Learning and Hedging the CVA

S. Crépey*, H. Nguyen†, B. Saadeddine‡

March 27, 2023

Abstract

We devise a CVA pricing and hedging scheme based on two runs of a simulation/regression engine: the first one for learning the conditional CVA at some hedging horizon of choice and computing the CVA at time 0, the second one for deriving CVA least-squares or expected shortfall regression sensitivities. Such sensitivities can be derived for market risk only, in the baseline setup of a calibrated model, or also account for the risk of shifts of model parameters, by randomization of the latter. The ensuing CVA regression sensitivities are two orders of magnitude faster than industry-standard bump-and-revalue sensitivities. As assessed numerically on synthetic swap and option data, they are also found to achieve a better CVA hedging performance.

Keywords: CVA, hedging, sensitivities, learning on simulated data, regression, training, least squares problem, expected shortfall minimization.

Introduction

The objective of this work is to provide fast and relevant CVA sensitivities with respect to diffusive market risk factors and (or not) model parameters. The idea is to regress these sensitivities by minimizing a loss function of choice of the hedged PnL, building on the CVA simulation/neural regression engine of Abbas-Turki, Crépey, and Saadeddine (2022). Related works include Buehler, Gonon, Teichmann, and Wood (2019) or, more precisely, its static hedging counterpart Buehler (2019). Another related paper that inspired our treatment of the sensitivities to model parameters, in a different context of SIMM computations, is Albanese, Caenazzo, and Syrkin (2017).

Acknowledgement: We are grateful to Claudio Albanese, Marc Chataigner, Noureddine Lehdili, and Moez Mrad for useful discussions.

*Email: `stephane.crepey@lpsm.paris`. LPSM/Université Paris Cité, France. Corresponding author. The research of S. Crépey has benefited from the support of the Chair *Capital Markets Tomorrow: Modeling and Computational Issues* under the aegis of the Institut Europlace de Finance, a joint initiative of Laboratoire de Probabilités, Statistique et Modélisation (LPSM) / Université Paris Cité and Crédit Agricole CIB.

†Email: `hoangdung.nguyen@natixis.com`. LPSM/Université Paris Cité. The research of H.D. Nguyen is funded by a CIFRE grant from Natixis.

‡Email: `bouazza.saadeddine2@ca-cib.com`. Quantitative Research GMD, Credit Agricole CIB, Paris.

Market practitioners typically hedge their CVA on the basis of bump(-and-revalue) sensitivities, mostly, or, recently sometimes, adjoint algorithmic differentiation (AAD) sensitivities à la Huge and Savine (2017). We aim at an approach applicable to refined CVA metrics encompassing, for instance, a dynamic initial margin (DIM) as per Albanese, Crépey, Hoskinson, and Saadeddine (2021, Eqn. (61)), computed by means of optimization (training) schemes that are difficult to differentiate using AAD. Hence we focus on bump sensitivities as our industry benchmark in this paper. In any case, AAD like bump sensitivities are sensitivities in the sense of partial derivatives (respectively exact and approximate), which may be too local for CVA hedging purposes. Our regression sensitivities, instead, compress a hedging criterion at a hedging horizon of choice: (6)-(8) or (9) in our baseline formulation below, then made Bayesian-robust to the evolution of model parameters in (11) and (12). Moreover, a key, computational time consideration is that, in the case of gamma hedging, computing a whole matrix of second order bump sensitivities is too time-consuming, so that practitioners usually limit their calculations to diagonal gamma bump sensitivities. Hybrid AAD and bump second order sensitivities can be computed in the same asymptotic time as diagonal gamma bump sensitivities[1], but this is already two orders of magnitude slower than our regression crosss-gamma sensitivities (see Table 1 on page 8)—and they are not applicable to CVA with initial margin. The possibility to account for cross-gamma CVA sensitivities at an affordable cost is another appealing feature of our regression sensitivities.

All our computations are performed under the probability measure which is the blend of physical and pricing measures advocated for XVA computations in Albanese, Crépey, Hoskinson, and Saadeddine (2021, Remark 2.3), with related expectation operator denoted below by $\mathbb{E}$. In particular, the PnLs of the different business desks of the bank are centered under this measure (Crépey, 2022, Remark 2.4). All equations below are written using the risk-free asset as a numéraire.

1 CVA Learning and Hedging Schemes

Our CVA computations rely on the following default-based and intensity-based formulations of the[2] CVA of a bank with clients c, at a simulation time $t = ih$[3]:

$$\begin{aligned}\mathrm{CVA}_t &= \mathbb{E}\Big[\underbrace{\sum_c \sum_{j=i}^{n-1} (\mathrm{MtM}^c_{jh})^+ \mathbb{1}_{jh<\tau^c\le(j+1)h}}_{\mathrm{LGD}_{nh}-\mathrm{LGD}_{ih}} \Big| X_{ih}, Y_{ih}\Big] \\ &= \mathbb{E}\Big[h\sum_c \sum_{j=i}^{n-1} (\mathrm{MtM}^c_{jh})^+ (e^{-\sum_{\imath=i}^{j-1}\gamma^c_\imath} - e^{-\sum_{\imath=i}^{j}\gamma^c_\imath})\mathbb{1}_{\{\tau^c>ih\}} \Big| X_{ih}, Y_{ih}\Big], \quad (1)\end{aligned}$$

where MtM^c is the counterparty-risk-free valuation of the portfolio of the bank with its client c, τ^c is the client c's default time with intensity γ^c, X is the vector default indicator process of all clients of the bank, Y is a diffusive market risk factor vector

[1] Asymptotically with respect to the number of variables, a second layer of AAD has no advantage over a bump-and-revalue approach in terms of computation time (Savine, 2018, page 402).

[2] uncollateralized, for notational simplicity, and time-discretized.

[3] cf. (Abbas-Turki, Crépey, and Saadeddine, 2022, (25)-(27)).

process such that each MtM_s^c and γ_s^c are Borelian functions of (s, Y_s), and

$$\mathrm{LGD}_{ih} = \sum_c \sum_{j=0}^{i-1} (\mathrm{MtM}_{jh}^c)^+ \mathbb{1}_{jh<\tau^c\leq(j+1)h}. \tag{2}$$

