Does initial margin eliminate counterparty risk?

It is widely believed that mandatory posting of initial margin should effectively eliminate counterparty risk from bilateral trading. Here, Leif Andersen, Michael Pykhtin and Alexander Sokol apply a new framework for collateralised exposure modelling that shows this is not necessarily true. The authors demonstrate that time lags in the credit support annex between trade payments and margin reposting can produce exposure spikes that substantially exceed VAR-based initial margin levels, then propose ways to mitigate the effect

Collateralisation has long been a standard technique of mitigating counterparty risk in over-the-counter bilateral trading. The most common collateral mechanism is variation margin (VM), which aims to keep the exposure gap between portfolio value and posted collateral below certain, possibly stochastic, thresholds. However, even when the thresholds for VM are set to zero, there remains residual exposure to the counterparty's default; this results from a sequence of contractual and operational time lags, starting at the last snapshot of the market for which the counterparty would post in full the required VM, and finishing at the termination date after the counterparty's default. The aggregation of these lags results in a time period, known as the margin period of risk (MPoR), during which the gap between the portfolio value and the collateral can widen.

The posting of initial margin (IM) to supplement VM provides banks with a mechanism to reduce the residual exposure resulting from market risk over the MPoR. Until very recently, IM in bilateral trading has mostly been reserved for bank counterparties deemed as high risk (eg, hedge funds), and it typically involves a trade-level independent amount set according to a simple deterministic schedule.

Recently, new uncleared margin rules (UMR) for bilateral trading have started to take effect, as laid out by the Basel Committee on Banking Supervision and International Organization of Securities Commissions in 2015 (BCBS-Iosco 2015). Under UMR, VM thresholds are forced to zero, and IM must be posted bilaterally (into segregated accounts) at the netting set level, either through an internal model or by look-up in a standardised schedule. If an internal model is used, the IM must be calculated dynamically as the netting set value-at-risk for a 99% confidence level. To reduce the potential for margin disputes, and to increase overall market transparency, the International Swaps and Derivatives Association has proposed a standardised internal model known as the standard initial margin model (Simm) (see Isda 2016). As a practical matter, it is expected that virtually all banks will use Simm for their day-to-day IM calculations.

In this article, we discuss credit exposure modelling in the presence of dynamic IM. Leaning on recent results in Andersen *et al* (2017), we start by formulating a general model for exposure in the presence of VM and IM. Applying the resulting framework to a simple case where no trade flows take place within the MPoR, we derive an asymptotic scaling factor for processes with Gaussian increments (eg, an Itô process) that converts IM-free expected exposure (EE) to IM-protected EE, for sufficiently small MPoR. The scaling factor depends only on the IM confidence level and the ratio of the IM horizon to the MPoR; it equals 0.85% at the BCBS-Iosco confidence level of 99%, provided that the IM horizon equals the MPoR. While some deviations from this universal limit value due to a non-infinitesimal MPoR are to be expected, the reduction of the EE by about two orders of magnitude is, as we demonstrate, generally about right when no trade flows are present within the MPoR.

For those periods where trade flows do take place within the MPoR, however, any scheduled trade payment flowing away from the bank will result in a spike in the EE profile. In the absence of IM, these spikes often constitute a moderate part of the overall credit valuation adjustment (CVA),[1] which mostly originates with the EE level between the spikes. We show that while IM is effective in suppressing EE between spikes, it will often fail to significantly suppress the spikes themselves. As a result, the relative contribution of the spikes to CVA is greatly increased in the presence of IM. For example, for a single interest rate swap, the spikes' contribution to CVA can be well above 90% for a position with IM. Accounting for spikes, we find that IM may reduce CVA by much less than the two orders of magnitude one might expect, with the reduction for interest rate swaps often being less than a factor of 10.

Exposure in the presence of IM and VM

Let us consider a bank B that has a portfolio of OTC derivatives contracts traded with a counterparty C, covered by a single netting agreement with VM and IM. Quite generally, exposure of B to default of C measured at time t (assumed to be the early termination time after C's default) is given by:

$$E(t) = (V(t) - \mathrm{VM}(t) + U(t) - \mathrm{IM}(t))^+ \tag{1}$$

where $V(t)$ is the time t portfolio value from B's perspective; $\mathrm{VM}(t)$ is the VM available to B at time t; $U(t)$ is the value of trade flows scheduled to be paid by both B (negative) and C (positive) up to time t, yet unpaid as of time t; and $\mathrm{IM}(t)$ is the value of IM available to B at time t. We further assume that B and C post VM with zero thresholds and are required to post BCBS-Iosco-compliant IM to a segregated account. We then deal with the modelling of each of the terms VM, U and IM in turn.

