p_\star}(C_0 + c_{p_\star}\,C_{p_\star})(M \wedge M')^{-\frac{(p_\star/2)\wedge(p_\star-1)}{1+p_\star}}, \tag{A.6}$$

where c_p only depends on p (see its definition in Lemma A.4). If, in addition,

(iv) there exists $C_\infty > 0$ such that for any $\delta > 0$ and any positive integer M,

$$\mathbb{P}\left(\left|\frac{1}{M}\sum_{m=1}^{M}\varphi_m - \mathbb{E}[\varphi_1|\mathcal{B}]\right| \geq \delta\right) \vee \mathbb{P}\left(\left|\frac{1}{M}\sum_{m=1}^{M}\varphi'_m - \mathbb{E}[\varphi'_1|\mathcal{B}]\right| \geq \delta\right) \leq e^{-C_\infty M\delta^2},$$

then, for any integers $M, M' \geq 3$,

$$\sup_{\xi \in \mathbb{R}} \mathbb{E}\left[|\mathbb{1}_{V>\xi} - \mathbb{1}_{V'>\xi}|\right] \leq 2\left(1 + \frac{C_0}{\sqrt{2\,C_\infty}}\right)\left(\frac{\log(M \wedge M')}{(M \wedge M')}\right)^{1/2}.$$

Proof. We have $|\mathbb{1}_{V>\xi} - \mathbb{1}_{V'>\xi}| = \mathbb{1}_{V>\xi\geq V'} + \mathbb{1}_{V'>\xi\geq V}$. Let $\delta > 0$. On the set $\{|V - V'| < \delta\}$, it holds

$$\{V > \xi \geq V'\} \subset \{V > \xi > V - \delta\}, \qquad \{V' > \xi \geq V\} \subset \{V + \delta > \xi \geq V\},$$

so that

$$\mathbb{E}\left[|\mathbb{1}_{V>\xi} - \mathbb{1}_{V'>\xi}|\mathbb{1}_{|V-V'|<\delta}\right] \leq \mathbb{E}\left[\mathbb{1}_{|V-\xi|<\delta}\mathbb{1}_{|V-V'|<\delta}\right] \leq \mathbb{P}\left(|V - \xi| < \delta\right).$$

By using (i), it holds

$$\mathbb{E}[|\mathbb{1}_{V>\xi} - \mathbb{1}_{V'>\xi}|] \leq \mathbb{P}[|V - \xi| < \delta] + \mathbb{P}\left(|V - V'| \geq \delta\right) \leq 2C_0\delta + \mathbb{P}\left(|V - V'| \geq \delta\right). \tag{A.7}$$

By (ii), (iii) and Lemma A.4, and by using $(x + y)^{p_\star} \leq 2^{p_\star-1}(x^{p_\star} + y^{p_\star})$ for any $x, y \geq 0$,

$$\mathbb{E}\left[|V - V'|^{p_\star}\right] = \mathbb{E}\left[\mathbb{E}\left[|V - V'|^{p_\star} \mid \mathcal{B}\right]\right] \leq \frac{2^{p_\star}\,(1 \vee C)^{p_\star}\,c_{p_\star}\,C_{p_\star}}{(M \wedge M')^{(p_\star/2)\wedge(p_\star-1)}}.$$

The Chebyshev inequality implies

$$\mathbb{P}(|V - V'| \geq \delta) \leq \frac{2^{p_\star}\,(1 \vee C)^{p_\star}\,c_{p_\star}\,C_{p_\star}}{\delta^{p_\star}\,(M \wedge M')^{(p_\star/2)\wedge(p_\star-1)}}. \tag{A.8}$$

We then obtain (A.6) from (A.7) and (A.8) applied to $\delta \leftarrow (M \wedge M')^{-\tilde{p}_\star/(1+p_\star)}$ with $\tilde{p}_\star := (p_\star/2) \wedge (p_\star - 1)$.

Under (iv), the second term in (A.7) is upper bounded by

$$\exp\left(-C_\infty M \delta^2/(4C^2)\right) + \exp\left(-C_\infty M' \delta^2/4\right) \leq 2\exp\left(-\frac{C_\infty}{4(1 \vee C)^2}\delta^2(M \wedge M')\right).$$

Hence we have, by setting $\widetilde{C}_\infty := C_\infty/(4(1 \vee C)^2)$ and $\overline{M} := M \wedge M'$,

$$\mathbb{E}[|\mathbb{1}_{V>\xi} - \mathbb{1}_{V'>\xi}|] \leq 2C_0\delta + 2e^{-\widetilde{C}_\infty \delta^2\overline{M}} \leq \sqrt{2}\frac{C_0}{\sqrt{\widetilde{C}_\infty}}\sqrt{\frac{\ln \overline{M}}{\overline{M}}} + \frac{2}{\sqrt{\overline{M}}},$$

where the last inequality is obtained by choosing $\delta \leftarrow \sqrt{(\log \overline{M})/(2\widetilde{C}_\infty\overline{M})}$. This concludes the proof since $\sqrt{\ln \overline{M}} \geq 1$ for $\overline{M} \geq 3$. ■

Proof of Lemma 2.1. We apply Lemma A.5 with $\mathbb{P} \leftarrow \mathbb{P}_0$, $\mathcal{B} \leftarrow \sigma(Z_0, Z_1)$, $C_i \leftarrow C_i(z)$ for $i \in \{0, p_\star, \infty\}$, $C \leftarrow c_\beta$ and $p_\star \geq 2$. This yields the inequalities (2.3) and (2.5). Since $|a^+ - b^+| \leq |a - b|$ and $p_\star \geq 1$, we have

$$\mathbb{E}_0 \left[|(L - \xi)^+ - (L' - \xi)^+| \right] \leq \left(\mathbb{E}_0 \left[|(L - \xi)^+ - (L' - \xi)^+|^{p_\star}| \right] \right)^{1/p_\star} \leq \left(\mathbb{E}_0 \left[|L - L'|^{p_\star}| \right] \right)^{1/p_\star}.$$

We conclude the proof of (2.4) by Lemma A.4.   ∎

Proof of Theorem 2.2. As in the proof of Theorem 1.2, we first establish the results on the sequence $\{\xi_k, k \geq 0\}$ by application of Theorem A.3. We then prove the results on the sequence $\{\chi_k, k \geq 0\}$ by application of Lemma A.2.