From a numerical viewpoint, the second line in (1) entails less variance than the first one (Abbas-Turki, Crépey, and Saadeddine, 2022). Hence we rely for our numerics on this second line or, equivalently to it, $\mathrm{CVA}_{ih} \approx \varphi_i(X_{ih}, Y_{ih})$ with

$$\varphi_i \in \operatorname*{arg\,min}_{\varphi\in\mathcal{B}} \mathbb{E}\Big(h \sum_c \sum_{j=i}^{n-1} (\mathrm{MtM}_{jh}^c)^+ (e^{-\sum_{\imath=i}^{j-1}\gamma_\imath^c} - e^{-\sum_{\imath=i}^{j}\gamma_\imath^c}) \mathbb{1}_{\{\tau^c>ih\}} - \varphi(X_{ih}, Y_{ih})\Big)^2, \tag{3}$$

where $\mathcal{B}$ is the set of measurable functions on the state space of all (market and default indicator) risk factors. We denote by CVA_t^θ the conditional CVA at time $t = ih > 0$ learned by a neural network with parameters θ on the basis of simulated CVA originating cash flows, replacing $\mathbb{E}$ by a simulated sample mean $\widehat{\mathbb{E}}$ (also used for computing the initial CVA based on the second line of (1) for $t = 0$ there) and $\mathcal{B}$ by a (positive) neural net search space in the optimization problem (3)[4]. The latter is then addressed numerically by mini-batch stochastic gradient descent, using the optimized GPU implementation of the hierarchical simulation/regression scheme of Abbas-Turki, Crépey, and Saadeddine (2022) available on `https://github.com/BouazzaSE/NeuralXVA`. This CVA learning engine, implemented in PyTorch with custom CUDA kernels for the generation of the CVA originating cash flows at the forward simulation stage, is validated numerically by a brute force nested Monte Carlo approach in Abbas-Turki, Crépey, and Saadeddine (2022, Section 4).

By loss L_t^θ, we then mean the following proxy for the hedging error (PnL) of the CVA desk of the bank on $[0, t]$:

$$\begin{aligned} &L_t^\theta = \delta\mathrm{CVA}^\theta + \mathrm{LGD}_t - (\delta Y)^\top \Delta - \frac{1}{2}(\delta Y)^\top \Gamma \delta Y - t\Theta, \text{ where} \\ &\delta Y = Y_t - Y_0,\ \delta\mathrm{CVA}^\theta = \mathrm{CVA}_t^\theta - \mathrm{CVA}_0, \text{ and } \mathbb{E}L_t^\theta = 0, \end{aligned} \tag{4}$$

where the constraint $\mathbb{E}L_t^\theta = 0$ stems from the choice of the prevailing probability measure mentioned at the end of the Introduction. The contribution

$$(\delta Y)^\top \Delta + \frac{1}{2}(\delta Y)^\top \Gamma \delta Y + t\Theta \tag{5}$$

to (4) is a quadratic approximation of the appreciation over $[0, t]$ of the hedging side[5] of the position sold by the bank for hedging its CVA. This representation (5) is in line with a market practice of only hedging the diffusive risk factors of the CVA, as opposed to clients' jump-to-defaults, fault, in particular, of suitable and sufficiently liquid hedging assets. In the numerics of Section 2, Δ and Γ in (4) are treated as free parameters, while Θ is deduced from the latter through the constraint that $\widehat{\mathbb{E}}L_t^\theta = 0$.

[4] in the Bayesian-mixture variant of the approach, model parameters (which are implicit in (1)) randomly drawn at time 0 may be added to the regressors (inputs of the neural net) (X_{ih}, Y_{ih}), the way detailed in Section 1.2.

[5] hedge assume not to pay any intermediary cash flows, for notational simplicity.

1.1 Baseline Regression Sensitivities

By CVA expected shortfall (ES) regression sensitivities (with gammas or not, which, if not, are then set to 0 in (4)), we mean[6]

$$(\Delta^{es}[,\Gamma^{es}]) = \underset{\Delta\in\mathbb{R}^d[,\Gamma\in\mathbb{R}^{d\times d}]}{\arg\min}\ \mathbb{ES}\left(L_t^\theta\right), \tag{6}$$

where $\mathbb{ES}(L_t^\theta)$ is the expected shorfall of L_t^θ at the confidence level $\alpha = 95\%$, i.e. the expected hedging loss L_t^θ given this loss exceeds the quantile (value-at-risk) of level 95% of its distribution (see e.g. Fissler, Ziegel, and Gneiting (2016)). The inclusion of gamma features reflects the case of a hedging portfolio encompassing nonlinear (optional) instruments, dubbed nonlinear hedging case hereafter, as opposed to the linear hedging case of a hedging portfolio without convexity. The interest of an expected shortfall (ES) hedging criterion is to protect against extreme events, but also to meet regulatory FRTB obligations, which increasingly involve an expected shorfall to validate a hedging strategy (Basel Committee on Banking Supervision, 2013).

An expected shortfall at the hedging horizon of $t = 1$ year also corresponds to the notion of economic capital that underlies the KVA metric in Albanese, Caenazzo, and Crépey (2016), Albanese, Crépey, Hoskinson, and Saadeddine (2021), or Crépey (2022). In fact, as discussed after Eqn. (33) in Albanese, Caenazzo, and Crépey (2017), one has

$$\mathrm{KVA}_0 \approx r\int_0^T e^{-rt}\mathrm{EC}(t)\mathrm{d}t, \tag{7}$$

where EC is an economic capital term structure, defined at each time $t > 0$ to be an expected shortfall (which is positive) of the trading PnL of the bank over the time window $[t, t + 1 \text{ year}]$, r (e.g. 10%) is a hurdle rate representing the compensation shareholders expect for their capital at risk, and T is the horizon of the portfolio of the bank, assumed to be held on a run-off basis. Hence, if one adapts the criterion in (6) to address a sum of expected future expected shortfalls, one would in effect be minimizing an approximation of KVA_0, either using a static hedge as in this paper, or a dynamic hedge using deep hedging.

Following Rockafellar and Uryasev (2000), (6) can be recast in the setup of the following convex optimization problem:

$$(\Delta^{es}, [\Gamma^{es},]c^{es}) = \underset{\Delta\in\mathbb{R}^d[,\Gamma\in\mathbb{R}^{d\times d}],c\in\mathbb{R}}{\arg\min}\ c + \frac{1}{1-\alpha}\mathbb{E}\left[\left(L_t^\theta - c\right)^+\right], \tag{8}$$

where c^{es} is then the corresponding value-at-risk of the optimally hedged PnL. As a slightly simpler variation on the sensitivities (6)-(8), we also consider the following least-squares (LS) hedging ratios:

$$\left(\Delta^{ls}[,\Gamma^{ls}]\right) = \underset{\Delta\in\mathbb{R}^d[,\Gamma\in\mathbb{R}^{d\times d}]}{\arg\min}\ \mathbb{E}[(L^\theta)^2]. \tag{9}$$

[6] By transposition, $(\delta Y)^\top\Gamma\delta Y = (\delta Y)^\top\Gamma^\top\delta Y$, which is hence also equal to $(\delta Y)^\top\tilde{\Gamma}\delta Y$, where $\tilde{\Gamma} = (\Gamma + \Gamma^\top)/2$. Hence minimizing over all the matrices Γ, in (6) or all the analogous expressions below, is equivalent to minimizing over the symmetric matrices only.