■ **Modelling VM.** The length of the MPoR, denoted by δ_C, defines the last portfolio valuation date $t_C = t - \delta_C$ prior to the termination date t (after C's default) for which C delivers VM to B. A common assumption, which we here denote the classical model, assumes that B stops posting VM to C at the exact same time C stops posting to B. That is, $\mathrm{VM}(t)$ in

[1] *Say, 20% of a vanilla interest rate swap's total CVA may originate with spikes; see, for instance, the first row of table A.*

(1) is the VM prescribed by the margin agreement for portfolio valuation date $t_C = t - \delta_C$. Ignoring minimum transfer amount and rounding, the prescribed VM in the classical model is thus simply:

$$\mathrm{VM}_{\mathrm{cl}}(t) = V(t_C) = V(t - \delta_C) \tag{2}$$

In the advanced model of Andersen *et al* (2017), operational aspects (and gamesmanship) of margin disputes are considered in more detail, leading to the more realistic assumption that B may continue to post VM to C for some period of time, even after C stops posting. The model introduces another time parameter $\delta_B \leqslant \delta_C$ that specifies the last portfolio valuation date $t_B = t - \delta_B$ for which B delivers VM to C. For portfolio valuation dates $T_i \in [t_C, t_B]$, B will post VM to C when the portfolio value decreases, but will receive no VM from C when the portfolio value increases. This results in a VM of:

$$\mathrm{VM}_{\mathrm{adv}}(t) = \min_{T_i \in [t_C, t_B]} V(T_i) \tag{3}$$

Equation (3), of course, reduces to (2) when one sets $\delta_B = \delta_C$.

Modelling U. In the most conventional version of the classical model, here denoted 'classical+', it is assumed that all trade flows are paid without incident by both B and C for the entire MPoR, up to and including the termination date (ie, in the time interval $(t_C, t]$). This assumption simply amounts to setting:

$$U_{\mathrm{cl}+}(t) = 0 \tag{4}$$

One of the prominent features of the classical+ model is that the time 0 expectation of $E(t)$, denoted $\mathrm{EE}(t)$, will contain upward spikes whenever there is a possibility of trade flows from B to C within the interval $(t_C, t]$. The spikes appear because, by the classical model's assumption, C makes no margin payments during the MPoR and consequently will fail to post an offsetting VM amount to B after B makes a trade payment to C. For banks calculating exposure on a sparse fixed exposure time grid, the alignment of grid nodes relative to trade flows will add numerical artifacts to genuine spikes, causing spikes in the EE profile to appear and disappear as the calendar date moves. As a consequence, an undesirable instability in EE and CVA is introduced.

An easy way to eliminate exposure spikes is to assume that neither B nor C makes any trade payment within the MPoR. The resulting model, here denoted 'classical−', consequently assumes that:

$$U_{\mathrm{cl}-}(t) = \mathrm{TF}^{\mathrm{net}}(t; (t_C, t]) \tag{5}$$

where $\mathrm{TF}^{\mathrm{net}}(t; (s, u])$ denotes the time t value of all net trade flows payable on the interval $(s, u]$. The classical− model avoids the EE and CVA instabilities in the classical+ model, and it is used by a significant number of banks.

It should be evident that neither classical+ nor classical− assumptions on trade flows are entirely realistic. In the beginning of the MPoR both B and C are likely to make trade payments, while at the end of the MPoR neither B nor C are likely to make trade payments. To capture this effect, the advanced model (Andersen *et al* 2017) adds two more time parameters, δ'_C and $\delta'_B \leqslant \delta'_C$: these specify the last dates, $t'_C = t - \delta'_C$ for C and $t'_B = t - \delta'_B$ for B, at which trade payments are made prior to closeout at t. This results in an unpaid trade flows term of:

$$U_{\mathrm{adv}}(t) = \mathrm{TF}^{C \to B}(t; (t'_C, t'_B]) + \mathrm{TF}^{\mathrm{net}}(t; (t'_B, t]) \tag{6}$$

where an arrow indicates the direction of the trade flows and $C \to B$ ($B \to C$) trade flows have a positive (negative) sign.