We check the assumptions of Theorem A.3 with $\theta_k \leftarrow \xi_k$ (so that $d = 1$), $\mathbb{P} \leftarrow \mathbb{P}_0$, $V_k \leftarrow L_k$, the distribution of $V_\star$ is $Q(z, \cdot)$, $\mathcal{G}_k \leftarrow \sigma(L_j, j \leq k)$ and

$$H(\theta, V) \leftarrow 1 - \frac{1}{1 - \alpha} \mathbb{1}_{V > \theta}, \qquad h(\theta) \leftarrow 1 - \frac{1}{1 - \alpha} \mathbb{P}_0(L > \theta).$$

Under H5 and H3, the conditions (i), (ii), (iii) and (v) hold; the proof is on the same lines as in the proof of Theorem 1.2 and is omitted. We establish the condition (vi) (which also implies the condition (iv)) with

$$r_{k+1} \leftarrow \mathbb{E}_0 \left[H(\xi_k, L_{k+1}) | \mathcal{G}_k \right] - \mathbb{E}_0 \left[H(\xi_k, \tilde{L}_{k+1}) | \mathcal{G}_k \right]$$

where

$$\tilde{L}_{k+1} := \phi^{k+1} + \beta^{k+1} \Psi(Z_1^{k+1}) - \Psi'(z).$$

Note that since the r.v. $(\phi^{k+1}, \beta^{k+1}, Z_1^{k+1})$ are independent from $\mathcal{G}_k$ and, conditionnally to $\mathcal{G}_k$, have the same distribution as the processes (ϕ, β, Z_1), then the distribution of $\tilde{L}_{k+1}$ given $\mathcal{G}_k$ is $Q(z, \cdot)$. Hence,

$$\mathbb{E}_0 \left[H(\xi_k, \tilde{L}_{k+1}) | \mathcal{G}_k \right] = h(\xi_k). \tag{A.9}$$

We write

$$r_{k+1} = \mathbb{E}_0 \left[H(\xi_k, L_{k+1}) - H(\xi_k, \tilde{L}_{k+1}) \Big| \mathcal{G}_k \right] = \frac{1}{1 - \alpha} \mathbb{E}_0 \left[\left(\mathbb{1}_{\tilde{L}_{k+1} > \xi_k} - \mathbb{1}_{L_{k+1} > \xi_k} \right) \Big| \mathcal{G}_k \right]$$

so that, by Lemma 2.1, there exists a constant c such that for any $k \geq 1$, $\mathbb{P}_0$-a.s.

$$|r_{k+1}| \leq c \left(M_{k+1} \wedge M'_{k+1} \right)^{-p_\star/(2(1+p_\star))};$$

this implies, by H5, $\sum_{k \geq 1} \gamma_k^{1-\kappa} \mathbb{E}_0[|r_k|^2] < +\infty$, thus proving (vi). This concludes the proof of the results on the sequence $\{\xi_k, k \geq 1\}$.

For the results on the sequence $\{\chi_k, k \geq 1\}$, we check the assumptions of Lemma A.2 with $\theta_k \leftarrow \chi_k$ (so that $d = 1$) and $U_{k+1} \leftarrow \xi_k + (1 - \alpha)^{-1} (L_{k+1} - \xi_k)^+$. We write

$$U_{k+1} = \xi_k + (1 - \alpha)^{-1} \int (x - \xi_k)^+ Q(z, \mathrm{d}x) + \tilde{e}_{k+1} + \tilde{r}_{k+1},$$

$$\tilde{r}_{k+1} := (1 - \alpha)^{-1} \left((L_{k+1} - \xi_k)^+ - (L'_{k+1} - \xi_k)^+ \right),$$

$$\tilde{e}_{k+1} := (1 - \alpha)^{-1} \left((L'_{k+1} - \xi_k)^+ - \int (x - \xi_k)^+ Q(z, \mathrm{d}x) \right).$$

As in the proof of Theorem 1.2, we prove by using H5, H3, that

$$\lim_{k\to\infty} \frac{1}{S_k} \sum_{j=1}^{k} S_j \gamma_j \xi_j = \xi_\infty \qquad \mathbb{P}_0 - \text{a.s.},$$

$$\lim_{k\to\infty} \frac{1}{S_k} \sum_{j=1}^{k} S_j \gamma_j \int (x - \xi_j)^+ Q(z, \mathrm{d}x) = \int (x - \xi_\infty)^+ Q(z, \mathrm{d}x) \qquad \mathbb{P}_0 - \text{a.s.}$$

$$\lim_{k\to+\infty} \frac{1}{S_k} \sum_{j=1}^{k} S_j \gamma_j \tilde{e}_{j+1} = 0, \qquad \mathbb{P}_0 - \text{a.s.}$$

By Lemma 2.1 and H5, $\sum_k \gamma_k \mathbb{E}_0 \left[\|\tilde{r}_{j+1}\| \right] < \infty$ so that by the Kronecker Lemma,

$$\lim_{k\to+\infty} \frac{1}{S_k} \sum_{j=1}^{k} S_j \gamma_j \tilde{r}_{j+1} = 0, \qquad \mathbb{P}_0 - \text{a.s.}$$