Once the CVA at time t has been learned as explained after (3) and the time-0 CVA obtained as a sample mean approximation for (the second line in) (1), the CVA expected shortfall regression sensitivities (6) are computed by stochastic gradient descent based on the empirical[7] version of (8); the CVA least-squares regression sensitivities are computed by performing a linear regression[8] corresponding to an empirical risk minimization version of (9). Bump sensitivities are computed by relaunching the CVA simulation engine for values shocked by 1% of the parameters of interest and taking the differences between the bumped CVA_0s and the baseline one. Note that bump sensitivities necessarily ignore the LGD_t term in (4) (or (13) alike in the Bayesian-mixture setup below), while the (least-squares or expected shortfall) regression sensitivities by definition take it into account (even if it is not perfectly hedgeable).

1.2 Bayesian-Mixture CVA and Its Sensitivities

For improving the robustness of the hedge with respect to shifts in model parameters denoted hereafter in vector form by ρ, the CVA simulation/regression engine is slightly modified so as to randomize these parameters. Namely[9], at time 0, we randomly draw sets of model parameters among a number of prespecified scenarios around the baseline set of parameters. The distribution used for the model parameters can be assessed historically, as done in the different context of Albanese, Caenazzo, and Syrkin (2017), or using a simulation device such as (14) below. Each drawn set of model parameters is then used for simulating a batch of trajectories of the market risk factors. Specifically, if our CVA engine generates m exposure trajectories and we want to use $p \leq m$ sets of model parameters (with m a multiple of p), then we generate $\frac{m}{p}$ p batches of exposure paths. Each batch is associated to one out of p shocks, randomly drawn at time 0, on the baseline model parameters ρ_0. Moreover, the model parameters ρ are added to the inputs of the neural networks capturing the conditional CVA at time $t > 0$, according to the following variation on (3): $\mathrm{CVA}_{ih} = \tilde{\varphi}_i(X_{ih}, Y_{ih}, \varrho)$, with

$$\tilde{\varphi}_i \in \operatorname*{arg\,min}_{\tilde{\varphi} \in \tilde{\mathcal{B}}} \mathbb{E}\big(h \sum_c \sum_{j=i}^{n-1} (\mathrm{MtM}^c_{jh})^+ (e^{-\sum_{\imath=i}^{j-1} \gamma^c_\imath} - e^{-\sum_{\imath=i}^{j} \gamma^c_\imath}) \mathbb{1}_{\{\tau^c > ih\}} - \tilde{\varphi}(X_{ih}, Y_{ih}, \varrho)\big)^2, \quad (10)$$

where $\tilde{\mathcal{B}}$ is the set of measurable functions on the extended state space of all risk factors and (relevant) model parameters, while a shock $\delta\varrho = \varrho - \rho_0$ on the baseline model parameters ρ_0 is randomly drawn at time 0. Note that the MtM^c, γ^c, X and Y now all depend on the random model parameters vector ϱ (so that MtM^c_s and γ^c_s are functions of $(s, Y_s; \varrho)$).

Our Bayesian-mixture variation on (8) (or (9)) then corresponds to substituting the baseline CVA by the Bayesian-mixture CVA in (8) (or (9)), in order to derive the ensuing Bayesian-mixture CVA sensitivities, and to widen the range of hedging instruments reflected by the hedged PnL (4), in order to also "hedge model shifts",

[7] baseline or Bayesian-mixture in Section 1.2.

[8] implemented using a truncated singular value decomposition (SVD) approach, see e.g. Golub and Van Loan (2013, Theorems 2.5.2 and 5.5.1).

[9] cf. Albanese, Caenazzo, and Syrkin (2017) for an analogous technique used in a context of nested Monte Carlo conditional value-at-risk computations.

i.e. hedge shifts of model parameters. In this extended setup, (8), (9) and (4) are respectively replaced by

$$(\Delta^{es}[,\Gamma^{es},\mathcal{V}^{es}],c^{es}) = \underset{\Delta\in\mathbb{R}^d[,\Gamma\in\mathbb{R}^{d\times d},\mathcal{V}\in\mathbb{R}^{\tilde{d}}],c\in\mathbb{R}}{\arg\min}\; c+\frac{1}{1-\alpha}\mathbb{E}\left[\left(L_t^\theta-c\right)^+\right], \tag{11}$$

$$\left(\Delta^{ls}[,\Gamma^{ls},\mathcal{V}^{ls}]\right) = \underset{\Delta\in\mathbb{R}^d[,\Gamma\in\mathbb{R}^{d\times d},\mathcal{V}\in\mathbb{R}^{\tilde{d}}]}{\arg\min}\; \mathbb{E}[(L_t^\theta)^2], \tag{12}$$

and

$$\begin{aligned} &L_t^\theta = \delta\mathrm{CVA}^\theta + \mathrm{LGD}_t - (\delta Y)^\top\Delta - \frac{1}{2}(\delta Y)^\top\Gamma\delta Y - (\delta\varrho)^\top\mathcal{V} - t\Theta, \text{ where} \\ &\delta Y = Y_t - Y_0,\ \delta\mathrm{CVA}^\theta = \mathrm{CVA}_t^\theta - \mathrm{CVA}_0, \text{ and } \mathbb{E}L_t^\theta = 0, \end{aligned} \tag{13}$$

in which CVA_t^θ is the Bayesian-mixture CVA based on (10). Algorithm 1 summarizes the derivation of the Bayesian-mixture CVA regression sensitivities.

name : BayesRegSensi
input : A set of baseline model parameters ρ_0, initial diffusive risk factor values Y_0, a hedging horizon t, a number of exposure paths m, a number p of shocks on the model parameters.
output: Estimated sensitivities $\Delta(,\Gamma,\mathcal{V})$.
Draw p i.i.d. shocked model parameters $\mathrm{P} = \{\rho_1,\dots,\rho_p\}$
for $\rho\in\mathrm{P}$ **do** `// loop over model parameter set`
  Simulate $\frac{m}{p}$ exposure paths (X_t,Y_t) corresponding to the model parameters ρ
end
Compute the related LGD_t as per (2).
Compute the related CVA_t^θ based on (10), by neural net regression the way detailed after (3), and compute CVA_0 as a sample mean approximation for the second line in (1).
Define the hedging error (13) (with $\Gamma,\mathcal{V}$ set to 0 in the linear hedging case)
if *ES regression* **then**
  Solve (11) (with $\Gamma,\mathcal{V}$ set to 0 in the linear hedging case) by Adam stochastic gradient descent.
else if *LS regression* **then**
  Solve (12) (with $\Gamma,\mathcal{V}$ set to 0 in the linear hedging case) by SVD of the covariance matrix of the corresponding least squares problem.
end

Algorithm 1: Derivation of the Bayesian-mixture CVA regression sensitivities.