Modelling IM. We model IM at a netting set level assuming BCBS-Iosco specifications. Following the BCBS-Iosco restrictions on diversification, we define IM for a netting set as the sum of IMs over K asset classes:

$$\mathrm{IM}(t) = \sum_{k=1}^{K} \mathrm{IM}_k(t) \tag{7}$$

Let $V_k(t)$ denote the value at time t of all trades in the netting set that belong to asset class k and are subject to IM requirements. For an asset class k, we define IM as the quantile at confidence level q of the 'clean' portfolio value increment over the BCBS-Iosco IM horizon δ_{IM} (which may or may not coincide with δ_C), conditional on all the information available at time t_C. That is:

$$\begin{aligned}\mathrm{IM}_k(t) = \mathrm{Q}_q(&V_k(t_C + \delta_{\mathrm{IM}}) \\ &+ \mathrm{TF}_k^{\mathrm{net}}(t_C + \delta_{\mathrm{IM}}; (t_C, t_C + \delta_{\mathrm{IM}}]) - V_k(t_C) \mid \mathcal{F}_{t_C})\end{aligned} \tag{8}$$

where $\mathrm{Q}_q(\cdot \mid \mathcal{F}_s)$ denotes the quantile at confidence level q, conditional on information available at time s. Note (8) assumes that C stops posting IM at the same portfolio observation date (t_C) for which it stops posting VM; hence, IM is calculated as of $t_C = t - \delta_C$.

Summary and calibration. To summarise, we have outlined three different models for the generic exposure calculation (1): classical+, classical− and advanced. Collecting the results, we have:

$$E_{\mathrm{cl}+}(t) = (V(t) - V(t - \delta_C) - \mathrm{IM}(t))^+ \tag{9}$$

$$E_{\mathrm{cl}-}(t) = (V(t) - V(t - \delta_C) + \mathrm{TF}^{\mathrm{net}}(t; (t_C, t]) - \mathrm{IM}(t))^+ \tag{10}$$

$$\begin{aligned}E_{\mathrm{adv}}(t) = \Big(V(t) - &\min_{T_i \in [t_C, t_B]} V(T_i) + \mathrm{TF}^{C \to B}(t; (t'_C, t'_B]) \\ &+ \mathrm{TF}^{\mathrm{net}}(t; (t'_B, t]) - \mathrm{IM}(t)\Big)^+\end{aligned} \tag{11}$$

where for all three models $\mathrm{IM}(t)$ is computed as in (7) and (8).

In practice, the calibration of the time parameters of the advanced model should be informed by both the bank's legal rights and its aggressiveness in pursuing them. For the former, we refer to the detailed discussion in Andersen *et al* (2017) of the Isda Master Agreement and the 'suspension rights' in its Section 2(a)(iii).

With four time parameters, the advanced model allows a bank great flexibility in modelling its risk tolerance and choosing procedures for exercising its suspension rights. A bank may, in fact, calibrate these parameters differently for different counterparties, to reflect its risk management practices towards counterparties of a given type.[2] Additional discussion can be found in Andersen *et al* (2017), which also provides some prototypical parameter settings: the aggressive calibration (B can always sniff out financial distress in its counterparties and is swift and aggressive in enforcing its legal rights) and the conservative calibration (B is deliberate and cautious in enforcing its rights, and acknowledges the potential for operational errors and for rapid, unpredictable deterioration of counterparty credit). For the numerical results in this article, we set the values of the time parameters to be between aggressive and conservative.

[2] *One may also make the time parameters stochastic, eg, by making the lags a function of exposure magnitude. This way, one can, say, model the fact that a bank may tighten its operational controls when exposures are high.*

The impact of IM: no trade flows within the MPoR

Here, we examine the impact of IM on EE when there are no trade flows within the MPoR. For simplicity, we work with the classical model, and assume the entire netting set is covered by IM and comprised of trades all belonging to the same asset class.

In the absence of trade flows on $(t_C, t]$, (9) and (10) show the classical model computes expected exposure as:

$$\mathrm{EE}(t) = \mathrm{E}[(V(t) - V(t_C) - \mathrm{Q}_q(V_k(t_C + \delta_{\mathrm{IM}}) \mid \mathcal{F}_{t_C}))^+] \tag{12}$$

where $\mathrm{E}[\cdot]$ is the expectation operator. In the absence of IM, this expression would be $\mathrm{EE}_0(t) = \mathrm{E}[(V(t) - V(t_C))^+]$; we are interested in establishing meaningful estimates of the IM 'efficiency ratio':

$$\lambda(t) \triangleq \frac{\mathrm{EE}(t)}{\mathrm{EE}_0(t)} \tag{13}$$

Suppose the portfolio value $V(t)$ follows an Itô process:

$$\mathrm{d}V(t) = \mu(t)\,\mathrm{d}t + s(t)^\top\,\mathrm{d}W(t)$$

where $W(t)$ is a vector of independent Brownian motions, and μ and $s(t)$ are well-behaved processes (with $s(t)$ being vector-valued) adapted to $W(t)$. Note that both μ and s may depend on the evolution of multiple risk factors prior to time t. For convenience, define $\sigma(t) = |s(t)|$. Then, for some sufficiently small horizon δ, the increment of portfolio value over $[t_C, t_C + \delta]$, conditional on $\mathcal{F}_{t_C}$, is well approximated by a Gaussian distribution with mean $\mu(t_C)\delta$ and standard deviation $\sigma(t_C)\sqrt{\delta}$.