By Lemma A.2, we obtain that $\mathbb{P}_0$-a.s., $\lim_k \chi_k$ exists and solves

$$\lim_k \chi_k = \xi_\infty + \frac{1}{1 - \alpha} \int (x - \xi_\infty)^+ Q(z, \mathrm{d}x).$$

∎

A.5. **A Central Limit Theorem for Stochastic Approximation Algorithms**

We recall in this section sufficient conditions for a central limit theorem (CLT) to hold for random variables $\{\theta_k, k \geq 0\}$ defined through a stochastic approximation algorithm: given a deterministic sequence $\{\gamma_k, k \geq 1\}$, a function $h : \mathbb{R}^d \to \mathbb{R}$, $\theta_0 \in \mathbb{R}^d$ and $\mathbb{R}^d$-valued random variables $\{e_k, k \geq 1\}$ and $\{r_k, k \geq 1\}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$, define for $k \geq 0$,

$$\theta_{k+1} = \theta_k + h(\theta_k) + \gamma_{k+1} e_{k+1} + \gamma_{k+1} r_{k+1}. \tag{A.10}$$

Theorem A.6 corresponds to [16, Theorem 2.1.]. It provides sufficient conditions for a CLT along a converging sequence $\{\lim_q \theta_q = \theta_\star\}$ where $\theta_\star \in \mathbb{R}^d$ is fixed (deterministic). On the mean field h and the limit point $\theta_\star$, it is assumed

C1. *a) The mean field $h : \mathbb{R}^d \to \mathbb{R}^d$ is measurable and twice continuously differentiable in a neighborhood of $\theta_\star$, where $h(\theta_\star) = 0$.*
b) The gradient $\nabla h(\theta_\star)$ is a Hurwitz matrix. Denote by $-\ell$, $(\ell > 0)$, the largest real part of its eigenvalues.

The sequence $\{e_k, k \geq 1\}$ satisfies

C2. *a) $\{e_k, k \geq 1\}$ is a $\mathcal{G}_k$-adapted $\mathbb{P}$-martingale increment sequence: $\mathbb{E}[e_k | \mathcal{G}_{k-1}] = 0$ $\mathbb{Q}$-a.s. for any $k \geq 1$.*
b) For some $C > 0$, there exists $\tau > 0$, such that

$$\sup_k \mathbb{E}\left[|e_{k+1}|^{2+\tau} \mathbb{1}_{|\theta_k - \theta_\star| \leq C} \right] < \infty.$$

c) There exists a symmetric positive definite matrix $D_\star$ and a sequence $\{D_k, k \geq 1\}$ of $\mathbb{R}^d$-valued random variables, such that $\mathbb{P}$-a.s.

$$\mathbb{E}\left[e_{k+1} e_{k+1}^T | \mathcal{G}_k \right] = D_\star + D_k, \qquad \lim_k D_k \mathbb{1}_{\lim_q \theta_q = \theta_\star} = 0.$$

The sequences $\{r_k, k \geq 1\}$ and $\{\gamma_k, k \geq 1\}$ satisfy

C3. r_k is $\mathcal{G}_k$-adapted and $\gamma_k^{-1/2}|r_k|\mathbb{1}_{\lim_q \theta_q = \theta_\star} \leq X_k Y_k$ where $\mathbb{P}(\sup_k |X_k| < \infty) = 1$ and and $\lim_k \mathbb{E}[|Y_k|] = 0$.

C4. One of the following condition holds

a) $\gamma_k \sim \gamma_\star/k$ and $\gamma_\star > 1/(2\ell)$.

b) $\gamma_k \sim \gamma_\star/k^c$ where $c \in (1/2, 1)$.

Theorem A.6. *[16, Theorem 2.1.]* Let $\{\theta_k, k \geq 1\}$ be the sequence given by (A.10) for some $\theta_0 \in \mathbb{R}^d$. Assume C1, C2, C3 and C4. Let Γ be the positive definite matrix satisfying $\mathbb{Q}$-a.s. on the set $\{\lim_q \theta_q = \theta_\star\}$

$$\Gamma(\mathrm{Id} + 2\gamma_\star \nabla h(\theta_\star)^T) + (\mathrm{Id} + 2\gamma_\star \nabla h(\theta_\star))\Gamma = -2\gamma_\star D_\star \qquad \text{under C4a,}$$

$$\Gamma \nabla h(\theta_\star)^T + \nabla h(\theta_\star)\Gamma = -D_\star \qquad \text{under C4b.}$$

Then under the conditional probability $\mathbb{Q}(. \mid \lim_q \theta_q = \theta_\star)$ the sequence $\{\gamma_k^{-1/2}(\theta_k - \theta_\star), k \geq 1\}$ converges in distribution to a random variable with the characteristic function

$$\frac{1}{\mathbb{Q}(. \mid \lim_q \theta_q = \theta_\star)}\mathbb{E}_0\left[\mathbb{1}_{\lim_q \theta_q = \theta_\star}\exp(-\frac{1}{2}t^T \Gamma t)\right].$$

Theorem A.7 corresponds to [16, Theorem 3.2.]; it provides sufficient conditions for a CLT for the averaged sequence

$$\bar{\theta}_k := \frac{1}{k}\sum_{l=1}^{k}\theta_l,$$

along a converging sequence $\{\lim_q \theta_q = \theta_\star\}$ where $\theta_\star \in \mathbb{R}^d$ is fixed (deterministic). It is established under essentially the same assumptions as in Theorem A.6 that, if C3 is strenghtened into

C5. r_k is $\mathcal{G}_k$-adapted and $\gamma_k^{-1/2}|r_k|\mathbb{1}_{\lim_q \theta_q = \theta_\star} \leq X_k Y_k$ where $\mathbb{P}(\sup_k |X_k| < \infty) = 1$ and $\lim_k \mathbb{E}\left[|Y_k|^2\right] = 0$. In addition, $k^{-1/2}\sum_{l=1}^{k} r_l \mathbb{1}_{\lim_q \theta_q = \theta_\star}$ converges to 0 in probability,

then:

Theorem A.7. *[16, Theorem 3.2.]* Assume C1, C2, C4b, and C5. Then for any $t \in \mathbb{R}^d$,

$$\lim_k \mathbb{E}\left[\exp(i\sqrt{k}t^T(\bar{\theta}_k - \theta_\star))|\lim_q \theta_q = \theta_\star\right] = \exp\left(-\frac{1}{2}t^T \nabla h(\theta_\star)^{-1}D_\star(\nabla h(\theta_\star)^{-1})^T t\right),$$

where $D_\star$ is introduced in C2.