2 Numerics

In our case study, we consider the CVAs on two portfolios: one of interest rate swaps, as in Abbas-Turki, Crépey, and Saadeddine (2022, Section 3.3), and one of European options, with a final maturity of 10 years in each case, and for two different CVA hedging horizons: $t = 1$yr, as a targeted reference case of interest (a horizon of 1yr being natural for counterparty credit risk applications), but also 0.1yr. All the numerical results are out-of-sample with a repartition training set 70%, test set 30%. All training sets have $4,480$K data points (i.e. model paths) and batches of size 50K. More precisely, we report below on the two following configurations:

1yr train $t = 1$ / test $t' = 1$, the reference case of interest;

0.1yr train $t = 0.1$ / test $t' = 0.1$, to assess the impact on the results of the CVA nonlinearity, which is more sizeable with longer hedging horizons (and with the option portfolio), so for 1yr rather than 0.1yr;

Hedging is performed by means of LS or 95% ES regression sensitivities, as well as Bayesian-mixture variants of the latter. Baseline industry-standard bump-and-revalue sensitivities with respect to market risk factors, and also model parameters in the setup of Section 1.2, are used as a benchmark. For bump sensitivities with market gamma features, we consider only the diagonal of the matrix Γ, as computing the whole matrix would be too time-consuming (see the Introduction).

This is implemented in Abbas-Turki, Crépey, and Saadeddine (2022, Section 3.3)'s setup of a bank trading derivatives in 10 economies with 8 clients. The corresponding market and credit model entails 10 interest rates, 9 cross-currency rates and 8 default intensities, yielding a total of 27 diffusive risk factors and 8 default indicator processes. The swap portfolio consists of 500 interest rate swaps with random characteristics (notional, currency and counterparty). All swaps are priced at par at inception. The option portfolio comprises 500 near-the-money European vanilla options written on randomly selected cross-currency rates, with other random specifications (maturity, notional, strike, counterparty, and call or put option). All the swaps and options have analytic counterparty-risk-free valuations (aggregated into the corresponding clients MtM^c processes) in our market models.

We use $n = 10$ (respectively $n = 100$) pricing time steps for $t = 1$ (respectively $t = 0.1$) and 25 simulation sub-steps per pricing time step. For comparability purposes, all the losses L_t^θ in (4) or (13) are standardized before reporting, i.e. normalized by the standard error of the corresponding[10] unhedged loss with $\Delta(=\Gamma)(=\mathcal{V}) = 0$ before inclusion in any table or figure below. For all the runs of the simulations, whether it is for training or testing, we use 10^5 paths[11] for the market risk factors, along with a hierarchical over-simulation factor for defaults of 64, i.e., given each market path, we simulate 64 default paths of the clients c, the way detailed in Abbas-Turki et al. (2022).

Table 1 shows the high computation time of the bump sensitivities compared to those obtained via (least squares or expected shortfall) regression. On our machine[12], one derivation of the bump sensitivities takes about 15 minutes for $t = 1$ (20 minutes for $t = 0.1$, which uses a smaller simulation time step from 0 to $T = 10yr$) in the baseline case and the runtime is doubled in the Bayesian-mixture case. One derivation of the LS sensitivities using SVD only takes 33 seconds. The ES sensitivities take 24 seconds more as we carefully do up to $4,500$ gradient steps by looping over 50 epochs of 90 batches (of size 50K each).

In the context of this paper where we focus on hedging the CVA at a fixed horizon t (such as one year, whereas the final maturity of the portfolio of a bank may reach 50 years or more), we opted for learning the CVA at time t only, using many epochs for this single learning task. An alternative detailed in (Abbas-Turki

[10] for otherwise same specifications.

[11] due to the time constraint and numerical stability of CVA_0, only 5×10^4 paths are used for computing the bump sensitivities.

[12] a server with an Intel(R) Xeon(R) Gold 5217 CPU and a Nvidia Tesla V100 GPU.

	Train t/ Test t'	Bump sensis hedge	LS hedge		ES hedge	
			Simulation and CVA regression time	Sensis regression time	Simulation and CVA regression time	Sensis regression time
Baseline approach	1/1	958	17	(1.5) 33	17	(7) 57
	0.1/0.1	1189	69	(1.5) 33	69	(7) 57
Bayesian-mixture approach	1/1	2006	21	(1.5) 33	21	(7) 57
	0.1/0.1	2468	190	(1.5) 33	190	(7) 57

Table 1: Computation times (in seconds) for CVA training and hedging, averaged on the swap and option portfolios. The times in parentheses correspond to regressions with delta features only. In the cases including (market risk) gamma features, bump sensitivities are restricted to diagonal gammas, whereas the LS and ES regression approaches use full gamma matrices.

et al., 2022) is to learn the CVA with less epochs at each time step of a CVA pricing time grid, starting from the final time step and going backward, benefiting from the transfer of the weights learned at each previous pricing time step to initialise the weights at the next one. The corresponding CVA computations are not necessarily longer, may even be more precise at time t, and result in a CVA learned at all (pricing) times, which can then be plugged in refined XVA computations, accounting for instance for the feedback impact of reserve capital (that encompasses the CVA) into the FVA and the KVA as per (Albanese, Crépey, Hoskinson, and Saadeddine, 2021, Eqns. (51)-(52) and top left panel of Figure 13), (Crépey, 2022, Eqn. (6.28)).

2.1 Results in the Baseline Case

When regressing with Δ features only, some regressed sensitivies are close to their estimates by bump sensitivites, especially for cross-currency rates (see the left subfigures). With the addition of Γ features, the regression deltas change significantly and they may become very far away from the bump ones (see the middle subfigures). Large discrepancies between the regressed and the bump diagonal gammas are noticeable (see the right subfigures). Such behavior is consistent with the observation concluding Section 1.2. However, additional investigations not reported in the paper for length sake seem to indicate that these discrepancies are mainly explained by numerical issues related to the normalizations that we use as a regression pre-processing step. However, we prefer to keep these normalizations as they help the learnings converge faster (especially the Adam stochastic gradient descent in ES regression) and result in more stable hedging scores.

Figure 1 summarizes the hedging performance of different approaches. As a sanity check, we observe in Figure 1 that, consistent with their definitions, the LS regression sensitivities yield the lowest standard error of the hedging error, while ES regression sensitivities focusing on the tail of the distribution have the lowest ES score. In the reference 1yr case, LS and ES regression sensitivities efficiently reduce the standard error of the hedged PnL down to 37% (a 4% boost down to 33% by the inclusion of Γ features) for the swap portfolio and 53% (a 14% boost down to 39% by the inclusion of Γ features) for the option portfolio, and approximately cut the expected shortfall of the loss by one half[13] (and another half by the inclusion of

[13] A reduction from 1.7 to 0.6 for the swap portfolio and from 2.2 to 1.2 for the option portfolio