Assuming $\sigma(t_C) > 0$, we may ignore the drift term for small δ. Under the Gaussian approximation above, it is then straightforward to approximate the expectation in (12) in closed form:

$$\mathrm{EE}(t) \approx \mathrm{E}[\sigma(t_C)]\sqrt{\delta_C}[\phi(z(q)) - z(q)\Phi(-z(q))]$$

where $z(q) \triangleq \sqrt{\delta_{\mathrm{IM}}/\delta_C}\,\Phi^{-1}(q)$, and ϕ and Φ are the standard Gaussian probability density function and a cumulative distribution function, respectively. Similarly, $\mathrm{EE}_0(t) \approx \mathrm{E}[\sigma(t_C)]\sqrt{\delta_C}\,\phi(0)$, so that λ in (13) is approximated by:

$$\lambda(t) \approx \frac{\phi(z(q)) - z(q)\Phi(-z(q))}{\phi(0)} \tag{14}$$

Interestingly, the multiplier in (14) is independent of t and depends only on two quantities: the confidence level q used for specifying IM, and the ratio of the IM horizon to the MPoR. For the value of the confidence level $q = 99\%$ specified by the BCBS-Iosco framework, (14) results in a value of $\lambda = 0.85\%$ when $\delta_{\mathrm{IM}} = \delta_C$, ie, IM is expected to reduce expected exposure by a factor of 117.

We emphasise that (14) was derived under weak assumptions, relying only on a local normality assumption for the portfolio value. A special case for which (14) becomes exact is when the portfolio process follows a Brownian motion, a result derived in Gregory (2015). More generally, however, we have shown that (14) constitutes a small δ limit, and a useful approximation, for a much broader class of processes. In fact, we can extend to jump-diffusion processes, provided that portfolio jumps are (approximately) Gaussian. The analysis in Andersen *et al* (2016a) suggests that the local Gaussian approximation is quite robust for a 10-day horizon,[3] lending support to the notion that introducing IM at the level of $q = 99\%$ should generally result in about two orders of magnitude reduction of EE, when no trade flows occur within the MPoR.

The impact of IM: trade flows within the MPoR

In what follows, we consider the efficacy of IM in the more complicated case where trade flows are scheduled to take place within the MPoR. In practice, many portfolios produce trade flows nearly every business day, so this case is of considerable relevance.

Unless one operates under the classical− model assumptions of no trade flows paid by either B or C within the MPoR, any possibility of B making a trade payment to C in the future results in a spike in the EE profile. These spikes appear because the portfolio value will jump following B's payment, but C would fail to post or return VM associated with this jump.

As an example, let us consider the impact of IM on the EE profile for a two-year interest rate swap, where B pays the fixed rate semiannually and receives quarterly floating payments.[4] We assume a flat initial interest rate of 2% under lognormal dynamics with volatility 50%. Panel (a) in figure 1 shows EE profiles for the classical± and advanced models under a two-way credit support annex (CSA), with zero-threshold daily VM but without IM. As expected, both the classical+ and advanced models produce spikes around the payment dates: the spikes are upwards at the semi-annual dates (when B makes a payment) and downwards at the quarterly dates between the semi-annual dates (when B does not make a payment). In panel (b), we add dynamic IM (defined by UMR in (8) and calculated exactly by taking advantage of the single-factor dynamics) to this swap and show the resulting EE profile on the same scale as the no-IM result in panel (a). In the areas between EE spikes, IM reduces the EE by a factor of approximately 100,[5] so no finite exposure is visible on the graph's scale. However, the spikes are barely reduced by IM, as the level of IM is calculated without the effect of trade flows, and the height of the spikes greatly exceeds the 'baseline' level of diffusion-driven exposure to which IM is calibrated according to (8). Note that the degree of reduction of the spikes decreases with the simulation time: the IM amount on all simulation paths shrinks over time due to the amortisation effect, while the size of the swap payments remains roughly the same.