A.6. **Proofs of the Results of Section 2.2**

Throughout this section, set $\theta := (\xi, \chi)$, and

$$h(\theta) := -\begin{bmatrix} 1 - \frac{1}{1-\alpha}\mathbb{P}_0(L > \xi) \\ \chi - \xi - \frac{1}{1-\alpha}\mathbb{E}_0\left[(L-\xi)^+\right] \end{bmatrix}, \quad H(\theta, X) := -\begin{bmatrix} 1 - \frac{1}{1-\alpha}\mathbb{1}_{X>\xi} \\ \chi - \xi - \frac{1}{1-\alpha}(X-\xi)^+ \end{bmatrix}. \tag{A.11}$$

We start with a preliminary lemma.

Lemma A.8. Assume H7. Then

$$\mathbb{E}_0\left[(H(\theta, L) - h(\theta))(H(\theta, L) - h(\theta))^T\right] = D_\star + D(\theta), \tag{A.12}$$

where

$$D_\star := \frac{1}{(1-\alpha)^2}\begin{bmatrix} \alpha(1-\alpha) & \alpha\mathbb{E}_0\left[(L-\xi_\star)^+\right] \\ \alpha\mathbb{E}_0\left[(L-\xi_\star)^+\right] & \mathbb{V}\mathrm{ar}_0\left[(L-\xi_\star)^+\right] \end{bmatrix} \tag{A.13}$$

and $\lim_k D(\theta_k)\mathbb{1}_{\lim_q \theta_q = \theta_\star} = 0 \; \mathbb{P}_0$-*a.s*

Proof. By H7, $L \sim Q(z, \cdot)$ under $\mathbb{P}_0$ so $\mathbb{E}_0 [H(\theta, L)] = h(\theta)$. We have

$$\mathbb{E}_0 \left[(H(\theta, L) - h(\theta)) (H(\theta, L) - h(\theta))^T \right] = \mathbb{E}_0 \left[H(\theta, L)H(\theta, L)^T \right] - h(\theta)h(\theta)^T.$$

Denote by $\mathcal{H}(\theta)$ the first 2×2 symmetric matrix on the RHS. Then

$$\mathcal{H}(\theta)_{1,1} = 1 + \frac{\mathbb{P}_0 (L > \xi)}{(1 - \alpha)^2} - \frac{2}{1 - \alpha}\mathbb{P}_0 (L > \xi)$$

$$\mathcal{H}(\theta)_{1,2} = \chi - \xi - \frac{1}{1 - \alpha}\mathbb{E}_0 \left[(L - \xi)^+ \right] - \frac{1}{1 - \alpha}\mathbb{P}_0 (L > \xi) (\chi - \xi) + \frac{1}{(1 - \alpha)^2}\mathbb{E}_0 \left[(L - \xi)^+ \right]$$

$$\mathcal{H}(\theta)_{2,2} = (\chi - \xi)^2 + \frac{1}{(1 - \alpha)^2}\mathbb{E}_0 \left[\left((L - \xi)^+ \right)^2 \right] - \frac{2}{1 - \alpha}(\chi - \xi)\mathbb{E}_0 \left[(L - \xi)^+ \right].$$

Under H7, $\theta \mapsto h(\theta)$ and $\theta \mapsto \mathcal{H}(\theta)$ are continuous at $\theta_\star$. This implies (A.12), where the expression (A.13) for

$$D_\star = \mathcal{H}(\theta_\star) - h(\theta_\star)h(\theta_\star)^T.$$

follows by using that $\theta_\star$ satisfies (1.6) (which implies that $h(\theta_\star) = 0$). ∎

Proof of Theorem 2.4. The proof consists in applying Theorem A.6. We check its assumptions with $\mathbb{Q} \leftarrow \mathbb{P}_0$, $\theta_k \leftarrow (\xi_k, \chi_k)$, $\theta_\star \leftarrow (\xi_\star, \chi_\star)$, the function h given by (A.11). The random variables e_k, r_k are set equal to

$$e_{k+1} \leftarrow H(\theta_k, \tilde{L}_{k+1}) - h(\theta_k) \qquad r_{k+1} \leftarrow H(\theta_k, L_{k+1}) - H(\theta_k, L'_{k+1})$$

where h and H are given by (A.11) and

$$\tilde{L}_{k+1} := \phi^{k+1} + \beta^{k+1}\Psi(Z_1^{k+1}) - \Psi'(z).$$

With these definitions, note that Algorithm 1 updates the parameter θ_{k+1} by

$$\theta_{k+1} = \theta_k + \gamma_{k+1}H(\theta_k, L_{k+1}).$$

Since $\theta_\star$ satisfies (1.6), we have $h(\theta_\star) = 0$. By H7, the function h is twice continuously differentiable in a neighborhood of $\theta_\star$ and the gradient is given by

$$\nabla h(\theta_\star) = -\frac{1}{1 - \alpha} \begin{bmatrix} f(z, \xi_\star) & 0 \\ -(1 - \alpha) + \mathbb{P}_0(L > \xi_\star) & 1 - \alpha \end{bmatrix} = - \begin{bmatrix} \frac{1}{1-\alpha}f(z, \xi_\star) & 0 \\ 0 & 1 \end{bmatrix} \tag{A.14}$$

where we used (1.6) in the last equality. Hence, by H7, the condition C1 is verified.