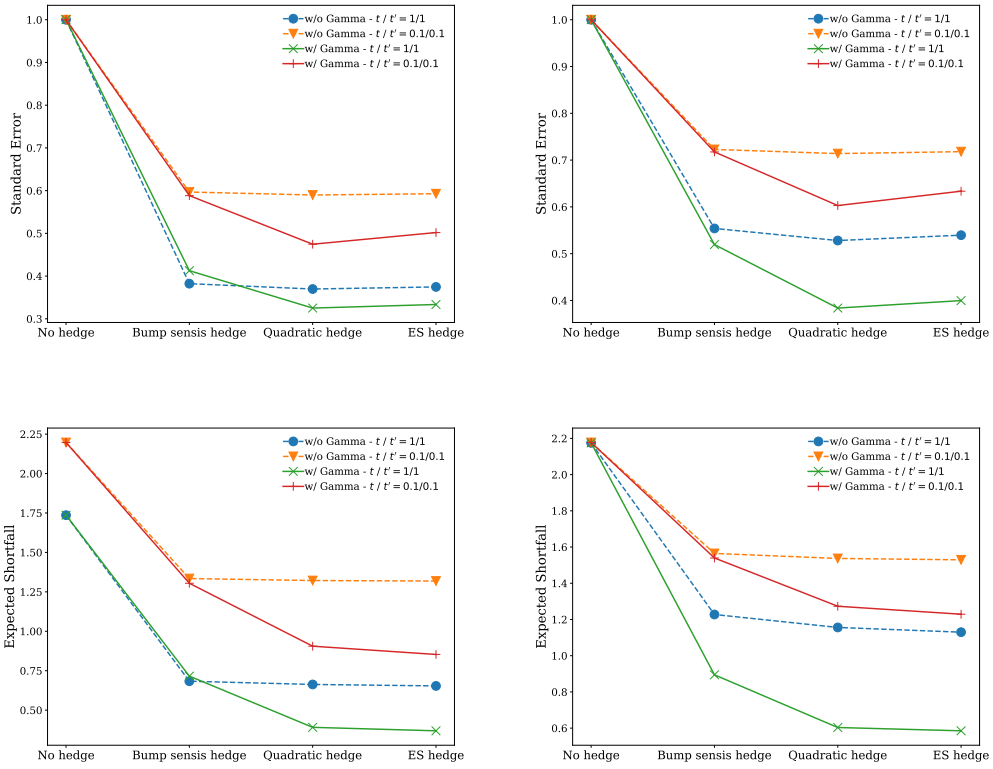

Figure 1: Baseline score results: (*dashed*) linear hedging case, (*solid*) nonlinear hedging case, (*left*) swap portfolio, (*right*) option portfolio.

Γ features). Except in the pathological[14] cross case, bump sensitivities are always the worst performer of all, for either risk score LS or ES (with, regarding the latter, qq plots in Figure 2 consistently below the diagonal of the delta bump sensitivities benchmarks); regression-based hedges are always significantly improved by the addition of Γ features. For the swap portfolio, including diagonal gammas even degrades the bump hedging performance (but it is not the case for the nonlinear option portfolio).

In the 0.1yr case, the results are similar, but the differences between the performances of the different approaches are less pronounced. Going to the cross case, the performance of (LS or ES) regression sensitivities is severely impacted by the mismatch between the hedging horizon and the horizon on which the sensitivities are regressed. Bump sensitivities provide a more robust hedge in this respect: even in this cross case, the bump sensitivities are just different estimates of the derivatives of CVA_0 with respect to the risk factors. The baseline sensitivities learned from a training data set, instead, do not perform well on the test data set associated with a different distribution (corresponding to a different time step $t' \neq t$). Adding second-order Γ regression terms is then even heavily counter-productive, ruining the performance of the corresponding hedges.

Figure 2 shows the qq plots of the PnL hedged by the different baseline ap-

[14] as detailed below.

proaches against the benchmark provided by the PnL hedged by bump delta sensitivities (which therefore corresponds to the diagonal on each plot). Consistent with their definition, ES sensitivities do not only hedge the quantile of the loss at the corresponding confidence level $\alpha = 95\%$ (corresponding to the nodes circled in red in the panels), but also at higher quantile levels. Both LS and ES hedges are significantly improved by the addition of Γ features, with a distribution of the loss, especially in the right tail, shifted downward by the inclusion of the corresponding second-order terms in Figure 2. The bump-sensitivity hedges, instead, cannot benefit from the inclusion of the cross-gammas, which are too heavy to compute numerically (see the Introduction). The comparison between the performances of the hedges with and without Γ features in Figure 1 (or with diagonal Γ features only in the case of of the bump-sensitivity hedges) shows that this is a severe limitation of the bump-sensitivity hedges with respect to the regression ones.

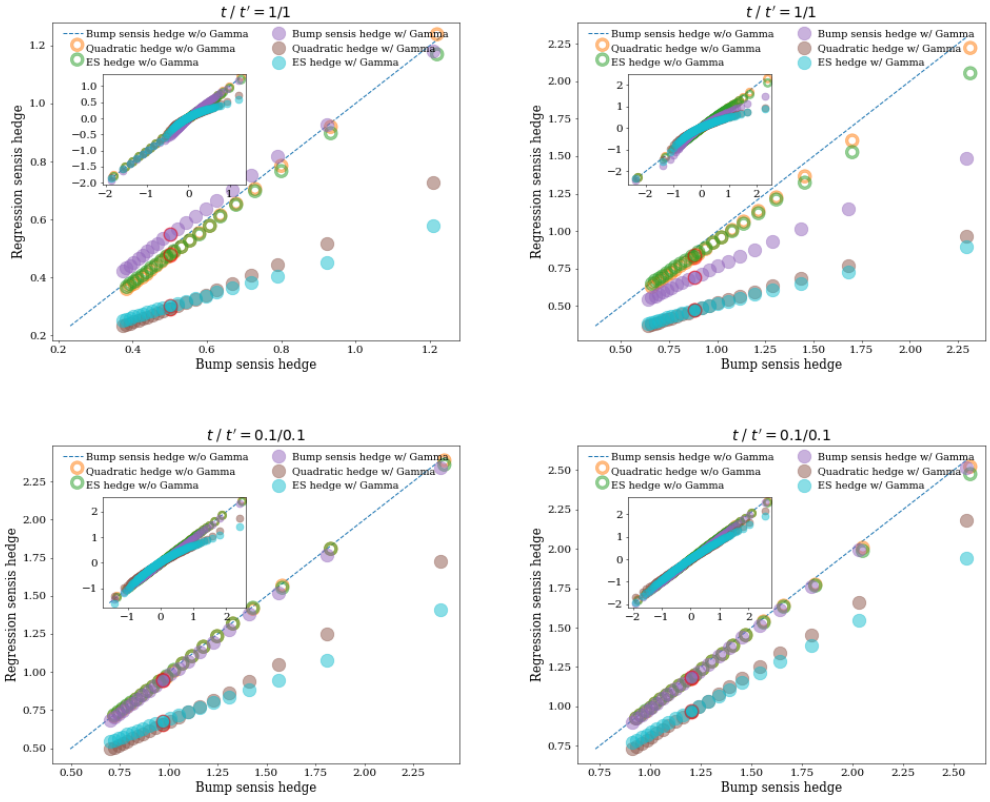

Figure 2: Baseline case: qq plots of the hedged PnL (4) against the benchmark provided by bump delta sensitivities: (*left*) swap portfolio, (*right*) option portfolio, (*cored disk*) linear hedging case, (*full disk*) nonlinear hedging case.
The main plot corresponds to the tail of the distribution starting from $\alpha = 90\%$, the subplot corresponds to the whole distribution. The red circled point corresponds to the confidence level optimised in (8), i.e. 95%.