To demonstrate the impact of IM on CVA for the swap example above, table A shows CVA calculated from the EE profiles in panels (a) and (b)

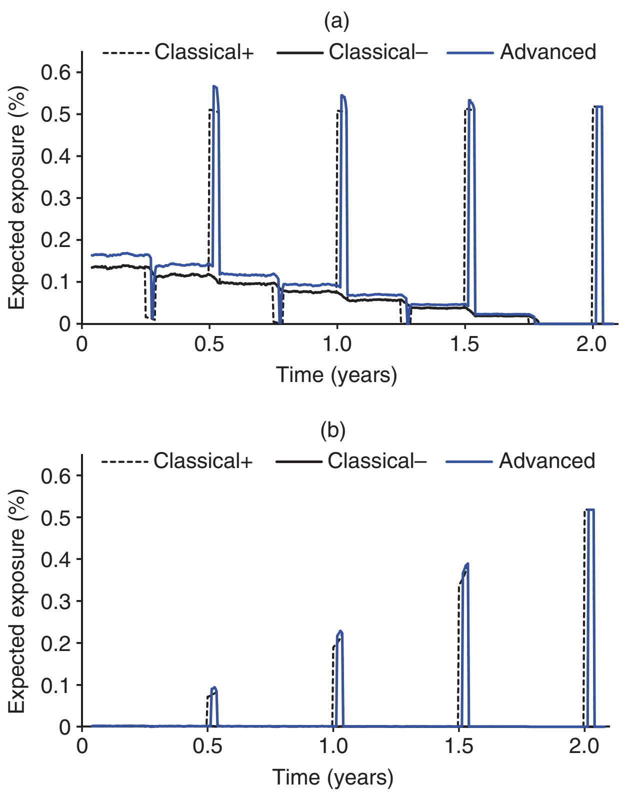
---

[3] *The accuracy of the local Gaussian approximation was checked in Andersen* et al *(2016a) by calculating λ analytically for a lognormal portfolio value. At an IM horizon equal to* MPoR = 10 *business days, deviations of λ from the local Gaussian value of* 0.85% *are noticeable and grow in the magnitude of the lognormal volatility (eg, if the volatility is equal to* 50%*, one computes values of $\lambda = 1.10\%$ for a long portfolio and $\lambda = 0.65\%$ for a short portfolio). While these deviations are certainly significant in a relative sense, they do not change the two-orders-of-magnitude scale of the EE reduction.*

[4] *Results for a swap with identical floating and fixed frequencies can be found in Andersen* et al *(2016a).*

[5] *The reduction factor is slightly less than the local normal approximation predicts. This is due to our use of lognormal interest rates, which violates the local normality assumption at finite horizons (see Andersen* et al *(2016a) for details).*

1 Expected exposure profiles for a two-year swap



EE profiles of a two-year interest rate swap, as percentage of swap notional. B pays 2% fixed (semi-annually), receives floating (quarterly). Interest rate curve flat at 2% quarterly compounded, with 50% lognormal volatility. CSA: zero-threshold VM, no MTA, no rounding, daily posting. MPoR assumptions: $\delta_C = 10$ business days, $\delta_B = 8$ business days, $\delta'_C = 6$ business days, $\delta'_B = 4$ business days. Exposure computed from 5,000 Monte Carlo simulations on a daily simulation grid. (a) Without IM. Note that the classical− EE profile has no spikes; away from spikes, the classical+ and classical− EE profiles perfectly coincide. (b) With IM, at the confidence level $q = 99\%$ for $\delta_C = \delta_{\mathrm{IM}} = 10$ business days

A. CVA of a two-year swap

Quantity	Classical−	Classical+	Advanced
CVA (no IM)	0.267	0.391	0.408
CVA (with IM)	0.003	0.094	0.061
Ratio	1.05%	24.0%	14.8%

The impact of IM on CVA under different models of collateralised exposure. All CVA values are given relative to the CVA value for the case of no VM and no IM. The third row lists the ratio of CVA with IM to CVA without IM. For CVA calculations, the recovery rate was 50% and the credit intensity was 150 basis points. Exposure profiles for the CVA calculations were imported from figure 1.

of figure 1.[6] Since the classical− model does not have spikes, CVA in this model is reduced by two orders of magnitude by IM, as one would expect from (14). The presence of spikes, however, reduces the effectiveness of IM significantly: the CVA with IM is about 24% of the CVA without IM for the classical+ model, and 15% for the advanced model.

[6] *To show the impact of VM on uncollateralised CVA, all CVA numbers are shown relative to the CVA of an otherwise identical uncollateralised swap.*

Overall, our swap example demonstrates that when trade flows within the MPoR are properly modelled, IM at the 99% VAR level may not be sufficient to achieve even a one-order-of-magnitude reduction of CVA. Also, since spikes dominate CVA in the presence of IM, accurately modelling the trade flows paid within the MPoR is especially important when IM is present. Here, the advanced model is clearly preferable, as neither the classical+ nor the classical− model produce reasonable CVA numbers for portfolios with IM.

Finally, we should note that even if one uses only the advanced model, the impact of IM on CVA may vary significantly, depending on the trade/portfolio details and the model calibration. It is well beyond the scope (and space constraints) of the current article to pursue this topic in detail, but we can make the following general observations.