Set $\mathcal{G}_k := \sigma (L_j, j \leq k)$; note that $h(\theta_k) = \mathbb{E}_0 \left[H(\theta_k, \tilde{L}_{k+1})|\mathcal{G}_k \right]$ – see (A.9) in the proof of Theorem 2.2. Hence, $\{e_k, k \geq 1\}$ is a $\mathcal{G}_k$-adapted $\mathbb{P}_0$-martingale increment sequence. Since $\theta \mapsto h(\theta)$ is continuous at $\theta_\star$ and $\theta_\star$ is fixed (deterministic), for fixed $C > 0$, there exists a constant C' such that $\mathbb{P}_0$-a.s., $\sup_k |h(\theta_k)|\mathbb{1}_{|\theta_k - \theta_\star| \leq C} \leq C'$. In addition

$$|H(\theta, X)| \leq (1 + 1/(1 - \alpha)) + 2|\theta| + (1 - \alpha)^{-1}(X - \xi)^+.$$

By H7, since $\tilde{L}_{k+1}$ has the same distribution as L under $\mathbb{P}_0$, there exists a constant C'' (depending upon C) such that

$$\sup_k \mathbb{E}_0 \left[|H(\theta_k, \tilde{L}_{k+1})|^{2+\tau_\star}\mathbb{1}_{|\theta_k - \theta_\star| \leq C} \right] \leq C''.$$

Hence, the conditions C2a-b are verified. The condition C2c follows from Lemma A.8. We write

$$\mathbb{E}_0\left[|r_{k+1}|\right] = \mathbb{E}_0\left[\mathbb{E}_0\left[|r_{k+1}|\,|\,\mathcal{G}_k\right]\right]$$

$$\leq \frac{1}{1-\alpha}\left(\sup_{\xi\in\mathbb{R}}\mathbb{E}_0\left[\left|\mathbb{1}_{L_{k+1}>\xi} - \mathbb{1}_{\tilde{L}_{k+1}>\xi}\right|\right] + \sup_{\xi\in\mathbb{R}}\mathbb{E}_0\left[\left|(L_{k+1}-\xi)^+ - \left(\tilde{L}_{k+1}-\xi\right)^+\right|\right]\right).$$

By Lemma 2.1, under H7, the LHS is upper bounded by

$$O\left((M_k\wedge M_k')^{-p_\star/(2(1+p_\star))} + (M_k\wedge M_k')^{-1/2}\right) = O\left((M_k\wedge M_k')^{-p_\star/(2(1+p_\star))}\right).$$

Hence, by H8, the condition C3 is verified. Finally, the condition C4 holds by H8 and (A.14). This concludes the proof of the theorem.

Proof of Theorem 2.5. The proof consist in an application of Theorem A.7. We use the same notations as in the proof of Theorem 2.4; it was already proved that C1 and C2 hold. We check C5: we have

$$\mathbb{E}_0\left[|r_{k+1}|^2\right] \leq \frac{1}{(1-\alpha)^2}\sup_{\xi\in\mathbb{R}}\left(\mathbb{E}_0\left[\left|\mathbb{1}_{L_{k+1}>\xi} - \mathbb{1}_{\tilde{L}_{k+1}>\xi}\right|\right] + \mathbb{E}_0\left[\left|(L_{k+1}-\xi)^+ - (\tilde{L}_{k+1}-\xi)^+\right|^2\right]\right).$$

By Lemma 2.1, there exists a constant C such that for any $k\geq 0$,

$$\mathbb{E}_0\left[|r_{k+1}|^2\right] \leq \frac{C}{(M_k\wedge M_k')^{p_\star/(2(1+p_\star))}}.$$

Therefore, the condition on $\gamma_k^{-1/2}r_k$ is satisfied by (2.8). In addition, by Lemma 2.1 again, there exists a constant C' such that for any $\delta > 0$,

$$\mathbb{P}_0\left(k^{-1/2}\left|\sum_{l=1}^k r_l\right| > \delta\right) \leq k^{-1/2}\delta^{-1}\mathbb{E}_0\left[\left|\sum_{l=1}^k r_l\right|\right] \leq C'k^{-1/2}\delta^{-1}\left(\sum_{l=1}^k \frac{1}{(M_l\wedge M_l')^{p_\star/(2(1+p_\star))}}\right).$$

Therefore, the condition C5 holds by (2.8).   ∎

A.7. **Sensitivities of Value-at-Risk and Expected Shortfall to Perturbations of the Input Distribution**

We develop in this section some estimates relative to the perturbation of the value-at-risk and expected shortfall that arise when we use different distributions for the underlying loss variable Z. We use the notation $\mathbb{V}a\mathbb{R}^\alpha(Z)$ and $\mathbb{ES}^\alpha(Z)$ for the $\mathbb{P}$ value-at-risk and expected shortfall of Z

$$\mathbb{P}(Z > \mathbb{V}a\mathbb{R}^\alpha(Z)) := 1-\alpha, \qquad \mathbb{ES}^\alpha(Z) := \frac{1}{1-\alpha}\int_\alpha^1 \mathbb{V}a\mathbb{R}^a(Z)\,\mathrm{d}a,$$

where $\mathbb{V}a\mathbb{R}^\alpha(Z)$ defined on the left is the infimum of such values.