2.2 Results in the Bayesian-Mixture Case

Under the Bayesian-mixture variation on our approach, we randomize the volatilities of the market risk factors[15]. To sample random scenarios of these volatilities, σ^l say, we use 16 realizations of an Euler scheme at the time horizon t (when training) or t' (when testing) for Ornstein–Uhlenbeck processes

$$d\sigma_t^l = a^l(b^l - \sigma_t^l)dt + \nu^l dW_t^l, \tag{14}$$

where the superscript l relates to the lth market risk factor, σ^l and W^l are the corresponding volatility and driving Brownian motion, and a^l, b^l and ν^l (common to the 16 realizations of the process σ^l) are chosen randomly subject to the Feller condition that guarantees positive volatilities.

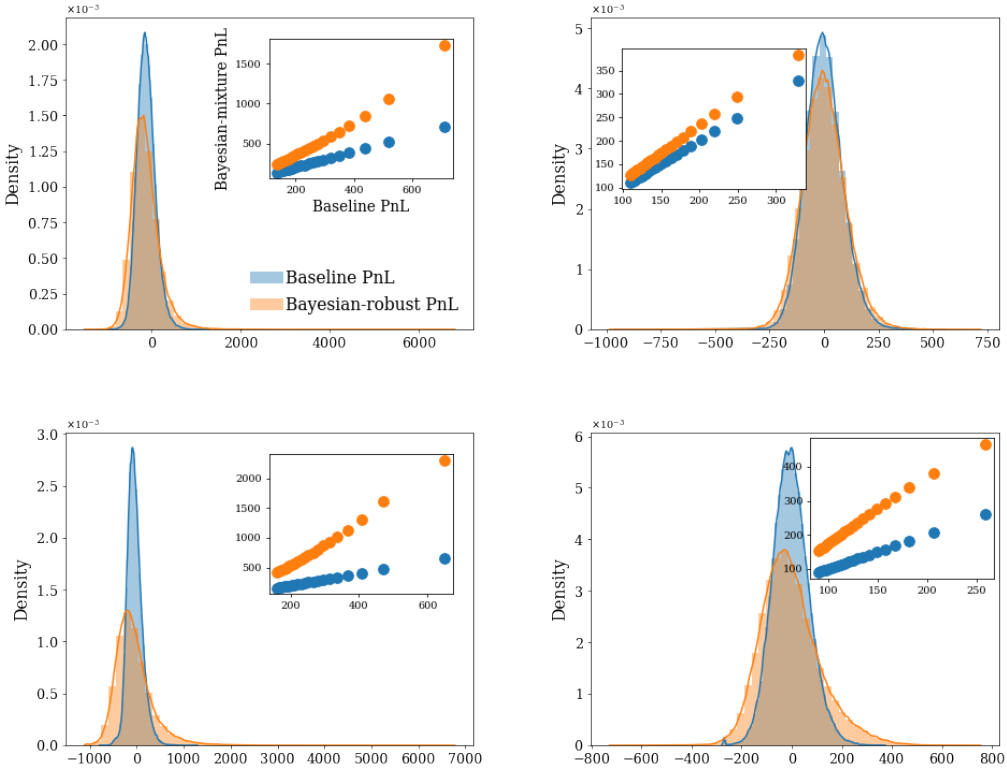

Figure 3: Density plots (*main plot*) and qq plots (*subplot*) of the unhedged PnL in the baseline and Bayesian-mixture cases. The qq plots sketch the tails of the distributions above their 90% quantiles.

Figure 3 compares the distributions of the unhedged CVA PnL, with all deltas, gammas and vegas set to 0 in (4) or (13), in the baseline and Bayesian-mixture cases. We notice the heavier tail of the distribution in the latter case. This effect is more significant for the option portfolio (see in particular the qq plot in Figure 3),

[15] all the model parameters are detailed in Abbas-Turki, Crépey, and Saadeddine (2022, Appendix B).

suggesting that the Bayesian-mixture approach would be especially needed in this case.

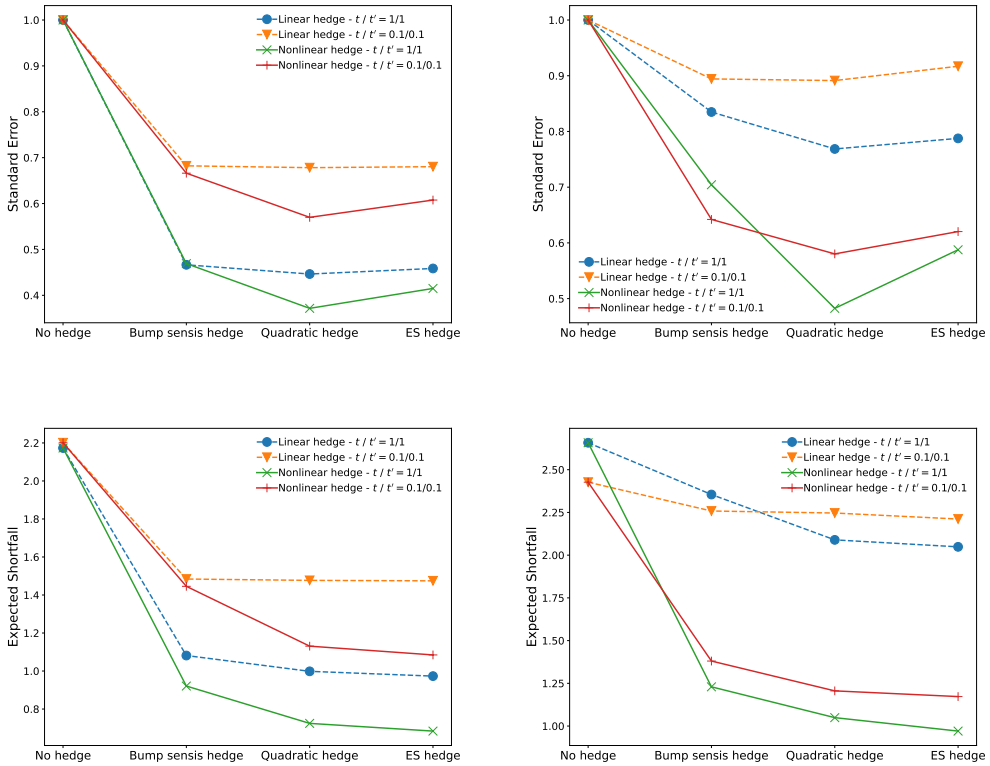

Figure 4: Bayesian-mixture score results: (*dashed*) Δ hedging, (*solid*) $\Delta, \Gamma, \mathcal{V}$ hedging, (*left*) swap portfolio, (*right*) option portfolio.