- *Payment frequency.* A higher frequency of trade payments results in more EE spikes, thus reducing the effectiveness of IM.
- *Payment size.* A higher payment size relative to trade/portfolio value volatility results in higher EE spikes, thus reducing the effectiveness of IM.
- *Payment asymmetry.* EE spikes are especially high for payment dates where only B pays, or where B is almost always the net payer. The presence of such payment dates reduces the effectiveness of IM.
- *Model calibration.* In the advanced model, the width of the spikes is determined by the time interval within the MPoR where B makes trade payments, ie, $\delta_C - \delta'_B$. Thus, larger δ_C (ie, the MPoR) and/or smaller δ'_B (ie, the time interval within the MPoR when B does not make trade payments) would result in wider spikes and, thus, reduced effectiveness of IM.

Numerical techniques

Daily time grid. The results described above make evident the importance of accurately capturing exposure spikes from trade flows with a daily resolution: something that, if done by brute-force methods, will likely not be feasible for large portfolios. In Andersen *et al* (2017), we discuss a fast approximation that produces a reasonably accurate EE profile without a significant increase of computation time relative to standard (coarse grid) calculations. This method requires the simulation of risk factors and trade flows with a daily resolution, but the portfolio valuations (which are normally the slowest part of the simulation process) can be computed on a much coarser time grid. Portfolio values on the daily grid are then obtained by Brownian bridge interpolation between the values at the coarse grid points, with careful accounting for trade flows. We refer the reader to Andersen *et al* (2017) for a detailed description of the numerical implementation.

Calculation of pathwise IM. As per (7) and (8), the calculation of IM requires dynamic knowledge of the distribution of portfolio value increments ('P&L') across K distinct asset classes. Since the conditional distributions of P&L are generally not known, one must rely on numerical methods to calculate IM. In Andersen & Pykhtin (2015) and Andersen *et al* (2016a), we discuss various regression techniques suitable for these calculations. For each asset class, a reasonable and practical approach is to base the regression on a single regression variable: the portfolio value on a given simulation path at the beginning of the MPoR. Either parametric or kernel regression can be used to calculate the standard deviation of the P&L conditional on a path. The IM amount on a path can then be calculated by applying the local normality assumption.

We should note that if a bank uses an out-of-model margin calculator (eg, the Simm method in Isda (2016)), an adjustment will be needed to capture the difference between the in-model IM computed (as above) by regression methods and the out-of-model margin calculator actually used.[7]

■ **Calculation of pathwise exposure.** Once IM on a path is calculated,[8] exposure on that path can then, in principle, be obtained directly by subtracting the calculated IM value from the no-IM exposure. This approach works well when the trades covered by IM represent a reasonably small fraction of the overall counterparty portfolio. However, when the netting set is dominated by IM-covered trades, this approach suffers from two issues that have an impact on the accuracy of EE (and, therefore, CVA) calculations: excessive simulation noise and potentially significant errors resulting from the local Gaussian assumption in calculation of pathwise IM.

Both of these issues can, as we shall see, be remedied by calculating the time t_C pathwise expected exposure for a time t default, rather than the exposure itself.[9] By assuming local normality for the P&L and that trade flows within the MPoR on a path are known at the beginning of the MPoR, it is shown in Andersen *et al* (2016a) that the time t_C EE on a path can be calculated as:

$$\mathrm{EE}^{(m)}(t) = \sigma^{(m)}(t_C)\sqrt{\delta_C}\,(d^{(m)}(t)\Phi(d^{(m)}(t)) + \phi(d^{(m)}(t))) \quad (15)$$

$$d^{(m)}(t) \triangleq \frac{-\mathrm{PTF}^{(m)}_{\mathrm{adv}}(t,(t_C,t]) - \sum \mathrm{IM}^{(m)}_k(t)}{\sigma^{(m)}(t_C)\sqrt{\delta_C}}$$

where (omitting arguments):