Definition A.9. The Kolmogorov distance $d_{\mathrm{kol}}(X,Y)$ between two scalar random variables X and Y is the sup norm between their cumulative distribution functions:

$$d_{\mathrm{kol}}(X,Y) := \sup_{\xi\in\mathbb{R}}|\mathbb{P}[X\leq\xi] - \mathbb{P}[Y\leq\xi]|.$$

We show that if X, Y are integrable scalar random variables with a continuous density, then for any $\alpha > 0$ fixed, the difference $|[\mathbb{V}a\mathbb{R}^\alpha(X), \mathbb{E}\mathbb{S}^\alpha(X)] - [\mathbb{V}a\mathbb{R}^\alpha(Y), \mathbb{E}\mathbb{S}^\alpha(Y)]|$ is bounded, up to a multiplicative constant depending of α and the density of X, by the L^1 and the Kolmogorov distances between X and Y.

Our first proposition has to do with the relationship between the Kolmogorov distance and the behavior of $\mathbb{V}a\mathbb{R}^\beta(\cdot)$ as a function of β.

Lemma A.10. *Let X and Y be scalar random variables having a continuous cumulative distribution function. Then for any $\alpha \in (0,1)$ and every $\mathbb{V}a\mathbb{R}^\alpha(Y)$ we have*

$$\mathbb{V}a\mathbb{R}^{\alpha - d_{\mathrm{kol}}(X,Y)}(X) \leq \mathbb{V}a\mathbb{R}^\alpha(Y) \leq \mathbb{V}a\mathbb{R}^{\alpha + d_{\mathrm{kol}}(X,Y)}(X). \tag{A.15}$$

for some elements from the respective $\mathbb{V}a\mathbb{R}^\alpha(X)$ sets and with the convention $\mathbb{V}a\mathbb{R}^\beta = -\infty$ (respectively $\mathbb{V}a\mathbb{R}^\beta = \infty$) if $\beta < 0$ (respectively $\beta > 1$).

If X and Y are also integrable then

$$|\mathbb{E}\mathbb{S}^\alpha(X) - \mathbb{E}\mathbb{S}^\alpha(Y)| \leq \frac{1}{1-\alpha}\mathbb{E}|X - Y|. \tag{A.16}$$

Proof. Let $\alpha \in (0,1)$ be given, and let $d := d_{\mathrm{kol}}(X,Y)$. From the definition of the Kolmogorov distance it follows that for every $\xi \in \mathbb{R}$

$$\mathbb{P}[X \leq \xi] - d \leq \mathbb{P}[Y \leq \xi] \leq \mathbb{P}[X \leq \xi] + d,$$

so that for every ξ^α such that $\mathbb{P}[Y \leq \xi^\alpha] = \alpha$ (i.e. for every $\mathbb{V}a\mathbb{R}^\alpha$ element of Y) we have

$$|\mathbb{P}[X \leq \xi^\alpha] - \alpha| \leq d,$$

thus showing that $\mathbb{V}a\mathbb{R}^{\alpha - d}(X) \leq \xi^\alpha \leq \mathbb{V}a\mathbb{R}^{\alpha + d}(X)$ for some elements in the respective $\mathbb{V}a\mathbb{R}$ sets.

For the second equality note that, by the characterization in [7, Section 2.1]- [24, Theorem 1] of $\mathbb{E}\mathbb{S}^\alpha(Z)$, for $Z = X$ or $Z = Y$ (under the assumptions of continuous c.d.f.),

$$\mathbb{E}\mathbb{S}^\alpha(Z) = \inf_x \left(x + \frac{1}{1-\alpha}\mathbb{E}\left[(Z - x)^+\right] \right). \tag{A.17}$$

Now consider the function

$$G(x, z) := x + \frac{1}{1-\alpha}(z - x)^+$$

and note that, for fixed x, the function $G(x, \cdot)$ is a uniformly Lipschitz function of z with Lipschitz constant $1/(1-\alpha)$. This implies in particular, by taking $Z = X$ and $Z = Y$, that for every x

$$\mathbb{E}[G(x, X)] - \mathbb{E}[G(x, Y)]| \leq \frac{1}{1-\alpha}\mathbb{E}|X - Y|.$$

Taking the inf in x in the above and using (A.17), we get (A.16) as desired. ∎

Inspired from [6], we develop further estimates on the Kolmogorov distance between X and Y that might depend on higher moments for the difference between these random variables. We apply these estimates to the error analysis of Algorithm 2, in which the bias due to fixing an approximation procedure for the samplings of Ψ has to be controlled in order to have useful criteria for the choice of the parameters of the algorithm.

Corollary A.11. *Assume that the scalar random variable X has a c.d.f. F which is continuously differentiable and strictly increasing in a neighborhood of $\mathbb{V}a\mathbb{R}^\alpha(X)$, let $f := dF/d\lambda$ be the respective density (where it exists), and let δ be such that the inverse F^{-1} of F exists in an δ−neighborhood of α. Then for any scalar random variable Y and any $0 < r, s < \delta$, the condition*

$$\mathbb{V}a\mathbb{R}^{\alpha - r}(X) \leq \mathbb{V}a\mathbb{R}^\alpha(Y) \leq \mathbb{V}a\mathbb{R}^{\alpha + s}(X)$$

implies that

$$|\mathbb{V}a\mathbb{R}^{\alpha}(X) - \mathbb{V}a\mathbb{R}^{\alpha}(Y)| \leq \sup_{x \in [\alpha-r,\alpha+s]} |f(F^{-1}(x))|^{-1} d_{\mathrm{kol}}(X,Y).$$

Proof. This follows from the fact that, under the given hypotheses

$$\mathbb{V}a\mathbb{R}^{\beta}(X) = F^{-1}(\beta)$$

whenever $|\beta - \alpha| < \delta$. The hypotheses on r and s imply now that

$$|\mathbb{V}a\mathbb{R}^{\alpha}(X) - \mathbb{V}a\mathbb{R}^{\alpha}(Y)| \leq \max\left(\mathbb{V}a\mathbb{R}^{\alpha}(X) - \mathbb{V}a\mathbb{R}^{\alpha-r}(X), \mathbb{V}a\mathbb{R}^{\alpha+s}(X) - \mathbb{V}a\mathbb{R}^{\alpha}(X)\right) \qquad (\text{A.18})$$

$$= \max\left(F^{-1}(\alpha) - F^{-1}(\alpha - r), F^{-1}(\alpha + s) - F^{-1}(\alpha)\right).$$

It follows from the mean value theorem and the inverse function theorem that

$$|\mathbb{V}a\mathbb{R}^{\alpha}(X) - \mathbb{V}a\mathbb{R}^{\alpha}(Y)| \leq \sup_{x \in [\alpha-r,\alpha+s]} |f(F^{-1}(x))|^{-1} \max(r,s)$$

which implies the desired conclusion if $\max(r,s) < d_{\mathrm{kol}}(X,Y)$.