We focus on the hedged CVA PnL later on. Figure 5 shows results analogous to the ones of Section 2.1, adapted to the Bayesian-mixture setup. In the Bayesian-mixture case, we put into perspective the linear hedging case, i.e. delta hedging market risk only, with the nonlinear hedging also accounting for gamma (with respect to market risk) and vega risk. Most remarks done in the baseline case without randomization are still valid in this enriched setup. The regressed sensitivities consistently outperform the bump sensitivities, both in terms of computation times (see Table 1) and of hedging performances with, in particular, the corresponding qq plots consistently below the diagonal of the bump delta sensitivity benchmarks in Figure 5. The observation on the estimated deltas from Figure **??** in Section **??** is similar as in the baseline case of Figure **??**: the regressed sensitivities may be very different from the bump sensitivities. Figure 5 shows that adding nonlinear Γ and $\mathcal{V}$ features only marginally improves the hedging performance for the swap portfolio; in this case, the loss is linearly explained by the linear changes in the risk factors, which is coherent with the nature of the product. By contrast, $\Gamma/\mathcal{V}$ hedging significantly improves the hedging scores for the option portfolio. For example, in

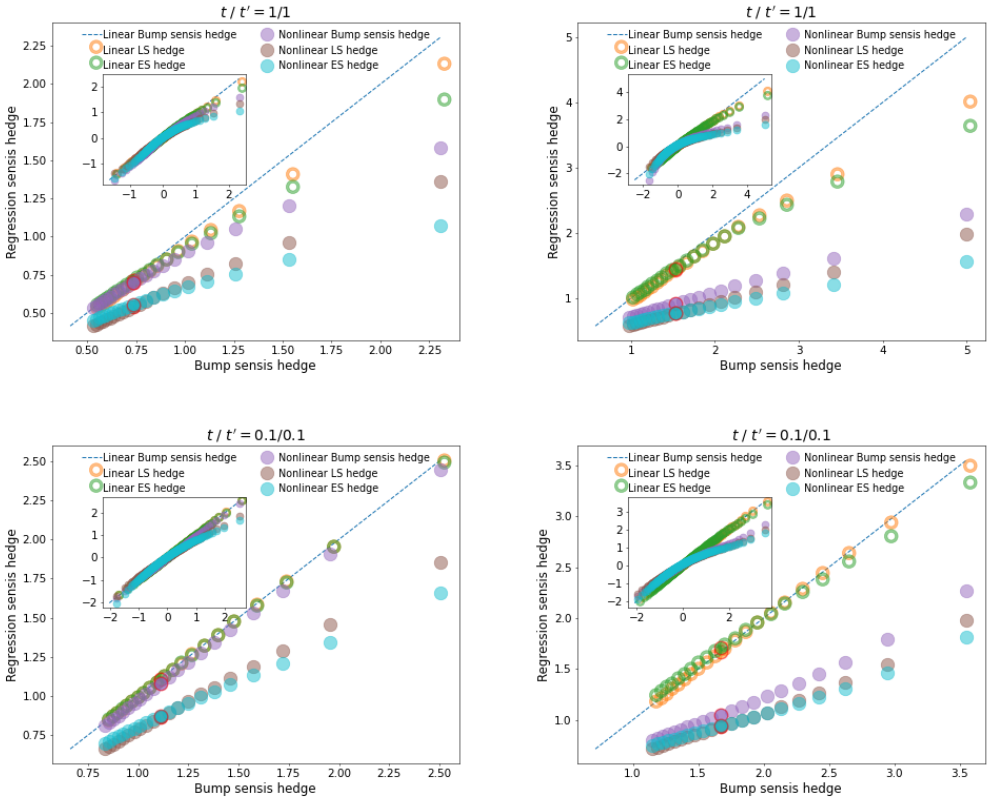

Figure 5: Bayesian-mixture case: qq plots of the hedged PnL (13) against the benchmark provided by delta bump sensitivities; (*cored disk*) Δ hedging, (*full disk*) $\Delta, \Gamma, \mathcal{V}$ hedging, (*left*) swap portfolio, (*right*) option portfolio.

The main plot corresponds to the tail of the distribution starting from $\alpha = 90\%$, the subplot corresponds to the whole distribution. The red circled point corresponds to the confidence level optimized in (11), i.e. 95%.

the reference 1yr case, $\Gamma/\mathcal{V}$ hedging reduces the LS score of the LS hedge by 29%[16] and of the ES hedge by 20%[17], and it reduces their ES scores by around 50%[18]. This boosting effect of nonlinear hedging features on the hedging performance is still visible in the 0.1yr case. The results also recommend to take into account vegas and diagonal gammas when hedging by bump sensitivities for options.

2.3 Robustness test

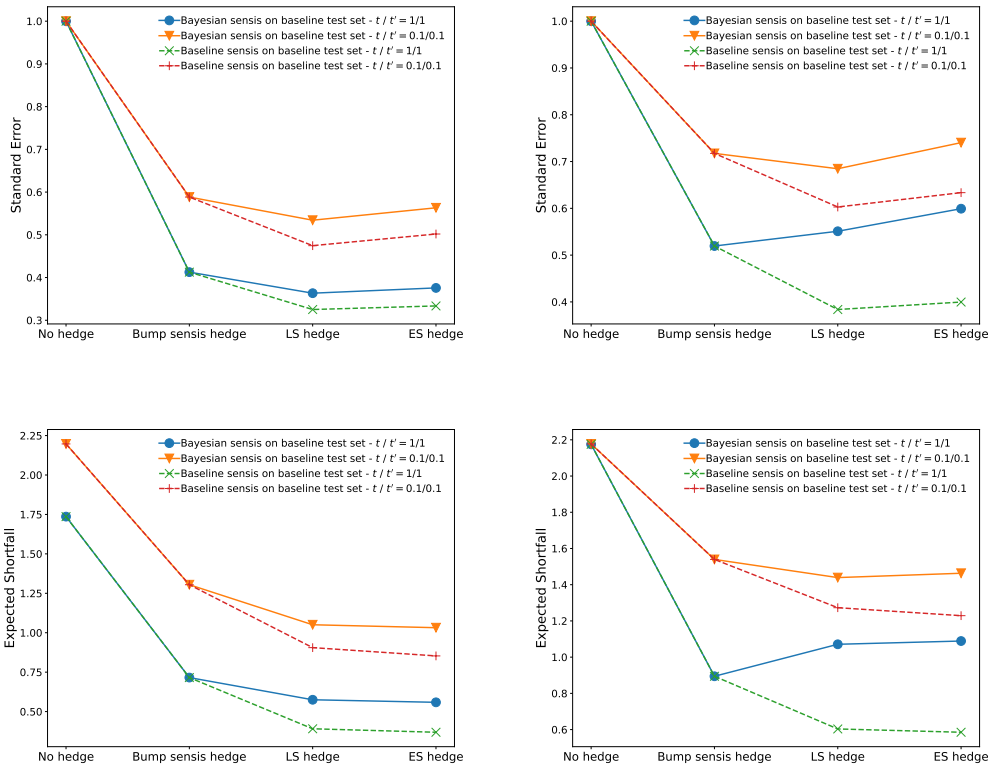

Figure 6: Hedging performance on the baseline test set: (*dashed*) hedging by baseline sensitivities, (*solid*) hedging by Bayesian-mixture sensitivities, (*left*) swap portfolio, (*right*) option portfolio.