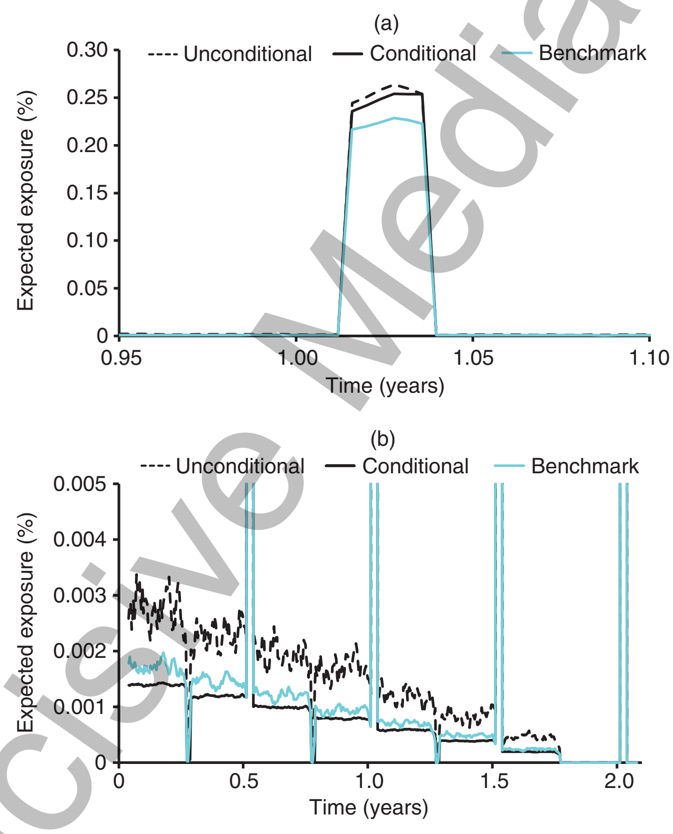
$$\mathrm{PTF}^{(m)}_{\mathrm{adv}} = \mathrm{TF}^{(m)} - U^{(m)}_{\mathrm{adv}}$$

are the net trade flows scheduled within the MPoR that are actually paid, according to the advanced model. Note that (15) was derived for the simplified case of the advanced model with $\delta_B = \delta_C$. In Andersen *et al* (2016a), we show how this result can be extended to the more general case $\delta_B < \delta_C$.

To illustrate the benefits of the proposed conditional EE simulation method, we turn to the two-year interest rate swap considered earlier. EE profiles from several computational approaches are shown in figure 2.

In figure 2(a), our attention is focused on the upward EE spike produced by trade payments at the one-year point. Both unconditional and conditional EE estimators here produce an almost identical spike, slightly exceeding the benchmark spike height; this is a consequence of using kernel regression and a Gaussian distribution to estimate IM. Part (b) of the figure shows the EE profiles at a fine exposure scale, allowing for clear observation of EE between the spikes. The advantages of the conditional EE approach can then be seen clearly:

2 Expected exposure profiles for a two-year swap



EE profiles of a two-year interest rate swap in the advanced model, as percentage of swap notional. Swap terms, CSA and model setting as in figure 1. The graphs contain three different exposure computations: 'Unconditional', direct computation with 5,000 daily simulation paths, and the IM on each path computed by kernel regression; 'Conditional', (15) with 5,000 daily simulation paths, and the IM on each path computed by kernel regression; and 'Benchmark', with 50,000 simulations and IM computed exactly, using the same technique as in figure 1. (a) Focus on the one-year spike. (b) Focus on exposure between spikes

■ *Reduction in simulation noise.* While both methods used Monte Carlo simulation with 5,000 paths, the simulation noise in the conditional EE approach is substantially less than that in the unconditional EE approach. In fact, the conditional EE noise is even less than in the benchmark EE results that were calculated with 50,000 paths.

■ *Reduced error from non-Gaussian dynamics.* We used a high-volatility lognormal interest rate model in our example to produce significant deviations from local normality over the 10-day horizon. This non-normality is the main reason for the deviations of the conditional and unconditional EE curves from the benchmark. In estimating (12), the conditional estimator uses a Gaussian distribution to approximate both IM and the portfolio increment, resulting in partial error cancellation. Such error cancellation does not apply to the unconditional estimator, which uses the empirical distribution for the portfolio increment yet estimates IM from a Gaussian distribution. Therefore, between spikes, the EE errors for the unconditional estimator are here significantly larger than for the conditional estimator.

---

[7] *Many possible methods could be contemplated here, eg, the common idea of using a multiplicative adjustment factor that aligns the two margin calculations at time* 0.

[8] *Note that pathwise IM results allow for straightforward computation of margin valuation adjustments (MVAs) (see, for example, Andersen* et al *2016b).*

[9] *If our target exposure measure is the unconditional (time* 0*) EE, this substitution is, of course, valid by the law of iterated expectations.*

Conclusion

The new BCBS-Iosco IM rules will lead to substantial margin postings into segregated accounts. These will be accompanied by inevitable increases in the funding costs (MVA) banks will face when raising funds for IM. According to conventional wisdom, these postings, while expensive, should effectively eliminate counterparty credit risk.

