The European Banking Authority

20 Avenue André Prothin

92400 Courbevoie

France

Subject: Response to the public "Consultation on draft RTS on the assessment methodology under which competent authorities verify an institution's compliance with the internal model approach" (EBA/CP/2023/04)

We are grateful to the European Banking Authority for opening to the public the debate around validation of internal models (IMA) in the Fundamental Review of the Trading Book (FRTB) rule.

We respond only to a single question, Q9, which we consider particularly important and falls in our area of expertise.

Motivations and background

Since its first publication, in May 2012, the FRTB was met with a serious model validation challenge: the replacement of 1% Value at Risk (VaR) with 2.5% Expected Shortfall (ES) as the new standard for capital adequacy, although welcomed by the academic community and motivated by strong theoretical arguments (... *"ES, a risk measure that better captures tail risk"*), immediately faced a major difficulty: no established ES backtest was known at the time.

The Basel Committee on Banking Supervision (BCBS) nonetheless took the decision to adopt ES as the new capital charge risk measure and chose to validate internal models by testing something else, namely by a mixture of a) backtesting VaR at 1% and 2.5% and b) by the so-called "P&L Attribution Test" (PLAT)[1]. This validation scheme is still unchanged today, as set forth in the final 2019 FRTB rule.

The goal of the backtest for a capital standard for prudential regulation is to ensure that the predictions of a bank's risk model do not underestimate the actual risk that the bank is running, as observed from realized losses. This has

[1] The PLAT is a test meant to verify the close alignment between the front and the risk pricing libraries, that produce the Hypothetical P&L (HPL) and the Risk-Theoretical P&L (RTPL), respectively. It was reformulated in March 2018 in terms of Spearman Correlation and Kolmogorov-Smirnov test.

been the role well-covered by the VaR backtest from Basel 1 throughout Basel 2.5, when VaR was the designated capital standard.

The problem with the current, ES-based, FRTB framework is that passing the proposed tests is no guarantee that ES predictions are correct, namely sufficient; nor failing these tests is necessarily a sign that ES predictions are wrong, namely underestimated. In other words, the FRTB IMA validation scheme is intrinsically prone to Type II errors (a false negative: acceptance of a wrong model) and Type I errors (a false positive: rejection of a good model) simply because what you test is not what you should test.

The PLAT, in particular, has nothing to do with assessing the quality of ES predictions. In fact, it has nothing to do with assessing any type of risk predictions: it is not a comparison between ex-ante risk predictions (of whatever sort) and ex-post loss realizations, but – by design – one between *ex-post* loss realizations as measured by the front office systems and (again) *ex-post* loss realizations as measured by the risk systems. As such, it is simply a wrong question when it comes to validating the adequacy of a model's risk predictions. Combined with the fact that it is hard and erratic to pass, and that it can fail for multiple insignificant reasons, the PLAT is often seen to be responsible for false positives, hindering the stability of IMA projects and distracting huge resources (qualified workforce and systems) from the real questions that a risk department should be focusing on.

The ES backtestability puzzle

The question as to whether ES is backtestable at all, has been the subject of intense and somehow controversial research over the past 20 years. The good news is that it has now come to conclusive results.

That backtesting ES is at least less easy than VaR, was clear long ago already, given the sporadic and unconclusive literature on the subject, but it was only in 2011 that a solid mathematical result was obtained (Gneiting, 2011) showing that, indeed, the possibility of testing ES is faced with fundamental limitations.

Gneiting's mathematical proof that "ES is not elicitable", means, in layman's terms, that *no pure backtest of ES exists*, where ES predictions alone can be directly checked against loss realizations[2]: any attempt to backtest ES will inevitably also carry some spurious sensitivity (an inevitable, unobservable bias) to some other metrics predicted by the risk model. In such a situation, the bias, generally two-sided, could tilt the ES backtest toward either undue acceptance or undue rejection of the model, the former being a serious obstacle for the adoption of ES backtests in the context of prudential regulation.

[2] This is what the good old VaR backtest does: for counting VaR exceptions, you just need to know the sequence of VaR predictions and corresponding loss realizations: VaR is in fact elicitable.

Another important piece of the puzzle was however the discovery that there exists a unique ES backtest (the *"ridge backtest"*) introduced in (Acerbi & Szekely, 2017) (Acerbi & Szekely, 2019), for which, the unavoidable bias is

- One-sided, and always in the prudential direction, hence never the cause of acceptance of wrong models
- Minimal, because vanishing at first-order

These are model-independent mathematical results, hence valid under any circumstance.

The bias of the ridge ES backtest is linked to the quality of VaR predictions: it is zero when these are correct, and very small otherwise, in which case it tilts the ES backtest slightly in the prudential direction.

The ridge backtest is therefore the unique backtest for ES that permits drawing prudential conclusions. All other (published or conceivable) ES backtests feature a bias that may lead to Type II errors and are therefore not suited for tests of ES for prudential regulation. The point, unfortunately, is omitted in many papers that propose new ES backtests[3].