Table 2 shows the the hedging performance assessed by plugging into the empirical test set (4) the Bayesian-mixture sensitivities Δ and Γ regressed by (11) or (12) (the corresponding $\mathcal{V}$ are also computed but they have no impact on (4)). By construction, the Bayesian regressed sensitivities cannot beat the baseline regressed sensitivities, which are estimated on a training set of the same distribution as the baseline test set. Still, the former are found to reduce significantly the standard error and expected shortfall of the loss. They even provide a better hedge than the

[16] a reduction from 77% to 48%
[17] A reduction from 79% to 59%.
[18] A reduction from 2.09 to 1.05 for the LS hedge and from 2.05 to 0.97 for the ES hedge

baseline bump sensitivities in most cases (except for the option portfolio in the 1yr case). As Table 1 demonstrates, they are also ways faster.

Portfolio	Train t/ Test t'	Standard Error				Expected Shortfall			
		No hedge	Bump sensis hedge	LS hedge	ES hedge	No hedge	Bump sensis hedge	LS hedge	ES hedge
swap portfolio	1/1	1.0	0.4127	0.3634	0.3757	1.7361	0.7157	0.575	0.5588
	0.1/0.1	1.0	0.5884	0.5342	0.5632	2.1971	1.3044	1.0505	1.0319
option portfolio	1/1	1.0	0.5195	0.5510	0.5992	2.1751	0.8944	1.071	1.0889
	0.1/0.1	1.0	0.7174	0.6846	0.7404	2.1773	1.5396	1.4393	1.4632

Table 2: Hedging performance on the baseline test set of the Bayesian-mixture Δ, Γ regressed sensitivities (estimated on the Bayesian-mixture training set).

Conclusion

Armed with the CVA simulation/regression engine of Abbas-Turki, Crépey, and Saadeddine (2022), one can not only price the CVA dynamically, but also hedge CVA diffusive fluctuations[19], on the basis of least-squares or expected shortfall regression sensitivities. Such CVA regression sensitivities are not only ways faster than industry standard bump sensitivities, but they are also found to provide a better hedge in practice. Our choice to benchmark the proposed approach on purely simulated data in this paper is for ease of implementation. To use it in production, one would need to have a large, real dataset, identify the financial products likely to calibrate each of the model parameters, and implement the corresponding calibration routines, as explained with more details in the different context of SIMM computations in Albanese, Caenazzo, and Syrkin (2017).

As opposed to bump sensitivities, our CVA regression sensitivities could easily be extended to account for transaction costs as in (Buehler, Gonon, Teichmann, and Wood, 2019) or Buehler (2019) (cf. also Burnett (2021); Burnett and Williams (2021)). Intermediate rebalancing hedging times could also be inserted in (4), in either setup (6)-(8) or (9), at a heavier simulation cost, in the direction of a deep hedging approach à la Buehler, Gonon, Teichmann, and Wood (2019).

As a final note, hedging model shifts, as Section 2.2 shows to be doable in the Bayesian setup of Section 1.2, refers to market practice such as vega hedging, which in our opinion does not necessarily qualify for "hedging model risk". In fact, the thesis developed in Albanese, Crépey, and Iabichino (2021) is that model risk cannot be hedged, but only compressed by upgrading the local models used by traders to a global fair valuation model that would be used consistently for all its purposes by the bank. A more systematic study of the connections between "hedging model shifts" and "hedging model risk" could be an interesting perspective of future research.

[19] we recall from Section 1 that, in line with market practice, we did not introduce hedging components for default losses in our study.

References

Abbas-Turki, L., S. Crépey, and B. Saadeddine (2022). Pathwise CVA regressions with oversimulated defaults. *Mathematical Finance*, 1–34. DOI: 10.1111/mafi.12368.

Albanese, C., S. Caenazzo, and S. Crépey (2016). Capital and funding. *Risk Magazine*, May 71–76.

Albanese, C., S. Caenazzo, and S. Crépey (2017). Credit, funding, margin, and capital valuation adjustments for bilateral portfolios. *Probability, Uncertainty and Quantitative Risk 2*(7), 26 pages.

Albanese, C., S. Caenazzo, and M. Syrkin (2017). Optimising VAR and terminating Arnie-VAR. *Risk Magazine*, October.

Albanese, C., S. Crépey, R. Hoskinson, and B. Saadeddine (2021). XVA analysis from the balance sheet. *Quantitative Finance 21*(1), 99–123.

Albanese, C., S. Crépey, and S. Iabichino (2021). A Darwinian theory of model risk. *Risk Magazine*, July pages 72–77.

Basel Committee on Banking Supervision (2013). Consultative document: Fundamental Review of the Trading Book: A revised market risk framework. http://www.bis.org/publ/bcbs265.pdf, accessed on Feb 1 2022.

Buehler, H. (2019). Statistical hedging. ssrn.2913250.

Buehler, H., L. Gonon, J. Teichmann, and B. Wood (2019). Deep hedging. *Quantitative Finance 19*(8), 1271–1291.

Burnett, B. (2021). Hedging value adjustment: Fact and friction. *Risk Magazine*, February 1–6.

Burnett, B. and I. Williams (2021). The cost of hedging XVA. *Risk Magazine*, April.

Crépey, S. (2022). Positive XVAs. *Frontiers of Mathematical Finance 1*(3), 425–465. doi: 10.3934/fmf.2022003.

Fissler, T., J. Ziegel, and T. Gneiting (2016). Expected Shortfall is jointly elicitable with Value at Risk—Implications for backtesting. *Risk Magazine*, January.

Golub, G. H. and C. F. Van Loan (2013). *Matrix computations*. The Johns Hopkins University Press.

Huge, B. N. and A. Savine (2017). LSM reloaded-differentiate xVA on your iPad mini. SSRN.2966155.

Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value-at-risk. *Journal of risk 2*, 21–42.

Savine, A. (2018). *Modern Computational Finance: AAD and Parallel Simulations*. Wiley.