If this is not the case, for instance if $r \leq d_{\mathrm{kol}}(X,Y) < s$, a similar argument using (A.15) (available since X has a continuous c.d.f.) to replace the second term in the maximum (A.18) gives

$$|\mathbb{V}a\mathbb{R}^{\alpha}(X) - \mathbb{V}a\mathbb{R}^{\alpha}(Y)| \leq \sup_{x \in [\alpha-r,\alpha+d_{\mathrm{kol}}(X,Y)]} |f(F^{-1}(x))|^{-1} \max(r, d_{\mathrm{kol}}(X,Y))$$

$$\leq \sup_{x \in [\alpha-r,\alpha+s]} |f(F^{-1}(x))|^{-1} d_{\mathrm{kol}}(X,Y).$$

The other cases are treated similarly.   ∎

In order to pass to controls that depend only on the L^1 distance, we present now two estimates of $d_{\mathrm{kol}}(X,Y)$ that are related to the actual difference between X and Y. These will be combined to estimate the expected error induced by the application of the stochastic approximation procedure to the sequence of samplings produced via regression.

Lemma A.12. *Assume that the scalar random variable X admits a density which is bounded by C_0. Then for any scalar random variable Y and any $\delta > 0$*

$$d_{\mathrm{kol}}(X,Y) \leq 2C_0\delta + \mathbb{P}\left[|X - Y| > \delta\right]. \qquad (\text{A.19})$$

Proof. The following argument was presented already in the proof of Lemma A.5, thus we give here a summarized version: for $\delta > 0$ given and any $\xi \in \mathbb{R}$, we have

$$|\mathbb{P}\left[X \leq \xi\right] - \mathbb{P}\left[Y \leq \xi\right]| \leq \mathbb{E}\left[|\mathbb{1}_{[X \leq \xi]} - \mathbb{1}_{[Y \leq \xi]}|\right]$$

$$= \mathbb{E}\left[\mathbb{1}_{[X \leq \xi < Y]} + \mathbb{1}_{[Y \leq \xi < X]}\right]$$

$$\leq \mathbb{E}\left[\mathbb{1}_{[-\delta+X \leq \xi \leq X+\delta]}\right] + \mathbb{E}\left[\mathbb{1}_{[|X-Y|>\delta]}\right]$$

$$\leq 2C_0\delta + \mathbb{P}\left[|X - Y| > \delta\right],$$

using the hypothesis.   ∎

Corollary A.13. *Assume that the scalar random variable X admits a density which is bounded by C_0. Then for any scalar random variable Y and any $p > 0$ we have*

$$d_{\mathrm{kol}}(X, Y) \leq (2C_0 + 1)(\mathbb{E}|X - Y|^p)^{1/(1+p)}.$$

Proof. For the case in which $\mathbb{E}|X - Y|^p = +\infty$ the conclusion is trivially true. For the $p-$integrable case, take $\delta = (\mathbb{E}|X - Y|^p)^{1/(1+p)}$ in equation (A.19) and apply Markov's inequality. ∎

A.8. **A Nonasymptotic Estimate for Regressions**

The following result is used to control the error due to the introduction of a regression procedure in Algorithm 2:

Theorem A.14 ([18, Theorem 11.5]). *Let (X, Y) be a random vector in $\mathbb{R}^d \times \mathbb{R}$, let $\mathcal{F}$ be a pointwise measurable set of functions $f : \mathbb{R}^d \to \mathbb{R}$, with finite Vapnik-Chervonenkis dimension $\mathbf{VC}_{\mathcal{F}} \geq 1$. Assume that the random variable Y is bounded by $B > 0$. If $\mathcal{D}_n = ((X_k, Y_k))_{k=1}^n$ is any vector of independent copies of (X, Y) and if we define the random function $f_{\mathcal{D}_n}$ by*

$$f_{\mathcal{D}_n} := \hat{f}_{\mathcal{D}_n} \mathbb{1}_{|\hat{f}_{\mathcal{D}_n}| \leq B} + B\mathbb{1}_{\hat{f}_{\mathcal{D}_n} > B} - B\mathbb{1}_{\hat{f}_{\mathcal{D}_n} < -B}$$

where

$$\hat{f}_{\mathcal{D}_n} := \arg\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^n |f(X_k) - Y_k|^2,$$

then there exists a universal constant $C_\star$ such that for any copy (X', Y') of (X, Y) independent of $\mathcal{D}_n$,

$$\mathbb{E}\left[|f_{\mathcal{D}_n}(X') - \mathbb{E}[Y' \mid X']|^2\right] \leq C_\star B^2 \mathbf{VC}_{\mathcal{F}} \frac{(1 + \ln(n))}{n} + 2\inf_{f \in \mathcal{F}} \mathbb{E}\left[|f(X) - \mathbb{E}[Y \mid X]|^2\right]. \tag{A.20}$$