In this article, we have examined the degree to which bilateral IM required by the BCBS-Iosco margin rules suppresses counterparty exposure. As we have shown in Andersen *et al* (2017), any trade flow to the defaulting party for which it does not return margin during MPoR causes a spike in exposure profile. These spikes are often ignored by banks as 'spurious' or as being part of 'settlement risk'. In reality, these spikes are an integral part of the exposure profile and represent real risk that has previously materialised in many well-documented incidents, notably the non-payment by Lehman of reciprocal margin to trade payments that arrived around the time of the bankruptcy filing.

We have shown that, under very general assumptions, the BCBS-Iosco IM specified as the 99% 10-day VAR reduces exposure between the spikes by a factor of more than 100, but it fails to suppress the spikes to a comparable degree. This happens because IM is calculated without reference to trade payments, and it is based only on changes of the portfolio value resulting from risk factor variability. As an example, we showed that IM reduces the CVA of a two-year interest rate swap with VM by only a factor of around 7. While VAR-based IM fails to fully suppress the contribution of exposure spikes to CVA and EAD, increasing IM to always exceed peak exposure would be impractical, and would require moving large amounts of collateral back and forth in a matter of days.

Another important property of CVA under full IM coverage is that it is dominated by exposure spikes: in our two-year swap example, spikes' contribution to CVA is about 95% in the presence of IM (compared with about 20% without IM). Thus, in the presence of IM, the focus of exposure modelling should be on capturing the impact of trade payments, which involves making realistic assumptions on what payments the bank and the counterparty are expected to make contingent on the counterparty's default. Furthermore, to accurately calculate CVA that is mostly produced by narrow exposure spikes, one needs to produce exposure on a daily time grid. A method for producing daily exposure without daily portfolio revaluations was discussed above, along with other useful numerical techniques.

A natural question to ask is why similar payment effects have not been recognised in trading through central counterparties (CCPs), which also require IM posting that is typically based on 99% VAR over the MPoR. As it turns out, CCPs already use a mechanism that amounts to netting of trade and margin payments; unfortunately, the same approach cannot be adopted in bilateral trading, as it would require changing all of the existing trade documentation, a practical impossibility.

While a trade payment and a reciprocal lagged margin payment cannot be netted in bilateral trading, the lag can be eliminated and two payments made to fall on the same day by making a simple change in the CSA. Specifically, if the CSA is amended to state that known trade payments due to arrive prior to the scheduled margin payment date must be subtracted from portfolio valuation for the purposes of margin (technically, this amendment effectively sets VM based on a two-day portfolio forward value), then the call for reciprocal margin will happen ahead of time, and it will arrive on the same day as the trade payment (a 'no lag margin settlement').

From an IT and back-office perspective, this change in the CSA is relatively easy to align with existing mark-to-market and cashflow processes, and is beneficial in several ways. First, it shortens the duration of exposure spikes and MPoR overall, reducing counterparty risk. Second, it makes margin follow mark-to-market without a two-day lag, thereby eliminating the need to use outside funds to fund hedging during this two-day period. Finally, with reciprocal trade and margin payments falling on the same day, payment-versus-payment services such as CLS Bank (Galati 2002) may be able to settle trade and margin payments together, reducing residual counterparty risk even further. ■

Leif Andersen is the New York-based global co-head of the Quantitative Strategies Group at Bank of America Merrill Lynch. Michael Pykhtin is a manager in the quantitative risk team at the Federal Reserve Board in Washington, DC. Alexander Sokol is the New York-based head of quant research at CompatibL. The opinions expressed in this paper are those of the authors, and do not necessarily represent the views of their employers.
Email: leif.andersen@baml.com,
michael.v.pykhtin@frb.gov,
sokol@compatibl.com.

REFERENCES

Andersen L and M Pykhtin, 2015
Accounting for dynamic initial margin in credit exposure models
Presentation at RiskMinds International 2015

Andersen L, M Pykhtin and A Sokol, 2016a
Credit exposure in the presence of initial margin
Working Paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2806156

Andersen L, D Duffie and Y Song, 2016b
Funding value adjustment
Working Paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746010

Andersen L, M Pykhtin and A Sokol, 2017
Rethinking the margin period of risk
Journal of Credit Risk 13(1), pages 1–45

Basel Committee on Banking Supervision and International Organization of Securities Commissions, 2015
Margin requirements for non-centrally cleared derivatives
BCBS-Iosco, March

Galati G, 2002
Settlement risk in foreign exchange markets and CLS Bank
BIS Quarterly Review December, pages 55–66 (available at www.bis.org/publ/qtrpdf/r_qt0212f.pdf)

Gregory J, 2015
The XVA Challenge: Counterparty Credit Risk, Funding, Collateral, and Capital
Wiley (3rd edition)

International Swap and Derivatives Association, 2016
ISDA SIMM methodology
Isda, April (version 3.15)