Conclusive results on ES backtesting

These results represent the point of arrival of decade-long research on ES backtesting. We would hesitate to be so self-referential, if it weren't that these are incontrovertible mathematical results, and not personal opinions.

We summarize:

- ES is not elicitable (Gneiting, 2011), and strictly speaking not backtestable either (Acerbi & Szekely, 2017). All conceivable backtests for ES necessarily suffer an uncontrollable bias relative to mispredictions of some other metrics (Acerbi & Szekely, 2017).
- There exists however a unique ES backtest (the "ridge backtest" (Acerbi & Szekely, 2019)) for which the bias is muted and one-sided, under any distributional assumption, and the side is of prudential nature. The bias of all other ES backtests is non-negligible and may induce Type II errors, making them unsuited for prudential regulation.

In addition, the ES ridge backtest provides not only probabilistic estimates (p-values) for model acceptance, but also monetary discrepancy measures, between the predicted ES and the actual, realized ES of the portfolio (Acerbi & Szekely, 2019). For this reason, the output of the ridge backtest is a rich dashboard for model rectification, and not only a mere RAG indicator for model validation, like the usual VaR backtest.

For all these reasons, the ridge backtest is rapidly being adopted by multiple primary banks, regardless of its possible future adoption in banking regulation. It

[3] A critical review of the ES backtesting literature is provided in (Acerbi & Szekely, 2023)

represents the most advanced possible validation methodology for ES-based risk models.

In the light of the above considerations, we now provide our:

Response to Q9:

Q9. What are your views in relation to the assessment method to verify that the internal validation process includes a direct back-testing of the expected shortfall, as per Article 21(b)? Do you expect this requirement to put significant burden on institutions? Which of the methods available in the literature do you expect credit institutions to use to back-test their expected shortfall? Please elaborate.

We think that the EBA proposal to introduce ES backtesting methodologies in the model validation framework for FRTB IMA, represents an opportunity for a fundamental advance in model risk methodologies for banking regulation and supervision. We applaud the EBA, first among financial regulators for such a proposal, which denotes foresight.

We agree with the sentence at point (12) pag. 24 and its motivations *"It is therefore appropriate to include a requirement for institutions to directly back-test their expected shortfall measures as part of the internal back-testing programme required in Article 325bj of that Regulation*".

We question on the contrary what is stated immediately after *"As there is not yet an established methodology among market participants for back-testing an expected shortfall measure, …".* As discussed in this document, the research on ES backtesting has now reached conclusive results: there are some known limitations on the extent to which ES can be backtested, and there exists a unique ES backtest methodology (the ridge backtest) that is suited for prudential regulation purposes. The message that there doesn't exist an established methodology is today outdated and gives the wrong impression that this subject is still prone to controversies as it was years ago, generating confusion among institutions and supervisors. Reference to the conclusive results on this subject should be made, to avoid perpetuating this misconception.

The introduction of ES backtesting methodologies certainly puts some additional burden on banks, for the simple reason that it requires new procedures, new algorithms, new knowhow, and new data collection. But apart from the novelty aspect, these methodologies (at least in the case of the ridge backtest), per se, are not burdensome at all, and don't involve major investments (of workforce, systems and processes) as compared with standard risk methodologies necessary to construct and validate a bank risk model.

The EBA limits to propose ES backtesting *in addition to* the FRTB validation tests, and for internal model validation purposes, and not *in replacement of* the FRTB validation tests, because this would constitute a departure from the international

standards set forth by the BCBS in the FRTB. But the ultimate goal, that certainly some institutions were hoping for, would be to adopt a single, all-encompassing ES test, instead of the current proposed tests, in particular in the light of the difficulties encountered by many banks with the PLAT.

The introduction of ES backtesting, even if just "in addition to", would in any case represent a first fundamental step in the right direction: it would provide banks with new, more advanced model validation tools and knowhow and it would permit regulators to collect new data and evidence from banks, which are necessary for studying these methodologies on real data, in view of a possible more ambitious cross-jurisdictional reform in the future.

For this reason, the relatively small additional burden that is put on banks today, may be in fact a good investment for reducing, one day, the much heavier burden created by the ill-designed FRTB tests and the inconsistencies and problems they generate.

We hope that the decisions on this consultation will be driven by farsightedness, with the long-term goal of equipping banking regulation with the most advanced model validation tools, in the interest of all the players of the financial system, institutions and regulators alike.

Best regards,

Carlo Acerbi,

Genève, 26/06/2023

LARIX risk consulting
Ch. De Tattes-Fontaines 43, Vandœuvres, 1253 Genève - CH Switzerland
carlo.acerbi@larixriskconsulting.com

References

Acerbi, C., & Szekely, B. (2017). General Properties of Backtestable Statistics. *SSRN preprint*.

Acerbi, C., & Szekely, B. (2019). The minimally biased backtest for ES. *RISK Magazine*.

Acerbi, C., & Szekely, B. (2023). Backtestability and the ridge backtest. *to appear on Frontiers of Mathematical Finance*, preprint available upon request.

Gneiting, T. (2011). Making and Evaluating Point Forecasts. *Journal of the American Statistical Association*, Volume 106, - Issue 494, pp. 746-762.