Remark A.15. It may be useful to recall the meaning of (A.20): if $\mu_X = \mathbb{P}X^{-1}$ is the law of X, then for any $\mathcal{D}_n-$measurable nonnegative (random) function $g = g_{\mathcal{D}_n}$ and any copy X' of X independent of $\mathcal{D}_n$,

$$\mathbb{E}\left[g_{\mathcal{D}_n}(X') \mid \mathcal{D}_n\right] = \int_{\mathbb{R}^d} g_{\mathcal{D}_n}(x)\mathrm{d}\mu_X(x). \tag{A.21}$$

Now, if m is a (deterministic) function with the property that

$$m(X) = \mathbb{E}[Y \mid X], \qquad \mathbb{P} - a.s.$$

(and therefore $m(X') = \mathbb{E}[Y' \mid X']$ because $(X, Y) \sim (X', Y')$), and if we apply (A.21) to $g_{\mathcal{D}_n} := |f_{\mathcal{D}_n} - m|^2$, we get that

$$\mathbb{E}\left[|f_{\mathcal{D}_n}(X') - \mathbb{E}[Y' \mid X']|^2 | \mathcal{D}_n\right] \geq \inf_{f \in \mathcal{F}} \mathbb{E}\left[|f \circ X' - m \circ X'|^2\right] = \inf_{f \in \mathcal{F}} \mathbb{E}\left[|f(X) - \mathbb{E}[Y \mid X]|^2\right],$$

and therefore (A.20) *tells us that, up to a factor of 2 (which can be improved by looking carefully at the proofs), the accuracy of $f_{\mathcal{D}_n}$ as a predictor constructed from $\mathcal{F}$ of Y as a function of X deviates from the optimal L^2-accuracy*

$$\inf_{f \in \mathcal{F}} \mathbb{E}\left[|f(X) - \mathbb{E}[Y \mid X]|^2\right]$$

for no more than

$$CB^2 \mathbf{VC}_{\mathcal{F}} \frac{(1 + \ln(n))}{n}$$

units, on $\mathrm{L}_{\mathbb{P}}^2-$ expectation.

About estimating the Vapnik-Chervonenkis dimension of a Neural Network, see for instance [4]. When $\mathcal{F}$ is a vector space, we can replace the factor $\mathbf{VC}_{\mathcal{F}}$ by the dimension of the vector space plus one.

References

[1] L. Abbas-Turki, S Crépey, and B. Diallo. XVA principles, nested Monte Carlo strategies, and GPU optimizations. *International Journal of Theoretical and Applied Finance*, 21:1850030, 2018.

[2] C. Albanese and S. Crépey. XVA analysis from the balance sheet. Working paper available at https://math.maths.univ-evry.fr/crepey, 2018.

[3] F. Anfuso, D. Aziz, K. Loukopoulos, and P. Giltinan. A sound modelling and backtesting framework for forecasting initial margin requirements. *Risk Magazine*, pages May 1–6, 2017.

[4] M. Anthony and P.L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 2009.

[5] Y. Armenti and S. Crépey. XVA Metrics for CCP optimization. Working paper available at https://math.maths.univ-evry.fr/crepey, 2018.

[6] R. Avikainen. On irregular functionals of SDEs and the Euler scheme. *Finance and Stochastics*, 13:381–401, 2009.

[7] O. Bardou, N. Frikha, and G. Pagès. Recursive computation of value-at-risk and conditional value-at-risk using MC and QMC. In *Monte Carlo and quasi-Monte Carlo methods 2008*, pages 193–208. Springer, Berlin, 2009.

[8] O. Bardou, N. Frikha, and G. Pagès. CVaR hedging using quantization-based stochastic approximation algorithm. *Math. Finance*, 26(1):184–229, 2016.

[9] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990.

[10] M. Broadie, Y. Du, and C. Moallemi. Risk estimation via regression. *Operations Research*, 63(5):1077–1097, 2015.

[11] Committee of European Insurance and Occupational Pensions Supervisors. QIS5 technical specifications, 2010. `https://eiopa.europa.eu/Publications/QIS/QIS5-technical_specifications_20100706.pdf`.

[12] S. Crépey, T. R. Bielecki, and D. Brigo. *Counterparty Risk and Funding: A Tale of Two Puzzles*. Chapman & Hall/CRC Financial Mathematics Series, 2014.

[13] S. Crépey and S. Song. Counterparty risk and funding: Immersion and beyond. *Finance and Stochastics*, 20(4):901–930, 2016.

[14] Y. Elouerkhaoui. Pricing and hedging in a dynamic credit model. *International Journal of Theoretical and Applied Finance*, 10(4):703–731, 2007.

[15] Y. Elouerkhaoui. *Credit Correlation: Theory and Practice*. Palgrave Macmillan, 2017.

[16] G. Fort. Central limit theorems for stochastic approximation with controlled Markov chain dynamics. *ESAIM Probab. Stat.*, 19:60–80, 2015.

[17] M.B. Gordy and S. Juneja. Nested simulation in portfolio risk measurement. *Management Science*, 56(10):1833–1848, 2010.

[18] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer Series in Statistics. Springer-Verlag, New York, 2002.

[19] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1980. Probability and Mathematical Statistics.

[20] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

[21] B.T. Polyak. New stochastic approximation type procedures. *Automation and remote control*, 51:98–107, 1990.

[22] B.T. Polyak and A.B. Juditsky. Acceleration of stochastic approximation by averaging. *SIAM J. Control and Optimization*, 30:838–855, 1992.

[23] H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics (Proc. Sympos., Ohio State Univ., Columbus, Ohio, 1971)*, pages 233–257. Academic Press, New York, 1971.

[24] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

[25] Swiss Federal Office of Private Insurance. Technical document on the Swiss solvency test, 2006. `https://www.finma.ch/FinmaArchiv/bpv/download/e/SST_techDok_061002_E_wo_Li_20070118.pdf`.